CONTENTS
2011 Spring Technical Forum Proceedings

What's On TV? Techniques and Technologies for Content Discovery

Moderator: Kevin Leddy, Time Warner Cable

Malnati, Rob - Motorola
Capitalizing on the Multi-Screen Opportunity….......1

Papish, Michael; Powers, Adam - Rovi Corporation
A Method for Implementing Unique Identifiers in the Entertainment Supply Chain…………......................7

Sandoval, Frank - CableLabs
Social Networking Applications for Cable…...........19

Schapira, Agustin - Comcast
Infrastructure and Tools to Support Secure, Scalable, and Highly Available APIs………………………...27

Tranter, Steve - NDS
Bringing a Centralized Service Delivery Platform for Advanced Navigation and Search to Companion Devices and the Existing STB……………………..39

Space Craft: Transforming the Headend with CMAP

Moderator: Tony Werner, Comcast Cable Communications

Chapman, John; Chopra, Rakesh; Montini, Laurent - Cisco
The DOCSIS Timing Protocol (DTP) - Generating Precision Timing Services from a DOCSIS System..50

Finkelstein, Jeff - Cox Communications
Salinger, Jorge - Comcast
DOCSIS to CMAP Evolution……………................86

Ulm, John - Motorola Mobility
New Converged Access Architectures for Cable Services…………………………………..………...94

Villarruel, Fernando; Hirt, Fred; Wu, Kuang-Yi - Cisco
Pluggable Optics for HFC……………................114

White, Greg - CableLabs
Woundy, Richard; Lee, Yiu; Williams, Carl – Comcast
Rignt-Sizing Cable Modem Buffers For Maximum Application Performance………………………...125

HFC-ing the Light: Advancements in Next-Gen Network Architecture

Moderator: Daniel Howard, SCTE

Emmendorfer, Michael - ARRIS
Next Generation - Cable Access Network An Examination of the Drivers, Network Options, and Migration Strategies for the All-IP Next Generation – Cable Access Network..139

Howald, Robert; Miguelez, Phil - Motorola
Digital Due Diligence for the Upstream Toolbox..197

Howald, Robert; Miguelez, Phil - Motorola
Upstream 3.0: Cable's Response to Web 2.0.........210

Salinger, Jorge - Comcast
Next-Generation Access Architecture Options for Additional Upstream and Downstream Transmission on HFC...237

Sniezko, Oleh - Aurora Networks
RFoG: Overcoming the Forward and Reverse Capacity Constraints…….....................................252

Packet Change: IP Video and Cable's Architectural Evolution

Moderator: Ken Wright, ARRIS

Ansley, Carol; Brooks, Jeff; Lotts, Ed; Shupe, Scott - ARRIS Group
Mapping a Cost Efficient Transition to Converged Services: Business Modeling Tools and Outcomes to Guide a Path to IP Video......................................267

Brooks, Jeff; Ansley, Carol; Cloonan, Tom; Kraiman, Stephen; Torbet, Dan - ARRIS
Architectural Alternatives for Cable IP Video…277

Cannella, Gene; Ogozaly, R. Wayne - Cisco Systems
Next Generation Video Infrastructure: Media Data Center Architecture……………………..…...…...292

Cloonan, Tom; Allen, Jim; Ansley, Carol; Arnold, Rick; Brooks, Jeff; Bugajski, Mark; Cheevers, Charles; Cotter, Tony; Howe, Jeff; Hanks, Bill; Torbet, Dan; Uecker, Matt; Wheelock, Ian - ARRIS
Architectural Approaches to Help Circumvent the "Simulcast Roadblock" of IP Video Deployments..304

Francisco, Mark - Comcast
Considerations When Delivering Cable TV to IP Connected Consumer Electronics………………..326

Business Plan: Architectural Considerations for Commercial Services Delivery

Moderator: Craig Cuttner, Home Box Office, Inc.
Hart, Gaylord - Infinera
Evolving Optical Transport Networks to 100G Lambdas and Beyond……………...…………......334

Miguelez, Phil - Motorola Mobility
Optical Transmitter Technology for Next Generation Access Networks………..342

Park, Dave - BelAir Networks
Architecting Small Cell Licensed and Unlicensed Networks……………...357

Wade, Kevin - Cyan
Carrier Ethernet SLAs: Technology Advancements to Improve Operational Efficiency for Cable Providers……………………..………………......363

Hey You, Get Onto My Cloud: Implications of New Delivery Approaches

Moderator: Dan Pike, GCI Cable

Broberg, Robert - Cisco Systems
Agapi, Andrei; Kielmann, Thilo - Vrije Universiteit
Birman, Ken; VanRenesse, Robert - Cornell
Comer, Douglas - Purdue
Cotton, Chase - University of Delaware
Lehr, Bill - Computer Science and Artificial Intelligence Lab
Clouds, Cable and Connectivity: Future Internets and Router Requirements......................................368

Denis, Xavier; Duzett, Robert - ARRIS
Cooperative Load Sharing Among Loosely Coupled or Independent Video Servers…………….............381

Hasek, Charles - Time Warner Cable
Krishnan, Santosh - Verivue
Approaches to Integrating CDN Technologies into Classic Cable VOD Platforms…………….............398

Holden, Dan - Comcast Innovation Labs
Syed, Yasser - Comcast
Video Ingest for the Cloud: Common Video Subscription Support for Advancement of Multi-Video Service Architectures..................................408

Ramaley, Alan; Rossi, Nick - thePlatform
Evaluating Best-of-Class Web Service APIs for Today's Multi-Platform Video Management Solutions…………………………….................415

ABCs of QoE: Managing Quality of Experience in a Changing Video Environment

Moderator: John Chapman, Cisco Systems

Adams, Michael - Ericsson
Will HTTP Adaptive Streaming become the Dominant Mode of Video Delivery in Cable Networks?.........430

Appelquist, Göran; Potter, Duncan - Edgeware AB
Adaptive Streaming and Converged Management Strategy In Multiscreen Video Service Implementation…………………………………...440

Bennington, Jeremy - Cheetah Technologies
QOE Issues Relevant to Video Streaming in Cable Networks……………...448

Liu, Xiaomei - Cisco system
Adapting Adaptive Streaming to Cable Access…..454

Tooley, Matthew - Sandvine
Adoption Trends of Over-the-Top Video from a Network Perspective………….................................466

CAPITALIZING ON THE MULTI-SCREEN OPPORTUNITY

Rob Malnati
Motorola

SERVING THE ANYWHERE CUSTOMER

The Internet has demonstrated its power to transform industries. In the telecom industry, Cable companies used VoIP to capture significant market share from service providers, and Internet and value-added service providers like Netflix have now moved onto the turf of paid TV providers. As broadband services have become widespread, these over-the-top (OTT) IP-based video service providers are attracting advertising dollars and harnessing imagery and personalization to make video consumption a compelling experience over the PC and directly on the television—giving consumers alternatives to paid TV services.

These providers are able to use the interactivity and viewer profiling abilities of IP networks without having to incur the heavy investments into the network. Pay TV service providers are naturally concerned that OTT providers will be the ones capturing the value of video over IP, while they are left holding the bill for their IPTV networks.

The good news is, however, incumbent providers have the assets, support organizations and the marketing power to stave off the competition and do more than just one-up the offerings of OTT players. They have the power to become the entertainment provider of choice, offering a video browsing experience with subscriber-personalized bundle marketing that spans across a multitude of mediums—the television, Internet and mobile device.

THE CORD CUTTING THREAT LOOMS

It's no secret that the popularity of Internet video is exploding and attracting advertising dollars. Non-traditional players such as Netflix and Hulu along with connected devices like Wi-Fi-enabled Blurays have become compelling providers of video content because of the personalized nature of their offerings and the discounted pricing.

There is much debate about whether consumers are cutting the pay TV cord and opting for OTT plays at this point. Pay TV subscribership is down, but some analysts attribute the loss to a poor economy. Others, such as SNL Kagan, conclude that OTT substitution is impacting subscribership of paid TV services.

Kagan said the cable, DBS and telco video offerings, the U.S. multichannel segment, decreased by 119,000 subscribers in the third quarter 2010. This is compared with a gain of 346,000 subscribers reported in the same quarter in 2009.[i]

—It is becoming increasingly difficult to dismiss the impact of over-the-top substitution on video subscriber performance, particularly after seeing declines during the period of the year that tends to produce the largest subscriber gains due to seasonal shifts back to television viewing and subscription packages," noted SNL Kagan analyst Ian Olgeirson.

Credit Suisse recently downgraded its outlook for a number of media companies, primarily because it sees pay-TV share eroding as younger consumers turn to OTT options such as Netflix. The firm found that 37 percent of Netflix users between the ages of 25 and 34 use Netflix streaming services rather than pay TV. And another 30 percent of subscribers between 18 and 24 have also cut the cord.

Meanwhile, Netflix reported that its third-quarter income, revenue and subscriptions all increased and announced that its streaming video business is the growth driver for the company.[ii]

The company reported that some 66 percent of all subscribers watched at least 15 minutes of streaming video from the company in the quarter, up from 41 percent a year ago, and 61 percent for the second quarter of 2010. Netflix predicted that its subscribers in the fourth quarter will watch more content streamed from the company than delivered on DVD.

A recent survey from Strategy Analytics concludes that 13 percent of current U.S. pay TV subscribers say they are —somewhat” or —very” likely to cancel their current subscription in the next year and not sign up with another provider.[iii]

—While it may represent only a relatively small percentage today, we anticipate the number of cord cutters to increase going forward,” said Ben Piper, director in the Strategy Analytics Digital Consumer Practice.

Smartphones and now wirelessly connected tablets such as the Apple iPad now enable consumers to watch video anywhere. The emergence of larger screen tablets may spur the TV everywhere phenomenon to move even faster.

According to survey from The Diffusion Group, iPad owners are more likely to actively consider cutting back on pay TV services as the iPad now incorporates content from a host of providers such as Netflix, Hulu and iTunes.[iv] According to the survey, 33.9 percent of iPad users are to different degrees likely to cancel their pay TV service in the next six months. Nearly 13 percent of iPad owners are —highly” likely to drop pay TV services, TDG said.

"Despite the fact that cord-cutting remains more widely discussed than carried out, forward-looking research continues to accumulate in support of the hypothesis that specific groups of consumers are quickly warming to the idea," noted Michael Greeson, TDG founding partner and director of research.

The bottom line is: There is a war for the video consumer’s time. Watching OTT video and streaming over the PC, mobile device and TV has become a larger part of people’s lives, and as the trend continues, the value of pay TV erodes. As such, pay TV providers must use their existing assets to their advantage to come out on top in the new world of TV 2.0.

TV ANYWHERE: PAY TV’S ANSWER TO THE OTT THREAT

Pay TV providers are by no means standing still as alternative video services come to market. Comcast and Time Warner in 2009 launched TV Everywhere to increase the number of TV shows available to watch online for free for those who already subscribe to their regular TV services. Verizon FiOS introduced FlexView as the core offering of FiOS TV, which provides Premium VOD across Smartphones, Tablets , PCs and the TV. The concept is open and non-exclusive, meaning cable, satellite and telco video distributors can enter into agreements with other programmers.

The purpose of TV Everywhere is to address the changing viewer habits of consumers while maintaining profitability for program producers and the networks that carry the programming. As such, TV Everywhere has been a focus of the pay TV service provider community throughout 2010 as a way to stave off competition from OTT video plays.

During the company's third-quarter 2010 earnings call, Time Warner CEO Jeff Bewkes said TV Everywhere has been adopted faster than expected, with a host of new programmers signing on to the initiative.[v] AT&T recently announced that its U-verse customers would have unlimited access to more than 1,200 hours of premium content from HBO and Cinemax through the programmers' online portals at anytime.[vi]

Comcast recently re-launched its TV Everywhere initiative called Xfinity Online TV after a previous lukewarm reception because of the lack of content and problematic authentication. Comcast is now promising a deeper breadth of content.[vii]

MULTI-SCREEN VIDEO CONSUMPTION DEMAND IS ALIVE

(General comment – is there also a way we could add stats about tablets as well, given this is a high-growth and sizeable segment?)

According to Nielsen Company's Three Screen Report for Q1 2010, the amount of video Americans consume continues to rise—and consumers are adding new screens and applications to accommodate their video consumption desires.[viii]

High-speed broadband services have paved the way for better user experiences, while nearly a quarter of households have smartphones, which enables consumers to ―palce shift" their video consumption. According to the report, mobile video consumption grew by 51.2 percent year-over-year as did time-shifted TV (18.1 percent), PC (17.3 percent) and TV itself (1.3 percent).

Coverage of the 2010 FIFA World Cup offers some valuable insights into the demand for multi-screen services via a single brand. ESPN Research+Analytics recently revealed that fans spent 4.9 billion gross minutes with ESPN.com and ESPN Mobile properties. Translating that figure into the platform-agnostic metric of average audience, the company found that 110,000 individuals used ESPN digital media to consume World Cup content in the average minute—a figure that was greater than the audience for 23 cable networks during the 31 days of the event.[ix]

ESPN learned that multiple screens served a complementary purpose in boosting audience share. Multi-platform users accounted for 26 percent of all users but they consumed about 47 percent of the content on the average day. Moreover, multi-screen users also spent two hours and 24 minutes of this time watching TV, significantly more than the average TV-only consumer.[x]

Perhaps most importantly, ESPN said its cross-platform offering brought new exposure to its brand, while users were engaged with advertisers more often because of their use of multiple screens.

SERVICE DELIVERY REQUIREMENTS OF THE MULTI-SCREEN OFFERING

Indeed, the era of multi-screen TV delivery has arrived, but current offerings remain fledgling services. In this new era, pay TV providers are challenged to offer an extensive collection of video on demand (VOD) catalogues in a cost effective manner while creating a streamlined way for their customers to discover and consume that content.

Ultimately, a successful multi-screen video offering will require:

- One common infrastructure that allows pay TV providers to deliver thousands of video episodes over a variety of mediums and operating platforms.
- A single catalog of content delivered over a TV set, PC and mobile device with the same look and feel.
- A customizable service delivery platform that generates recommendations based on available information (we can just end at ―personalized recommendations" – because in reality, the driver is not just ―avail info about subs prefs")
- The ability to expand the catalog through the inclusion of federated content from third parties, who also fulfill the delivery of this content.
- Flexible billing options complete with a rating engine that enables pay TV providers to offer bundled and promotional pricing.

THE CASE FOR A COMMON INFRASTRUCTURE

Perhaps the largest challenges pay TV providers face in implementing a multi-screen strategy center on how to create simple and compelling services. Consumers are overwhelmed by the wide variety of offerings available through various providers and easily put off if they must constantly log into various devices or maintain different customer profiles. Most consumers would rather buy their services from a handful of trusted providers who understand their specific needs and interests and deliver these services in a customer-friendly manner. Service providers, therefore, have a key asset in their long-standing consumer relationships.

The traditional model, in which services were delivered to individual devices in a silo fashion, has evolved. With convergence, blended services can be delivered across multiple devices and networks, thus maximizing ubiquity, customer ease-of-use and pay TV revenue opportunities. Thus, service providers must establish themselves across all devices as the preferred landing page for targeted services using a flexible content Management System (CMS). In addition to providing scalable and easy to use Administrator capabilities, the CMS needs to offer subscribers the ability to quickly search, browse, preview and purchase relevant content on any device for consumption on the same or another device.

Enabling the content experience across multiple delivery platforms in an integrated manner will be critical. Pay TV providers must use tools that simplify the delivery of interactive video services across any platform – including IPTV, PC and mobile. Consolidating operations into a single CMS also allows for more cost-effective integration to exciting third party services, which can then be available across all screens, allowing for faster time to market with new customer experiences.

PERSONALIZED CONTENT WILL RULE

To secure customer loyalty, as many nimble Internet/web competitors have done, pay TV providers must build upon their core network assets by continually refreshing their services and delivering them in a personalized way across one or more storefronts, along with targeted advertisements, promotions and personalized/relevant incentives.

The burden cannot be on the subscriber to search for the content they want. Pay TV providers must bring a higher level of interactivity and rich presentation that consumers already see on the Internet. The key is to pro-actively generate recommendations based on available information about the subscriber's devices, subscriptions, preferences and location. Providers can pre-search all available services and personalize the presentation of available material in the electronic program guide and video on demand storefront to present only those services and advertisements that are relevant to the customer's criteria.

For instance, a customer buys a significant number of action movies. The interfaces paid TV providers employ should understand this and make recommendations, and position content of interest in the high level folders/navigation on the TV screen in a compelling way to make movie and merchandize recommendations. Likewise, if a customer buys a particular movie, merchandizing offerings should come attached. A movie-related ringtone or game could be sent to customers' mobile devices or customers could have the opportunity to buy early release movie trailers.

An important piece of personalized content is a federated profile management capability. A pay TV provider must use available information about the subscriber's devices, subscriptions, preferences and location to pre-search all available services and present only those services that are relevant to the customer. Doing so provides a more compelling experience for both the subscriber and the advertiser.

MULTIPLE REVENUE MODELS ENSURE VIDEO CONSUMPTION

Pay TV providers all desire to evolve their revenue models as they expand their portfolio of services. One price does not fit all! As subscribers shift around their video consumption, flexible pricing and discounting is required to maximize subscriber take rates. Subscriptions will require support for daily, weekly, monthly, annual and metered usage, while transaction rating is require to support a wide range of per-use rating and volume discounting. Promotional price rating ensures that pay TV providers can manage a subscriber's lifecycle through initial and ongoing target marketing as well as the life cycle extension through discounting and promotion such as target market discounts, volume discounts and buy-one-get-one free pricing. And finally, a flexible revenue model allows pay TV providers to align service subscribers with ads that are relevant.

THE MOTOROLA ADVANTAGE

Motorola's Multi-Screen Service Management Software Suite Merchandiser, known as the Medios Merchandiser, provides the flexibility needed to increase pay TV service provider average revenue per user (ARPU) by differentiating On Demand video-based offerings. Through advanced metadata management tools and automation, the Medios Merchandiser offers a highly scalable multimedia content marketing system built to effectively market the ever-expanding on-demand catalog, including video, music, games, and applications. The Medios Merchandiser solution enables providers to successfully provide consumers with a media-rich bundling experience enhanced with flexible pricing and discounting, as well as personalized recommendations that drive impulse purchases and your return on investment (ROI).

Medios Merchandiser providers pay TV providers with the following business differentiating functions:

- Bundle Definition — Discover or ingest asset metadata, even from federated partner/third party catalogs, to populate the three-screen multimedia on demand catalog. Medios Merchandiser provides GUIs to allow pay TV providers to build bundles of video, music, games, and physical goods. It also enables them to combine TV, web and mobile in the same bundle.
- Multi- Screen Marketing — Medios Merchandiser's APIs are exposed to existing TV, web, or mobile portals, preserving these investments, while enabling cross-domain marketing and fulfillment.
- Price and Discount — Medios Merchandiser's rating engine enables flexible pricing and discounting to maximize subscribers take rates. Its GUI- based operations make it simple to build and associate pricing and discount plans to individual assets, bundles and categories of assets.
- Recommend and Target — Using profiles and inputs, such as subscriber preferences and social network inputs, Medios Merchandiser positions the assets, offers, and pricing most likely to result in a purchase one-to-two clicks from the subscriber's home portal screens.
- Fulfill and Settle — Once an order is placed Medios Merchandiser generates the grants, DRM license terms, billing records, payment gateway events, and settlement records to fulfill the transaction. It also handles reversing all this in the event of a customer reversal.

Motorola's Medios Merchandiser is a game-changing platform for the pay TV provider's three-screen personalization and multimedia bundling needs. It brings relevant, targeted content to your consumers with custom bundling capabilities across TV, web, and mobile —- increasing your revenue opportunities and ROI and lowering your operational costs.

---

[i] —SNL Kagan Analysis Shows U.S. Multichannel Video Subscribers Drop for Second Straight Quarter," SNL Kagan press release, Nov. 17, 2010.

[ii] —Netflix CEO: We are now primarily a streaming company; Q3 earnings, revenue, subscribers increase," FierceOnlineVideo, Oct. 20, 2010

[iii] –Strategy Analytics: 13% of Americans Likely to 'Cord Cut' from Pay Television in the Next Year," press release, Sept. 13, 2010

[iv] iPad Users and Intenders More Likely to —Cut the Cord," press release, The Diffusion Group, Nov. 11, 2010

[v] —Bewkes: 'TV Everywhere' Making Progress," Multichannel News, Nov. 3, 2010

[vi] —AT&T ups ante for TV Everywhere play, gives U-verse customers HBO, Cinemax access online," FierceIPTV, Nov. 29, 2010

[vii] —Comcast dismisses cord-cutting threat; says it's relaunching Xfinity... again," FierceIPTV, Sept., 23, 2010

[viii] Nielson Three Screen Report, Volume 8, US Q1 2010

[ix] —ESPN Presents Results of World Cup Cross-Platform Research Project – ESPN XP," ESPN Media Zone, Sept. 27, 2010

[x] —ESPN XP World Cup Dispatch #5—Ten Things ESPN Learned During the World Cup." ESPN MediaZone, July 16, 2010

A METHOD FOR IMPLEMENTING UNIQUE IDENTIFIERS IN THE ENTERTAINMENT SUPPLY CHAIN

Michael Papish, Solutions Architecture Director, Rovi Corporation
Adam Powers, VP, Technology Strategy, Rovi Corporation

Abstract:

The emergence of digital technologies has transformed every aspect of the professional audiovisual supply chain from content creation and post-production to distribution and consumption and created new opportunities for all stakeholders. With these opportunities also come challenges, such as more complex value chain interactions and an explosion in the number of assets relevant to commerce.

In an effort to make the most of the new opportunities and address the challenges, major content producers have redesigned workflows to make them content-centric, putting in place digital media management infrastructure to help streamline operations. While these efforts help reduce inefficiencies, the role of unique identifiers in the movie and television supply chain can also facilitate and improve the efficiency of automated workflows.

In late 2010, the Entertainment Identifier Registry (EIDR) was launched by a group of companies — including Comcast, Disney, Warner Bros, Rovi, MovieLabs, and CableLabs — as a non-profit trade organization acting as a centralized registry for uniquely identifying video works. EIDR is built on the flexible data model established by the Digital Object Identifier standard.

However, with digitization has also come a fragmentation of the entertainment supply chain. As the industry adopts cloud-based systems, a web services model built on open APIs will soon follow. Small app-developers will vie with large hardware manufacturers to deliver the most engaging consumer entertainment experience using the same platforms and content items. We need to translate our centralized ID space into a technology service layer which can be accessed by a distributed ecosystem.

This paper provides an overview of the implementation of and interfaces to EIDR, including registering and looking up records, metadata schema, and deduplication process. We then extend this registry into the distributed ecosystem by describing technical methods developers can use to keep in sync with the ID space as it grows and morphs. Web service-based mechanisms for matching and translating content IDs across catalogs are described. We introduce the concept of ID stability and a time to live (TTL) model that uses readily understood concepts of internet architecture to implement EIDR and EIDR-like ID spaces into common software frameworks. We conclude with an investigation of potential implications and implementations for content management, advertising, interactive applications, and EPG metadata.

SECTION 1 – WHY UNIQUE IDS?

The past decade has seen the beginning of a dramatic shift in how content is created, distributed and consumed. Theatrical production costs have dropped while production workflows have become increasingly more complex and fragmented; broadcast workflows are making the switch to be file based; MSOs are reaching out to every screen with broadcast, video on demand, and new forms of engaging consumers; and advertisers are under pressure to deliver lower CPMs, higher targeting and more engagement.

This total upheaval of the entire supply chain has created many challenges, but also presents many opportunities as well. Whether a company's dream is to provide the myriad of new features and functionalities that are creating buzz on the market, or even achieving the long desired dream of total workflow automation, the devil is frequently in the details when reaching for those dreams. One of those devils is reaching a scale that allows those dreams to be economically viable, and the human factor is usually the largest hindrance. Even after several decades of personal computers and more than a decade of widespread Internet usage, humans are all too often involved in every step of processing and distributing content – manually creating avail notices, manually ingesting content, manually associating interactive assets with linear content.

In order to transition away from the manual infrastructure that is in place today to the automated infrastructure that enables tomorrow's applications and scale, new processes and techniques will need to be created that allow assets to be understood and processed by computers. Perhaps the largest hindrance is enabling computers to understand the assets with which they are dealing. In the manual world, humans have a great deal of context and outside information when working with assets. ―Which version of Robin Hood is this?," can often be answered with some understanding of whether a human is working with new releases, animations, or back catalog – although mistakes still happen. Computers, lacking context and nuanced decision making capabilities, must rely on unique identifiers for assets.

But all unique identifier systems are not created the same, nor can they all be applied the same. This paper attempts to lay out a framework for both designing and analyzing unique identifier systems. It then applies that analysis to two different systems: the unique identifier systems that have broad adoption in the market today, and the Entertainment Identifier Registry (EIDR) that was launched late last year.

SECTION 2 – THE DIFFICULTIES IN FORMING IDS

Whereas previously we could aspire to create a single, perfectly structured and completely clean database of entertainment content, we must now admit that this task is Sisyphean. Not only has the pace of content creation increased but the rise of semi-professional forms of audiovisual content such as YouTube complicates the matter. In addition, we now have multiple, always-on streams of information *about* media to parse such as Twitter, Flickr, blogs, check-ins, reviews, etc. We must treat our target as always moving and adopt technical methods which give us maximum flexibility in describing the content space.

The Four C's of Catalogs

Our goal is to identify each entertainment item (e.g. movie, TV episode, actor, etc.) with a numerical ID which allows each participant in the ecosystem to uniquely refer to the item throughout the distribution chain. The space of IDs is known as a *catalog*.

There are four attributes of a catalog ID space which we want to optimize:

- Coverage – How much content does the ID space address? The larger the

content space, the more useful the catalog.
- Cleanliness – How close does our ID space provide a 1-to-1 mapping from ID to real-world entity (i.e. no duplicates, no ambiguities). The cleaner the ID space, the better utility it provides the ecosystem.
- Churn – How quickly are existing IDs changing meaning[i] or becoming deprecated? We want to minimize churn in the catalog.
- Convenience – How easy is it for us to add new content to the catalog? What about new ―types‖ of metadata which we want to layer on top of the IDs?

Creating a catalog structure which optimizes these four tenets is difficult. The two obvious catalog construction techniques are: complete editorial control or fully automated construction. In an editorially controlled catalog, all new items are verified by a human gatekeeper before they are added to the catalog. In a fully automated system, a matching algorithm without human supervision is used to recognize uniquely novel entertainment works, and automatically creates a new ID and entry from them in the catalog.

These are how the two methods compare:

	Editorially Curated Catalog	Fully Automated Catalog
Coverage	Poor	Excellent
Cleanliness	Excellent	Poor
Churn	Excellent	Poor
Convenience	Poor	Excellent

In Section 4, we describe a hybrid method of catalog construction which helps overcome this contradiction.

Who Needs to use ID Catalogs?

From the perspective of a catalog maintainer, there are two groups of stakeholders in the ecosystem.
- Upstream Users. This group consists of content creators and owners who want to associate unique IDs with each of their entertainment works. In Section 5 we describe EIDR, a method to allow multiple upstream users to interact with a well-constructed catalog.
- Downstream Users. This group consists of content users (distributors, application developers, viewers, etc.) who want to access and describe entertainment content. These users may interact with multiple parties in the ecosystem and need to be assured that a common ID can be used cross-system and cross-device. In addition, downstream users want to be able to persist additional data (e.g. customer ratings, reviews, personal cloud

lockers, etc.) on top of the ID space and therefore need a stable catalog. In Section 4, we describe a catalog which has these properties and the methods used to keep a downstream ecosystem in sync as the catalog evolves over time.

Key for both groups of users is dealing with the constant cleansing and curating work a catalog maintainer such as Rovi conducts on the ID space. Two common issues which can cause problems are ID merging and splitting. As we describe above, the pace and variety of content creation and the need for a catalog with high coverage levels ensures that there will always be ID mistakes introduced into the catalog which are fixed post-entry. We define these operations as:

- Merges – When a single real-world entity is represented by more than one ID, we perform a merge to compress 2 through n IDs to a single ID.
- Split – When a single ID mistakenly represents more than one unique, real-world entity, we perform a split to turn a single ID into 2 through n separate IDs.

Since downstream consumers have a need to persist data on the IDs, we must invent a method to communicate merge and split operations throughout a distributed ecosystem. Our goal is to construct a methodology for the ID space which harnesses the contributions of upstream and downstream consumers but doesn't create chaos in the catalog. We describe this method in Section 3.

Multiple Stakeholders and Conflicting Design Goals

When trying to create a global unique identifier system that can be adopted on a global scale, one of the largest challenges in designing a unique identifier system is having a broad enough view of the entertainment industry to understand the needs of stakeholders at every step during an asset's lifecycle. Design pressures come from multiple, and potentially conflicting, business models, as well as individual stakeholder concerns around privacy and control. These issues are explored as part of Section 5.

SECTION 3 – TECHNICAL METHODS TO KEEP AN ECOSYSTEM SYNCHRONIZED

In a catalog ID space similar to that described in Section 2, we have upstream providers who want to register and contribute metadata to the catalog and downstream consumers who want to access the catalog and IDs to power applications. While we attempt to maintain a stable ID space, the dynamics of today's entertainment ecosystem ensure that cleaning, updating and curating are ongoing activities.

It is important to note that these activities occur simultaneously on both the IDs themselves and the underlying attributes of metadata associated with these IDs. For purposes of this paper, we will concern ourselves with the task of synchronizing the IDs themselves.[ii]

For downstream consumers, stable and synchronized IDs are required due to the need to persist data on top of the IDs. For example, assume you are an MSO providing a Video On Demand (VOD) service to your viewers. As your viewers watch VOD content, you want to be able to recommend them new movies and shows they may also like based on their previous viewing history (e.g. “You really liked *Meet the Fockers* last year and now *Little Fockers* is available on demand in HD”). In order to satisfy this use case, the MSO needs to store the household’s viewing history in a profile. The ideal method for accomplishing this is to keep a list of the IDs of video items previously purchased.

In an unstable ID space, it is possible that the ID for *Meet the Fockers* may “mean” something different now than it did when the MSO originally stored the data. This is a problem.
As described in Section 2, there are two situations where a change may occur to an ID post-creation: merges and splits. Ideally, a message could be broadcast to all downstream consumers whenever one of these actions occurs and the consumer could update their records. However, this approach is technically infeasible for several reasons:

- The ecosystem is fragmented and open. Previously, we could operate under the assumption that all downstream users were known subscribers by the catalog maintainer. However, in a world of interoperable APIs and dynamic data applications, we want to allow a large and distributed ecosystem to utilize the ID catalog without necessarily requiring centralized authorization.
- It is onerous to require downstream consumers to update records. If the MSO has stored IDs throughout their data systems, it is onerous to require them to propagate ID changes every time the catalog updates.
- It is brittle. If downstream consumers miss an update message, they will not be able to reconstitute their records and will have a corrupted ID space. Given the distributed nature of the ecosystem, we cannot rely on a synchronization protocol which is brittle.

An elegant solution which allows the resolution of splits and merges but avoids the problems described above is as follows: The catalog maintainer continuously monitors the ID space looking for errors. When the maintainer identifies two IDs referring to an identical entity which need to be merged, it:

1. Identifies the “dominant” ID. This is either the ID on which the most amount of activity has taken place or was the first to be created.
2. Creates a link from the subordinate ID to the dominant ID.
3. Stores the subordinate ID as a “deprecated” ID for the dominant ID.

Similarly, when the maintainer finds a single ID which erroneously refers to more than one real-world entity, it invokes a split mechanism which:

1. Identifies the “dominant” real-world entity, if possible. If not, selects the dominant as the first entity to be linked with this ID.
2. Creates new IDs for each entity.
3. Stores a link from the original ID to the two new IDs.

Now, how do we communicate these changes to the downstream consumers? First, we make the following assumptions:

- Downstream consumers have an incentive to know about ID updates.
- However, downstream consumers are not required to know about updates in order to operate. If they are unaware of an update and refuse to edit their stored data, the ID space should provide ID accuracy no worse than as if the update never occurred.

Given these assumptions, we adopt a system of implicit notification. Every time a consumer submits a request for metadata about a given ID, we include in the response the most up-to-date ID for this item. If no cleaning has occurred, this ID will be identical to the ID submitted. However, if there has been a change, the consumer will receive the new ID link. The consumer can choose whether to update their records and silently discard this new information. Since the catalog maintainer always knows both the previous meaning of each ID and the new meaning, the consumer will never run into ―dead IDs" which return no associated data.

An optional implementation includes an on-demand, ID lookup web service. In this implementation, the consumers can submit an ID and the maintainer will return its history and current state. This service is useful if the downstream consumer wants to check its records and perform larger-scale updates on its ID space data.

SECTION 4 – ROVI ID SPACES: A LOOK AT MULTIPLE ID CATALOGS

Now that we have a method to keep an ID space in sync throughout the ecosystem, how do we construct a catalog which maximizes coverage, cleanliness and convenience while minimizing churn?

Historically, there are two methods which could be used to maintain a metadata ID catalog:

1. Editorially controlled catalog. In this construction method, each addition of an ID to the catalog is controlled by a human editor. The human editor ensures that the ID does indeed represent a new, unique real-world entity. While this process maximizes catalog cleanliness, it is extremely resource intensive and slow to add new content (e.g. difficult to quickly grow coverage).
2. Fully automated catalog. In this construction method, a computer algorithm evaluates each new entity for uniqueness and then either matches it to an existing ID or makes a new entry to the catalog. This method makes it quick to ingest new content into the catalog. However, since it can be difficult to automate the matching of different sources of metadata, cleanliness will suffer with many duplicate ID entries.

Since neither catalog construction method is ideal for the purposes of creating a single entertainment catalog, we have developed a new, hybrid catalog construction model which allows for the dynamic aggregation of multiple metadata sources into a single ID space. We call this construction a Dynamic Aggregate Catalog (DAC). In addition, our method allows for both human and algorithmic cleaning methods and assumes these curation activities will be ongoing as opposed to only occurring on ingestion.

First, we describe our motivations for creating DAC. Then detail the methods underlying its use.

Over several years, Rovi has acquired multiple metadata databases including TV Guide, All Media Guide (AMG) and Muze. While each database describes overlapping content they do so for different purposes and in differing manners. For example, TV Guide had complied data on movies for display in a listings grid whereas AMG had longer descriptions geared for the retail setting. And, while TV Guide specialized in TV-related data, AMG also had deep information on television series provided they were made available in DVD boxsets for retail.

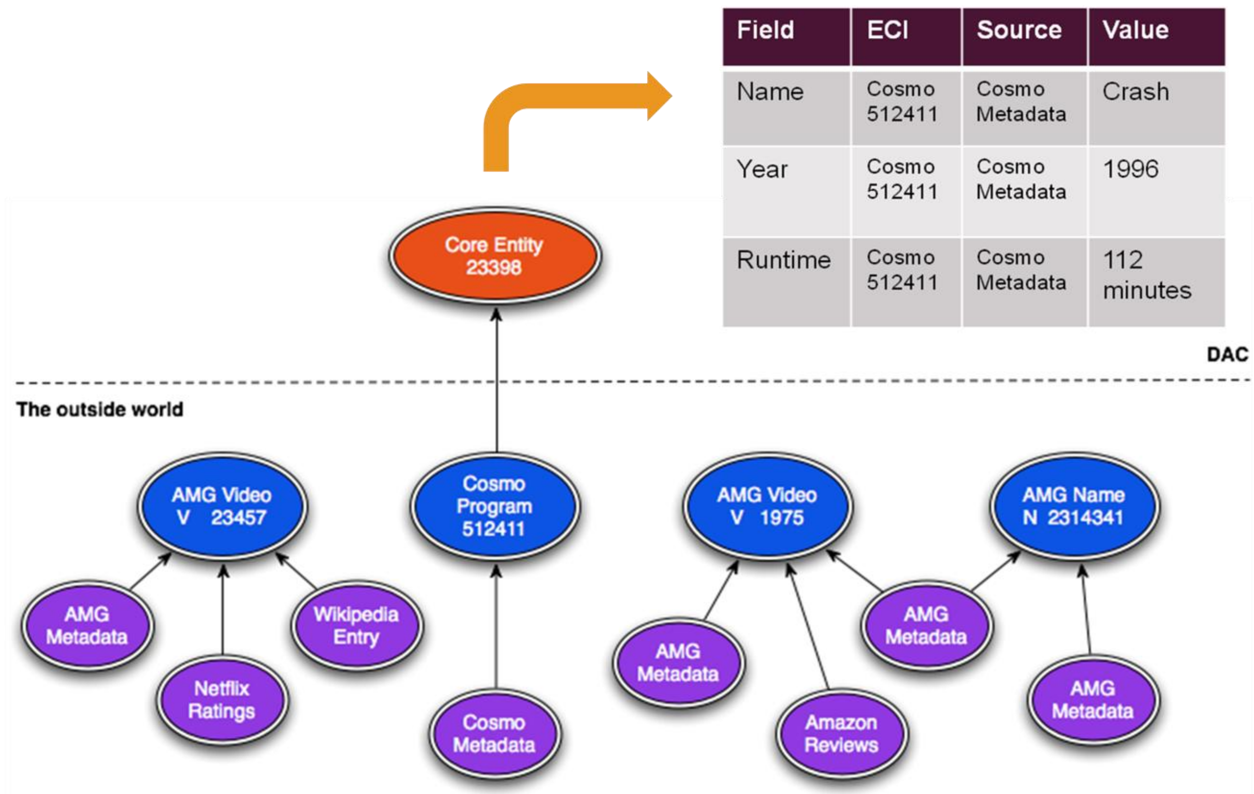
Even though these datasets had different conventions, our goal was to create a single, normalized catalog ID space which would allow us to access all available data. This is important because going forward, different types of applications will want data in different formats and optimized for different use cases. However, we do not want to create individual database silos locking away access to valuable information.

Rovi's motivation is not unique. As the pace and quantity of digital media consumption increases, the ecosystem will have access to multiple sources and catalog of entertainment content. They will need a way to organize and centralize this data around an ID space. Our DAC model can serve as a template for this process.

Inside the DAC

Figure 1 describes the DAC. The building blocks for the catalog are individual external catalogs of metadata such as AMG or Netflix. These catalogs can be growing (e.g. active) or fixed and can be either editorially controlled or fully automated. Each external catalog consists of individual items (represented by blue circles in Figure 1). External catalog items (ECI) have metadata layered on them. This metadata can be from metadata in the catalog itself or from 3rd-party sources persisted on the ID space itself.

Next, we create DAC items (orange) which link to an ECI. The DAC item inherits all source data from the ECI and as the source data grows and improves it is automatically synced with the DAC item. The data itself is stored referencing its original source. The link between a core DAC entity and an ECI has an inherent confidence value associated with it.



Field	ECI	Source	Value
Name	Cosmo 512411	Cosmo Metadata	Crash
Year	Cosmo 512411	Cosmo Metadata	1996
Runtime	Cosmo 512411	Cosmo Metadata	112 minutes

Figure 1: An overview of the DAC

Multiple ECIs may link to the same DAC item. We use software logic to reconcile conflicting values amongst the source data.

The DAC system allows for editorial curation **on top** of the catalog. The links between core entities and ECIs can be created automatically or via editorial means. When a link changes the underlying metadata follows automatically. This property is what renders the DAC a **dynamic** catalog.

SECTION 5 – THE ENTERTAINMENT IDENTIFIER REGISTRY (EIDR)

In section 3, we described the concerns of downstream consumers when interacting with an ID space. Simultaneously, upstream consumers have their own unique requirements for registering and identifying assets in the catalog. In this section we describe the EIDR system which meets these requirements.

In October 2010, a coalition of companies from the professional video ecosystem announced a new organization that would specifically address the creation of an identifier space for video assets. The organization, called the Entertainment Identifier Registry (EIDR), is a non-profit, centralized registry that was founded by companies such as Comcast, Disney, Warner Brothers, CableLabs, MovieLabs, and Rovi, and operates as an open organization for the standardization and adoption of EIDR.

The design of EIDR is based on another international standard and unique identifier system that has been around for over a decade – the Digital Object Identifier (DOI). This system uses a federated approach that links together unique identifier catalogs for widely disparate systems, ranging from identifying academic papers to uses for military applications to identifying video assets for EIDR. The core concept of DOI is to assign permanent unique identifiers that can be resolved using a URL-style scheme.

Figure 2 – Example of an EIDR unique identifier with full DOI notation

The DOI implementation for EIDR was developed by the Corporation for National Research Initiatives (CNRI) based on their open-source Handle System software. The modifications that EIDR made to the Handel System software embody a number of design trade-offs and considerations.

Opaque Identifiers

It should be noted that the identifier in Figure 2 is opaque – there is no information conveyed by the identifier itself about the asset that it is describing. The design tradeoff underlying this decision is one between permanence and human readability. While humans would like to use identifiers that they can read (such as —doi/10.5240/Desperate Housewives/Season 1/Episode 1”), this only leads back to the reference-by-title confusion that led to the creation of identifier systems in the first place. Using an opaque identifier also means that any errors or changes, whether to spelling, hierarchy, or some other aspect of the underlying asset, will not result in changes to the identifier. This is key in enabling the identifier to remain permanent and unchanged throughout its life. The obvious downside is that opaque identifiers cannot replace labeling on physical assets that don’t otherwise have distinguishing marks or labels on them.

Centralized Registry

One of the initial design decisions of EIDR was to create a centralized registry where each work is registered and assigned a unique identifier as part of a centralized repository. This design decision is a tradeoff of key business concerns such as accuracy, control, and privacy, where a centralized registry gains accuracy at the expense of the control and privacy of the organizations that are performing the registrations.

In order to realize the accuracy gains of a centralized registry, EIDR also employs a deduplication process that is a combination of automatic and manual comparison of records to determine the uniqueness of the record. The automatic comparison uses a scoring system to calculate the distance between two records and sets thresholds to determine whether the records should be considered the same or not. That distance is measured by the difference between the elements of the registered asset, which in turn had great influence on the definition of the metadata schema (as described below). There are cases where the distance between two records may be indeterminate as to their uniqueness, resulting in the need for human intervention to make a final determination.

To contrast this centralized model and its accuracy against the alternative, there exist other registries, such as ISRC for music, that use a distributed registry model. In a distributed model, multiple stakeholders receive an allocation of numbers and assign numbers to assets at their discretion. In a strongly distributed registry, the resolution of these numbers is also performed by the organizations that make the assignments; in a loosely distributed registry, the assignments are eventually communicated back to a centralized database without any checking to ensure the uniqueness of the records. In a strongly distributed registry, the organizations have ultimate control around the assignment of records and can determine when they are made public. In a loosely distributed registry, there is an opportunity to improve accuracy and reduce errors of multiple duplicate submissions through deduplication and other mechanisms; however, this comes at a cost of cleanliness from multiple erroneous records being published and having to be corrected after they are made public.

Metadata Schema

Another fundamental design decision for EIDR was the coverage of the registry, which is manifested through the definition of the schema. With the complexities of today's global reach of video assets and increasingly fragmented distribution channels, it was quickly decided that the registry should cover a hierarchy that includes three main groups: abstract, variations and encodings.

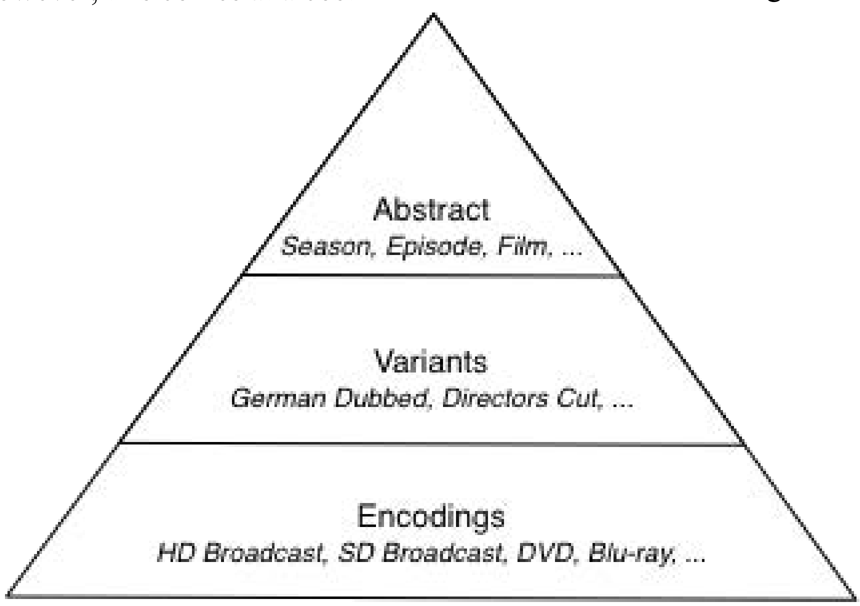


Figure 3 – Hierarchy of assets that are described by the EIDR metadata schema

Abstract records describe a group of assets at the ―title" level. These assets do not actually exist in one viewable form, but exist as multiple variations of a real asset. Title level assets include both series, which are a collection of television episodes, as well as episodes and films, which are typically a collection of variations of the video asset.

The variations are the next group down in the hierarchy and describe the different cuts and edits that happen during the post-production process. Examples include movies that have multiple edits, such as original theatrical release, directors cut, approved for broadcast television cut, approved for UK distribution cut (e.g. – no headbutts), safe for airline cut, the German dubbed version, the

French subtitled version, and so forth. Another example would be a television series, which is
an abstract of multiple related television episodes.

At the lowest level of the hierarchy is encodings. EIDR has the capability to uniquely identify and describe each individual encoding of an asset, based on video codec, audio codec, codec resolutions and bitrates, and other encoding attributes. Examples include both digital media, such as a 1080 MPEG2 Transport Stream with a Dolby AC3 audio codec, and physical media, such as a DVD or Blu-ray.

Each group in the hierarchy has its own metadata schema and nomenclature used to describe that group of assets. While metadata is a vague and broad term, the design criteria for EIDR was to capture the necessary and sufficient fields required to ensure the uniqueness of each record, which ultimately supports the design goal of being able to provide deduplication of records across the entire registry.

A schema for EIDR in XML format can be found through the EIDR website: http://eidr.org.

Identifier Interoperability

A final design criterion came about through the desire for flexibility — in the realization that no single identifier system could have total and absolute coverage of all stakeholder concerns. To that end, EIDR included the ability to cross-reference other unique identifier systems. Each record in EIDR may reference one or more third-party unique identifiers. For the sake of near term design simplicity, these identifiers are treated as opaque objects that have few syntax restrictions and are not subject to the requirements around cleanliness (i.e. – deduplication) that the rest of the registry is subject to. Ultimately this gives EIDR a way of incorporating both existing standards and proprietary systems, as well as being able to expand through the cross-referencing of other unique identifier systems.

SECTION 6 – MSO APPLICATIONS AND ADOPTION

MSO Identifier Applications

Whereas the rest of this paper attempts to describe the formation and application of unique identifiers independent from individual stakeholders in the entertainment industry, this section will focus specifically on how unique identifiers can be applied to solve problems and enable new functionality and automation in MSO systems.

At a high level, the application of identifiers in MSO systems can be broken down into three categories:

1. Enabling the automation and distribution of video asset
2. Use in business systems surrounding the video assets
3. Use in providing features and functionalities that aren't currently widespread

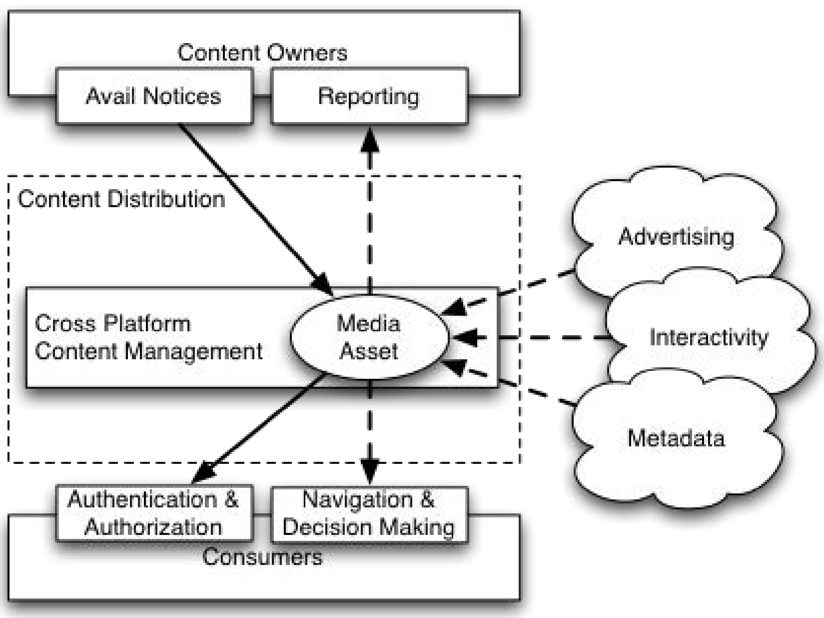


Figure 4 – Relationships between media assets and their uses that immediately benefit from a unique identifier system. Solid lines represent the applications of unique identifiers that directly facilitate the distribution of assets; dotted lines represent the application of unique identifiers in supporting business systems or in augmenting applications.

Asset Distribution

As studios and broadcasters integrate unique identifiers into their systems and move towards standard interfaces for the distribution of assets, unique identifiers will become ubiquitously available starting with the ingestion of the content. Until then, identifiers must be associated after content has been received, either through fingerprinting, watermarking, or metadata matching techniques.

Assuming that content has been received and an identifier has been associated with it, a content management system can then utilize the asset by associating it in any number of ways and especially to distribute it to the end consumer. In distributing it to the end consumer, the content management platform can use the unique identifier to check against distribution rules (such as the rights to distribute to various platforms, or the windows of when content is available in specific distribution channels) or to check a consumer's right to access a specific piece of content as part of a subscription.

Business Systems

Unique identifiers also play a role in the supporting business infrastructure, including finance and auditing systems. For example, when a consumer watches a VOD asset the unique ID of that asset may be transmitted back to the finance reporting system along with other billing information. Finance departments can then use this unique identifier for their own internal auditing, or the number may be shared with other departments that are responsible for sharing consumption information with the content owner.

Another common application of unique identifiers in business systems is for antipiracy. The unique identifier may be embedded into assets as an audio or video watermark, which enables forensic teams to recover information about which asset was pirated (for example, which edit or which encoding).

Augmenting Applications

The final and most interesting use of unique identifiers is their application to emerging technologies that require scale and automation beyond what can be done with a manual labor force.

One example of this is advertising. As the number of video assets and the platforms that they are viewed on both increase, so do the number of advertisements that can be placed. This creates both a larger market for advertising and a problem of cross-platform measurement. In order to solve the challenges around placing ads and aggregating the viewership metrics to be reported back to the ad buyers, a common unique ID system must be established between all stakeholders.

Along the same lines, interactive assets such as EBIF are becoming more common for both augmenting consumer experiences and creating more engaging advertising with new calls to action. While interactive assets are sometimes embedded into video streams by broadcasters, they may also take a peripheral route, requiring that they be associated and reattached with intended video asset(s).

Another class of problems is presented by the metadata associated with video assets. Both schedule and program data are delivered by third-party sources to the MSOs, requiring that they be re-associated with the broadcast programming and VOD assets in order to assist with both content management and ultimately with the navigation and decision making experiences at the hands of the consumer.

Adoption

The ability to adopt a unique identifier varies based on the criteria of each application. While some applications are highly dependent on the ―four C's" described in Section 2, others may have secondary adoption criteria. Those adoption criteria include:

- Number of assets that are currently identified, as related to the total population of assets
- Legal terms of use or intellectual property restrictions around the use and propagation of the identifiers
- Network effects from adoption by other applications, systems, or actors
- Ability to associate identifiers with assets

---

[i] What does an ID ―mean"? For our purposes, an ID should map to a real-world entity. For example, the AMG ID "V50435" refers to the original release of the movie *Top Gun*. As the ID space is updated, we always want this ID to point to the movie *Top Gun* which is something people in the real-world can watch, purchase, and have opinions.

[ii] Ensuing the ecosystem has access to the most accurate and clean metadata is, of course, also an important design goal. However, the technical challenge is mitigated once the ID space issue is removed. Once a downstream consumer has access to a stable ID, they can access on-demand services (e.g. a RESTful metadata service) to obtain the freshest possible metadata or receive regularly scheduled updates in bulk.

SOCIAL NETWORKING APPLICATIONS FOR CABLE

Frank Sandoval

Abstract

To paraphrase Wikipedia"s definition: A social network is a set of individuals tied together through some form of interdependency. By this loose definition, I submit that the set of all cable TV subscribers does in fact form a social network. The connections between subscribers are implicit, as they simply share a connection to a cable service provider. But their interdependence is real in that the audience as a whole directs what content flourishes and what advertising is presented.

The introduction of programmable receivers, through ETV and tru2way®, the use of consumer-owned devices such as iPads to extend aspects of cable service, and the popularity of social networking applications on the Internet combine to suggest that the time is nigh for cable to consider strategies to create valuable applications based on its own implicit social networks.

Social networking apps on cable could help to extend the cable experience to more touch-points is a person"slife, integrating entertainment into communications for instance. A key effect of such services would be to generate data about user activity, and this data in turn dives ever more innovative and valuable services.

There are two aspects to consider: what is the nature and expression of cable social networks, and how might applications based on these networks be developed?

A natural instinct for each operator may be to develop these services individually, but the well known value of the "network effect" argues that these services should include the entire cable audience. These services are more useful to subscribers as more subscribers join in; i.e. a „Top Ten on Cable"is arguably of more utility to viewers than a top ten on a given cable system.

At CableLabs we always look for ways to streamline the adoption of new technologies in cable through development of interoperable interfaces, content formats, and other technical elements. While one theme of this paper is to suggest areas where common definitions might be valuable, the greater goal is to encourage thinking about new ways to enhance the cable experience with social networking applications.

IT'S ALL ABOUT APPLICATIONS

From a user"s perspective social networking is simply a matter of using an application that includes some kind of social networking feature. I won"t attempt to derive some theoretically precise definition of exactly what attributes a social networking application must or must not have. For our purposes, we might consider social networking applications as those that either support direct communications between users, or that incorporate some aspect of dynamic collective user behavior. This eliminates applications that simply reference static user databases, but otherwise leaves the field pretty open.

Observe that social networking applications do not necessarily rely upon explicit user selection of connections. For instance, while many aspects of Facebook, like viewing a profile and posting to someone‟s Wall requires you to be friends with that someone, other functions, such as friend recommendations, are performed without users actively making connections.

Note also that explicit user registration with the application is not always necessary. For instance, users register with Netflix to get videos, but the Netflix content recommendation engine is a social networking app that leverages the viewing history of the entire set of Netflix subscribers.

This scenario is analogous to how we might view cable social networking. People voluntarily sign up for cable service, and along with that may come a whole host of applications that embody social networking features.

Because some popular Internet social media applications publish open web APIs, integration of cable applications with these services is possible. For instance, a widget on a tru2way® set-top box or a cable iPad app could allow a viewer to send pre-formatted Tweets about what they‟re watching or about what content they like, or post similar messages on Facebook.

However, cable could develop its own applications that serve to draw viewers deeper into the video experience. Facebook on its own can‟t truly integrate with the cable experience by presenting content.

Examples of applications that might flourish on cable include:

- Content sharing – push a button to send a TV show link to someone. They in turn can simply push a button to view the content
- Chat – open a chat window and join in conversation in real-time while watching TV
- Content Ratings – push a button to rate what you‟re watching. The ratings could be used by your cable provider, shared with other cable providers for a nationwide cable ratings service, shared with Rotten Tomatoes, sent to Twitter, and more.
- Content recommendations and personalization – navigation systems can learn from the viewing history, preferences, and ratings of your friends and the network at large to explicitly recommend content or configure the presentation of content

It appears there are two broad categories of social networking functions – those that entail direct one-to-one, one-to-many, and many-to-many communications among members, and those that harness the intelligence of the network to provide value added services to members.

UNLOCKING THE INTELLIGENCE OF THE NETWORK

At its most simple, the set of cable subscribers is defined and instantiated within the billing systems of each operator. However, lots of extenuating data are or could be associated with the basic billing records. No generalized model has been developed to describe such information, nor are interoperable interfaces available to potential application developers. We are not suggesting open interfaces, available to any would-be developer (although over time this might be valuable), but rather interfaces that make internal cable operator application development efficient and that allow applications to be portable across systems, a key cost saver.

One means of capturing the value of billing records and other data sets is to define a generalized data model. A logical data model can provide a so-called abstraction layer separating the specific encodings of data from the logical structure of the data. A data model describes data entities and their relationships, and serves as a description of data independent of its physical representation. This is a critical enabler for applications, as all applications can be written against the data model, while data storage, transfer, and encoding formats can vary or be modified without breaking the apps. This creates a loose binding between apps and data, which is a good strategy in managing complex systems.

A data model might be organized around the concept of a User Profile through which a number of discrete data sets are associated. Differing data sets can be subject to different policies and protections, with differing access controls. For instance, highly sensitive Personally Identifiable Information (PII), such as name, address, phone number, can be isolated into its own data set and wrapped in the strictest security and access control policies. By separating this information from other data, these other data, by themselves, remain anonymized.

Other data sets included in a generalized User Profile might include:

- Viewing history
- Entitlements (services and programming packages)
- Devices (the STBs, PCs, tablets and mobile devices from which someone can access cable services)
- Preferences (opt-ins, parental controls, etc)
- Audience Qualifiers (personalization variables)
- Contact List (for explicit social connections)

The following diagram illustrates the key concepts for a generalized User Profile data model. This is simply a sketch of an idea and does not represent an implementation or specification.

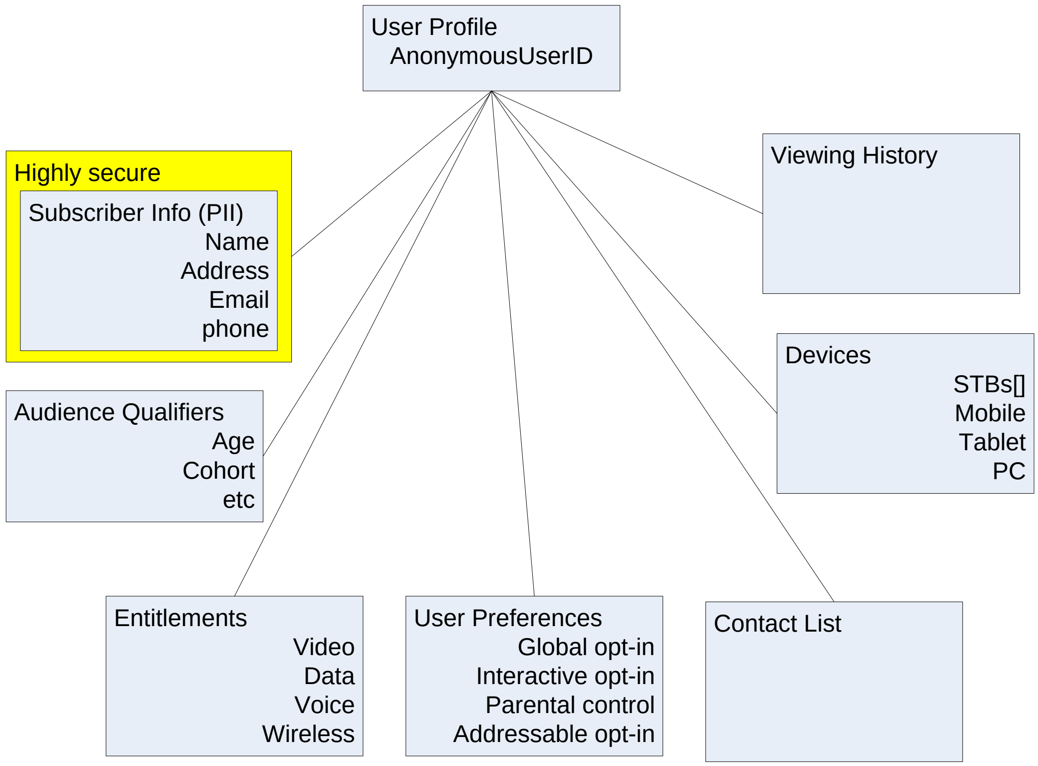


Figure 1: Sketch of User Profile data model

By isolating different types of information into separate buckets, operators can carefully manage access to each data set for every application. This is a good way to ensure privacy while enabling a wide variety of applications to access only the various pieces of information they need. By separating PII from other data sets, those data sets are by nature anonymized.

A User Profile also provides a means to organize and incorporate data sets not maintained by the service provider. Examples might include browsing history and viewing history on non-cable devices, demographic information, so-called affiliate qualifiers, such as status of home and car loan, credit card status, and so on. Also, touch-points with other social networking applications can be established, integrating the cable experience with Facebook, Twitter, and all the rest in a controlled and private way.

We suggest that the industry adopt a single data model for User Profile. This would serve a number of worthwhile goals.

It would provide transparency to regulators, subscribers, and partners. Given the potentially incendiary concerns over privacy, a simple and consistent approach adopted industry-wide might provide the best response.

A single model could also lead to efficiencies and cost savings, as application developers, both internal and external to operators, could work off of a well-known framework.

Finally, interoperability of applications would be possible. Again, this could lead to cost savings, since applications could be reusable on different systems. But more importantly, it could lead to the emergence of a social network comprised of all cable subscribers. The well-documented network effect is thereby amplified, enhancing the value of applications to users and the operators.

A single data model definition does not imply that operators necessarily modify their existing systems. As described above, a logical data model simply provides a description of data entities and their relationships, and does not describe their physical representation. Underlying implementations may vary widely, and many existing systems may already conform to a uniform model, to a greater or lesser degree.

The model is supported where a deterministic mapping between a systems data representation and the model is defined. Also, a given system might not have to support the entire model. A service provider may not collect or maintain certain bits of information defined in the model. This simply means that the system cannot fully support those applications that access that data, but may support a wide variety of other applications..

BUILDING SOCIAL INTELLIGENCE INTO APPLICATIONS

With well-formed data that can be generated by and made available to applications, we may now consider how applications interface with such data.

We might consider two broad categories of apps, those that are solely implemented within the bounds of a particular service provider – either through internal development teams or through partnerships – and those that might be interoperable across providers. While it‟s solely the purview of an individual operator to tackle the first set, the latter set requires coordination among operators.

A useful paradigm might be to consider that User Profiles, and their constituent parts, can be exposed through a set of data services Application Programming Interfaces (APIs). These APIs could be proprietary and made available only to applications that are internal to a service provider. This approach may appear to lead to a faster time to market since internal development teams can simply write directly to whatever access methods might be made available by internal systems. It might also appear that this approach benefits from "security through obfuscation," as external applications can‟t know the access methods to set or get data elements.

Another approach is to develop a set of data service APIs as a companion to the User Profile data model. It must be emphasized from the start that a commonly defined set of APIs does not mean that any and all applications may access them. Just as with any other cloud service or other enterprise-level interface, access is granted only to registered licensees and policy is securely enforced at runtime. The benefits of common APIs are that it collectively saves operators time as one design process leads to the API definition, a single test kit that can be used by everyone, and allows partners to port apps to multiple operators, allowing for the creation of a community that includes all cable subscribers.

Proprietary and standardized implementations of a data services API are not mutually exclusive. Particular operators may choose to implement one or the other, both, or neither.

The exact nature of a common interface is immaterial at this stage. Whether they are RESTful web services or based on some other technology can be determined by a technical committee tasked with their specification. Our concern at the moment is to explore the utility of such an API, whether proprietary or common.

A data services API available to multiple applications provides the means to both grow the User Profile data set and to extend the functionalities of applications.

An example is a content ratings application. Such an application might span multiple operators, therefore requiring and utilizing a common data services API implemented by multiple operators. There also may be one-off ratings apps provided by individual operators.

A ratings application is largely implemented as a network utility, or so-called cloud service, with any number of client instantiations. For example, user facing clients in ETV on legacy receivers, Java on tru2way, JavaScript on IPTV STBs and consumer-owned devices, or native apps on Android or iOS, not to mention web pages on your PC, could all connect to the ratings service. The application supports two primary functions –allowing viewers to rate content, and displaying the average rating of a piece of content.

With our User Profile model, any given user‟s ratings can be associated with that viewer. This enables applications to access an individual‟s ratings to provide recommendations or otherwise personalize their service, but it also allows the ratings service to harvest ratings from any number of users to calculate an average rating.

Once again, I‟ll stress that the data model is designed to protect privacy, and access would be securely regulated on a per application basis.

Examples of client apps that set ratings include a very simple slider bar that pops up on your TV, allowing you to use the right and left buttons on the remote to give the current programming a 1-10 rating, or an iPad app that allows one to set a ratings value for a content selection displayed on a navigation screen.

Any number of apps can then access the ratings database to display ratings for a given piece of content, from within the operators‟ core navigator on TV, from a tablet, web page, etc. Of course, with the proper licensing, an operators‟ app might connect to Rotten Tomatoes or other Internet ratings service and display their ratings. Perhaps the optimal ratings service blends cable generated ratings with other ratings and other metadata services.

If one were to quibble that while a content ratings service utilizes dynamically generated user data, it might not really be a social networking app, then we can extrapolate the basic functions a bit. Imagine now that I want to share my ratings with my cousin in Phoenix, or I wanted to see what my group of friends known as the „TV addicts„ have collectively rated as the best show on tonight‟s broadcast line-up.

A network application built in similar fashion as the ratings app could allow viewers to build up contacts and groups lists in a contacts data entity within their User Profile. Again with many clients able to access the service, I might edit my contacts from my PC, tablet, or mobile device, rather then trying to do so with the remote on my TV.

A service provider could extend the access permissions available to the ratings service to allow it to read contact lists, thereby enabling it to incorporate the sharing functions mentioned above.

The following diagram illustrates the key elements of a network application that accesses User Profile data and empowers a multitude of client applications.

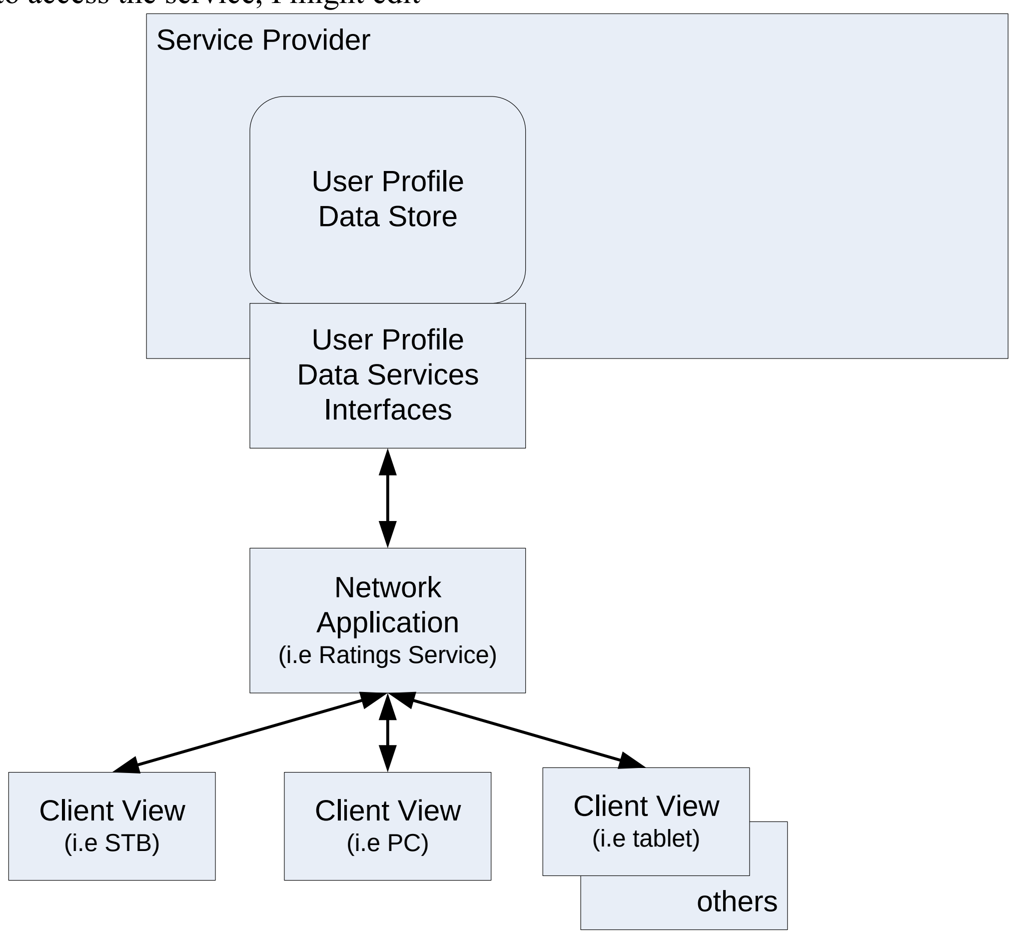


Figure 2: Application model for cable Social Networking applications

BUSINESS CONTEXT

There are many potential benefits to cable for pursuing social networking applications either on their own or in a federated manner. Not only would cable viewers get many of the same features available from Internet social networking apps, but these features also can incorporate cable capabilities, such as integration of video.

Perhaps an alternative approach would be to develop technologies to empower independent Internet applications to leverage cable assets. While this would certainly be good for the Internet app providers, and could improve the user experience for those apps, it‟s not clear where the value is for cable. It‟s conceivable that such a model might work to cable‟s benefit, perhaps by leveraging these services to support consumption of cable services.

But perhaps the greatest value to cable in pursuing a social networking strategy is to support a robust set of data services. Just as Google, Amazon, and Facebook view the data they collect about their users as key strategic assets, cable can also position itself to use social networking apps to extend and to utilize subscriber generated data. Ultimately, this can lead to cable playing a more intimate role in peoples‟ lives.

As new social networking applications become available through a cable provider, on the TV, the web, tablets, and Smartphones, the cable provider can play a more central role as not only a go-to entertainment source, but as a provider of all sorts of communications and utility applications.

With social networking apps, cable could provide compelling user experiences that enrich its value proposition, and increase subscribers contact. Ultimately, keeping customers focused on cable services, rather than sharing them with other service providers, is of tremendous benefit. Rather than ceding ground to Internet social networking sites, cable can retain mindshare and consumer data by providing social network functions directly.

Infrastructure and tools to support secure, scalable, and highly available APIs

Agustin Schapira
Comcast

Abstract

This paper describes a layered HTTP infrastructure designed to support web-based APIs in a standard, secure, scalable, and highly available fashion. The efforts described here are based on the observation that large scale distributed systems exhibit greater robustness, flexibility, and extensibility when they are conceived, built, and operated as a set of small independent but interconnected components. Leveraging the power of DNS routing and the HTTP protocol, we have built a platform that makes it easy for engineering teams within the organization to expose their services to other teams as HTTP APIs, and in turn to build their solutions based on other teams' APIs. The adoption of the API platform has reduced duplicate efforts, increased the overall security of our systems, provided greater control and visibility of how components are being used, and ultimately helped us innovate more quickly.

INTRODUCTION

It is by now a widely accepted tenet of web-based application development that de-coupling application logic and core data services from presentation layers enables the independent development of the two pieces and therefore reduces development cycles, increases the maintainability of the code, and allows for the quick migration to new platforms and devices (Burbeck, 1987). More general software engineering principles, such as decomposition (Parnas, 1972) and separation of concerns (Dijkstra, 1982), enhance overall architectural robustness, reduce the duplication of efforts, facilitate systems integration, and encourage the sense of ownership and responsibility over an individual component. The benefits of these practices go beyond pure engineering and into the business realm, as the same mechanisms that encourage the sharing of solutions across engineering teams and subsidiaries may also enable partners to further extend a company's core offerings through syndication arrangements, encouraging development of new, unforeseen solutions based on the existing, highly focused components (Benslimane, Dustdar, & Sheth, 2008). Ultimately, these strategies enable companies to innovate quickly in a technological landscape constantly in flux (Papazoglou & Georgakopoulos, 2003).

In the world of Internet applications and services, reusable solutions are most commonly implemented as components that expose an Application Programming Interface, or API, which other components (internal or external) may invoke in order to extend their own functionality. An API is a well-defined contract that specifies how a consumer should exchange information with a service provider in order to access and use its services (Papazoglou & Georgakopoulos, 2003). In the case of the web, the particular set of rules and data formats specified by the API usually rely on general communications and architectural patterns, such as the SOAP (W3C, 2007) and REST (Fielding, 2000) styles, to specify how the actual messages should be exchanged and interpreted (e.g., how to represent the space of possible actions, how to encode the messages, how to report errors, etc.). Ultimately, all of these architectural styles use HTTP as the underlying communications protocol.

This paper describes a platform to encourage and support, within an organization, the creation and use of web APIs. The starting point for this work was the question: "*What are best ways to encourage the creation, operation, documentation, maintenance, and use of APIs in a large organization, where tens of engineering groups and hundreds of developers are working on their own problems and solutions?*" In other words, if we believe that this particular engineering practice can improve our agility, robustness, and time-to-market, then how can we encourage its adoption in a large organization with multitudes of product lines, deadlines, engineering practices, and goals?

To guide our work, we established five core principles, and then sought to provide answers and solutions to address them:

1. **Exposing APIs should not impose a big burden on the owner of the service** --and, in fact, it should have obvious benefits. To address this issue, we built shared infrastructure that solves the common difficulties involved in supporting APIs (security, access policies, and scalability), and established standard protocols and procedures for incorporating an API onto the infrastructure. By simply "plugging in", API providers get all those issues solved and are freed to focus on their core competencies.

2. **Leveraging existing APIs should not impose a big burden on developers** --the ultimate goal being that a developer who learns how to use one API will immediately also know how to access all other internal APIs. To address this issue, we established well-known protocols for making calls through the infrastructure onto the APIs, and based them on widely used open standards. Furthermore, we created and distributed libraries for several programming languages to take care of the details of making HTTP requests according to the infrastructure's requirements.

3. **API owners should not lose control of their services** --and, in particular, they should be able to make all the choices afforded by the HTTP protocol, as long as those decisions do not interfere with the core security requirements. To address this issue, we chose a highly distributed operations and control structure. We ensured that the API infrastructure imposes as few demands as possible (and that those demands are based on open standards), placed the operational responsibilities on API owners, and gave them complete freedom on issues such as data formats, URL patterns, cache control directives, etc.

4. **The use of APIs should not introduce unnecessary obscurity** in the underlying communication channels --and, in particular, the API infrastructure should preserve as much as possible the transparency and ease-of-use of HTTP and of the architectural styles built on top of it. To address that issue, we introduced HTTP headers that trace the processing of requests as they travel through the infrastructure, and built tools to test and debug APIs from a web browser.

5. **Delays for getting on board, either as an API provider or a consumer, should be minimal** --because people lose interest very quickly when faced with too many hurdles to test a new technology. To address this issue for API providers, we created a sandbox version of the infrastructure --with very few limitations compared with the production environment-- where it's possible to get a new API up and running in less than 30 minutes. For API consumers, on the other end, we built a Developer Portal where they can find an API catalog, read detailed documentation, and even access live versions of APIs directly from their web browser, so they can quickly learn about the inputs and outputs of an API without writing a single line of code.

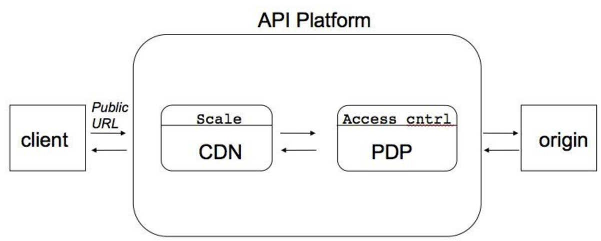
The remainder of the paper describes in more detail the shared API infrastructure that we built, the communications protocols that we imposed on it, and the Developer Portal where we centralize the documentation of existing APIs.

API INFRASTRUCTURE AND ENVIRONMENTS

The most common issues that engineering teams have to face when exposing APIs are security, access control policies, and scalability. In order to address these issues, we built an API infrastructure that enables service providers to simply "plug in their APIs" (following some basic protocols) and effectively delegate those responsibilities to the shared platform.

An API integrated with this infrastructure offers its resources via *public URLs* that bind (via Layer 3 / DNS routing) to the platform infrastructure. Through this level of indirection, an HTTP request from an API client is first routed through the platform so that, by the time it reaches the API's origin servers, most of the scalability and security decisions have been made.

Following the separation of concerns principle, we conceived of the infrastructure as a layered architecture, with a set of HTTP intermediaries tackling the different issues. Internally, the request is routed (through a combination of Layer 3 and Layer 7 mechanisms) through a small set of specialized HTTP intermediaries. Scalability and high-availability are the responsibility of an initial load-bearing layer, implemented through a **Content Distribution Network** (leveraging its very-large-scale edge topology, caching services, and high-availability and failover mechanisms). This outermost layer also acts as a **Policy Agent**, enforcing policy decisions made by a second layer, the **Policy Decision Point**, to which the request is further routed, and which is responsible for authenticating the client, determining whether it has access to the requested resource, applying business policies (such as rate limiting), and logging detailed information about the request for reporting and monitoring. If the request is valid, it is finally routed to the API origin. The response from the origin traverses back downstream though the intermediaries, which may decorate it, and finally back to the client that initiated the request. Figure 1 illustrates this flow.



An important requirement while designing this infrastructure (reflected in the fifth guiding principle) was that developers should be able to "plug" an API and start offering it to others in less than 30 minutes –the underlying assumption being that longer setup times would discourage developers from trying out the platform and would therefore reduce the levels of adoption throughout the organization. This requirement led to the creation of two almost identical but separate environments on top of the platform: a ***Sandbox environment***, with an almost self-service model of API provisioning and where developers can test their APIs in development and QA scenarios, and a ***Production environment***, with SLAs and complete branding for operationally ready APIs --but with a necessarily slower and more involved setup procedure.

The internal structure and functioning of the two environments is almost identical. The major difference resides in the fact that sandbox APIs are exposed with *unbranded* public URLs (e.g., as a sub-domain of a generic *api_plaform.example.com/*), whereas production APIs get their own branded hostnames (e.g. *voice.example.com/*). This is due to the fact that the procedures for establishing new DNS domains and acquiring SSL certificates for them are much more involved than simply creating a sub-domain for an existing domain (and, in particular, require lengthier setup procedures with the load-bearing CDN partner). But beyond that distinction, there are no other differences. This combination of environments allows API owners to get started very quickly (simply by filling out a form with data about the desired edge URL and the location of the origin server and waiting a few minutes for operations staff to initialize an endpoint and turn the API on), while at the same time retaining the ability to easily migrate to a production environment when the APIs and their owners are ready.

The following sub-sections detail different communications protocols imposed on top of the layered architecture in order to guarantee security and offer scalability. Since the differences between the two environments are minimal and have already been detailed above, no further distinctions between them will be made.

Internal Security

The API infrastructure allows API origins to delegate decisions of security. This means that when an API origin receives an incoming request, it should be free to assume that the security and access control policies have already been enforced, and that it may therefore process the request without having to concern itself with those issues. Because API resources are offered via public URLs accessible through the open Internet, however, the API origin must ensure that the request is indeed arriving from a component in the shared infrastructure --otherwise, a malicious client could access it directly and effectively bypass all the security restrictions. For that reason, all layers in the architecture, including the API origins and the Policy Decision Point, are asked to implement what we have called the "*Intermediary Authentication Protocol*". In simple terms, the protocol requests a) that all intermediaries attach HTTP headers to their requests with a ***message authentication code*** (MAC) (Federal Information Processing Standards, 1985) signed using ***shared keys***, and b) that both intermediaries and origins verify the existence and the validity of the code in all incoming requests. (The outermost layer receives requests directly from clients, not intermediaries, and is therefore not expected to verify incoming headers, but it should, on the other hand, provide headers identifying itself as an intermediary before forwarding the request upstream to the next layer).

The protocol requires intermediaries and origins to maintain a list of well-known ***Intermediary IDs*** and their corresponding ***secret keys***. Two internal HTTP headers are attached and verified by all intermediaries and origins involved in the upstream processing of a request. The first header provides information about the request from the intermediary (an ID for the intermediary, a timestamp, and a nonce). The second header provides a hashed MAC (HMAC) computed by concatenating the information from the first header with details about the request (e.g. the *base URL)*, and signing it with a shared key in order to ensure that the information in the other header cannot be tampered with.

Each intermediary should validate an incoming request by 1) making sure that the headers exist and are in the right format, 2) retrieving the secret key that corresponds to the Intermediary ID, 3) creating a signature using the same function as above, 4) comparing the new signature to the value of the signature received in the second header, 5) retrieving the timestamp and ensuring it falls within a certain window (past & future), and finally 6) retrieving the nonce and ensuring that the request is unique (within that window). If any of these steps fails, the intermediary or origin should immediately return an HTTP 403 response, with a body detailing the reason (e.g. "*signature invalid*"). Otherwise, it should continue the processing of the request in regular fashion, but previously replacing the *Intermediary Authentication* headers that it received with new values, so as to identify itself to the next layer upstream.

Access Policies - Consumers

Once it is guaranteed that origins will *only* accept requests coming from the shared API infrastructure, it is easier to model and apply security and access control guarantees to APIs. In the most basic model, which we call the "*Consumer Authentication Protocol*", API owners are able to choose who gets access to their services, and under which circumstances. The enforcement of these restrictions (i.e., the guarantee that only authorized users are allowed to access the API) is the responsibility of the Policy Decision Point component of the platform. Through configuration dashboards and out-of-band operations, API owners can grant **Consumer Keys and Secrets** to individual consumers**,** and specify the set of policies (e.g., rate limits, time and geographic restrictions, etc.) that must be respected when each of those consumers issues a request against their API. API consumers, in turn, are asked to sign their requests using those credentials, following in particular the 2-legged version of the widely adopted OAuth 1.0a standard (Internet Engineering Task Force, 2010).

The original OAuth specification should be consulted for full details, but in summary the protocol requires that HTTP requests be decorated with an ***authorization string*** provided in the *Authorization* HTTP header. The authorization string includes the consumer's key along with a hashed message authentication code (based on the method, the hostname, the path, and the GET/POST parameters of the request) signed with a concatenation of the consumer's key and its secret (not to be confused with the Intermediary's key and secret as described in the previous section). The protocol also requires a timestamp and a nonce, to avoid replay attacks.

To authenticate an incoming request, the Policy Decision Point retrieves the shared secret corresponding to the consumer's key provided in the Authorization HTTP Header, and uses it to build the OAuth signature that it should expect given the request that it is actually processing (its hostname, method, path, and parameters). If the expected and received signatures don't match (or if the authorization string is missing or incomplete, the timestamp is older than a certain allowed time window, or the nonce parameter has already been received within that window), the Policy Decision Point instructs the outer layer in the infrastructure (the Policy Enforcement Point) to DENY the request, with an HTTP 403 code and an explanation in the body of the response (e.g., "*Missing Consumer Key*", "*Invalid Signature*", etc.), without forwarding the request on to the API origin. If the signature is correct, on the other hand, the Policy Decision Point considers the consumer properly authenticated, and then proceeds to apply business authorization rules such as rate limiting. Eventually, if the request is fully authorized (i.e., if it's coming from an authorized consumer, and all the access policies for that consumer are successfully passed), then it is forwarded to the API origin with an internal header that reveals the identity of the consumer, should the origin need it for tracking or logging purposes.

Access Policies - Resources

The Consumer Authentication protocol ensures that only requests from authorized consumers are allowed, and is an effective solution when the consumer can be trusted to keep its credentials securely. For cases when the security of the credentials cannot be guaranteed (e.g., when the consumer is an application running on a mobile device), our platform also supports a "*Resource Authorization Protocol*", which gives a particular instance of that consumer access to **a single resource** (e.g., a single user account). With this protocol, if an instance of a consumer application gets compromised (e.g., a particular mobile device gets hacked), then *only the particular resource for which that device had been authorized* is compromised.

The protocol is identical to the 3-legged version of the OAuth 1.0a standard (Internet Engineering Task Force, 2010), and a slight variation over the Consumer Authentication Protocol. In summary, in addition to Consumer Keys and Secrets, valid requests must be signed with an **Access Token and Secret** as well. The protocol defines a set of flows that enable a consumer to acquire an Access Token and Secret for a particular resource (for example, if the resource in question is a user account, the protocol defines a flow by which the consumer application should redirect the user's web browser to an endpoint where the API platform will issue a **Request Token** -- potentially, but not necessarily, first redirecting to a login page and requiring that the user for whose account access will be granted explicitly authorizes the access--, which the consumer application then exchanges for an Access Token and Secret via a backend call to the API platform). Regardless of the flow chosen, the API platform internally correlates the Access Token to the identity of the particular user or resource to which access is being granted. Once the authorization tokens have been acquired, the consumer application will submit API requests signed using a similar mechanism to the one required by the *Consumer Authentication* protocol, with the additional requirement that the Access Token be also provided in the Authorization string, and that the signature be computed using both the Consumer Secret and the Access Token Secret.

The Policy Decision Point, in turn, will receive the request, verify the identity of the consumer, validate the signature, and apply access policies to the consumer, just like in the Consumer Authentication Protocol. Additionally, the Policy Decision Point will retrieve the identity of the resource that corresponds to the Access Token in the Authorization HTTP Header, and apply access policies that apply specifically to that resource. Finally, if all checks pass, the Policy Decision Point will forward the request to the API origin, modifying the original URL to include the identity of the requested resource. For example, if the consumer makes a request to *www.example.com/myaccount* with a given Access Token, and the Access Token corresponds to a user with GUID *123*, then the request to the API origin will be *api_origin.example.com/accounts/123* (the details of URL re-writing can be configured on a per-API basis, of course).

Because access permissions are attached to a single resource, and because it effectively hides the actual identity of the resource, the "Resource Authorization" protocol is also useful in cases where the API owners want to give access to the API to an external partner, without having to reveal internal resource identifiers such as UIDs. It is also useful when business policies demand the explicit approval of the user on whose behalf the consumer application will be issuing requests.

Scalability

In addition to the issues of security and access controls described above, API owners also have to address problems of scalability when they choose to expose their services for access beyond the applications for which they were originally intended. Fortunately, scalability is another issue that may be well addressed by a common API infrastructure. In particular, our API platform contains a load-bearing tier, implemented on top of a commercially available CDN, which is setup to be the first layer to handle incoming requests (and the last to process outgoing responses). Responses from API origins may be temporarily stored at this layer, which may later choose to respond to subsequent requests directly from its cache, without the need to first forward the requests to the Policy Decision Point and the API origin.

Serving content directly from the outer load-bearing tier, however, introduces a new security issue: if new requests don't have to travel to the Policy Decision Point, then how can the platform guarantee that only authorized consumers will be allowed to access the protected services? The "*Edge Authorization Protocol*" addresses that problem, effectively offering the option of trading fine grain policy decisions for massive scale, high availability, and optimal performance. In this model, the Policy Decision Point is allowed to optionally issue cryptographically secure authorization assertions in the response to requests from a consumer. Those assertions act as logically sessioned tokens: if and only if the consumer replays that authorization token in subsequent requests may the load-bearing edge network assume that the consumer has already been authorized, and that therefore the request may be processed without further engaging the Policy Decision Point.

In terms of actual implementation, the Edge Authorization protocol specifies that the Policy Decision point may decorate the downstream response with a cookie that encodes the IP of the consumer for which authorization should be assumed, the duration of the authorization, a list of paths for which the authorization should be valid, and a signature. Before sending the final response to the client, the edge tier will check for the existence of the cookie and, if found, it will encrypt its value using its own secret key (to make it opaque). The client may then issue another request to the same API. On its end, all the standard rules apply; in particular, the request must still be signed using the *Consumer Authentication* protocol, through which the consumer identifies itself to the API infrastructure. However, the client may also replay the Edge Authorization cookie that it received in the previous response; in that case, the edge processor will decrypt its value, validate it (IP address, expiration time, path, and signature) and, if all checks pass, return content directly from its cache. It is also possible that the edge tier does not find the desired content in its cache (or that the stored response has expired); in that case, the edge tier will have to forward the request all the way to the origin, but it will be allowed to *bypass the Policy Decision Point* and engage the Service Provider interface directly, since the request has already been authorized.

From the point of view of the client, there shouldn't be any difference in the way that it issues the request (with the exception of the extra cookie) or the way it receives a response (again, with the exception of the cookie). The same is true for API origins: how the request is routed through the API infrastructure is of no concern to them. With minimal effort, then, the shared caching and high-availability layer (along with the Edge Authorization protocol) enables API owners to take advantage of infrastructure that may be too expensive or too complicated to maintain for a single API (thereby reducing the burden of exposing their services, per the first principle) while, at the same time, giving them full control, through the well established HTTP cache control directives, to decide which sections of their APIs should be cached, which ones require refreshes, which ones may be served in a stale state in the case of origin downtimes, etc. (following the third principle).

It must be noted that the Edge Authorization protocol introduces an obvious trade-off, in which scalability and performance are gained at the expense of fine-grained security decisions. If responses are served directly from the cache or the origin, bypassing the Policy Decision Point, one loses the ability to immediately revoke access to a consumer (i.e., the edge will continue to grant access until the edge authorization cookie expires). The rate limiting and usage reporting capabilities afforded by the Policy Decision are also lost for requests that are edge authorized. For these reasons, API owners must retain control about which consumers may participate in this Edge Authorization protocol, and for how long the authorizations should be granted. In practice, the Edge Authorization protocol is only used with selected APIs and consumers, and the expiration times for the edge authorization cookie usually set to no more than 1 or 2 minutes.

Traceability

For purposes of traceability and operational awareness, necessary given the various layers involved in the processing of a request through the API platform, intermediaries (including the API origins) are asked to decorate upstream requests (that they process themselves or forward to other layers) and downstream responses (that they return to upper layers) with values that they append to an internal HTTP header. The details of the header are specified by our "*Message Exchange Fingerprint Protocol*".

Specifically, an HTTP header is added by all systems that participate in the handling of a request. Values from different components are separated by a single space, and each value in turn consists of a dash-separated set of a) a direction prefix, either UPSTREAM ("u") or DOWNSTREAM ("d"), b) a component identifier, granular enough to provide uniqueness to the fingerprint (e.g., in a Java system this could some arbitrary "system ID" plus a "process ID" plus a "thread ID", and c) a timestamp, recorded in Unix time (milliseconds since the UTC epoch).

An example value for the internal header, as received by the consumer with the full HTTP response, is shown below (pretty-printed to show processing brackets):

```
u-cdn75209+72.246.30.14+138036545-1287680639000
  u-proxyworkeri6866f301.pdp.com8735-1287680639532
    u-t24005274509060@API-QA-1287680639596
    d-t24005274509060@API-QA-1287680639597
  d-proxyworkeri6866f301.pdp.com8735-1287680639591
d-cdn75209+72.246.30.14+138036545-1287680639800
```

This shows which components participated in the processing of the request (a CDN server at 72.246.30.14, a Policy Decision Point worker with the ID 6866f301, and thread ID t24005274509060 on the API-QA origin). The header is also useful to show the time each bracket took in processing the request (including the calls to the lower layers): 1ms at the origin, 59ms on the Policy Decision Point, and 800ms total at the edge layer (including the time spent waiting for a response from the next layer down).

DEVELOPER PORTAL

The multi-layered architecture of the API infrastructure, along with its protocols and the procedures that providers and consumers must follow, address the first four principles that guided our efforts to encourage the creation of APIs throughout the organization: *minimize effort for API providers*, *minimize effort for API users*, *leave as much control as possible with API providers*, and *do not obscure the communication protocols*. The sandbox platform, a copy of the production environment but running under a default edge URL, tackles the fifth principle of *minimizing unnecessary delays* for API owners. In order to increase the adoption of existing APIs and address the fifth principle for **API consumers**, we created a Developer Portal, a central place where service providers advertise and document their APIs, and where potential users quickly find out about the APIs offerings and learn how to use them.

Access to the Developer Portal is limited to employees within the organization, and their access credentials tied to the organization's LDAP servers, to minimize the effort required to participate in the portal. Once logged in, visitors can access Wiki pages that explain the inner functioning of the API platform and its protocols, and in particular provide detailed instruction for signing requests using the Consumer Authentication and Resource Authorization protocols (libraries for several programming libraries are also provided here).

More importantly, visitors to the Developer Portal have access to a full API Catalog, listing all the available APIs categorized by group or function, and with links to find further information about each. This is where developers learn about the APIs that they can use to solve the problems they are working on. Each API, in turn, has its own set of documentation pages (which are carved out from the Developer Portal's URL namespace when the API is first exposed through the infrastructure), and where we encourage API owners to provide extensive documentation. Most commonly, the documentation includes the basic definition of endpoints, URLs, path structures, and query parameters, along with the kinds of responses that API consumers should expect (and the error codes that might be returned).

So that future consumers may get as real a taste for what the API can do, moreover, we allow API owners to attach what we call "*LiveDocs*" to their API's pages. LiveDocs are small, inline boxed forms that can be easily embedded in documentation pages, and through which potential consumers may specify parameters and issue *live* calls (from the browser only) to the corresponding API, without having to worry about getting API keys or having to learn how to sign and issue HTTP requests (a responsibility delegated to the LiveDocs plugins). In addition to displaying the responses returned by the APIs, LiveDocs also present detailed information about how the request was issued (so that potential users can learn about how to sign their requests), the headers that were returned (so they can learn about special protocols such as Edge Authorization), and the intermediaries that participated in the handling of the request (so that they can learn to use the Message Exchange Fingerprints protocol and explore the inner workings of the platform). In this way, general documentation and LiveDocs allow potential consumers to learn as much as possible, and with very little effort, about the set of APIs that are available through the platform, and how they can use them.

The LiveDocs boxes are limited to the parts of the API that owners want to bring attention to in their documentation pages (e.g., they only allow for a small subset of parameters, or provide closed choices for a parameter in the form of a drop-down list). Sometimes developers need more than that. For that purpose, and once they have requested and acquired credentials to access those APIs, developers can access a more powerful on-line tool (also available from the Developer Portal), which we call "*codebug*". Through codebug, developers have full control over all the details of their request: they can specify the full URL of the request, use HTTP verbs other than GET, provide specific headers, add Pragma directives that control the behavior of the platform, and even change signature methods. The codebug console takes care of signing the requests using the developer's credentials for the API, and then presents the same information returned by LiveDocs: the request and response headers and parameters, the actual response body returned by the API origin, and tracing information showing how the request was handled by the platform.

We believe that being able to interact directly and issue live calls to the API, changing parameters and observing the XML or JSON that the API returns, makes a big difference in terms of the adoption of APIs. Developers can effectively interact with an API without having to write a single line of code, and therefore are able to explore and adopt existing APIs quicker (and based on first-hand experience).

LESSONS LEARNED - FUTURE WORK

Since the release of the API platform within the organization, we have seen an important increase both in the number of APIs offered (for services which would have normally remained closed), and in the adoption of APIs across organizational boundaries (to reuse solutions created by teams from very different groups). More APIs are being incorporated every month, and we may be getting close to a tipping point, from which all (or most) teams will be expected to have solutions running on the platform. Ultimately, this is enabling us to reduce the duplication of efforts, increase the security of our applications, gain deeper understanding and control of who accesses each service, and innovate with shorter development cycles and reduced time to market.

The experience of building and operating the API infrastructure has also given us data to verify our assumptions about what was required to encourage the production and use of APIs, and to discover ways in which we could improve. In particular, we sense the need for improvement in the following areas:

1. **Intra data-center calls need to be treated differently**. The API platform, as detailed in previous sections, consists of an initial CDN layer followed by a Policy Decision Point. In our initial implementation, these two layers are implemented and operated by existing commercial offerings, outside of our internal data-centers. This architecture imposes a significant (and unnecessary) latency overhead when both the API consumer and the API origin are located within the same data-center: requests from the consumer have to travel all the way out of the data-center into the CDN, and then to the Policy Decision Point somewhere in the open Internet, before coming back to the data-center to the processed by the API origin (and the responses need to follow the inverse path).

To address this issue, we are evaluating solutions to bring these layers in-house.

2. **Intermediary Authentication is sometimes hard to deploy**. As we engaged internal teams to incorporate APIs onto the platform, we discovered that some of those services are owned and/or operated by third parties, and it is either difficult or expensive to access their source code. As a consequence, adding the necessary filters so that the origins conform to the Intermediary Authentication protocol is a non-trivial exercise.

To address this issue, we deployed an internal gateway whose sole purpose is to front those closed systems: the gateway handles Intermediary Authentication, and the API origins sit behind a firewall that only enables requests from the gateway. While adding (minimal) latency, this enables us to expose services that otherwise would have to remain closed (or would be very expensive to open up without security risks).

3. **API owners need to have complete access to traffic logs**. When an API origin sits behind the API platform, its web server access logs lose richness. For example, it is no longer possible (or easy) to find our which IP addresses are hitting it (since they are all

coming from the Policy Decision Point), or filter out requests that had to be dropped because of security violations (since they are dropped by the Policy Decision Point), or which ones resulted in cache hits (since caching is handled by the CDN layer). This is useful data that API owners should not lose access to.

To address this issue, we are currently working on a prototype solution that fetches and combines log files from the three infrastructure layers (the CDN, the Policy Decision Point, and the API origin) and makes them available to API owners. In the future we will consider integrating this tool with the existing API management dashboards.

4. **Exposing APIs implies a cultural shift, which takes time and education.** Early adopters of the API platform were easy to get on board: they were already interested in exposing their APIs, they felt comfortable with the security implications, they understood the details of layered HTTP architectures, and they were able and willing to use low-level tools (and cumbersome procedures) to configure their endpoint. As we push for adoption beyond this initial group, however, it is becoming more important to be able to explain in less technical terms the advantages of opening up APIs, to detail more explicitly the security guarantees that the platform provides, and in general to provide more guidance about the best ways to go about doing it.

To address these issues, we are working on Best Practices documents, open security reports and assessments, easier to use (and more accessible) forms and procedures for creating API endpoints in the sandbox, and simplified key management policies. We are discovering that, in order to gain wide acceptance and adoption, the existence of these materials is as important as the technical merits of the platform.

ACKNOWLEDGEMENTS

This API project was initiated at Comcast by Matt Stevens, who produced the original design of the architecture and core protocols. It was later maintained and expanded by Benjamin Schmaus, with further help from Peter Cline, Michajlo Matijkiw, and Matt Hawthorne.

During the implementation phase, we got invaluable help from Philip Grabner, Satish Narayanan, and Cory Sakakeeny from Akamai, and Bill Lim and Peter Nehrer from Mashery. Also fundamental were the contributions of Bahar Limaye, from Software Security Consultants, who performed in-depth reviews of the security of the platform.

Many people within Comcast pushed for the adoption of the infrastructure throughout the organization, and I'd like to thank in particular Rich Woundy, Rich Ferrise, and Hai Thai.

REFERENCES

W3C. (2007). SOAP version 1.2. Retrieved from W3C: **http://www.w3.org/TR/soap/**

Burbeck, S. (1987). *How to use Model-View-Controller (MVC).* Retrieved from http://st-www.cs.illinois.edu/users/smarch/st-docs/mvc.html

Benslimane, D., Dustdar, S., & Sheth, A. (2008). Services Mashups: The New Generation of Web Applications. *IEEE Internet Computing , 12* (5), 13-15.

Dijkstra, E. W. (1982). On the role of scientific thought. In *Selected writings on Computing: A Personal Perspective* (pp. 60–66). New York, NY: Springer-Verlag New York, Inc.

Federal Information Processing Standards. (1985). *Publication 113 - Computer DataAuthentication.* Federal Information Processing Standards.

Fielding, R. T. (2000). *Architectural Styles and the Design of Network-based Software Architectures.* Irvine, CA: University of California, Irvine.

Internet Engineering Task Force. (2010). *RFC 5849: The OAuth 1.0 Protocol.* (E. E. Hammer-Lahav, Ed.)

Papazoglou, M., & Georgakopoulos, D. (2003, October). Service-Oriented Computing. *Communications of the ACM , 46* (10), pp. 24-28.

Parnas, D. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM , 15(12)*, pp. 1053-1058.

BRINGING A CENTRALIZED SERVICE DELIVERY PLATFORM FOR ADVANCED NAVIGATION AND SEARCH TO COMPANION DEVICES AND THE EXISTING STB

Steve Tranter
NDS

Abstract

Cable technologists are facing pressure to ensure their networks and in-home experiences include compelling and intuitive access to the most content possible on as many devices as possible, while relying on the limiting capabilities of deployed set-top boxes (STBs). Guides for cable systems are generally restricted to the abilities of the least capable STB deployed, due to the increased effort it takes to integrate and support variants of the user interface (UI) across many STB types. Thus the guide is often overlooked in favor of limiting the integration effort.

This paper will present technology considerations for decoupling the guide appearance (skin) from the underlying STB architecture, such that updates and re-skins of the user experience can be quickly achieved and optimized for the STBs on which they are running. Adopting this service layer approach also enables the operator to move some of the complex UI features, such as search, from the limiting processing environment of the STB to the headend, simplifying the STB integration further. Migrating the guide functionality from the STB to the headend not only extends the capabilities of the STB, but also provides a centralized architecture for supporting companion guides and services on consumer purchased devices, such as PCs, iPads, etc.

This paper will review the various service layer approaches that can be adopted within the STB software architecture, such as EBIF (XML) and OnRamp (Java), highlighting the advantages and disadvantages of each. Furthermore, the data transport architecture will be discussed to show the infrastructure and related standards required to centralize guide functionality in the headend (DOCSIS and DSG implementations, etc.).

Finally, considerations of the centralized metadata headend will be explored to highlight the technology options that can be adopted to extend the guide and MSO service offering to companion CE devices on which the MSO wishes to have a presence. This includes architectural decisions regarding the application server environment and the inclusion of a service delivery platform to provide standard, pre-certified application programming interfaces (APIs) into the TV ecosystem for application developers, such that applications can be contextual to and control the TV viewing experience (channel change, DVR recording, etc.) as well as

taking advantage of in-home networking standards, such as DLNA.

INTRODUCTION

Since the first deployment of digital STBs the guide and user interface have been tied to, and hence limited by, the STB on which they are running. The user interface has evolved since its first deployment to cater to new services being deployed by MSOs, such as video on demand (VOD), digital video recording (DVR) and simple interactivity through add-on user-agents (EBIF). Every new service and corresponding user interface enhancement has had to run on the same STBs that were originally deployed. As such, each new user interface enhancement has been implemented as a disconnected application that independently provides an interface to the STB's new feature and not to the overall service being offered by the MSO. For example, this is seen in separate user interfaces for linear, VOD and DVR content, even though the same content can be found across all three.

Further, guide user interfaces have been limited by the hardware and networking capabilities of the STB platforms. Because of limited memory, processing and graphics capabilities, most legacy guides are built in native C code using 640x480 resolution screens, with little to no animation, limited color palette, and basic video scaling capabilities. These traditional guides have limited networking capabilities for providing real-time data updates and therefore limit the amount and frequency of guide data updates that can be accessed by the STB. Adding new features to the guide requires lengthy development and integration cycles. As a result, a great deal of integration work is required for every new guide feature and new STB model that is deployed. This is necessary to ensure that the guide and supporting data run correctly on every STB and with each disconnected application that is resident on these STB platforms.

The legacy STB dependency that exists today has resulted in:

- Restricted user experiences, dated by 10 year old technology; and
- Complex development, integration and testing cycles, requiring excessive manpower and long delays in getting even minimum guide changes to deployment.

Now, with the proliferation of broadband connectivity (DOCSIS 2.0 and 3.0) to the home, the STB restrictions described above are being eroded away. Having a managed broadband connection means that much of the data storage and processing that has traditionally been performed on the STB can now be performed in a centralized data center and the resulting information delivered into the home when required. In addition, not only can the data center support guide services for DOCSIS-enabled STBs, but also any broadband IP connected device, such as PCs, tablets or mobile devices. Thus, enhancing the way an MSO promotes their services and providing control to viewers through companion devices.

Figure 1: Example Guides for Broadband IP Connected Devices Supported by a Centralized Data Center

This new cloud-based architecture provides greater efficiencies for guide deployment – five months to deploy an iPad guide application compared to 18 months for a legacy STB guide. These dramatic development cycle reductions are as a result of eliminating much of the regression testing that is required for legacy STB guides. As a result, the development staff needed is also reduced considerably; typically only 10 developers are used on an iPad application compared to 300 for a legacy guide across multiple STB models.

ARCHITECTURE

MSOs are gravitating towards cloud-based architectures for video navigation and delivery, which enables them to keep up with the accelerated product cycles required to stay competitive. As more and more functionality is moved from the STB to the cloud, Service Delivery Platform (SDP) infrastructure is being utilized to enable these cloud-based services. SDP architecture provides the benefits of delivering advanced guides and services to companion devices as well as integrating with the existing MSO headend to provide improvements to the guide on DOCSIS enabled STBs.

The SDP architecture is a multi-tier, web service-based platform architecture that integrates with MSO's TV headend infrastructure, as well as new Internet-based social networking, enhanced data, and content providers. Internally, SDP architecture consists of a web-services presentation tier, an aggregation tier to Internet-based services and an adaptation tier to existing operator backend capabilities, which includes bridging components for certain legacy environments.

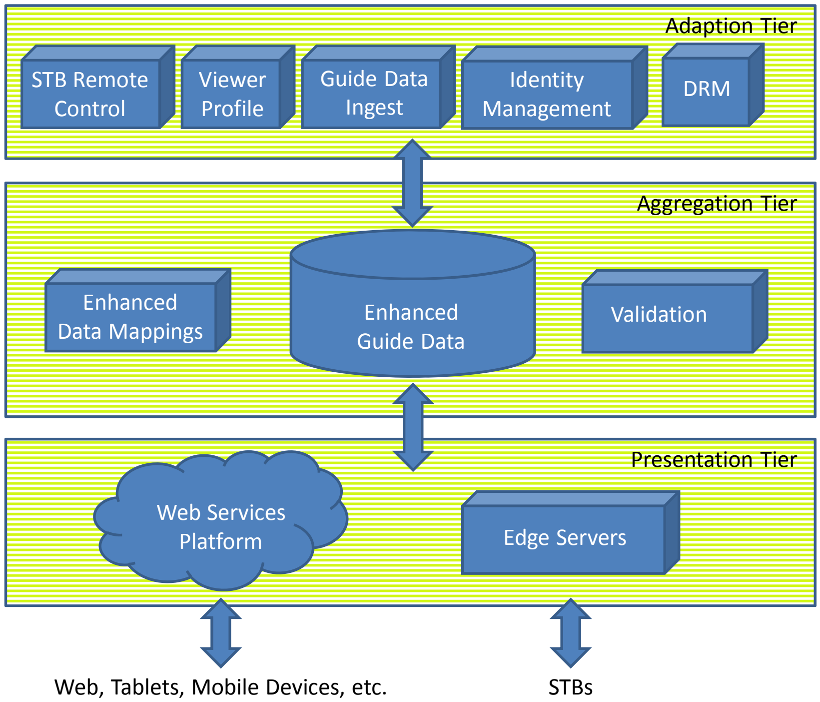


Figure 2: Centralized Service Delivery Platform Architecture

PRESENTATION TIER

The presentation tier is a standard Web services platform (WSP) that can be deployed in the cloud or via edge servers in an operator network. It includes security via standards such as HTTP and OAuth. The user authentication and authorization is tightly integrated with the operator's identity management systems via the adaptation tier, permitting operators to further control which web services are available to which applications. The presentation tier also includes monitoring and analytics such that operators can be aware of the usage of their services across different applications, devices and demographics.

When a device requests information through an SDP architecture web services API, the SDP can use service tokens to identify device type and user identity, which are used to filter service metadata and content references appropriately to the application, user and/or device type. Service accessibility rights are exposed to the presentation tier either through the adaptation tier, which utilizes the operator's own identification management systems to determine application or user profile information, or through the aggregation tier to access similar third party account information. Applications using the presentation layer may leverage this service accessibility information to modify the user experience. For instance, by determining IP address and geo-location, an application may act differently when accessed within the home or outside the home.

AGGREGATION TIER

The aggregation tier provides the majority of the improvements to the SDP-delivered user interface, compared to that of legacy STB guides. The process is predominantly focused on aggregating the metadata for all of an MSO's services, enhancing the metadata relationships and validating the resulting relational data structure. The result is a superset of metadata that can be queried, filtered and utilized by guide applications on the various client devices.

In the aggregation tier, operator-specific metadata, user profile and identity information are aggregated with third party enhancements such as additional content references or social network linkages. This enhanced metadata may be available from public sources on the Internet or through federations between network operators and other service providers. By aggregating metadata, the third party information may be used in a contextually appropriate manner, for instance by providing references to critical reviews on Rotten Tomatoes, to a YouTube video for a live linear program, or a viewer's Internet identity, on sites such as Facebook.

The aggregation tier enables the MSO's guide application to aggregate access to content from a variety of sources (linear, VOD, DVR, over-the-top (OTT), etc.) into a single, coherent guide experience. This means that the guide can now present a wealth of information for content from a variety of sources, providing greater choice and satisfaction to viewers. Also, by storing this guide information in the headend, STBs no longer need to store program information beyond the next day or two, which frees up memory in the STB, for caching other information. One way this can be utilized is for the STB guide to dynamically request and cache images and data in advance, related to the current guide view, such that guide screens can be quickly populated, with no delays as the viewer navigates the guide.

The aggregation tier performs much of the data manipulation, searching and filtering that are traditionally performed by legacy STB guides. In addition, the aggregation tier generates relational links between the various sources of content, such as linear, VOD and DVR as well as enhancing the relational links beyond those defined in the ingest process. For example, relating content by genre, cast and crew, etc. In addition, the aggregation tier provides the ability to associate content metadata with external web sources, linking the ingested, structured content identification mappings to external metadata sources, such as poster art, reviews, biographies, etc. As long as the source for this supplemental data is trusted, this provides a controlled method by which to enhance the information available in the guide, while maintaining a structured format that maps to the source of the content.

The validation role of the aggregation tier is important to ensure the guide data always directs viewers to the correctly associated content source and that any supplemental metadata is correctly associated with the content. For example, Jack Black is assigned as part of the cast for the 2005 version of King Kong and not the 1933 version. This is in addition to ensuring regulatory information is present, such as parental ratings and audio descriptions. This role is predominantly automated, based on pre-defined rules and checks, which ensure that the resulting guide metadata being offered to client applications is complete and accurate.

By centralizing this process in an SDP architecture, it makes the task of adding new data sources and mappings to new content sources much more efficient and less problematic than if it were being done within the guide application on the STB.

ADAPTATION TIER

The adaptation tier provides the bridge from the operator's TV headend into the SDP architecture. Functionally, the adaptation tier includes silos such as the following:

- Content metadata for linear, VOD;
- DVR recording information and management;
- User identity, profile and personalization;
- STB remote control;
- STB viewing history; and
- Recommendations.

The means of adaptation depend on the capabilities available in a particular operator headend. If an operator has existing components for a given functionality, the adaptation tier includes simply an adapter from the interfaces provided by the operator's existing components into an SDP. However, if the function does not exist in the operator's headend, an SDP deployment must include new components to perform that functionality. In other cases, the operator's existing components provide a basic functionality, but the SDP requires more advanced capability, so a hybrid architecture for adaptation is employed.

As an example, the adaptation tier relies on an ingest process, in much the same way as is done with legacy guides today. TV listings data for linear and VOD content is imported from known, trusted sources to ensure there is a structured data set based on unique identifiers for the content, channels, program series, etc., each mapped to supplemental information such as descriptions, cast and crew, etc. This results in a structured set of metadata that is mapped to content sources and their delivery/access parameters, which describe how the client application accesses the referenced content over the MSO's delivery plant. This ensures that the source of the content is mapped correctly to the program metadata and avoids broken links or attempting to access incompatible content formats.

The ingest process is an offline process, occurring periodically. However, the retrieval of data from the resulting database of metadata is a continuous process, with higher rates of concurrency than experienced with legacy guides, due to the greater real-time demand for data with the cloud-based architecture. For this reason, it is important to separate the ingest process from the dynamic query process to ensure that data access from client applications remain unaffected.

VIEWER PROFILE MANAGEMENT

A Viewer Profile Management component can store profile information entered by viewers through the guide application, such as preferences, parental controls, favorite channels, etc. The profile information can also be enhanced with subscriber information (billing location, subscription entitlements, etc.) to enable applications to make logical decisions about how the content metadata is to be displayed on a viewer's guide, based on their profile. Through integration of the viewer profile management with the digital rights management (DRM) domain management system, MSOs can tailor the content choices presented to viewers based on an MSO's carriage rights for the content (such as geographic blackouts or in-home/out of home viewing). Another obvious benefit of storing the viewer profiles centrally is that each viewer's preferences are accessible from whichever STB or device they are using and do not have to be managed independently per device. In addition, it provides a gateway for external applications, such as remote DVR recording and TV Everywhere services.

STB REMOTE CONTROL

One of the new features provided by guides on companion devices is the ability to control the TV experience being watched on the STB. The STB remote control component is required to adapt the commands received via an SDP API into protocols that can be passed over the operator TV network to the STB inside the home. To deploy this function to legacy STBs requires either modification or enhancement of the existing resident guide (not likely to be achieved very easily), or the creation of an unbound EBIF control application deployed on the STB. Through this mechanism the SDP can change channels, manage DVR recordings and synchronize to the channels and programs being viewed on the STB.

COMMON BRAND EXPERIENCE

Today, all MSO guides are developed as applications that reside and run on their respective devices. This means that multiple versions of an application are required (even for different STB models), each integrated and tested on their respective device. This also means that to ensure common branding and usability, the guide application feature set is based on the least capable device being supported. This means that capabilities of some devices are not used to their full advantage.

NEW UI TECHNOLOGIES

Consumer electronic devices in the home can now take advantage of new UI technologies. These technologies are more feature rich, easier to develop and perform better than the traditional technologies used to implement guides on legacy STBs. New UI technologies available today support both browser-based applications and native applications.

UI technologies that appear to be making the most traction today include the following:

- HTML5
  An open and evolving W3C standard, which, while still maturing as a standard, promises to offer rich graphical functionality and is becoming widely implemented on many CE devices via web browsers.
- Adobe Flash
  A mature, proprietary technology that is widely available on IP STBs and PCs and available as a plugin to a web browser or as a standalone environment.
- Qt
  A proprietary framework technology that enables comprehensive, rich UI implementations. Although Qt requires a greater depth of integration on client devices, the results are superior to that of browser-based rendering. Qt can also be used in conjunction with browser technologies, such as HTML5, to provide acceleration enhancements while minimizing integration effort.
- Android
  A java-based and relatively feature rich environment, which is becoming widely deployed on mobile platforms. It supports native Android applications which run on a Dalvik VM.
- iOS
  The operating system on Apple devices, such as iPad, iPhone, and iPod, that provides a simple and pleasing user experience, given the controlled UI design requirements, widget set, and UI effects. A native application environment using Objective-C.

REMOTE UI

By using an SDP architecture, the guide appearance and functionality can be managed by the edge servers in the presentation tier in the same way that the metadata is tailored for each device. By implementing remote UI technology, a design framework for the guide that defines the layout, navigation and functionality can be delivered in an appropriate format and resolution to a connected device without needing to have an integrated application on the device. Various remote user interface technologies exist today that rely on a browser or rendering application on the device, which renders the guide framework to the screen. HTML5, Flash and Qt can be used as remote UI technologies. These can also be supported by edge servers in conjunction with native applications running on other client devices, where typically a native application still provides optimum performance, such as the iPad.

Whether utilizing remote UI technologies or native applications on the various client devices, MSOs now have the opportunity to optimize the look and feel of the guide based on the device on which it is running. A common guide look and feel based on the lowest common denominator of functionality is no longer necessary, especially when the guide application will be one of many other applications on a CE device, to which it will be compared. Therefore, the guide design is evolving to become a design framework that takes into account differentiating functionality across devices, while maintaining a consistent brand and user experience to the viewer. The UI design framework must take into account a number of varying attributes, which may be implemented differently on devices, such as aspect ratio, human interface controls (remote control, touch screen, mouse/keyboard, etc.), screen rotation (portrait/landscape), resolution, etc. Support for these unique features can be built into native applications, developed for specific devices or managed by the edge servers along with the metadata filtering.

DEVELOPMENT TOOLS

A key advantage of utilizing an SDP architecture is the decoupling of the UI application on the client device from a good portion of the guide engine, which now resides in the cloud. SDP architecture also opens up the possibility of developing the application as a native application for a specific device or using a remote UI technology, which can be rendered in an optimum manner by a number of devices. This decoupling means that development of the client application no longer requires specialist knowledge of the cable infrastructure and associated complex integration. This opens up the development task to a wide resource of developers familiar with the respective devices and/or remote UI environments.

Although development tools exist for popular environments such as HTML5, Flash, Qt, Android, and iOS, certain aspects of the guide application require creating features and functions specific to the TV environment. Some of these features control the TV viewing experience, such as changing channel, managing DVR recordings, changing settings, etc. Other features provide context about what the user is watching on their TV or has watched recently. Finally, enhanced metadata about the TV programming or related to buddies in the social network complement the TV experience.

These TV-control and context features all require integration and testing with the cable system and present a level of control to the application. It is important that these functions are separated from the main UI application development as a set of pre-developed, integrated and tested APIs, which can be included as part of the SDP software developer kit (SDK) made available to developers. This lets the MSO maintain control and integration of the service, leaving the developer community to concentrate on the application development.

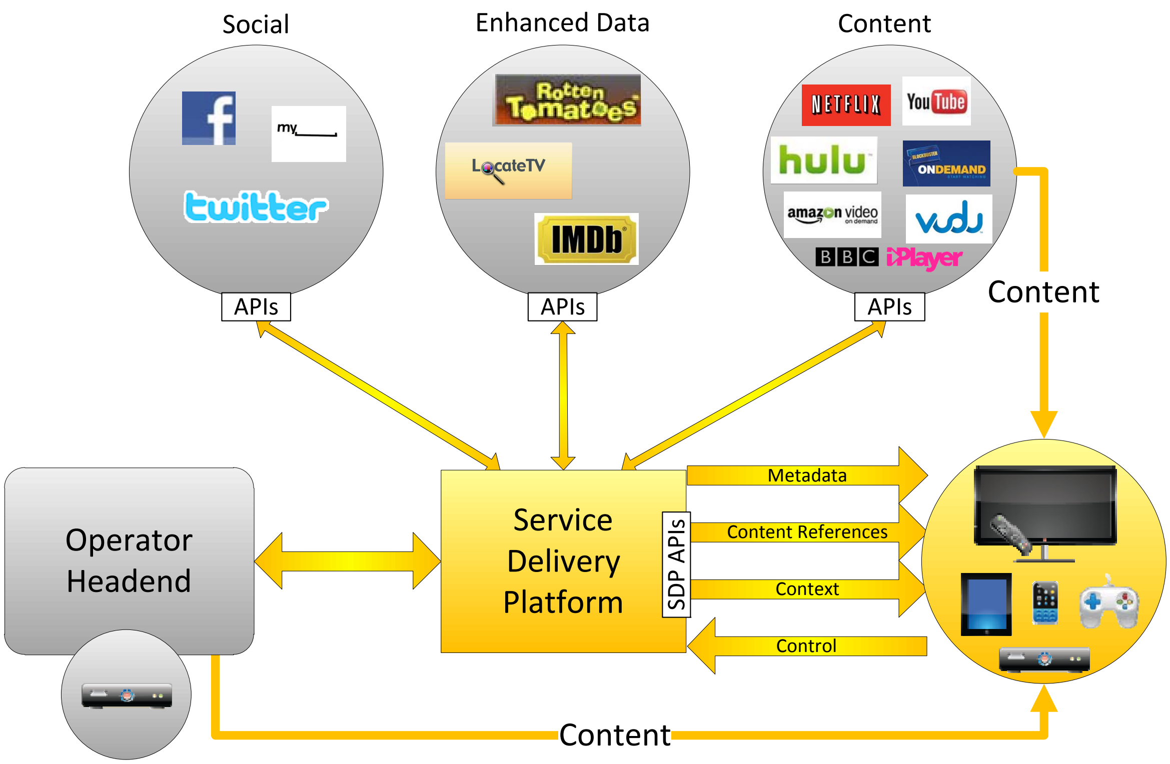


Figure 3: Service Delivery Architecture: TV-control and Context Features

The SDK APIs made available to the developer community are at the discretion of the MSO with full access controls in place, but there is obvious opportunity to standardize at least a core set of these APIs to ensure commonality of control, context and metadata features and speed of integration across platforms. For example, APIs to EBIF servers in the headend for the control and monitoring of TV viewing functions on legacy STBs from companion guide applications, such as channel changes and DVR recordings, is one such area being looked at by CableLabs®.

The SDP architecture also lends itself to convenient and controlled testing without the need for expensive lab systems. As the edge servers provide authenticated access to the aggregation tier, dedicated access can be defined on a per developer basis, which provides a filtered set of test data designed to exercise the guide application to its fullest and would not be available to regular client devices.

STANDARDS

By pushing much of the guide functionality into the cloud, the standards related to an SDP-architected guide solution are focused mainly on communication and not on hardware or implementation. The following are key standards bodies that are essential for an interoperable SDP-based guide solution:

- **uPnP** - Essential for companion device discovery on the home network.
- **DLNA** – Enables monitoring of the STB status and to send commands to control the ST for the communication between companion devices and STBs within the home.
- **HTML5** – A remote UI environment that can render the guide in an optimized format across a variety of devices. Needs to be included as the UI technology in an updated **CEA-2014**.
- **MPEG7** – For the standardization of program metadata and unique identities, by which programs can be identified across different systems and avoid re-mapping of unique IDs within the Aggregation tier.

CONCLUSION

Companion devices, such as tablets, are positioned to revolutionize the way TV is experienced by consumers. One way that operators can enable these new devices and capabilities is via a service delivery platform architecture that allows new features and functionality well beyond the capabilities of legacy guides. It can also be developed in a time frame significantly reduced from the legacy guides development and deployment cycle.

Furthermore, an SDP architecture can service DOCSIS-enable STBs to enhance the legacy guide experience by delivering the enhanced features and services over the broadband connection to the STB either for interpretation by the resident guide or through an EBIF user agent.

Additionally, building a standardized service delivery platform API and infrastructure across multiple operator systems can also create a more economically viable application development environment with the ability to deploy applications across operator networks and tie together users with different service providers.

The fast paced world of application development and social media mobility lead to challenges for MSOs to keep pace with the consumer electronics domain. Providing a controlled means of accessing platform services, protecting customer identity and confidentiality while enabling thousands of potential third party developers will ensure that an MSO's services are accessible from the widest variety of devices. The most successful applications will engage customers and ensure that the services are always relevant on the current generation of devices.

The DOCSIS Timing Protocol (DTP)
Generating Precision Timing Services from a DOCSIS System

John T. Chapman, CTO ATTG & Fellow, jchapman@cisco.com
Rakesh Chopra, Principal Engineer, rakchopr@cisco.com
Laurent Montini, Technical Leader, lmontini@cisco.com

Abstract

New market opportunities for DOCSIS include applications such as cellular backhaul of femtocell, picocell, microcell, and macrocells. These applications may require network timing in terms of time and frequency. With the deployment of IP and Ethernet based networks, PTP (IEEE 1588) and Synchronous Ethernet have become popular approaches for distributing carrier-class network timing over a network.

DOCSIS and the HFC plant present many challenges why it is difficult to propagate network timing information from the headend, through a DOCSIS network and into a CPE device with any degree of accuracy. These challenges include:

- *HFC plant asymmetry,*
- *DOCSIS asymmetry due to the upstream scheduler variability,*
- *unknown asymmetrical plant delay between CMTS and CM,*
- *unknown delay of CMTS and CM PHYs,*
- *uncalibrated ranging.*

This paper proposes a solution called DOCSIS Timing Protocol (DTP) and discusses how DTP can address these challenges in a DOCSIS system and what specification and product changes are needed to the DOCSIS CM, CMTS, and DTI Server. The resulting design can support the generation of precision timing protocols such as NTP, PTPv2 (IEEE Std1588-2008) and Synchronous Ethernet that can serve new and evolving CPE devices with traceable time and frequency synchronization requirements.

TABLE OF CONTENTS

Table of Contents 1

Introduction 2
Introducing DTP 2
Wireless Basics 2
Mobile Backhaul Synchronization 3
Radio Frequency Synchronization 4
Phase/ToD Synchronization 5
What about GPS? 6
Alternative to GNSS solutions 6

Synchronous Ethernet 7

IEEE 1588 8
Evolution History 9
Network Node Types 9
PTP Protocol Overview 11
Achieving Frequency Synchronization 13
Achieving Time Synchronization 14
Time Synchronization Error Sources 15
Improving Packet-Based Timing Accuracy 15

DTP Operation 16
System Description 16
CM Frequency Synchronization 16
CM Time Synchronization 18

Time Offset Technique 20
DOCSIS Path Latencies 20
DOCSIS Ranging 20
Measuring Round Trip Delay 21
An Example 23

Caveats 23
First Pass Approximation 24
Measuring DOCSIS Asymmetry 24
Offset Math 25
What about DPV? 26
DTI Server Recap 26
Additional Sources of Error 27
Reference CM Precise Timing Output 27
DTI Server Propagation Delays 27
Differences in CM Hardware & Software 27
Differences in CMTS Hardware & Software 27
Ranging Accuracy 27
Upstream Interleaver 28
Methodology 28
Deployment Scenarios 29
Scenario 1 30
Scenario 2 31
Scenario 3 32
Summary of Design Changes 33
CM 33
CMTS 33
DTI Server 34
DOCSIS Protocol Changes 34
DTI Protocol Changes 34
Conclusion 34
Acknowledgements 35
References 35

INTRODUCTION

Introducing DTP

The DOCSIS Timing Protocol (DTP) is the proposed name for a series of hardware, software, and protocol modifications to the DOCSIS system to support the highly accurate and traceable generation of precision timing from the CM.

The DOCSIS system in the context of DTP includes the CM, the CMTS, and the DTI Server.

Precision timing in the context of DTP refers to protocols such as Network Time Protocol (NTP), Precision Time Protocol (PTP), or hardware interfaces such as a Pulse Per Second (1PPS) output and Synchronous Ethernet.

Any specific reference to supporting a PTP output on the CM in this white paper inherently could include other protocols such as NTP and/or 1PPS-style interfaces and their associated protocols. A standard telecom 1PPS output is not defined at this time but is under study by the ITU-T.

All references to PTP and IEEE 1588 imply the latest version of PTP that is PTPv2 as defined by IEEE Std 1588-2008.

Wireless Basics

For clarification, it is useful to review the basics of wireless terminology for this paper.

Wireless includes cellular technology such as LTE, GSM/UMTS and others as well as non-cellular technologies such as Wi-Fi and DECT. This white paper targets cellular wireless technologies since they generally require highly precise time and frequency synchronization.

Each wireless technology generally has a base station that acts as a coupling point between the wireless and wired network. The classification of base station is related to its coverage and usage. Common classifications used today along with typical coverage and typical usage areas are listed on the next page. Each technology and cell site produces slightly different results.

Standard GSM macrocell range is limited to 35 km but an extended range cell may go up to 60-100 km in certain areas. Range depends on various parameters including technology, power, area or coexistence with other cells of same or distinct radio technology. The example ranges below are for WCDMA in open air (reduced coverage inside of buildings)

- Macrocell
  - WCDMA: 43 dBm/30m = 1 km
  - Rural areas or along highways
- Microcell
  - WCDMA: 33 dBm/20m = 400 m
  - Malls, hotels
- Picocell
  - WCDMA: 24 dBm/10m = 200 m
  - Transportation hub, airplane
- Femtocell
  - WCDMA: 24 dBm/1m = 71m
  - Actual coverage area is usually less due to being inside of a building.
  - Residential home

Mobile Backhaul Synchronization

In the last few years, synchronization in access networks has become an important topic because of the evolution from TDM-based to packet-based networks. In particular, the mobile wireless operators are struggling to increase their backhaul capacity that is required by the newest radio technologies in order to provide greater bandwidth and improved services.

Although the introduction of smaller capacity base stations (namely microcell, picocell and femtocell) permits mobile and broadcast operators to improve the wireless service by providing better coverage, it also demands increasing the number of network connections.

Such trends lead to optimization of the mobile backhaul infrastructure.

Packet-based networks allow the operators a cost-effective way to fulfill the necessary improvements in bandwidth, coverage and access. Today, IP over Ethernet is the most utilized transmission option for the aggregation networks.

The choices for the last mile access transport technologies include Ethernet (fixed and microwave), Passive Optical Network (PON), Digital Subscriber Line (DSL) or legacy TDM line such as T1/E1 or SONET/SDH (either fixed or microwave). Cable HFC (hybrid fiber coax) networks are being considered also as the bandwidth of DOCSIS (Data over Cable Service Interface Specification) based systems continue to increase.

The first generation of this mobile backhaul network evolution called for the support of legacy TDM circuits (e.g. for 2G GSM base stations). Replacing these circuits created a market for circuit emulation service over packet networks. Because T1/E1 requires accurate and stable clocking, circuit emulation services (or TDM pseudo-wires services) over packet networks inherited the need for frequency synchronization.

DOCSIS 1.1 introduced circuit emulation support and was able to leverage inherent DOCSIS frequency transfer such as NCR (Network Clock Recovery) via Symbol Clock Lock. Other access technologies such as SHDSL or GPON also have bit timing (physical layer) capability. But this requires equipment at both ends to be able to either receive or to retransmit clock signal. Timing distribution must then be accordingly planned.

Classic Ethernet has no such synchronous clocking capability. Adaptive

Clock Recovery (ACR) – that is, recovering frequency from a packet flow – from Circuit Emulation Services (CES) traffic was the first method developed to support TDM pseudowires [G.8261]. If such a solution was sufficient in some cases for CES application, it appeared to be sub-optimal to support base station radio interface requirement.

Radio Frequency Synchronization

Indeed, one critical aspect of mobile base stations (from GSM to LTE or WiMAX) and broadcast transmitters is their need for synchronization of their radio interface. Accurate frequency synchronization between base stations allows user handsets to seamlessly handover between base stations, reduces interference between cells and optimizes radio bandwidth capacity.

To improve timing services available from networks, particularly Ethernet based, ITU-T Question 13 in Study Group 15 took the leadership on investigating solutions and defining the appropriate specifications. Focus was first given to frequency distribution because of the CES application and 2G/3G base stations.

This focus led to adopting Synchronous

Radio Technology or Service	Cell (Base Station) Type	Frequency Read: better than…	Phase or Time Synchronization Read: less than…
GSM	Macro Pico	±50 ppb ±100 ppb	N/A
WCDMA (and LTE) FDD	WideArea Medium/LocalArea (micro/pico-cell) Home BS (femtocell)	±50 ppb ±16 ppb (OBSAI) ±100 ppb ±250 ppb	N/A
WCDMA TDD	WideArea LocalArea	±50 ppb ±100 ppb	±2.5 μs between base stations
TD-SCDMA	WideArea LocalArea	±50 ppb ±100 ppb	±3 μs between base stations
LTE TDD	WideArea LocalArea	±50 ppb ±100 ppb	±3 μs between base stations (may range from ±0.5 μs to ±50 μs)
CDMA2K	Macro Pico and Femto	±50 ppb ±100 ppb	ToD (UTC) sync *should* be less than 3 μs and *shall* be less than 10 μs
WiMAX Mobile		Up to ±1 ppb (with an average target of ±15 ppb)	Usual values between ±0.5μs and ±5 μs
MB SFN Service		±50 ppb	±1 μs
LTE-Advanced Services (CoMP, relaying function, carrier aggregation…)		Up to ±5 ppb (CoMP)	±0.5 μs [±1 μs] may be < ±0.2 μs (TBC)
DVB SFN		Up to ±1 ppb	General agreement: ±1 μs

Table 1 – Cellular Accuracy Requirements

Ethernet, a physical layer method that demands hardware changes in Ethernet equipment as well as to define recommendations to support packet-based frequency transfer with no hardware changes in packet network elements.

As mentioned earlier, frequency synchronization (also named syntonization) is a common requirement for base stations. Refer to Appendix IV of [G.8261]

Table 1 presents a summary of the main wireless applications driving the standard development for synchronization in telecom operations. These telecom network requirements are currently the tightest known and therefore provide guidance for what accuracy levels DTP should target.

Because of smaller impact of the Doppler effect, smaller cells sites can tolerate more relaxed requirements as shown in Table 1.

Phase/ToD Synchronization

Table 1 also presents another critical aspect of some mobile base stations: the need for phase or time of day (ToD) synchronization. Most of the values in Table 1 are based upon publicly available information and standard references. Some values depend on radio parameters and a few, such as CoMP, have to be confirmed by the appropriate organization. ITU-T WG15 Q13 is currently working on [G.8271] that will further describe the requirements and point to appropriate references. Similar phase or ToD requirements are also seen in other market segments such as: broadcast operators, power utilities or Smart Grid, real-time applications as for audio video bridging, or more conventionally for better network performance measurement.

As for frequency synchronization, those wireless phase/ToD requirements apply to the radio interfaces, particularly if Time Division Duplexing (TDD) is being used (e.g., WCDMA or LTE TDD). TDD is a method allowing radio interface to transmit and receive in different time slots on the same media or frequency band.

The phase/ToD synchronization requirements for radio services can be independent of the radio transmission methods. For instance, Single Frequency Network (SFN) is used for Multicast and Broadcast Services and MBS (also named Multicast and Broadcast Multimedia Services –MBMS– in 3GPP). Phase synchronization allows the simultaneous transmission of the same frame by multiple base stations or transmitters in the same SFN domain.

In most cases, the synchronization accuracy expected from the network for large cells will typically be in the sub-microsecond range.

Other applications are less demanding. We could then categorize all these applications at different levels based upon their timing performance requirements. The lowest level may not require the same network changes as the higher levels. Table 2 proposes such a performance scale.

The level of time synchronization that can be achieved over DOCSIS without DTP depends on multiple variables such as the location of timing source, the HFC plant configuration, the protocol and its setup or the receiving clocking servo. For instance, a software-based standard NTP implementation is not expected to provide sub-millisecond time accuracy.

Level of Accuracy	Typical Applications	Range of Requirements
1	Billing, Alarms	> 1 ms
2	IP Delay monitoring (range depends on network and applications)	few µs to hundreds of µs
3	Radio interfaces requirement (range depends on technology and radio configuration) Power Utilities, SmartGrid, Real–Time Audio and Video (Broadcast, AVB)	1 µs to few µs
4	Wireless services (e.g., CoMP, LBS or E911)	< 1 µs

Table 2 - Ranking Different Applications

What about GPS?

Before network-based precise timing distribution in a telecom network became a critical development topic, the only solution was to utilize over-the-air PNT (Positioning Navigation and Timing) solutions, particularly the well-known GPS (Global Positioning System). GPS receivers were embedded in the base stations (e.g., CDMA™2000) or installed at the cell site to feed the base station (e.g., WiMAX TDD or DVB).

GPS (or equivalent GNSS – Global Navigation Satellite Systems), despite being ubiquitous, have some drawbacks. For carrier-class timing purposes, such a system can be expensive because of the installation and operation costs. Indeed, the increased number of cells would just augment the number of GPS receiver installation or call for further cabling requirements in order to distribute the GPS clock signal within a building (as for picocells and femtocells).

Moreover, multiple governmental, industry or engineering organizations have pointed out the usual over-reliance on GPS for critical public services, while highlighting that. GPS signals are susceptible to threats such as jamming (intentional or not) and spoofing. Hence, when relying on a GNSS solution, proper backup mechanisms are desirable. Currently, most of the time, an expensive oscillator or an atomic clock provides the necessary stable local reference required for carrier-class timing.

For these reasons, GPS and other GNSS, cannot be considered the only alternatives anymore. Therefore stronger attention has been given to network-based timing solutions for backing up or replacing GNSS receivers.

Alternative to GNSS solutions

For instance, one alternative way to backup a GPS receiver is to provide a stable frequency reference such as a signal traceable from a PRC (Primary Reference Clock – ITU-T) or PRS (Primary Reference Source –Telcordia/ATIS) device. Such a signal can replace the expensive local oscillator that would take over in case of GPS signal failure. The last valid time information would be maintained with a physical layer PRC/PRS-traceable signal stability.

Another option would be to provide another time reference from the network, complementing or replacing the PRC/PRS-traceable frequency source. Eventually, this network timing reference may become the principal and unique time reference available to the applications.

In summary, multiple applications with different requirements may benefit from timing services from packet networks. For the most demanding applications, such as mobile wireless, the network must provide specific support leading to improvements in transmission technologies. For naturally asymmetrical access networks, specific techniques must be developed and adopted as part of timing network engineering.

The next section will present new technologies developed mainly for Ethernet networks that can be used as part of the cable operator aggregation network.

SYNCHRONOUS ETHERNET

TDM networks were designed and optimized to carry continuous rate traffic. Over time, they have been adapted to carry packetized IP traffic. Time Division Multiplexing (TDM) network deployments relied on Layer 1 frequency distribution techniques to synchronize multiple network elements together allowing for slower speed interfaces to be multiplexed together from multiple sources.

The master/slave synchronization architecture of hierarchical TDM networks allowed network providers to rely on all nodes of their network to be synchronized to a PRC/PRS. Over time they built up infrastructure and deployment models around this capability.

Ethernet, meanwhile, was designed to carry packet data across the shared medium of a local area network. Within each network element that linked isolated local area network segments, packets are buffered and retransmitted. This removes the requirements for frequency synchronization allowing each Ethernet node to run asynchronously from all other Ethernet nodes.

Owing to the shared medium nature of Ethernet, transmitters only send data when necessary. This technique frees the medium for other nodes to transmit. Because of this, the nodes are not constantly driving bits onto the wire, thus precluding the Ethernet link partner from continuously recovering the frequency of the transmitter.

Modern day Ethernet provides a non-shared medium with full-duplex options over

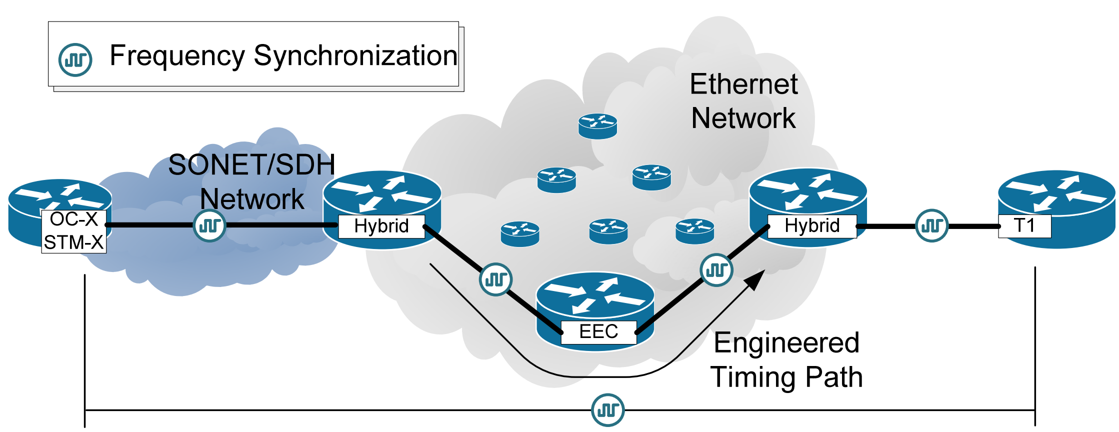


Figure 1 – Synchronous Ethernet Frequency Distribution

both fiber and copper. With these changes, the receiver can see a continuous bit stream and then can reliably recover the frequency from the transmitter allowing for Layer 1 frequency distribution.

First defined in G.8261 (2006), then complemented by ITU-T G.8261(2008), G.8262, G.8264 and a new release of G.781, Synchronous Ethernet specifies not only the method and requirements for frequency recovery and transmission but also provides standardization on advertisement of clock quality through the network.

Like all Layer 1 frequency synchronization techniques, all network elements between network segments must be capable of recovering and passing the frequency downstream. Therefore, changing a path from Ethernet to Synchronous Ethernet requires all nodes in-line to be changed to Synchronous Ethernet Equipment Clock (EEC). However, unlike traditional TDM networks where all nodes must be synchronized, only the non-Synchronous Ethernet network elements involved in the engineered timing path need to be upgraded.

Figure 1 provides a high level view of a Synchronous Ethernet network being used to frequency synchronize a SONET/SDH network to a T1 node.

With Synchronous Ethernet, network providers can replace old TDM equipment with more cost effective, higher performance, IP optimized Ethernet equipment while still enabling deployments that require frequency traceability.

For early deployment of frequency transfer, the drawback of Synchronous Ethernet is to ask for some hardware changes. Before approval of Synchronous Ethernet technology, packet-based solutions were already investigated.

IEEE 1588

IEEE 1588 standardizes the Precision Time Protocol (PTP) which is a two-way time transfer (TWTT) protocol. Another example of an earlier TWTT protocol is the IETF NTP (Network Time Protocol).

A TWTT protocol uses bi-directional traffic flow between a master or server and slave or client to exchange four timestamps, T_1, T_2, T_3 and T_4 (see Figure 2).

A slave or client clock servo will use those timestamps to synchronize as accurately as possible to the master or server clock. Refer to Appendix XII "Basic Principles of Timing over Packet Networks" of [G.8261] for further details.

While Synchronous Ethernet provides frequency traceability with Layer 1 connectivity, IEEE 1588 (like NTP) can provide time synchronization between two nodes across a packet network without mandating all intermediate nodes being replaced. Because time advances at a specific rate it is possible to also synchronize frequency with IEEE 1588 (or NTP).

IEEE 1588 also specifies system properties necessary to PTP for optimized time recovery.

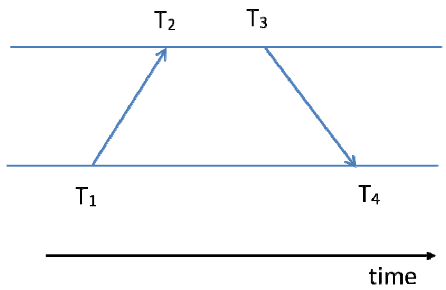


Figure 2 - Two-Way Time Transfer Principle

Evolution History

The first release of IEEE Std 1588 (PTP Version 1) was approved as a standard in 2002 and is used primarily today for industrial automation and test and measurement fixtures. The second release of IEEE Std 1588 (PTP Version 2), started in 2005 and approved in 2008, provided several key enhancements and added flexibility to the standard, enabling it to be adapted to other industries such as telecommunications.

Some of the main improvements in IEEE Standard 1588-2008 are:

- Higher packet rates for increased frequency accuracy and resiliency against packet delay variation (PDV) per G.8260
- Support for unicast transmission
- Support for redundant configurations to allow for increased fault tolerance
- Introduction of transparent clocks
- Configuration options and profiles.

PTP is quickly becoming the industry standard for highly accurate time distribution when other sources such as GPS are not available.

Network Node Types

To help define the equipment that 1588 protocol messages traverse, the standard specifies the following node types:

1. Grandmaster Clock (GM): The ultimate master of time for clock synchronization within a single PTP domain.
2. Ordinary Clock (OC): A node with a single PTP port in a domain that maintains the time used within that domain. There are two states of ordinary clocks:
   - Grandmaster Clock: A node that sources time to one or more slaves.
   - Slave Clock: A node that receives time from a master port.
3. Boundary Clock (BC): Multiple PTP ports in a single PTP domain with one slave port and at least one master port. A boundary clock can become a grandmaster clock.
4. Transparent Clock (TC): A device that modifies PTP event messages as they traverse through it. The transparent clock calculates the time the PTP event message takes to pass through the node and stores this value into the message. By doing this the node can look "transparent" and the node's contribution to PDV can be compensated for by the slave port.
5. Management Node: A device that

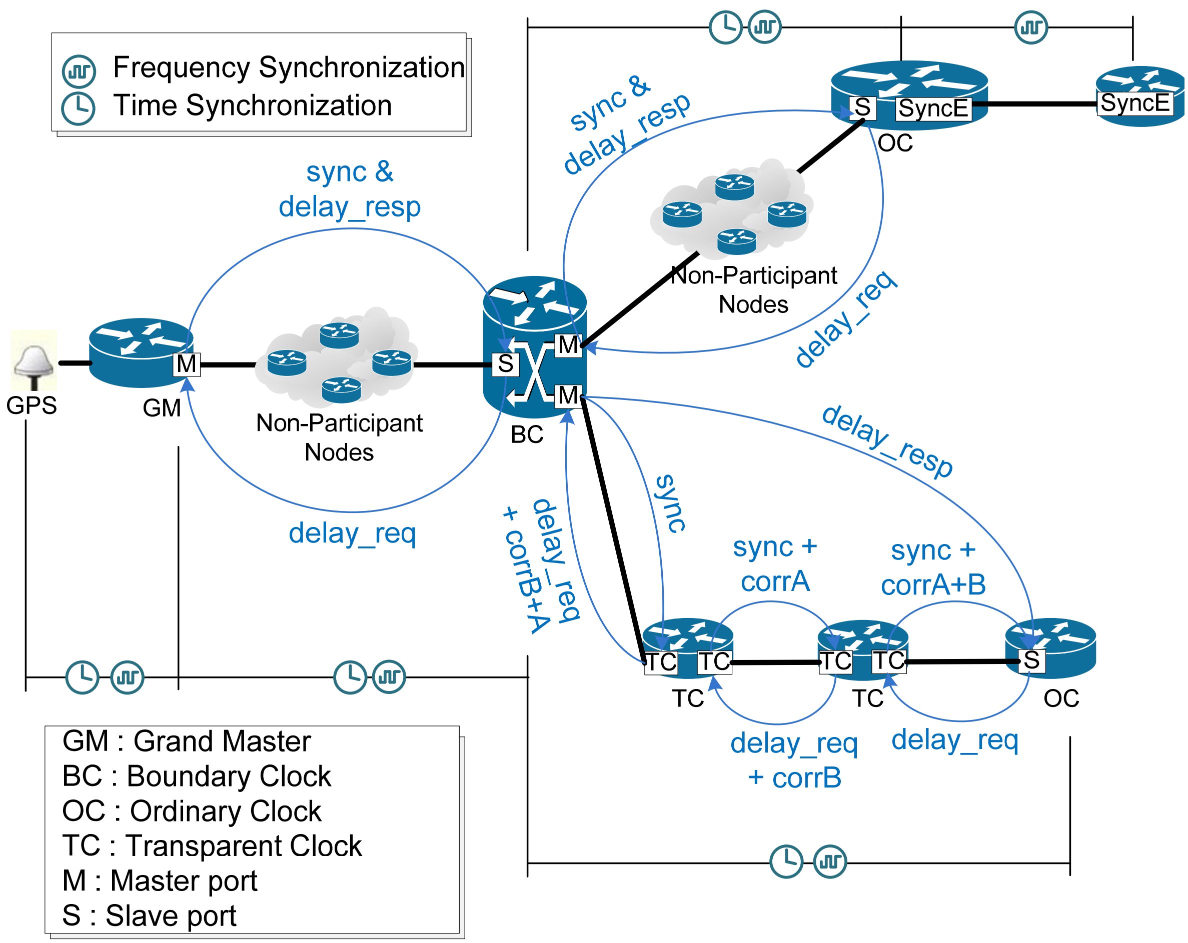


Figure 3 - 1588 Time and Frequency Synchronization

configures and monitors clocks.

Despite not being specifically defined by IEEE Std 1588, the protocol implicitly allows messages to pass through a network element that does not generate, modify, or consume PTP messages. These nodes are commonly referred to as Non-Participant nodes. These can have an impact on the recovered timing signal at a slave port due to large PDV.

Examples of these node types are illustrated in Figure 3. In the example, the grandmaster clock (GM) on the left receives its time and frequency source from a GPS receiver. It uses PTP to synchronize the slave port of the downstream boundary clock (BC) across a non-1588 aware network. The boundary clock recovers the time and frequency from PTP messages thereby roughly synchronizing it to the GM.

In addition to the slave port, the BC provides two master ports. The top master port uses PTP to allow the top OC to synchronize to its clock across another network built with non-participant nodes. The top OC recovers the time and frequency from PTP and it has a transmit reference for its Synchronous Ethernet port. The GPS and the top right Synchronous Ethernet nodes are now frequency synchronized to within a small margin of error in the short term but highly accurate and stable in the long term.

The lower master port uses PTP to synchronize the lower OC to itself. Along the path, the TC modifies the time critical events (Sync, Delay_Req) with the time the message took to pass through the TC node. The lower right OC can then recover the time and frequency from the BC and utilize the information provided by TC nodes to compensate for the PDV.

Once up and running the GM, BC, and the two OC clocks are all time and frequency synchronized to within a small margin of error. The Synchronous Ethernet node is roughly frequency synchronized to the GPS receiver.

PTP Protocol Overview

PTP defines many different message types to achieve time synchronization and node management. These are:

1. Sync, Follow_Up, Delay Request (Delay_Req) and Delay Response (Delay_Resp): These messages are utilized by PTP ports on OC and BC to synchronize time and frequency.
2. Path Delay Request (Pdelay_Req), Path Delay Response (Pdelay_Resp), Path Delay Response Follow Up (Pdelay_Resp_Follow_Up): These messages are used to measure delays between adjacent nodes when peer delay mechanism is used between master and slave ports.

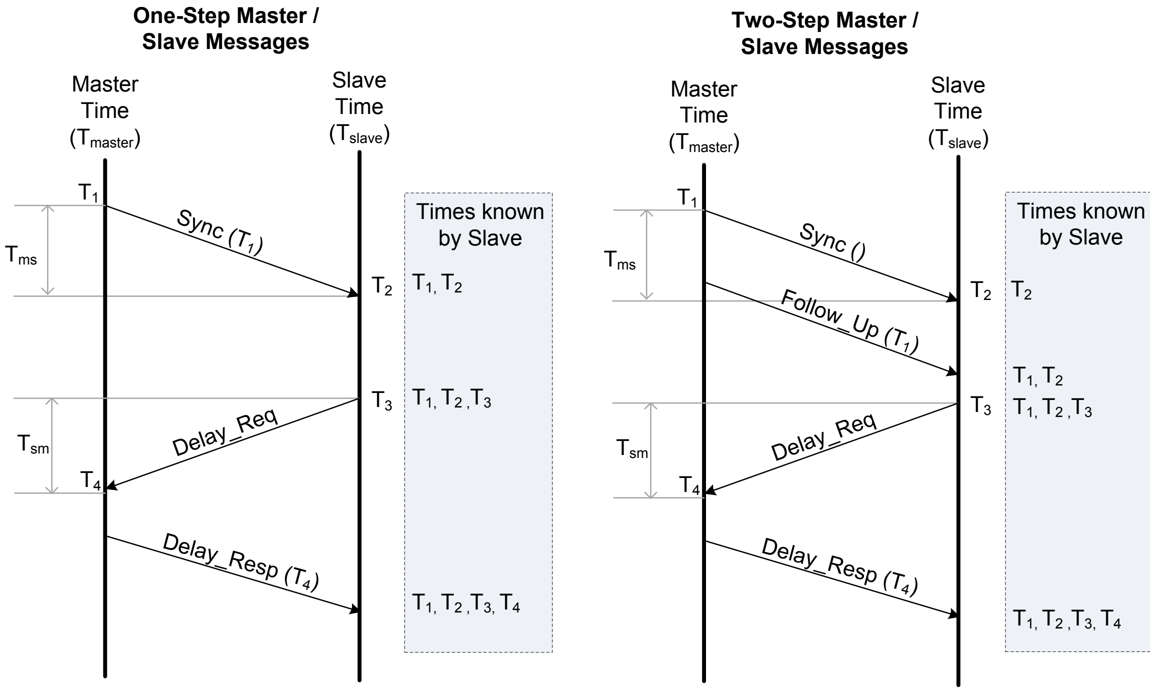


Figure 4 – PTP Message Exchange between Master and Slave

3. Announce: These messages are used as part of the clock selection algorithm.
4. Management: These messages are used to configure and monitor the PTP nodes.
5. Signaling: These messages are used for communication between PTP nodes.

We will focus primarily on the Sync, Follow_Up, Delay_Req, and Delay_Resp messages in this white paper since these messages provide the method for synchronizing time and frequency between the master and the slave nodes.

Figure 4 provides a high level logical view of the network elements and the messages used for achieving clock and frequency synchronization.

Figure 5 provides a high level view of the usage of these messages for both one-step and two-step masters.

The following provides an overview of the message usage:

1. The master port sends a Sync message containing the time T_1 when it left the master. There are two options for providing the time T_1 to the slave port.
   - Placing the timestamp T_1 into the Sync message (one-step). This requires hardware modification of the packet near the physical port to achieve a high level of synchronization.

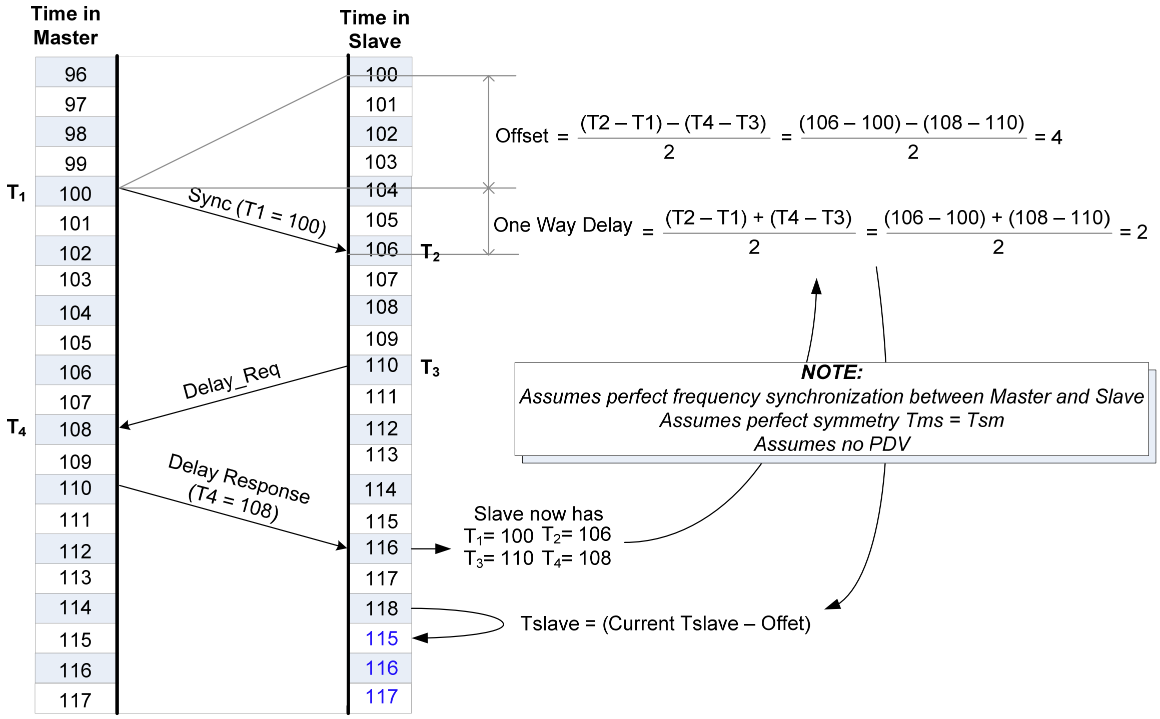


Figure 5 – Idealistic example of Slave Time Synchronization

- Placing the timestamp T_1 into a Follow_Up message (two-step). This allows the master to simply record the departure time of the Sync message. Ideally this timestamp location is near the physical port to achieve high levels of synchronization. A Follow_Up message is then generated with the Sync message departure time. This alleviates the need to do on the fly packet modification.

2. The message takes T_{ms} to travel through the network to be received at the slave.
3. The slave records the time T_2 that the Sync message arrives at its input port. If the master is a one-step master, the slave knows the time T_1 with the decoding of the Sync message. If the master is a two-step master, the slave receives the time T_1 in the Follow_Up message.
4. At this point, the slave knows the time at the master but with an unknown offset of T_{ms}.
5. To measure the delay, the slave device sources a Delay_Req message to the master. The slave records the time T_3 that the Delay_Req left the slave device.
6. The message takes T_{sm} to travel through the network to be received at the master.
7. The master records the arrival time T_4 and relays the received time back to the slave in the Delay_Resp message.

This set of message exchange provides enough information to the slave to approximately synchronize the slave to the master, as described in the following sections. However, multiple sources of error complicate the slave time recovery process.

Achieving Frequency Synchronization

Without frequency synchronization, the master and slave nodes' clocks will drift between message updates. Because these Sync messages are sent repetitively, the slave is able to calculate the drift between the master clock and slave clock with the following formula:

$$Drift = \frac{\triangle T_2 slave - \triangle T_1 slave}{\triangle T_1 slave}$$

Where ΔT_2slave is the time T_2 between multiple Sync messages being received at the slave, and ΔT_1slave is the difference in time T_1 between the same Sync messages received at the slave.

By comparing drift over time, the slave can synthesize a frequency that tracks the master frequency and keeps the time synchronized in between message updates.

Frequency recovery is only required if alternate frequency traceability methods – such as Synchronous Ethernet – do not exist and the necessary levels of time accuracy require it. If Layer 1 frequency traceability is available, it should be used since it provides a higher level of frequency accuracy and stability.

Achieving Time Synchronization

Once frequency synchronization has been achieved, time synchronization can begin. [1588 Applications] [1588 Tutorial] Before time synchronization has occurred, there is a natural offset between the master and slave devices which can be represented with two equations:

$$T_2 - (T_1 + T_{ms}) = Offset$$

$$T_4 - (T_3 + T_{sm}) = - Offset$$

Where T_{ms} is the master to slave delay and T_{sm} is the slave to master delay. Combining these two equations we get:

$$Offset = T_2 - (T_1 + T_{ms}) = -(T_4 - (T_3 + T_{sm}))$$

We now have two unknowns, T_{ms} and T_{sm}, and a single equation. To solve the equation, PTP assumes that the delays from the master to the slave and from the slave to the master are perfectly symmetrical ($T_{delay} = T_{ms} = T_{sm}$).

$$T_2 - (T_1 + T_{delay}) = -(T_4 - (T_3 + T_{delay}))$$

Solving the above equation for the one way delay (T_{delay}) we get:

$$Tdelay = \frac{(T2 - T1) + (T4 - T3)}{2}$$

By substituting in T_{delay} from above, we can solve the original offset equation:

$$Offset = T_2 - (T_1 + T_{ms})$$

$$Offset = T_2 - (T_1 + T_{delay})$$

$$Offset = T2 - \left(T1 + \frac{(T2 - T1) + (T4 - T3)}{2}\right)$$

$$Offset = \frac{(T2 - T1) - (T4 - T3)}{2}$$

The time at the slave time (T_{slave}) can then be set at some point later by adjusting the current time (Current T_{slave}) with the calculated offset:

$$T_{slave} = Current\ T_{slave} - Offset$$

Alternately, the slave time could be adjusted on the next T_2 time with:

$$T_{slave} @ T_2 = T_1 + T_{delay}$$

Time recovery can require a very complex algorithm that is affected by many real world effects such as slight variations in frequency, PDV, and network asymmetry. However in order to simplify the understanding, these real world effects are ignored in the example in Figure 5 to show how the previous math can be applied.

Time Synchronization Error Sources

There are three main sources of error for time synchronization in any TWTT protocol including PTP. These are:

Fixed Path Asymmetry

Because PTP assumes that the master to slave and slave to master paths are perfectly symmetrical, any asymmetry in the paths will result in a time offset between the master and slave nodes equal to the following basic formula:

$$Terror = \frac{Tms - Tsm}{2}$$

In the previous overly simplistic example, if $T_{ms} = 4$ and $T_{sm} = 2$ the slave would have calculated the offset as 5 instead of 4, thus leading to a time shift of 1 at the slave.

The asymmetry can arise from many sources, including but not limited to:

- network topology differences
- timestamp location differences within the master, slave, or transparent clock nodes
- node delay asymmetry through non-participant nodes.

Packet Delay Variation (PDV)

Because time synchronization relies on constant flight time between the master and slave, any variability in packet delivery in either direction will make it more difficult for the slave to accurately recover time and frequency. Each calculation of drift, offset and one-way delay will produce unique results based on the PDV in the network.

Therefore, slaves use a slave servo algorithm to integrate the results to determine the true offset and one-way delay measurements over time. Alternatively a slave servo algorithm could pre-process the time values before calculating the offset, drift, and one-way delay looking for minimum packet delays.

However as mentioned earlier this algorithm is left to the implementer and not standardized as part of IEEE 1588.

Frequency Drift Between Master and Slave

In between time updates from the slave servo algorithm, PTP time is advancing based on the slave's holdover frequency. If the frequency at the master and the slave are not perfectly synchronized the time at the slave will drift away from the master's time. The rate of drift is proportional to the frequency difference.

If the frequency on the slave is recovered from packet timing flow, as with PTP (but also true for CES or NTP), then the accuracy of the frequency recovery will be impacted by the PDV through the network.

Improving Packet-Based Timing Accuracy

IEEE 1588 transparent clocks can be used in-line between a master port and slave port to provide PDV information to the slave. This enables an increase to the maximum number of nodes between a master and a slave with the same accuracy, or to increase the accuracy of time alignment between them with the same number of nodes. Because transparent clocks record the residency time of a packet, any error in timestamp location, frequency offset or drift will have a negative impact on their performance.

IEEE 1588 boundary clocks (or NTP stratum servers) can be used to divide the PDV effects into smaller segments and to increase the scale of 1588 deployments by distributing the burden of packet generation to multiple nodes within the network. Unfortunately, because boundary clocks are susceptible to the same error contributions as a slave, they may actually have a negative impact on time alignment through the network. In between message updates, the boundary clock operates in holdover and therefore induces time and frequency error proportional to their onboard oscillator quality and slave servo algorithm.

Even with these techniques, achieving highly accurate results with IEEE 1588 requires very careful planning and implementation.

DTP OPERATION

The DOCSIS Timing Protocol (DTP) is a series of extensions to the DOCSIS protocol and the implementations of the DOCSIS CM, CMTS, and DTI Server that are intended to support protocols like PTP with a much higher degree of accuracy by leveraging the internal precision timing of the DOCSIS system.

The basic design of DTP involves synchronization of frequency and time (phase).

- Frequency is addressed by coupling the cable modem (CM) Ethernet timing to the DOCSIS downstream baud clock.
- Time is addressed by coupling the CM PTP timestamp message to the DOCSIS SYNC message timestamp.
- Time offset and asymmetry will be addressed through measurement, signaling, and ranging.

The CM would have an Ethernet output that support synchronous Ethernet [G.8261], that would have an output circuit for precision packet time stamping [802.3bf], and would support a network timing protocol such as PTP [1588].

System Description

The DTP system is shown in Figure 6. The system consists of four main components

- The CMTS. This can be an integrated CMTS (I-CMTS) or a modular CMTS (M-CMTS).
- The remote CM. This CM provides precision timing to an external entity.
- The reference CM. This is a reference CM that is identical (same manufacturer, model number, and software load) to the remote CM. It is co-located with the CMTS and is used for comparative measurements.
- The DTI Server. The DTI Server is common in M-CMTS systems. It may be external or embedded in the CMTS. In DTP, the DTI Server serves as a source of clock. It also provides measurement functions using a PTP slave port.

The additional functionality defined by DTP for the DTI Server may also be native to the CMTS.

Figure 6 also introduces various system delays that are further defined in Table 3.

CM Frequency Synchronization

DTP specifies that the CM design will synchronize the Synchronous Ethernet port to the baud clock of the downstream QAM signal. Since the jitter of the downstream DOCSIS baud clock generally exceeds the jitter requirements for Ethernet, a PLL with

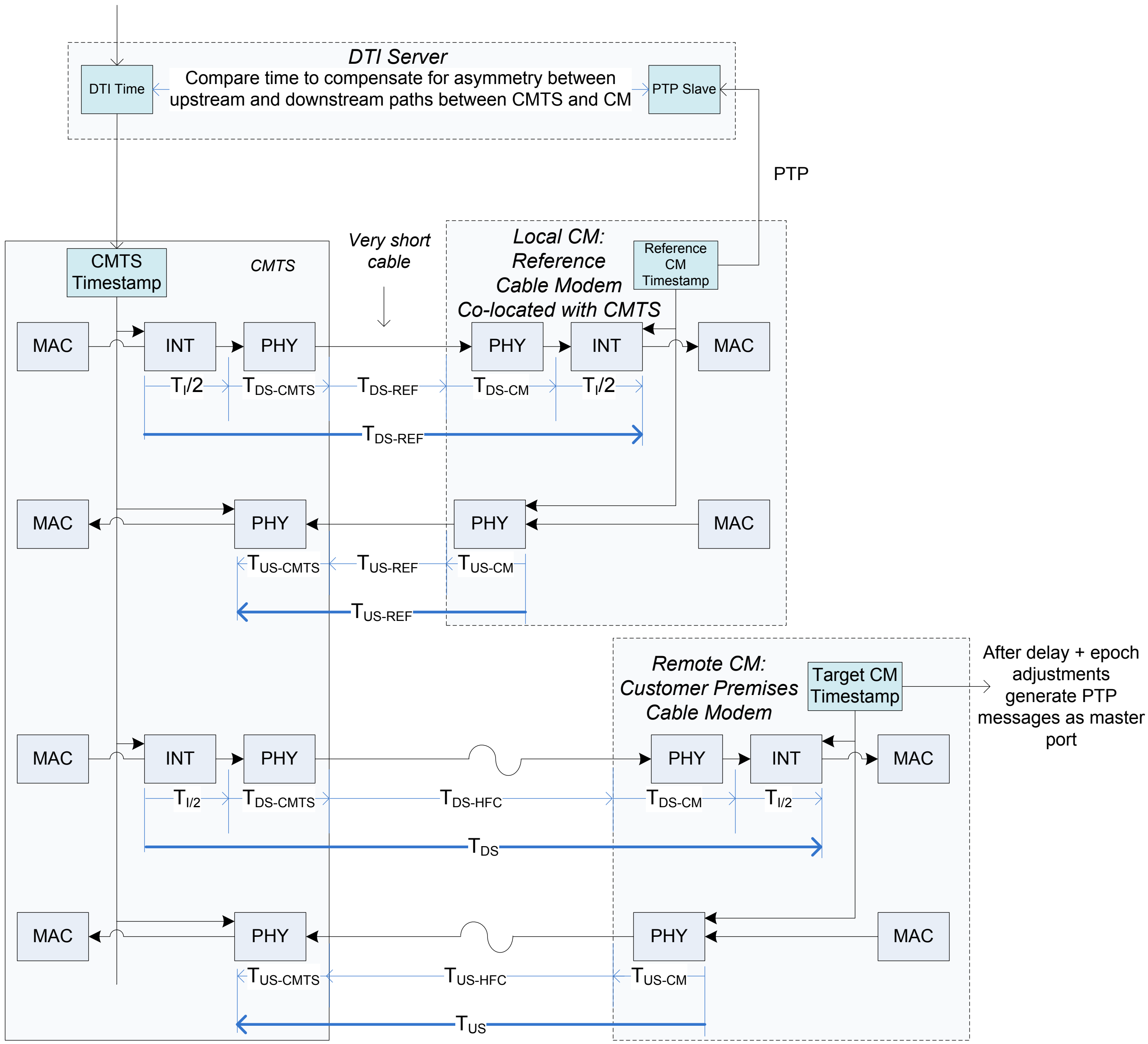


Figure 6 – DTP System Diagram

M/N frequency correction and jitter filtering will be needed.

Prior to DOCSIS 3.0, a CM in advanced time division multiple access (ATDMA) mode was not required to lock to the downstream baud clock. In TDMA or ATDMA, the CM timing was derived from entirely from the DOCSIS timestamp. In synchronous code division multiple access (SCDMA), the CM always locks to the downstream baud clock. In DOCSIS 3.0, the CMTS publishes a bit in the MAC Domain Description (MDD) message called the symbol clock-locking indicator. If this bit is set, then the CM must lock to the downstream baud rate clock.

Reference Section 6.4.28 “MAC Domain Descriptor (MDD)”, subsection 6.4.28.1.10

“Symbol Clock Locking Indicator” and section 7.1.2 “CM Synchronization” of [DOCSIS MACUP] for more detailed information.

For Synchronous Ethernet, the CMTS must set this bit and the CM must lock to the downstream baud clock for any upstream multiple access type in use, including TDMA, ATDMA, and SCDMA.

The base clock frequency of a gigabit Ethernet port is 125 MHz. The base frequency of a 100BaseT port is 25 MHz. The base frequency of the DOCSIS QAM signal is 10.24 MHz. The mathematical relationships between these clocks are:

- 10.24 MHz * 3125 / 256 = 125 MHz
- 10.24 MHz * 625 / 256 = 25 MHz

A fractional M/N PLL on the CM would implement this function.

CM Time Synchronization

Time synchronization refers to the generation of a PTP compatible timestamp at the Ethernet interface that is synchronous and phase aligned with the DTP timing source. The DTP timing source is then offset from a defined epoch. An epoch is the origin point in time of a time scale.

The DOCSIS timestamp in a stand-alone CMTS has an arbitrary value. If the CMTS is connected to a DTI Server and the DTI

Variable	Known	Comments
Tds	No	Total downstream delay from CMTS timestamp reference point to the CM timestamp reference point. This is the ultimate value that needs to be determined.
Tus	No	Total upstream delay from the CM timestamp reference point to the CMTS timestamp reference point.
Trtt	Yes	Trtt = Tds + Tus. This is a measured value.
Ti	Yes	Total interleaver delay in the downstream path. The delay is equally shared between the CMTS and CM implementation.
Tds-cmts	Yes	Delay from CMTS timestamp reference point to CMTS output. This does not include the interleaver delay.
Tds-hfc	No	Delay of the HFC plant in the downstream.
Tds-cm	Yes	Delay from the input port of the CM to the CM timestamp reference point. This does not include the interleaver delay.
Tus-cm	Yes	Delay from the CM timestamp reference point to the CM output.
Tus-hfc	No	Delay of the HFC plant in the upstream.
Tus-cmts	Yes	Delay from the CMTS input port to the CMTS timestamp reference point. This delay should take into account the difference from where the CM timestamp was inserted into the upstream packet and the reference point used by the CMTS timestamp that the CMTS US PHY attaches to the packet.
A	Yes	An assigned variable that expresses the upstream to downstream asymmetry. This does not include the downstream interleaver delay or the upstream queuing delay or scheduler uncertainty. Asymmetry may come from differences in propagation delay at different frequencies and if there are any differences in path length between the downstream and upstream paths.

Table 3 – System Delay Definitions

Server is connected to GPS, then the DTI Server can align the DOCSIS timestamp with the GPS Epoch. The GPS epoch is January 6, 1980.

PTP references time as the number of nanoseconds after the epoch of the beginning of the day of Jan 1, 1972. PTP can also have arbitrary epochs.

In practice, the DTP approach is to first compensate for the time offset between the DOCSIS timestamp at the CM and the DOCSIS timestamp at the CMTS. Then, the DOCSIS timestamp at the CM is transformed into a PTP timestamp.

There are three tasks to be accomplished at the CM:

1. The least significant bits of the PTP timestamp are derived from the DOCSIS timestamp (and potentially the fractional timestamp extension).
2. The most significant bits of the PTP timestamp are derived from a signaling message
3. The offset that represents the delay from the CMTS to the CM is measured, calculated, signaled and then applied to the timestamp.

The PTP timestamp is defined in [1588] as seconds and nanoseconds from the original chosen epoch (which can be PTP or Arbitrary). The first field is the seconds field and is 48 bits long. The second field is the nanoseconds field and is 32 bits long. The nanosecond field never exceeds 10^9. This means that the PTP timestamp is referenced to 1 ns with a possible further resolution down to 15 femtoseconds by using a correction field (NTP has a 232 picoseconds resolution).

The DOCSIS timestamp is defined in [DOCSIS DRFI] as a 32-bit binary counter that is clocked with the CMTS 10.24 MHz master clock. This means that the DOCSIS timestamp is referenced to 97.65624 ns.

Because SCDMA allows for up to 128 CMs to transmit in the same timeslot with 128 orthogonal codes, the CMs must be aligned to within a fraction of a timeslot to avoid packet corruption. To accomplish this, an additional 8-bit fractional field advertised with a TLV is used providing a resolution of 1/16384 or 0.3814 ns.

Usage of the fractional field enables higher resolution time to be represented in SCDMA. PTP has the ability to express timing accuracy in fractional nanosecond resolution. It may be useful to include this fractional time field in the DOCSIS protocol or DTP extensions to increase PTP accuracy, even when using ATDMA.

Note that within the CM electronics, the DOCSIS timestamp is in a different clock domain than where the PTP timestamp is generated. The timestamp value must be transferred across the CM internal boundary in a consistent manner across multiple implementations.

Due to the limited size of the DOCSIS timestamp, it rolls over to zero approximately every 7 minutes. This means that the upper bits of the PTP timestamp should be sent more frequently than 7 minutes and that the CM mechanism must deal with the rollover when attaching the upper bits.

There are various ways of construction signaling. In one method, the various system offsets are measured and collected by the CMTS. The CMTS then sends a final correction value to the CM. In another method, the CMTS would publish any offsets it has, the CM would measure its internal offsets and then perform the final math.

The derivation of this offset is described in the next section.

TIME OFFSET TECHNIQUE

DOCSIS Path Latencies

The round trip DOCSIS path delay is inherently asymmetrical and can contain a moderate to high amount of jitter. Asymmetry and jitter introduce error into any timing protocol that might traverse the DOCSIS network. DTP mitigates these two factors by modifying the DOCSIS hardware design and deriving timing information directly from the DOCSIS system at the CM for use by NTP/PTP.

The packet transport delay in the downstream path of DOCSIS is relatively stable. It has a variety of fixed delays in the equipment and some variable propagation delay on the plant depending upon wind and temperature. The downstream interleaver delay is a programmable value and for DOCSIS is typically 0.68 ms (for 256-QAM). The length of the DOCSIS plant can be from zero to 100 miles. As a result, the one-way transit delay of the HFC plant can be up to 800 µsec. The PHY delays are unknown. The actual time that a bit passes through the external RF interface is indeterminate. Transit time will also depend upon other configuration parameters such as modulation order and FEC type.

It is not necessary to launch a separate signaling message with a timestamp in it in the downstream. DOCSIS already has a SYNC MAC Management containing a timestamp that is delivered from the CMTS to the CM with less than 500 ns of jitter, as specified in Section 6.3.9 of [DOCSIS DRFI]. The SYNC message bypasses the downstream output queues and their associated jitter and latency. The CM synchronizes itself to the timestamp in the SYNC message. The DOCSIS system will have to ensure that the jitter from the DOCSIS timestamp is sufficiently filtered so that it does not contribute error to the PTP timestamp.

The upstream DOCSIS path has much more uncertainty. To send a packet upstream, the CM must launch a request packet in a contention slot. If that fails, it keeps trying at longer and longer time intervals until it gets through. The CMTS then schedules a data transmission slot and issues a MAP MAC Management message. MAP messages tell CMs when to start and stop upstream transmissions, and what modulation profile to use.

This mechanism is actually quite efficient, but is not very predictable. This is one reason why timing protocols that are run over the top layer may see large variation in their results.

There are other scheduling techniques in DOCSIS, such as unsolicited grant service (UGS) or real time polling (RTP) services that can make the transmission opportunities more predictable. However, due to the natural jitter (on the order of 1 ms) in CMTS scheduling resolution, these alternative scheduling techniques do not provide enough accuracy to provide precision timing.

DTP relies on the DOCSIS system to take a series of measurements and then to supply the appropriate correction factor to the CM timestamp to arrive at the PTP timestamp.

DOCSIS Ranging

DTP relies partly on the DOCSIS ranging mechanism, so it is important to describe how that works.

It is important to realize that the CM receives the CMTS timestamp and then uses it directly as the CM timestamp. Thus, the CM timestamp is delayed with respect to the CMTS timestamp. In fact, the entire delay chain of the downstream, including portions of the CMTS, the HFC plant and the CM, contributes to the delay of the CM timestamp.

If the CM used this timestamp to transmit an upstream packet, that packet would arrive late at the CMTS. In fact, it would arrive late by an amount approximately equal to the entire downstream delay and the entire upstream delay.

To solve this problem, the CMTS has a two-part process. The first part is known as initial ranging and the second and re-occurring part is periodic maintenance.

The CMTS sets up a large (usually 2 ms) upstream Initial Ranging window. This window is contention based and any unregistered CM can attempt to register.

The CM sends a ranging request packet. The CMTS measures the arrival time error and sends back the error in a ranging response message. This continues until the system is working with specification limits.

The process is then repeated every 30 seconds or less. This is called periodic maintenance and is unicast in nature since the address of the CM is now known.

The net result is that the CM will calculate a ranging offset. It will then subtract this ranging offset from its timestamp to figure out when it really needs to transmit a packet.

There are two specific characteristics to note. The first is that the CMTS is not formally told what the ranging value is. The second is the ranging value held in the internal CM register can be unique to each implementation. There are many delay elements in the CM upstream design. The ranging offset is just one of them.

To recap, when the CM is told to transmit a packet to the CMTS at a particular time, it must send it earlier. To figure out how much earlier, the CM goes through a ranging process with the CMTS to create a ranging offset. It uses this ranging offset to transmit a packet earlier than the timestamp indicated in the DOCSIS MAP message. If ranging was done correctly, then the packet will arrive at the CMTS in the correct timeslot that it is supposed to.

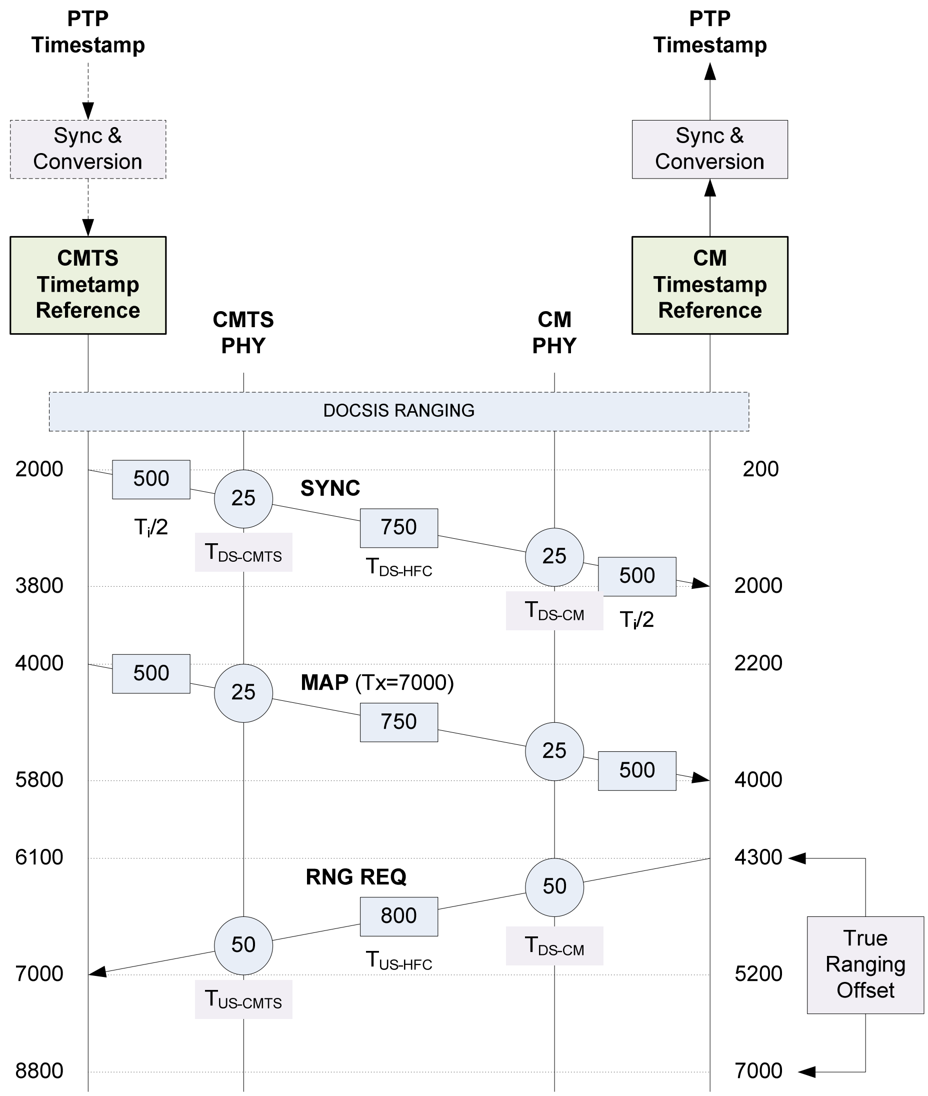
Measuring Round Trip Delay

The first DTP system measurement is a round trip delay. There are several different ways that this could be done. One way is to leverage the ranging process.

Since the actual ranging offset used in a CM is implementation specific, DTP makes a measurement. That measurement in DTP is called the true ranging offset (TRO).

That measurement is taken between the two reference points that matter to DTP – the CMTS timestamp, as referenced in the MAP message, and the CM timestamp.

The TRO of the CM in DTP is defined as the difference between the time the first bit of a packet is transmitted in the upstream from the CM in terms of the CM timestamp and the time the first bit of the packet is expected to arrive at the CMTS.



Where:
- All values are in arbitrary time units for sake of example.
- Upstream HFC delay is set to slightly more than downstream HFC delay
- DOCSIS ranging process determines internal offset for upstream tx time.

Thus:
Round Trip Delay (calculated) = 500 + 25 + 750 + 25 + 500 + 50 + 800 + 50 = 2700
True Ranging Offset (measured) = 7000 – 4300 = 2700

Formula Results:

Actual Offset Needed:	PTP Offset = 500 + 25 + 750 + 25 + 500 = 1800
Formula (1) approximation:	PTP Offset = (2700 – 1000) / 2 + 1000 = 1850
Formula (9) with Tus-off = 0:	PTP Offset = 1050 +(2700 – 1150) / 2 = 1825
Formula (9) with Tus-off = 50:	PTP Offset = 1050 +(2700 – 1150 - 50) / 2 = 1800

Figure 7 – True Ranging Offset

The expected packet arrival time at the CMTS is listed in the MAP message. So, the CM has to store that value, capture the local CM timestamp value when the correct packet transmits, and subtract the two.

In essence, *the total round trip delay is equal to the true ranging offset of the CM.*

The ranging offset of the CM is intended to correct for the round trip delays in the DOCSIS system. Thus, it should be possible to measure the operating state of the CM and reverse engineer what those network delay values are.

An Example

An example of this process is shown in Figure 7. Somewhat arbitrary values are used for illustration purposes only. The PHY delays were intentionally set different, and the upstream was given more delay than the downstream.

This example shows how the CMTS can either receive network timing or be self-contained for timing. The CM timestamp is synchronized to the DTP timing source and converted to a PTP timestamp.

DOCSIS Ranging occurs. By a process of trial and measurement, the CM arrives at a ranging offset that works for its particular implementation.

When the SYNC message traverses from the CMTS to the CM, the CM uses the value of the timestamp it receives. In this example, a value of 2000 was sent and received. Due to the downstream delay, when the CM timestamp is at 2000, the CMTS timestamp has already advanced to 3800.

Next a MAP is received that tells the CM to transmit a packet at the time of 7000. Using its ranging offset, the CM launches this packet when the CM timestamp is 4300. Because of the ranging process, the CMTS receives the packet at the time of 7000.

The true ranging offset can be measured after the ranging process is complete by taking the difference of the timestamp in the MAP and the timestamp in the CM corresponding to the start time of upstream transmission in the MAP. Note that the true ranging offset will generally have a fixed offset from the actual ranging offset used in the CM.

It can be seen that the ranging circuitry of the CM picked an actual ranging offset that caused the true ranging offset to be equal to the round trip delay.

Caveats

The true ranging value is a offset that can be measured by the CM and reported to the CMTS. Note that this value may not exactly equal the actual ranging value used in current implementations, since there are other circuit delays involved in the use of this value.

Any portion of the round trip that is outside of the measurement path cannot be included in the measurement. However, if it can be defined, a correction factor can be applied. For example, if there is a delay in the CMTS between the receive timestamp and the transmit timestamp, the CMTS will have to provide a correction factor.

In theory, the true ranging offset could change with every ranging interval. Ranging intervals occur every 25 to 30 seconds. Such changes can occur if the delay of the plant increases due to temperature shifts (reference Appendix VIII of [DOCSIS 2.0]. As a result, the CM may choose to time average the true ranging offset a finite amount of time to remove this uncertainty. Too long a period of time should be avoided as it would impact the CMs ability to react to network changes that would then impact the value of the PTP timestamp at the CM.

The timestamp is also being constantly updated with the SYNC message at least every 200 ms. Even though the CMTS and CM are frequency locked through the downstream PHY, if there was enough of a change in the delay of the downstream path, the timestamp value at the CM would be adjusted over time to the new value from the CMTS. This will also impact the PTP timestamp value.

The true ranging offset can only changed during periodic ranging. The natural packet to use for upstream measurements is then the DOCSIS MAC Ranging Request message.

Alternatively, other approaches could be used that focused on a different MAC management message. For best results, the upstream packet should be contained within a single carrier (no bonding). Fragmentation, concatenation, and CCF (continuous concatenation and fragmentation) should be disabled. These requirements are met with the Ranging Response message.

Sometimes, a DOCSIS system uses a different upstream PHY profile for different upstream operations (ranging vs. data for example) requests. A different modulation profile could result in a different upstream path delay. This may be okay for this particular application. Further analysis is needed. However, if the upstream delay is needed for other applications, this mechanism may need a correction factor, a different upstream message with which the measurement is made, or a ranging packet with the same PHY profile as the upstream data path.

First Pass Approximation

At this point, the round trip delay is now known and an approximation can be made of the time offset needed for PTP. However, there is still some information missing. The delay through the PHY circuitry at the CMTS and CM transmit and receive is not known. Further, the asymmetry of the downstream path and upstream path is not known.

The approximation would be to subtract out the downstream interleaver delay, assume all four PHY delays are symmetrical, and that the remaining downstream and upstream DOCSIS paths are symmetrical. Then divide the measured path by 2.

$$\textit{Approximate Offset} = (T_{rtt} - T_i) / 2 + T_i \quad (1)$$

But, what if this is not accurate enough? If we could determine the total asymmetry of the DOCSIS path, then a more accurate offset could be calculated. The next step is to derive the one-way delay in the downstream.

Measuring DOCSIS Asymmetry

Even better accuracy can be achieved if the CMTS has access to a reference CM that is identical in build (same manufacturer and same model) to the CM in the field. It can then compare measurements on the reference CM to the remote CM.

If there is more than one type of CM deployed that shall provide precise time downstream, there may have to be more than one local reference CM. If there is more than one type of CMTS line card, then there may have to be a duplicate reference CM on each unique CMTS line card.

The CMTS will program the reference CM with the same PHY configuration as the remote CM. The same software should be loaded as well (although that is generally not under the control of the CMTS). The DOCSIS system then performs two measurements.

1. It makes a round trip measurement.
2. It makes a one-way downstream path delay measurement.

The downstream path delay measurement is made by connecting the PTP or 1PPS output port of the reference CM into a PTP slave input port or a 1PPS input port on the DTI Server or on the CMTS.

If the external DTI Server is used, then it measures the delta between the PTP timestamp and the DOCSIS timestamp and reports it over the DTI interface to the CMTS.

If the reference CM has the right offset, then the timestamp delta will be zero (PTP timestamp is converted to a DOCSIS timestamp to perform the math) and the total downstream delay will be represented by the PTP offset used by reference CM.

One approach is to adjust the PTP timestamp offset of the local CM until the error between the local CM output and the CMTS timestamp, as measured externally, is minimal.

Offset Math

There are several ways to put the numbers together. Further, the calculations could be done at the CMTS or CM that will impact the approach slightly. Here is one basic method.

The system diagram for this example is in Figure 6. The definition of the variables is in Table 3.

The downstream delay for the reference CM is a measured value with a near-zero (and therefore ignored) HFC plant path length, and is defined as follows:

$$Tds\text{-}ref = Ti + Tds\text{-}cmts + Tds\text{-}cm \quad (2)$$

The downstream delay for the remote CM differs by the path length of the HFC plant downstream.

$$Tds = Tds\text{-}ref + Tds\text{-}hfc \quad (3)$$

The round trip time for the reference CM is a measured value with a near-zero HFC plant path length, and is defined as follows:

$$Trtt\text{-}ref = Ti + Tds\text{-}cmts + Tds\text{-}cm + Tus\text{-}cm + Tus\text{-}cmts \quad (4)$$

The round trip time for the target CM differs by the path length of the HFC plant downstream.

$$Trtt = Trtt\text{-}ref + Tds\text{-}hfc + Tus\text{-}hfc \quad (5)$$

Let’s assign linear correction factor to the HFC plant asymmetry called Tus-off. Tus-off expresses the additional amount of the upstream delay when compared to the downstream delay. Tus-off would be assigned based upon the operator’s knowledge and characterization of the plant. For example, Tus-off could account for group delay differences between the DOCSIS downstream and upstream carrier frequencies. Note that Tus-off does not represent any asymmetry within the hardware of the CMTS or the CM itself since the reference cable modem removes that asymmetry.

For example, if Tus-off = 50 ns, then the upstream path would have 50 ns more latency than the downstream path.

$$Tus\text{-}off = Tus\text{-}hfc - Tds\text{-}hfc \quad (6)$$

$$Tus\text{-}hfc = Tds\text{-}hfc + Tus\text{-}off \quad (7)$$

Applying equation (7) to (5) and solving for the downstream hfc delay,

$$Trtt = Trtt\text{-}ref + Tds\text{-}hfc + Tds\text{-}hfc + Tus\text{-}off$$

$$Tds\text{-}hfc = (Trtt – Trtt\text{-}ref – Tus\text{-}off) / 2 \quad (8)$$

Applying equation (8) to (3) yields the final equation for the offset of the downstream timestamp.

$$Tds = Tds\text{-}ref + (Trtt - Trtt\text{-}ref - Tus\text{-}off) / 2 \quad \dots (9)$$

Applying formula (9) to the example in Figure 7 with no HFC correction factor (Tus-off = 0 ns) and where the plant length is 0 for the reference CM yields:

Tds = 1050 + (2700 – 1150) / 2 = 1825

and correcting for HFC plant asymmetry using the value from the example for Tus-off = 800 – 750 = 50,

Tds = 1050 + (2700 – 1150 – 50) / 2 = 1800

As an alternative calculation, the asymmetry could be expressed as a ratio of Tds-hfc and Tus-hfc.

What about DPV?

DOCSIS 3.0 has a MAC management message called DOCSIS Path Verify (DPV). DPV allows the beginning and ending timestamp in each direction of the link to be captured and analyzed by the CMTS.

DPV has similar goals but less accuracy than the measurement technique discussed in DTP. DPV packets will see queuing delays in the upstream path where as the current DTP proposal does not. This is because DPV is using a timestamp (reference point U_1) generated prior to queuing. Refer to Section 10.5.2 “DPV Reference Points” in [DOCSIS MACUP].

As an alternative implementation of DTP, DPV could be improved if the timestamps were provided directly by a hardware mechanism after packet queuing and just prior to transmission (theoretical reference point U_1’).

Measuring the true ranging offset may be an easier implementation for a CM design than modifying an upstream DPV packet.

DTI Server Recap

Here is a recap of the system requirements from the point of view of the DTI Server.

The DTI existing functionality supports the generation, maintenance and distribution of precision time. The DTI server function shall be extended to support the precision PTP monitoring function. The PTP monitoring function is that portion of the DTI server that measures any timing offset between the reference CM and the CMTS.

The precision PTP monitoring function includes the following elements:

1. The PTP monitoring function provisioning is controlled externally. The control function will reside in the CMTS that also manages the pool of co-located reference CMs.
2. The PTP monitoring function supports establishment of monitoring sessions. In a monitoring session the DTI server shall be operated as an ordinary client in the 1588 protocol exchange.
3. The PTP monitoring function shall collect timestamp data for each session and extract an estimate of time alignment error with respect the DTI server precise 1PPS reference.
4. The PTP monitoring function shall perform the time error estimation task based on a schedule. The schedule of the start and duration of each measuring session is provided by the external control function.

5. The DTI PTP monitoring function shall support a minimum of eight simultaneous sessions.
6. The DTI PTP monitoring function may support a calibration function to mitigate asymmetry in the Ethernet connection between the co-located reference cable modem and the DTI server. In this mode one physical Ethernet port on the DTI server shall operate as a master and the Ethernet cable that normally terminates on the reference CM will be temporarily connected to the calibration master port. The calibration session control shall be supported in the DTI server using existing SNMP and CLI user controls.
7. The PTP monitoring function required a minimum of two physical Ethernet ports to support the calibration function. Optionally, supporting one port per simultaneous session will provide the highest achievable measurement accuracy.

ADDITIONAL SOURCES OF ERROR

There are a series of minor errors that either can be ignored or compensated for.

Reference CM Precise Timing Output

There is a delay from the reference CM Ethernet Master Clock output to the adjacent DTI Server Slave Clock input. Using the PTP delay_req and delay_resp messages, the symmetrical delay between the reference CM and the DTI server can be automatically removed by PTP. If using a 1PPS instead of PTP to connect the reference CM to the DTI Server the cable length could be pre-configured by the user to remove the delay.

DTI Server Propagation Delays

The DTI Server must make a difference measurement between the Reference CM and the CMTS Timestamp. If there are any offsets in the DTI circuit, it must compensate for them.

The DTI Server could be separate or embedded in the CMTS. Alternatively, the CMTS could host a PTP or 1PPS input and do the delta measurement with its own timestamp, even if there is an external DTI server.

Differences in CM Hardware & Software

To minimize measurement error, the two CMs under measure should be identical models from the same manufacturer and be running the same firmware and configuration. Under certain circumstances, it may be necessary to have a common manufacturing lot number to ensure the same performance in items such as tuners or LFE which are external to the CM silicon.

Differences in CMTS Hardware & Software

In the ideal case, the same CMTS linecard should be used for the remote CM and the reference CM. If the reference CM is on a different line card, then any differences between the two line cards can contribute to error.

This may not be practical if there are many remote modems all running PTP. At the very least, there should be one local reference CM per type of CMTS line card used.

Ranging Accuracy

A CM is ranged to a particular degree of accuracy. This should be analyzed to see if the residual ranging error, if any, would impact the accuracy of the PTP timestamp.

Upstream Interleaver

There is an interleaver that operates at the packet or burst level in the upstream direction [DOCSIS DRFI]. If the interleaver is implemented prior to packet queuing at the CM and after the packet queuing at the CMTS, then it is outside the transmission path and can be ignored. If it is created in real time and is part of the transmission path, then it could be included in asymmetry calculations.

METHODOLOGY

DOCSIS has the same goal as PON and DSL in that they all provide an access network technology that can be used for the last mile connectivity to end user. All of these last mile technologies use aggregation networks that tend to be based on IEEE 802.3 Ethernet technology.

PON, DSL links, and DOCSIS over cable plants introduce large PDV and asymmetry that are detrimental to packet-based timing distribution. This issue has been recognized and in the ITU-T, relevant Questions 2 and 4 in Working Group 15 have worked on developing their own time distribution mechanism.

By using a method targeted at the access network technology, the time transfer can be optimized, leveraging information specific to the access technology. As such, PDV and asymmetry can be reduced or compensated for, allowing better time transfer than simple transmission of PTP messages over the access network.

Such specific time distribution methods cannot be used over other media. For example, it is very difficult to provide precise timing with PTP over the top of a DOCSIS network to Ethernet based equipment located beyond a cable modem. Moreover, if the timing source is not co-located with the line termination equipment, such as a CMTS, the time shall be transferred to the DTI server or to the CMTS by other means.

At both ends of the access network, at least one other mechanism such as IEEE 1588 PTP or NTP would have to be used with some adaptation layer to transfer the time between the two network sections. If the continuity of the time signal can be maintained, the traceability of the time from the source may be broken without a specific adaptation function. This paper introduces such solution, named DOCSIS Timing Protocol (DTP).

In addition, the access network, that is, the pair PON OLT/ONU, DSLAM/DSL modem or CMTS/CM, through the utilization of their specific time distribution (e.g., refer to [GPON], [XG-PON] or clause 13 of [802.1AS] for EPON; for other access networks such as VDSL2, work is still in progress), may assist other packet-based timing protocol such as NTP or PTP.

In cases when timing packets must pass through the access link, the access network devices may simulate or act as virtual or distributed IEEE 1588 boundary clock, NTP stratum server or IEEE 1588 transparent clock. In such a design model, the PTP or NTP communication path would not be broken but the access network would provide correction for the PDV and asymmetry.

Hence multiple design scenarios based on DTP can be evaluated.

DEPLOYMENT SCENARIOS

Some customers, such as residential customers (e.g., for home femtocell), may just need a high quality timing source without being concerned about its traceability.

Other customers, such as power utility companies, mobile or broadcast operators may need information about the timing source in order to correlate time between distinct end devices that must be synchronized. This level of information may be requested for traceability or accountability.

Moreover, the operator providing a timing service to its customers may also have its own constraints that drive them to distinct timing distribution architecture.

The following figures depict three main deployment scenarios with some variants (a, b…). These are summarized in Table 4.

Those scenarios provide distinct timing infrastructures from the end user's and/or operator's viewpoint and offer flexibility to the timing infrastructure. However, each scenario may impose its specific

#	Description
1	Timing source at DTI Server/CMTS location
1a	CM acts as IEEE 1588 grandmaster (or NTP server)
1b	CM acts as IEEE 1588 grandmaster faking the DTI Server/CMTS
1c	DTI Server/CMTS acts as IEEE 1588 grandmaster (or NTP Stratum 1 server) CM acts as IEEE 1588 boundary clock (or NTP stratum server)
2	Timing source and IEEE 1588 grandmaster (or NTP server) is upwards the DTI Server/CMTS location
2a	DTI Server/CMTS fakes the IEEE 1588 grandmaster (or NTP Stratum 1 server)
2b	DTI Server/CMTS/CM acts as a distributed IEEE 1588 boundary clock (or NTP servers)
2c	DTI Server/CMTS and CM are virtual IEEE 1588 boundary clocks (or NTP servers)
2d	CMTS and CM are distributed or virtual IEEE 1588 boundary clocks (or NTP servers); DTI server is removed from timing communication path
3	CMTS/CM acts as a distributed IEEE 1588 transparent clock

Table 4 – Possible Scenarios of Timing Transfer

requirements and configurations at the CMTS and CM.

The nomenclature used for the following diagrams is shown in Figure 8:

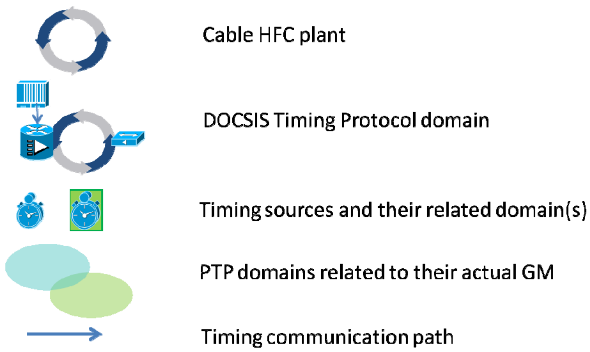


Figure 8 - Scenario Nomenclature

Scenario 1

Scenario 1a represents the baseline configuration.

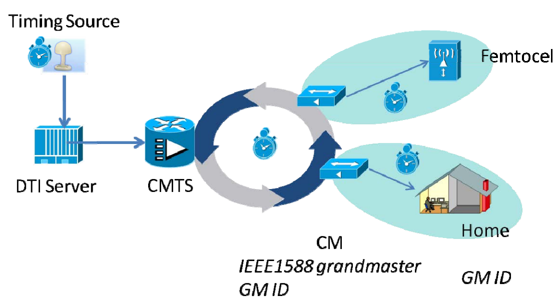


Figure 9 - Scenario 1a

The CMTS is provided precise timing by the DTI server. CM acts as a IEEE 1588 grandmaster (GM) using timing sourced by the DTI server/CMTS via the DTP method. The DTI server/CMTS shall also provide some timing source information so that the CM's GM function can populate the PTP dataset members transmitted to the downstream PTP slaves. The CM would provide its own clockID for grandmaster identifier (GM ID).

Utilizing the same DTP distribution variants 1b and 1c allow the PTP clocks beyond CMs to trace the real-timing source and identify the DTI server/CMTS as GM. In such scenarios, the CMs would have to use a DTI server/CMTS PTP clockID for the GM ID.

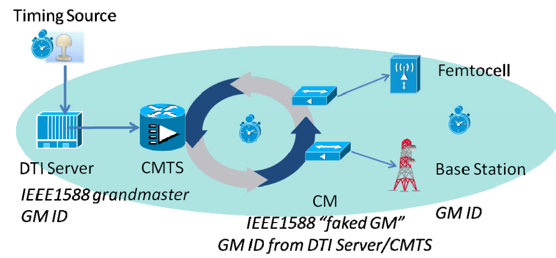


Figure 10 - Scenario 1b

In scenario 1b, the CM remains the real IEEE 1588 GM but replaces its own clockID by the DTI server/CMTS ClockID.

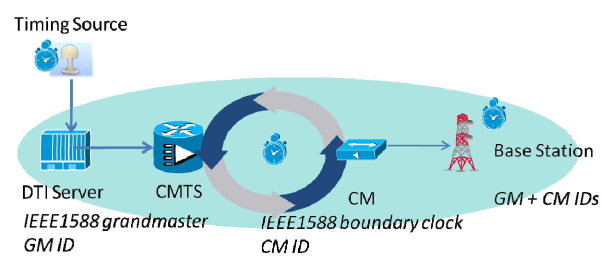


Figure 11 - Scenario 1c

In scenario 1c, despite the fact that there are no PTP messages being sent over the cable path to the CM, the CM simulates a boundary clock. It generates PTP messages with its own clockID but uses the DTI server/CMTS PTP clockID for GM ID. As a result, all CMs connected to this DTI server/CMTS would be related to same "GM".

Scenario 1c may be useful for providing PTP traceability to the "GM" by providing the distinct IDs. In contrast, in scenario 1b, the CM would substitute for or fake the DTI server/CMTS as GM.

For those scenarios, the CMs can be managed by PTP management or through

MIBs. The DTI server/CMTS can be a PTP management node or a Node Manager MIB with MIB extensions for the CMs acting as grandmaster or boundary clock as depicted in Figure 12.

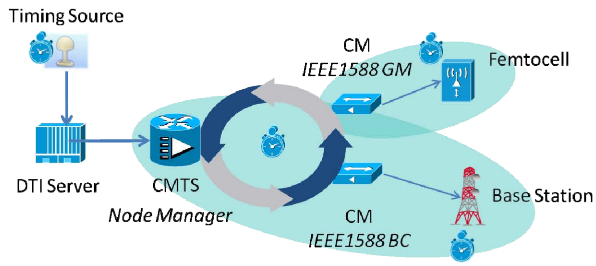


Figure 12 - Scenario 1 Management

Scenario 2

When the primary or a backup timing source and IEEE 1588 GM are remote to the DTI Server/CMTS locations (for instance somewhere in the aggregation or core network), other scenarios are conceivable.

From a timing domain viewpoint, the main difference would be to have the same source/GM for multiple customer clocks. Those clocks might be spread over multiple CMTSs, HFC plants and, also accessible from other access networks (e.g., Ethernet).

In such scenarios, we have to consider the transmission of the timing references towards the DTI Server/CMTS location then to the PTP clocks beyond the CMs. As shown in Figure 13, the DTI server and CMTS have no local timing source. The timing source (e.g., GPS) and GM are remote.

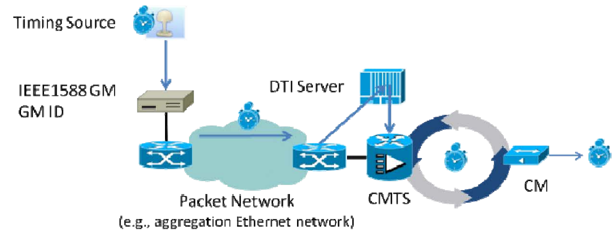


Figure 13 - Scenario 2

Transmission of timing signals from reference(s) over a packet network towards the cable plants can be achieved by the normal TWTT protocol with network assistance as described earlier (e.g., Synchronous Ethernet for physical-layer frequency, hardware assistance to PTP or NTP packets).

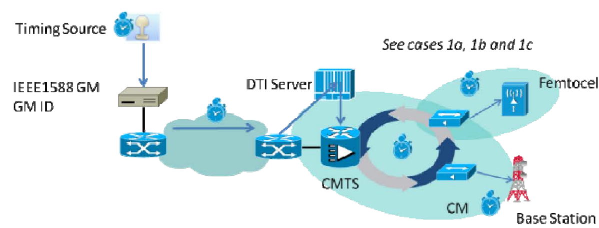


Figure 14 - Scenario 2a

In scenario 2a, the DTI Server/CMTS/CM conceptually can behave as in any previous scenarios 1a to 1c. The difference is that the DTI server will use the network timing signal(s) instead of a local timing source. This scenario does not provide traceability to the actual timing source and IEEE 1588 GM.

Hiding or removing the traceability to the central GM may be intentional, for instance, for timing and management domain delimitation/creation towards the customer network. The information at the CM is defined at the DTI server/CMTS. But because the information sent by the CMTS to the CM comes from the GM via DTI, the operator can trace back to the real GM.

Conceptually the next scenarios may look like one distributed boundary clock (scenario 2b) or two virtual boundary clocks (scenario 2c).

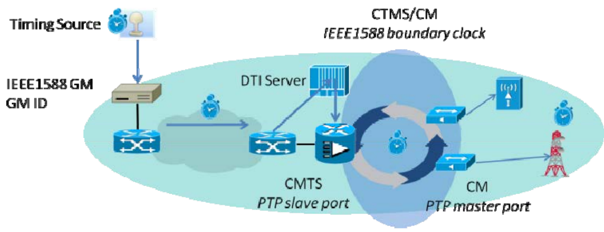


Figure 15 - Scenario 2b

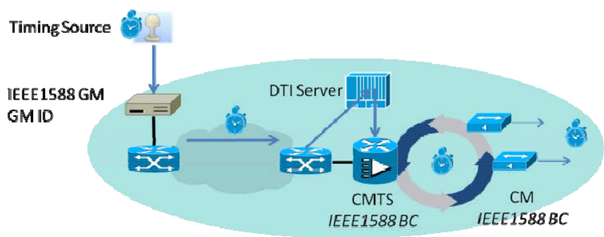


Figure 16 - Scenario 2c

The DTI/CMTS recovers the time from central GM, i.e., the DOCSIS time is synchronized to central timing source. The DTI Server/CMTS and the CM must modify the information from the central GM before being delivered to clocks beyond the CM via PTP.

Similar to scenarios 1b and 1c, the customer clocks can trace back to the real central timing source and GM, not the DTI server/CMTS fake GM.

Differences between the one BC and two BCs scenarios come from the clockID being used for traceability and from the count of PTP hops. In case of one BC scenario, the clock ID may be the clockID of the DTI Server/CMTS, the one from the CM or a distinct clockID that would represent the BC.

The BC hop count (PTP "removeStep" dataset member) would be incremented by one. For a two BC scenario, the DSIT Server/CMSTS and CM would use their respective clockID and the BC hop count would have to be incremented by two. This is again a distinction related to PTP communication path management and monitoring.

Note that in scenario 2c, the CMTS and CM are not actual IEEE 1588 BC because they do not receive PTP messages. We might consider that the CMTS may receive directly the PTP messages and recover time (PTP slave port) without going through the DTI server and necessary associated manipulations. This is depicted by Scenario 2d.

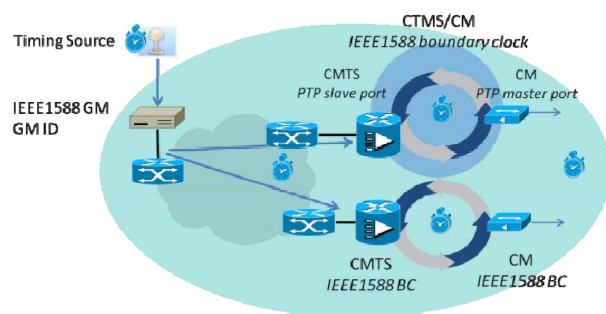


Figure 17 - Scenario 2d

Scenario 3

A final alternative would allow the HFC plant to be independent of the timing source located in the network.

Conceptually this scenario may look like a distributed transparent clock.

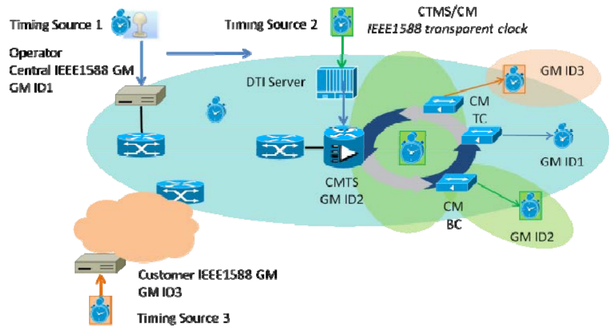


Figure 18 - Scenario 3

In this scenario the DTI Server and DTP can use distinct timing references. This scenario allows timing source 2 to be transmitted to CM as depicted by scenarios 1a to 1c. However, the PDV correction provided by DTP could be a benefit to the PTP traffic sent "over" the cable path just as an IEEE 1588 transparent clock would do. Such a mode would permit correcting any time reference (such as timing source 1) distributed over the cable plant, allowing the support of multiple timescales.

From PTP management viewpoint, the CMTS can still play as node manager for the connected CMs. A more centralized approach can also apply.

A mix of solutions may be useful because different application, customer and timing domain management. However only a subset of the presented various scenarios should be considered.

SUMMARY OF DESIGN CHANGES

The authors of this paper will be pursuing these proposed changes with the appropriate standards committees and interested vendors.

CM

Hardware

- MUST be able to lock to the DOCSIS downstream baud clock, regardless of the upstream modulation type.
- MUST filter jitter from the DOCSIS downstream baud clock to a level acceptable for an Ethernet clock.
- MUST support Synchronous Ethernet.
- MUST have a PTP output circuit on its Ethernet port that inserts a PTP timestamp.
- MUST be able to measure the difference between the transmit time of an upstream packet in terms of the local CM timestamp and the CMTS timestamp in the MAP.
- MAY have a 1PPS output.

Software

- MUST have a PTP Stack
- MUST support DOCSIS protocol changes for DTP.

CMTS

Hardware

- MAY have a PTP slave input for network sync.
- MAY have a PTP slave input and/or a 1PPS input for connecting to reference timing CMs.
- MUST synchronize downstream baud clock to CMTS master clock.

Software

- MUST support DOCSIS protocol changes for DTP

- MAY support DTI protocol changes for DTP
- MUST support the coordination of DTI Server and CM operations for DTP.

DTI Server

Hardware

- MUST have a PTP Slave port for connectivity to CM under measurement.
- MUST have a PTP Slave port for network sync connectivity. These two slave ports MAY share a common physical port.
- MAY have a 1PPS input.
- MUST have the ability to compare the timestamp from the PTP slave port when received directly to the DOCSIS timestamp.

Software

- MUST support DTI protocol changes for DTP.

DOCSIS Protocol Changes

DS MSG:

- Upper bit PTP prefix (for operation greater than 7 minutes)
- Lower bit PTP suffix (for resolution down to 1 ns) based upon the fraction time field originally intended for SCDMA-only use.
- Translation of DOCSIS timestamp with suffix and prefix to EPOCH
- Clock ID

US MSG

- True ranging offset.

DTI Protocol Changes

CMTS to DTI Server

- DTP measurement request

DTI Server to CMTS

- DTP measurement response

CONCLUSION

The DOCSIS system is already based upon highly precise timing. Rather than running timing protocols such as NTP and PTP independently over-the-top of DOCSIS, better performance can be achieved by leveraging the precision timing already native to the DOCSIS system that allows the CM to be time and frequency synchronized to the CMTS.

Using this as the cornerstone, the time and frequency synchronization of the CMTS and CM can be used to correct any path delay variation through the network when acting as a transparent clock thus allowing more effective PTP over-the-top of DOCSIS. This approach would even allow the support of multiple timing domains.

Similarly, the DTP system can accurately generate time and frequency from the CM to natively support different protocols and interfaces like PTP, NTP, Synchronous Ethernet or 1PPS.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of George Zampetti, the Chief Scientist of Symmetricom and the primary author of the DTI Specification [DTI] in the proving out of this concept and in the review of this white paper. The authors would also like to thank Michael Overcash of Cisco for his contributions.

REFERENCES

[1588]
IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, IEEE Std 1588-2008, http://standards.ieee.org/findstds/standard/1588-2008.html

[1588 Applications]
Recent Advances in IEEE 1588 technology and Its Applications, John C. Eidson, July 19, 2005

[1588 Tutorial]
IEEE 1588 Tutorial, Prof. Hans Weibel, October 2, 2006

[802.1AS]
IEEE 802.1AS-2011 Timing and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks

[802.3bf]
IEEE P802.3bf TimeSync Task Force, 2011 http://www.ieee802.org/3/bf/public/index.html

[G.8261]
ITU-T G.8261-Y1361-200804, Timing and Synchronization Aspects in Packet Networks (and related Corrigendum, Amendment and Addendum)

[G.8262]
ITU-T G.8262-Y.1362 (201007), Timing Characteristics of Synchronous Ethernet Equipment Slave Clock (EEC) (and related Corrigendum, Amendment and Addendum)

[G.8264]
ITU-T G.8264-Y.1364-200810, Timing Distribution Through Packet Networks, (and related Corrigendum, Amendment and Addendum)

[G.8271]
ITU-T SG15 Q13 work in progress recommendation which will define the Time and Phase synchronization requirements and network limits

[DOCSIS 2.0]
CM-SP-RFIv2.0-C01-081104, DOCSIS Radio Frequency Interface Specification, CableLabs, August 11, 2004

[DOCSIS DRFI]
CM-SP-DRFI-I11-110210, DOCSIS Downstream RF Interface Specification, CableLabs, issued Feb 10, 2011

[DOCSIS MACUP]
CM-SP-MULPIv3.0-I15-110210, DOCSIS 3.0 MAC and Upper Layer Protocols Interface Specification, CableLabs, Issued Feb 2, 2011

[DTI]
CM-SP-DTI-I05-081209, DOCSIS Modular Headend Architecture Timing Interface Specification, CableLabs, issued Aug 12, 2009, and ITU-T J.211-200611, Timing Interface for Cable Modem Termination Systems

[GPON]
ITU-T G.984.4-200911-Amd2

[XG-PON]
ITU-T G.987.3-201010 (Pre-published)

DOCSIS To CMAP Evolution

Jeff Finkelstein
Senior Director, Network Architecture
Cox Communications

Jorge Salinger
VP, Access Architecture
Comcast

ABSTRACT

As the need for additional video channels increases for legacy MPEG-TS delivery and new and evolving IP delivery, the amount of equipment needed to provide that new capacity may be greater than current facilities can support. Building more head-end or hub locations may be an option, but it is one of the most expensive and invasive steps that operators may undertake. As an industry, we have realized that to wait until the time you need the bandwidth is too late, and that we need to be proactive in order to be prepared for this eventuality.

Denser Edge QAM technologies may be used to help solve the growth from an edge perspective, but they do nothing to solve the need for more backend processing to handle the IP video streams. DOCSIS bypass has been proposed as one method for solving the IP handling and avoiding being forced to add more CMTS equipment, but in turn, DOCSIS bypass forces operators into a non-standard solution.

The Converged Multi-Service Access Platform (CMAP) provides the combined functionality of legacy Edge QAM, data processing of CMTS, and IP video processing. CMAP is able to provide full downstream spectrum through a single port that allows operators to have complete flexibility in deployment of services throughout the full range of channels.

The purpose of this paper is to discuss methods for using CMAP to solve these problems by using the new architecture as an evolutionary step in moving towards a completely converged edge solution across all services. There is no one-size-fits-all solution in this area as it crosses technology and business units, but CMAP provides a mechanism allowing each to continue managing their services as we do today and evolve into the future by incremental steps. CMAP also allows us to take a revolutionary approach and leap into the future by immediately converging services at the edge.

BACKGROUND

Much has been written and said recently regarding the need to leverage the more efficient delivery and lower cost features of QAM technologies in support of broadcast video, unicast video, and high-speed data services. This has led to the development of specifications that have collectively become known as the Converged Multi-Service Access Platform (CMAP) by a Comcast-led team of MSO and vendor partners.

Rather than delving into why there is a need for CMAP and of what it consists of, this paper focuses on explaining the how. In particular, we will look into the planning involved in migrating from current modular CMTS deployment to the CMAP architecture.

PLANNING FOR CHANGE

As an industry, we strive to develop architectures that allow us to add features without requiring a complete redesign or replacement of existing equipment. While this is a good business goal, at times it does not remove the need for an equipment forklift out of necessity. We may be able to reduce the parts in need of replacement by changing out line cards, but there will be decision points for complete chassis replacements when the required capacity processed by each constituent device exceeds the backplane or overall chassis capabilities.

These inflection points tend to come at the worst possible time, which is typically when we are rolling out new services. With each new service come additional complexities in operations, administration, and management (OA&M) functions. The resulting outages caused by chassis replacement on customer service and operational effectiveness is at times devastating, and while that can be mitigated with planning and staging, the impact is real and never quite goes as expected. When dealing with chassis forklifts, Murphy is an optimist. If something can go wrong, it not only will go wrong, it likely already has but we have not noticed it as yet.

Much thought has been given to how our future needs converge video, voice, and data. IP video services, including both unmanaged over the top services and managed services, change how we view the network. We realize our service groups must come into alignment at some level as well to maximize the intrinsic value of our equipment spends. With new silicon development creating the potential to provide the entire forward path from a single port, our architectural view of the HFC ecosystem is on the verge of a paradigm shift.

With the opportunity to architect a new HFC network comes the chance to consider a new method that allows incremental changes to our existing infrastructure. We may also plan to reduce complexities inherent in our current HFC network by taking advantage of this unique opportunity in converged services. Space, power, heating, and cooling savings are key drivers as well.

The architecture must be simple and flexible in its design with built-in growth options. The cost should be significantly lower than that of existing DOCSIS solutions by leveraging technology developments via significantly improved

QAM density while maintaining hardware cost constants. Operationally, the new platform must provide us with a more reliable and manageable product that has integrated redundancy and reduces the amount of individual components being managed.

As we consider the need for this new platform, one thing becomes evident to us; it would be beneficial to leverage portions of the current architecture with deployed Modular CMTS (M-CMTS) and Edge QAM networks to reduce the complexity of transitioning to the new technology. Taking these small steps may minimize the overall impact of the new technologies.

CMAP ARCHITECTURES

CMAP was designed to support a primary architecture of a single integrated chassis, where high-level processing and physical line cards for both downstream and upstream channels are developed in a single enclosure.

For the purpose of discussion in this paper, we are focusing on Modular CMTS and how we may utilize existing technologies to provide an evolutionary path to a full CMAP deployment.

To begin the discussion we need a background on Modular CMTS technologies.

MODULAR CMTS BACKGROUND

In the current Modular CMTS architecture, the CMTS Core has one or more downstream network interfaces that communicate with one or more Edge QAM devices to provide the downstream channels sent to the fiber nodes for distribution to customer premises. Upstream receivers are integrated into the M-CMTS Core to simplify MAC level processing. A timing interface is required between the MAC layers contained in the M-CMTS Core and the PHY layer resident in the Edge QAM to provide the precise synchronization needed in scheduling upstream burst transmission by cable modems.

In such Modular CMTS head-ends, MSOs will continue to grow their access networks to support capacity needs for which it will be necessary to add downstream QAMs and the corresponding Edge QAM ports (in addition to new interface cards and routing/switching equipment to provide the communication path between them). We will also need to add more M-CMTS Core processing chassis' to house the new line cards needed and more Edge QAM chassis to handle the QAM modulation. It is expected that we will all optimize these purchases recognizing that not moving towards CMAP will eventually result in running out of *"brick and mortar"* space before having sufficient capacity farther down the road for future service needs.

As we on the CMAP core team discussed the alternatives, one option stood out. As the interface between M-CMTS and Universal Edge QAM was already defined within the CableLabs [MHA] specifications and had been successfully implemented by a number of vendors, couldn't CMAP also take advantage of that work?

The CMAP team discussed many of these options and decided to formally incorporate into the CMAP specification the appropriate CableLabs specification references to explicitly provide support for Universal Edge QAM functionality. This allows the integrated CMAP equipment to function as universal Edge QAM devices. By taking this step, MSOs would be able to utilize new CMAP equipment with existing Modular CMTS equipment, providing a transition roadmap from today's architecture to the CMAP future, without requiring a forklift of existing equipment.

MODULAR HEAD-END ARCHITECTURE SPECIFICATION SUPPORT IN CMAP

There are a number of interfaces defined in the CableLabs Universal Edge QAM specifications including:

- DOCSIS Timing Interface [DTI]
- Edge Resource Management Interface [ERMI]
- Downstream External PHY Interface [DEPI], including L2TPv3 over IP, L2TVv3 over UDP/IP, MPT and PSP modes, etc.
- Edge QAM Provisioning and Management Specification [EQAM PMI]
- Edge QAM Video Stream Interface [EQAM VSI]
- M-CMTS Operations Support System Interface [M-CMTS OSSI]

While achievable, adding all the above specifications and options into CMAP might not be necessary given the original objective of adding Edge QAM functionality and would potentially add delays to the implementation timeline. As MSOs reviewed the specifications it was determined that the minimal set of M-CMTS requirements would be:

- DOCSIS Timing Interface
- Downstream External PHY Interface with MPEG Transport (MPT) support

The DTI and DEPI MPT are the primary specifications/features used by MSOs deploying modular CMTS today, and many vendors have already developed products that meet these CableLabs specifications. Therefore, it made sense to maintain that support in the CMAP specifications for the Integrated CMAP device.

BRIDGE YEARS TO CMAP

For those MSOs that have extensive deployments of modular CMTS, we are investing in additional capacity both for CMTS and Edge QAM over the next few years to support the anticipated growth resulting from market forces. We are looking at CMAP as a long-term solution towards solving the increasing costs involved both capital and operational expenditures.

But as we work towards extending the bandwidth and capacity of our high-speed data and video services, our goal of maximizing the usable life of our currently deployed equipment is challenged by the needs of new services. We risk running out of capacity for line cards or Edge QAM devices, which will require us to add more equipment. By adding more equipment during the bridge years leading to CMAP, we are potentially making less than optimal choices on the devices we are deploying.

What we need is the new CMAP platforms available in the marketplace sooner rather than later so that we may reduce the amount of capital expenditures for new services. But, the availability of new CMAP platforms are still far enough away in time that we need to make these purchases now to provide this new bandwidth.

So what do we do?

EVOLUTION, NOT REVOLUTION

In order to simplify the initial deployments of CMAP devices, we have been working with vendors on a downstream-only version of CMAP that some have dubbed "CMAP-lite". This allows a CMAP device to be used as a Universal Edge QAM by the CMTS core and provide a very dense solution for unicast and broadcast traffic.

While the density of CMAP functioning as an Edge QAM adds more downstream capacity for high-speed data using Modular CMTS interfaces, there is an additional benefit of higher density for video QAM channels as needed by narrowcast services e.g. SDV and VOD. Being able to leverage these increased densities for all unicast services simplifies the eventual transition to CMAP by reducing complexity in the head-end combining network.

An additional benefit is taking advantage of replication within the CMAP chassis, which delays the need for converging video and data service groups. In today's head-end architectures, we have mismatched service group sizing between narrowcast and broadcast groupings, and between video and data. The ability to replicate streams internal to the CMAP device simplifies the combining network in the head-end. Figure 1 shows an example of how using the CMAP replication feature we are able to delay alignment between data and video service groups.

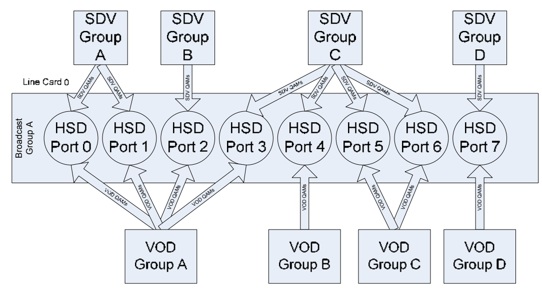


Figure 1 - Service Replication Internal to CMAP

EXAMPLE ARCHITECTURAL EVOLUTION

The following figures show how an operator deploying Modular CMTS today might be able to take advantage of Modular Headend Architecture [MHA] support in CMAP. Figure 2 shows how a Modular CMTS deployment looks today. In figure 3, CMAP may be used to augment existing video Edge QAM equipment and migrate to a full CMAP deployment. Both show how CMAP may be used to add capacity for a modular CMTS deployment and begin migrating all services to CMAP.

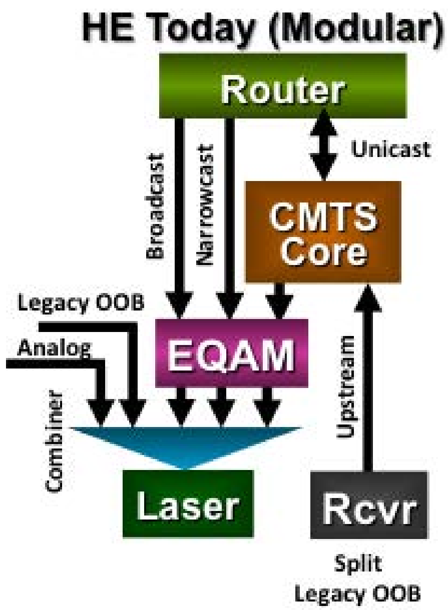


Figure 2 - Modular CMTS Architecture Today

As is shown in figure 2, the Edge Router sends broadcast and narrowcast video through the Edge QAM to be transmitted by the laser to the fiber node. Unicast data traffic is sent through the M-CMTS Core and down to the laser via the Edge QAM. Upstream traffic is routed through the M-CMTS Core and out through the Edge Router. All data handling is provided by the M-CMTS Core and sent to the Edge QAM for transmission to the customer.

The next few examples show methods for using an integrated CMAP for providing Edge QAM like services.

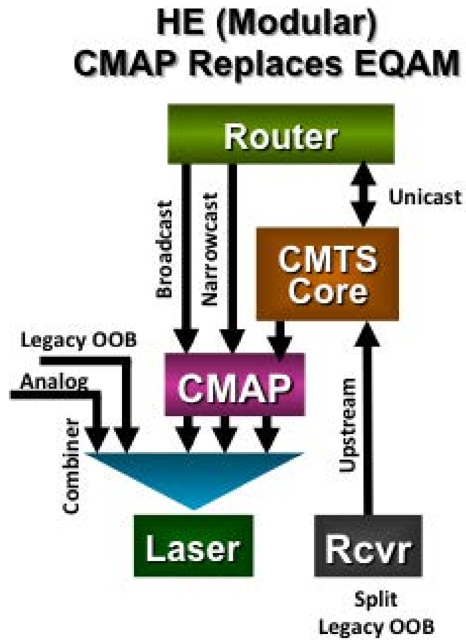


Figure 3 - CMAP replacing existing Edge QAM

In figure 3, using CMAP as an Edge QAM provides for denser QAM deployments without requiring any changes in the DOCSIS network. This allows growth in downstream capacity for both video and data services without impacting the M-CMTS Core devices deployed in Modular CMTS head-ends being used today.

If we need to scale legacy video, instead of adding more Edge QAM devices, we may simply reassign QAM channels to video services. If more data channels are required, we can do the same reassignment.

MOVING TO A FULL CMAP SOLUTION

At this point we can now replace the M-CMTS Core with CMAP to take over the routing and high level MAC processing. The CMAP device now becomes an integrated CMTS core and Edge QAM all in one chassis by having all linear video, narrowcast video, and data traffic sent to it by the edge router.

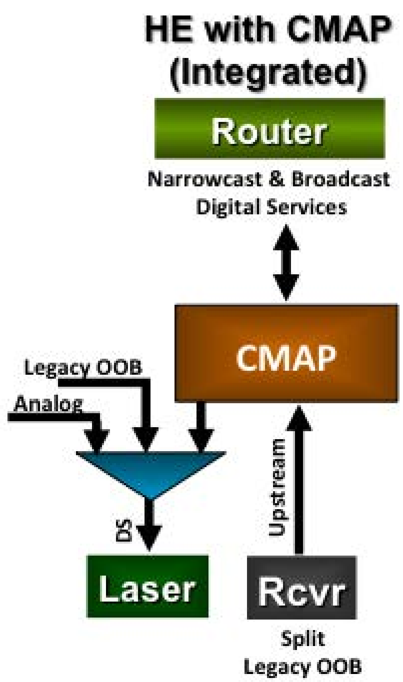


Figure 4 - Integrated CMAP Fully Deployed

Figure 4 shows how an integrated CMAP may be used to provide all capacity for narrowcast and broadcast services, whether video or data. This step is taken when the capacity provided in figure 3 has exceeded the CMTS core device capacity and would require more equipment to be installed.

We have now removed all existing legacy video and data equipment from the traffic path. Benefits to this architecture are many, the most significant being the power, heating, cooling, and rack space savings realized by replacing multiple video and data equipment by a CMAP chassis.

CONCLUSIONS

The CMAP platform provides the next step in access technology evolution. As it provides for full forward spectrum from a single connector, most head-end wiring and resulting complexity become obsolete. With all services including SDV, VOD, broadcast and HSD provided from a single edge device, the points of management in the network are reduced. Environmentally floor spaces, power consumption, UPS capacity, heating, and cooling savings are significant.

The challenge that MSOs face with CMAP is one of evolution toward deployment without having to revolutionize their network and headend design to take advantage of this next step in the access network technology life cycle. By taking advantage of the work done to date by CableLabs and the CMAP team, vendors and MSOs are able to provide a transition for the bridge years by using CMAP as a super dense Edge QAM while progressively retiring existing equipment to optimize expenditures prior to a broad deployment of CMAP. This simplifies the transition and allows services to be migrated as needed, which improves capabilities of the current spend in providing a longer usable life for equipment being deployed today.

ACRONYMS

BC: Broadcast
CAS: Conditional access system
CLI: Command line interface
CMTS: Cable modem termination system
dBmV: Decibel referenced to millivolt
DOCSIS: Data over cable service

interface specification
DRFI: DOCSIS radio frequency interface
FPGA: Field programmable gate array
GHz: Gigahertz
GigE: Gigabit Ethernet
HE: Headend
HFC: Hybrid fiber-coax
HSD: High-speed data
MAC: Media access control
MCX: Multi commodity exchange
MHz: Megahertz
MPEG: Moving Picture Experts Group
MSO: Multiple system operator
NC: Narrowcast
OA&M: Operations, administration, and management
OTN: Optical termination node
PHY: Physical
PMI: Provisioning and management interface
PON/EPON: Passive optical network/Ethernet passive optical network
QAM: Quadrature amplitude modulation
RF: Radio frequency
RFI/RFQ: Request for information/request for quote
SCTE: Society of Cable and Telecommunications Engineers
SDV: Switched digital video
SNMP: Simple network management protocol
VOD: Video on-demand
VSI: Video stream interface
XML: Extensible markup language

REFERENCES

[DEPI] DOCSIS Modular Headend Architecture Downstream External PHY Interface Specification, CM-SP-DEPI-I08- 100611, June 11, 2010, Cable Television Laboratories, Inc.

[DRFI] DOCSIS Downstream RF Interface Specification, CM-SP-DRFI-I10-100611, June 11, 2010, Cable Television Laboratories, Inc

[DTI] DOCSIS Timing Interface Specification, CM-SP-DTI-I05-081209, December 9, 2008, Cable Television Laboratories, Inc

[ERMI] DOCSIS Edge Resource Manager Interface, CM-SP-ERMI-I03-081107, November 08, 2007, Cable Television Laboratories, Inc.

[MHA] EQAM Architectural Overview Technical Report, CM-TR-MHA-V02-081209, Cable Television Laboratories, Inc.

[MHA Prov] Edge QAM Provisioning and Management Interface Specification, CM-SP-EQAM-PMI-I01-081209, December 9, 2008, Cable Television Laboratories, Inc.

[MULPI] MAC and Upper Layer Protocols Interface Specification, CM-SP-MULPIv3.0-I14-101008, October 8, 2010, Cable Television Laboratories, Inc.

[PHY] Data-Over-Cable Service Interface Specifications Physical Layer Specification, CM-SP-PHYv3.0-I09-101008, October 8, 2010, Cable Television Laboratories, Inc.

New Converged Access Architectures for Cable Services

John Ulm, Gerry White
Motorola Mobility

Abstract

CMTS architectures have been evolving over the years, from the initial platforms with fixed upstream and downstream ratios to flexible integrated and modular CMTS. Today, we are seeing the emergence of a new generation of converged CMTS + UEQAM called CMAP. This paper reviews various CMTS/CMAP architectures including integrated and modular options. We then discuss migration strategies from today's silo infrastructure to a single converged platform. From here we investigate potential future evolutions as the industry prepares for all IP video delivery. We also look at the inclusion of application level services into these converged edge platforms to support new added value services.

INTRODUCTION

The authors have had the privilege of participating in the genesis of the cable broadband industry at LANcity and then author the first DOCSIS specifications. We have watched DOCSIS grow and change over the last 15 years. DOCSIS 1.0 enabled data services for cable operators to compete with DSL. DOCSIS 1.1 introduced QoS capabilities to enable voice services. DOCSIS 2.0 enhanced the upstream bandwidth while DOCSIS 3.0 enabled wider bonded pipes to withstand FTTP competition. Now, the next major impetus for DOCSIS evolution is IP video delivery.

IP video delivery over cable has been slowly gathering critical mass, but to date has always just been on the edge of taking off. Numerous architectures and other technical proposals such as M-CMTS and CMTS Bypass have been discussed over the years to enable IP delivery; but these have not received widespread adoption. With the proliferation of intelligent video capable devices such as gaming consoles, smart phones and tablets there is a renewed emphasis on IP video. When combined with the latest industry efforts around a converged next generation CMTS + EQAM platform known as CMAP; these may be the final ingredients needed for widespread IP video deployment.

Below we dive into the various CMTS & CMAP architectures. Then we look to how to enable existing equipment to be integrated with CMAP equipment to provide a smooth migration from early IP video deployments to IP video everywhere. Finally, we delve into the future evolution of CMAP and the integration of application level services into these converged edge platforms.

THE EVOLUTION OF DOCSIS ARCHITECTURES

In the early days of DOCSIS, CMTS systems tended to have fixed upstream to downstream ratios since the early generation silicon required a tight coupling between upstream and downstream. The price per CMTS channel then was measured in tens of thousands of dollars. These early CMTS systems enabled broadband data service and voice services to get off the ground and become well established. As internet traffic continued to skyrocket, people in the industry started to look at ways to significantly reduce the cost of DOCSIS downstream channels so that DOCSIS could expand to meet this growth. These efforts eventually led to the Modular CMTS (M-CMTS) and DOCSIS 3.0 specification work at CableLabs.

M-CMTS

While DOCSIS 3.0 emphasized larger data pipes via bonded channels, the M-CMTS effort was focused on breaking the logjam on cost per DOCSIS channel. The first critical concept it introduced was the ability to decouple the downstream from the upstream. This would allow operators to add downstream capacity independent from upstream. In earlier systems with upstreams and downstreams bundled together, you paid handsomely for the upstreams even if you only needed downstream capacity.

The second important aspect of M-CMTS was the desire to ride the EQAM cost curves for the downstream PHY (i.e. QAM + Upconverter technology). EQAM costs were on the order of one tenth the cost of a DOCSIS channel at that time. So the Downstream External PHY Interface (DEPI) was developed as part of the M-CMTS work. However, this required a change in existing EQAM. They needed to include a DOCSIS Timing Interface (DTI) to support critical DOCSIS timing. This created a new category of EQAM devices called Universal EQAM (UEQAM).

Modular vs. Integrated

A number of the CMTS vendors recognized that the M-CMTS benefits could also be implemented inside their single integrated CMTS (I-CMTS) system. It was possible to decouple the upstream and downstream inside existing CMTS and apply the underlying technology that was driving down EQAM costs to the CMTS downstream cards.

A pictorial view of I-CMTS and M-CMTS is shown in Figure 1 which illustrates some of their key differences. From a block diagram level, the main difference for M-CMTS is that the downstream PHY has been moved over to the UEQAM. From a cost perspective, M-CMTS was intended to create a healthy vendor ecosystem with lots of competition.

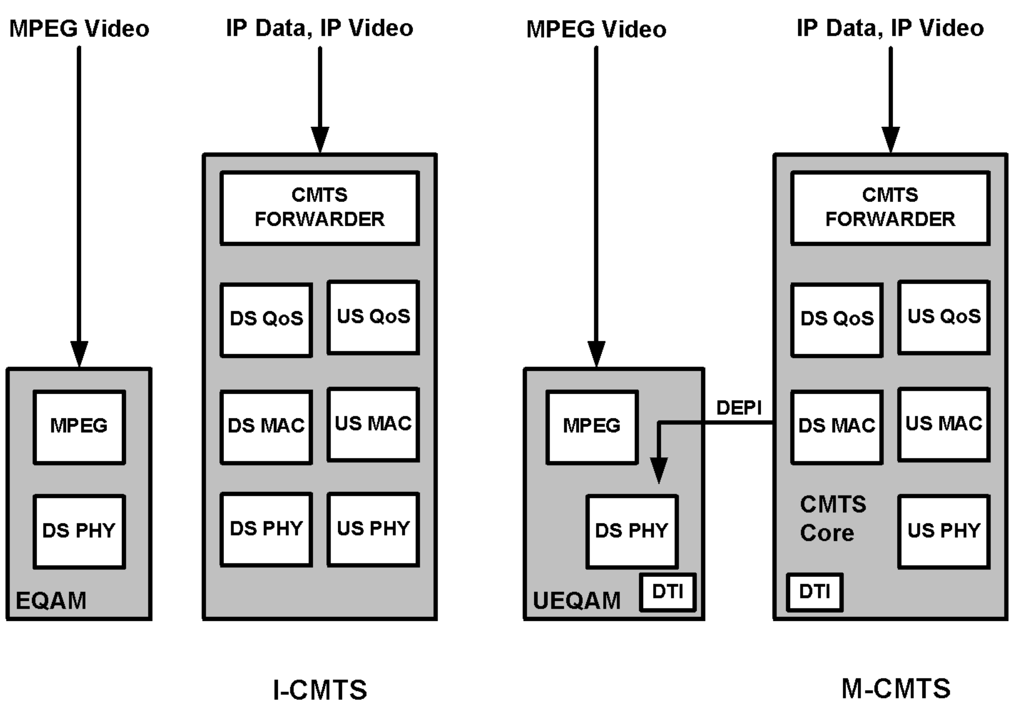


Figure 1 I-CMTS and M-CMTS Components

While multiple UEQAM products have come to market, this is balanced by the limited number of CMTS core vendors that chose to implement the DEPI functionality.

The M-CMTS approach has additional hidden costs. M-CMTS requires a DTI server that both the CMTS and UEQAM need to implement. There are also interconnect costs that may include an Ethernet switch. The costs are compounded as redundancy is factored into the system.

On the operational side, the I-CMTS has a clear advantage. It offers a single vendor system integration that leads to reduced operational costs. For M-CMTS, the operator needs to be the system integrator. The M-CMTS approach also took a longer time to get to market since it was a new multivendor interface. Early adopters went through numerous integration challenges.

DOCSIS IPTV Bypass Architectures

Competition between I-CMTS and M-CMTS helped drive down DOCSIS channel costs quickly. However, EQAM costs continued to drop at an equally quick pace. As the cost of DOCSIS channels headed towards $1,000, the EQAM channel costs drove towards $100. This led to the development of a DOCSIS IPTV Bypass Architecture. This was first published in 2007 [1]. Other variants of CMTS Bypass have appeared over the following years.

While CMTS Bypass approaches promised more cost effective delivery of IP Video, it was optimized for Constant Bit Rate (CBR) video using multicast delivery. It has several drawbacks when considered for more general purpose delivery. There are issues which have prevented wide scale deployment including unicast delivery, VBR, DOCSIS 3.0 bonding, mixing data & video in the same channel and privacy.

THE DAWN OF A NEW ERA

NGAA + CMAP Highlights

In 2009, Comcast expanded its Next Generation Access Architecture (NGAA) initiative to include other MSO's and vendors in an effort to define the requirements for a converged CMTS, EQAM + PON product. This would be called the Converged Multiservice Access Platform (CMAP). Several papers on CMAP may be found in the references [2, 3, 4].

CMAP provides a highly integrated system with all services in a single box. It supports voice, data and IP video services along with legacy MPEG video, digital broadcast and commercial services. A major tenet of CMAP is that all narrowcast services for a given serving group are delivered from a single CMAP RF port, greatly collapsing the RF combining network. This leads to significant Head End space savings and extensive operational savings. Hence, the CMAP needs to function as both a CMTS and an EQAM. The CMAP based solution offers significant operational advantages. It provides a single management point for the entire system, including video, voice and data services.

This level of integration is now possible thanks to the on-going advances in silicon technology. CMAP targets 64 narrowcast channels per port with up to 96 digital broadcast channels shared across multiple ports. The goal is to have downstream blades which are based on EQAM technology. The same downstream technology that is driving EQAM costs lower will also drive CMAP downstream blade costs as well.

Given the large number of downstream channels in a CMAP device, another key CMAP attribute is its ability to reassign any channel to be either a legacy MPEG video or a DOCSIS data channel. For example, an operator could initially deploy CMAP with 80% of the downstream channels being legacy video; then over time adjust the mix until they reach 100% IP delivery. This will be one of the key CMAP features that enable the transition to IP Video services.

With the downstream PHY technology increasing density tenfold, the rest of the DOCSIS solution including CMTS forwarder and downstream QoS must scale as well. DOCSIS traffic in a typical CMTS often needs significant processing. This includes classification to map packets to a service flow followed by complex traffic scheduling. IP Video traffic has some key characteristics such as large packet sizes and fixed delivery intervals that will allow the CMAP to process these packets with lower overhead than the HSD traffic on today's CMTS. This has sometimes been referred to as a "DOCSIS Lite" approach. Effectively, the bypass concepts are now being implemented inside a single box rather than across several boxes. This along with new generations of network processor silicon means that the processing overhead of IP Video will be essentially the same as legacy MPEG video transport.

CMAP will enable a "standards" based solution. All IP traffic moves transparently over the DOCSIS infrastructure. No special control protocols are needed for IP Video transport in contrast to bypass options which require a proprietary protocol at the EQAM and Cable Modem Gateway. With a generic DOCSIS solution, you can simultaneously support multiple IPTV protocols over the same infrastructure. You also have flexibility to change over time as new protocols such as adaptive streaming take hold. A CMAP solution is ideally suited for generic unicast IP video delivery such as Over The Top (OTT) video or any other managed IP video traffic coming from the "cloud".

The CMAP solution also allows you to take advantage of the statistical gains from large DOCSIS bonding groups and VBR video delivery. CMAP can enable bonding groups of 16 or 32 channels that deliver up to 1Gbps downstream service rates. Because of the large bonding groups, the CMAP solution can mix high speed data with the IP video and provide full DOCSIS services over a single pipe. This provides for a very flexible network transport since it appears as a single IP pipe to the system. This is another feature that enables the IP Video migration.

Integrated and Modular CMAP Overview

The CMAP architecture supports both an integrated (I-CMAP) and a modular approach (M-CMAP). We will look closely at these two CMAP approaches and discuss their pros and cons in rolling out an IP Video delivery system. The major components of both are shown in Figure 2 below.

The I-CMAP components look almost identical to the I-CMTS shown earlier. The ability to process MPEG transport streams is added to the downstream MAC block. This is the additional capability required to implement EQAM functions. The other key addition is a common CMAP OSSI component that enables multiple devices to appear as a single configuration and management entity. While the I-CMAP functional component blocks are the same as I-CMTS, the I-CMAP capacity is tenfold larger than today's CMTS. Given the similarity of component integration to today's CMTS, it is expected that most existing CMTS vendors will opt for this approach, at least out of the gate.

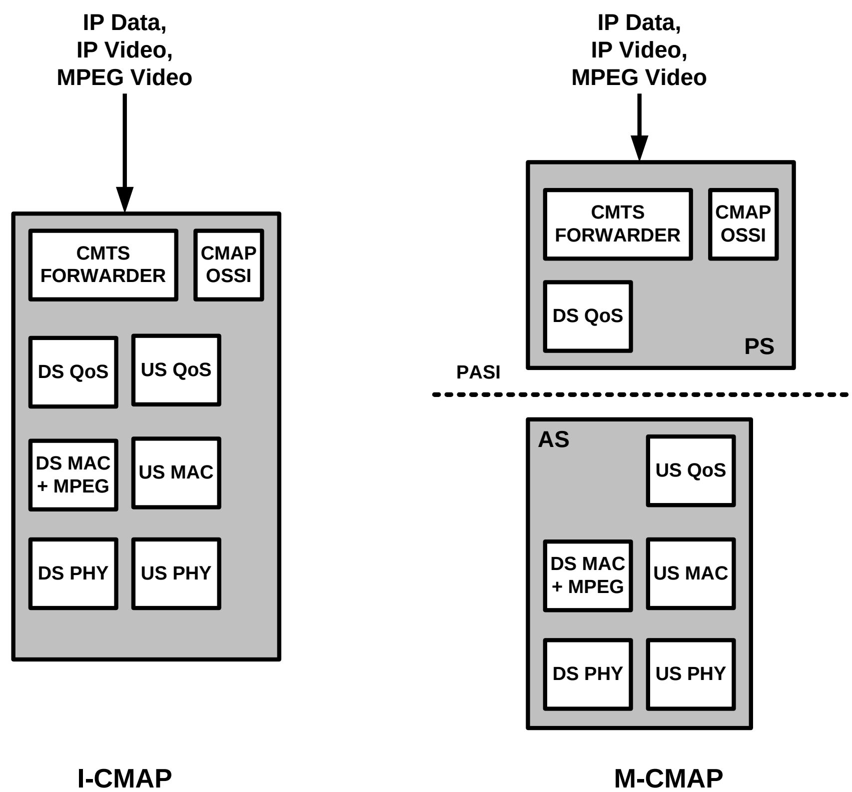


Figure 2 I-CMAP and M-CMAP Components

The M-CMAP approach partitions the CMAP functions into two pieces: a Packet Shelf (PS) and an Access Shelf (AS). A comparison of Figures 1 and 2 shows that this is a radically different partitioning than M-CMTS. The Packet-to-Access Shelf Interface (PASI) specification will become a new CableLab's standard to define the communication between the PS and AS.

At a high level, the PS contains all of the L3 and higher functions including classification, routing and traffic management for the downstream. The PS also includes the DOCSIS Downstream QoS and a large part of the DOCSIS control plane. This partitioning will lend itself to implementation by traditional router vendors, a number of whom participated in the CMAP specification development. One PS may control several separate AS.

The AS will primarily contain the PHY and lower level L2 MAC functions. The AS also contains the DOCSIS upstream QoS and the remainder of the DOCSIS control plane. Unlike M-CMTS, the AS will integrate both the upstream and downstream components. This eliminates the need for an external DTI. This split also has the benefit that the AS can be physically remote from the PS. The PS could be at a centralized Head End serving multiple AS, with some of the AS located in remote hubs. Many of today's EQAM vendors will look to AS products as they migrate to this new converged EQAM + CMTS world.

While the components in Figure 2 give a good representation of the data plane split in M-CMAP, it does not touch on the control plane. The DOCSIS 3.0 MULPI spec (MAC and Upper Layers) assumed an integrated approach like that used in I-CMAP. In

PASI, DOCSIS control and state machines have been split between the PS and AS. In some cases, they are shared between the two. A key tenet of M-CMAP is that the PS is the sole focus for the operator for configuration and management. The entire M-CMAP system appears as a single management entity with the PS coordinating with the multiple AS below it.

Modular vs. Integrated . . . Again

Once again, the cable industry is setting itself up to support two competing architectures: integrated vs. modular. Below is a discussion on several topics on how modular stacks up against an integrated approach for CMAP.

Expanded Vendor Ecosystem – One of the key rationale for a modular approach is expanding the vendor ecosystem. This is certainly true in the CMAP world as M-CMAP will enable both traditional router vendors and EQAM vendors to participate in this new converged CMTS + EQAM market. Having both architectures will lead to a healthy vendor ecosystem with the corresponding competition and innovation. However, an expanded vendor ecosystem by itself is not an indication of one approach being better than the other.

Best of Breed Components – In general, one of the advantages of a modular architecture is that you can mix and match to pick the "Best of Breed" for each individual component in the system. In M-CMAP, you split CMAP into a routing & packet processing engine and an Access shelf. The Best of Breed philosophy is only as good as the selections within each component. As we witnessed in M-CMTS, there was no choice selection when it came to the CMTS core. An unknown question for M-CMAP will be how many router vendors implement the Packet Shelf. If you do not get all of the major routing vendors, then you may not get your Best of Breed choice for routing.

In the CMTS world, some of the most complex components deal with the DOCSIS control plane and the upstream technologies. The existing major CMTS vendors have invested over a dozen years and hundreds of man-years of effort into DOCSIS interoperability testing and certification waves. If the current CMTS vendors opt for I-CMAP, then the Best of Breed DOCSIS and upstream implementations will not be available in an M-CMAP system.

M-CMAP does allow for other access technologies to be easily introduced. This may be well suited for introducing EPON technologies. In today's PON world, most OLT are L2 or simple L3 devices that would be well suited to the PS/AS split in M-CMAP. An EPON AS also does not need to worry about the complexities of splitting the DOCSIS control plane. This is a much simpler partition. Later down the road, the modular split may also enable other wireless technologies like WiFi hotspots or maybe 4G wireless technologies.

Rate of Innovation – Often a modular architecture enables vendors to focus on their area of expertise. This can offer the best likelihood for innovation provided the innovation rests completely within that modular component. The modular approach may actually inhibit innovation if the innovation must cross component boundaries such as PASI. The desire of an operator to maintain flexibility in selecting vendors often leads to the "least common denominator" of features being deployed. If a PS vendor and AS vendor combine to provide added value, then this is simply a two vendor integrated system rather than a truly modular one.

For I-CMAP, it is easier and faster for a single vendor to innovate in an integrated system without the need to coordinate changes with external partners and execute PASI interoperability tests. This may lead to better feature enhancements from integrated vendors. More innovation is possible when you own the entire solution.

Scaling from small to large hub sites – With an M-CMAP split, it's possible to make a simpler AS to fit into small hub sites. The PS based on traditional router platforms does not easily scale down to small hub sites. This leads to an M-CMAP scenario where a single PS supports multiple small hubs with remote AS. This is especially important as operators look to support small remote hub sites in a "lights out" manner. With the PS remotely located at a central Head End, these remote hubs can now be managed from a single location. This may turn out to be one of the most appealing attributes of a modular approach.

Outside of these small remote hubs, how does Modular stack up as we scale to medium and large scale systems? Conventional wisdom may say that the AS can achieve higher densities than I-CMAP because it has less functional components inside. I-CMAP must implement the CMTS Forwarder and downstream QoS functions above and beyond the AS. For medium size systems, it is not clear if this advantage will be enough to outweigh the cost/space/power of adding a PS into the system. So, the modular approach may be more suitable for very large sites where the overhead of the PS can be distributed across many AS.

However, there can be hidden costs with M-CMAP as we scale to very large sites. Every AS needs sufficient network resources to connect to the PS. These need to be redundant links as well. Some head end sites may need to connect multiple AS through an Ethernet switch network in order to conserve PS ports or to simplify head end wiring. Thus, a M-CMAP approach adds the cost of many additional 10G or even 100G network ports to the PS, AS and external switch that are above and beyond what is needed in the I-CMAP approach.

In reality, scaling is more dependent on individual designs than on integrated or modular split. Innovative I-CMAP designs can scale just as well or better than M-CMAP. With today's ASIC silicon technologies, an intelligent network processor can implement the CMTS Forwarder and downstream QoS functions in roughly the space that an AS requires for an internal switch to interconnect its PASI to its downstream components. We envision that I-CMAP densities will rival AS densities.

The overall system costs of an I-CMAP design may turn out to be noticeably lower as well. As mentioned above, the M-CMAP system will have added costs for the network interconnect and switches between the PS and AS. There is also the factor of the CMAP packet processing costs. Historically, the cost per port drops as functions are pushed to the edge of the network; as seen by comparing core router with edge router port costs. This begs the question as to whether it is better to push the CMAP packet processing to the edge in an I-CMAP or to implement it in a PS within a core access router? It remains to be seen which can achieve the most cost effective results.

Operational Complexities – One of the most important focuses of CMAP is the operational simplification. This is a major goal of CMAP. From a configuration and management perspective, the PS will make the system appear as a single CMAP entity.

Thus the M-CMAP can be managed and configured as simply as the I-CMAP system.

Beyond configuration and management, we need to consider other operational impacts. With I-CMAP, the operator has a single vendor support model. One vendor is completely responsible for the system and provides a single contact point for troubleshooting and diagnostic support. Overall, there are fewer components in the system which therefore has overall less complexity.

With M-CMAP, the operator has a multi-vendor support model. The operator must become the system integrator. This is especially critical while the modular technology is still very immature. Troubleshooting problems across multiple vendors significantly increases complexity and is substantially more challenging. The M-CMTS DEPI spec is an order of magnitude simpler than the CMAP PASI spec and we can see how long it took to iron out all the DEPI interoperability issues. The multivendor integration will also be complicated in that one or more of the vendors participating may have never received DOCSIS CMTS certification before.

Another important operational impact is managing the availability of features and bug fixes and coordinating software updates. With a single I-CMAP vendor, the operator can be assured that a SW release has been tested as a complete system. With a multivendor M-CMAP system, the operator needs to coordinate the various SW roadmaps between the different vendors. If a bug requires a simultaneous patch on both the PS and the AS, the operator needs to ensure it works and coordinate the change. As the number of vendors increases and each vendor generates multiple software releases, the interoperability matrix the operator must manage to maintain a functioning system can become very large. M-CMAP becomes a more complex system to operate and manage over time.

Risk and Time to Market – A key item to consider when looking at I-CMAP and M-CMAP system is the risk and time to market impacts of each approach.

The M-CMAP approach needs to create a new Cablelabs spec called PASI. Following the spec creation, the testing procedures and certification infrastructure must be put into place at CableLabs. The entire ecosystem for M-CMAP needs to be created from scratch. Once the ecosystem is in place, the time to deliver new features will require additional time for PASI qualification before being ready for market.

The PASI spec also introduces new risks. It is partitioning DOCSIS functionality that has been operating in the field for a dozen years as a single entity. The downstream QoS and the DOCSIS control plane have become more complex in M-CMAP. On top of the split, new signaling is being introduced to try and manage a multivendor system as a single entity. The amount of industry effort and time to qualify PASI may match or exceed the multiyear efforts required for DOCSIS 1.0, 1.1 and 3.0 certifications and qualifications.

The one area where M-CMAP may reduce risk is in the area of new access technologies such as EPON. The PASI interface partition is well suited for the functions currently supported in EPON OLT.

The I-CMAP approach can leverage existing CMTS designs to get to market faster. I-CMAP devices will run through existing CMTS qualifications. You do not need to wait for the PASI ecosystem to get into place. Going forward, new features can then be added quicker with single vendor integration since you eliminate the added steps for multivendor interoperability and PASI qualifications.

The integrated approach will have fewer boxes leading to a simpler implementation. An example of this is the downstream QoS and DOCSIS control integrated in a single product. Thus, single Vendor integration can mitigate risk compared to complex multi-vendor interoperability around brand new standard.

IP Video Implications – CMAP will be a disruptive change in space, cost and power of a head end. It will enable the continued growth of legacy narrowcast video in the short term followed by the transformation from legacy MPEG to IP video. With its tenfold increase in QAM channels per port, CMAP will enable both the number of channels needed to roll out these services and the cost per channel to make this economically viable.

For the most part, I-CMAP and M-CMAP are identical in their ability to support IP video as described above. However, there is one key area that we need to investigate where they may differ, and that is in the area of multicast distribution. This topic was discussed in detail in reference [6].

From that paper we saw an example of an 80 service group system, For the I-CMAP system, ~2Gbps of multicast video needs to be delivered to each I-CMAP. For a M-CMAP system, the PS might perform the multicast replication. This would require ~50Gbps of bandwidth between PS and AS plus additional links for redundancy. If the AS performs the multicast replication, then QoS problems may be introduced if this traffic is mixed with other traffic that has already been groomed by the PS.

As operators look to make their transition plans to IP Video services, it will be critical that they understand the issues of unicast vs. multicast delivery and the resulting impact on M-CMAP vs. I-CMAP systems.

CMTS EVOLUTION CONTINUES

We do not envision that CMAP will be the end of the CMTS evolution. It will continue to morph and change to meet customer needs. The next section of the paper discusses some possible directions that the CMTS/CMAP evolution might take.

Mixed I-CMAP and M-CMAP Systems

While a lot of the previous discussions focused on comparing and contrasting modular with integrated CMAP, in actuality it makes sense to use the two together. In Figure 3 below we show an AS that is subtended to an I-CMAP device.

The I-CMAP device acts as the PS for the AS. Thus an operator could choose an I-CMAP initially for any of a variety of reasons like reduced risk, faster time to market, appropriate sizing for its head end. Then later when the operator needs to expand beyond the capacity of that I-CMAP, this can be done with the addition of an AS. For example, the I-CMAP may initially support 40 Service Groups. After splitting nodes, the operator could add an AS to the I-CMAP to support the additional 40 SG.

Instead of having two separate I-CMAP boxes to manage, this is configured and managed as a single box. This approach of an I-CMAP with an AS may also offer the best way for an operator to introduce new access technology such as EPON for business services by adding an EPON AS.

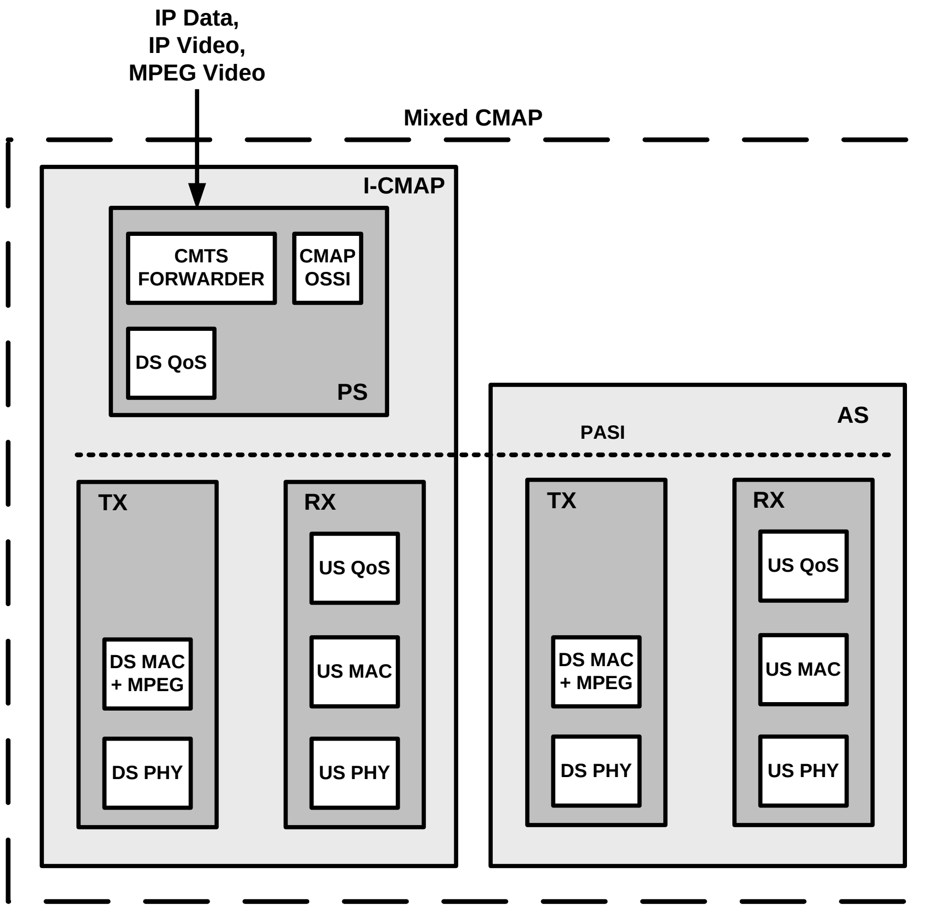


Figure 3 System with mixed I-CMAP and M-CMAP

Distributed CMTS Architecture

The CMTS architectures will continue to evolve as we witnessed when I-CMTS approaches evolved after the M-CMTS specs were written. The M-CMAP PASI specification was created to enable router vendors to become Packet Shelf providers and EQAM vendors to become Access Shelf providers. As discussed previously, this creates a brand new split inside the 750 page DOCSIS MAC and Upper Layer (MULPI) interface spec. This split was not based on any existing implementation and has associated risks with it. Existing I-CMTS vendors have developed systems over the years to independently scale the Transmit (TX) and Receive (RX) portions of CMTS. This work enables a different functional split than a pure M-CMAP System using PASI.

In figure 4 below, the AS components have been separated into separate TX and RX components. We shall call the interface between the PS and TX/RX components the *Distributed CMTS Interface*. The TX and RX components can be replicated as needed to increase downstream and upstream capacity while sharing a common PS.

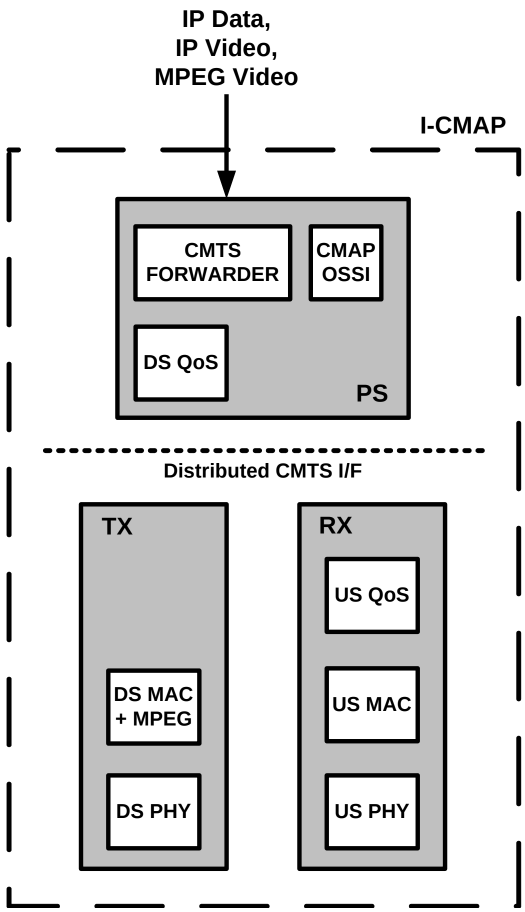


Figure 4 Distributed CMTS Architecture

The PS functions provide the common CMAP capabilities including packet processing, downstream QoS and OSSI. From the operator's perspective, the entire system still looks like a single I-CMAP box. But this alternative split in architectural components enables new possible configurations. The TX and RX components could be located in adjacent chassis. We shall refer to the chassis containing the PS component as the primary chassis while any chassis with just TX &/or RX components will be called a subtended chassis.

Distributed CMTS – Upstream Only

Previously, we discussed a mixed CMAP system where an AS was added to an I-CMAP. In the example stated, this allowed expansion from 40 to 80 Service Groups. Note that this requires a significant amount of interconnect between these components; perhaps as much as 100 Gbps. This could then be doubled if you added extra links for full redundancy.

With a Distributed CMTS system, you can attack the problem differently. Since the TX and RX components are separated, you could put all the RX components in a subtended chassis while the TX components stay closely coupled with the PS functions inside the I-CMAP. This is depicted in figure 5 below.

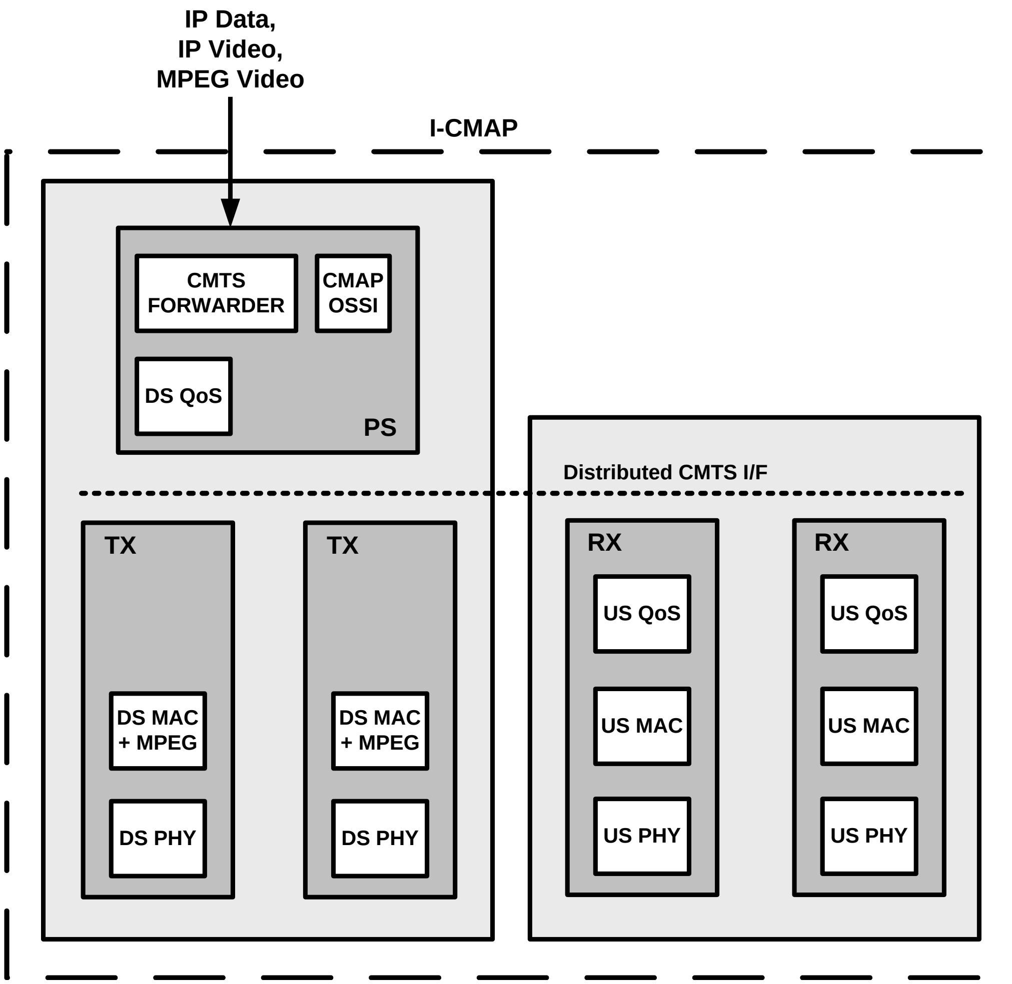


Figure 5 Distributed CMTS Architecture with Upstream Only

This partitioning of the architecture has a significant impact on the interconnection bandwidth required. In CMAP, the upstream bandwidth may be one tenth that required for the downstream. This means that the Distributed CMTS architecture in the above example would only need a single 10 Gbps link to support 80 upstream Service Groups. This compares to the 100Gbps plus redundancy that the M-CMAP approach requires as shown in the previous example.

Another important benefit of this Distributed CMTS Architecture is the reduction in overhead costs for redundancy. The mixed I-CMAP + AS system requires TX and RX redundancy in both boxes. For a 14-slot I-CMAP product, it would implement 5+1 redundancy for its upstream and downstream cards for 20% overhead.

The Distributed CMTS Architecture example above only requires TX redundancy in the primary chassis, and only requires RX redundancy in the subtended chassis. Using the 14-slot example, the primary chassis could implement 11+1 TX redundancy while the subtended chassis implements 11+1 RX redundancy. This is less than 10% redundancy overhead. Thus the Distributed CMTS Architecture has cut the redundancy overhead by more than half.

MIGRATION STRATEGIES

CMAP systems promise a disruptive change in the space/cost/power of future Head Ends. But to get these benefits implies a fork lift upgrade of existing EQAM and CMTS equipment. Since the operators will be investing in EQAM and CMTS equipment up until the day CMAP systems are delivered, a critical question is how can operators leverage existing equipment and transition to a full CMAP system?

The CMAP spec team has partially addressed this by stating that initial CMAP systems should be capable of operating as a UEQAM downstream only product. With this reduced functionality, it is hoped that products come to market quicker to meet existing needs for legacy video expansion of VOD and SDV. Then later, this product can be upgraded to a fully compliant CMAP product. This migration story can apply to either an I-CMAP device or an AS.

I-CMAP + M-CMTS

The above strategy assumes that vendors provide a chassis that operates as an UEQAM today and is upgradable to CMAP in the future. This ignores a very large segment of dense UEQAM that will be deployed before the availability of these new chassis systems. Another strategy would be to use M-CMTS concepts combined with I-CMAP to leverage existing UEQAM. This is shown in the figure 6 below.

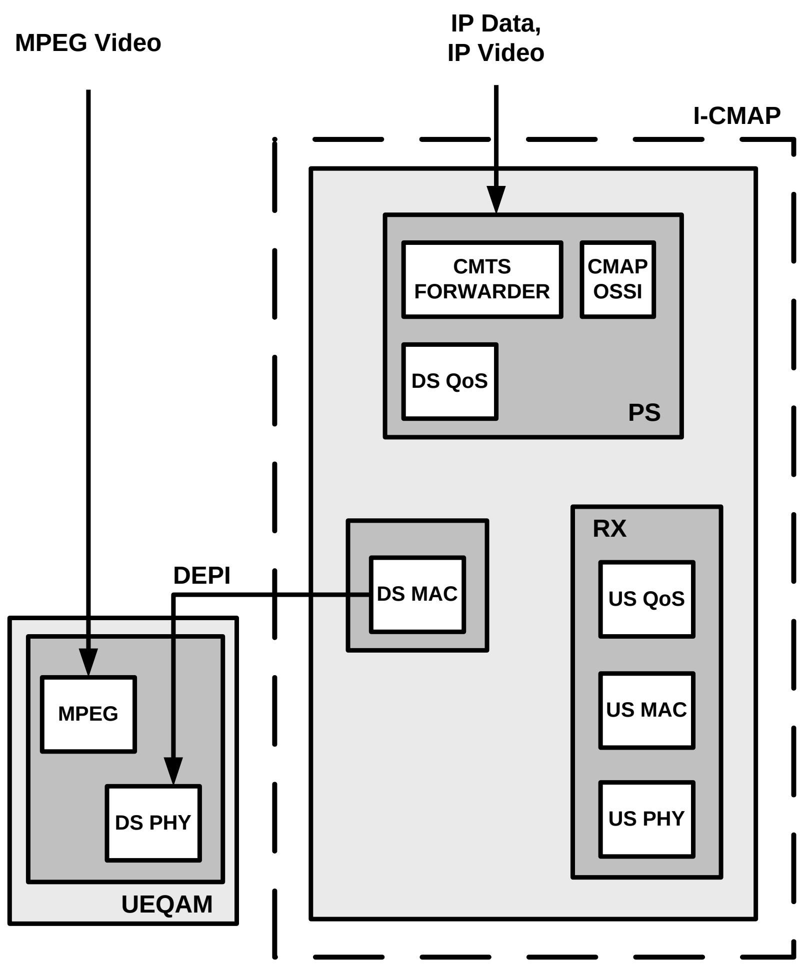


Figure 6 System with I-CMAP and M-CMTS

This M-CMTS + I-CMAP approach is not fully integrated from a configuration & management perspective. This is fine in the short term as legacy video is often managed as a separate silo from the data. But this could become a problem as operators start to converge into a single OSSI model.

In the above, note that the I-CMAP needs a downstream MAC component that then outputs over a DEPI interface to the UEQAM device. The other interesting aspect of the above split is that the legacy video (i.e. VOD and SDV) go directly to the UEQAM and bypass the PS. This seems a bit ironic since our early CMTS evolutions considered IP video bypass of the CMTS core. Now future evolutions could see legacy video bypass the CMAP PS.

Distributed CMTS Migration

A Distributed CMTS architecture can solve this problem in a different way. The UEQAM device may be enhanced to support the downstream MAC functions and become a TX component in the Distributed CMTS architecture. This is shown in figure 7.

By taking this approach, the enhanced UEQAM becomes a subtended chassis and the entire system is now managed as a single I-CMAP entity. This solves the long term migration problem to a single converged configuration and management system.

This approach is more cost effective by eliminating the need of the downstream MAC component from the primary chassis.

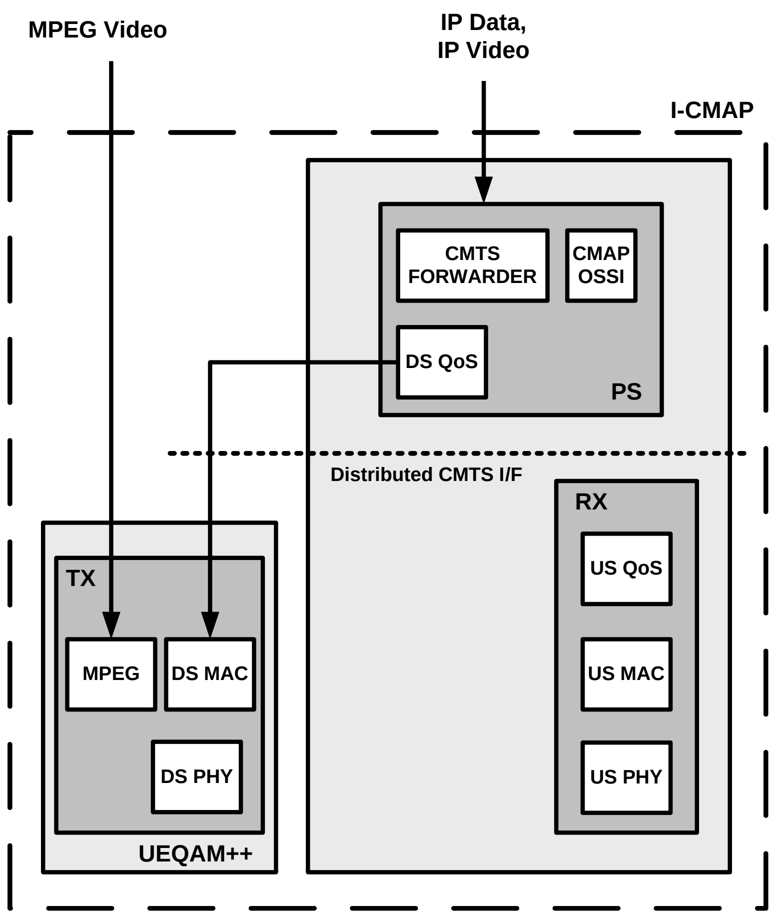


Figure 7 Distributed CMTS Architecture with Upstream Only

APPLICATION LEVEL SERVICES AND VIRTUAL ENVIRONMENTS

To date, CMTS systems have focused on providing data transport and offering L2 and L3 capabilities to service providers. This has enabled traditional data and voice services. As the CMTS evolution continues, we will start to see more capabilities added to these converged edge platforms. They will extend beyond data transport and begin to reach up to the application layer.

Offering basic data services to residential customers allows service providers limited opportunity for differentiation. In order to better compete, the provider should be able to offer additional services layered on top of the data transport. Examples of these include subscriber based services (rather than CM or STB based), business services, application acceleration and caching.

Application level services are typically delivered in a general purpose server environment rather than in an embedded platform. There are however situations when these services must be implemented in the embedded platform, such as when they need access to the data stream or need to have a packet by packet impact on how data should be forwarded.

Trying to determine ahead of time which application level features to embed in a system is to some extent an exercise in frustration. By the time the system goes into operation the applications have changed. In order to resolve this "catch 22" like situation the embedded platform needs to offer a development environment similar to that found on servers, so that functions can be developed quickly and added to the platform post deployment. This has not been practical in the past. Infrastructure systems must provide very high levels of availability and to preserve this have a restricted environment in which any hardware or software changes are the preserve of the system vendor and must be extensively tested prior to deployment.

Recent technology developments offer a potential resolution to this conundrum. The development of multi-core CPUs, hypervisors and virtual machine environments enable infrastructure equipment to offer an application programming environment which appears similar to a general purpose server but which can be isolated from the rest of the system. This can provide a sandbox in which application code can run but with the protection that any failures are limited and do not propagate through the rest of the system. Standard interfaces can similarly be used to enable the addition of hardware processing with filtering engines in the mainstream data forwarding path identifying and diverting packets for specialized processing.

While this appears to offer a utopian future in which equipment vendors, service providers and third party application developers can all add applications and even hardware to in service platforms there are of course several caveats. The applications are in the embedded platform because they need to influence the platform behavior to a greater or lesser extent. The equipment vendor must provide interfaces to enable this but must do it in a fashion so that they can be confident that operation will not be compromised for the rest of the platform.

This introduces the requirement for isolation between application level services and on any impact they have to the data and control planes of the system. The platform must evolve into a virtual environment where multiple service contexts (software and hardware) run in parallel with no knowledge of other contexts and more importantly with no access to them or impact on them. These service contexts must be capable of extending beyond the application sandbox into the rest of the system so that in effect a set of virtual systems is supported. References 7 & 8 describe academic and commercial work along these lines.

The potential value of an application service delivered in this manner can be illustrated by the following example. A cable MSO may offer WiFi hot spot service using DOCSIS cable modems (CM) for backhaul. The hot spots may be open to subscribers from two different wireless service providers (SP1 and SP2) who have negotiated different SLA's with the operator. Subscribers from SP1 may simply receive best effort service with a maximum rate limit while subscribers for SP2 receive an enhanced service set and higher rate limits.

To differentiate between the types of subscriber the system cannot use the CM address (as for conventional DOCSIS services) as it is a shared resource. Thus the subscriber must be identified with a wireless service provider and placed in the appropriate context. In the SP2 context the subscriber may then have access to services such as higher data rates, enhanced QoS, video caching or even local video processing hardware. This identification and service logic may be provided by SP2 and loaded into the converged edge platform.

BUSINESS SERVICES

Business services are a good example of the need to provide a set of virtual systems. Each business customer should have the impression that their service runs in its own virtual context and that this context is not vulnerable to disturbance from other business or residential contexts which share the common resources. For business service delivery the context is not simply a virtual application space but must include the control and forwarding planes. This can be delivered as VLANS, MPLS or L3 VPNs or as a full function virtual router depending on needs.

Next generation converged edge platforms may be deployed primarily as a residential service platform but must be capable of providing a range of business services as well. These services are typically delivered in compliance with well defined service level agreements (SLAs) and place additional requirements on the base platform. An SLA may define bounds on features such as downtime, packet loss, latency and jitter. The platform must not only deliver business services in compliance with these SLA's but must be able to monitor service delivery to confirm compliance. Thus to provide SLAs to each customer and continue to offer residential services the ability to provide virtual contexts and isolation between them is essential.

SERVICE CONTEXT ISOLATION

In order to maintain commercial SLA's the CMAP platform must isolate high value business traffic from the unpredictable loads created by best effort residential services. Ideally this isolation will be provided "end to end" and should be supported for both the data and control planes of the platform.

Data Plane

In the data plane, isolation is typically implemented with three components as the packet moves through the system. Packet classification on reception, internal buffering and queuing while it is stored and scheduling for transmission as it leaves the system.

Classification

Incoming packets must be classified to determine the service context in which they will be processed and the forwarding treatment they will receive in this context. The classification system must be flexible as core and regional networks use multiple options to segregate business traffic. These can vary from basic packet priority, through layer 2 virtual LANS to sophisticated MPLS and layer 3 virtual private networks (VPNs). The classification system needs to recognize each of these mechanisms in order to identify packets for priority treatment, mark them as such and hand them off to the queuing system.

Queuing & Buffering

The queuing system receives packets from the input classifiers and places them into buffers while they wait for transmission. The system is located at a point in the network where there is a very significant discontinuity in interface speeds. Fiber based interfaces to the regional networks operate at data rates which are three orders of magnitude greater than the HFC interfaces. Thus buffer space in the system must be large enough to accommodate packet bursts from the regional networks which can overwhelm HFC capacity. At busy times the buffers may be filled with a backlog of business service and best effort traffic. In order to maintain the SLA when buffers overflow low value best effort packets should be discarded rather than high value business service packets. Overflow in one service context must be contained within that context and not impact other contexts operating in parallel.

Scheduling

When there is backlog in the system packets must be scheduled for transmission based on their "value" but while continuing to provide a reasonable level of residential service. Business traffic should have priority but not cause starvation for residential services. In order to achieve this, a relatively sophisticated scheduler is required for both upstream and downstream traffic, which should be capable of arbitration between services contexts and service differentiation within a context.

Control Plane

The previous section describes the operations required in the data plane to support business services. The control plane must provide the mechanisms to configure and manage these operations. This requires both relatively static provisioning mechanisms such as configuration files and command line interfaces and more dynamic mechanisms such as routing and session control protocols to establish and maintain virtual private networks. Just as the data plane is required to isolate business service traffic the control plane should provide similar isolation. Thus a system may provide multiple control plane contexts. Each context has its own static and dynamic control plane mechanisms so that the control plane for each major business service may be self contained. This avoids problems such as configuration errors from propagating between independent business service offerings.

SLA VALIDATION

To successfully provide business services an operator must not only be able to deliver them according to the SLA, but must also be able to measure this delivery and be able to provide metrics showing how the service is operating. These metrics are needed both for internal use by operations staff and also potentially for delivery to the customer in order to confirm service delivery. Thus the converged edge platform must measure, store and export a number of variables on a per customer basis. These will typically include average and burst data rates, service outages and service impacting errors such as packet loss. They will need to be measured in granular time intervals and exported periodically using a mechanism such as IPDR for off line compilation and analysis. As with data and control plane operations the monitoring and reporting must be cognizant of the service contexts.

UPSTREAM BANDWIDTH

Traffic for residential services is dominated by multimedia downloads and as a result is very asymmetric. Business services are typically much more symmetrical and require higher upstream data rates. Thus service providers need to offer higher upstream peak rates to business customers but more significantly must be prepared to support higher average upstream data rates.

All versions of DOCSIS since 1.1 provide the capability to isolate residential and business services (by defining service flows with specific QoS which can be scheduled independently). This ability to schedule the upstream traffic independently is only useful if sufficient bandwidth is available for both residential and business customers.

Fortunately current DOCSIS 3.0 systems have the potential to operate up to 85MHz. Using the full 5-85MHz spectrum along with the DOCSIS S-CDMA capabilities can enable up to 400 Mbps of upstream bandwidth once the operator needs it. Existing residential services can stay below 42MHz while multiple new 100 Mbps services can be offered in the 40 to 85MHz band. CMAP will enable future downstream rates approaching 1Gbps. These downstream rates will require increased upstream burst rates on the order of 200 to 300Mbps. Next generation CMAP devices must be capable of extracting all possible upstream bandwidth to meet these demands.

MULTIPLE ACCESS NETWORKS

Not all business services will be delivered over DOCSIS. Ethernet and EPON based access networks will be used to offer data rates beyond those practical in the DOCSIS network so that a range of business services can be offered using multiple technologies.

EPON is a point to multipoint technology providing similar services and operational issues to DOCSIS. In order to simplify the operation and provisioning of EPON services for cable operators, the DOCSIS Provisioning of EPON (DPoE) standard [reference 9] has been developed. With this in place integration of EPON into a CMAP like platform is a practical proposition.

While existing EPON deployments are typically based on a 1Gbps version, newer 10Gbps PON technology is quickly becoming feasible and cost effective. For many service providers, it will make sense to go directly to 10G PON services and obtaining a significant competitive advantage. Integrating 10G PON capability into a CMAP system will become a key differentiator in the near future.

Point to point Ethernet will continue to be used for the highest business customer tier. Next generation converged edge platforms can easily support Ethernet interfaces and can offer the same set of enhanced application level services over Ethernet as well as DOCSIS and EPON networks.

SUMMARY

Over the years, we have witnessed the evolution of CMTS architectures as services were added and bandwidth needs increased. From this, we saw the introduction of M-CMTS and competing I-CMTS systems. The move to IP video over cable will be the next major driver of CMTS evolution. It will drive systems to increase downstream port density and reduce costs per downstream channel.

The CMAP architecture is the next major step in this CMTS evolution and has widespread support from multiple operators and vendors. It is currently being standardized through CableLabs. CMAP supports both integrated and modular solutions. Either mode is capable of supporting IP video. Operators will have plenty of choice to select vendors in either camp.

However there are significant tradeoffs between the modular and integrated solutions which operators will need to consider carefully. The areas that need careful consideration include:

- Expanded Vendor Ecosystem
- Best of Breed components
- Operational complexities
- Scaling from small remote hub sites to large Head Ends
- Risk and Time to Market
- IP Video implications
- Migration Strategies
- Rate of Innovation

CMAP systems promise a disruptive change in the space/cost/power of future Head Ends. But to get these benefits implies a fork lift upgrade of existing EQAM and CMTS equipment. Since the operators will be investing in EQAM and CMTS equipment right up until the day CMAP systems are delivered, a critical question is how can the operator leverage existing equipment and transition to a full CMAP system? This will include solutions such as Access Shelves and I-CMAP that can initially operate as UEQAM and integration of M-CMTS concepts with integrated CMAP.

Will I-CMAP and M-CMAP be the end of the DOCSIS CMTS Architecture evolution? Definitely not. We envision that there will be a blending of the two systems where I-CMAP devices act as the PS for other AS. This could give the operators the best of both worlds.

Beyond that, innovation will keep moving forward and both I-CMAP and M-CMAP will morph to adapt. We gave one potential example with a Distributed CMTS architecture that could allow the downstream TX component to be separated and scaled independent from the upstream RX component.

Past this, these new converged edge platforms will need to start integrating application level services to empower the service providers and keep them competitive. This will include expanded business services and expansion into new access technologies, including wireless.

In any respect, this is an exciting time for the cable broadband industry as we begin a new seismic shift.

REFERENCES

[1] M. Patrick, J. Joyce, *"DIBA - DOCSIS IPTV Bypass Architecture"*, SCTE Conference on Emerging Technology, 2007.

[2] J. Salinger, *"Proposed Next Generation Cable Access Network Architecture"*, SCTE Conference on Emerging Technology, 2009.

[3] J. Salinger, *"Understanding and Planning CMAP Network Design and Operations"*, SCTE Cable-Tec Expo, 2010.

[4] J. Finkelstein, J. Salinger, *"IP Video Delivery using Converged Multi-Service Access Platform (CMAP)"*, SCTE Canadian Summit, 2011..

[5] J. Ulm, P. Maurer, *"IP Video Guide – Avoiding Pot Holes on the Cable IPTV Highway"*, SCTE Cable-Tec Expo, 2009.

[6] J. Ulm, G. White, *"Evolving Architectures for Cable IP Video Delivery"*, SCTE Canadian Summit, 2011.

[7] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner.OpenFlow: enabling innovation in campus networks, ACM SIGCOMM Computer Communication Review, 38(2):69–74, April 2008.

[8] OpenFlow switch Specification, www.Openflow.org

[9] DPoE™ Architecture Specificatio, www.cablelabs.com DPoE-SP-ARCHv1.0-I01-110225

PLUGGABLE OPTICS FOR HFC

Fernando Villarruel, Fred Hirt, and Kuang-Yi Wu
Cisco

Abstract

Our time is that of an insatiable appetite for fresh, new information, taking and giving, alone and in unison; from grandma who Skypes with the grandkids, to the boy that has never known linear television, to the masses who can Tweet a country to political change in a matter of days, and everyone else on Facebook.

Access platforms are built for their time and the successful ones are a complement of technologies that work towards a particular goal. So what is the access platform of our time? What does it do? How does it scale? What key technologies does it integrate to facilitate the services the cable industry aims to provide?

In this paper we propose an optical RF transmitter that allows integration of HFC optics into next generation CMTS platforms facilitating all-IP transmission and allowing power and footprint reductions through the use of 10 Gbps optics, such as is found in XFP packages.

INTRODUCTION

Cable operators are actively creating road-maps toward end-to-end all-Internet protocol (IP) service functionality. Simultaneously, recent technical developments have resulted in various schemes for access delivery via IP video to complement existing Data over Cable Service Interface Specification (DOCSIS). Yet, while there is agreement on IP video delivery as a goal for access transport, there is no broad agreement on the specifics of the access plant hardware required to accommodate IP video delivery *without discarding major portions of the hybrid fiber coax (HFC) network*. Thus, a greatly desired part of any service transition is key technology advances that allow an increase in capacity and performance while maintaining as much of the sunken investment, particularly in the outside plant.

In this paper, as a key part of next generation access platforms, we propose a simple, cost-effective, power and footprint efficient method to enable IP traffic *utilizing pluggable RF optics*. The RF pluggable optics we propose maintain the cable operator's significant investment, from fiber spans, through the node and into the RF plant. This is in contrast to Ethernet-over-coax techniques or schemes to remote the cable modem termination system (CMTS) or RF gateway physical-layer hardware out to the node, all of which significantly impact the nodes and RF plant.

Particularly, the technology advance we propose is integrating the HFC optical transmission into higher-layer platforms, such as Comcast's Converged Multiservice Access Platform (CMAP) or Time Warner Cable's Converged Edge Services Access Router (CESAR) via RF-modulated small-form-factor pluggable optic modules. We focus our attention here on the forward path, but we also advocate integrating the reverse path optics as well, leaving that specific treatment for another opportunity.

PLUGGABLE OPTICS

The concept of pluggable RF optics logically follows from an understanding of the two demarcation points that define HFC access architectures. These are the transition point from baseband digital content transmission to RF modulated transmission

[RF gateway or cable modem termination system (CMTS)] and the transition point at which fiber ends and coaxial cable transmission begins (typically the optical node). We will look these two demarcation points from the perspective of an access network designed for end-to-end IP video transport.

Our discussion includes the architectures expected and technology specific to modulation formats. Within this broader discussion we will answer questions about the viability of our proposition such as,

- Are pluggable RF optics feasible, can they even be built?
- Under what circumstances can they be produced?
- Would other network components be required to change?
- Can pluggable RF optics be fabricated cost-effectively?
- Is a standard attainable, and if so how?
- Are there line power utilization benefits, i.e. is it green?

We propose answers to these and other questions to paint a clear and undisputable picture of pluggable RF optics as the rapid-deployment, cost-effective means of achieving end-to-end IP video delivery.

HFC AND IP

Transmission formats create boundaries and opportunities. In the case of HFC, its strength and flexibility is that it leverages both frequency division multiplexing (FDM) via the RF spectrum and simultaneously time division multiplexing (TDM) via DOCSIS. This unique combination has allowed it to scale from very basic services including broadcast analog transmission, to narrowcast video services, high-speed data, and voice over Internet protocol (VoIP) telephony, without any fundamental changes to methodology of transmission. Now in the wake of an IP services boom, it merits re-evaluating if this FDM/TDM combination is still useful and practical.

In the last few years the capacity of HFC architectures has increased significantly with a migration to large numbers quadrature amplitude modulation (QAM) channels. This migration creates a scenario where the capacity of an all-QAM signal lineup can be competitive with that of any other architecture, even FTTH [1]. Specifically, with a usable data rate of about 38 Mbps per 6 MHz bandwidth 256-QAM channel, the RF spectrum in the forward path as a whole can very easily grow to an aggregate 5.8 Gbps. Nevertheless, it is not only raw capacity, but the simultaneous use of spectrum partitioning and timed availability that multiplies HFC's effectiveness in comparison to other TDM- or FDM-only applications.

A conservative future example is that of 200 homes sharing a full all-QAM forward path spectrum for an all-IP service offering. In this example the leveraging of multiple bonding groups within the RF spectrum and including bandwidth accommodation for reverse path growth via a mid-split segmentation easily allow a very competitive transmission rate of 1 Gbps downstream and 100 Mbps upstream [2]. Most importantly, an IP deployment of this sort has the ultimate benefit that the node transition point and function remains. Thus, when evaluating what to keep and what to leave behind in IP architectures, it is hard to put aside the current combination of RF modulation and DOCSIS.

The previous example and others showing migration to smaller service groups, in conjunction with maintaining RF transmission [3], are in line with optical segmentation techniques deployed during the last few years.

Scaling for future optical wavelengths in service could result in anywhere from four to eight times as many as there are now, leaving on the table the very real and pertinent question of footprint and power availability for projected new hardware. This is a fundamental, practical question that equipment providers must answer as cable operators migrate to all-IP networks.

COMPLEX RF MODULATION (CRM)

In order to achieve maximum bandwidth efficiency in the physical transport layer, high-order (64 through 1024) QAM transport is required. We refer to such *all-QAM signaling as complex RF modulation (CRM)*, as distinct from that of traditional analog video (NTSC, PAL, etc.), quasi-constant-envelope digital signaling (QPSK, O-QPSK, etc.), base-band digital transmission techniques (e.g. OC-192, 10 G, etc.) or mixes thereof. This prevents ambiguities associated with the more general terms "analog" or "digital" transport, which can variously refer to amplitude-modulated vestigial sideband (AM-VSB), various orders of QAM transmission, or the baseband digital format of some digital return path, backbone, metro, and cable's transnational links.

CRM loadings are fundamentally different from the mixed AM-VSB/QAM loads which comprise the majority of current deployments. Intuitively, one expects a uniform CRM loading to be "easier" to transport than mixed analog/QAM content, with CRM yielding more robust signaling, and greater noise tolerance. All of this leading to greater link budget. This is true, but what makes it so?

Access Link Performance Requirements at the Node Output for Existing & Future Access Payload Modulation Schemes		
Performance Parameter [1,2]	**Existing (Analog/QAM)** **• 78 Analog Carriers** **• 75 Carriers, 256-QAM**	**Future (CRM)** **• 153 Carriers of All 256-QAM** [2]
CNR (dBc)	> 50	>40
CSO (dBc)	< 63	< 55
CTB (dBc)	< 63	< 55
MER (dB) [3]	> 37	> 37
BER, Pre-FEC [3]	$< 10^{-9}$	$< 10^{-9}$
BER, Post-FEC [3]	$< 10^{-12}$	$< 10^{-12}$

[1] Analog measurements according to ANSI/SCTE 06 2009, ANSI/SCTE 17 2007

[2] 153 QAM carriers in continuous-wave (CW) mode to measure CNR, CSO, CTB

[3] Equalized QAM, measurements of ANSI/SCTE 121 2006, ITU J.83 Annex B source.

Table 1− Performance Requirements: Existing and Future Access Payloads

Transmission Characteristics for a CRM Payload

To understand we begin by examining fundamental access link performance parameters to see what must be maintained and what can be relaxed when transitioning from analog-rich to analog-free link payloads. Table 1 details performance parameters expected for a current access optical link, as measured at the HFC node. It compares performance for a mixed modulation loading of 75 AM-VSB channels with 75 channels of 256-QAM (representing an existing case for many access networks) to a load consisting of 153 channels of 256-QAM.

The network performance differences between the access loading schemes shown in Table 1 are both subtle and significant. Of particular note are the differences in the carrier-to-noise ratio (CNR), composite second order (CSO), and composite triple beat (CTB) values required for unimpaired transmission in each case. Let us examine the details relating these parameters in order to differentiate between the requirements for existing, mixed analog/QAM HFC access networks and an end-to-end IP video network employing CRM transmission for access.

First, in order to measure "analog" parameters such as CNR/CSO/CTB for a 153 channel 256-QAM CRM load, all QAM modulators used during measurement must be set to continuous wave (CW) operation. Further, such CW level must be calibrated at a level corresponding to modulated carriers yielding a minimum 37 dB equalized modulation error ratio (MER) for the 153 channel QAM load. Although obvious, this is necessary in order to differentiate among linear and nonlinear impairment mechanisms which result in the noise component of the QAM MER. Essentially analog measurements are used to give a more detailed description of the mechanisms responsible for impairing (or limiting) the QAM MER values of an access link.

Second, in general it is understood that analog parameters are necessary, but not sufficient, to yield robust MER values. This is due to the fact that they do not fully account for the effects of phase noise or "quasi-phase noise"-like effects. Thus, there exist instances in which phase noise components greatly determine the QAM MER performance, particularly in the case of very high carrier-to-composite noise (CCN) and CNR values. Such cases occur frequently in mixed analog/ QAM links when large (> -40 dBc) analog distortion products fall near or under a QAM carrier and are resolved by the customer premises equipment's (CPE's) demodulator as non-coherent single frequency components. This quasi-phase noise degrades modulation recovery, thus reducing MER. *This impairment is unique to mixed analog-QAM transmission*, due entirely to the high energy analog carriers producing discrete distortion products. It is important to note that such effects do not exist in all-QAM CRM transmission.

Thus, robust optical links themselves do not contribute noticeable phase noise to QAM signals. Any residual phase and delay impairments beyond the access optical link, due to RF impedance mismatches for example, are adequately compensated for by the QAM receiver's adaptive equalizer.

A third point is that the optics be *approximately* linear for amplitude and phase transmission. This implies no clipping and no compression in transmission along with the avoidance of excessive, variable timing delays. That is, no variable delays on the order of multiple milliseconds, as is the case in route redundancy switching between greatly differing time-of-flight routes. Such changes adversely affect upstream ranging (for CMTS) and downstream latency in VoIP applications.

In the case of clipping and distortion, such issues are routinely dealt with in proper optical transmitter design and calibration. In the case of redundant link delays, they can be accommodated by approximately-matched delays in the redundant link layout. Ultimately, well-designed optical link delays are limited by dispersion, a sub-nanosecond phenomenon which does not contribute significant phase noise to QAM signals at access optics link lengths (sub-100 km).

In addition to the hardware and link considerations, a final point in understanding the performance of CRM signals over HFC access optical links is the nature of the nonlinear components generated. In mixed analog and QAM access transport links, impairments consist of noise, discrete distortion products due to nonlinearities such as CSO and CTB, as well as optical RF crosstalk and beating effects. As previously mentioned, such discrete distortion products can lead to tones lying near or under QAM carriers, which result in degraded MER, while still exhibiting low noise and excellent CNR.

In contrast, a CRM payload's nonlinear impairments manifest themselves as Gaussian noise-like components. That is, the second and third order products are noise-like rather than clusters of composite beats, and *can be considered additions to the noise floor under a QAM carrier*. Further, this principle also extends to multi-wavelength crosstalk components such as optical cross-phase modulation (XPM) and four-wave mixing (FWM) [4].

In summary, a CRM loading offers a new set of choices for the network designer due to the more forgiving nature of the distortion impairments appearing as Gaussian noise-like components. End-to-end IP functionality benefits from both the increase bandwidth efficiency of high order, all-QAM CRM transmission, as well as the relaxed transmission requirements as compared to that of mixed analog/QAM channel loads, shown in Table 1. IP video transport takes maximum advantage of a CRM loading which does not suffer from the out of band discrete distortion beats created by analog channels and exacerbated as drive levels reach non-linear peaking or compression. In mixed loads, those beat clusters falling near or in QAM channels stress decision boundaries and are ultimately problematic for demodulation routines to withstand and correct [5, 6.]

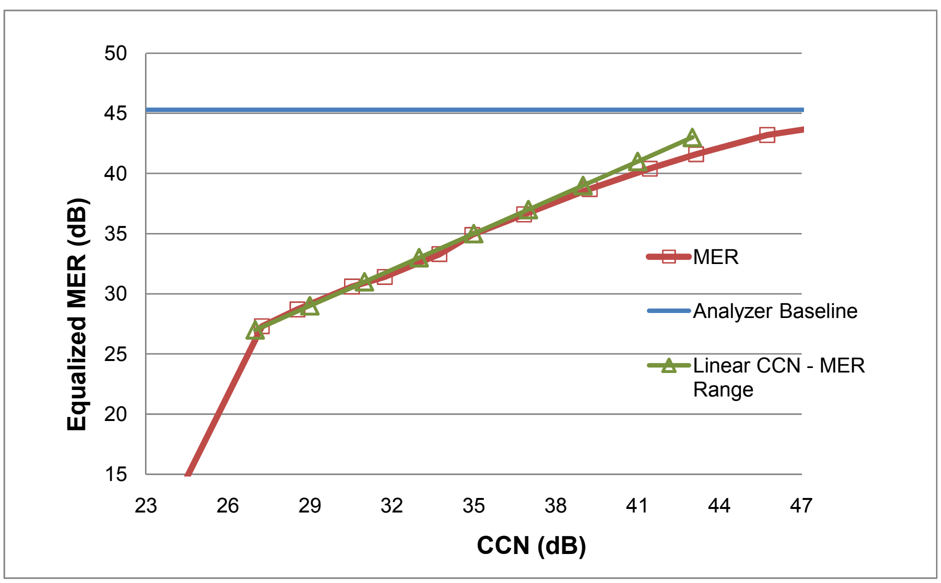


Figure 1 – Equalized MER as a Function of Carrier to Composite Noise (CCN) Ratio

<u>Exploiting the CRM Advantage in DOCSIS Data and IP Video Access Links</u>

The Gaussian noise-like nature of distortion products generated in an all-QAM access link allows the CCN of a CRM loading *to be near-linearly related to the MER*, see Figure 1. In this plot, the CCN and equalized MER are seen to have a linear relationship within the operating range of a Rhode +

Schwarz EFA QAM signal analyzer. The CCN varies linearly form a 28 dB CCN lower analyzer acquisition locking limit (for 256-QAM), to a 42 dB CCN upper limit due to the analyzer's 46 dB maximum MER measurement capability. This observation of a linear relationship between CCN and MER leads to the following two advantages of an end-to-end IP access network based upon all-QAM, CRM transmission.

The first is that the linearity necessary for unimpaired optical transmission is *greatly* decreased when dealing with CRM network payloads, specifically relaxing CSO and CTB for both optical and RF domains, as well as relaxing XPM and FWM requirements in the optical domain. Specifically, CNR is reduced by nearly 10 dB, with CSO and CTB reduced by 8 dB. This opens a number of possibilities for new links and radically different RF transmitter designs, *greatly* simplifying the design, manufacturing and tuning challenges for CRM optical transmitters over those of their mixed-payload analog predecessors. Pluggable RF optics leverages these differences to reduce optical transmitter size and power dissipation while tightening its integration within the headend or hub's content-generation hardware. These reductions greatly reduce shelf space and power dissipation while improving ease of use and sparing issues.

The second benefit is that the linear relationship between CNR/CSO/CTB along with optical cross-talk (XPM and FWM), and MER (recalling that nonlinearities map into CCN which is linearly related to MER) allows the use of the same design approaches traditionally used to make hardware decisions for access hardware links, with the advantage of relaxed noise and distortion goals. That yields positives all the way from design and manufacturing, to deployment and turn-up. From the selection of components used in the transmitters to their manufacturing and tuning, to the link designer, headend tech, node techs and maintenance staff, the CRM transmitter relies on well understood, tested transmitter technology. Despite being end-to-end digital, with DOCSIS data and IP video delivery traffic, networks utilizing 256-QAM designs can target specific CNR, CSO and CTB goals, with the expectation of well-defined QAM MER performance.

One caution regarding the data shown in Table 1 must be raised. It might be implied that if a heavily-analog channel loading does not permit the advantages of a CRM payload, then half as many analog channels would "spilt the difference" and be a good compromise. The reality is that any inclusion of analog channels has the negative effect of producing *discrete* clusters of distortions which can and do peak within the bandwidth of a QAM channel, degrading overall performance. The degradation that occurs from the presence of analog distortion products affects MER in a fashion *which CNR and CCN will not reveal*.

It is necessary that next generation hardware employ all-QAM, not mostly-QAM, transmission in order to fully leverage the benefits which CRM loading brings over mixed analog/QAM loadings. As a practical matter, provision can be made for the addition of two or three service tones, spread throughout the operating bandwidth, without degrading the CRM signal. In summary, CRM payloads allow simplified, straightforward design rules for access links, which can be exploited to greatly improve access transmitters for end-to-end IP video delivery.

Link Considerations

Another benefit of CRM is that it can increase the link budget, the exact amount of which depends upon link parameters such as equivalent Optical Modulation Index (OMI) per CW channel and the desired RF output at the node. Switching to CRM can result in nearly a 3 dB reduction in optical input power to the node receiver since in current networks the QAM channel powers are already 6 dB (RF) lower than the AM-VSB channels. Note that the full 3 dB optical power decrease is not *always* achievable by loading change alone, since the 153 channel QAM load is not being compared to a 153 analog load but rather to a mixed analog/QAM loading.

A further reduction to receiver input power can be made, however, if the QAM loading utilized exhibits a high CCN. In such cases the receiver input can be reduced to the point where the shot and thermal noise components of the receiver dominate at this lower input power. Such a link budget improvement can be used to lower launch power and so lower the non-linear dynamics occurring in the fiber. Since fiber nonlinearities are launch-power-dependent even a 2 dB launch power reduction can yield a significant reduction in crosstalk and four-wave mixing [7].

In cases where the operator has some leeway in accounting for RF power, throughout the RF chain to the home via unused amplification potential or node segmentation, the optical input power into the receiver can be reduced by more than 3 dB, down to -10 dBm or lower depending on several performance factors.

The implication of higher optical link budgets also creates ample space for the reduction of stimulated Brillouin scattering (SBS) suppression, which has been one of the daunting challenges for cable optical transmitters since their inception. SBS is a scattering effect that takes place in fiber when the launch power of a wavelength is approximately greater than 7 dBm. Special and very complicated circuitry has been created to compensate for this issue

IS PLUGGABLE RF OPTICS A REALITY?

So the question remains, can small form factor pluggable transmitters be a part of the HFC landscape moving forward? The answer is a resounding yes. The upshot of reduced linearity requirements for all-QAM channel loads is that it creates a potential tangible shift from the hardware employed to make legacy cable optical transmitters, to making future IP ready transmitters. In particular, there are new opportunities in the mix of components that can be used to reach the desired performance values, in addition to reduced size and power consumption.

For many years now the ability to make cable transmitters has been determined by a few key factors. In the case of directly modulated transmitters (DMTx), their lasers have had to have a minimum necessary linearity and stability dependent both on the growth characteristic and packaging structure, ultimately creating a specific pool of usable lasers and a size threshold for the optical package, only relevant to the cable space. While some deviation has come from the typical butterfly “analog” laser package in the last few years, the gains have been minimal. Also, legacy DMTx have the necessity for electronic harmonic distortion correction, both for residual CSO and CTB from the analog laser, and for fiber induced CSO, ultimately the extent to which these corrections are utilized also creates a power consumption and size threshold in the electronics.

In the case of externally modulated transmitters (EMTx), where they are highly desired for their low noise capability and lack of high CSO accumulation over fiber, their size, power draw and price has typically made them unattractive in comparison to DMTx. For an EMTx, a high power CW laser, external modulator and SBS suppression circuitry typically define the size and power consumption.

The ability to simplify the technical requirements for CRM allows for the concept of creating a small form factor pluggable transmitter. There are two key requirements: the first is finding the smallest possible optical component packages that would be able to meet CRM linearity requirements and the second is to collapse the new necessary RF electronics into integrated circuitry. These two steps create a framework under which one could envision smaller packaged transmitters.

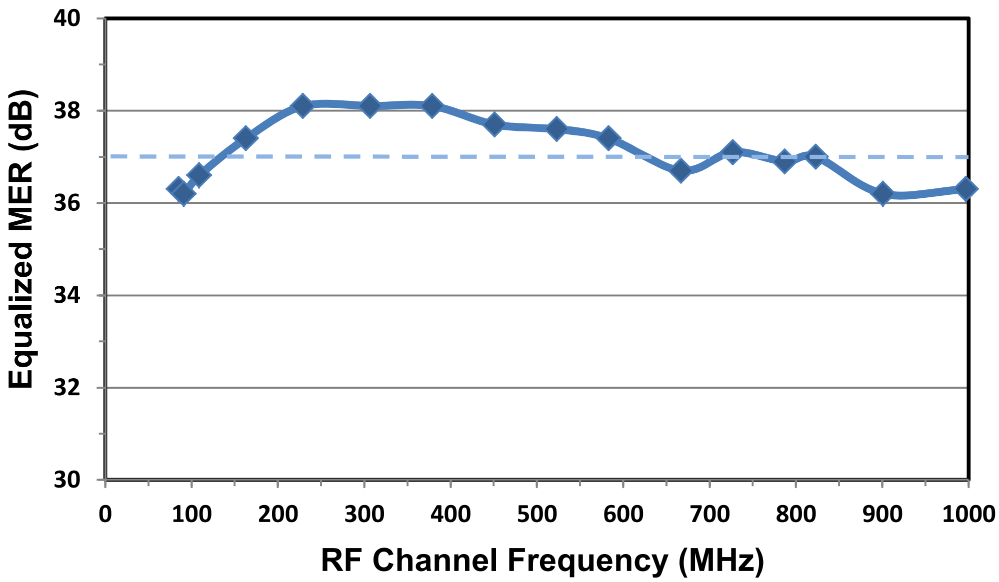


Figure 2 – Experimental Results for a Rudimentary, Proof-of-Concept XFP Transmitter

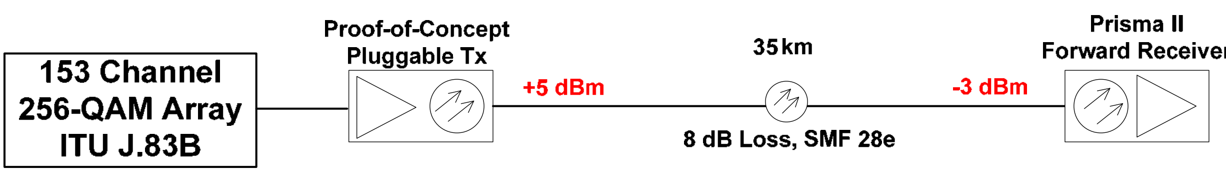


Figure 3 − Optical Test Link

Proof of Concept

Figure 2 shows the equalized MER performance of a rudimentary proof of concept transmitter whose optical package and pertinent RF electronics can fit comfortably in a 10 Gigabit small form factor pluggable XFP package. Figure 3 depicts a representation of the test link set up, where the proof-of-concept transmitter with an output of 5 dBm is followed by 35 km of fiber into a Prisma II forward receiver with an input power of -3 dBm. The input channel loading was 153, 256-QAM, ITU-T J.83 Annex B channels. The channel loading spanned 1 GHz to 82 MHz, leaving room for expected growth in the return path. We expect that integration and optimization of components will improve both the performance and reach.

A Standards-Based Pluggable Transmitter: An XFP Form Factor CRM Transmitter for the Cable Space

A question arises as to what would constitute an acceptable small form factor package, keeping in mind that the cable industry wishes to adopt, where possible, industry standards for form, fit and function without hampering the competitive aspects of functionality and added value. In the baseband digital space there are various multi-source agreements (MSAs) for small form factor transceiver packages. For example, XENPAK, 10 Gbps small form factor pluggable (XFP), small form factor pluggable (SFP), and small form factor pluggable 10 Gbps (SFP^+) are all popular such multi-source agreements.

MSAs are by definition industry-accepted by producers and users; they have been in place for some time, benefitting from mature components and predictable cost reduction curves. Note that the majority of the MSA documentation is focused on interface standards such as powering, signaling, monitoring, DC power limits and thermal dissipation, physical outline and mechanical specifications. It is absolutely conceivable that the cable space can leverage such an MSA interface, while internally maintaining functionality specific to CRM operation.

As an example of such a usable standard see Figure 4 which details the physical connector interface for a 10 Gbps XFP, as defined in the XFP MSA, [8.] The XFP specification, as currently defined, gives a CRM transmitter the opportunity to use several already existing, industry-standard interface lines (those NOT highlighted in a red circle) for powering, communication, control, and modulation inputs. For example, the differential input signal interface specified for the XFP (TD^+ and TF- on pins 29 and 28) lend themselves well to RF QAM input. Various powering options exist, including 1.8 V (Vcc2), 3.3 V (Vcc3), and 5 V (Vcc5), with nine ground pins specified for excellent RF and DC connectivity.

Module control is equally straightforward with an industry-standard two-wire serial interface (serial clock, SCL, and serial data, SDA), along with interrupt, module de-select, power-down, module numbering and presence detection. Communication protocols to the pluggable module are called out in the XFP specification, along with allowable DC dissipation limits for each power line and for the module as whole. One difference between the XFP as specified in the MSA and the transmitter under discussion here is that we refer to a transmitter, not a transceiver. Only the forward path transmitter will be present; the return path can be located separately. This releases 5 pins specific to the receive portion of a transceiver, received data (RD+, RD-), reference clock (REFCLK+, REFCLK-), and receiver loss of signal (RX_LOS), further easing host PCB layout.

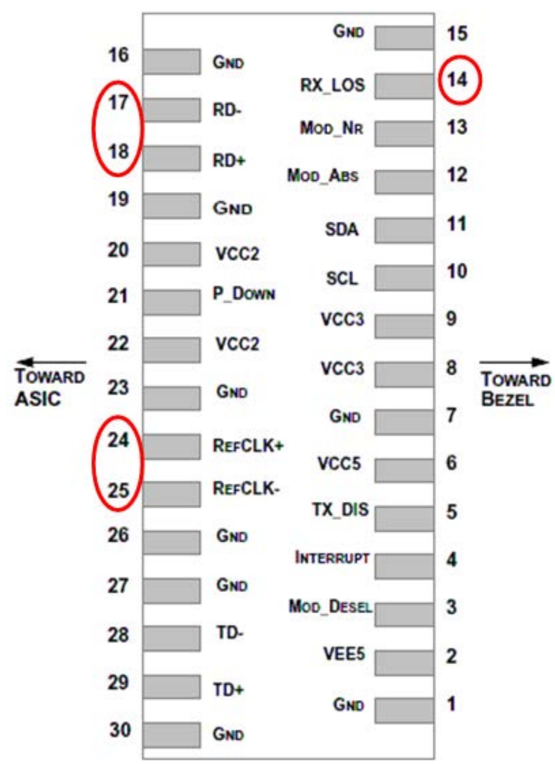


Figure 4 –XFP Interface: MSA-Defined Specification

Transmitter Functionality Options

Regarding the operation of small-form-factor transmitters, there are two options that have been expressed so far. The first is where the functionality of the pluggable transmitter is totally self-contained. This means that the burden of linearity for the transmitter is internal and independent of the incoming signal. The only interaction with its host would be for powering and status monitoring. As is the case now with most, if not all, contemporary HFC platforms, where an RF broadband signal comes in to a transmitter, it is analyzed and corrected for distortions and optically modulated to exit. The second option is that the transmitter linearity is not self contained, and is dependent on an input signal that has already been analyzed and corrected for distortions, leaving the pluggable transmitter to have only the function of optical modulation [9].

While both options are technically feasible and have their advantages and disadvantages, we prefer the self-contained transmitter approach for various reasons: It allows for a potential standardization based purely on interfaces and not tying performance parameters across active components, beyond what exists already at the RF level, such as the DOCSIS Downstream Radio Frequency Interface Specification (DRFI). It frees potential higher layer host platforms from having to carry calibration data for pluggable optics. And finally it is fundamentally important that host platforms not be transmitter-specific, that is, DMTx, EMTx, 1310 nm, DWDM, etc. Conversely, the option that it is not self-contained could possible lead to a better cost structure up front, though not including deployment issues involving cross vendor reliability and traceability of faults when failures occur.

IT'S EASY BEING GREEN

The XFP MSA specification [8,] defines the maximum DC power dissipation levels for the XFP package. The DC dissipation for a Power Level 3 device is specified at not more than 3.5 W per module. *We propose adopting this dissipation limit for RF optic pluggable transmitters*. At a 3.5 W maximum DC dissipation per optical transmitter, the DC power dissipation reduction from current, state-of-the-art transmitters which consume anywhere from 7 – 15 W, is a very green ***50 to 75% per transmitter***.

As large as the power consumption reduction is, it still does not completely capture the power savings due to status monitoring and control being directly absorbed into the content generation hardware, CMTS, or RF gateway. No external processors or aggregators are necessary to control the pluggable XFP, outside the content generation hardware, whether it be a CMAP box, a CESAR router, or a CMTS. Optics are now truly part of the "smart box" with high-speed backbone inputs and access optical outputs.

Existing headend architectures still create channel lineups *manually* via lossy RF combiner structures which aggregate content generation outputs before applying them to the access optics. As content generation evolves, generation devices which synthesize entire channel lineups via direct digital synthesis (DDS) are becoming available. Such devices can or will generate 135 to 153 channels per port for direct transmission to the node. Such "full-spectrum" port synthesis creates an additional opportunity for a significant power savings within the content generation boxes themselves, *on the order of 3-5 watts per port*, by reducing the port's output levels generated when attached to a pluggable optical module.

Programmable port power can reduce some >350 W per 96-downstream signals per port device, just by leveraging the lower output levels required of an RF XFP directly driven by a content generator's output port. Under the tightly integrated control of the content generator, the XFP also becomes part of an "agile" channel lineup, allowing sparing, redundancy, and idling of unused functionality to optimize power consumption versus bandwidth requirements. Whether by itself or integrated into the content generator, an XFP-based RF optical transmitter can save between 5 to 12+ watts per port in DC power dissipation. Considering an even optimistic 85% off-the-wall efficiency improvement yields a 6 to 15+ watts PER PORT savings.

CONCLUSIONS

We have shown that pluggable RF optics within a next generation access IP platform are possible, without triggering changes to the outside plant infrastructure. Use of an all-QAM complex RF modulation payload allows simplified, relaxed design rules in the access link to the customer. We use this to create a new means to recapture space, power, and cost by the use of specification-based pluggable optical transmitters. Pluggable RF XFP optical transmitters move the electrical to optical transition from a separate chassis to an integral part of the IP platform, tightening network control, lowering total power requirements, and saving rack space.

Acknowledgement

We would like to extend our gratitude to John Chapman, Ron Hranac, Steve Condra, Hisham Saad and Dan Ott of Cisco, for their support and helpful discussions.

References

[1] Chris Bowick, *"The Winning Network,"* Multichannel News, November 1, 2010.

[2] Steve Condra, Discussion, Director of Network Architecture, Cisco.

[3] Arris Group Inc., *"IP Video…An Architecture in Transition,"* Presented to the Greater Chicago SCTE Chapter, May 2010, http://www.gcscte.org/presentations/2010/IP%20Video%20arch%20Chicago%20SCTE.pdf

[4] Sheryl L. Woodward, Mary R. Phillips, *"Optimizing Subcarrier-Multiplexed WDM Transmission Links,"* Journal of Lightwave Technology, Vol. 22, Nr. 3, March 2004.

[5] Gèrard Terrault, *"QAM Signal Impairments And Their Effects on MER and BER,"* Sunrise Telecom, White Paper, Rev 1.04, 2-Jan-03, http://docsis.beckitrue.com/documents/sunrise/QAM_Impairment Effects_on_MER_BER_104.pdf

[6] Ron Hranac (Cisco) and Bruce Currivan (Broadcom), *"Digital Transmission: Carrier-to-Noise Ratio, Signal-to-Noise Ratio and Modulation Error Ratio,"* Cisco/Broadcom, White Paper, http://www.cisco.com/en/US/prod/collateral/video/ps8806/ps5684/ps2209/prod_white_paper0900aecd805738f5.html

[7] T. Brophy, F. Villarruel, K-Y. Wu, *"The Delicate Balance,"* Cisco, White Paper, http://www.scientificatlanta.com/products/customers/white-papers/g1667a_thedelicatebalance.pdf

[8] XFP, MSA, ftp://ftp.seagate.com/sff/INF-8077.PDF

[9] David Piehler, *"Next-Generation Components for Optical Access Networks,"* Neo-Photonics USA, Optical Fiber Conference Proceedings, 2011.

Author's Contact Information

Fernando Villarruel: villarf@cisco.com

Fred Hirt: frhirt@cisco.com

Kuang-Yi Wu: kuanwu@cisco.com

RIGHT-SIZING CABLE MODEM BUFFERS FOR MAXIMUM APPLICATION PERFORMANCE

Greg White (CableLabs®)
Richard Woundy, Yiu Lee, Carl Williams (Comcast)

Abstract

The sizing of the cable modem (CM) upstream buffer can significantly impact application performance and end-user quality of experience. Too little buffering will result in poor TCP throughput in the upstream direction. Too much buffering will have a negative impact on latency sensitive applications when the upstream link is heavily utilized. Studies have indicated that cable modems in the field today have perhaps an order of magnitude greater buffering than is needed to ensure TCP throughput, with the result being a sluggish user experience for web browsing, and outright failure of VoIP and real-time games at times when the upstream is being used simultaneously by TCP.

This paper provides a background on the network dynamics involved, describes a new feature in the DOCSIS® 3.0 specification that allows the operator to tune the upstream buffer, and provides some initial experimental work that seeks to provide guidance to operators in the use of this feature.

INTRODUCTION

Buffering of data traffic is a critical function implemented by network elements (e.g. hosts, switches, routers) that serves to minimize packet loss and maximize efficient use of the next-hop link. It is generally thought that for the Transmission Control Protocol (TCP) to work most efficiently, each element in a network needs to support sufficient buffering to allow the TCP window to open up to a value greater than or equal to the product of the next hop link bandwidth and the total round trip time for the TCP session (the bandwidth delay product). In many residential broadband scenarios, the cable modem is the head of the bottleneck link in the upstream direction, and as a result its buffering implementation can have a significant impact on performance – either imposing limits on upstream data rates or increasing delay for latency sensitive applications.

Historically, upstream buffer implementations in CMs have been sized (statically) by the modem vendor, and the network operator has had no control or visibility as to the size of the buffer implementation. In order to ensure optimal performance across the wide range of potential upstream configured data rates, modem suppliers have sized their upstream buffers such that there is sufficient buffering for TCP to work well at the highest possible upstream data rate, and at the greatest expected round trip time (i.e. the greatest possible bandwidth delay product). In the majority of deployment scenarios, the configured upstream data rate is significantly less than the highest possible rate, and the average TCP round trip time is significantly less than the greatest expected. The result is that in the majority of cases the cable modem has an upstream buffer that greatly exceeds what is necessary to ensure optimal TCP performance.

In today's mix of upstream TCP and UDP traffic, some of which is latency sensitive, some of which is not, the impact of a vastly oversized upstream buffer is that the upstream TCP sessions attempt to keep the buffer as full as possible, and the latency sensitive traffic can suffer enormously. The result is poorer than expected application performance for VoIP, gaming, web surfing, and other applications. Studies have shown upstream buffering in deployed cable modems on the order of multiple seconds.

Recent work has established a standardized means by which a cable operator can set the upstream buffer size at the cable modem via the modem's configuration file (alongside other service defining parameters, such as the maximum traffic rate). This paper provides an introduction to this new capability, investigates the performance impact of buffer size on a range of traffic scenarios, and seeks to provide some initial guidance for operators to properly configure buffer size for their service offerings.

RESEARCH AND BEST PRACTICES RELATING TO BUFFER SIZING

The TCP [2] is a connection-oriented, end-to-end protocol that enables the reliable transmission of data across the Internet. The TCP is designed to recover data that is damaged, lost, duplicated, or delivered out of order by the network.

The original TCP specification provides reliability and end-to-end flow control through the use of sequence numbers and the use of a “receive window”. All transmitted data is tracked by sequence number, which enables the TCP receiver to detect missing or duplicate data. The receive window enables flow control by identifying the range of sequence numbers that the receiver is prepared to receive, to prevent overflow of the receiver’s data buffer space.

However, these controls were insufficient to avoid “congestion collapse” in the Internet in the mid-1980s [4]. Van Jacobson [1] developed the initial mechanisms for TCP congestion control, which many researchers have continued to extend and to refine [6]. In particular, TCP congestion control utilizes a “congestion window” to limit the transmission rate of the sender in response to network congestion indications.

Two key congestion control algorithms are “slow start” and “congestion avoidance”. In slow start, the sender increases the congestion window by one segment for each TCP acknowledgement (ACK) received; this action doubles the transmission rate for each roundtrip time, leading to exponential growth. Once in congestion avoidance, the sender increases the congestion window by one segment per roundtrip time; this action increases the transmission rate linearly. If a segment is dropped due to a timeout, the threshold for entering congestion avoidance is set to one-half the current congestion window, and the congestion window itself is reset to one segment. This “sawtooth” behavior is illustrated in Figure 1 [16].

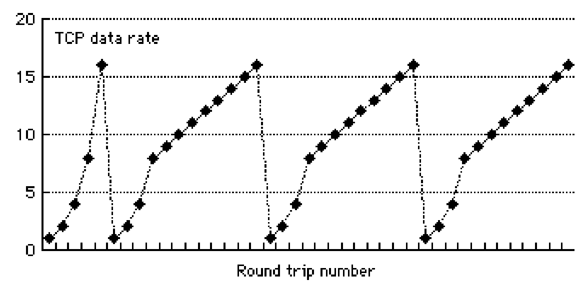


Figure 1

Modern TCP implementations also implement “fast retransmit” and “fast recovery” algorithms. According to fast retransmit, when the sender receives three duplicate ACKs, the sender assumes a segment has been lost and re-transmits that segment. If that segment is subsequently transmitted successfully, then under fast recovery the sender cuts its congestion window in half (rather than resetting to one segment).

Using modern congestion control, the TCP sender is constantly attempting to increase its transmission rate to the maximum possible, and it decreases the transmission rate in the presence of segment loss. As a result of this behavior, the TCP sender will send packets at a continuously increasing rate, filling the buffer of the network device at the head of the bottleneck link, until one or more packets are dropped. (This discussion ignores the presence of Active Queue Management [3] or Explicit Congestion Notification [5], neither of which are commonly implemented in broadband modems.)

Since the TCP sender's behavior is to keep the buffer full at the bottleneck link, it is helpful to review the research recommendations for the size of network buffers.

Back in 1994, Villamizar and Song [7] provided performance test results using up to eight TCP flows over a national backbone infrastructure. Their results led to the rule-of-thumb that bottleneck links need a buffer size equal to the bandwidth delay product (BDP) – that is, equal to the available bandwidth at the bottleneck multiplied by the roundtrip time of the TCP sessions – for the TCP flows to saturate the bottleneck link. A decade later, Appenzeller, Keslassy and McKeown [8] confirmed the BDP rule-of-thumb for a single long-lived TCP flow, which may apply best to a bottleneck link at the edge of the network, such as a broadband modem's upstream link.

Since that time, the topic of buffering at network hops has become an active one, with some beginning to look at the impact that the rule-of-thumb has on non-TCP traffic, as well as whether the rule-of-thumb holds up under more real-world conditions. Vishwanath, et al. [10] provide a good survey of recent work in this area.

If the optimal buffer size for TCP performance is presumed to be equal to the bandwidth delay product, then one might inquire how that compares to the buffer size implemented in current broadband modems. In 2007, Dischinger et al. [9] measured the performance of a number of DSL and cable operators in Europe and in North America. While the measured broadband modem minimum delay was typically under 15 milliseconds, the upstream buffer sizes were found to be quite large – resulting in 600 milliseconds of queuing delay for most DSL links, and several seconds for many cable modems. These results suggest that broadband buffer sizes are much larger than the BDP rule-of-thumb.

Gettys [17] shows that such large buffer sizes, or "bufferbloat", is an issue today with broadband modems both for wireline and for wireless networks. As one example, some have observed six seconds of latency in 3G wireless networks related to queuing. The issue of bufferbloat has also been observed in other consumer equipment, such as home routers and laptops. Gettys points to excessive buffering at the head of the bottleneck link as being disruptive to many user applications.

CABLE MODEM UPSTREAM BUFFERING AND IMPACTS ON APPLICATION PERFORMANCE/USER EXPERIENCE

In the upstream traffic direction, the cable modem is generally the entry point of the bottleneck link, both topologically and in terms of link capacity. As a result of this, its buffer size can have a significant impact on application performance. As noted previously, TCP upload performance can be sensitive to buffer size on the bottleneck link, and furthermore will seek to keep the bottleneck link buffer full. When latency-sensitive or loss-sensitive traffic is mixed with TCP traffic in the same cable modem service flow, issues can arise as a result of the modem buffer being kept full by the TCP sessions.

One impact will be that applications will see an upstream latency that can be predicted by the buffer size divided by the upstream configured rate. For example, in the case of a 256KiB buffer and a configured upstream rate of 2Mb/s, the upstream latency would be expected to be on the order of 1 second. More buffering will result in a proportional increase in latency.

Another impact will be that upstream applications will experience packet loss due to buffer overflow. Bulk TCP sessions are immune to packet loss resulting from buffer overflow, and in fact their congestion avoidance algorithm relies on it. However, other applications may be more sensitive. In contrast to the latency impact, the packet loss rate will increase if the buffer is undersized.

How Much Buffering Should There Be?

Commonly used VoIP and video conferencing applications such as Vonage, Skype, iChat, FaceTime, and umi, provide a better user experience when end-to-end latency is kept low. The ITU has developed models and recommendations for end-to-end delay for voice and other interactive services. The ITU-T Recommendation G.114 [11] suggests one-way latency be kept below 150ms in order to achieve "essentially transparent interactivity", and that delays greater than 400ms are unacceptable for interactive applications. Furthermore it has defined the "E-model" (ITU-T G.107 [12], G.108 [13] & G.109 [14]) used for predicting digital voice quality based on system parameters. The E-model has been reduced to a simpler form for VoIP applications by Cole & Rosenbluth, which points to significant degradation in voice quality beginning when the one-way latency exceeds ~177ms. Packet loss also degrades quality, with losses of less than 2.5% being necessary to achieve "Best" quality. Included in the packet loss statistic are any packets that exceed the ability of the receiver de-jitter buffer to deliver them isochronously.

Multiplayer online games can also be sensitive to latency and packet loss. Fast-paced, interactive games are typically the most sensitive to latency, where some games require that players have less than 130ms round-trip time between their host and the game server in order to be allowed to play, and lower is generally better.

Even more latency sensitive are games that are run on a cloud server, with a video stream sent to the player, and a control stream sent from the player back to the server. Services such as these may require a maximum of 20-50ms round-trip time across the cable operator's network.

While not typically considered to be a latency sensitive application, web-browsing activities can also be impacted by upstream latency. Consider that a webpage load consists of a cascade of perhaps a dozen or more round trips to do DNS lookups and HTTP object fetches. An upstream latency on the order of 500ms or more is likely to cause the web browsing experience to be much slower than a user might find acceptable, particularly if they are paying for a downstream connection speed measured in tens of megabits per second.

In addition, any downstream-centric TCP application (such as web browsing) has the potential to be throttled due to the latency experienced by the upstream TCP-ACK stream from the receiver. In practice this is not an issue, because many cable modems support proprietary means to accelerate TCP-ACKs by queuing them separately from bulk upstream data.

How Much Queuing Latency Is There?

Historically, cable modems have implemented static buffer sizes regardless of upstream data rate. Evidence suggests that cable modems in the field today may have buffers that are sized (using the BDP rule-of-thumb) for the maximum possible upstream data rate (~25Mb/s for DOCSIS 2.0 CMs and

~100Mb/s for current DOCSIS 3.0 CMs) and the maximum coast-to-coast Round Trip Time (RTT) that might be experienced (100ms or more).

Our observations indicate that most conventional cable modems are built with buffer size between 60KiB and 300KiB.

Since the majority of modems (particularly those in residential service) are currently operated with upstream rates in the range of 1 to 2Mb/s, the result (as corroborated by Dischinger [9]) is that the modems have buffering latencies on the order of several seconds, which may be 1 or 2 orders of magnitude greater than would be ideal.

DOCSIS 3.0 BUFFER CONTROL FUNCTIONALITY

A recent addition to the DOCSIS 3.0 specifications provides, for the first time, the ability for cable operators to tune thetransmit buffers for cable modems and CMTSs in their networks. This new feature, support for which became mandatory for cable modems submitted for CableLabs certification in early April 2011 (and is optional for CMTSs), gives the operator control of the configured buffer size for each DOCSIS Service Flow. The feature controls upstream buffering in the cable modem, and downstream buffering in the CMTS.

The new feature is referred to as Buffer Control, and is defined as a Quality of Service (QoS) Parameter of a DOCSIS Service Flow. Specifically, the Buffer Control parameter is a set of Type-Length-Value (TLV) tuples that define a range of acceptable buffer sizes for the service flow, as well as a target size. This allows the CM/CMTS implementer some flexibility on the resolution of its buffer configuration. For example, a CM implementer might choose to manage buffering in blocks of 1024 bytes, if that offered some implementation advantages. When presented with a Buffer Control TLV, such a CM would then choose a number of blocks that results in a buffer size that is within the range of acceptable sizes, and is as close to the target buffer size as possible.

The Buffer Control TLV is comprised of three parameters:

Minimum Buffer - defines the lower limit of acceptable buffer sizes. If the device cannot provide at least this amount of buffering, it will reject the service flow. If this parameter is omitted, there is no minimum buffer size.

Target Buffer - defines the desired size of the buffer. The device will select a buffer size that is as close to this value as possible, given its implementation. If this parameter is omitted, the device selects any buffer size within the allowed range, via a supplier-specific algorithm.

Maximum Buffer - defines the upper limit of acceptable buffer sizes. If the device cannot provide a buffer size that is less than or equal to this size, it will reject the service flow. If this parameter is omitted, there is no maximum buffer size.

Each of these parameters is defined in bytes, with an allowed range of 0 - 4,294,967,295 (4 GiB).

As noted, each of the three parameters is optional. If all three parameters are omitted, the interpretation by the device is that there is no limitation (minimum or maximum) on allowed buffer size for this service flow, and that the device should select a buffer size via a vendor specific algorithm. This is exactly the situation that existed prior to the introduction of this feature. As a result, if the operator does not change its modem configurations to include Buffer Control, the operator should see no difference in behavior between a modem that supports this new feature and one that does not.

In many configuration cases it will be most appropriate to omit the Minimum Buffer and Maximum Buffer limits, and to simply set the Target Buffer. The result is that the modem will not reject the service flow due to buffering configuration, and will provide a buffer as close as the implementation allows to the Target Buffer value.

In certain cases however, the operator may wish to be assured that the buffer is within certain bounds, and so would prefer an explicit signal (i.e., a rejection of the configuration) if the modem cannot provide a buffer within those bounds. Hard limits are provided for these cases.

The cable modem is required to support buffer configurations of up to 24KiB per service flow, but is expected to support significantly more, particularly when a small number of service flows is in use. Put another way, if the Minimum Buffer parameter is set to a value less than or equal to 24576, the cable modem is guaranteed not to reject the configuration due to that parameter value. If the Minimum Buffer parameter is set to a value that is greater than 24576, then there is some risk that a modem implementation will reject the configuration.

Since they are part of the QoS Parameter Set, the Buffer Control TLVs can be set directly in the cable modem's configuration boot file; they can be set indirectly via a named Service Class defined at the CMTS; and they can be set and/or modified via the PacketCable^{TM} Multimedia (PCMM) interface to the CMTS.

In order to use this new feature to control upstream buffering in DOCSIS 3.0 cable modems, it is necessary that the CM software be updated to a version that supports it. Furthermore, CMTS software updates are necessary in order to ensure that the CMTS properly sends the TLVs to the CM, regardless of which of the above methods is utilized.

EXPERIMENTAL WORK

Methodology

The new Buffer Control parameters described in a previous section enable us to provision the buffer size in the cable modem, and thus provide the opportunity to investigate the application performance impact in various scenarios. For the purposes of this paper, we ran three sets of tests:

In the first set of tests, we correlate the buffer size and upstream speed. The purpose of the test is to understand the impact buffer size would have on the ability of the customer to utilize his/her provisioned upstream bandwidth. As an illustration, if the round-trip time for a TCP session is 40ms and the upstream service is provisioned for 5Mb/s, the rule-of-thumb for buffer size would indicate 5000Kb/s * 0.04s = 200Kb or 24.4KiB. If the buffer size was then set to 16KiB, the expectation is that a single TCP session would not be able to utilize the 5Mb/s service. In this set of tests, we validate that expectation, and we additionally look at the effect when multiple simultaneous TCP sessions are sharing the upstream link.

In the second set of tests we correlate the buffer size and QoS of real-time applications during upstream self-congestion (i.e. saturation of the user's allotted bandwidth by a mix of applications). During congestion, packets from real-time applications may queue up in the buffer. If the buffer size is too large, this will increase the latency and may severely impact real-time applications such as online games and Voice over IP (VoIP). This set of tests examines whether changing the buffer size can improve real-time applications' QoS or not.

The third set of tests examines the implications of buffer size when the token bucket feature defined in the DOCSIS 3.0 specification is utilized to provide the user a high burst transmit rate (as is currently the practice by several operators). In this test, the token bucket is set to 5MB and, after a suitable idle period, a 5MB file is transferred using a single TCP session (FTP). The TCP session runs through the process discussed earlier, beginning with slow start and transitioning to congestion avoidance. Since the token bucket size is equal to the size of the file, the rate shaping function in the CM doesn't come into play, and thus the limits to the transfer speed in this test are: a) the maximum speed of the upstream channel, which in this case is approximately 25Mb/s; and b) ability of a single TCP session to utilize the link in the presence of limited buffering.

We used upstream buffer sizes of 8KiB, 16KiB, 32KiB, 64KiB, 128KiB, and 256KiB, although in some test scenarios we limited the choice of buffer sizes due to time constraints. In the experiments the upstream bandwidth was provisioned to 1Mb/s, 2Mb/s, 5Mb/s, or 10Mb/s. These upstream speeds are indicative of services currently available in the market. The tests simulated RTTs of 40ms and 100ms. 40ms represents a local file transfer within a metropolitan network region (perhaps from a cache). 100ms represents a coast-to-coast session.

Network Setup

Figure 2 depicts the experimental setup, which emulates a typical home network behind a CM. The home network consisted of PCs running test applications and other hardware for measuring network characteristics and performance. Typical home networks would include a home gateway between the CM and the rest of the home network equipment. To eliminate any bias that might be introduced by the packet forwarding performance of a home gateway, we instead used a Gigabit switch to connect the home network equipment to the CM. We then configured the CM to allow multiple devices connecting to it.

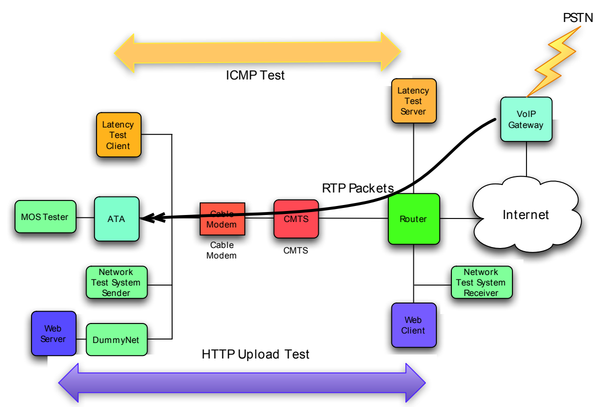


Figure 2

The CM was installed with the new DOCSIS 3.0 firmware that provides the capability of using the Buffer Control feature so that specific buffer queue sizes might be set. A CMTS with knowledge of such settings was connected to the CM. Inside the home network, there are four applications: (1) a PING test client, (2) a VoIP Analog Terminal Adaptor (ATA), (3) a Network Test System, and (4) a Web Server. The PING test client measured the RTT to the PING server in the test network. This was set up to validate the PING latency while the network was congested. The VoIP ATA was set up for the QoS test for real-time applications. A VoIP tester was connected to the ATA. This tester would measure the Mean Opinion Score (MOS) when the ATA made a phone call to the VoIP gateway in the Internet. The Network Test System allowed an easy way to simulate multiple TCP devices simultaneously. The Web server was setup to generate packets over TCP by uploading a large file via HTTP to the client behind the CMTS. In order to simulate the network latency, Dummynet was used. Dummynet [18] is an internal packet filter for FreeBSD which can inject artificial delay in the network. It was connected between the CM and Web server. Delay was injected in both upstream and downstream. For example: we configured Dummynet to inject 20ms in both directions for the HTTP upload test to simulate 40ms RTT delay.

Baseline Results

As an initial test, we configured the CM with an upstream rate limit of 35Mb/s (effectively no rate-limiting), and investigated the achievable TCP data rate for a range of buffer sizes, both for the 40ms RTT case and for the 100ms RTT case. Figure 3 shows the resulting achieved data rate for 10 simultaneous upstream TCP sessions, along with the data rate that would be predicted based on the BDP. The maximum data rate that a single DOCSIS upstream channel could support is approximately 25Mb/s, so the achieved rates for 256KiB, 128KiB and perhaps 64KiB should be considered to be artificially limited by this constraint. It is interesting to note that in nearly all cases, when multiple TCP sessions were utilized, we were able to achieve a higher data rate (and in some cases a much higher data rate) than would be predicted based on the rule-of-thumb. Our conjecture is that the multiple TCP sessions were desynchronized (that is, the TCP sawtooths were not synchronized across sessions), and Appenzeller, Keslassy and McKeown [8] demonstrate that significantly less buffering than the bandwidth delay product is needed in this scenario.

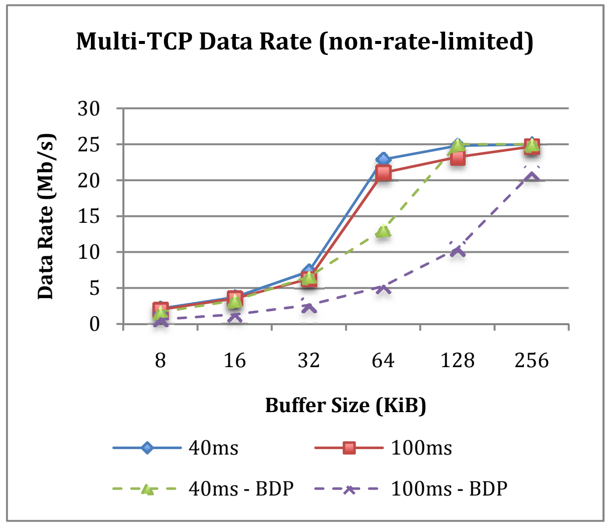


Figure 3

Test 1 Results

Figure 4 and Figure 5 show the results of data rate in various buffer sizes. Figure 4 shows the data rate of 1Mb/s and 2Mb/s upstream services. Figure 5 shows the data rate of 5Mb/s and 10Mb/s upstream services. The tests were executed with either 40ms or 100ms delay. All the tests were run with a single TCP stream.

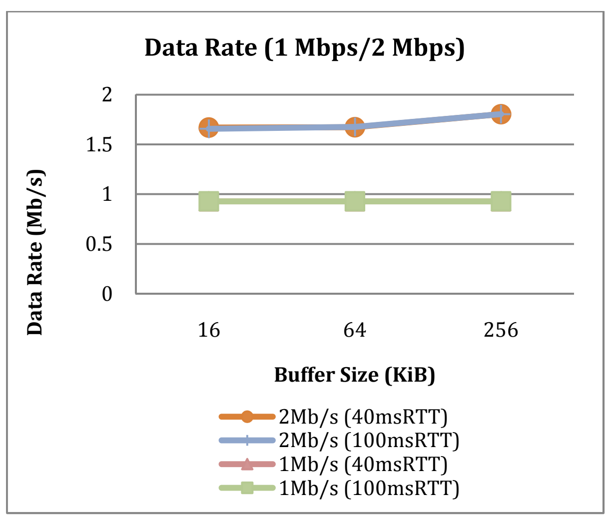


Figure 4

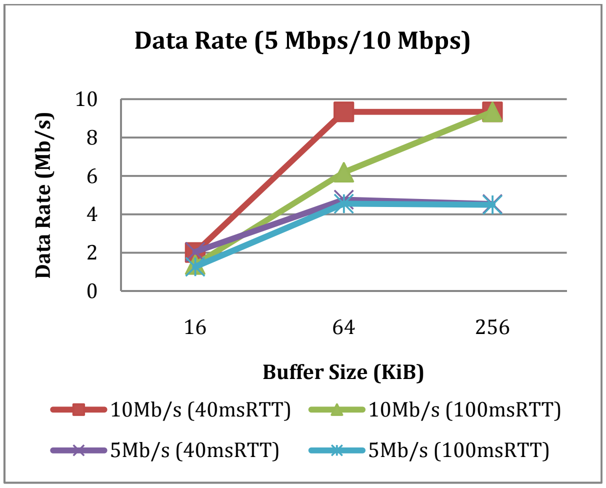


Figure 5

For 1Mb/s and 2Mb/s services, 16KiB buffer size was sufficient to utilize the upstream bandwidth with either 40ms and 100ms delay. For 5Mb/s, results showed that at least 64KiB buffer was needed when delay was 100ms. For 10Mb/s, results showed that 64KiB buffer was needed for 40ms delay and 256KiB buffer was needed for 100ms delay.

From the previous section, the tests performed with 10Mb/s upstream service for 16KiB and 64KiB showed significant bandwidth under-utilization. The question becomes: is it possible to achieve utilization of the whole bandwidth with multiple parallel TCP sessions? In additional experiments we used 10 TCP sessions established in parallel recording individual data rates and provided the aggregation plots in the graph provided in Figure 6a and Figure 6b for 10Mb/s upstream service with buffer sizes of 16KiB and 64KiB.

Figure 6a shows that with a 16KiB buffer, a single TCP session with 40ms delay is unable to achieve the full upstream rate, with only 1.97Mb/s of the 10Mb/s upstream service being used. By aggregating 10 TCP sessions and keeping other factors fixed, the channel utilization improved by a factor of four to 7.11Mb/s. Still, full utilization wasn't realized. When 64KiB buffer size was used, we saw no difference in throughput between a single TCP session and 10 TCP sessions. Both were able to achieve the full 10Mb/s rate limit.

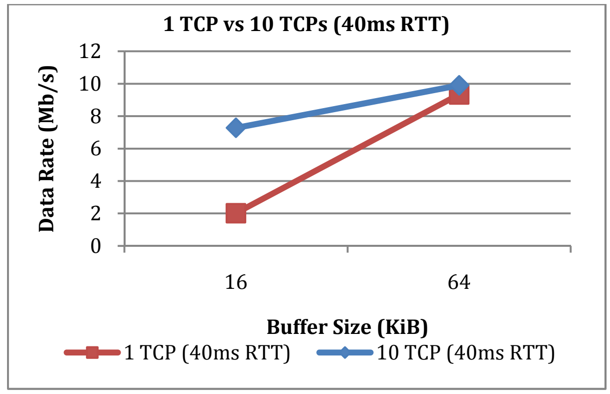


Figure 6a

Figure 6b shows further results with a 100ms delay. Bandwidth utilization could be improved for both 16KiB and 64KiB buffers when aggregating 10 TCP sessions. The test using 16KiB buffer size with a 100ms delay resulted in 4.95Mb/s upstream utilization, a three times improvement over the test that was conducted using a single TCP sessions. However, 5Mb/s of capacity remains unused. When using a 64KiB buffer size with 10 TCP sessions, there was improvement with a data rate of 8.77Mb/s compared to 6.05Mb/s for the single TCP flow test. Here again 1Mb/s of capacity was unused.

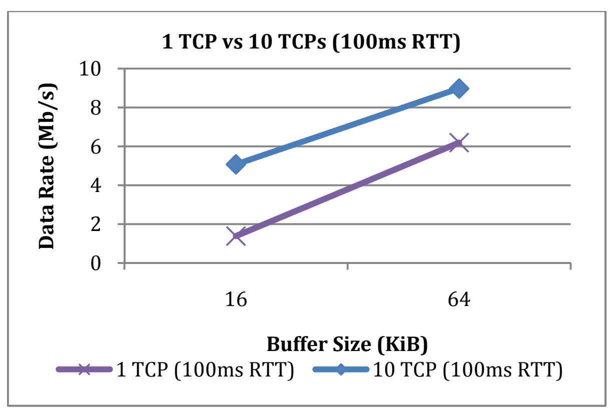


Figure 6b

Figure 7 and Figure 8 show the results of packet latency when an upstream link was congested with a single TCP session. Figure 7 shows the latency of 1Mb/s and 2Mb/s upstream services. Figure 8 shows the latency of 5Mb/s and 10Mb/s upstream services when the upstream link was congested with a single TCP session.

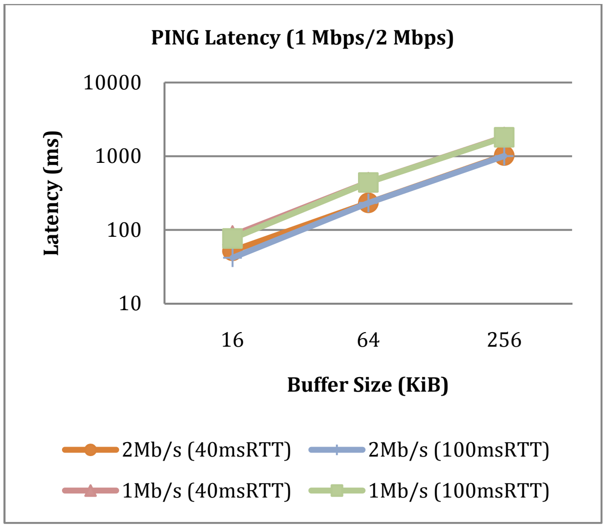


Figure 7

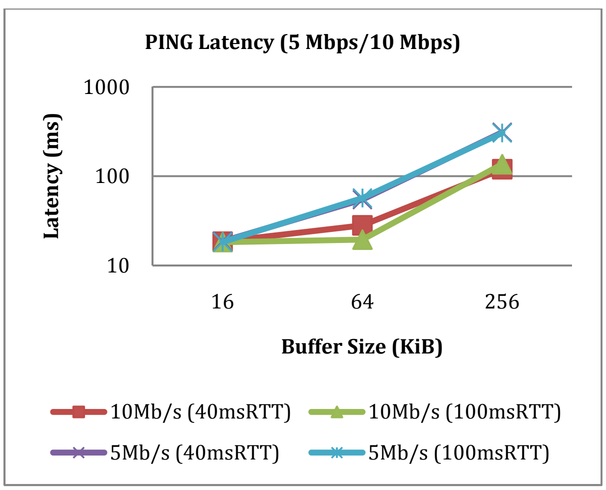


Figure 8

For all upstream services, the pattern is that the bigger the buffer, the higher the PING latency. The observed pattern of the test results show (as expected) that the latency is directly proportional to the buffer size. If the buffer size is increased to four times the size, the latency would increase roughly four times. For example, when testing with a 2Mb/s upstream service, the PING latency was 52.04ms for a 16KiB buffer. The PING latency increased to 200ms for a 64KiB buffer and 1019.17ms for a 256KiB buffer. However, this pattern was not observed when using a 10Mb/s service for 16KiB and 64KiB buffer sizes. This is because a single TCP session was unable to utilize the full 10Mb/s service with 40ms and 100ms delay. Therefore, the PING packets proceeded without congestion impact.

<u>Test 2 Results</u>

In this test, we wanted to understand how buffer size would impact the VoIP MOS under congestion. The VoIP ATA was connected into the simulated home network in our test environment. We started a large upstream file transfer (using a single TCP session) to simulate a self-congested upstream. The Dummynet test set injected either 40ms or 100ms round-trip latency to the TCP stream. The two latency values impact the dynamics of the TCP congestion avoidance algorithm differently, and thus result in a different pattern of CM buffer usage. For each bandwidth/buffer/delay permutation, we ran 10 tests and computed the average MOS.

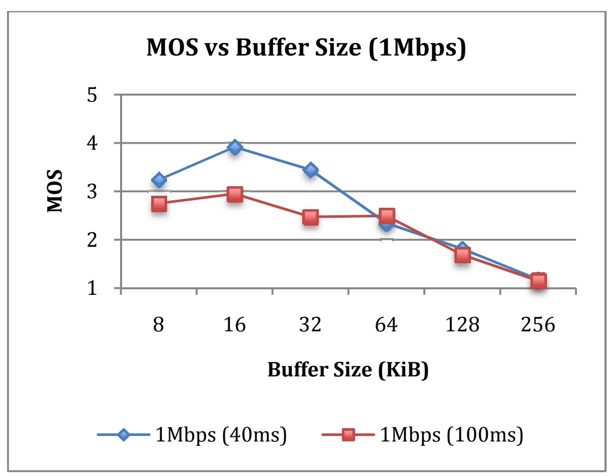


Figure 9a

Figure 9a shows the results when the upstream was configured to 1Mb/s. In this case, the 16KiB buffer gave the best MOS result regardless of the TCP RTT. When the buffer size increased to 64KiB, the MOS started to drop. In fact, when the buffer size was set to 256KiB with 100 ms delay, many tests couldn't be completed. The test set reported a Timeout. The timeout could be caused by long delay introduced by the large buffer.

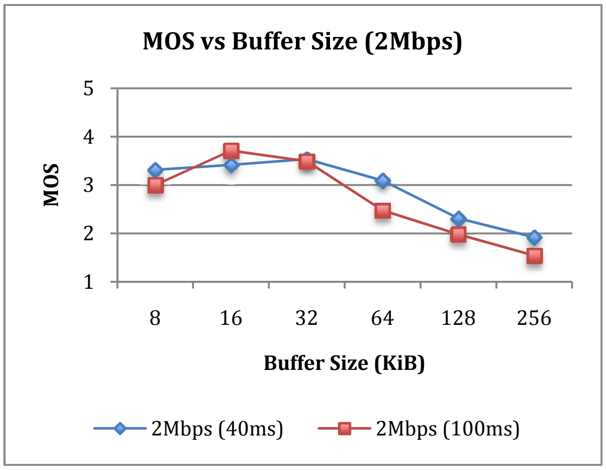


Figure 9b

Similar results were obtained for 2MB/s, as shown in Figure 9b.

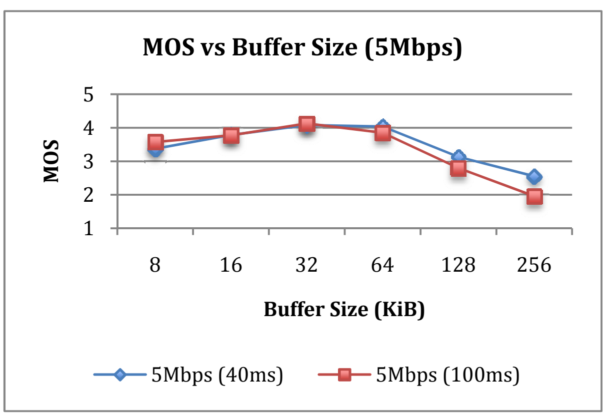


Figure 9c

Figures 9c and 9d show the results of 5Mb/s and 10Mb/s upstream bandwidths. For both of these cases, the best MOS results were recorded when the buffer size was set between 32KiB and 64KiB. What was interesting is when the buffer was set below 64KiB, there should not be any congestion because the buffer size was too small for the competing TCP session to utilize the full upstream bandwidth. Also when the buffer size was set to 8KiB or 16KiB, the VoIP traffic should see very low buffering latency (between 6ms and 26ms). Both of these factors should result in high MOS scores. However, the MOS score was degraded when the buffer size was set to 8KiB or 16KiB, and this low MOS is likely to be caused by packet loss due to the periodic saturation of the small CM buffer by the competing TCP congestion window.

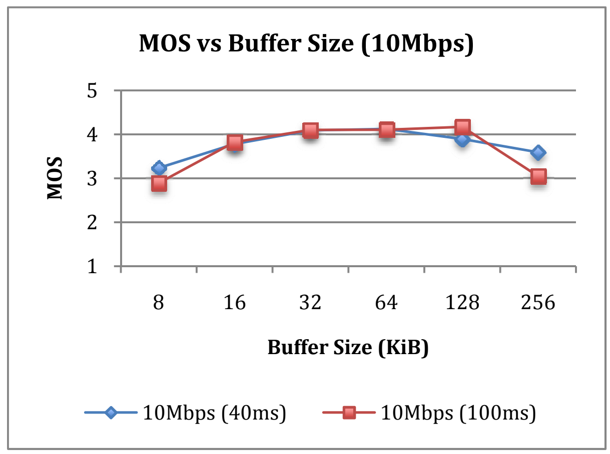


Figure 9d

When we increased the buffer size above 128KiB, the MOS started to deteriorate. In fact, the test set failed to complete some tests and required us to re-run the test when we set the upstream bandwidth to 10Mb/s with 256KiB buffer and 100ms delay. This could be caused by the long delay of the VoIP signaling packets for the call setup.

Another interesting observation is that, while adding latency to the TCP session did in some cases affect the VoIP MOS, it didn't point to a different optimal buffer size.

Test 3 Results

In this test, the CM's rate-shaping token bucket depth was set to 5MB to mimic the configuration that some operators use to boost performance for bursty interactive traffic. The test set was then used to transit a 5MB file in the upstream direction using FTP. The test measured the transfer time and, from this, an average data rate was calculated. Due to the fact that the file size was equal to the token bucket depth, there was effectively no upstream rate-shaping in play. Figure 10 shows the result of Average Transfer Time for the scenario with 40ms RTT and the scenario with 100ms RTT.

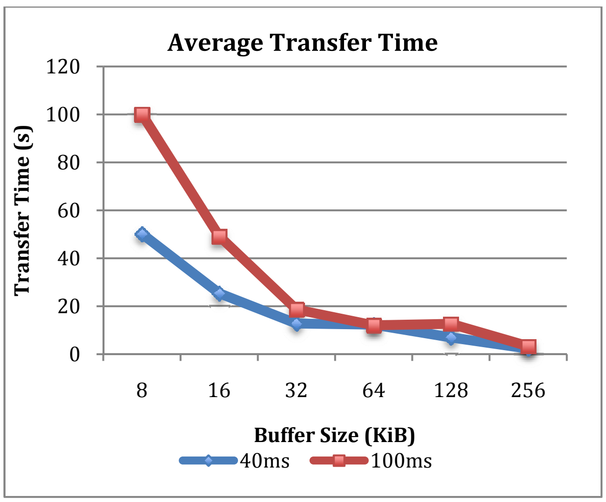


Figure 10

Figure 11 shows the calculated average throughput based on the average transfer time and the file size. It is clear that the choice of buffer size can have a dramatic impact on achievable throughput during this performance boost period. It is important to note that the average throughput shown in Figure 11 is for the entire file transfer, and so includes the slow-start effect as well as the initial FTP hand-shaking.

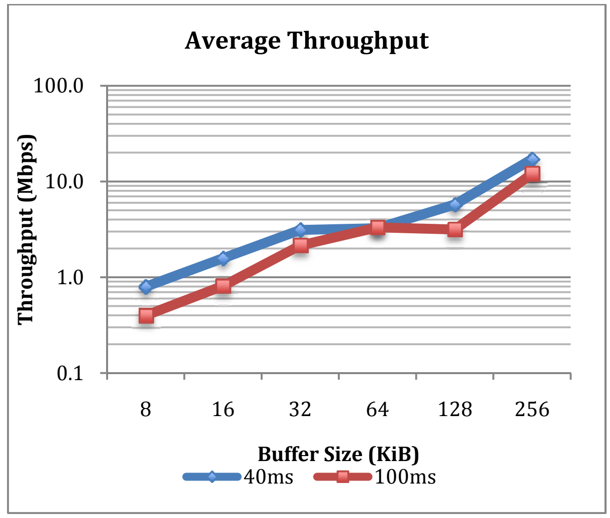


Figure 11

FUTURE WORK

Additional work is necessary to more fully understand the impact of CM buffering on application performance and user experience. The overarching goal of this work is to enable cable operators to optimize network performance and user experience across a wide range of application and usage scenarios.

The experimental work described in this paper focused both on upstream bulk TCP performance, and on the impact that bulk TCP traffic would have on a VoIP session across a range of buffer sizes in fairly isolated and controlled scenarios. Further work will seek to examine the effect of buffering on some more real-world application scenarios, such as a scenario in which TCP sessions are numerous and short-lived, such that many of them stay in the slow-start phase for their entire duration, and never (or rarely) enter the congestion avoidance phase. Additionally, future work should assess the impact of buffer sizing on other latency or loss sensitive applications beyond VoIP.

CONCLUSIONS

We have shown that the size of the upstream service flow buffer in the cable modem can have significant effects on application performance. For applications that use TCP to perform large upstream transmissions (i.e. upstream file transfers such as uploading a large video clip), insufficient buffering capacity can limit throughput. Applications utilizing a single TCP session see the most pronounced effect. Applications that are latency sensitive, on the other hand, see degraded performance when too much buffer capacity is provided.

Cable modems deployed today appear to have significantly greater buffering than is needed to sustain TCP throughput, potentially causing high latency when the upstream is congested. The result is poorer than expected application performance for VoIP, gaming, Web surfing, and other applications that are latency-sensitive.

A new feature for DOCSIS 3.0 CMs allows operators to configure the upstream buffer size for each upstream service flow in order to optimize application performance and to improve user experience. In choosing the buffer size, the operator will need to consider the upstream QoS parameters for the service flow, the expected application usage for the flow, as well as the service goals for the flow.

Service features that utilize a large token bucket size (in order to provide high throughput for short bursts) complicate matters since the buffer size cannot realistically be resized in real time. Thus a buffer configuration that may be optimized to provide a good balance in performance between TCP uploads and real-time services for the configured sustained traffic rate, may result in poorer than expected burst speeds.

Further work is needed to evaluate the performance impact that buffer size has on a wider range of application scenarios.

REFERENCES

[1] V. Jacobson, Congestion Avoidance and Control, ACM SIGCOMM '88, August 1988.

[2] J. Postel, Transmission Control Protocol, STD 7, RFC793, September 1981.

[3] Braden, B., et al., Recommendations on Queue Management and Congestion Avoidance in the Internet, RFC 2309, April 1998.

[4] S. Floyd, Congestion Control Principles, BCP 41, RFC 2914, September 2000.

[5] K. Ramakrishnan, S. Floyd and D. Black, The Addition of Explicit Congestion Notification (ECN) to IP, RFC 3168, September 2001.

[6] M. Allman, V. Paxson and E. Blanton, TCP Congestion Control, RFC 5681, September 2009.

[7] C. Villamizar and C. Song, High Performance TCP in ANSNet. ACM CCR, 24(5):45-60, 1994.

[8] G. Appenzeller, I. Keslassy, and N. McKeown, Sizing Router Buffers, ACM SIGCOMM, USA, 2004.

[9] M. Dischinger, et al., Characterizing Residential Broadband Networks, Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, Oct. 24-26, 2007, San Diego, CA, USA.

[10] A. Vishwanath, V. Sivaraman, M. Thottan, Perspectives on Router Buffer Sizing: Recent Results and Open Problems, ACM CCR, 39(2):34-39, 2009.

[11] ITU-T Recommendation G.114, One-way Transmission Time, http://www.itu.int

[12] ITU-T Recommendation G.107, The E-model, a computational model for use in transmission planning, http://www.itu.int

[13] ITU-T Recommendation G.108, Application of the E-model: A planning guide, http://www.itu.int

[14] ITU-T Recommendation G.109, Definition of categories of speech transmission quality, http://www.itu.int

[15] R. G. Cole and J. H. Rosenbluth, Voice over IP performance monitoring, *ACMCCR,* 31(2), 2001.

[16] Internetworking lectures from La Trobe University, http://ironbark.bendigo.latrobe.edu.au/subjects/INW/lectures/19/, 2011.

[17] J. Gettys, Bufferbloat: Dark Buffers in the Internet, http://mirrors.bufferbloat.net/Talks/PragueIETF/IETFBloat7.pdf, April 2011.

[18] L. Rizzo, Dummynet: a simple approach to the evaluation of network protocols, ACM Computer Communication Review, Vol. 27, No. 1, pp. 31-41, January 1997.

NEXT GENERATION - CABLE ACCESS NETWORK

AN EXAMINATION OF THE DRIVERS, NETWORK OPTIONS, AND MIGRATION STRATEGIES FOR THE ALL-IP NEXT GENERATION – CABLE ACCESS NETWORK

Authors:M. Emmendorfer, S. Shupe, D. Cummings, T. CloonanContributors:Z. Maricevic, M. Schemmann, B. Dawson, V. Mutalik, J.Howe, A. Al-Banna,and F. O'Keeffe

ARRIS

ABSTRACT

The Cable Industry is facing a decade of unprecedented change in the areas of video and high-speed Internet services. This change,driven by competition and consumer demand, will transform the cable network end-to-end. This paper will focus entirely on what we are calling the Next Generation Cable Access Network,examining the business drivers, network options, and migrations strategies in the access layer of the data and HFC network to provide more IP-based capacity to and from the home. The document covers in-depth the core business drivers and the technical options spanning animmense area of network disciplines and technologies, thus we have included a comprehensive executive summary at the conclusion of the report.

The analysis includes the allocation of existing spectrum and possible future spectrum expansion to accommodate consumer demand. Cable Operators and their competitors are enabling consumers to change their viewingoptions for video services and the usage of the high-speed Internet network.In the area of high-speed Internet service, competition and consumer demand is increasing the service speed tiers offered,and network traffic usage continues to rise at an alarming rate. Cable Operators like the United Kingdom's *Virgin Media*announced in April 2011 an Internet speed trial of up to 1.5 Gbps downstream and 150 Mbps upstream [1].The cable competitor *Verizon*is reportedly exploring plans to upgrade its FiOS system to XG-PON, the 10 Gbps downstream and 2.5 Gbps upstream technology [2]. New entrants in the video distribution space are capitalizing on the network investments made by the telecom industry, forcing changes in their video delivery network as well as the high-speed data network. A key challenge the cable industry will face in the future will be offering PON-like IP-based capacity in the downstream and the upstream to consumers, while leveraging their existing coaxial network.

Some of the most often asked questions by cable industry forward-looking planners reflect the key challenges the industry is facing for this decade and beyond. Some of these challenges and questions include: 1) How long willthe current spectrum split and 500 HHP nodelast? 2) What are the network technology and architecture options and what are the pros and cons? 3) How long will each of these new network architecture options last? 4) What are the financial

impacts of the options?5) What are the best ways to leverage previous, current and future investment?

This paper will seek to provide some visibility and answers to these questions and key challenges. The paper will focus entirely on the network aggregation and access layer including the CMTS, HFC and home network. This paper will provide some predictions for service tier and traffic growth, which serve as the drivers for network capacity and network utilization forecasts that are used to predict the timing of the network changes and investment. We will examine the network technology and network architecture options from spectrum splits, data MAC and PHY technologies as well as network architecture options. This paper will consider the capabilities of a drop in upgrade with an effort to maintain a 500 HHP service group and typical number of actives and passivesto determine the viability and impact for upstream spectrum expansion. The funneling effect must be considered in the analysis for the NG Cable Access Network. The paper provides analysis and comparison of some of the network elements under consideration. The paper introduces a term called Digital Fiber Coax (DFC) as a next generation architecture, which may augment the HFC media conversion style architectures that utilize centralized data access/aggregation layer equipment.

We considered a couple of migration strategies as more viable than others and while not picking a particular end-state approach, our position is to examine the options and document the pros and cons of each network architecture and technology, so industry leaders may make an informed decision.These topics under consideration comprise several network technology disciplines, which are often separate areas of concentration. This paper is by no means conclusive; some of the areas under examination have not had significant study or have the absence of products to sufficiently examine and forecast the best path.There are also timing considerations and business trade-offs that will need to be considered.

INTRODUCTION – PLANNING FOR THE NEXT GENERATION – CABLE ACCESS NETWORK

A major challenge the cable industry will face in the future will be meeting the needs of the consumer and addressing the competitive threats of PON/FTTH systems all while leveraging the existing coax to the home. This will mean significant changes in the use of network technologies, spectrum allocation, and overall network architectures. Planning for the Next Generation – Cable Access Network is extremely difficult as this spans across several network disciplines within the cable industry and even technologies outside or not widely deployed in cable, such as PON, Wireless, and EoC. The span of network technology disciplines also reaches into the network elements and underlying sub-systems such as MAC layer, PHY layer, HFC optical transport components, as well as several access radio frequency technologies end-to-end such as amplifiers, passives, and coaxial cable. What is proving to be a significant challenge is the increase dependency between all of these traditionally separate network disciplines as part of the new cable access network architecture. In years past these technologies functioned in many ways independent of one another. This next generation cable access architecture will likely migrate to more IP based spectrum in the downstream cannibalizing existing technologies which are non-IP based creating a more efficient and competitive network transport platform to compete with PON on the downstream and simply have the versatility of IP based technology. The upstream will need more spectrum and it is the overall spectrum allocation and placement of this new spectrum, whichwill have the greatest impact on the cable industry for decades to come.

The challenge we have is predicting the timing of the change in the network and how long each change will last. Additionally and most importantly what are the impacts of each of the upstream spectrum options that may be considered for the future. This paper will provide predictions, such as the drivers for the use of the spectrum in the downstream and upstream. The paper considers the spectrum allocation options and predicts how long each will last beginning with the current sub-split options and several spectrum splits which add new upstream capacity and how long these will last. The report provides a technical comparison of the upstream spectrum options and the impacts that each has from services to overall network architecture and cost.

Our Goals for Next Generation – Cable Access Network include:

- Achieve upstream bandwidth requirements through this decade
- Achieve downstream bandwidth requirements through this decade
- Continued versatility to accommodate advances in networking technology without massive changes to the outside plant network.
- Flexibility to accommodate incremental allocation of IP/Bandwidth for smooth transition strategy and pay as you grow or just in time network planning
- DOCSIS Backwards Compatibility leverages MAC/PHY channel bonding groups previously deployed and occupying spectrum yielding investment protection delaying or avoiding significantly costly approaches to find new spectrum
- Investment protection by re-using spectrum already in service, DOCSIS, HE lasers/receivers, and CPE (STB/Data) as much as possible
- Leverage network passives the most numerous OSP element
- Avoid costly and unnecessary fiber builds
- Keep the OSP as Simple as Possible for as Long as Possible
- Leverage High densities and economies of scale

Importance of Backward Compatibility with DOCSIS 3.0 and Any Successor

The authors of this paper believe that DOCSIS and any successor should consider the value of backwards compatibility especially across channel bonding groups. This assures previous and future investment may be applied to create a large IP based bandwidthnetwork while not stranding previous capital investment and spectrum. The use of channel bonding leverages every MHz, which are finite and not free, this is all towardsan effort to create one large IP pipe to and from the home. The use of backwards compatibility has benefitted the cable industry as well as other industries which use technologies like IEEE Ethernet, WiFi, and EPON creating consumer investment protection, savings,and a smooth migration strategy. The adoption of backward compatibility simply allows the MSOs to delay and perhaps avoid major investment to the network such as adding more data equipment, spectrum, node splits, or running fiber deeper.

Overview of Our Methodology For Network and Capacity Planning

In our analysis and in the structure of this paper we have examined the Next Generation-Cable Access Network in several steps as captured in the illustration below (figure 1). As shown in the illustration our process was to first determine the future requirements. The first step examined the service tier and traffic growth estimates based on a model that captured a thirty-year history to make an attempt to predict the future network needs for perhaps the next two decades. In the second phase we considered the technology and most importantly the spectrum allocation options to forecast network capacity. Then after considering the Service and Traffic growth we measured this against the network capacity options, in the section referred to as Network Utilization and Capacity Planning. In this section we forecast the timing and duration of each network step. In step four we examine several of the network technology and architecture options under consideration. The Network Migration Analysis and Strategies consider all of the factors of the aforementioned steps and provides some analysis of possible migrations strategies to address the competitive threats and consumer drivers. The migration strategies selected by the cable operators are dependent on many factors, and there may not be a consistent approach selected across all MSOs. In fact, within a given MSO the analysis may vary by market. Our analysis measures the costs of several network options.

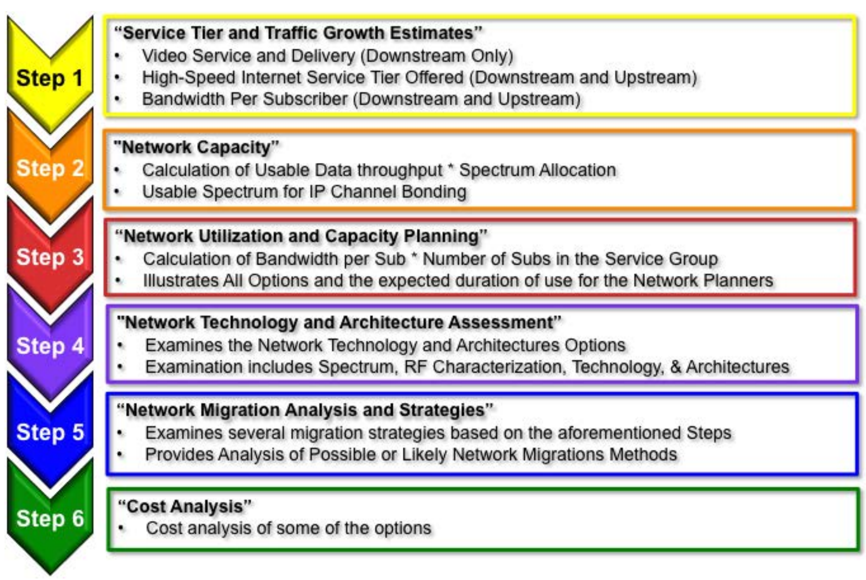


FIGURE1: METHODOLOGY FOR NETWORK AND CAPACITY PLANNING

SERVICE TIER AND TRAFFIC GROWTH ESTIMATES

Consumers and Competition Are Driving Change

The MSO's competitive landscape has changed rapidly in just the last 12 months especially from Over The Top (OTT) video providers such as Apple TV, Amazon, Hulu, Netflix and others entering the On-demand video market. In many ways the consumer electronic companies like Appleare becoming service providers enabling the video experience across all platformsand across any carriers' network. The OTT competition affects the MSOs in lost revenues for On-Demand services and perhaps a reduction in the subscription service. Adding to the lost revenue is increased costs to the high-speed data network due to increased consumer usage.

The recent completion of Verizon's FiOS roll out will undoubtedly remain a threat to the MSO's triple play offering. Additionally it was reported that Verizon will consider an upgrade to their FiOS network to the next generation Passive Optical Network (PON) technology known as XG-PON, the 10 Gbps downstream and 2.5 Gbps upstream system [2]. This could replace the earlier generation B-PON (622 Mbps down and 155 Mbps up) and the G-PON (2.5 Gbps down and 1.25 Gbps up) systems. The Verizon FiOS network also uses what is known as the video overlay network along with the PON technology. The video overlay network provides broadcast video services using technology similar to cable systems. The video overlay may employ a 750 MHz to 1002 MHz systemequivalent over 4.3 - 6 Gbps of downstream capacity but it is unknown if all of this capacity is used. The PON network is used for IP based services like Internet, telephone and perhaps on-demand unicast video transmission. If we consider both the PON system as well as the video overlay system, the FiOS network capabilities may reach ~14 Gbps+ of downstream throughput (XG-PON 10 Gbps + 750 MHz at approximately 4 Gbps+) and upstream reaching 2.5 Gbps). This capacity may be more throughput than is needed for many years or even decades to come based on the modeling in the following sections. This level of capacity may not be needed until the year 2025-2030.

The cable network has a massive amount of capacity perhaps up to 6 Gbps to the home and perhaps 100 Mbps from the home. The cable industry is making investments in IP based video delivery technology and expanding the high-speed Internet IP capacity as well. The coaxial network is very nimble and may increase the spectrum allocation beyond the current levels in either direction. This important fact is covered in detail in this paper. The amount of capacity needed in each direction is projected over a period of nearly two decades as well as several technical options are explored.

Upfront Disclaimer on Service Tier and Traffic Growth Estimates

In this report we will be making network traffic predictions for the next two decades and we acknowledge that these numbers are highly debatable.Theseforecasts may not match any particular cable or telecom provider.The modeling for the Internet portion of the traffic is based on modeling, which goes back nearly thirty years. This model illustrates Data Service Tiers offered to consumers increase at about a 50% compound annual growth rate (CAGR) and this model also is used to forecast actual consumer traffic usage which also grows at roughly a 50% CAGR. The data service portion of the model is predictable but at some point as with Moore's Law, the growth rate for Service Tiers Offered to Consumers as well as traffic usage may not continue on this trajectory for another 20 years.We are only using these Service Tier and Traffic Growth Estimates as ―rough ballpark numbers" to allow discussion and forward planning.The Network forecast will include Video Services offered by the cable provider as well asHigh-Speed Internet Services.

Video Service and Delivery Assumptions(Downstream Only)

We could have considered many factors for the video service network requirements. We could have done a year-by-year prediction of the allocation of linear programming, VoD, SDV, SDTV, HDTV, 3DTV, amount of in-home pre-caching, and service group size and number of tuners, etc, but we did not consider all of these areas individually as these may vary widely among MSOs and over time.

We simply will assume that Video Services will use all available capacity not being used by the High-Speed Internet Services. We will however make some forecast for what could be considered a minimum allocation of capacity for an MSO deliveredvideoservice, below are our Video assumptions and traffic forecast.

Video Assumptions	
Take-rate of the service	60%
Viewers are actively watching a program during the busy-hour/busy-day	60%
Average video viewers per active home	2
Linear Service (Broadcast)	0%
On-Demand Service (Unicast) (this worst-case assumption creates the biggest BW challenge)	100%
Average program bandwidth (assumes mix of SD, HD, and 3D in MPEG4)	10 Mbps
HHP Fiber Node or Service Group (SG)	250

FIGURE 2: VIDEO ASSUMPTIONS FOR FUTURE CAPACITY PLANNING

FIGURE 3: VIDEO TRAFFIC ALLOCATIONS FOR FUTURE CAPACITY PLANNING

Video Calculatio	
250 HHP/Node * (0.6 take-rate) * (0.6 active) * (2 viewers/active home) * (10 Mbps/viewer)	1.8 G or 3.6 G

The video service is projected be a unicast offering and the model essentially will always reserve or allocate 12 Mbps per video subscriber (1800 Mbps / 150 video subscribers) as illustrated in the figure 2 and 3. However, like today video services will dominate the spectrum allocation compared to High-Speed Internet for nearly the entire decade. The modeling in the remainder of this paper assumes that video services will consume all of the bandwidth that High-Speed Internet does not require, however the model reserves the 12 Mbps per video sub as a minimum allocation for a video service offered by the MSO, unless otherwise stated. Certainly the MSO's high-speed data subscribers may use the data network to view video content on devices like tablets, handhelds, TVs, PCs, and other devices.

High-Speed Internet Service Tier Offered (Downstream and Upstream)

The network traffic estimates need to consider the downstream and upstreamhigh-speed Internet service tier, in other words the data speed package the MSO offers to consumers. The highest data speed offered in either direction is a determining factor for sizing the network. TheHigh-Speed Internet service tier and traffic will grow considerably during this decade moving from perhaps four6 MHz channels downstream, which is less than 4% of the MSO'stotal spectrum allocation and may grow to perhaps 40-50% in the next 10 years. The high-speed Internet service tier offering will be a key contributor to overall bandwidth drivers. The figure below shows a thirty-year history of the max bandwidth offered or available to consumers. This figure also attempts to predict the max service tier we may see in the future, if the growth trend aligns with the preceding years. Perhaps we will allocate the entire 750 MHz downstream spectrum or equivalent to Internet services by 2023. As illustrated in the figures below, the downstream and upstream modeling began with the dial-up era, moving into the broadband era and now the DOCSIS channel bonding and PON eras. This model assumes a 50% CAGR for the Internet service tier.

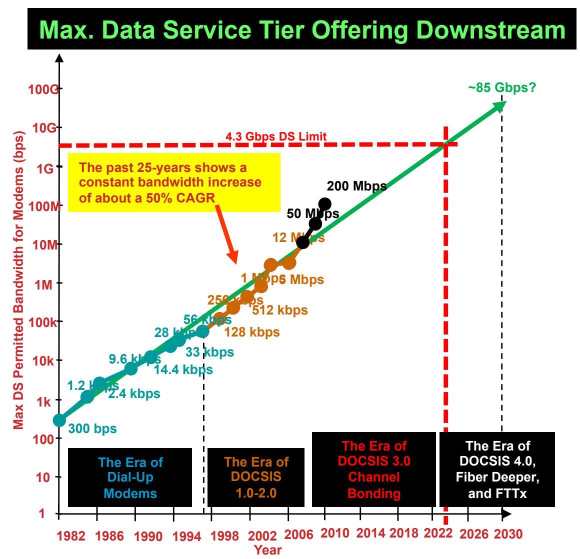


FIGURE 4: MAX INTERNET DATA SERVICE TIER OFFERING DOWNSTREAM

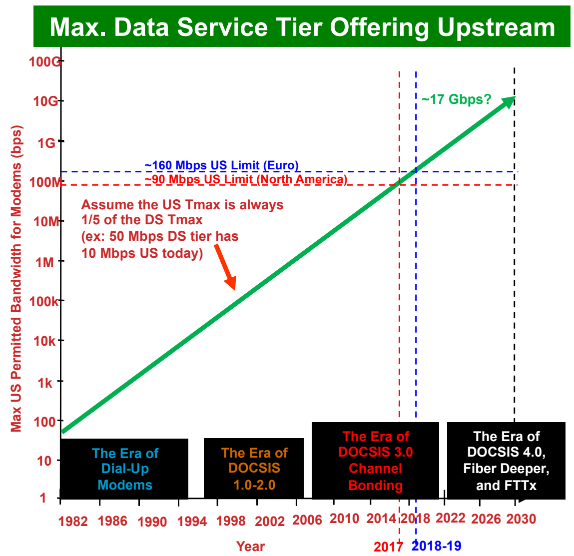


FIGURE 5: MAX INTERNET DATA SERVICE TIER OFFERING UPSTREAM

The table below captures the year-by-year predictions of the downstream and upstream service projections from the figures above. This table will be used for the capacity requirements found in the Network Utilization and Capacity Planning section later in this document. It is uncertain if the Max Service Tier trends will continue for the next 15 years at a 50% CAGR. The service offerings will, from time-to-time, not maintain alignment with the projections. Typically leaps above the line happen when there are major technology advances, such as dial-up to cable modem/DSL, then to channel bonding and PON.So, if we analyze where the telecom industry is today with their max downstream and upstream service offerings this may not be in alignment with the predictions.

Year	Downstream Max Service Tier	Upstream Max Service Tier
2010	26	5
2011	38	8
2012	58	12
2013	86	17
2014	129	26
2015	194	39
2016	291	58
2017	437	87
2018	655	131
2019	983	197
2020	1,474	295
2021	2,211	442
2022	3,317	663
2023	4,976	995
2024	7,464	1,493
2025	11,196	2,239

FIGURE6: COMBINED INTERNET MAX SPEED PREDICTIONS

There has been a significant increase in the services offered resulting in an up tick off the linear progression. Additionally, announcements from cable operator Virgin Media of an Internet speed trial of up to 1.5 Gbps downstream and 150 Mbps upstream [1] and Verizon reportedly exploring plans to upgrade its FiOS system to XG-PON, the 10 Gbps downstream and 2.5 Gbps upstream technology may further move the model higher [2]. The rollout of downstream channel bonding was a key contributor to the expansion of the service offering as well as PON. As upstream channel bonding is deployed in the near term we expect an expansion of the upstream max service tier to increase as well, perhaps initially at a higher rate that the 50% CAGR as the model has captured over the last thirty years. This is critical information for the network planners; any acceleration in the service tier offered would change the predictions we have captured in this paper affecting the estimated migration timeline. The expansion of service tier often leads to higher per customer bandwidth usage or network traffic.

High-Speed Internet Bandwidth Per Subscriber (Downstream and Upstream)

In addition to the service tier offered to consumers, the actual usage of the network by the consumers is a critical factor for network planners. This is known as the bandwidth per subscriber (BW per Sub). The determination of bandwidth per sub, is a measurement of the total amount of bandwidth or traffic in a serving area divided by the number of consumers in the serving area, this may be measured during busy hour(s).The bandwidth per subscriber is measured in the downstream and upstream direction. The downstream is currently measured at a 220 kbps per subscriber and the upstream at 36 kbps per subscriber, as illustrated in figure 7 and 8. The rate of growth is projected at a 50% CAGR. The bandwidth per subscriber and the CAGR may vary, however these numbers seem reasonable for the North American market. These numbers are used for planning purposes in this paper.

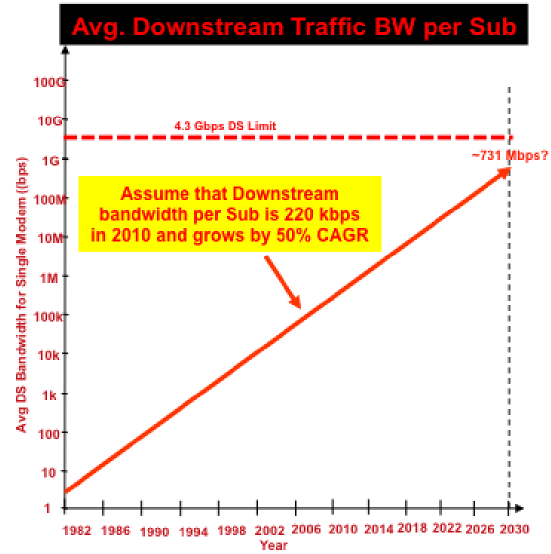


FIGURE 7: DOWNSTREAM BANDWIDTH PER SUBSCRIBER

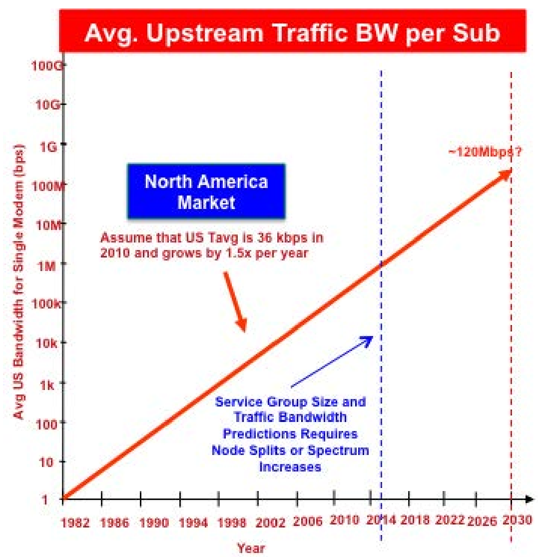


FIGURE 8: UPSTREAM BANDWIDTH PER SUBSCRIBER

Summaries for Service Tier and Traffic Growth Estimates

The video service offering will evolve over time from broadcast to unicast. The model plans for 3.6 Gbps of video traffic in a 500 HHP service group and 1.8 Gbps in a 250 HHP serving group. The model will use both High-Speed Internet projections, like the Service Tier Offering and bandwidth per subscriber to predict Network Utilization and Capacity Planning.

NETWORK CAPACITY

The network capacity of the cable access network is determined by the amount of spectrum available and the data rate possible within the spectrum. There are many factors that determine the amount of spectrum and data rate possible such as the location of the spectrum, noise, PHY/MAC layer efficiencies possible, and several other factors. We have modeled several spectrum allocation options as well as data rate possibilities. The analysis below captures the PHY layer throughput assumptionsof DOCSIS QAM (Quadrature Amplitude Modulation) forthe downstream and upstream. The analysis also considers a DOCSIS OFDM (orthogonal frequency-division multiplexing) based system that could emerge in the future. These are again PHY layer efficiency estimates additional MAC layer overhead has not been calculated. The authors wish to express that ARRIS is not aware of industry plans to adopt OFDM in the DOCSIS standard. Additionally, ARRIS has conducted internal studies for years examining the possibilities of an OFDM based systembut we have no plans to incorporate OFDM based systems into our products.

PHY Layer Throughput Assumptions

DOCSIS QAM Based

There are three figures which captures the assumptions of DOCSIS QAM based system. The first calculates the DOCSIS 256QAM downstream, figure 9. The remaining two tablesmodels the upstream using DOCSIS 64QAM and DOCSIS 256QAM each assumes ATDMA, figure 10-11. These tables measure the PHY layer spectral efficiency of DOCSIS QAM based solutions. These are used to calculate the network capacity of the cable network considering several spectrum options.

Downstream (assume Annex B 256QAM)
Bits/symbol = 8
In 1 Hz of bandwidth (with a raised cosine roll-off factor of 0.12), we can transmit 1/1.12= ~0.89 symbols/sec
We must include J.83 bandwidth overhead reduction factors:
•J.83 FEC addition factor = 122/128
•J.83 Trellis addition factor = 38/40
•MPEG-TS Header addition factor = 184/188
Resulting spectral efficiency = (8 bit per symbol)*(0.89 symbol per sec/Hz) *(122/128)*(38/40)*(184/188)= 6.3 bps/Hz

FIGURE9: DOWNSTREAM DOCSIS 256QAM

Upstream (assume 64QAM)
Bits/symbol = 6
-In 1 Hz of bandwidth (with a raised cosine roll-off factor of 0.25), we can transmit 1/1.25= ~0.8 symbols/sec
-We must include other bandwidth overhead reduction factors (assume 175-byte packet = 234 symbols):
-Guard-band= ~8 symbols
-Preamble = ~26 symbols
-FEC = ~6 symbols
-Bandwidth reduction factor = [1-(8+26+6)/(234+8+26+6)] = 0.854
-Resulting spectral efficiency = (6 bit per symbol)*(0.8 symbol per sec/Hz)*0.854 = 4.10 bps/Hz

FIGURE 10: UPSTREAM DOCSIS 64QAM

Upstream (assume 256 QAM)
Bits/symbol = 8
-In 1 Hz of bandwidth (with a raised cosine roll-off factor of 0.25), we can transmit 1/1.25= ~0.8 symbols/sec
-We must include other bandwidth overhead reduction factors (assume 175-byte packet = 175 symbols):
-Guard-band= ~8 symbols
-Preamble = ~26 symbols
-FEC = ~4 symbols
-Bandwidth reduction factor = [1-(8+26+4)/(175+8+26+4)] = 0.82
-Resulting spectral efficiency = (8 bit per symbol)*(0.8 symbol per sec/Hz)*0.82 = 5.248 bps/Hz

FIGURE 11: UPSTREAM DOCSIS 256QAM

A key take away is performance gap between 256QAM PHY and 64QAM layer efficiencies. The assumptions for 64QAM at 4.10 bps/Hz would require 28% more spectrum and DOCSIS channels to maintain the equivalent PHY layer throughput. The use of DOCSIS 256QAM for the upstream is not part of the DOCSIS standards, however some CMTS and CM products support this modulation profile in hardware. ARRIS believes that the DOCSIS specifications should be modified to include 256QAM upstream as well as 1024QAM in the upstream and downstream.

DOCSIS OFDM Based

For analysis purposes the paper provides measurements using OFDM/OFDMA, again OFDM is not part of the DOCSIS standards.

OFDM/OFDMA 1024 QAM Analysis	
FFT size	4096
Subcarriers	3800
Subcarriers spacing	50 KHz
Bandwidth	190 MHz
Modulation	1024 QAM (10 bits/symbol)
Synchronization overhead	8.33% (1 symbols per 12 symbols frame (11))
Cycle Prefix	1/8
OFDMA Symbol time	22.5 us
FEC	0.85
Aggregate PHY throughput	1315 Mbps
PHY Efficiency bits/Hz	6.92
Computation:	
3800*10*11= 418,000 bit per frame	
418,000 / (12*22.5) 1588 Mbps	
1548 * .85 = 1315 Mbps	
1315/190 = 6.92	

FIGURE 12: OFDM 1024QAM ANALYSIS

OFDM/OFDMA 256 QAM Analysis	
FFT size	4096
Subcarriers	3800
Subcarriers spacing	50 KHz
Bandwidth	190 MHz
Modulation	256 QAM (8 bits/symbol)
Synchronization overhead	8.33% (1 symbols per 12 symbols frame (11))
Cycle Prefix	1/8
OFDMA Symbol time	22.5 us
FEC	0.85
Aggregate PHY throughput	1052 Gbps
PHY Efficiency bits/Hz	5.54
Computation:	
3800*8*11= 334,400 bit per frame	
334,400 / (12*22.5) 1238 Mbps	
1238 * .85 = 1052 Mbps	
1052/190 = 5.54	

FIGURE 13: OFDM 256QAM ANALYSIS

In the figures above 256QAM was analyzedusing estimates for PHY layer efficiency comparing DOCSIS single carrier 256QAM and DOCSIS OFDM 256QAM. The analysis for the OFDM based approach shows a slightly higher PHY layer efficiency. The actual performance of either in real-world deployments is unknown. There are many attributes and assumptions than can be modified; we used an estimate that we considered to be fair for single carrier QAM and OFDM. These are subject to debate.

Downstream Capacity

The most critical determination for the capacity of the network is the amount of spectrum available. The determination of the downstream capacity will assume the eventual migrations to an all IP based technology. The migration to all IP on the downstream which will optimize the capacity of the spectrum providing the versatility to use the network for any service type and provide the means to compete with PON and the flexibility to meet the needs of the future. This table provides capacity projections considering: 1) the upstream spectrum split, 2) the use of DOCSIS QAM or DOCSIS OFDM, 3) several downstream spectrum allocations from 750 MHz to 1002 MHz. Certainly there are other spectrum options that could be considered such as moving the downstream above 1 GHz and other spectrum options for the upstream. This table will calculate the estimated downstream PHY layer capacity using several spectrum options.

Split Type	MSO Downstream Channel Bonding Bandwidth Summaries	Spectrum Summaries				Technology Data Rates		Total Capacity Data Rate Usable (Mbps)
		Total Downstream Spectrum Available	Total Downstream Spectrum Usable for Channeling Bonding	Spectrum Usable for DOCSIS QAM	Spectrum Usable for DOCSIS OFDM	DOCSIS QAM Usable Data Rate Per MHz 256 QAM)	DOCSIS OFDM Usable Data Rate Per MHz (OFDM w/ LDPC))	
Mid-split	750 MHz (DOCSIS QAM) with Mid-split	645	645	645	0	6.3	7	4064
	750 MHz DOCSIS OFDM OFDM w/ LDPC with Mid-split	645	645	0	645	6.3	7	4515
	860 MHz (DOCSIS QAM) with Mid-split	755	755	755	0	6.3	7	4757
	860 MHz DOCSIS OFDM OFDM w/ LDPC with Mid-split	755	755	0	755	6.3	7	5285
	870 MHz (DOCSIS QAM) with Mid-split	765	765	765	0	6.3	7	4820
	870 MHz DOCSIS OFDM OFDM w/ LDPC with Mid-split	765	765	0	765	6.3	7	5355
	1002 MHz (DOCSIS QAM) with Mid-split	897	897	897	0	6.3	7	5651
	1002 MHz DOCSIS OFDM OFDM w/ LDPC with Mid-split	897	897	0	897	6.3	7	6279
High-Split (200)	750 MHz (DOCSIS QAM) with High-Split (200)	492	492	492	0	6.3	7	3100
	750 MHz DOCSIS OFDM OFDM w/ LDPC with High-Split (200)	492	492	0	492	6.3	7	3444
	860 MHz (DOCSIS QAM) with High-Split (200)	602	602	602	0	6.3	7	3793
	860 MHz DOCSIS OFDM OFDM w/ LDPC with High-Split (200)	602	602	0	602	6.3	7	4214
	870 MHz (DOCSIS QAM) with High-Split (200)	612	612	612	0	6.3	7	3856
	870 MHz DOCSIS OFDM OFDM w/ LDPC with High-Split (200)	612	612	0	612	6.3	7	4284
	1002 MHz (DOCSIS QAM) with High-Split (200)	744	744	744	0	6.3	7	4687
	1002 MHz DOCSIS OFDM OFDM w/ LDPC with High-Split (200)	744	744	0	744	6.3	7	5208
Top-split (900-1050)	750 MHz (DOCSIS QAM) with Top-split (900-1050)	696	696	696	0	6.3	7	4385
	750 MHz DOCSIS OFDM OFDM w/ LDPC with Top-split (900-1050)	696	696	0	696	6.3	7	4872
Top-split (1250-1750)	1002 MHz (DOCSIS QAM) with Top-split (1250-1750)	948	948	948	0	6.3	7	5972
	1002 MHz DOCSIS OFDM OFDM w/ LDPC with Top-split (1250-1750)	948	948	0	948	6.3	7	6636

FIGURE 14: DOWNSTREAM NETWORK CAPACITY ESTIMATES

The model used a lower order modulation assumption for QAM but high order modulations are certainly possible. The spectrum capacity of single carrier QAM and OFDM may actually be similar, however more real-world analysis is needed to accurately measure the performance of both technologies.

Upstream Capacity

The upstream capacity measurements are more complicated and not as straightforwardas the downstream capacity projections. In the table below, many of the spectrum split options were evaluated considering several PHY layer options and modulation schemeswithin each spectrum split.

These are some key assumptions about the upstream capacity estimates:

- Sub-split spectrum region considered 22.4 MHz eligible for channel bonding
- Sub-split spectrum was calculated with only DOCSIS 3.0 64QAM
- Sub-split channel bonding spectrum counted in capacity summaries with any new spectrum split
- All estimates use PHY layer efficiency estimates additional MAC layer overhead has not been calculated.

An important assumption is that the upstream capacity measurements assume that spectrum blocks from the sub-split region and any new spectrum split will all share a common channeling bonding domain. This is essentially assuming that backwards compatibility is part of the upstream capacity projections. The upstream capacity projections for each split will assume DOCSIS QAM and if adopted in the future DOCSIS OFDM based systems will all share the same channel-bondinggroup. This will allow for previous, current, and future investments made by the MSO to be applied to a larger and larger bandwidth pipe or overall upstream capacity. If backward compatibility were not assumed the spectrum options would have to allocate spectrum for DOCSIS QAM and separate capacity for any successor, resulting in a lower capacity throughput for the same spectrum allocation and would compress the duration of time the same spectrum may be viable to meet the needs of the MSO.

The upstream capacity measurements found in figure 15compares various spectrum splits, modulation types as well as single carrier QAM and OFDM. The spectrum splits found in the table include Sub-split, Mid-split, High-split (200), Top-split (900-1050), and Mid-split with Top-split (900-1050). The Top-split options above 1.2 GHz were not calculated in this table.

The spectrum split, PHY, and modulation type are examined in figure 15 to determine the ―Total PHY Channel Bond Capacity Usable", found on the last column. This was intended to delineate between single carrier QAM and OFDM omitting the MAC layer throughput calculations. Traffic engineering and capacity planning should consider the MAC overhead and headroom for peak periods. Similar to the examination of the downstream capacity projections above, the upstream projections illustrate that OFDM has more capacity compared to QAM; this may not be the case in real-world deployments.

Split Type	MSO Upstream Channel Bonding Bandwidth Summaries	Spectrum Summaries					Technology Data Rates Per MHz			Channel Bond Data Rate Capacity				Total PHY Channel Bond Capacity (Usable)
		Total US Spectrum	Total Upstream Spectrum Usable for Channeling Bonding	Sub-split Spectrum Likely Only used for DOCSIS 3.0	NEW Spectrum Usable for DOCSIS QAM	NEW Spectrum Usable for DOCSIS OFDM	DOCSIS QAM 64 Usable Data Rate Per MHz	DOCSIS QAM 256 Usable Data Rate Per MHz	DOCSIS OFDM Usable Data Rate Per MHz	DOCSIS QAM 64 Sub-split Spectrum Likely Only used for DOCSIS 3.0	DOCSIS QAM 64 Total Capacity Data Rate Usable (Mbps)	DOCSIS QAM 256 Total Capacity Data Rate Usable (Mbps)	DOCSIS OFDM Total Capacity Data Rate Usable (Mbps)	
Sub-split	DOCSIS QAM 64	37	22.8	22.4			4.1	5.2	6.28	92	-	-	-	92
Mid-Split	DOCSIS QAM 64	80	65.8	22.4	43.2		4.1	5.2	6.92	92	177	-	-	269
	DOCSIS QAM 256	80	65.8	22.4	43.2		4.1	5.2	6.92	92		225		316
	DOCSIS OFDM	80	65.8	22.4		43.4	4.1	5.2	6.92	92			300	392
High-split (200)	DOCSIS QAM 64	195	180.8	22.4	158.4		4.1	5.2	6.92	92	649	-	-	741
	DOCSIS QAM 256	195	180.8	22.4	158.4		4.1	5.2	6.92	92		824		916
	DOCSIS OFDM	195	180.8	22.4		158.4	4.1	5.2	6.92	92			1,096	1,188
Top-split (900-1050)	DOCSIS QAM 64	187	172.8	22.4	150.4		4.1	5.2	5.54	92	617	-	-	708
	DOCSIS QAM 256	187	172.8	22.4	150.4		4.1	5.2	5.54	92		782		874
	DOCSIS OFDM	187	172.8	22.4		150.4	4.1	5.2	5.54	92			833	925
Mid-Split + Top Split (900-1050)	Mid-split + Top-split (DOCSIS QAM)	230.4	216.2	22.4	193.6		-	-	-	92	617	225		933
	Mid-split + Top-split (DOCSIS QAM and OFDM)	230.4	216.2	22.4		193.6	-	-	-	92		225	833	1,150

FIGURE 15: UPSTREAM NETWORK CAPACITY ESTIMATES

A very important point is that the network architecture and performance characteristics of the plant in the real world will determine the spectrum capacity to be used. The determination of the network architectures that may work at various spectrum splits, modulations, and number of carriers was a critical finding of this report. We have modeled the network architecture and performance assumptions to estimate the modulation and capacity possible for each spectrum split. This allowed us to determine the overall requirements and impacts to cost of the various split options and the ability for the spectrum split to meet the business needs of the MSO. The network architecture requirements and impacts for each spectrum split will be found in the sections called ―Network Technology and Architecture Assessment" and the cost assessment section called ―Cost Analysis".

NETWORK UTILIZATION AND CAPACITY PLANNING

If you are wondering how long a spectrum split may last or the sizing of the service group in the downstream or upstream this sections will provide some estimates for consideration. In this section of the report the network utilization estimates and capacity planning forecasts are examined. This section will predict the year and potential driver for network change. The information found in this section will be based on the findings of the preceding sections, which forecasted the service usage for video and High-Speed Internet as well as network usage on a per-subscriber basis. Additionally this section will use the network capacity estimates for the downstream and upstream.

An important attribute of cable systems is that the HFC optical and RF network as well as the data access layer network like the DOCSIS CMTS allows for upstream and downstream capacity upgrades may be made separately, where and when needed per service group. The report separates the utilizations and capacity planning results for the downstream and upstream to take advantage of this key feature.A key factor for the calculations will be the service tier growth forecast and the per subscriber usage, which have been separated as well. As stated previously these are just predictions and there are many factors that may influence change and the rate of change, so these findings should just be used for discussion purposes only.

The Downstream

The downstream network capacity drivers will be separated into High-Speed Internet Max Service Tier plus Video Traffic Predictions and another measurement will be forEstimated Bandwidth per Service Group.

Capacity Planning for High-Speed Internet Max Service Tier plus Video Traffic

The upstream and downstream High-SpeedInternet service tier growth by year and direction is used to forecast the date when the downstream may be at capacity, see figure 16. The HFC downstream capacity assumptions will usethe equivalent to a 750 MHz system, approximately 700 MHz of usable downstream spectrum to measures the date the capacity threshold is reached. The table shows that the MSO may offer a 2.2 Gbps Downstream High-Speed Data Internet service tier and support capacity for a managed video service package of about 1.8 Gbps, the year 2021. Additionally, if the High-Speed Internet growth rates remain at a 50% CAGR, that by about the year 2023 the existing downstream spectrum would be entirely needed for High-Speed Internet Services. It again should be stated that these are just prediction for the next decade or more, it is uncertain if speeds would be desired or offered at the levels shown.

Year 2021: HSD Max Service Tier Prediction of 2.2 Gbps Plus 1.8 Gbps of MSO Video Traffic is sustainable (HSD Service Tier may reach 2.5 Gbps)

Year 2023: HSD Max Service Tier Prediction Consumes All Downstream Spectrum thus additional Spectrum Above 1 GHz is required or FTTx

Year	Downstream Max Service Tier	Upstream Max Service Tier
2010	26	5
2011	38	8
2012	58	12
2013	86	17
2014	129	26
2015	194	39
2016	291	58
2017	437	87
2018	655	131
2019	983	197
2020	1,474	295
2021	2,211	442
2022	3,317	663
2023	4,976	995
2024	7,464	1,493
2025	11,196	2,239

Figure 16: Downstream Service Tier and Network Capacity Estimates

Estimated Bandwidth Per Service Group (Downstream)

There are several contributing factors used to forecast the capacity for a service group. They include, the size of the service group, take rate of the services, estimated per subscriber data usage, and the allocation of capacity for an MSO managed video service offering. The model defines a service group as a collection of HHP beginning at 1,000 HHP to 63 HHP. We use the modeling projections from the previous section and apply the capacity capabilities of the 750 MHz system or equivalent. The analysis predicts that a 500 HHP service group will meet the capacity needs for the future and a migration to each 250 HHP service group will last a full decade. The estimated bandwidth per service group is a measures based on the high-speed Internet user traffic, the call outs in figure 17 capture the video allocation estimates used for the 500 HHP and 250 HHP service group to estimate the date of the migration to a smaller service group.

Year 2017-18: Downstream 500 HHP Node is at capacity (assuming 250 HSD Subs with 2.1 Gbps of Traffic Plus 3.6 Gbps of MSO Video Traffic)

Year 2021: Downstream 250 HHP Node is at capacity (assuming 125 HSD Subs with 2.3 Gbps of Traffic Plus 1.8 Gbps of MSO Video Traffic)

DOWNSTREAM								
ESTIMATED BANDWIDTH PER SERVICE GROUP (Shown in Mbps)								
Year	Avg. Downstream Bandwidth Per Sub (Megabits)	HHP Per SG	1000	500	250	125	62.5
		% of HSD per SG	50%	50%	50%	50%	50%
		Customers Per SG	500	250	125	63	31
2010	0.22		110	55	28	14	7
2011	0.33		165	83	41	21	10
2012	0.50		248	124	62	31	15
2013	0.74		371	186	93	46	23
2014	1.11		557	278	139	70	35
2015	1.67		835	418	209	104	52
2016	2.51		1,253	626	313	157	78
2017	3.76		1,879	940	470	235	117
2018	5.64		2,819	1,410	705	352	176
2019	8.46		4,229	2,114	1,057	529	264
2020	12.69		6,343	3,172	1,586	793	396
2021	19.03		9,515	4,757	2,379	1,189	595
2022	28.54		14,272	7,136	3,568	1,784	892
2023	42.82		21,408	10,704	5,352	2,676	1,338
2024	64.22		32,112	16,056	8,028	4,014	2,007
2025	96.34		48,168	24,084	12,042	6,021	3,011
2026	144.50		72,252	36,126	18,063	9,032	4,516
2027	216.76		108,379	54,189	27,095	13,547	6,774
2028	325.14		162,568	81,284	40,642	20,321	10,161
2029	487.70		243,852	121,926	60,963	30,482	15,241
2030	731.56		365,778	182,889	91,445	45,722	22,861

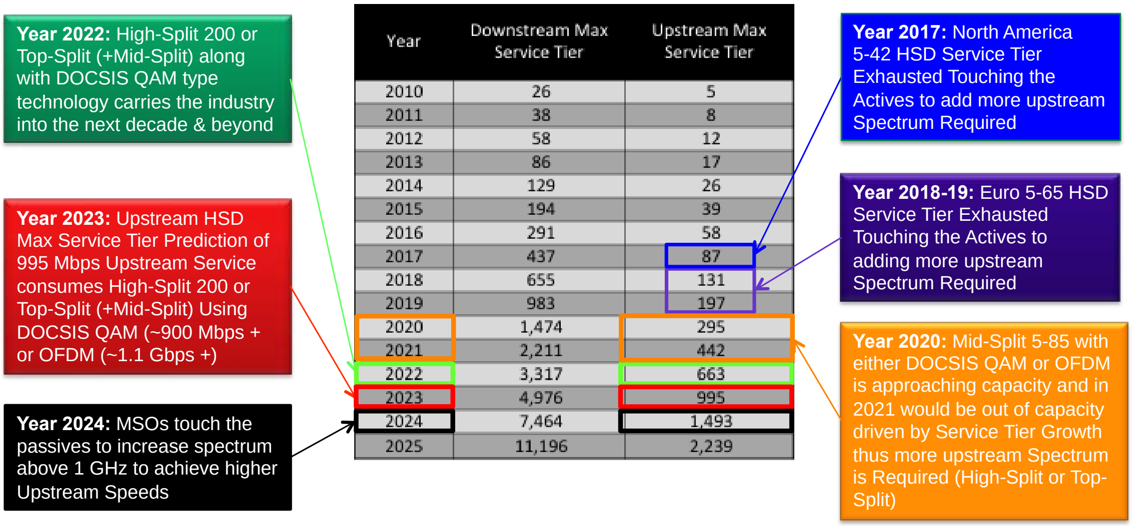
Figure 17: Downstream Service Group Capacity Planning Estimates

The Upstream

The network utilization and capacity planning forecast of the upstream may meet the capacity limits of the sub-split 5-42 in North America and Europe's 5-65 within this decade. Surprisingly, it could be the network utilization or traffic at a 500 HHP service group which meets the throughput capacity of the sub-split spectrum. The section below examines the spectrum split options and the timing impacts.

Capacity Planning for High-Speed Internet Max Service Tier

This section captures the duration of time each of the upstream split options under examination will last. In figure 18, the upstream max service tier used in the Service Tier section earlier in the paper is assessed with the network capacity data estimates for the split options found in the immediate preceding section. The service tier estimates along with the capacity estimates for each split option is used to predict the year each upstream split option will be at or near capacity.We again wish to point out that these are just estimates used for planning purposes.



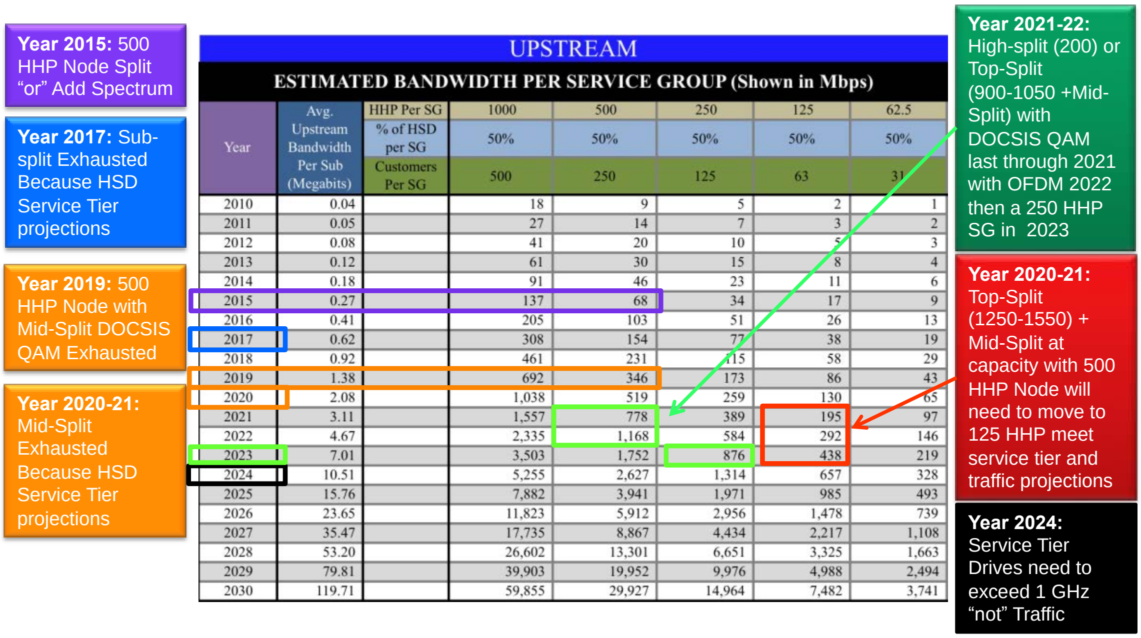
Year	Downstream Max Service Tier	Upstream Max Service Tier
2010	26	5
2011	38	8
2012	58	12
2013	86	17
2014	129	26
2015	194	39
2016	291	58
2017	437	87
2018	655	131
2019	983	197
2020	1,474	295
2021	2,211	442
2022	3,317	663
2023	4,976	995
2024	7,464	1,493
2025	11,196	2,239

Figure 18: Upstream Service Tier and Network Capacity Estimates

Estimated Bandwidth Per Service Group (Upstream)

It is estimated that the major factor that may cause pressure on the upstream split options will be the capacity caused by traffic usage or bandwidth per service group. The table below estimates the traffic generated by the users in a service group. The network capacity estimates of each split option is then used to determine the year and service group size that may sustain a given split option. In the table below and discussed later in this paper are some assumptions to the usage and indeed relation of the spectrum options to the service group size at the upstream optical domain level. This table highlights the year, split option, and service group size that may sustain the traffic load and this table will also mention when service tier projection will force a split or spectrum increase.

It should be observed that a 500 HHP service group with 250 subscribers will last for a decade or more, however this assumes that spectrum increase like that to mid-split may happen in the year 2015 driven by traffic growth, this will be 2 years before sub-split is projected to run out because of service tier growth. Thus moving to mid-split in 2015 would allow the 500 HHP service group to be leverage until about the year 2019. If high-split (200) is added in 2019 this may allow the 500 HHP service group to remain until 2021-2022.



UPSTREAM

ESTIMATED BANDWIDTH PER SERVICE GROUP (Shown in Mbps)

Year	Avg. Upstream Bandwidth Per Sub (Megabits)	HHP Per SG	1000	500	250	125	62.5
		% of HSD per SG	50%	50%	50%	50%	50%
		Customers Per SG	500	250	125	63	31
2010	0.04		18	9	5	2	1
2011	0.05		27	14	7	3	2
2012	0.08		41	20	10	5	3
2013	0.12		61	30	15	8	4
2014	0.18		91	46	23	11	6
2015	0.27		137	68	34	17	9
2016	0.41		205	103	51	26	13
2017	0.62		308	154	77	38	19
2018	0.92		461	231	115	58	29
2019	1.38		692	346	173	86	43
2020	2.08		1,038	519	259	130	65
2021	3.11		1,557	778	389	195	97
2022	4.67		2,335	1,168	584	292	146
2023	7.01		3,503	1,752	876	438	219
2024	10.51		5,255	2,627	1,314	657	328
2025	15.76		7,882	3,941	1,971	985	493
2026	23.65		11,823	5,912	2,956	1,478	739
2027	35.47		17,735	8,867	4,434	2,217	1,108
2028	53.20		26,602	13,301	6,651	3,325	1,663
2029	79.81		39,903	19,952	9,976	4,988	2,494
2030	119.71		59,855	29,927	14,964	7,482	3,741

Figure 19: Upstream Service Group Capacity Planning Estimates

The figure 19, does not capture all of the options but may be used for planning based on traffic forecasts. The selection of Top-split 900-1050 with a 500 HHP service group has an estimated capacity of ~700 Mbps coupled with the sub-split. The Top-split 900-1050 option with Mid-split may yield a capacity of ~930 Mbps in a 500 HHP service group given the assumption documented in the following section, see figure 23. The use of the Top-Split options at 1250-1550 will not be able to use the high order modulation if we assume a 500 HHP service group, however there is more spectrum available. The Top-split 1250-1550 option with Sub-split is estimated to have a capacity of ~500 Mbps and with Mid-split ~725 Mbps. All of the Top-split options are viable for a 500 HHP service group but at lower order modulation when compared to the low frequency return options of Mid-split or High-split. If we assume Mid-split is a first step and Top-split 900-1050 is consider yielding ~930 Mbps this has slightly more capacity than High-split (200) and the passives are not touched with this Top-split option and avoids the STB out of band communications challenge.

Summary of Capacity Planning

It is very important that the reader understands that our assumptions use HHP per service group, this may be a physical node or a logical node which uses segmentation to meet the sizing level. Another very important consideration is that the model assumes that over time that full spectrum would be allocated to the service group to meet the capacity projections for user traffic in the downstream. The upstream split options will have a direct relationship to the network architecture to include the size of the service group, number of actives, passives, cable portion of the network.

NETWORK TECHNOLOGY AND ARCHITECTURE ASSESSMENT

The goal of any cable operator is a drop in upgrade to add spectrum capacity when needed. This saves time and money in resizing the network such as node and amplifier location and spacing. Adding network elements or changing network element locations will impact cost for electrical powering requirements. Ideally, the upgrade would touch the minimum number of network elements to reduce cost and time to market. In the section, the technologies, systems and architecture options are explored. The paper will examine some of the pros and cons of several technologies and architectures, which could be used to provide additional capacity.

Overview of Important Considerations and Assumptions

This report has highlighted some important areas for network planners to consider while making the decisions for the next generation cable access network.

Avoidance of Small Node Service Groups or FTTLA

The analysis and conclusions found in this report indicates that the need for smaller node groups with few actives and passives such as Node +3 or even Fiber to the Last Active (FTTLA) is not required to meet capacity, service tier predictions or network architecture requirements for this decade and beyond.

500 HHP Node Long-Term Viability

Our analysis finds that upstream and downstream bandwidth needs may be met while leveraging a 500 HHP node service group for a majority of this decade and even beyond. The maintaining of a 500 HHP service group is of immense value to the MSOs. The ability to solve capacity changes while maintaining the node size and spacing enables an option for a drop-in capacity upgrade.

If the goal is to achieve 1 Gbps capacity upstream this may be achieved using a typical 500 HHP node service group with 30 actives and 200 passives, and over 6 miles of coax plant in the service area as fully described later in this paper, see figure 23.

The existing 500 HHP node has long-term viability in 750 MHz or higher systems providing enough downstream capacity to last nearly the entire decade. In the upstream a 500 HHP node is predicted to last until mid-decade when the sub-split spectrum may reach capacity and then a choice of node split, node segment or add spectrum like mid-split to maintain the 500 HHP service group are options. The physical 500 HHP node service group may remain in place with High-split (200)or Top-Split with Mid-split providing 900 to 1 Gbps capacity.

1 GHz (plus) Passives - A Critical Consideration for the Future

The industry will be considering several spectrum splits and special consideration should be made to the most numerous network elements in the outside plant, the passives. Avoiding or delaying modification to the existing passives will be a significant cost savings to the MSO. Below are key factors about the 1 GHz passives:

- Introduced in 1990 and were rapidly adopted as the standard
- This was prior to many major rebuilds of the mid-late 90s and early 2000s
- Prior even to the entry of 750 MHz optical transportandRF amplifiers/ products in the market place
- Deploymentof 1 GHz passives that would have more capacity than the electronics would have for nearly 15 years
- Passives are the most numerous network element in the Outside Plant (OSP)
- Volumes are astounding perhaps as many as 180-220 behind every 500HHP Node or about 30 per every plant mile (perhaps 40-50 Million in the U.S. alone)
- 1 GHz Passives may account for 85% of all passives in service today
- Vendor performance of the 1GHz Passives will vary and some support less than 1 GHz
- Our internal measurements indicate that most will support up to 1050 MHz
- Taps in cascade may affect capacity, thus additional testing is required

Assessment of the Passives

The authors believe that special consideration should be given to solutions that leverage the existing passive. This will avoid upgrades that may not be needed until the 2020 era when the MSOs may pursue spectrum above 1 GHz. If the 1 GHz passives are considered and the desired use is over 1 GHz we believe that 1050 MHz is obtainable. There will be challenges with AC power choke resonances, which may impact the use of these passive greater than 1050 MHz with predictably.

The Value of Time

The legacy STB out of band (OOB) communications which uses spectrum in the High-split area will be a problem for this split options; however a mid-split as the first step will provide sufficient capacity for nearly the entire decade according to our service and capacity predictions. The thinking is that another decade goes by and the legacy STBs may be few or out of the network all-together. If the STBs still remain in service another consideration is that these legacy STB may be retrieved and relocated to markets than may not need the advanced upstream spectrum options. Yet, another consideration is a down conversion of the OOB communications channel at the homes that have legacy two-way non-DOCSIS set-tops.

Overview Of Spectrum Splits

The spectrum allocation options should consider the impact to the overall end-to-end system architecture and cost. The solutions should also consider the timing of these changes as this may impact cost. The end-state architecture should be considered for this next touch to the HFC. We do not need to solve next decades problems now, however we should consider them as part of the analysis. The MSO has several spectrum split options available and some are examined in this paper. The figure below is an illustration of some of the spectrum split options; it also depicts a few other options, such as Top-split with Mid-split. In figure 20, the Top-split (900-1050) optionshas a 150 MHz block of spectrum allocated for guard band between 750-900 MHz and 150 MHz block of spectrum between 900-1050 MHz for upstream.

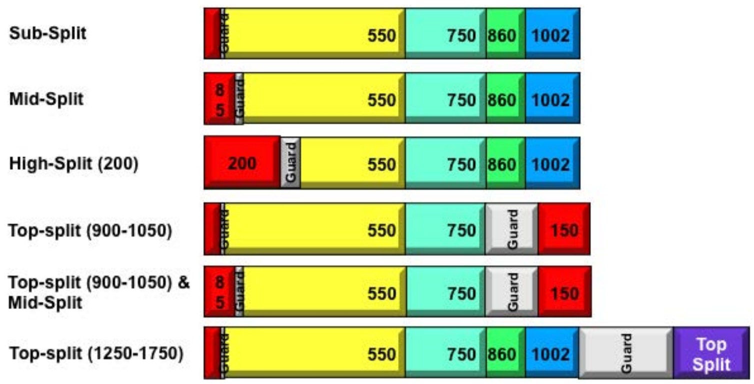


FIGURE 20: SPECTRUM ALLOCATION OPTIONS

Mid-split

Overview

The Mid-split Architecture is defined as 5-85 MHz upstream with the downstream starting at approximately 105 MHz; this may also be referred to as the 85/105 split. The mid-split architecture essentially doubles the current upstream spectrum allocation however this may triple or even quadruple the IP based capacity. The capacity increase in data throughput is a result of the high-order modulation and all of the new spectrum may be used for DOCSIS services, which is not the case with the sub-split spectrum that has generally accepted unusable spectrum and legacy devices consuming spectrum as well.

Pros

- Sufficient bandwidth to last nearly the entire decade
- DOCSIS QAM capacity approaching ~316 Mbps
- Avoids conflict with OOB STB Communications
- Lowest cost option
- High order modulation possible 256QAM perhaps higher
- The use of 256QAM translates to fewer CMTS ports and spectrum (using 64QAM would require approximately 28% more CMTS ports and spectrum)
- DOCISIS systems already support this spectrum (5-85)
- Some amplifiers support pluggable diplexer filter swap
- Some existing node transmitters and headend receives may be leveraged
- Does not touch the passives
- Upstream path level control is similar to the Sub-split (~1.4 times the loss change w/temp);

ThermalEqualizers EQT-85 enables +/-0.5 dB/amp delta

Cons

- Impacts Video Service (in low channels)
- Reduces low VHF video spectrum
- Throughput over 300 Mbps is less than the newer PON technologies

Assessment

The selection of Mid-split seems like an excellent first step for the MSOs. This split option has little impact to the video services and does not impact the OOB STBcommutations. This spectrum split may last nearly the entire decade, allowing time for the MSOs to assess future splits, if required, and the impacts to other split optionat that time.

High-split (200)

Overview

The High-split (200) Architecture is generally defined as 5-200 MHz with the downstream starting at approximately 250-258 MHz. Though other High-split options may be considered above 200 MHz these were not part of the examination. High-split is being considered because full or partial analog reclamation is underway or planned by cable operators. This will allow a smoother transition when considering consumption of existing analog spectrum. As with mid-split DOCSIS 3.0 specifications systems may be used; however, to take advantage of thespectrum, additional development is required.

Pros

- Operates effectively at a typical 500 HHP node group using 256QAM (see details in the sections later in this paper)
- The use of 256QAM translates to fewer CMTS ports and spectrum (using 64QAM would require approximately 28% more CMTS ports and spectrum)
- DOCSIS QAM capacity approaching ~916 Mbps
- DOCSIS OFDM capacity exceeds 1.1 Gbps
- Very low cost spectrum expansion option, especially considering similar capacity Top-split options (STB OOB cost was not considered in the analysis)
- Lowest cost per Mbps of throughput
- Some existing HFC Equipment supports High-split like node transmitters and headend receivers
- DOCISIS systems already support some of this spectrum (5-85)
- Passives are untouched
- High-split provides sufficient upstream capacity and the ability to maximize the spectrum with very high order modulation
- High-split (200) does not waste a lot of capacity on guard band
- Level control using Thermal Equalizers EQT-200 (~2.2 times Sub-split cable loss)

Cons

- Conflicts with OOB STB Communications if DOCSIS Set-top box Gateway (DSG) is not possible

- Takes away spectrum from Video Services (54-258 MHz)
- Takes away spectrum from Video devices (TVs and STBs)
- Potentially revenue impacting because of spectrum loss supporting analog video service tier
- Downstream capacity upgrade from 750 MHz to 1 GHz to gain back capacity lost to upstream

Assessment

The use of high-split would impact OOB Set-top Box communications for non-DOCSIS Set-top Gateways were not possiblein the upgraded service area. If the deployment of High-split (200) is planned later in time,this may allow these older STBs to be phased out. This split provides lots of bandwidth with a minimal amount of prime spectrum wasted for guard band.

If the main challenges with the use of High-split are overcome, this seems like the ideal location for the new upstream(technically). The economics are also compelling for High-split (200) against the other split options considering just the network access layer. If the STB Out of Band (OOB) and analog recoveryneed to be factored into to the High-split, the cost analysis will change.

Top-split (900-1050)with Sub-split or Mid-split

Overview

A new spectrum split called Top-split (900-1050)defines two separate spectrum bands, which may either use sub-split or mid-split plus the new spectrum region of900-1050 MHzfor a combined upstream band. The total upstream capacity may be either 187 MHz or 230 MHz depending on the lower band frequency return selected.The downstream would begin at either 54MHz or 105 MHz and terminate at 750 MHz in the current specification.All of these architectures will share a 150 MHz guard band between 750-900 MHz, this may vary in the end-state proposal howeverthese defined spectrum splits will be used for our analysis. The Top-split (900-1050) with Sub-split and with the Mid-spilt option are compared in a table called Spectrum Allocation Comparison, figure 21. The placement of additional upstream atop the downstream has been considered for many years. The Top-split (900-1050) approach may be similar to a Time Warner Cable trial called the Full Service Network in the mid 1990's, which is believed to have placed the upstream above the 750 MHz downstream. These are some of the pros and cons of Top-split (900-1050):

Pros

- Operates effectively at a typical 500 HHP node group but with no more than 64QAM (see details in the sections later in this paper)
- Top-split with Sub-split DOCSIS QAM capacity ~700 Mbps given a 500 HHP Node/Service Group
- Top-split with Mid-split DOCSIS QAM capacity ~933 Mbps given a 500 HHP Node/Service Group (equal to High-split)
- Top-split with Mid-split DOCSIS OFDM capacity exceeds 1.1 Gbpsgiven a 500 HHP Node/Service Group (equal to High-split)

- With Sub-split “no” video services, devices, and capacity is touched
- With Mid-split has a “low impact” to video
- STB OOB Communications are not affected
- Passives are untouched (only Top-splitthat avoids touching passives)
- Existing 750 MHz forward transmitters are leveraged

Cons

- No products in the market place to determine performance or accurate cost impacts
- The analysis estimates that Top-split (900-1050) is about 1.39 times the costof High-split (200) with a 500 HHP node architecture
- The analysis estimates that Top-split (900-1050) with Sub-split is about 2.4 times the costof High-split (200) with a 125 HHP node architecture
- Achieving similar capacity of High-split (200) and with Top-split (900-1050) with “Sub-split” will require a 125 HHP service group (node) which is a major cost driver.(Note the use of Mid-split with Top-split will provide 900+ Mbps more than High-split)
- Spectrum Efficiency is a concern because of guard band (wasted spectrum) and lower order modulation (less bits per Hz) resulting in lower throughput when measured by summing the upstream and downstream of Top-split (900-1050) and High-split using similar spectral range(see figure 21).
- High-split has 12% or more capacity for revenue generation when compared to Top-split (900-1050) plus Mid-split, this is because the guard band requirements waste bandwidth
- Will require more CMTS ports and spectrum when lower order modulation is used, perhaps 28% more CMTS base on the efficiency comparison estimates found on figures 10 and 11.
- Upstream is more of a challenge compared to using that same spectrum on the forward path
- Upstream is more of a challenge compared to using that same spectrum on the forward path (cable loss ~5x Sub-split, 2.3x High-split; ~+/-1 dB/amp level delta w/EQTs is unknown)
- Interference concerns with MoCA (simply unknown scale of impact but may affect downstream in same spectrum range)

Assessment

The Top-split (900-1050) options are being considered because option keeps the video network “as is” when considering sub-split and has marginal impact if mid-split is used. The Top-split 900-1050 option has additional benefits in that the Set-top box out of band (OOB) challenge is avoided and this optiondoes not touch the passives. This Top-split is estimated to cost more than the High-split; estimated at 1.3 to 2.4 times depending on the architecture selected. The MSOs will just begin to evaluate this option against the others.

Top-split (1250-1550) with Sub-split or Mid-split

Overview

The Top-split (1250-1550) Architecture will be defined as part of the 1250 – 1750 MHz spectrum band. In our analysis we limited the amount of spectrum allocated for data usage and transport to 300 MHz and defined the placementin the 1250–1550 MHz spectrum band. The allocationof 300 MHz provides similar capacity when compared to the other split option. The main consideration for this Top-split option is that it avoids consuming existing downstream spectrum for upstream and avoids the OOB STB communication channel.

Pros

- May operate at a typical 500 HHP node group but estimated to use QPSK, unless HHP is reduce to 125 then it is estimated the 16QAM may be used (described in more detail in the network architecture and cost sections of this report)
- Top-split 1250-1550 with Sub-split DOCSIS QAM capacity ~500 Mbps given a 500 HHP Node/Service Group
- Top-split 1250-1550 with Mid-split DOCSIS QAM capacity ~725 Mbps given a 500 HHP Node/Service Group
- Top-split 1250-1550 with Sub-split DOCSIS QAM capacity ~912 Mbps given a 125 HHP Node/Service Group
- Top-split 1250-1550 with Mid-split DOCSIS QAM capacity ~1136 Mbps given a 125 HHP Node/Service Group
- With Sub-split “no” video services, devices, and capacity is touched
- With Mid-split has a ―low impact” to video
- STB OOB Communication is not affected
- Placing the upstream spectrum beginning at 1250 MHz and up allows for the expansion of capacity without impacting the downstream

Cons

- Passives must be touched
- Smaller nodes or upstream Service Groups perhaps a 125 HHP will be required to approach or exceed the 1 Gbps speeds comparable to High-split (200) and Top-split (900-1050)
- Highest cost solution compared with High-split and Top-Split.
- The Top-split (1250-1550) with Sub-split is about 3 times the cost of High-split (200) for similar capacity without consideration to the DOCSIS layer.
- Will require more CMTS ports and spectrum when lower order modulation is used, perhaps 90-100% more CMTS base on the efficiency comparison estimates found on Figure 11 compared to the use of 16QAM estimated at 2.7 bps/Hz.
- No products in the market place to determine performance or accurate cost impacts.
- Return Path Gain Level Control: (cable loss >6x Sub-split, 2.8x High-

split; +/-2 dB/amp w/EQTs is unknown)

- Interference concerns with MoCA (simply unknown scale of impact but may affect downstream in same spectrum range)

Assessment

The Top-split (1250-1550) with Sub-split is about 3 times the cost of High-split (200).The placement of the return above 1 GHz requires the passives to be replaced or upgraded with a faceplate change. There are approximately 180-220 passives per 500 HHP node service group. It is estimated that the node service group of 500 HHP may be leveraged initially, however the requirements for higher capacity will force smaller node service group, which will addto the cost of the solution. The use of lower order modulations will require more CMTS upstream ports and more spectrum, which will impact the costs of the solution as well. Additionally, the conditioning of the RF components to support above 1 GHz may add to the costs of the solution. However determining the financial impacts of performing ―Above 1 GHz plant conditioning" is unknown and was not considered in the financial assessment found later in this report. If we consider the service and network capacity requirements for the upstream and downstream for the next decade and beyond, the cable industry should have sufficient capacity under 1 GHz, which is the capacity of their existing network.

Name of Spectrum Split	Mid-Split	High-Split (200)	Top-Split (900-1050)	Top-split (900-1050) & Mid-split	Top-split (1250-1550)	Top-split (1250-1550) & Mid-split)
Upstream Spectrum Range	5-85	5-200	5-42 & 900-1050	5-85 & 900-1050	5-54 & 1250-1550	5-85 & 1250-1550
Downstream Spectrum Range	> 105 MHz	> 258 MHz	54-750	105-750	54-1002	105-1002
Upstream Spectrum Bandwidth in MHz	80	195	187	230	337	380
Downstream Spectrum Bandwidth in MHz	897+	744+	696	645	948	897
Guard band Spectrum Allocation (wasted spectrum between US/DS)	20	58	162	170	260	268
Upstream PHY Layer Spectral Efficiency Capacity (assume QAM & 500 HHP SG) in Mbps	316	916	708	933	500	725
Downstream PHY Layer Spectral Efficiency Capacity (assume QAM) in Mbps	5651	4687	4385	4064	5972	5651
Total PHY Layer Spectral Efficiency Capacity Usable (Up+Down)	5967	5603	5093	4997	6472	6376
Video Service Impact	Yes	Yes	None, Assumes existing 750 System	Yes because Mid-split used in Lowband	None	Yes because Mid-split used in Lowband
Video Spectrum Loss Location (assuming 54 MHz-860MHz usable)	54-105	54-258	750-860	54-105	None	54-105
Video Spectrum Loss in MHz (assuming 54 MHz-860MHz usable)	51 Mhz	204 MHz	Assumes existing 750 System	51 MHz	None	51 MHz
OOB STB Communications ANSI/SCTE 55-2 2008 [4] (70 - 130 MHz) ANSI/SCTE 55-1 2009 [5] (70 - 130 MHz) Some STBs may be hard coded within the mid-split range (75.5 and 104.25 MHz)	Not likely, however some STBs may be hard coded within the mid-split range (75.5 and 104.25 MHz)	Impacted	No	No	No	No
Estimated Node Service Group Size per Return Laser in HHP (estimate & may vary)	500	500	500	500	500	500
Maximum number of Actives Supported (estimate & may vary)	30	30	30	30	30	30
Maximum number of Passives Supported (estimate & may vary)	200	200	200	200	200	200
Headend Optical Transmitter (Requirements, Replacement, Leverage)	Up to MSO	Up to MSO	Up to MSO	Up to MSO	Up to MSO	Up to MSO
Headend Optical Receivers (Requirements, Replacement, Leverage)	May Be Leveraged	May Be Leveraged	Replace	Replace	Replace	Replace
Nodes Optical Side (Requirements, Replacement, Leverage)	May Be Leveraged	May Be Leveraged	Replace	Replace	Replace	Replace
Node RF Side (Requirements, Replacement, Leverage)	Replace	Replace	Replace	Replace	Replace	Replace
Amplifiers (Requirements, Replacement, Leverage)	Best Case: Amp is removed from service if pluggable diplexer filter swap is supported (this is not a field upgrade). Worst case replace the Amp	Replace	Replace	Replace	Replace	Replace
Passives (Requirements, Replacement, Leverage)	Leverage	Leverage	Leverage	Leverage	Replace	Replace
House Amplifiers (Requirements, Replacement, Leverage)	Replace	Replace	Replace	Replace	Replace	Replace
Achieve Cable Modem Value not to exceed 65 dBmV	Yes	Yes	Yes	Yes	Yes	Yes
CPE Cost Impacts	$	$$	$$$	$$$	$$$	$$$
Interference Concerns with MoCA [6]: MoCA 1.0 and 1.1 Operating frequency 850 – 1500 MHz MoCA 1.1 Annex Expanded operating frequency of 500 MHz—1500MHz MoCA 2.0 Expanded operating frequency from 500 MHz – 1650 MHz	No Interference Concerns with New Upstream	No Interference Concerns with New Upstream	Yes Interference Concerns with New Upstream	Yes Interference Concerns with New Upstream	Yes Interference Concerns with New Upstream	Yes Interference Concerns with New Upstream
FM Radio Band, DTV and Aeronautical frequencies - avoidances of these bands reduces the overall spectrum bandwidth available for data services. Areas affected may have lower order modulation and smaller service group to attain desired capacity level.	No Interference Concerns with New Upstream	Yes Interference Concerns with New Upstream	No Interference Concerns with New Upstream	No Interference Concerns with New Upstream	No Interference Concerns with New Upstream	No Interference Concerns with New Upstream

[4], [5], [6]

FIGURE 21: SPECTRUM ALLOCATION COMPARISON

Characterization of RF Components

The network components that most affectsignals carried above 1 GHz are the coaxial cable, connectors, and taps. The characteristics of these components are critical, since the major goal in a next generation cable access network is to leverage as much of the existing network as possible.

Before getting into the specifics about the RF characterization and performance requirements, it is worthwhile to establish the quality of signals carried above 1 GHz and below 200 MHz. The bottom line is that while return path signals can be carried above 1 GHz, they cannot be carried with as high order modulation as is possible at lower frequencies. For example, if the goal is to meet similar return path data capacity the signal carriage above 1 GHz is possible using 16QAM with about 300 MHz of RF spectrum(47 channels of 6.4 MHz each). Whereas below 200 MHz 256QAM is possible (due to lower coaxial cable loss) and only 24 channels occupyingabout 180 MHz spectrum is required, using rough estimates. Additionally, the over 1.2 GHz solutions may require a 125 HHP service group to support 16 QAM, where as the 200 MHz solutions may use a 500 HHP service group, this is a key contributing factor to the cost deltas of the split options.

Path Loss and SNR

In a typical HFC Node + N architecture, the return path has many more sources for extraneous inputs, ―noise‖ than the forward path. This includes noise from all the home gateways, in addition to all the return path amplifiers that combine signals onto a single return path (for a non-segmented node). For now we will ignore the gateway noise, since in principle it could be made zero, or at least negligible, by only having the modem return RF amplifier turned on when the modem is allowed to ―talk‖.

The RF return path amplifier noise funneling effect is the main noise source that must be confronted; and it cannot be turned off! This analysis is independent of the frequency band chosen for the ―New Return Band‖ (e.g., Mid-split 5-85 MHz; High-split 5-200 MHz; or Top-split with UHF return), although the return path loss that must be overcome is dependent on the highest frequency of signals carried. For a first cut at the analysis, it suffices to calculate the transmitted level from the gateway required to see if the levels are even possible with readily available active devices. The obvious way to dramatically reduce the funneling noise and increase return path capacity is to segment the Node. That is not considered here to assess how long the network remains viable with a 4x1 configuration, a 500 HHP node service group.

The thermal mean-square noise voltage in 1 Hz bandwidth is kT, where k is the Stefan-Boltzmann constant, 1.38x10^-23 J/deg-K, and T is absolute temperature in degrees Kelvin. From this we have a thermal noise floor limit of -173.83 dBm/Hz. For a bandwidth of 6.4 MHz and 75-ohm system, this gives -57.0 dBmV per 6.4 MHz channel as the thermal noise floor. With one 7 dB noise figure amplifier in the chain, we wouldhave a thermal noise floor of -50 dBmV/6.4 MHz channel.

Two amplifiers cascaded would give 3 dB worse, four amplifiers cascaded give 6 dB worse than one. And since the system is balanced to operate with unity gain, any amplifiers that collect to the same point also increase the noise floor by 10*log(N) dB, where N is the total number of amplifiers in the return path segment. For a typical number of 32 distribution amplifiers serviced by one node, this is five doubles, or 15 dB above the noise from one RF Amplifier, or -35 dBmV/6.4 MHz bandwidth. The funneling effect must be considered in the analysis for the NG Cable Access Network.

If the return path signal level at the node from the Cable Modem (CM) is +15 dBmV, it is clear that the Signal-to-Noise Ratio (SNR) in a 6.4 MHz bandwidth is 50 dB; very adequate for 256QAM or even higher complexity modulation. But if the Return path level at the node port is 0 dBmV, the SNR is 35 dB; this makes 256QAM theoretically possible, but usually at least 6 dB of operating margin is desired. If only -10 dBmV is available at the node return input, the SNR is 25 dB; and so even the use of 16QAM is uncertain.This illustrates (figure 22) the very high dynamic range of "Pure RF" (about 15 dB higher than when an electrical-to-optical conversion is involved).

Modulation Type	Theoretical C/N	Desired C/N
QPSK	14 dB	20 dB
16-QAM	21 dB	27 dB
64-QAM	27 dB	33 dB
256-QAM	34 dB	40 dB

FIGURE 22: MODULATION AND C/N PERFORMANCE TARGETS

The table below, figure 23, documents many important assumptions and assumed node configuration conditions. An important assumption is the CM maximum power output level of +65 dBmV into 75 ohms. What this means is that if many channels are bonded (to increase the amount of data transmitted), the level of each carrier must be decreased to conform to the CM maximum power output constraint. Two channels bonded must be 3 dB lower each; four channels must be 6 dB lower than the Pout(max). Since the channel power levels follow a 10*log(M) rule, where M is the number of channels bonded to form a wider bandwidth group. For 16 channels bonded, each carrier must be 12 dB lower than the Pout(max).

For 48 channels bonded, each must be 16.8 dB lower than the Pout(max). So for 48-bonded channels, the level per channel is at most 65 dBmV -17 dB = +48 dBmV. If there is more than 48 dB of loss in the return path to the node return input, the level is <0 dBmV and 64-QAM or lower modulation is required. The node and system configuration assumptions are as follows.

Typical Node Assumptions		
Homes Passed	500	
Home Passed Density	75	hp/mile
Node Mileage	6.67	miles
Amplifiers/mile	4.5	/mile
Taps/Mile	30	/mile
Amplfiers	30	
Taps	200	
Highest Tap Value	23	dB
Lowest Tap Value	8	dB
Largest Amplifier Span	2000	ft
Largest Feeder Span	1000	ft
Largest Drop Span	150	ft
Home Split Loss to Modem	4	dB
Maximum Modem Power	65	dBmV

FIGURE 23: TYPICAL NODE ASSUMPTIONS

Cable Loss Assessment

Two different lengths of 1/2" diameter hardline coax were tested for Insertion Loss and Return Loss (RL). The loss versus frequency in dB varied about as the square root of frequency. But as can be seen below, the loss at 2 GHz is about 5% higher than expected by the simple sq-rt(f) rule. The graph below illustrates a slightly more than twice the loss at 2 GHz compared to 500 MHz, see figure 24.

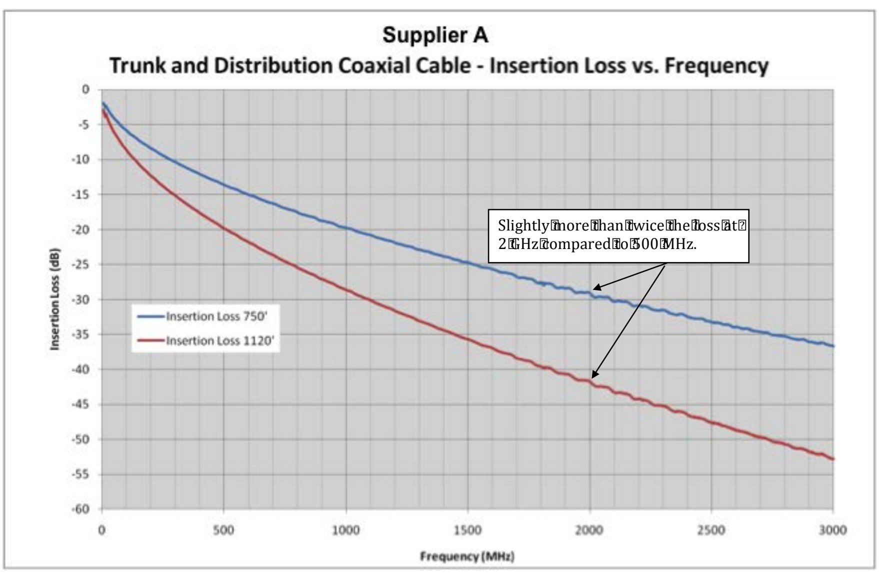


FIGURE 24: DISTRIBUTION COAXIAL CABLE – INSERTION LOSS VS. FREQUENCY

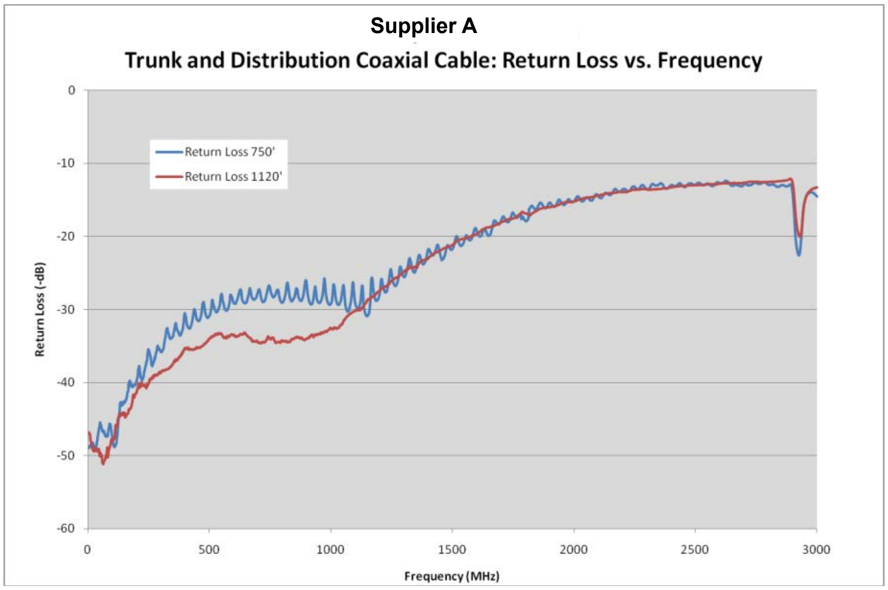


FIGURE 25: DISTRIBUTION COAXIAL CABLE – RETURN LOSS VS. FREQUENCY

In the plot (figure 25),the coax Return Loss (RL) did not vary as expected above 1200 MHz. This appears due to an internal lowpass matching structure in the hardline-to-75N connectors (apparently for optimizing the 1-1.2 GHz response). The connectorsare an important element to return loss with signals above 1 GHz.

Tap Component Analysis

Taps are the components with the most variability in passband characteristics, because there are so many different manufacturers, values, and number of outputs. Most were designed more than ten years ago, well before >1 GHz bandwidth systems were considered. One of the serious limitations of power passing taps is the AC power choke resonance. This typically is around 1100 MHz, although the "notch" frequency changes with temperature. Tap response resonances are typical from ~1050 to1400 MHz.A limitation of power passing taps is the AC power choke resonance. This is an important finding when leveraging the existing passives; therefore the use above 1050 MHz may not be predictable or even possible.

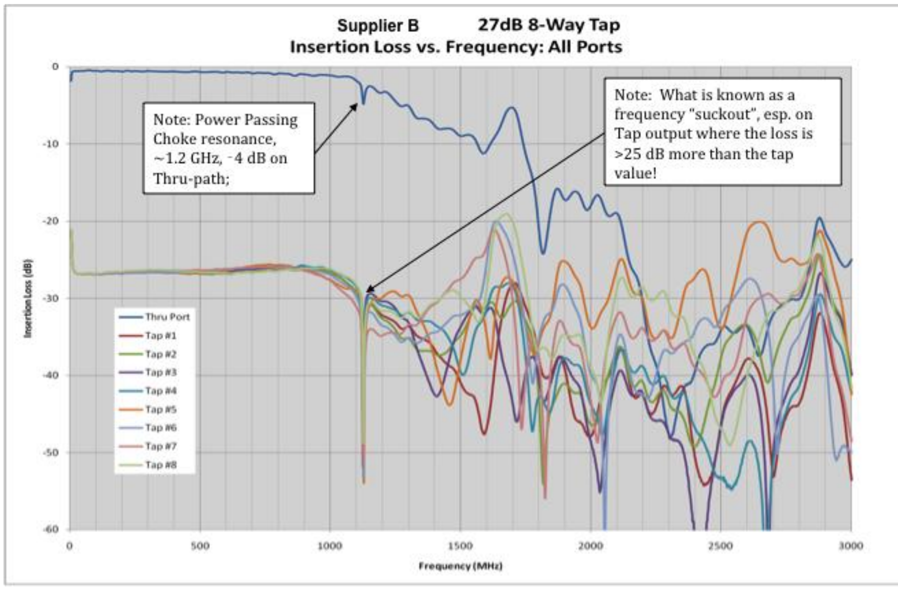


FIGURE 26: 27 DB X8 TAP - INSERTION LOSS VS. FREQUENCY: ALL PORTS

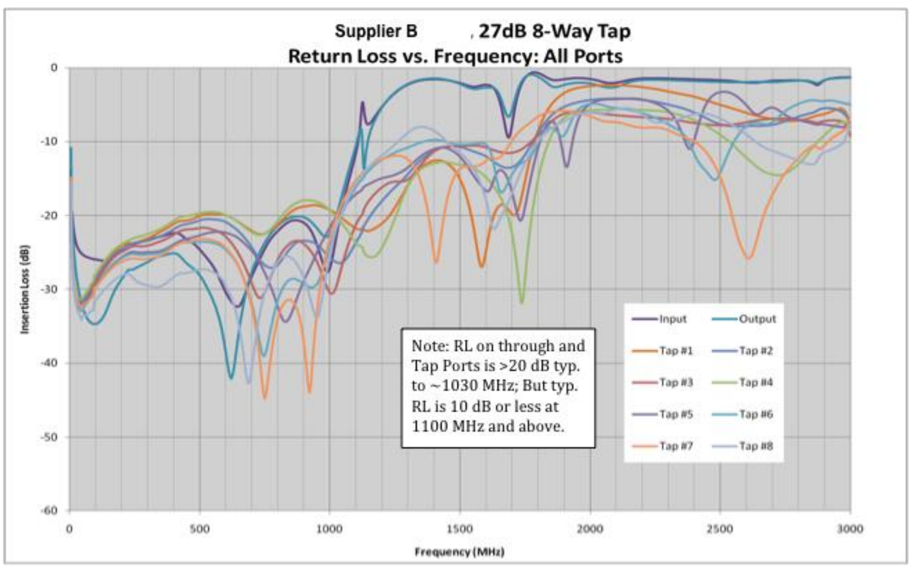


FIGURE 27: 27 DB X 8 TAP - RETURN LOSS VS. FREQUENCY: ALL PORTS

Even the newer, extended bandwidth taps, with passband specified 1.8 GHz or 3 GHz, the taps usually have power choke resonances (or other resonances, e.g., inadequate RF cover grounding) resonances in the 1050 MHz to1300 MHz range. Especially on the tap coupled port. However, most Taps work well to ~1050 MHz.

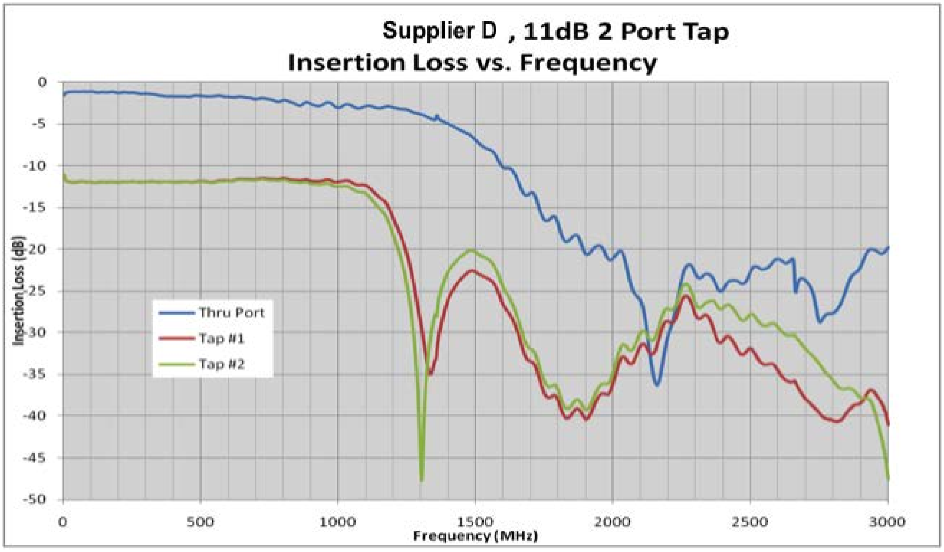


FIGURE 28: 11 DB X 2 TAP- INSERTION LOSS VS. FREQUENCY

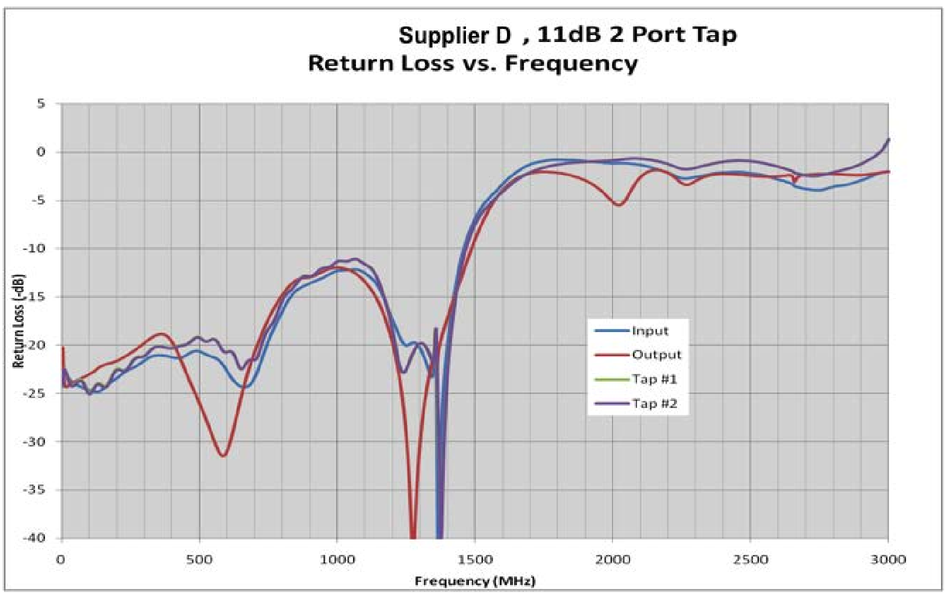


FIGURE 29: 11 DB X 2 TAP - RETURN LOSS VS. FREQUENCY

Nearly all taps exhibit poor RL characteristics on all ports above 1400 MHz. Some are marginal for RL (~12 dB), even at 1 GHz. Therefore tap cascades must be tested and over temperature to verify theactual pass band response due to close by tap reflections.

HFC Optical Return Path Transport Architecture

As we have analyzed several areas of the network to assess the impact and requirements to support additional upstream capacity we will now turn to the optical portion of the HFC network. The optical layer will be examined to support the additional upstream capacity. We will look at two classes of HFC optical transport, analog return path and the second type is digital return, which is commonly referred to as Broadband Digital Return (BDR).

Overview - Analog Optical Return Path

Analog return path transport is accomplished with a Distributed Feedback (DFB) laser located in the node housing and an analog receiver located in the headend or hub. Analog return path transport is considered as a viable option for sub-split, mid-split, and high-split returns.

Pros

The chief advantage of this method is its cost effectiveness and flexibility. If the analog return optics are in use in the field today, there is a good chance that they will perform adequately at 85 MHz and even 200 MHz loading, if required in the future. This would allow an operator to fully amortize the investment made in this technology over the decade.

Cons

There are drawbacks to using analog optics. Analog DFB's have demanding setup procedures. RF levels at the optical receiver are dependent on optical modulation index and the received optical level. This means that each link must be set up carefully to produce the desired RF output at the receiver when the expected RF level is present at the input of the transmitter. Any change in the optical link budget will have a dramatic impact on the output level at the receiver unless receivers with link gain control are used. Also, as with any analog technology, the performance of the link is distance dependent. The longer the link, the lower the input to the receiver, which delivers a lower C/N performance. The practical distance over which an operator can expect to deliver 256QAM payload on analog return optics is limited.

Assessment

The analog return transmitter will work well for the low and high frequency return. Analog return path options should be available for the higher frequency return options at 900-1050 MHz and 1200-1500 MHz. However the cost vs. performance at these frequencies when compared to digital alternatives may make them less attractive. There will be distance limitations and EDFAs will impact the overall system perform noise budgets. The distances of 25-30 km are reasonable and longer distance would be supported.

Overview - Digital Optical Return Path

Digital return path technology is commonly referred to as broadband digital return (BDR). The BDR approach is ―unaware" of the traffic that may be flowing over the band of interest. It simply samples the entire band and performs an analog to digital conversion continuously, even if no traffic is present. The sampled bits are delivered over a serial digital link to a receiver in the headend or hub, where a digital to analog conversion is performed and the sampled analog spectrum is recreated.

Pros

There are a number of advantages to the BDR approach. The output of the receiver is no longer dependent on optical input power, which allows the operator to make modifications to the optical multiplexing and de-multiplexing without fear of altering RF levels. The link performance is distance independent – same MER (magnitude error ration) for 0 km as for 100 km. The number of wavelengths used is not a factor since on/off keyed digital modulation only requires ~20dB of SNR; thus fiber cross-talk effects do not play a role in limiting performance in access-length links (<100 km)

The RF performance of a Digital Return link is determined by the quality of the digital sampling rather than the optical input to the receiver, so consistent link performance is obtained regardless of optical budget. The total optical budget capability is dramatically improved since the optical transport is digital. This type of transport is totally agnostic to the type of traffic that flows over it. Multiple traffic classes (status monitoring, set top return, DOCSIS, etc) can be carried simultaneously.

Cons

The chief drawback to BDR is the fact that nearly all equipment produced to date is designed to work up to 42 MHz. Analog receivers are not useable with Digital Return transmissions. Further, the analog-to-digital converters and Digital Return Receivers aren't easily converted to new passbands. It requires ―forklift upgrades" (remove and replace) of these optics when moving to 85 MHz and 200 MHz return frequency. There is currently no standardization on the Digital Return modulation and demodulation schemes, or even transport clock rates.

Assessment

It is more difficult and therefore more costly to manufacture BDR products. This may be a driver to use DFB products for the new returns. The selection of BDR products may be driven by distance and performance requirements. Another driver to move to BDR will be when there is near cost parity with DFB, today this is the case with the 5-42 MHz optical transport systems and this may be the case in the future with the new spectrum returns.

Review of HFC / Centralized Access Layer Architecture

The HFC has been around for nearly two decades and has evolved to include many technologies and architectures. Some of these include analog and digital optical transmission technologies and HFC architectures such as Node+N, Node+0, RFoG, QAM overlay, full spectrum, etc. But regardless of HFC technology and architecture selectedthe function of an HFC class of network remains constant, it has always been a ―media conversion technology" using analog, and today digital methods,to joindissimilar media types like fiber and coaxial. The HFC architecture being a ―conversion technology" allows the outside plant to remain relatively simple and very flexible to changes at the MAC and

PHY layers. The HFC architecture is also a —centralized access layer architecture" where all of the MAC/PHY processing takes place at the headend.

The HFC architecture has proven to be a valuable asset for the MSO, enabling the evolution to next generation access layer technologies while avoiding changes to the HFC layer of the network, with the exception of adding spectrum and capacity. Examples of this transition include analog video systems, digital video systems, EQAM, UEQ, SDV, CBR voice systems, pre-DOCSIS data systems, DOCSIS 1.0, 1.1, 2.0, 3.0 and so on. This entire multi decade transition did not fundamentally change the HFC architecture. HFC remained simple and carried the next generation data technology through it transparently; these are clear examples of its flexibility. Additionally, HFC may have carried all of these technologies simultaneously to support a seamless migration to the next generation; a clear example of versatility! The evolution of the cable network primarily is achieved by changing the bookends and not the plumbing in-between. It is hard to imagine the impact if the outside plant, such as nodes, needed to be changed to support each next generation MAC/PHY technology that came along over the last 20 years.

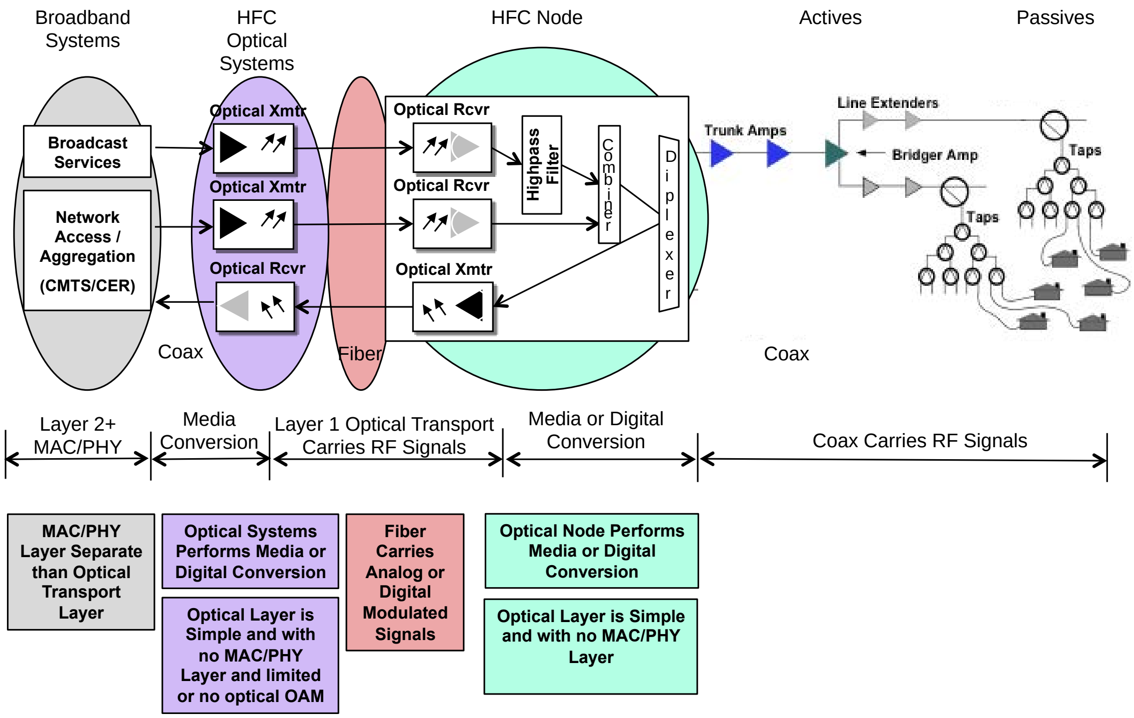


FIGURE 30:HFC A "MEDIA OR DIGITAL CONVERSION ARCHITECTURE"

Introduction to DFC (Digital Fiber Coax) / Distributed Access Layer Architecture

As we examine the future to support higher IP upstream data capacity and a transition of the downstream to more IP capacity for data and IPTV, the underlining architecture of HFC and centralized access layer may be placed into question. This is certainly nothing new. With each major shift in technology or major investment planned, the question of centralized vs. distributed appears. We will examine this class of architecture we are calling Digital Fiber Coax (DFC).

The Digital Fiber Coax Architecture label is for a network class which differs from HFC in that MAC/PHY or just PHY processing is distributed in the outside plant (node) or MDU and also uses ―purely digital” optical transport technologies such as Ethernet, PON, or others to/from the node. The industry may determine to call this class of architecture something else, but the functions, technology choices and architectures are different than HFC.

As described in the following sections there are many technologies and architectures that could all be categorized asthe class of architecture we defineas Digital Fiber Coax (DFC). This term may certainly change and is just used for discussion purposes within this document, however it is clear that the functions of DFC are not similar to HFC, the industry should consider naming this class of architecture.

The underpinning of this style of distributed access layer architecture is not new and goes back to the CMTS in the node discussions in the late 1990’s and early 2000’s for the cable industry. These concepts arose when the DOCSIS technology was emerging to replace proprietary data and CBR voice systems and was also considered during a period when HFC upgrades were still underway or planned. These distributed access layer architectures were discussed again in the late 2000’s when DOCSIS 3.0 was emerging, this time referred to as M-CMTS (modular-CMTS) or P-CMTS (partitioned CMTS), which could place some CMTS functions in the node, perhaps just the PHY layer. Again, in late 2010 with the announcement of DOCSIS-EoC (Ethernet over Coax) from Broadcom, placing the CMTS in the node or MDU revived the industry debate. Though DOCSIS-EoC is mainly focused on the China and worldwide MDU market. In addition to the CMTS in the node, the cable industry has considered QAM in the Node as well.

The diagram below illustrates an example of Digital Fiber Coax Class of Network Architecture using CMTS as the coax technology and PON or Active Ethernet as the optical transport to the node. The illustration is a node but could be an MDU. Additionally any optical technology or coaxial data technology could be employed, as discussed in detail below.

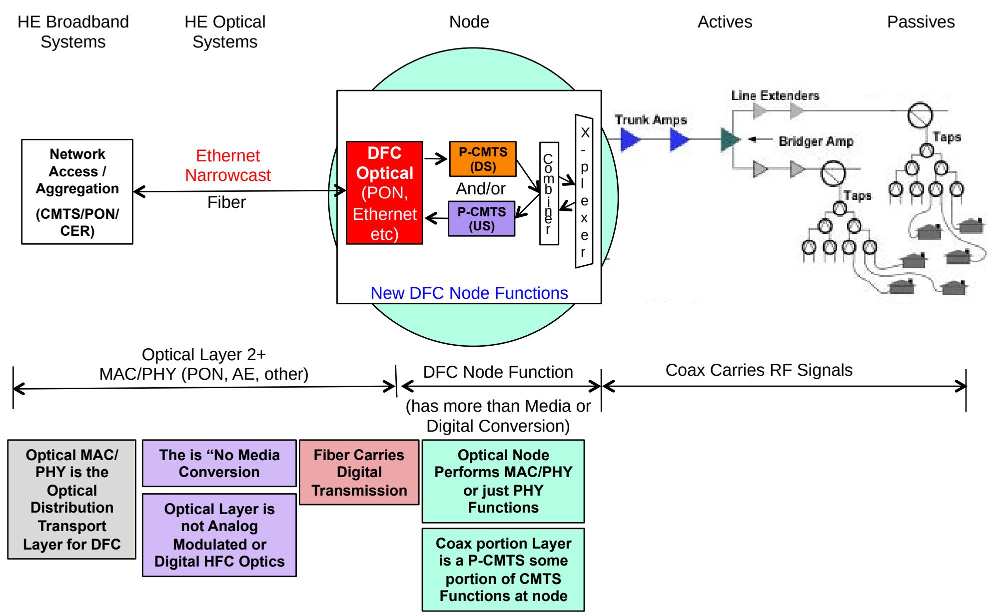


FIGURE 31: DIGITAL FIBER COAX A “PHY OR MAC/PHY PROCESSING ARCHITECTURE”

The actual development of a CMTS in the node, referred to as DOCSIS EoC by Broadcom, is a direct response to several other technologies morphing to be used as a coaxial access layer technology. While DOCSIS was designed from the ground up to be a cable access layer technology, the only architecture that a DOCSIS system was designed for was a centralized access layer approach, to be carried over an HFC network. The centralized access layer approach is a valuable approach to MSOs that have full two-way HFC and customers spread over vast outside plant areas. However, not all MSOs worldwide have full two-way, high capacity, and suitable HFC networks for data service; thus making a centralized CMTS architecture a challenge. A network architecture or suite of technologies used over coax referred to as ~~-~~Ethernet over Coax” (EoC) emerged to compete against DOCSIS; and the architecture was distributed, placing the CMTS-like functions in the node or MDU gateway.

The functions of EoC technologies are similar in many ways to DOCSIS, as most have a device, which functions like the CMTS, a central controller for scheduling network resources in a multiple or shared access network with end points like modems. The EoC architecture uses a fiber connection, likely Ethernet or PON to the node or MDU, where this transport is terminated and the data is carried to/from the CMTS-like function in the node/MDU and to/from customers over the coax.

Ethernet over Coax may be considered as an access layer technology where many consumers gain access to the service provider’s network. However, some of the technologies in the EoC space may have started as home networking technologies such as MoCA, BPL, HomePlug, HPNA, G.hn, HiNOC, WiFi over Coax, and more. The placement of any of these technologies in a node to interface with coax in our view is not HFC style architecture, but rather DFC style architecture, as the MAC and the PHY processing takes place at the node.

Overview - DFC is a New Architecture Class for Cable

There are two different Fiber to the Node (FTTN) architectures, which utilize coax as the last mile media. If we consider HFC an architecture class with several technologies and architectures that may be employed, the same could be applied for the DFC architecture class.

To simply summarize the delta between HFC and DFC Architecture Classes:

- HFC is a ―Media or DigitalConversion Architecture”
- DFC is a ―PHY or MAC/PHY Processing Architecture”

Technology Optionsfor DFC

The DFC class of architecture could use several optical transport technologies to/from the Headend link to the node; this is called ―Ethernet Narrowcast”. The optical technologies could employ PON, Active Ethernet, G.709, or others to carry data and management communications to/from the node.

The DFC class of architecture could use several coaxial-based MAC/PHY technologies such as DOCSIS, Edge QAM MPEG TS, MoCA, BPL, HomePlug, HPNA, G.hn, HiNOC, WiFi over Coax.

Architecture Options for DFC

The DFC architecture could consist of MAC/PHY or simply PHY functions in the Node. The architecture could support downstream and upstream functions or just a single direction.

Examination of DFC with EPON and P-CMTS

We have examined many layers of the network architecture and considered many approaches for upstream spectrum expansion and performance as well as optical transport in HFC style architectures. We wish to consider a distributed access layer architecture approach the digital fiber coax (DFC) style architecture.

The differences have been defined already between HFC and DFC. In addition, several technology and architecture choices that could be grouped under the DFC Class of Architecture were also covered. This section examines the use of DFC style architecture.

The DFC Architecture selected as an example in figure 32, illustrates 10G/10G EPON as the optical transport placing the optical MAC/PHY in the optical node and the second selection is DOCSIS as the RF technology. The architecture is an upstream only PHY in the node.

Consider then the use of HFC and DFC to support the legacy and new architecture simultaneously, this approach may be referred to as the DFC Split Access Model, as illustrated in figure 32. The HFC is usedto support legacy transport technology, services, and most importantly the centralized access architecture for the downstream such as the very high capacity CMTS/UEQ and the plus side is the massive and existing downstreamoptical transport is leveraged. There is no need to place downstream RF MAC or PHYs in the node in most configurations.

In the DFC split access model, the HFC upstream optical transport is leveraged as well, which may include the Sub-split 5-42 MHz band and even perhaps Mid-split. The HFC upstream optical transport will support a centralized access layer. The HFC with centralized access layer may be considered as the high availability architecture, because the OSP performs just media conversion and the centralized access layer systems, like a CMTS have highly redundant systems.

The DFC style architecture is used for two-way high capacity optical transmission to the node (like 10G Ethernet or 10G EPON) however initially this architecture will just consider using the upstream for the expanded coax upstream, see figure 32. The Partitioned CMTS (aka P-CMTS) using upstream only is examined in this paper, however additional downstream capacity could use the existing optical connection and the placement of a future P-CMTS for the downstream could be added later if needed, perhaps over 1 GHz spectrum range.

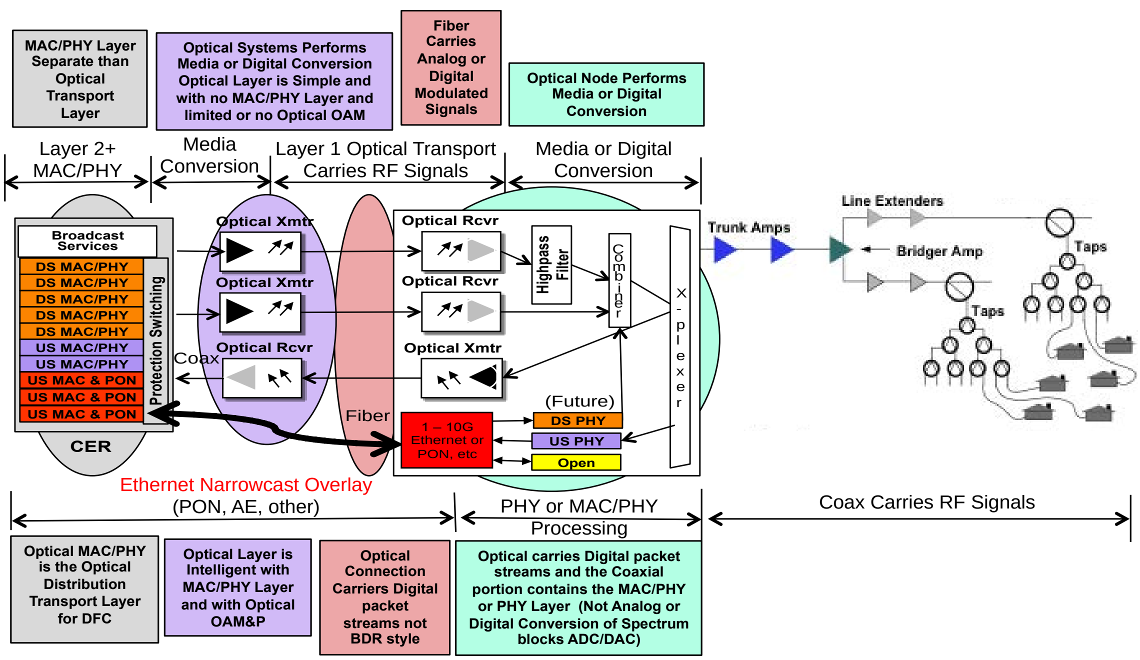


FIGURE 32: DIGITAL FIBER COAX (DFC) SPLIT ACCESS

Defining an architecture that placed the downstream PHY in the node was not financially prudent since the HFC forward already exists and has massive optical capacity. Centralized access layer architecturesover HFC have proven to be flexible, economical and will keep the outside plant (nodes) as simple as possible. The P-CMTS using just the PHY and just the upstream was selected to keep costs down as much as possible and may only be used ifconditions would require a distributed architecture. An example could be in situations where the use of an extremely long distancelink above 50 km or much higher is required. We believe the MSOs will want to keep the outside plant (OSP) as simple as possible for as long as possible; this has proven to be a very valuablecharacteristicfor 50 + years. Placing the entire CMTS in the node was not considered prudent because of the very high capital cost and it is least flexible for the future.

Description of P-CMTS (upstream only)

This paper has created a few new labels for cable networking architectures just for use within the paper. Another term that was created in a 2008 whitepaper but perhaps not well known is the Partitioned CMTS (aka a P-CMTS) [6]. The P-CMTS proposes to remote the DOCSIS PHY sub-system from the core CMTS chassis. The primary reason for the P-CMTS is that it can potentially permit the PHY sub-systems to be located a long distance from the MAC sub-systems, such as the node, and use the optical transport options defined above for Digital Fiber Coax to/from the node. The digital packet streams carrying the DOCSIS-encapsulated packets are transmitted across the optical link between the external PHY (node) and the core CMTS chassis.

Pros

The advantages of this approach are that the system would perform as if there were no optical link at all. The RF plant would essentially connect directly to the CMTS upstream port in the node as if it would have been in the headend. The advantages of this approach would be the same as the advantages of BDR with the exception of begin agnostic to the traffic. Its performance would be slightly better than BDR since the A to D and D to A conversions are eliminated.

The DFC style architecture would use optical technology that would have less configuration and two-way transport capacity with optical monitoring. The other advantage is distance and performance when compared with HFC style optical transmission. The DFC style architecture could be a QAM narrowcast overlay competitor, where typically optical links are long, costly, and challenging to optically configure.

Cons

The disadvantages of the DFC style architecture in this case using P-CMTS approach are that the solution would only work with DOCSIS returns (or the specific demodulations for which it was programmed) and would be ―unaware" of any other traffic that may exist on the network.

Placing MAC/PHY or just PHY functionsin the node may be difficult to change as new technology becomes available. This should be a very important consideration for MSOs as they reflect on the MAC/PHY technology changes in just the last 10-15 years, as described in the preceding HFC architecture section. The thought of touching every node to make a MAC or PHY change may be unthinkable to some operators.

Costs is an additional concern, when placing MAC/PHY or just PHYs in the node this means each node may need to be configured with enough capacity to meet the servicetier offeringand traffic capacity estimates up front. Another option is to make the node configurable to add capacity, this would mean visiting each node when additional capacity is needed.The MSOs have typically allocated capacity in the largestserving area possible, to gain economies of scale and additional capacity.

There may also be performance concerns with TCP latency with distributed architectures. The reliability and redundancy is also a consideration, there are more active components in the field.

Overall Assessment of DFCStyle Architectures

The HFC Architecture enables centralized access layer architectures; and DFC enables distributed access layer architecture. As discussed above, the industry since the 1990s has examined the placement of intelligence in the nodes; like CMTS and since then different part of the CMTS has been considered for the node, Edge QAM, PON, and many other technologies have been considered.

The main consideration for DFC style architecture may be the very long links of QAM overlay in the forward and in the future the new high capacity upstream. In those markets, there may be benefits from a DFC style architecture, but more study is needed and there are trade-offs.

MSOs using HFC and DOCSIS have benefited from leveraging DOCSIS capacity across many service groups expanding and contracting service group size at the DOCSIS layer where and when needed. Placing the CMTS functions of any kind or any MAC/PHY or PHY technology may have higher start-up costs and total cost of ownership could be a challenge. More study in needed to determine the viability of this type of architecture. Placing the intelligence in the node and distributing the architecture may limit future flexibility. Additional concerns of power and space will need to be explored if remote DOCSIS P-CMTS Upstream is explored.

NETWORK MIGRATION ANALYSIS AND STRATEGIES

This section provides analysis of someof the migration strategies. It is very important to note that the starting points of the MSO will greatly influence the selection of a particular network technology and architecture path. Additionally, the network utilization and capacity planning forecast in the local market will be a major driver for the migration strategy selected and timing by the cable operator. These are some of the factors that may influence the operator's selection path:

- Competitive and user consumption levelHigh-Speed data
- Video services offering
- Deployment level of STB which use Proprietary OOB or DSG
- Deployment level of DTA in market
- All Digital Offering
- Desire to offer analog service tier

- Current system spectrum capacity level (750 MHz or 1 GHz)

Downstream Migration Analysis

The findings of this report illustrate the MSOs existing downstream capacity is sufficient for this decade and beyond at the spectrum level 750 or equivalent assuming the upstream split options. There will be reclamation of the analog service tier and spectrum, reduction in the service group size, and reallocation of the distribution network from MPEG TS to IP.The downstream migration will be managed mainly from the headend systems and CPE migration to IP. Additionally, investment in full spectrum to each node will be need this decade and an additional forward transmitter to the node service group based on our service and capacity projections.

Upstream Migration Analysis

The cable industry has existing spectrum capacity and channel bonding capability in the upstream that will meet their service and capacity needs for many years to come, perhaps 2015-2017.

A migration to Mid-split first in the 2015 timeframe or when the capacity of Sub-split is exhausted by the technical and business analysis a good first step.

The next choice is High-split or a Top-split option. All of these solutions may start with 500 HHP node service group, but this will depend on the capacity requirement and split option selected.The service tier, customer traffic, and the node service group network elements will influence the service group size.

The Top-split analysis for a 500 HHP node has 64QAM as possible for Top-split (900-1050) and QPSK for Top-split (1250-1550) but neither achieve the capacity of High-split. Top-split (900-1050) with Mid-split in a 500 HHP node may reach the capacity of High-split.

The Top-split migration to a 125 HHP Service Group, reduces the funneling noise by 6 dB, which permits 256QAM for Top-split (900-1050) and 16QAM for Top-split (1250-1550) and both of these Top-split options coupled with Sub-split will yield the capacity of High-split.

Below are some high-level considerations for the migration and split selection.

- Consider Mid-split first (This buys about a decade and churns out old STBs to avoid the OOB)
- Consider an eventual High-splitupgrade path capable of 200-300 MHz
- If there are concerns with video service, STB OOB, and want to avoid touching the passives, consider Top-Split (900-1050)<u>with</u> Mid-split and a 500 HHP node as this has same capacity as High-split (200).
- Consider reserving the above 1 GHz for the next decade.
- A migration to Top-split (1250-1550) with Sub-split will require a 125 HHP service group to have the capacity of High-split.

COST ANALYSIS

The report has covered some of the key inputs and levers to begin to assess the costs of the upstream spectrum options. The underlining requirementsof the network architecture to meet the capacity targets have been documented. In this section some of those key technical assumptions will be covered to provide perspective as to the drivers and considerations in the cost analysis. The cost analysis makes some estimates to cost for products that have not yet been invented, so these are rough estimates used for discussions purposes only. The actual relationship between the upstream migration options may vary.

The HFC network allows operators to employ a number of methods to manage the abundant downstream spectrum. Some of these options include analog reclamation, switched digital video delivery of multicast content and service group size management, mainly achieved by node segmentation or node splitting to achieve the desired ratios. The relative costs and benefits of these downstream augmentations are well understood and used extensively today. As shown in the previous sections, the upstream spectrum has been sufficient to meet the demand, but before the end of the decade, additional spectrum may be required, in some MSO markets.

Downstream Cost Analysis

- Converged Edge Router DOCSIS/Edge QAM device will enable an effective migrations
- The Downstream may leverage a 6MHz by 6MHz channel investment which supports a smooth and economical transition while assuring revenue targets per 6 MHz channel for the MSO are met
- High-Split 200 may requirea forward laser upgrade to 1 GHz to gain the spectrum lost to upstream
- Top-split (900-1050) or Top-split 1250-1550 will leverage existing lasers for downstream
- Full Spectrum to a 500 HHP Node to 250 HHP segmentation is expected within the decade
- QAM overlay solutions may need to migrate to full spectrum
- Passive changes may be avoided for entire decade and beyond
- FTTLA may be avoided for entire decade and beyond

UpstreamCost Analysis

The following analysis is focused on a number of upstream options. The feasibility and relative cost of each path is compared. The analysis assumes a ―typical" HFC node has the following characteristics shown in Figure 33.

Beginning with a 500 home passed node, the first approach was to determine what might be possible without having to disturb the layout of the physical plant.

Figure 34 shows what the gain requirements (excluding port, EQ losses) would be for an upstream amplifier at the ranges of operating frequencies reviewed earlier in this paper. For this analysis, 0.75" PIII class cable was assumed for express amplifier spans and 0.625" PIII class cable was assumed for tapped feeder spans.

Typical Node Assumptions		
Homes Passed	500	
Home Passed Density	75	hp/mile
Node Mileage	6.67	miles
Amplifiers/mile	4.5	/mile
Taps/Mile	30	/mile
Amplfiers	30	
Taps	200	
Highest Tap Value	23	dB
Lowest Tap Value	8	dB
Largest Amplifier Span	2000	ft
Largest Feeder Span	1000	ft
Largest Drop Span	150.0	ft
Home Split Loss to Modem	4	dB
Maximum Modem Power	65	dBmV

FIGURE 33: GENERAL NODE ASSUMPTIONS

		Sub-Split	Mid-Split	High-Split 200	Top-Split (900-1050)	Top Split (1250-1550)
Upper Frequency	MHz	42	85	200	1050	1550
Typical Maximum Cable Loss (Amp to Amp 70 deg F)	dB	7.1	10.1	15.5	35.5	43.1
Aditional Gain Required for Thermal Control (0 to 140 deg F)	+/-dB	0.5	0.7	1.1	2.5	3.0
Total Reverse Amplifier Gain Required	dB	**7.6**	**10.8**	**16.6**	**38.0**	**46.1**

FIGURE 34: RETURN AMPLIFIER GAIN CALCULATION

It is worth noting that the Sub-split, Mid-split and High-split gain requirements can be satisfied with commonly available components that are currently used in amplifier designs today and would likely involve no cost premium. However, the Top-Split options would likely require multistage high gain amplifiers to overcome predicted losses, which would be more costly. It is also important to note that thermal control would likely become a major issue in the Top-split designs. Figure 34 shows seasonal temperature swings of 5 to 6 dB loss change per amplifier span would be likely in the top split solutions. Reverse RF AGC systems do not exist today, and could be complex and problematic to design. Thermal equalization would be sufficient to control the expected level changes at 200 MHz and below, but it is not certain that thermal equalization alone will provide the required control above 750MHz. This needs more study.

Figure 35 is a summary of path loss comparisons from home to the input of the first amplifier, which will ultimately determine the system operation point. It is interesting to note that as soon as the upper frequency is moved beyond the Sub-split limit, the maximum loss path tends toward the last tap in cascade as opposed to the first tap. There is a moderate increase in expected loss from 42 to 200 MHz, and a very large loss profile at 1000 MHz and above. The expected systemperformance can be calculated for each scenario.

		Sub-Split	Mid-Split	High-Split 200	Top-Split (900-1050) with Sub-split	Top Split (1250-1550) with Sub-split
Upper Frequency	MHz	42	85	200	1050	1550
Worst Case Path Loss	dB	**29.4**	**30.8**	**35.6**	**53.1**	**59.8**
Hardline Cable Type		0.625 PIII	0.625 PIII	0.625 PIII	0.625 PIII	0.625 PIII
Cable Loss/ft	dB/ft	0.0042	0.0060	0.0092	0.0211	0.0256
Drop Cable Type		Series 6	Series 6	Series 6	Series 6	Series 6
Cable Loss/ft	dB/ft	0.0134	0.0190	0.0292	0.0669	0.0812
Path Loss from First Tap	dB	*29.4*	*30.5*	*32.3*	*39.1*	*41.7*
Hardline Cable to First Tap	ft	100	100	100	100	100
Cable Loss	dB	0.42	0.60	0.92	2.11	2.56
Tap Port Loss	dB	23	23	23	23	23
Total Drop Loss	dB	2.0	2.9	4.4	10.0	12.2
In Home Passive Loss to Modem	dB	4	4	4	4	4
Path Loss from Last Tap	dB	*28.2*	*30.8*	*35.6*	*53.1*	*59.8*
Hardline Cable to Last Tap	ft	1000	1000	1000	1000	1000
Cable Loss	dB	4.21	5.99	9.19	21.06	25.59
Tap Insertion Loss	dB	10	10	10	10	10
Tap Port Loss	dB	8	8	8	8	8
Total Drop Loss	dB	2.0	2.9	4.4	10.0	12.2
In Home Passive Loss to Modem	dB	4	4	4	4	4

Figure 35: Path Loss from Home to First Amplifier

Figure 36 shows the compared performance calculations for the 500 home passed node outlined in Figure 33. The desired performance target is 256QAM for each scenario; if it can be achieved, the throughput per subscriber will be maximized. For each approach, it is assumed that a CPE device is available with upstream bonding capability that can use the entire spectrum available at a reasonable cost. The number of bonded carriers transmitting must not exceed the maximum allowable modem transmit level, so the maximum power per carrier is calculated not to exceed 65 dBmV total transmitted power. The maximum power, along with the worst-case path loss, yields the input level to the reverse amplifiers in the HFC Network. If the return level was greater than 15 dBmV, it was assumed that it would be attenuated to 15 dBmV.

Armed with the input level and station noise figure, the single station amplifier C/N is calculated and then funneled through the total number of distribution amplifiers serving the node to yield the C/N performance expected at the input of the node. Worst case performance and lowest cost for return optical links was assumed to be obtained from analog DFB lasers up to 200 MHz which suggests staying with Analog Return; but cost parity between analog and broadband digital return (BDR) systems at 1000 MHz and above is now possible.

The results show that the solutions up to 200 MHz have sufficient performance to support 256QAM modulation at a 500 HHP node.The top split options suffer from cable loss, not to exceed +65 dBmV, and noise funneling.The Top-split (900 -1050) may operate at 64QAM and Top-split (1250-1550) may operate at QPSKand stay within margin budget using a 500 HHP node. Cost projectionsfor the various solutions show that High-split 200 MHz return delivers the highest throughput per subscriber at the lowest relative cost.

		Sub-Split	Mid-Split	High-Split 200	Top-Split (900-1050) with Sub-split	Top Split (1250-1550) with Sub-split
Upper Frequency	MHz	42	85	200	1050	1550
Homes Passed		500	500	500	500	500
HSD Take Rate		50%	50%	50%	50%	50%
HSD Customers		250	250	250	250	250
Desired Carrier BW	MHz	6.4	6.4	6.4	6.4	6.4
Modulation Type		**64-QAM**	**256-QAM**	**256-QAM**	**64-QAM**	**QPSK**
Bits/Symbol		6	8	8	6	2
Desired C/N	dB	33	40	40	33	20
Number Carriers in Bonding Group		3.5	11	29	23	47
Max Power per Carrier Allowed in Home	dBmV	60	55	50	51	48
Worst Case Path Loss	dB	**29.4**	**30.8**	**35.6**	**53.1**	**59.8**
Maximum Return Amplifier Input	dBmV	30	24	15	-2	-11
Actual Return Amplifier Input	dBmV	15	15	15	-2	-11
Assumed Noise Figure of Amplifier	dB	7	7	7	7	7
Return Amplifier C/N (Single Station)	dB	65	65	65	48	39
Number of Amplifiers in Service Group		30	30	30	30	30
Return Amplifier C/N (Funneled)	dB	50.4	50.4	50.4	33.7	23.9
Optical Return Path Technology		DFB	DFB	DFB	BDR	BDR
Assumed Optical C/N	dB	48	45	41	50	50
System C/N	dB	**46.0**	**43.9**	**40.5**	**33.6**	**23.9**
Expected Maximum Datarate after Overhead	Mbps	91.8	370.2	975.9	603.2	455.3
Extra Datarate from Sub/Mid Bands					91.8	91.8
Total Datarate from all Bands		91.8	370.2	975.9	694.9	547.1
Throughput/Customer	Mbps	0.37	1.48	3.90	2.78	2.19
Cost Scale			100%	100%	139%	218%
Solution Figure of Merit (Throughput/Cost Scale)			1.48	3.90	2.00	1.00

FIGURE 36: PERFORMANCE OF 500 HP NODE

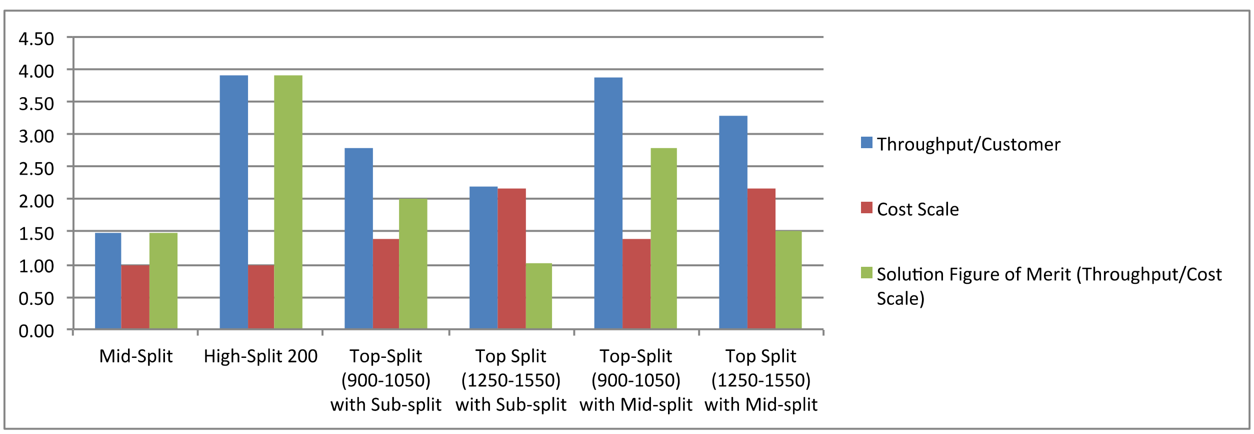


FIGURE 37: RELATIVE COST AND THROUGHPUT COMPARISON 500 HP NODE SOLUTIONS

Further analysis of the Top-split options concludes that reducing the node size, and thereby the funneled noise in the serving group could yield higher modulation capability. Figure 38 shows the comparison again with the Top-split scenarios including a four way node split. The noise funneling is reduced to a level where higher order modulations are possible. The costs of the additional node splits seem to scale appropriately with the additional throughput per subscriber.

		Sub-Split	Mid-Split	High-Split 200	Top-Split (900-1050) with Sub-split	Top Split (1250-1550) with Sub-split
Upper Frequency	MHz	42	85	200	1050	1550
Homes Passed		500	500	500	125	125
HSD Take Rate		50%	50%	50%	50%	50%
HSD Customers		250	250	250	62.5	62.5
Desired Carrier BW	MHz	6.4	6.4	6.4	6.4	6.4
Modulation Type		**64-QAM**	**256-QAM**	**256-QAM**	**256-QAM**	**16-QAM**
Bits/Symbol		6	8	8	8	4
Desired C/N	dB	33	40	40	40	27
Number Carriers in Bonding Group		3.5	11	29	23	47
Max Power per Carrier Allowed in Home	dBmV	60	55	50	51	48
Worst Case Path Loss	dB	**29.4**	**30.8**	**35.6**	**53.1**	**59.8**
Maximum Return Amplifier Input	dBmV	30	24	15	-2	-11
Actual Return Amplifier Input	dBmV	15	15	15	-2	-11
Assumed Noise Figure of Amplifier	dB	7	7	7	7	7
Return Amplifier C/N (Single Station)	dB	65	65	65	48	39
Number of Amplifiers in Service Group		30	30	30	7	7
Return Amplifier C/N (Funneled)	dB	50.4	50.4	50.4	40.0	30.2
Optical Return Path Technology		DFB	DFB	DFB	BDR	BDR
Assumed Optical C/N	dB	48	45	41	50	50
System C/N	dB	**46.0**	**43.9**	**40.5**	**39.6**	**30.2**
Expected Maximum Datarate after Overhead	Mbps	91.8	370.2	975.9	774.0	863.8
Extra Datarate from Sub/Mid Bands					91.8	91.8
Total Datarate from all Bands		91.8	370.2	975.9	865.8	955.6
Throughput/Customer	Mbps	0.37	1.48	3.90	12.38	13.82
Cost Scale			100%	128%	240%	302%
Solution Figure of Merit (Throughput/Cost Scale)			1.48	3.06	5.17	4.58

FIGURE 38: COMPARISONS WITH TOP SPLIT ONLY AT 125 HP NODE

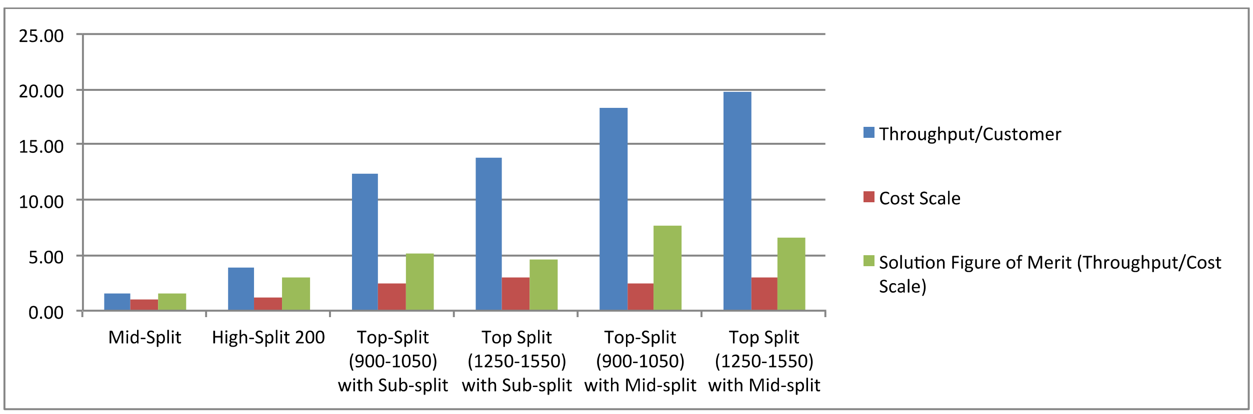


FIGURE 39: COST AND THROUGHPUT COMPARISON WITH TOP SPLIT ONLY AT 125 HP NODE

Summary of Cost Analysis

As stated in the opening of this section, the cost analysis providesestimates for discussion purposes only. The actual cost relative to one another may vary widely when products have been developed and released to market.

Perhaps a reasonable way to consider the data found in figures 36 and 37is that these may represent ―initial costs" to provide capacity above Mid-split. The cost analysis material in figures 38 and 39 may represent the ―end state cost analysis" to achieve the capacity requirements for this decade and beyond (approximately up to 2021 to 2023 timeframe). Moreover, these last analysesalso provide apple-to-apples cost estimates of High-split and the two Top-split options to reach similar network capacity targets.

The cost analyses capture the equipment and labor cost estimates for the optical transport and HFC systems. The cost analysis predicts that Mid-split and High-split will share similar costs on the return path systems, however the loss of downstream spectrum will have to be replaced if High-split is selected, captured in figures 38 and 39. The costs to solve the STBOut of Band (OOB), impacts to analog service tier, and loss of video spectrum for STBs and TVs was not calculated in the analysis of High-splits.

The two Top-split options do not account for the cost ofMid-split this isassumed to be the first upstream augmentation selection by the MSOs. Additionally, the High-split option uses the spectrum of Mid-split thus we considered this a sunk cost. The use of either Top-split (900-1050) option with Mid-split should achieve similar or greater upstream capacity compared to High-split. Either Top-split optionwith Sub-split will need to move to a 125 HHP node service group to achieve the capacity of High-split. The Top-split options do not account for the additional DOCSIS capacity because more channels are needed to achieve the capacity level of High-split (200).

The cost analysis intentionally excluded the items related to High-split and Top-split mentioned above in order to illustrate the relative cost for the spectrum split. We have additionalanalyses, which includesome of the items which were excluded from High-split and the Top-split options.

These are the findings of the cost analysis comparing four spectrum split options. The costs of the spectrum option to yield similar data capacity consideringHigh-split (200) and both Top-split options are illustrated in figure 40.

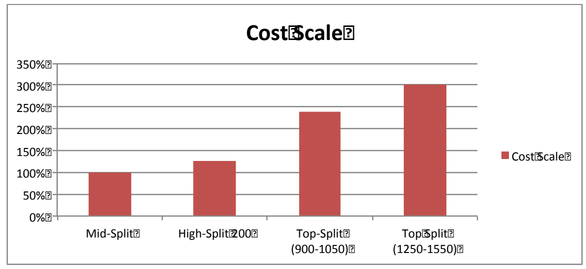


FIGURE 40: COST ANALYSIS

The Top-split (900-1050) with Mid-split given a 500 HHP may reach the capacity of High-split. The analysis estimates that Top-split (900-1050) is about 1.39 times the costof High-split (200), in this HFC topology. Again the Mid-split cost is not considered.

The Top-split options that avoid using Mid-split spectrum but does use the existing Sub-split spectrum will need to move to a 125 HHP node service group to achieve similar capacity to High-split. The analysis estimates that Top-split (900-1050) with Sub-split is about 2.4 times the costof High-split (200). The Top-split (1250-1550) with Sub-split is about 3 times the costof High-split (200).

EXECUTIVE SUMMARY AND CONCLUSIONS

The cable industry will transition and evolve their existing networks from largely broadcast video to unicast, likely using IP-based delivery technology. The High-Speed Internet Service and networks will continue to expand, meeting the needs for higher service tiers and more capacity downstream and upstream.

The paper provides assumptions and predictions for the evolution of advanced video and high-speed Internet services. The report assesses the timing and drivers of network change, possible technologies, architectures, migration options, and a cost vs. performance analysis.

The comparisons found in the report are at nearly every layer of the MSO's Next Generation Cable Access Network. These areas include spectrum splits, data network technology, and network architecture options. The examination will include a look at the traditional centralized access layer data architecture used over cable HFC networks.

We also define a possible alternative architecture to HFC, which we refer to as Digital Fiber Coax (DFC). This is a distributed access layer architecture using EPON or Ethernet style optical transport to the node where a coax based MAC/PHY or PHY resides supporting downstream and upstream; or just one direction (upstream).

Additionally, as the industry considers the future evolution of the network the report examines the importance of backward compatibility and the drivers behind this methodology.

Video and High-Speed Data Service Estimates

- Unicast services like On-Demand Video and High-Speed Internet will dominate the MSO service offering and spectrum allocation for the coming decade.
- Today, less than 4% of the MSO's downstream spectrum is allocated to High-Speed Data; however, it is forecasted that this might be 50% within the next 10 years.
- An industry accepted modeling tool based on a thirty-year history of Data Service offerings and capabilities predicts a ~ 2.5 Gbps Downstream and 500 Mbps Upstream Internet service offering in 10 years.
- The recent announcements from the United Kingdom's cable provider *Virgin Media* of a 1.5 Gbps Down / 150 Mbps Up trial and *Verizon's* FiOS reports of upgrades to 10 Gbps Down / 2.5 Gbps Up, is disrupting this thirty-year industry benchmark study of data service growth.

Cost Analysis & Performance Summaries

The upstream spectrum split options reviewedhave vastly different HFC topology and data network layer requirements, which have significant impact to the cost estimates. The cost analyses capture the equipment and labor cost estimates for the optical transport and HFC systems.

The cost analysis predicts that Mid-split and High-split will share similar costs for return path electronics, however the loss of downstream spectrum will have to be replaced if High-split is selected. The High-split option will have some impacts to network functions and services. The costs to solve the STB Out of Band (OOB), impacts to analog service tier, and loss of video spectrum for STBs and TVs was not calculated in the analysis of High-split.

The two Top-split options do not account for the additional DOCSIS channels needed to achieve the capacity level similar to High-split (200). The Top-split (900-1050) with Mid-split (given a 500 HHP service area) may reach the capacity of High-split. The analysis estimates that Top-split (900-1050) is about 1.39 times the costof High-split (200), in this HFC topology. Again, the Mid-split cost is not considered.

The Top-split options that avoid using Mid-split spectrum, but does use the existing Sub-split spectrum, will need to move to a 125 HHP node service group to achieve similar capacity to High-split. The analysis estimates that Top-split (900-1050) with Sub-split is about 2.4 times the costof High-split (200). The Top-split (1250-1550) with Sub-split is about 3 times the cost of High-split (200).

The cost analysis intentionally excluded the items related to High-split and Top-split mentioned above in order to illustrate the relative cost for the spectrum split. We have additional analyses, which include some of the items that were excluded from High-split and the Top-split options, that were not included in this report.

Top-split costs are driven by the network characteristics including: cable loss that progressively increases as frequency increases, modem maximum power output composite not to exceed +65 dBmV and the funneling effect of a large number of return path amplifiers. These critical factors andkey findings of the report illustrate the impacts to the HFC network topology, such as the size of the node service group. Additionally, the Top-split options must use lower order modulations, resulting in more spectrum and more CMTS ports needed to sustain equivalent capacity of High-split.

Key Network Performance Factors

- Network data capacity parity between High-split and the Top-split Options have vastly different network topologies and costs.
- The major reasons why lower frequency return (Sub-, Mid-, and High-split) and higher frequency return (Top-split options) have such performance differences which impact network architecture, capacity, and cost are as follows:
- First Major Reason: Cable loss progressivelyincreases as frequency increases, thus a major factor when considering higher frequency return.
- Second Major Reason: Modem maximum power output composite not to exceed +65 dBmV (to minimize power and cost, andmaintain acceptable distortion)

- Third Major Reason: Funneling effects of a large number of return path amplifiers. This is not a factor at low frequency because the cableloss is low enough that a cable modem can provide adequate power level to maintain high C/N.
- Existing 5-42 / 750 MHz system with a 500 HHP node may remain unchanged until 2015 and then a series of network migration steps that are defined in this paper may occur
- Existing 1 GHz Passives do not have to be touched until perhaps the year 2023
- Existing Passives may support up to 1050 MHz for additional upstream or downstream capacity
- A limitation of power passing taps is the AC power choke resonance. This is an important finding when leveraging the existing passives; therefore the use above 1050 MHz may not be predictable or even possible
- Passives represent approximately 180-220 devices per 500 HHP node group
- Small Nodes and FTTLA are not required until perhaps the second half of the 2020's decade.
- Downstream spectrum recovery methods will support the transition from broadcast to unicast service delivery and will help solve the downstream capacity challenge
- Downstream (assuming an equivalent 750 MHz system) will need to support full spectrum per node service group within the decade
- Mid-split is an excellent first step: low cost, small spectrum, high data capacity which lasts about a decade
- Mid-split in place of a node split may enable a 500 HHP node to last about a decade
- HFC Conclusions 1: architecture remains viable well through this decade and beyond
- HFC Conclusions 2: existing technology, costs, flexibility and versatility to support transport of virtually any new MAC/PHY technology remains a core benefit and value
- HFC Conclusions 3: Optical transport supported with DFB analog lasers and BDR last throughout the decade
- HFC Conclusions 4: HFC allows the outside plant to remain simple, just performing media and/or digital conversion (for BDR), thus no MAC or PHY layer processing is required in the node
- HFC Conclusions 5: enables a centralized access layer for economies of scale and just in time investment in capacity
- Digital Fiber Coax (DFC)
  DFC Conclusion 1: A distributed MAC or PHY architecture that would compete or complement HFC is not a viable replacement in nearly all cases, except extremely long distance to the node,as discussed in the report, however more study is required.

- DFC Conclusion 2: A distributed architecture and its risks include: stranding capital, low flexibility and limited versatility
- DOCSIS Conclusion 1: enables the full spectrum migration to IP in both the downstream and upstream
- DOCSIS Conclusion 2: enables a smooth channel-by-channel migration, this is key for the MSO to maximize revenue per MHz and allowing just in time investment while converting to IP
- DOCSIS Conclusion 3: DOCSIS will support new upstream spectrum, increase modulation schemes, may add MACand PHY layer improvements (perhaps OFDM)
- Figure 41:DOCSIS QAM Estimates for:HFC Topology, Spectrum Split and PHY Capacity Comparison

DOCSIS QAM	Sub-split	Mid-split	High-split 200	Top-split (900-1050) with Sub-split	Top split (1250-1550) with Sub-split	Top-split (900-1050) with Mid-split	Top split (1250-1550) with Mid-split	Top-split (900-1050) with Sub-split	Top split (1250-1550) with Sub-split	Top-split (900-1050) with Mid-split	Top split (1250-1550) with Mid-split
Node Service Group	500	500	500	500	500	500	500	125	125	125	125
Capacity from Sub-split	92	92	92	92	92	92	92	92	92	92	92
Capacity from Mid-split		225	225			225	225			225	225
Capacity from High-split			599								
Capacity from Top-split				617	410	617	410	782	820	782	820
Total PHY Channel Bond Capacity (Usable) in Mbps	92	316	916	708	502	933	726	874	912	1,099	1,136

FIGURE 41: HFC TOPOLOGY, SPECTRUM SPLIT AND PHY CAPACITY COMPARISON

Network Evolution Prediction Summary

- **Year 2015:**Upstream 500 HHP Node Split/Segment —ởr add new spectrumupstream& keep node size
- **Year 2017:** North America 5-42 is exhausted because of High-Speed Internet Service Tier, more upstream spectrum required.
- **Year 2017-18:** Downstream 500 HHP Node is at capacity (assuming 250 Subs with 0.9-1.4 Gbps of Traffic Plus 3.6 Gbps of MSO Video Traffic)(reduce SG or add spectrum)
- **Year 2018-19:**Euro 5-65 is exhausted because of High-Speed Internet Service Tier,more upstream spectrum required.
- **Year 2019:** Upstream 500 HHP Node with Mid-split DOCSIS is exhausted (reduce SG or add spectrum)
- **Year 2020:** Mid-split (5-85) with either DOCSIS QAM or OFDM is approaching capacity and in year 2021 would be out of capacity driven by Service Tier growth; thus more upstream Spectrum is Required (High-split or Top-split)
- **Year 2021:** HSD Max Service Tier prediction of 2.2 Gbps + 1.8 Gbps of MSO Videoapproaching capacity of 750 MHz system
- **Year 2021:** Downstream 250 HHP Node is at capacity (assuming 125 HSD Subs with 2.3 Gbps of Traffic

Plus 1.8 Gbps of MSO Video Traffic) change SG size or service

- **Year 2021-22:** Upstream 500 HHP Node nears capacity with High-split 200 or Top-split (900-1050)+Mid-split Using DOCSIS QAM or OFDM (use of OFDM bought a few months of capacity)
- **Year 2023:**Upstream HSD Max Service Tier Prediction of 995 Mbps Upstream Serviceconsumes High-split 200 or Top-split (+Mid-split) Using DOCSIS QAM or OFDM
- **Year 2023:** Downstream HSD Max Service Tier Prediction Consumes All Downstream Spectrum thus additional Spectrum Above 1 GHz is required or FTTx
- **Year 2023 - 2024: MSOs touch the passives to increase spectrum above 1 GHz to achieve higher Downstream and Upstream capacity based on HSD predictions**
- **If these do not materialize** (i.e. High-speed data service prediction over 4 Gbps Down and 1 Gbps Up in the year 2023) nodes splits/node segmentation will solve the traffic growth projections for many more years
- **Finally,**if neither the Service Growth Rates nor Traffic Utilization Growth Rates are maintained at a 50% CAGR, then the timing and drivers for investment will change and the HFC will last far longer.

Importance of Backward Compatibility

- DOCSIS 3.0 QAM based and any successorshould consider that every MHz should all share the same channel bonding group, this maximizes the use of existing spectrum and delays investment
- Sharing channel bonding groups with DOCSIS 3.0 and Any Successor creates "one" IP Network (cap and grow networks hang around awhile)
- Sharing the same bonding group assures previous and future investment may be applied in creating larger IP based bandwidth and not stranding previous capital investment
- Backward Compatibility has benefitted industries like the IEEE Ethernet, WiFi, and EPON saving the entire eco-system money
- Backward Compatibility simply allows the MSOs to delay and perhaps avoid major investment to the network such as adding more spectrum or running fiber deeper.
- All of our analysis in this report assumes backward compatibility with DOCSIS 3.0 QAM and any successor, like DOCSIS OFDM; thus creating a larger and larger IP bonding group with each year's investment. If this is not the case the investment in HFC upgrades will pull forward. It is uncertain of the exact level of financial impact but the total cost of ownership may be higher when deploying two separate IP based network technologies.

These are the major takeaways of this paper, however additional information is contained in the report providing perspective and details to the conclusions cited above.

The examination of the next generation cable access network spans several disciplines and this report is not a complete analysis of all of the possibilities.

REFERENCES

[1] Communications Technology, ―Virgin Media Trials 1.5 Gbps Broadband”, April 21,

[2] CED Magazine, ―Verizon open to 10G PON bids in 2011”, June 23, 2010

[3] 2011Tom Cloonan, ―On the Evolution of the HFC Network and the DOCSIS® CMTS - A Roadmap for the 2012-2016 Era,” *SCTE Cable Tech Expo*, 2008.

[4] ANSI/SCTE 55-2 2008, ―Digital Broadband Delivery System: Out of Band Transport Part 2: Mode B”, SCTE

[5] ANSI/SCTE 55-1 2009, Digital Broadband Delivery System: Out of Band Transport Part 1: Mode A.

[6] MoCA, "The Only Home Entertainment Networking Standard In Use By All Three Pay TV Segments—Cable, Satellite, IPTV, MoCA Presentation 2010, www.mocalliance.org.

LIST OF ABBREVIATIONS AND ACRONYMS

BPON	Broadband PON
CAGR	Compound Annual Growth Rate
CBR	Constant Bit Rate
DBS	Digital Broadcast System
DFC	Digital Fiber Coax
DOCSIS	Data Over Cable Service Interface Specifications
DSG	DOCSIS Set-top Gateway
DTA	Digital Terminal Adapter
EoC	Ethernet over Coax
EPON	Ethernet Passive Optical Network
FTTH	Fiber To The Home
FTTLA	Fiber to the Last Active
FTTP	Fiber to the Premise
FTTx	see (FTTH, FTTP, etc)
Gbps	Gigabits per Second
GPON	Gigabit PON

HFC	Hybrid Fiber Coaxial
HHP	Households Passed
HPNA	HomePNA Alliance
HSD	High Speed Data
IP	Internet Protocol
IPTV	Internet ProtocolTV
MAC	Media Access Layer
Mbps	Megabit per Second
MoCA	Multimedia over Coax Alliance
MSO	Multiple Systems Operator
OFDM	Orthogonal Frequency Division Multiplexing
OSP	Outside Plant
OTT	Over The Top
P2P	Peer-to-peer
PHY	Physical Layer
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RFoG	RF Over Glass
SDV	Switch Digital Video
UHF	Ultra High Frequency
US	Upstream
VoD	Video on Demand

DIGITAL DUE DILIGENCE FOR THE UPSTREAM TOOLBOX

Phil Miguelez and Dr. Robert Howald
Motorola Mobility

Abstract

As upstream services have become more sophisticated, so too has the technology that supports it. Simple, robust modems were developed to enable STB communications. Later, more sophisticated DOCSIS® modems complemented these for high-speed data services. Similarly, in the plant, return optics adequately supported the robust physical layer of STB communications using low-cost Fabry-Perot (FP) lasers. For DOCSIS®, laser technology advanced to meet the new challenges. This included higher power FP lasers, isolation techniques to mitigate noise, and the introduction of Distributed Feedback (DFB) lasers into the return transmitter portfolio.

DOCSIS®, of course, has continued to advance. Deployment of DOCSIS® 3.0 began in earnest in 2010 and will continue through 2011 and beyond. A key challenge is the demands that this places on the return channel – yet more sensitive modulation of wider bandwidth, the addition of new DOCSIS® channels, and potentially the addition of new spectrum. Together, it adds up to a much more aggressive use of available dynamic range. Nonetheless, the promise of DOCSIS® 3.0 must be met upstream for consumers whose speed and QoE expectations continue to rise. Next generation systems also may ultimately impose additional high performance requirements on the HFC upstream.

As the deployments of DOCSIS® 3.0 have increased, so too has new confusion over technology options to cost-effectively support this new DOCSIS® era. The new crossroads, replacing the debate over FP and DFB lasers, is around DFB technology options and Digital Return systems. Digital Return, developed now over ten years ago, offers an intriguing set of benefits that all MSOs should consider when planning future HFC migration. However, it also comes with some important constraints. This paper will take a comprehensive look at the use of DFBs and Digital Return technologies for supporting DOCSIS® and new potential upstream requirements in the context of anticipated traffic growth and expansion. We will delve into the comparable digital return parameters that will help operators compare these systems to the capabilities of their analog counterparts, as well as among the various incarnations of Digital Return systems themselves. Finally, we will weigh the pros and cons of each and provide recommendations.

UPSTREAM GROWTH MANAGEMENT

Quantifying Traffic Growth

DOCSIS service rates have continued to increase with time, as users find more bandwidth-consumptive ways to enhance their Internet experience. The applications driving the growth vary. Because of the futility in trying to pick the next "killer app," analysis of Internet traffic trends often rely on "Nielsen's Law" which quantifies consumer traffic demand as a function of time. Most importantly, the law recognizes that the unmistakable historical growth trend over the life of mass Internet access has a compounding nature to it. There was a day when the parallel to interest-bearing savings accounts resonated better than it does in 2011, but it is that same principle at work – a multiplication each year by a value we refer to as Compound Annual Growth Rate, or CAGR. While each year the actual growth rate varies, over a number of years the CAGR trend can be smoothed into an average which can be used to portend future requirements. Because of the sensitivity of long-term end results to the CAGR used, a strong sense of past behavior is valuable. For the same reason, a mindset that errs on the side of aggressive ensures being prepared with the proper network resources to handle growth and manage long-term capex spending levels. Or, conversely, to avoid not being able to meet the demands of new traffic growth.

Figure 1 shows a sample of growth projection using CAGR analysis. It shows three CAGR trends – 30%, 40%, 50% – and three thresholds of upstream throughput for 5-42 MHz systems – 60 Mbps, 100 Mbps, and 150 Mbps. The chart represents a serving group aggregate, so along the way service group splitting is factored in, effectively doubling the average bandwidth per subscriber. The latter two represent rough estimates of the maximum available in a 5-42 MHz system for A-TDMA (only) and A-TDMA + S-CDMA systems.

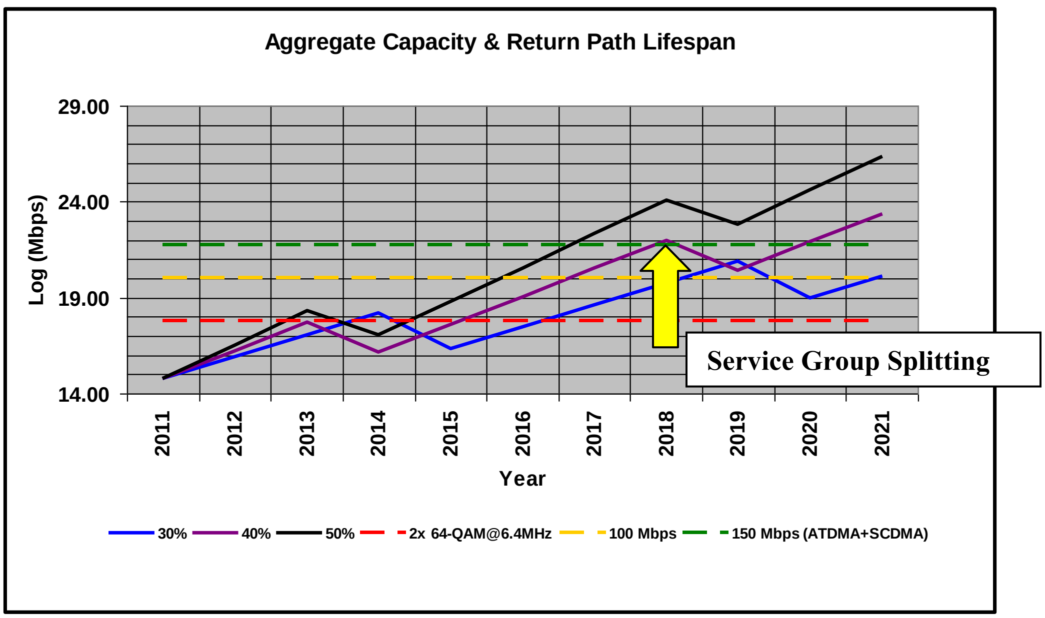


Figure 1 – Use of CAGR to Project Future Upstream Traffic Needs

A key point of Figure 1 is to recognize that it projects that next generation upstream technology will be required at some point in the near future lest available bandwidth be exhausted. The key question of "when" is a matter of CAGR, current level of service (starting point of plot), service group size, and efficiency of the use of the return bandwidth. For example, at 50% CAGR (black) using A-TDMA only (yellow), the upstream becomes fully consumed in 4.5 years with one service group split. At 30% CAGR – closer to actual for the past couple of years – it is exhausted in roughly seven years. For more details on traffic growth and considerations, please refer to [1], [2], and [3].

A few important comments should be made about use of Figure 1. First, it represents demand-based growth beginning with aggregate “demand” being what is offered today. Latent unused capacity of today’s services – how far below alarm thresholds as a measure, for example – adds margin to the calculation. Second, and perhaps more important, is that demand is but one driver of new service rates. Other market forces also intervene. There may be a competitive need to offer higher peak rates and deliver averages that provide a competitive level of experience outside of mathematical demand. Limitations of today’s available bandwidth come into play in both total capacity and peak rates long-term.

Finally, while upstream channel bonding (UCB) provides the key tool necessary to increase peak service rate offerings, it is important to recognize that bonding does nothing inherently to improve total upstream capacity. In implementation, UCB may involve lighting up new return spectrum that was not previously utilized. However, this is not new capacity enabled by UCB. It is merely latent capacity being exploited because there is available upstream spectrum – with or without UCB.

<u>Where Do We Find the Bits?</u>

Because of the potential to exhaust capacity and the upstream limitations on peak service rates, even with bonded channels, the next generation of upstream technology must consider the limitations of the current architecture and how to overcome them. Can they handle the inevitable traffic growth ahead? Can they support more sophisticated modulation profiles that create more bandwidth efficient use of upstream spectrum? Can they support the addition of new spectrum itself? These are the critical next generation system items that we look to address as we examine future upstream technologies and provide more bit-per-second

Theoretical capacity instructs us on where to go to find more bits. Capacity, C, of an additive Gaussian white noise-corrupted channel relies on two variables through a very simple equation. The variables are bandwidth available (BW) and Signal-to-Noise Ratio (SNR). The relationship is Shannon’s well-known capacity theorem,

$$C = BW \cdot \mathrm{Log}_2 (1+SNR) \qquad (1)$$

Note SNR is in linear (not dB) format in (1). While we certainly have more than this type of noise to deal with on the return path, it is the starting point of any attempt to add new bits per second. Other receiver technologies are designed to manage additional impairments (equalizers, ingress cancellation, forward error correction, spreading).

The addition of upstream bandwidth has been part of industry discussion for many years. Once DOCSIS services grew to be widespread, it became clear that the bottleneck of 5-42 MHz could have long term implications to the traditional forward/return frequency domain duplex (FDD) architecture. What was once a “long-term” proposition is now winding down in quantifiable ways, as shown in Figure 1. Tools exist to further optimize 5-42 MHz, in particular 64-QAM and S-CDMA. While these will indeed buy some time, ultimately the limit itself is that it is only 37 MHz of spectrum, as equation (1) points out. The resulting industry discussion has been around the various split architectures used or specified elsewhere, such as the European split using 65 MHz of return, and the DOCSIS 3.0 identification of the N-Split of 85 MHz. There also has been industry discussion about an extended return split, foreseeing a goal to support a potential 1 Gbps of capacity and/or service rate. Some proposals suggest that this may be most easily done, and with the least disruption, by exploiting unused spectrum on the coax above the forward band in a triplex architecture. The bulk of our discussion will be return path growth and technologies associated with low/high diplex arrangements.

A way to use any bandwidth most effectively is to provide the highest SNR possible, per (1). More SNR means more bandwidth efficient modulation profiles can be used. When this is the case, the digital symbols can be more tightly packed within the same average transmit power because the risk of incorrect decision making due to noise has decreased. The DOCSIS specification sets a minimum for the upstream at 25 dB. Table 1 compares what this SNR means to theoretical capacity versus what the DOCSIS physical layer transport rate delivers assuming 64-QAM. 64-QAM itself represents an important step in modulation evolution of the return path, increasing by 50% what the return could do relative to the 16-QAM limitations of first generation DOCSIS. Table 1 also compiles theoretical rates as we consider variations to the diplex split for future growth.

Maximum Capacity for Each Bandwidth

Return Bandwidth	DOCSIS	Maximum Capacity
5-42 MHz	150 Mbps	300 Mbps
5-65 MHz	270 Mbps	500 Mbps
5-85 MHz	360 Mbps	650 Mbps
5-200 MHz	900 Mbps	1.6 Gbps

Table 1 – Bandwidth, DOCSIS, and Capacity Limitations

It is clear that there is room between the DOCSIS transport rate, lowered further by forward error correction (FEC) overhead, and theory. While reaching theoretical capacity is impractical, advances in communications and information theory in the 10+ years since DOCSIS emerged have continued to close the gap. What has become clear using modern DOCSIS and HFC technology is that another step closer towards channel capacity can be enabled with additional evolution of the modulation profile to higher density constellations such as 128-QAM and 256-QAM [4].

The relationship between SNR, offering the potential for new bits per second, and modulation efficiency is observed most readily in Figure 2.

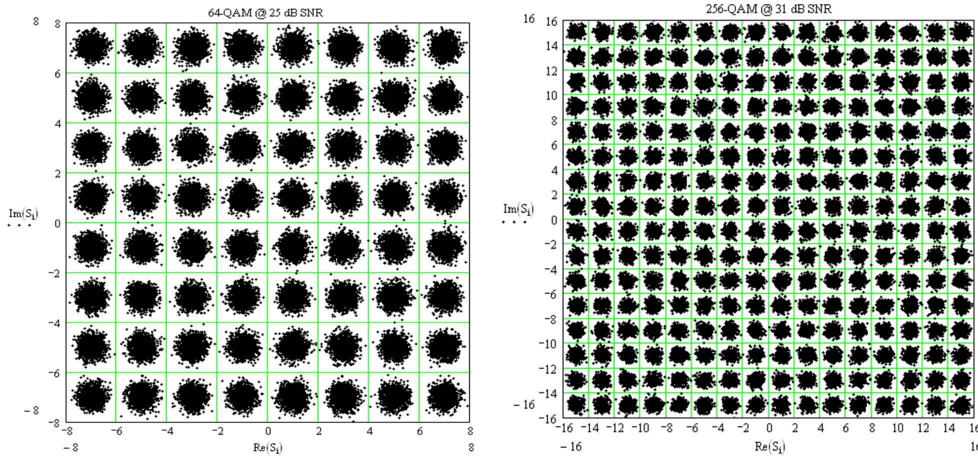


Figure 2 – 64-QAM @ 25 dB SNR and 256-QAM @ 31 dB SNR

The similarity of the relative relationship of the symbol "clouds" to the decision boundaries is readily apparent. Clearly, in absolute terms, the 256-QAM symbols could be misidentified with less noise power, and the amount less is precisely the 6 dB between these two constellations shown at SNRs of 25 dB and 31 dB. The actual receiver benefits from FEC and uses sequence detection algorithms, but Figure 2 properly identifies the obstacles to getting closer to capacity via increasing the modulation efficiency.

The evolution of the bandwidth allocated to return spectrum, and in the physical layer techniques such as advanced modulation are the tools by which new upstream capacity will be extracted, as capacity equation (1) easily instructs us. Given the inevitable growth of traffic, both of these variables become important elements for the evolution of any new upstream technologies. Each will be discussed, along with other practical and legacy hurdles, in the sections that follow.

UPSTREAM LINK COMPARISONS

Analog Return Solutions

Analog laser technology provides a low cost link connection that is expandable as bandwidth needs increase, and interoperable with a wide range of cable plant equipment. DFB lasers are widely available covering an array of optical output power levels and ITU grid CWDM plus DWDM wavelength options. DFB laser packaging and internal matching networks are typically capable of contiguous bandwidth exceeding 3 GHz. In practice, the usable bandwidth for upstream analog transmitters is usually determined by the RF drive amplification gain blocks and added filters used to reduce ingress noise or increase isolation between forward and return path signals. The RF driver hybrids and MMIC's in most legacy upstream transmitters have upper frequency capability of 150 to 200 MHz. The node diplex filter is used to establish the upper edge of the active return bandwidth split. In some cases changing the diplexer is all that is needed to expand the upstream capacity.

SNR for a DFB enabled upstream link is determined by a combination of the inherent laser noise or RIN (Relative Intensity Noise) caused by spontaneous non-coherent laser emissions, along with optical output launch level and the optical modulation index (OMI) of the RF drive signals. At the Head End / Hub side of the link the optical receiver noise performance also contributes to link SNR.

The reach limit for an analog link is determined by the optical link budget. This includes the laser output level, fiber loss, and passive losses due to muxing multiple upstream wavelengths onto a common fiber plus demuxing at the receiver. The upstream optical receiver sensitivity is a critical element in the link. Typical PIN diode detectors that are widely deployed in legacy Hub receiver platforms have a lower optical sensitivity limit of -16 to -18 dBm. Newer receiver designs have extended this threshold to > -26 dBm achieved by improving the detector EINC (equivalent input noise current) and limiting the total bandwidth of the receiver.

The different analog return solutions that are available represent a range of link reach capability. Table 2 illustrates a few common configuration examples and the link budget reach for each case. The reach capability of these options cover the characteristic maximum link distances needed for the majority of network serving area deployments by the major MSO's.

Band	Link Design	Reach
1310	P2P	50 km
1550	8λ CWDM	50 km
1550	8λ DWDM	80+ km*

* Reach extended with EDFA

Table 2 – Wavelength vs Reach

One of the main advantages of analog return systems is conversely identified as its biggest disadvantage. Analog transport allows the interoperable use of a broad assortment of equipment from different vendors since the RF upstream bandwidth is modulated onto the laser without manipulation. This allows cable operators to maintain existing, multiple source networks and acquired properties without the need to rework the Hub and node equipment. The downside of this flexibility is that laser transmitter gain and modulation level (OMI) varies from manufacturer to manufacturer. This requires the cable operator to verify and adjust, if necessary, the return path link gain during the initial node installation and alignment.

Digital Return Solutions

The advantages of digitizing data streams to provide improved signal to noise performance and extended reach are well known and certainly apply to Digital Return (DR) solutions for HFC upstream transport. Digital Return equipment has been successfully deployed for several years, enabling long link networks that are beyond the reach of analog lasers and without the need for expensive O-E-O regeneration.

The bandwidth capability of a digital return transmitter is determined by the system sampling rate and the data rate capability of the selected laser. Nyquist theorem dictates a minimum sampling rate of 2X the highest frequency or 2X the bandwidth for IF systems. This minimum number of samples is multiplied by the bit sample rate of the A/D chip set for each stream that will be transported.

Example: 5 to 42 MHz BW, 12 bit A/D, 2 channel transmitter (2 RF streams)

The minimum sample rate needed for this example is

$(2 * 42) * 12 * 2 = 2.016$ Gbps

In order to transport this data rate with some margin a digital laser with bandwidth capability of at least 2.5 Gbps is needed. Lasers with this and similar data rates are available in a wide range of form factors including pluggable SFP optics covering the full spectrum of ITU grid wavelengths.

SNR capability of a channel in all digital networks is best represented through use of Noise Power Ratio (NPR). This is basically a measurement of the noise floor rise of a vacant (RF) channel resulting from noise and distortion generated by the other channels of a fully loaded system. All digital return systems in use today are based on Time Division Multiplexing (TDM). As a result, the dynamic range performance of the selected A/D chip set as well as the bit sampling rate chosen for the design will determine the achievable NPR. Figure 3 shows the theoretical NPR peak and dynamic range differences corresponding to changes in the bit rate resolution of the A/D.

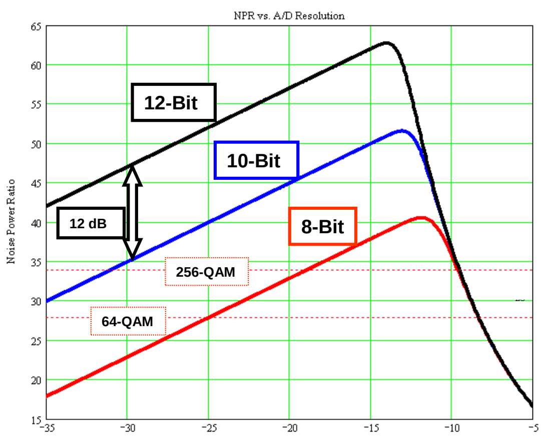


Figure 3 – NPR vs. A/D Resolution for Digital Returns

Digital return systems are not affected by optical noise from an NPR perspective for a properly functioning link. Digital transport is bit-error limited according to well-understood rules for the optical SNR of binary transport. Error correction makes the optical links themselves very robust. The result is digital return transmissions can be amplified over long link spans of 100 km or more with no impact on NPR performance.

A drawback for digital return systems is that the design implementation is different for every manufacturer. The consequence is that each vendors design becomes de-facto proprietary and not interoperable with other existing digital return equipment. This same disadvantage is conversely considered one of the main advantages of digital return by most cable operators. Since the DR transmitter and receiver at each end of the link are matched by design there is no uncertainty about the link gain parameters so digital return is considered to be a Plug-N-Play technology.

UPSTREAM CAPACITY PLANNING FOR THE FUTURE

Upstream data rate usage today is very asymmetrical when compared to downstream data rates. This is primarily due to the popularity of VOD content and services such as Netflix movie downloads. This trend is expected to continue for the foreseeable future. However, upstream rates are still increasing at a steady rate and with the deployment of DOCSIS® 3.0 the usable portion of the current 5 to 42 MHz return bandwidth will eventually be exhausted, and can be quantified by analysis such as shown in Figure 1. The earlier CAGR example predicted this could happen within five years, and perhaps sooner if a new smart CPE device or consumer application takes hold.

There are a number of options that the cable operator can choose to meet the potential data capacity requirements of their customers. As pointed out by equation (1), these boil down to two primary avenues – increasing RF bandwidth, or moving to a higher QAM modulation format can provide significant improvements in bandwidth efficiency. Options such as creating the new 1 Gbps symmetrical return band above the active downstream spectrum in a triplex architecture, or moving to a non-DOCSIS modulation format are still in the early study phase. In this segment we will explore the effect of the near term options on both analog and digital return systems.

Impact of Upstream Bandwidth Expansion

RF bandwidth expansion (mid split, high split) is a potential solution that is available today or in the near future. Table 1 (repeated below for convenience) quantified the maximum data rates for several of these RF bandwidth split cases comparing current DOCSIS 3.0 rates and the upper limit predicted by the Shannon capacity theorem.

Maximum Capacity for Each Bandwidth

Return Bandwidth	DOCSIS	Maximum Capacity
5-42 MHz	150 Mbps	300 Mbps
5-65 MHz	270 Mbps	500 Mbps
5-85 MHz	360 Mbps	650 Mbps
5-200 MHz	900 Mbps	1.6 Gbps

*DOCSIS Capacity @ 64-QAM. Assumes 25 dB minimum SNR limit

As described in the overview of analog systems, analog return lasers and RF driver stages already accommodate these bandwidth extension options with minimum if any changes needed to existing deployed transmitter modules or Hub receivers. For digital return the sampling rate and laser data rate requirements for a typical 2X RF stream transmitter become increasingly difficult and expensive as the bandwidth increases. Using the same back of the envelope calculation as shown previously, Table 3 shows the optical line rates resulting from various combinations of A/D resolution and RF upstream bandwidth.

Return BW (MHz)	10 bit A/D Sample Rate	12 bit A/D Sample Rate	Laser BW Requirement
5 - 42	1.90 Gbps	2.28 Gbps	2.5 Gbps
5 - 85	3.60 Gbps	4.32 Gbps	4.5 Gbps
5 - 125	5.20 Gbps	6.24 Gbps	8 Gbps
5 - 200	8.40 Gbps	10.08 Gbps	12 Gbps

Table 3 – Digital Return: A/D Resolution, Upstream BW, and Optical Link Bit Rate

The implication here is that each incremental increase in bandwidth will require a new design iteration replacement of the current DR transmitter / receiver pair. The A/D and laser cost also increases with each iteration, driven by the increasingly higher sampling speeds.

Another factor that must be considered when expanding the upstream bandwidth is the noise increase due to the larger bandwidth and the corresponding decrease in NPR / SNR. Figure 4 and Table 4 quantify the NPR reduction resulting from the expanded bandwidth for analog optical return paths. Figure 4 shows modeled NPR performance of DFB return links, using minimum specified laser and return path receiver (RPR) performance. Table 4 shows explicitly the loss in dB attributed to uniformly sharing the fixed laser drive power across a broader signal bandwidth of constant noise power density.

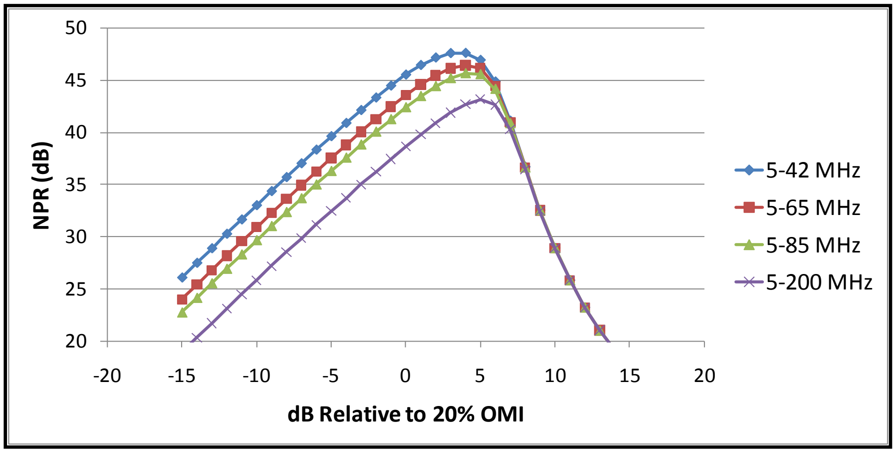


Figure 4 – Calculated NPR versus BW for Upstream DFB Links

Bandwidth Increase	NPR(SNR) Reduction
5-42 to 5-65	2.10 dB
5-42 to 5-85	3.35 dB
5-42 to 5-200	7.22 dB

Table 4 - NPR Loss vs BW

For Analog return systems the reduction in NPR due to expanded bandwidth could potentially be compensated with the substitution of a higher output power laser and/or by restricting the HFC cascade depth.

For digital returns, an increase in bandwidth often comes with a penalty to the A/D performance in terms of Effective Number of Bits (ENOB). This loss of NPR, combined with the desire to support more bandwidth efficient modulations, may dictate the use of a 12-bit A/D. In addition, any subsequent DSP operations outside the core function of the SerDes become more difficult and costly to implement in real time.

For a comparative perspective of current analog and digital technology, Figure 5 shows measured typical NPR performance of a 2 mw DFB-RPR return of nominal link length. A measured DR system using (post-processed) 10-bits of transport is shown overlaid, in each case using a 65 MHz (European) split. Note that there is link length dependence for the analog link, and the associated wavelength vs loss dependence. These variables are not drivers of NPR performance for optical fiber lengths within the digital optical link budget of the DR system, such as is commonly the case for HFC applications. Nonetheless, this data confirms the general equivalence of a digital return system achieving a full ten bits of performance to nominally performing higher power DFB returns over average HFC link lengths. It is apparent also how both technologies show comfortable margin to the higher order QAM thresholds shown.

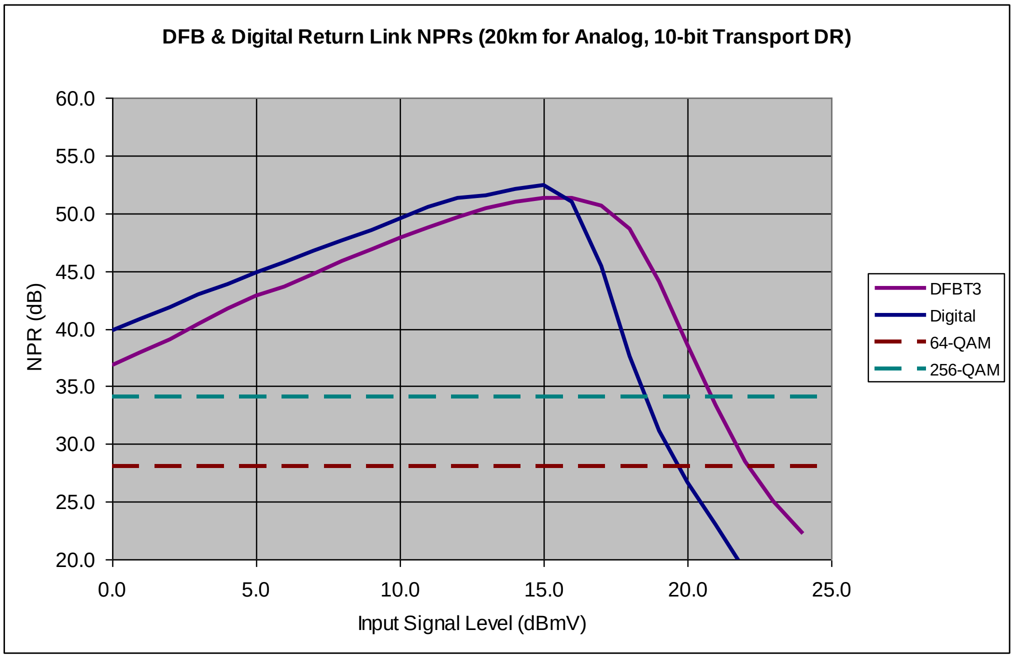


Figure 5 – Typical NPR Performance, Existing DFB and Digital Returns (65 MHz)

Impact of Higher Modulation Formats

The CableLabs DOCSIS 2.0® specification defined and expanded the approved modulation formats for downstream and upstream transport. In the upstream QPSK through 64-QAM are the accepted formats. For a single data channel 64-QAM provides roughly 28 Mbps data capacity (6.4 MHz channel). The DOCSIS 3.0 standard defined channel bonding to allow even higher data throughput rates. With four bonded 6.4 MHz channels of 64-QAM it is now possible to provide symmetric 100 Mbps data service to business and premium customers. However, as discussed, bonding does not add *new* capacity. In fact, use of bonding to create higher peak rates channels may also exhaust the available clean spectrum in many systems.

An alternative solution to RF bandwidth expansion for increasing data capacity is to move to yet higher order QAM modulations, as Table 1 indicates should be possible. While DOCSIS does not yet recognize modulation formats higher than 64-QAM for upstream transport, it is generally known in the industry that the most commonly used CMTS and cable modem DOCSIS chip sets already have these higher QAM formats built in. The benefit of 256-QAM modulation over 64-QAM is a 33% increase in data throughput for the same channel bandwidth. As described earlier and shown in Figure 2, the drawback of 256-QAM is the 6 dB higher SNR needed. Table 5 shows the 1e-8 BER-SNR relationships and the range of practical operating margins to compensate for HFC system variations for each of the higher order modulations we have discussed. The orders of M-QAM are necessary steps to delivering more bits per second on the return path, closer to full capacity, and support the associated traffic growth.

M-QAM	SNR (dB) Docsis Min	SNR (dB) Typ System
64	28	30 - 32
128	31	34 - 36
256	34	38 - 40

Table 5 – SNR Relationships for Higher Order M-QAM Formats

On the HFC plant side of the equation, the return transmitter laser peak power (NPR) and dynamic range is critical to providing the headroom needed for multi-channel loading and higher modulation rates. As discussed previously, the same DFB or DR solution enhancements suggested for expanded bandwidth would also apply in the case of providing the margin needed for higher order modulation. The higher SNR requirement for 256-QAM may also impose a limit to amplifier cascade depth, such as no more than N+3.

With high-performance HFC return paths in place, the noise performance of the CMTS upstream receiver becomes critical to making higher order modulation successful. A recent demonstration by Motorola and Cox Communications [4] used new generation CMTS receiver modules that provide significantly lower noise figure which increased usable dynamic range by greater than 10 dB. Taking advantage of this enhancement, the analog DFB, N+3 link was loaded with three 6.4 MHz, 256-QAM channels plus additional S-CDMA channels to fill the 5 to 42 MHz bandwidth, providing 141 Mbps of usable data capacity. A further test demonstrated the potential of 85 MHz bandwidth expansion by showing that with modified modems that support 85 MHz return and 256-QAM channel upstream loading a record breaking 400+ Mbps of upstream data capacity could be achieved.

Figure 6 shows the combined NPR performance of analog HFC optics and new CMTS receivers. Both legacy DOCSIS receivers and new generation cards are shown. Two important conclusions can be drawn from Figure 6. First, legacy DOCSIS equipment, while capable of 64-QAM when combined with high performance return optics, is not well-suited to supporting 256-QAM upstream, as expected. This is evident from the red trace in Figure 6 relative to the 256-QAM threshold (purple). The additional 6 dB required is simply not available with any reasonable operating margin. The second key point, conversely, is that the *new* generation of CMTS receivers, over the same high performance DFB, *can* support 256-QAM with plenty of operating margin to be practical. The newly considered 42-85 MHz spectrum tends to be considerably pure compared to the 5-42 MHz band, and in particular the 5-20 MHz band. This makes 256-QAM in this band yet more robust.

A final related point is that, theoretically, digital return optics should also be capable of supporting 256-QAM on 85 MHz return architectures. Figures 3 and 5 suggest this prospect. However, 85 MHz DR systems are still in development, so verification of this projection is not possible at this stage. This once again points to the issues of upgradability for DR systems. While analog returns can relatively easily be repurposed for extended bandwidth returns, DR systems typically require design iterations to do so due to A/D converter and sampling rate impacts of added RF bandwidth.

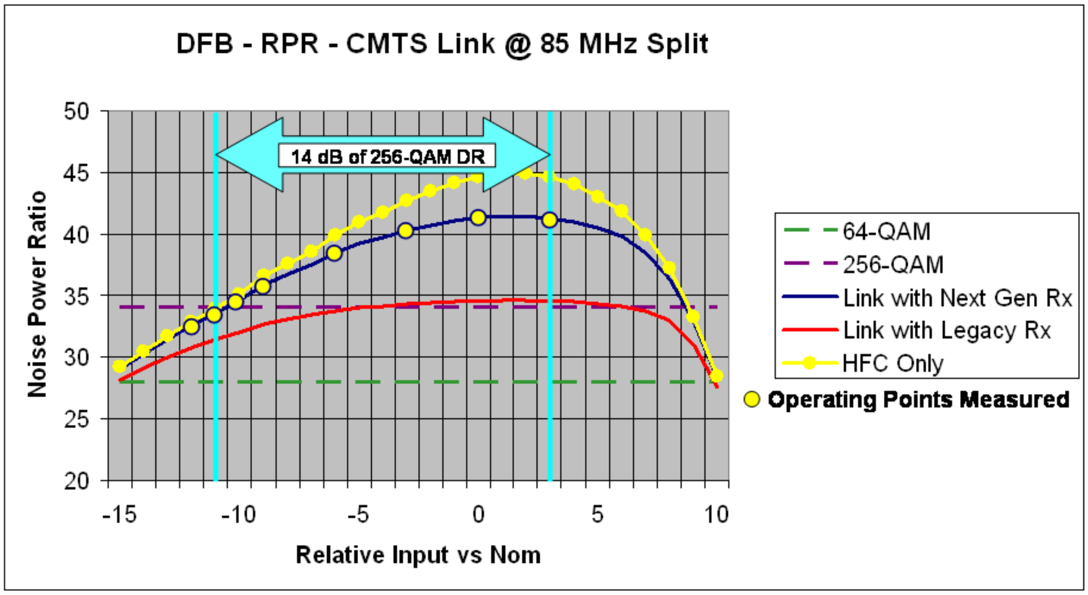


Figure 6 – HFC + CMTS NPR @ 85 MHz Split

Further testing of digital return systems is planned, although at this time only 5-42 MHz or 5-65 MHz digital return hardware is available. Actual deployments of 85 MHz mid-split systems are for the most part only planning exercises since many operators still maintain a basic analog tier of downstream video channels and commercially available cable modems supporting mid splits are not yet available. This is expected to change starting sometime in 2012.

Conclusions

Both analog and digital return laser transmitter technology are well suited to provide the wavelength option flexibility, bandwidth, and noise performance needed for the vast majority of current D3.0 deployments. This parity in operational performance and the closing gap in cost differential between the two technologies is expected to continue into the near future even as some cable operators take advantage of the mid-split upstream bandwidth option defined in the D3.0 specification.

As upstream bandwidth needs continue to grow the differences between analog and digital become more noticeable. DFB analog laser transmitters have considerable built in RF bandwidth expansion and dynamic range capability. Digital return transmitter designs, and in some cases, the matching digital receiver, must be replaced as a result of A/D chip set changes to meet sampling rate requirements. As bandwidth expands past the 85 MHz mid-split, the laser data rate requirements for digital return transmission could drive much higher costs compared to current designs and especially in comparison to analog DFB laser alternatives. Digital return solutions for 1 Gbps out of band proposals such as the above 1 GHz top split approach are not feasible due to both the A/D and laser requirements in the current designs.

Increasing data capacity by using a higher order modulation format while maintaining existing HFC upstream bandwidth allocations at 5–42 MHz to 5–85 MHz would appear to be the most cost effective potential solution for new and existing cable network deployments. New generation CMTS equipment and modems are becoming available that will make this option possible in the relatively near future. The experience gained deploying D3.0 today will allow cable operators to take advantage of 256-QAM upstream transmission if it becomes approved for use in the future. In this case the higher SNR requirements for 256-QAM may exceed the peak and dynamic range capability of lower cost 8 and 10 bit digital return solutions that are adequate for current HFC systems.

Except for extremely long reach links, DFB analog return path transmitters and receivers have the flexibility to meet a wide range of possible future data capacity enhancements in the upstream network. Technology to make analog return plug-n-play is possible but at the cost of making these systems completely proprietary and not interoperable with other deployed equipment.

Acknowledgements

The authors would like to thank Steve Frederick for measurements taken in support of this paper.

References

[1] Howald, Robert, "Calibrating the Crystal Ball for the Next Decade of Growth," 2010 Cable Show, Los Angeles, CA, May11-13.

[2] Howald, Robert and Floyd Wagoner, "Right-Sizing Your Fiber-Deep Architecture," Communications Technology Webcast, August 10, 2010

[3] Howald, Robert and Phil Miguelez, "Upstream 3.0: Cable's Response to Web 2.0," 2011 Cable Show, Chicago, IL, June 14-16, 2011.

[4] Robuck, Mike, "Cox, Motorola lay claim to new return path speed record," CedMagazine.com, March 01, 2011.

UPSTREAM 3.0: CABLE'S RESPONSE TO WEB 2.0

Dr. Robert Howald and Phil Miguelez
Motorola Mobility

ABSTRACT

It has been a little over ten years since Upstream 1.0: Video Service Support migrated to Upstream 2.0: Internet Access. Continuing traffic growth coupled with consumer expectations mark the next phase of planning. In particular, today's web experience includes new behavioral and service type accelerants that continue to apply pressure on upstream capacity. On the behavioral side, the Web 2.0 paradigm is always-on, real-time, high-speed access that delivers the cloud of connectivity that consumers expect. The percentage of the population documenting their everyday lives minute-by-minute and tracking others doing the same has grown rapidly. On the services side, we have seen staggering growth in user-generated content (UGC) tied to the social networking phenomenon and in multi-media uploads for content sharing on sites such as YouTube. These image and video-centric experiences represent a substantial increase in bit volume. As a result, MSOs now must plan accordingly and deliver Upstream 3.0: Web 2.0.

In this paper, we explore the implications of Upstream 3.0 and options for successfully delivering on these requirements. The two components to consider are optimizing the efficiency of available spectrum, and preparing an upstream migration plan that can support an aggressive compound annual growth rate (CAGR). Optimizing available spectrum includes utilizing key DOCSIS® 3.0 tools such as wideband 64-QAM, S-CDMA's newest features, and pre-equalization. Positioning the network for anticipated CAGR involves recognizing its implications on the upstream lifespan and developing a strategy to address these implications. As CAGR marches steadily ahead, operators are at risk of running out of capacity within the business planning horizon. In this discussion, we will focus on the ability to harvest new DOCSIS® 3.0 capable spectrum up to 85 MHz and subsequently beyond. We will quantify the potential upstream capacity, illustrate its ability to deliver on long-term growth projections, and describe the implications in the plant and in the home. We will describe steps to ensure that these gains are fully realized, and to guarantee that potential additional expansion is cost-effectively achievable. Finally, we will describe recent HFC and in-home environment measurements that can guide operators in implementing the migration.

INTRODUCTION

With the introduction of high-speed data services over DOCSIS over a decade ago, signaling the end of the dial-up age, broadband users have easily found ways to consume the new bandwidth. At first, it was simply a faster web surfing experience. But shortly thereafter, the web experience itself was modified as creative entrepreneurs rapidly made use of the bandwidth in a surge of development to create the most compelling web pages, and large scale e-commerce opportunities emerged. While surfing is primarily a downstream experience, DOCSIS introduced a significant jump in upstream speeds as well, ushering in new behaviors and activities – in particular peer-to-peer file sharing. It was this peer-to-peer beast that drove the period of bandwidth growth beginning early in the last decade and lasting several years, in part as legal battles wound down.

As peer-to-peer wound down, however, important new social experiences were well underway, growing up from low-bandwidth but highly popular chat (AOL messenger, chat rooms) to media sharing experiences, online games, and user generated content (UGC) driven by the standard set by YouTube. The social and community networking experience was yet one more element created by these pioneering network speeds, but also by the responsiveness and what was gradually becoming ubiquitous access. Ubiquity began as finite set of anywhere's – coffee shops, airports, restaurants, bookstores, etc. outfitted with WiFi. Today, it extends to truly anywhere because of the universality of the smart device. Virtually every phone today qualifies as a Smartphone relative to years gone by. Along with the needs of businessmen and women, some Web 2.0 habits in particular (Facebook/Twitter Updates, Sports GameCenter/Fantasy, mobile photo uploads) have led to a 24/7 connected paradigm and ubiquitous, portable access. This always-on, always-connected, always immersed in social networks, experiences, and media swapping, is the driving force of the New Era of Internet usage, Web 2.0.

Web 2.0 is simply about how the modern Internet is used – augmenting classic search and consume browsing with usage behaviors around UGC publishing, social networking, media sharing, media-centric (Flickr) and community-oriented web services and applications (Groupon, ESPN Fantasy, PokerStars.net), etc. Real-time broadband two-way IP sessions are the basis of these activities. Along with a newly emerging interest in home automation and associated services, these activities are the new residential pressure on upstream bandwidth. DOCSIS is cable's IP connectivity system, and its capabilities are in large part responsible for there being the evolution we now call Web 2.0.

UPSTREAM GROWTH MANAGEMENT

Consumer Expectations and Trends

Authoring and publishing of content locally – UGC – in the form of pictures, music, and video, requires a robust upstream data service. DOCSIS enables a high-speed upstream; however, tremendous pressure will be placed on current DOCSIS capability to keep pace with continued growth. That the social networking aspect of web life has grown is not a secret, and that UGC has become a major part of it is also no surprise. However, Figure 1 puts into perspective the scale of that growth. It represents that "yellow" or "red" flag for operators trying to grasp what Web 2.0-type activities create for the HFC upstream that simple web page (browsing) requests did not. And, this is ahead of any surge of home automation services that may lie ahead.

Figure 1 compares various aggregates of professional media content as measured in Petabytes (10^15 bytes). The right-most bar represents an estimate of the largest single day of photo uploads to the cloud in the United States in 2008. This one *day* (July 4? Mother's Day?) of upload activity represented more than twice the entire *year* of original television content generated worldwide. The photo upload bar also towers over the universe of all-time movie releases (somewhat impressive) and music (more apples and oranges). Fortunately for cable, a significant amount of these uploads are synchronized with Smartphone universality and Web 2.0 behaviors, not simply family photo albums moving between homes and clouds. As such, a large burden for this upstream activity is also being placed on wireless networks. However, as wireless networks, and in particular 3G networks, become crowded with new traffic, offloading of the traffic to nearby WiFi infrastructure often will ultimately have some of it ending up on cable networks. Furthermore, MSOs are actively courting new WiFi traffic with distributed plant access points for broader wireless coverage, and to take advantage of the rapidly scaling count of portable IP devices with sophisticated processing capability.

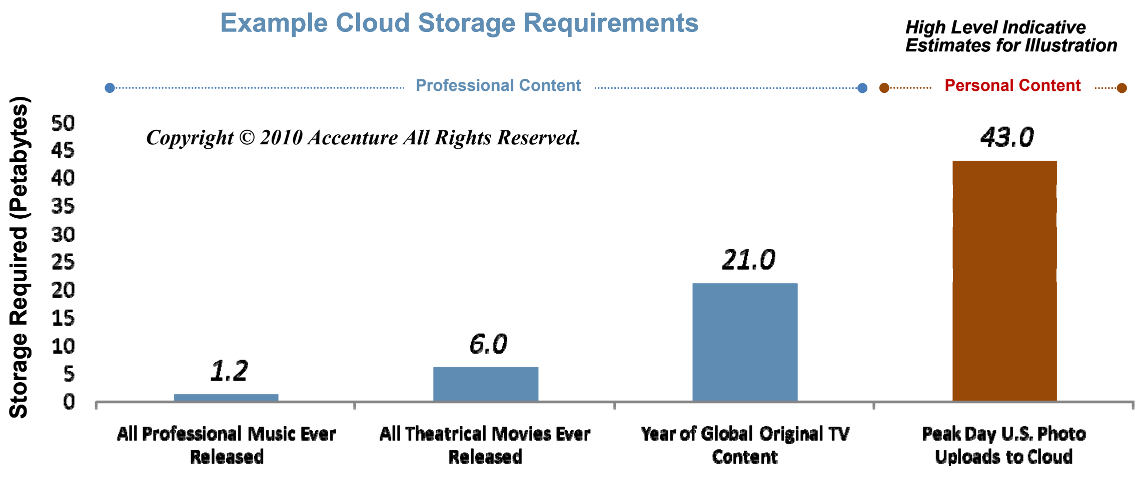


Figure 1 – Web 2.0: Upstream UGC Demand Two-Way Broadband

Tracking Bandwidth Growth

Traffic models based on Nielsen's Law, which uses a compound annual growth (CAGR) methodology, have been shown to represent well historical traffic growth trends. While actual traffic tends to grow and stagnate in a staggered fashion over the years, Nielsen's law has historically been a sound way to mathematically describe it, with the understanding that it works best as a long-term representation. As such, there is an element of "placing your bet" with regard to observing the historical trends and usage in the context of the new paradigm delivered by Web 2.0 experiences. Because of the short term fluctuations, there is a need to engineer ahead of the curve so as not to be waylaid by an unexpected step function in growth. There are several potential applications that we could ponder that could create such a step, but history has also shown that trying to guess the winners and the timing has been difficult. A new service or application often scales quickly and often catches us by surprise, creating "Napster" moments.

Figure 2 uses Nielsen's law under three different CAGR assumptions – 30%, 40% and 50%. It is Nielsen's law applied to a node or service group aggregate, under the assumption that average per-user increases are reflected similarly by the aggregate. The three trajectories are interrupted by two breakpoints over the next ten years. These represent node and/or service group splits – effectively 3 dB (best case) offsets, or a doubling of average bandwidth per home. Note that the 3 dB would be a step *straight* downward by 3 dB at implementation, so that by the time the next year comes around, some of that has been consumed, and the year-to-year step is less than 3 dB.

These trajectories are plotted against three different HFC upstream thresholds, using raw physical layer transport rate, as we will do throughout for consistency:

- 60 Mbps – Approximately two 64-QAM DOCSIS channels at 5.2 Msps
- 100 Mbps – Approximate available bit rate in 5-42 MHz with only A-TDMA
- 150 Mbps – Approximately a fully utilized 5-42 MHz using both A-TDMA and S-CDMA

With this information juxtaposed on Figure 2 with the CAGR trajectories, we can estimate at which point the various CAGRs will exhaust the available upstream throughput. The starting point of the trajectory – a key point from which all subsequent growth takes place – is an assumption of providing a maximum tier of 5 Mbps, while traffic engineering (TE) at 50:1 oversubscription on a 500 hhp node at 60% penetration. This nets out to providing 30 Mbps of total upstream to that service group, or one fully utilized wideband 64-QAM channel when hitting the 2% concurrent use metric at peak service hour.

The working assumption, then, can be looked as keeping pace with tier rate increases that follow demand-based CAGR, or managing the TE aggregate to keep pace with CAGR. It is a simple matter to shuffle the starting point on the chart up or down based on specific cases – not yet offering a 5 Mbps tier, averaging 5 Mbps with the economy tiers for a net average service rate, higher or lower rates of oversubscription and penetration, etc. The entire trajectory simply scales up or down by the same amount of dB offset. In addition, in keeping pace with user demand, more upstream comes with the price of delivering increases in downstream tiers. These increases must be supported by some consistent level of up/down asymmetry to be effective.

Finally, it is important to emphasize that it is a demand-based consumption analysis. Competitive market forces also have an important impact on the service rates offered.

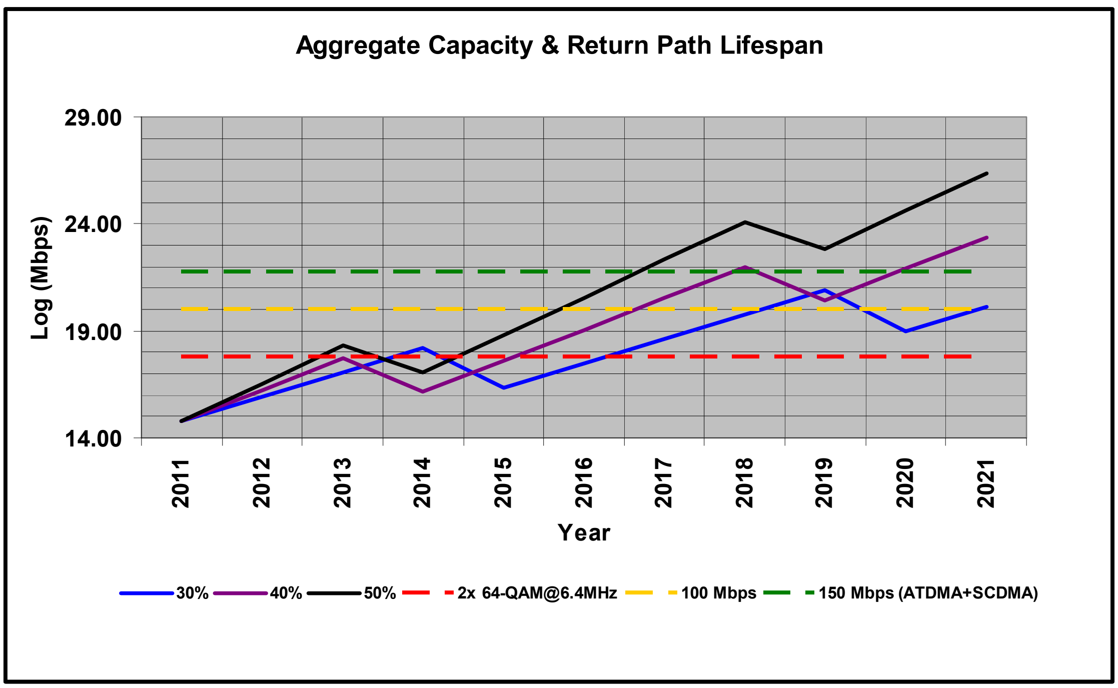


Figure 2 – Using CAGR to Project Upstream Growth Thresholds

Some key conclusions can be drawn from Figure 2. Clearly, a couple of 64-QAM DOCSIS channels get exhausted within a few years without a service group split. MSOs generally have recognized this and are planning a generation of splits or executing them on an as-need (traffic demand) basis already. More important, however, is to estimate when 5-42 MHz itself gets exhausted. While node splits are costly and intrusive, they are business-as-usual (BAU) type activities operationally. When demand or market needs push beyond what 5-42 MHz can provide, a more significant change must be considered. And, while further node splitting will indeed provide more average bandwidth, the maximum service rate limit also come into play as we look further into the future and project towards 100 Mbps upstream service groups.

For example, in an A-TDMA only case, a best case scenario may be providing a 100 Mbps peak service rate from the throughput available, with a collection of bonded 64-QAM channels providing the best effort service. One such service group could be enabled for upstreams well-behaved enough to support the bandwidth and modulation profile (64-QAM) required. Service rates significantly above 100 Mbps would be unattainable

Referring again to Figure 2, note that a single service group split gets us through 4-7 years of growth, based on CAGR assumptions shown, considering 100 Mbps as the 5-42 MHz throughput boundary. The former (4 yrs) represents a "should be planning what's next" time frame, while the latter registers less urgency. Again, a second service group split would buy more time, but would not address peak service rate growth objectives. In particular, the ability to offer 100 Mbps to commercial customers would be limited in a 5-42 MHz only architecture, and shared with residential users. With diurnal variations of commercial and residential, this may be reasonable.

Note that through use of S-CDMA (green) and an assumption of relatively robust (40%) average growth, the upstream could last the decade. Recent data [1] suggest short term CAGRs have slowed, registering at about 25%. This represents, approximately, a three year traffic doubling cycle. Critical variables to the 40% result, however, include the aforementioned commercial services, and the lingering possibility of an aggressive CAGR (50%) bandwidth-buster type application, either of which could break this attractive conclusion. The 50% CAGR would pull the bandwidth exhaustion point in to 7 years. Finally, well before the 5-42 MHz is exhausted, planning will have had to begun, and in fact steps taken, to be prepared with new bandwidth. This is simply because the nature of any new steps involves some intrusive changes in the plant and home environments.

Figure 2 acts as a useful guide to managing growth versus time, and is relatively easy to extrapolate into other circumstances more specific to an operator or region.

OPTIMIZING LEGACY 5-42 MHZ RETURN SPECTRUM

Given the inevitability of a bandwidth bottleneck, with the possibility that it could be a near-term concern, it is clearly important to understand how to squeeze every last bit per second out of the 5-42 MHz upstream. Important underlying assumptions were made in the prior section in order to add up all of the possible return path bits per second and to determine thresholds to mitigate CAGR. In particular, we assumed 64-QAM upstream links at 5.2 Msps (6.4 MHz) were possible, that A-TDMA would limit the ability to extract throughput out of the 5-42 MHz return paths, and that S-CDMA opened up extra bandwidth where A-TDMA would not. Much has been written and discussed on these topics [2][3][4]. We summarize some of those findings here.

Optical Link Performance

Figure 3 singularly summarizes the HFC element of return path optics relative to their ability to support increasingly sophisticated modulation profiles. The return path optics, which typically sets the performance of the plant itself (Home and HE not included), are displayed from least capable to most capable, beginning with Isolated Fabry-Perot (IFPT, red trace), in each case using minimum specified performance. Note that yet older Fabry-Perot lasers exist, which are non-isolated. They have even lower performance. These are not shown here because they play no useful role in the enhancement of the return path for high-speed data services. They were deployed in their era for supporting interactive STB traffic, which is very narrowband and implements a very robust modulation profile.

From the IFPT, the additional traces identify minimum performance of the latest generation of FPs (EIFPT), two Distributed Feedback (DFB) lasers – 1 mw and 2 mw – and a digital return system based on 10-bits of transport (for this case only, measured).

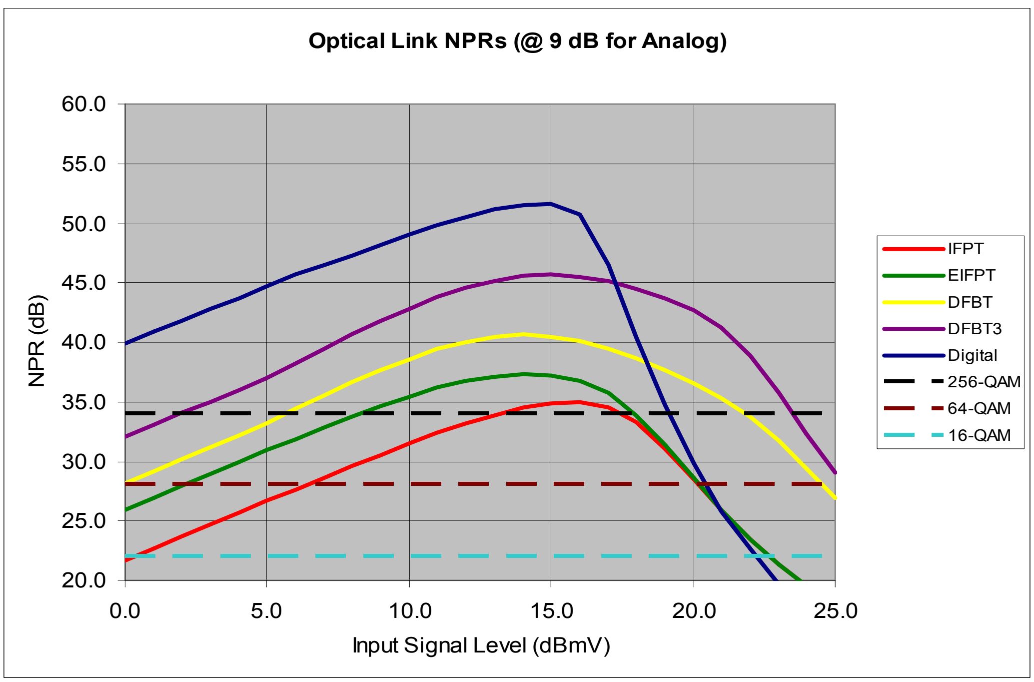


Figure 3 – Optical Link Noise Power Ratio vs Technology

The QAM thresholds shown represent 16-QAM, 64-QAM, and 256-QAM, the latter being non-DOCSIS. The values shown represent SNR values of 22 dB, 28 dB, and 34 dB, respectively. On the left-hand side of the NPR curve, away from the peak, NPR and SNR are one in the same. The area around the peak and to its left is the practical operating region of the curve. The concept of Noise Power Ratio as it applies to HFC returns is described in great detail in [5]. The basic NPR concept is that it represents the performance of the return link if it were fully loaded with digital channels, with the total power of those channels being the value on the x-axis.

The QAM thresholds chosen represent approximately 1e-8 *Bit* Error Rate values. From this, one can consider that there is forward error correction (FEC) coding gain that means several dB *less* is acceptable as a performance threshold, as is done for the DOCSIS downstream (23.5 dB/30 dB for 64-QAM/256-QAM post-FEC). Conversely, one can reserve coding gain for the unknown that often is symptomatic of the return channel, and add several dBs above the shown thresholds to identify what is a comfortable operating margin. Those policies vary across operations and system architects, so we use a value that comes with no gray areas to consider.

For reference, Motorola has taken the additional step of correlating measured *codeword* error rates (i.e. with FEC) to packet error rates (PER) – that which would reflect end user experience. The results indicate that an uncorrected *codeword* error rate of 1e-3 approximates the threshold beyond which a steady increase in packet errors ensues beyond an acceptable level, as shown in Figure 4. Many bit errors are required to make a codeword not correctable, however. Suffice to say that a 1e-8 BER would be error-free post-FEC. So, in that sense, use of the above SNR thresholds is conservative relative to packet loss. The interested reader is referred to [2] for more detail.

Referring back to Figure 3, we quantify in Table 1 the operating dynamic range for the two higher order modulation profiles provided by each of the return path optical technologies shown.

Packet Error Rate vs Uncorrectable Codewords

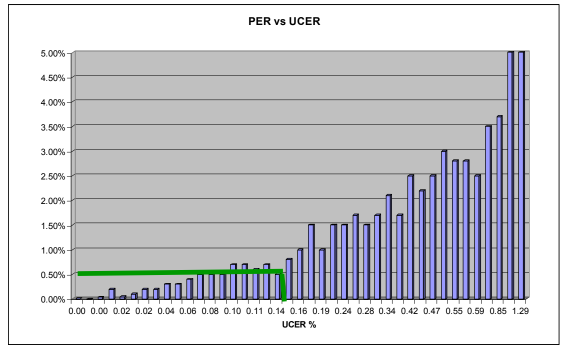


Figure 4 – Codeword Error-Packet Error Correlation

Table 1 – M-QAM vs Dynamic Range

	Dynamic Range	
Return Optics	64-QAM	256-QAM
IFPT	8	2
EIFPT	11	5
DFB (1mw)	15	9
DFB (2mw)	19	13
Digital (10-bit)	25	19

Several conclusions can be drawn from Figure 3 and Table 1:

- First, Fabry-Perot (FP) lasers start to lose comfortable margin levels when 64-QAM is deployed. Moving from 16-QAM to 64-QAM is at minimum 6 dB of lost margin. Link performance shown is further compromised by other built-in system noise, such as deep cascades and receiver noise figures, further degrading margin. We will quantify some of this loss in a subsequent section.

- 64-QAM can be supported over modern FP lasers, but the lost margin is more likely to make deployment challenging. It could result, for example, in inconsistent performance and/or maintenance due to impairments that do not bother 16-QAM. As such, getting started with existing FPs is quite possible and allows a gradual migration to DFBs to take place, but there must this migration must take place to fully maximize throughput with robustness.

- DFB optics and digital return optics comfortably support DOCSIS 64-QAM performance. Furthermore, these technologies offer the potential to increase the modulation order to yet more bandwidth efficiency using 256-QAM. In the case of the 2 mw DFB and 10-bit digital returns, double digit (i.e. comfortable) margin exists for supporting the additional 33% new bandwidth efficiency of 256-QAM over 64-QAM.

The moral of the optical link part of the "optimizing" story is simple – migrate towards DFB optics or digital return to fully utilize the 5-42 MHz return. Which direction to go includes various other factors that weight the pros and cons of analog and digital [6], in particular the capability to continually support more new RF spectrum to be discussed herein.

S-CDMA

It has long been understood that the low end of the return band is a messy place to live, fraught with short wave radio interference and home-induced impulsive noise from a variety of common sources. As such, most MSOs write-off the region of spectrum below 15 MHz for DOCSIS services, perhaps even below 20 MHz. Figure 5 shows a classic example of the kind of muck a signal may find itself suffering through in the lower half of the spectrum. In this case, the burst is wideband enough to intrude on a portion of the DOCSIS channel in a way that impacts short term Modulation Error Ratio (MER). Resulting MER behavior from such an event is shown in Figure 6.

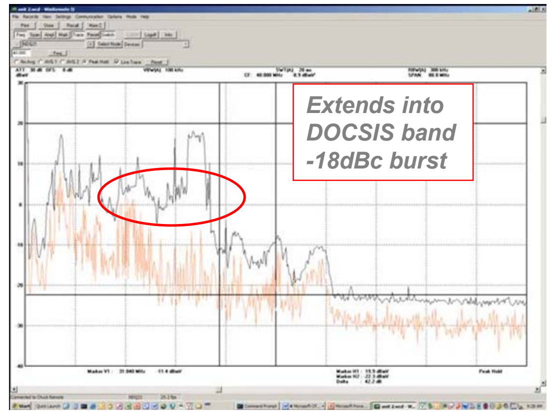


Figure 5 – Spectrum Capture at HE of an Impulse Noise Burst

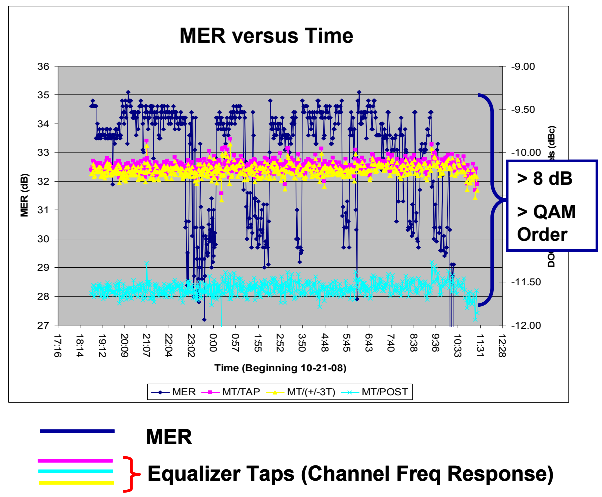


Figure 6 – MER Behavior Under Severe Impulse Noise

While inherently unfriendly territory, DOCSIS 2.0 put in place the mechanism to deal with this portion of the band. DOCSIS 2.0 includes the requirement to support S-CDMA. Figure 7 shows a recommended channel distribution [4] using S-CDMA to squeeze every last bit out of 5-42 MHz. In this case, S-CDMA offers 33% more capacity. Under complete avoidance of the low end of the band, up to 50% more throughput can be added compared to today's usage. We will utilize these S-CDMA conclusions to determine its role in extending the lifespan of the return path for the legacy 5-42 MHz bandwidth, as well as for subsequent new bandwidth growth.

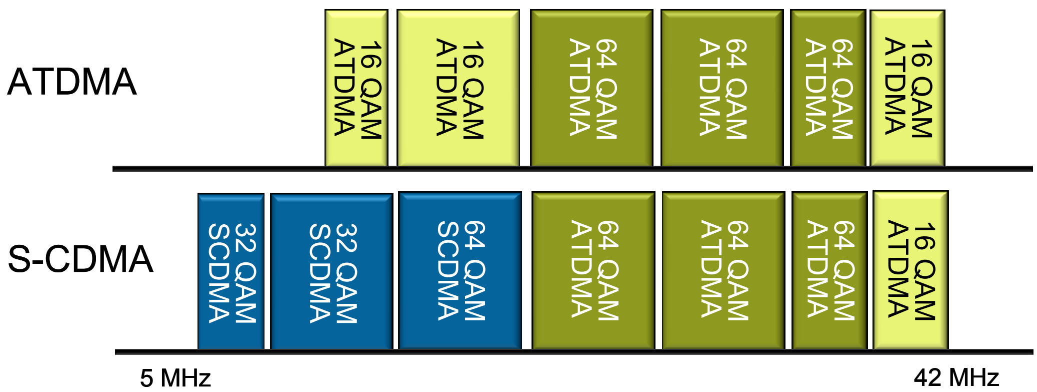


Figure 7 – Maximizing Available Throughput Using S-CDMA

<u>Capacity and Lifespan</u>

In Figure 2, we showed upstream CAGRs of 30%, 40% and 50%, and showed how those rates crossed various thresholds of available throughput in the 5-42 MHz band. We pointed out the critical nature of the CAGR variable itself – something anyone that has an interest-bearing savings account (is there such a thing anymore?) or had a securities account in the 1990's understands well. Dollars or Megabits, compounding math works the same way.

In Figure 8, we have displayed the same information in a different fashion, allowing us to understand the sensitivity of the exhaustion of the 5-42 MHz HFC plant relative to the CAGR assumption. The same finite set of threshold conditions are displayed, and we plot the curves from the same starting point as discussed for Figure 2 (5 Mbps peak, 50:1, 500 hhp@60%). In Figure 8, service group splits are instead represented by dashed traces for the 100 Mbps and 150 Mbps cases (only, for clarity).

The two crosshairs on the figure are positioned to help understand the interpretation between Figure 2 and 8. For example, observe the trajectory of the 50% CAGR and note the point at which it exhausts a 100 Mbps maximum throughput channel in Figure 2. This occurs slightly more than 4.5 years into the future. We can see this same point represented by the leftmost crosshair on the 100 Mbps (pink dashed) curve in Figure 8. Similarly, we can observe in Figure 2 that the 40% CAGR trajectory crosses the 150 Mbps threshold in seven years. This matches the crosshair marking the 150 Mbps threshold curve (dashed yellow) in Figure 8. In both cases, in Figure 2, these trajectories also crossed the thresholds mentioned after undergoing a service group split, thus the correlation to the dashed curves in Figure 8.

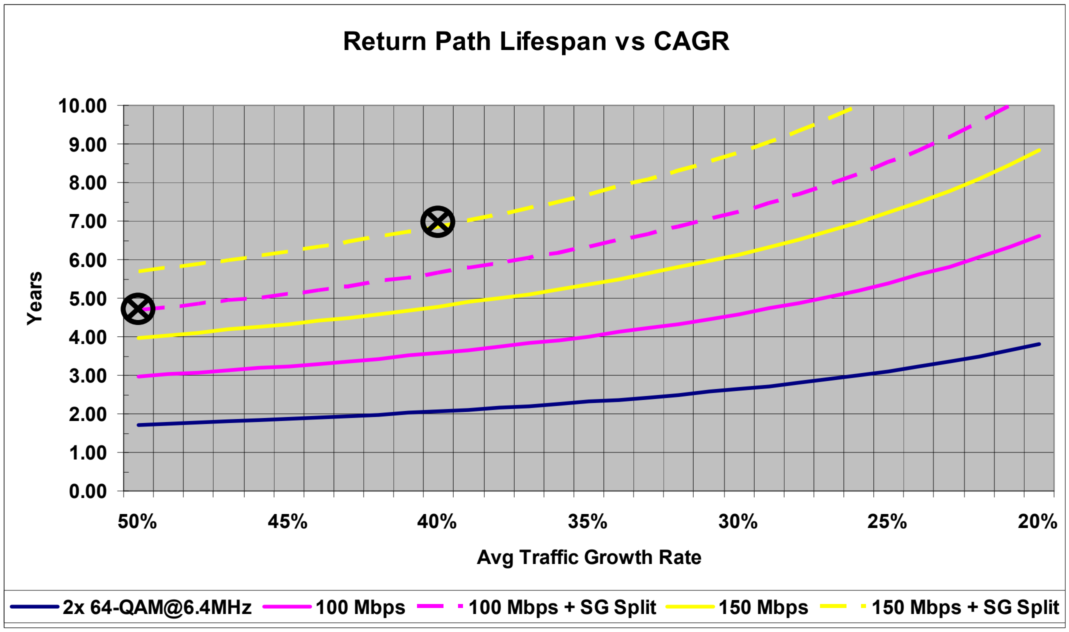


Figure 8 – Lifespan of 5-42 MHz vs CAGR

Figure 8 can be used as an excellent guide to the state of urgency (or not) of an operator's upstream situation based on historical and projected CAGRs. Its usefulness stems from the granularity provided by the perspective of Figure 8 when considering the sensitivity to the variable of most impact (CAGR).

Some of the core conclusions from Figure 2 can be drawn as well from Figure 8 under the given introductory assumption. A near-term need to add DOCSIS channels upstream will continue. At least one service group split is most assuredly on the horizon to provide support for average bandwidth growth, even under relatively modest CAGRs of 25-30%. This is the case in particular if operators continue to avoid the low end of the spectrum where S-CDMA can buy additional time.

As with Figure 2, Figure 8 quantifies average bandwidth growth from a demand standpoint. The same principles discussed around supporting increasingly high residential peak rates, 100 Mbps commercial service opportunities, and higher rates due to "market push," still apply.

DELIVERING NEW DOCSIS CAPACITY

Benefits of 85 MHz Return (N-Split)

An unavoidable conclusion from Figure 8 is that the end is in sight for upstream capacity. It may not be near (or it may be), but it is certainly to the point where planning for what comes next is prudent. For practical reasons, there is no intent to remake the 5-42 MHz modem technology to squeeze out any latent Mbps that may be possible given that DOCSIS is 10 years old, the RF section of the HFC plant has gotten shorter, and serving group sizes have gotten smaller. In fact, we have begun to take advantage of some of these HFC performance benefits within DOCSIS today by turning on 64-QAM modulation, which is a step closer to optimum use of the upstream channel compared to 16-QAM. And, any new modem technology developments that take place in the future for HFC and take advantage of the latest developments in communications technology and information theory to close the gap between current performance and theoretical capacity can always call out frequency range support for the legacy DOCSIS bands.

Table 2 illustrates the available DOCSIS transport rate for various split architectures, and the theoretically available capacity at the DOCSIS-specified minimum of 25 dB. While it is impractical to achieve theoretical capacity, the gap has indeed closed over time between practice and theory.

Table 2 – Bandwidth, DOCSIS, and Theory @25 dB SNR

Maximum Capacity for Each Bandwidth

Return Bandwidth	DOCSIS	Maximum Capacity
5-42 MHz	150 Mbps	300 Mbps
5-65 MHz	270 Mbps	500 Mbps
5-85 MHz	360 Mbps	650 Mbps
5-200 MHz	900 Mbps	1.6 Gbps

Note once more we using transport rate as the basis for all number unless otherwise mentioned. Actual user throughput upstream is highly dependent on upper layer (MAC) parameters, scheduling, packet sizes, and burst overhead.

Working down Table 2 from top to bottom, one obvious place to look for new capacity is simply new bandwidth. One straightforward way to exploit new bandwidth and remain compatible with DOCSIS is use of the 85 MHz return band, referred to as the N-Split. The limit set at 85 MHz was wisely chosen to maximize clean low band return without impinging on the FM band that could render some of the band difficult and unpredictable to use.

The advantages of considering expansion to N-Split are numerous. The primary benefits are listed below, and we will quantify several of them in subsequent sections.

- N-Split is supported by DOCSIS 3.0 for cable modems and CMTS
- Existing silicon is already capable
- Very clean new spectrum 42-85 MHz offers the potential for higher order modulations
- Legacy STB out-of-band signals (tunable across 70-130 MHz) are supported
- Entails minimal encroachment into the downstream band
- Has similar cable loss versus frequency properties as legacy band
- Multiple 100 Mbps tier serving groups are possible
- Support for traffic growth lasts through the decade under aggressive CAGR assumptions
- Architecture remains a diplex-based frequency domain duplex (FDD)
- Systems are possible with analog or digital return technology; analog already supports

We will compare additional properties of the 85 MHz approach to alternatives for supporting Gbps capability in later sections.

<u>New Return Spectrum: HFC Performance</u>

The opportunity to light up new, clean spectrum has the advantage of elimination of the impairments commonly seen in the lower half of the 5-42 MHz spectrum and shown in Figure 5. However, this added bandwidth does come at the expense of increased load to the return laser, and its fixed available allocation of total power. With the available bandwidth more than doubling, this nets out to 3.3 dB of SNR loss to the signals sharing the load over the fixed noise spectral density of the link. This loss is shown in Figure 9.

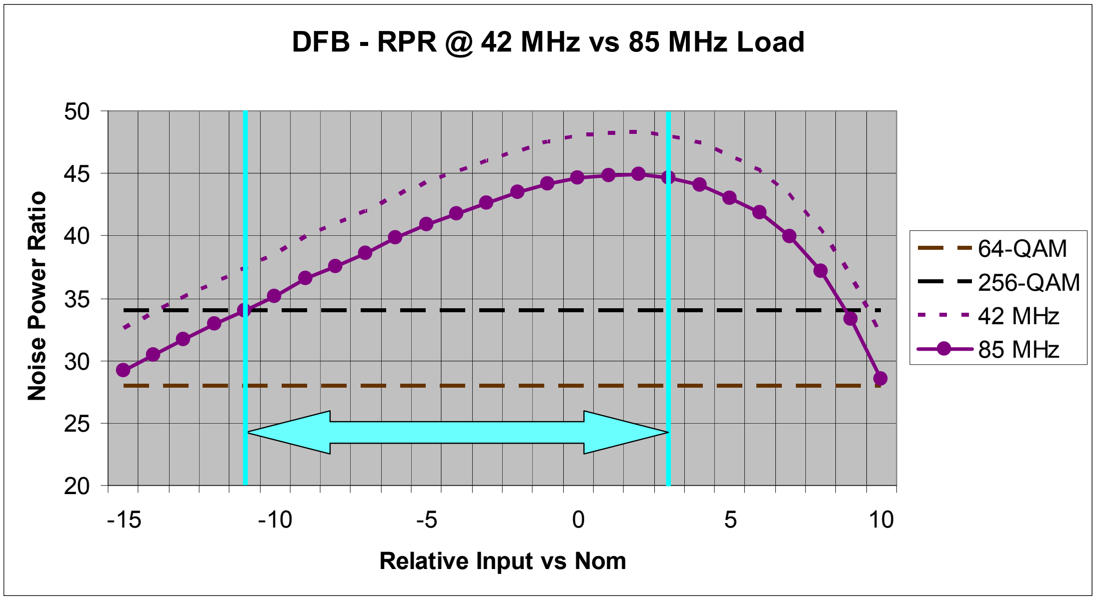


Figure 9 – Bandwidth Loading Effect (2 mw DFB Link @ 20 km)

Despite this loss in available SNR, it is apparent from Figure 9 that high quality return path optics – in this case, standard 2 mw, 1310 nm DFB lasers – have significant margin to support both 64-QAM and 256-QAM over typical HFC link lengths. In fact, we would estimate from Figure 9 that the HFC part alone has 13-14 dB of dynamic range over which 256-QAM could be supported in this scenario, quite operationally practical. As we shall see, this does not mean it is necessarily simple to flip the switch to 256-QAM. The HFC optical link represents but one component of the system, although it is the dominant contributor to performance of the plant itself.

Figure 10 shows a snapshot of a recent trial of an N-Split architecture, where the upper half of the band was used to support 256-QAM channels. In this case, supporting this modulation profile while co-existing with a maximum legacy 5-42 MHz load based on 64-QAM, was under evaluation. A mid-band test channel was left unoccupied for monitoring the most probable location of maximum distortion build-up as dynamic range was exercised, as in an NPR test.

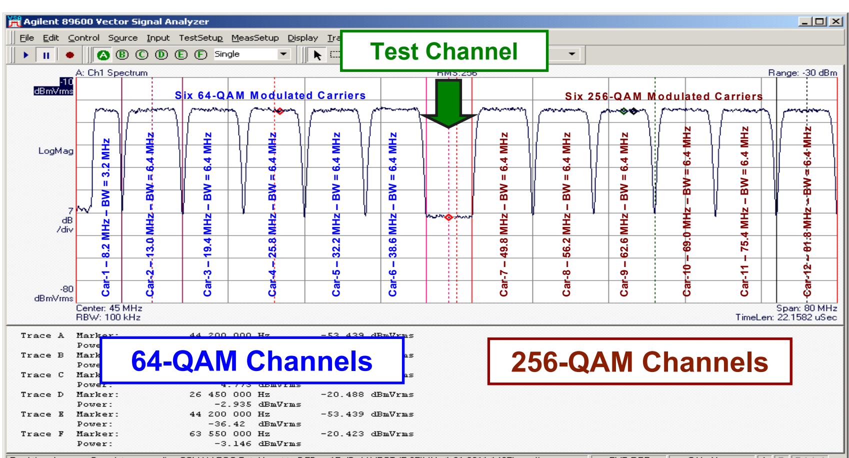


Figure 10 – 64-QAM and 256-QAM Channels Over N-Split

Evident from Figure 10 is the high available SNR delivered by the HFC link using existing analog DFB return optics at nominal input drive. The available SNR as measured at the input to the CMTS receiver of about 45 dB is consistent with the performance shown in Figure 9. In the case of Figure 10, the complete upstream link included a three amplifier RF cascade (i.e. an N+3 system). This performance is also consistent with support for higher order modulation profiles, such as 128-QAM and the 256-QAM example shown in Figure 9. Each represents a further increment towards closing the gap between theoretical capacity and practical data rates.

Considering the 5-85 MHz line in Table 2 and Figure 10, we can estimate that the gap between actual and theory is closed from 55% of capacity to 73% of capacity if migration across the band to 256-QAM were implemented. Of course, and as the 5-42 MHz line in Table 2 implies, extracting more out of the lowest end of the band is difficult. S-CDMA obviously helps considerably, as shown in Figure 7. However, with ideal channel capacity itself being a theoretical additive white Gaussian noise (AWGN) construct, it is to be expected that in the region where the noise is no longer AWGN-dominated that closing the gap would be more difficult. S-CDMA was leveraged at the low end of the band for the 85 MHz results shown above. Both with and without S-CDMA cases are quantified for N-Split analysis in subsequent discussion

Clearly, an inescapable conclusion from Figure 9 and 10 is that high performance analog optics – today's vintage of DFB-RPR links – are proven capable of supporting N-Split return links with higher order modulations for next generation upstreams.

New Return Spectrum: System Performance

As indicated, the HFC link is very capable of supporting higher order modulations over wider bandwidths than currently deployed. However, early generation CMTS equipment was designed to support 16-QAM as the maximum modulation profile requirement. Most vendors took this guidance and provided enough margin in their systems to enable 64-QAM, which was embraced in DOCSIS 2.0. However, enabling 256-QAM (an additional 12 dB of performance, minimum, over 16-QAM), was not cost effective to consider in early stages of DOCSIS deployment. This is not the case, however, in some newer receivers.

Given DOCSIS 2.0 requirements to support 64-QAM, as well as the requirement to support S-CDMA, which entails another degree of fidelity associated with its synchronous operation, new receivers coming to market, such as Motorola's RX48, have increased margin and dynamic range. Because of this, 256-QAM can now be comfortably supported. DOCSIS does not yet call out 256-QAM. However, much of the existing silicon base already supports this mode using the basic physical layer architecture of DOCSIS 3.0. Figure 11 quantifies the performance of the end-to-end system.

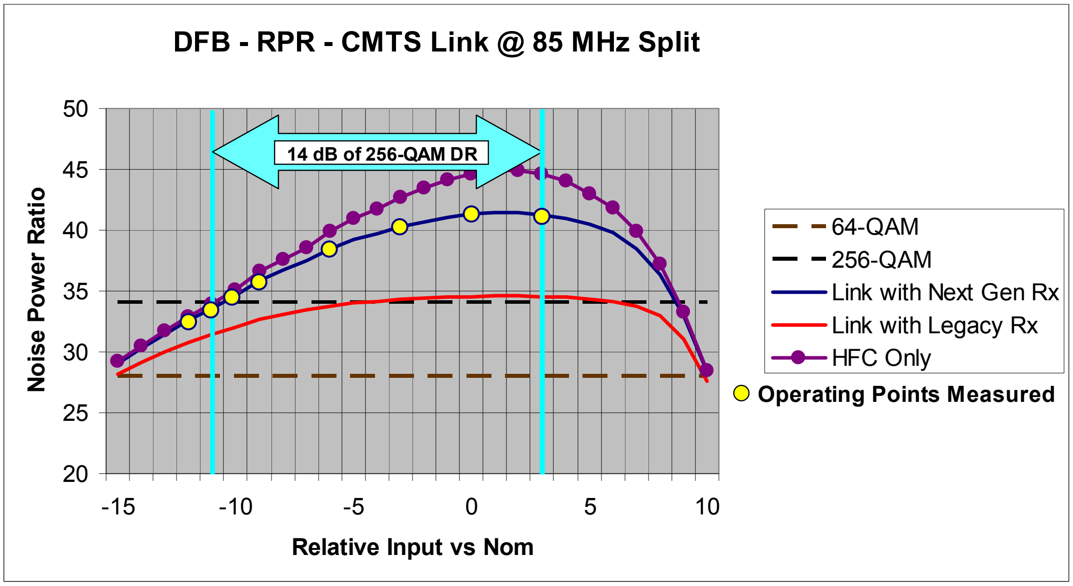


Figure 11 – 256-QAM Dynamic Range Over N-Split

Several key points can be taken from Figure 11. First, DFB HFC optics plus today's CMTS receivers comfortably support 64-QAM with sufficient, practical, operating dynamic range. This lesson is being proven everywhere DOCSIS 3.0 is being deployed. As previously described, in some cases newer, high quality FP lasers such as shown in Figure 3 can support 64-QAM as well. However, moving forward with more channels loaded and higher bandwidth, it is recommended that these eventually be replaced with DFBs. Nonetheless, it is comforting to know, given the magnitude of this task, that newer FPs can get 64-QAM started while the large task of exchanging lasers methodically takes place.

A second key point is that, using new CMTS receivers with extended dynamic range, such as the RX48 previously described, sufficient performance exists for 256-QAM to be practical. In fact, it is supported with nearly the same dynamic range that existing receivers provide for 64-QAM. Finally, comparing the HFC (purple) NPR trace to the HFC+CMTS (blue) trace, it is apparent also how little loss of NPR is incurred by new high fidelity CMTS receivers. Note that the yellow marked points on the composite NPR trace of Figure 11 represent low packet loss, *measured* loading points during the testing phase.

To complement the above field trial work performed with Cox Communications [7], Motorola chose to optimize loading further to determine the maximum throughput supported under the same HFC optics and the new generation of CMTS receivers. Both S-CDMA and A-TDMA were utilized, using 12 carriers employing modulations from 32-QAM to 256-QAM across the band. As before, however, 256-QAM was implemented only in the band above 42 MHz (and as such, still leaving some possible additional bits per second on the table). The 32-QAM recognizes that at the low end of the return band, even S-CDMA is challenged to overcome the noise and interference that congregates in that area. Table 3 quantifies the results of this evaluation.

Table 3 – Fully Loaded N-Split, A-TDMA + S-CDMA

5 MHz to 85 MHz Throughput Performance

	Frequency	Bandwidth	Symbol Rate	Modulation	Bits/sym	Data - SR	MOD	FEC-T	FEC-K	DOCSIS OH	ETH TP	MOD-PRO#
Car-1	11.4	6.4	5.12	32	5	25.60	S-CDMA	4	232	0.8242	21.10	431
Car-2	17.8	6.4	5.12	64	6	30.72	S-CDMA	4	232	0.8236	25.30	432
Car-3	24.2	6.4	5.12	64	6	30.72	A-TDMA	12	232	0.8724	26.80	522
Car-4	30.6	6.4	5.12	128	7	35.84	A-TDMA	8	232	0.9040	32.40	523
Car-5	37.0	6.4	5.12	128	7	35.84	A-TDMA	12	232	0.8705	31.20	524
Car-6	43.4	6.4	5.12	256	8	40.96	A-TDMA	10	232	0.8887	36.40	525
Car-7	49.8	6.4	5.12	256	8	40.96	A-TDMA	10	232	0.8887	36.40	525
Car-8	56.2	6.4	5.12	256	8	40.96	A-TDMA	8	232	0.9058	37.10	526
Car-9	62.6	6.4	5.12	256	8	40.96	A-TDMA	8	232	0.9058	37.10	526
Car-10	69.0	6.4	5.12	256	8	40.96	A-TDMA	8	232	0.9058	37.10	526
Car-11	75.4	6.4	5.12	256	8	40.96	A-TDMA	8	232	0.9058	37.10	526
Car-12	81.8	6.4	5.12	256	8	40.96	A-TDMA	8	232	0.9058	37.10	526
						445.44					395.10	

Raw Data Rate

Ethernet Throughput

The results of Table 3 indicate a maximum of about 400 Mbps of *Ethernet* throughput under the packetized traffic conditions used. For an apples-to-apples perspective with the other scenarios herein, and staying consistent with the rest of the paper, we will recognize the transport data rate of 445 Mbps as the value for comparative analysis.

Capacity and Lifespan

Having roughly doubled the amount of available spectrum, but also proven 256-QAM capability and turned on S-CDMA, what kind of additional lifespan can we expect out of a fully utilized N-Split, or at least as much as it has been proven to support? Using Table 3 results along with the 5-42 MHz thresholds identified in Figure 8, we add the new capabilities of N-Split to the mix in Figures 12 and 13.

Beginning with Figure 12, the gap between the set of 5-42 MHz options and the maximized N-Split is readily apparent. Even for the case where S-CDMA is not turned on to exploit the lower half of the return band, which most MSOs have been hesitant to implement to date, the difference between 5-42 MHz and 5-85 MHz capability is noteworthy.

The gap identified by the red arrow shows the lifespan impact for the CAGR choice of 30%. This value is lower than what has been seen during past periods of high growth, and more aggressive than has been seen during other periods, including the past couple of years. It could be considered a reasonable average over the past 5-7 years, if not slightly on the aggressive side. If S-CDMA is fully leveraged in both cases (42 MHz and 85 MHz), then Figure 12 predicts over four additional years of growth is available. Without use of S-CDMA (dashed green for N-Split) for either case, that buyback expands to nearly 5.5 years. In both cases, the transition to N-Split pushes the lifespan of the return path architecture such that it supports a full decade of new growth – a very comfortable chunk of next generation network planning time.

Figure 13 postulates that, given the time frames we are discussing in Figure 12, it is likely and even planned in many cases that node and/or service group splitting will occur, allowing for an increase in average bandwidth per home for the same total capacity. Both cases – with and without a split, are shown for two threshold cases in Figure 13 – the 100 Mbps use of 5-42 MHz (no S-CDMA), and the N-Split, also with no S-CDMA.

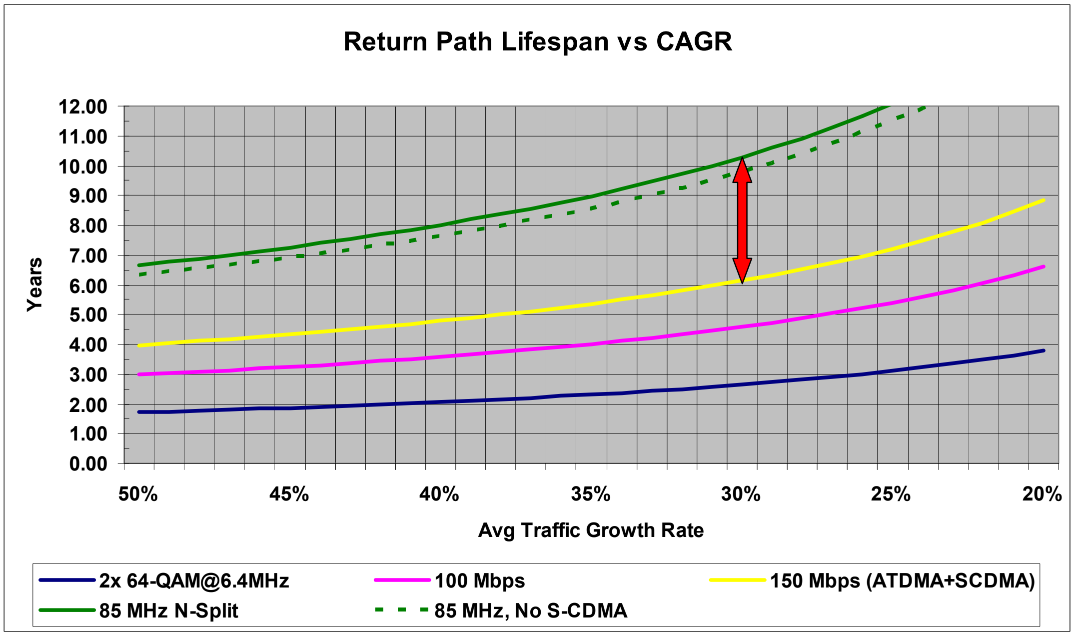


Figure 12 – N-Split Years of Growth vs. 5-42 MHz Use

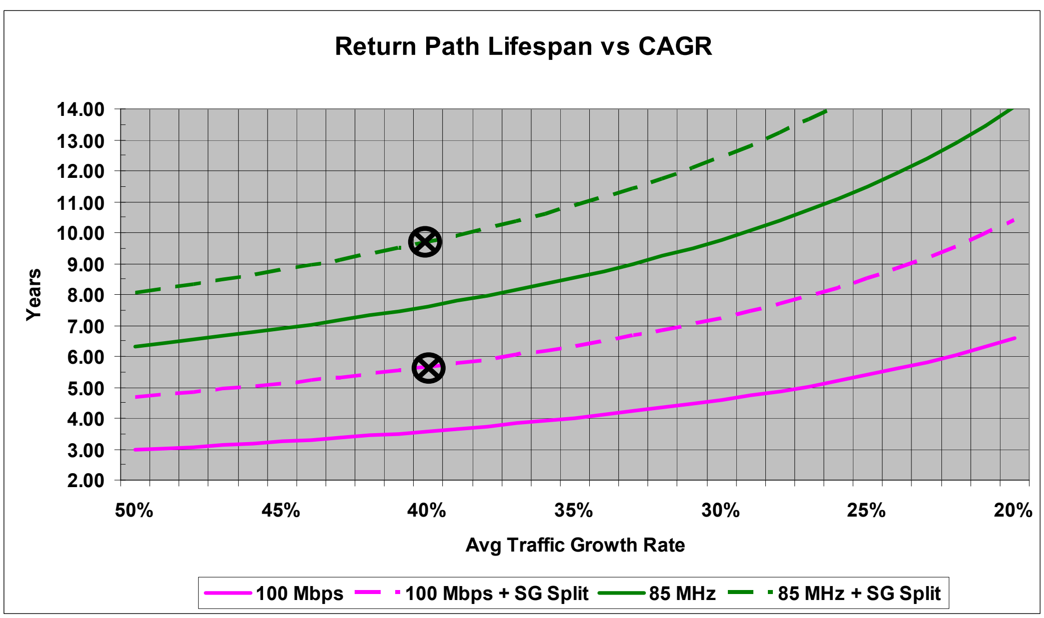


Figure 13 – N-Split Years of Growth vs. A-TDMA Only Plus Segmentation

Figure 12 showed the added lifespan for a CAGR of 30%. With a node split factored in within the next few years, as a prudent way to extend upstream support over 5-42 MHz, we can see in Figure 13 that the N-Split architecture now extends the life of the network to over 12 years of new growth at a 30% CAGR – beyond a reasonable period of planning the next implementation step given the speed at which technology change takes place. Because of this, when combined with a service group split, N-Split represents, for practical purposes of demand-based growth, a *long-term* solution.

Figure 13's bull-eyes highlight a more aggressive planning CAGR – in this case 40%. Again, under an N-Split upgrade, and without use of S-CDMA, almost a full decade of new growth is supported assuming a node split occurs. Obviously, with a second segmentation factored into the 10-yr span, more than a decade is then covered. Given the cyclical nature of plant investment and historical segmentation patterns, this may be simply a continuance of BAU bandwidth remediation. Alternatively, given the introduction of more cost-effective fiber architecture migrating into HFC, perhaps further HFC upgrades instead begin to give way to fiber-to-the-premises investment before the decade is out. An attractive feature of the N-Split is that it offers so much additional observation and planning time to consider the next phases of infrastructure investment.

The bull's-eyes themselves identify the basic mathematical relationship underlying compounding growth for a 40% CAGR. This value represents approximately a two-year traffic doubling period. Considering that we have used 100 Mbps to represent the A-TDMA only case within 5-42 MHz, and that the N-Split minus the S-CDMA contribution in Table 3 is almost 400 Mbps, we can back-of-the-envelope calculate the lifespan effect. The ~400 Mbps possible is 4x the 100 Mbps, so it is simply two traffic doubling periods. Since a doubling period at 40% CAGR is two years, two doubling period is four years. This is exactly the difference shown in Figure 13 as the lifespan bought moving from 5-42 MHz to N-Split at 40% CAGR.

Legacy CPE Challenges

The effects in the plant are easily understood for an N-Split migration. They are not necessarily operationally attractive, given that actives in the plant employ diplex filtering that must be changed to support 85 MHz. However, the steps to doing so are easily defined and represent no technological challenge. Instead, it represents primarily cost and logistics challenges. Managing the when and how in conjunction with fiber deep upgrades or node segmentations is an effective way to "kill two birds" and smooth the path for N-Split.

While plant migration is "blocking and tackling" steps, the implications within the home itself of modems transmitting in the new band is less well understood. Prior efforts have quantified the effect of 5-42 MHz modems on TV tuners when DOCSIS was first introduced [8], and recognized that, unsurprisingly, you could indeed overload a TV tuner even given an FDD architecture designed to keep forward and return apart from one another. With transmit frequencies now potentially in-band of tuners on CPE devices, it becomes yet more important to quantify the behavior of receiving STBs and TVs in order that new N-Split deployments do not disrupt current video services. As is the case in xDSL environments, it is anticipated that filtering in the home will be required in some cases to prevent degradation to video services. This would logically occur with a new modem deployment supporting N-Split, and would become a normal operation during such an install. However, the amount of filtering required is important to quantify, as well as the most sensitive frequencies for operational planning. An understanding of the nature of the observable impact for future customer support would also be valuable.

Figure 14 shows a block diagram of a test setup used to evaluate the sensitivity of different STB families and video loads to single and bonded channel transmit frequencies in the 42-85 MHz band.

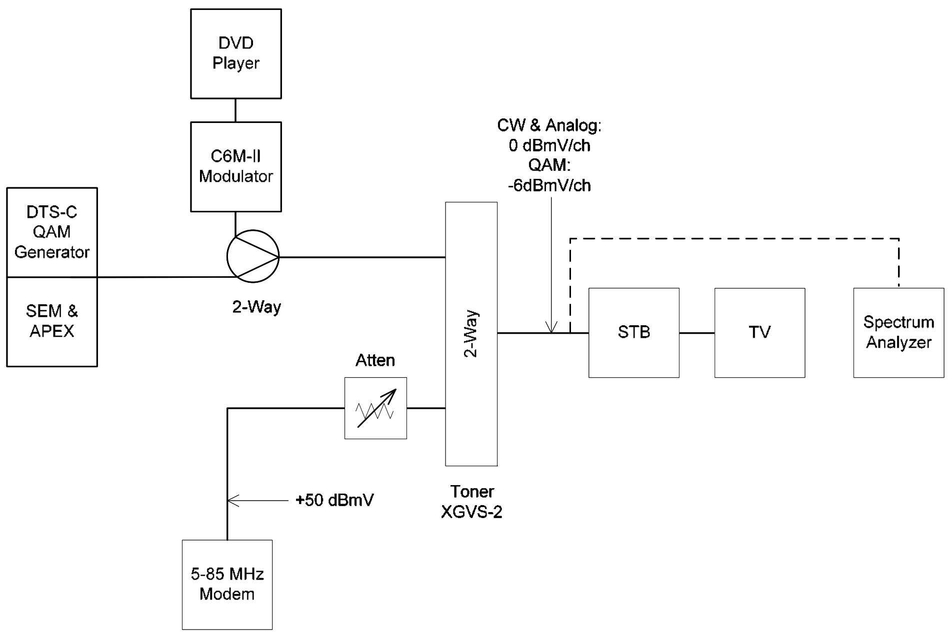


Figure 14 – CPE Testing for N-Split Home Environments

Figure 15 summarizes coarse results of the evaluation of video services that have analog, and those with only QAM. Based on observable pixilation or analog video picture distortion, it shows how much CM transmit power and what bands have the potential to cause interference concerns in the home, as a function of the RF isolation between the CM and the CPE. Of course, the amount of isolation is unpredictable in the home, and off-the-shelf retail splitters can have very poor port-to-port isolation. Figure 15 identifies (blue dashed lines) the range offered by splitters isolating CM and CPE of 10-30 dB, representative of a very poor or very good single splitter. Also identified is a typical range of CM transmit powers, up to the maximum allowed value (red dashed lines)

As to be expected, the STB is more robust against interfering signals when the channel load is QAM, allowing CPE signals as high as +30 dBc above the tuned video level in most cases. For analog services, the degradation threshold varied, but was on the order of 15 dB worse than in the QAM-only case. There is of course a frequency dependence between CM transmit frequency and CPE tuned frequency. There are many more permutations of this evaluation that will take place to comprehensively understand the impacts in the home, but these early results are illustrative of common tuner sensitivities. An important early conclusion, as can be seen in Figure 15, is that even under the most troublesome interference case of highest CM transmit power to most sensitive analog tuned input, the filter required to mitigate the interference requires a very reasonable 40 dB rejection value. This represents a relatively modest low pass filter design, capable of being easily manufactured at very low cost at these frequencies.

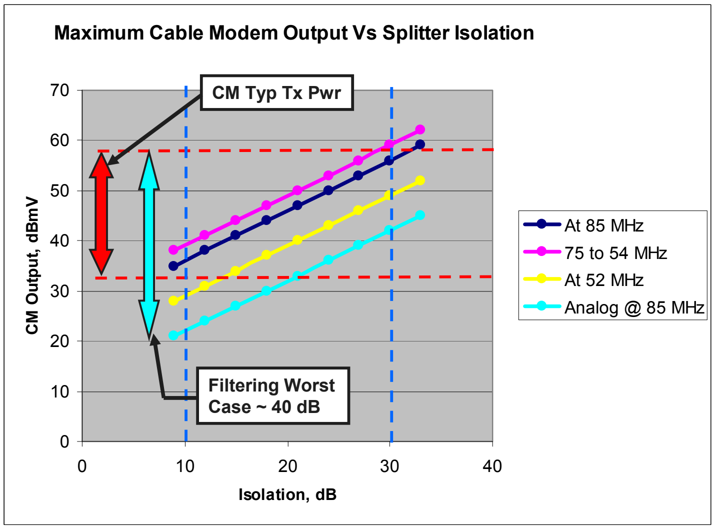


Figure 15 – CPE Testing for N-Split Home Environments

NON-DOCSIS SPECTRUM ENHANCEMENT

Gigabit Services

We discussed at length the ability of the N-Split architecture to support upstream traffic growth. We pointed out how, when coupled with BAU node splitting, how it should be considered a *long-term* traffic growth solution, not an upstream band-aid. The one feature that simply cannot be accomplished with the N-Split architecture is support of a full 1 Gbps of capacity, or naturally, of a Gbps tier rate. This is absolutely the case with DOCSIS use of the band (360 Mbps), and also the case even should theoretical capacity be achieved under DOCSIS SNR assumptions (650 Mbps). A theoretical 1 Gbps in the N-Split architecture alone would require a 38 dB return path SNR.

In Figure 11, we can see that with a new generation CMTS receiver and analog DFB optics, this 38 dB value is actually achieved over about 9 dB of dynamic range. In practice, however, that range essentially evaporates completely as theory gives way to actual implementation, and HFC variations eat away at what is left. However, it does point out that we are entering a new realm of possibilities on the return, where very high fidelity has never been the strong suit of the channel. Now, with 85 MHz of spectrum, modern HFC optics, and new CMTS receivers, many new dBs have been freed up that get us closer to capacity, and more importantly provide avenues that support new bits per second in the future.

Table 2 points out that a 1 Gbps threshold requires the split to move up to about the 200 MHz range. That bandwidth supports well over 1 Gbps of theoretical capacity, but under the reasonable assumption that DOCSIS remains in use in the 5-85 MHz band, the 85-200 MHz region can be exploited with something more aggressive that puts the total capacity above 1 Gbps. DOCSIS' maximum profile today (64-QAM@6.4 MHz) itself filling the band entirely out to 200 MHz falls short, although if 256-QAM were employed, even ignoring the low end of the legacy return, this would no longer be the case.

In the case of using split technologies (5-85 MHz of DOCSIS and 85-200 MHz of something else), a shortcoming that could come into play is the inability of that architecture to support a 1 Gbps *peak* service rate. It may be a complex endeavor in any case, but the complexity would be significantly magnified if any new technology developed to exploit 85-200 MHz needed to be integrated with the DOCSIS band to deliver 1 Gbps tier rates.

As was done in Figure 9, the additional bandwidth load must be quantified to understand exactly what a 200 MHz system might deliver. Figure 16 shows how the 200 MHz loaded performance compares to the 85 MHz and 42 MHz analysis previously discussed. The loss is again easily predictable, as simply the dB relationship among total bandwidths. And, similar to Figure 9, we can observe for this case that 10-11 dB of dynamic range for 256-QAM exists across the HFC optics – again an operationally practical amount of margin to accommodate alignment and plant behaviors.

Figure 17 is the analogous figure to Figure 11 for N-Split, showing, in this case, projected performance on a 200 MHz "high" split when factoring in an "equivalently performing" CMTS receiver (DOCSIS does not extend to 200 MHz). As would be expected, with the receiver performance equivalent to legacy CMTS receivers, inherently not equipped for 256-QAM, performance does not even breach the threshold. However, with a new generation of high fidelity receivers that achieves the noise performance of today's new cards, which already support 85 MHz, system analysis projects that there remains the full 10-11 dB of operational dynamic range to 256-QAM performance over a fully loaded 200 MHz return path.

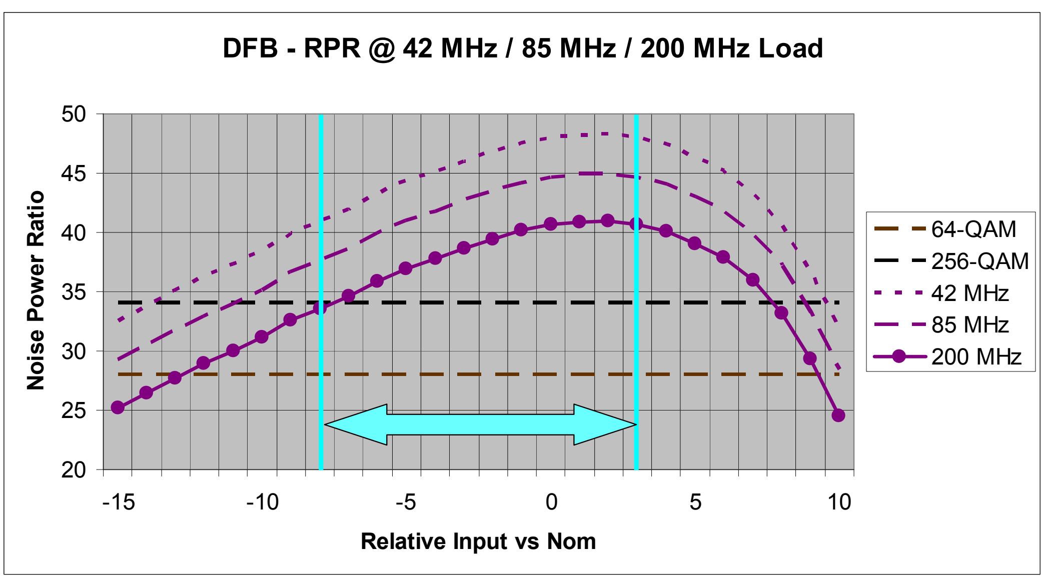


Figure 16 – Bandwidth Loading Effect, 200 MHz (2 mw DFB Link @ 20 km)

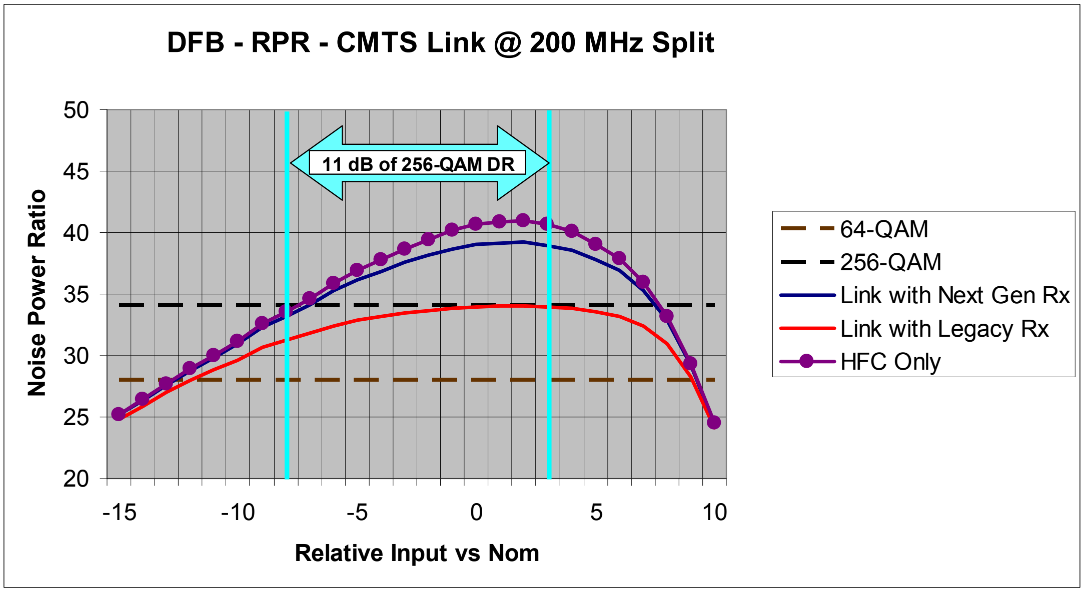


Figure 17 – Projected 256-QAM Dynamic Range Over 200 MHz Split

Alternative architectures exploiting coaxial bandwidth above 1 GHz have been around for many years. New iterations of these approaches could be leveraged to turn on currently unoccupied spectrum for adding upstream. The advantage of this approach is that in principle it does not interrupt legacy services, making a transitional path in theory non-intrusive to customers. Disadvantages include the need to work around legacy plant devices that are incapable of processing signals in this band, high losses at frequencies above 1 GHz, translating to significantly more power required from CPE devices, and new technology and operational hurdles that may arise in the new band. And, of course, the complete CPE itself becomes entirely new, or at a minimum requires the addition of block converters to support frequency translation.

On the passive network, coaxial cable and even some current 1 GHz taps are indeed capable of supporting useful bandwidth above 1 GHz [9]. However, the frequency dependence of cable loss quickly attenuates signals above 1 GHz when we consider it relative to the low band upstream. Given well-understood loss versus frequency relationships, for nominal trunk and drop spans of a passive segment, we can anticipate almost twice the loss (in dB) extending the return band to 200 MHz. However, above 1 GHz, the loss is increased by roughly a factor of five compared to legacy return for such a span. CPE devices must make up for that loss, and also must deliver additional total power associated with the wider bandwidth they would occupy to enable peak rates of a Gbps, relative to today's maximum of 6.4 MHz single channel power. Bonded channels as implemented today do not increase the transmit power.

Nonetheless, as technology continues to advance and HFC networks continue to get fiber deeper and RF shorter, the opportunities to exploit bandwidth above 1 GHz, whether for downstream or upstream, improves [9].

Capacity and Lifespan

So, what does all of this effort around making it to a Gbps of capacity and/or a Gbps of peak service rate gain for us relative to long term traffic growth? The answer to this question can be examined in Figure 18. It shows three threshold cases – 100 Mbps (A-TDMA only), N-Split (in this case, including use of S-CDMA), and 1 Gbps of capacity, however we manage to achieve it. Figure 18, as with any of the lifespan charts, does not care about implementation.

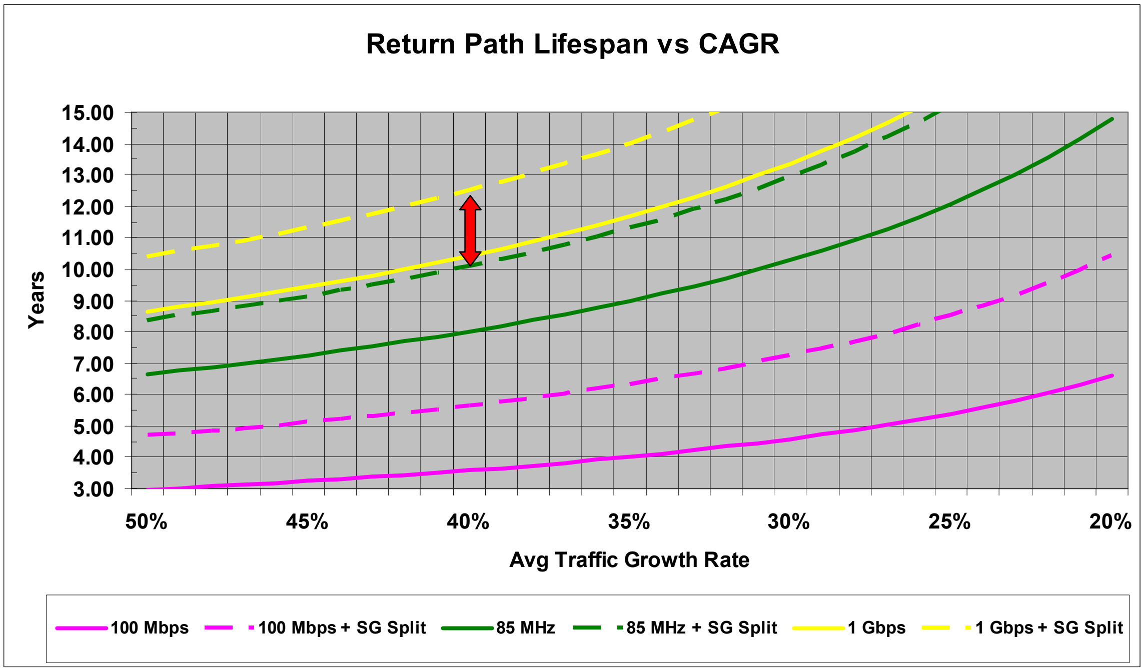


Figure 18 – Years of Growth: A-TDMA Only, N-Split, 200 MHz High Split

Zeroing in on the red arrow identifying the gap between N-Split and 1 Gbps at 40% CAGR, in each case with a node split assumed in the intervening years, we see that there exists about 2.5 years of additional growth. When we think of 1 Gbps, this intuitively seems odd. Why does migrating to N-Split buy a decade or more of traffic growth coverage, yet implementing a 1 Gbps system offers only a couple more years of survival on top of that decade? Unfortunately, this is simply how compounding works. It is up to our own judgment and historical experiences to consider how valid it is to be guided by the rules of CAGR and Nielsen, and if so what reasonable year-on-year (YOY) behavior assumption to assume. However, the mathematical facts of CAGR-based analysis are quite straightforward: with CAGR behavior, it takes many YOY periods to grow from, for example, 5 Mbps services today, consuming or engineered for perhaps tens of Mbps of average return capacity, up to the 445 Mbps we saw delivered by N-Split in Table 3. Once there, the subsequent annual steps sizes are now large.

As an example, if 20 Mbps of average capacity is needed to satisfy demand today, then traffic can double four times and not eclipse 445 Mbps. It eclipses 445 Mbps in the 5th traffic doubling period. For 40% CAGR (recall, two years doubling), that's a total of ten years. However, once the upstream capacity *has* consumed the N-Split, it is not very long via compounding before the 1 Gbps threshold is eclipsed. In fact, traffic can only double once and stay within this threshold. This is what Figure 18 is pointing out graphically. As such, relative to a solution that provides 1 Gbps, N-split gets us through 80% of that lifespan under the assumption of 40% CAGR and an intervening node split.

N-Split Plus

Based on analyzing the lifespan of the network to support traffic growth, the above result argues that N-Split is a reasonable next step that supports many, many years of new growth. It cannot achieve a Gbps of capacity or peak rate, at least as we can conceive of its use over HFC under reasonable system assumptions. However, it is not clear that Gbps service tiers will be required for residential upstream in a time frame reasonable for business planning purposes, certainly from a consumption demand perspective. Given the upstream lifespan offered by N-Split, evaluation can continue around what should follow, with technology changes in the intervening years likely to have a strong impact on the direction to head so many years down the road.

Nonetheless, while developing N-Split architectures, it seems prudent to enable a subsequent HFC step that can be implemented in a way that does not require another round of equipment visits to the physical plant. This can be used should Gbps capacity become critical as a marketing tool or for suddenly accelerated demand over a longer than normal period of time (i.e. a prolonged acceleration of CAGR). It is then also in place if it turns out that the best longer term evolution answer is that the HFC network lives on further, without a need to turnover technology such as to an FTTP architecture. If this is the case, and the decade plus of N-Split becomes exhausted, a natural "phase 2" RF step for upstream would be to increase the split once more. When considered as a possible phased solution, such an architecture would have selectable diplex options as part of the migration to N-Split. Multiple options could be made available under the control of the operator.

Another reason to consider a phased approach is for the practical reason of the legacy OOB downstream STB signals. These are tunable only to 130 MHz, so an upstream that engulfs this band obsoletes many existing STB. However, over the period of time that is bought by N-Split, these devices can be managed down, either through natural attrition or an active effort to reduce reliance on legacy OOB (vs DSG, for example), or via a more wholesale transition to a new generation of home gateways and/or DOCSIS-based IP video architecture.

Should 1 Gbps become part of the playbook for residential upstream, the approach to delivering this by extending the return band on the low side has many technical advantages over an approach using above 1 GHz. We have mentioned some of them already:

- Much better loss properties supporting more cost effective CPE
- No technology or plant hurdles such as housing, connectorization, bypassing
- Less guard band lost spectrum – no "triplex" guard band
- No bookending of downstream

Forward Relief

Moving to N-Split adds 43 MHz of return bandwidth, but does so at the expense of forward bandwidth. When factoring in the new guard band, possibly nine or ten 6 MHz slots in the traditional analog band are eliminated. Mathematically, converting these channels to digital allows them to all fit into one slot. As such, as analog reclamation continues, this forward loss does not represent a significant capacity concern. The primary operational concern is that the nature of the channels in this region often represent a basic service tier, and therefore cannot simply be transitioned into the digital tier and off of the analog tier, as perhaps some of the longer tail of the analog service could. Instead, some channel re-mapping and/or more aggressive deployment of digital adaptors would be required. In any case, given the powerful set of tools available to provide downstream capacity, 85 MHz presents no significant imposition on the forward bandwidth in terms of capacity loss.

In the case of a 200 MHz extension, however, this is not necessarily the case. This can easily shown to be so, for example, in 750 MHz systems that are trying to accommodate aggressive CAGR, such as to support OTT growth, where extensive analog exists and may continue. The issue is magnified further when considering new trick play video services and alternate screens that result in more unicast delivery, and when considering the addition of more HD, 3D, and possibly even higher resolution services.

As previously indicated, in the case of 1 GHz, there is significant "free" bandwidth available above the specified 1 GHz value. Figure 19 shows the frequency response on the "through" port of a particular 1 GHz tap – the port that would be in series with other taps on the way to a connected home. The response on the tapped port also has essentially parasitic, low-loss properties over the first 200 MHz above 1 GHz. Though not as perfectly flat, it creates no significant burden to RF signals in the band, and in particular when considering a new generation of modem technology, such as multi-carrier [9]. The same is the case for some families of 750 MHz taps (available bandwidth exists above 750 MHz) and 870 MHz taps (available bandwidth exists above 870 MHz).

The amount of useful bandwidth and loss properties are vendor dependent, but MSOs already often use slots above these limits. Conveniently, as Figure 19 shows, the amount of available new bandwidth simply trickling over the top of the band is virtually the same the amount of bandwidth that would be removed from the forward by 200 MHz systems, when considering guard bands for that extension. Note, however, that there is no forward/reverse guard band involved here without an upstream system contending for spectrum. Also, this "replacement" bandwidth amount provides adequate spectrum to facilitate downstream Gbps services. The ability to fully exploit this bandwidth in the passive plant obviously depends heavily on the band coverage of the actives themselves and the depth of the cascade. Clearly, this is where shortening cascades and "N+small" continue to payoff for HFC evolution.

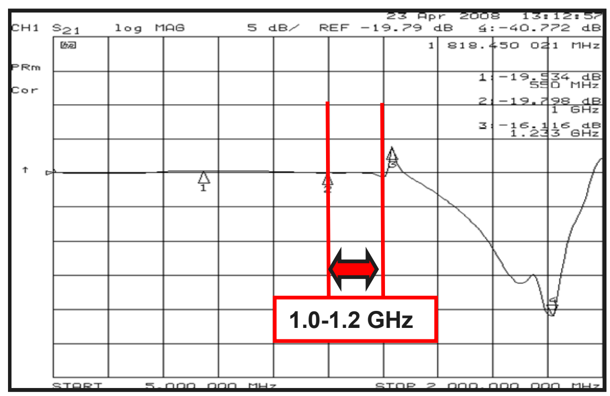


Figure 19 - Above 1 GHz Frequency Response of the Thru Port of a 1 GHz Tap

There are three other compelling advantages to considering use of the band over the end of the defined tap bandwidth for forward services, as opposed to reverse. First, the forward path is inherently designed for high fidelity in support of analog video. As we know, the reverse path was never architected with high fidelity in mind. Over time, technology has been introduced to enable a high-speed data channel, but the low noise and high linearity architected into the forward path is orders of magnitude above the return path. This difference translates to a much more straightforward exploitation of bandwidth with high performance on the downstream.

Second, the forward path levels are designed for RF path losses out to 1 GHz. Because of this, the parasitic losses above 1 GHz of the coax, and the minimal additional attenuation, are not a stretch to achieve when extending the forward path. It is an entirely different case in the return, where the architecture has relied on the low loss end of the band, which increases only modestly as it is extended to 85 MHz or even 200 MHz.

The third point, related to the first two, addresses the issue of CPE transmit power to overcome these high losses. Forward path RF systems, being design for similar losses, have seen investment in broadband RF hybrids drive higher and higher levels over increasing forward bandwidths, still based on supporting a full analog and digital multiplex. As a result, the output levels of these hybrids and nonlinear characteristics have continued to improve. However, investment in these premium devices for the forward path is spread over the number of homes serviced by the actives. More broadly, investment in the number of premium RF hybrids from the node to the final amplifier is shared by the number of homes passed by the node. In the reverse path, each home needs a high power, linear transmitter (though less than an octave) in this higher frequency band.

Figure 20 quantifies what is ideally available exploiting the frequency response above 1 GHz for forward band purposes based on the passive segment only, or representing effectively and N+0 situation.

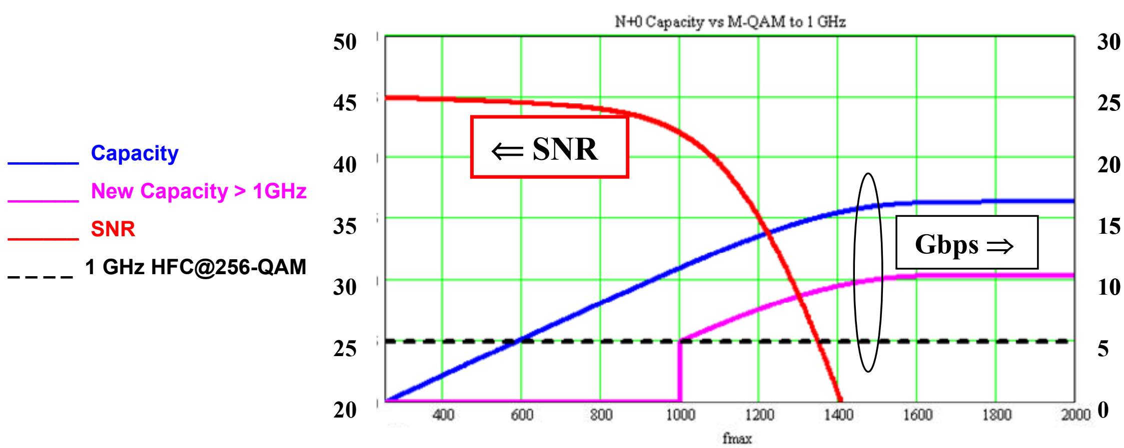


Figure 20 – Gbps: Current Use, Ideal Use, and Additional New Above 1 GHz

In Figure 20, a full forward band throughput of 256-QAM is shown, along with the theoretical capacity in Gbps (blue, right vertical axis), for a given maximum upper edge of the band shown on the x-axis. These capacities are shown along with the SNR vs. frequency delivered from a 5-tap cascade made up of taps such as that shown in Figure 19, and one coupled port from the same. The final trace (pink) recognizes the 256-QAM legacy spectrum as a given, and above that identifies new theoretical capacity potentially that can be exploited above 1 GHz in the passive segment as a function of the maximum upper frequency used. Clearly, within the first 200 MHz above 1 GHz, a Gbps of throughput can be extracted. Also apparent is how much latent capacity still exists as the cascades shrink and open up new RF bandwidth potential, considering that 256-QAM is today's maximum modulation profile. Many recent analysis have proposed use of 1024-QAM and perhaps even higher order modulations [10][11]. There are plenty of available bits per second left to be exploited in the passive infrastructure, above and beyond the current use of the bandwidth to use as "replacement" capacity should a phase 2 migration above N-Split be required.

CONCLUSION

Operators are dealing with the inevitable charge forward of traffic demand. With historical trends as a guideline, the traffic increase can be quantified and used to develop timelines and strategies for dealing with the growth. The first step of optimizing the existing 5-42 MHz is underway, with multiple operators moving ahead with DOCSIS 3.0 capabilities, including its most bandwidth-efficient modulation profile, 64-QAM. The additional tool for extracting everything possible in 5-42 MHz is S-CDMA. While it has yet to be embraced fully in North America, it is a field proven, powerful technique to make productive use of currently vacant spectrum.

Despite these 5-42 MHz optimizations, the upstream is ultimately limited by its hard cap in total bandwidth under today's diplex architecture, which highly favors downstream. Using CAGR analysis, we can estimate when this obstacle needs to be removed. A straightforward and relatively *long-term* solution, providing a decade or more of potential growth, is the 5-85 MHz N-Split architecture already called out in DOCSIS 3.0. Equipment is available now that supports N-Split, and more will become available in the very near future. Furthermore, the current generation of HFC optics using analog DFB returns is *already* capable of supporting the added bandwidth. Better yet, the new bandwidth is exceptionally clean. This fact, combined with the HFC link performance and a new generation of high fidelity CMTS receivers, makes 256-QAM usage practical with solid operating margins in the upstream. This, too, has been proven in the field using existing hardware. When applied to CAGR analysis, the results show that N-Split can capably support a decade or more of new upstream demand-based consumption. This extended timeframe positions the MSO well, offering a lengthy opportunity to evaluate technology shifts over the next ten years and plan next generation architecture steps. These benefits are derived through a low-risk transition to N-Split.

Though a long-term solution has been identified, should the next steps beyond turn out to involve simply many more years of HFC-flavored evolution, support for continued upstream traffic growth beyond the above time frame can continue by further plant segmentation or implementing a second phase of diplex adjustment that extends the band to 200 MHz or beyond, should that be necessary for larger peak service rates. If developed in a forward-looking way, new phases of diplex shifting can be done with minimal plant disruption and operational complexity. As we have shown, this bandwidth extension *also* should be capably supported, and with higher order modulation profiles such as 256-QAM, using today's generation of HFC optics and receivers on par with today's generation of low noise CMTS'. Of course, these complementary tools – new bandwidth and splitting of nodes – can and should both be considered.

Finally, if necessary, new "replacement" forward bandwidth may become easily accessible above the top end of today's forward band if the additional return imposes on downstream growth. Above the top end of the forward band, it is much simpler and more bandwidth efficient to create additional *forward* capacity than to try and push upstream signals against their architectural will in the face of many obstacles.

HFC has a long and impressive history of technology and architecture evolution, and of new services. It also has a long and undoubtedly impressive future ahead of it, capable of much more capacity exploitation, and full of potential for many new exciting services that maintain and delight customers. It is hoped that this analysis is found useful as a guideline for planning a migration strategy that fully realizes this latent potential in today's HFC networks.

ACKNOWLEDGEMENTS

The authors would like to thank Jack Moran, Robert Thompson, Charles Moore, Chris Chiu, and Jim Weineck for their invaluable contributions to many of the measured results shown in this paper.

REFERENCES

[1] Finkelstein, Jeff, "Upstream Bandwidth Futures," 2010 SCTE Cable-Tec Expo, New Orleans, LA, Oct. 20-22.

[2] Howald, Dr. Robert L., Phillip Chang, Robert Thompson, Charles Moore, Dean Stoneback, and Vipul Rathod, "Characterizing and Aligning the HFC Return Path for Successful DOCSIS 3.0 Rollouts," 2009 SCTE Cable-Tec Expo, Denver, CO, Oct 28-30.

[3] Ulm, John, Jack Moran, Daniel Howard, "Leveraging S-CDMA for Cost Efficient Upstream Capacity," SCTE Conference on Emerging Technologies, Washington DC, April 2, 2009.

[4] Howald, Dr. Robert L., "Maximizing the Upstream: The Power of S-CDMA" Communication Technology Webcast, Sept. 9, 2009.

[5] Howald, Dr. Robert L. and Michael Aviles, "Noise Power Ratio the Analytical Way," 2000 NCTA Show, New Orleans, LA.

[6] Miguelez, Phil, and Dr. Robert Howald, "Digital Due Diligence for the Upstream Toolbox," 2011 Cable Show, Chicago, IL, June 14-16.

[7] Robuck, Mike, "Cox, Motorola lay claim to new return path speed record," CedMagazine.com, March 01, 2011.

[8] Stoneback, Dean, Dr. Robert L. Howald, Joseph Glaab, Matt Waight, "Cable Modems in the Home Environment," 1998 NCTA Show, Atlanta, Ga.

[9] Dr. Robert L.,"Fueling the Coaxial Last Mile," SCTE Conference on Emerging Technologies, Washington DC, April 2, 2009.

[10] Dr. Robert L.**,** Michael Aviles, and Amarildo Vieira, "New Megabits, Same Megahertz: Plant Evolution Dividends," 2009 Cable Show, Washington, DC, March 30-April 1.

[11] Hasse, Philipp, Dirk Jaeger and Joerg Robert, "DVB-C2 – The Second Generation Technology for Broadband Cable," 2009 Cable Show, Washington, DC, March 30-April 1.

NEXT GENERATION ACCESS ARCHITECTURE OPTIONS FOR ADDITIONAL UPSTREAM AND DOWNSTREAM TRANSMISSION ON HFC

Jorge D. Salinger
VP, Access Architecture
Comcast Cable

NOTE: The concepts and proposals presented in this paper are for discussion purposes only and do not reflect actual plans from Comcast. Similarly, all examples presented are only provided for illustrative purposes.

Abstract

HFC networks currently have an US capacity limited by the available spectrum. Typically, North American networks are limited to the 20-30 MHz of spectrum these systems have available between 5 and 30 or 42 MHz, and European systems have an additional 25 to 45 MHz of available spectrum. At 6.4 MHz of channelization and QAM64 modulation, assuming complete use of the available spectrum, this capacity allows transmissions of 100-300 Mbps.

The potential migration of North American systems from the current 42 MHz-split to 65 MHz, 85 MHz, or even higher split systems, and of European systems from 65 or 85 MHz to higher splits, are complex to implement, service impacting and financially and operationally onerous. The capacity gain from such migration is limited to approximately doubling, or perhaps tripling, the maximum transmission capacity. But, an indirect impact of performing such spectrum allocation change is a reduction of the downstream capacity of the HFC systems, which is larger as the split is moved further up into the downstream spectrum used today.

Taking into account other factors may result in concluding that other solutions may be more appropriate and no more operationally impacting than just changing the spectrum split to increase upstream transmission.

Such other factors may include:

- *Use of a different portion of the spectrum for US transmission to simplify the migration and avoid impacting currently available HFC capacity.*
- *Enable additional DS transmission capacity, not just additional US capacity, which will have the additional benefit of alleviating the expected peak capacity needed when new services (additional programs, additional technologies such as IP video simulcast, etc.) are added to an already fully allocated HFC network.*
- *Establish a new, more efficient HFC transmission mechanism for coupling with home gateway devices, which would be based on existing technologies, such as those being considered for advanced video service planned by most MSOs.*
- *Solve certain limitations of the current transmission mechanisms, such as enabling more native support for IP-based, enhanced business, wireless and other services and applications.*

When considering the above factors in addition to increasing US transmission capacity, alternate solutions to just changing the US/DS spectrum split may become more attractive and valuable to MSOs because the burden of such implementation could be attributed to the deployment of services not otherwise possible, with the resulting potential increase in both ARPU and RGUs.

This paper will then presents an analysis of various such options, and expands on one particular alternative that offers the following characteristics:

- *Leaves all current services untouched in the currently allocated HFC spectrum.*
- *Unleashes new spectrum in the HFC network.*
- *Could enable transmission of up to an additional 3 Gbps combined between upstream and downstream.*

To do so, the proposed approach makes use of traditional and already available technology and network strategies in a new way, such as:

- *Continue pushing fiber deeper into the HFC network, up to the current last active, without changing the current HFC architecture, such as not requiring cascade reductions.*
- *Use of existing baseband fiber-based technologies, such as PON, and EPON in particular, and expansion in the use of WDM, for transmission from the headend to the current location of the last active,*
- *Implementation of existing technologies for higher-order PHY modulation and encoding and light-weight MAC for transmission from the last active to the home gateway using silicon devices already under development.*
- *Superimpose the new transmission method in a portion of the spectrum above the band currently being used, such as above 1GHz, and still utilize spectrum allowing the use of passive devices such as below 1.8 GHz.*
- *Use of a new, simpler home gateway to terminate the HFC network in the home, bridging the new transmitted capacity described above, and using consumer network technologies inside the home.*

The proposed approach described above leaves existing services untouched, enables considerable new capacity in the HFC network (almost doubling the overall current HFC capacity), solves the current US capacity limitation, leverages existing technologies, builds upon the current MSO strategy of segmenting the HFC network to increase overall capacity, and should be neither more costly for initial deployment nor operationally more impacting than simply changing the HFC US/DS spectrum split to provide a marginal increase of US transmission capacity.

TYPICAL HFC NETWORKS

Most MSO's HFC networks have been designed to either 750 or 860 MHz of spectrum capacity. If not fully utilized, it is expected that use of their capacity will be increased to the point of exhaustion as the use of DOCSIS increases for high HSD service tiers, additional HD programs for both broadcast and especially narrowcast services such as VOD and SDV are deployed, or new services such as network-based DVR are added.Proportionally few HFC networks have been deployed to operate up to 1 GHz, although all equipment available today can support the use of spectrum up to 1 GHz and even 3 GHz for some components.

In recent years the growth in, and demand for, HD programming has resulted in the need for allocation of large numbers of EIA channels for HD services, both for BC and NC, which has filled every available portion of the spectrum. This is especially true for BC, where large numbers of programs are offered in HD format, while simultaneously the need for distributing the SD version has persisted. This has resulted in the need for use of 3x to

5x the number of EIA channels than previously required. For example, a typical digital multiplex including 10 to 15 programs would require an additional 3 to 5 EIA channels for the HD equivalent streams, even assuming the newer, more sophisticated multiplexing schemes available in the market. Of course not every program is available, or still sought by subscribers, in HD format. But very large numbers of them are, including 100 to 150 BC programs.

The above is also applicable to a great extent in systems utilizing SDV technology for distribution of its content. The difference is that the SD version of the program is not distributed unless a subscriber is requesting it, which reduces the marginal increase in capacity. Assuming that all programs are distributed in only one format, which is certainly a valid expectation for programs of low viewership, then the increase in capacity for a conversion from SD to HD would just be the increase in capacity required for the transmission of the HD program without requiring the simultaneous use of bandwidth for both formats.

Additionally, considerable spectrum is needed to deploy high-capacity narrowcast legacy videiservices, especially nDVR, and a full-array of HD video-on-demand services. For the former, initial observations suggest that network requirements for nDVR may be as high as 4x to 5x that of VOD, and that peak utilization overlaps, at least partially, with that of peak use for other narrowcast services.

Finally, the growth in HSD services shows no sign of letting up. As shown in Figure 1, network operators have responded to competition by increasingspeed tier offerings, which over the long term amounts to year-over-year compounded growth of 40% to 60%, as observed for well over a decade now. The applications have changed, but the demand has continued to increase. One thing MSOs have learned, as have network operators of more traditional Internet network services, is that it is very difficult to predict what subscribers will use the additional capacity for, but that the bandwidth growth is somewhat predictable.

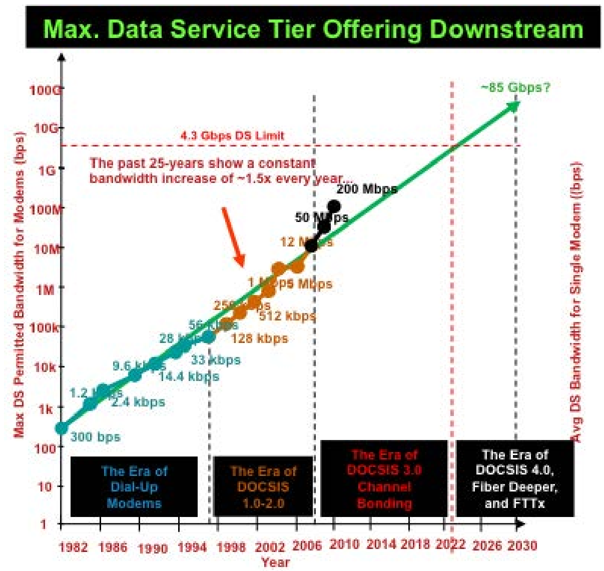


Figure 1: HSD Service Tier Trends[i]

In fact, such increase in demand for HSD capacity shows no evidence of decreasing. Quite to the contrary, all the evidence gathered over many years indicates that the growth will continue, and perhaps even increase when considering the observed increase in the consumption of video-based content. When projecting this experienced growth into the future, as also shown in Figure 1, it is easy to see that should the usage growth pattern continue networks will be required to provide services in the range approximating 1 Gbps within the next few years.

To support this growth, MSOs have deployed, or are considering deployment of, bandwidth reclamation tools such as SDV for digital broadcast, DTAs for analog services, or a combination of both. These tools have been extremely valuable to MSOs, which have seen their operational complexity and cost to be well justified.

In the case of SDV, early predictions several years back from industry analysts projected that the efficiency of SDV would reach 40% (e.g., programs requiring 10 EIA channels could be carried in 6). This has proven to be understated, since it was based on the use of SDV for reduction in bandwidth required for existing services. As SDV's role in the network grew, the efficiencies have been even greater, especially as SDV has been used to introduce niche services that have low viewership and would have otherwise been difficult to deploy.

The benefit of DTAs has been just as, or perhaps even more, striking. MSOs deploying DTA devices are able to eliminate the need to distribute the analog channels in the network. Even if DTAs are distributed to top analog tier customers, such as only to subscribers of the traditional expanded basic subscribers, such deployment would reduce a channel line up from perhaps 50 EIA channels dedicated to 50 analog programs to perhaps as little as 4 EIA channels dedicated to transport the 50 programs in their equivalent digital transport. Using the same comparison method as the above SDV case, this is a >90% efficiency. If extended to the entire analog tier the efficiency gains are very significant.

Despite the availability of these tools, they are not universally applicable. With respect to SDV, in general it is not likely that all broadcast programs will be switched since experience shows that many broadcast programs are constantly viewed by someone in the service group during peak hours, which will leave a large portion of the spectrum still used for broadcast. Similarly, not all analog channels can be removed in the short term due to operational and/or cost constraints. Additionally, while many MSOs will use one or both tools, in general these tools won't be used by every MSO for all applications.

Finally, there are also significant potential gains to be achieved from the use of AVCs and VBR. In the case of AVCs, coding efficiencies of approximately 50%, depending on implementation and content type, can be obtain with H.264[ii] and/or MPEG-4 Part 10[iii]. And the use of VBR could result in a capacity efficiency gain of as much as 70% versus CBR[iv].The combined gains from using both approaches could be very significant. However these are difficult tools to take advantage on the network since proportionally relatively few set-tops still support AVCs and VBR, especially the latter tool. However, because of the lack of widespread support across set-top populations, MSOs are only able to use these tools for unicast-based services, such as VOD, for the small minority of the set-tops that support the tools. These tools will likely enjoy significant support in newer, IP-video based services equipment moving forward.

NEW TREND: IP-BASED ADVANCED VIDEO SERVICES

Industry-wide, MSOs are now considering deployment of video services supporting IP devices. For this, 3 approaches are possible:

a. Reuse of existing legacy content by re-encapsulating at the edge,

b. Use of the existing legacy infrastructure to distribute IP-based content, and

c. Leverage the DOCSIS infrastructure to distribute IP-based content.

The first approach makes very efficient use of existing spectrum for a new application, but has significant draw-backs, such as: limits content to that currently distributed, need for CAS termination in a gateway device, complex and seemingly expensive gateway, proprietary implementation, etc.

The second requires the development of special-purpose software in the gateway, inability to benefit from certain network efficient DOCSIS features such as channel bonding, lack of standardization across potential suppliers, etc.

With the advent of lower cost DOCSIS CMTS gear in recent years and especially moving forward, the third has become the one garnering the most attention. It offers significant network efficiency factors that can't be simultaneously achieved with the other two methods, such as immediately benefiting from VBR gains, immediate use on all applicable clients of AVCs, multiplexing gains from channel bonding, standardization of the gateway implementation, benefit from multi-industry approaches for web-based video distribution, multicast and caching, and a very broad availability of development talent and supplier ecosystem.

When considering the options and their trade-offs, all of the above is believed to make the third approach, that of a DOCSIS-based gateway, significantly more efficient and long-term beneficial than that of the other two approaches.

But, this approach will require additional capacity on the network. This is especially true when considering that the deployment of these advanced video services will result in an additional simulcast of video programs, at least initially, which is expected since its deployment will not at least initially replace the currently deployed services. Furthermore, ubiquitous support for such devices would require considerable spectrum if the legacy services are maintained for an extended period, as it is expected since legacy devices are and will continue to be deployed. This increase in simultaneous use of advanced video services while maintaining legacy services will be especially impacting over time as its penetration increases.

Initial target for most MSOs seems to be 2nd screen devices, such as PCs, tablets, game consoles and non-traditional customer-owned devices. This target subscriber use and device base likely only require low-resolution video services (e.g., 1.5 to 3 Mbps streams), possibly expanding to higher stream data rates for TV-attached devices and larger screens.

It is likely that if the deployment of these advanced video services is successful, a possible additional target would be to deploy IP-based set-tops as part of the mainstream MSO device profile, for which higher bandwidth full-resolution services would likely be used (e.g., 9 Mbps and higher resolution streams).

ALL-IP VIDEO CAPACITY EXAMPLE

What follows is an example of capacity required in a network to support an extensive deployment of advanced video services.

Assume the following scenario and circumstances:

- Typical service group of 400 HHP, which is an accepted average across the industry.
- 50% penetration of IP-based devices, which is typical in legacy video today.
- 2.5 devices per home, which is a high average for legacy video, but considered likely with IP-devices.
- 70% of homes active at peak, which is a typical industry average for video services
- 70% of secondary devices are active at peak, which is also a typical average
- Assume all-unicast video at 6 Mbps. This assumption is based on a number of factors, including: initial deployment of IP video will likely not benefit from multicast; over time it is expected that video viewing will trend towards more individualized viewing; rate could be

considered a high average, but video quality is constantly being enhanced requiring additional bandwidth; the use of ABR would cause the simultaneous transmission of content at multiple rates for different subscribers; etc.

Using the above assumptions, peak capacity could be established by the following formula:

400 * 0.5 * (1 + 1.5 * 0.7) * 0.7 * 6 Mbps

The above would yield a total capacity of over 1.7 Gbps.

Assuming an additional capacity of 1 Gbps for HSD, the total capacity required would be over 2.5 Gbps, or an equivalent of over 70 DOCSIS channels.

When all services and all devices in the network are based in IP transport, a 750 MHz plant should suffice to support such services. In fact, a 550 MHz plant would suffice as well.

And, even a higher US capacity could be supported. A 200MHz split could yield >1 Gbps US, which could be accommodated in a 750 MHz system. In such case the remainder of the spectrum implementing 10 b/Hz yields >5 Gbps. Furthermore, implementing any change in split would likely be accompanied by an expansion of the spectrum to 1 GHz, yielding significantly more DS capacity than would seemingly be required.

SIMULTANEOUS IP AND LEGACY

At issuemay not be whether the existing network would support a set of all-IP services, but rather that, if implemented, a transition to an all-IP set of services will likely take considerable time, and during such transition it would be necessary that both advanced video services and current legacy services be provided simultaneously.

For financial and operational reasons, the migration to all-IP services will likely take a long time to be completed. The reasons for this are the same as always: technologies take long to be developed and longer to be adopted;operational readiness is a long road for any new approach, especially in replacement cases; amortization of equipment dictates a need for preserving the investment; and even when equipment is amortized it is by no means trivial to fund a replacement.

Moreover, as support for advanced video services and IP-based devices is deployed and grows over several years, additional legacy devices will continue to be deployed. For example, viewership of HD VOD will increase, requirements for HD services in SDV increase, and nDRV with its expected high usage may be accompanied by a high use of content in HD, all of which will expand the need for spectrum for legacy services,which must be preserved and expanded.

Therefore, both legacy and IP-based services would need to be supported simultaneously. In that case, which appears likely, the deployment of IP-based services will occur while supporting a full array of legacy services, which is still growing.
For that, more HFC capacity than is used today would be needed. Current spectrum utilization for legacy services will likely not decrease, but rather need to increase as described above. In parallel, spectrum would need to be allocated for IP-based services, which is likely to grow over time while the capacity allocated to legacy services has not yet decreased.

ADDITIONAL SERVICE TARGETS

In addition to the likely need to simultaneously support existing legacy services, for which capacity needs continue to expand, and the deployment and parallel growth of advanced video services to IP-based devices, it might make sense to consider additional service targets.

For example, a list of desirable targets could include the following:

- Native support for TDM and IP wireless services, which are currently supported with fiber-based technologies, mainly for now via MetroE using mostly dedicated fiber for the time being, and migrating over time, at least in some cases, to PON technologies.

- Higher-capacity IP-based commercial services than those available via DOCSIS cable modems, which are normally provided via dedicated fiber-based network services (again, MetroE).

- As described in the above sections, an expansion of<u>both</u>the upstream and downstream capacity, not just the upstream.

- Continue to simplify operations and increase reliability as all cable-based services have become primary and fundamental for subscribers. For this, fiber deeper into the HFC network has been a goal and driver for quite some time.

LIKELY DESIRABLE PREFERENCES

If industry operators and vendors could just freely do what they thought would be best, especially irrespective of financial considerations, what might they decide to do?

At very first blush one might argue that the most desirable path would be to expand network capacity without any cost or any change in the plant. While the first one could possibly done (i.e., no cost), the second is realistically unlikely (i.e., expand capacity without any change).

So, assuming that some change needs to be made, but still the financial aspects of doing so could be disregarded for the time being, then the most logical outcomes would be:

- Unleash more capacity in the HFC network to expand its use further

- Continue the current path towards smaller service groups since fiber deeper increases capacity and reliability and is already business-as-usual for MSOs

- Leave all current equipment and services unaffected, including leaving STBs in place for the services these provide until a natural/organic transition takes place, and not forcing the removal of analog and/or broadcast channels in favor of deploying DTAs or SDV where not already planned, etc.

- Develop technology that could be deployed on a success basis, incrementally, minimizing cutover changes, and deployed to the extent needed and justified.

- Assign any new DS and/or US spectrum in a flexible way, such as: starting with some spectrum and grow over time as needed, enable future change even as new infrastructure is deployed, change DS:US ratio as needed without service interruption, etc.

OPTIONS BEING CONSIDERED

Let us review the 3 categories of options being considered throughout the industry, and evaluate how each one fulfills the above desirable targets.In the process, let us review the key implementation aspects of each option, leaving for another opportunitythe details of theoptions and on how these could be deployed.

1. Increase US capacity by moving the US/DS split to a higher portion of the spectrum and simultaneously expand DS to 1 GHz

   From an equipment perspective, this option is generally readily available to MSOs. From a network perspective, this option involves the change of the diplexers throughout the network such that the frequency division crossover is moved from the 42-50 MHz up to a higher portion of the spectrum, plus the simultaneous expansion of the network capacity to 1 GHz via a retrofit of the active components with minimal changes to the plant spacing and passive components.

   However, from an operational perspective, this option requires perhaps the most operational change to existing services, such as the removal of analog channels in that portion of the spectrum. That may not be possible for many MSOs that are either required to maintain support for analog TVs directly (e.g., without DTAs), or are unable to remove the analog channel for contractual reasons, or some combination of the above two reasons.

   Even if removing the analog channels is possible, this option seems to require the installation of CPE filters in most or perhaps all home CPE devices (e.g., TVs, VCRs, etc.) to both protect that portion of the spectrum from emissions from such home devices and to protect the devices themselves from the levels of transmission of the new CPE that would use that portion of the spectrum for transmission.

   And, even if removing the analog channels and deploying the necessary filters were possible, this solution alone provides limited additional US capacity in the network, as follows:

   - A move of the split to 65 MHz provides an additional capacity of just 15 MHz, which less than doubles the current capacity. By all accounts, this is a change not worth embarking on.

   - A move of the split to 85 MHz almost triples the US capacity, and the simultaneous expansion of the DS network capacity to 1 GHz would add a net 15-30 new DS QAMs (this calculation considers the combined effect of expanding the capacity of the network to 1 GHz from 860 MHz or 750 MHz respectively, and the loss of DS spectrum with the move of the split into the current DS region).

   - The shift of the split up to the 200 MHz is also being considered, but while this change would provide much more US capacity, it would reduce the next number of DS capacity significantly and would require the change of large numbers of non-DSG STBs (most of the STBs deployed to date) because the existing and extensively deployed OOB carriers would become inoperable since the region of the spectrum these utilize would be used for the US. Additionally, this change has other plant implications, such as the US equipment currently deployed would not support such extensive US, and thus a new HFC return strategy/equipment would be required.

2. Additional US via top-split, up to 1 or ~1.2 GHz

   Unlike the 1st option, this approach involves equipment not currently available. Instead, the preferred implementation of this option will require the development of network components and corresponding equipment that would make use of the existing forward spectrum but would use an unused portion of the spectrum, above 750 or 860 MHz and up to 1 or ~1.2 GHz, for US capacity.This new equipment could be built in the form of a new network gateway that would be installed in the vicinity of the node, which would provide the 'translation' from optical transmission from the headend up to the node location into electrical signals, and RF transmission from the node location through the coaxial portion of the HFC network. Additionally, this option would require the deployment of WDM equipment where not already installed to enable the use of the existing node backhaul fiber for communication back to the HE.

This approach would increase US capacity considerably, likely providing an additional 1 Gbps of net additional US bandwidth. In the process it leaves legacy services and existing CPE untouched, but it does not provide any additional DS capacity.

In order to implement this approach,new actives and return equipment would be required and some passive changes might be necessary, especially for older 750 MHz systems that may not have been built with 1 GHz capable actives and passives.

Because of the necessary equipment development and plant impacts, this option is not readily available to MSOs for execution as the first option was. In fact, if standardization were desirable, which all parties considerto almost be a requirement, this option requires considerable development before it would be ready to MSOs for deployment.

3. Overlay DS & US network above ~1.2 GHz up to ~1.7 GHz

   Like the 2nd option, this approach will require considerable equipment development before it would become available for deployment. Such equipment would use spectrum above that being used today for both additional US and DS capacity. Also like the 2nd option, this approach would require the development of a network gateway that would convert signals from electrical to optical to bypass the analog optical link from the headend to the node.

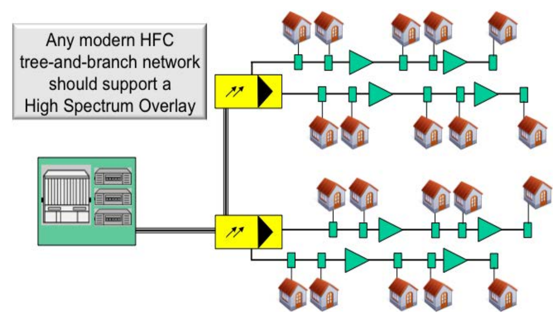


Figure 2: Typical HFC Network

This option could be implemented in two fundamental ways: one where the network gateway is located throughout the HFC network, and the other where the network gateway is deployedin the vicinity of the node.

In the first case, the network gateway would be installed in the vicinity of each active component where advanced services are to be provided. Therefore, this option would require the deploymentof additional fiber beyond what's already installed in the network, namely between the existing node and each of the active components in the HFC network. In that way WDM would be used to carry baseband signals up to the node, from which traditional PON technology would be used to interconnect each of the new network gateways back to the HE.

As shown by Figure 2, any modern HFC network should support a High Spectrum Overlay.

Figure 3 depicts an initial deployment of High Spectrum gateways, for which PON equipment is deployed in the headend, a separate optical wavelength is used in the trunk fiber to carry PON signals up to the node (shown in dashed blue lines), additional fiber is deployed in the distribution portion of the network (shown in solid blue lines), and new Network Gateways that provide optical-to-electrical signal conversion are installed to provide the overlay within an HFC segment between amplifiers.

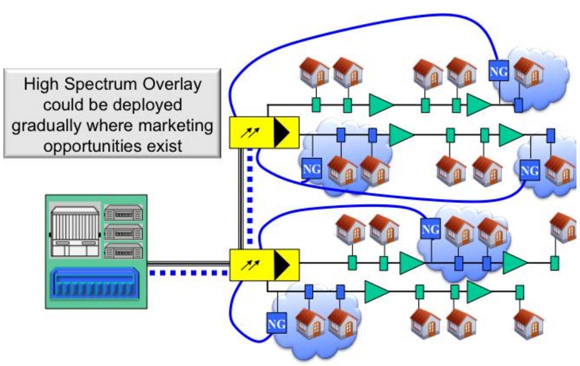


Figure 3: Initial High Spectrum Overlay

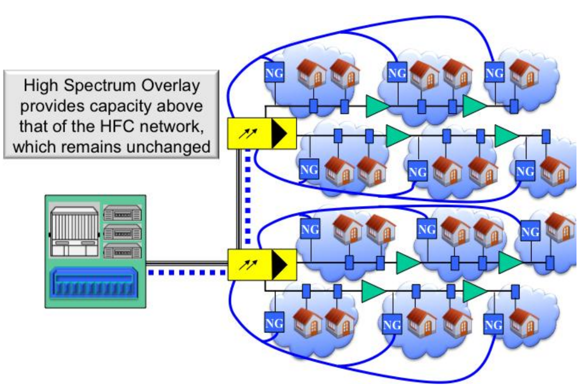


Figure 4: Complete High Spectrum Overlay

In the second case, signals in the new portion of the spectrum would be transmitted through the HFC amplifiers, for which these amplifiers would need to be modified. Figure shows a diagram for this approach.

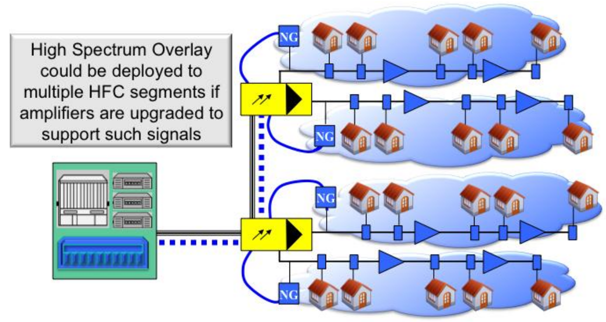


Figure 5: Multi-segment High Spectrum Overlay

Similar toOption 2, Option 3 leaves existing legacy services and current CPE untouched, and would take considerable amount of time to be developed.

However, unlike Option 2, this approach would enable more US andmore DS capacity.

COMPARISON AND APPLICABILITY OF EACH OPTION

For ease of evaluation of the options as versus the previously outlined ‘ideal criteria’, Table 1 below depicts a comparison of each of the 3 options considered above as a function of the ‘likely desirable preferences’ outlined before.

If only additional US is needed, #1 and #2 are good options. Option 1 will likely beless costly and quicker to implement, but could prove to become operationally more complex. Option 2 is probably 2x more expensive that Option 1, but it is operationally easier to deploy and offers greater US capacity.

If instead more DS is needed, then only #1 and #3 are viable. Option 1 is also less costly than option 3, but offers less peak capacity for both US and DS. Instead, Option 3 is probably 4x more expensive than Option 1, but offers significantly more capacity and long-term headroom and is the only option that can be deployed entirely on a success basis. Additionally, it could be argued that Option 3 is the only one that could be implemented with true flexibility in the allocation of US and DS, while Option 2 would be much more restrictive in achieving this goal.

IMPLEMENTATION ALTERNATIVES FOR OPTION 3

While multiple alternatives could exist for implementation of Option 3, only 2 of themare outlined in this paper. In addition, a combination of these 2 alternatives and its benefits is described.

A. Fiber to every active

Option	Added capacity	Path to smaller SG	Legacy unaffected	Success-base deployment	Flexible US/DS allocation
1. Increase US and expand DS to 1 GHz	~ 200 Mbps US	Yes	No	No	No
2. Increase US w/top split	~ 1 Gbps US	Yes	Yes	Partial	Yes
3. Overlay above ~1.2 GHz up to ~1.7 GHz	~ 1 Gbps US plus ~2 Gbps DS	Yes	Yes	Yes	Yes

Table 1: Evaluation of Options vs. Ideal Solution

This first implementation alternative consists of using fiber to the last active to transport US and DS signals optically between the headend and the coaxial part of the HFC network, converting these signals into RF via a network gateway at the location of the last active, combining the resulting RF signals onto the coaxial plant after the active HFC component, distributing thesignals via the cascade of passives to homes corresponding to that portion of the network (e.g., the signals from one network gateway are combined into the coaxial plant after an HFC active and do not traverse the following amplifier in the cascade), and finally terminating the RF signals in a home gateway where the HFC network is bridged onto the home using standard home networking technologies (e.g., Ethernet, MoCA, WiFi).

This approach should not be construed as resulting in a Node + 0 HFC cascade reduction. This is because the cascade of HFC actives is not modified. Instead the RF output operating between ~1.2 and ~1.7 GHz of the gateways deployed in the HFC network are combined with the RF signals existing in the coaxial network which operate below 1 GHz, much in the same way as narrowcasting a set of signals on a per service group basis where the other signals are broadcasted to the set of service groups.

The following categories of work would need to be performed in the plant in order to achieve the above:

- WDM could be used from the headed to the location of the node to reuse the existing long-haul fiber.
- To provide the remaining optical link from the node to the location of each active, additional fiber would be over-lashed to the distribution coaxial hardline cable, which is generally a short to medium length span.
- Finally, in order to pass RF signals between ~1.2 and ~1.7 GHz on the distribution network, it is likely that a large proportion of the tap faceplates would need to be replaced, although it is expected that the tap housing will likely support these new faceplates, and that only faceplates serving subscribers using this new portion of the spectrum would need to be replaced.

Assuming a high-bandwidth optical network from the headend to the network gateway, such as 10 Gbps EPON,and a high-order modulation and encoding scheme, it is expected that a transmission achieving ~6-8 b/Hz might bepossible, therefore resulting in a combined US/DS payload transport capacity exceeding 3 Gbps.

B. Fiber to the node

This second alternative consists of only using fiber to the node, andinstead of over-lashing fiber onto the coaxial distribution hardline RF signals would be used on the coaxial plant from the node through the active components of the HFC network. The same tap faceplate replacement would be required as in the previous implementation alternative.

This approach needs little to no fiber construction, but as a trade-off requires the development and installation of modified actives which are likely complex to develop and expensive to deploy.

Assuming currently available modulation and encoding techniques, and a reduced operating spectrum, an effective transmission of~4-6 b/Hz is expected, which could result in a payload transport capacity of 2-3 Gbps.

C. Combination of alternatives A. and B.

An alternative approach could be that of combining the two alternatives previously described. This could be done for a number of reasons, but perhaps the most important one could be the relative value of each approach. The first approach has the benefit of its increased capacity and lower RF complexity, but has the drawback of requiring the fiber over-lash throughout the coaxial hardline in the HFC network. The second has the converse benefits and drawbacks. Perhaps it might make sense to utilize the first approach where aerial plant exists, and deploy the second approach in cases where underground plant exists. In this way the best performance/cost trade-off can be achieved.

Furthermore, it might make sense to progressively deploy fiber deeper into the network up to the last active in areas where construction costs are relatively high as compared to the cost of the modified actives and their installation. Fiber could then be added progressively as additional capacity is needed over time, likely coinciding with the continuedcascade reduction business-as-usual strategy, eventually reaching the last active, or deploying to last active when additional capacity is needed in the narrowcast portion of the spectrum.

COMBINING OPTIONS

As a further refinement of the approaches suggested, it may make sense to develop a technology strategy that implements the benefits of each option in various stages, and progressively leverages them as these become necessary.

For example, while US capacity beyond 200 – 300 Mbps in the upstream is sufficient and no significant additional DS spectrum is needed, deployments of Option A (e.g., move split to 85 MHz and extend forward spectrum to 1 GHz) might be sufficient, and could be followed by deploying network gateways from Option C using either of he proposed approaches as additional capacity is required.

OVERALL ACCESS ARCHITECTURE

The new edge platform devices currently under development by vendors, as specified by the CCAP architecture, will support either of the approaches described above. The CCAP architecture already supports the modularity necessary to upgrade line cards progressively as new technologies become available. Furthermore, the CCAP architecture provides support for EPON, such that even the Option 3 is supported in the overall access architecture.

SILICON DEVELOPMENT

One important consideration in evaluating the benefits of each approach is the need and availability of silicon components, or on the flip side the need for its development.

This is critical for the following fundamental reasons:

a. When silicon exists the availability of the system solution is quicker, whereas when it needs to be developed the timeline is significantly longer, and

b. If silicon devices, or at least some of their components, are used for multiple purposes, especially for multiple industries, then their production increase rapidly and costs decrease considerably.

Option 1 might likely not need silicon development, but the other two options would, for which technology design decisions would be important.

CONCLUSIONS

HFC networks currently have an US capacity limited by the available spectrum, typically limited to the 20-30 MHz of spectrum available between 5 and 30 or 42 MHz in North American systems. At 6.4 MHz of channelization and QAM64 modulation, this capacity allows transmissions of ~100 Mbps.

While US capacity is limited, analysis of DS capacity needs indicates that additional DS capacity will likely be necessary, and perhaps be needed even before additional US capacity is required.

Additional US capacity could be achieved via a change in split to 65 MHz, 85 MHz, or even higher split systems, but these are complex to implement, service impacting and financially and operationally onerous. The capacity gain from such migration is limited to doubling or perhaps tripling the maximum transmission capacity, and has an indirect impact in the reduction of the downstream capacity of the HFC systems.

Taking into account other factors may result in concluding that other solutions may be more appropriate and no more operationally impacting than just changing the spectrum split to increase upstream transmission.

Such other factors include: use a different portion of the spectrum for US,enable additional DS transmission capacity, establish a more efficient HFC transmission mechanism for coupling with home gateway devices, enabling more native support for IP-based, enhanced business, wireless and other services and applications.

When considering theseadditional factors, alternate solutions to just changing the US/DS spectrum split may become more attractive and valuable to MSOs.

This paper presented an analysis of various such options, and expanded on one particular alternative that offers the following characteristics: leaves all current services untouched in the currently allocated HFC spectrum, unleashes new spectrum in the HFC network, and could enable transmission of up to an additional 3 Gbps combined between upstream and downstream.

To do so, the proposed approach makes use of traditional and already available technology and network strategies in a new way, such as pushing fiber deeper up to the current last active, use of existing baseband fiber-based technologies, implementation of existing technologies for higher-order PHY and light-weight MAC, superimpose the new transmission in a portion of the spectrum above the 1 GHzband currently being used, utilizing spectrum allowing the use of passive devices such as below 1.8 GHz, leaving existing services untouched, enables new capacity in the HFC network, solving the current US capacity limitation, and increasingDS capacity.

ACKNOWLEDGEMENTS

Thanks to my colleagues Mitch Bowling, Doug Jones, John Leddy, Saif Rahmanand Dave Urban from Comcast for their contributions to the discussions leading to the proposals and conclusions presented in this paper.

Also,many thanks to my colleagues from CableLabs®, the many vendors and advisors of Comcast, and especially fellow MSOs, for the many discussions leading to the materials presented.

ACRONYMS

The following acronyms are used within the paper without being previously defined:

ABR: Adaptive Bit-rate

ARPU:Average Revenue per User

AVC: Advanced Video CODEC

BC: Broadcast

CAS: Conditional Access System

CBR: Constant Bit-rate

CCAP: Converged Cable Access Platform

DOCSIS: Data over cable service interface specification

DTA: Digital transport adapter

Gbps: Gigabit per second

GHz: Gigahertz

GigE: Gigabit Ethernet

HE: Headend
HFC: Hybrid fiber-coax

HSD: High-speed data

MAC: Media Access Protocol

MetroE: Metro-Ethernet

MHz: Megahertz

MPEG: Moving Picture Experts Group

MSO: Multiple system operator

NC: Narrowcast

nDVR: Network DVR, sometimes referred to as RS-DVR for remote storage DVR

OOB: Out of Band

PHY: Physical

PON/EPON: Passive optical network/Ethernet passive optical network

QAM: Quadrature amplitude modulation

RGU: Revenue Generating Unit

RF: Radio frequency

SDV: Switched digital video

STB: Set-top Box

VBR: Variable Bit-rate

VOD: Video on-demand

WDM: Wave Division Multiplexing

---

[1] Presented at Cable Congress 2011 by Arris

[ii] ITU-T Recommendation H.264: 2005, Advanced Video Coding for generic audio-visual services
[iii] ISO/IEC 14496-10: 2005, Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding
[iv] Capacity, Admission Control, and Variability of VBR Flows, CableLabs Winter Conference, February, 2009

RFoG: OVERCOMING THE FORWARD AND REVERSE CAPACITY CONSTRAINTS

Oleh J. Sniezko, Chief Technical Officer
Aurora Networks, Inc.
5400 Betsy Ross Dr, Santa Clara CA 94054
Phone: 408-235-7000 | osniezko@aurora.com

Abstract

The paper will present technical analysis and performance modeling for forward transmitters to show that in the absence of RF loss in the access distribution networks, bandwidth in excess of 1.6 GHz and capacity in excess of 10 Gbps can be supported at nominal incremental cost. Hence, in this case, RFoG has the potential of exceeding 10G PON/EPON forward capacity without the need to add overlay systems.

For the reverse path, the paper will show how FDD can take advantage of forward and reverse wavelength separation (WDD) to support full duplex communication in RFoG systems with very high upstream path capacity. Ultimately, WDD enables the capacity of an entire single wavelength to be dedicated to the forward or the reverse path. This increased capacity in the upstream can exceed 1 Gbps and be provisioned in a manner that provides full compatibility with the deployed network edge equipment; equipment in the headend and even more importantly consumer electronics and the operator's terminal equipment on the customer premises.

INTRODUCTION

The industry has agreed – there will always be the need for more bandwidth driven by higher data speeds, over-the-top services, 3D-TV, more HDTV – the list goes on. There is no such concept as 'too much bandwidth.' However, the industry's newest architecture, RFoG, has until very recently suffered from quite the opposite – it was unable to match the capacity of proven HFC networks. Initially RFoG deployments were motivated mostly by the desire to provide FTTH networks based on business considerations unrelated to the cost or capability of the technology. Given the investment required, it was important for RFoG to match the capabilities of HFC networks. The technology has successfully accomplished this objective within the first four years of deployment.

The technology today allows for all of the below:

1) Reach of 20 km in a passive configuration with 32-way field split;
2) Support for full forward capacity of the HFC network;
3) Support for reverse capacity of 27-30 MHz load with up to four 6.4 MHz 64-QAM signals with adequate operational margins;
4) All of the above with a combining group of 500 customers (1,000 HP) for highly interactive services;
5) And with analog modulated reverse laser technology to provide the widest possible interoperability.

Extended reach for customers located further than 20 km from headends and hubs was also added to this list of accomplishments.

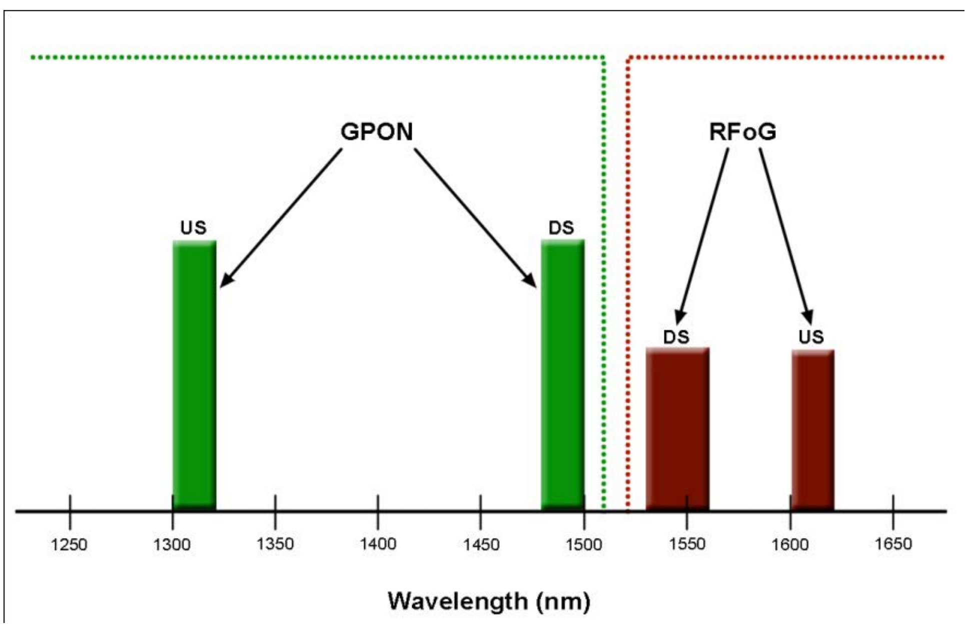
To complete the list, the RFoG technology was enhanced with the capability to prevent OBI when non-TDMA reverse access protocols are used (e.g., S-CDMA) or several (multiple MAC domains), non-synchronized multiple TDMA services are activated or ingress is high enough to break the reverse transmitter squelch[1, 2].

What happens next, when even more bandwidth is required? This is the question that all broadband telecommunications network operators are pondering.

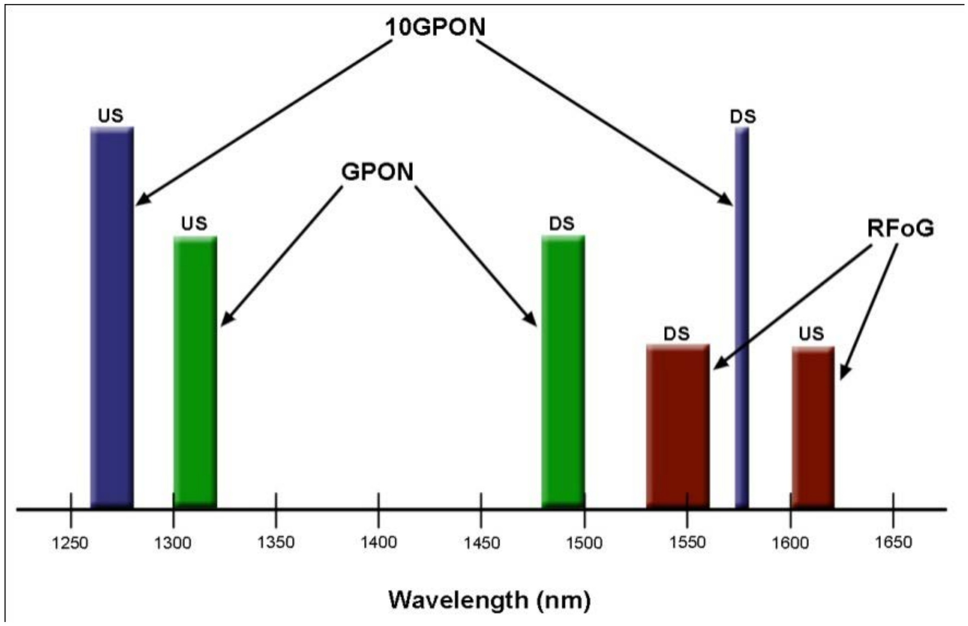
RFoG networks offer some comfort to the operators who installed them based on other considerations. These networks offer options for expanding both forward and reverse capacity in an easy to implement manner. In a world of changing and growing demand for capacity and continuously changing traffic parameters (with asymmetrical capacity demand in the forward and reverse directions continuously changing based on the traffic characteristics of an application "du jour"), the ease of capacity expansion for installed RFoG networks is a critical factor adding to their benefits.

RFoG FTTH AND PON OVERLAY

An RFoG network is a specific case of an FTTH system. The downstream and upstream wavelength selection can enable compatibility of the RFoG network with other FTTH technologies such as GEPON or GPON. The selection of wavelengths for achieving such compatibility with GPON and 1G EPON (a.k.a. GEPON) has been defined by the industry and documented in the SCTE RFoG standardization effort[3]. This wavelength allocation as currently defined by the industry is presented in Figure 1 and enables compatibility with GPON and GEPON systems by addition of a relatively simple optical diplex filter. However, to also achieve compatibility with 10G PON and 10G EPON systems, a more complex, and expensive optical filter would have to be deployed. This translates into higher cost of the R-ONUs (RFoG ONUs).



a) Wavelength Compatibility with GPON/GEPON Overlay (with Optical Filter Overlay)



b) Wavelength Allocation for RFoG Compatible with GPON, GEPON and 10G PON/EPON
Figure 1: Existing Recommended Wavelength Allocation for RFoG System Compatible with GPON/GEPON Systems

Modified wavelength allocation remedies this situation. Figure 2 depicts a wavelength allocation for compatibility with GPON/GEPON and 10G PON/EPON systems by allowing a simple optical filter to separate RFoG and PON wavelengths. An RFoG system with wavelengths presented below is simple to build with a low incremental cost for an optical filter. The simplicity of the internally integrated RFoG filter for separating forward and reverse signals at the R-ONU is also maintained.

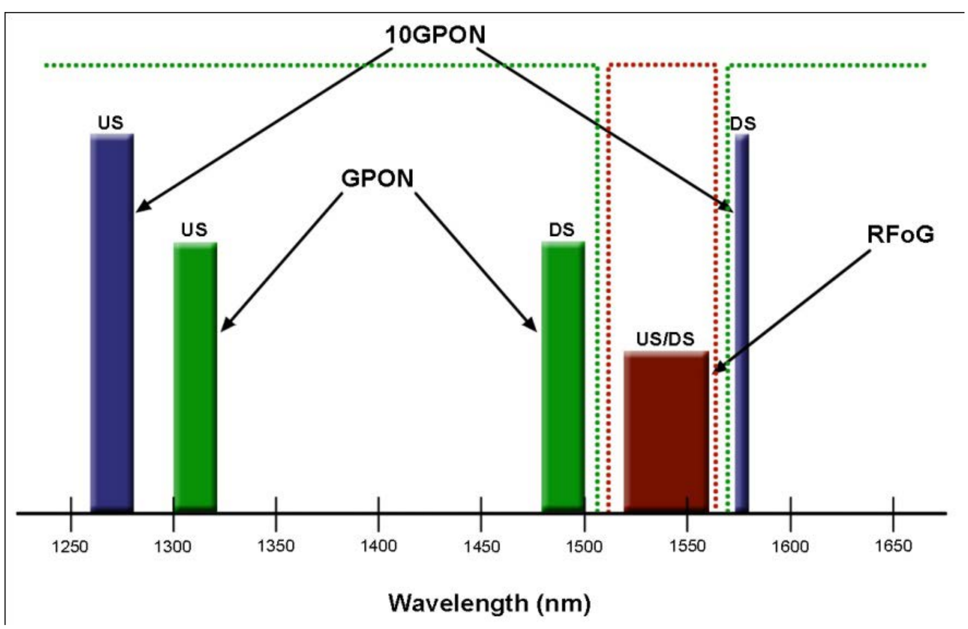


Figure 2: Proposed RFoG Wavelength Allocation for Compatibility with GPON/GEPON and 10G PON/EPON Systems

PON overlay over RFoG network permits fast expansion of the network capacity in both the forward and reverse paths beyond that readily achievable in HFC systems. Forward capacity can be enhanced by 10 Gbps in the forward direction and by 1, 2.5 or 10 Gbps in the reverse direction.

Although forward capacity expansion can be accomplished in HFC networks, this comes at a significant upgrade cost. Traditional HFC network of 1 GHz can support approximately 6 Gbps capacity with conversion to digital-only load. Further expansion of forward capacity would require both an upgrade of the network performance and, related to it, increased maintenance cost (to support 1024-QAM signals) or bandwidth expansion of the optical transmitters, nodes, RF amplifiers and passives (with the associated labor cost). The expansion of the reverse capacity in HFC networks is even more difficult and costly.

Although both forward and reverse capacity expansion is less burdensome and costly to address in Fiber Deep HFC N+0 networks, a PON overlay for RFoG network provides an easier path. However, easier might not necessarily mean less costly. This expansion of capacity with PON overlay requires the addition of a new set of edge devices with duplication of optoelectronics. The cost, including additional headend components, optical filters, R-ONUs compatible with PON overlay and PON ONUs can range between $200 and $500/customer today even with just a moderate increase in capacity by deploying GPON or GEPON components.

Our industry elected to use FTTH RFoG systems for residential services instead of converting to FTTH PON systems to preserve the deployed base of and sizeable investment in headend and hub equipment and consumer electronics and consumer terminal equipment through the preservation of the FDM structure of the transported signals. The RFoG network delivers the same FDM signals as the traditional HFC network. But the most important fact is that with the technological advances of the last five years, RFoG systems can be built to provide capacity expansion paralleling (if not exceeding) the capacity the expansion offered by PON overlay at a cost projected to be lower (possibly significantly lower) than the cost of additional PON overlay equipment and overlay filters and without the need for any other RFoG upstream wavelength than 1310 nm.

RFoG WITH EXPANDED CAPACITY

The RFoG network, as already explained above, is an FTTH system. It supports FDM signals while *x*PON systems today deliver services by transporting digital baseband (TDM) signals. This difference stems from historical differences in the transport media used by operators favoring RFoG systems and operators favoring *x*PON systems. RFoG is favored by operators that historically used coaxial cable with RF amplifiers for reach extension to deliver telecommunication services to their customers while *x*PON systems are favored by operators that historically used passive copper pairs. These two systems differ in the way they support bidirectional communication: active coaxial systems achieve it through FDD (efficient for long distances supported by active coaxial networks) while passive copper system support it with TDD suitable for shorter distances and lower speeds. Passive coaxial network in HFC N+0 systems can also support bidirectional communication using TDD.

Fiber can easily transport either FDM RF signals (and has been deployed in these applications since the late 1980s) or baseband digital signals. Fiber based FTTH systems, whether RFoG or PON, can support FDM RF signals and baseband digital signals as well. Despite being passive in nature for the last several miles, both RFoG and *x*PON systems support bidirectional communication based on perfect FDD technology: WDD. In the case of *x*PON networks, forward and reverse signals occupy overlapping frequency bandwidths and full duplex service is enabled by WDM technology. In the case of traditional RFoG systems that mimic HFC networks, the FDD used in the active coaxial network has been preserved to support bidirectional communication. This arrangement allows compatibility with the existing consumer equipment and terminals and headend/hub equipment.

However, FDD is not technically required to support full duplex communication in RFoG PON because forward and reverse signals are carried on separate, not conflicting and well isolated wavelengths. Figure 3 depicts the simple fact that as long as the FDM RF signals are kept from overlapping each other before EO and after OE conversions, they can occupy overlapping RF bandwidths in the fiber where they are carried on separate wavelengths.

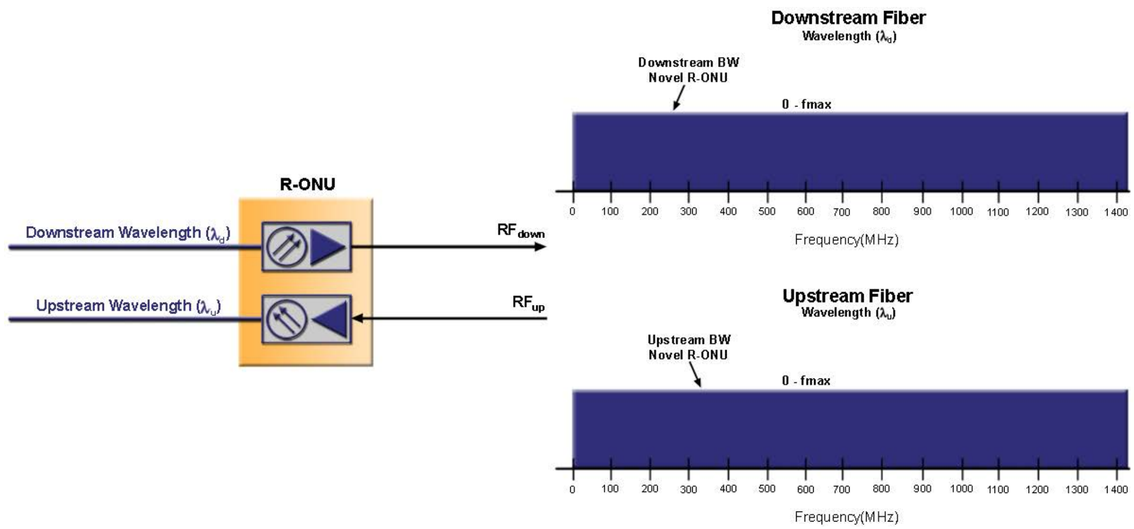


Figure 3: WDM as Replacement for FDD

<u>Forward Capacity</u>

WDM technology enables dedication of the entire single wavelength capacity either to forward or reverse signals. How much wavelength capacity can be used in a particular system will depend on edge equipment quality and many other non-trivial factors. In the forward direction, several solutions for increasing forward capacity are available. One of the most straightforward is increasing capacity of the forward transmitter.

The analysis below is based on the assumption that before additional capacity is added to an RFoG network, the transition to digital-only load has been completed. The following performance of the RFoG network forward components and parameters of the carried signals are also assumed:

1) Link loss budget of 19.5 dB at 1550 nm:
   a. Fiber length of 6 km (1.5 dB loss)
   b. Passive loss of 18 dB (splitters for 32 and filters)
2) Transmitter:
   a. Directly modulated
   b. 10 dBm optical output power
   c. Transmitter load of 149 channels with 256-QAM signals
3) R-ONU receiver:
   a. Thermal noise performance of 5 pA/√Hz (after the filter separating forward and reverse wavelengths)

Under these assumptions with predistortion circuitry implemented to compensate for dispersion/chirp related second order CIN, CNR for the RFoG optical link will equal approximately 40 dB. The test results for this fiber length and received input level confirm the modeling numbers.

The same transmitter can be used to carry signals above 1,002 MHz for additional capacity. Let us make assumptions about these signals:

1) Additional load of 600 MHz
2) Modulation of 64-QAM
3) Level relative to the load below 1,002 MHz at −6 dBc.

For this transmitter type, two factors will affect the link CNR after the additional load is implemented:

1) Lower OMI/channel at the same composite OMI,
2) Higher CIN caused by additional second order intermodulation products due to interaction between the laser chirp and fiber dispersion.

The OMI correction for CNR is simple to calculate and for the assumptions listed above will approximate to 0.7 dB CNR degradation for 256-QAM channels based on the following equation:

$$10 \times \log \left({}^{N_1}\!/_{N_2} \right)$$

where:

N_1 – channel count for the original load (149 channels)

N_2 – equivalent (in level to 256-QAM channels) channel count for the new load equivalent (149 + 100/4)

The CNR correction for second order CIN is more difficult as it depends on predistortion circuitry efficiency. Assuming no predistortion, the following equation can be used to calculate noise floor relative to additional 64-QAM channel levels.

$$N(f)_{64-QAM} = 10 \log \left[N_{CSO} \left(2\pi f \, m \frac{\lambda^2}{c} D \, L \, \upsilon \right)^2 \right]$$

where:

N_{CSO} – number of CSO beat products at RF frequency f

m – OMI (per channel) of the added 64-QAM channels

D – fiber dispersion

L – fiber length

λ – optical wavelength of the transmitter

ν – laser chirp at 100% peak OMI.

Figure 4 presents the beat count plot for beats generated by the additional load between 1,002 and 1,602 MHz and the level of CIN relative to frequency of the beats. Figure 5 presents the noise (as defined by the equation above) generated by the beats for lasers of different chirp at the frequency of the worst performance (~400 MHz). The noise floor is referenced to the level of 256-QAM channels (6 dB higher than the channels within 1,002 and 1,602 MHz frequency range).Although each beat power is spread over double the channel width, they overlap except for the extreme frequencies of beats hence the correction for the peak value of noise floor can be neglected (noting that a small downside correction could be applied due to the fact that the QAM channels is only 5.3 MHz wide).

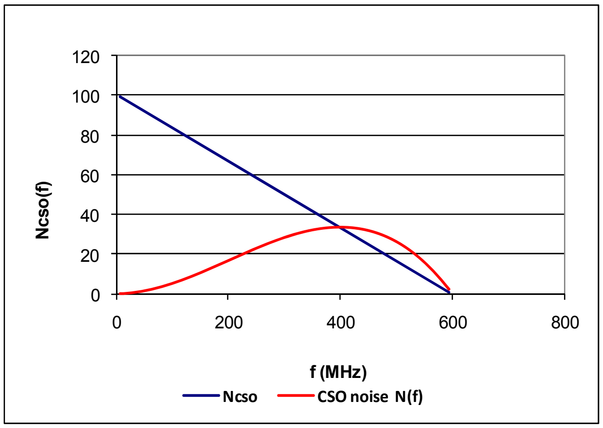


Figure 4: CSO_ beat count distribution and their relative level as a function of frequency

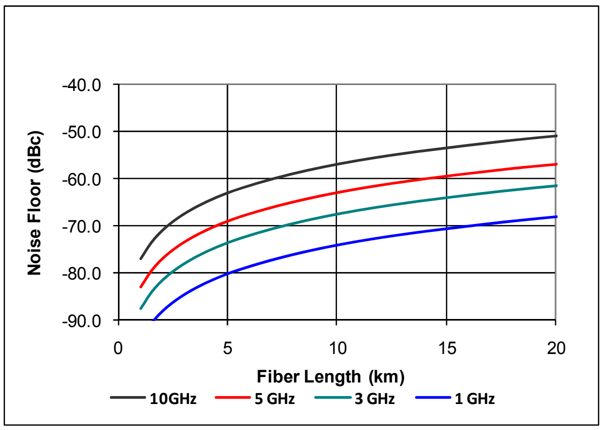


Figure 5: CIN noise caused by the load between 1,002 and 1,602 MHz at the worst frequency relative to the level of 256-QAM signals for lasers with different chirp

The noise levels depicted in Figure 5 are negligible for the distances used in the model (6 km) even without predistortion. It is therefore safe to assume that the CNR degradation caused by additional load of the forward transmitter for the link described and under the assumptions presented above will be limited to 1 dB across the 108 to 1,002 MHz bandwidth.

The assumptions above define only one of many possible sets of conditions but there are many tools at the disposal of optical link engineers to extend the reach and to allow different channel loading. The performance modeling is indispensable during the selection of these tools and the topology of the network.

In the example described above, forward capacity was expanded by approximately 2.8 Gbps without significant additional cost (except for the cost of NRE to be recovered by the vendors). Many other technologies and approaches are available to expand capacity even further. Tools for expanding RFoG network bandwidth up to 2 GHz with 256-QAM load across all frequencies are readily available. These tools can more than double the forward bandwidth from approximately 6 Gbps to 12.5 Gbps. This bandwidth can scale from supporting 1,000 HP to being shared among 32 HP (and even fewer customers). It provides unbeatable and scalable offer in terms of cost and capacity in all RFoG systems where there is no RF loss heavily dependent on frequency as is the case in coaxial networks.

Reverse Capacity

The forward capacity expansion in RFoG systems is quite dramatic and achieved at relatively low incremental cost. But even more dramatic is the capability for expansion of the reverse capacity way beyond the capacity of today's 5-42, 5-65 and even 5-85 MHz HFC networks. The capacity of these systems can at best reach 120, 240 or 360 Mbps respectively. RFoG capacity is not limited by bandwidth. The bandwidth limitation is the legacy having its origin in HFC RF active networks that required FDD to support bidirectional traffic over relatively long distances with the help of amplification of both forward and reverse signals. The asymmetry of the FDD arrangement was created due to limited demand for reverse capacity and due to the fact that most of the spectrum above 50 MHz was allocated to off-air broadcast services, with consumer electronics designed to receive those signals.

In RFoG systems where the entire wavelength is dedicated to carry reverse signals between electrical interfaces, the bandwidth limitation disappears. How much bandwidth can be used and how much capacity can be provided depends on the system architecture and performance of its components.

The tradeoffs[4] among several network parameters can significantly increase RFoG network potential for expanded capacity. In this paper, two of them will be explained in detail.

One of the two factors is the level of aggregation. The HFC networks show high level of flexibility in aggregation levels. Fortunately, thanks to the technological advances in RFoG reverse receivers, this aggregation flexibility has been maintained. Figure 6 depicts a design of a high aggregation receiver that avoids the penalty of RF combining of typical HFC receiver outputs. A careful compromise between thermal noise performance of the front end of this receiver and parasitic parameters of the photodiodes used in this example enable expansion of the available bandwidth to between 300 and 400 MHz at reasonable thermal noise performance (see Figure 7) and aggregation levels of 128 HP. When the capacity available in this design becomes insufficient, de-aggregation (see Figures 8 and 9) can lead to higher bandwidth and hence increased aggregate (burst) capacity and capacity-per-user in the reverse path (see Figure 10). This last de-aggregation step provides RFoG customers with the ultimate bandwidth that photodiodes analyzed in this example can support.

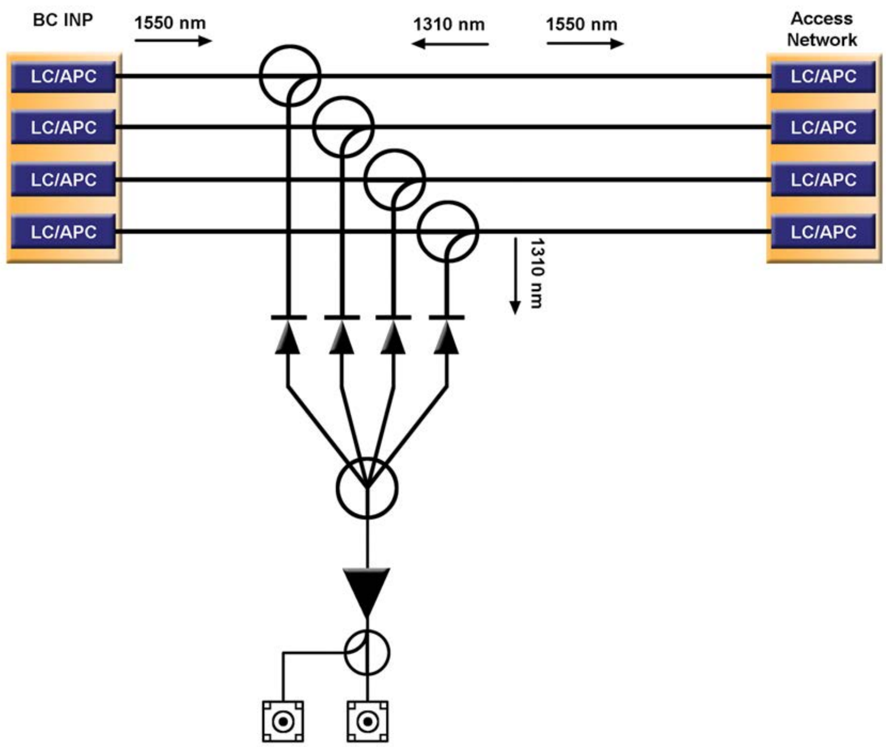


Figure 6: Design of High Aggregation Receiver without Noise Aggregation Penalty

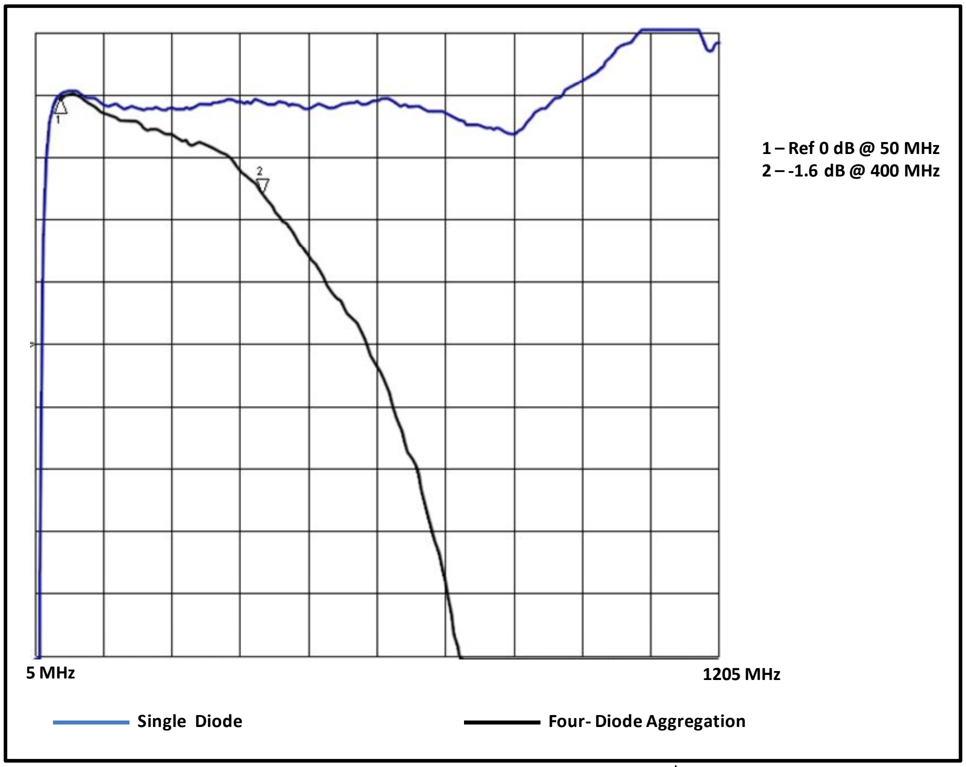


Figure 7: Bandwidth Capacity for 128 HP Aggregation Level and 4 pA/√Hz Equalized Thermal Performance

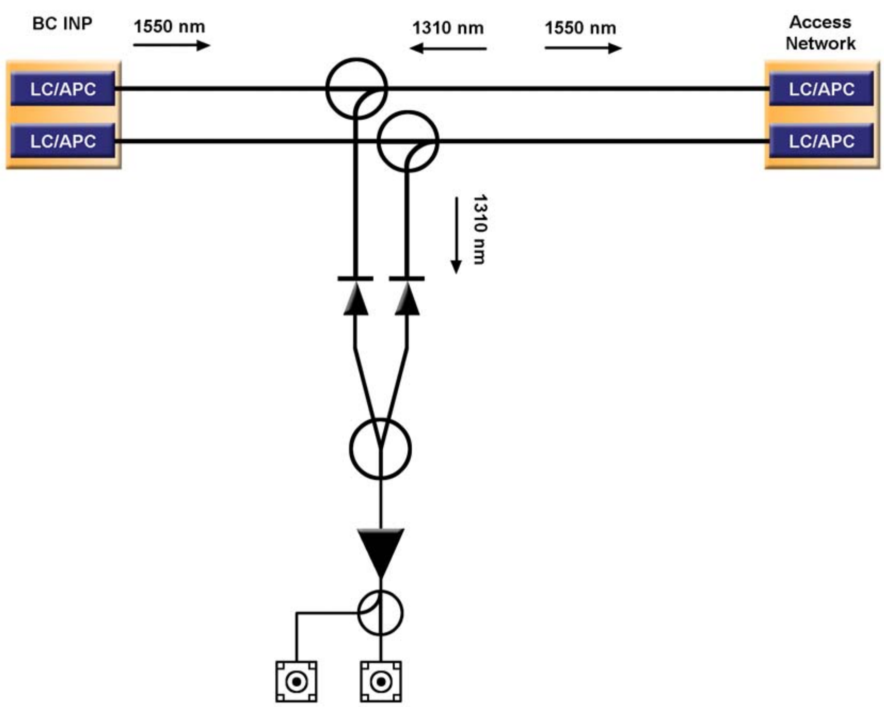


Figure 8: Reverse Receiver Design with 64 HP Aggregation Level

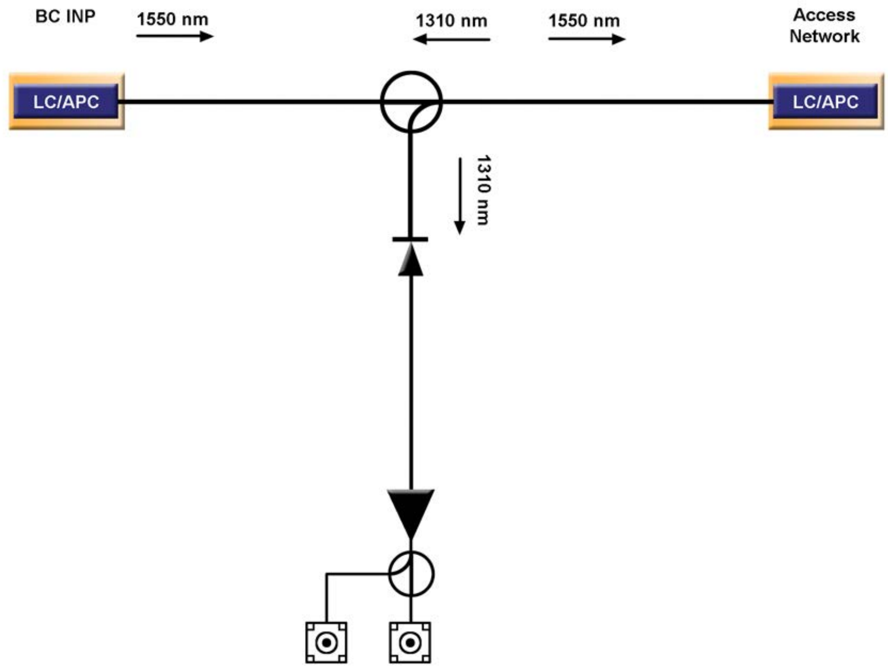


Figure 9: Receiver for 32 HP without Aggregation

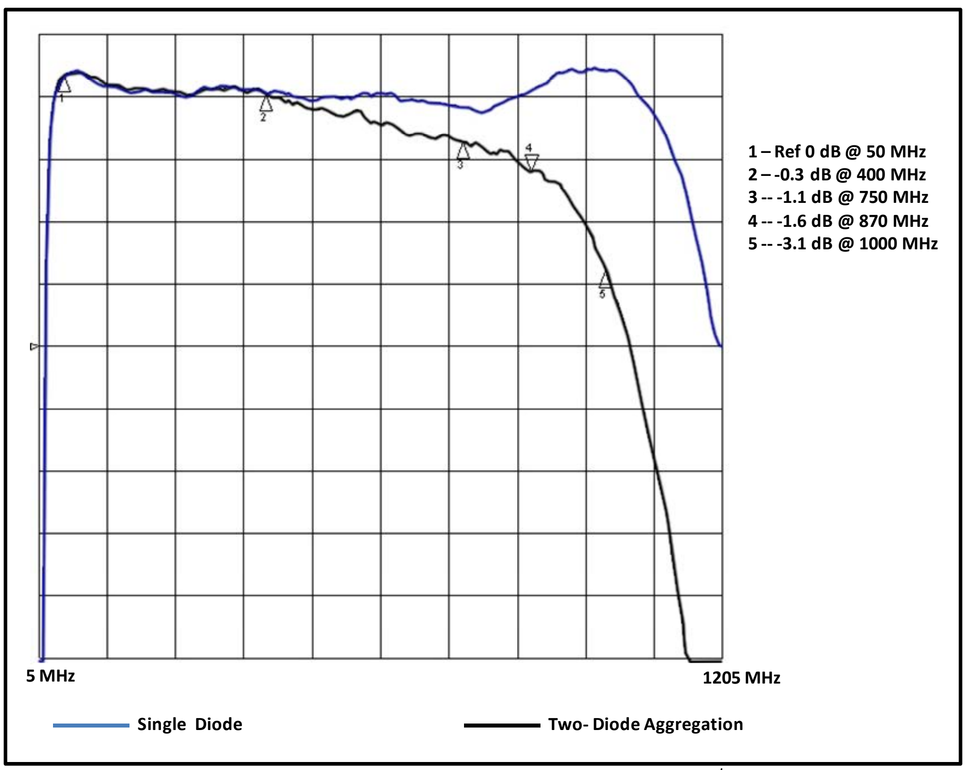


Figure 10: Bandwidth Capacity for 64 and 32 HP Aggregation Levels and 4 pA/√Hz Equalized Thermal Performance

Thermal noise performance plots in Figures 7 and 10 indicate that at 4 pA/√Hz equalized thermal noise performance, the available bandwidth for:

- 128 HP aggregation level is 320 MHz
- 64 HP aggregation level is 640 MHz
- 32 HP aggregation level is 960 MHz.

Based on the available bandwidth and digital signal requirements, several capacity configurations were modeled under the following assumptions:

1) Link loss budget of 19.0 dB at 1310 nm:
    a. Fiber length of 6 km (2.0 dB loss)
    b. Passive loss of 17 dB (splitters for 32 and one filter)
2) R-ONU transmitter:
    a. Directly modulated
    b. 2 dBm optical output power
    c. Transmitter load as in analyzed load capacity models
3) Headend receiver:
    a. Thermal noise performance of 4 pA/√Hz (after the filter separating forward and reverse wavelengths)
    b. Bandwidth as in analyzed aggregation examples
4) The channels are assumed to be similar to DOCSIS® channels (channel bandwidth, modulation and symbol rates).

<u>*Example 1:*</u>

At an aggregation level of 128 HP, the network is bandwidth limited and the capacity is optimized with 50x6.4 MHz channels of 64-QAM modulation. The modeled performance (see Figure 11) indicates just sufficient operational dynamic range for 10E-6 uncoded BER performance. The total reverse capacity is 1.5 Gbps. With these channels bonded, the burst capacity would exceed the burst capacity of GEPON networks or traditional RFoG capacity with GEPON overlay.

<u>*Example 2:*</u>

At an aggregation level of 64 HP, the network is performance limited and the capacity is optimized with a mix of 64-QAM and 16-QAM channels. Maintaining the same dynamic range for 64-QAM channels as in example 1, the optimal capacity would be supported with approximately 38x6.4 MHz channels of 64-QAM modulation and 62x6.4 MHz channels of 16-QAM modulation for a total reverse capacity of 2.4 Gbps. However, for operational simplicity at this level of aggregation, loading with 100x16-QAM would be preferred. It would provide wider operational dynamic range and uniform loading across the entire bandwidth.

<u>*Example 3:*</u>

At an aggregation level of 32 HP (single RFoG group), the network is performance limited and the capacity is optimized with 150x6.4 MHz channels of 16-QAM modulation. The modeled performance (see Figure 12) indicates comfortable operational dynamic range for 10E-6 uncoded BER performance. The total reverse capacity is 3 Gbps. Up to 12x6.4 MHz channels of 64-QAM modulation to support legacy modems at the highest possible capacity can be added at lower frequencies with the remaining load of 16-QAM channels. Although the capacity gain in this case is minimal, it could be beneficial to support legacy modems at higher rates. Alternatively, wider operational dynamic range can be traded for longer reach (extended to 10 km).

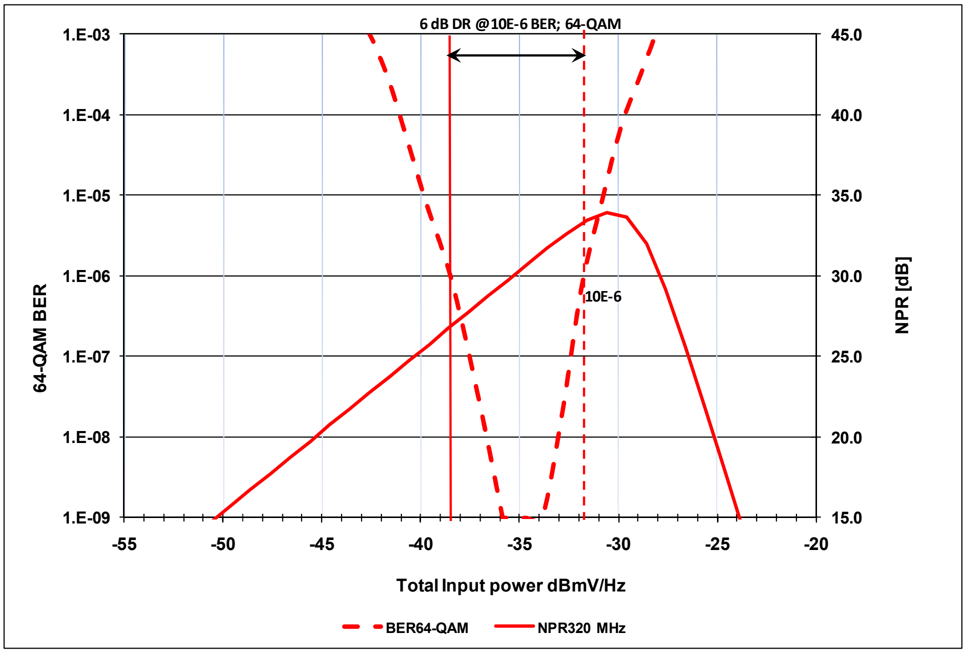


Figure 11: Reverse RFoG System Performance at 128 HP Aggregation Level with 320 MHz Load of 64-QAM Signal

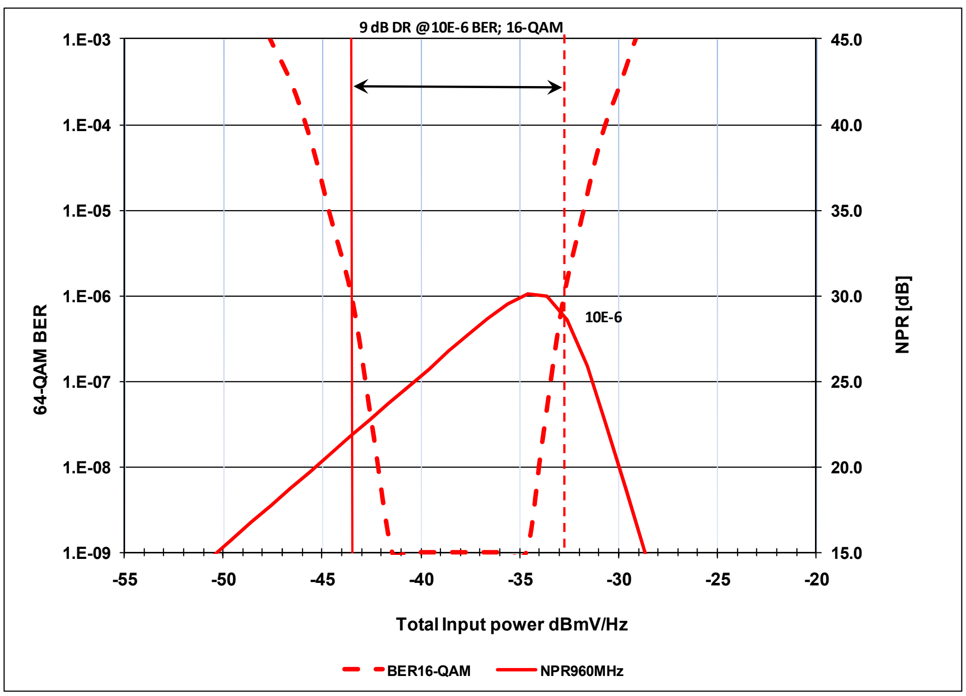


Figure 12: Reverse RFoG System Performance at 32 HP Aggregation Level with 960 MHz Load of 16-QAM Signal

The examples described above document that even with quite standard components, RFoG FTTH networks can significantly expand reverse capacity available to operators. Free of the FDD bandwidth constraints of coaxial active networks, RFoG FTTH networks can match and exceed capacity of the PON networks while preserving operators' investment in headend/hub equipment and consumer electronics whether owned by the operator or by customers. This is due to the fact that the RFoG networks preserve the FDM nature of signals.

The capacity expansion implementation of must take this desire to preserve investment into account to enable the legacy equipment to operate without interference while providing the extended capacity in the forward and reverse directions.

RFoG Capacity Potential: Recapitulation

RFoG FTTH systems are capable of supporting full-duplex (bidirectional) communication without maintaining reverse and forward signals in different frequency ranges. The FDD required in coaxial active networks to separate these signals is replaced with WDM where different wavelengths can carry overlapping RF bandwidths, especially in a counter-propagating mode. This fact enables RFoG networks to provide capacity matching capacity of GPON/GEPON and 10G PON/EPON systems as explained in the paper. Moreover, given that there is no RF distribution network to speak of, the levels in the forward and the reverse bandwidth do not have to be excessively high to secure adequate performance. Finally, the required isolation between the legacy equipment and extended bandwidth signals can be accomplished within the R-ONU.

Figure 13 presents an example of an R-ONU implementation that separates legacy signals in a manner compatible with all consumer electronics and consumer terminal equipment using built-in isolation between the legacy HFC signals and the signals supporting enhanced capacity. It also provides a separate port for the FDD RF signals with enhanced (symmetrical or asymmetrical) capacity service. This R-ONU is transparent to signals appearing on the second port as long as they observe the frequency allocation of that FDD system. The frequency allocation, channel scheme, modulation type and level, coding and protocols of these signals are at the discretion of the system operator assuming that the link (downstream and upstream) performance is adequate to transport these signals unimpaired.

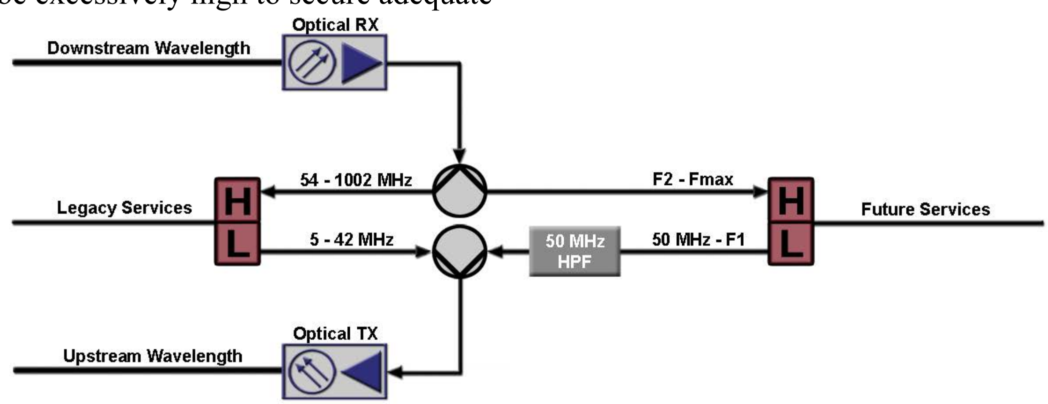


Figure 13: Example of High Capacity R-ONU with Legacy and Enhanced Capacity FDD Bi-Directional Communication (F1<F2)

Many other implementations[4] are feasible, including implementations with integrated cable modems, advanced cable modems for expanded capacity, home networking circuitry, EMTA for voice services and other service terminals.

CONCLUSIONS

RFoG systems are the networks of choice in green field scenarios for operators who prefer the FDM signal delivery method for communication services, especially in places where the construction cost of such a network is lower than the construction cost of an equivalent HFC network. As reported by the industry, this is usually the case where population density falls below approximately 50 HHP/mile. Other business considerations may also influence RFoG technology selection for deployment in a particular project. RFoG systems match HFC network capacity and flexibility but more importantly, they can exceed it. Indeed, operators who deploy these systems can be assured that the system can readily be provisioned or upgraded at any time in the future for capacity well in excess of any HFC network deployed today. They can match and exceed the capacity of GPON/GEPON and 10G PON/EPON networks at minimal incremental cost while preserving all the benefits of HFC networks. The capacity expansion can be accomplished with PON overlay. However, significant cost reduction can be accomplished by deploying enhanced capacity RFoG systems at minimal incremental cost if any. The choice of options will be dictated by the operators' preferences and cost considerations. Modern RFoG systems are basically future-proofed.

ACKNOWLEDGEMENTS

Krzysztof Pradzynski and Sudhesh Mysore contributed critical material and analysis to this paper to enhance its value. The author wishes to thank them for their contributions.

ABBREVIATIONS AND ACRONYMS

10G PON	ITU-T's next-generation broadband transmission standard with 10 Gbps throughput, a.k.a. 10G-PON or 10GPON
10G EPON	IEEE 802.3 Ethernet PON standard with 10 Gbps throughput, a.k.a. 10G-EPON or 10GEPON
BER	Bit Error Rate
CIN	Composite Intermodulation Noise
CNR	Carrier-to-Noise Ratio
DOCSIS®	Data over Cable Service Interface Specification
DR	Dynamic Range
EMTA	Embedded Multimedia Terminal Adapter
EO	Electro-Optical
FDD	Frequency Division Duplex
FDM	Frequency Division Multiplexing
FTTH	Fiber to the Home
GEPON	IEEE 802.3 Ethernet PON standard with 1 Gbps throughput, a.k.a. 1G-EPON, G-EPON or 10GEPON
G PON	ITU-T G.984 Gigabit PON standard with 2.488 Gbps throughput, a.k.a. 1G-PON or G-PON
HDTV	High Definition Television
HFC	Hybrid Fiber Coaxial
HHP	Households Passed
HP	Households Passed
MAC	Media Access Control
Mbps	Mega Bits per Second
NRE	Non-recurring Expense
OBI	Optical Beat Interference
OE	Optical-Electrical
OMI	Optical Modulation Index
ONU	Optical Network Unit
PON	Passive Optical Network
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
RFoG	Radio Frequency over Glass
R-ONU	RFoG Optical Network Unit

S-CDMA	Synchronous Code Division Multiple Access
SCTE	Society of Cable Telecommunications Engineers
TDD	Time Division Duplex
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
WDD	Wavelength Division Duplex
WDM	Wavelength Division Multiplexing
*x*PON	Any of a family of passive optical network standards (e.g., GPON, GEPON, 10G PON)

[1] Oleh J. Sniezko, RFoG Promise to Support High Reverse Bandwidth (Reverse Bandwidth Bottleneck Elimination), SCTE Canadian Summit 2010

[2] Oleh J. Sniezko, RFoG – How to Make It Work and How to Expand It, SCTE Conference on Emerging Technologies 2009

[3] Radio Frequency over Glass Fiber-to-the-Home Specification, IPS SP 910 Rev 17 Amended, October 19, 2010

[4] Oleh J. Sniezko, RFoG: Matching and Exceeding HFC and PON Performance and Capacity, SCTE Canadian Summit 2011

MAPPING A COST EFFICIENT TRANSITION TO CONVERGED SERVICES: BUSINESS MODELING TOOLS AND OUTCOMES TO GUIDE A PATH TO IP VIDEO

C. Ansley, J. Brooks, E. Lotts, S. Shupe

ARRIS

ABSTRACT

ARRIS has developed a comprehensive cable operator business modeling tool with over 300 input variables that can assist the operator in selecting a low risk, economical transition plan to an IP-enabled converged network. Through tailoring of the inputs and evaluating the various service, network and component choices along with review of their current competitors' offerings, operators can evaluate if they should;

- *First deploy reliable converged in home services or TV anywhere services*
- *Delay or accelerate their IPTV initiatives in response to market pressure*
- *Reclaim bandwidth with DTAs or deploy SDV*

In addition, the model can quantify effects of operational improvements from deploying a single network for all services and potential new revenue streams. The tradeoffs between sophisticated set top box deployments versus enabling increasing amounts of consumer-purchased equipment in their networks can also be evaluated. In a separate paper from ARRIS we will also be looking at the bandwidth transitions and traffic engineering associated with the support of all services during the transition to a converged services network.

INTRODUCTION

Operators today are faced with updating their service offering both in the quality of service and the diversity of service to meet the ever increasing competitive offers from Telco, Satellite and Over the top service providers. Cable Operators have been presented many options to achieve the endpoint of a robust network offering an attractive range of converged services. Each operator also has their own unique network built up over the years and faces unique local market dynamics in terms of competition and consumer expectations. This wide array of variables makes selecting the most economical path pure guess work in the absence of a modeling tool. A modeling tool can significantly reduce the range of variables by identifying the key variables and thereby improve the operators' decision process.

Background on Transition to IP Video Services

Multiple System Operators (MSOs) are beginning to plan and deploy architectures which will ultimately be used to expand the range of Video services delivered to their subscriber base to include video over IP to non-settop boxes within the home and eventually to all video devices. In general, subscribers will experience IP Video as a delivery system that permits them to steer video over their home IP network to any video-enabled device including: IP STBs, connected TVs, handheld devices (like smartphones or tablets), and PCs.

Why the IPTV Transition is Needed

The video entertainment marketplace is changing rapidly. The traditional business model of user content subscriptions combined with leased settop boxes augmented by local ad sales is facing competition. New entrants are competing for both user subscription fees and advertisers' dollars. New consumer electronic equipment such as connected TVs and tablets are changing the consumer's expectations for video services and potentially reducing the MSO's brand presence to just another app on the screen. The average consumer is beginning to explore new ways of accessing video content that only a few years ago were not available.

For CATV MSOs to retain their position as the premier way to access and experience the best video content, the traditional CATV network needs to change. The new devices sweeping the CE market support OTT (over the Top) internet providers at least as well if not better than the MSOs' services. The MSOs need to alter and augment their networks to deliver content to these new devices as well as find ways to expand the offerings they make available directly to compete with the offerings of telcos and satellite providers.

As ARRIS evaluated how to best to support the industry as it confronts these sweeping market changes, we found that the questions posed by this monumental shift in market and technology in the video industry touched many facets of planning and implementation. While most operators' networks share basic technologies and architectures, the actual networks implemented by each operator, and sometimes by each region within larger operators, vary widely. So, a solution evaluated from one set of starting assumptions could be completely invalid for another set of starting assumptions.

We also found that the range of possible transition paths was wide with proprietary and non-proprietary solutions being proposed. Some vendors suggested separate overlay networks, while others proposed transition strategies reshaping current networks in various ways.

Finally, the relative importance of various other factors, such as the coming availability of connected TVs for example, has been hotly debated because there was not a simple way to compare the various scenarios objectively and evaluate their relative merits.

GOALS OF THE MODEL

After analyzing the above problems, we decided to build a model that we could use internally and while working with our customers to efficiently evaluate and understand the tradeoffs between the many paths available to the industry.

The model's goals were:

- Provide a structured set of input factors that could be tailored to represent various network configurations including analog and digital channels, SDV if deployed, and DTAs if deployed;
- Evaluate a transition strategy to IP Video comparing different industry proposals with the default option of no significant change;
- Allow various options to be explored such as adoption of IP devices in lieu of standard STBs, or deployment of streaming solutions to generic IP devices such as tablets;
- Evaluate deployment of Network-based DVR and Network-based Time Shift Buffered TV;
- Allow costs to be calculated and compared for various options, including both capital estimates and operational costs estimates;
- Allow revenue estimations, including subscriptions, and ad revenues, that reflect the latest analysts' projections;

All of the above goals were a formidable set, so some simplifying assumptions were made.

- Operational expenses were at least partially calculated as a percentage of revenues
- Four scenarios were chosen as the most likely:
    - Transition from today directly to an All-IP network,

- Transition from today to an model with IP only in the home,
- Transition from today to mixed model that uses both MPEG and IP in the HFC network and IP in the home,
- and finally a scenario that leaves the current network in place and only adds incremental services over IP.

STRUCTURE OF MODEL

The model was implemented within Excel™ to take advantage of its portability to multiple computing platforms, and facilities for generating graphical output. The model was configured to accept a set of network starting conditions, and a set of assumptions about future network trends. An example of this part of the model is that it can accept a current channel lineup with analog, SD and HD channels delivered by broadcast and/or SDV. The user then selects the behavior of this channel lineup over the simulation period, i.e. will the analog channels decrease each year, will the SD channels increase or remain constant, similarly the HD channels.

	Y-Y Growt	Year 0	Year 1	Year 2
System Bandsplit		750/42	750/42	750/42
Analog Channels		70	70	70
Total Linear SD Program Streams	1%	365	369	373
Total Linear HD Program Streams	2%	200	204	209
Broadcast SD Program Streams	-10%	65	59	54
Broadcast HD Program Streams	-10%	30	27	25
SDV SD Programs		300	310	319
SDV HD Programs		170	177	184
SDV 3DHD Programs	5%	5	6	7

Figure 1 - Portion of Channel Lineup

Also set as inputs to the model are cost estimates for equipment and operations as well as revenue estimates from subscriptions and other sources, such as advertising.

The inputs described above were then applied to the set of scenarios listed above. Each scenario began with the same starting conditions and applied the specified growth curves. Within the model, network bandwidth was evaluated periodically as the growth curves played out and necessary node splits and frequency upgrades were added in to ensure that the scenario remained realistic.

The model provides graphs and charts to communicate the results. Results are generated for each scenario showing capital and operational investments as well as the network bandwidth allocation that is predicted by the model. The model also adds in the estimated revenue figures to reach an estimate of the free cash flow for each option.

IPTV Transition Scenarios Modeled

Four scenarios were selected for inclusion in the model:

- No Change,
- Hybrid Home,
- IP Transitional,
- All-IP.

For all scenarios there are some common parameters, such as the STB replacement rate. The STB replacement rate models the normal churn and breakage of subscriber devices and may also adjusted upwards to model a deliberate increase in STB upgrades to speed IP video. For the examples shown, a gradual replacement rate of 10% per year was used. To ensure fair comparisons, all scenarios also provide the same set of services such as VoD, and DVR, HSD and voice, as well as following the same set of service growth curves.

<u>No Change</u>

The No Change model assumes that the MSO does not actively pursue folding IP technology into their network. The STB replacement activity still runs, but new boxes only provide MPEG4 as an added capability.

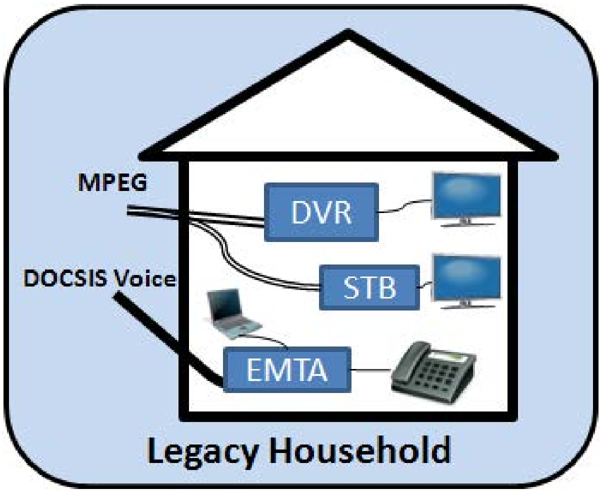


Figure 2 - Legacy Household

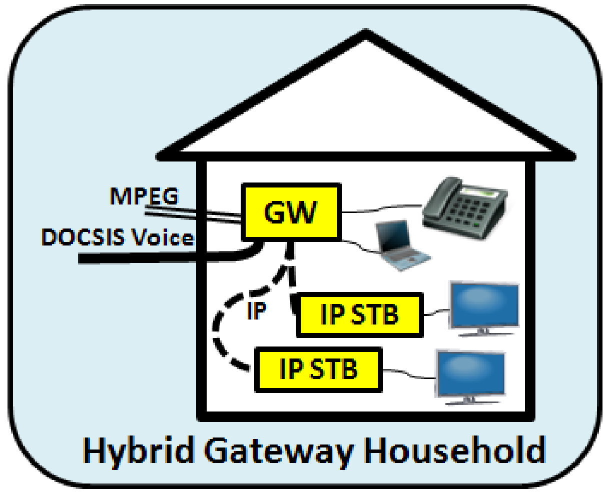


Figure 3 - Hybrid Gateway Household

The model assumes by default that some form of DVR is provided by at least one of the new boxes in the home – so one box is a higher price than the others. Alternatively, it can be set up to model a single type of box deployed in the home.

If SDV is enabled, the model evaluates the probability of simulcasting SDV programs in MPEG2 and MPEG4 and allows MPEG4 simulcast when the predicted channel usage is less for the mixed MPEG4 and MPEG2 program delivery than for video delivery using only MPEG2.

Hybrid Gateway

The Hybrid Gateway model assumes that old STBs are replaced with headless Media Gateways, not capable of directly feeding a television, and IP STBs, but no video traffic is placed over DOCSIS until the transition is complete. The model assumes that the Hybrid GW has a cable card, but no DRM expenses are included until the legacy units are all replaced and traffic moves to DOCSIS. The Media Gateway replaces the cable modem and/or EMTA for the household.

The Hybrid Gateway model also allows MPEG4 simulcast with SDV scenarios when it is more efficient.

IP Transitional

The IP Transitional Model assumes that STBs are replaced with headless Media Gateways and IP STBs. The Media GW also replaces the cable modem and/or EMTA. The Media GW is capable of accessing video program content from either DOCSIS or traditional MPEG channels. These units are deployed with CableCards, but are also capable of supporting IP DRMs. The model moves all Video on Demand traffic to DOCSIS using new CAS/DRM, but assumes that linear program delivery does not move to DOCSIS until the last of the legacy units are removed.

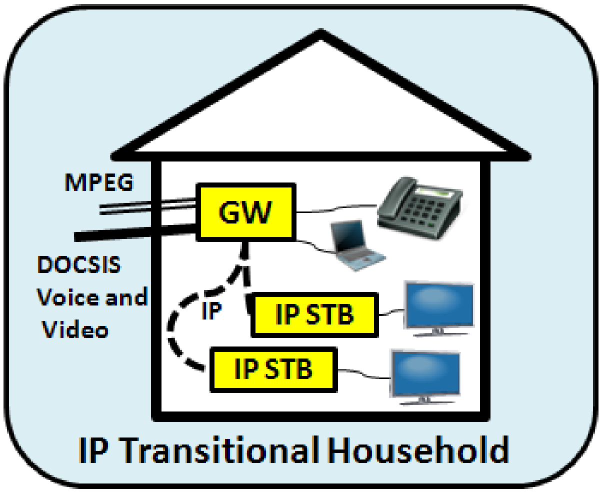


Figure 4 - IP Transitional Household

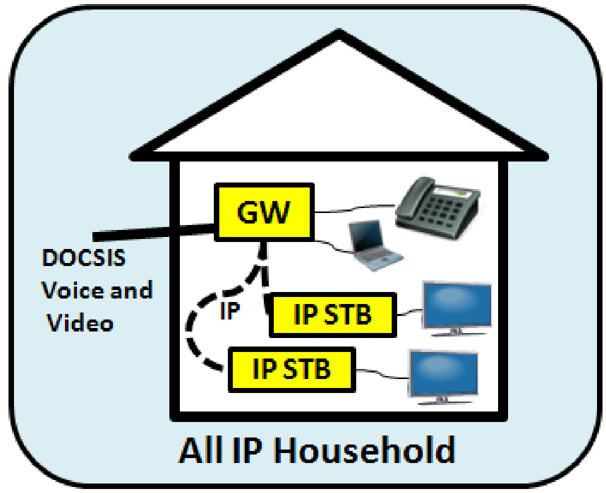


Figure 5 - All-IP Household

The model also assumes the opportunistic use of DOCSIS and MPEG4 when SDV is enabled. The model evaluates the probability of simulcasting SDV programs in MPEG2 and DOCSIS/MPEG4 and allows DOCSIS/MPEG4 simulcast with MPEG2 when the predicted channel usage is less for the mixed MPEG4 and MPEG2 program delivery than for video delivery using only MPEG2.

All-IP

The All-IP model assumes that as STBs are replaced, the MSO provides a new IP Gateway and IP STBs for that household. The IP Gateway is assumed to be headless, not capable of directly feeding a television, and providing HSD and telephone service. The IP GW also replaces the cable modem and EMTA in the household.

All video distribution to All-IP households is over DOCSIS bonded channels. The model assumes that the modem in the GW is capable of handling a large enough bonding group for good bandwidth efficiency. For most simulations, the video bonding group was found to require less than 14 channels. Video on Demand is assumed to be sent unicast, while linear channel viewing that is not N-DVR (Network Digital Video Recording) or N-TSB (Network Time Shift Buffer) is assumed to be multicast.

Since all programs are now being carried over IP, new CAS/DRM costs are added, and CableCards are not included in the IP GW cost, though residual maintenance fees are still included to support the legacy boxes until they are all replaced. The IP GW is headless and replaces the CM and/or EMTA.

DATA FOR MODEL

The data used to construct the model came from many sources. The cost projections for system components were based on our internal estimates as well as an evaluation of the overall marketplace. Similarly the operational expenses were derived partially from other operational studies ARRIS has sponsored as well as from industry publications.

For the revenue estimates, industry studies were consulted to estimate the operational expenses as a percentage of the revenues. Other operational expenses were estimated based on industry studies and internal experience.

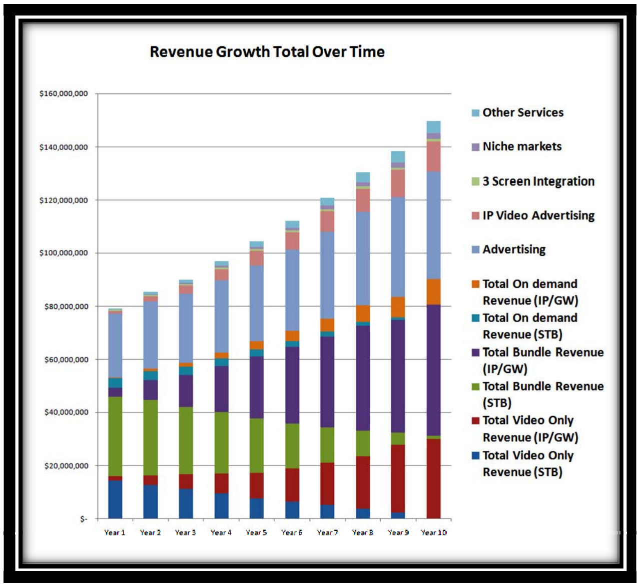


Figure 6 - Example Revenue Projection

The SDV model was developed from a study of actual user channel change behavior. The High Speed Data model is based on historical experience and trends.

The model has been shown to various industry experts for confirmation that its basic assumptions are reasonable and that its predictions are generally in accord with other models.

SOME RESULTS FROM MODEL

Sample analysis

A sample analysis was done as an example for this paper. It uses a starting network that has deployed SDV and has a section of analog channels. The channel lineup is slowly shifting to add more HD channels over the transition period. At the end of the transition period, analog channels are also reduced. The No Change scenario keeps the digital MPEG broadcast channels through the transition, but the other scenarios taper off the MPEG broadcast channels under the assumption that the MSO would be trying to motivate their customers to move to the newer technology.

In all scenarios, the network also supports non-STB devices with unicast streaming over DOCSIS. DVR services are being provided through traditional DVR STBs and Gateways.

The transition path replaces 10% of the STBs each year.

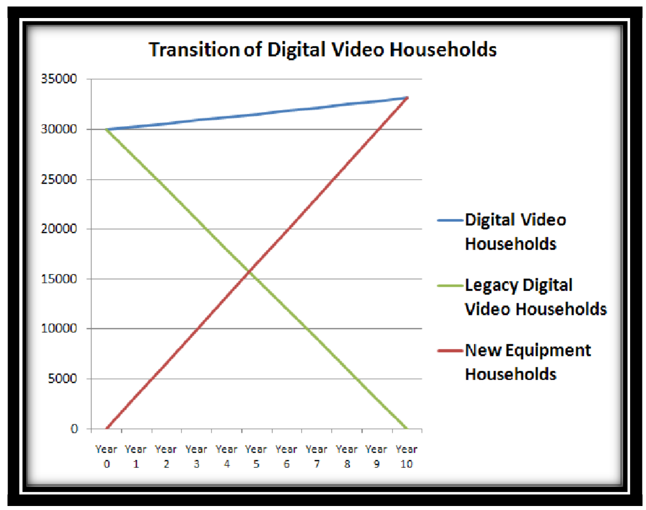
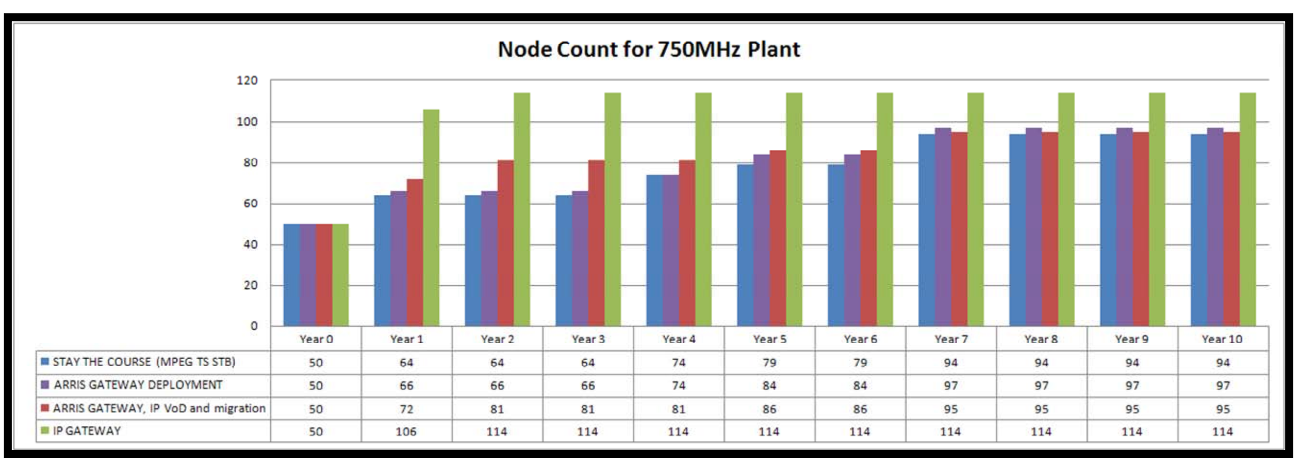


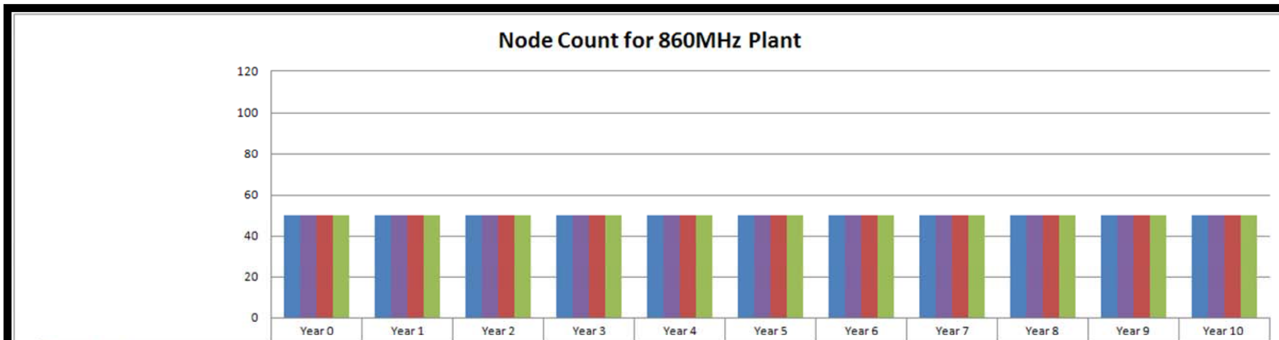
Figure 8 - Transition of Digital Video Households

The model computes bandwidth utilization for each scenario and automatically simulates node splits to ensure that each



	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
STAY THE COURSE (MPEG TS STB)	50	64	64	64	74	79	79	94	94	94	94
ARRIS GATEWAY DEPLOYMENT	50	66	66	66	74	84	84	97	97	97	97
ARRIS GATEWAY, IP VoD and migration	50	72	81	81	81	86	86	95	95	95	95
IP GATEWAY	50	106	114	114	114	114	114	114	114	114	114

Figure 7 - Example Node Calculations - 750MHz Plant



	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
STAY THE COURSE (MPEG TS STB)	50	50	50	50	50	50	50	50	50	50	50
ARRIS GATEWAY DEPLOYMENT	50	50	50	50	50	50	50	50	50	50	50
ARRIS GATEWAY, IP VoD and migration	50	50	50	50	50	50	50	50	50	50	50
IP GATEWAY	50	50	50	50	50	50	50	50	50	50	50

Figure 9 - Example Node Calculations for 860MHz Plant

service group remains within pre-defined channel limitations. The graphs below compare node splits required for a 750MHx plant with an 850 MHz plant with all other variables held constant.

The 860MHz plant obviates the need for any node splits for this example configuration. At first, this may seem counter-intuitive because the number of additional channels provided by an 860MHz upgrade does not double the available channels, while the node splits required in a 750MHz plant equate roughly to a doubling of nodes. The answer understood after considering the next diagram.

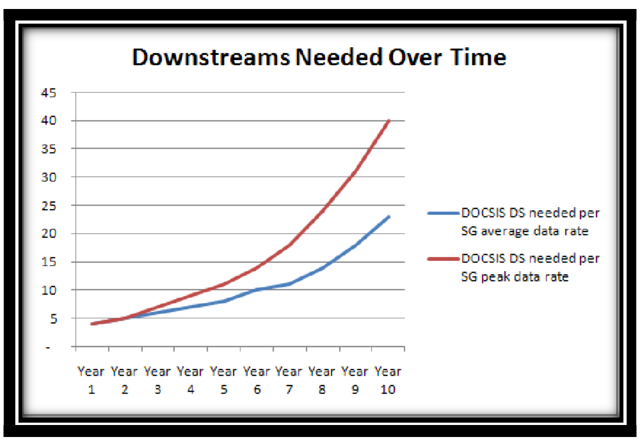
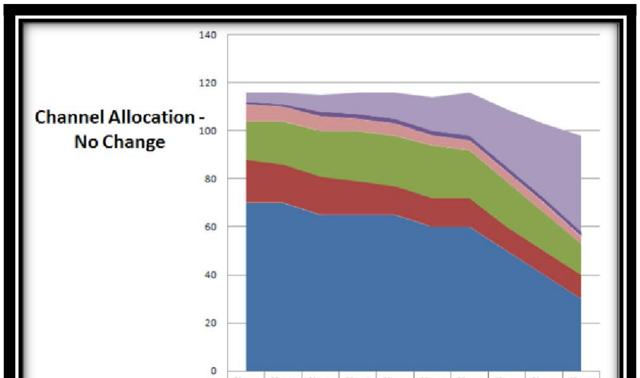


Figure 10 - Comparison of DOCSIS Peak and Average Downstream Requirements

The amount of true narrowcast bandwidth in these networks that varies with the size of the node is quite a small percentage, just SDV and VoD. Broadcast analog and digital video are easily understood as a constant in the bandwidth allocation per node. In most current HFC networks, DOCSIS High Speed Data is considered a narrowcast service with bandwidth expectations that parallel the node size. but Figure 10 shows that expectation is no longer true with projected increases in DOCSIS peak rates and average consumption. From a bandwidth perspective, while the other services scale with node size, DOCSIS HSD becomes invariant to the size of the node as the promised peak speed increases. Even if there are not enough users on that node to fill the DOCSIS downstream on average, the channels are required to ensure that the promised advertised speed is available at any instantaneous time. The headend resources may be shared among several nodes, to make more efficient use of the overall downstream channel group, but each node must have enough bandwidth to satisfy the premium user's speed tests.

The next diagram illustrates the bandwidth usage in a typical node as the transition progresses in a No Change, or Stay the Course, scenario.

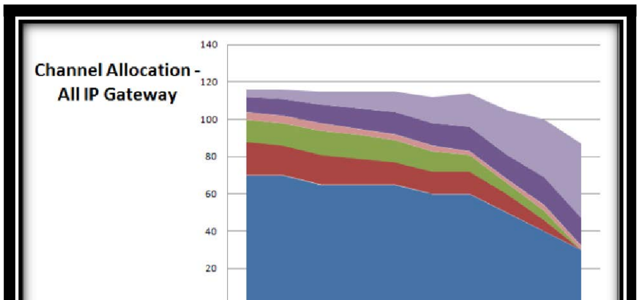


	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Network DVR MPEG2/4	-	-	-	-	-	-	-	-	-	-
DOCSIS HSD	4	5	7	9	11	14	18	24	31	40
DOCSIS IP Video	1	1	2	2	2	2	2	2	2	2
VOD MPEG 2/4	7	6	6	5	5	4	4	4	4	3
SDV MPEG2/4	16	18	19	21	21	22	20	19	16	13
Broadcast MPEG2	18	16	16	14	12	12	12	10	10	10
Analog	70	70	65	65	65	60	60	50	40	30

Figure 11 - Channel Allocation - No Change

The two bottommost wedges show a progression where analog channels decrease with time, but the digital broadcast channels remain, perhaps reflecting renegotiation of analog carriage agreements to migrate into the digital tier, and other channels moving to the SDV tier. The next bar up shows SDV that moves with time from all MPEG2 to mixed codecs then to all MPEG4.

The diagram below provides a view of the All IP scenario from the same starting point as the previous scenario.



	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Network DVR MPEG2/4	-	-	-	-	-	-	-	-	-	-
DOCSIS HSD	4	5	7	9	11	14	18	24	31	40
DOCSIS IP Video	8	9	10	11	12	12	13	13	15	15
VOD MPEG 2/4	4	4	4	3	3	3	2	2	3	2
SDV MPEG2/4	12	12	13	13	12	11	9	6	5	-
Broadcast MPEG2	18	16	16	14	12	12	12	10	6	-
Analog	70	70	65	65	65	60	60	50	40	30

Figure 12 - Channel Allocation - All IP

Notice for both scenarios the DOCSIS HSD bar on the top is the same width due to the Peak rate issue discussed earlier. In the All IP diagram the second band from the top shows the bandwidth used for IP video. Also the third band from the bottom shows the SDV tier which continually shrinks in this scenario as subscribers migrate to IP.

General Observations

The model of the various scenarios yielded results that allow a number of general observations. One general observation is that the exact transition mode deployed does not appreciably affect the overall cash flow analysis. The magnitude of operational expenses with their recurrence every year far outweighs the one-time capital expenditures for any model. The most significant factor coming out of a transition is that the operator is changing their network to accommodate new services that also introduce new revenue streams, or reduce their operational expenses. In this case since a portion of the opex is derived from the revenues, the larger revenue slice is partially offset by increased operational expenses, but the new revenues enabled by IP Video technology, such as increased personalized advertising, can result in a net positive business case.

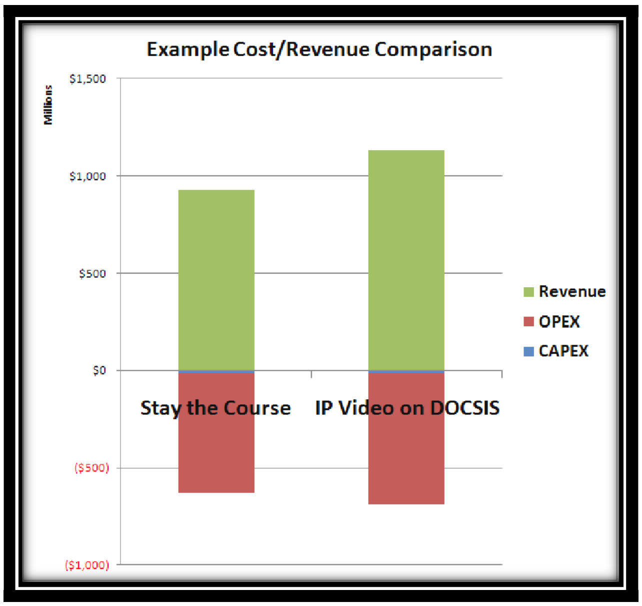


Figure 13 - Example Cost - Revenue Comparison

Another general observation, that in retrospective may seem obvious, is that any cost or income that is calculated on a per subscriber basis will be significant. For example, over 75% of the capital expenditures are CPE related for most models, regardless of the transition path taken simply due to the volume of equipment required. Additional revenue streams from subscribers can be individually small, but collectively can make a real difference in the overall cash flow.

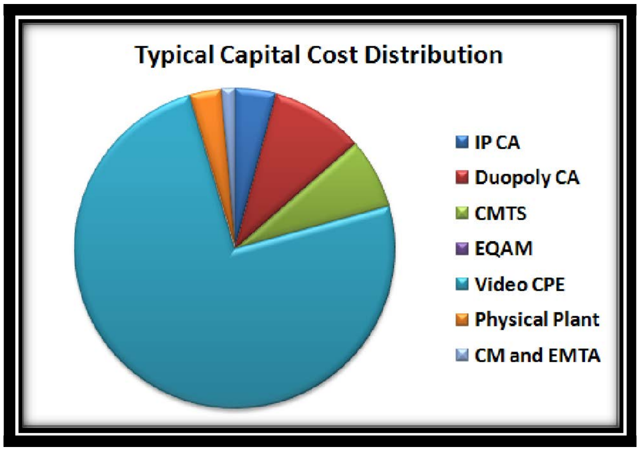


Figure 14 - Example Capital Cost Distribution

Sensitivity analysis for N-DVR

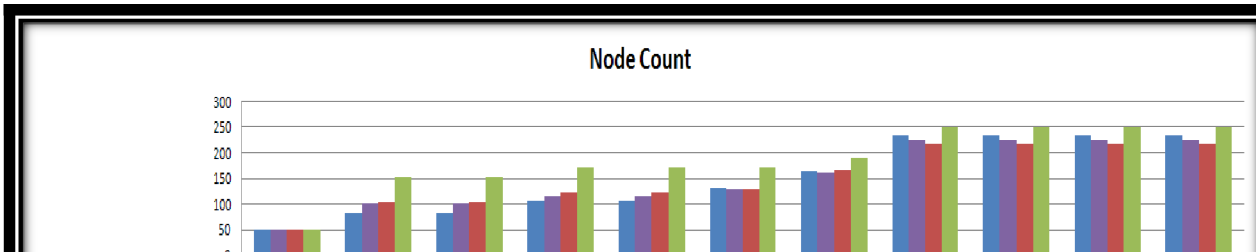
As another example, a network-based DVR deployment was added to the previous example. This technology is under evaluation for both MPEG-based architectures as well as IP-based architectures. This example analysis uses MPEG transport of N-DVR streams for the No Change scenario and the Hybrid Gateway scenario. IP over DOCSIS transport is used for the All-IP and IP transitional scenarios.

There are two possibilities for N-DVR deployment. An MSO could offer these new capabilities to only subscribers receiving new equipment, or they could offer it to all their subscribers. If an MSO has already deployed DVR technology within their network, it would make sense to allow the deployed DVRs to perform their designed function until they have to be replaced. This approach makes efficient use of that sunk capital, using it to allow the bandwidth increases due to N-DVR to grow gradually, minimizing network disruptions. On this theory in this example, newly deployed MPEG STBs or GWs utilize network DVRs while legacy equipment does not.

The N-DVR usage rate in this simulation is taken from ARRIS MOXI experiences regarding the use and frequency of DVR recordings. The bandwidth model has had normal DVR recording rates subtracted from the SDV demand totals, and average playback rates added in as unicast traffic.

The following graph shows the node count changes required to support N-DVR as it gradually rolls out with the STB replacements. The presence of additional unicast traffic forces additional node splits as compared to the model without N-DVR.

While the CPE costs decreased by an average of 14% compared to the original model, that decrease was offset by the increased nodes



	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
STAY THE COURSE (MPEG TS STB)	50	83	83	108	108	131	164	233	233	233	233
ARRIS GATEWAY DEPLOYMENT	50	101	101	115	115	129	162	227	227	227	227
ARRIS GATEWAY, IP VoD and migration	50	105	105	123	123	128	165	217	217	217	217
IP GATEWAY	50	153	153	173	173	173	191	249	249	249	249

Figure 15 - Example Node calculations for N-DVR Deployment

splits and increased headend equipment needed to provide the additional downstream channels. There were no changes made to the operational costs assumed, and perhaps additional saving might be found there, but published information was not found to substantiate that assumption.

CONCLUSIONS

In conclusion, this paper introduces a comprehensive business model built to examine and quantify the many factors that can affect an MSO's evaluation of the plethora of IP Video options. It does not attempt to evaluate every possible option, but concentrates on factors that are frequently under discussion in the industry currently.

ARCHITECTURAL ALTERNATIVES FOR CABLE IP VIDEO

Jeff Brooks, Carol Ansley, Tom Cloonan, Stephen Kraiman, Dan Torbet
ARRIS

ABSTRACT

A transition to IP Video is being seriously considered by many MSOs and since many different architectural approaches are possible, it seems timely to begin analyzing the advantages and disadvantages of many of the architectural approaches to IP Video. This paper will begin by presenting an overview and brief analysis of the many different architectural approaches that are being considered for IP Video Delivery.

INTRODUCTION

Many Multiple System Operators (MSOs) within the Cable Industry are beginning to plan architectures which will ultimately be used for the deployment and delivery of Internet Protocol (IP) Video services to their subscriber base. Different MSOs will move toward IP Video at different times and at different rates, and different MSOs may also choose slightly different architectures as they unveil these new services. However, the authors believe that once the transition to IP Video begins within any particular MSO network, it will likely occur quite rapidly as MSOs work to offer competitive responses to both traditional video service providers and Over-The-Top (OTT) Video content providers.

By offering this new type of IP Video service to their subscribers, most MSOs are hoping to accomplish several important goals, including:

- Gaining access to a broader audience through all 3 screens in the home (TV, PC, and handheld devices)
- Building a direct conduit to the 15-30 year-old demographic (through their handheld devices)
- Creating new means of further monetizing their high-quality video content with new subscription fees
- Providing an opportunity to enter the growing "Internet advertising market" through directed advertising in IP-based videos
- Allowing themselves to become the "organizers" of all IP Video content (MSO-based and Web-based)
- Reducing the high costs traditionally associated with legacy STBs

The basic goal behind any IP Video Architecture is to deliver a common experience for all video services over a managed or unmanaged broadband access network via Internet Protocol to a consumer's TV/PC/handheld device. For MSOs, this broadband access network is typically DOCSIS over Hybrid-Fiber Coax (HFC) plant. However, MSOs would also like the flexibility to offer IP Video services over the Internet so that their subscribers can access the MSO-managed video content even when they are on the road. The former IP Video delivery service (over the HFC plant) is referred to as an "On-Net" IP Video delivery service by many MSOs, implying that the delivery occurs on the MSO's managed HFC network. The latter IP Video delivery service (over the Internet) is called an "Off-Net" IP Video delivery service (a.k.a. TV Everywhere) by many MSOs, implying that the delivery occurs off of the MSO's managed HFC network.

From a subscriber's point of view, the method of delivering these video services is less important than several other attributes, including:

- The ability to route the video data around the home network to any IP-enabled endpoint using Internet Protocol
- The availability of bandwidth capacity permitting them to access all the video content that they want
- The Quality of Service (QoS) mechanisms to ensure that the video content is delivered reliably without pixilation or halts (which can only be guaranteed for On-Net IP Video delivery services)
- The availability of large quantities of high-quality live and on demand video content, both popular and long tail
- The ability to search through the available video content, regardless of source, in a fast, easy-to-use fashion
- The ability to access the video content at any time and in any format required by any desired endpoint device
- The ability to perform trick modes (pause, rewind, and fast-forward, progress bar) on the viewed video content

This paper will begin by presenting an overview and brief analysis of the many different architectural approaches that are being considered for IP Video Delivery. This paper will not address in detail the market drivers, bandwidth calculations or network cost analysis associated with the cable industries transition to IP technologies for the delivery of video services. Two of these topics (bandwidth and cost analysis) are being covered by papers submitted by Tom Cloonan and Carol Ansley of ARRIS.

IP VIDEO TECHNOLOGY MATURITY BY INDUSTRY

Before addressing architecture details, it would be beneficial to discuss some of the factors that will impact decisions to be made. Technology maturity and standards availability are important factors associated with network architecture decisions. All technologies progress through a basic lifecycle from concept to ubiquitous deployment. Some make it through the entire cycle, while others never make it out of the concept stage. It would be an oversimplification to group all IP Video technologies into a single category. The telecommunications, OTT video providers and cable industry are each at a different point in their evaluation and adoption of IP technologies for video delivery. The OTT provider started with IP but do not own the network. This makes each of their approaches to implementing IP technologies for video delivery unique.

The telecommunications companies in North America began IPTV deployments in 1999[1]. Today there are specifications in development or available from ATIS, Open IPTV Forum, ITU-T, and IEEE for Video delivery over an IP network.

The internet or over-the-top video delivery technology has grown through several streaming technologies since its beginnings in the mid 1990's [2]: progressive download, real time streaming protocol, and most recently adaptive bitrate delivery. Much of this technology was proprietary for the initial deployments. This industry now has a number of standards initiatives underway with a number of standards bodies including W3C, IEEE, ISO/ICE, and Ultraviolet.

The cable industry is just beginning to deliver IP Video to subscribers. They have the advantage of leveraging the available standards and experience from these other industries in developing standards through CableLabs to address their specific needs.

Cable specific IP Video Standards activities include;

CableLabs draft release Multimedia Gateway Device Architecture Technical Report **https://www.cablelabs.com/doczone/cross_project_specs/requirements/tech_reports**

CableLabs released OCLA (OnLine Content Access) specifications **http://www.cablelabs.com/specifications/CL-SP-AUTH1.0-I01-101029.pdf**

The TV Everywhere deployments by Cable operators are currently an overlay service that leverages the technologies and suppliers used by the OTT streaming media service providers.

Figure 1 illustrates the typical lifecycle and an estimate of the relative maturity of the technology for each industry.

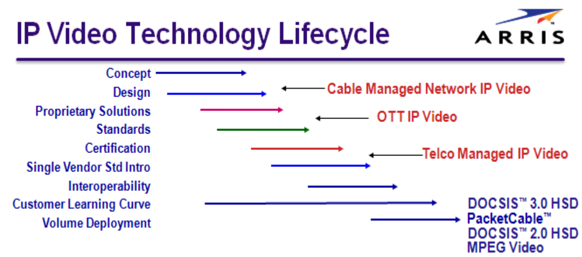


Figure 1: IP Video Technology Lifecycle

TECHNOLOGY TRENDS AND CLASSIFICATIONS AS A GUIDELINE FOR ARCHITECTURAL CHOICES

Before a discussion on the architectural choices and their associated merits, it is important to establish a definition of the various video delivery types and some high level measures to guide the operator in the decision process. These guidelines should be similar to an organization's core values. The guidelines should be non-specific, however if operators stay true to these guiding principles, the end objective will be met.

Proposed Architectural Guiding Principles

1. Open standard solutions are highly preferred to proprietary implementations.

   Open standards allow greater participation by suppliers. This in turn creates competition to reduce costs. Open standards also eliminate the risk of being stranded with a solution from a supplier that fails.

2. Internet Protocol (IP) technologies are preferred over regional or industry segment technologies.

   Selection of IP technologies allow the service provider to take advantage of new developments generated across multiple industry segments versus only those from their single industry segment.

3. Converged networks are preferred when compared to networks segmented by service type or distribution technology.

   Converged networks yield efficiency in utilization of resources throughout a service provider's business. Capital is more efficiently utilized; the statistical gains of convergence increase the utilization of equipment used for processing, storage, and distribution of content. Operations costs are contained; the simplification of the delivery network reduces the variety of equipment needed with cost savings available in spares, service agreements, and support personnel. Most importantly, maintenance and troubleshooting are simplified since a complete view of the network and subscriber is available when a service call is required.

4. Cloud based (centralized) networks are preferred over client based implementations. Much like the statistical gains of a converged network, cloud based systems can reduce the number of devices in the network. However, and possibly more importantly, cloud based networks can reduce the cost and complexity of the end points. For instance, an Operator can put an edge cache device in the network or use Network DVR to reduce the size and type of memory needed in a CPE device. This reduction in complexity of the end point can aid in speeding the deployment of new services without the need to update each CPE device individually. This may in turn extend the useful life of CPE devices capable of operating in this environment.

These four guiding principles will need to be balanced against the realities of today's technology availability, network reach and organizational structure. These realities may present a good reason for not immediately achieving the optimum solution. However, when used as a vision to guide the implementation, the network approaches the optimum over time.

CLASSIFICATION OF VIDEO DELIVERY TYPE

It may be useful to establish a taxonomy of video delivery technologies to measure against the above guiding principles. Figure 2 was developed by ARRIS to assist in the development of an IP video transition plan.

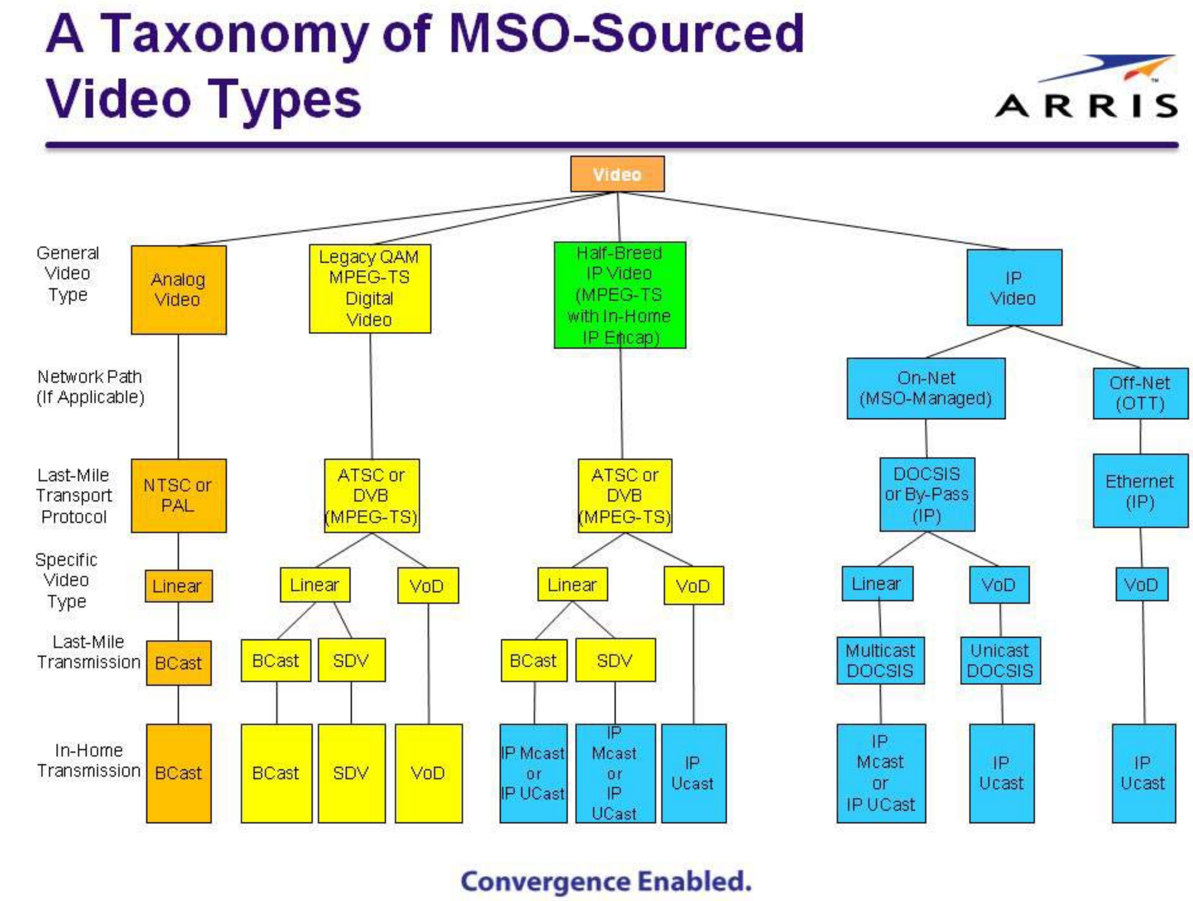


Figure 2: Taxonomy of MSO-Sourced Video Types

The far left side of Figure 2 illustrates analog video. Analog distribution has been in existence for over 50 years and arguably will be around for a few more years. The industry developed digital video delivery to increase the capacity of the infrastructure and improve the security of the content. Video On Demand (VOD) was also developed to extend the service offering. More recently some operators have implemented Switched Digital Video (SDV) to further expand the capacity of the network. For purposes of this discussion, the section on the far left of Figure 2, terminating in the 4 boxes at the bottom left will be named "Digital MPEG Distribution".

While the cable industry was making these service improvements, a host of organizations were developing the technology on the far right of the diagram to deliver video over unmanaged network connections. This will be referred to as "IP Unicast Distribution."

Some cable operators have begun to use IP Unicast Distribution to delivery video to devices other than the TV. This divergence from traditional delivery methods illustrates why it becomes important to look at the long term guiding principles. Do operators simply layer on the new network or do they begin to map a path similar to the center section of the diagram, "Hybrid Distribution", which leads to a converged network? A review against the guiding principles suggests operators should map a path to a "Convereged IP Distribution" network.

OVER-THE-TOP / STREAMING TECHNOLOGIES

In order to develop a plan to convert the networks to an IP transport, we must first understand the differences between the old and new technologies. There are three basic types of internet video/streaming technologies.

- Progressive download - Very robust to network impairments - Relatively long wait before start period - has been used in the past for OTT video delivery.

- Real Time Streaming Protocol (RTSP) with UDP video transport – Very susceptible to network impairments - Little wait time to start - used in today's cable networks to deliver MPEG-TS video to Edge QAMs for transmission.

- Adaptive Streaming (AS) technologies - Robust to network impairments - Some wait time to start - newer streaming technology used to provide additional robustness on congested networks and underpowered hosts.

Much of the cable industry still has analog video in the network. The Internet and mobile industry will support multiple types of streaming for some transition period. However, the race by Microsoft, Adobe, Apple [3, 4, 5, 6, 7, and 8] and others to enhance their adaptive streaming solutions suggest cable should concentrate on the new growing technology.

There are a number of significant differences between today's digital MPEG video distribution technologies and the newer adaptive streaming video technology. There are also similarities from an abstract perspective.

Starting with the similarities: a VOD system and a Content Delivery Network (CDN) perform the same basic network operation of managing servers and the content on those servers to maximize the subscribers' experience and minimize network resources required for distribution. One of the greatest differences, and most difficult paradigms to break, is the difference between MPEG-TS technology and Adaptive Streaming technology. This is partially due to naming conventions.

The Digital MPEG Distribution system uses UDP transport in the routed network and then modulates each stream and places them into an RF carrier over the HFC network and all "streams" are transmitted to the home.

Adaptive Streaming is not actually streaming but is TCP file transfer. A video stream is packaged into small play time fragments. These fragments are then delivered when the client sends a simple "HTTP GET" command. Each fragment of the video sent may also be cached in the network. The repeated requests by the first user, in effect, distribute the content within the network's caches and when the next client requests the same content, the cache closest to the subscriber holding this content may respond to that request. This implementation only transmits the stream that is requested by the home

The above Adaptive Streaming process impacts the processing, storage and delivery resources required in the network. The processing at ingest is much more than the current one MPEG-TS stream for HD and one for SD. It may require as many as 12 different bit rates for each format to be distributed. The encoded content is then packaged based on the devices to be supported and delivered to an origin server. The Tables below provide one example of this process to size this effort for a small cable head end for just the linear programs.

Network Assumptions

- 300 live channels
- MPEG4 encoding (H264)
- HD bit rates; 5.0, 3.0, 2.0, 1.0, 0.5 Mbps
- SD bit rates; 1.5, 1.0, 0.5 Mbps
- 50% HD / 50% SD mix
- Device / Format Supported
    - MPEG 4 Cable STB
        i. HDTV
        ii. SDTV
    - Apple HLS
    - Adobe HTTP Dynamic
- Rewind/ Start over window of 15 minutes

Legacy processing for Delivery to MPEG STBs	(300 channels x (50% HD) x (5 Mbps)) + (300 channels x (1.5 Mbps)) = 975 Mbps
Output from video packaging process	2streaming formats x ((300 channels x (50% HD) x 5 bit rates(5.0, 3.0, 2.0, 1.0, 0.5 Mbps)) + (300 channels x (50% SD) x 4 bit rates(1.5, 1.0, 0.5 Mbps))) = 2 x (1755+450) Mbps = 4,350 Mbps
Objects Stored in 15 Minutes	15 min x 60 sec/min / 2 seconds per fragment Apple HLS x (150 channels HD x 5 bitrates / HD + 150 channels SD x 3 bitrates SD) + 15 min x 60 sec/min / 10 seconds per fragment Adobe HTTP Dynamic Streaming x (150 channels HD x 5 bitrates / HD + 150 channels SD x 3 bitrates SD) = 540,000+108,000 = 648,000
Origin/Caching Server Capacity	4350 Mbps x 15 min x 60 sec/min / 8 bit/byte / 1024 = 478 GB

Figure 3: Sample Calculations

	Storage Server (MPEG System) Serving only SD & HD TVs	Origin Server (Adaptive System) Serving HD/SD TV's and mobile devices
Encoder output (Mbps)	975	2,175
Packager Output (Mbps)	n/a	4,350
Objects Saved	300	648,000
Server Capacity Required (GB)	107	478

Table 2: Capacity Requirements

The content is now available for IP Unicast Delivery over any internet connection. A sample calculation to understand how this new service might consume network resources in the cable operators' DOCSIS/HFC is provided only to help the reader understand the scale. A detailed analysis of bandwidth is addressed in the paper entitled "**Architectural Approaches to Help Circumvent the "Simulcast Roadblock" of IP Video Deployments"** authored by T. Cloonan, J. Allen, C. Ansley, R. Arnold, C. Cheevers, T. Cotter, J. Howe, B. Hanks, D. Torbet, & I. Wheelock of ARRIS.

Delivery to portable device Analysis

Network Assumptions

- It is desirable to deliver the best quality video (highest bitrate) to the subscriber. The network should be robust enough that the client shouldn't have to request lower quality segments due to congestion.
- 3.0 mbps to a tablet/pc screen
- 500 Home Passed per Node
- 8 DOCSIS channels per node
- 15% penetration of subscribers
- 25% concurrent use at peak usage hours

Bandwidth Utilization Calculation	(500 HHP x .15 x .25 x 3.0 Mbps) / (8 x ~ 40 Mbps) x 100% = ~ 56 / 320 = ~ 18% of the available DOCSIS capacity

Table 1: Delivery to "Not-The-TV" Analysis Bandwidth Utilization Calculation

If an extreme situation is modeled, the results of the analysis are even more dramatic.

Delivery to HDTV Analysis

Assumptions

- It is desirable to deliver the best quality video (highest bitrate) to the subscriber.
- 8 mbps to an HDTV
- 500 Home Passed per Node
- 8 DOCSIS channels per node
- 60% penetration of subscribers
- 50% concurrent use at peak usage hours

Bandwidth Utilization Calculation	(500 HHP x .6 x .5 x 8 Mbps) / (8 x ~ 40 Mbps) x 100% = 1200 / ~ 320 = ~ 375% of the available DOCSIS capacity

Table 2: Delivery to HDTV Analysis Bandwidth Utilization Calculation

While there will be some statistical gain associated with a large deployment, it will not be enough to fit within an eight channel DOCSIS band. The existing digital network cannot supply the services desired by the subscribers without substantial changes and the existing headend network architecture does not support full adoption of OTT delivery.

THE TRANSITION

An implementation of an overlay architecture for TV Everywhere services results in the addition of a number of new functions/elements in the network and the duplication of some functions due to technical differences. Figure 1 illustrates a Digital MPEG Distribution network that has had an IP Unicast Distribution overlay.

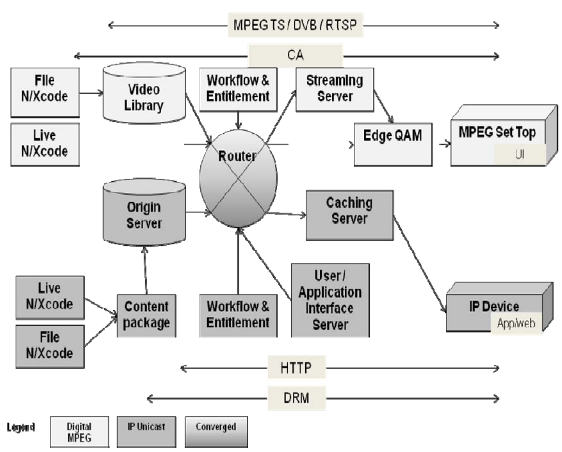


Figure 4: Digital MPEG Distribution network that has had an IP Unicast Distribution overlay

CONSOLIDATION OF WORKFLOW MANAGEMENT

One of the first steps an operator can take toward assuring a smooth transition to IP video services is the implementation of a highly automated asset management system that can serve as a single point of control over content ingestion, processing and distribution. Such a system must, at a minimum, be able to leverage existing operational assets, including back-office components, storage centers, advertising management systems, metadata repositories, and policy servers, while providing the means for operators to efficiently provide multi-screen IP services.

Linear and VOD already require multiple viewing rules along with various modes of capture, distribution and advertising associated with time-shifted content to establish a holistic management approach. Metadata, interactive applications and advanced advertising must be managed across linear and on-demand outlets. TV Everywhere has added still another sphere of operations. With the onset of IP migration, the range of processes to be managed is even greater.

Operators must be able to efficiently manage usage and metadata policies, subscriber and device authentication, multiple encryption, transcoding, streaming, advertising formats, and other processes unique to each IP content category and each IP device. And they must be able to manage all processes essential to assuring accurate billing and payment disbursements to third-party suppliers and advertisers.

These processes start with the ability to maintain an inventory of all the digital copyrights associated with all the assets and to make sure those policies are accurately embodied in the Digital Rights Management System that assigns specific rights to a particular content element going out to a particular outlet. The workflow system must be able to collect and transfer the usage data to the back-office systems that confirm policy enforcement, perform billing processes and orchestrate payments on all contracts.

To do this, the operator has a number of choices including maintaining separate systems, expanding the current MPEG system to address IP deliveries, or expanding the current IP system to address the MPEG requirements, or developing and adopting a new system that meets all of the guiding principles.

The tools and processes in use today for MPEG TS /DVB are not adequate for IP delivery and the systems for IP delivery are not adequate for MPEG TS streams so you are forced into a world where you have to manage the two workflows separately for some period of time during the transition

Alternatively an operator could build application programming interfaces (API) between one of the existing systems and the rest of the network to achieve convergence. Evaluation of each of the existing MPEG and IP systems based on the guiding principles suggests augmenting the IP Unicast Distribution systems to address this network requirement.

CONTENT PROCESSING

The existing content processing requirements for the Digital MPEG Distribution network are simple when compared to the content processing requirements of a network capable of OTT service. A review of processing required, by use cases, will assist in illustrating the processing necessary for each type of network distribution.

The Digital MPEG distribution network must ingest and process "live" content and prepare it for delivery via MPEG Transport Stream (MPEG-TS) for both SD and HD Television consumption. This process was calculated in Table 2. Some operators also ingest this live content and process it for file storage on a library server for use in start over and look back type of service. In addition to the live content ingest, Digital MPEG distribution networks ingest mezzanine files and create compressed files for storage to a library server for delivery of VOD service.

Table 5 depicts the input and output video processing requirements for Digital MPEG. Two (2) types of input Live or Files and six (6) types of output. Two (2) output types for files and four (4) types of output for live content.

Inputs	Outputs			
Live	1.5 Mbps SD MPEG-TS	8 Mbps HD MPEG-TS	1.5 Mbps SD File	8 Mbps HD File
Mezanine File			SD File	HD File

Table 3: Input and output video processing requirements for Digital MPEG

The IP Unicast distribution network must ingest and process "live" content and prepare it for AS delivery for both SD and HD Television consumption and the myriad of other devices an operator chooses to support. Example calculations of this process are shown in Table 6. The first review of this process may appear to simplify the processing as the content is only processed to file. However, the multiple bitrates and formats complicate the output for live content processing. Similar to the MPEG processing above the IP Unicast Distribution network may also ingest mezzanine files and create compressed files for storage to an origin server for delivery of VOD service. In addition to the encoding of the material, the IP Unicast Distribution network requires packaging for each format supported.

Table 6 depicts the input and output video processing requirements for IP Unicast. Two (2) types of input Live or Files and twenty-seven (24) types of output for each live and file input in this example.

Inputs	Outputs		
Live	3 SD bitrates	5 HD bitrates	3 packaging formats
Mezanine File	3 SD bitrates	5 HD bitrates	3 packaging formats

Table 4: Input and output video processing requirements for IP Unicast

Operators must evaluate their current MPEG encoding systems used for live and file ingest to determine if they are capable of the increase in output streams, output resolution and encoding formats. Operators should also consider how they want to group encoding and packaging in the converged network of the future. While it may be possible to serve the IP unicast delivery and the MPEG delivery simultaneously from the encoder output, network transition and component utilization may lead to maintaining separate resources for at least the live feeds processing for the legacy MPEG set tops and the emerging IP Unicast delivery clients.

Operators that have deployed encoders for file processing of IP Unicast Distribution must evaluate the current systems ability to support the real time requirements for live feed processing.

Figure 5 depicts a network that separates the encoding and packaging functions thus allowing the potential for one encoder to feed both the MPEG distribution and IP Unicast distribution requirements. This diagram could have easily been draw with segmented functionality.

CONTENT DELIVERY NETWORK ARCHITECTURE

Operators may be able to rely on the traditional Internet Content Delivery Network (CDN) infrastructure to accommodate the performance requirements of their managed IP video services. However, it is important they also plan their CDN architectures to accommodate legacy VOD service during the transition period.

The new Internet CDN architecture represents a significant departure from the traditional approaches to VOD content distribution. Most CDN solutions, based on commercial off-the-shelf servers and network-attached storage devices, rely on storage sub-systems where CPU speeds and processing modes make it impossible to satisfy new cable operations requirements by merely adding more storage and server capacity. Moreover, there are hidden costs that are incurred as capacity on these traditional solutions is expanded, which requires more load balancers and ESRs (Ethernet switch/routers) along with increases in power consumption and space allocations.

Traditional VOD systems may not yet support adaptive streaming technologies. New scalable CDN solution architectures are required to serve all on-demand content distribution requirements, whether content is ultimately distributed to end users via MPEG or IP transport streams. These new CDNs must fully leverage the capabilities of the cable-managed IP network, affording operators complete flexibility to ingest, store and deliver growing volumes of contents to all types of devices.

Sophisticated simulation methods can be used to determine optimal architectural set-ups for any given operating environment. With expert assistance in this critical design arena, operators will find they can implement CDN architectures that greatly out-perform and out-scale traditional HTTP (Hypertext Transfer) -based approaches.

The above evaluation clearly indicates a move to CDN architectures. Operators that have VoD systems that can support the new functionality of adaptive streaming should look to upgrade instead of wholesale replacement of their installed base.

SECURING CONTENT THROUGH EFFICIENT USE OF DRM TECHNOLOGY

Adding to the complexities of mounting a multi-device premium TV service is the fact that the client software associated with various streaming systems supports different DRM systems. Fortunately, a growing number of backend content management systems are designed to help automate the implementation of different Adaptive Streaming CODEC and DRM combinations. The fact that transcoders as well as other functional components are now available as software engines running on Common off the Shelf (COTS) hardware helps immensely with the flexibility required to keep pace with the changing environment.

There are at least three ways to structure the protection system under discussion within cable industry circles. One is the end-to-end approach, where device-specific DRM encryption is applied at the headend and content is decrypted at the client device.

The other two entail encryption of all traffic using either a single DRM mode or the CMTS-supplied DOCSIS BPI+ protection system between the headend and the home media gateway, where video is decrypted at the gateway, then re-encrypted for distribution over the home network. Typically, these strategies anticipate employing the DTCP-IP (Digital Transmission Content Protection over IP) protocol in conjunction with the use of DLNA (Digital Living Network Alliance) home networking standard to discover and connect devices.

In addition to the three protection schemes used for IP Unicast Distribution, the operator must maintain the current Conditional Access (CA), link layer, protection for the legacy Digital MPEG Distribution portion of their network. The use of CA is an important consideration in Hybrid Distribution. The CA must be terminated somewhere and the content protected in some other acceptable mode for delivery to devices other than MPEG Distribution Set Top Boxes. The options are to bridge the protection to a DRM or terminate the CA in a gateway at the home and use DCTP-IP for protection over the in home network.

It may be required to support multiple content protection schemes simultaneously, prior to standardization of content protection schemes. It may be possible to use content protection frameworks or API's to minimize the complexity of the client and its interface to the network when this is required.

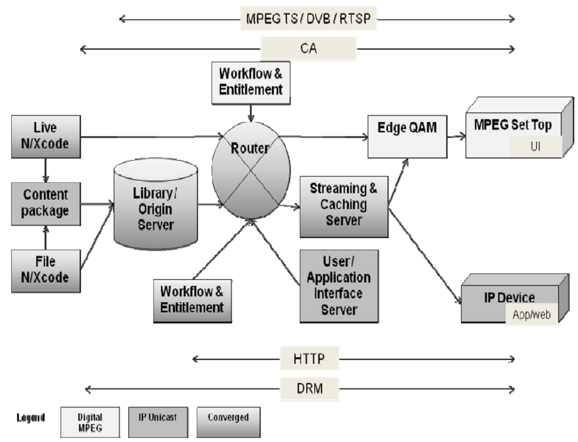


Figure 5: network that has converged the content process, storage, streaming, caching and developed the IP Unicast workflow and entitlement system to interface with the Digital MPEG workflow and entitlement system

Figure 5 illustrates a network that has converged the content process, storage, streaming, caching functions and developed the IP Unicast workflow and entitlement system to interface with the Digital MPEG workflow and entitlement system. This step can facilitate the transition to either a Hybrid Media Gateway or an IP only Media Gateway.

MEDIA GATEWAYS AS A TOOL FOR IP SERVICE MIGRATION

Remembering that neither the IP Unicast Distribution network nor the Digital MPEG Distribution network can meet all demands of today's services, the home media gateway has emerged as a linchpin in IP cable TV migration. It is a major point of intelligence and distribution that can be used to support all video delivery technologies during the migration period.

The gateway serves as the termination point of the operator's network and the primary interface for the home network. It serves as the source for all content entering the home, including legacy MPEG-2 as well as IP video. As a DOCSIS 3.0 sub-system with an embedded modem supporting high-capacity channel bonding, the gateway may also serve as the interface for IP telephony and high-speed data. The gateway is one of the first steps to implement a Hybrid Distribution network.

A gateway can support interfaces to the various home networking platforms required to distribute content to devices in the home, including MoCA, Wi-Fi 802.11n, Ethernet and other standardized physical interfaces suited to premium video transport. The open DLNA protocol stack can support client discovery and connection of all consumer devices. The gateway must be able to interact with all device clients in support of functions such as channel selection, multicast joins, authentication and authorization, distribution of encryption keys, management of AS flow rates and DOCSIS QoS.

Operators must review the functions and features being considered for gateways using the same principles for determining how to transition other elements of the network. Gateways and supporting middleware systems must be flexibly designed to fit virtually any architecture and may need to support:

- IP encapsulation of MPEG transport streams for distributing content to IP devices in the home
- server and proxy functionalities essential to running a common advanced user interface across all devices
- multi-room DVR;
- whole-home media management;
- content protection systems conversions

- conversion of multicast streams to unicast streams
- interactive applications for TVs and companion devices
- dynamic ad insertion
- transcoding of streams to multiple device formats
- value-add services such as energy management and home security

The operators' ability to leverage MPEG and DOCSIS networks and the web based software platform to introduce IP services of all types into the integrated navigational experience represents not only a value-add incentive to introduce gateways, it also allows the transition of the user interface to a cloud based service. Operators can address changing market conditions by offering a branded over-the-top service featuring certain aggregators' offerings, or they can go further by providing consumers a wide-open approach to accessing Web content through the cable branded unified User Interface (UI).

Advanced caching intelligence is another potentially important attribute for gateways. Intelligent algorithms can be employed to manage the video content from the network. Real-time selection of multi-screen streams for gateway storage based on content popularity and local usage patterns can serve to improve channel-change performance and increase bandwidth efficiency. This would present the gateway as an extension of the operator's CDN.

The detailed analysis of each choice within the gateway and the impact to the network architecture is not addressed in this paper, as the author believes the topic should be addressed at some length in a separate stand-alone technical paper.

Figure 6 builds upon Figure 5 with the addition of the Hybrid Distribution Network gateway.

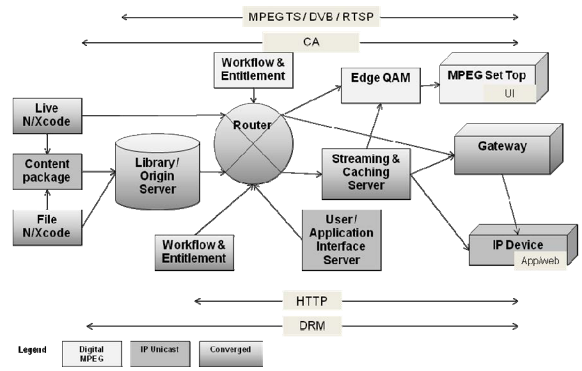


Figure 6: builds upon Figure 5 with the addition of the Hybrid Distribution Network gateway

CONCLUSIONS

This paper has provided a classification of three existing and emerging video distribution technologies. The paper then provided an example comparison of the two corner case distribution methods; Digital MPEG and IP Unicast to a) expose their technical differences and b) propose that neither architecture is capable of meeting all of the operator's requirements for delivery of video services. A framework of guiding principles was then used in the comparison of the various distribution technologies and to evaluate the major functional components of each of the video distribution technologies in an attempt to assist the operator in developing a strategic path to transitioning (or not) their networks to a converged all IP distribution architecture.

Using the distribution technology classifications and guiding principles the authors believe, operators will find it difficult to meet the emerging subscriber service demands through the simple implementation of new technologies on the existing infrastructure. The adoption of the existing solutions for these new services could easily result in a new network overlay. The addition of a new overlay would increase the operators' cost and the complexity of service delivery and support. This outcome is contrary to the ultimate objective.

During the transitional phase operators must take care in the selection of new components that do not limit their ability to achieve their guiding principles. They must also engage with suppliers of their current network elements to determine the ability to upgrade current components to meet new requirements. The use of the guiding principles should allow progress while minimizing the risk of limited flexibility.

REFERENCES

[1] BCE Jan 24, 2000 Press Release "NBTel leading the way in North America with Aliant's new interactive information and entertainment television service – VibeVision"

[2]FundingUniverse.com http://www.fundinguniverse.com/company-histories/RealNetworks-Inc-Company-History.html

[3] REDMOND, Wash., and SEATTLE — April 8, 2011 Press Release "Microsoft and thePlatform Enable Delivery of Protected, High-Quality, Premium Online Video"

[4] REDMOND, Wash., and CAMBRIDGE, Mass. — Oct. 28, 2008 Press Release "Microsoft and Akamai Innovate on Consumer Video Experiences Using Silverlight"

[5] LAS VEGAS — April 20, 2009 Press Release "Microsoft Smooth Streaming Provides True High-Definition Video Delivery"

[6] SAN JOSE, Calif. — March 23, 2011 Press Release "Adobe Pass for TV Everywhere: New Premium Video Authentication Solution Now Available"

[7] SAN JOSE, Calif. and SEATTLE, Wash. — Jan. 25, 2011 Press Release "Adobe and thePlatform Announce Alliance to Advance the Delivery of Premium Online Video Content

[8] IETF R. Pantos, Ed., W. May

, Apple inc., November 19, 2010, HTTP Live Streaming, draft-pantos-http-live-streaming-05

NEXT GENERATION VIDEO INFRASTRUCTURE: MEDIA DATA CENTER ARCHITECTURE

Gene Cannella, Cisco Systems
R. Wayne Ogozaly, Cisco Systems

Abstract

Service providers seek to deploy next-generation interactive, immersive, and profitable video services that reach any device with any content, anytime, anywhere. This requires a managed evolution to a much more powerful, scalable, end-to-end video infrastructure. Here we present a Media Data Center architecture for a scalable, video-aware delivery model that offers virtualization of services and subscriber applications, employs content delivery networks and adaptive bit rate methods, unifies support for MPEG and Web protocols, employs a cloud-based network design, and leverages video intelligence throughout the network, while interworking with legacy MPEG TV systems.

THE NEED FOR A MEDIA DATA CENTER

Video service providers are seeing competitive pressure to deliver the next wave of video services to more devices, in more formats, across a national footprint, and to millions of subscribers. These next-gen video services are more interactive and involve a wide range of subscriber applications, widgets, and streaming techniques that are more closely aligned with Internet protocols than they are with traditional MPEG TV systems. In many ways these new applications are very disruptive, because they impose a new set of operational requirements and a unified architecture that can handle both MPEG and Web-based protocols. Traditional TV and Internet services are converging and the video infrastructure must also evolve to support a new set of “TV Everywhere” services and cloud-based applications.

"TV Everywhere" services deliver premium entertainment to PCs, tablets, and mobile devices. Adaptive Bit Rate (ABR) streaming technologies such as Apple HTTP Live Streaming and Microsoft Silverlight Smooth Streaming are typically used to reach these additional Internet devices. This technology disrupts existing service provider infrastructure by creating dramatically increased compute and storage demand, multiplying file counts and file management requirements with fragmented file formats, and by increasing streaming load with an inherently unicast delivery model. The process to convert live content or on-demand titles into various ABR formats is very compute and storage intensive. ABR workflows typically include a number of processing stages, including acquiring content, transrating content into multiple streaming profiles at different bit rates, encapsulating each profile into the appropriate ABR formats, applying Digital Rights Management (DRM), and publishing the formatted content to origin servers. Storage and processing requirements can multiply as service providers begin to support several ABR formats, numerous streaming profiles, and diverse DRM requirements

In many cases, service providers are required to deploy more infrastructure to handle each of these new video service requirements. With each new service, operations and systems become more complex, channel lineups and VoD loads increase, and client types proliferate. Typically, infrastructure to support new services has been overlaid on legacy facilities and managed as individual service-delivery islands. Each video service requires new dedicated resources, network capacity, and drives increased operational costs. Replacing this siloed infrastructure approach with a more unified approach is advocated to evolve the traditional headend into a more powerful delivery system, namely the "Media Data Center."

FOUNDATIONAL TENETS OF THE NEXT-GENERATION MEDIA DATA CENTER

Support for a Multiservice Environment

The Media Data Center design should support continued scaling of traditional MPEG video services (linear broadcast, switched video, VoD, IPTV, etc.), but must also now support a broader set of Internet-based streaming and user applications. The design should support delivery of these services over any access network (cable, DSL, FTTx, mobile) to multiple end device types. For even if a service provider operates only one type of access network, the provider may offer its subscribers the ability to access their content as they move into or through other providers' access networks.

IP Early Content Acquisition

The Media Data Center design should provide a common content-acquisition process that moves video streams into an IP encapsulation early in the acquisition process. These encapsulated streams can then be shared across all service modes (MPEG, ABR, progressive download, etc.) with fewer ingest points and improved resiliency.

Service-Independent Scaling

Each video service, whether it be a legacy service such as linear broadcast, switched digital broadcast, and video on demand, or, one of the emerging next generation services such as IPTV, multi-resolution streaming, or cloud enabled subscriber applications, should independently accommodate scaling of infrastructure capacity. Existing services should not be disrupted when capacity is added, new services are deployed, or upgrade and maintenance operations are performed.

Cloud-based Network Infrastructure

The Media Data Center should support network assisted video delivery, proximity, load balancing, and cloud-based service orchestration. The design should optimize resiliency and content delivery over a distributed cloud enabled network. Video resources can then be managed, scaled, and secured independently using proven data center cloud innovations.

Virtualization of Content, Services, and Resources

Data center efficiencies are derived from the virtualization of servers, storage, and unified network elements. Content is abstracted, transcoded into different formats, and stored in multiple locations, for use across many devices.

Fault Containment and Resiliency

Video control plane and data plane faults should be contained to the Media Data Center. Fault identification and repair mechanisms are inherent to the design. Redundant data center locations should be used to provide high availability across a national and regional footprint.

Consolidation of Operations and Management

The consolidation of operations and management is achieved through the use of a unified computing platform and converged network technologies. Stream visibility and MPEG quality monitoring is provided across all services at each resource.

ENTERPRISE DATA CENTER DESIGN AS A MODEL

Data center computing, storage, and networking technologies are undergoing rapid change. Virtualization practices are creating a more scalable and secure data center. Massive remote storage, migration to 10-Gigabit Ethernet for edge connectivity, and higher-performance servers are changing design standards. These data center industry innovations can now be applied foundationally to create an advanced video infrastructure.

Enterprise data center networks are built hierarchically to provide high availability and continuing scalability. Hierarchical data center design provides the operator with the flexibility to create logical topologies that use traditional Layer 2 and Layer 3 network configurations, service module insertion, and network device virtualization to create architectures that scale orthogonally at each layer within dynamic application requirements.

As is shown in Figure 1, the enterprise data center network is typically modeled as Core, Aggregation, and Access network layers, with a Services layer for both security and network services.

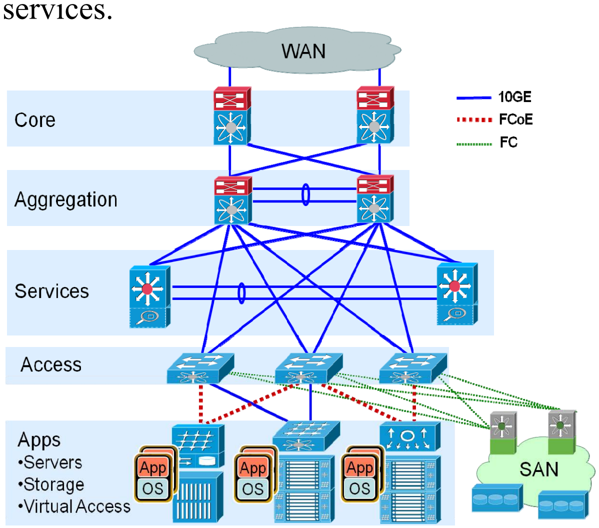


Figure 1. Enterprise Data Center Layered Network Design

This hierarchical, layered data center model employs redundant switches at each layer of the network topology for device-level failover, creating a highly available transport between end nodes over the network. The Services layer appears in this model because data center networks ordinarily require additional services beyond basic packet forwarding. These include, for example, server load balancing, firewalls, and intrusion prevention. These services can be introduced as standalone appliances or as modules that populate a slot of one of the switching nodes in the network. For each of these services an independent decision may be made whether to deploy redundant hardware based on the availability requirements associated with that service.

Experience with enterprise data center operations has shown the "Core, Aggregation, Services, and Access" layered model to be an enabling paradigm for scalability, performance, flexibility, resiliency, and maintenance. Here we will adapt and extend this model to next generation service provider video infrastructure, focusing on the Media Data Center while recognizing many of the architectural principles are equally applicable to other service provider facilities, such as next generation Hubs or Pod deployments.

ADAPTING ENTERPRISE DATA CENTER INNOVATIONS TO THE NEXT GENERATION SERVICE PROVIDER MEDIA DATA CENTER

In Figure 2, video resources are folded into the hierarchical data center model, providing functionality for acquisition and other traditional MPEG workflows as well as next generation IP-based subscriber apps and content distribution. Subsequently, we will develop the ideas of virtualization of infrastructure at each layer and the unification of compute and storage networking into a logical representation of this physical topology model.

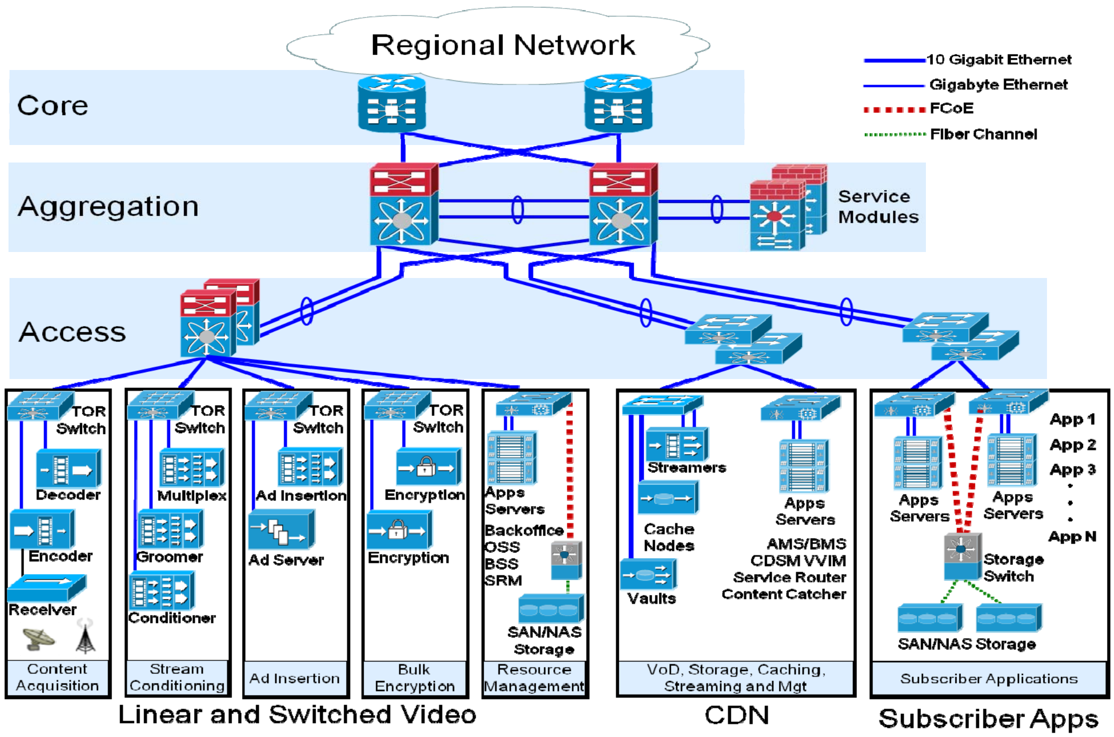


Figure 2. Media Data Center Components

Hierarchical Networking

As in the enterprise data center, the Media Data Center is implemented via hierarchical network layers. The design provides distinct layers for Core, Aggregation, Network & Security Services, and Access. With this layered approach, the Media Data Center eases the operational challenges of managing resource access, security, and scaling across a diverse set of video elements.

The Core network layer provides a Layer 3 connection to the national or regional transport network.

The Aggregation network layer performs load balancing and aggregation of video flows and management data across various service contexts.

The Network and Security Services Layer implements firewalls, server load balancers, intrusion prevention systems, application-based firewalls, and network analysis modules.

The Access Layer (a switching layer in the Media Data Center not to be confused here with the service provider access network or outside plant) provides efficient access to a range of modular video resources.

Compute, Network, and Storage Unification and Virtualization

Emerging best practices in the data center will unite compute infrastructure, storage access, virtualization, and the networking of both compute elements and storage in cohesive systems that simplify operations and reduce total cost of ownership.

Flexible and advanced x86 architecture server technology integrated in blade form factor enables the running of compute infrastructure as virtual machines. Server virtualization increases the total utilization factor for the compute infrastructure, enables application mobility, and increases service availability, all while generating power savings and total cost reduction. Virtualization becomes the new data center best practice with the emergence of hypervisor applications such as VMware's ESX/ESXi Server, Microsoft's Hyper-V, and the open source hypervisor Xen, leveraging the availability of powerful, multi-core x86 CPUs and virtualization technology from both AMD (AMD Virtualization) and Intel (Intel Virtualization Technology).

Stateless computing is used to enable seamless server mobility. Service profiles are used to abstract server attributes and decouple them from physical hardware attributes. Applications will run on virtual machines that are instantiated in response to dynamic conditions. Applications and processes can rapidly and easily migrate between hardware instantiations. This will allow server capacity planning that leverages the statistical characteristics of both application and service load and improves overall utilization accordingly.

As mentioned briefly above, virtualization of servers is achieved using virtualization applications known as hypervisors. These are virtual machine monitors that enable multiple operating systems to run concurrently on a host computer. Native hypervisors are software systems that run directly on the host's hardware as a guest operating system monitor, and provide the ability to scale virtual machines to the degree required in the Media Data Center. Multiple virtual machines can be created to run in isolation, side by side on the same physical server. As shown in Figure 3, each virtual machine has its own set of virtual hardware (RAM, CPU, NIC) upon which an operating system and applications are loaded. The operating system when loaded sees a consistent and normalized set of hardware without regard to the actual physical hardware components upon which it is instantiated. This ability to run multiple virtual machines abstracted on a single set of hardware sees it greatest benefit in applications with high peak to average CPU load, or simply low average load. Examples of these in the Media Data Center might include management or control applications, or video on demand applications, but likely extend to a surprising number of other applications.

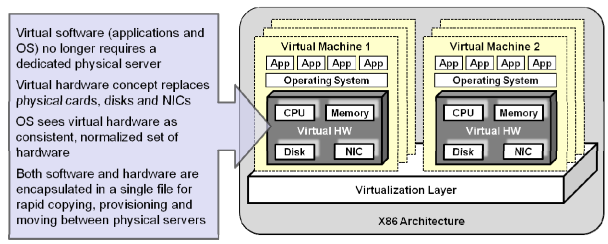


Figure 3. Virtual Machine Concept

Yet another feature of the virtual machine approach is the ability to encapsulate virtual machines into files for rapid saving, copying, and provisioning. Full systems, including operating systems, BIOS, virtual hardware, and fully configured applications can be moved within seconds from one physical server to another in response to changing service loads or for zero down time maintenance. This "*service orchestration*" feature gives the server infrastructure the ability to dynamically apply additional resources to new service loads, yielding operational and utilization benefits across a wide range of use cases, but offering benefits of particular interest in applications with high average CPU load but variable service load. In the Media Data Center examples of such applications might include video processing operations for ingest, live or on-the-fly transcoding, time-shifted streaming, or many others. Together, the two capabilities of 1) running multiple virtual machines on the same physical hardware, and 2) dynamically moving file-encapsulated virtual systems onto new hardware, give the virtualized server infrastructure the simultaneous ability to dynamically scale in *two* dimensions in response to both dynamic CPU loads and dynamic service loads. Improvements in utilization factor can be dramatic. In a use case deploying a home security monitoring application, server hardware requirements were reduced by a factor of 4:1. In another use case, a hardware reduction factor of 8:1 was seen for a session & resource management application such as is commonly deployed by service providers to manage QAM sessions for VoD, switched broadcast, or data services.

In the next generation Media Data Center design, the formerly separate LAN, SAN, and high performance computing networks are consolidated onto a single network and unified fabric which uses 10G Ethernet as the fundamental link interconnect. Legacy data center storage network designs typically use Fibre Channel interfaces for storage and associated switching. These Storage Area Networks (SANs) are built and operated independently of the data center Ethernet switching networks, creating a duplication of cabling, switching, and other resources. The Fibre Channel over Ethernet encapsulation, as shown in Figure 4, below, is a standards based approach that converges these two networks onto a single more cost effective Ethernet infrastructure.

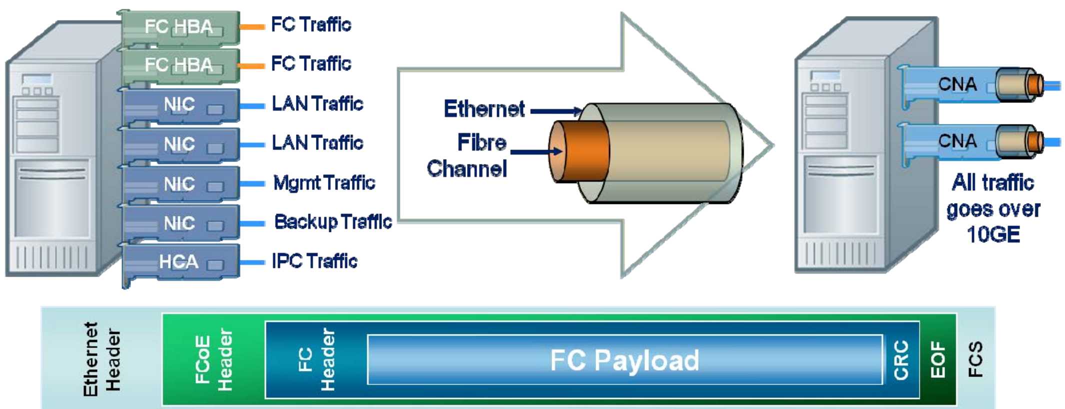


Figure 4. Consolidating LAN/SAN Networks using Fibre Channel over Ethernet

Devices that formerly attached to Fibre Channel networks using Host Bus Adapters (HBAs) will now attach to Ethernet networks via Converged Network Adapters, which integrate the functionality of both the HBA and an Ethernet Network Interface Card (NIC). These Converged Network Adapters, or CNAs, perform the Fibre Channel over Ethernet encapsulation and allow storage systems to use native storage protocols over the unified fabric at lower cost than traditional SAN switching. Figure 4 depicts a single pair of 10 GigE CNA ports replacing the multiple adapter types and multiple links of the legacy configuration. This scheme provides a fully redundant transport that supports both Fibre Channel SAN and Ethernet LAN topologies over a single physical link,

The unified fabric approach eases requirements for cabling, switching, power, cooling, and network adapters, with particular savings derived as a result of the migration from 1 Gbps Ethernet and Fibre Channel cabling to 10 Gbps Ethernet connections. In addition, the unified fabric provides a simplification of the switching network architecture. Compute servers will use a FCoE connection to the access layer that supports both LAN and SAN interfaces. This allows a single pair of switches that support FCoE with 10G Ethernet interfaces to replace four switches (two LAN and two SAN) that were present in the legacy topology. As a result, a complete layer of Fibre Channel switches and cabling can be removed from the topology as is illustrated in Figure 5.

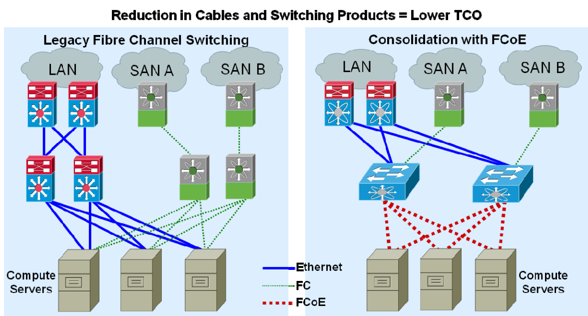


Figure 5. Unified Fabric Switching Topology

Another dimension of virtualization available in the next generation Media Data Center is virtual access switching. In a dynamic environment, individual and total application requirements for compute and storage will scale up and down based on subscriber behavior. Network access for the responsible virtual machines will need to scale correspondingly. Providing a virtual access switching layer is a powerful method to virtualize the interfaces for virtual machines and unified storage networks.

A Virtual Ethernet Module is installed into each server hypervisor kernel, while a distributed virtual switch is created by installing a Virtual (Switch) Supervisor Module on an additional virtual machine. Network LAN policies are implemented at this layer in port profiles, much as server profiles were created for the stateless computing mode of operation of the virtual machines. A typical port profile identifies, for example, MAC, World Wide Name (the Fibre Channel unique element ID), Boot Order, and Firmware, as well as network (LAN) and storage (SAN) policies. In this way operators may dynamically and easily apply policy-based configuration and operation of network services for each virtual machine; they can also easily manage virtual machine connectivity across physical servers to balance server workloads for optimized application performance or to provide improved availability during routine hardware maintenance.

Each virtual machine connects to its own Virtual Ethernet port on the access switch. This creates a key capability; the network administrator is given traffic visibility and policy control on a *per virtual machine* basis. Virtual machines can be managed as though they were physical servers from the perspective of network connectivity. If applications need to be moved or to grow, port profiles will follow the application to the new resources. Operators will simply modify service profiles to add or subtract capacity for video applications, eliminating multiple complex configuration operations and the need to separately touch the configurations of servers, adapters and interfaces, LAN/SAN switches, and storage devices.

The Media Data Center design should also include embedded management systems that enable management of the unified compute infrastructure (servers, storage, network, virtual machines) as a single entity. To fully leverage the virtualized and unified environment of the new Media Data Center, management GUIs, CLI interfaces, and robust programmatic interfaces will allow the definition of service profiles that logically encapsulate desired physical configurations and provisioning of unified resources.

<u>Device Virtualization and Virtual Device Contexts</u>

Increasingly, routers are offering advanced features allowing the virtualization of network device features. The possible dimensions of device virtualization include: forwarding plane, control plane, management plane, and the partitioning of software and hardware components into those planes. The Media Data Center architecture here described leverages each of these dimensions of virtualization in the aggregation switch layer, creating partitions that we will subsequently call “Virtual Device Contexts.” In this way, aggregation switches become virtual devices that present as independent and unique logical entities to users connected within each Virtual Device Context. Each Virtual Device Context will maintain its own set of software processes, have its own configuration, and be administered via its own management context.

The Media Data Center uses Virtual Device Contexts to partition families of services and applications *according to the operational practices of the service provider.* An example partitioning might delineate the resources associated with legacy linear, switched digital broadcast, and IPTV in one Virtual Device Context from the resources associated with CDN-based services such as VoD and Internet streaming in a second context, while creating yet a third context for management applications and advanced subscriber interactive applications.

USE CASE: INSERTING THE “TV EVERYWHERE” MEDIA POD

This section describes an insertion strategy for a "TV Everywhere" Media Pod, or design module. This business use case describes how the Media Data Center architecture can accommodate the insertion of new features such as adaptive bit rate video streaming. The following description provides a vision of how new services can be integrated smoothly into the overall architecture.

Common Cloud Infrastructure

Service providers have embraced "TV Everywhere" as a way to bring TV-quality video programming to devices beyond the home television. The "TV Everywhere" initiative develops the infrastructure and applications to deliver premium live and on-demand content to many IP devices, including the PC, tablet, and mobile devices. Service providers are in various stages of system deployment and system trials.

With the Media Data Center and the Media Pod modular design, service providers can implement a distributed cloud based infrastructure that supports “TV Everywhere” services. As shown in Figure 6, this infrastructure includes a combination of National Media Data Centers, Regional Media Pods, and a distributed Content Delivery Network to support streaming video services. The “common-cloud” infrastructure utilizes data center infrastructure that has proven performance and known scaling for the relevant video workflows. These workflows can be replicated across regional Media Pods and National Media Data Centers. Given this approach, service provider operations teams can develop a level of certainty in their infrastructures, increase service velocity, and focus on the deployment of video applications and subscriber services.

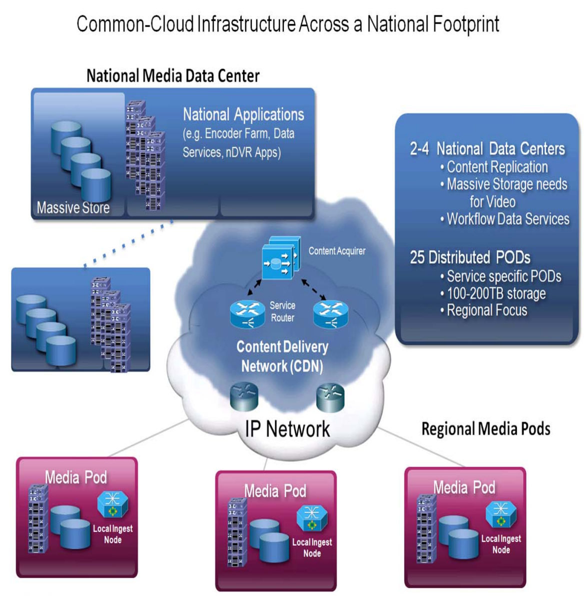


Figure 6. Common Cloud Infrastructure

Adaptive Bit Rate Workflows

Adaptive Bit Rate video delivery plays a critical role, because video content is likely to traverse an unmanaged or congested network to reach devices in the home or on the road.

ABR delivery models can adapt the streaming rate in response to real-time network conditions, in order to maintain the video experience. The data center infrastructure needed to acquire, create, manage, and deliver ABR video streams includes a mix of traditional video components along with advanced data center products. ABR workflows contain multiple processing stages in which video is acquired, transrated, encapsulated, packaged in a DRM, and stored in an origin server. These stages must support linear and on-demand titles, as well as various ABR formats and bitrate profiles. Sample workflows for an adaptive bit rate Linear, On-Demand, and Time-shift services are provided in Figure 7 below. As video content passes through each stage of the workflow, a combination of compute, network, and storage resources will be consumed. Data center infrastructure elements must support these dynamic processing requirements and scale to vast content sources and wide ranging delivery formats.

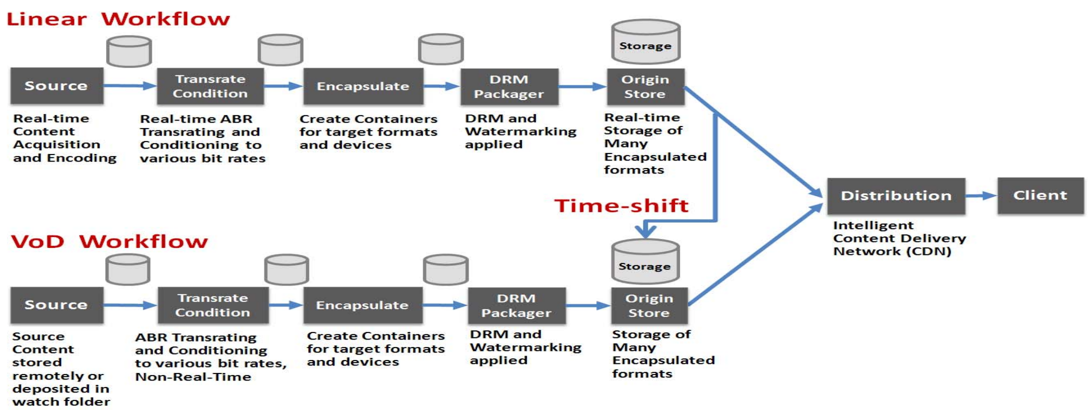


Figure 7. Workflows for Adaptive Bit Rate Services

<u>Media Pod Components and Applications</u>

The Media Pod includes the processing elements and infrastructure required to manage and produce ABR content. A number of Media Pod components and applications required to support a "TV Everywhere" service are shown in Figure 8. These ABR applications include content transcoders, packagers, DRM and watermarking applications, as well as a wide range of content management and publishing tools.

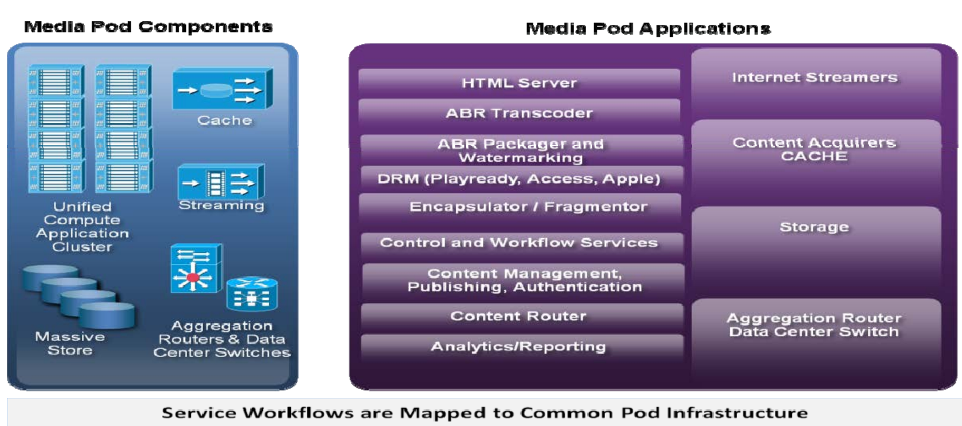


Figure 8. Media Pod Applications and Components

Service providers can map video data plane workflows and content management functions to the Media Pod infrastructure. The Media Pod is designed to provide a scalable compute, network, and storage architecture required by next-gen video services and cloud-based applications. The Media Pod is depicted in Figure 9 below and utilizes the Media Data Center architectural principles and innovations to significantly reduce operational complexity and reduce total cost of ownership.

Combined SAN / NAS storage devices supporting virtualized storage with block and NFS storage technologies are used in the Media Pod. Unified storage scales easily to support massive centralized storage or smaller scale regional storage.

The Media Pod incorporates unified storage switches delivering 10GE based FCoE technology. These switches enable 10GE FCoE network connections, saving significant cabling costs and achieve a consolidation of switching devices.

The Media Pod employs a powerful x86 based compute environment that can implement a "bare metal," or a virtualized compute environment. Virtual compute servers provide more efficient use of compute resources, and deliver superior operational management. Video workflow applications, content management, and subscriber applications reside on this compute environment.

Access switching in the Media Pod uses virtual access-switch technology that supports cloud based operations, application movement across different compute resources, and profile based configuration of network and storage requirements.

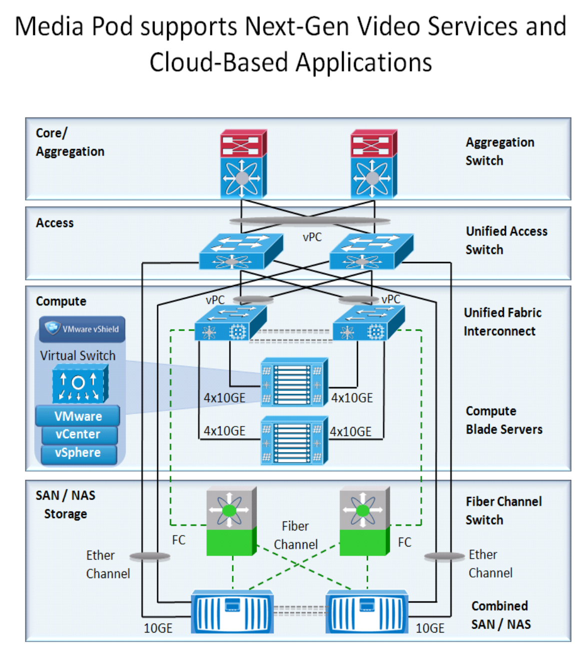


Figure 9. Media Pod Architecture

Video applications are mapped to virtual machines and unified storage elements in the Media Pod. Different ABR workflow stages (transrate, fragment, DRM, watermark) are overlaid onto blade servers and storage elements within the Pod architecture. Application profiles that define the operating system, compute requirements, network interfacing and QoS policies, and storage requirements are created for each workflow application. These video application profiles are overlaid onto Media Pod resources as described in Figure 10 below.

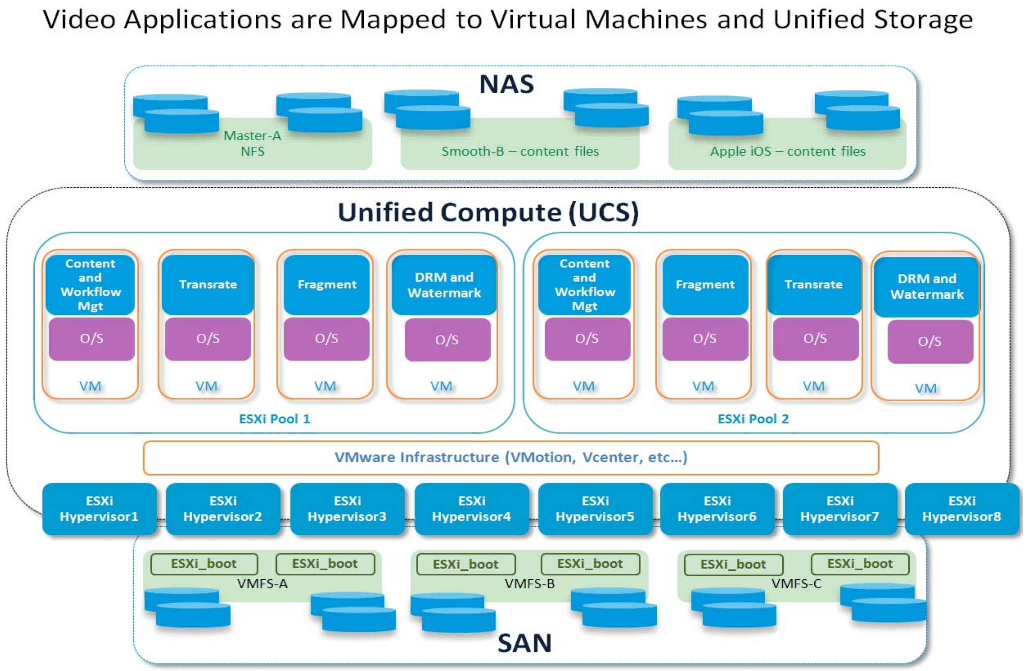


Figure 10. Video Applications are mapped to Virtual Machines and Unified Storage

Media Pod Insertion

The Media Data Center has been architected to allow the smooth insertion of design modules such as the Media Pod. Figure 11 depicts how the Media Pod can be inserted into the architecture to manage and deliver ABR content for "TV Everywhere" applications. It is likely that the Media Pod and other data center design modules will share some certain processing workflows such as the live content acquisition workflow. The content acquisition workflow is designed to support this shared service environment. In the example below, both traditional TV and the ABR Media Pod share the content acquisition workflow. In addition, the CDN system will be updated to include ABR Internet streaming components. Because the Media Data Center has been designed to leverage enterprise data center principles and innovations in accordance with the foundational tenets discussed earlier, the Media Pod and Internet CDN components insert easily into the architecture.

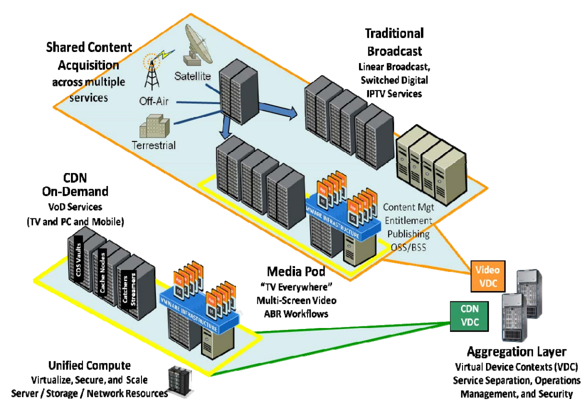


Figure 11. Media Pod Insertion

Figure 12 depicts the infrastructure for the new TV Everywhere ABR service overlaid on the Media Data Center logical topology. The Media Pod folds into the design as a secure partition, and can efficiently use new computing, storage, and network resources. Similarly, the CDN partition can accommodate new storage, cache, and streaming components as needed for the Internet streaming function.

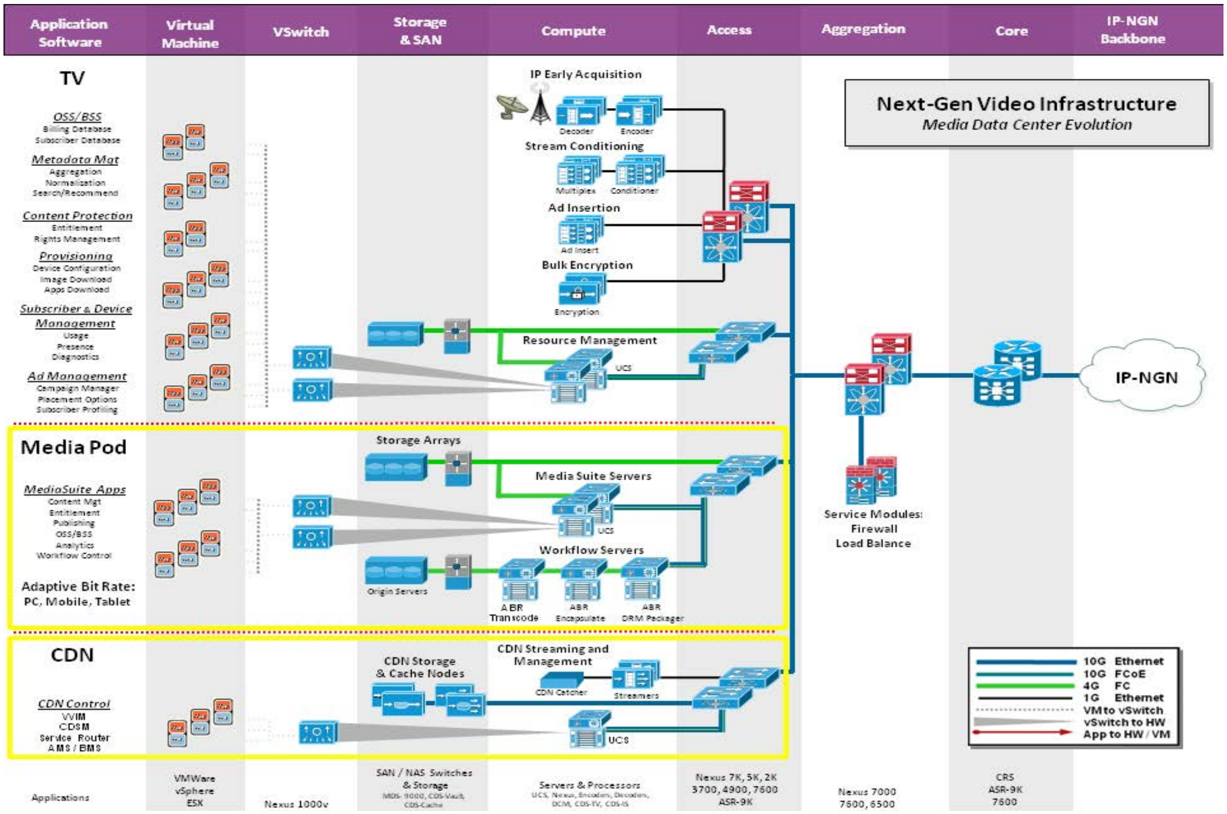


Figure 12. Media Pod Partition

CONCLUSION

An architecture for a Media Data Center is developed that leverages enterprise data center principles and innovations for a scalable video infrastructure. Virtualization is leveraged in the compute domain as well as in all layers of network connectivity and administration. Storage, compute and network resources are unified to achieve powerful operational benefits and cost reductions. Finally, a use case is demonstrated, with the implementation of a Media Pod design module that provides scalable infrastructure for adaptive bit rate streaming and the emerging "TV Everywhere" service.

Architectural Approaches to Help Circumvent the "Simulcast Roadblock" of IP Video Deployments

T. Cloonan, J. Allen, C. Ansley, R. Arnold, J. Brooks, M. Bugajski, C. Cheevers, T. Cotter, J. Howe, B. Hanks, D. Torbet, M. Uecker, & I. Wheelock

ARRIS

Abstract

Multiple System Operators (MSOs) throughout the Cable Industry are planning to roll out IP video services. Many MSOs hope to eventually migrate these IP Video Services to their DOCSIS 3.0 infrastructure in an effort to capitalize on many benefits, including the statistical multiplexing gains of channel bonding, the QoS that permits partnerships with other Internet content providers, and the simplicity of a single delivery system.

One particular challenge to this planned migration is created by the fixed amount of bandwidth available on the HFC plant to support simultaneous deployment of legacy services and the new DOCSIS-based IP video services during the transition. This "Simulcast Roadblock" problem will be described within this paper, and a toolkit of potential solutions will be identified.

INTRODUCTION

Background on IP Video Services

Multiple System Operators (MSOs) are beginning to plan architectures which will ultimately be used for the deployment and delivery of Internet Protocol (IP) Video services to their subscriber base. From the subscriber viewpoint, IP Video will manifest itself as a video delivery system that permits their video content to be distributed over their home IP network to multiple types of devices: IP STBs (with accompanying TV display devices), IP-enabled TV"s, game players, DVD players, handheld devices (such as tablets and smartphones), and PCs.

Different MSOs will move toward IP Video at different times and at different rates, and different MSOs will also choose slightly different architectures as they unveil these new services. However, the authors believe that once the transition to IP Video begins within any particular MSO network, it will likely occur quite rapidly. MSOs will work to offer competitive responses to the Over-The-Top Video (OTT) content providers who have begun to thrive on the Internet. In many ways, this OTT content has created a new challenge to the MSOs" legacy video delivery systems [Ins1].

By offering this new type of IP Video service to their subscribers, most MSOs are working to accomplish several important goals, including:

1) Gaining access to a broader audience through delivery to multiple screens in the home
2) Building a brand identity with the 15-30 year-old demographic (through their handheld devices)
3) Creating new means of further monetizing their high-quality video content with new subscription fees (for the various multiple screen types in the home)
4) Providing an opportunity to enter the growing "Internet advertising market" through directed advertising in IP-based videos
5) Ring-fencing their subscriber base by becoming the popular organizers and

aggregators of all IP video content (MSO-based and Web-based)

6) Reducing the high CPE costs traditionally associated with legacy STBs

Three Modes of IP Video Delivery

Three fundamentally different delivery paths are currently being architected and studied by MSOs, and the services associated with each delivery path may be rolled out at different times. These delivery paths include:

1) Internet-based, unicast delivery of MSO-owned VoD and Linear video content to subscribers not directly connected to the MSO‟s HFC plant. This is known as "off-net" delivery of the video content. Depending on where the content is actually hosted, this service could even be a cloud-based IP video service for the MSOs.

2) On-Net HFC-based, unicast delivery of MSO-owned VoD and Linear and Remote Storage DVR (RS-DVR) video content to subscribers directly connected to the MSO‟s HFC plant. This is known as MSO-managed or walled-garden or "on-net" delivery of VoD and Linear and RS-DVR video, because the MSO can better manage the link utilization and Quality of Experience associated with the service (since it is offered entirely on the MSO‟s links).

3) On-net HFC-based, multicast delivery of MSO-owned Linear video content to IP-based subscribers directly connected to the MSO‟s HFC plant. This is known as MSO-managed or walled-garden or "on-net" delivery of Linear video. The use of multicast delivery for Linear video provides bandwidth efficiency gains over the use of unicast delivery by reducing the number of stream replications appearing on the HFC plant.

Three Transport Methods for On-net Delivery

The actual on-net delivery of the content across the HFC network can be implemented in at least three different ways:

1) MPEG/IP Encap

The video content could be transmitted to the home using the legacy QAM-based MPEG-TS digital video distribution infrastructure (which is normally used to transmit video signals to legacy STBs). The in-home receiving device receives the MPEG-TS content, re-encapsulates it into IP packets, and then transmits those packets throughout the home IP network where they can be accessed and decoded by any IP-enabled device with the appropriate client

2) DOCSIS

The video content could be transmitted to the home using the legacy DOCSIS distribution infrastructure, with an advanced in-home cable modem (called a media gateway) receiving the IP packets. Depending upon the home video distribution chosen, the video packets may be stored in the media gateway to provide whole-home DVR functionality and time shift buffers, and/or processed to accommodate lower bit rate clients, and/or forwarded directly to a subtending client on the home network. Depending on how the MSO chooses to implement multicast Linear distribution, a multicast to unicast conversion of the IP streams might also be required within the media gateway before forwarding the packets through the home network. This multicast to unicast conversion is merely one of several important features, such as device discovery, encryption, digital rights management, and caching, which are found in media gateway devices. These media gateway devices have been called cable modems on "video steroids." When used for Linear video, a DOCSIS video delivery service promises to offer all of the improved bandwidth efficiencies of legacy switched digital video services, because IP multicast streams will likely be transmitted to a service group only when one or more clients in that service group have joined multicast sessions for long-tail content.

3) Bypass

The video content could be transmitted to the home using a proprietary, bypass-enable EQAM with a subset of DOCSIS MAC functionality. The video streams would then be fed through the proprietary EQAM over a DOCSIS channel to a proprietary cable modem, which then forwards the video streams through the home network (assuming no media gateway functionality).

Each approach has its own list of potential issues. The authors believe that the negative issues for the proprietary bypass approach are quite constraining, and these issues have already been outlined in other papers (ex: [Clo1]). As a result, proprietary bypass approaches will not be considered within this paper, and the authors will instead focus on IP video solutions based on MPEG/IP Encap transport and IP video solutions based on DOCSIS transport. The authors believe that many MSOs will eventually want to chart a course that takes them to the particular solution that uses DOCSIS for IP video. Reasons for this belief include:

1) DOCSIS would provide the simplicity and low cost of a single, fully-converged infrastructure for all services; voice, data, and video.

2) DOCSIS channel-bonding provides statistical multiplexing gains allowing more video programs to be delivered to subscribers on a given number of channels. It is expected that many MSOs will also want to capitalize on the further benefits that can be realized if Variable Bit Rate (VBR) IP video streams are utilized instead of Constant Bit Rate (CBR) video streams in a channel-bonded IP Video environment. As an example, with four Downstream Bonded Channels in a particular Downstream Channel Set, it has been shown that an MSO can transmit ~25% more programs on those four Bonded Channels than when the four Channels are used in a Non-Bonded fashion (see **Fig. 1**). To account for the lower bandwidth capacity requirements for VBR IP Video on bonded channels (**Fig. 1**), we will utilize the average bandwidth of each video stream to determine the amount of capacity consumed whenever those IP video streams are flowing over a reasonably-sized Bonding Group with three or more DOCSIS Downstream Channels. However, we must assess a ~25% "No-Stat-Mux-Gain tax" whenever video programs are multiplexed together on small Downstream Channel Sets (or non-bonded channels). In particular, within this paper, it will be assumed that video programs which are multiplexed together on a non-bonded channels will each require an increased effective bandwidth capacity given by 1.25x(the average bandwidth value for the video stream). This "No-Stat-Mux-Gain tax" will account for the extra bandwidth headroom that must be made available in a single Downstream Channel to accommodate the burstiness and larger peak bandwidth-to-average bandwidth ratios of IP Video stream aggregates that do not contain many programs. Thus, we will assume that channel bonding systems with N bonded DOCSIS channels will support 25% more video programs than N-channel systems that do not support channel bonding (whenever N>=3). It should be clear that legacy MPEG-TS video delivery does not

provide bonding, and therefore it is subject to the 25% tax as well.

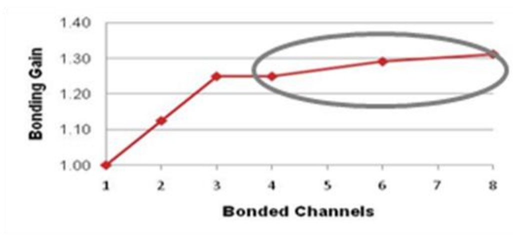


Fig. 1- VBR Stat-Mux-Gain (Bonding Gain) as a Function of the Number of Bonded Channels carrying 4 Mbps Streams

3) DOCSIS provides advanced QoS that is necessary to safely source IP video content from Internet content partners

4) Many DOCSIS CMTSs offer the high availability needed for future services

5) Per-channel pricing on DOCSIS CMTS equipment is dropping as demand for channels increases

As a result of these benefits, many MSOs are likely to pursue one of several paths in the future as they make their way toward an ultimate DOCSIS IP Video delivery solution. Two of these paths are outlined below.

Two Basic Transition Strategies

MSOs must carefully plan their paths as they migrate towards IP Video Architectures in the future. Two different paths are likely to be followed by different MSOs. One is the 2-Step Transition Strategy, and the other is the 1-Step Transition Strategy.

In the 2-Step Transition Strategy, MSOs convert their video delivery methods in two steps. They begin with their current video delivery methods, incorporating technologies such as analog transport of broadcast video, digital video transport of broadcast video, digital video transport of SDV, and digital video transport of VoD. They transition to Phase 2, IP in the Home, as they deploy an MPEG/IP Encap delivery system, using digital video transport of broadcast, SDV, and VoD over the HFC plant, and then converting to IP packets for transmission over the home network to low cost IP-based STBs or other IP-based connected devices. IP in the Home can be implemented as a replacement for the current in-home delivery system, or it can be implemented as an augmentation to the current delivery system. They finally transition to Phase 3, All IP, by moving their video streams to the DOCSIS infrastructure. The All IP Phase can be implemented as a replacement for the MPEG Delivery Phase delivery system, or it can be implemented as an augmentation to the MPEG delivery system. In other words, an MSO may choose to introduce the IPTV experience with an IPTV tier, which may or may not contain channels that overlap with the conventional tiers. Any overlap of the programming between MPEG and DOCSIS transport will require simulcasting of video streams on both the analog/MPEG infrastructure and the DOCSIS infrastructure.

In the 1-Step Transition Strategy, MSOs convert their video delivery methods in one step. They begin with their current video delivery methods, incorporating technologies such as analog transport of broadcast video, digital video transport of broadcast video, digital video transport of SDV, and digital video transport of VoD. They transition to the final All IP Delivery phase, Phase 2, when they move to the DOCSIS infrastructure. The All IP Delivery Phase could be implemented as a replacement for the current MPEG delivery phase, but in most real-world scenarios, there will likely be a window of time during which the MPEG and DOCSIS delivery systems are running in parallel. This requires simulcasting video streams on both the analog/MPEG infrastructure and the DOCSIS infrastructure.

In both the 2-Step Transition Strategy and the 1-Step Transition Strategy, there is a possibility that simulcasting of some video streams may be required on both the analog/MPEG infrastructure and the DOCSIS infrastructure. This simulcasting requirement could lead to challenging demands on the HFC plant bandwidth capacities, which could in turn become a roadblock to the deployment of IP video services. This paper will attempt to describe the potential Simulcast Roadblock problem and propose useful solutions.

PHILOSOPHIES ON THE FUTURE AND AN EXAMPLE OF THE SIMULCAST ROADBLOCK PROBLEM

Regardless of which Transition Strategy is selected, MSOs may run into the Simulcast Roadblock problem. The severity and magnitude of this problem will be different for each MSO. To some extent, the magnitude depends on which of two differing philosophies on IP Video bandwidth growth ultimately proves to be true. These two philosophies might be labeled the "zero-sum-game" philosophy and the "new-growth" philosophy. According to the "zero-sum-game" believers, MSOs who add MSO-managed IP video delivery to all three screens will not increase the total bandwidth consumption of their subscribers. Instead, they will merely ensure that a good portion of the IP video content viewed on PCs and hand-helds will be sourced from servers owned by the MSOs (instead of being sourced by Over-The-Top providers).

According to the "new-growth" believers, MSOs who add MSO-managed IP video delivery to all three screens and couple it with the high-end content in their content libraries (which is much better than the content in the content libraries of Over-The-Top providers) will create a new demand for bandwidth capacity that did not exist for the Over-The-Top providers in the past. This will ultimately increase the total number of eyes watching IP Video and increase the total bandwidth consumption of their subscribers. Recent deployments of iPad applications that can receive Live programming content from the MSOs" DOCSIS networks seem to be displaying this type of "new-growth" behavior, but it is still too soon to tell.

Regardless of which philosophy is correct and regardless of which Transition Strategy a particular MSO selects, the period of time when simulcasting is required may lead to challenges for the HFC plant bandwidth- especially if MSOs are reluctant to reduce the size of the existing service offerings. We will perform an example analysis below assuming the more conservative numbers of the zero-sum-game philosophy. However, it should be recognized that the magnitude of the simulcast bandwidth problem could be much larger if the new-growth philosophy proves to be at all valid in the future.

While it is not possible to present all of the different scenarios that MSOs are likely to encounter, it may be beneficial to examine the magnitude of the simulcast roadblock problem for a mythical MSO whose HFC plant and channel characteristics might be "typical." Defining a "typical" HFC plant for an analysis is always a challenging task, because the characteristics of different HFC plants can vary quite extensively from MSO to MSO, and the validity of any assumptions will undoubtedly be heavily debated. Nevertheless, for this paper, we will assume the following to be "typical" HFC plant characteristics for a mythical MSO in the future:

1. The conservative "zero-sum-game" philosophy will be assumed, so it is possible that bandwidth capacity demands would be even greater if the "new-growth" philosophy were used
2. The HFC plant is a 750 MHz plant that supports 115 channels
3. Each digital channel carries a 6 MHz, 256 QAM signal that supports 42 Mbps of raw bandwidth
4. For video distribution prior to the arrival of IP Video, no digital

broadcast programs are being used. Only analog Linear programs, digital SDV Linear programs, and digital VoD programs are being used prior to the arrival of IP Video. (Note: This approach may not be common, but it simplifies the analysis to assume that all legacy digital, Linear programs are delivered via SDV).

5. For IP Video distribution, no nailed-up programs are being used. Only Switched Digital multicast IP Linear programs and unicast IP VoD programs are used. (Note: It is possible that digital Linear programs could be nailed up in an IP Video distribution system, but it simplifies the analysis to assume that all IP Linear programs are delivered via Switched IP services).

6. Analog Broadcast programming is used to fill out the spectrum after all other services have been allocated their required number of channels. It is assumed that these Analog channels would only carry the basic tier programming content.

7. The number of selectable SDV programs offered by the MSO to their subscribers in the MPEG-TS pool = 250

8. The number of selectable Switched Digital multicast IP programs offered by the MSO to their subscribers in the IP Video over DOCSIS pool = 250

9. Required SDV (or Switched Digital multicast IP) Blocking Probability = 0.01%

10. Assume all programs transmitted over the HFC plant have been converted to MPEG4 encoding (whether used for MPEG-TS delivery or DOCSIS delivery)

11. The mix of MPEG4 programs on the HFC plant might be 50% TV-HD programs at an average of 8 Mbps, 30% TV-SD or PC-SD programs at an average of 1.5 Mbps, and 20% Handheld programs at an average of 300 kbps, which results in an average program bandwidth given by: (0.5)*(8 Mbps) + (0.3)*(1.5 Mbps) + (0.20)*(300 kbps) = 4.51 Mbps. For channel-bonded DOCSIS delivery systems, the bandwidth required per program will be that 4.51 Mbps value. For MPEG-TS delivery systems (where the stat-mux benefits of channel bonding are not provided), the "No-Stat-Mux-Gain" 25% tax must be applied, increasing the effective bandwidth per program to a value of (4.51 Mbps)(1.25) = 5.64 Mbps. (Note: This again includes a mix of traffic resolutions ranging from 8 Mbps for TV-HD to 300 kbps for Handheld video).

12. The initial HFC Channel Map for the "typical" 115-channel plant at the beginning of the transition is defined as shown below:

    Analog Broadcast = 75 channels

    SDV = 33 channels

    VoD = 4 channels

    HSD/VoIP = 3 channels

13. The Fiber Node size is 500 House-Holds Passed (500 HHP)

14. The Video take-rate will remain fixed at 60% of the HHP (500*0.6=300 Video subscribers per Fiber Node)

15. The Percentage of Video subscribers with Digital STBs (or IP Video Gateways) in their homes will remain fixed at 75% (300*0.75=225 subscribers with Digital STBs or IP Video Gateways per Fiber Node) … only these subscribers can access SDV and VoD services

16. The Percentage of Video subscribers who consume only Analog programming will remain fixed at 25%.

17. The HSD/VoIP take-rate will remain fixed at 40% of the HHP (500*0.4=200 HSD/VoIP subscribers per Fiber Node)

18. The Service Group size for each of the Narrowcast service tiers will begin the transition period as shown below:

    SDV Service Group size = 4 Fiber Nodes (300*4*0.75=900 SDV subscribers)

    VoD Service Group size = 4 Fiber Nodes (300*4*0.75=900 VoD subscribers)

    HSD/VoIP Service Group size = 4 Fiber Nodes (200*4=800 HSD/VoIP subscribers)

19. Busy-Hour/Busy-Day Utilization (as a function of SDV subscribers) for SDV Video Services = 60% (Thus there are 900*0.6 = 540 active viewer homes per SDV Service Group)

20. Busy-Hour/Busy-Day # of active SDV viewers per active viewer home = 1.5 (Note that 540*1.5 = 810 active viewers per SDV Service Group)… from a separate SDV usage analysis with program popularity curve defined by the Power Law with alpha=0.7 and with 250 selectable programs, this requires ~225 transmitted programs (as illustrated in **Fig. 2**).

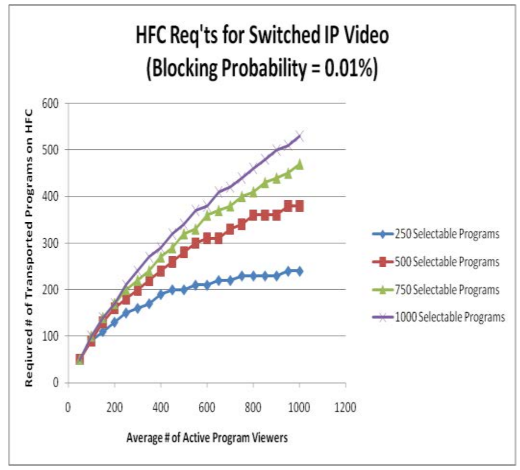


Fig. 2- Required Number of Transported Programs vs. Number of Active Program Viewers

(Note: The shape and height of the curves in **Fig. 2** are very sensitive to the assumptions made on the associated program popularity curves, and field data has indicated that the program popularity curves can vary quite extensively with the demographics of the service group and with the nature of the available programs. Thus, the program popularity curves can change, and the values in **Fig. 2** would also change).

21. For MPEG-TS transmission, if each viewer watches a non-bonded program at 5.64 Mbps, the total bandwidth transmitted = 1269 Mbps total… if each MPEG-TS channel provides 38.8 Mbps of usable bandwidth, this requires 1269/38.8 = 33 channels… For DOCSIS transmission, if each viewer watches a bonded program at 4.51 Mbps, the total bandwidth transmitted is 1015 Mbps total… if each bonded DOCSIS channel provides ~36 Mbps of usable bandwidth, this requires 1015/36 = 29 channels)

22. Busy-Hour/Busy-Day Utilization (as a function of VoD subscribers) for VoD Video Services = 3% (27 active

viewers per VoD Service Group… For MPEG-TS transmission, if each viewer watches a non-bonded program at 5.64 Mbps, the total bandwidth is 152 Mbps… if each non-bonded MPEG-TS channel provides 38.8 Mbps of usable bandwidth, this requires 152/38.8 = 4 channels… For DOCSIS transmission, if each viewer watches a bonded program at 4.51 Mbps, the total bandwidth is 122 Mbps… if each bonded DOCSIS channel provides ~36 Mbps of usable bandwidth, this requires 122/36 = 4 channels)

23. At the beginning of the transition, Busy-Hour/Busy-Day HSD Bandwidth per HSD/VoIP subscriber = 100 kbps (80 Mbps per HSD Service Group… adding 25% head-room, as many MSOs do, yields a need for 100 Mbps of HSD Bandwidth per Service Group… if each DOCSIS channel provides ~36 Mbps of usable bandwidth, this requires 100/36 = 3 channels)
24. DOCSIS HSD Bandwidth demands will increase by 50% every year, so the DOCSIS HSD Bandwidth demands on a yearly basis during the transition to IP Video over DOCSIS are described in **Fig. 3** below for five successive years:

	Avg HSD BW per Sub (kbps)	Avg HSD BW per Serv. Group (Mbps)	Avg HSD BW (w/ 25% headroom) per Serv. Group (Mbps)	# Reqd DOCSIS Channels per Serv. Group
Year 1	100	80	100	3
Year 2	150	120	150	5
Year 3	225	180	225	7
Year 4	338	270	338	10
Year 5	506	405	506	15

Fig. 3- HSD Bandwidth Trends

25. IP Video Gateways can access all of the DOCSIS Downstream Channels associated with IP Video… this eliminates the complications resulting from multiple viewers within a home (behind a single IP Video Gateway) trying to access different Linear multicast streams on different bonding groups when all of the IP Video Gateway tuners are already in use
26. No node-splits will be performed during the period of transition. (Note: This assumption may not be a valid assumption for many MSOs, but it will be valid for some).
27. Extra bandwidth required to support the transition from legacy MPEG-TS video to IP Video over DOCSIS will be obtained by reclaiming channels from the Analog TV service tier.
28. Static, separated application groups will be assumed (i.e.- HSD/VoIP traffic will use one set of DOCSIS channels and IP Video traffic will use another set of DOCSIS channels)
29. Static, separated IP Video tiers will be assumed (i.e.- VoD IP Video traffic will use one set of DOCSIS channels and Switched Linear (SDV) IP Video traffic will use another set of DOCSIS channels)
30. The transition will span a period of five (5) years

Using these assumptions, we can now determine the number of channels required when using the 1-Step Transition Strategy. Using this strategy, the MSO would continue to support its legacy MPEG-TS Video transmission system to its legacy STBs while simultaneously supporting the installation of a new IP Video over DOCSIS transmission system to its DOCSIS IP Video Gateways. As a result, simulcasting of the most popular video streams to both types of end-points is obviously required.

The key attribute of this transition plan is that the MSO will attempt to offer a full array of video services and a complete channel line-up to its customers in both the legacy MPEG-TS camp and the new IP Video over DOCSIS camp. As a result, it is typically assumed that a given home will be either receiving its video feeds through the MPEG-TS Video transmission system (with legacy STBs or Hybrid IP Video Gateway/IP-STB), or the home is receiving its video feeds through the IP Video over DOCSIS transmission system (with IP Video Gateway/IP-STB). In other words, the assumption is that no home utilizes both types of services. However, it is also assumed that a particular Service Group will have to support both types of homes (legacy STB-based homes and IP Video over DOCSIS-based homes). Finally, it is assumed that there will be a gradual transition in the number of homes that support IP Video over DOCSIS, with the IP Video over DOCSIS pool growing from zero homes in the beginning of the transition to 100% of the homes at the end of the transition. Thus, all of the homes that are added to the IP Video over DOCSIS pool are homes that have been removed from the MPEG-TS pool. Traffic Engineering can take advantage of this fact by appropriately adjusting the number of channels to match the expected number of SDV (Switched Linear) and VoD sessions for both legacy MPEG-TS service and IP Video over DOCSIS service. Due to the legacy-to-DOCSIS transition effect, the number of channels required for SDV and VoD sessions within the legacy MPEG-TS service will decrease while the number of channels required for switched video and VoD sessions within the IP Video over DOCSIS service will increase. However, there will typically be an increase in the total number of channels required for SDV (Switched Linear) and VoD programs being sent to the MPEG-TS service and the IP Video over DOCSIS service during the middle years of the transition. This is primarily due to the shape of the SDV curve in **Fig. 2**, which illustrates that the number of channels required to transport SDV services to a pool of subscribers does not rapidly decrease until the number of viewers within a given pool is reduced to a fairly small number. Thus, we need to move most of the MPEG-TS subscribers over to IP Video over DOCSIS service before we can experience a rapid drop in the number of required channels associated with MPEG-TS SDV service.

A Traffic Engineering analysis for the 1-Step Transition Strategy was performed using the simplifying assumptions listed above. This particular analysis assumed that the transition occurred slowly over a five-year period. During each of the latter four years, we assumed that the MSO converted 25% of the digital video homes from MPEG-TS STBs to IP Video Gateways running only the DOCSIS feature capability (no Hybrid MPEG/IP capability). After the fifth year, all of the digital video homes within the Service Group had therefore been converted into IP Video Gateway homes running IP Video over DOCSIS. The tables shown in **Fig. 4** below illustrate the number of channels required for each of the different Narrowcast service types carried over the HFC plant on a year-by-year basis while this 1-Step Transition Strategy is being carried out. The final row (in yellow) shows the total number of required Narrowcast channels, and the “hump” that occurs in years 2, 3,and 4 is quite apparent. It is this “hump” that could cause MSOs to struggle with the Simulcast Roadblock problem.

The “hump” in the last row of **Fig. 4** is a direct result of the need to simultaneously support legacy SDV MPEG-TS channels, legacy VoD MPEG-TS channels, Switched (Linear) IP Video over DOCSIS channels, and VoD IP Video over DOCSIS channels. It is interesting to note that the hump disappears by the end of the transition period, because by then all of the Linear MPEG-TS channels and VoD MPEG-TS channels have been removed from the Service Group.

If the MSO is unable to turn off Analog and/or Digital Broadcast channels during the middle portion of the transition period, then this simulcast tax could become a serious impediment to the deployment of IP Video over DOCSIS. As a result, it might be beneficial to explore techniques for reducing the amount of bandwidth required during the simulcasting period- i.e., techniques for getting "over the hump." These techniques can be beneficial even if simulcasting is not performed, because they can generically be used to help conserve channel bandwidth in an IP Video environment- with or without simulcasting. However, the techniques are definitely helpful if simulcasting is implemented by the MSO.

			Year		
Per Service Group	1	2	3	4	5
# Homes Accessing Legacy MPEG-TS SDV	900	675	450	225	0
# Homes Accessing Legacy MPEG-TS VoD	900	675	450	225	0
# Homes Accessing DOCSIS Switched Video	0	225	450	675	900
# Homes Accessing DOCSIS VoD	0	225	450	675	900
# Homes Accessing DOCSIS HSD	800	800	800	800	800

			Year		
Per Service Group	1	2	3	4	5
Legacy MPEG-TS SDV Bandwidth (Mbps)	1269	1184	1072	733	0
Legacy MPEG-TS VoD Bandwidth (Mbps)	152	114	76	38	0
DOCSIS Switched Video Bandwidth (Mbps)	0	586	857	947	1015
DOCSIS VoD Bandwidth (Mbps)	0	30	61	91	122
DOCSIS HSD Bandwidth (Mbps)	80	120	180	270	405

			Year		
Per Service Group	1	2	3	4	5
Legacy MPEG-TS SDV Bandwidth (# channels)	33	31	28	19	0
Legacy MPEG-TS VoD Bandwidth (# channels)	4	3	2	1	0
DOCSIS Switched Video Bandwidth (# channels)	0	17	24	27	29
DOCSIS VoD Bandwidth (# channels)	0	1	2	3	4
DOCSIS HSD Bandwidth (# channels)	3	5	7	10	15
Total Required Narrowcast Channels (# channels)	40	57	63	60	48

Fig. 4- Traffic Engineering Tables for a 1-Step Transition Strategy

TOOLS FOR SOLVING THE SIMULCAST ROADBLOCK PROBLEM

There are actually several potential tools in the toolkit that can be employed to help reduce the total amount of bandwidth capacity used by IP Video over DOCSIS feeds. Many of these tools would likely be implemented in the CMTS, but some would be implemented in other IP Video sub-systems.

Bandwidth reclamation

Some traditional techniques for salvaging bandwidth capacity on the HFC plant may come in handy when IP Video simulcast issues begin to develop. A common technique that might be employed by many MSOs is the simple bandwidth reclamation process. In this process, channels previously being used for legacy service are retired and the channel is re-injected into the MSO channel map with a new service association. A typical example has some basic tier, legacy analog channels being retired. These retired legacy channels can still be fed into the subscriber homes if they are replaced by digital feeds and DTA deployments within the homes… or the retired legacy analog channels could merely be removed from the program list without replacement, leading to a reduction in the available programming to those analog subscribers). In either case, the retired legacy channels could then be re-assigned to carry DOCSIS IP Video services.

Service Group Size Reductions

Another traditional technique for creating extra bandwidth capacity within a service group is the use of service group size reductions. This process can be carried out using either node segmentation or node splits.

With node segmentation, each of the four coaxial distribution legs emanating from a fiber node is assigned its own set of upstream and downstream channels (instead of sharing a single set of upstream and downstream channels between all four distribution legs). This results in each coaxial distribution leg having a quarter of the subscribers managed by the original fiber node.

With node splits, the homes passed by a single fiber node are re-assigned to two fiber nodes, with each of the new fiber nodes having half of the subscribers as the original fiber node.

In both cases, the number of subscribers sharing a set of upstream and downstream channels is reduced. Since the amount of bandwidth capacity required for narrowcast services (ex: SDV, VoD, HSD) is related to the number of subscribers sharing the bandwidth, these service group size reductions provide an effective means of reducing the bandwidth capacity required to satisfactorily support those narrowcast services. If the number of subscribers sharing a set of narrowcast services is cut in half, then the number of required VoD and HSD channels can roughly be cut in half. If the number of subscribers sharing a set of narrowcast services is cut in half, then the number of required SDV channels can roughly be reduced according to the non-linear curves shown in **Fig. 2**. As a result of the narrowcast channel count reductions that result from the service group size reduction, narrowcast channels are freed up and can be used by the new DOCSIS IP Video services. Unfortunately, SDV and IP multicast achieve their most impressive channel gains for larger groups of users, so a tradeoff exists between reducing the demand for unicast services by reducing the number of subscribers and reducing the efficiency of the SDV model by retreating from the most efficient service group size. One possibility may be to combine nodes for the SDV/IP multicast groups, while narrowcasting the DOCSIS HSD and VoD groups.

Connection Admission Control

Connection admission control (or CAC) is a powerful feature available in many CMTSs that can help ensure a good quality of experience for DOCSIS HSD subscribers and DOCSIS IP Video subscribers whenever the available bandwidth capacity becomes scarce.

Connection admission control is the important functionality that checks how much bandwidth is currently being utilized on a particular Downstream Service Group before a new service request is honored. The connection admission control algorithm compares that bandwidth to a threshold to determine if there is enough spare bandwidth capacity to permit a new service flow to be added to the Downstream Service Group. When the service flow is used to transport the packets for an IP Video stream, then the connection admission control algorithm is actually determining if there is enough spare bandwidth capacity to permit a new IP Video stream to be passed through a particular Downstream Service Group. This functionality ensures that a set of active IP Video streams do not over-subscribe the available bandwidth capacity within a Downstream Service Group. A rejected service flow will manifest itself as a message on the IP Video STB or client indicating that the user should try accessing the channel later.

Intelligent Load-Balancing

Load Balancing is a CMTS feature that identifies when the traffic loads on the Downstream Channels or Downstream Channel Sets are not evenly distributed. Operation with an unequally-distributed traffic load is undesirable, because it can lead to exceptional Quality of Experience levels for subscribers on the lightly-loaded Channels and unacceptable Quality of Experience levels for other subscribers on the heavily-loaded Channels. Experience has shown that it is prudent to re-distribute the traffic loads to create a more equal distribution of traffic loads and a more uniform (and hopefully acceptable) Quality of Experience level for all subscribers.

Once an uneven traffic load is identified by the Load Balancing algorithms within a CMTS, the CMTS can re-distribute the traffic load across the Downstream Channels of Downstream Channels Sets so that heavily-loaded Channels experience somewhat lighter loads, and lightly-loaded Channels experience somewhat heavier loads. This adjustment helps ensure that the bandwidth of all channels is evenly utilized.

Client-Controlled Adaptive Streaming for Unicast Programs

Adaptive Streaming is an interesting approach for trading off between IP Video bandwidth requirements, client device processing requirements, and the Quality of Experience levels for the subscribers. Adaptive Streaming techniques are being developed by most of the Video Delivery solutions on the IP Video market (ex: Adobe, Microsoft, and Apple), and MSOs may also decide to develop their own algorithms within their proprietary clients. All of these Adaptive Streaming solutions tend to recognize the fact that there may be times when a particular resolution video feed being sent to a subscriber is not the ideal resolution at that instant in time.

Sometimes, the video resolution (and associated bit-rate) may need to be temporarily reduced to adjust for heavy network congestion or to adjust for a heavy, transient processing load being placed on the client device. This reduction in video resolution will reduce the bandwidth demands on the shared network resources and will also reduce the processing load on the client device that must render the video. However, the resultant video display will also be lower quality.

At other times, the video resolution (and associated bit-rate) might be allowed to be temporarily increased to react to a lighter level of congestion on the network or a lighter processing load at the client device. This increase in video feed resolution will increase the bandwidth demands on the shared network resources and will also increase the processing load on the client device that must render the video. However, the resultant video display will be higher quality.

Most Adaptive Streaming algorithms are based on the client monitoring the arrival rate of MP4 fragments or the client‟s processing load to trigger changes in the requested stream resolution. The different stream resolutions can be accessed by having the client perform HTTP GETs for MP4 fragments from different content files with different resolutions whenever Adaptive Streaming adjustments are required. As a result, in its traditional form, Adaptive Streaming is best suited for unicast feeds- either unicast VoD feeds or unicast Live feeds. When managed by the clients, it is not well-suited for IP multicast feeds.

<u>Head-End-Controlled Adaptive Streaming for Unicast Programs</u>

While client-controlled adaptive streaming is the most well-known form of adaptive streaming, there is another form of adaptive streaming that may provide even more benefits to MSOs. This alternative form will be called head-end-controlled adaptive streaming within this paper. The goal behind head-end-controlled adaptive streaming is similar to the goal behind client-controlled adaptive streaming: it is a technique for dynamically adjusting the resolution and bit-rate of individual video streams in an effort to ensure that the video streams are successfully delivered in a timely fashion to the video clients, even in the presence of network congestion. Head-end-controlled adaptive streaming does not consider or adjust resolutions and bit-rates as a result of changes in the processing loads placed on the clients. In essence, it assumes that the clients have adequate processing resources to render the video streams that they have requested.

Head-end-controlled adaptive streaming requires an omniscient control element in the MSO head-end to monitor and manage the bandwidth for each of the video flows being transmitted over the HFC plant. This omniscient control element could instantiated in one network element or a combination of network elements such as the CMTS, a policy server, an application manager, an IP Session Resource Manager, or some other intelligent node in the MSO head-end. In general, the omniscient control element should be cognizant of several things that could impact IP video Quality of Experience, including:

1) The topology of the HFC plant (including which DOCSIS channels are accessible by which video subscribers)
2) The available content and utilization levels of each of the IP Video servers
3) The real-time congestion levels on the HFC plant
4) The attributes of each video subscriber (such as the maximum processing capability, the video screen size, the subscription level, and the priority)
5) The policy rules that the MSO would like to enforce during periods of congestion (such as which subscribers or which screen sizes should be adaptively throttled first)
6) The triggering rules that the MSO would like to use to initiate the adaptive throttling of video stream resolutions and bit-rates

All dynamic streaming client devices feature a sizeable play-out buffer which “buys” time that may be needed for effective, pro-active intervention on the flow of adaptive streams by the head-end-control element. The size of these buffers vary from 7 to 70 seconds across the available HTTP delivery schemas, and some of this buffer time can be used to remedy any delay issues caused by the use of a head-end-based control element.

Once the need for throttling is identified by the omniscient control element, there are many ways to manage the throttling. For example, the omniscient control element in the head-end could communicate with either the video client or the video server to initiate the throttling functions.

Omniscience and the enforcement of intelligent policy rules and triggering rules are the keys to making a head-end-controlled adaptive streaming system work well. When using head-end-controlled adaptive streaming, the ability to recognize congestion, the ability to identify the causes of the congestion, and the ability to determine how to intelligently throttle a sub-set of the video streams should lead to improved Quality of Experience and improved bandwidth utilization within the end-to-end IP Video solution.

VBR Channel Bonding with Stat-Mux Gains

The effect of using bonded VBR streams has already been mentioned within the traffic engineering analysis above, but its value is worth mentioning again. If MSOs are willing to use VBR encoding for the future video streams and are also willing to send those video streams across a channel-bonded environment with at least three or four bonded channels, then they will experience an increase in the total number of programs that they can transport across their HFC plant (when compared to the number of video streams that could be transported if CBR encoding and un-bonded channels were utilized). [Bug1]

As an example, simulations with a specific set of twenty-one 20 Mbps, MPEG4 video streams have shown that the move from CBR encoding on non-bonded channels to VBR encoding on non-bonded channels helped reduce the total number of required channels from 11 channels to 7 channels. The move from VBR encoding on non-bonded channels to VBR encoding on bonded channels helped to further reduce the total number of required channels from 7 channels to 4 channels. These numbers included some "breakage effects," whereby the high-bandwidth, 20 Mbps video streams did not nicely fill the available bandwidth of the channels. But nevertheless, the overall effect is still quite significant.

What further improves the performance of VBR in an adaptive streaming environment is the buffering provided inside of the IP video client device. The buffers bring a new dimension of elasticity into content streaming. The duration of delivery of each fragment can be shifted, „stretched-out" or accelerated by the throttling mechanism during periods of congestion, and as long as the buffer is not "starved" by this process, all video sessions will be uninterrupted.

Fig. 5 and **Fig. 6** illustrate how the same fragments are being successfully delivered to the client device in both uncongested and congested downstream channels. The packet transfer curves may look differently in these two instances but the same fragments are successfully delivered to the client devices in time for smooth playout.

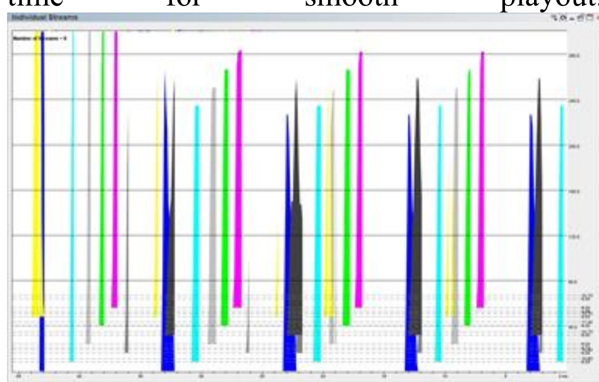


Fig. 5- Streaming of Fragments in an Uncongested Network

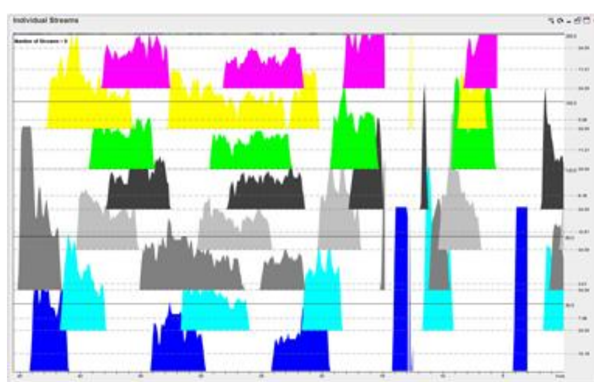


Fig. 6- Streaming of Fragments in a Congested Network

It is therefore possible to achieve a very efficient usage ratio on the HFC bandwidth (**Fig. 7**), measured by the Peak to Average Ration (PAR) of the composite (sum) of all streams. Such a high efficiency is not possible with the legacy MPEG2TS set-tops, because legacy MPEG2-TS set-tops do not tolerate much delay, jitter or wander within the stream without performing expensive, time consuming, quality-affecting transcoding for stat-muxing (typically 3 or 4 into one) at the head-end.

Our studies showed that a simple rule can be devised to calculate how many fragmented VBR streams can be reliably sent to the serving area over bonded DOCSIS3.0 channels. For calculations, the Average Bit Rate value must known for each of the streams. This value is typically set in the encoder for the live streams, and the actual average bit rate value can be easily calculated for VoD streams that are stored in servers. The Actual Average Bit Rate value of any VoD asset (called ActBR) can be derived by dividing the VoD asset size (in MB) by the duration of the asset (in seconds). The result needs to be then converted to Mbps (multiplying by 8) and is comparable to the ABR numbers of live streams.

The simple rule for calculating the required bandwidth inside of a <u>quad-bonded downstream</u> pipe (as shown on **Fig. 7**) is as follows:

(*Sum of ABR values of all live streams + sum of ActBR of all VoD streams) x 1.1*.

The 1.1 overhead factor, or 10% "tax" is a result of many simulations using the actual MSO content and allows the system to support fast „trick-play" on a limited number of devices, and also allows the system to support fast admission/start of several new IP video sessions.

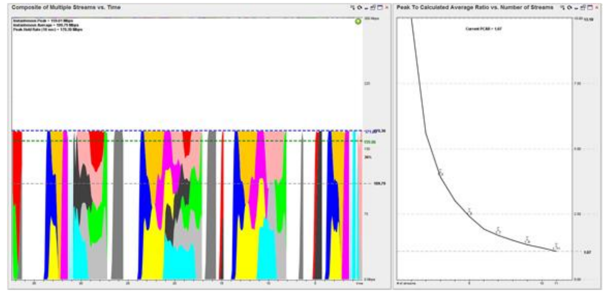


Fig. 7- Composite bit rate and Peak to Actual Average (decreasing as more Streams are added and combined) inside a Fully Loaded Quad-bonded Downstream

For a set of <u>octal-bonded downstreams</u>, simulations showed that the overhead factor ("tax") can be lowered to 1.05 (5%) as effects of self-averaging of VBR streams are more pronounced in the "fatter" delivery pipe.

<u>Hybrid Media Gateways with Both QAM and DOCSIS Receivers</u>

If the in-home gateway that receives the IP Video over DOCSIS signals has enhanced functionality that permits it to simultaneously receive both DOCSIS signals and MPEG-TS signals, then that enhanced gateway offers MSOs many new opportunities to deal with the Simulcast Roadblock issue.

One technique for reducing the number of simulcasted channels between the MPEG-TS tier and the DOCSIS tier is to only simulcast a fraction of the programs within the DOCSIS tier until all of the homes in a service group have been equipped with media gateways. For example, during the transition period, an MSO may choose to only transmit VoD streams and special interest Switched IP programs (ex: foreign language programming) to their media gateway subscribers over the DOCSIS feed. Those VoD streams and special interest Switched IP programs would not be sent over the MPEG-TS feed. As a result, the special interest Switched IP programs would not even be available to subscribers who still have legacy STBs that only receive MPEG-TS feeds. This could be used to encourage subscribers to sign up and pay an extra fee to get the new media gateway access which provides access to the special interest programming. Using this technique, the broadcast and SDV programs would only be moved to the DOCSIS tier when all subscribers in a service group were equipped with media gateway boxes, and the programs would no longer have to be transmitted on the MPEG-TS tier after that transition. It is possible and even probable that MSOs may actually leave a small set of broadcast or SDV programs in the MPEG-TS tier after the transition to support legacy STBs or DTAs, or to honor content or franchise contracts still requiring some analog or clear QAM channel transmissions. However, the availability of a hybrid media gateway can be used to greatly reduce the amount of simulcasting that occurs during the transition period.

Media Gateways with Large Numbers of DOCSIS Receivers

A subtle, but important problem that can lead to an increased demand for IP Video bandwidth can result if IP Video unicast feeds (often used for VoD service) and IP Video multicast feeds (often used for Switched IP service) are distributed across the channels in a Downstream Channel Set in an undesirable fashion. To give an example, consider the scenario shown in **Fig. 8**.

The IP Video streams are distributed across four Downstream Channels, but each media gateway has only two channels dedicated to accessing the IP Video streams. Media gateway x is accessing multicast streams A and B on Downstream Channels 1 and 2, respectively, while media gateway y is accessing multicast streams C and D on Downstream Channels 3 and 4, respectively.

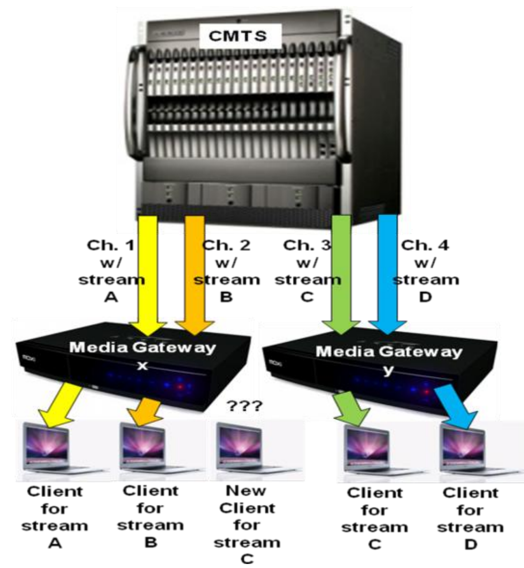


Fig. 8- Scenario with Multiple Video Multicast Streams on Different Channels

If a new video client behind the first media gateway (x) were to request the opportunity to Join and view multicast stream C, then the CMTS would have to perform some rearrangements to give media gateway x access to Stream C. The CMTS could (for example) create another replication of Stream C on either Downstream Channel 1 or Downstream Channel 2 to make it readily accessible to media gateway x, but this approach would tend to waste bandwidth on the HFC plant by transmitting multiple replications of the same Stream C content on the HFC plant. Alternatively, the CMTS could use DCC or DBC messages to force the media gateway to channel change its first receiver from Downstream Channel 1 to Downstream Channel 3 and then move already-flowing multicast video stream A to Downstream Channel 3 where the desired multicast stream C is found. However this channel change and flow movement could create a temporary disruption to the in-process video rendering of the video stream A. In addition, the movement of video stream A from Downstream Channel 1 to Downstream Channel 3 could have a domino effect if other media gateways were also receiving multicast video stream A, because their receivers would also be forced to change channels, and any flows besides video stream A that are being carried over Downstream Channel 1 to one of those media gateways would also have to be moved. Thus, many layers of problems can be identified with this proposed movement of stream A.

All of the problems mentioned in the previous paragraph can be easily circumvented if all of the media gateways were to be designed with more receivers. Broadband or full-band capture capabilities are becoming more accessible with the advent of higher-speed digital-to-analog converters, so many media gateway devices will be increasing the number of received channels in the future. In a perfect world, the media gateways would always have enough receivers to be able to tune to the number of DOCSIS Downstream Channels (IP video and HSD) that might be accessed by a typical home during the busy hour (which would typically be less than the total number of DOCSIS channels). As a result, it is clear that media gateways with larger numbers of DOCSIS receivers can help reduce multicast replications and can also help reduce the required bandwidth for IP video on the HFC plant.

Intelligent Assignment of Multicast Linear Programs to Specific Channel Sets

The problem outlined in the previous section illustrates how multicast video streams can cause exacerbated HFC bandwidth issues whenever extra replications of the multicast stream must be created to provide access to media gateways that may not be tuned to the channel on which the multicast video stream is being transmitted. As illustrated in the previous section, increasing the number of receivers on the media gateways can help to minimize this problem. However, if media gateways or modems are used that do not have these increased receiver counts, then other solutions must be found.

Another solution that could help minimize this extra multicast replication problem is described in this section. Basically, MSOs might decrease the number of extra multicast replications (and the amount of required HFC bandwidth) if they assign their most popular multicast streams to a specific set of channels in the HFC plant. In fact, it might be ideal to actually limit the most popular multicast streams to being placed on single, non-bonded service flows, while less popular multicast streams can be placed on bonded service flows to capitalize on statistical multiplexing gains. The theory behind this approach is that the popular Linear multicast programs are likely to be viewed by many subscribers, so if all of those programs are available on the same set of channels, then there is a high probability that a media gateway that is tuned to that set of channels to receive one popular multicast stream might also be asked (by another client in the home) to collect another popular multicast stream. If that additional multicast stream happens to appear on the same channel set as the first multicast stream that is already being received within the home, then extra replications are unlikely to be required.

As an example, the scenario shown in **Fig. 8** would not have been problematic if all of the program streams (A, B, C, and D) were all available on Channel 1, because both media gateways would have had the capability to be tuned to Channel 1. When the new client behind media gateway x requested access to program stream C, media gateway x would have easily accessed it on Channel 1 with the popular programs all assigned to Channel 1.

Intelligent Mixing of Compressible Traffic with Non-Compressible Multicast Linear Traffic

In most of the early-stage IP video deployment scenarios currently being planned by MSOs, the multicast Linear video streams may not be adjusting their resolutions and bitrates the way that unicast video streams might. (Note: Unicast video streams are expected to capitalize on techniques such as adaptive streaming right away, but complications in the delivery of multicast video streams may preclude the use of adaptive streaming in early deployments. As a result, the multicast streams might be considered to be "incompressible" streams during these early deployments). As a result, channels carrying only multicast Linear streams would not be able to throttle down the bit-rate on the different streams if the aggregate bit-rate from the multiple multicast streams ever burst to higher-than-normal levels.

In an effort to provide some ability to absorb and adapt to the transient bit-rate bursts that might develop on a channel carrying multiple multicast Linear streams, there may be a benefit to including some "compressible" streams within the channel. Compressible streams on a channel can include:

1) Unicast IP video streams with very deep receiver buffers
2) unicast IP video streams with adaptive streaming capabilities enabled
3) unicast HSD streams whose data can be delayed to lower the overall bit-rate.

Simulations were created to explore the effectiveness of this technique. In particular, a high-utilization (93.6%), 4-channel bonding group was simulated with different mixes of multicast Linear IP video traffic and unicast IP video traffic. The buffers for the unicast IP video traffic were designed to hold up to 10 seconds of video and the buffers for the multicast IP video traffic were designed to hold only 500 milliseconds of video. The percentage of errored video seconds was monitored, where an error occurred within a given second if a video packet was not available in the receiver‟s buffer at the time the packet was needed for rendering. The output of this simulation is shown in **Fig. 9**. It can be seen that with 15% (or more) of the video program bandwidth being unicast, the percentage of errored video dropped to negligible levels. Thus, it is apparent that this approach can be quite effective.

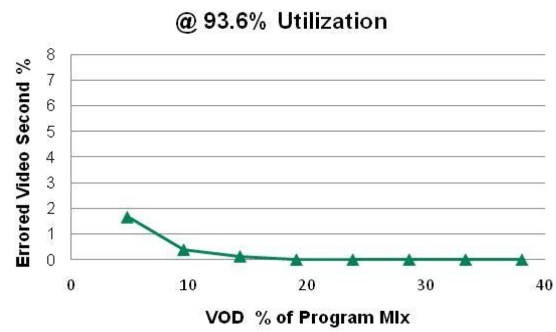


Fig. 9- Effect of Video Program Mix on % Errored Seconds

Acceptable Buffering of Unicast Video Streams at the Client

The examples in the previous section illustrate the generic benefits of using deep packet buffers in the video client. These buffers can act as "shock absorbers," providing IP video packets to the rendering engine whenever needed. If the buffer depth is made large enough, then the IP video delivery system can flywheel through transient periods of congestion leading to delayed packets and dropped packets (whose re-transmissions are delayed).

Low-bandwidth recovery mechanisms for corrupted multicast Linear programs

Linear multicast IP video streams will often be terminated at high-resolution TVs, so video impairments due to lost packets would undoubtedly reduce the viewer Quality of Experience level. Unfortunately, the use of multicast will typically require the Linear video streams to be transported using UDP instead of TCP, so re-transmission of lost packets (if implemented at all) will have to be handled by an upper-layer protocol.

If an upper-layer protocol packet retransmission scheme is developed, then it could be problematic, because the loss of a single multicast packet on the HFC plant prior to the arrival at the multiple receivers who have joined the multicast would cause each and every one of those receivers to initiate a separate re-transmission of the packet. These semi-synchronized requests for re-transmission could lead to a semi-synchronized burst of re-transmitted packets, which could cause a transient bandwidth spike on the HFC plant, which would be undesirable.

Several techniques have been proposed to reduce this problem. One technique requires the receivers to use a randomized delay before sending their re-transmission requests, which should serve to spread out the re-transmissions. Another technique requires the head-end server to recognize the first request for a re-transmission and to re-transmit the requested packet using multicast so that all of the receivers would receive the single re-transmitted packet. The other requests for re-transmission of the same packet would be ignored, since it can be assumed that they would have been satisfied by the single re-transmission.

Temporal Spreading of DVR Program Tuning Changes at the Half-Hour Epochs

Whenever a program tuning change occurs, there are often bursts of video packet transmissions that occur as each client attempts to pre-load its receive buffer. At half-hour epochs, many program tuning changes can occur in a short window of time, causing the video packet transmission bursts to become semi-synchronized. Unfortunately, DVR devices may produce a large percentage of the semi-synchronized program tuning changes at the half-hour epochs. One way to help minimize this transient problem is to ensure that DVR boxes managed by the MSO use a randomized delay before sending their tuning requests, which should serve to spread out the bursts at the half-hour epochs.

Intelligent Pre-Loading of Media Gateway Memory with VoD Content That is Statistically Likely to Be Viewed

HSD channels and IP video channels will likely cycle through periods of high utilization and low utilization throughout every day, and in most cases, the peak periods of high utilization will typically be at the same time for both the HSD channels and the IP video channels. (Note: This peak period usually occurs between 7pm and 10pm). The real HFC bandwidth challenge for MSOs will usually occur during the periods of high utilization, so most of the techniques for reducing the channel bandwidths will be aimed at performing their duties during the peak utilization periods.

Obviously, this implies that other types of activities can be supported during the low utilization periods. Perhaps techniques that help reduce the bandwidth during the peak utilization periods can take advantage of the extra bandwidth available during the low utilization periods.

One possible approach might attempt to pre-load VoD video content into DVR memories within the media gateway during the low utilization periods so that the VoD would not need to be transmitted over the HFC plant during the peak utilization periods. The trick is to try to predict which particular VoD-based video content might be viewed in the near future by a particular home. Information such as the successive episodes of a series previously watched in VoD can provide hints that help identify likely VoD candidates for the future.

Intelligent Pre-Loading of Media Gateway Memory with Linear Content That is Statistically Likely to Be Viewed

The content pre-loading methods described in the previous section can be applied (with slight variations) for other applications. For example, many MSOs are interested in ensuring that their IP video subscribers will have an above-average viewing experience. This will typically require support for features like:

1) Fast Linear feed channel changes (when the viewer tunes from one multicast Linear program feed to another multicast Linear program feed)
2) Start-over features that permit the viewer to re-start a Linear program feed if they join the program within a half hour (or so) of the feed‟s starting time
3) Trick-mode features that permit the viewer to pause, rewind, and fast-forward (up to the present time) a Linear program feed

There are many ways to provide the features listed above, but one common technique involves the launching of unicast IP video feeds to each viewer requesting the use of those features. This launching of separate unicast IP video feeds to each viewer can lead to a significant increase in the bandwidth capacity required for IP video delivery over the HFC plant. This would obviously exacerbate the Simulcast Roadblock problem. As a result, finding alternatives to the launching of separate unicast IP video feeds would be beneficial.

One alternative method requires the media gateway to pre-cache the Linear video streams as they are distributed via efficient multicast IP video feeds, so that the content is locally available in the home whenever the viewer initiates a channel change or start-over or trick-mode operation. This pre-caching obviates the need for the launching of additional unicast IP video feeds. Depending on the storage technologies (ex: DRAM, Flash, rotating disk) used in the different models of media gateways, there may be different memory speeds (which dictate the number of IP video streams that can be simultaneously written in parallel), different endurances (repeatable write cycles, which dictate the lifetimes for the media gateways), and different capacities (which dictate the total number of IP video streams that can be stored).

Since there will be these finite storage limits, the most challenging part of pre-caching multicast Linear IP video streams is determining which Linear streams to pre-cache and when to pre-cache them.

For example, a simplistic design might attempt to pre-cache every multicast Linear IP video feed that is passing down to the service group, but this design could suffer from many problems. Accessing all multicast Linear IP video feeds could place costly requirements on the number of tuners that must be supported in the media gateway. It could also place costly requirements on the speed and endurance and size of the storage technology that would be holding all of those streams. These added costs may not be acceptable to the MSO.

Thus, a carefully-planned design may be required for the pre-caching algorithms used in media gateways. As an example, one might consider a design that only pre-caches Linear IP video feeds during periods of time when HFC bandwidth congestion is expected, because use of pre-caching at other times may not be beneficial. There may be more than enough spare bandwidth to permit the use of unicast IP video feeds to provide fast channel changes, start-overs, and trick-modes during those periods of low HFC bandwidth congestion.

A carefully-planned design may also use history-based heuristics within each media gateway to try to predict the Linear IP video feeds that are likely to be accessed within each half-hour window. These heuristic algorithms might (for example) search back through the records to identify the particular Linear IP video feeds that were viewed by that home during the same half-hour window of time in previous weeks to try to predict the Linear IP video feeds that might be viewed during the same half-hour window this week. If that Linear IP video feed is requested by other subscribers within a service group, then it probably makes sense for the media gateway to pre-cache that feed.

Data analysis on real-world viewing habits was performed in an effort to determine the efficacy of heuristics-based algorithms of this nature. One algorithm used the most popular programs watched by this particular home in the same half-hour window 7 days ago and 14 days ago. This algorithm then added the most popular programs watched by the entire service group in the same half-hour window 7 days ago and 14 days ago. The results (describing the cache miss ratio) for media gateway caches of different sizes are illustrated in **Fig. 10**.

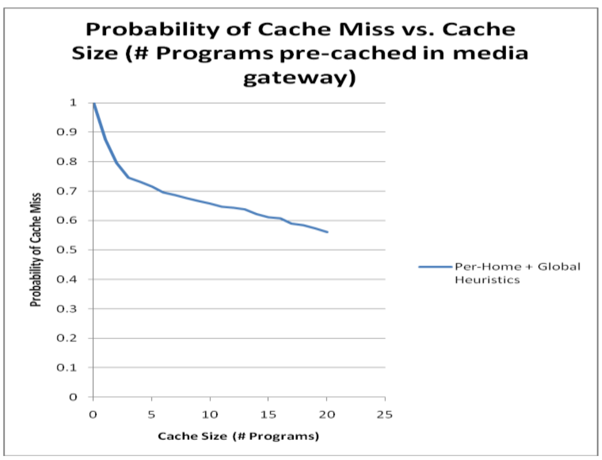


Fig. 10- Probability of Cache Miss vs. Cache Size (# Programs in media gateway)

As can be seen from the plot, these heuristics-based algorithms can give some assistance in the attempt to predict the Linear IP video programs that will be viewed in a particular half-hour window. If the media gateway can pre-cache up to 20 programs and the programs are selected using the aforementioned heuristics-based algorithm, then the data analysis predicts that the pre-cache programs would yield a 55% cache miss ration (=45% cache hit ratio). If the programs that are viewed during a particular half-hour period within a home are already pre-cached in the media gateway 45% of the time, then there is a good chance that channel change, start-over, and trick-mode events will find their desired content already stored in the gateway about 45% of the time. This could greatly help minimize the Simulcast Roadblock problem, because the use of pre-caching will reduce the total amount of required HFC bandwidth. In general, the additional unicast IP video feeds that would be launched if the content was not pre-cached in the gateway would not be launched if the pre-caching is performed in the media gateways.

CONCLUSIONS

In the first half of this paper, we outlined the various transition paths that MSOs might follow as they begin to migrate their networks to support new IP video services. The paper has also described and quantified (via an example) some of the challenges that MSOs may face as they try to deploy both legacy video service and IP video service on a single HFC plant infrastructure. In general, MSOs will need to find intelligent ways to navigate through bandwidth challenges that may face them as they begin to deploy both types of services. A particular problem known as the "Simulcast Roadblock" was defined to be the HFC bandwidth challenges that may result if MSOs try to transmit the same video streams in both their legacy MPEG-TS delivery system and their new IP video delivery system.

In the second half of the paper, we outlined a lengthy list of tools that MSOs can use to help them reduce the amount of HFC bandwidth required as they begin to deploy IP video services. These tools can be used to help solve the "Simulcast Roadblock" problem or they can be used in a general fashion to simply reduce the amount of bandwidth required to deliver a given subset of IP video program feeds to a service group.

These tools can be used in isolation or they can be used together to create a blended end-to-end solution for the IP video delivery problem. It is unlikely that any MSO will choose to use all of the tools mentioned in the second half of this paper, but MSOs may choose to use a sub-set that best suits their needs and the constraints created by their own HFC plant.

It is hoped that MSOs will find this toolkit to be a valuable resource as they begin their transition towards the deployment of IP video delivery systems in the upcoming future.

REFERENCES

[Bug1] M. Bugajski, "Use of Variable Bit Rate Video in Digital CATV Interactive Services," in Broadband, Vol. 31, No. 2, August, 2009.

[Clo1] Cloonan, Tom, "Bullpen: DIBA Reconsidered," Communications Technology, January 1, 2009, http://www.cable360.net/ct/strategy/businesscases/33178.html.

[Ins1] "The US Market for OTT Video Services", Instat Research, August 2010

[Pin1] Pinson, Margaret, et. al., "HDTV Subjective Quality of H.264 vs. MPEG2, with and without Packet Loss," http://www.its.bldrdoc.gov/pub/n3/video/ieee_10.pdf.

CONSIDERATIONS WHEN DELIVERING CABLE TV TO IP CONNECTED CONSUMER ELECTRONICS

Mark Francisco
Comcast Cable Corporation

Abstract

The number of connected consumer devices is expected to grow to more than 2.5 billion worldwide by 2014. The ability to directly deliver traditional Cable TV services to the most relevant classes of devices can provide greater opportunity, value and consumer satisfaction. The variation across and within multiple device classes present challenges when adapting content and services to the network link and capabilities of the video player.
This paper will introduce a variety of relevant device classes, their user interaction presentation and playback ecosystems. It will catalog cable TV features (i.e. closed-captioning, iTV, SAP and others) and discuss mechanisms to adapt to the device classes. An explicit objective is to maintain a good quality television viewing experience equivalent to that of a STB connected TV.

OBJECTIVES

Television as an Application

Cable television and similar Multi-channel Video Programming Distributors (MVPD) are accustomed to the concept of an Additional Outlet (AO) delivery model. AO's are typically extensions of the primary screen television service to additional televisions in the home and carry the same programming and features. The AO may offer a subset of the primary outlet services if the Cable Receiver or Television Receiver has limited capabilities. The growing popularity of online video viewing on connected consumer devices suggests direct support of an AO model to these devices is of consumer interest and benefit.

Connected consumer electronics include one or more IP-connection technologies, and the ability to integrally display video or support an external display through HDMI or similar digital interface. The devices include a user interface controlled by a remote control, keyboard/mouse, integrated buttons or touchscreen. The devices navigate and play video through an applications interface. The application environment may be standard or proprietary, with open or controlled access by service providers. Some of the most notable and popular video-supporting application environments today include (but not limited to):

1) Adobe Flash running within a PC browser
2) Apple iOS
3) Android
4) HTML 5
5) Microsoft Silverlight running within a PC browser

Adapting cable services to these environments requires a change of the traditional definition of television as a device or service to television as an application.

Principles of Television

With the objectives of extending MVPD services to connected consumer electronics, the characteristics that define television should be discussed. The following characteristics should be adapted in keeping with the traditional delivery model:

Television is available continuously and without interruption – Consumers expect television to arrive with the powering on of the device and remain until it is powered off. This characteristic will stress battery-operated devices and IP networks without broadcast capability. Heuristics should be employed to provide an instant, "always on" experience while conserving power and network capacity when the user is not actively viewing.

Television is Live – Live events are the cornerstone of television and maintain its highest viewership. Live television has explicit scheduling, minimal propagation delay and may experience high concurrent viewership. Balancing robust delivery methods, which may employ buffering against the desire for minimal delay from live, is a challenging task, particularly when delivering over unmanaged (capacity) network segments. The number of live television events delivered over the Internet using unicast delivery has increased, aided by emergent adaptive delivery methods.

Television is Multi-channel - The predominant behavior of television viewers is to watch programming at the first time of airing on the host network. Despite the lack of relevance of channel numbers identifying the means of tuning in the programming, established norms have maintained identity of networks via their historical frequency assignment. In other words, channel surfing, while increasingly supplanted by more effective discovery methods such as electronic program guides, search and recommendations, remains core to television viewing behavior. The requirement to zap, or rapidly change channels may be addressed through application or delivery techniques or more likely obsoleted through more effective discovery methods.

Television is Immersive – In the 20th century, television emerged as a focal point for living room gatherings of family and friends to watch live radio shows adapted to video. It subsequently offered prime-time entertainment, news and live sports in appointment-based viewing events that individuals would consider when planning their day. Only recently has television viewing become somewhat personal and associated with multi-screen, multi-tasking consumption. To satisfy traditional viewing habits, television should occupy the user's primary attention and deliver an immersive soundstage.

Television has Features and Control Requirements – Television programs are delivered with synchronous data and multi-program audio. The synchronous data may include teletext, alternate audio, program metadata, enhanced applications, alternate programming insertion triggers and control instructions. Programming may be copy protected. These features may include closed-captioning, emergency alerts, advertising insertion, descriptive audio, alternative language, parental advisories, rights management and interactive television. Several elements may be required through programmer contractual agreements and applicable transmission regulation.

While the television experience may be tailored to the specifics of a consumer device, usage and environment, adaptation of the core principals is possible and worthy of technical definition.

TELEVISION APPLICATION PLATFORM

Television has historically been delivered to NTSC or ATSC-based receivers with minimal variation. Viewer operation has moved from mechanical elements to remote controls and only recently been augmented with web-based discovery and control. Consumer devices offer challenges to traditional interaction and consumption and provide the opportunity to accelerate creation of new experience models.

Environmental Considerations

A number of factors differ between conventional HFC-based cable distribution of QAM video and IP delivery to consumer electronics. Most notable is the difference between managed and predictive capacity of MPEG-2 MPTS delivery over QAM versus the varying and contention-based method of delivery of IP networks. The second key difference is the variation in consumer electronics regarding video display capability, including resolution, frame rate, aspect ratio, and audio/video CODEC. Translation of video and audio formats is a requirement for most device categories. This may involve both spacial and temporal changes for video, and dynamic and encoding changes for audio.

User behaviors with connected devices may differ significantly from traditional television. Personal portable equipment such as smartphones and tablets are designed for mobile activity in short, but frequent intervals. Brief start-up time is a key requirement, as well as the ability to adapt to frequent changes in network and occasional loss of connectivity.

The third and perhaps most challenging area of adaptation is the concept of delivery to unmanaged devices. Consumer devices have a variety of operating systems, software stacks, application provisioning methods and native video pipelines. Ensuring application integrity and content security may involve detection of the characteristics of the device as configured and creation of secure enclaves within the device for provisioning of security related elements such as content keys.

TECHNICAL REQUIREMENTS

The reference architecture of Figure 1 proposes a system to realize the principles and environmental conditions. Elements are either under service provider control or customer supplied; therefore the two categories of requirements will be discussed separately. The requirements discussed below are derived in support of linear video applications, which may include broadcast content or content originating from video on demand servers. File-based delivery of content to consumer devices is a mature application and not described. While the technical solutions are proposed and discussed, the implementation timeframes and costs are not.

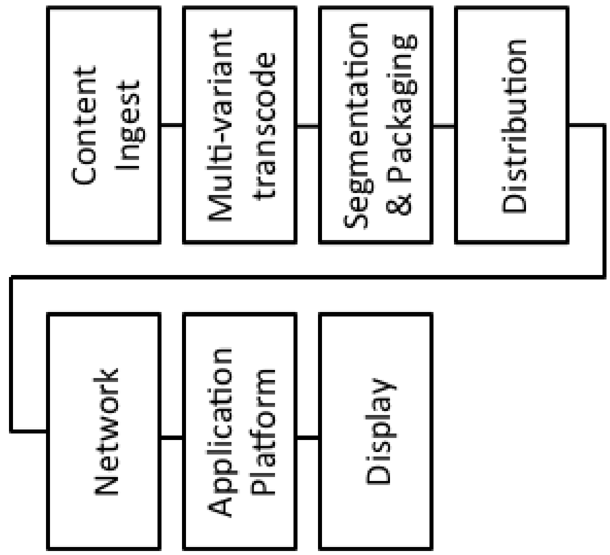


Figure 1. Reference Architecture

Service Provider Elements

Content arrives from many sources with most formatted for broadcast or Cable VOD distribution. This primarily MPEG-2 content requires transcoding into multi-variant H.264 profiles to reach the widest variety of devices and network conditions. Live content delivery across unmanaged networks requires adaptive delivery. The number and range of stream variants will be discussed in conjunction with the Technology Options section.

To maintain all associated metadata oriented features of the incoming video streams, the stream-associated information must be mapped from incoming transport to delivery technology to the client player protocol. Multiple translations will be presented within the Technology Options section. Content encoding is the element of the network with the lowest scalability and greatest use of custom hardware, lending to its separation from other elements in the delivery path.

The delivery path for video to consumer devices is expected to involve a home network, often wireless. TCP delivery of video content will be used, as home routers do not adapt well to UDP or multicast transport. HTTP delivery is widely adopted for video delivery to Internet connected devices as it is stateless, error resilient and traverses NAT.

Content security is an integral part of segmentation & packaging. Most content security clients rely on three main elements; content encryption, key management and authentication. Elements that need to be considered when defining a content security solution include integrity, ability to cache or store the content and device compatibility. Many popular content security systems rely on one of two AES-128 block cipher methods, cipher-block chaining (CBC) or counter (CTR). Both modes need to be considered due to fragmentation of support in the most popular connected devices.

Choice of key management and authentication systems is often dictated by the capability of the client. While a single key management / DRM system may be considered, an alternative is to unify the content encryption and adapt the key delivery to the client's native capability.

Customer Supplied Elements

The most widely deployed Internet connected video devices are PCs. This is followed by and soon to be eclipsed by smartphones. The third most popular connected video device is the game console. Other quickly growing categories include tablets, connected TVs and connected Blu Ray Players. Digital Media Receivers such as the Apple TV and Roku round out the most popular and relevant device list. With little exception, all include either an integrated display or support an HDMI-connected television.

Categorically, small-screen devices such as smart phones support low resolutions of 480p30 or less and require content encoded in H.264 baseline profile. Mid-size devices such as PCs, Tablets and digital media receivers typically support up to 720p30 and support H.264 baseline or main profile. Some high end PCs, digital media receivers and most connected TVs and BluRay players support up to 1080p60 resolution, which requires H.264 High Profile.

Content security in consumer devices is implemented as a link protection such as DTCP-IP or TLS (SSL), through application-based DRM or natively in the platform. Very few connected devices have native DRM capability that is available to service operators.

The application environment is an important aspect of adapting services to the consumer device. Most devices tie their application-provisioning environment to their native OS. The most popular application environments include Apple iOS, Android OS, Windows and MacOS. The application store process defined by Apple iOS and game consoles are the most constraining due to the requirement for certification prior to launch, while others are very flexible. In addition to application framework, many devices support presentation frameworks such as Silverlight or Flash.

Consumer-supplied elements provide the target platforms for delivery of Cable TV. The following sections "thread the needle" with content transformation, delivery method and application to complete the adaptation.

TECHNOLOGY OPTIONS

The core element required to deliver cable TV to connected consumer devices is the adaptive delivery technology. With the derived requirements stating the delivery will utilize HTTP, a number of ecosystems are candidates and will be introduced here.

HTTP Live Streaming

HTTP Live Streaming (HLS) was designed by Apple and submitted for standardization to the IETF. It is the most commonly used smartphone and tablet media framework in the United States and has been supported on all Apple iOS devices since the introduction of the iPhone 3G in July, 2008. HLS was initially used for streaming of file-based assets from sites such as YouTube and ABC but has since been applied to live video delivery. All video delivered to iOS devices must be presented using HLS in order to gain approval for 3G network use. Other mechanisms used for WiFi delivery include progressive MP4 download and PIFF delivery. Apple additionally limits content security to AES-128 CBC stream encryption with TLS key delivery.

HLS provides a robust delivery mechanism that can traverse most networks, but suffers from considerable latency and lacks features such as seamless content splicing and trick modes. It has not seen widespread adoption beyond iOS and Mac OS. HLS should be considered where necessary, as iOS devices are currently the most relevant for live video delivery, based on deployment and user interest.

A key benefit of HLS is the use of MPEG-2 transport stream for its container. This provides easier conversion from cable TV services but more importantly provides compatibility with all methods of carrying metadata such as captioning, ad triggers, and content advisories.

DLNA

The Digital Living Network Alliance (DLNA) specifies a set of device and media profiles that allow sharing of media between content sources and playback devices within the home. The Alliance assembles externally defined standards into interoperability guidelines and provides a certification program for manufacturers to receive approval to use the DLNA logo.

The DLNA is working on additional guidelines that will allow MVPDs to adapt subscriber content for delivery to DLNA players. These guidelines, which are being developed by the Alliance with service provider input, are planned for imminent release.

More than 8,500 devices have been DLNA certified but currently few support DLNA content protection. Adoptions of future guidelines are necessary to ensure DLNA certified devices are able to receive subscriber content. Cablelabs' OpenCable Home Networking Tru2Way (OCHN) extensions are DLNA device compatible, offering a good transition from in-home gateway delivered content to network IP sources of video.

The DLNA does not currently specify any adaptive delivery methods. It is anticipated that DLNA would include adaptive formats in updates to its guidelines.

Flash Streaming

Flash adaptive streaming is currently the most widely deployed Internet video technology as it is supported on 99% of PCs. In most cases, the stateful RTMP streaming mechanism is not usable through firewalls. The system defaults to a progressive download model in these cases. Flash adaptive streaming has found limited adoption in smartphone and tablet products. Progressive download methods are not well suited for live content delivery.

Adobe has recently released HTTP Dynamic Streaming (HDS), a version of Flash Streaming that supports dynamic streaming over HTTP connections. HDS requires Flash

10.1 or AIR v2 or later and is incompatible with RTMP origin servers. A content security technology called Flash Access is available for HDS streaming.

Flash support must be considered due to its widespread availability on PCs and integration into the Flash graphics presentation environment, although Silverlight is seeing growing adoption as an alternative. Advertising is almost exclusively distributed using Flash technologies.

WebM

Google has created an open source media framework entitled WebM. WebM is supported in Chrome Browsers, Android OS, and the Gstreamer open source media player. It includes support for an open source CODEC created by On2 Technologies called VP8. WebM has a plug-in structure with a relatively small group of components available for file-based on demand streaming. Live tools and features are not available at this time.

MPEG-DASH

MPEG-DASH is a multi-media delivery platform based on HTTP. DASH will deliver both MPEG-4 file and MPEG-2 TS based content. It is likely that HLS, DECE, 3GPP, and PIFF / SmoothHD will be supported by the DASH standard when completed. The MPEG-DASH specification is currently an ISO Draft International Standard (ISO/IEC DIS 23001-6) with an anticipated completion in July 2011 and release by end of year.

MPEG-DASH addresses HTTP delivery of streaming video/audio with adaptive features and supports both live and file-based streaming. It standardizes the container description information to ensure interoperability between servers and clients. DRM is not explicitly defined although support of DRM metadata is included in the description, making this a complement to the DECE content security framework.

MPEG-DASH shows promise as a useful and flexible method of adaptive delivery to the widest variety of consumer devices. DASH is proposing an HLS compatible profile in addition to a PIFF compatible profile.

IIS Smooth Streaming

IIS Smooth Streaming is a component of Microsoft's Protected Interoperable File Format (PIFF), a common file structure and adaptive delivery method for Silverlight and other HTTP clients. While PIFF is an open standard as per Microsoft's Community Promise license, Smooth Streaming is currently an element bound to IIS Origin servers. PIFF is based on the ISO MPEG-4 file format specification. Metadata may be carried as timed tracks. This will require mapping elements such as captions, ad triggers and advisory data from the traditional MPEG-2 transport stream mechanism to XML-based timed tracks.

PlayReady is the default DRM for IIS Smooth Streaming and supported in the Silverlight environment. Currently, Silverlight is implemented in Windows and MacOS, Xbox and PS3, many connected TVs and Windows Mobile Smartphones, but currently lacking in iOS and Android devices.

Technology Option Summary

A number of viable technology ecosystems have been introduced here, all with comprehensive feature sets but incomplete market availability and in some cases incomplete definition. The market is undergoing rapid, evolution and introduction of candidate technologies, further complicating selection. PIFF and MPEG-DASH show the most promise from a broad-scale adaptability but their ultimate adoption is unknown. Due to difference in core encryption methods between HLS, which specified AES-128 CBC and PIFF, which specified AES-128 CTR, common encryption will be a challenge. It may be possible to use a common key delivery method to decrypt content delivered in the device's native cipher mode.

PERFORMANCE CHARACTERIZATIONS

To achieve the goals of fast video acquisition time and uninterrupted playback, tradeoffs between elements related to video quality and elements related to robust delivery are required. Elements related to video quality include resolution, compression profile, frame rate, and bitrate. The elements related to robust delivery include GOP size, segment size, and the step size between variants. Ultimately, the client will determine the limits on the values for these items and will contain internal heuristics, which may dictate the most favorable combinations.

A set of experiments was conducted using a transcoder, web server and iPad to explore the impact these parameters have on an actual device. The results included in Figure 2 and 3 show a direct relationship between elements that increase the size of segments and the stream acquisition time. While results tend towards improved performance more frequent segmentation and smaller file size, other issues may result when serving a larger number of small files, given overhead requirements on a per file basis. In order to balance the requirements of video quality and scalability, it may be possible to direct the initial content acquisition to use a fast-access profile and allow it to adapt to higher quality profiles a short time after the start of streaming.

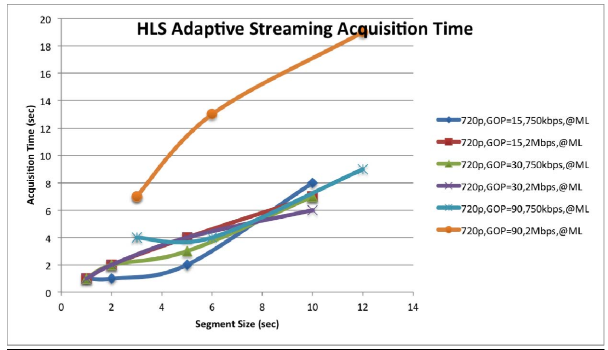


Figure 2. HTTP Streaming Acquisition Time – 720p Resolution

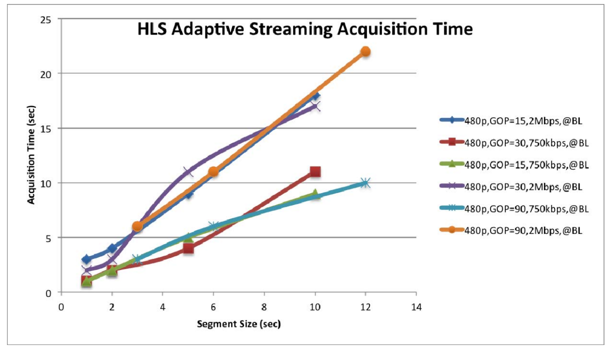


Figure 3. HTTP Streaming Acquisition Time – 480p Resolution

Other performance topics to discuss include captions and secondary audio programming. The method to convert captioning differs between transport stream-based (i.e. PIFF) and file-based (i.e. HLS) adaptive technologies. When using a transport stream-based adaptive technology, ATSC-A53 or SCTE-21 carriage of broadcast EIA-708 captions is possible. File-based adaptive technologies require conversion of captions to W3C timed-text. In either case it will be the responsibility of the client video player to display the captions. The client video player may require specific treatment in the client application to enable captioning. In a similar fashion, SCTE-18 ad splicing or SCTE-35 emergency alert triggers may be maintained in the transport stream, or mapped to a specific track in the MPEG-4 file structure. The application data delivery is not currently standardized in any file-based adaptive technology. Again, the client application is responsible to act upon the application data and take necessary actions, such as to display an ad or change to an emergency broadcast channel.

Alternate audio is standardized in transport-stream delivery although not currently handled in Apple's iOS player. Audio is limited to a single alternate in MPEG-2. PIFF offers a mechanism called late binding, where one of any number of audio alternatives may be joined to the video stream within the player environment. This feature is of interest particularly to support descriptive audio for hearing impaired viewers.

SUMMARY

A method of adapting Cable TV delivery to IP connected consumer devices is presented. Due to fragmentation in the video delivery technologies supported by popular devices, a multi-ended architecture is recommended. With the scalability and simplicity of container adaptation, a set of encoding variants can readily be packaged on a per stream basis. A replication of Cable TV on connected consumer devices can be achieved with appropriate mapping of channel associated metadata and application handling in the client.

REFERENCES

[1] HTTP Live Streaming, draft-pantos-http-live-streaming-06, March 31, 2011, http://tools.ietf.org

[2] Portable encoding of audio-video objects The Protected Interoperable File Format (PIFF) http://go.microsoft.com/?linkid=9682897

[3] HTTP Dynamic Streaming, Adobe Website, http://tinyurl.com/25mtnwj

[4] DLNA Interoperability Guidelines (member only), http://www.dlna.org

[5] DLNA Enables Premium Commercial Content Across Home Networks, http://tinyurl.com/3jownmf

[6] Webm, an open web media project, http://www.webmproject.org/

[7] Information technology — MPEG systems technologies — Part 6: Dynamic adaptive streaming over HTTP (DASH), http://www.iso.org

[8] US Connected Devices Sales (2009-2015), Parks Associates

Evolving Optical Transport Networks to 100G Lambdas and Beyond

Gaylord Hart
Infinera

Abstract

The cable industry is beginning to migrate to 100G core optical transport waves, which greatly improve fiber utilization while lowering transponder count for equivalent transport bandwidth. However, transporting 100G waves requires complex optical modulation to preserve performance and increase spectral efficiency. These modulation methods require several additional discrete optical components per lambda, so the migration to 100G waves alone does not fully address the scalability issues of increasing cost, power, space, and heat as bandwidth requirements continue to grow.

Optical networks are rapidly approaching the point where continual scaling of higher bit rate optical lambdas using discrete optical components is reaching its limits. Photonic integration, which combines multiple optical subsystems on a single IC, can efficiently support complex modulation schemes without increasing component counts and can offer significant improvements over discrete designs, providing a scalable path for future growth.

This paper provides an overview of complex optical modulation methods for 100G waves and higher. An analysis of discrete component solutions and photonic integration solutions is provided which demonstrates the impact of photonic integration in reducing component counts, heat, space, and optical coupling requirements.

INTRODUCTION

In 2014, the Internet is predicted to handle four times more traffic than it did in 2009 (see Figure 1, below). This is largely driven by the explosive growth of Internet video services, which are anticipated to be over fifty percent of all consumer Internet traffic by 2012, and by mobile data services, which are predicted to grow at a 108% compounded annual growth rate (CAGR) through 2014.

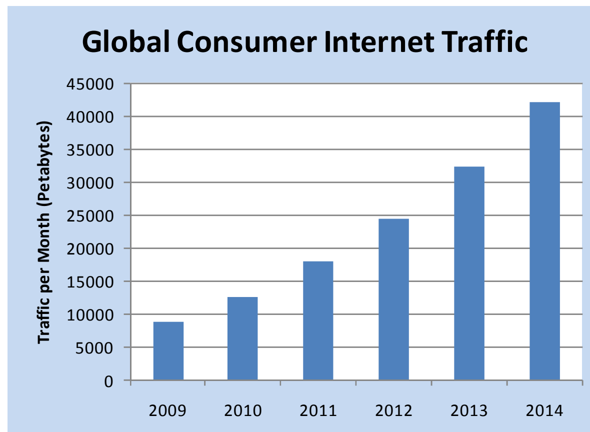


Figure 1—Global Consumer Internet Traffic

Cable subscription Video on Demand (VoD) traffic will also grow significantly over the next few years, more than doubling from 2011 to 2014 (see figure 2, below).

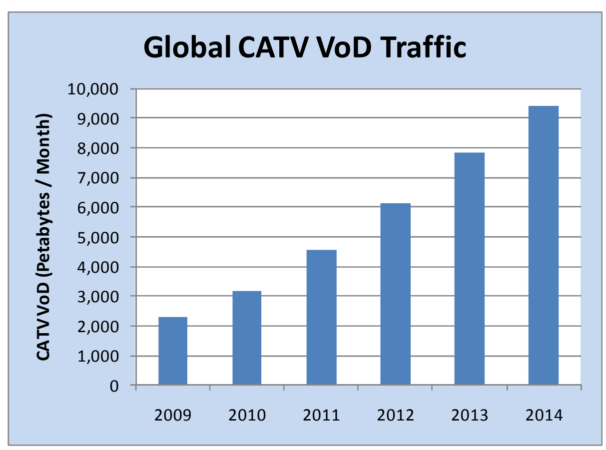


Figure 2—Global CATV VoD Traffic

The combined growth of high speed Internet access and VoD traffic alone will require an extensive expansion of transport capacity in cable networks. Currently, most high-speed optical transport is carried over 10 Gb/s waves using DWDM on a pair of fibers (Tx and Rx). Using 160 x 10G DWDM lambdas using double density spacing at 25 GHz, a single fiber pair can support 1.6 Tb/s capacity, or half that if 50 GHz spacing is used.

Typically, if additional bandwidth is required, another fiber pair is lit up in parallel to the existing lit fibers. This is relatively easy to accomplish if sufficient fiber pairs are available for growth, but lighting up new fibers requires additional chassis and common equipment to support the new fiber link, and incurs additional installation and operational costs associated with increased power, space, and cooling requirements. If additional fiber pairs are not available, considerable additional costs and time delays are usually incurred when new fiber must first be laid.

This linear approach to bandwidth growth, while workable in the short term, does not fundamentally lower the cost per bit for transport over time, nor does it improve scalability by making more efficient use of existing resources (e.g., fiber) or reducing resource consumption (e.g., space and cooling requirements). While fiber relief can be achieved by increasing the DWDM channel count per fiber or by expansion into the L-band, this does not sufficiently address the broader issues of scalability.

As cable operators' optical transport bandwidth requirements continue to grow at a compounded 50-100% per year, network growth cannot continue to scale on a linear basis simply by adding more optical transponders and add/drop multiplexers. Aside from equipment costs, the required space, power, cooling, and optical couplings rapidly become unmanageable when using a linear network expansion model to support geometric bandwidth growth.

SCALABILITY CONSIDERATIONS

A scalable transport solution accomplishes two things over time. It increases the transport capacity per fiber and lowers the total transport cost per Gb/s. The transport cost per Gb/s is comprised of several components, including power, space, hardware, and reliability costs. Additional factors which may contribute to these costs include bandwidth efficiency, the ability to minimize stranded bandwidth in the network, and the ability to simplify and accelerate operational process associated with the network. An ideal solution will make improvements across all these cost components as technology evolves and the network scales to higher transport bandwidth.

INCREASING TRANSPORT CAPACITY

To increase fiber transport capacity, two approaches are generally available: increase the DWDM channel count available on the fiber or pack more bits into the existing channels. Options for bandwidth expansion in the C-band are limited due to amplification requirements and the fact that most of the ITU-grid C-band is already used by existing transport systems. Expansion in the L-band is also possible, but this requires the use of special amplifiers. Consequently, most vendors have focused on increasing the capacity per DWDM channel.

Transport platforms supporting 40 Gb/s lambdas have been available for some years now, and 100 Gb/s lambdas have also recently been deployed. However, according to industry analyst Ovum, the cost per bit for 40G line side lambda transponders is not expected to drop below the 10G lambda cost until 2015, and the cost per bit for 100G line side lambda transponders is not expected to drop below the 10G lambda cost until 2014. So for the near term, 10G transport costs will remain below 40G and 100G unless fiber exhaust is a factor.

The cable industry has recently begun migrating to 40 and 100 Gb/s transport waves, which has been accelerated by the ratification in June 2010 of the IEEE 802.3ba standard for 40 and 100 Gigabit Ethernet. 40G waves can be deployed using 25 GHz spacing, which means a 160 channel system can support 6.4 Tb/s per fiber pair. 100G waves are typically deployed using 50 GHz spacing, which means an 80 channel system can support 8 Tb/s per fiber, but with half the transponders as the 40G approach.

One important consideration for cable operators is the ability to seamlessly migrate to 40G and 100G transport lambdas on today's 10G networks without having to re-engineer or upgrade the network, including leaving existing dispersion compensation and amplifiers in place. To meet the dual objectives of increasing bandwidth per DWDM channel and operating these lambdas on existing 10G networks, most DWDM equipment uses higher order modulation methods for 40G and 100G lambdas to retain the current channel spacing and increase spectral efficiency.

HIGHER ORDER OPTICAL MODULATION

Most early DWDM systems used 1G or 2.5G lambdas which were modulated using NRZ OOK modulation (non-return-to-zero, on-off keying) applied to the laser light source. This is the simplest form of amplitude shift keying in which a one is represented by the presence of light and a zero by its absence (or vice versa). This modulation technique uses a minimal number of components, but is not very spectrally efficient. The migration from 2.5G line side waves to 10G was fairly straight forward and did not require major changes or developments in modulation technology. The NRZ OOK modulation was simply speeded up to accommodate the higher line rate, which only required faster components. This is still the most common modulation used for 10G waves. Figure 3, below, shows a block diagram of the components required to transmit and receive a single NRZ OOK optical signal.

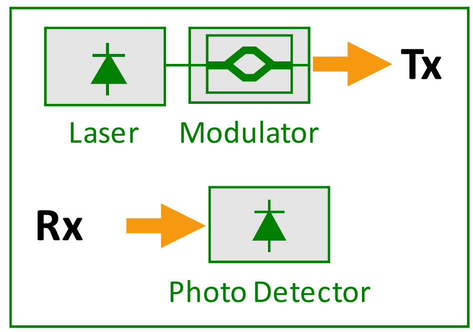


Figure 3—NRZ OOK Transmitter and Receiver Optical Components

As shown in the block diagram, the transmitter consists of a continuous wave laser and a simple OOK modulator. The receiver is even simpler, requiring only a photo detector.

One side effect of using OOK for 10G modulation is that the transport spectrum is wider than for 2.5G, while the bit period is shorter, which makes chromatic dispersion (CD) a more significant factor for 10G transport than for 2.5G. Depending on the transport distances encountered, it is common to use dispersion compensation at regular intervals along the fiber path to mitigate the effect of CD on 10G transport.

Similarly, higher bit rate transmission is more susceptible to other impairments as well, and many of these effects increase with the square of the bit rate. Two of these limiting factors are polarization mode dispersion (PMD) and optical signal to noise ratio (OSNR), which must be taken into account in optical network design for optimal BER and reach performance.

Using NRZ OOK modulation above 10G line rates, these optical impairments begin to severely impact DWDM transport performance. Because these impairments are related to symbol rate rather than bit rate, most DWDM vendors are using higher order modulation methods for 40G and 100G wave transport, which effectively reduces the symbol rate while increasing the bit rate, and thus minimizes the negative impact of these impairments. Higher order modulation also is spectrally more efficient, which means the information capacity of the fiber is improved as well.

There are many higher order modulation formats which can be used for optical transport, but the most commonly used today rely on some form of phase shift keying (PSK) to translate bits into optical phase states for transport across the network. Table 1, below, provides the bits per symbol for common optical modulation schemes. The more bits per symbol encoded in transmission, the more efficient the modulation format is. As seen in the table below, BPSK encodes one bit per symbol, so the baud rate is the same as the bit rate. QPSK, which is twice as efficient as BPSK, encodes two bits per symbol, so the baud rate is half the bit rate.

Modulation Format	Bits / Symbol
BPSK	1
QPSK	2
8 PSK	3
8 QAM	3
16 QAM	4

Table 1—Bits per Symbol for Common Modulation Formats

Further improvements in spectral efficiency and reduction of symbol rate can be achieved using polarization multiplexing on the optical signal. In this case, the transmitting laser's output is split into two signals, and the polarization of one of these is shifted 90 degrees before being modulated. Each polarized signal is then modulated separately and then combined at the transmitter output for transport over a single fiber. A block diagram of the polarization multiplexed transmitter and receiver is provided below in Figure 4.

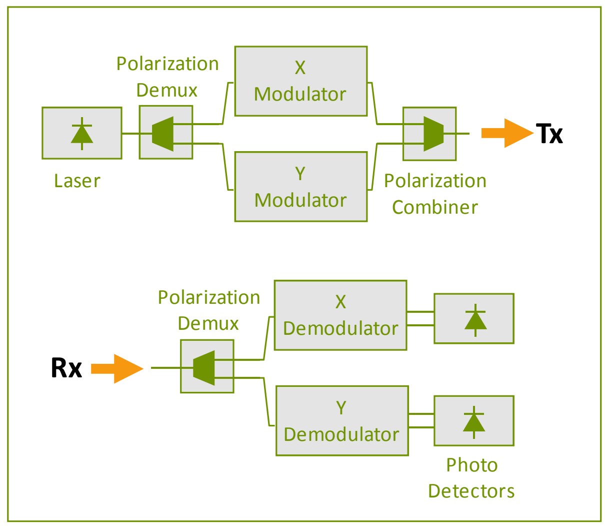


Figure 4—Polarization Multiplexed Transmitter and Receiver

On the receive side, the incoming optical signal is split into the original two polarized signals using a polarization demultiplexer, and then each is independently demodulated. The two recovered bit streams are then combined to reproduce the original data stream used to modulate the transmit laser. Polarization multiplexing may be used with any of the common modulation formats, and for a given transport data rate it effectively increases the bits per symbol for a given modulation format by a factor of two.

Most DWDM transport equipment uses QPSK modulation for 40G and 100G transport waves. While QPSK is roughly twice the complexity of BPSK and requires about twice as many components to implement, it represents a good tradeoff between spectral efficiency and OSNR performance. For even better performance, polarization multiplexed QPSK (PM-QPSK) is typically used, though at the expense of greater complexity and increased component count.

Further transmission performance enhancements can be achieved if coherent detection is used at the receiver. In differential QPSK (DPQSK), detection is performed by measuring changes in the phase of the received signal rather than the absolute phase itself. Because a copy of the originating reference signal is not required, DQPSK receivers are much simpler to implement. However, DQPSK receivers provide a lower level of performance which translates into higher BER or shorter reach when compared to coherently detected QPSK. This difference is usually sufficient to justify the use of coherent detection, and the Optical Internetworking Forum (OIF) has standardized on coherent PM-QPSK for 100G wave transport.

(sometimes referred to as synchronous detection) requires a local copy of the transmitter's original CW carrier to perform direct phase detection. In the optical domain, this requires a separate local laser at the receiver (equivalent to a local oscillator in the RF domain) which is phase locked to the incoming optical signal to create a copy of the original optical carrier.

Coherent detection is usually implemented with an ASIC which integrates the A/D conversion for the optical detector outputs as well as a Digital Signal Processor (DSP), AGC controller, and other functions on the ASIC. Using digital signal processing in the coherent receiver enables other features as well, including electronic dispersion compensation (EDC). EDC is far more tolerant than fiber based dispersion compensation and can readily provide compensation for +/-50,000 ps/nm without the use of bulky dispersion compensation modules and their inherent attenuation. Similarly, PMD performance is greatly improved when coherent detection is used with digital signal processing, and a DGD tolerance of 200 ps peak can be achieved for 100G waves.

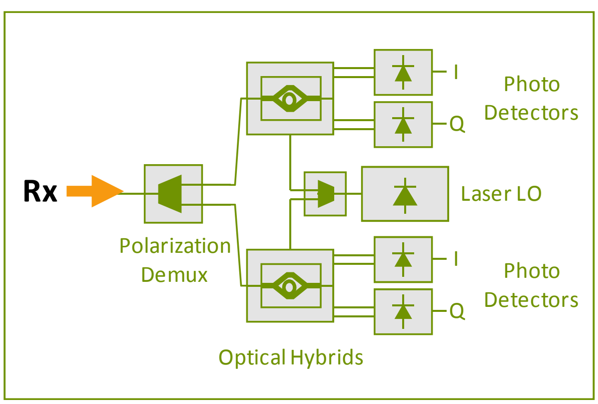


Figure 5—PM-QPSK Coherent Receiver Block Diagram

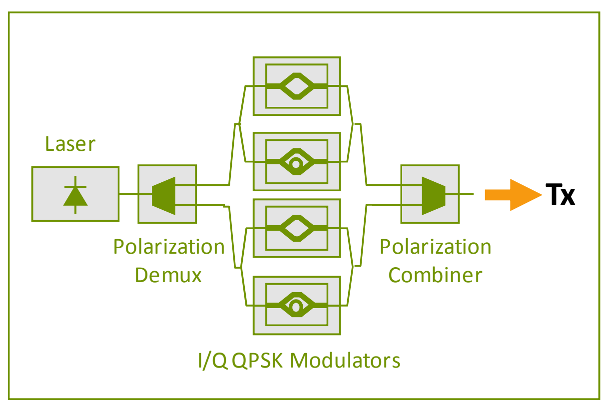


Figure 6—QM-PSK Transmitter Block Diagram

Figure 5, above, shows a block diagram of a PM-QPSK receiver. Coherent detection

Figure 6, above, shows a block diagram of a PM-QPSK transmitter. The source laser signal is split into two polarized sources 90 degrees out of phase. These are then QPSK modulated, creating I and Q modulation output components, and these are then combined for transport on the fiber.

One additional benefit of DSP-based coherent detection is the ability to provision the modulation format in software. Thus the transmitter and receiver pair can be designed to be operated using either BPSK or QPSK, for example, and the operating mode may be selected by the operator based upon link requirements. BPSK supports longer reach at a lower data rate for ULH applications, where QPSK is optimized for higher data rates with some compromise in reach.

PHOTONIC INTEGRATION

As one can see comparing the number of components required for NRZ OOK transmission of a single lambda in Figure 3, above, with the number required for PM-QPSK in Figures 5 and 6, also above, it is apparent that a significant number of additional optical components and fiber connections are required per lambda to support 40G and 100G wave transmission. If one considers transmission of 80 x 100G or 160 x 40G lambdas per fiber, the number of required additional components and fiber couplings is quite large. This has a direct and negative impact on reliability, power consumption, space, and cooling requirements, all of which lead to higher transport costs and a lower degree of scalability.

In practice, some of these discrete components may be combined in a single package using either hybrid construction (where two or more discrete components are placed in the same package) or small-scale integration (where two or more components are monolithically manufactured and packaged together). However, this approach offers only limited relief since it does not address the entire receiver and transmitter subsections or the DWDM system as a whole, which requires multiple lambdas as well as mux/demuxes.

Ideally, as optical transport networks scale in bandwidth by using higher order modulation techniques, new technology should be available to mitigate the effects of the increased complexity and component counts. Fortunately, this is indeed the case. Modern photonic integration allows multiple optical subsystems to be monolithically manufactured on an Indium Phosphide (InP) chip using large-scale integration. It is possible today to put all the optical components necessary to support multiple lambdas, including the mux/demux functions, on a single photonic integrated circuit (PIC) and address integration at the system level rather than the component level.

Commercially deployed PICs for 10G transport have been available for NRZ OOK modulation since 2005. These currently support 10 x 10 Gb/s DWDM wavelengths on a pair of PICs less than 5 mm square each (complete TX and RX systems, including mux/demux functions, 100 Gb/s per PIC). PICs which support 5 x 100 Gb/s DWDM wavelengths using PM-QPSK modulation and coherent detection have already been demonstrated in long haul networks and are planned for commercial availability in DWDM systems in 2012.

PICs provide the unique benefit of integrating not only the discrete optical components on the IC, but the optical connections between them as well. Table 2, below, provides a summary of the currently available 10 x 10G PICs, including the number of functions integrated on the Tx and Rx PICs, the number of equivalent discrete component “gold box” packages eliminated, and the number of fiber connections eliminated.

10 x 11G NRZ OOK PICs	100G PICs
InP monolithic ICs (Tx & Rx)	2
Integrated optical functions	62
“Gold box” replacements	20-50
Fiber coupling reduction	>100

Table 2—100G PIC Large-Scale Integration

Modeling an 80 x 10G lambda terminal configuration node (800 Gb/s), photonic integration, when compared against discrete transponder based solutions, can save up to 45% on power costs while reducing the required rack space from three 7 foot bays to one 7 foot bay.

The benefits of photonic integration only increase as more complex modulation methods are used to provide more bandwidth per lambda. Table 3, below, shows the corresponding integration summary for 500G PICs supporting 5 x 100G lambdas using PM-QPSK. This represents a fivefold increase in bandwidth and a fourfold improvement in fiber coupling and discrete component reduction when compared to today’s 100G PICs.

5 x 114G PM-QPSK PICs	500G PICs
InP monolithic ICs (Tx & Rx)	2
Integrated optical functions	>600
“Gold box” replacements	>100
Fiber coupling reduction	>400

Table 3—500G PIC Large-Scale Integration

Moore’s law indicates significant improvements are yet possible in future generations of PIC technology, and 10 x 100G Terabit PICs have already been produced and tested successfully in the lab. Current models predict PIC bandwidth will double about every three years, keeping pace with bandwidth growth while lowering the overall cost per bit.

EVOLUTION OF THE TERABIT PIC AND MULTI-CARRIER SUPER-CHANNELS

As the industry migrates to Terabit optical transport, there will be an ever-increasing need to squeeze more bandwidth from fiber and yet to retain flexibility on how that bandwidth is used. Modulation formats will continue to evolve, with 8QAM and 16QAM being next in line for extending transport bandwidth. At the same time, it is desirable to retain the flexibility to support multiple modulation formats on the same fiber, which will allow seamless migration to higher bandwidth and allow suitable tradeoffs to be made for optical reach versus total fiber capacity.

Current DWDM systems are based on an ITU channel grid which provides for lambda spacing of 100, 50, or 25 GHz. Inherent in this spacing are dead zones to allow for optical filtering of individual wavelengths, and these dead zones limit channel density, which in return can result in up to 50% of the available fiber spectrum being unusable.

In future DWDM transport systems, it will make more sense to move beyond the current ITU channel plan and implement multi-carrier super-channels that eliminate the dead zones for the carriers within the super-channel, but which preserve a guard band at the edges of each super-channel for filtering purposes. Multiple sub-rate carriers can readily be implemented within the super-channel using PIC technology which also allows multiple modulation formats and flexible channel spacing to be supported and provisioned in software. Such DWDM systems should enable transport capacities of up to 25 Tb/s per fiber, well beyond the capacity of today’s systems.

SUMMARY AND CONCLUSIONS

As transport bandwidth requirements grow at a 50-70% CAGR, DWDM transport systems will have to migrate to complex modulation methods such as PM-QPSK to conserve fiber spectrum and increase the transport capacity per lambda. Complex modulation requires substantially more optical components per lambda than the current 10G transport lambdas require. This increase in components is not sustainable in the long run due to the cost, space, heat, power, and reliability requirements associated with discrete component implementations of DWDM optical transport systems.

Photonic integration, which combines onto a single VLSI chip the entire optical subsystems needed to transport multiple lambdas using complex modulation, provides a scalable technology that enables future migration to Terabit PICs which will lower the overall transport cost per bit while reducing space, power, and cooling required per bit.

NOTES

Source for bandwidth growth forecasts: "Cisco Visual Networking Index: Forecast and Methodology, 2009-2014," June 2. 2010.

"Optical networks volumes and revenue history and forecast, 2006-2015," Ovum, September 21, 2010.

Author Contact Information:

Gaylord Hart
Infinera
303.840.9381
ghart@infinera.com

Optical Transmitter Technology for Next Generation Access Networks

Phil Miguelez
Director, Access Network Architecture
Motorola Mobility, Inc.

Abstract

Initial Docsis 3.0 deployments increased during 2010 and are expected to accelerate rapidly over the next two to three years putting pressure on cable operators to allocate bandwidth in both the downstream and upstream paths. The number of added narrowcast QAM channels continues to steadily increase as a result of SDV deployments and the shift to HD content. The anticipated introduction of CMAP (Converged Multi-media Access Platform) equipment estimated for 2012 is also expected to dramatically amplify the need for even higher numbers of additional QAM channels and with it the requirements for a cost effective means to deliver these channels to smaller node serving areas and targeted customers. On top of this, new business services opportunities, cell tower backhaul, and WiFi access point deployments require a growing share of plant access network capacity.

The need for more bandwidth never sleeps which causes more than a few cable system operators to lie awake at night as they try to determine which of the numerous alternatives to meet this never ending challenge is the most cost effective and future proof. The consensus has gradually shifted away from the traditional path of expanding plant RF bandwidth and is now moving to take advantage of the broad WDM segmentation capacity of existing fiber. Facilitating this change is a wide range of fiber architectures and Headend optical transmitter technologies ranging from expanded capability QAM lasers to new 10 Gb devices that provide a bridge between today's analog / QAM transport requirements and the high speed IP delivery needs of the future.

This paper provides a comparison of these currently available laser transmitter technologies identifying the differentiating features and limitations of each design type along with link application examples. The impact of the all digital channel loading transition and converged services on the performance, reach, and network cost using these technologies will also be examined.

INTRODUCTION

The common complaint of most cable operators today regardless of size is that they are nearly out of downstream bandwidth. The growth of DOCSIS® services for small business and home office customers and the popularity of streaming video and downloading movies over the internet is stretching the available data capacity of existing nodes. At the same time the forecasted requirements for new Narrowcast (NC) channels is 2 to 3 times the number allocated in current systems.

The traditional path for expanding cable access bandwidth has always been to manipulate the available RF spectrum. Until recently this was accomplished through expensive wholesale upgrades of the access plant equipment extending the high frequency edge of the band to the current 750 MHz, 870 MHz and even 1 GHz networks that exist today. This approach was driven primarily by the broadcast analog video channel loading that is still considered by many MSO's to be a major positive differentiator between HFC cable and competitor systems.

The introduction of digital QAM channels provided more efficient use of the 6 MHz RF channel space allowing 10 to 15 standard definition digital video programs or 2 to 3 HD programs per downstream channel. Digital technology also opened the door to video on demand (VOD), high speed data (HSD), and other narrowcast services that could be targeted to specific subscribers or smaller serving areas within the MSO network. This actually caused a further increase in the number of possible channels and the pressure for more access plant bandwidth.

The cost and disruption of cable plant upgrades in order to increase bandwidth has created an adverse environment for the shares of publically traded cable companies. This has driven cable operators to implement a number of alternative, lower cost, incremental solutions including reclaiming analog channels, deploying Switched Digital Video (SDV) systems, and when necessary, selective node splits. These options have helped to extend the life of the current legacy HFC networks but the need for more bandwidth continues to accelerate. The potential cost / benefit gains from further RF bandwidth enhancements are limited. As a result, cable operators have shifted their focus and are now actively evaluating optical segmentation solutions that will provide increased BW per subscriber plus enable business services growth by reclaiming fiber for point to point applications such as cell tower backhaul. The technologies that offer these benefits are described in the following sections.

The Evolution of HFC Optics

Today's HFC optics are primarily dominated by point to point fiber links transporting the full downstream program channel load from hub to node over a single 1310 nm wavelength. Upstream traffic utilizes a second fiber in most common configurations. The selection of 1310 laser transmitters to transport a full spectrum of analog video and QAM signals is not a coincidence. The commonly deployed fiber in HFC networks is SMF28. This fiber type is a recognized standard with characteristics defined by ITU-T G.652 and numerous other standards organizations. An important property of SMF fiber is its zero dispersion value at 1310 nm. By operating in a zero dispersive media the detrimental effects of DFB laser chirp are completely mitigated improving the distortion performance of the link.

1310 DFB lasers are the workhorse of current HFC networks. The linearity of these devices has been optimized over many years such that the analog distortions are well controlled and with simple predistortion correction, meets and exceeds the requirements for end of line performance. 1310 DFB lasers are available in a wide range of output levels typically up to 15 dBm. This is sufficient for link reaches of at least 40 km based on the average 0.35 dB / km attenuation of SMF28 fiber. As long as additional fiber is available, segmenting traditional 1310 HFC networks is accomplished by adding new hub transmitters and node receivers connected using dedicated fibers as depicted in Figure 1 below.

Figure 1 – Typical 1310 Hub to Node Downstream Optics

For link distances longer than 40 km or areas that have limited available fiber, Broadcast / Narrowcast overlay networks have been deployed. In an overlay network all broadcast channels are carried by one transmitter on a dedicated fiber. A second fiber transports multiple DWDM wavelengths each carrying a unique set of narrowcast channels. The broadcast and demuxed narrowcast wavelength are combined at the node to reconstruct the full RF spectrum of channels. Figure 6 shows the connection details for the overlay architecture. The overlay design while complex is very fiber efficient. With relatively light narrowcast channel loads (up to ~30 channels) the DWDM overlay architecture can easily support 40 nodes on only 2 fibers. The overlay design also takes advantage of the lower (0.25 dB / km) fiber attenuation at 1550nm and the availability of EDFA's to further extend the reach compared to 1310 nm links.

As forecasted growth of VOD and internet traffic continues to accelerate, cable operators once again need to increase data capacity available per subscriber. Competition from FTTX suppliers also is driving many systems to enhance data capacity of their networks. To accomplish this, node serving areas have to be reduced. Additionally, many operators have begun to explore commercial services opportunities by addressing businesses that are within a few kilometers of an existing node. Each of these requirements potentially need additional allocated fiber. Unfortunately, a limiting factor in all HFC networks is the amount of available dark fiber. Meeting the network segmentation and business services needs without the major expense of pulling new fiber spurred the development of multi-wavelength optics.

Multi-wavelength Analog Optical Impairments

Optical Distortion

In order to understand the advantages and trade-offs of different multi-wavelength schemes it is important to recognize the various fiber distortions that can occur in these applications. The following is a brief explanation for the primary fiber induced distortions that affect WDM optical networks.

Stimulated Brillouin Scattering

Stimulated Brillouin Scattering (SBS) is a nonlinear interaction between laser light and the molecular structure of the fiber which generates acoustic waves causing a variation in the index of refraction corresponding to the intensity of the wave. This causes partial scattering of the light in the backward direction from the resultant index diffraction gratings. This can produce an avalanche effect if the intensity of the light is high enough, resulting in high attenuation and induced noise in the forward direction. This acts as a limiting factor as to how much power can be launched into fiber for single wavelength transport. Since the bandwidth in which this scattering process can take place is very narrow, the threshold power needed to initiate this effect can be raised significantly by widening the optical linewidth of the source. This can be accomplished through various methods of dithering the laser, either directly or indirectly, causing a spread in the optical spectrum beyond that of the Brillouin bandwidth (tens of MHz depending on the fiber characteristics). Since this linewidth spread can result in performance degradation when operating in the highly dispersive 1550 nm region of standard fiber, most externally modulated transmitters use some method of phase modulation using single or multiple high frequency tones to effectively breakup the optical signal into a number of separate carriers, each at a reduced level from the original spectra in which case the highest of these individual modes sets the SBS threshold.

Taking advantage of the mitigation techniques described above and with other more significant effects highly contingent on wavelength parameters, SBS would not be a major factor in determining optimal multi-wavelength schemes.

Raman Crosstalk

Stimulated Raman Scattering (SRS) is a nonlinear parametric interaction between laser photons and the molecular structure of the fiber which causes partial inelastic scattering of the light signal due to excitation. The scattered light is shifted downward in frequency (upward in wavelength), corresponding to the molecular vibration frequency, which results in energy transfer between the original wavelength and the generated scattered wavelengths. If additional wavelengths are within the range of the newly generated scattered photons, crosstalk will occur. The triangular shape of the Raman gain (excitation) profile peaks at a wavelength spacing of approx. 100 nm so while the magnitude of the Raman coefficient is much smaller than that of the Brillouin coefficient, it's bandwidth

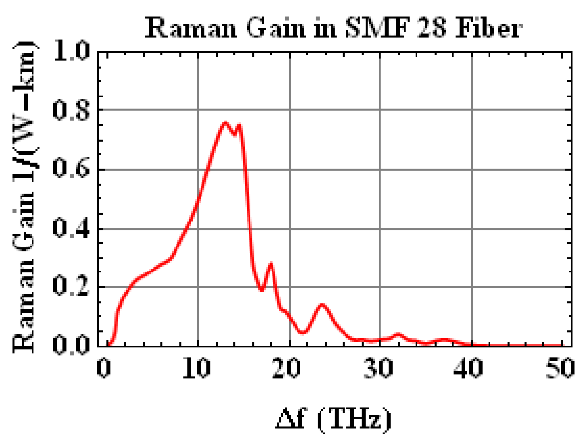


Figure 2 – Raman Gain Profile.
20 nm $\simeq$ 3.4 THz

of influence is much wider. Since the ITU grid DWDM wavelengths are usually spaced 100 or 200 GHz apart (approximately 0.8 and 1.6 nm respectively at 1550 nm), it's a major source of crosstalk in a multi-wavelength system.

Chromatic Dispersion

Chromatic dispersion or group velocity dispersion is caused by a variation of the group velocity in fiber as a function of optical frequency. The chromatic dispersion of standard SMF-28 fiber at 1550 nm is approximately 17 ps/nm/km. When an intensity modulated transmitter with high laser chirp (change in optical frequency vs. modulation) is exposed to dispersive media, the incidental frequency modulation is converted to intensity modulation, which mixes with the original intensity modulation and leads to the generation of intermodulation distortion with 2nd order distortion being the most harmful. The impact of dispersion is greatly reduced if the transmitter has very low chirp. Additionally, the effects can be removed through the use of electronic delay circuit compensation or dispersion compensating fiber (DCF) with equivalent and opposite dispersion characteristics.

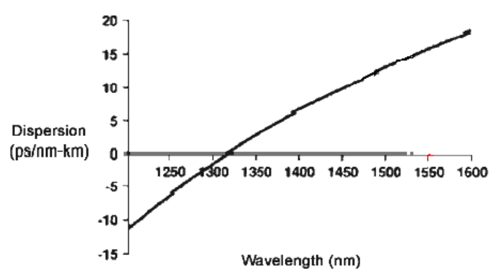


Figure 3 – Dispersion as a function of wavelength for single mode fiber

Crossphase Modulation

Cross-Phase modulation crosstalk is due to the non-linear index of refraction of fiber. The modulation power from one channel causes a small change in the index of refraction which results in a phase modulation of each channel traveling through the fiber. Chromatic dispersion due to the fiber then converts the phase modulation into an amplitude modulation. Cross-phase modulation tends to increase as the spacing between wavelengths decreases and the distance traveled increases.

Four Wave Mixing

Four-wave mixing (FWM) is a 3rd order non-linearity, comparable to the CTB intermodulation effect exhibited in electrical systems, due to the power sensitive refractive index of optical fiber. FWM occurs when multiple wavelengths interact and generate mixing products that fall at one or more of the existing channels, which in turn generates crosstalk at those channels. Four-wave mixing is most troublesome in systems that launch at high powers and utilize a large number of densely packed wavelengths in low dispersion environments.

Raman 2nd Order Distortion

Another potential limitation to using single transmitters for full spectrum BC and NC channels in a multi-wavelength transport has been the enhanced CSO distortions generated due to second order multiplicative effects of the parametric interactions between the various wavelengths due to Raman scattering.

Raman generated CSO is a function of the wavelength spacing, fiber distance and optical launch levels into fiber. The effect is typically strongest at the lowest frequency channels with higher frequencies having the mitigating benefit of dispersion induced walk-off effects.

Modeling shows that even with only two wavelengths, there is a need to carefully set the broadcast launch parameters in order to limit the induced CSO to tolerable levels. Assuming the native CSO distortion of the optical plant is typically around the -66 dBc level, the Raman induced CSO would need to be a maximum of -63 dBc in order to achieve a final CSO distortion contribution of -60 dBc due to the optical plant.

O-Band (1271 – 1371nm) WDM Technology

The first HFC analog full spectrum WDM solutions focused on familiar 1310 (O-Band) optics. The transmitter technology for these designs is identical to the standard point to point 1310 nm CATV transmitters that have been reliably deployed for many years.

There are a number of obstacles to analog multi-wavelength transport in the O-band. The usable bandwidth is bounded by water peak attenuation at 1383 nm on one side and increasing fiber attenuation below 1260 nm on the other side resulting from the loss profile of SMF28 fiber. Since the zero dispersion point of SMF fiber is centered near 1310 nm, dispersion effects are low and increase slowly up to the edges of the usable optical bandwidth. One anomaly in the O-band is that the polarity of dispersion reverses below 1310 nm further complicating multi-channel system designs. The primary distortion impacts are caused by either Raman effects or four wave mixing depending on the choice of wavelength spacing employed.

ITU standard grid wavelengths in the O-Band are only defined for CWDM channels spaced at 20 nm. CWDM optical passives are readily available and provide low ripple, flat passband response. However, systems with three or more 20 nm spaced wavelengths experience high Raman gain effects resulting in significant crosstalk degradation. Without applying unique correction techniques, the Raman impacted CNR and CSO performance severely limits the link reach of these systems.

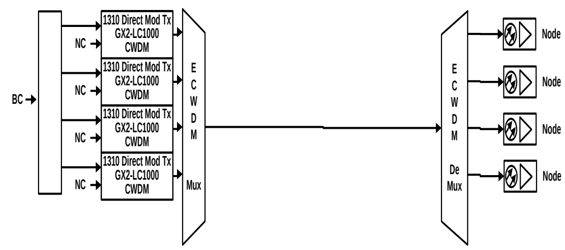


Figure 4 – Motorola ECWDM Multi-wavelength System

As shown in Figure 4 the broadcast channel signal is split to feed all transmitters launched into the same fiber. Most O-Band systems require the broadcast channel load to be identical for all transmitters in order to minimize crosstalk impairment on the analog channels.

Some equipment manufacturers took a completely different approach to reducing crosstalk impairments by designing O-Band WDM systems with closely spaced wavelengths which minimize Raman gain interactions. Unfortunately, given that one of the primary properties of SMF fiber is zero dispersion at ~ 1310 nm, a different fiber distortion (FWM) is strongly enabled by the use of dense, equally spaced wavelengths. Dispersion helps to de-correlate signals, preventing coherent signal beats. FWM is especially severe in fiber with low dispersion. To avoid

FWM degradation unique uneven wavelength spacing plans are required along with a shift away from the zero dispersion point of the plant fiber.

DWDM equivalent wavelength spacing in the 1310 band is not defined by ITU or other standards organizations therefore custom optical passives are required. Another challenge that had to be addressed is the passband flatness response of these custom passives. DWDM optical filters have narrow pass bands which can result in higher wavelength tilt response. DFB laser chirp creates an optical frequency modulation which can interact with the ripple tilt response of the Mux and Demux filters. The optical FM to intensity modulation conversion caused by this interaction can cause relatively high additive second order distortion depending on the amount of tilt encountered through all of the optical passive elements in the system. Specifying low passband tilt passives introduces another layer of customization to these filter devices.

Figure 5 shows the measured tilt response of a 1310 band CWDM and DWDM mux filter. The broad low tilt ripple of the CWDM filter dramatically reduces any laser chirp generated second order effects. Filter slope tilt greater than +/-0.1 dB per nm is usually not acceptable unless laser chirp is extremely low.

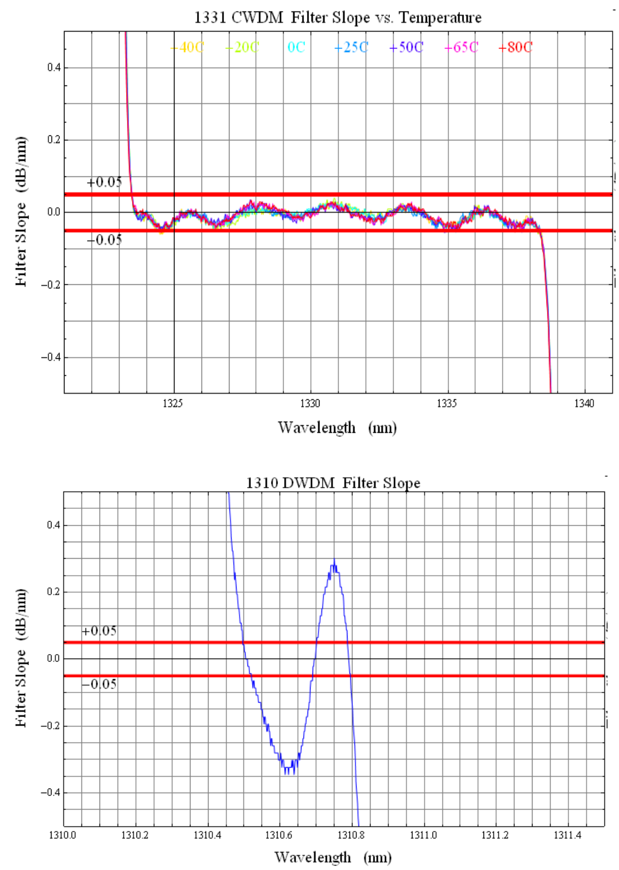


Figure 5 – CWDM and DWDM Slope Tilt Comparison

Due to the significant fiber induced distortions encountered in designing a multi-wavelength system at O-Band, each HE equipment vendor solved the problem in a unique way creating a different, proprietary solution.

Higher fiber attenuation in the 1310 nm band along with crosstalk and FWM analog distortions restrict the link reach of a four wavelength WDM system to roughly 25 km. This limited reach is adequate for approximately 65% of the total links served by the major MSO's.

C-Band (1530–1565nm) Broadcast – Narrowcast Overlay Technology

Fiber attenuation at 1550 nm is 30% lower than 1310 nm. Erbium Doped Fiber Amplifiers (EDFA) make optical amplification possible and practical in the C-Band. These advantages along with ITU grid DWDM wavelength channel standards make the C-Band the logical choice for long reach and limited fiber serving areas.

Along with the fiber distortion impairments described previously in 1310 WDM systems, fiber dispersion at C-Band wavelengths must now be considered. The high chirp of DFB lasers makes full spectrum (55 MHz – 1 GHz) loading impractical for long link networks. Dispersive 2nd order distortion generated by the interaction of the fiber dispersion and laser chirp would dramatically limit the distortion performance of the system.

Externally modulated laser transmitters eliminate the laser chirp issue. ExMod transmitters combine a high power DWDM DFB CW laser source with a Lithium Niobate ($LiNbO_3$) Mach Zehnder (MZ) modulator. The MZ modulator produces zero chirp. To avoid SBS impairment and allow amplified high optical output level for extended reach the CW laser is dithered with one or more high frequency tones. MZ modulators are not inherently linear so a number of correction circuits and multiple feedback loops must be used to optimize 2nd and 3rd order distortion. The $LiNbO_3$ modulator also has a relatively high optical through loss of 3 to 5 dB. A higher power CW laser is usually selected to compensate for this loss although at a higher cost. The $LiNbO_3$ modulator which is designed for CATV linear applications is also a high cost component. All of these elements that make up a typical HFC ExMod transmitter cause this design to be extremely complex and expensive. As a result muxing multiple ExMod transmitters as a full spectrum multi-wavelength system would be cost prohibitive.

The alternative architecture which has been deployed for several years combines a single ExMod transmitter with multiple directly modulated DWDM DFB QAM laser transmitters in an overlay configuration. This architecture allows the high cost of the ExMod transmitter to be spread across a large number of nodes and subscribers. The ExMod transmitter carries the entire broadcast analog and broadcast QAM channel load. The lower cost digital QAM transmitters carry unique channel loads of narrowcast content targeted for specific node serving areas. The content of each QAM transmitter is also carried on a separate DWDM wavelength and muxed together at the hub. The optical launch level of the narrowcast wavelengths must be adjusted based on the OMI per RF channel to assure that a – 6 dB derate between analog and QAM channels is maintained at the node receiver.

The ExMod and muxed QAM laser outputs can be transported on separate fibers or combined onto a single fiber. EDFA's are used to extend the reach up to 100 km or more. At a remote hub or splice housing the narrowcast QAM wavelengths are demuxed for transport to the individual node serving areas. The broadcast wavelength is split and combined with each narrowcast stream either at the demux point or in the node.

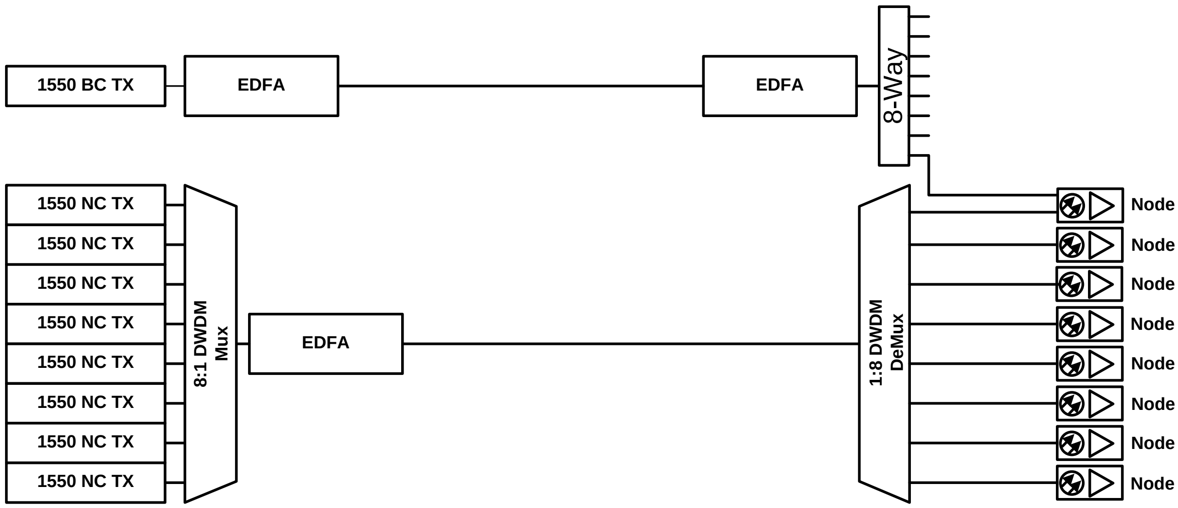


Figure 6 – Broadcast – Narrowcast Overlay Representative Design

Although complex, the BC–NC overlay configuration is very fiber efficient. Up to 40 nodes can be served using only 2 fibers. Until recently most cable networks used very few narrowcast channels. This minimized the level of crosstalk and second order distortion generated by the DFB QAM laser chirp. The low number of NC channels also meant that the OMI per channel was high which also improved distortion performance and subsequently caused the optical launch level of the NC wavelengths to be set ~ 7 dB lower than the BC wavelength in order to maintain the proper RF derate at the node receiver. In many systems a single node receiver detects the incoming BC-NC optical stream.

Narrowcast content is steadily increasing and is expected to triple or quadruple over the next several years. There are a number of consequences to the overlay configuration as a result of this increased NC loading.

As the number of NC channels significantly increases, the OMI per channel correspondingly decreases. This requires a re-balancing of the optical launch levels between the BC and NC transmitters. In a system with 64 NC QAM channels the optical power delta between the BC and NC wavelengths at the node receiver would decrease to only 3 dB raising the detector noise level in single receiver systems as well as the low frequency RF CIN cumulative noise contribution. The higher number of NC channels also means that the chirp performance of the DFB QAM laser becomes critically important. Additional dispersion compensation may be required to reduce crosstalk impairments to acceptable levels. CIN distortion due to the higher NC channel load will increase impacting the lower frequency BC channels. This increased level may challenge the acceptable CCN performance requirement in single receiver systems. Adding a second receiver to separate the BC and NC detection plus filtering the NC RF output will appreciably improve the low frequency noise performance.

C-Band DWDM Full Spectrum Technology

The reach limitations in addition to proprietary issues with 1310 WDM and the higher complexity of BC-NC overlay networks has caused system operators to seek a low complexity, more flexible system that can be expanded where needed on a pay as you grow basis. Development of low cost analog capable full spectrum transmitters for C-Band DWDM has proceeded rapidly. As always, each manufacturer has approached the problem from a slightly different angle and came up with a distinctive solution. The following is a description of the technologies employed in these designs.

Analog DFB Lasers

A directly modulated DFB laser design is always the lowest cost transmitter solution. However, there are significant issues in using a DFB that ultimately limit the useful reach and distortion performance. Laser manufacturers continue to make improvements in the analog distortion performance of high power DFB lasers. The problem at 1550 wavelengths is the fiber induced distortion generated by the interaction with directly modulated DFB laser chirp. The use of pre-distortion correction and tightly controlled electronic dispersion compensation can provide usable performance even with multi-wavelength configurations. Trade-offs between optical launch power, link length, and distortion, particularly CSO, considerably limit WDM reach compared to other available technologies. Since precise dispersion compensation is required to correct for laser chirp the compensation must match the link distance within 5 km to achieve good analog distortion performance.

ExMod Technology Solutions

Long reach external Mach Zehnder modulator transmitters were described earlier for BC-NC overlay applications. The high cost component driver in this design is the $LiNbO_3$ MZ modulator itself. In shorter reach configurations it may be possible to reduce the cost of the modulator and supporting circuitry with some tradeoffs in distortion performance. However, even with the achievable cost reductions the price of a MZ based ExMod transmitter will typically be 2X to 3X the cost of a reference 1310 DFB laser transmitter.

An alternative design uses a $LiNbO_3$ Phase Modulator to cancel the chirp produced by a directly modulated analog DFB laser. The phase modulator pass through loss is also 3 to 5 dB requiring a higher output analog capable DFB laser to compensate. The chirp level of the DFB is somewhat sensitive to the channel loading. The phase modulator and supporting loop circuitry must be optimized to cancel the worst case chirp generated by the laser.

As with other ExMod designs, the $LiNbO_3$ phase modulator component is the highest cost element. One novel approach to mitigate the modulator cost penalty is to share a single phase modulator with multiple analog DFB laser sources. This approach distributes the modulator and supporting circuitry cost across typically four wavelengths, lowering the cost per stream of the full spectrum WDM system. A caveat in this design is that the phase modulator can only be aligned for a single operating condition. Therefore all of the shared lasers must be identically matched to the same operating point (phase, chirp, OMI, etc) to achieve optimum system distortion performance. This approach lowers the cost per stream of the WDM system but requires the operator to absorb the full cost of the multi-wavelength system on day one even if the designed segmentation deployment is planned over several years.

Figure 7 – Full Spectrum WDM System Design

EML Technology Solution

A completely different approach to full spectrum C-Band laser transmitter design utilizes devices that are widely deployed in 10 Gbps telecom optical transport networks. Electro absorption modulated lasers (EML's) are semiconductor devices that combine an E-A Modulator with a CW DFB laser integrated on a single chip. The E-A modulator is processed as a waveguide structure that uses an electric field to control (modulate) the intensity of light passing through it. E-A modulators require low control voltages and are capable of operating at very high speeds making them ideal for digital communications over optical fiber. EML's produce near zero chirp giving them the same immunity to dispersion as MZ based transmitter designs.

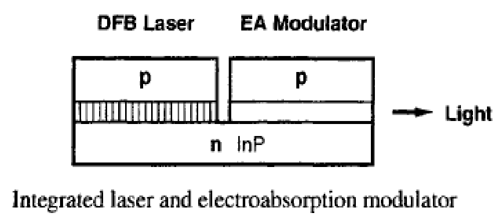


Figure 8 – EML chip design

While the intended application for EML devices is high speed baseband digital over fiber, many of these devices have very linear response characteristics making them capable of transporting analog video and QAM signals with excellent distortion performance. Digital links do not require high optical power levels so the range of available EML device output levels is from -1 dBm to +6 dBm. Higher output levels are possible by using a higher level DFB laser source but this would result in the creation of a custom version with relatively low volume potential compared to the commercially available devices.

The design of an EML based full spectrum analog transmitter is very similar to a typical 1310 DFB laser transmitter. One notable unique difference is that the EML has an impedance controlled GPO / SMP RF connector built into the standard butterfly laser package to facilitate 10 Gbps input signals. During the 2011 CableLabs Winter Conference held in Atlanta, GA a dual mode (HFC analog / 10 Gbps digital) modular laser transmitter was demonstrated to the attending MSO's. The basic 10 Gbps performance characteristics of the EML are unaltered by use in analog HFC applications, making these devices a candidate for eventual migration to 10 Gb optical transport in the access fiber plant.

Figure 9 – 10 Gb EML Device Example

Since EML's are wafer scale devices the cost is dramatically lower than competing lithium niobate modulators. The use of EML's also leverages the much larger volumes of the telecom market putting these devices on the favorable side of the cost curve.

Full Spectrum WDM Link Performance

Full spectrum distortion performance and usable link reach for DWDM multi-wavelength systems is largely determined by the inherent optical fiber distortion impacts rather than the base transmitter performance. For directly modulated DFB laser transmitters, dispersion is the primary distortion performance limiter. In the case of ExMod and EML transmitter designs Raman crosstalk and parametric 2nd order distortion dominate.

Four wave mixing has previously been regarded as a less serious impairment issue for C-Band WDM systems. In HFC BC – NC overlay architectures the DWDM NC transmitter modulation consists of a limited number of channels each loaded with unique un - correlated content streams. High dispersion in SMF fiber provides another strong de-correlation mechanism that has been supported by a number of white paper studies and system analyses. Also, the number of FWM beats falling on any particular channel is usually low considering the limited number of wavelengths deployed in typical applications. Empirical testing however has shown that full spectrum DWDM transmitter configurations do confirm slight but measurable improvement when FWM distortion is avoided. Shifts in wavelength and phase offset of the FWM beats can reduce the observable distortion impact. Therefore aligning all transmitters exactly on ITU consecutive channel centers would represent the worst case impairment condition.

Unique, non-consecutive, or shifted wavelength plans have been proposed to avoid even the chance of FWM distortion. While these solutions are feasible when the number of combined wavelengths is small, they become increasingly difficult to manage in larger scale multi-wavelength networks. The cost / benefit of these solutions are also uncertain since components needed to implement these strategies make volume pricing and repair sparing logistics more difficult.

Full Spectrum multi-wavelength link reach capability varies depending on several factors such as the number of deployed wavelengths, the wavelength spacing, and the channel loading. Ultimately, the link reach limit is determined primarily by the fiber induced distortions that these factors generate.

Figure 10 below provides a guide to the achievable link reaches for different wavelength configurations and two channel load examples.

λ's	79 Analog 75 QAM	30 Analog 124 QAM
2	50 km	60 km
4	40 km	50 km
8	30 km	40 km
16	15 km	25 km

Figure 10 – Full Spectrum Link Reach

All Digital Migration

Analog video distortion performance is always the most stringent. As analog gradually gives way to digital QAM the performance of the remaining channels will significantly improve, particularly CSO and CTB due to the lower number of 2^{nd} and 3^{rd} order beat counts. The robust tolerance of QAM to noise and beat interferers allows transport systems with only QAM channel loads to achieve better end of line performance and in most cases longer link reach than systems with analog video loading.

Each of the full spectrum analog transmitter technologies described in the preceding sections of this paper are also capable of operation with an all QAM channel load. One possible application that may benefit is the traditional BC – NC overlay. If the current directly modulated DFB NC QAM transmitters used today were replaced with full spectrum transmitters much of the dispersion related distortion impacts would be eliminated. This could potentially improve the CCN performance of the low frequency broadcast channels which are the most vulnerable to dispersive CIN distortion particularly with older legacy equipment that are not capable of increased NC loading. Full spectrum transmitters are higher cost than DM QAM transmitters but the performance gains that could be achieved without the expense or disruption of added node receivers and filters or external dispersion compensation may justify the cost differential.

Link reach for fully loaded all QAM WDM systems can increase but not dramatically when compared to systems with a small tier of analog channels. CIN distortion due to the increased channel load along with fiber induced Raman distortion effects still dominate the final system performance. This result implies that the BC – NC overlay network is still an essential solution for long reach networks.

System Cost Comparison

To compare the relative cost of each multi-wavelength architecture the bill of material for a 40 km, 8 wavelength downstream aerial fiber link was prepared. Three designs were analyzed; 1310 home run, 1550 BC–NC Overlay, and 1550 Full Spectrum multi-wavelength corresponding to Figures 1, 6, and 7 as illustrated in this paper. The O-Band multi-wavelength design shown in Figure 4 was excluded because it is not capable of supporting a 40 km links.

The following component costs were included in each design:

- Hub Transmitters (8) plus a BC Tx in the case of the overlay design
- Node cost plus optical receivers
- Mux and Demux passives where required
- EDFA's where required
- Optical Splitters where required
- Aerial 12 count fiber construction (40 km) at $3000 per kilometer

The chart in Figure 11 summarizes the relative cost difference of each architecture in a Greenfield application. As expected the BC–NC overlay is the most expensive due to the high ExMod laser cost and multiple EDFA's used to overcome optical splitting losses. The other result is the comparable cost of 1550 Full Spectrum to traditional 1310 system designs while providing an 8:1 improvement in fiber utilization preserving additional dark fiber for other revenue generating applications. The slightly higher premium for the full spectrum solution is due to the additional optical passives and EDFA required to combine wavelengths onto a single fiber.

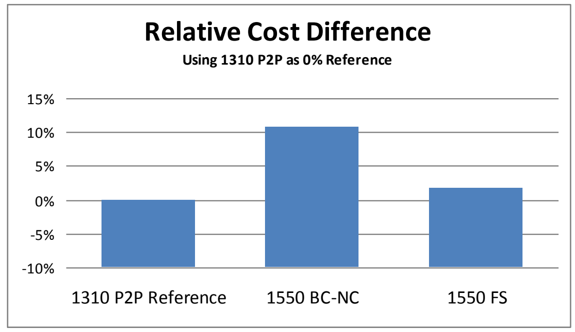


Figure 11 – Multi-wavelength system cost comparison

Summary

The continuing expansion of new narrowcast HD, VOD, and HSD channels is putting pressure on available downstream bandwidth. To meet this demand cable operators need a solution that does not require expensive fiber deployments or the need to touch every active in the network. Virtual segmentations and node splits are part of the answer but to make these viable solutions a fiber efficient WDM design is needed that offers the flexibility to adapt to the increasing channel load while providing the link reach capability required to cover 95% or more of the network footprint.

Broadcast – Narrowcast overlay designs are still the most fiber efficient architectures for long haul access networks and systems with very limited available fiber. But the overlay scheme is complex, requiring optical level rebalancing as the ratio of narrowcast channel loading changes. Overlays are also expensive, especially if only a small numbers of nodes are connected.

Attempts to leverage 1310 laser technology for WDM segmentation has had limited success. The high chirp levels generated by directly modulated 1310 DFB lasers in conjunction with the particular fiber distortion mechanisms present in the 1290 to 1370 nm wavelength region limit the link reach capability to 25 km with four wavelengths. While this is certainly valuable for most near term applications the link reach constraint means that multiple solutions will be needed to cover the full network footprint area.

The advances made with 1550 DWDM analog + QAM Full Spectrum laser transmitter designs has opened the door to a number of realizable near term solutions and potential future migration possibilities. The Full Spectrum designs available today eliminate laser chip and take advantage of lower fiber attenuation losses and EDFA amplification to address link reach coverage that is compatible with existing 1310 deployments. This allows operators to harvest fiber from existing nodes for use by commercial services customers or for new node serving area reduction targets.

The introduction of Electro Absorption modulated laser technology for HFC analog transport provides a new lower cost solution that can ultimately support digital data rates up to 10 Gbps for future generation system applications.

Acknowledgements

The author would like to acknowledge the invaluable support and contributions provided by Fred Slowik (Dir Systems Marketing), Mike Aviles (System Engineering) and Jun Wang (R&D Engineering) toward completing this paper.

Architecting Small Cell Licensed and Unlicensed Networks

David Park
BelAir Networks

Abstract

The advent of the smart phone and more recently the rise of the tablet has transformed the demand side of mobile networks. The end user's relationship with their mobile operator is driven by the data user experience they experience. This drives a need to fundamentally change the mobile network architecture from a macrocell based system to a small cell architecture. The fundamental challenge of increasing the cell density is where to put the basestations, how to power them and how to backhaul them. Cable Operators have a unique opportunity to leverage the Hybrid Fibre-Coax network and enable the deployment of 100s of thousands of cell sites. This opens up new business models and new capabilities for delivering wireless services for MSOs and Mobile Operators that could transform the mobile world.

THE RATIONAL FOR SMALL CELLS

The Data Explosion is Here!

It is now universally acknowledged by both analysts and mobile operators that the data consumption is rising exponentially while data revenues are growing linearly. AT&T [1] has made statements that wireless data traffic grew by a factor of 80 times over the period of 4 years and most sources show a double digit annual growth rate. Gabriel Brown from Heavy Reading put this data usage into perspective when he stated that "95 percent of mobile data traffic is best effort Internet." So, the problem becomes how to deal with this deluge of data as cost-effectively as possible, while delivering a high quality user experience. This challenge will come as no surprise to most mobile carriers, especially those who were first movers with the iPhone, iPad and other smart devices.

Much has already been written over the last few years about the demands that this increased data traffic is placing on the mobile carriers" backhaul requirements, but that is a simplification of the overall problem, which is the need to place many more cell sites, closer to the end users. One of the ways that mobile carriers are dealing with their requirements is by leveraging fibre infrastructure from the local cable operator to backhaul their mobile traffic.

In their May 2010 report [2] Visant forecasts that cable operator share of mobile wireless backhaul market will grow by more than five times by 2016, generating more than $3 Billion in annual service revenues to cable operators by 2015. So, mobile carriers will be doing more backhaul on cable networks, but is there more that mobile carriers can be doing to leverage these networks?

What Do the Economics Look Like?

…..While data consumption has been growing exponentially, growth in mobile carrier‟s data ARPU (average revenue per user) hasn‟t come anywhere near the levels required to offset declines in voice ARPU. CTIA‟s semi-annual survey [3] shows that US monthly post paid ARPUs have remained between $47 and $50 since mid 2002, and that during the start of the massive data growth period from 2006 onwards, ARPU was flat.

Of course, 4G technologies, such as Long Term Evolution (LTE), hold the promise of increased mobile network capacity. But, it‟s important to remember that an LTE network deployed in today‟s standard macrocell architecture, with base stations installed on towers and rooftops, will at best increase capacity by 2 to 4 times (depending on the carrier‟s spectrum allocation). Compare this with the exponential growth in data usage that is driving the requirement for mobile network capacity increases on the order of 10 to 100 times, and it becomes clear that a simple 3G to LTE macrocell swapout won‟t address the capacity challenge. To sufficiently increase capacity, cell sizes need to get smaller. Essentially, more base stations need to be deployed in more locations. But, remembering that this increased capacity must be accomplished within an environment of reduced ARPU, the time and cost involved in building more towers (if you can even get approval for them) and securing more rooftops could result in installed costs that just can‟t deliver a reasonable return on investment (ROI) for the carrier. And that assumes that increasing macrocell density is feasible from a spectrum reuse point of view.

The Capacity vs. CapEx Challenge

The conundrum of more capacity for less capital expenditure (CAPEX) is beginning to be addressed by the category of small cell base station known as outdoor metropolitan picocells. If you‟ve been following the small cell market, you may have noticed that there‟s some naming confusion out there. Cell size naming (and associated base station equipment) follows a hierarchical methodology with the traditional large cells being macro and reduced cell sizes denoted as micro then pico and, finally, femto. While that‟s generally the case, you may hear outdoor metropolitan picocells also being referred to as Class 3 femtos. One important distinction is that a femtocell is usually associated with an in-home device that is deployed to improve coverage, whereas a picocell is an inherent part of the radio access network that includes that ability to support full handoff, from pico to pico and from pico to macro and back again.

…..According to In-Stat [4] outdoor metropolitan picocells "will allow an operator to provide excellent coverage, capacity, and data speed to users in dense areas and other public spaces". In-Stat‟s report also points out the critical role that these picocells will play in delivering on the promised benefits of 4G:

….."Where 4G technologies differ from older technologies is that a very strong signal and a close proximity to a base station, and few users per cell are all requirements if users are to experience the ultra-fast broadband speeds and high-capacity that wireless operators have been promising, and this can only be accomplished with small cells." The fundamental reason for this is that the promised higher speeds are delivered by higher order modulation rates such as 64QAM and those require higher signal to noise ratios that mean being closer to the end user.

…..The concept of a small cell architecture has been proposed before, but has never been successfully exploited until now. When you consider that many mobile carriers in North and South America, Europe, Middle East and increasingly in Asia, already find themselves with a pressing need to augment capacity in areas of high user concentration, it may seem strange that picocells have not yet achieved widespread adoption. The reason for this is less technical than it is operational. Picocells have typically presented deployment challenges for mobile carriers in terms of how to mount, power and backhaul the base stations. The typical tower and rooftop installations with which mobile carriers are most familiar don‟t suit the picocell architecture which aims to bring coverage closer to the user to mitigate the service degradation that occurs, in macrocell coverage, as users move towards the outside edge of a cell.

So, the question becomes, how can the mobile carrier find and secure appropriate mounting sites that offer readily available power and backhaul? But, wait, there‟s more. It‟s not enough to find appropriate mounting sites if it means that the mobile carrier has to negotiate with multiple different property owners, utilities or government departments to secure them. The cost and time associated with these sites will have a negative impact on the success of the deployment.

THE STRAND PICOCELL

Leveraging the HFC

As we‟ve already noted, many mobile carriers currently leverage cable operators‟ fibre plant to backhaul their macrocell base stations and are expected to do more of that in the future. So, why not leverage the HFC to solve the problem of where to put small cells as well?

This leads us to the strand-mounted picocell, a new class of outdoor metropolitan picocell that leverages available HFC infrastructure (aerial, pedestal, cabinet or vault) for power, mounting and backhaul. The strand picocell can incorporate licensed or unlicensed radios. In fact, a combination of licensed cellular technology (3G or 4G) with Wi-Fi, in a single picocell, as illustrated in Figure 1, presents a very compelling business proposition for both the mobile carrier and the cable operator.

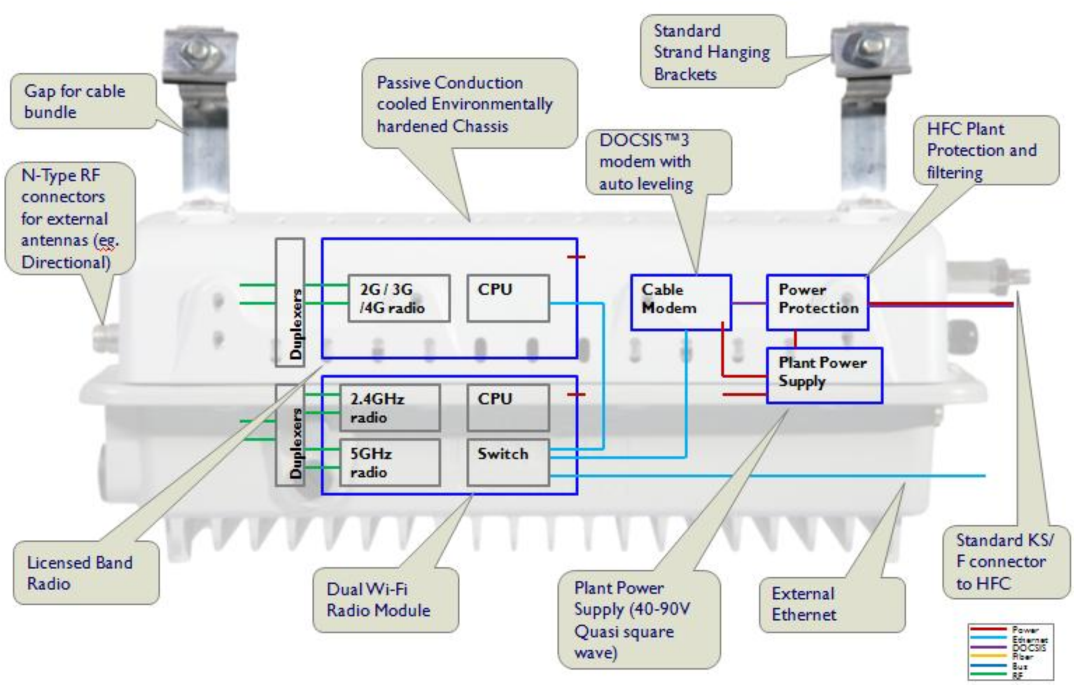


Figure 1: Strand Mounted picocell

The strand picocell is fed from the HFC from a standard power passing tap. This feeds power and DOCSIS to the unit. A power conditioning circuit provides surge protection and prevents and of the RF from the picocell leaking back onto the plant. Power is split from DOCSIS inside the unit and is fed to a plant power supply that uses the standard 40-90V quasi square wave that is used to power nodes and amplifiers. The DOCSIS signal goes to a specialized modem that is designed for operation on the main line. That feed Ethernet to a control card containing a managed Ethernet switch and dual band Wi-Fi radios. This card provides power and control to a picocell radio. The control card performs autoconfiguration of the picocell, and provides prioritization of the cellular traffic over the Wi-Fi. The various frequency bands are duplexed together and fed to the antennas.

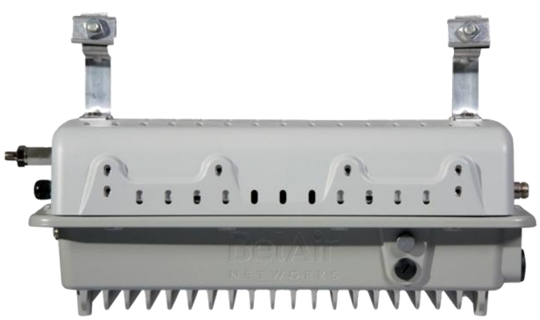


Figure 2: Strand Mounted Pico Cell

The strand picocell can be mounted directly on to the messenger wire of aerial HFC plant, or attached to the HFC inside cabinets, pedestals or vaults. These installation variants are shown in Figure 3.

Figure 3: Installation locations

By leveraging the ubiquity of HFC plant, the mobile carrier is able to quickly and cost effectively add picocell capacity to areas where the high volume of mobile users and their associated data traffic is creating service-affecting congestion in the macrocell network. These high usage areas tend to be busy downtown traffic corridors and shopping districts, parks, arenas and special events, hotels and convention centers, university campuses, and mass transit stations.

Figure 4 shows a strand mounted picocell being installed. A power passing tap is used to drop both power and DOCSIS to the unit and the messenger wire is used to physically hang the unit

Figure 4: Installing a Pico Cell

In the traditional 3G macrocell network we see the NodeB connecting to the RNC and then diverting voice and data traffic to the PSTN and Internet, respectively. In addition to the operational issues that it solves, another benefit of integrating with the cable network is that it‟s an all IP network, end to end. The cable network also includes features such as auto configuration servers that simplifies the process of deploying picocells.

The DOCSIS ® 3.0 cable modem in the picocell plugs directly into the cable network where it autoconfigures like any modem. Once connected to the cable operator‟s backhaul network, the red arrow in Figure 6 shows how 3G voice and data traffic is handled as it travels between the picocell, the cable network and the mobile network, effectively bypassing and offloading traffic from the existing 3G or 4G radio access network (RAN). By leveraging the unlicensed Wi-Fi radios in the picocell, data traffic can also be offloaded from the operator‟s core network, as indicated by the green arrow.

Policy enforcement in the picocell, in conjunction with policy management in the cable network and the mobile network, enables intelligent, real-time decisions to be made with regard to switching traffic between the networks, depending on network loading, QoS service parameters of the customer and other variables.

The cable operator can also leverage the network to deliver branded Wi-Fi services to their own customers, as indicated by the light blue arrows. In fact, multiple cable operators can utilize the same Wi-Fi infrastructure while providing unique, differentiated services to their respective customers.

So, with one picocell platform a whole range of service options are enabled for delivering a true mobile broadband experience while offloading both the RAN and the 3G core.

Figure 5: Strand Pico Cell in service

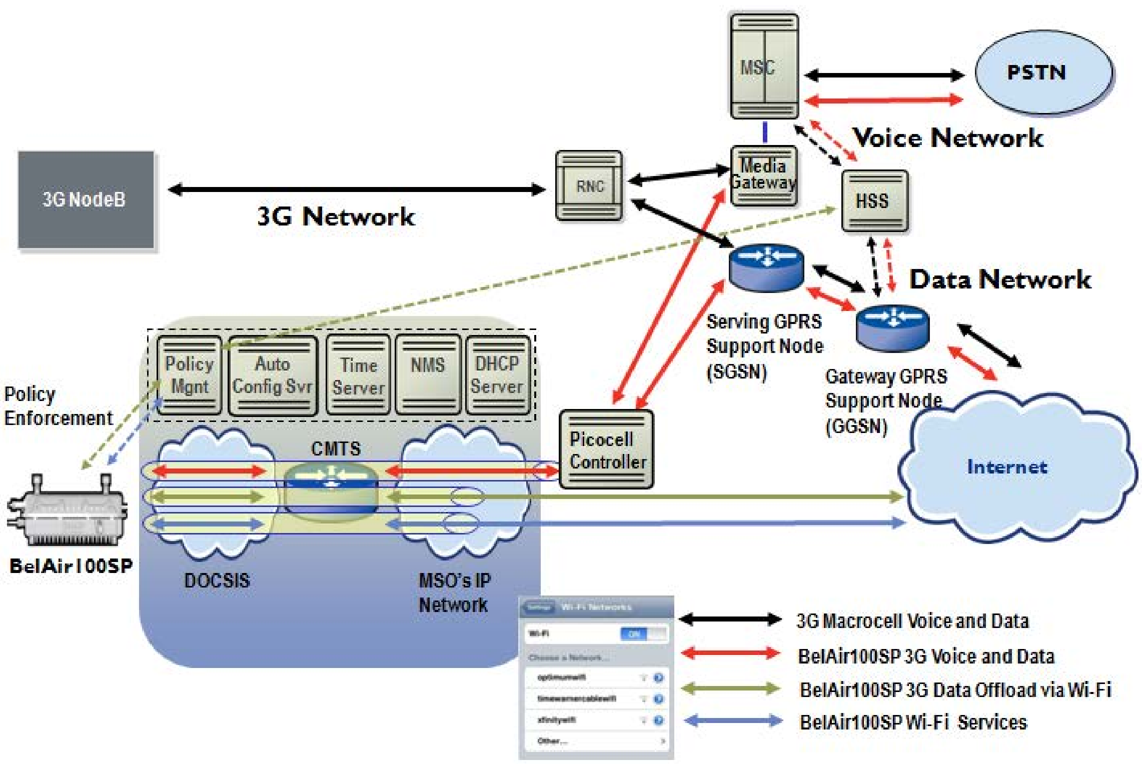


Figure 6: Network Architecture

THE BUSINESS MODEL

How Does the Mobile Carrier Benefit?

For the mobile carrier, a strand picocell deployment represents a future proof architecture to address the huge and continued growth in mobile data demand and resultant network traffic. By solving the traditional challenges of small cell deployment, the strand picocell enables the mobile carrier to establish the new network architecture that will be critical to their LTE success, while also addressing the capacity challenges they face in their network today. As already noted, the concept of mobile carriers leveraging the cable network is not new – mobile carriers already partner with cable operators, leveraging the hybrid fiber coax infrastructure to backhaul mobile traffic from 2G and 3G macrocells.

The deployment of strand picocells just extends this mutually beneficial relationship to enable widespread small cell deployments that enhance the capacity of 3G and LTE macrocell coverage. Of course, there may also be common ownership of the mobile and cable entities, which facilitates this cooperation. In either case, the mobile carrier benefits from a network model and a business model that are very compelling.

Meanwhile the mobile carrier‟s customers benefit from a better user experience with consistently higher speeds and broadband throughput throughout the cell coverage area.

How Does the Cable Operator Benefit?

For cable operators, the BelAir100SP establishes a new business model whereby the cable operator operates as a managed wireless service provider with a hosted base station offering for the mobile carrier – a higher value and more differentiated offering than basic mobile backhaul. And, because the strand picocell requires no additional cable infrastructure, it effectively leverages existing HFC assets for additional revenue and subscriber retention benefits. Strand picocells are fast and easy to deploy, commission, operate and manage, so time to revenue is quick while CapEx and OpEx are low.

Of course, the cable operator can also leverage the strand picocell deployment to provide a mobile broadband complement to their residential subscribers, to encourage new subscriptions and reduce churn. And for the MSO with spectrum, the Strand picocell offers a rapid way to deploy and maximize the use of limited spectrum resources.

SUMMARY: THE FUTURE IS A SMALL CELL WORLD

With the increased data demand and the advent of high order modulation advanced wireless systems, the need for change to a „small cell‟ architecture is evident. The traditional barriers to adoptions have been location, power and backhaul and the MSO has the unique asset to solve that problem.

The strand picocell offers a transition to a new architecture that is forward looking and moves wireless networks to the next generation.

REFERENCES

[1] Mobile broadband explosion and pressing demand for spectrum AT&T March 2011.

[2] “US Mobile Wireless Backhaul 2011” Visant

[3] CTIA‟s Semi-Annual Wireless Industry Survey.

[4] “Worldwide Femto, Pico, and Microcell Market Analysis” Instat

THE END

CARRIER ETHERNET SLAs: TECHNOLOGY ADVANCEMENTS TO IMPROVE OPERATIONAL EFFICIENCY FOR CABLE PROVIDERS

Kevin Wade
Cyan

Abstract

High-capacity, low-latency connectivity is important for the new generation of Ethernet-based retail and wholesale commercial services, including ultra high-bandwidth business Ethernet services and wireless backhaul. As network performance increases in importance to customers, overall characterization of the network and service level agreement (SLA) considerations also become more critical. Network planning, management and verification tools that apply to all layers of a cable provider's packet-optical transport network – starting at the fiber layer and rising up through the wavelength, SONET, connection-oriented Ethernet (COE) and Ethernet services layers – are essential to supporting these requirements.

This paper provides an overview of advanced tools and technologies that allow cable providers to optimize their networks for end-to-end SLA performance across multiple layers, resulting in more intelligent utilization of resources, faster resolution of network issues and improved operational efficiency.

TRANSITIONING TO ETHERNET SERVICES

Ethernet is increasingly being adopted by cable providers as the foundational technology used to deliver the new generation of commercial services. Wireless backhaul, a mainstay wholesale service offering for many cable providers, is transitioning to Ethernet to support the rapidly expanding Long Term Evolution (LTE) network deployments by Tier-1 wireless operators. Unlike prior generation (2/2.5/3G) wireless networks that are supported by TDM or hybrid TDM/Ethernet backhaul networks, LTE mandates that the backhaul architecture be Ethernet/IP-based. A similar transition is taking place with retail business Ethernet services, where cable providers currently hold the Nos. 4 and 7 US market share rankings respectively, in terms of the number of enterprise ports installed*. Here, ultra high-bandwidth switched Ethernet services (GbE and higher) are experiencing the fastest growth, primarily to support the financial, data center, education and state/local government vertical markets.

Recognizing that poor network performance negatively impacts their businesses, end-customers' quality expectations associated with these and other Ethernet-based services are at a high level, and are continuing to rise. As applications across all vertical markets become more cloud- and video-based, higher levels of bandwidth and availability are being demanded, in addition to lower delay, jitter and packet loss. Ethernet, as is well known, does not inherently provide the levels of service availability, reliability and QoS found with traditional TDM/SONET services. So while Ethernet's flexibility and low cost make it attractive to cable providers, profitable delivery of Ethernet-based services on a broad scale is directly related the cable providers' ability to support Ethernet SLAs. Cable providers who are effective in this area will have a competitive advantage.

KEY ETHERNET SLA AND OAM DEVELOPMENTS

The ability to support Ethernet SLAs has improved significantly in recent years as a key focus area for the evolution of Carrier

Ethernet. Concerted efforts by the IEEE, ITU, MEF and other organizations have resulted in a mature set of standards for Ethernet Operations, Administration & Maintenance (OAM) that provide monitoring at the Ethernet link level, as well as end-to-end

(continue reading to the right)

service-level performance monitoring and connectivity fault management. These key Ethernet OAM standards include IEEE 802.3ah, IEEE 802.1ag and ITU Y.1731 as illustrated in Figure 1.

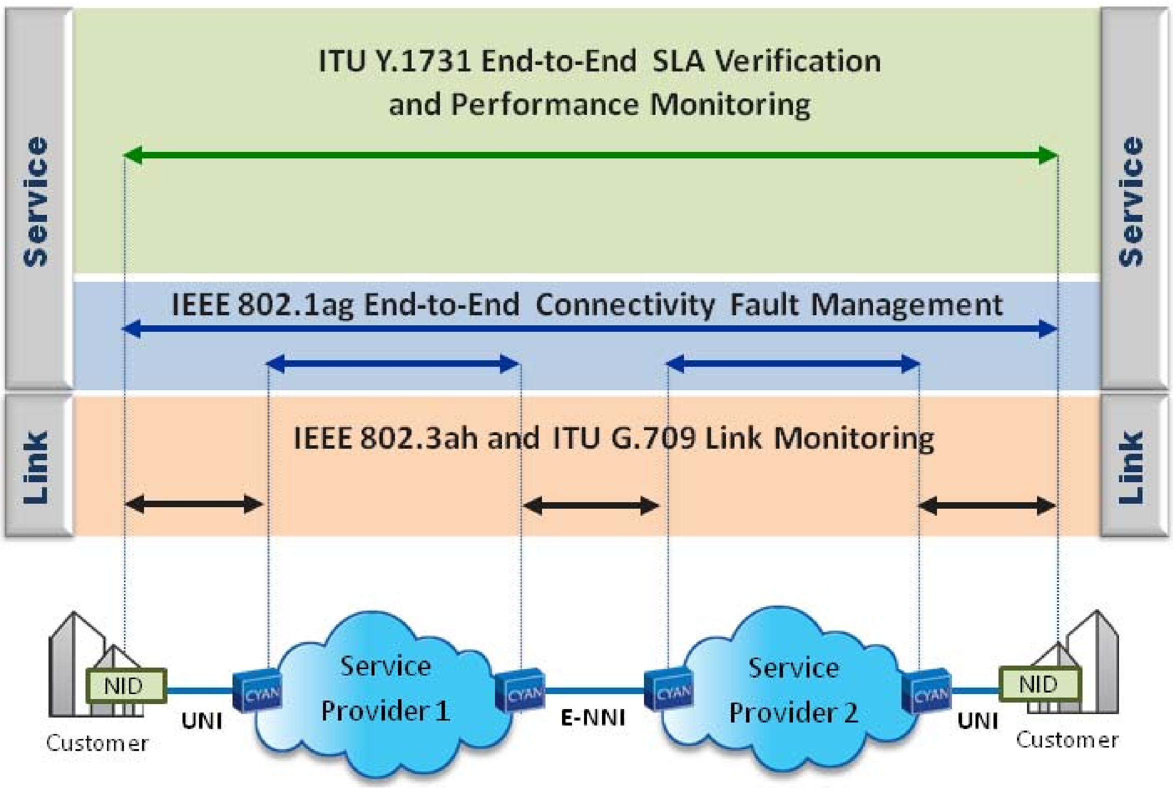


Figure 1: Ethernet OAM Standards

<u>802.3ah Link OAM</u>

The IEEE 802.3ah OAM standard enables cable providers to monitor and troubleshoot at the physical Ethernet link level. 802.3ah is particularly valuable in the “last-mile” connection to the Ethernet access or customer demarc network interface device (NID), where the majority of link events occur. Link OAM allows the cable provider to monitor a link for critical issues, and then if necessary, perform a "loopback" with the remote device in order to test the Ethernet link.

<u>802.1ag Service OAM</u>

The IEEE 802.1ag Connectivity Fault Management standard allows cable providers to manage individual customer connections at the service-level. A customer service instance, also defined by the MEF as an Ethernet Virtual Connection (EVC), is the service that is sold to a customer and is typically designated by a Service-VLAN tag on the User-to-Network Interface (UNI). 802.1ag enables the cable provider to know if a service instance/EVC has failed, and if so, provides the utilities to rapidly isolate the failure within the overall network management framework.

Y.1731 Service OAM

The ITU Y.1731 standard builds on the concept of 802.1ag by allowing cable providers to monitor the performance of individual customer connections at the service-level. Y.1731 provides real-time and historical EVC monitoring including measurements of Ethernet frame delay, frame delay variation, frame loss and throughput. This data is collected from network devices to support end-to-end SLA verification and assurance. *(continue reading to the right)*

MULTI-LAYER, SERVICE-LEVEL AWARE NETWORKS

Service OAM enables performance monitoring and fault management at the Ethernet services layer, but network performance is not determined by this layer alone. Cable providers' networks consist of multiple, independent network layers, each of which plays a role in delivering services from end-to-end, as well as in determining network performance. Depending on the provider, the network can include a physical fiber layer and rise up through the DWDM, Optical Transport Network (OTN), TDM/SONET, connection-oriented Ethernet (COE) transport and Ethernet service layers, as shown in Figure 2.

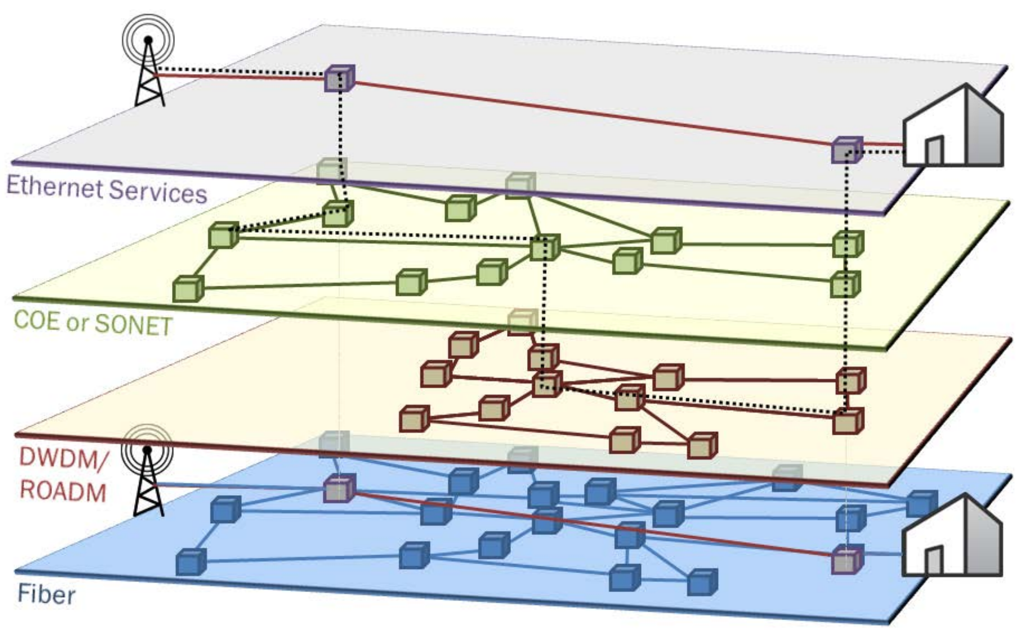


Figure 2: An End-to-End Service Traversing Multiple Network Layers

Historically, each of these layers has been managed and operated independently, without awareness of adjacent layers or services. This inefficient method of operation makes it difficult for cable providers to optimize their networks for end-to-end Ethernet SLA performance. "Service-level aware" tools and technologies that concurrently extend across multiple layers are required to meet these challenges. Key multi-layer, service-level aware building blocks that have recently been introduced to the market include the following:

Multi-Layer Network Planning Software

Advancements in cloud-based network planning software enable cable providers to quickly and cost-effectively design multi-layer networks that are optimized for SLA performance. Advanced algorithms are utilized to ensure the most efficient and economical network design for any given traffic requirement. In addition, the software auto-generates a complete bill of materials that includes equipment configuration and installation checklists to accelerate the planning and deployment process.

Multi-Layer Network Management Software

Innovations in multi-layer management simplify the process of implementing, monitoring, maintaining and operating networks to improve SLA performance for cable providers. Multi-layer network management incorporates object-based intelligence, which provides the software with an awareness of network resources such as switching/transport nodes, connections, and network layers, as well as their inter-dependencies, to optimize the efficiency of network planning and operations. Multi-layer network management also provides advanced 3-D visualization that delivers intuitive, graphical views of the network and network resources, as shown in Figure 3.

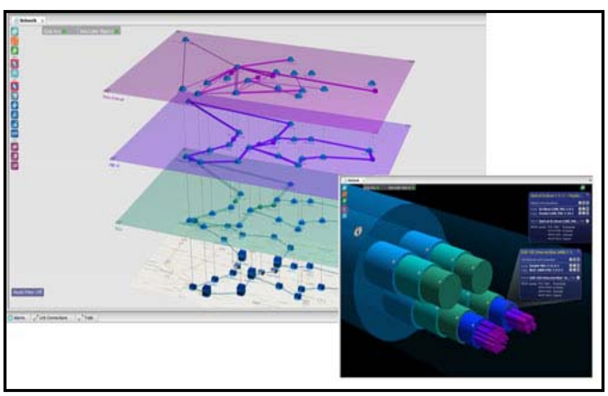


Figure 3: Multi-layer Network Management -- Network View and Cable/Connection View

In addition, multi-layer network management supports "virtualized" network views based on customer, service type, region or responsibility, allowing cable providers to partition network visibility in a manner that is consistent with their operational procedures – accelerating the introduction of new services and lowering costs.

Multi-Layer SLA Verification Software

Developments in cloud-based performance monitoring software allow cable providers to make network performance data available via customer-specific web portals for SLA verification purposes. Each customer-specific SLA verification portal provides real-time and historical reporting of Ethernet network performance data for parameters such as availability, delay, jitter, throughput and packet loss that is collected from network devices based on Y.1731 or other standards-based protocols. The cable provider maintains full administrative control over the level of SLA/network performance visibility provided to individual customers via secure, partitioned network views, as illustrated in Figure 4.

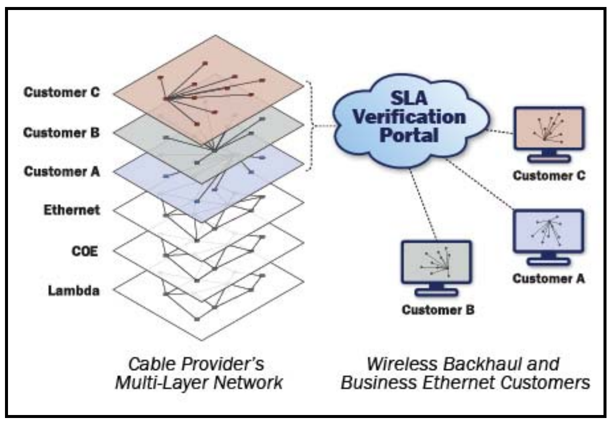


Figure 4: SLA Verification Portals Partitioned to Individual Ethernet Customers

Packet-Optical Transport Hardware

Evolutions in equipment design have led to the development of the packet-optical transport platform (P-OTP). P-OTPs integrate service- and link-layer OAM-enabled 1/10

Gigabit Ethernet switching and connection-oriented Ethernet transport to deliver deterministic bandwidth with carrier-class protection switching (50 ms or less), and support stringent SLA requirements associated with availability, resiliency, throughput and low-latency. Further, the P-OTP integrates SONET and DWDM switching and transport, and is capable of supporting future 100 GbE services as well as 40G and 100G optical transport, allowing cable providers to scale their multi-layer networks with improved operational and capital efficiency.

CONCLUSION

The new generation of Ethernet-based commercial services mandate predictable performance. As a result, Carrier Ethernet SLAs are increasingly required to ensure that these services can be delivered profitably on a broad scale. While important SLA and OAM monitoring standards have matured, cable provider's networks are multi-layer in nature. "Service-level aware" tools and technologies have recently been introduced that allow cable providers to optimize their networks for end-to-end SLA performance concurrently across multiple layers. These advancements include scalable packet-optical transport platforms and powerful software tools that enable management of the entire network life cycle, from service-level aware planning to management/operation and verification. For cable providers, these new tools and technologies result in more intelligent utilization of network resources, faster resolution of issues and improved operational efficiency.

* Vertical Systems Group's 2010 U.S. Business Ethernet Leaderboard

CLOUDS, CABLE AND CONNECTIVITY: FUTURE INTERNETS AND ROUTER REQUIREMENTS

Robert M. Broberg (Cisco), Andrei Agapi (Cisco), Ken Birman (Cornell), Douglas Comer (Purdue), Chase Cotton (University of Delaware), Thilo Kielmann (Vrije Universiteit), Bill Lehr (MIT), Robbert VanRenesse (Cornell), Robert Surton (Cornell), Jonathan M. Smith (University of Pennsylvania)

Abstract

The "computing utility" vision of cloud computing posits a future Internet that offers a universal infrastructure capable of providing on-demand access to computing, storage, and communication services. Clouds will support a diverse range of user/usage contexts, ranging from the delivery of advanced television and content to support for corporate enterprise networks, from controlling smart grids to remote control of an insulin pump.

This paper discusses the research program for a future Internet that is being undertaken by the Nebula project, with support from the National Science Foundation, and in collaboration with Cisco in its on-going efforts to rethink the software architecture for large multi-processor router platforms. The Nebula architecture is comprised of three elements: NCore for tying together cloud data centers and core routing infrastructure, NVent for implementing a flexible and extensible control plane, and NDP for fine-grained, policy-based end-to-end control of network flows. Herein, we concentrate our discussion on NCore. We also briefly highlight some of the non-technical policy and business challenges posed by migrating to this new, more capable and robust architecture.

1. INTRODUCTION

Cloud computing means many things to many people. We adopt a computing utility vision [Fano65] of the cloud as a universal infrastructure capable of providing on-demand access to computing, storage, and communication services over the Internet to support a wide array of user-needs and applications. These may range in diversity from entertainment television delivery, to support for public safety emergency calling, or even the ability to remotely control a diabetic's insulin pump. This vision is analogous to the collective relationship between the data centers constituting an electricity grid of power generating facilities, the long-haul transmission grid manifest as a core Internet routing infrastructure, and the local distribution facilities that provide access networks. Supplying electric power to businesses and consumers, this ensemble is equally responsible for maintaining flexibility towards all different respective requirements regarding performance, security, and end-user control.

This computing utility vision posits the existence of virtualization software, capable of supporting the illusion of dedicated capacity while sharing computing and storage resources located across multiple providers. To end-users, this vision implies a shift of intelligence, data, and services into the network "cloud." An increase in energy and administrative costs, the growth of data-intensive applications, and a proliferation of new usage contexts (including thin clients, mobile computing, and machine-to-machine applications) are all contributing to an inevitable adaptation to some form of network-centric, cloud-based resource sharing. In light of a surge in video and other rich media traffic precipitating the search for new content delivery strategies, and their critical role in providing last-mile broadband, we believe that traditional broadcast/cable companies should be lead players in guiding the design and migration to a more capable, flexible, and secure Internet.

The Nebula research project (see http://nebula.cis.upenn.edu) supported by the United States National Science Foundation (NSF) and complemented by Cisco's on-going research effort (see http://r3.cis.upenn.edu), is exploring an Internet architecture to foster this cloud computing vision. The Nebula architecture will embody three components: a high-speed core that interconnects data centers and enables direct transfer among them, a set of wired and wireless access networks that provide connectivity to individuals and enterprises, and a transit layer that allows an individual to connect to the nearest data center over a path with guaranteed properties, such as security. Early goals include continuous availability for routing components such as BGP, even when the processor on which BGP is running fails; peers will be completely insulated from failure/restart events, avoiding route flaps, black holes and other transients seen when BGP fail-over occurs on today's core network routers. As our effort moves forward, we believe we can do even more. The team is exploring novel options for securing routes and protecting against attacks, for creating new kinds of router-hosted services, and broadly, for transforming the modern router into a better partner for the evolving cloud. The development of this new architecture is consciously motivated to address real-world deployment issues, such as compatibility with regulatory policies, and scalable deployment within today's evolving industry value chain.

This paper is organized into the following sections. Section 2 presents an overview of the Nebula architecture. In Section 3 we examine some of the technical and non-technical challenges posed by this vision of cloud computing. Section 4 concludes.

2.0 A SECURE ARCHITECTURE FOR CLOUD COMPUTING

Traditional Internet services [Comer06] are built on a best effort packet delivery service. What is ultimately desired is the ability to deliver the best features of today’s Internet architecture, combined with new architectural features to support applications with *beyond best-effort* real-time and policy requirements. For stored or dynamic content, the HTTP/TCP/IP model seems adequate, but for an expanding range of service offerings, the conventional architecture could be much improved. Consider, for example, the highly reliable video delivery services offered by Cable providers. This traditional “Cable TV” service is still the basis for many service subscriptions, and continues to evolve in fidelity as end-user equipment evolves (e.g., HDTV). An ideal Internet solution would allow for this traffic to be treated in isolation throughout the network, and for the resources necessary for subscriber fidelity to be acquired and maintained as needed. Further, the reliability would be such that the illusion of classic coaxial service could be maintained, while at the same time capitalizing on the operational advantages of a universal IP infrastructure. Conversely, multiple internal IP infrastructures are often used today. This split acutely illustrates the challenge for the future: a universal infrastructure that affords extensibility for new services, policy enforcement and ultra-high availability. To meet the demands of scalable performance growth, as well as the variety of security and trust requirements implicit in managing healthcare applications, smart grids, delivery of entertainment television or bulk data delivery, new architectural elements are clearly needed.

2.1 The Nebula Project

Nebula is a project that was founded in 2010, supported by both Cisco and NSF funding which was awarded to a number of institutions. It is an outgrowth of an earlier research effort, Router Reliability Research (R3), an investigation initiated by some of the authors into new software architectures for large-scale multiprocessing core routing systems, which will be discussed later in this paper.

Nebula is intended to encourage the *cloud-based future Internet*. In doing so, it addresses many challenges for which solutions would be useful even in today's world. Nebula is based on three architectural elements:

- Nebula Core (NCore), that provides a highly-connected graph of ultra-reliable routers to interconnect data centers;
- Nebula Virtualizable and Extensible Network Techniques (NVENT), that provides an extensible future control plane with more transparency and control for applications; and
- Nebula Data Plane (NDP), which provides robust fine-grained policy enforcement and subsumes many roles that are now filled by a zoo of middleboxes.

Figure 1 is a conceptual illustration of the Nebula architecture with these three elements in place:

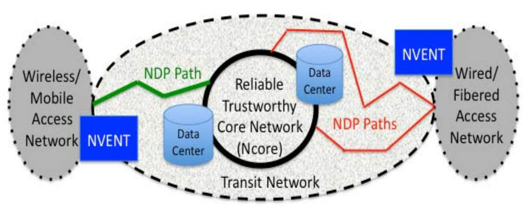


Figure 1: The Nebula Network Architecture: NCore, NVENT and NDP

We will give a very brief description of the three components here; more information is available on the Nebula project web site, including a white paper and project overview slides.

The NDP [Naous10, Popa09] is a new packet format incorporating cryptographic tokens that demonstrate each "realm" (roughly equivalent to an autonomous system, but realm borders are defined by public keys). Each realm must consent to carry the traffic under some specified policy, e.g., HIPAA-compliance for health care data mentioned in the introduction. Two things required for a forwarder to forward the packet are a proof of consent (that the packet would be passed) and a proof of path (that the path was taken). These cryptographic tokens require 42 bytes per realm at present – based on our analysis, about 200 bytes will be required per packet. We are trying at this stage to resist premature optimization, as the major goal is robust policy enforcement. Some preliminary experiments with an FPGA-based prototype have shown roughly 4 Gbps performance, so high performance is achievable.

The purpose of the NVENT control plane is to furnish an application programming interface (API) that can specify attributes for [Birman03], and to acquire [Loo05], paths through the network. For example, Figure 1 illustrates that an NVENT system for an attached access network is providing two paths using NDP – this might satisfy a reliability requirement (e.g., graceful degradation of service), an increase in capacity (through network striping [Traw95]) or some other desired property. This form of reliability requirement might exist for a medical application utilized by the potential insulin pump mentioned in Section 1, where a continuous glucose monitor might sample every five minutes or so. In such a situation, the data rate would remain fairly low, but the reliability must be extremely high. NVENT nodes also fill the role of policy and consent servers for NDP.

The NCore architecture uses striped connections between a data center and core router, and then again stripes amongst a set of core routers. Figure 2 illustrates:

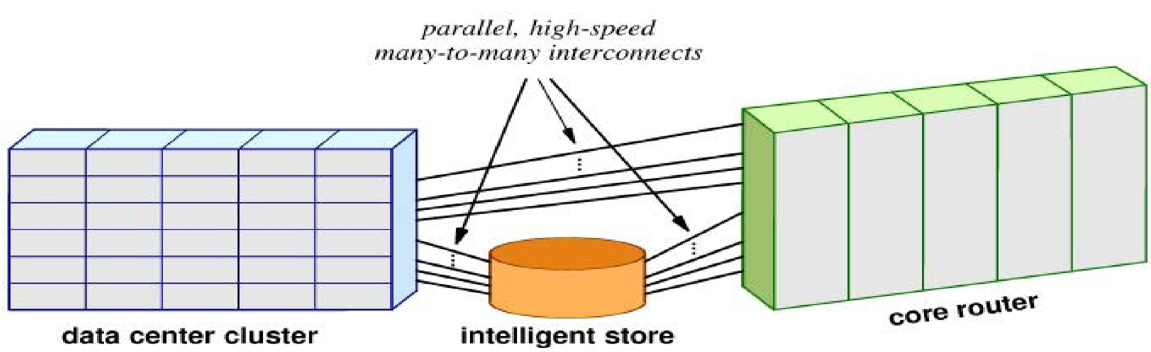


Figure 2: NCore has a rich connection graph for high reliability and performance

In our view, the rich graph connectivity depicted offers many benefits, amongst which are the capacity to resist denial of service (DoS) attacks, resistance to failures and the ability to load-balance or physically isolate nodes for reliability. The use of network striping techniques allows aggregation of links for higher capacities; this might permit, for example, the migration of virtual machines (VMs) for load-balancing, latency or to "follow the sun" (*i.e.,* open stock exchanges).

A key part of NCore is the use of multi-chassis core routers. These can be viewed as large-scale multiprocessors or cluster computers, and can be transformed into ultra-reliable routing systems with the type of software architecture enhancements discussed in the next section.

2.2 New Software Architecture for Cloud Routers

The Cisco research effort is seeking an innovative response to the heightened demand for reliable content distribution over the Internet, while simultaneously encompassing legacy systems (through emergency services such as 911/telephony, with five nines expectation). The purpose is to maintain significant existing investment to date, constituted currently by delivery over highly custom embedded machines, in conjunction with a migration onto a newly created environment. Relying on the isolation of applications from the fault tolerant infrastructure, the result is a fully distributed fault tolerant system, which has also been designed to provide a platform for the smooth integration of future system developments.

The complexity and performance demands of the modern Internet have made core routers into large-scale parallel processing devices. Each line card can be realistically viewed as a high performance processor that is primarily tasked with managing multiple high-speed I/O streams (*i.e.*, the packets). Much of the line card's hardware functionality is devoted to offloading and accelerating packet processing activities such as flow identification and forwarding, and may include features such as access control, rate policing and queue management schemes. Topologically, the line card manages a link layer interface to another line card, a WAN connection, or a host interface. Line cards are interconnected through an internal *switching fabric*. The fabric is a specialized router-internal network that optimizes communication amongst the line cards for high throughput and minimal collisions. Control processors for the router aggregate and share adjacency and policy information across the switching elements, as determined by routing protocols (e.g. BGP) and user-applied policies.

A modern core router can be configured with hundreds of line cards, distributed across multiple chassis interconnected by fiber optic links. While often (naively) thought of as a processor with some line cards attached to its I/O bus, large router configurations are necessary to minimize hop counts, consolidate management and minimize cost, energy and real estate footprint. The analogies to scalable cluster computing are very strong.

An important issue with scale is the likelihood of failed components, which for a given constant component reliability increases with scale. Since the incentives to scale configurations are compelling, failures become more of an issue and motivate the application of fault tolerant computing techniques to modern routers.

Particular goals include an overall "always-on" model that allows for multiple concurrent software versions, live upgrades [Hicks05] and robust failover for processes. While line cards connecting customer equipment with computer host interfaces cannot recover all state, fault tolerant protocols to interconnect core routers (such as a fault tolerant BGP based on a fault tolerant TCP protocol) can overcome many intermittent failures. The live upgrade / versioning issue can be addressed with virtualization technologies similar to those which enable cloud computing. Fault-tolerant storage of state in the router (*e.g.,* information contained in tables) can be made robust with new data structures such as distributed hash tables across multiple independent compute elements. More generally, replication to achieve redundancy is a very powerful strategy, and can be used to provide the software equivalent of a "hot spare" capability to the router control software. Each of these techniques is being pursued with the overall goal of downtime, for one or more routers in a "routing complex."

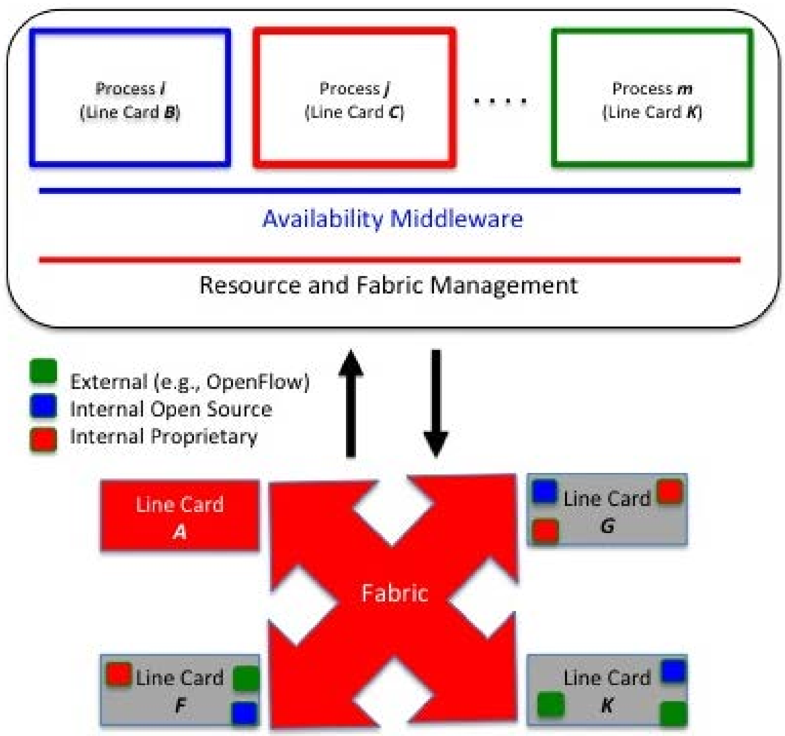


Figure 3: An advanced software architecture leverages core router hardware

Figure 3 illustrates a logical division of software layers (in the top box), with multiple applications instantiated as processes operating on line cards, and its further instantiation shown below on a set of line cards. The colors indicate the type of software, with a green box indicating external software (which could be loaded while the router is in operation), a blue box indicating Nebula/R3 open source, and a red box indicating vendor proprietary software. The virtualization [Birman03] in this instance protects not only resources but isolates software constrained by different intellectual property regimes. It should be clear that specialized functionality (e.g., per-customer services

[Alexander98]) is only one business model direction enabled by this approach.

3. IMPLEMENTATION CHALLENGES FOR NEBULA/CLOUD VISION

In the following sub-sections we consider some of the technical and non-technical challenges to developing and migrating to new architecture.

3.1.0 Technical Implementation Challenges: Fault Tolerant State Store (FTSS)

Some of the crucial steps in implementing this Nebula/Cisco vision for a cloud-enabled network involve actualizing fault-tolerant protocols and fault-tolerant state storage in the router. As with many other aspects of our system, we accomplish this by using the current Internet fabric and improving its resilience, while preserving backward interoperability.

Intra- and inter-domain routing protocols represent typical classes of latency-sensitive, critical, "online" applications. Protocols such as BGP compose the foundation for a functional Internet, which therefore demand quick failover, failure resilience, and speed. Yet BGP, as it is currently deployed, exhibits serious resilience and availability shortcomings. For example, complete recovery from BGP process crashes on routers is basically now done by remote synchronization of full Internet tables from potentially distant peers, which is a sub-optimal feature. However, we believe deficiencies of this nature to be inherent to the traditional resilience models executed in practice, perhaps most significantly 1+1 redundancy, rather than to the protocol itself.

Incrementally improving certain BGP characteristics, such as resilience, stability and Mean Time To Recovery (MTTR), while preserving the current protocol and requiring only minimal modifications to its existing codebase, is a highly desirable scenario to both Internet carriers and equipment vendors. To reiterate the relevance of the cloud model, we are able to achieve our goal largely by building on a paradigm that supports the separation of data from processing. In addition, we realize fault tolerance through a focus on safeguarding the application state data.
The major architectural element that allows for this approach is called FTSS (Fault Tolerant State Store), a distributed, resilient, high-performance, in-memory data store running across router components. It is designed to rapidly store, replicate, and retrieve arbitrarily structured application state. Our FTSS prototype is essentially a performance-optimized 1-hop distributed hash table (DHT); being malleable, it sustains failures, additions and replacements of underlying storage elements, while also providing automatic load balancing. The purpose of this format is to prevent overall unavailability for any subset of stored application data, in the event of multiple failures across any K storage elements. The store itself is specifically optimized for write intensive operations, both latency- and throughput-wise. It is tailored to scale well, takes advantage of underlying testbed size and characteristics, while adapting its communication and routing optimizations.

An example of the use value for such a fast store is the need to checkpoint online applications, expected to be highly responsive while operating at high throughputs and low latencies, and allowing for fast recovery as processes fail. We have successfully used FTSS to create a resilient BGP, with no modifications to the protocol itself, and minimal adjustments to off-the-shelf BGP implementations for protocol machinery. This method permits the resulting system to be deployed or migrated on computer clusters, Internet routers (*e.g. as prototyped on* Cisco CRS), or a combination thereof.

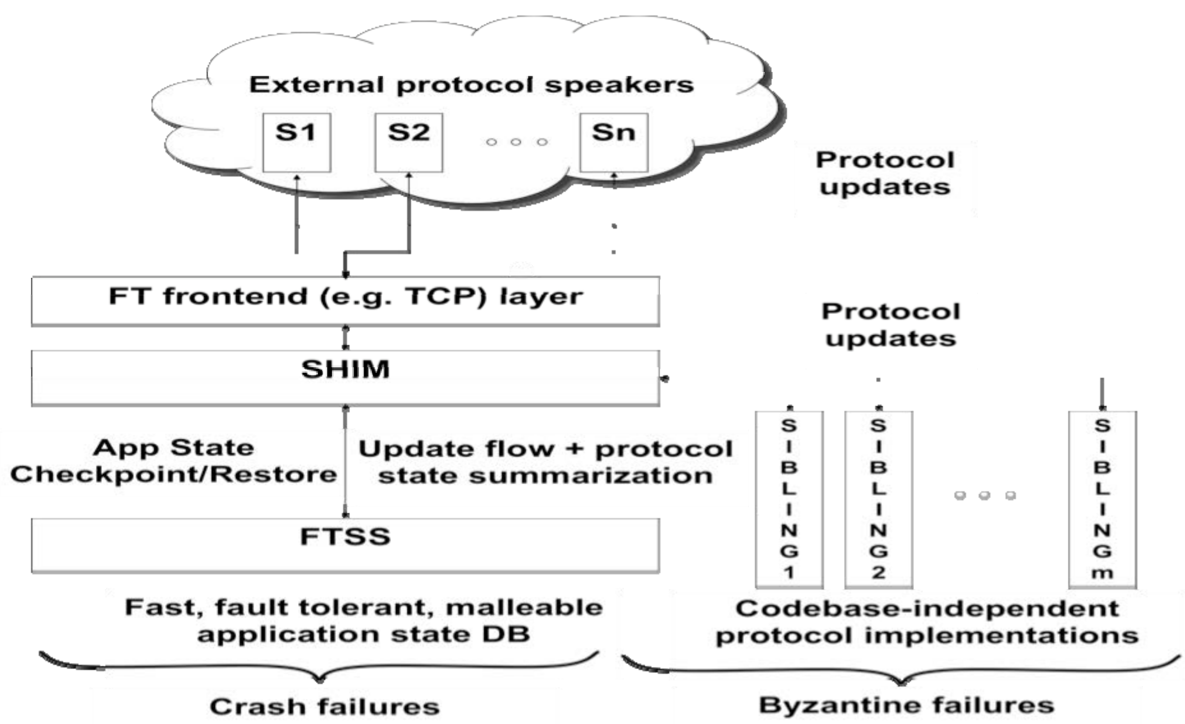


Figure 4: Resilience-enabling online processes such as BGP

In Figure 4, we outline FTSS use for fast, K-redundant state replication. Byzantine failures (*e.g.* implementation bugs) are handled by separated modules, such as through feeding independent implementations of the same process state machine in parallel. The state stored inside FTSS, received from a protocol-specific module called the *shim*, should therefore be as "raw" as possible so as to avoid any contamination from processing in the state machine. The shim delays update TCP ACKs to the remote side until replication has occurred; therefore, in any failure scenario, updates have either been persisted or will be retransmitted. In this fashion, correct K-redundancy against failures can, in principle, be achieved with only one running copy of each of the involved process types. While more than one process copy can still clearly be deployed (*e.g.* to maintain "hot spare" processes), there is never the need to maintain up to K copies of each process type, and the latter scheme displays much more efficient resource usage.

3.1.1 Technical Implementation Challenges: TCP with Session Recovery (TCP-R)

To return to the diagram in Figure 4, we will now discuss the "fault-tolerant front end" (i.e. TCP) layer situated between the shim and external protocol speakers, i.e. BGP, which is designed to address further vulnerabilities inherent in current BGP deployment concerning router availability and recovery. We call this layer TCP-R, for "TCP with session recovery," and it is structured to target the following issues.

BGP servers communicate with their remote peers via TCP sessions, and if these sessions happen to be disrupted (perhaps in the case of a fail-over within our fault-tolerant BGP implementation), those remote BGP servers may react in unexpected ways that can harm availability, such as by routing around a failed router. Slow resynchronization can subsequently occur when the sessions are re-established, causing routing instability delays of up to several minutes for core Internet routers. These routers operate under very high loads, and may well have tens or hundreds of remote BGP peers. During these periods, problems such as route flaps, "black holes" or routing "loops" can arise.

As a preventative measure, BGP servers often implement a "graceful restart" to handle such events. The value of a restarted BGP server's initial state is empty, but it restarts on a hardware router that has maintained active hardware routing tables. Proceeding under the assumption that those tables are mostly correct, the BGP itself recovers, but the routing tables are left in place and neighboring routers may continue to send traffic. The intended goal is for BGP, which is functioning based on increasingly stale routing state, to resume active control quickly enough for this period of inconsistency to be brief. However, this method has not proven to effectively eliminate the problems visible in ungraceful restarts.

The solution we have developed to approach this involves masking a BGP fail/restart from all neighboring routers. We basically graft a new TCP connection onto an old one, in such a way that this event is made invisible to the remote endpoints holding the old TCP connection. Technically, it is possible to compare this to the behavior of a standard network address translating (NAT) box, which effectively grafts a TCP endpoint that believes itself to be connected to, i.e. server X on port P, while in reality those values are different. TCP-R achieves similar results through a TCP session's internal sequence numbering, which is used to identify bytes within TCP's sliding window. If BGP fails over, the new server restarts with the same state prior to its crash, which has been stored in FTSS. It restarts in a state prepared to finish any interrupted send of a BGP update, and ready to read the next byte in sequence of input from a peer's update. Because its per-flow state is only a few bytes of session-related data, TCP-R can handle tens of thousands of concurrent flows. The system that comprises FTSS, FT-BGP, and TCP-R, allows for router failures to be completely concealed from remote peers, and maintain the appearance of always-on, non-stop routing.

As for our technical requirements, no changes were made to the O/S kernel or the TCP stack used, other than recompiling the kernel with a standard Linux packet-filter package. Regarding speed, we are in the process of measuring numbers for the time delay required for BGP's migration from node to node, but best-scenario results currently suggest they are in the tens of milliseconds.

3.1.2 Technical Implementation Challenges: System IS-IS (SIS-IS)

Another issue we have been compelled to address, relevant to constructing a large distributed system, is that of automatically organizing and configuring a system that will comprise many processes and process types spread across many execution elements. Furthermore, these processes must be able to quickly and reliably find each other across the system.

System IS-IS (SIS-IS) is a lightweight system used to register processes within a distributed system, based on the use of Link-State routing protocols. Link-State protocols such as IS-IS [Oran90] and OSPF [Moy98], with their ability to reliably synchronize global system knowledge, have become the foundation of many carrier and enterprise routing networks. SIS-IS, prototyped using Linux and Quagga [QUAGGA], exploits these strengths so as to allow a large group of distributed processes to easily identify each other, both by type and location, across a set of processing elements. In addition to building arbitrary communication meshes between both sibling and other cooperating processes, knowledge of the global process state also allows any individual process to enter new processes into the overall system. The goal of these process additions is to enable both the desired scale across the whole system and the desired physical distribution across available processing elements, so as to meet overall system reliability requirements. This reliability is achieved by instantiating a specific process activity (*e.g.* routing, statistics collection, management, etc.) into a set of identical, cooperating sibling processes, all executing on different hardware components. The results from these sibling process groups are then compared prior to evaluating, and consequently selecting, them based on their validity, with the intent of removing any results that contain errors due to software or hardware faults.

3.2 Non-Technical Challenges

Success in resolving the technical issues will not be sufficient to ensure realization of the Nebula/R3 vision. From a value-chain perspective, we expect that the Internet services will be supported over facilities owned and controlled by multiple complementary and competing cloud service providers. Ensuring the security and reliability of end-to-end services while fostering open and vigorous facilities-based competition poses significant commercial and regulatory challenges.

The problem of ensuring interconnection across multiple networks is hardly new and underlies a history of extensive telecommunications regulation, but the technical, business, and policy challenges become immensely more complex in a world of cloud computing. First, the rise of cloud computing does not imply the decline of edge-based computing any more than the rise of electronic communications implied the end of paper-based communications. Second, the range of resources that need to be transparently shared and integrated is greatly expanded (transport, storage, computing cycles, and power). Third, the range of capabilities to be supported is much more ambitious (increased need for diverse QoS and security to support both sharing of video entertainment, health records, and public safety communications on much faster time-scales and across more diverse physical infrastructures ranging from fiber to ad hoc wireless). By focusing on a few prototypical challenges, we expect to be able to better highlight the challenges. One of those core challenges is the need to interconnect core routers across disparate ASes (where from an economic perspective, what we mean by AS is a centrally-managed cluster of networking resources, where the central management refers to the economic management of those resources – property rights to manage CAPEX/OPEX decision-making, including contracting for wholesale/retail services). The saliency of these issues was highlighted in a recent talk by Vint Cerf wherein he noted that the networking community with respect to interconnecting cloud resources is confronting a situation that is comparable in challenge and import to that which prevailed at the creation of the Internet [Cerf11]. The interconnection issue is attracting current attention relative to the discussions over Network Neutrality (network management) regulation [Stelter10] and as a consequence of the interconnection flap between Comcast and Level 3.

Clearly as we migrate more socially and economically diverse and important applications onto cloud resources, the challenges of ensuring appropriate reliability expand. The Nebula/R3 vision anticipates enabling an ultra-reliable Internet routing core via a fully distributed, high performance, fault-tolerant software platform that provides virtualized access to distributed data centers. Since different applications have different requirements and abilities to pay for security/reliability, it will be challenging to design fair and efficient resource allocation and cost recovery mechanisms.

The core switching fabric of modern telecommunications networks and electricity grids are designed to meet the requirements of "5 9's" reliability – implying availability asymptotically approaching 100%. This is viewed as a requirement for critical basic infrastructure. The Nebula/R3 goal is to achieve a similar level of highly reliable core routing functionality. A better understanding of how the incremental costs of enhanced system reliability might be shared across competing service providers is needed.

4.0 CONCLUSIONS AND NEXT STEPS

The cloud computing model provides a new form [Armbrust09] for networked computation and is a rich source of new applications. The challenges posed by the cloud computing model include high performance, high availability, flexible configuration and policy enforcement. The Nebula project is attempting to address these challenges in a comprehensive and coherent way, including thinking about regulatory policy and economic implications of new technologies, as well as the regulatory hurdles to their adoption.

The cable industry has been characterized by rapid introductions of new services. The rethinking of core router software architecture we have described here enables rapid deployment of these new services and capabilities, allowing for concurrent execution of multiple software versions, possible run-time updating of software systems and an "always-on" availability mode. More generally, the cloud computing model is well-suited to what can be perceived as in-network computing and data, lessening the burden on set-top box and cable modem technologies, thus reducing technology transitions at customer premises.

We expect the next steps to be deployment of software bundles made up of open source and proprietary software that result in an ultra-reliable router. Over the long-term, as our router software model evolves, we expect a software marketplace to emerge, with vendor communities competing to deliver novel products to service providers and their customers.

5.0 REFERENCES

[Alexander98] D. S. Alexander, W. A. Arbaugh, M. W. Hicks, P. Kakkar, A. D. Keromytis, J. Moore, C. A. Gunter, S. M. Nettles, and J. M. Smith, ''The SwitchWare Active Network Architecture,'' *IEEE Network Magazine, special issue on Active and Programmable Networks* **12**(3), pp. 29-36 (May/June 1998).

[Armbrust09] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", *Technical Report No. UCB/EECS-2009-28*, Electrical Engineering and Computer Sciences, University of California at Berkeley, February 10, 2009.

[Birman03] Ken Birman, "The League of SuperNets", *IEEE Internet Computing*, **7**(5) 2003, pp. 92-96.

[Cerf74] V. Cerf and R. Kahn, "A Protocol for Packet Network Intercommunication", *IEEE Transactions on Communications*, Vol. **COM-22**, No. 5, pp 637-648, May 1974.

[Cerf11] V. Cerf. "Re-thinking the Internet," talk presented at Stanford, February 2, 2011, available at: http://www.youtube.com/watch?v=VjGuQ1GJkYc.

[Comer06] Douglas Comer, "Internetworking with TCP/IP: Volume 1: Principles, Protocols and Architecture, 5th Edition", Prentice-Hall 2006.

[Fano65] R. M. Fano, "The MAC System: The Computer Utility Approach," *IEEE Spectrum*, vol. 2, pp. 56-64 (Jan. 1965).

[Hicks05] M. Hicks and S. Nettles, "Dynamic software updating", in *ACM Trans. Program. Lang. Syst*. 27, 6 (Nov. 2005), pp. 1049-1096.

[Jacobson88] V. Jacobson, "Congestion Avoidance and Control", in *Proc. SIGCOMM 1988*, Stanford, CA., pp. 314-329.

[Loo05] Boon Thau Loo, Joseph M. Hellerstein, Ion Stoica, and Raghu Ramakrishnan, "Declarative Routing: Extensible Routing with Declarative Queries", ACM SIGCOMM Conference on

Data Communication, Philadelphia, PA, Aug 2005.
[Moy98] J. Moy. "OSPF Version 2," *RFC 2328*, April 1998.

[Naous10] Jad Naous, Arun Seehra, Michael Walfish, David Mazières, Antonio Nicolosi and Scott Shenker, "Defining and enforcing transit policies in a future Internet," *Technical Report TR-10-07*, Department of Computer Science, The University of Texas at Austin, February 2010.

[Oran90] D. Oran. "OSI IS-IS Intra-domain Routing Protocol," *RFC 1142*, February 1990.

[Popa09] Lucian Popa, Ion Stoica, and Sylvia Ratnasamy. "Rule-based Forwarding (RBF): improving the Internet's flexibility and security", in *Proc. ACM Workshop on Hot Topics in Networks (HotNets),* October 2009.

[QUAGGA] "Quagga Routing Suite," available at: http://www.quagga.net

[Stelter10] Brian Stelter. "F.C.C Faces Challenges to Net Rules," *The New York Times*, December 22, 2010.

[Traw95] C. Brendan S. Traw and Jonathan M. Smith. ''Striping within the Network Subsystem,'' *IEEE Network*, pp. 22-32 (July/August 1995).

COOPERATIVE LOAD SHARING AMONG LOOSELY COUPLED OR INDEPENDENT VIDEO SERVERS

Xavier Denis
Robert Duzett
ARRIS

Abstract

As video delivery systems are tasked with larger and larger content libraries, significantly larger streaming loads across increasingly broader geographies, and a wide diversity of client devices, the need for more dynamic scaling and virtualized provisioning of large on-demand resource pools is apparent.

This paper examines strategies and policies for enabling a logically-shared video delivery load across multiple independent but networked video servers, and provides insight into practical applications such as efficient and scalable models for the origin, mid-tier, or edge caching points of a Content Delivery Network (CDN).

INTRODUCTION

The fast growth of video traffic to computers, tablets and other mobile devices illustrates the profound changes affecting the ways in which people access video content. To compete with an increasing array of entertainment options, MSOs must harness the strength of their programming and work to facilitate access to it across the wide array of devices used by consumers.

These market forces are bound to have a significant impact on infrastructure and operations. Content providers have more options today to reach their target audiences. To keep the upper hand, MSOs need more efficient content ingest / packaging operations to enable fast capacity growth and quickly turn around assets that can be consumed on multiple platforms and client devices. As content continues to proliferate, consumption will increase, driving networks to near capacity. In fact, experts predict managed IP video traffic will quadruple over the next 4 years.

Instead of locally-dedicated servers and storage systems with bounded growth limits and limited load sharing, today's video loads call out for networked resource pools (processing, storage, and streaming resources) that can be easily deployed at discrete locations within a Content Delivery Network (CDN) architecture. These networked resource pools respond to changes in load levels and format mixes with agility, provisioning the resources of semi-independent and loosely-coupled servers and systems in an organic and efficient way.

This paper first discusses the requirements associated with next generation video delivery. After introducing a server pool architecture enabling multiple loosely-coupled independent video servers to operate as one virtual load sharing system, we study the central issues of this proposed approach, including content allocation, load balancing and asymmetric load behaviors. Analysis is supported by simulations driven by real-world field data. In addition, we discuss the advantages of the proposed server pool architecture in terms of scaling / upgrading and review a practical application of the architecture to caching. Conclusions and key findings as well as topics suggested for further research are found in the final section.

VIDEO DELIVERY APPLICATION REQUIREMENTS

The challenges highlighted in the previous section imply new requirements to ensure efficient and scalable video delivery:

- Cost-Efficiency – The need for cost-efficient scalability puts unsustainable pressure on proprietary platform design models. Instead, video delivery solutions must leverage commoditized hardware to exploit the advantages of Moore's Law.
- Low latency and reduced network load – CDN network demands must be minimized to enable high quality delivery, especially for live video applications. Employed as a cache, the video delivery system has a role to play in ensuring low latency and reduced network load.
- Modular and agile scaling – There are multiple dimensions of resource scaling, including streaming capacity, content storage capacity, and ingest capacity. Streaming and storage provisioning requirements vary widely across different deployments and even across different hierarchical levels within a given network. For instance, the two core elements of a CDN, the origin and the cache, dictate nearly opposite sets of provisioning requirements, the former needing lots of inexpensive storage, the latter requiring high I/O performance and streaming density, but reduced storage overhead costs. In other words, the video delivery system architecture needs to support a provisioning model that is flexible while consistently/uniformly efficient, even for fairly asymmetric deployments and upgrades.
- Maintenance / upgrade agility – Upgrades must be straightforward, simple, and quick as possible, having as little impact on the existing system as possible, i.e. modular. Re-building, re-loading, or replacing elements of the existing configuration, in order to perform an upgrade, should be avoided.
- High Availability – The video delivery infrastructure must maintain constant service uptime. This implies redundancy and load balancing across separate resources with fail-over policies.
- Client device independence – modern video delivery platforms must offer flexible and dynamic resource provisioning mechanisms to transparently accommodate different delivery formats and varying workloads to a wide array of devices.

Today's video delivery applications demand increased IO performance, provisioning flexibility, scalability and reliability in addition to accommodating a wide array of content access patterns and diverse file formats and sizes, all of which must be achieved with greater cost-effectiveness. In the following section, we define a flexible framework fulfilling the above requirements while offering great design freedom, including the ability to leverage commoditized hardware and standard networking protocols for optimal cost-performance.

ARCHITECTURAL SCOPE

The architectural scope of this paper encompasses a "server pool" approach to video delivery, in which the server pool is defined as a collection of loosely-coupled, semi-independent servers in a common network, with common reach-ability to video subscribers. Each server is a "node" of the interconnected delivery system. Both the available video content and the streaming load are distributed across the nodes.

The server pool architecture borrows the following key principles from Cloud Computing [1]:

- Shared resource pooling for increased utilization and efficiency
- Reliability by distributing load across multiple hardware instances and eliminating single points of failure
- Scalability by enabling capacity expansions through server instance additions without service disruption
- Maintainability by enabling upgrades and hardware maintenance without service disruption

While the descriptions and analysis in the sections below refer to "content" in general and thus seem to imply a stand-alone content library, a la CDN origin server, these concepts are equally applicable to a pool of cache content nodes. The scalability, redundancy, storage efficiency, and load-balancing efficiency of this solution can serve to optimize the cost-performance and operational efficiency of any appropriate location of a CDN, including the origin server, caching edge sites, or mid-tier caching or library sites. For simplicity of description and analysis, generic references to "content" and "objects" are used, with little mention of caching. However, near the end of the paper, after the core analysis sections, a "Caching Model" section describes methods for mapping these concepts onto a pull-thru cache pool.

Some general principles that are imbued in this architecture include:

- Virtualization – multiple nodes appear, and work together, as one. This applies to the content, whether as a library or as an aggregated cache, as well as the streaming resources and the node pool network as a whole. The pool concept is especially well-suited for exploiting shared end-user connectivity and pooled resources to efficiently aggregate and amplify the unified cache performance. All nodes in the pool have shared affinity with, or common reach-ability to, the video end-user. Policies designed to ensure maximum content heterogeneity across multiple servers' caches drastically increase effectiveness of the aggregate cache, thus positively affecting cache hit performance at the edge and reducing traffic in the network [2]. Once the algorithms are in place that enable the servers in the pool to organically manage their respective resources, the aforementioned policies are easy to implement.

- Network and storage tradeoffs – inter-node communication, hierarchical network loads, and content propagation can all be reduced by judicial provisioning of additional content storage; and vice-versa.
- COTS – in support of the need for easy and large-scale modular resource scaling, using inexpensive and ubiquitous modules, the hardware server platform and underlying system software elements (OS) are assumed to be COTS and the network elements are standard commodity Ethernet devices. Furthermore, this COTS foundation encourages and enables the development of hardware-independent value-add software to implement the concepts introduced in this paper.

Architectural Objectives

- Balance - Balance streaming loads efficiently according to the capacities (content and streaming) of the various nodes. Establish rules and algorithms for provisioning nodes and resources, allocating content objects to nodes, and directing streams to nodes. Inter-node content movement to correct imbalances should be kept to a minimum.
- Scaling - Configure, scale-up, and upgrade overall capacities with minimum disruption to system operation, and asymmetrically (capacities of the various nodes may differ) if necessary. Establish policies and guidelines for adding nodes & capacity.
- Redundancy - Maintain services at rated capacities even in the face of a full node failure. Establish redundancy methods to fail-over lost content & streaming. Ensure:
    - Continued accessibility to all content titles
    - Continued full rated streaming capacity
    - In the case of a failed node in a caching pool, minimal network traffic devoted to content re-acquisition from the CDN origin server

With these objectives in mind, some key architectural elements are proposed to enable pool-based video delivery.

General Approach

The redundancy requirement calls for full rated streaming capacity and continued access to all content in the face of one off-line node, i.e. n+1 redundancy. In this, the architecture shares many of the goals of other RAIN (Redundant Array of Independent Nodes) architectures. This requirement is the most visible driver of the architecture's general methods, briefly described here:

- Store at least two copies of every video content object, with the copies on different nodes. Provision storage for this extra content. Dual-copy content allocation strategies are further described in the next section below, "Allocating Content". It should be recognized here that, in the case of a caching pool, policies may be applied that dynamically vary the level of redundancy across the objects, some with no copies or 1 copy (relying on the origin server for backup) and some with 2 or more, depending on the popularity or streaming load of the object (see the Caching Model section later in this paper).

- Provision video streaming capacity across the system in such a way as to absorb one node going totally off-line. De-rate the streaming capacity of a node to account for the failover streaming capacity that must be reserved in case one of the other nodes fails. In other words, to cover the potential loss of one node, the system will ideally require no more than the streaming capacity of n+1 nodes to ensure an effective streaming capacity of n nodes. Capacity de-rating strategies and equations will be described in a later section, "Directing Streams".
- When a stream is requested, direct it to one of the nodes that has a copy of the requested object. Provisionally reserve an equivalent streaming load at one of the other copies, for possible node failover coverage. Selection of the streaming and failover nodes must consider the current streaming loads and provisional failover streaming reservations of the set of nodes with copies of the object. Stream-direction (node-selection) algorithms will be discussed below.

ALLOCATING CONTENT

In this architecture, content allocation entails replicating each content object on two nodes, being prepared to add a third copy dynamically as needed. This approach is in essence a heterogeneous distribution with limited (the minimum) replication, forming a judicial synergy of network and storage resources. Total provisioned storage is effectively 2 * rated content size, the minimum to meet the node failover redundancy requirement. One could extend this to the extreme and put a copy of all objects on all nodes, i.e. full content replication. This would maximize flexibility and simplicity in stream allocation and failover, as well as the ability to absorb streaming load asymmetries by stream direction only, but at the cost of significantly more storage, especially as the pool's node count increases.

So, again assuming only two copies per object, every object will belong to a specific content node pairing or "object group", which consists of all the objects whose two copies are allocated to the same pair of nodes. An object will belong to only one object group, but a node will intersect multiple object groups, one for each of the other nodes in the system. In a system of n nodes, there will be n(n-1)/2 object groups. In a graph of nodes, if one were to draw an arc between every pair of nodes, those arcs correspond to the object groups. This is illustrated in the diagrams here:

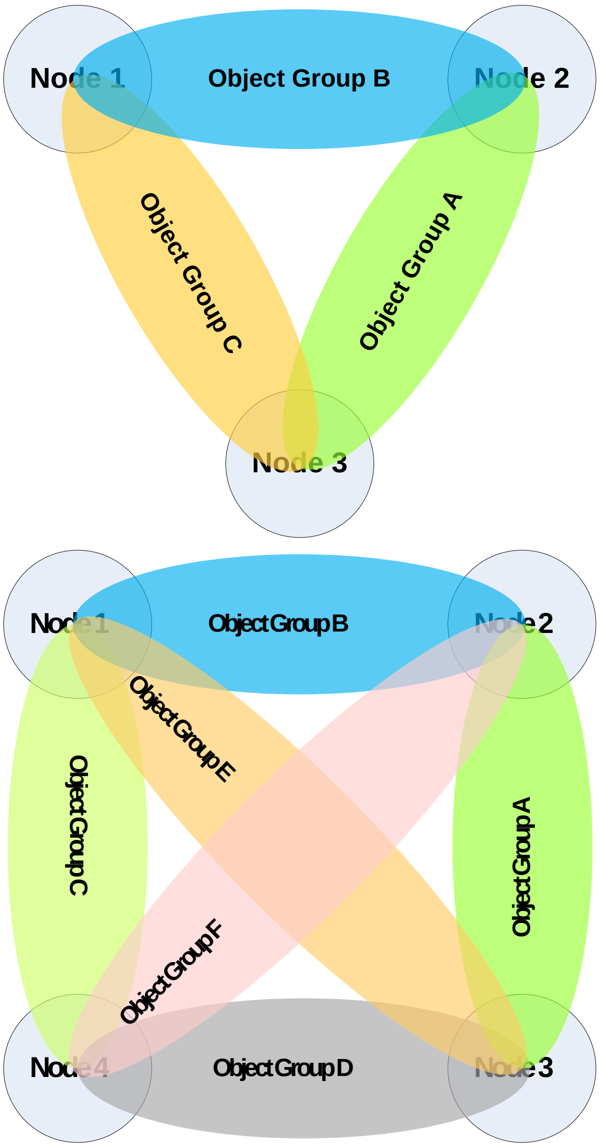


Figure 1 - Object group definitions for 3-node and 4-node server pools

Note that this object group concept can be extended to encompass a variety of node group sizings in the same system (see the "Exceptional Asymmetries" section for a description of 3-copy object groups). For now, the focus will remain on two-node object groups.

Content should be allocated across the nodes in as balanced a way as possible.

Ideally, all the nodes will end up with very near the same amount of stored content (weighted for relative capacities), and all the object groups will also be nearly equal in their storage allocation. A careful, even allocation of content objects across the nodes will lay a smooth, flat foundation for stream allocation – a sea of object copy pairs spread randomly across the array of nodes, ready for a streaming load and failover reserve to be carefully mapped onto it in as balanced a way as possible. Content should be allocated in a way that distributes objects to storage in an apparently random way, without regard to expected popularity, thus naturally mixing popular and less-popular objects within and among the object groups, maximizing the opportunity to absorb streaming hot spots and balance loads using the redundant resources provided (redundancy expands choice and flexibility). In effect, randomized but even content allocation tends to flatten the apparent content usage profile from the perspective of node and network utilization.

Another aspect of content allocation is content pre-placement. For the case of caching pools, pre-positioning of content may or may not be desirable or practical, depending on the specific implementation. Significant reductions in network loading have been shown for metadata-directed pre-placement of cache content. Regardless, one can pre-place none, some, or all of the content while following the allocation scheme mentioned above; and then place new content as it arrives in the same manner; or direct/re-direct streams for pull-thru to result in the same desired placements (more details on pull-thru approaches are given in the "Caching Model" section further below).

DIRECTING STREAMS

Stream requests should be directed to nodes in such a way as to minimize the de-rating of system streaming capacity, absorbing in an optimal way the failover streaming load of any off-line node as well as asymmetries in streaming demand (popularity hot spots). See examples of streaming load asymmetries in the table and diagram below. The streaming load asymmetries are expressed as the largest fraction of total system streams sourced from any object group (multiple "trials" are shown, in which the random elements of the content allocation mechanism are re-seeded). The data in the table was generated with simulations driven by real-world field data. More details on "absorbing asymmetries" are given in a section further below.

Asymmetry - Largest streaming load on an object group	3 nodes	4 nodes	8 nodes	16 nodes
Ideal (perfectly balanced)	.333	.167	.036	.0083
Trial #1	.363	.210	.103	.059
#2	.416	.224	.116	.059
#3	.383	.208	.085	.056
#4	.359	.215	.113	.075
#5	.423	.211	.086	.073

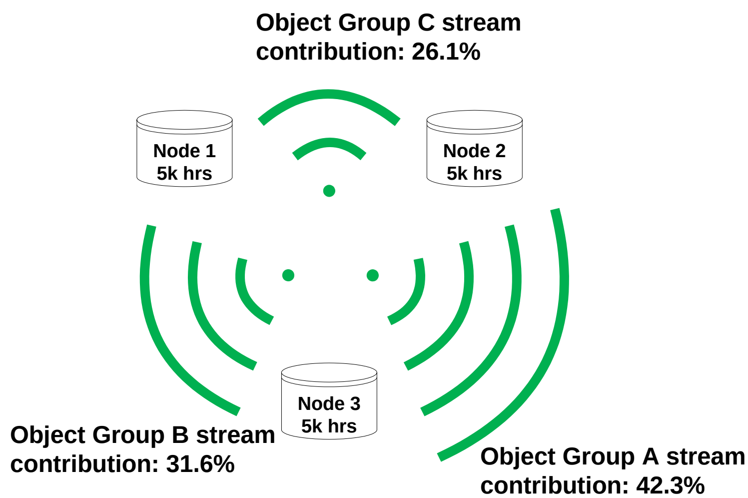


Figure 2 - Asymmetric streaming loads across object groups

When a new stream request occurs, a node must be selected to source the stream. Actually, two nodes are selected for every stream request – one to stream and one to provisionally reserve failover resources for the stream. Both of these nodes must have access to the content object requested and must be able to reach the appropriate transport network with the stream. If content has been allocated in the manner described in the section above, there will be exactly two nodes with copies of the desired object, so selecting one node automatically selects the other. Effectively, the object picks the node pair and the "stream director" merely decides which node of the pair will source the stream and which will shadow the stream for possible fail-over. These two nodes together uniquely describe the identity of the "object group" containing this object and others with their two copies on these same two nodes.

Since every object is uniquely assigned to an object group, the streaming load allocated to the object group at any given time is the aggregated streaming load at that time of all the objects of the group. In addition, an equivalent failover load is also allocated to the group. More specifically, a portion of the object group's streaming load is assigned to one of its associated nodes while the other portion is assigned to the other node in the pairing. The associated failover loads are apportioned to the opposite nodes of the pairing, so the total streaming-plus-potential-failover loads assigned from this object group to each of the two nodes hosting it are always equal.

Every node intersects a number of object groups, each of those groups being uniquely associated with this node and one of the other nodes. Therefore, the set of object groups of one node will overlap exactly one group belonging to another node, but each node's set is unique. Therefore, the total collection of object copies of one node are never the same as that of any other node. Likewise, the streaming loads and provisional failover loads of one node are unlikely to match those of another node.

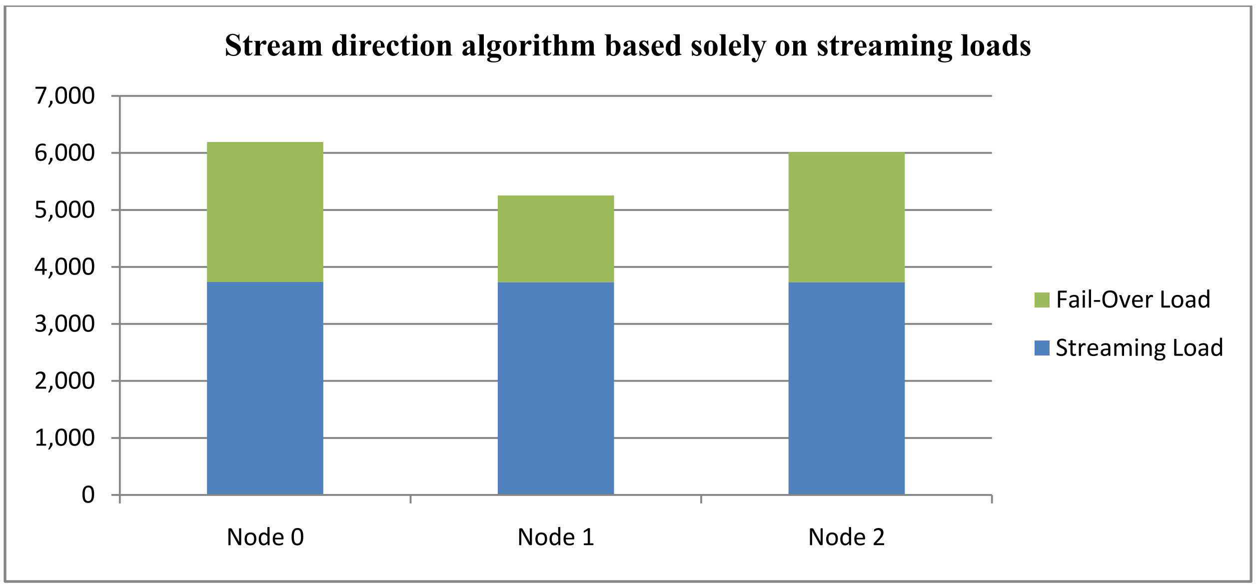
The streaming capacity of a node must be sufficient to cover both its expected nominal streaming load and its worst-case allocated failover streaming load. At a given node and a given point in time, each of the object groups associated with that node contribute to the node a portion of the current streaming load of the object group and a complementary (corresponding to the other portion of the object group's streaming load) potential failover load. The same object group contributes the opposite loads to the other node associated with the object group. The worst-case total potential streaming load of a node at a given time is the **sum** of the actual streaming loads allocated to that node from all the object groups for that node, **plus** the **maximum** of the potential failover streaming loads allocated to that node from its set of object groups. The maximum failover load from among the object groups is used instead of the sum because only one of the nodes in the system is expected to be off-line at any one time and the worst-case scenario for this node is the failure of the complementary node of the object group that contributes the largest potential failover load to this node. So, for time t, the potential load l(t) at a node is given by:

l(t) = total_allocated_streaming(t) + max_allocated_failover(t)

and the required minimum streaming capacity that must be provisioned for that node is the maximum l(t) over the lifetime of the node's current configuration. A simple formula relating node capacity with the peak system streaming load and system size (node count) is given in the "Absorbing Asymmetries" section further below.

The objective in choosing one node over another to source a new stream is to maintain a balanced maximum streaming load across the nodes of the system, i.e. to match a node's worst-case load against its relative streaming capacity in the system. Since the worst-case streaming load, and thus the required capacity, of a node, is the total current streaming load plus the maximum current potential failover reserve, this is the metric that should be compared when selecting one node over another to direct a stream. Selecting a node to source a stream based solely on the current actual streaming load of the candidate nodes will NOT result in a balanced system but will in fact lead to an extremely off-balance system that will not be able to fully fail-over the streaming load of a lost node. Nominal streaming loads will be balanced, but not the maximum potential load (i.e. after failover).

The graphs below compare two approaches to balancing loads, one based solely on the nodes' current streaming loads and one based on both streaming and maximum failover loads. Results are shown for a 3-node system, showing maximum streams and maximum streams+failover for all 3 nodes. The data was generated from simulations driven by actual field data.



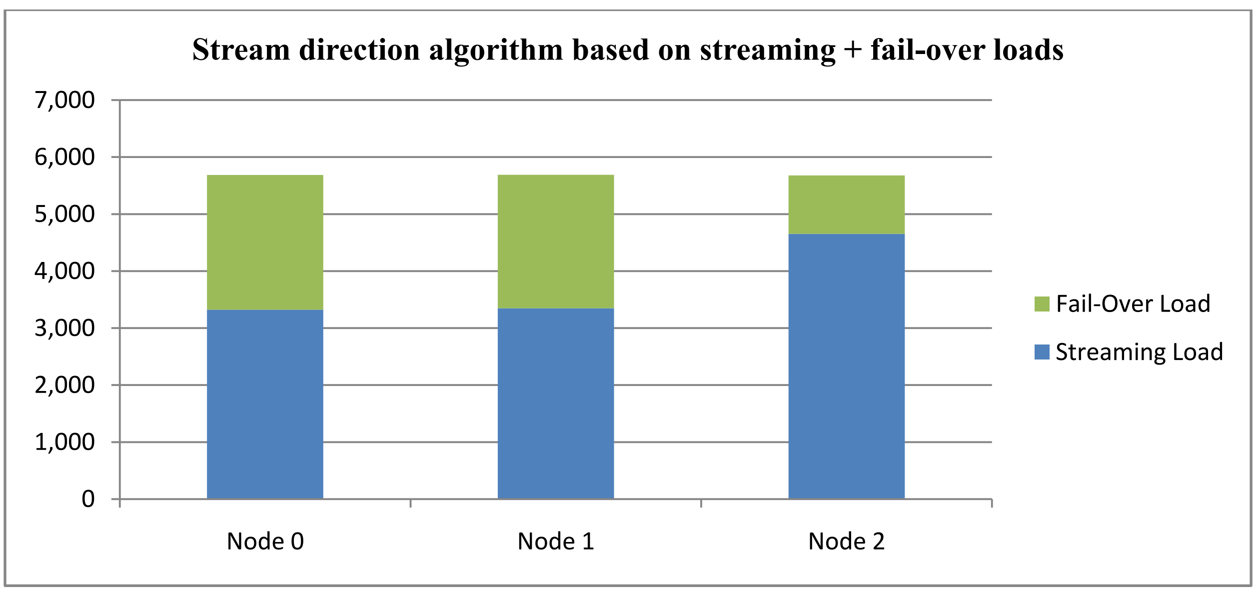


Figure 3 – Comparing load distributions with different direction algorithms

When only streaming loads are considered, the resulting maximum streaming loads are very well balanced, but total possible loads including failover are unbalanced and worst-case is quite high. When streaming+failover loads are considered, however, the streaming loads alone are unbalanced but the total load is well-balanced and lower. Streaming load alone means nothing when maximum potential failover streaming must also be reserved.

When an object is identified to source a new stream, that stream will be assigned to one of the nodes of the pair associated with the requested object's two copies. The node selected should generally be the one with the lowest current potential load, i.e. total streaming load plus maximum potential failover load. This is because the incremental streaming load will always translate completely to additional load on a node, while the incremental potential failover load may or may not add to a node's maximum failover load (a node's potential failover load from this object group may not be the current maximum for the node). So, the stream is directed to the node, of the pair, that has the lowest [streaming plus maximum failover] load, and the failover role for the stream is assigned to the other node.

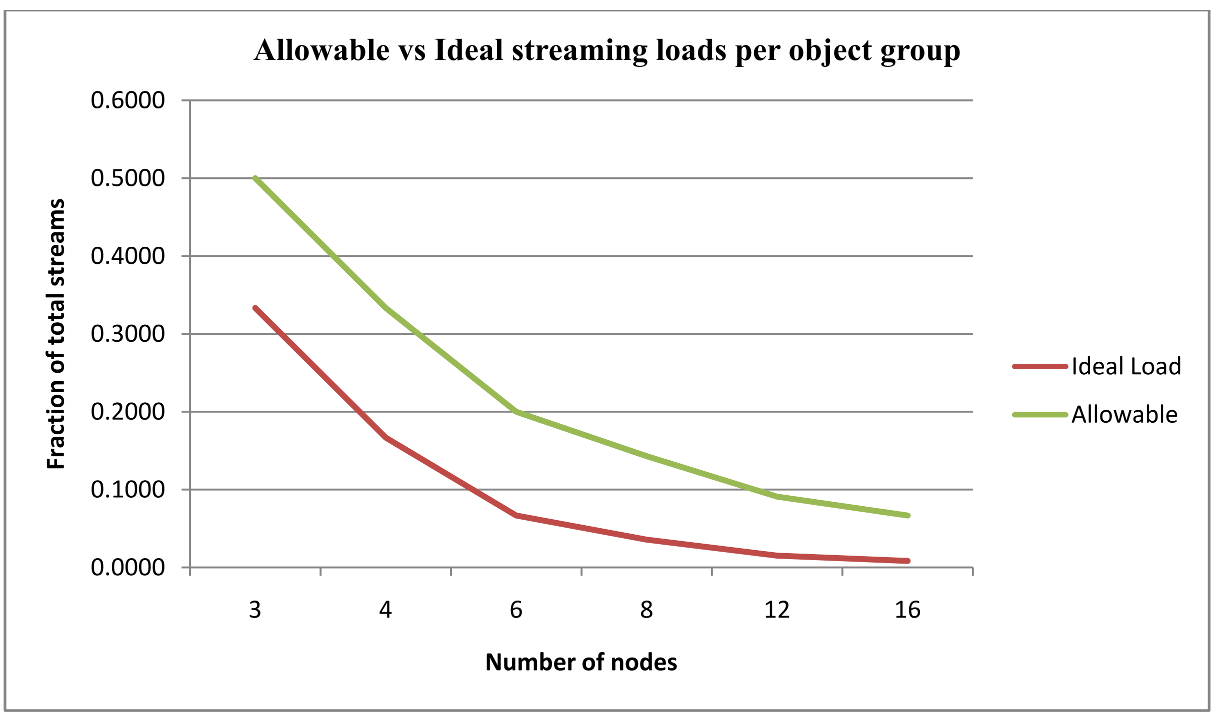
ABSORBING ASYMMETRIES

The asymmetry introduced by node failover has been addressed by the object storage and stream load redundancies described above. By storing two copies of an object on different nodes and accounting for failover streaming loads when de-rating a node's capacity and when assigning a stream, the possibility of a node failure is anticipated and provisioned for.

Note that the architecture's provisioning for content and streaming redundancy to cover a node failure also provides natural flexibility in stream allocation that will support efforts to balance uneven streaming loads across the nodes and object groups. Under most conditions, rules guiding content replication and the over-provisioning of streaming capacity will be sufficient to maintain balance in the face of dynamic load asymmetries (shifting popularity profiles) as well as to absorb a node fail-over, without having to move or adjust content. This has been verified by analysis and by simulation driven by real-world field data (just one example is shown in the graphs above).

In a system in which content objects and streams are allocated as described in the sections above, and given a peak total streaming load "S", the maximum streaming+failover load experienced by any of the nodes should nominally be $1/(n-1) * S$ (this equation simply represents the system streaming load spread across all the nodes but one, possibly failed, node; this is the de-rated capacity of a node). This per-node maximum will hold for a range of streaming load asymmetries described as follows: if each node of a system is provisioned to support at least $1/(n-1) * S$ streams, and the worst-case streaming demand on any object group of the system is less than $1/(n-1) * S$, the nodes will fully absorb the streaming load as well as the failover load of any node failure. The determination of this upper limit to object group streaming load is based on the observation that each node of a pair must be able to absorb the full streaming load of their associated object group because the group's streaming+provisional_failover load is double the streaming load and is evenly allocated to the two nodes. A streaming load, on an object group, greater than the capacity of either of its paired nodes is thus guaranteed not to be absorbable by the nodes. Note that a perfectly even distribution of streaming load would allocate 2/(n(n-1)) * S streams to each object group (there are n(n-1)/2 object groups (node pairings) in a system of n nodes). This means the maximum absorbable object group load is n/2 times the perfectly even (ideally balanced) load. See the table and diagrams below for streaming load asymmetry ranges for various node counts.

#nodes (n)	#object groups (arcs connecting node pairs) 2/(n(n-1))	Minimum Rated Node Capacity – fraction of system streams 1/(n-1)	Range of allowable max streaming load per object group n(n-1)/2 – 1/(n-1)	Allowable/Ideal ratio n/2
3	3	1/2	1/3-1/2	1.5
4	6	1/3	1/6-1/3	2.0
6	15	1/5	1/15-1/5	3.0
8	28	1/7	1/28-1/7	4.0
12	66	1/11	1/66-1/11	6.0
16	120	1/15	1/120-1/15	8.0



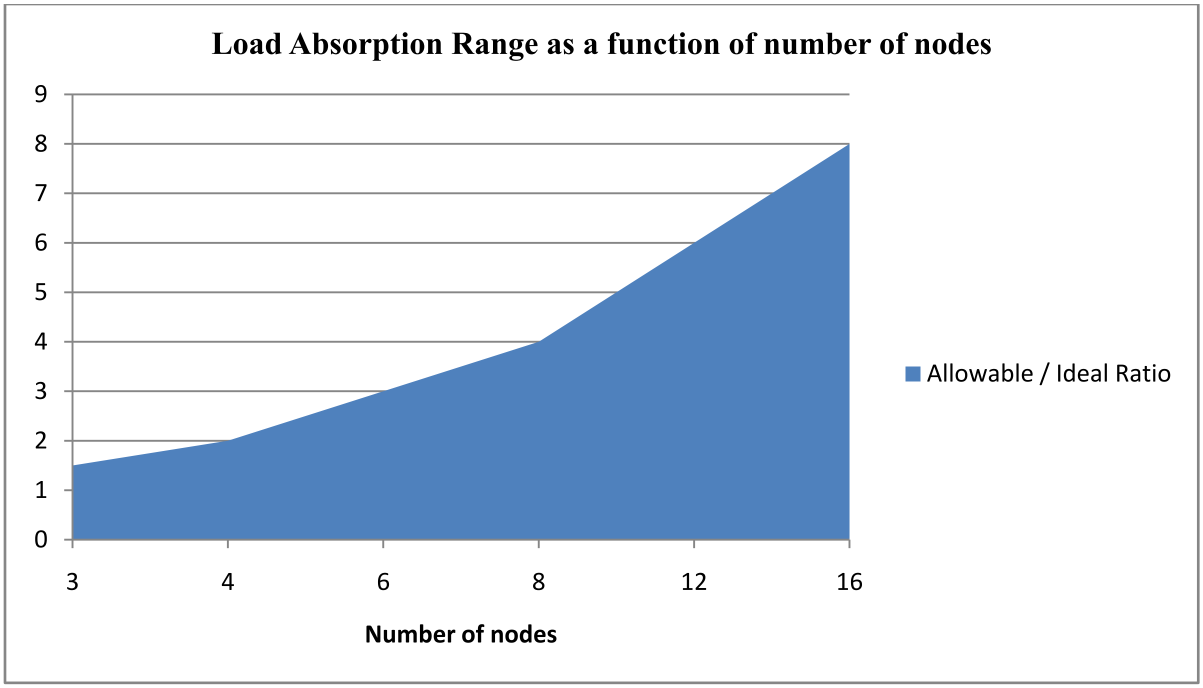


Figure 4 – Headroom for streaming load asymmetries

Simple n+1 provisioning, combined with 2-copy content provisioning and efficient allocation, is sufficient to absorb most practical asymmetrical situations if the stream direction algorithm is also effective (balance streams + max-failover, NOT just streams).

Given this most basic provisioning, a smart content allocator, and a smart stream director, how much asymmetry can be absorbed? The table and graphs above show the theoretical bounds of asymmetry for various system sizes. Below is a graph containing some examples of various load asymmetries, from simulation models driven by actual field data. Shown are the streaming loads of the object groups of a 4-node system, these loads all clustered around the ideal balanced load level and all below the absorbable limit calculated above.

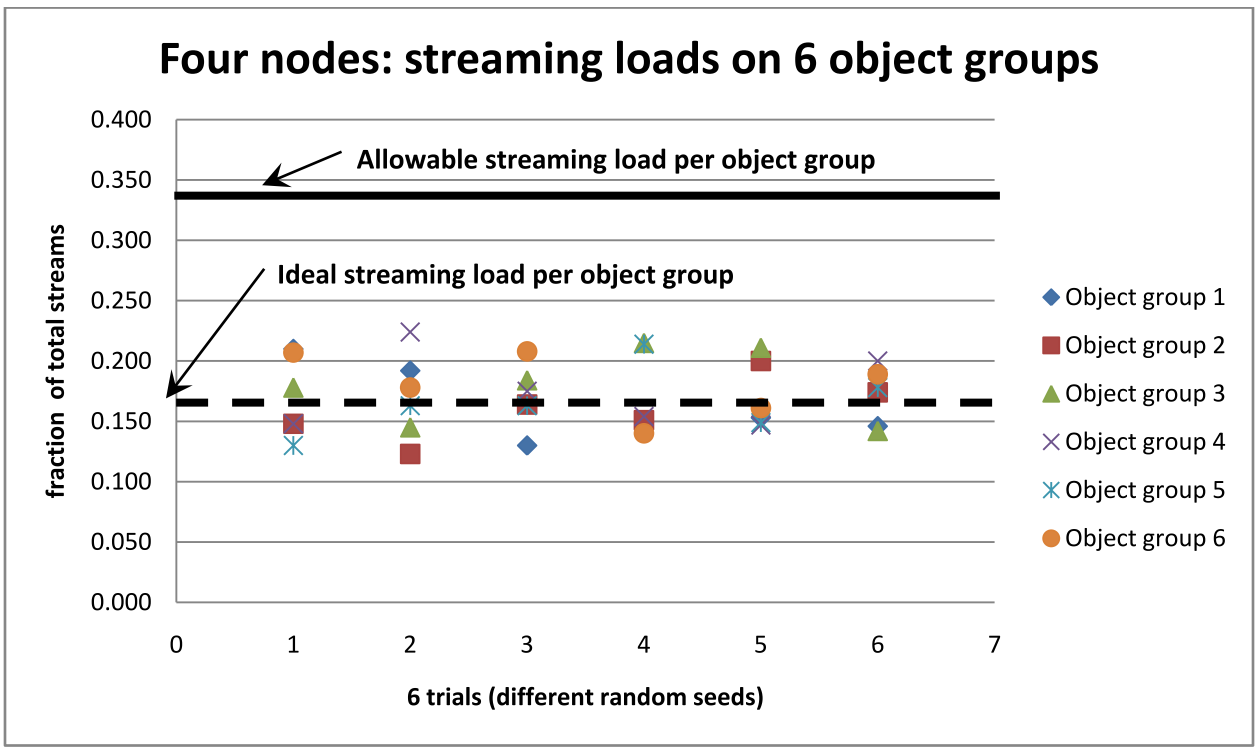


Figure 5 – Absorption of asymmetric streaming loads

Normal everyday hot spots are generally exhibited by a group of popular objects that are scattered randomly across the object groups and so tend to exhibit themselves as minor imbalances in demand. These migrate over time, exhibiting normal fluctuations in user demand but staying within the absorption range defined above. However, shorter-term, faster-ramp and higher-magnitude spikes in demand for isolated objects can also occur. If these are not too severe and/or they occur while no nodes are off-line, they are also generally absorbed successfully.

The exceptional scenario is the sudden demand for an uber-popular object that soaks up a significant fraction of total streaming capacity for some period of time, i.e. the so-called “super bowl” scenario, especially if, unlike the Super Bowl, it’s unplanned and unpredictable. Depending on the number of nodes in the system, an object group that suddenly accounts for $1/4^{th}$, $1/3^{rd}$, or ½ of all streams because of one or two super-hot objects could easily exceed the bounds given in the table above. This is when dynamic propagation of extra object copies to other nodes becomes important.

Note from the tables and graphs above that, although the ranges of relative allowable streaming asymmetry is much higher for large node-count systems, the absolute maximum loads are much smaller than for small node-count systems. Thus, as systems grow to higher and higher node counts, they are actually more vulnerable to isolated uber-popular objects.

Be aware that the inherent absorption range of the node pool can be expanded to virtually any reasonable level by further de-rating the streaming capacity of the nodes. This is an alternative that can be traded-off against more content redundancy (more copies to begin with or more extra-copy dynamic object propagations). Another alternative is to go the other direction and simply plan on absorbing a fixed but reduced level of asymmetric and/or failover loads, as one’s risk tolerance dictates, and accept that some capacity may be unavoidably curtailed (or provisioned as extra load on the origin server and the intervening network).

Exceptional Asymmetries

The system must be able to handle asymmetries that exceed de-rated capacities. As indicated above, the streaming load on an object group should not approach or exceed $1/(n-1) * S$. When imbalances or load fluctuations overtax an element of the existing configuration, dynamic adjustments must be made to rebalance the load. The approach taken by the proposed architecture is to dynamically propagate or pull-thru additional copies of the problematic object to nodes with unused content & streaming capacities.

Adding a third copy of an object creates an effective triangle of nodes and thus three node-pairings to which the object’s streaming and provisional failover assignments can be made. This in effect creates a new object group of 3 nodes tiered above the two-node object groups triangulated by those nodes. It gives the stream director three possible node pairings from which to choose when assigning a new stream, rather than just one. The additional node pairs are available to absorb the excess loads being experienced by the original 2-copy object group.

To accommodate a reasonable 3^{rd}-copy capability, the nodes of the pool must be provisioned with a fractional increment of unallocated content storage. Exceptional asymmetries are generally caused by just a few objects, so the required incremental capacity is relatively small.

SCALING AND UPGRADING

A prominent feature of a networked server pool architecture is its promise of easily scaling up or scaling down the pool's resources by adding or removing nodes or modules. This eases the operational load and costs of system capacity upgrades & maintenance, all with minimum disruption to active operations.

Adding new content objects to the system, when there remains existing storage capacity, is straight forward. The content allocation method described earlier in this paper should continue to work as long as no node fills its storage nor steals from 3rd-copy reserve. If, however, storage is full or near full, additional storage should be added to the system. One could add incremental storage, i.e. independent storage volumes or shelves, to each (or some) of the nodes, or one could add additional node(s) to the existing ones. Either way, the new storage capacity should be evenly and smoothly integrated into the system by, for example, randomly selecting existing object copies to be migrated to the new storage, and disabling the old copy from further stream allocation and ultimately deleting it. This should continue at an acceptable pace until the storage utilization is once again even (by weight) across the volumes.

To add streaming capacity, it is generally easier to add standard nodes to the pool than to add CPU and IO capacity to a node.

While arbitrary resource asymmetries cannot always be well-balanced or efficiently exploited, systems can be incrementally scaled-up with nodes that are provisioned with resource capacities different from those of the original nodes.

CACHING MODEL

This node pooling architecture can be applied to a cache pool as well as it can to a content library pool. In a pool of pull-thru cache nodes, for example, cache content could be provisionally allocated by directing a primary stream pull-through at one node and a secondary pull-through (for potential failover coverage) at another, letting the cache logic and state of the individual nodes determine whether and how long the content stays in the cache. The stream's failover node does everything the streaming node does, including provisionally reserving the streaming bandwidth, except it doesn't actually stream the content. Future stream requests for the same object are directed at the same pair of nodes with both nodes hitting and/or updating their caches accordingly and one being chosen to stream while the other provisionally reserves bandwidth for failover.

While some objects will take up storage space in two caches, this is a minimum redundancy ensured by the disciplined content allocation mechanism. If the streaming activity of an object is sufficient to cause it to naturally hold a place in the two caches, the potential impulse load on the network and other nodes caused by a failed node will be reduced, because another copy of the object is already cached. On the other hand, the limited redundancy approach of the content allocation scheme actually maximizes the uniqueness and heterogeneity of content across the caches, thus improving the cache efficiency of the server pool over ad-hoc methods that allow caches to all pull from the same large library .

Note that policies may be applied that vary the level of redundancy across the objects, some with no copies or one copy (relying on the origin server for backup) and some with two or more, depending on the popularity or streaming load of the object.

Another approach would initially assign a single node (and provisionally its cache) to an object and its associated streaming load, until the demand for that object warrants an additional caching node to offset the load and provide valuable failover capacity (again, avoiding an impulse load on the other nodes and the upstream network if a node fails).

CONCLUSIONS, IMPORTANT FINDINGS

- An efficient server pool architecture can be effectively applied to optimize the cost-performance of any location in a CDN, including the origin server, a caching edge site, or a caching mid-tier site.
- Storage costs can be reduced significantly by provisioning content storage (whether library or cache) for minimum inter-node redundancy (i.e. 2 copies). Even this limited redundancy provides the stream director with significant headroom and flexibility for absorption of both failover demand and asymmetric streaming loads (hot spots). Exceptional events are then sufficiently handled by propagating extra copies of selected objects.
- A careful, even allocation of content objects across the nodes will ensure a flat foundation for stream allocation, minimizing the effects of streaming load asymmetries occurring on top of the content pool.
- Provisioning for an off-line node (n+1 redundancy) is not sufficient in itself to ensure smooth or successful failover. Load balancing and stream direction logic must also come into play to anticipate, and allow for, a worst-case node loss (instantaneous demand spikes), not just to balance current streaming loads. Failover allowances cannot be made arbitrarily at the system-wide level nor by a fixed amount applied equally to all nodes globally. Appropriate allocation levels for streaming and failover will be node-specific, dynamic, unpredictable, and highly variable.
- Re-active content propagation can be avoided or minimized with smart provisioning. Relying solely on intelligent but straight-forward provisioning guidelines and stream direction algorithms, a multi-node pool can be statically configured and provisioned to maintain optimum balance while absorbing node failover and/or significant demand asymmetries, with minimum redundancy and cost. A minimum of dynamic churn (e.g. content propagation, and other dynamic load balancing methods) is required to handle exceptional outlier cases.
- The role of a stream director is simply to decide which node will source a stream vs which will provisionally reserve fail-over bandwidth, but this decision has major impact on load balance and resource utilization. Significant streaming load asymmetries are intelligently absorbed while piggy-backing on the

minimum node failover provisioning (both content & streaming).

FURTHER RESEARCH

- 3rd-copy and 4th-copy dynamics
- Asymmetric upgrades, including >2 different capacities in the server pool
- Command, control, & communications to enable & support decision-making
- N+2 redundancy
- Other affinities & dimensions thereof, including network loads & capacities
- Bit-rate striping
- OTT video cache-ability and how to increase it?
    - How much is currently cache-able on a network?
    - How different is it within an operator's network?
- Transport & storage cost models

REFERENCES

1. *NIST.gov – Information Technology Laboratory – Cloud Computing Program (http://www.nist.gov/itl/cloud/index.cfm)* – Retrieved 5/2/2011
2. *Managed CDN – Optimizing the Behavior of Hierarchical VOD* - Robert Duzett (ARRIS), Jeremy Craven (ARRIS)

APPROACHES TO INTEGRATING CDN TECHNOLOGIES INTO CLASSICAL CABLE VOD PLATFORMS

Charles Hasek
Principal Architect
Time Warner Cable

Santosh Krishnan
Vice President, Product Strategy
Verivue

Abstract

Classical cable video on-demand (VOD) systems have been based on traditional client-server architectures, in which content is replicated on several streaming servers in each geographical location. More recently, cable operators have turned their attention to distributed content delivery networks (CDN) as a solution for expanded content libraries that can no longer be economically addressed by replicated client-server systems. In this paper, we provide justifications for embracing an open standard-based CDN architecture, based on a foundation of HTTP and caching, i.e., technologies that are now widely recognized to have scaled content delivery over IP-based networks. A challenge that remains for classical VOD delivery is to adopt the benefits of such CDN technologies without fork-lift upgrades of the entire existing ecosystem. Here we enumerate a few key modifications to existing components that can enable the creation of hierarchical cache-based architectures. In summary, the proposed modifications can be used as a practical recipe for integrating existing VOD ecosystems into an HTTP-based CDN.

INTRODUCTION

Classical cable video on-demand (VOD) systems were constructed using a traditional client-server approach. In essence, this paradigm consisted of silos of content storage and streaming within each geographical area, typically defined as a head-end. The VOD ecosystem, broadly consisting of streaming servers, storage, back-office software and application portals, would then be replicated from head-end to head-end, essentially replicating content libraries at each location. As content offerings have grown, so has the amount of storage at the edge of the networks.

Because of such a silo-based approach, content is propagated (or "pushed") and replicated at each location, regardless of the number of views or popularity. Such replication consumes an inordinate amount of storage, power and space to store very low-use such as unpopular content and long-tail catalogs. Hierarchical storage architectures would allow for much more content to be stored in very inexpensive storage at centralized locations and only moved to servers at the edge when needed, i.e., based "on demand."

A couple of key technology drivers in the content delivery network (CDN) space are now being utilized in the classical VOD space to enable efficient growth of storage and streaming capacities (e.g., see [1], [2]). The first key component is the usage of standard HTTP for content requests and propagation [3]. In fact, in recent years, HTTP-based protocols are already being used to deliver video content end-to-end, in both real-time streaming (e.g., HTTP live streaming [4]) and progressive download fashions. Such capabilities can be applied towards classical VOD delivery as well. A second key component is use of hierarchical caching components and models to efficiently store and move content. This capability can, once and for all, end the storage proliferation issue associated with traditional architectures.

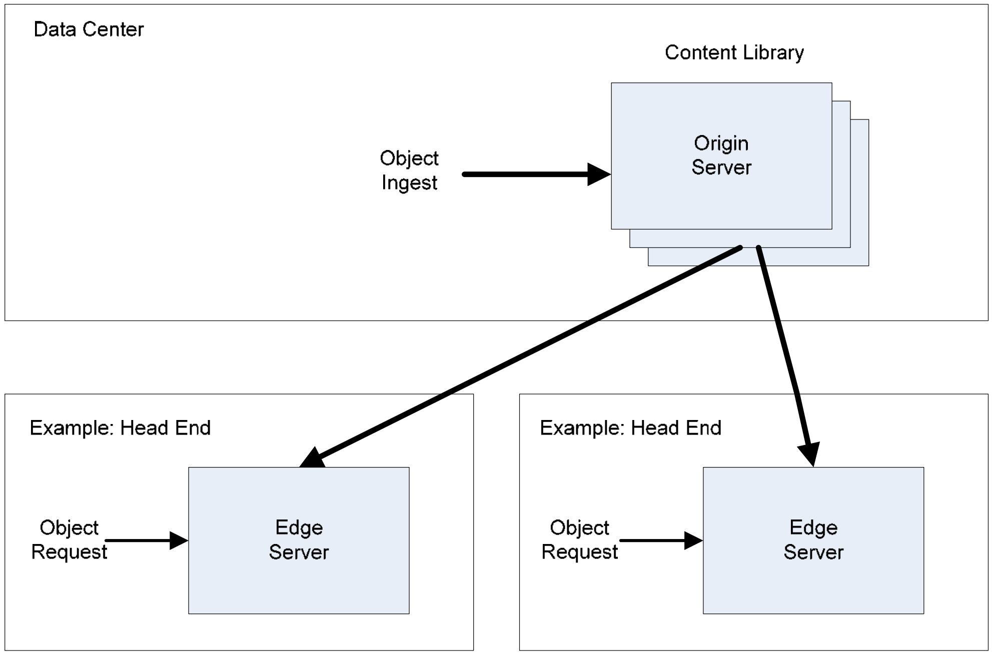


Figure 1: Example of a two-tier hierarchical content delivery network

The adoption of CDN technologies in classical VOD itself is not new. However, as cable companies look to deploy this technology, one has to consider the trade-offs between a complete re-architecture of the network to build a best-of-breed HTTP CDN from scratch and building less-than-standard VOD-specific distribution networks with no leverage beyond classical VOD. In order to resolve such trade-offs, it is important to look at existing components that may be modified to be integrated into hierarchical cache-based architectures. Components such as the back-office and streaming servers will need minimal but critical modifications to support such an architecture.

The purpose of this paper is to discuss use-cases and architectures where the above technologies, namely HTTP and caching, can be utilized to create a highly scalable VOD platform. As part of the discussion, the paper will explore advantages that can be gained with very little modification to existing platforms. The proposed modifications may then be used as a practical recipe for integrating existing VOD platforms into a HTTP CDN.

In the rest of the paper, we first discuss the underlying technology drivers, namely HTTP and caching, and the VOD ecosystem components that need modifications to benefit from those drivers. The, we will discuss details and use-cases of how such enhanced VOD ecosystems interface with an HTTP-based CDN, as well as the enhancements themselves.

In the following section, we describe the VOD-specific design considerations, i.e., how to bring the well-known classical VOD application into the CDN model.

TECHNOLOGY DRIVERS: HTTP AND CACHING

A primary goal of a content delivery network is to scale content libraries by utilizing distributed caching as opposed to replicating entire libraries at each serving location. Figure 1 illustrates a two-tier hierarchical CDN for reference. The intent of such a CDN is to enable caching of the most popular objects in the edge servers, using optimum caching techniques, and to enable the transfer of content between the origin server and edge server, as popularity changes, utilizing well-chosen transfer techniques.

We posit that content delivery networks represent a strategic infrastructure investment for operators, a layer 7 interconnect for transfer of objects akin to a layer 3 interconnect for transfer of IP packets. Therefore, a careful choice needs to be made on transfer and caching techniques, a choice that lays the foundation for multiple content delivery applications. From that viewpoint, bringing CDN technologies to address classical cable VOD applications is more about bringing VOD ecosystems to a well-chosen CDN infrastructure than the other way around. In other words, it makes economical sense to avoid designing content delivery networks specifically for VOD.

Content Transfer

We propose a standard HTTP-based [5] content delivery network as the foundation for integrating VOD ecosystems, in accordance with principles of Representational State Transfer (REST) [6]. In other words, we narrow down on a standard usage of HTTP for our choice-of-content transfer technique, a choice that has been proven to scale the internet across various content delivery applications. Existing VOD ecosystems form the periphery of a core HTTP CDN. For example, in the figure above, edge servers are VOD streamers, repurposed as caches that employ HTTP to fetch content from multiple tiers of HTTP caches (only one tier, the origin server, is shown in the figure). Object ingest and object request commands are repurposed VOD back-office commands. The following summarizes some aspects of the choice-of-standard HTTP for content transfer:

- *Naming:* VOD objects (media files, metadata) are named using universal resource identifiers (URI)

- *Client intelligence:* The client of the CDN, i.e., the VOD streamer, retains all intelligence regarding when (e.g., cache miss, background fill) and how (e.g., entire files, blocks) to make HTTP requests

- *Media awareness*: All media awareness is encompassed at the peripheries of the CDN. For example, random-access, trick-modes and other rich-media operations are facilitated by the client and the origin. Manifest or index files may be used by clients to make HTTP requests in such a fashion that the core of the CDN remains media unaware.

- *Limited use of extended headers*: Extended HTTP headers should not be used as object modifiers, but may be used in a limited fashion to facilitate auxiliary tasks, such as authentication and bandwidth allocation.

- *Standard DNS/HTTP-based request routing:* Request routing, i.e., determining which specific node in the CDN responds to a VOD streamer request for objects, is a natural consequence of resolving a virtual host name.

REST, in essence, denotes an architectural style that imposes a set of constraints (on the usage of HTTP in this case) to induce desired architectural properties. The desired properties here include keeping the core CDN unencumbered by VOD media specifics to allow unrestricted scale and extensibility, and to maintain cacheability of generic named objects without the need for special application logic. Some of the key REST constraints most applicable to integrating VOD ecosystems include:

- *State:* VOD streamers, i.e., clients of the CDN, maintain session state, while server nodes within the CDN remain stateless. To enable such statelessness, all requests into the CDN are self-describing (using merely the URI of the desired object) and idempotent, i.e., the order of requests does not modify the identity of the returned objects.

- *Layering:* Components interact only with their immediate neighbors, thus virtualizing the rest of the network. For example, VOD streamers only interact with their immediate parents in the CDN. This kind of layering enables all kinds of clustering, load balancing, redirection, and hierarchies to be hidden from the client.

Caching Techniques

HTTP-based content delivery networks primarily employ distributed caching nodes that pull content on a cache miss, directly from origin servers or from intermediate caches en route to the origin server (multi-tiered cache architectures). This allows caching intelligence to be retained in the clients (the requesting entity) as opposed to centralized tracking systems that may have difficulty scaling as the same CDN is employed for multiple applications. The centralized origin servers provide a "single point of ingest" to push or place content into the CDN. The rest of the CDN then leverages pull-based distributed caching, with algorithms in each node determining the subset of the content library that finds itself in the cache at each given point in time.

A pull-based approach does not restrict the specific caching algorithm that a node may employ to optimize its cache-hit ratio. Some broad examples include:

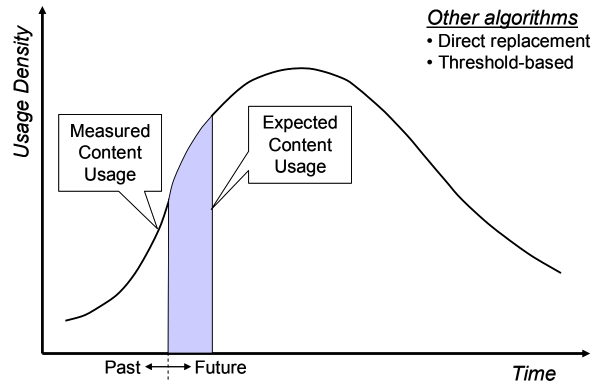


Figure 2: Trend analysis based caching

- *Direct replacement:* Every cache-miss leads to filling the cache with the requested object and the eviction of a well-chosen object already in the cache. Examples of eviction policies include the well-known least-recently-used (LRU), least-frequently-used (LFU), and adaptive replacement caching (ARC), which combines aspects of both. Since every request goes through the cache, this is purely a "stream-through" scheme.

- *Threshold-based replacement:* The *N*-th cache-miss for an object within a fixed interval of time leads to filling the cache with the requested object and the eviction of another. The first $N-1$ cache misses just result in a proxy ("stream-through") or redirect ("stream-around"), on each occasion, to a parent node.

- *Background replacement using trend analysis*: Based on local access patterns and trend analysis, as illustrated in Figure 2 (a content popularity curve for each object in the library), a cache may request content in the background. In this case, all cache misses result in a proxy or redirect to a parent node.

Irrespective of the caching algorithm used, the actual content transfer (and subsequent cache fill if applicable) may occur in one of several modes. A caching node may request content as continuous portions of files (including entire files), e.g., using byte ranges. Alternatively, it may request content on a segment-by-segment basis, where a segment is defined as any well-defined range of the file. In order to let the client exercise its intelligence so as to optimize its local goals, the CDN itself must not restrict either the caching algorithm or the mode of transfer. The analysis of such algorithms themselves is a rich area of research and is outside the scope of this paper.

We propose that while the content transfer methodology must adhere to industry-wide open standards, caching techniques, including the replacement algorithm and mode of transfer, must be left to specific node implementations so as to promote vendor innovation.

VOD Ecosystems and HTTP Caching

In order to utilize a best-of-breed CDN, in accordance with the guiding principles described above, legacy VOD ecosystems will typically require these key enhancements:

- *Back-office modifications:* Many existing back-office systems have typically presumed a replicated model of deployment, i.e., one back-office instance controls one or more replicated sites that contain the entire content library. Due to this assumption, the back-office carefully orchestrates the ingestion of content to specific servers and the subsequent request of content to the respective servers. This link between content ingestion and delivery, which made sense in classical client-server ecosystems, must be decoupled. For example, instead of explicitly ingesting content on each VOD streamer, the back-office may now provide the URI for the ingested asset. A second set of enhancements may be related to the centralization of the back-office functions, which can allow a complete virtualization of the CDN from the point of view of the back-office.

- *VOD streamers as caches:* Legacy VOD streamers have typically been based on the same presumption as above, i.e., content is explicitly pushed into such streamers prior to delivery. VOD streamers which now form the edge ecosystem to the core HTTP-based CDN must be enhanced to include the content transfer and caching techniques described above.

- *Media-related operations:* Operations such as random-access (based on time or chapter numbers) and trick-mode s must now be supported in the context of a CDN. Typically, this is accomplished by generating the necessary meta-data, e.g., manifest files or index files, which can be used by the entire base of VOD streamers. This in essence may require additional elements to generate such meta-data, and enhancements to VOD streamers in order to use it.

Now, we turn to a discussion of how such enhanced VOD ecosystems interface with an HTTP-based CDN and the enhancements themselves.

CDN Standards

As mentioned above, we strongly espouse the usage of industry-wide open standards for content transfer. Standardization in this area will help with interoperability and an ability to leverage already deployed systems. While many of the foundational items of such a CDN, namely protocols such as HTTP [5] and DNS [7], have long been standardized, the industry has been lacking in the wider adoption of VOD CDN standards.

A promising new standard specification by the IPTV Interoperability Forum (IIF) may address that gap. The IIF Content on Demand specification [8] defines several reference points between components of an HTTP-based video on-demand CDN. For our purposes, the C2 reference point in the specification provides a template for the content transfer interface between edge streamers and an HTTP CDN. Also relevant is the C2 index file specification that aids streamers to perform media-related operations, such as trick mode. While specification of interfaces is out of the scope of this paper, we note here that C2 is an HTTP interface, including URI and header conventions, to pull content from a network. In line with our goals, the specification allows streamers to exercise any caching algorithm or mode of their choosing.

VOD CDN: DESIGN CONSIDERATIONS

The goal of integrating CDN technologies into classical VOD platforms is to leverage the capabilities of a best-of-breed HTTP-based CDN architecture for content library expansion, while at the same time, maintaining as much of the current VOD infrastructure as possible. In addition to maintaining infrastructure components, such as streamers and the VOD back-office, media-related functions commonly offered in classical VOD, such as trick mode, must be maintained. Similarly, critical back-office functions such as session management, catalogs and billing support must also be maintained.

Figure 3 illustrates how existing VOD platforms may be migrated to the new architecture. HereVOD streamers accustomed to explicit asset ingest (at each head-end) are repurposed as caches that use HTTP (e.g., IIF C2 reference point) to pull content on-demand from the CDN. The back-office ingest and delivery commands are essentially decoupled by centralizing asset ingest. As part of asset ingest, a centralized asset preparation server generates the necessary metadata and media files required to support media-related operations, such as trick mode. Both media and metadata are ingested into a centralized origin server. Instead of explicit asset ingest into the streamer, the streamer is merely notified of the URI of the ingested asset to be used on a cache miss.

Streamers as Caches

In traditional VOD deployment scenarios, storage and streaming have been tightly integrated at each head-end, as shown at the top of Figure 3. With the advent of offerings such as network DVR and Start-Over, highly scalable ingest mechanisms were added to these systems to allow for many linear channels to be ingested into the VOD streamers. Each such ecosystem then grows atomically and separately from other VOD ecosystems in the network. As such, with growing storage and ingest capacity requirements, replicated expansion becomes expensive and cumbersome. Consequently, the best use of the current VOD ecosystems is to tie them into a larger CDN network, allowing for expansion of content, as shown at the bottom of Figure 3.

Wherever possible, the best option for an operator is to upgrade (e.g., via software update) the standalone VOD clusters so as to enable them to pull files (or fragments of files) from a hierarchical CDN network using HTTP, e.g., using a protocol such as IIF C2. The existing disk subsystems of the standalone cluster would now function as an efficient caching element at the edge. In some

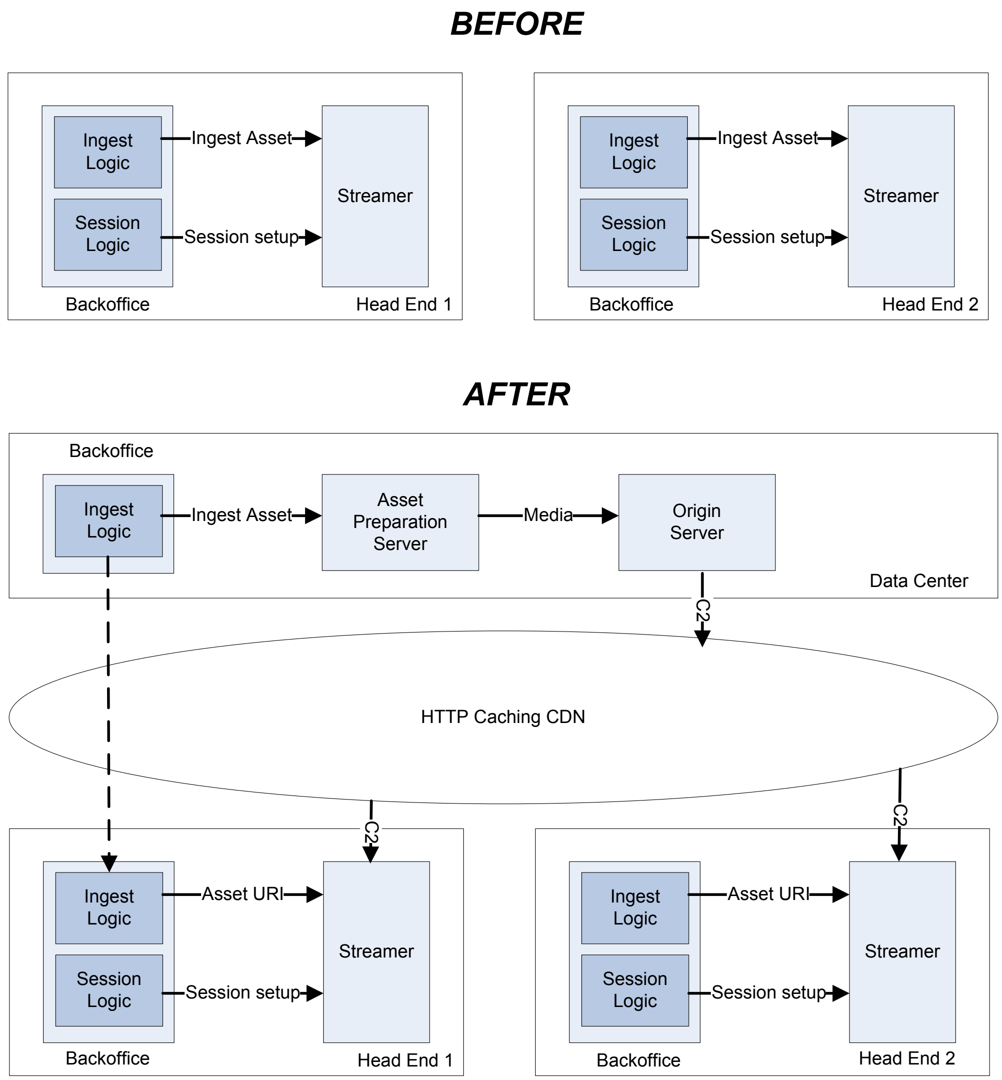


Figure 3: Classical VOD ecosystem modified for CDN integration

Cases, depending upon the design of the existing system, storage previously located at the edge may be able to be repurposed towards a central library.

Careful consideration is needed to ensure that the repurposed VOD streamer can handle ingest requirements for an optimized edge cache. For example, if the content library yields an 80-20 popularity curve (80% of requests are of the top 20% of the content library), the streamers must be able to handle an ingest bandwidth of 20% during peak times. Also, the system must be able to support trick mode operations either through real-time processing or via support of common trick mode streams (e.g., as specified by the IIF C2 index file specification). As a cache, the repurposed streamers need to be optimized simultaneously for streaming and ingest performance.

As opposed to proactively ingesting content based upon an ingest command, the VOD streamers must support the ability to provision HTTP-locatable content (URI) to be used to pull content on demand. The VOD streamer would need to be able to make requests to a DNS or HTTP-based request router (e.g., using 302 redirect) to determine the server location within the CDN responsible for satisfying each on-demand content request.

Well-chosen caching and content life cycle management techniques must be used by the VOD streamers to manage the content (files and file segments) populated as part of the content playback requests. As we have described in the previous section, a number of different caching methods may be used, such as direct replacement with LRU, LFU or ARC eviction policies, to maximize the cache hit ratio. As part of the chosen caching technique, stream-around or stream-through methods may be considered.

The edge VOD complex acts as a termination system for such protocols as Session Setup Protocol (SSP) [9] and Lightweight Stream Control Protocol (LSCP) [9]. By allowing the VOD complex to manage these protocols required for classical VOD, the CDN does not need any awareness of customers, session or state. All these components and states are managed by the edge streamer, thus requiring no changes to the set-top box (STB) client or the rest of the VOD ecosystem. The CDN network is also protected and firewalled from the STB using such an edge VOD system.

<u>Back-office Modifications</u>

A number of back–office systems exist in the cable community today, many of which are built around the popular Interactive Services Architecture (ISA) [9] or Next-Generation On-Demand (NGOD) protocol suites. With a few modifications to such back-office systems, a standalone VOD ecosystem can be integrated into a CDN.

One initial area that needs to be addressed is the ability to provision "CDN" content on the back office and VOD streamers. In the traditional architecture, content is provisioned on streamers via the back office (using an asset ingest command), which immediately loads the content onto the streamers for playback. When using a CDN, the back-office system must instead use indicators (URI) to the VOD streamers to provision the availability and authoritative location of content (e.g., using a modified asset ingest command). The content itself must not be proactively loaded.

One method for achieving this combination of static provisioning and subsequent dynamic retrieval is using an HTTP flag in the asset ingest provisioning process. The presence of the HTTP flag conveys the provisioning of a URI to the VOD streamer for future HTTP-based retrieval from the CDN. In addition to a modified provisioning command to the streamer, the asset management component of the back office is essentially centralized. Assets are provisioned into a central HTTP-based origin server, using either the existing (e.g., ISA ingest) or a new asset ingest command. As part of such provisioning, a central asset preparation server may generate index and trick files, which are also placed into the central origin server (as shown in Figure 3).

The back-office system must also be able to manage larger catalogs, even though all content does not persist at the edge VOD complexes. A related area of note is the required presentation mechanisms for large content catalogs, including a robust navigation system supporting the larger catalog. Techniques such as web-based navigation, play-listing and reservation lists, and tablet/smart-phone navigation apps could all be utilized to allow for additional navigation ease.

The modified back office must also provide the necessary content lifecycle management to ensure that files (or file segments) are properly accounted for and removed as needed (e.g., using CDN-wide purging of content, if necessary). The back-office must continue to honor updates to license and offering windows.

Beyond catalog management, ingest provisioning and content lifecycle management, the back office does not need to have any awareness of where the actual content files reside. As such, this allows the operator to minimize the amount of changes necessary in the back-office system for CDN integration, thereby simplifying operation and design.

Trick Mode and Media-related Operations

One of the popular features of a classical VOD platform is the ability to fast forward, pause and rewind content, much as with an in-home DVD player. When content is pro-actively ingested into a VOD system, as in the traditional architecture, there are several simple and well-known processes for locally creating trick mode files and/or indexes in support of such features. With CDN integration, the edge streamer may not be able to proactively create trick and index files, since the content is typically not pre-analyzed. This may in some cases lead to a failure to support trick mode features, especially on a cache miss, e.g., a request for a 32x fast forward as content is being streamed through the edge VOD streamer at normal speed. Therefore, it is critical to explore new methodologies to supporting trick modes in a CDN environment.

One method to support trick modes is to provide an index file (constructed by the asset preparation server in Figure 3) that outlines the structure of the content, such as the location of I, P and B frames for MPEG-2 content. Such a file can be small enough to be transferred during the initial phase of the content transfer, thus giving the edge streamer a "hint file" to assemble trick mode streams. The streamer then could pull the appropriate frames, as needed, to build out the trick mode stream on the fly.

Another method relies on pre-generated trick mode files, created centrally (e.g., by the asset preparation server in Figure 3) during the asset ingest process. The edge streamer would then pull the appropriate pre-generated trick mode file, based on the selected trick speed. A companion index file is typically used to correlate files of different speeds. The edge streamer may then cache the trick mode file segments much like the normal-speed files. When using this approach, it is critical to support a standardized trick mode file format for interoperability.

Another potential method that eliminates the need for both index and trick mode files relies on retrieving content faster than real-time and using the existing local process. However, if the cache-miss ratios are expected to be non- trivial, coupled with a non-uniform arrival of trick mode requests during cache misses, this method becomes highly impractical because of a multiplicative effect on the required cache-miss bandwidth, as well as the unnecessary retrieval of portions of files that may never be viewed. Not to mention, the I/O subsystems of many VOD systems may have difficulty maintaining the high ingest rate while attempting to create a trick mode stream to the customer. Multiple transfers at this high ingest rate may cause most disk I/O sub-systems to perform poorly.

From a practical standpoint, an operator should provide as much flexibility as possible since different VOD streamers may employ different options for trick mode support in a CDN. As older servers are aged out and replaced, it could be an opportunity to harmonize methodologies and technologies.

CONCLUSIONS

The ability to support cloud and CDN technologies for classical VOD delivery platforms is available now and can be leveraged to allow for service growth. We have illustrated how operators can both build upon a foundation of best-of-breed CDN technology and, at the same time, retain existing infrastructure with a few key modifications. It is critical for operators and their partners to jointly develop open and publishable standards to allow for interoperability and for best-of-breed technologies to take hold. A modular approach to design allows the system to grow organically for expanded content offerings, new technology adoptions, and rapid deployment of new products to customers.

REFERENCES

[1] W. Mao, "Building Large VOD Libraries with Next Generation on Demand Architecture," NCTA Technical Papers, 2008

[2] W. Mao, "Key Architecture and Interface Options for IP Video Over Cable," SCTE Conference on Emerging Technologies, 2009

[3] S. Krishnan, W. Mao, "Open Content Delivery Networks for Managed Video Delivery to Multiple Screens," SCTE Cable-Tec Expo, 2010

[4] R. Pantos, ed., "HTTP Live Streaming," draft-pantos-http-live-streaming-01, Internet Draft (work in progress), Internet Engineering Task Force, Jun 2009

[5] R. Fielding et al., "HyperText Transfer Protocol—HTTP/1.1," Request for Comments (RFC) 2616, Internet Engineering Task Force, 1999

[6] R. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," Doctoral Dissertation, Chap. 5, UC Irvine, 2000

[7] P. Mockapetris, "Domain Names – Implementation and Specification," Request for Comments (RFC) 1035, Internet Engineering Task Force, 1989

[8] ATIS IPTV Interoperability Forum, "ATIS-0800042: IPTV Content on Demand Service," Dec 2010

[9] Interactive Services Architecture, www.interactiveservices.org, Time Warner Cable

Video Ingest for the Cloud: Common Video Subscription Support for Advancement of Multi-Video Service Architectures

By Yasser Syed, Ph.D.
Dan Holden
Comcast

ABSTRACT

This paper discusses how a video model can be developed and evolve from preparing content for a single specific service (e.g. QAM) to an architecture where multiple types of video services (e.g. SDV, Adaptive Streaming, Mobile) can subscribe to a common set of video feeds. Moving towards this architecture can increase scaling efficiencies in transcoding and monitoring, improve content quality consistency across platforms and allow for innovation in services/platform/devices. Legacy services and new multi-streaming technologies will be discussed. Additionally, this paper will cover common encoding approaches for acquisition, transcoding and service wrapping. These types of architectures will be needed to transport content efficiently for existing and new delivery infrastructures.

INTRODUCTION

Video services have always determined how content sources have been acquired. Until recently, cable video services have been a single linear service for delivery to the television. Initially, it was an analog transmission in 6Mhz channels, then digital information (also over an analog transmission), but intended for delivery to a television display. The video was encoded for linear delivery and the costs were burdened on the service.

Since then services have changed into non-linear VOD, switched digital video (SDV) and Internet streaming, but still the burden of the costs for video encoding has been taken on by the individual services.

With cloud-based delivery networks, there is one ingest point for content, but there are many services and platforms being addressed: VOD, linear broadcast, PCs, tablets, TVs, smartphones and personal devices. With the traditional connection between services and content acquisition so strong, we have to examine if there is a better model that could save costs and allow us to design and initiate services in a faster manner.

A NEW APPROACH

In this section, we examine the benefits of designing content acquisition systems to beg more loosely tied to the service. With recent advancements in networking and technology, convergence in content acquisition is possible. Though there are more services today, the end customer experience is nearly the same.

There is a convergence of content in many of the services being planned or offered. It’s either the same linear feeds or the same VOD content. What determines which content it is delivered to the service is dependent on physical availability and license determination. Physical availability is how much more effort in design and costs it takes to acquire the content in the service. In addition, the new services are negotiating licensing fees for sources and files. It is easier to negotiate for content that already exists and is provided for customer consumption rather than create new content for the service. There is little new content for any specific service due to the cost to exclusively produce and promote the content. From this, there is a convergence of type content demanded by video services.

Even with this, the licensing negotiation process can determine how, and in what format, source linear feeds and content files are acquired. This is determined by an acceptable content acquisition expense and expected ROI for that service. The negotiated costs can often dictate the final quality of these feeds and files.

The new approach separates the content acquisition expense from the expected ROI of any particular service by creating a distribution network for the content to which services can subscribe and deliver. The quality of the content sources will not be dependent on the negotiated licensing fees, but on existing distribution output delivery channels. It is expected that the content sources will be the same for most services and that standardizing content acquisition and processing for all services will mitigate and distribute the costs across services. It creates a content distribution channel system that can be matched to multiple services including a CDN delivery architecture. Also, quality and customer experience can be consistently maintained across all services.

There have been a number of recent technology improvements and convergences that have made designing standardized content acquisition and processing practical. One is network multicast distribution which allows multiple processing devices to pull from the same source. An early contributing factor was the development of an edge QAM. This also simplifies the physical planning where an in-place GIG-E network does not mean adding an entire separate physical infrastructure for each new candidate service. Additionally, there has been a convergence in the video technology area where standardization in the industry has reduced the number of supported video service codec formats needed. This is also happening at the transport level where standards and specifications have been developed that can support these popular video formats (e.g. SCTE 128, 14496-15).

There are also convergences happening on the consumer device side. There are far fewer video codecs that need to be supported in these devices. Resolutions, bit rates and frames rates are approaching fuller video rates. And delivery bandwidth is approaching rates that are acceptable to carrying a full screen video. Applying these same advantages and convergences with a common content acquisition and distribution structure will result in a reduction of costs for getting the same content (or subset) to services sourcing from these devices. It will also allow newer devices to anticipate and design for these types of signals. Lastly this architecture will be able to maintain content for an expected quality and customer experience for all services delivering to these devices and separate these issues from delivery factors

Legacy Platforms

Applying these principles can have multi-generational advantages with legacy services receiving immediate benefits. We define legacy services as MPEG-2 distribution structures for Broadcast, SDV and VOD for High Definition (HD) and Standard Definition (SD) sources.

Multicast distribution allows the same source encodes to be used for SPTS and MPTS solutions since the core elementary stream is MPEG-2. Development of a distributed multiplexer allows MPTS streams to be created from a set of SPTS sources. The distributed mux can statistically multiplex media services together in an efficient way either at a cable headend or in a national feed. The MPTS combined streams feeds broadcast to MPEG-2 STBs, while the separate SPTS feeds service SDV and HITS where the channel line-up can be determined at the headend side while the SPTS are distributed by fiber or satellite.

Other generational improvements are to create SD streams from the same corresponding HD sources. This is accomplished through processes such as center-cuts/ pan and scan, and adding in progressive to interlace conversions. With improvements to encoders, more than one output stream can be developed from the same input source stream.

These and other first generational improvements benefit the existing distribution infrastructure while enabling a common content acquisition and distribution infrastructure.

Multi- Streaming Platforms

Two developments occurring now aid in creating the next stage in this approach. One is the advent of adaptive streaming technology and the second is the plethora of personal video devices emerging that can operate in managed and unmanaged networks.

With the emergence of Over The Top (OTT), a video service delivered over the Internet to the PC, TV, handheld pads and mobile devices, content providers like You Tube, Hulu, Netflix, Blockbuster, Vudu, Boxee, Sony, Apple, Samsung, and a long list of others, are providing media content options to consumers in a session based, streaming delivery format. The next step is delivering live content to these same devices via the same adaptive streaming technology in a session-based delivery format. This session-based delivery format is an opportunity to compete with the OTT providers on these stand alone devices, and it opens up more programming opportunities..

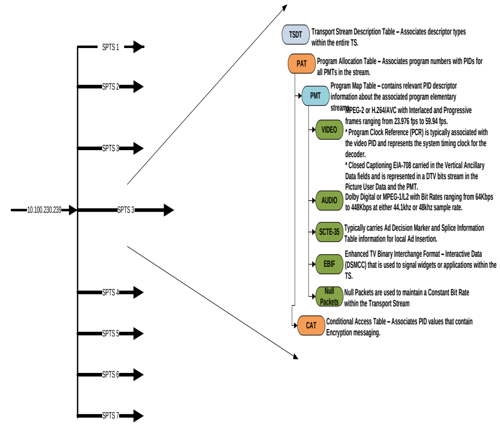
Adaptive streaming develops the idea that a single encoding process can create multiple aligned transcoded outputs. Before this, most encoding processes resulted in a single output, but now alignment between streams needs to be coordinated enough to switch between bit rates through the concept of a fragmentation structure (a grouping concept separate from a GOP (Group of Pictures) structure but can be used to create a session-based delivery format). To achieve this, a unified distribution network for specific content should be in place. Additionally, encoding platforms must be developed to handle multiple output streams that can extend beyond a single hardware box/ software set of processors.

Developing the content sources for this system involve working with the same content providers along with same limitations of present day content source delivery infrastructures (satellite, network) as before. The separate bandwidth generated by this process is quite large. It is not feasible to increase the incoming plant bandwidth just to provide the same content, but for an adaptive streaming service. It is more reasonable and achievable to transcode content that was already there for broadcast purposes. With adaptive streaming this became an early case of using a common mezzanine/contribution format that could be transcoded into a second distribution format.

There are also other advantages of using a common source format. For instance along with the broadcast content source, there is SCTE 35 information that the source stream contains, This can be reutilized by the adaptive streaming service to create its own ad-insertion features in the service.. If a separate content source were used instead, it would be difficult to justify the extra logistics and costs for putting this ad triggering signal in place. In this case, this important revenue feature is added to the service at minimal cost..

One consideration for designing the adaptive streaming service was the number of streams, bit rates and resolutions. In terms of transcoder performance, adding lower bit rate streams and resolutions is not a tremendous burden. There is commonalities between personal devices and television in terms of bit rates and resolutions; many devices can use a single stream or subset of streams for there own needs without having to use the entire adaptive streaming set. The resolutions, bit rates , video format (H.264/ AVC or MPEG-2) and the elementary streams structures are compatible, yet with changes above the ES (elementary stream) layer can make this suitable for different services. For adaptive streaming, this happens at the point of fragment creation where the layer above the ES layer is stripped out and replaced. Separating the transcoder processes from the fragmentation processes (service distribution wrapping) allows for the output of the transcoders to be reused including devices outside of the adaptive streaming or traditional broadcast technology. The transcoded streams could be readapted to use in managed or unmanaged networks through the use of lower-cost reformatting devices.

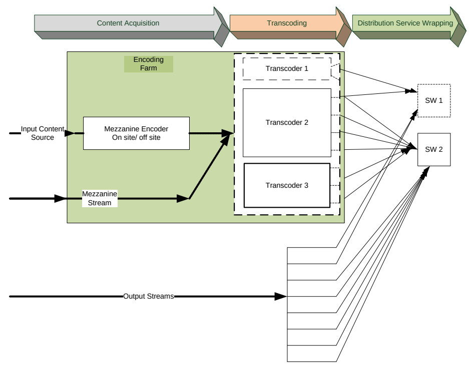
With the right selection of stream resolutions, bit rates, and codecs in a standardized multicast stream format, a number of services could ingest commonly demanded content by subscribing to these specific streams and reformatting them for their own purposes.



Universal Encoding Platform

Some of the basic groundwork for the Universal Encoding Platform has been laid with the development of multi-stream encoding platform. The idea of mezzanine/contribution sources, multiple transcoded H.264/AVC output streams and repurposing of elementary streams are established and proven that it can co-exist with an MPEG-2 distribution system using the same content sources in a multicast environment.

Expanding on this groundwork is facilitated by a set of convergence points occurring in acquisition, transcoding and distribution spaces. In each case where this does not happen for specific content channels, the content processes need to be individualized with extra costs in tools and formatting.



Acquisition

The acquisition convergence point standardizes the ingestion of content and affects the overall quality of the content as it is processed downstream. It can be the place to standardize feature for common customer experience across all services for that content channel. It also provides a natural demarcation point to distinguish between content source issues from downstream factors. Below are expected convergences:

1. Creating common specifications for Mezzanine/Contribution feeds

- Encoders can be placed at MSO or content provider sites
- Direct inputs into a transcoder system that can multicast into multiple devices.
- Formats are limited to standardized inputs such as AVC/H.264, MPEG-2 at bit rates up to 50-60 Mbps and common mezzanine audio format.
- Video would support resolution formats up to 720P or 1080P with the top tier HD resolution supporting frame rates at 60 (59.94), 24 (23.98), and 30 (29.97).

2. Set of Video & Audio pre-processing and filtering operations

- Convert all sources to native frame rates/picture formats to allow downstream service transcoders to adjust frame rates where needed.
- Normalize audio levels before passing down to service transcoders.

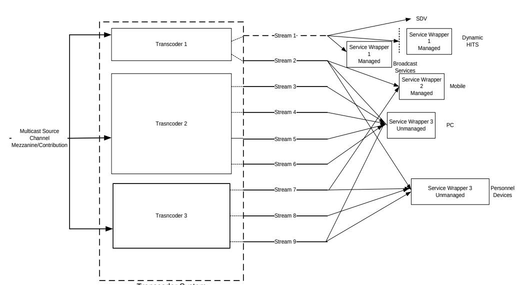
3. Video and Audio encoding processes to optimize quality rather than bit rate

- Video encoders output near-constant quality feeds rather than constant bit rate.
- Intelligent encoding that can signal less expensive downstream transcoders how to process/recover from errors.
- Can also be a single source point to correct content issues that can be sourced to all services.

4. Insert or pass-through auxiliary data information or trigger points

- Also a single source point to update data information to all downstream services.
- Downstream transcoders or service distribution wrappers only pass through information needed for that service.

Transcoders



The transcoding system creates the set of output elementary streams that the service wrappers subscribe to for distribution. The trick here is to establish a stable set of elementary streams based on resolution, frame rate, temporal compression structure and bit rate that are acceptable for multiple services in both a managed (QOS) and unmanaged (OTT) service environments. Some of the lower bit rate streams can also be used for auxiliary applications such as guides.

The adaptive streaming technology for live linear services is a driver because it establishes the concept of aligned multiple transcoded output streams that feed into a service wrapper (a multi-channel dense fragmentation/ encapsulation/ DRM process).
This transcoding point is also useful to retain higher output video quality by tailoring video filtering according to output bit rate and resolution. For example, reducing film grain noise through MCTF filters at lower bit rates allows the transcoder to spend fewer bits on processing noise (which would be lost anyways) and more on the actual video quality. Additionally, perceptual-based video filtering could be completed at lower video rates to help maintain video quality..

An important convergence change here is to treat this as a transcoder system, not a device, outputting a set of multiple bit streams, but can be created using more than one device. This will allow a subset of bit streams to be created by a set of cost effective optimized transcoders rather than limited by one device. It also breaks apart the performance limitation of restricting the number of output streams based upon a single device platform, which can then further reduce the costs of the encoding system through competitive activity. Additionally costs over time can be reduced by putting more functionality in the mezzanine/contribution feeds such that transcoders can operate more simply through cues in the source streams.

Lastly, the transcoding point provides an opportunity to standardize the carriage of these output streams such that service wrappers can subscribe and ingest streams and data in a common manner. A multicast distribution using an MPEG-2 transport layer is a suitable common interface for service wrappers.

<u>Service Wrappers</u>

Service wrappers adapt the elementary stream to the intended service and can be a point to add DRM/conditional access or XML messaging specific to that service. More than one service wrapper can subscribe to an elementary stream or set of elementary streams.

The advantages are each service does not singularly bear the costs for transcoding the stream. The devices designed for each service have lower costs per a stream because they are mostly wrapping the elementary stream (with some optional light transcoding functions) and more likely to be designed as a dense stream product. The reuse of the ES streams to multiple services is possible by focusing end-devices to support standardized video codecs like H.264/AVC. With the ability to output multiple aligned output transcoded streams, a small subset of streams can source multiple services.

Some examples of service wrapper functionalities are:

1. Groom multiple SPTS to set of MPTS streams
2. Convert a VBR stream to a CBR stream
3. Convert an SCTE 35 trigger to a suitable service equivalent

4. Match specific audio/data streams(s) to the service ES stream
5. Fragment the incoming conditioned ES stream
6. Apply DRM to the stream
7. Add service specific XML data

Additionally, the service wrapper developers do not have to be experts in video encoding, nor will companies that have video expertise have to learn a new set of domain skills for the new service.

IMPLEMENTATION

This architecture solution standardizes the content acquisition and transcoding processes. The output of the system is a set of multicast streams to which service wrappers can subscribe.

Transitioning to this concept is an expansion of existing infrastructure, operations and monitoring that already takes place at some MSOs for operationalizing video systems. Content sources are already ingested for today's broadcast services. Expanding this does not mean increasing the content selection, it means standardizing mezzanine contribution specifications to support the quality and data needs for multiple services. It also means developing transformaters to go from MPEG-2 to AVC/H.264. This will allow content providers to transition over instead of immediately cutting over. It will also allow for the transcoders to convert to and from AVC/H.264 and MPEG-2. Operations and monitoring are already taking place in a centralized system but need to be expanded to create a demarcation point between content source issues, transcoding issues and service issues.

Operation and monitoring should be viewed from three perspectives on both a content and service level:

1) Alerting to loss of content/service or content/service degradation;
2) Troubleshooting loss of content/service or content/service degradation; and,
3) Data and metrics gathering related to quality of content/service.

These three perspectives share common system touch points that provide existing content/service channel condition data. With the separation of content and services, the service operations and monitoring simplifies to handle the wrapping and adaptation modifications to fit the ES stream to the service. Some operations and monitoring features covered are:

1. Alerting to loss or degradation
   - System will monitor content loss and degradation for content to all services with each service only needing to monitor modifications
   - Existing MPEG tools can be reutilized to monitor loss and degradation in the system for content streams
2. Troubleshooting loss or degradation
   - Devices within the system will be capable of at least rudimentary measurements of media service conditions.
3. Metrics Gathering of Media Service Quality
   - Devices are able to calculate a Level of Service (LOS) that can be sent to collection points
4. Data Collection Points
   - All data (a normal state, an off-normal state, degradation, failure and quality level) will be gathered at a central point for processing and archiving
   - The Data Collection Point (DCP) is a system that will provide a user interface to system Operation Engineers (OE)

The benefits of common implementation is the costs of operations and monitoring will be centralized across services with each service only responsible for operations and monitoring for modifications completed at the service wrapper point and beyond.

TRANSITIONAL STRATEGIES

The number of output resolutions and rates supported by this video model should support playable video on devices intended for service. The initial set of streams should support devices for broadcast, VOD, SVC, adaptive streaming, mobile and personal devices for managed and unmanaged networks.

Future services need to justify additional stream formats based on ROI, or need to readapt the existing streams. Adding an additional stream would imply adding far more than the targeted service, which should available to all services and across all content selections.

In regard to existing streams, lower bit rates and resolutions will be more fluid over time since it is believed the device lifetime in these areas is much shorter than hardware desktop devices. Changes in the lower bit rate streams may lean toward increasing frame rates to make the set of output streams more homogenous and available across the platform. Changes in the number of streams may increase to accommodate new service needs or consolidate due to end of legacy services or incorporating new scalable video technologies.

SUMMARY

With new delivery and platform structures, creating a common set of video feeds for content that can span across multiple services has many benefits including a reduction in cost to develop and monitor new services. Without this type of architecture it will be harder to bring up new types of video services with a common set of service features because of the high ROI cost to support\common customer experience across services. With a new set of video streams for content, it will be easier to design and initiate new services that can already subscribe to one or more subsets of these streams

Evaluating Best-of-Class Web Service APIs for Today's Multi-platform Video Management Solutions

By Alan Ramaley, CTO, and Nick Rossi, VP Engineering
thePlatform for Media, Inc.

ABSTRACT

Video management and publishing platforms are evolving to meet the market's need for reaching consumers with reliable, high-capacity services – anytime, anywhere, on any device. As such, solution providers have to integrate their technology with a vast set of devices, systems, and environments—includingauthenticated syndication, third-party websites, mobile devices with vastly differing specs, set-topboxes, connectedTVs, smart over-the-top devices, andthird-party services, such as ad networks and content discovery engines.

Web service application programming interfaces (APIs) play an integral role in enabling content providers and distributors to succeed in a consumer driven market that's in constant flux. Developers at media companies and TV service providersneed flexibility and open APIs to adapt to changes in TV, online, and mobile video publishing.

This paper provides an in-depth evaluation of the most important features web service APIs should offer and explains why those features are important. It also examines the evolution of APIs and recommends best practices for a flexible, reliable and easily managed API set.

Several areasfor evaluation are examined and explained, all with an eye towards how APIs informed by service-oriented architecture (SOA) can be used to decoupleand safeguard business-critical services in a deployment and scale them independently.

Areas of focus will include:

- ***Breadth** – an API should expose all the functionality in the underlying service*
- ***Cohesion**– a given service should have a single area of responsibility*
- ***Security** – we will compare and contrast five common models*
- ***Web standards**– support for REST,Atom, RSS, and JSON for data services, and REST and SOAP for business services.*
- ***Data access** – APIs should provide very flexible read and write access to service data*
- ***Notifications** – with a comparison of push vs. pull notification models.*
- ***Extending the schema**– what to look for to make sure a service can support your custom data.*
- ***Scalability**– how to build scalability into an API at the core, to allow for a 99.99% read SLA*

Lastly, the paper focuses on some of the best developer support practices, including API clients and documentation.

INTRODUCTION

The recent introduction of Time Warner Cable's iPad application is just one example of the kind of services and applications that service providers and media companies can develop in-house by taking advantage of open web service APIs.

Going forward, web service APIs will continue to play a crucial role in enabling developers at content companies and TV service providers the flexibility to develop new services and respond to the changes in multi-platform video publishing.

When video management systems were in their infancy, few offered a set ofAPIs that anybody could use to build a media business.Most solution providers incorporated user interfaces on top of proprietary systems that could only expand when in-house developers felt like adding features. If anoutside user wanted to conduct their own development on top of such systems, they were out of luck.

The industry has since learned that web service APIs are a critical component for content providers and distributors, as it enables them to adapt to a fluid marketplace where consumer demand and IP-connected technologies are in constant flux. For this reason, APIs are now a standard part of every video management system. But despite the widespread adoption of APIs, not every system is equal. It begs the question: How good are a given system's APIs, and willthey continue to meet the needs of a media business as it grows?

This paper explores the most importantcapabilities to consider when evaluating the effectiveness of a system's APIs.

BREADTH

First, APIs should expose as much of the video management system's functionality as possible. It's very hard to predict what parts of your system you'll need to automate, based on where customer needs take your business. So,the more elements are available via the API, the more flexibility you have to respond to a changing marketplace.

Verification Process

A good ad-hocapproach to testing an API's breadth is to go to the management console or user interface and ascertain whether the technology vendor uses its own published API. If the vendor is not using it, not only is that a sign that they haven't built their system for maximum adaptability, but it also demonstrates that thevendordoesn't rely on its own APIs to support its product.

This can often be checkedby watching a network trace while using the system'smanagement console. If there areprivate protocols or undocumented payloads going back and forth, then it's likely the public APIs aren't complete enough or powerful enough for general usage.

COHESION

Each API endpoint should focus on a single area of responsibility within the system and use consistent operations and serialization methods for everyobjecttype. With one set of rules to interact with the system, developers can more easily integrate with it.

Multiple Services Versus a Single Monolithic Service

In a provider's API, if every call goes against a single "api.provider.com" or "services.provider.com" endpoint, with some kind of "command" or "service" parameter as a switchboard, that means the API provider has implemented a single monolithic endpoint that contains all APIs.

For example, you might see calls like this in a monolithic API:

- **http://api.provider.com/index.php?service=baseentry&action=list**
- **http://api.provider.com/index.php?service=multirequest&action=null**
- **http://api.provider.com/index.php?service=flavorparams&action=list**
- **http://api.provider.com/index.php?service=accesscontrol&action=list**
- **http://api.provider.com/index.php?service=partner&action=getInfo**

A monolithic API has several drawbacks:

First, it makes federated deployments very difficult,where some data is local to the content or service provider while other data is in the cloud. For example, youmight want to use an API cloud for most services, but store end-user transaction data locally for securitypurposes. A single, monolithic API endpoint does not have this capability.

Second, it puts a limiter on how fast the provider can extend the service. As the feature set grows, provider development will lag as internal teamsare encumbered with the increasing overhead ofcoordinating feature work and deployments in a single code base.

Finally, there's no single scalability strategy that works for all APIs: some get orders of magnitude more traffic than others, and the mix of read vs. write traffic varies, but the deployment of a single switchboard API is limited to an unhappy compromise between traffic capacity and cost.

One must be rigorous about dividing services into areas of responsibility to avoid these pitfalls. A good system will split its APIs into separate, focused services in which each API endpoint has a single job. This ensures that other services aren't affected if unexpected load hits one piece of the service, and the deploymentcan scale each endpoint as appropriate. For example, if there is an abundance of feed requests, administrators can simply add another feeds server instead of spinning up another instance of the entire API stack.

Data Services versus Business Services

There are two basic types of web services:

1. **Data services**, which handle stateful persistence of metadata.
2. **Business services**, which are stateless services with business logic that interactswith the data in data services.

The best approach is to look for a web-service framework that follows the principles of service-oriented architecture (SOA), which decouples data persistence from business logic so that services of each kindcan be deployed and optimized independently.

Base Objects

Optimally, every data service object has a base object with identically named properties for identifiers, modification history, and other common settings.

Properties such as ***title***, ***id***, ***guid***, ***added***, ***updated***, and ***locked***should beconsistent across all services. Consistently identified properties are beneficial, especially when querying for data objects, since the same kinds of queries can be used across all implementations.

If a framework implements these core properties, you can use similar queries across various services. For example, if "updated" is a base property, here's an example of a query you could use inanyservice to get items updated in the month of September 2010:

http://<service>/data/<objectType>?**byUpdated=2010-09-01T00:00:00Z~2010-10-01T00:00:00Z**

If "id" is common, the following query could be performed in order to get object IDs sorted by when they were added:

http://<service>/data/<objectType>?**fields=id&sort=added**

Finally, if "title" is common, in order to search for the first five items that have a title starting with "Test",one could execute the following query:

http://<service>/data/<objectType>?**byTitlePrefix=Test&range=1-5**

In a system that implements base objects and base queries, the only things that need to change in order to perform the queries in these examples are the host name and the object name. Because the pattern repeats across services, once you've learned how one service works, you've learned how all of them work.

SECURITY

APIs must be secure, and calls to APIs from end-user services (such as web form comments) must be completely separated from admin services (such as video publishing).

Admin Security

The level of admin security that is needed depends on what the user is trying to accomplish. Web service API authentication methods tend to fall into one of five models. See the table on the following page, which outlines each security type:

Security Type	When You'd Use It	Drawbacks
API freely available to any user	There are some cases where no security is desired. For example, read-only, highly constrained RSS feeds that are exposed to end users.	Not secure enough for admin authentication.
User name and clear-text password on every URL	Never	Credentials shouldn't get passed around on calls, as they can be intercepted and used indefinitely in ways that can't be controlled.
Vendor-provided API key	This is typically a random keybound to an organization with a particular set of rights. In many ways, this is similar to a user name and password on URLs, with the token acting as an obfuscated password.	If it's compromised, it can be used indefinitely to make additional calls until one discovers the breach and revokes the key.Revoking this key will typically disable legitimate uses as well, which will need to get updated with a new key.
Non-expiring API keyplus a signature on every URL	The signature is generated by hashing the URL parameters with a private/secret key. The hash is checked on the server before allowing the call. This works fairly well for server-to-server traffic or trusted clients: the URL can be used forever, but cannot be used to create different calls.	To do client-side end-user AJAX-style UI, one needs to push theprivate secret to the client to create this hash for each call, which makes the secreteasy to compromise.
Expiring token	This approach involves making a call to a secure API to generate an expiring token tied to a particular user's permissions, and then including that token on subsequent calls.	These can be captured via "man-in-the-middle" attacks, but this is mitigated by the expiration date.

While none of these options are impenetrable, the best approach is to use an expiring token. This solves the problem of tokens never expiring, so even if the user or system doesn't realize that a token is compromised, it can't be used after the expiration time specified when it's requested. And because the token is reusable across calls while it's still valid, there is no need to push signature secrets to the client.

API Types

There are three kinds of APIs that a video service should provide:

1. An admin read/write API that requires admin permissions to work with. For example, service providers would use an admin media API to publish or edit premiumvideos.

2. An audience read/write API that might also allow anonymous access. For example, viewers would use an audience API to add or edit comments or ratings.
3. A read-only, highly cached feed API for end-user guide data. For example, viewers would use a feedAPI through a video playerto retrieve lists of content or play videos.

All of these APIs must be separated. Audience users or viewers cannot be allowed to make admin API calls. For scalability, audience users must access different services altogether.

The primary concern is resource contention: if audience members are allowed into a service or content provider's admin APIs, even with an "audience-only" read-only token, they're competing with the provider's admin requests for API resources. This means that service providers run the risk of having a massive spike of audience-originating requests disrupt video publishing. Or, vice versa: the consumer experience could be degraded by admin API activity. Either situation can negatively impact your business with complaints of perceived outages.

Also, each type of API has a different usage pattern and should betuned separately. Admin APIs are used by a relatively small set of users, and get a relatively high volume of writes, while end-user APIs need to support a massive scale of read traffic with relatively few writes They all need to be configured differently to run optimally.

It's important to evaluate how a system implements a secure wall between these three kinds of services. One effective method is to have separate authentication services for administrators vs. audience members and configure a given API to run against the appropriate one. Another method is to physically separate the deployment of such services so that traffic on one cannot affect the others.

WEB STANDARDS

Avoiding any web service with a proprietary serialization format is preferred. When services support web standards, it's easier for developers to find clients and tools that can consume those services.

Data Services: REST withAtom, RSS, and JSON

It's imperative that any platformsupport a diverse set of Web standards, to increase the chance of interoperability with existing solutions. The following are specific examples from thePlatform's web services, but it should be straightforward to determine the pattern that any web service follows to deliver standards-based serializations:

Atom:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?form=atom

RSS:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?form=rss

JSON:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?form=json

JSON should alsosupport "callback" and "context" parameters for **JSONP**-style usage:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?form=json&callback=parseFeed&context=sample

Business Services: SOAP and RESTful XML /JSON

The service provider should support SOAP APIs for business services, including WSDL URLs for discovery of method signatures.

SOAP can be a heavy protocol to work with, so the service should also support some XML-RPC variant.For cases where you want the smallest serialization possible, the service should support JSON as well.

Finally, for cases where business service calls need to be made cross-domain in web browsers, the service should support a RESTful interface with JSONP responses.

REST Verbs (HTTP methods) and Just GET

Many REST implementations just support HTTP GET. Even when creating, updating, or deleting objects, one has to do a GET with parameters embedded in the URL.

However, following REST to the letter, this is incorrect. HTTP GET is supposed to be idempotent, so no matter how many times you call the same GET URL, it should make no change to the server state. Other verbs like POST, PUT, and DELETE are intended for state changes. That's why there is a prompt by browsers when refreshing a POST, but not when refreshing a GET.

If your code is making JSONP calls in a browser, the nature of cross-site scripting security requires that every request to a remote domain use a GET, and a good API should make an exception to idempotence in this case. But in less constrained clients, it's cleaner and more standards-based to use the available HTTP verbs with their typical interpretation: POST to create, PUT to update, and DELETE to delete.

For example, to delete a media with an ID of 1586532611, the following HTTP call would be made:

DELETE http://mps.theplatform.com/data/Media/1586532611?schema=1.2.0&token=...

One could also delete everything with a particular title prefix:

DELETE http://mps.theplatform.com/data/Media?byTitlePrefix=Old+Media&schema=1.2.0&token=...

If a given URL is too long for server gateways to handle, it's possible to convert this to a POST using the ***application/x-www-form-urlencoded***content type header, and put the parameters in the POST body, along with a ***method*** override:

POST
http://mps.theplatform.com/data/Media
Content-Type: application/x-www-form-urlencoded

method=delete&byTitlePrefix=Old+Media&schema=1.2.0&method=delete&token=...

Finally, if GET must be used for calls from a browser to avoid cross-domain issues, one can do a GET with a method parameter override to delete:

GET
http://mps.theplatform.com/data/Media/1586532611?method=delete&schema=1.2.0&token=...

DATA ACCESS

The ways in which solutions consume data are as varied as the data itself, so a service provider should offer flexible APIs for updating, querying, searching, sorting and paging lists of data, similar to what is possible withdatabase or search queries.

Combining Queries

Many web services don't let you combine queries: if you see method names like ***findByCategory*** or ***findByRating***, that means you'll never be able to do a query that searches for both category and rating.

Any given API should implement a set of base queries and then additional queries, and you should be able to use them in any combination. For example, our media API supports over 20 different queries. Here's a query for objects in the "Action" category:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCategories=Action

And here's a query for anything in ***Action*** OR ***Comedy***:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCategories=Action|Comedy

And here's a query for anything in ***Action*** AND ***Comedy***:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCategories=Action, Comedy

You can combine a category query with a content rating query:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCategories=Comedy&byRatings=G

And you can further combine these with a custom data query:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCategories=Comedy&byRatings=G&byCustomValue={year}{2010}

In a flexible API, there should be no limit to the number of queries you can combine.

API Queries Across Multiple Accounts

A larger organization often needsmultiple accounts in the web service for different groups. Optimally, these organizations will want the APIs to be able to fetch from multiple accounts at once, instead of having to switch users or tokens for each account.

Getting All Objects at Once

Often there is a scenario where the user wants to get every object in one call. For example, you might be synchronizing with a new content management system (CMS), and youwant to get every video in the account(s).

Most web services can't handle this, so they restrict the maximum "page" size a given API call can return, typically to 20 or 100 items. This puts the burden on client code to retrieve successive pages and deal with any errors that come up between them. It also means that the server keeps invoking larger and larger internal queries to skip over the results that were in previous pages, and each subsequent page will therefore be slower to fetch.

A more effective system avoids this by allowing unbounded result sets that stream out to the client. To get every item available in an API, you might make a call like this:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?range=1-*

Control Over the Sorting of Result Sets

Most web services allow control over API result sorting, but almost all put tight restrictions on what can be sorted and only allow one level of sorting. A more effective API allows sorting on any combination of fields. Here's an example of a sort by title:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?sort=title

Here's the same sort, but flipped to be in descending order:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?sort=title|desc

And here's a four-level sort, by locked, then approved (descending), then the ***year*** custom field, and finally publication date:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?sort=locked,approved|desc,:year,pubDate

NOTIFICATIONS

There are many cases where a client process needs to know when data changes in a service, such as when synchronizing datawith another content management system, invalidatinga customized caching layer, or keeping an audit trail of edits.

Notifications are the optimal way to keep things in sync. Suppose there is the need to synchronize a list of feeds between the provider's web service and the content management systemthat is being used for a provider's web site. When a feed changes, the provider's web service would create a notification that thesynchronization logiccould pick up. Thelogicwould then usethat notification to update thecontent management system. This wouldn't be a human-readable notification like an email, but a special payload designed to be consumed by an API client.

Support for Notifications

Some APIs don't support notifications, and their only mechanism for synchronizing with other systems is to use a polling approach against the actual objects. These APIs require you to check for changes every few minutes, typically with a ***modifiedSince*** query or something similar. If data has been modified, you have to figure out if the change matters to theclient.

This is an inferior approach when objects aren't constantly changing, becauseit forces you to make many pointless calls when there's no change just to make sure there is a timely update if something does. This polling taxes the client, the server, and the network.

The better approach is to supporttrue notifications that clients can register for. An API without notifications hasn't taken third-party integrations seriously.

Notifications for All Objects

Many web services are stingy with their notifications. They'll support notifications on some objects but not others, or they'll only send a notification when a video finishes processing, or they'll only notify on add and delete, but not update.

A more effective APIsupports a complete set of notifications. It should notify users on create, update, and delete for every object, and provides updates within seconds of the change being committed.

Notifications through a Firewall

A common but naïve approach to notifications isto let users register a "notification URL", and whenever something changes, the system sends an HTTP message to that URL.

There are limitations to this approach. If there are any hiccups at all between the notificationserver and the customer's server (e.g., network glitch, the customer's server is down for maintenance, exception in the customer's notification handler, etc.), notifications get lost. Most customers that use this approach have to run a monthly "resync everything" process to deal with these lost messages.

This method also requires the user to expose a public URL that anybody could hit. It could be locked down with passwords and IP exclusions, but it providesa vector for attack if hackers found out about it.

A better approach is to store all notifications on the server and serve them with a Comet-style "push" model. A typical exchange starts with a call like this from a client:

http://mps.theplatform.com/notify?token=...

This returns a payload with the ID of the most recent notification available. The programmer then uses this ID and makes another call to open up an HTTP request, which stays open until there arechanges to report:

http://mps.theplatform.com/notify?since=668017728&block=true&token=...

When an object changes, the response gets returned along with the latest notification ID, and the cycle continues. Because the client always initiates this exchange, it can function safely behind a firewall.

Notification Delivery

Another advantage of client-initiated, Comet-style notifications is that this approachcan guarantee that the client never misses a notification.

If clients go offline, for example, they can use their last remembered notification ID and request all notifications since that ID. For example, if the server storesnotificationsfor seven days (or whatever period of time is deemed acceptable), thenas long as the client does not go offline for longerthan that, there's no chance of any notification getting lost.

Notifications can also be delivered in the exact order they were committed to the data store, so they are always received in the proper order.

EXTENDING THE SCHEMA

Because a web service rarely has the exact schema needed for a solution, APIs must allow the providerto extend the object schema with custom fields and data types.

Custom Fields for All Objects

Many videoservices only support adding custom fields to the main media or video object. It's important to have the ability to add custom fields to as many objects as possible in the API. This ensures that in addition to adding custom fields to media, customers can also add them to players, feeds, categories, servers, etc.

Some technology vendors require you to contact their support organization to add custom fields. Others enable only a small number of custom fields to be added. Based on our experience with real-world schemas, a good upper limit is 100 custom fields per object type. It is important to allow users to administer custom fields themselves, so the vendor is never standing in the way of solution development.

Support for Custom Data Types

Many web services don't support the idea of typed custom fields. Instead, custom fields are always strings or string arrays. But typed custom fields are important in all layers of an application for the same reason they're important for native fields:

- The ability to do queries that take advantage of the data type (e.g., date range queries or numeric queries).
- Sorting that works correctly [e.g., the numbers 1, 2, and 11 would sort incorrectly as strings (1, 11, 2), but sort correctly as numbers].
- Data that is correctly serialized. This means that the user doesn't have to write any ***toString()*** and ***parseString()*** functions in their code.
- The ability to choose the right control types in the console UI (e.g., date types get a date picker).

For video services, it's important to look for an API that supports at least these custom data types:

- Boolean (true, false)
- Date (9/19/2010, 2/7/2011, etc.)
- DateTime (9/19/2010 3:27 PM, etc.)
- Decimal (1.01, 2.02, etc.)
- Duration (0:01.5, 1:34:23, etc.)
- Image (an object with an image URL and a hyperlink URL)
- Integer (-3, 5, 10000, etc.)
- Link (an object with a hyperlink URL and a title)
- String ("hello!", etc.)
- Time (9:15 AM, 3:27 PM, etc.)
- URI (any well-formed URI)

It's also important that an API support arrays and maps of anyof these types.

Serialization of Custom Data

Often, web services won't allow control over the namespace, namespace prefix, or tag name for custom data. Instead, they'll use a proprietary serialization.

But if you aredesigning a solution around web standards that require fields in a particular XML namespace, you need to be able to control all these aspects.

For example, suppose you want to add a "Latitude" custom field to media. If using Geo XML, then the custom field needs to be serialized as follows:

```
<feed
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
<item>
    <title>My Media</title>
    <geo:lat>51.51</geo:lat>
</item>
</feed>
```

Ultimately, custom data serialization should match the consuming client'sneeds: if you have an existing solution that's expecting custom data in a particular namespace, you don't have to change that solution.

Searching by Custom Data

It's also important that the API offers the ability to search by custom data. Otherwise, you would need to implementsolutions where the client pullsback more data than itneeds in order to filter the results, which is inefficient.

Here's how a solution could support this. For example, to see all movies with a "year" value of 2008, the query might look like this:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCustomValue={year}{2008}&fields=title,:year

To see all movies released between 2008 and 2010, one could perform a ranged query:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?byCustomValue={year}{2008~2010}&fields=title,:year

Ideally, you should be able to invoke range queries on any numeric or time-based custom field type, in addition to exact matches.

Sorting by Custom Data

Most web service solutions don't allow for sorting by custom data, but it's an important capability to have. The ability to sort on the server is usually faster than trying to sort on the client. It also allows the ability to page through multiple pages of results with a consistent sort order.

For example, to see all movies sorted by a "year" custom field"with a tiebreaker sort on the native "title" field, the query might look like this:

http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?sort=:year,title&fields=:year,title

SCALABILITY

APIs must be able to handle the provider's site traffic. They should be like a dial tone: always on. It's notoriously hard to support this with hosted services—especially multi-tenant services—and it's a rare organization that's been able to do it: that short list includes Amazon, Google, Salesforce, Yahoo, and a few others.

In order to scale hosted services, it's important that the system has a minimum of 99.99% guaranteed uptime in which the service is available for reads. If using a service with a 99.9% service level agreement (SLA) on reads, that means that the user might not be able to do reads for nearly forty-five minutes each month, and that's forty-five minutes a month during which thesite might be completely dark. A 99.99% read SLA means the service is guaranteed to have unscheduled read downtime of less than five minutes each month.

High Availability

Service reliability depends on a properly engineered deployment that includes redundancy, automatic failover, and 24/7 human response when services experience failures.

It's important to ask a service provider pointed questions about their deployment architecture and the systems they have in place to prevent and respond to outages. A provider should satisfy your questions with evidence of redundant infrastructure, API traffic management, quality engineering in the web services themselves, and a 24/7 support team.

For example, data service APIs should automatically failover to a read-only copy of the data when the primary data source experiences a failure. Data storage in general is unreliable enough that such failover systems are one of the many requirements for achieving 99.99% uptime.

Also, in multi-tenant systems, web services should be designed to prevent heavy traffic from any one tenant from interfering with others.

Response Cache

Most solutions involve repetition of a small set of read-only API calls. Thus, for performance,APIs should provide a response cache for read requests. An API call where the response comes from a cache will respond in a fraction of the time (often under 10 milliseconds) compared with a call that invokesqueries to a database or other remote service.

A good way to check if the API is using a response cache is to look for the ***X-Cache***header in HTTP GET responses. Here's an example from our console data service:

X-Cache: HIT from data.mpx.theplatform.com:80

This tells you whether or not the response came directly from the response cache; there should either be either***HIT*** or ***MISS***. Another indicator of a response cache is a ***Last-Modified*** and a ***Cache-Control*** header. For example, for GET calls, one should see headers like the following:

Last-Modified: Tue, 05 Oct 2010 22:54:57 GMT
Cache-Control: max-age=0

Also, watch what happens when you pass in an***If-Modified-Since*** header with the ***Last-Modified*** value from a previous call. If the service has a response cache, there should be a 304 response:

HTTP/1.1 304 Not Modified

If the API doesn't show any signs of having a response cache, it's not going to hold up under load.

UI Edits

Some vendors implement their console application separately from their API. The result is that when changes happen in their console, it can take some time—sometimes up to five minutes—for the changes to appear in the API response cache.

This doesn't occur when every part of the system is run off of the underlying API, including the console. If a change is made in the console, it will show up in the next admin API call. But if nothing changes, then the API will hold on to the cached response, and the console user willexperience fasterinteraction.

Support for Sparse Objects

A rich object definition will have many fields on it, but it's unlikely that youneed every field when you make a request. Minimizing the actual set of fields returned improves performance at the server (less to query and serialize), over the network (less to transmit), and on the client (less to parse).

It's important that APIs support sparse objects with a ***fields*** parameter or something similar. For example, the following would just return ***title*** and ***id*** fields.

> **http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?fields=title,id**

Asking for all fields in a particular namespace would return everything in the media RSS namespace:

> **http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?fields=media:**

If an API supports dependent objects—objects that are contained in other objects, like files inside of media—the system shouldalso support field lists for the dependent objects:

> http://feed.theplatform.com/f/ZlTfSB/wAeAAAKTtr_L?**fields=title,content.url,content.bitrate**

If these kinds of nested objects are supported, it means that youcan make fewer calls to get the data you need.

Sparse objects are important for create and update calls as well. An API should allow you to specify the fields to update in a call, rather than requirefull objects. Such sparse updates should be less expensive to invoke than a full update.

Multi-Item Create, Update and Delete

Many APIs only allow write operations on a single item at a time. For example, ifyou want to add 100 media, youneed to make 100 separate calls over the network. So, for every call, you are penalized for network transit time as well as the time to do an individual insertin the data store:

> **Total Time = (# of items) * ((network latency) + (time to add 1 item to data store))**

Some APIs try to work around this via "boxcarring," where you package multiple API calls into a single request. However, this methodonlyreduces the network latency because each call typically still gets evaluated separately on the server, with a separate data store add for each one:

> **Total Time = (network latency) + ((# of items) * (time to add 1 item to data store))**

A better solution is to allow for true multi-item create, update, and delete. With this method, youcan send multiple items in a single feed, and the web service updates them in the data storein a single atomic operation:

> **Total Time = (network latency) + (time to add N items to data store)**

Adding 20 items to a data store in a single operation is significantly faster than adding20 individual items separately.

DEVELOPER SUPPORT PRACTICES

APIs should be easy to develop against, and they should make iteasy for you to get help if youget stuck.

Browser-basedAPI Client

It can be tricky to build a REST URL for an API call. That's why companies like Flickr and Twitter have pages that offer help in accomplishing this.

It's best to look for a system where the data services have a built-in web client to help the user build your API calls. For example, in thePlatform's services, you can go to the root of anydata service and add ***/client***., This will return an HTML page that lets a programmer construct ad hoc REST calls and test functionality.

Supported API Clients

When you see notes on a technology vendor's web site like "API clients are not maintained or supported and are used at your own risk," or documentation that points you to community forums for support, you know that API clients have gotten short shrift. They've either been abandoned or crowd-sourced.Either way, if there are bugs in them, you'll need to depend on the community or fix them yourself.

Instead, you should look for a technology vendor whose API clients are maintained and officially supported. Optimally, these clients should get built as part of the core services, and not as an add-on by a different team.

For example, if you're a Java programmer using thePlatform's services, we provide JAR bundles with Java classes that implement calls to the service APIs, which can be downloaded from our Technical Resource Center. We also provide client DLLs for .NET.

Finally, if you're using another framework, you can use the web client to compose your particular REST calls and then add the URLs to your code.

API Documentation and Support

Everything in the system's APIs should be documented and available through an online technical resource center.

It's also important that any API have a team of support engineers who are trained in the API and capable of resolving even the most difficult problems.

CONCLUSION

Today, web service APIs are critical for enabling content companies and TV service providers to build systems that can toleratecontinuous change inIP-connected technologies and consumer behavior.

Despite the widespread adoption of APIs in video management systems, not every system is equal, and service providers must evaluate them carefully to ensure theyaccommodate the needs of developers working with a wide variety of technologies, partners, and types of content.This paper explores a baseline of characteristics that any robust API should provide, but only through a detailed evaluation of a system based on aprovider's unique requirements canyou fully determine the suitability of any technology.

WILL HTTP ADAPTIVE STREAMING BECOME THE DOMINANT MODE OF VIDEO DELIVERY IN CABLE NETWORKS?

Michael Adams
Ericsson Solution Area TV

Abstract

There is a great deal of interest in HTTP adaptive streaming because it can greatly improve the user experience for video delivery over unmanaged networks. Adaptive streaming works by adapting, in real-time, to the actual network throughput to a given endpoint, without the need for "re-buffering". So, if the network throughput suddenly drops, the picture may degrade but the end-user still sees a picture.

Although adaptive streaming was originally developed for "over-the-top" video, it brings significant advantages in managed networks applications. For example, operators could set session management polices to permit a predefined level of network over-subscription rather than blocking all new sessions. This flexibility will become more and more important as subscribers start to demand higher quality feeds (1080p and above) and 3D programming. Meanwhile, adaptive streaming increases transport overhead, requires multiple bit-rate encoding, additional buffering and synchronization, and two-way network connectivity.

Not very long ago, Internet Protocol (IP) was seen as a niche protocol best used for delivering datagrams over unreliable networks. Today, IP has become a ubiquitous transport protocol for every application over every possible physical layer. This transition happened rapidly despite the additional overhead and complexity of IP compared with protocols like SONET and ATM. Will the same become true for adaptive streaming protocols? Will they quickly dominate, as every new consumer electronics device ships with support for adaptive streaming? Will the ubiquitous nature of adaptive streaming trump any loss of efficiency that it brings?

This paper will describe what makes adaptive streaming different from other modes of video delivery, and how adaptive streaming works. It will discuss the pros and cons for adaptive streaming and analyze to what extent it will become the dominant mode of video delivery in cable networks.

INTRODUCTION

HTTP adaptive streaming is the generic term for various methods of adaptive bit-rate streaming over HTTP. These include:

- Adobe Dynamic Streaming for FLASH [1]
- Apple HTTP Live Streaming (HLS) [2]
- Microsoft Smooth Streaming for Silverlight [3]

Although each the above are different implementations of adaptive streaming, they have a set of common properties:

- Content is encoded at multiple bit-rates
- A point-to-point HTTP connection is used to deliver the content stream from a server to a client
- The bit-rate can be changed on the fly to adapt to changes in available network bandwidth
- The client is responsible for fetching multimedia data (‘client-pull’) from the network.

Standardization

MPEG Dynamic Adaptive Streaming over HTTP (DASH) is set to reach Final Draft

International Standard status in July, 2011 [4]. It is based on the 3GPP Adaptive HTTP Streaming specification.

Meanwhile HTTP Live Streaming has been submitted as a draft informational proposal to the IETF [1].

Nevertheless, it is unclear how long it will be before a common approach is widely adopted. The most likely prospect is that multiple adaptive streaming implementations will continue to co-exist and that cable operators will choose to support the most popular variants in order to support a broad range of consumer devices appearing in the marketplace. Apple appears to be an early winner in the tablet space where the iPad supports only HLS. Meanwhile Silverlight and Adobe FLASH are both being used for streaming to PCs.

A recent development at the time of writing (April 18, 2011) is that Adobe has announced support for Apple HLS in their latest Flash Media Server.

Advantages

Adaptive streaming brings a number of key advantages to the network operator when compared to its close cousin, progressive download. These include:

- The ability to change video bit-rate on the fly, allowing the client to select the best stream according to network throughput, which can be indirectly measured by monitoring the receive buffer.
- Only content that is actually watched traverses the network.
- Secure DRM based on content encryption rather than secure HTTP.
- A seamless mechanism for real-time ad insertion.
- Fast channel change implemented by selecting low bit-rate stream first.

Although these properties are important, this paper will focus on how adaptive streaming compares with MPEG-2 transport, which is the dominant mode today for delivering content to the cable set-top box.

HOW DOES ADAPTIVE STREAMING WORK?

The easiest way to understand how adaptive streaming works is to start with its close cousin, progressive download.

Progressive Download

Progressive download incorporates two functions that are coupled together:

1) Download - an HTTP session is established to transfer the content file from the server to the client device.

2) Playback - once the client estimates that the receive buffer is sufficiently full, it starts to play the file from the head. If the network bandwidth is constant, the playback will continue uninterrupted because playback of the file will always lag download of it.

A major flaw in progressive download is that if the playback rate exceeds the download rate eventually the buffer will be exhausted and playback will freeze. This 're-buffering' is extremely frustrating to the user.

Another flaw is that progressive download does not include any provision for flow control. In conventional (for example, MPEG-2 transport) delivery of video and audio packets, the rate of transmission is synchronized to the bit-rate of the payload. This is an important function that is necessary to prevent buffer underflow or overflow at the receiver. In contrast, progressive download treats the payload as just another file that should be downloaded as fast as the network permits. Thus progressive download may consume network resources in a burst of traffic until the file transfer is completed.

In practice, some progressive download servers implement a throttling mechanism to cap the maximum download rate to slightly more that the payload bit-rate to prevent this kind of behavior.

Adaptive Streaming

Adaptive streaming makes changes at the server and the client to increase the overall quality of experience of the end-user (viewer). These changes also directly impact the network characteristics of adaptive streaming. Finally, these changes pave the way for extensions to enable the delivery of live streams.

1) Multiple bit-rate encoding

To support adaptive streaming, the content must first be encoded at multiple bit-rates which must be pre-defined by the operator to provide an acceptable tradeoff between quality and bit-rate. In order to reduce the bit-rate, the content can be encoded at lower resolution and/or lower frame rates than the source.

2) Segmentation

Adaptive streaming sub-divides the encoded multimedia content into segments (or "chunks"). The segments are typically quite large containing between 2 and 10 seconds of multimedia content. Each segment can be delivered at a different bit-rate because it is aligned to the Group of Pictures (GOP) structure.

When combined with a set of encoding sessions of the same content, at different bit-rates and qualities, segmentation allows a client to switch from one stream to another seamlessly. Each stream is called a profile in streaming parlance.

In order for the client to be able to select segments from the appropriate profile for the stream, a manifest file is created. This is essentially a set of pointers, within a media file or list of media files, allowing the client to access the next segment at the desired bit-rate.

3) Adapting to Network Throughput

The adaptive part of adaptive streaming is enabled at the client rather than the server. The client continually monitors the available bandwidth and the media being delivered, and will dynamically switch to a higher or lower bit-rate session in order to keep the receive buffer within set limits. Seamless adaptive streaming means that the user sees no visible interruption in this process, because segments are aligned to closed GOP boundaries.

4) Flow control

Flow control is almost a misnomer as it is an indirect effect of segmentation and client behavior and not an explicit goal of adaptive streaming. However, the result is similar:

- The client is designed to download a sufficient number of segments in order to prevent buffer starvation in the event of network congestion. Clients may use different algorithms, but approximately 30 seconds of buffering is common.
- In the steady state, once the buffer is sufficiently full, the client will only request the next segment when a segment is played, essentially synchronizing the fetch rate to the play out rate.
- If network congestion occurs, the buffer will begin to empty, but the client will start to request segments encoded at a lower bit-rate to compensate.
- In the new steady state (if congestion persists), the client will once again synchronize to the (lower) play out rate.

It is important to note that all of the above assumes a given client behavior. Certain clients may be extremely conservative and attempt to maintain a much larger buffer (in fact, it would even be possible to modify a client to emulate a progressive download by continuing to fetch segments regardless of buffer fullness).

5) Live Content

The addition of segmentation means that adaptive streaming can be used to deliver live content in real-time. This is achieved as follows:

At the encoder:

- A set of real-time encoders is used to encode the source content at multiple bit-rates.
- As before, the output of each encoder is segmented according to the GOP structure of the video. Segments are buffered in memory (for a limited period of time).
- The manifest is updated in real-time, providing the client with an index of segments.

At the client:

- After the channel has been selected, the client will download the manifest file, and then starts fetching segments. The first segments are actually a few seconds behind the current time because a certain receive buffer size (typically in the order of 10 seconds) is needed to maintain a smooth play out.
- If network conditions are good, the client will rapidly fill its receive buffer but continue to play the content with significant delay.
- Once a steady state is achieved, the client will continue to fetch packets as soon as they are published by the server. In other words, it will frequently re-download and re-check the manifest file to see if new segments are available for download.
- In the case of network congestion, the client will fetch smaller segments (that is segments encoded at lower bit-rates) to ensure that the receive buffer doesn't underflow.
- In the case of severe network congestion, the client will have to pause play out and restart the process.

As can be understood from this description, "live" is a term that is loosely applied in this context. While it is true that there is a delay in the play out of MPEG-2 transport streams at the client, this is constant and tightly controlled by the specification. In contrast, the behavior in an adaptive streaming client is poorly specified and delays introduced are much larger.

APPLICATIONS IN CABLE NETWORKS

Adaptive streaming has different applications, each with different implications for cable operators. Beyond "over-the-top" (OTT), the most obvious fit is to supplement or replace on-demand services, and this leads to a fairly straightforward comparison with existing techniques. The second is live streaming for "broadcast" programming which has a much larger potential impact upon the network.

On-demand Services

On-demand services are implemented in cable systems to minimize the impact on the set-top box. In the first commercial deployments of on-demand (circa 1998), the most expensive part of the system was the set-top box, which was optimized for playing a standard MPEG-2 multiple program transport stream carrying MPEG-2 video and AC3 audio, the ATSC standard for digital broadcasts. Therefore on-demand systems were specified to emulate a broadcast stream as closely as possible. Some minor changes were made:

- A constant bit-rate format SPTS is specified. Initially at 3.75 Mbps for standard definition video, and later at 15 Mbps for high definition video.
- Initially, conditional access encryption was ignored and subsequently a fixed-key scheme was used (in contrast to broadcast streams where keys are updated continuously).

None of these changes made the set-top box more expensive. All the differences in operation were software changes related to the program guide and signaling.

Taking an existing on-demand system and extending it to provide the same service using adaptive streaming can be done by adding components to the existing system as follows:

- A content management system to manage encoding of assets into multiple bit-rate MPEG-4 AVC format.
- An offline encoding system that is aware of GOP boundaries and is able to create synchronized segments of video/audio payload.
- A server capable of servicing HTTP requests from the clients and delivering the multimedia payload in the chosen adaptive streaming format. The server must also publish a manifest that indexes each segment at each chosen bit-rate.

The result is that new clients that support adaptive streaming can now access the same on-demand library offered to the set-top box.

In practice, a Content Delivery Network (CDN) will be used to scale the system. This takes advantage of a property of HTTP that it can be cached. Thus a subsequent request for the same stream by a different client could be serviced by the CDN transparently.

Broadcast Services

Providing broadcast (that is "live") services to an adaptive streaming client is quite a different challenge for cable operators.

A broadcast channel must encoded in real-time into the adaptive streaming format since content is not available ahead of time.

Taking an existing broadcast system and extending it to provide the same service using adaptive streaming can be done by adding components as follows:

- A real-time encoding system that is aware of GOP boundaries and is able to create a synchronized encoded payload
- A system that segments the video/audio payload
- A server capable of servicing HTTP requests from the clients and delivering the multimedia payload in the chosen adaptive streaming format. The server must also update the manifest file in real-time.

In the live streaming case, a CDN that is optimized for streaming media is essential since the encoding system will be located in the core of the network.

SERVICE CHARACTERISTICS

With the introduction of adaptive streaming, the cable operator is moving into a new realm of operation with many more aspects of service delivery being now out of their control. For example, if we compare the buffer model of a set-top box, it is well defined and implemented according to MPEG-2 transport systems. Extensive testing and analysis of large-scale systems has been done to ensure unexpected side effects do not affect the delivery of video to the device. In contrast, the cable operator has little control over the algorithm implemented by a particular iPad or PC connected to their network. If one client implementation is not well behaved could it negatively affect performance of other well-behaved devices on the same network segment?

What are the potential impacts on the service, as perceived by the subscriber, as adaptive streaming is introduced into their viewing experience?

Delay

In most implementations of live streaming, the experience is still significantly more delayed than with current MPEG-2 transport delivered services. One study found that adaptive streaming introduced an 8 second delay [5]. This compares unfavorably with delays due to encoding/statistical multiplexing which are usually about 2 seconds at worst.

This means in practice that a subscriber may notice that the video on their iPad (for example) is significantly delayed compared to the video on their TV (while watching the same programming on both devices in the same room). In practice there is no technical solution to this as each client device may implement a different algorithm and attempt to maintain different playback buffer sizes, and therefore introduce differing playback delays.

Channel change

Adaptive streaming clients are typically designed to start streaming at the lowest acceptable bit-rate after a channel change and then rapidly increase the bit-rate selected according to network throughput. This provides a useful fast channel change mechanism.

Stability

If multiple adaptive streaming clients contend for limited network bandwidth, as one client reacts to congestion it has the effect of making more bandwidth available to the other clients. In certain circumstances, a feedback loop can be created leading to instability.

The effect of this is that the client may constantly switch between different bit-rates generating an annoying artifact, visible to the user as a repetitive cycle of softening and sharpening of the picture.

One solution is to avoid over-subscription, and therefore congestion, of the network. However, this means giving up a potential benefit of adaptive streaming, namely the ability to allow over-subscription during peak demand.

Customer support and Troubleshooting

Taking a scenario where the client fails to deliver a smooth, acceptable quality video stream, how can the trouble call be resolved? This represents one of the biggest challenges that will be faced by operators as adaptive streaming is widely deployed:

- Is the problem in the network or the client?
- If it is a network congestion issue, what kind of real-time trace can be performed to identify the source of congestion?

Ad Insertion

Client-side ad insertion has become the dominant model used with adaptive streaming. In this case the player makes a local decision to insert an ad before a requested video, or between videos clips in a play-list. The reason for this approach is targeting – the ad can be targeted to the individual subscriber based on known preferences or based on recent searches.

Adaptive streaming of broadcast content will, in many cases, include traditional ad spots. It would be possible to mark ad avails using SCTE 35 data in the manifest file, allowing ad insertion to be accomplished at the client (to replace the network ads). It is technically challenging to implement network ad insertion, because the CDN would have to be aware of SCTE 35 information and perform re-direction based on geographic or other parameters.

NETWORK CHARACTERISTICS

As adaptive streaming becomes a more significant source of video content how will this affect overall network utilization? What are the likely impacts to the cost of video delivery over IP networks compared to the traditional deliver using MPEG-2 transport streams over QAM?

Assumptions

On-demand content will continue to be delivered as a constant bit-rate MPEG-2 Single Program Transport Stream (SPTS) to existing devices such as deployed set-top boxes. The standard rates for these streams are 3.75 Mbps (Standard Definition) and 15 Mbps (High Definition).

As new devices, such as smart TVs, iPads, Tablets, etc. appear in homes, cable operators (notably Comcast, TWC, and Cablevision) are starting to support adaptive streaming to these devices. This trend will continue and more and more programming will be delivered using adaptive streaming.

In many cases the final connection to the device will be by WiFi, and is therefore subject to instantaneous fluctuations in throughput due to changes in propagation, including distance from the WiFi router, and other devices operating in the same limited frequency band.

Traffic Profile

Adaptive streaming has an interesting, and potentially very useful, traffic profile that makes it attractive to the cable operator:

- It plays well with other TCP-IP traffic since all TCP-IP traffic reacts to congestion is a predictable way.
- It takes advantage of the maximum network throughput available up to a limit, which is set by the highest bit-rate encoded version of the content.
- It automatically responds to network congestion by progressively reducing the bit-rate of the content (according to pre-defined bit-rates determined by the operator).

If adaptive streaming and general Internet traffic are distributed over the same network infrastructure, using differential quality of service mechanisms to tag video traffic as higher priority is recommended.

Overhead

How much overhead does HTTP adaptive streaming add when compared to MPEG-2 transport stream? The increased overhead in the forward direction is mainly due to the additional headers for HTTP and TCP/IP. Meanwhile, in the return direction, TCP-IP acknowledgements introduce an entirely new traffic flow.

However, adaptive streaming uses the latest, most-efficient compression algorithm (most likely, but not always, H.264, otherwise known as MPEG-4 AVC). In comparison to MPEG-2 compression the bit-rate is significantly reduced, approximately halved in fact.

Fairness

There is no guarantee that different streams over the same network segment will receive equal shares of the available bandwidth. It is quite possible that one device may hog the bandwidth (due to a more aggressive algorithm) while another may be starved (due to a more conservative algorithm). This behavior could cause disruption to other services.

Burstiness

Adaptive streaming is fundamentally different from MPEG-2 transport in that, for the duration of the download of a fragment, the HTTP transfer will consume as much network bandwidth as is available, generating a bursty traffic profile.

This behavior can be modified by limiting the maximum rate of each segment download to slightly greater than the maximum bit-rate. This technique requires more intelligence in the content distribution network (CDN).

Comparison with MPEG-2 Transport Streams

The current dominant mode of video and audio delivery in cable systems is based on MPEG-2 transport streams. How do MPEG-2 Transport Streams compare with adaptive streaming?

1) Broadcast

Statistical multiplexing is universally employed to pack more channels into each QAM channel on the network. Since zero packet loss can be tolerated, the multiplexer must analyze the video payload in real-time and, during peaks of traffic, reduce it by re-quantization of DCT coefficients. This makes statistical multiplexing a relatively expensive process that can only be justified for broadcast streams.

In comparison, adaptive streaming achieves a similar result by dynamically selecting the streaming profile according to traffic conditions. However, adaptive streaming does this by using a unicast delivery model. This means that each unique viewer of a given broadcast generates a dedicated stream in the access network.

On the other hand, when a broadcast channel is not being viewed (or recorded) by any subscribers within a service group, the bandwidth allocated to it is entirely wasted. In this case, an adaptive streaming approach would consume no network bandwidth for that service group.

For the above reasons, adaptive streaming is a very poor substitute for delivering *popular* broadcast services. Introducing adaptive streaming will cause a dramatic increase in the access network traffic because of the multiplicative effect of delivering a single broadcast channel as individual unicast streams to each client.

The underlying question is whether broadcast channels in the HFC network will continue to be an efficient delivery mechanism. As subscribers move to an on-demand mode of consumption, even of news and sports programming, will any channels remain that attract enough concurrent viewers (within a service group) to justify dedicating fixed bandwidth to them?

2) Switched Digital Video

Switched digital video (SDV) is a multicast delivery mechanism, and a channel consumes no network bandwidth when it is not being watched (or recorded).

Adaptive streaming is a poor substitute for switched digital video as long as the service is *popular* (as explained above in the broadcast case). However, niche programming, that is currently delivered using SDV, would be an excellent candidate for conversion to adaptive streaming.

3) On-demand

On-demand systems have been engineered to support MPEG-2 transport streams requirement for constant delay and zero packet loss:

- Bandwidth is reserved for the duration of the session (and that bandwidth is wasted if the session is paused)
- There is a hard limit to the number of sessions
- The number of QAM channels allocated to on-demand must be over-provisioned to minimize the probability of blocking
- In the normal case, QAM utilization is relatively poor.

Adaptive streaming is therefore a good candidate to completely replace on-demand services in cable systems over time. To provide the same quality of service as today's on-demand care would have to be taken to ensure that the network is not over-subscribed. Alternatively, differential QoS could be implemented to provide different service guarantees to ensure that pay and premium content is not degraded during periods of peak demand.

RECOMMENDATIONS FOR CABLE OPERATORS

Clearly operators have little choice when it comes to supporting new devices, like the iPad, in their networks. However, should a cable operator pro-actively consider a migration to adaptive streaming for devices within their control?

As newer set-tops are specified, inclusion of adaptive streaming could bring significant benefits in terms of network efficiency:

- On-demand QAMs could eventually be retired, liberating more RF channels for DOCSIS.
- Statistical multiplexing efficiencies could be gained by sharing the pipe with other traffic types.

CONCLUSIONS

Adaptive streaming was developed to provide the best user experience for streaming of content over an unmanaged network, like the Internet. As described in this paper, adaptive streaming cannot provide a service delivery quality that matches that of MPEG-2 transport systems. In particular, adaptive streaming compares unfavorably when it comes to delay, stability, and quality guarantees. In addition, because it is a purely unicast delivery mechanism, where the client pulls content from the network, no shared bandwidth efficiencies are gained from broadcast services.

Nevertheless, adaptive streaming brings with it a level of flexibility precisely because it was designed for an unmanaged network. It allows the operator to move away from the connection-oriented bandwidth reservation system required for MPEG-2 transport systems, and, eventually, to supporting a single IP network infrastructure for all services. This approach also allows new services to be deployed extremely rapidly, a well-proven result of network transparency from the Internet model.

Adaptive streaming is here to stay because of the appearance of popular client devices – tablets, smart phones and PCs – that support only adaptive streaming. Given this reality, cable operators are already moving rapidly to add adaptive streaming capabilities to their content delivery infrastructure.

Existing set-top boxes in the network will continue to function side-by-side until they eventually become obsolete. Newer set-tops will inevitably be designed to accept adaptive streaming formats as they become standardized. Eventually, an optimized future version of adaptive streaming will become the dominant mode of video delivery in cable networks.

REFERENCES

1. Dynamic streaming in Flash Media Server 3.5 – Part 1: Overview of the new capabilities, David Hassoun, Aug 16, 2010: http://www.adobe.com/devnet/flashmediaserver/articles/dynstream_advanced_pt1.html
2. HTTP Live Streaming, IETF Informational draft version 6, R. Pantos and W. May, Mach 31, 2011: http://tools.ietf.org/html/draft-pantos-http-live-streaming-06
3. Smooth Streaming Technical Overview, Alex Zambelli, March 31, 2009: http://learn.iis.net/page.aspx/626/smooth-streaming-technical-overview/
4. MPEG Dynamic Adaptive Streaming over HTTP (DASH) http://www.slideshare.net/christian.timmerer/http-streaming-of-mpeg-media
5. An Experimental Evaluation of Rate-Adaption Algorithms in Adaptive Streaming over HTTP, Saamer Akhshabi, Ali C. Begen, and Constantine Dovrolis, MMSys' 11, San Jose, CA.

ADAPTIVE STREAMING AND CONVERGED MANAGEMENT STRATEGY IN MULTISCREEN VIDEO SERVICE IMPLEMENTATION

Duncan Potter, Goran Appelquist
Edgeware AB

Abstract

With the massive proliferation of both video services and number of devices capable of consuming high quality video, the question of whether or not a cable operator should deploy a multiscreen environment is now one of survival rather than just opportunity.

This paper addresses 2 main areas of complexity:

1/ The use of HTTP Adaptive Streaming frameworks and their associated protocols in a distributed network environments

2/ The scaling of a management architecture to support this huge increase in both sessions, and traffic, and possible routes to further monetization

INTRODUCTION

Traditionally, telco and cable operator‟s video services have been distributed over managed access lines where the bandwidth required for a good quality of experience has been provisioned and is suitably robust. However, there are now a huge range of Internet connected devices available which are capable of high quality video playback. These include laptops and home media centers, smartphones such as the Apple iPhone, Blu-ray devices, and gaming consoles. These devices are typically connected to unmanaged access networks such as 3G, home networks and Wi-Fi hot spots.

In addition, video content owners are increasingly choosing to make their content available directly on the Internet via massively popular services such as the BBC iPlayer and Hulu™. Delivery of these services is typically handled by Content Delivery Networks (CDNs) such as Akamai and Limelight Networks® that deliver "Over The Top" of operator networks, leading to the description "OTT video services". Although CDNs optimize delivery over the transit network, these services are all affected by varying degrees of congestion when they reach the local operator‟s network – the so called "middle and last mile".

However, demand from consumers to watch video anytime, anywhere has led to an urgent requirement for operators and CDNs to be able to deliver video services to these devices with a high Quality of Experience (QoE). A number of leading companies have developed HTTP Adaptive Streaming technologies to specifically enable this including Microsoft®, Apple and most recently Adobe®.

In order to scale a system for providing services to this plethora of devices there are many components required.

1. HTTP ADAPTIVE STREAMING FRAMEWORKS AND PROTOCOLS

Issues Addressed by HTTP Adaptive Streaming

Network Connectivity and Traversal Assurance Requirement

When attempting network connectivity in an unmanaged network, there are a lot of unknowns. The unknowns include the potential existence of routers, firewalls as well as which ports are open. In a home network, there are personal firewalls, possible routers and security software scanning port activity.

In a Wi-Fi hot spot, the port access can be extremely limited due to security concerns.

This is a well-known hurdle with network applications and is overcome by using the HTTP protocol for communication. HTTP uses port 80 for requests. Requests to this port are most likely to be allowed through any firewall or router as they are used for all web surfing. As HTTP uses a state-full TCP connection, any issues that can be incurred by NAT based networks are also overcome.

Bandwidth Management Requirement

Bandwidth consistency is a major issue. If a user is watching a video and someone else on the same network suddenly decides to perform a file transfer, the available bandwidth for the video can be severely impacted. In order to maintain a good Quality of Experience, content therefore needs to be encoded at different bit rates and the delivery protocol needs to be able to dynamically switch the bit rate with no interruption in playback or action by the user.

The HTTP protocol is a synchronous client-to-server protocol so only one request can be made for a video file. Switching bit rates on the fly is therefore not possible in the middle of an HTTP transaction. To overcome this problem, the video must be sliced up into "chunks." Each chunk is typically between two to ten seconds of video. The chunk sizes are such that the reference IFrame at the beginning of each chunk is synchronized.

The delivery server can host several different bit rate encodings of the same video content. Each bit rate encoding has a separate playlist, which is defined by a master playlist. These playlists are typically in an M3U format and contain the list of chunks in order. When the client detects either insufficient bandwidth or more available bandwidth, it can switch to either the lower or higher bit rate playlist and download the chunks in that list. Since each chunk is synchronized with the other bit rate streams, there is a seamless transition between them so that the video playback is not interrupted. This maintains a high quality user experience.

Multiple Clients and Resolutions

As more and more Internet connected devices appear on the market every month, with greater and greater capabilities for video browsing and playback, the number of resolutions and bit rates required to support these devices increases exponentially. Historically, many devices have also communicated via a proprietary protocol. It is not viable to have separate encoders, DRM systems and delivery servers for each device that needs to be supported.

HTTP Adaptive Streaming enables delivery of multiple resolutions and bit rates over a common, open protocol. This enables consolidation of the encoder, encryption and delivery server infrastructure to a single manageable system. New devices can be added simply via a new encoding profile.

Content Security

The traditional broadcast method of restricting access to live content is achieved by encrypting the transmission stream. This method has been successfully applied to delivery of live video over the Internet: content is encrypted between the encoder and the server and then again between the server and the client. However, the risk with this approach is that content is being temporarily written to cache in the server before immediate retransmission and again is written to the client cache before immediate presentation and deletion. During these short periods, content is sitting „in the clear" and this presents a security risk. Also, for VOD and catch-up TV services, content will be stored on servers in the network or client devices and must remain protected to avoid piracy.

HTTP Adaptive Streaming enables encryption of the individual chunks of content for both live streaming and VOD / catch-up TV downloads. The content itself is encrypted during or immediately after encoding and remains so during transmission across networks and when stored on servers or client devices.

<u>HTTP Adaptive Streaming Ecosystem and Architecture</u>

<u>Support within an Asset Caching & Propagation System</u>

Microsoft® Silverlight®, Apple Quicktime and Adobe® Flash® are the leading frameworks for the creation and playback of Rich Internet Applications incorporating video. These include server programming tools for the creation of the presentation portals, tools for the creation of encoders to compress and format video for Internet delivery and client programming tools for creation of video players on different Internet connected devices.

All frameworks now support variants of HTTP Adaptive Streaming. The following describes how an asset caching and propagation system might support these variants:

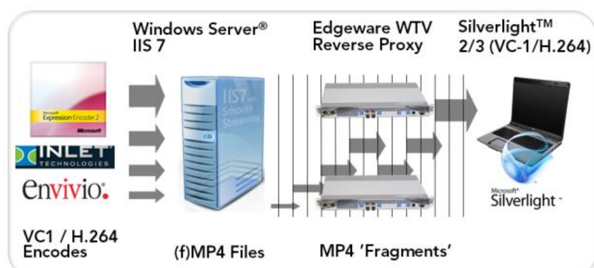
<u>Microsoft® Smooth Streaming</u>

The video delivery server acts as a reverse (server side) proxy to a Windows Server**®** running Internet Information Service 7 with the Media Service 3.0 extension. When running in this mode, caching replaces the need for proprietary asset propagation and this is therefore disabled.

The following is a description of the Smooth ecosystem and content flow:
Microsoft® or 3rd party encoders generate a single contiguous file per bit rate according to the ISO/IEC 14496-12 ISO Base Media File Format (MP4) specification. The files have the ***.ismv** extension and contain MP4 video fragments (and audio fragments if the video source also contains audio).

An XML-based server manifest file and a client manifest file is also generated. These describe available bit rates and other information required by the IIS 7 server and the Silverlight clients

For on-demand Smooth Streaming, following encoding, all these files are copied to the IIS 7 server or are published if WebDAV is enabled. For Live content, rather than storing the fragments in MP4 containers, encoders deliver the fragments directly to Live Smooth Streaming publishing points on the IIS 7 server. The server itself then generates a manifest for clients, based on information provided to it by the encoder.



Smooth Ecosystem and Content Flow

Requests for on-demand content stored on the IIS 7 server or for live content that is being delivered to a Live Smooth Streaming publishing point on the IIS 7 server should be directed to the distributed video server E.g. a request for an on-demand asset stored on the IIS 7 server with the following URL *http://iismedia7/BigBuckBunny/default.html* should be directed to the IP address of the Edgeware server.

In addition, the Client Cache Settings of the IIS 7 server should be enabled to specify the Cache-Control header. This is used in the IIS

7 server to specify any intermediate caches the conditions and restrictions for caching of content.

Requests for live and on-demand Smooth streams on the IIS 7 server are proxied by the distributed video server, ingested via HTTP and cached according to the specified cache Control directives. For a fragment that has been cached, subsequent requests are served directly from the distributed video server.

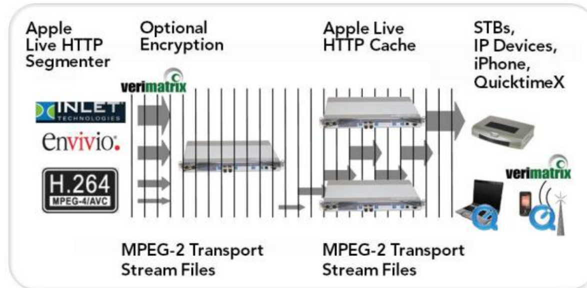
Apple Live HTTP Streaming

There are two modes of operation when supporting Apple Live HTTP streaming with distributed delivery servers: one for non-encrypted content and one for encrypted content.

For encrypted content, the distributed video server does not act as a reverse proxy but instead receives segmented content pushed directly from an encoder or content store. The architecture is very similar to Smooth Streaming except the (f)MP4 files are replaced with MPEG-2 Transport Streams (*.ts extension).

A segmenter creates and maintains an index file (*.M3U8 extension) containing a list of the media files. Apple provides a software segmenter but typically this is included as part of the encoder functionality. The index file and the stream segments are all pushed to the first server using WebDAV.

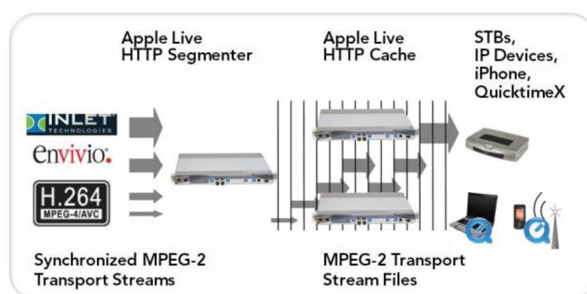
Additional distributed video servers can be added as edge caches to this first server and distributed asset propagation system can be enabled to download transport stream segments according to their popularity.



Mode 1 – Encrypted Apple HTTP Ecosystem and Content Flow

For non-encrypted content, encoders deliver a single Transport Stream per bit rate without segmentation but with synchronized IFrames. These can be ingested by the Web TV server via FTP, HTTP or even scheduled UDP multicast for live transmissions. This option significantly reduces core network bandwidth, especially if live content is to be distributed to multiple servers.

A centralized server can then perform the segmentation and create the index file. Additional Servers can be added as edge caches to this first server and asset propagation systems can be enabled to download transport stream segments according to their popularity. Alternatively, each edge server can ingest complete transport streams via scheduled multicast or according to popularity via the asset propagation and it can then perform segmentation on the fly.



Mode 2 – Non-Encrypted Apple HTTP Ecosystem and Content Flow

Requests for Apple Live HTTP files should be directed to the IP address of the appropriate server. Client software first reads the index file, based on a URL identifying the stream. This index specifies the location of the available media files, decryption keys (if applicable), and any alternate streams available. For the selected stream, the client downloads each available media file in sequence.

This process continues until the client encounters the #EXT-X-ENDLIST tag in the index file. If no #EXT-X-ENDLIST tag is encountered, the index file is part of an ongoing broadcast. The client loads a new version of the index file periodically. The client looks for new media files and encryption keys in the updated index and adds these URLs to its queue.

Conditional Access Solution for Enhanced Apple HTTP Live Streaming

The Apple HTTP Live Streaming Protocol provides optional encryption of the video chunks using the AES-128-CBC block encryption algorithm. During encryption, a 128-bit key is generated and placed in a "key file" on a server so it can be downloaded by the client. The location of the key file is given in the playlist. A new key can be given at any time. When a key file is encountered in the playlist, this key must be used to decrypt each subsequent chunk until another key file is encountered.

The URL provided to the client to retrieve the key file is HTTPS. This ensures that the connection to the server will at least be encrypted, even if it does not have server-side authentication. However, the protocol does not provide a way to ensure mutual authentication of the HTTPS session. There is no client-side certificate provisioning built into the protocol and there is no mechanism for the reporting of a unique client identifier from the device to the server. The lack of either of these features does not allow for enforceable entitlements against the video content encryption key.

To overcome these security limitations and enable HTTP lives streaming video delivery suitable for a secured pay-TV service, the distributed delivery supplier must have confirmed interoperability with a security solution such as that provided by Verimatrix. Using such a system, the 128-bit keys are managed and selectively distributed to the encoder and authorized clients only and can support HTTP live streaming in a standalone OTT service configuration, or can be utilized as part of a unified security head-end supporting multi-screen deployments pairing OTT delivery alongside IPTV and DVB content distribution. Currently, to support this additional security, the encoder must perform the segmenting of the chunks prior to encryption and distribution to distributed video delivery servers (Mode 1).

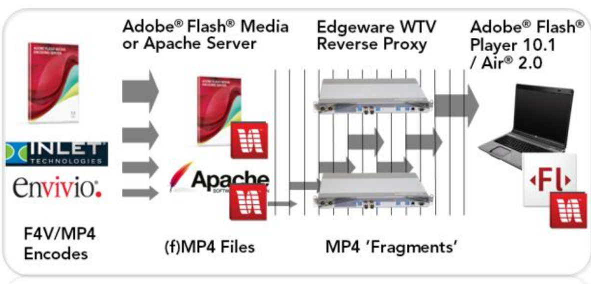
Adobe® HTTP Dynamic Streaming

The Web TV server acts as a reverse (server side) proxy to an Apache Server running the Adobe® Origin Module (VOD) or an Adobe® Flash® Media Server 4.0. When running in this mode, caching replaces the need for asset propagation system and this is therefore disabled.

The following is a description of the Adobe® HTTP Dynamic Streaming ecosystem and content flow:

Adobe® or 3rd party encoders generate a single contiguous file per bit rate using the following file types: F4V/MP4 compatible files and FLV. For VOD content, an Adobe® File Packager command line tool is used to parse the file and translate it into fragments. This tool can also be used to encrypt files for use with Adobe® Flash® Access 2.0 DRM. Once created (and encrypted if used) fragments are written to files with the ***.f4f** extension. Each file can contain multiple fragments and quantity and duration of fragments per file can be optimized.

An XML-based client manifest file is created with the filename of the input file. This contains information about the codec, resolution and the availability of multi-bitrate files. In addition, a server index file is also generated. This contains information on the specific location of fragments within a file for the Origin Module to translate to byte range requests.



Adobe® HTTP Dynamic Streaming Ecosystem and Content Flow

For VOD, requests for content stored on the Apache Origin server should be directed to the distributed video delivery server. E.g. a request for an on-demand asset stored on the server with the following URL http://localhost/media/webplayer.html should be directed to the IP address of the Edgeware server.

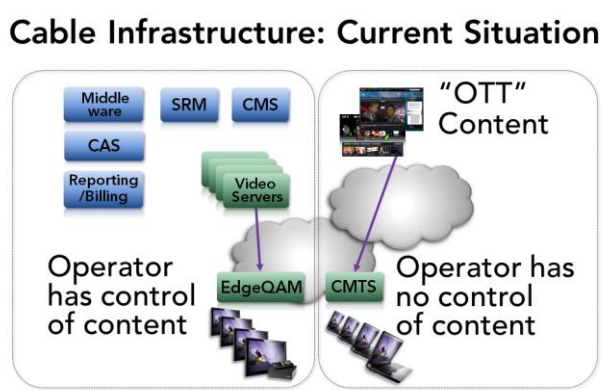
In addition, the Client Cache Settings of the Apache server should be enabled to specify the Cache-Control header. This is used in the Apache server to specify any intermediate caches the conditions and restrictions for caching of content.

Requests for on demand streams on the Apache Origin server are proxied by the distributed video delivery server, ingested via HTTP and cached according to the specified cache Control directives. For a fragment that has been cached, subsequent requests are served directly from the distributed video delivery server.

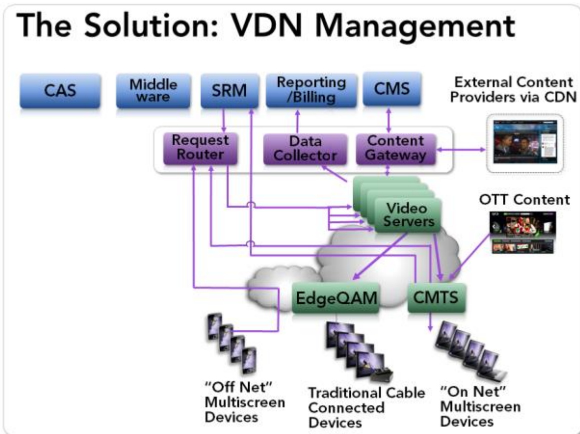
For live streams, the Apache Origin is incorporated into the Flash® Media Interactive Server (FMIS). Live streams are ingested over Real Time Messaging Protocol (RTMP) and segmented into *.f4f files. The FMIS server has a built in Apache HTTP Server and uses the Origin Module to deliver the live content over HTTP. Requests for live streams are proxied by the distributed video delivery server and cached in the same way as on demand streams.

2. CONVERGED SCALABLE MANAGEMENT ARCHITECTURE

In many cases the current cable environment provides sophisticated management of the video infrastructure but little opportunity to extend access to the video assets outside the closed cable environment. The challenge is not necessarily to provide access to the video assets but to provide access without building completely parallel infrastructure and management systems. To achieve this, it is important to identify the key management components of an existing system and then to propose the minimum changes to the architecture to provide integrated management of both traditional cable and multiscreen environments



In the new converged management environment, an additional layer is inserted that will be referred to as the Video Delivery Network (VDN) management layer.



It is made up of 3 main components:

Request Router (a.k.a. Session gateway)

The request router is responsible for receiving the request from one of 3 potential sources:

Existing Session Resource Manager (SRM)

An existing SRM system would forward a request from an existing Set Top Box (STB) using a standard protocol such as Real Time Streaming Protocol (RTSP) or (in the US) Streaming Control Protocol/ Lightweight Stream Control Protocol (SCP /LSCP).

"On Net" Multiscreen web device

On the operators own IP network, a device would access the video services through a middleware interface but in effect would connect to the video services through the Request Router (RR) using an HTTP request. The RR would then redirect the device to the nominated video delivery server, which would provide the requested content.

"Off Net" Multiscreen web device

There is also a possibility that devices connected to an alternative network such as a wireless operators network would also want to access the video services. Assuming that peering agreements are in place, the RR would also be responsible for ensuring access to the content for off net devices.

Decisions on which video delivery server to connect to could be made via a range of selection criteria, such as geographic lookup, IP range or service group depending on the operators own criteria.

Content Gateway

The content gateway is responsible for the interface between the Video Delivery Network subsystems, the existing CMS system and also providing an interface to external content providers such as CDNs.
It is responsible for content ingest for both LiveTV and VOD content

NPVR functionality is also provided by the content gateway with middleware providing the user interface to allow users to schedule their own recording schedules, which are then provided in a standard format to the content gateway.

Data Collection and Reporting system

The Data Collection and Reporting system is vital to the revenue generation element of the system. It is responsible for 4 main elements:

Usage statistics

This is the collection of statistics showing which assets are being viewed, by which subscribers, at what time, for how long and through which viewer.

Performance monitoring

In an adaptive streaming environment it is vital that not only viewing behavior can be monitored but also the delivered quality is monitored as well. Unlike a traditional cable environment, adaptive streaming works by providing different quality video and the ability to automatically adjust based on bandwidth availability. It is therefore essential to be able to understand the actual quality of the video being delivered by the video delivery system and the underlying network infrastructure.

Fault Monitoring

Monitoring of errors, faults and overall loss or degradation of the hardware and software components of the system.

Reporting

This part of the system is also responsible for providing all statistics either through a northbound interface to an ISA or NGOD compatible system, a proprietary API or through a provided graphical user interface.

Note: Security and Conditional Access

Security and conditional access falls outside the scope of this paper, being a significant subject in its own right. It is suggested that anyone looking in to this area must consider it as an integral part of the system and therefore should seek assurance of the interoperability of a system such as the Verimatrix VCAS [1] system with the chosen management framework

3. SUMMARY

This paper addresses two of the most common questions and challenges for a cable operator contemplating a multiscreen approach to serving new and existing subscribers. There are of course many other considerations that fall outside the scope of this paper.

Understanding of adaptive streaming protocols and their implementation is key to successfully planning an implementation. A converged management strategy is essential to maintain the lowest possible TCO and efficiency while launching, maintaining and growing profitable video services to an expanding subscriber base.

4. REFERENCES

[1] Verimatrix HTTP Live Streaming Interface Control Document (ICD).

QOE ISSUES RELEVANT TO VIDEO STREAMING IN CABLE NETWORKS

Jeremy Bennington
Praveen Mohandas

Cheetah Technologies, Sunnyvale, CA

Abstract

This paper explores how monitoring video quality in a streaming service environment is different from a traditional video network. It discusses technical differences as well as how quality assurance data must be correlated and reported so that it becomes operationally relevant and actionable for an operator. The paper also proposes a set of best practices and architecture for operators to consider for monitoring video quality in a streaming service environment.

INTRODUCTION

Emerging Internet video streaming platforms like Hulu, Netflix, YouTube, Amazon and iTunes are challenging the traditional video service business offered by Cable MSO's. In addition to providing on-demand service, they provide access to TV shows and events aired previously – providing a cloud-based video functionality across multiple mobile platforms, an experience appealing to many users. User generated content aggregators like Justin-TV, YouTube, Ustream, LiveStream and several other providers are also trying to get into the premium content space. Netflix currently has about 20 million customers; in few years they expect this number to rise to 60 million. MSOs, like Comcast through its Xfinity TV application, TimeWarner Cable and Cablevision with their iPad apps, are rapidly innovating their service offerings as well to provide a multiple screen service to subscribers.

While subscribers expect good video quality for multi-screen streaming services just like they do for traditional video services, the radically different architecture, distribution model and consumption behaviors of these new services calls for a new approach to qualifying and ensuring a high Quality of Experience (QoE). New ways of measuring, correlating and aggregating QoE data are required to take into account, at the most basic level, the disappearance of the household as the ultimate frontier. With streaming services, users are not limited to a single device, to a single operator network, or even to a device hard-wired to the network.

Monitoring video QoE of streaming services involves different codecs (MPEG4), wrappers (Adobe Flash, HTTP5, MS Silverlight) and distribution protocols. Note, with recent announcements, the adoption of HTTP Live Streaming (HLS / HTTP5) may become the dominant method. All these techniques pose a significant challenge due to segmentation of the video into small chunks with multiple bit rates for transmission. Segmentation, however, only represents a portion of the issue at hand.

The biggest challenge lies in the representation of the data and in making it operationally relevant to the operator in a rapidly changing business and technical climate. Some of the questions operators are trying to address include: how was the user's experience across multiple assets since the idea of a channel does not directly apply? How was the experience across multiple screens? How was the experience of the subscribers within a single household (ie. revenue generating unit)? How was the experience of a specific asset across a geographic region or across a type of device? How does the streaming experience of a broadcast correlate with the traditional delivery of a given channel?

STREAMING VIDEO ARCHITECTURE

There are several architectures for streaming of video. In general they all start with either a video broadcast stream from a programmer that may also be broadcast over a traditional distribution like CATV, or the streaming video starts with a fixed file asset like VOD or user generated content. This part of the architecture can be named "Content Ingest" (see figure 1).

The next step in the streaming video architecture is the "Content Encoding" or transcoding of the video into the codec and bit rate desired by the operator. Typically the encoding is MPEG4 and includes multiple bit rates for each program so a client can adapt to the network bandwidth available. There are two to four bitrates of each program in typical deployments. The next step in the "Content Encoding" is the segmentation of the video into small "chunks" of video typically about 2 seconds long. These segments are progressively requested and downloaded by the client application based on the m3u8 "playlist" file.

After the encoding and segmentation of the video the programs arrive at the "Origin Servers" where the information about the programs is made accessible by the clients. The client negotiate and communicate with the Origin Servers to gain access to the program and the different bit rates as bandwidth constraints are determined by the HTTP/TCP protocols and client interaction. From the Origin Servers the video is transmitted over a managed or unmanaged network. It may be a Content Delivery Network (CDN) specifically or purely third

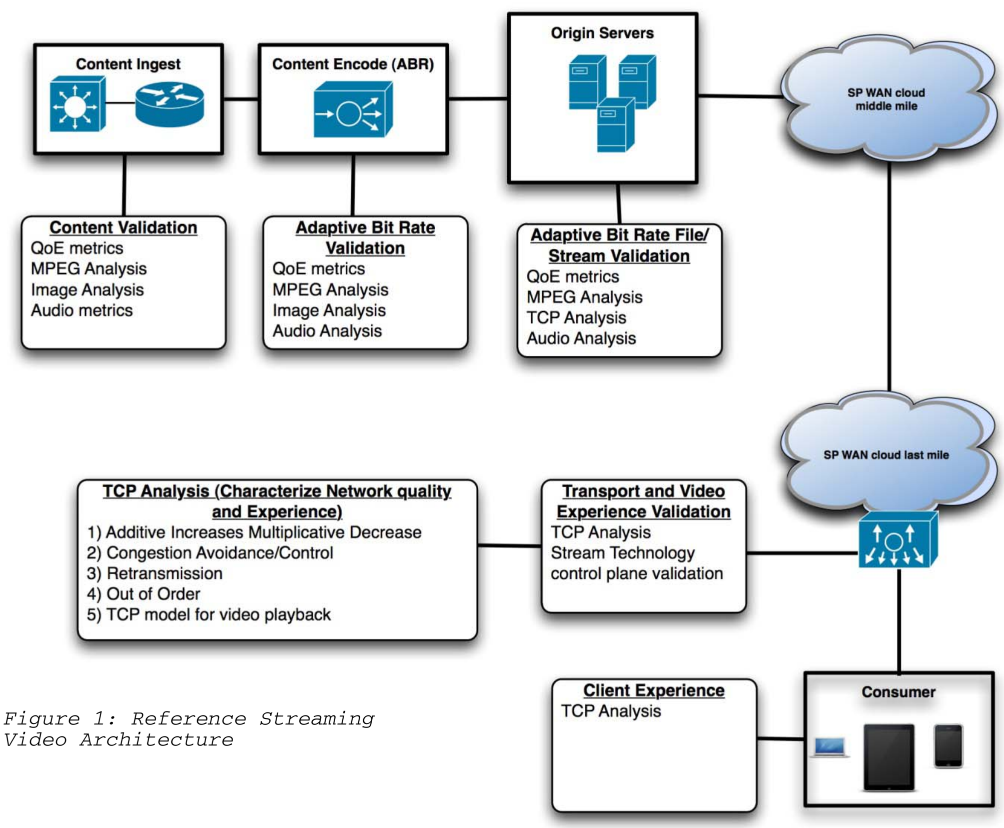


Figure 1: Reference Streaming Video Architecture

party unmanaged networks or any mixture. Once the video has traversed the network it is consumed by a client. The client may reside on a TV, PC, tablet or mobile device.

COMMON QOE METRICS FOR STREAMING & CATV

In both streaming video and traditional video in a cable network the content acquired from a programmer or user must be validated for quality before encoding. This "Content Validation" provides quality assurance and a baseline for the operator to compare QoE at all the subsequent steps of video processing and distribution. Content Validation at the core is not highly dependent on the MPEG and transport information and instead measures the decompressed video quality as if a subscriber were actually viewing the video.

Using human vision system models we can computationally model the subjective score of video quality, but instead of repeating subjective scores using personnel watching mosaics of video monitors one can automate the process. Thus making the video fidelity (image analysis) scores objective, repeatable, consistent and scalable to ensure that an operator always knows what the video quality was to begin with. This avoids finger pointing between programmer and operator, as well as solving answering the age old question of junk in / junk out or good in / junk out?

When video fidelity measurements are tracked over time operators and programmers can ascertain the overall quality ranking of programs, monitor changes in quality and short term events where the quality diverges. If this is then correlated with information about the MPEG coding and transport the programmer and operator can ensure the Content is validated before an operator processes the video for distribution.

Once the content has been validated, quality trends identified and quality baselines established the encoding/transcoding of the video must be monitored. Monitoring of the encoding process for streaming video and traditional video is very similar. The output of the encoder must measured for MPEG-TS metrics like PCR jitter and other ETSI TR101290 type metrics. The MPEG VCL (video coding layer) should be analyzed to determine the GOP structure, slice types, quantizer values, bit rates and other attributes of how the video was encoded. Last, but most important, the video content must be inspected for image artifacts commonly introduced by encoders like blockiness, frozen video, blackout and jerkiness.

All of the QoE measurements from the Content Encode can be correlated to provide quality assurance of the encoding/transcoding process across the service offering. This information should also be compared and correlated to the content ingest measurements to provide the operator with a measurement of the degradation of each program introduced by their processing of the video.

At this point in a traditional cable network the video would be multiplexed with other programs and sent over the network for transmission (with possibly ad-insertion). In a streaming network the video is not multiplexed. Instead it is segmented into two second chucks (typically) and transmitted to clients via HTTP/TCP by the origin server. At the origin server an operator should emulate a client to monitor the various bit rates of each service and the availability.

Another vast difference between streaming and traditional video distribution in cable networks is the end device and sometimes the presentation of the content. In traditional environments a tightly controlled STB decodes the video (at a fixed bit rate) and provides one of three or four fixed resolutions to the television. Even though there are many television suppliers it is more constrained than the display types and capabilities of mobile devices! The presentation of the video is also tightly constrained by a channel lineup and guide typically provided by the cable operator unless a CableCard device is being used. In the streaming environment there are multiple bit rates, a huge type of display types, presentation, etc.

UNIQUE METRICS FOR STREAMING

Unlike traditional video distribution, in the streaming architecture, video will be encoded into multiple bit rates. This is done to allow a client to adapt to bandwidth availability thus providing high quality video when bit rates are high and lower quality video when bit rates are low, but reducing the number of stalls which are more annoying to a user than image quality (in most cases). With multiple bit rates of video being produced by the encoder they each must be correlated with the content ingest and between each other to ensure consistency. This can also be used to ascertain the end-user quality experience when they look at a program and shift between different bit rates of the same program.

Unlike the traditional distribution and reception of video by the user in a streaming environment the quality may be changing on purpose! Traditionally a user tunes to a HD or SD channel and receives a fixed resolution and bit rate service. An operator monitors to ensure that the service meets that expectation to reduce support costs and maintain high customer satisfaction. In a streaming environment a user will have a variable quality experience. The causes of the variability may be within the control of the subscriber, operator or a third party. Regardless of the cause an operator should monitor several attributes to ascertain the subscriber QoE. Primarily they should look at the bit rates that are streamed for each asset to each subscriber. This correlated with the number of stalls provided a very good indication of the subscriber QoE. It is also very insightful to correlate this with the percentage of the asset viewed for a VOD asset or the length of view versus average length of view for a broadcast asset.

Unlike in a traditional distribution in a streaming architecture the transmission of video is over TCP/HTTP. This is a very flexible method, but it requires constant communication between the client and the server to provide the service. This increases the upstream bandwidth requirements of the service. If a cable operator chooses to provide a streaming video service they must consider not only the large downstream bandwidth requirements for their DOCSIS network, but also the upstream requirements. And, would it even make sense to build the upstream transactions into the QoS mechanisms of DOCSIS? Similarly a cable operator could provide a different SLA/QoS for streaming service, assuming the business case and law permits.

In m3u8 files there are bandwidth descriptors, however, monitoring the actually video bandwidth upstream and downstream provides a more accurate view of the network requirements. Moreover, aggregating the bandwidth usage based on bit rate types, program/asset and locations provides a better understanding of what programs are generating the most bandwidth demand, what locations are consuming the most bandwidth per stream, etc. Another important correlation is the bandwidth usage, stalls/buffering events, QoE and the duration/percentage of a program watched. From this an operator can gain insight into the affects of QoE on user behavior. For instance if a channel is only viewed for short periods or a VOD assent is not played completely is it because of the QoE and number of stalls, or was it ok and that is the normal behaviors for he particular piece of content?

BUSINESS MODEL IMPACTS ON QOE

The choice of business model and the choice of broadcast, VOD short-form and VOD long-form have a large impact on the QoE for streaming. For instance the user expectation of QoE for a subscription service is higher than a free service, but lower than that of a PPV service. Similarly the QoE expectation of long form video is higher than short form even though it may be more difficult to meet the short form QoE requirement (a conundrum).

The choice of a business model that adds streaming to an existing subscription versus a unique service also impacts the user expectations of QoE. In some cases linking the traditional subscription to a streaming one may create an expectation of the same QoE, which may be difficult to meet.

The choice of business model whereby the network/CDN is wholly owned and QoS can be guaranteed and QoE measured at all points is vastly different than an environment where the Origin Servers are the last asset that is owned. If the network is owned and tightly constrained by the operator they can monitor the origin server, client and also at the edge of the CDN where it interfaces with the broadband network. In this model it is easier to determine the quality across the service and the cause of issues when they arise. If the Origin Servers are the last owned part of the network then there is a greater need for monitoring within the client player/device to ascertain what the end user QoE is.

CONCLUSIONS

Video streaming in cable networks is a valuable service to expand a cable operator's value to subscribers by reaching beyond the TV to all media capable devices. Moreover it offers a different value and level of interaction with subscribers not available in a traditional TV model. When deploying a video streaming service QoE must be monitored to ensure that subscribers are receiving the QoE expected by the operator. Moreover, the operator needs to monitor QoE and be able to identify the root cause of any issues and determine if they are issues that are internal or external. Some of the QoE monitoring methods are the same between streaming and traditional video distribution, while there are several methods that are unique to streaming.

As operators launch a video streaming service the ability to correlate quality to viewing behaviors and across new and old distribution is vital to not only providing a good service and reducing cost, but also in finding new revenue opportunities. For instance the cable operator will now be able to differentiate different users within a "subscriber" address. In the traditional video environment a subscriber is a household with TVs. In the new paradigm of providing both a traditional TV service and streaming there is the possibility that each person in the household will consume their own content together or independently. Targeting ads to each person in the household becomes possible. Enabling interaction between the streaming services and the TV becomes possible as well. All of these possibilities hinge around ensuring a specific QoE for each service and correlating the data to reduce costs and ensure that behaviors are truly related to the content and not a poor QoE.

BIBLIOGRAPHY

[1] S. Winkler: *Digital Video Quality*. John Wiley & Sons, 2005.

[2] J. Deutscher: *Smooth Streaming Primer*. http://learn.iis.net/page.aspx/941/smooth-streaming-primer/

ADAPTING ADAPTIVE STREAMING TO CABLE ACCESS

Xiaomei Liu
Cisco Systems

Abstract

Cable operators are increasingly embracing a new wave of video delivery to extend video services to multiple screens and off-net subscribers. The latest and the most important technology breakthrough to enable this new paradigm of video delivery is adaptive streaming. With wide support on client devices, adaptive streaming holds the promise of a unified video service delivery architecture for unmanaged devices and off-net subscribers.

Adaptive streaming successfully tackles the challenges of delivering video services over unmanaged networks. However, delivering video services using adaptive streaming at scale in managed cable networks is demanding.

This paper examines various technical challenges of adaptive streaming in managed cable access networks and then explores a variety of optimization options. It proposes upstream optimization with TCP ACK QoS prioritization and TCP ACK suppression in addition to simple upstream bandwidth expansion. For downstream optimization, the coupling of VBR video and adaptive streaming is proposed to reduce downstream bandwidth consumption and to improve video quality significantly. Finally, a flexible downstream QoS design with bandwidth protection, stream prioritization and video quality optimization is introduced.

INTRODUCTION

The popularity of Internet video streaming and over-the-top (OTT) video services shows no signs of abating. Cisco's Visual Network Index research predicts that 91% of all consumer IP traffic will be video traffic by 2014 [1]. Netflix, a leading OTT video streaming service provider, reached a milestone of 20 million subscribers at the end of 2010. Statistics indicate that Netflix video streaming traffic already represents more than 20 percent of total downstream traffic during peak times in the United States [2]. The initial success of OTT video providers, such as Netflix and Hulu, not only supports the business case of online video but also proves the feasibility of the underlying technologies.

Behind the scene of this undeniable success of streaming video lies the technical foundation: Adaptive Bit Rate (ABR) Streaming. Despite implementation variations from major ABR streaming vendors, such as Microsoft, Adobe and Apple, the architecture essentials of ABR streaming are the same. A brief description of adaptive streaming architecture is given in the following paragraphs while a thorough introduction is referenced [3].

An ABR video delivery system consists of servers, networks and clients as depicted in Figure 1. At the server side, a video asset is encoded at multiple video quality levels using different bitrates. The higher the encoding bitrate of an asset, the better the video quality. Video content encoded at a particular bitrate profile is further segmented into small fragments. Each fragment corresponds to a few seconds of video playtime. There is also a manifest file for each ABR video asset. The manifest file stores metadata of the asset, such as bitrate information and fragmentation boundary information. ABR servers include processing and packaging servers and origin servers. The processing and packaging servers convert linear and on-demand video input to video fragments and generate manifest files. The origin servers host the video fragments and manifest files. To support linear video delivery, the processing servers and packaging servers continuously ingest video content and output packaged fragments to origin servers. At the same time, the manifest files for linear streams are updated continuously.

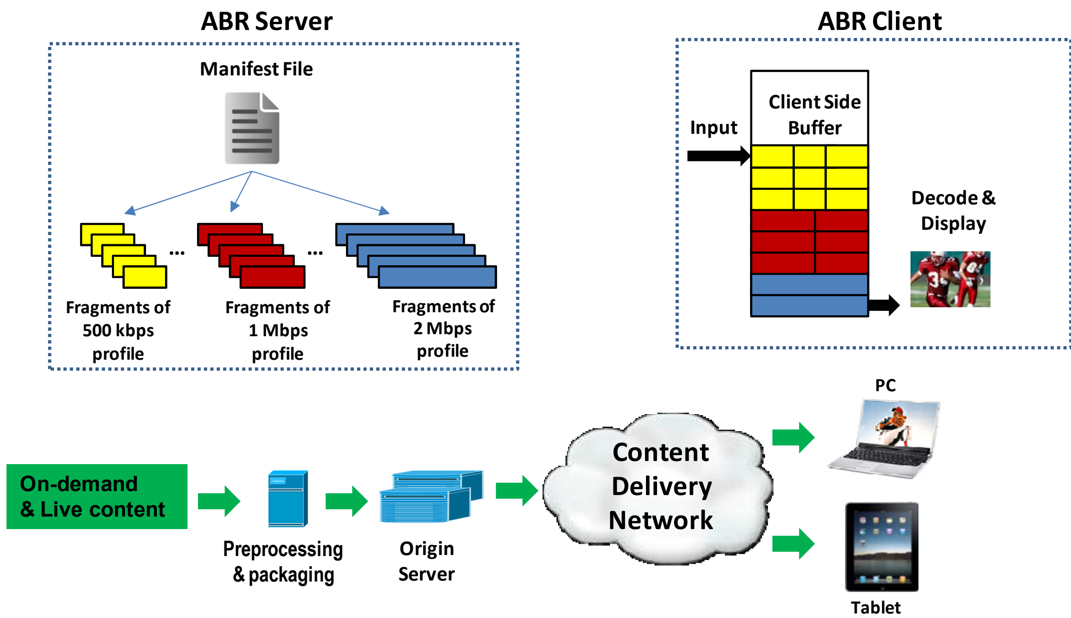


Figure 1. Adaptive Bit Rate Streaming

Since ABR was designed to support video delivery over unmanaged networks, ABR streaming presumes best effort networks without bandwidth guarantee and QoS protection throughout the ABR delivery process. Large scale deployments of ABR streaming typically employ Content Delivery Network (CDN) to reach clients across far reaching locations. Instead of having centralized ABR streaming servers to directly serve ABR clients, client requests are served by distributed CDN edge cache servers. The most popular content is cached at the CDN edge cache servers and can be served to clients immediately upon request. If content requested by a client is not in the cache of CDN edge servers, the CDN obtains the content from an origin server and then delivers the content to the requesting client. Given that ABR streaming utilizes HTTP for the delivery of video fragments, the standard HTTP caching capability of CDN can be readily leveraged without imposing special requirements on the CDN. The CDN system greatly reduces origin server loading while at the same time minimizing the bandwidth demands on regional and aggregation networks.

At the client side, an ABR client first obtains the manifest file of a video asset and then requests video content from streaming servers or CDN on a fragment by fragment basis. Each fragment is obtained via HTTP GET with the URL uniquely identifying the requested fragment. Depending on implementation, byte-range may also be used for fragment request. When the client requests a fragment, the client's desired bitrate profile is provided to the server. Based on multiple factors, such as client screen resolution, CPU load, power consumption, and network conditions, the client can adjust the bitrate profile requested to the server in real-time by sending a new profile request to the server. For the rest of the paper, the focus is on the bitrate adaptation to network conditions.

How does ABR streaming solve problems of video delivery over unmanaged or congested networks? The two major roadblocks to successful video delivery in unmanaged networks are unreliable packet delivery and lack of bandwidth guarantee for video traffic. Due to the fact that video content is highly compressed, any packet drop in a delivery network greatly impacts the end user experience. Common encoding technologies, such as MPEG, improve coding efficiency by removing temporal and spatial redundancy of video sequences. A video sequence is divided into groups of pictures. Each group has a reference frame and several other pictures that have decoding dependencies on the reference frame. A single packet drop can therefore impact a group of pictures if the dropped packet happens to be on the reference video frame. ABR streaming solves the unreliable network problem by first delivering video packets over the TCP protocol instead of the UDP protocol. TCP protocol is a reliable transport protocol that automatically repairs packet loss through retransmission. As a result, the video application layer benefits from a reliable transport and improved packet loss resilience.

In order to solve the second problem of non-guaranteed network bandwidth, ABR streaming applies an innovative approach dealing with network bandwidth variations. When networks experience congestion, an ABR client down-shifts the bitrate profile of requested video fragments to align with less available network bandwidth. Similarly, when the network load is light, the ABR client upshifts the bitrate profile of requested fragments to improve video quality. Video quality at the client side is self-adapting to network conditions such that end users can obtain the best video quality under the constraint of network resources.

An ABR client maintains a large client side buffer (e.g. 10-30 seconds). Network conditions are derived by monitoring the client side buffer fullness and the achieved throughput. Low buffer fullness implies network congestion while high buffer fullness suggests sufficient network bandwidth. To ensure good end user experience, the ABR client switches seamlessly from one fragment to another fragment even when the two fragments have different bitrate profiles. ABR video encoding guarantees fragment coding independency to support this seamless transition. This bitrate adaptation mechanism and smooth transition design cope with network congestion effectively and make video delivery over unmanaged network possible.

Just as important is the prevalence of ABR clients on Customer Premise Equipment (CPE) devices in consumer markets. With industry heavyweights Microsoft, Adobe and Apple all supporting ABR streaming, ABR streaming has become the technology of choice for many operators to reach unmanaged consumer devices.

In short, key advantages of ABR streaming are:

- TCP transport and reliable video delivery
- Bitrate and video quality self-adaptation to network conditions
- HTTP and efficient content caching with CDN
- Wide availability of ABR clients

ADAPTIVE STREAMING CHALLENGES IN CABLE ACCESS

Cable operators are under severe competitive pressure as consumers are choosing to view premium content on devices outside of the managed and limited TV experience. Cable operators are moving rapidly in response to these pressures and have begun to leverage the ABR streaming technology to extend video services to unmanaged CPE devices and to off-net subscribers (i.e. subscribers outside provider networks).

Despite the success of ABR streaming in OTT environments, deploying managed ABR based streaming video services at scale in provider networks has a few challenges. The first challenge is bandwidth consumption. ABR streaming may produce multiple bandwidth pressure points throughout cable networks, but the bandwidth bottleneck is most likely to be the cable last mile. In the downstream direction, the TCP based ABR streaming requires unicast transmission, so last mile bandwidth savings derived from efficient multicast delivery is not yet available. Delivering managed linear video services to ABR clients directly via unicast in cable access networks is not attractive, at least from a bandwidth perspective. Optimization of linear video delivery to ABR clients is an active area of research and innovations in the industry.

In the upstream direction, the upstream bandwidth required for ABR video delivery should not be overlooked. The dominating contributing factor of upstream traffic for ABR streaming is TCP Acknowledgement packets (ACKs). With delayed TCP ACK implementation, every two TCP packets downstream must have a TCP ACK packet upstream. This standard-based TCP behavior is defined by IETF RFC 2581. Assuming 60-byte TCP ACK packet size and 1500-byte downstream TCP video packet size, the TCP ACK overhead translates to approximately 2% of downstream bandwidth. For instance, if the total downstream bandwidth consumed by ABR video traffic is 300 Mbps, then 6 Mbps upstream bandwidth is required. A two-way delivery model is replacing the well-known one-way delivery model used in traditional cable video systems.

When one considers that hundreds of subscribers share the last mile upstream bandwidth of 27 Mbps (e.g. 64 QAM Annex B upstream) in a typical cable network today, this TCP ACK overhead is significant. Compounding the problem, cable plants are asymmetric and the upstream spectrum is extremely limited. DOCSIS 3.0 only supports upstream spectrum from 5MHz to 85MHz (about 300 Mbps per service group capacity) while downstream can reach 1 GHz (over 5 Gbps per service group capacity). Though the cable industry is actively exploring solutions to expand upstream capacity [4], the plant upgrade cost, the CPE cost and industry standardization process will limit the upstream capacity for the foreseeable future. Meanwhile, new upstream-intensive applications such as consumer telepresence, home monitoring and automation will significantly strain the DOCSIS upstream path.

In addition to the bandwidth requirements, an equally challenging fact is that ABR streaming video quality is sensitive to network congestion in both upstream and downstream directions. It is not surprising that downstream congestion causes ABR streams to downshift to lower bitrates and to reduce video quality. What about the impact of upstream congestion? Not so obvious is the effect of upstream congestion on ABR video quality. Even when the downstream has sufficient bandwidth to accommodate higher bitrate profiles of an ABR stream, any upstream congestion in the network may reduce TCP throughput and cause the ABR stream to downshift to a lower video bitrate profile and result in inferior video quality. The vulnerability of ABR streams to upstream congestion directly impacts end user experience and reduces bandwidth efficiency of cable access in the downstream direction.

UPSTREAM OPTIMIZATION

To improve upstream transport for ABR streaming, a multi-pronged approach is presented: leveraging upstream channel bonding to increase upstream bandwidth capacity, applying upstream Quality of Service (QoS) to prioritize ABR TCP ACK packets and enabling TCP ACK suppression to eliminate unnecessary TCP ACK packets.

Upstream Capacity Expansion

For deployments of managed ABR video streaming services in cable networks, DOCSIS upstream capacity must be carefully planned to accommodate the additional upstream bandwidth requirement for TCP ACKs. DOCSIS 3.0 introduces both upstream channel bonding and downstream channel bonding. Without channel bonding, an upstream channel capacity is limited to 27 Mbps using Annex B QAM64 modulation. With the latest cable modem channel bonding technology, up to four upstream channels can be bonded together to form a larger upstream pipe. The upstream capacity is expanded to over 100 Mbps, quadrupling the original upstream capacity.

Upstream Quality of Service (QoS)

The sensitivity of ABR video quality to upstream congestion makes upstream QoS critical to ABR streaming video. Given ample downstream bandwidth, operators desire to obtain the best ABR streaming video quality and minimize the impact of upstream congestion to video quality. The proposed design gives upstream TCP ACKs of ABR streaming higher priority over non real-time upstream traffic by applying standard based DOCSIS QoS. With DOCSIS, different priorities can be applied to service flows. TCP ACKs from ABR video can be classified to higher priority service flows to receive differentiated delivery service in the upstream direction.

Proof of concept work was conducted to evaluate the impact of QoS on ABR streaming video. In this proof of concept test, a Microsoft Smooth ABR video stream with two bitrate profiles (2.1 Mbps and 6 Mbps) was delivered over a DOCSIS 3.0 access network. The downstream peak capacity was about 300 Mbps with an 8-QAM (Annex B 256 QAM) bonded channel. The upstream peak capacity was about 100 Mbps with a 4-QAM (Annex B 64QAM) bonded channel. The upstream was congested by generated network traffic. The experiment recorded both bitrate profiles requested by the ABR client and ABR video fragment (chunk) download time.

The results of non QoS-assisted best effort delivery and QoS-controlled delivery are compared in Figure 2. In the first case, all traffic was delivered best-effort without special QoS treatments. The ABR stream under test was only able to reach 2.1 Mbps. In the second case, TCP ACKs of ABR streaming were given higher QoS priority than other high speed data traffic. This time, the ABR stream reached 6 Mbps. As demonstrated, in the situation of congested upstream and no upstream QoS, only the lower bitrate profile and the lower video quality were achieved even if downstream bandwidth was abundant. When QoS and higher delivery priority were applied to upstream TCP ACKs, ABR video delivery achieved the higher video quality.

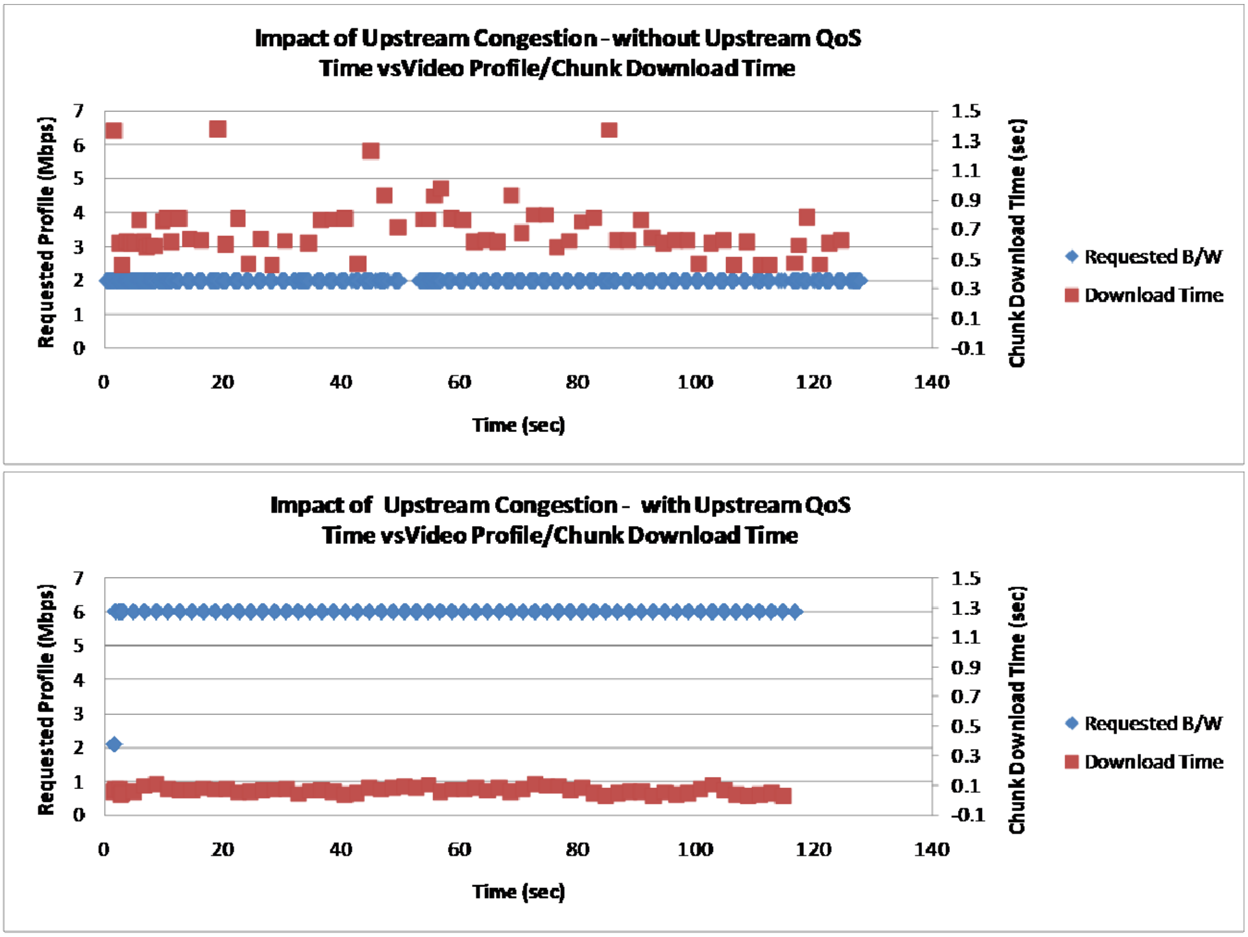


Figure 2. Applying QoS to Upstream TCP ACK

It is also interesting to observe the chunk download time in both cases. The chunk download time is the time a network takes to deliver an entire video fragment. Each fragment in the example was 2 seconds in video playing time. The size of the video fragment in the second case was larger than that in the first case because of the higher bitrate profile. Still, the chunk download time in the second case was much smaller compared with that in the first case. The reason has to do with the way video fragments are delivered in ABR streaming.

Unlike traditional video delivery, an ABR video fragment is always delivered at the current data rate of the network regardless of the bitrate profile of the fragment. In other words, a fragment with a bitrate profile of 6 Mbps will be delivered at 300 Mbps if the network throughput is 300 Mbps for the stream. Due to this bursty nature of ABR video delivery, the chunk download time is inversely proportional to the TCP throughput. The higher the TCP throughput, the less time is needed for a fragment to download. In the first case, upstream congestion and TCP ACK packet loss significantly reduce the TCP throughput; therefore, the chunk download time is longer. In contrast, QoS prioritization allows much higher TCP throughput in the second case. The increased TCP throughput and shorter chunk download time cause the profile adaptation mechanism of the ABR client to choose a higher bitrate profile. Since ABR streaming supports multiple TCP sessions per video stream, the TCP throughput refers to the aggregated TCP throughput for the stream.

This proof of concept study clearly demonstrates the importance of upstream QoS in delivering better quality of ABR video streams and utilizing downstream bandwidth more efficiently.

ACK Suppression

For a better understanding of TCP ACK suppression, DOCSIS upstream delivery is briefly reviewed here. In order to transmit a packet upstream in a DOCSIS network, a cable modem must request bandwidth from a CMTS. The CMTS then grants the bandwidth and schedules the packet delivery. The cable modem waits for its scheduled timeslot before it transmits the packet. This cycle is referred to as the request-and-grant cycle. Without any optimization, TCP throughput is limited by the request-and-grant cycle, because the modem can send only a single TCP packet in the upstream direction for each request-and-grant cycle.

One optimization technique to improve upstream throughput in DOCSIS is called concatenation. Concatenation allows a cable modem to combine multiple upstream packets in a single upstream transfer. Concatenation is applicable to all traffic types and is not limited to TCP traffic. The concatenation becomes even more efficient with DOCSIS 3.0 upstream channel bonding, when continuous concatenation is used and the concatenation occurs at sub-packet boundaries. All small packets, including ABR ACKs, benefit from concatenation and the improved upstream efficiency.

The other optimization technique to improve upstream bandwidth efficiency is specific to TCP traffic and is called TCP ACK suppression. In TCP transmission, each ACK packet contains an acknowledgment number acknowledging the last contiguous byte received successfully. All prior bytes are considered acknowledged. ACK suppression takes advantage of this cumulative nature of the TCP acknowledgement scheme and removes unnecessary TCP ACKs in the upstream direction. In the example shown in Figure 3, when the cable modem receives the first TCP ACK from the CPE, it sends a request for bandwidth equivalent to one TCP ACK. When the grant arrives, the cable modem already has three ACKs from the same TCP flow queued up. It is only necessary for the cable modem to send ACK#3 to acknowledge the receipt of all three TCP ACK packets. This scheme of sending only the last TCP ACK in the queue decreases the bandwidth consumption in the upstream direction.

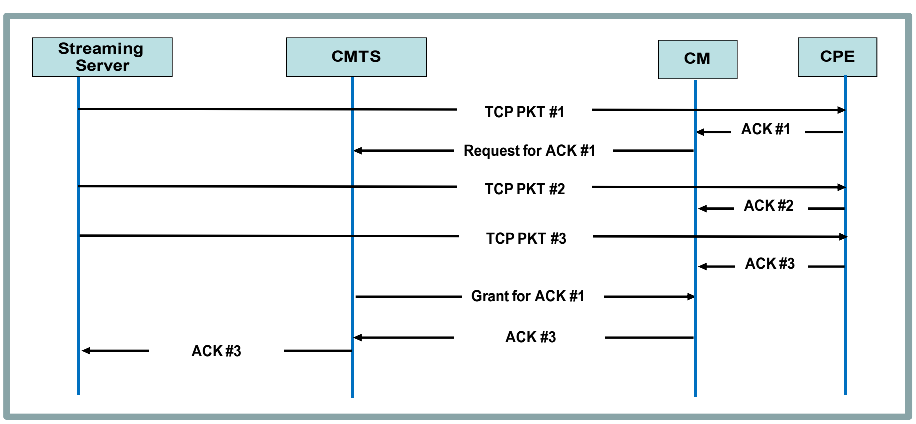


Figure 3. Upstream TCP ACK Suppression

TCP ACK suppression, when enabled in cable modems or residential gateways, has the potential to reduce upstream bandwidth required by ABR streaming. From Figure 3, it is apparent that the effectiveness of TCP ACK suppression is highly dependent on how fast TCP ACK packets are accumulated. In the above example, there are three ACK packets accumulated in the DOCSIS request-and-grant interval. Only one out of three ACK packets needs to be sent upstream. Therefore, an ACK suppression rate of 67% is achieved. At the other end of the spectrum, if there is only one ACK packet arriving in the request-and-grant interval, the ACK suppression rate will be zero.

To quantify the effectiveness of TCP ACK suppression, an experiment was carried out. An ABR video stream (Microsoft Smooth) was delivered over a DOCSIS 3.0 network with controllable available bandwidth. The efficiency of TCP ACK suppression was measured as the percentage of TCP ACKs that were suppressed by the cable modem. Traffic was captured by a Wireshark at both the client side and at the server side. The Wireshark capture was then analyzed to identify the suppressed ACK packets. From the results shown in Figure 4, it is clear that the efficiency of TCP ACK suppression is directly related to the available downstream bandwidth. When the downstream available bandwidth is 10 Mbps, the ACK suppression efficiency is only 30%. However, if the available downstream bandwidth is higher, namely 50 Mbps, the ACK suppression efficiency is 70%.

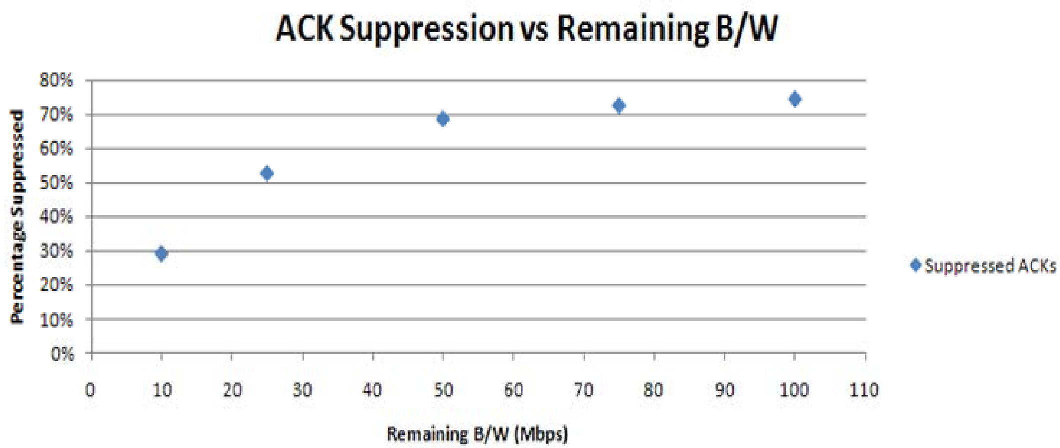


Figure 4. TCP ACK Suppression Efficiency

Taking a closer look at how ABR streams are delivered will provide insights into the ACK suppression efficiency results. ABR video fragments are delivered via HTTP/TCP. The fragment delivery speed is proportional to the TCP throughput of the transport. The faster the fragments are delivered, the more TCP ACKs are accumulated in each request-and-grant interval and thus the better TCP ACK suppression efficiency. Based on the test results above, it is clear that TCP ACK suppression is more effective when downstream bandwidth is higher.

In service provider networks, however, the TCP ACK suppression efficiency is further complicated by other factors. First, end-to-end network conditions may also impact TCP throughput. The proof of concept work described earlier is simplified to only consider the last mile cable access as the bottleneck for ABR video delivery. This is likely the case when CDN and content edge caching are used to facilitate the ABR video delivery. However, if edge caching is not involved, the delivery bottleneck can be anywhere in the network. The bottleneck reduces the end-to-end TCP throughput and can negatively impact TCP ACK suppression efficiency. Furthermore, when multiple ABR clients compete for last mile bandwidth, the available bandwidth for each client decreases. Consequently, the TCP ACK suppression efficiency can be reduced due to the lowered TCP throughput. Since the average download speed of the ABR fragments varies greatly throughout each day, it is difficult to accurately quantify the overall impact of ACK suppression.

DOWNSTREAM OPTIMIZATION

Two downstream optimization approaches are proposed. The first method lowers bandwidth consumption in the downstream direction by utilizing more efficient video encoding. The second approach optimizes bandwidth distribution under the constraint of existing network capacity to obtain best end user experience.

Variable Bitrate Video and ABR Streaming

The bandwidth advantages of Variable Bit Rate (VBR) video encoding over Constant Bit Rate (CBR) video encoding is well understood. With comparable video quality, VBR encoding saves 40%-60% bandwidth over CBR video [5]. However, VBR is typically used only in broadcast service in traditional cable video delivery. Since traditional cable video is transmitted over a narrow bandwidth pipe (e.g. 38 Mbps Annex B QAM channel), delivering VBR video requires MPEG statistical multiplexing to squeeze multiple VBR streams into the constant bandwidth pipe. Multiple drawbacks of MPEG statmuxing, such as high cost, long latency and video quality degradation, have prevented the adoption of VBR video in narrowcast video services such as switched video and on-demand video.

DOCSIS 3.0 video delivery introduces a new VBR delivery model by providing a wideband transport pipe and network statmuxing [6]. Wider pipes with DOCSIS 3.0 channel bonding and skinner streams with advanced video coding eliminate the need of MPEG statmuxing. A large number of VBR streams are statistically multiplexed naturally and efficiently by the DOCSIS transport. To address concerns about guaranteed VBR video delivery via DOCSIS network statmuxing, several methods have been proposed [6]: mixing VBR video and best effort HSD data in a converged DOCSIS pipe, applying VBR admission control and error retransmission. Despite these improvements, VBR video delivery using unreliable transport such as UDP/RTP over DOCSIS is still considered by many as unguaranteed delivery.

As ABR streaming is gaining traction in managed video services, new possibilities to maximize the potential of VBR bandwidth savings are on the horizon. The use of ABR streaming and VBR encoding not only solves VBR network delivery reliability issues, but also enhances the ABR streaming with improved bandwidth efficiency. In a typical ABR streaming implementation, each video quality profile is explicitly signaled using bitrate value, such as 500 kbps, 1 Mbps, 2 Mbps streams. If we consider these bitrates as average bitrate instead of constant bitrate, VBR encoding can be applied to each video quality level instead of CBR encoding (Figure 5). VBR is a superior encoding choice as it naturally keeps the video quality constant while varying the encoding bitrates.

Although typical ABR video implementations today have clients signal absolute bitrate values to identify video quality profiles, the video quality levels can alternatively be represented using relative terms. In the upcoming MPEG Dynamic Adaptive Streaming over HTTP (DASH) standard [7], relative quality ranking is used instead of absolute bitrate and VBR can be supported easily.

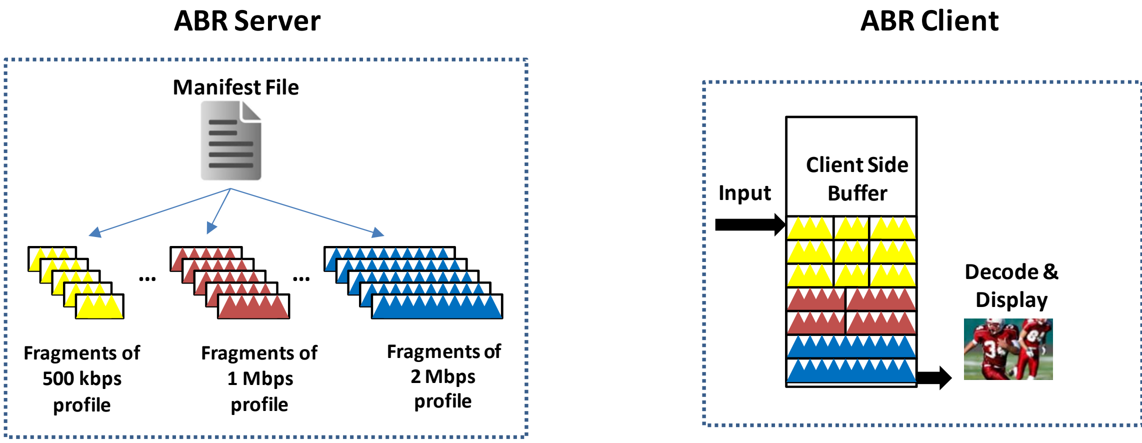


Figure 5. ABR Streaming with VBR Encoding

Thanks to TCP delivery and ABR streaming, VBR bitrate spikes are now handled gracefully and VBR delivery is guaranteed. In UDP based cable IPTV systems, the video delivery rate in the last mile is the same as the video consumption rate at the client side decoder. Namely, an 8 Mbps HD stream is delivered by a server at 8 Mbps to a client, excluding allowance of limited network jitter in the order of 100 ms. In contrast, ABR video is delivered at the maximum speed the access network permits. An ABR video fragment at 2 Mbps bitrate profile could be bursted at a speed of 100 Mbps in the last mile. ABR video fragments are downloaded to ABR clients, just like files. Therefore, the bandwidth variation of VBR streams is masked out by this fragment download operation and is no longer critical. In fact, it is the size or the average bitrates of video fragments that matter to the network. A more efficient encoding will yield a smaller size for each fragment. When fragments are smaller in size, they in turn consume less bandwidth and contribute to higher bandwidth utilization.

The powerful combination of ABR and VBR makes it possible to reap the bandwidth benefits of VBR video without dealing with the complexity of VBR rate variations in transmission pipes.

Downstream QoS

Given that ABR streaming adapts bitrate profiles to network conditions, are QoS and network controls still necessary? While ABR streaming provides a technique for OTT video providers to overcome bandwidth congestion issues in unmanaged networks, more can be done for service providers who own the network infrastructure. Taking advantages of network control to provide superior video experience to on-net subscribers is a key differentiator for service providers.

DOCSIS access networks provide two types of traffic delivery mechanisms. One is the Committed Information Rate (CIR) service with guaranteed delivery. The other type is the Best Effort (BE) delivery. A DOCSIS service flow can be delivered by means of CIR, BE, or a combination of both. In general, traffic without bandwidth guarantee and admission control is opportunistic traffic.

Although ABR video does not require CIR protection, there is no guarantee of video delivery quality. Network congestion may force an HD ABR client to receive video at a low-bitrate profile, e.g. 1 Mbps, even if higher bitrate profiles are offered. Usually, such a low bitrate stream may have unacceptable video quality for HD devices. Delivering content with unacceptable quality not only negatively affects users' perception of the quality of an operator's video service, but also wastes precious network bandwidth. It is beneficial to deliver ABR video with a CIR floor to protect the video with minimum acceptable quality. This minimum acceptable video quality or bitrate should be directly related to the client screen size and content type. If the network is not able to provide even the bare minimum acceptable video quality, the new ABR video request should be rejected by admission control and proper feedbacks can be sent to clients. Meanwhile, the design allows the ABR video to move to higher bitrate profiles when network bandwidth is available for better video quality. The ABR traffic beyond the CIR protection threshold is treated as BE traffic in DOCSIS access.

Opportunistic video traffic can be further optimized in cable access. DOCSIS access is designed to provide different QoS priorities to different types of traffic. Providing higher priorities to ABR streaming video over other types of non-realtime traffic is straightforward and can be easily accomplished. Marking managed ABR streaming traffic with a different DSCP value and then applying a higher DOCSIS priority to it in the last mile will prioritize managed ABR streaming against other traffic during network congestion. What is more complicated is traffic prioritization among ABR clients themselves. Research [8] shows that when two ABR clients competing for available bandwidth, the bandwidth allocation between the two clients are not deterministic and the video quality profile oscillates. Future research involving a larger number of ABR clients may shed additional light into the client behavior. More interesting questions are: What is the desired bandwidth allocation among competing ABR clients? How can the network facilitate better bandwidth allocation?

To obtain superior end user video experience, not all ABR clients should be treated equally. For instance, an ABR client is serving video content to an HD device at a video bitrate profile of 4 Mbps. Another ABR client is serving video content to a SD device at the same video bitrate profile of 4 Mbps. Downshifting the HD ABR client to a lower bitrate profile (e.g. 2 Mbps) will negatively impact user experience while the effect of downshifting the SD client to the next bitrate profile is less noticeable. Under network congestion, it is desirable to provide higher priorities to ABR streams at bitrate profiles critical to end user viewing experience.

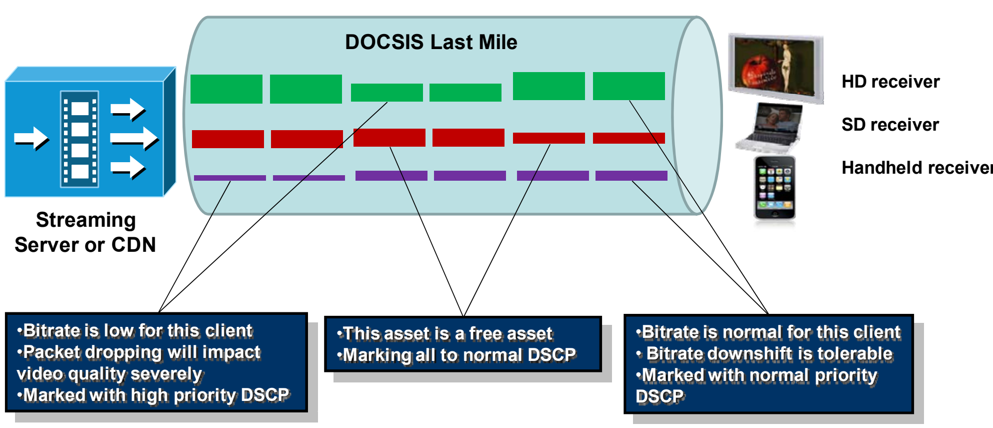


Figure 6. ABR Stream Prioritization in Downstream

The proposed QoS scheme is illustrated in Figure 6. When an ABR client requests a video segment via HTTP GET, it indicates to the network that it is operating in a normal or a critical condition through DSCP marking of the TCP traffic. An ABR client running in a bitrate profile critical to end user video experience is in a critical condition. Traffic with different markings can be sent using different TCP sockets. The DOCSIS network just provides a higher priority to traffic marked in critical conditions to minimize the impact of video quality degradation. Taking this prioritization scheme one step further, operators can potentially apply business rules. For example, free assets and free service may not be allowed to use the critical condition priority. The proposed scheme prioritizes ABR streams among ABR clients and optimizes end user viewing experience.

SUMMARY

The next wave of video entertainment is coming our way. With technology advancements in adaptive streaming, video content can be delivered to any device, on-net or off-net, managed or unmanaged. Adaptive streaming is becoming a new tool in the service providers' toolkit to deliver advanced video services. To overcome some of the technical challenges of scaled ABR streaming deployments in managed networks, a variety of optimization techniques can be applied to upstream and downstream DOCSIS networks. From upstream optimization of simple expansion of bandwidth, TCP ACK prioritization and TCP ACK suppression, to downstream optimization of VBR coding, stream protection and stream prioritization, these techniques enable cable operators to unleash the full potential of adaptive streaming and pave the way for future video services.

ACKNOWLEDGEMENTS

The author appreciates Jonathan Muslow's great proof of concept work. The author also thanks Li-shen Chin for his help on ABR stream analysis tools and his team's earlier work in the area of ACK suppression.

REFERENCES

[1] Cisco Visual Networking Index, June 2010 http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf

[2] Sandvine 2010 Fall Global Internet Phenomena report, http://www.sandvine.com/news/global_broadband_trends.asp

[3] A. Begen, T. Akgul and M. Baugher, "Watching video over the Web, part 1& 2: streaming protocols," IEEE Internet Comput., vol. 15/2 & 15/3, 2011

[4] J. T. Chapman, "Taking the DOCSIS Upstream to a Gigabit per Second", NCTA Spring Technical Forum 2010

[5] S. J. Huang, "Principle, Applications of Variable Bit Rate Coding for Digital Video Broadcasting, w. Statistical Multiplexing Extension", NAB 1999

[6] X. Liu, A. Bernstein, "Variable Bit Rate Video Services in DOCSIS 3.0 Networks", NCTA 2008

[7] Dynamic Adaptive Streaming over HTTP, http://mpeg.chiariglione.org/working_documents/mpeg-b/dash/dash-dis.zip

[8] S. Abhshabi, A. Begen, C. Dovrolis, "An Experimental Evaluation of Rate-Adaptation Algorithms in Adaptive Streaming over HTTP", Multimedia System, Feb, 2011

ADOPTION TRENDS OF OVER-THE-TOP VIDEO FROM A NETWORK PERSPECTIVE

Matt Tooley
Sandvine Incorporated

ABSTRACT

The video content delivery industry is in a state of flux due to the changing business models for delivering video to the consumer. The traditional business models used by the cable companies are being challenged by new models for delivering the content due to technology innovations in consumer devices that have been fueled by the wide-spread adoption of both wireline and wireless broadband services.

This paper looks at the effects of technological and market advances on subscriber usage behavior and network traffic.

INTRODUCTION

The adoption of broadband service with DOCSIS over the past ten years ignited a market for the delivery of video content using these same broadband networks. The wide spread adoption of broadband service, the ever increasing speeds of service, and the growing ubiquity of Wi-Fi enabled devices capable of receiving and playing video, has created a tidal wave of demand for content from subscribers.

For this paper, a study was performed to look at trends in over-the-top video and its impacts on North American wire-line broadband networks. The study included an examination of the trends in broadband connectivity speeds in North America, and of data collected over the last four years by Sandvine as part of its Global Internet Phenomena report in order to identify the effects of the changes in broadband connectivity speeds. Finally, the research also included a case study as an example of the type of on-line viewing events that have been enabled with the advances in technology.

In brief, the study showed

- Broadband speeds in North America continue to trend up
- The growth and adoption of over-the-top video correlates with the increases in the broadband connectivity speeds
- A growing trend towards alternate platforms for viewing video

This paper provides a detailed analysis and summary of the findings.

YEAR OVER YEAR TRENDS

Average Measured Connection

To understand what is enabling the growth of over-the-top video traffic, it is important to look at the average broadband connection speed. Akamai's *State of the Internet* which is published quarterly provides a snapshot of the measured average speed of Internet connections, by country on a global scale. As shown in Figure 1, the average connection speed in North America has been going up since 2008. The speed trend in North America roughly aligns with the wide spread offering of DOCSIS 3.0 by the North American cable operators.

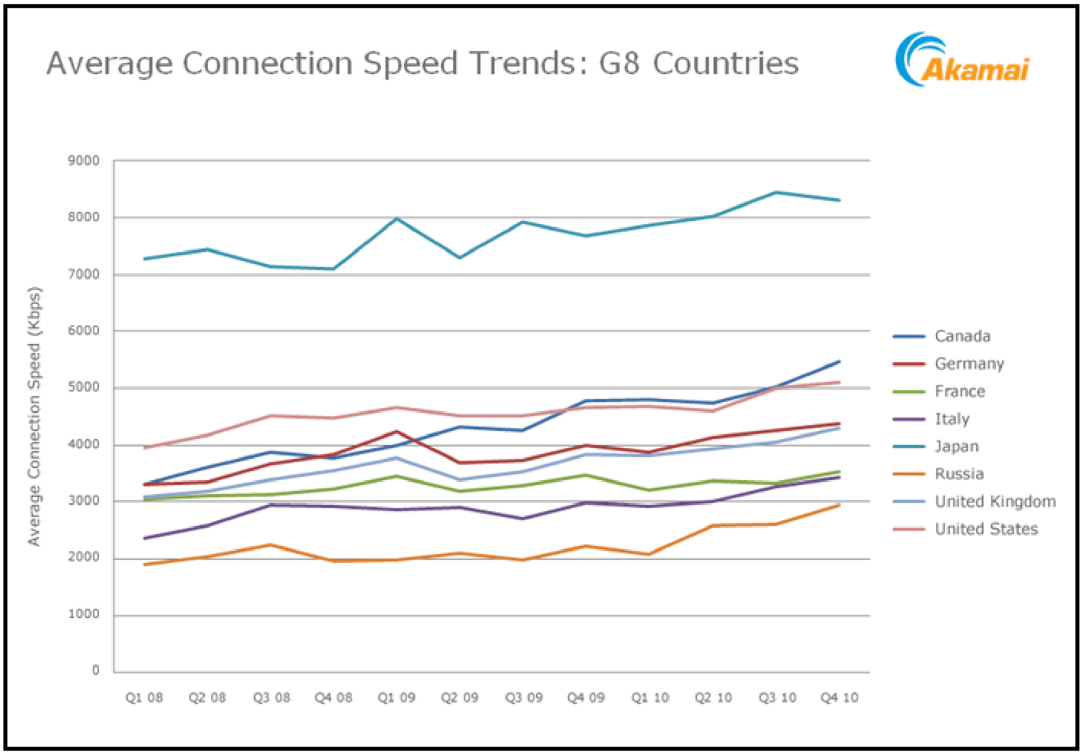


Figure 1: Average Connection Speed Trends: G8 Countries

Period	United States - Average Measured Speed
Q4 – 2008	3.9 Mbps
Q4 – 2009	3.8 Mbps
Q4 – 2010	5.1 Mbps

Table 1: Average Measured Connection Speed for Select Quarters[1]

Adoption of New Computing Devices

Over this same period, subscribers began adopting additional computing devices to access content using their broadband connection. The Apple iPhone was released June 29, 2007[2] followed by the release of the 3G version a year later in June 2008. Around the same time other manufacturers released similar smartphones. The wide spread adoption of smartphones led manufacturers to release tablet computers with a look and feel similar to smartphones. Apple released the first generation iPad in June 2010[3] which was quickly followed by similar releases from other manufacturers.

Subscriber Usage Behavior

What is the effect of the increases in broadband connectivity speeds and device choices? The key change is a shift in subscriber usage behavior. Sandvine conducts and publishes annually its *Global Internet Phenomena report*[456], wherein it looks at trends in subscriber usage behavior with respect to applications. The report is global in scope, and breaking down trends and usage by subscribers on a regional level. For this paper, data was gathered from the 2008-2011 studies with a focus on North American subscribers and content being delivered to the subscriber (which is commonly referred to as in the downstream direction). The report summarizes and categorizes the applications. The data for this study was grouped by application genres to illustrate the subscriber usage trends as it relates to over the top video. The content category definitions are as described in Table 2.

Content Category	Definition
Web	Web browsing protocols and specific websites such as HTTP
P2P	File-sharing applications and protocols that rely on peer-to-peer network models for content distribution such as BitTorrent, Gnutella, and eDonkey
Real-time Entertainment	Application and protocols that allow "on-demand" entertainment that is consumed as it arrives such as Flash video, Slingbox, PPStream.
Other	All other protocols

Table 2: Content Categories and Definitions

Table 3 and Table 4 show the aggregate traffic composition for each year, for the top four categories. The data show constant a year-over-year decline in web traffic, and a trending decline in P2P traffic. The declines in the web traffic and P2P traffic are matched by similar increases in real-time traffic. These downward trends in web and P2P traffic and the corresponding increase in real-time traffic illustrates the continued shift from a reliance on "download now, use later" content acquisition to a on-demand mentality where bytes are consumed as they arrive.

	2008	2009	2010	2011
Web	57%	39%	20%	19%
P2P	19%	16%	19%	13%
Real-time Entertainment	10%	29%	43%	53%
Other	14%	17%	18%	14%

Table 3: Aggregate Traffic Composition by Category

	2008	2009	2010	2011
Web		-19%	-19%	-1%
P2P		-3%	4%	-6%
Real-time Entertainment		18%	14%	11%
Other		3%	1%	-4%

Table 4: Year-Over-Year Relative Change

The overall usage is interesting in itself, but what is important to the network engineers is the distribution of the usage during the peak hour. The figures below show the downstream traffic profiles for 2009-2011. Examination of the traffic profiles again shows the shift to real-time entertainment, in particular during the "peak hours". There has been continued expansion of the window of peak hours as well as "how" subscribers use the network during peak hours. In 2009, the majority of the traffic during the peak hours was from web traffic and the second largest category of traffic was real-time entertainment. 2010 shows both a dramatic increase in the overall percentage of traffic that is categorized as real-time entertainment as well as the expansion of the peak hours.

This trend continues into 2011, with real-time entertainment continuing its expansion both in time and usage. The other trend that can be observed in the data is the change in the ratio of the peak-to-trough. From 2009 to 2011 we see the peak-to-trough increase; further evidence of a shift from "download now, use later" to on-demand. When P2P traffic was dominant, the peak-to-trough ratio was less due to the around-the-clock nature of P2P downloads. The shift to an on-demand mentality is causing the peak-hours to get busier and the off-hours to be less busy.

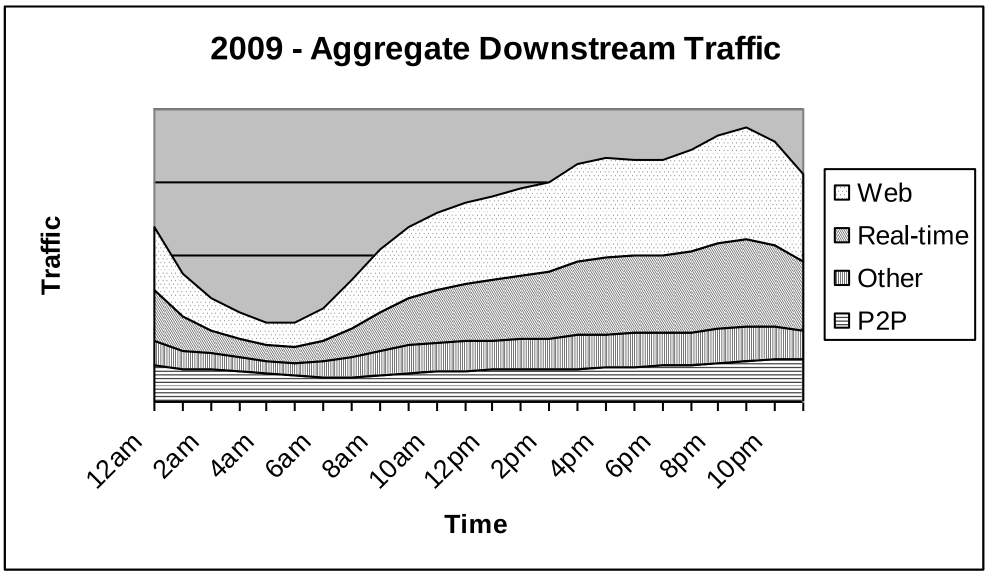


Figure 2: 2009 - Downstream Aggregate Daily Traffic

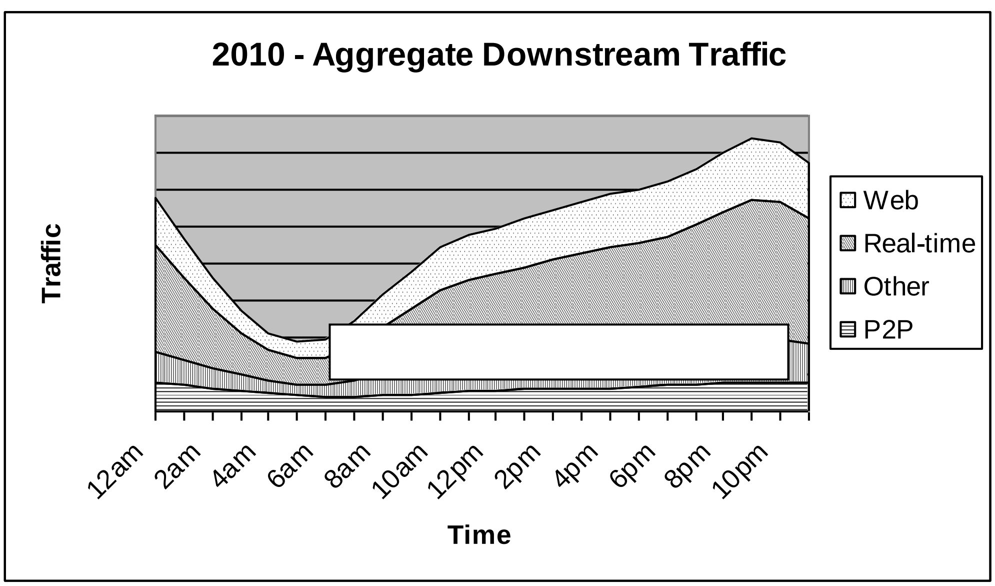


Figure 3: 2010 - Downstream Aggregate Daily Traffic

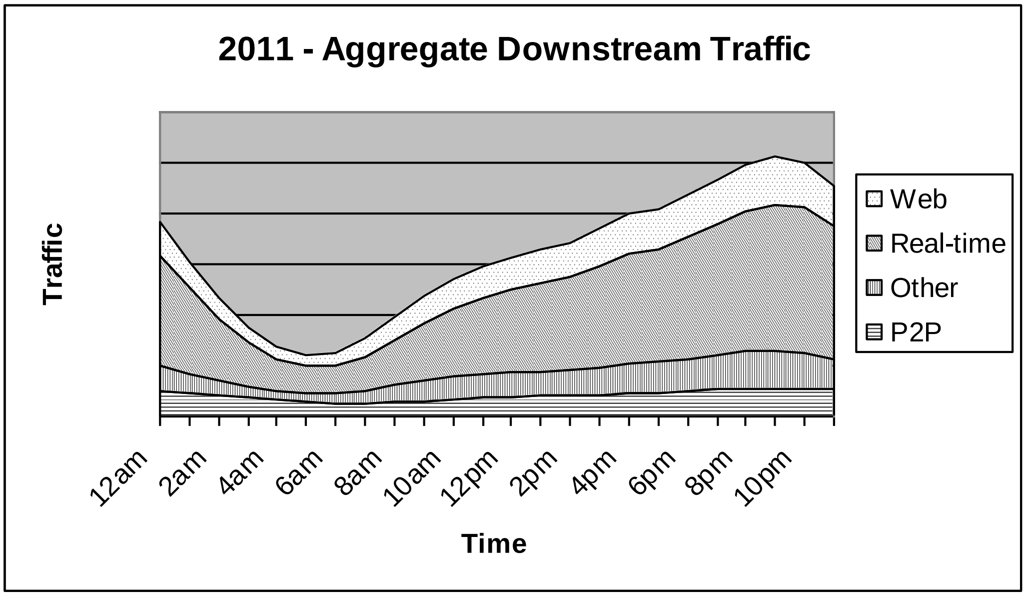


Figure 4: 2011 - Downstream Aggregate Daily Traffic

These drivers of "peak hour" are illustrated in Table 5. Downstream traffic during the peak period over the last three years is comprised increasingly of real-time entertainment.

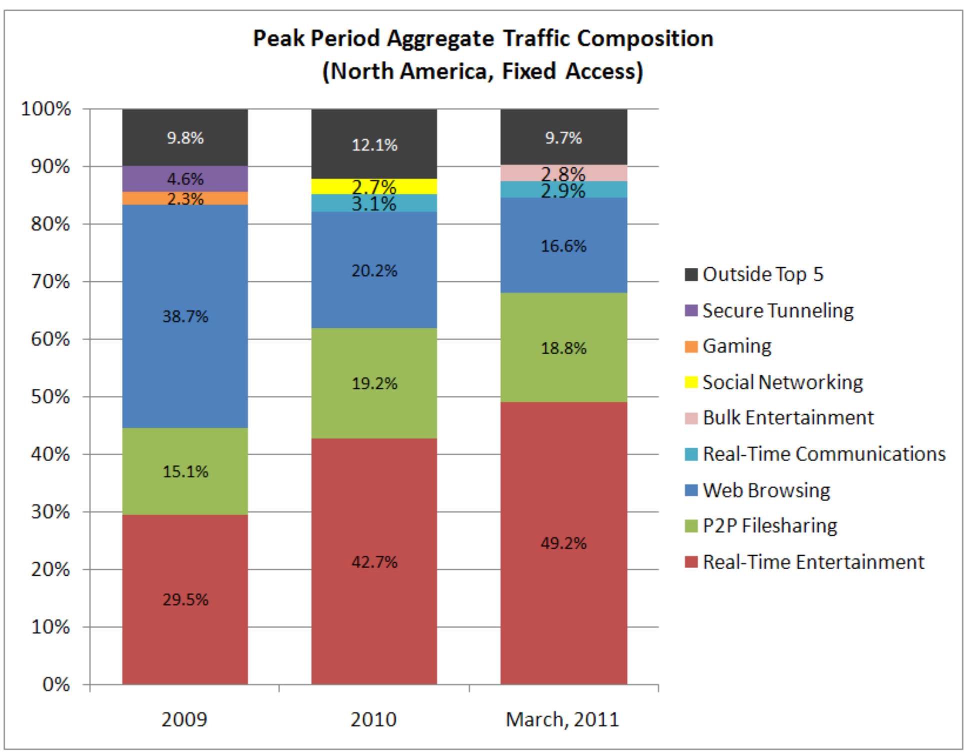


Table 5: Peak Period Aggregate Traffic Composition - North America, Fixed Access

CASE STUDY

Increases in broadband connectivity speed and the introduction of multiple platforms for receiving and viewing broadband content has fueled over-the-top video events, whether as on-demand content such as movies or live appointment content such as sporting events. For this paper, the 2011 NCAA Men's Basketball tournament served as a case study in order to explore possible implications of these technological and market changes for network traffic management.

Media coverage of the basketball tournament, as in year's past, included a live video stream, referred to as NCAA® March Madness® on Demand (MMOD) carried on the Internet by Turner Sports, CBS Sports and the NCAA. MMOD delivered a 63% increase in total visits compared to 2010 as well as 13.7 million total hours of streaming video consumed through iPad and iPhone applications, representing a 17% increase over 2010[7].

Simulcasting on the Internet and broadcast television provided a unique opportunity to measure subscriber preferences. For this study, the network probe was programmed to measure and record the number of concurrent video streams, total usage per subscriber, and the type of device the used to access the data. The data was collected from March 25-28 during the early rounds from a set of sites at residential, broadband wire-line with a sample size of 450,000 subscribers. The data collected provides a snapshot of how residential broadband subscribers used their service within their homes during a live media event also available via legacy television systems (CATV, broadcast, satellite).

Figure 5 graphs the aggregate number of minutes of streamed video during each 30 minute sample periods over the course of the case study. The peaks in the minutes watched closely align with when the games were being streamed live by the MMOD service. Interestingly enough, the number of minutes did not go to zero between games. One possible explanation for this is that viewers were utilizing this time to view the available video highlights and the option to rewatch the games.

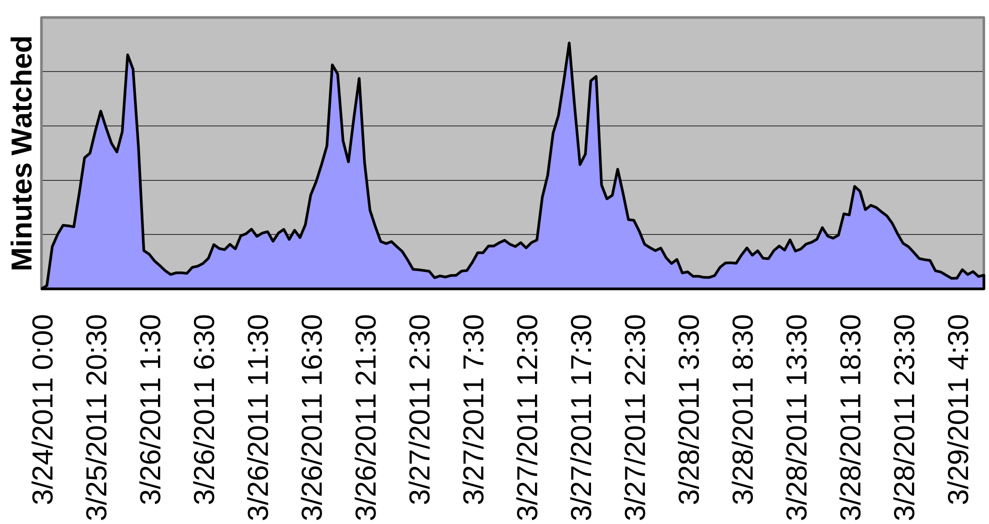


Figure 5: Minutes Watched

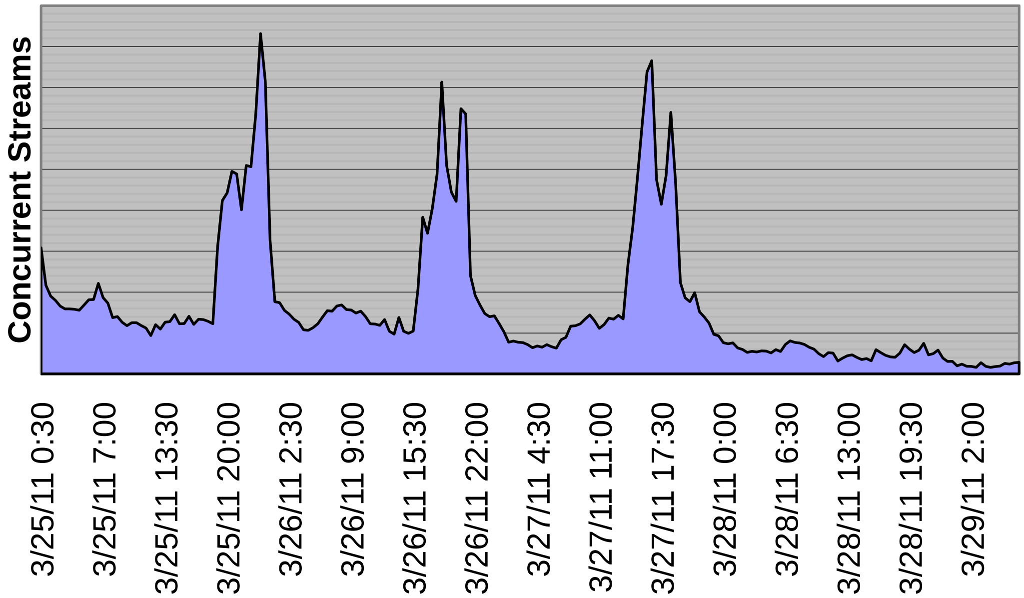


Figure 6: Concurrent Video Streams

Figure 6 shows the number of concurrent streams for the same period of time. Again the peaks align with the simulcast of the live games. Assuming one video stream per subscriber, this then shows that during the peaks 3-4% of the subscribers were accessing the MMOD service with their residential fixed broadband service.

As part of the study, the device type used to access the MMOD service were recorded and categorized as follows:

Figure 7 shows the number of views grouped by device category. For the residential subscriber, the device of choice when the games were being played was the desktop machine. As shown in Figure 8, between games there is increased usage of alternate devices such as the iOS and Android powered devices such as smartphones and tablets that have Wi-Fi capabilities.

Category	Devices
Desktop	Window PCs, Apple Macintosh, and Linux
iPhone/iPad	All iOS devices – iPhone, iPad, iTouch, iPod
Blackberry	All devices running a Blackberry OS based browser
Android	All devices running any version of the Android OS
Playstation	All Playstation devices

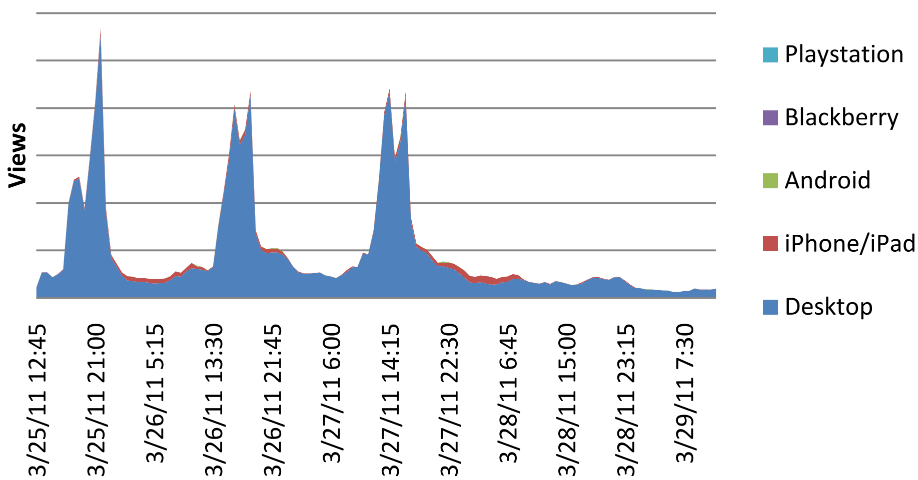


Figure 7: Views by Device Category

Of interest in the graphs is increased usage of devices other than the desktop computer in valleys in Figure 7. Figure 9 and Figure 10 show in greater detail the first valley from 1:00 AM on March 26th until about 2:00 PM on the March 26th. A closer examination of these valleys shows that during the periods when the games were not being transmitted live that there is an increase in the usage of the non-desktop device for accessing the MMOD content. During the period between games, subscribers opted to substitute the use of device with a smaller, lower resolution screen such as a Smartphone (iPhone) or Tablet (iPad). One possible explanation for the increase usage of iPhone/iPad between games could be that the subscribers were watching highlights from the earlier games or "snacking" on video chunks on the smaller screens.

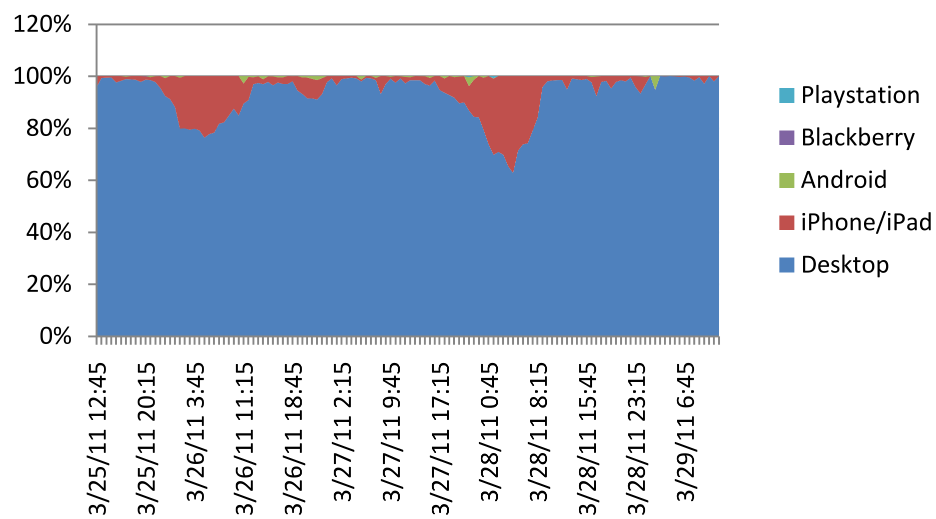


Figure 8: Relative Views by Device Category

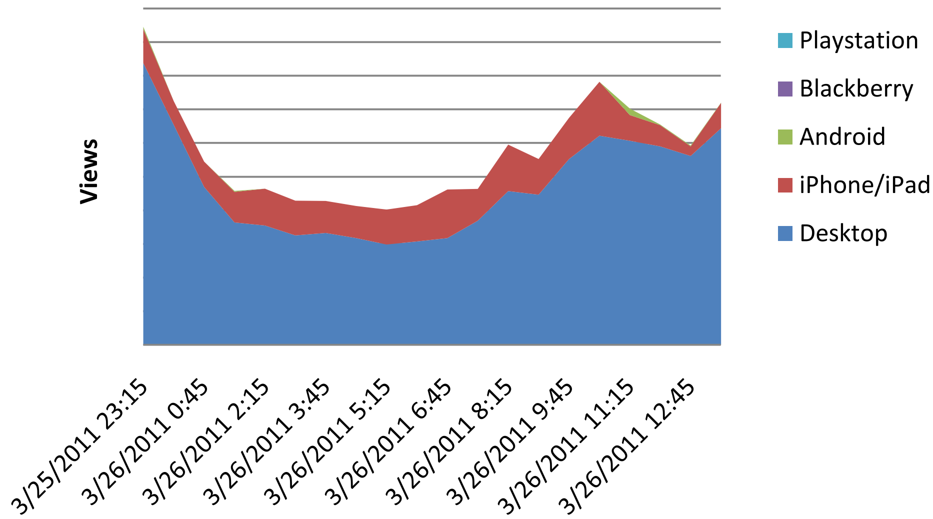


Figure 9: Views by Device Type March 25-26th

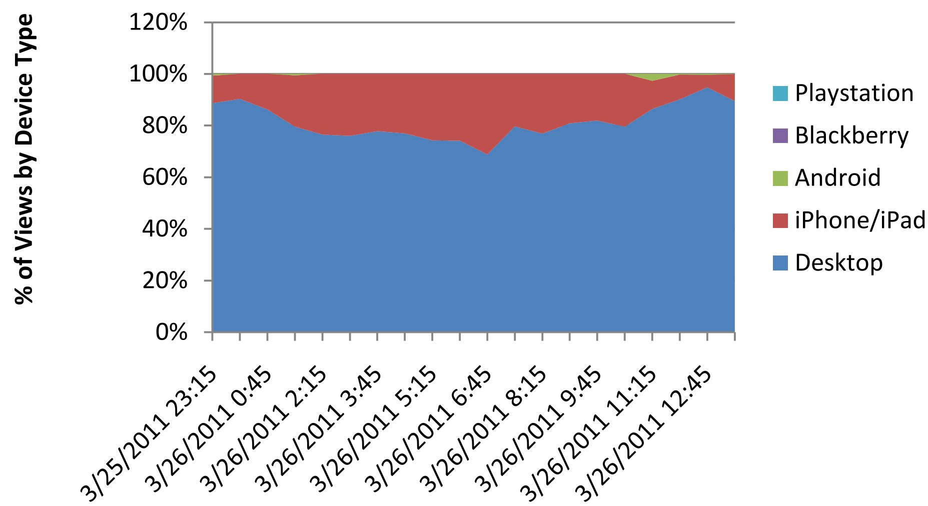


Figure 10: Relative Views by Device Type March 25-26th

The case study discussed here illustrates the effects of the adoption of smartphones and tablets on subscriber behavior and network traffic. Peak network traffic is still aligning with the live event, and during the periods between events, subscribers are filling this time with the use of a smartphone or tablet to watch on-demand content. The net effect of this is both an increase in the peak traffic as well as the overall expansion of peak traffic period due to traffic from the smartphones and tablets.

CONCLUSION

Although not exhaustive, this paper provides some interesting insights into trends with regards to subscribers and their online video viewing behavior. One possible conclusion from this study is, as broadband connectivity speeds continue to increase, this will continue to amplify the shift of traffic to peak periods as well as the expansion of the peak periods. The peak-to-trough ratio should continue to grow unless subscribers are incented to shift their behavior.

In summary:

- Broadband speeds continue to rise;
- Enabling a shift from "download now, use later" to on-demand for content;
- Peak-to-trough ratio for traffic is growing;
- And peak periods are expanding as subscribers use Wi-Fi enabled devices to access content on the shoulders of the peak hours.

---

[1] "The State of the Internet"; Akamai; Q4 2008, Q4 2009, Q4 2010; http://www.akamai.com/stateoftheinternet/

[2] "History of the iPhone", Wikipedia, http://en.wikipedia.org/wiki/History_of_the_iPhone

[3] "iPad", Wikipedia, http://en.wikipedia.org/wiki/IPad

[4] "2010 Global Internet Phenomena Report", Sandvine, http://www.sandvine.com/news/global_broadband_trends.asp

[5] "2009 Global Broadband Phenomena Report", Sandvine, http://www.sandvine.com/downloads/documents/2009%20Global%20Broadband%20Phenomena%20-%20Full%20Report.pdf

[6] "2008 Global Broadband Phenomena Report", Sandvine, http://www.sandvine.com/downloads/documents/2008%20Global%20Broadband%20Phenomena%20-%20Full%20Report.pdf

[7] "2011 NCAA March Madness on Demand Sees 63% Increase in Total Visits and 17% Increase in Video Consumption across Multiple Platform for the NCAA Division I Men's Basketball Championship"; Turner Sports, CBS Sports, and NCAA press release April 5, 2011 on Turner Sports web site, http://news.turner.com/article_display.cfm?article_id=5636