THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:

NCTA
Technical
Papers

"My (TV) Space" – Using Subscriber Management Systems to refine and redefine video-based social networking

Andrew Poole
Joe Matarese
ARRIS Group, Inc.

Abstract

Social networking sites like MySpace and Facebook have enabled on-demand user-generated content to millions of viewers. In this paper we ask whether it is possible to leverage social-networking web technology, user-generated video, and cable infrastructure to provide a "best of all worlds" experience for these millions of viewers.

INTRODUCTION

Glass Houses

Social networking sites have enabled, for better or sometimes worse, unfettered access to as much personal information as individuals dare make public. When Warhol uttered the line about everyone being famous for 15 minutes, he appropriately conjugated the verb with a "will" rather than a "may". One feels that it is fast becoming impossible not to be famous for 15 minutes. In this "glass house" environment, cable potentially has a role in allowing subscribers to selectively draw the shades. Subscriber Management Systems (SMSs) have long served as a means to provision specific services to subscribers. Such services typically include high speed data tiers and subscription VOD packages. Why not take this a step further and allow subscribers to act upon service and account information so as to offer personal content in a more selective manner?

The TV is King.

A further aspect of MySpace and Facebook video-sharing is that the display mechanism is on the PC. Cable technology has the easiest consumer access to the most preferred viewing device for video: the television. According to a 2008 CTAM study, 96% of adults who subscribe to cable or satellite services prefer to watch television on traditional TV sets. [Reference: CTAM] While it is possible for viewers to hook up a PC to the television, the process remains more difficult than accessing content via a cable set top box. Given Cable's widely deployed VOD infrastructure, can Cable leverage its customer-friendly on-demand access to the television to provide user-generated content via the television?

This paper investigates three scenarios for selective entitlement of subscriber-generated video content via cable: local sharing, global sharing within an MSO, and global sharing across MSOs, all of which may in fact overlap. In all of these scenarios, the solution acts as a trust broker, allowing subscribers to issue entitlements to other subscribers – friends, family and/or acquaintances. The cable operator stores the subscriber-generated video content, uploaded via a web portal over an MSO's high speed data service, and maintains the associated entitlements on the subscriber's behalf. The solution provides a mechanism for the subscriber to allow or revoke entitlements on any subset of the subscriber's content.

THE GROWTH OF USER-GENERATED CONTENT

The question could be asked, "Who cares about user-generated content anyway?" and the answer is a resounding "Users do."

In terms of user-generated video specifically, in 2005 there were 3.3 billion user-generated video views, growing to 34 billion in 2008. [Reference: MediaPost]

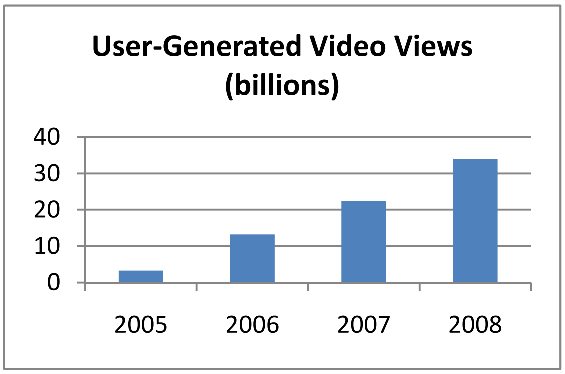


Figure 1. User-Generated Video Views

Cable has an opportunity to tap into this phenomenon to provide a value-add service that brings this wealth of user-generated content in a user-friendly manner all the way to the television set.

THE VISION

The "My (TV) Space" vision starts with a social networking web site providing all the features of the current "social sites" of today: account management, profile management, "friend" management, and content management. Content management today includes text, photos, and video. So, on the social networking web site users can upload user-generated content and share it with one or more of their friends. Note that this web site could be an existing social networking site, an existing MSO web portal, or a new web site.

Now, envision a connection between the social-networking web site and an MSO site. The social-networking solution would understand the connection between web users and MSO subscriber accounts, deliver user-generated content to the MSO site, and inform the MSO site of what subscribers have access to what content.

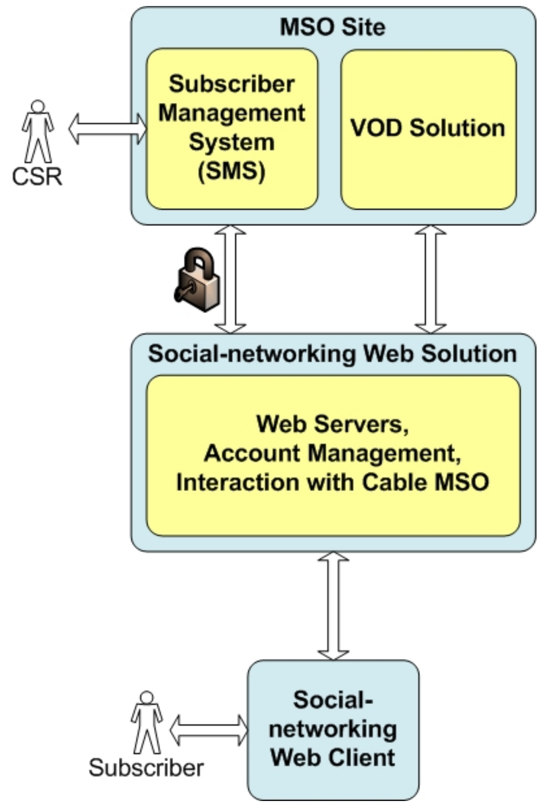


Figure 2. High-level System Overview

Subscribers at the MSO site can subscribe to the user-generated content subscription package in order to gain access to the user-generated content that their friends have shared with them. As in subscription Video-on-Demand (SVOD) packages, the user may subscribe to the package via an interactive screen on the EPG or through a web portal.

(Unlike SVOD packages, the user would not necessarily subscribe via a call to a CSR.)

The MSO's VOD system would be extended to understand the per-subscriber, per-content entitlements to conditionally allow subscribers to access the user-generated content. Titles will automatically appear in the subscriber's VOD portal—either a special social-networking VOD client application or a special category in the existing VOD client. Though outside the scope of this paper, there are many possible ways in which cable subscribers viewing user-generated content may provide feedback to the content author or other viewers about the video.

"MY (TV) SPACE, YOUR (TV) SPACE" – THREE SCENARIOS

At least three basic scenarios exist for selective entitlement of subscriber-generated video content via cable: local sharing, global sharing within an MSO, and global sharing across MSOs.

Think Local, Act Local

In the local sharing scenario, a subscriber might share content with schoolmates or with members of a neighborhood association. Another example is a subscriber sharing video content of little league games with the families of the other little league players. These groups are geographically local by nature and thus could be served by a single MSO site (e.g. a single city). The cable operator can easily manage entitlements and video content within the confines of the local VOD system.

Think Local, Act Global

In the global sharing case within an MSO, a subscriber might post video to extended family, to dispersed college friends, or to remote colleagues. In this situation a cable operator with a national footprint must be able to exchange entitlement information and content across its markets. So in our little league example, the Malibu-based grandparents of the little league player in St. Louis could watch the video of the game. Of course, even cable operators without national footprints can participate in this use case through integration with a web portal incorporating streaming video.

Think Global, Act Global

An enhancement of the above case extends the solution to share video content across markets that span multiple cable operators. In this solution friends can share user-generated content via Cable television regardless of the city or cable operator. This service could be provided only as long as the viewer's home market is part of a coordinated effort among the cable operators.

A SOLUTION – THE UNITED FEDERATION OF CABLE SYSTEMS

One solution option to implement the "My (TV) Space" vision by necessity involves coordination of subscriber information across MSO sites, coordination of content across MSO sites, and coordination of entitlements across MSO sites.

While a "united federation of cable systems" with coordinated entitlements, content, and subscriber information may seem to be somewhere between the implausible and the impossible, the introduction of social-networking solutions which can leverage SMS interactions may simplify some aspects of the solution.

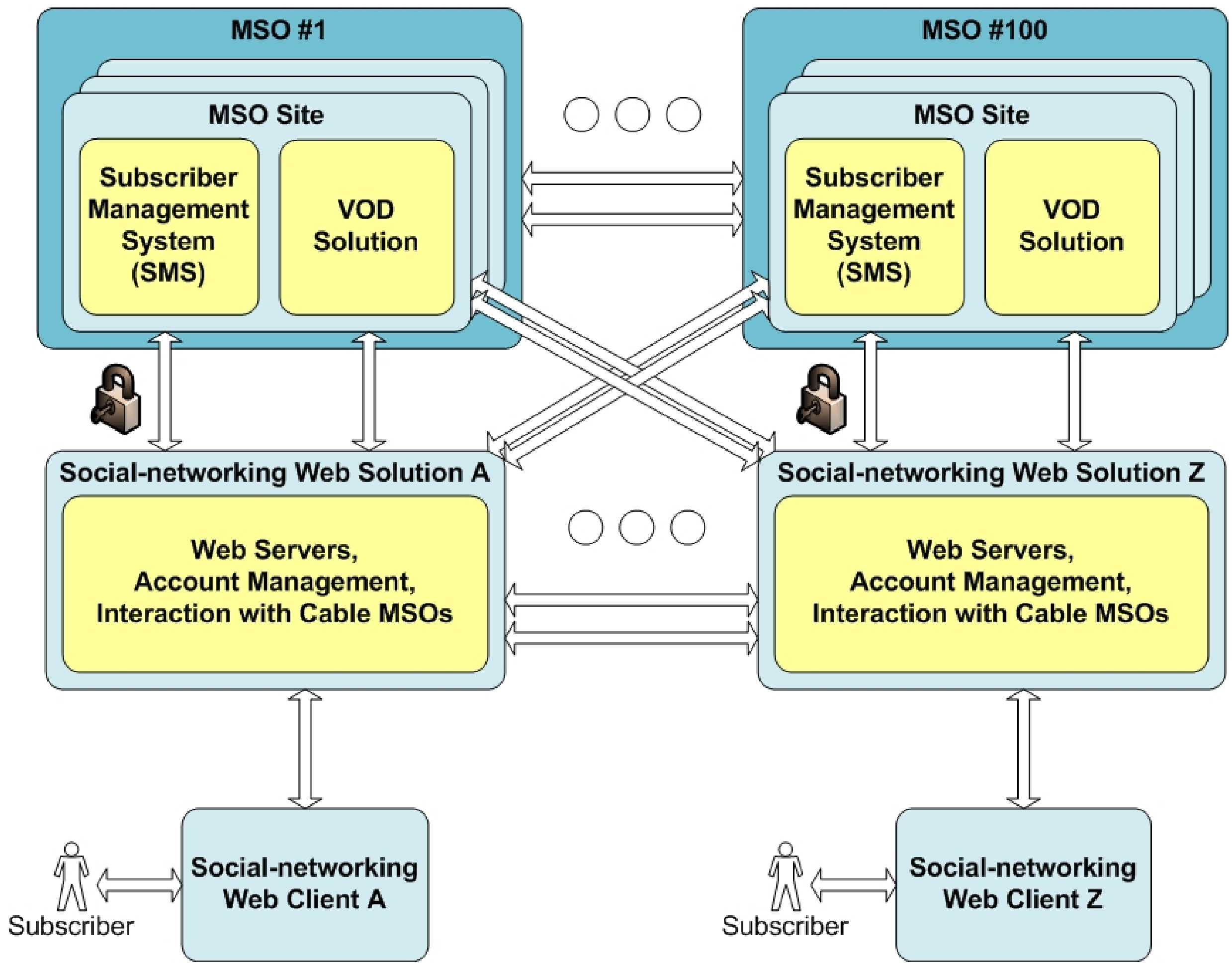


Figure 3. Cross-MSO System Overview

TECHNICAL ASPECTS OF THE SOLUTION

Now let's explore some technical aspects of the solution.

The Social-Networking Web Sites

The social-networking web sites will provide a web-based control interface to users of the system. Capabilities to manage profiles, connect friends, upload videos, and assign viewing access to videos are key aspects of this component. Technical considerations here involve using web development technologies or perhaps partnering with existing social networking sites. Note that this site could be an MSO-branded site or a social-networking site which is independent of MSO interactions. Also note that multiple social-networking sites could be supported by the overall solution.

One new function of the "My (TV) Space" web solution will be back-end transcoding of video to a format conducive to display on cable set top boxes. Existing CableLabs content encoding standards and existing MPEG transcoding technologies may be leveraged for this purpose.

Federated Cable Subscriber and Cable Service Information

In order to enable content access to only those cable viewers who have subscribed to the user-generated content package, the solution as a whole will need to understand cable subscriber information (and package entitlement) across cable operator sites and even across the cable operators themselves. This will require interaction with existing Cable Subscriber Management Systems (SMS's) to coordinate the information. Several industry-standard or existing SMS-vendor interfaces may be leveraged for this interaction.

Federated Assets (Metadata and Content)

A key element of the solution is to allow user-generated content to flow to cable operator networks. This will include both the user-generated videos and the metadata associated with the videos such as title, description, length, and the user who created the content. This data will need to move across MSO sites and across multiple MSOs as well. One approach to getting this content and metadata to the service provider site is to leverage existing CableLabs Asset Distribution and Metadata interfaces.

Federated User-specific Asset Entitlements

Once the correct assets are on the MSO site and the VOD solution recognizes that the account has access to the user-generated content package, the solution still needs the entitlement information of which users are entitled to which specific assets. This information will need to be provided across the cable operator sites and potentially across the cable operators.

Federated User-generated Content Session Feedback

As cable viewers interact with user-generated content there exists an opportunity for feedback to the social network. This feedback could be as simple as the number of views for a given asset. Furthermore, the cable viewer could send messages back to the user who generated the content such as "Great Video!" or "Very funny!" Again, this aspect of the system is a subject unto itself, but the basic feedback mechanism must be acknowledged.

Social Networking VOD Applications

The cable operator's VOD Application Server and VOD Client may or may not change depending on the approach taken to the solution.

One approach is to leverage the existing VOD Application. By enhancing the existing VOD application server and client, the VOD system could display a "User-Generated Content" category in the existing VOD Guide. This category would contain the titles of the user-generated content that the subscriber is entitled to view.

Another approach is to develop a new VOD Application. There are advantages to providing a new VOD Application (server and client), for example, leveraging tru2way, to form the solution. The viewer could be provided with a branded custom "User-Generated Content" user interface with features specifically designed for interacting with social networking user-generated content. The VOD client experience then becomes an extension of the web-based experience. This approach might further benefit cable operators by providing a means to "upsell" set-top devices capable of supporting more sophisticated user interaction.

NON-TECHNICAL CHALLENGES

There are some non-technical challenges that need to be overcome. Among these are potential legal challenges. Most social networking web sites have site guidelines prohibiting upload of copyrighted or illegal material. Will this be sufficient in the domain of cable networks? Will the social-networking web site need to provide moderation of video uploads? These challenges will need to be overcome prior to adoption of user-generated content solutions on cable networks.

WHAT THE FUTURE HOLDS

Given the infrastructure described above the possibilities for new functionality are endless.

Cable operators can extend the approach to create appealing and easy-to-use new services, including:

- Video albums containing personal content and links to entertainment content to be shared among a limited audience.
- Personal TV channels, programmed by the subscriber
- “Internet TV” – bringing web-based series programming to the TV. Today, these series can be found at web sites such as blip.tv and Crackle.com.
- Video mail – either multicast or unicast
- Real-time subscriber interaction across cable systems – chat with your friend in a different city via the TV screen – even if your friend is serviced by a different MSO
- Multi-screen – TV, PC, Mobile content availability

Such services amount to a natural offering, in keeping with cable operators’ position as a trusted provider.

CONCLUSION

With the ever-increasing ways for people to consume video, Cable operators face ever-increasing competition for the eyes viewing that video. The “My (TV) Space” concept, as described here, explores some of the technical aspects of providing user-generated content in a user-friendly manner to cable-enabled televisions across the globe. The challenge of cable operators will be to further leverage the efforts of those individual content creators and to move those user-generated content viewers from small-screen PCs back into the big-screen living room.

REFERENCES

[CTAM] CTAM (6/30/08) “TV Web Sites Grow More Popular, but Viewers Still Prefer Their TV Sets”

[MediaPost] MediaPost (1/31/08) “User Generated Video Expects 34 Billion Views in 2008”

A NEW CONCEPT FOR ROBUST VIDEO MARKING

Niels Thorwirth
Verimatrix, Inc.

Abstract

Digital watermarking is the concept of permanent and imperceptible tagging of video data. User-specific watermarking embeds tracking information that accurately identifies the last legal recipient of a video. This type of watermark does not enforce content use restrictions but it enables identification of sources of content abuse, acting as a deterrent and encouraging responsible consumer behavior. These types of techniques represent a way to maintain control over video content while permitting convenient content consumption that can compete with free but inconvenient and illegal content distribution.

Required is a secure, robust and imperceptible marking technology that does not degrade consumer experience and yet establishes a reliable trace that survives attacks or passages through the analog hole. This paper presents our implementation approach, optimized to create a commercially significant capability incorporating new concepts tailored to digital video.

A key element of the approach may appear to be counter intuitive, in that the embedded information is not designed to be machine readable but it can be extracted into a human readable form. While it requires human interaction, it allows use of well honed human perception processing to aid the extraction task - which is simply superior to machine recognition and thereby solves the problem of content synchronization. It overcomes the challenges of content misalignment that fools machine readable approaches, which are often unable to read the mark after small geometric content transformations.

This paper outlines the research results that enable human readout and the challenges to make the mark robust yet invisible. The result allows for wide content distribution, ensures the mark is secure and can establish strong evidence linking the content to a consumer while maintaining privacy.

DIGITAL WATERMARKING 101

Over the last few years, the representation, storage and distribution of digital video have grown massively in diversity and popularity, and are now approaching the level of digital music in its ease of use and ubiquity. The advantages of this type of distribution and consumption, just as with music, also fuel the growth of large scale piracy. The rightful owners of video content are deprived of their revenues and could be discouraged from investing in the creation of content in the future.

Several approaches have been applied to secure digital media:

Digital encryption technology is effective to enable secure delivery. However, once decrypted and presented in a human visible format, it can be re-recorded to obtain and distribute an illegal, unsecured copy. No secure technology currently exists to prevent re-recording using a camcorder.

Amongst other applications, the marking of media can help investigation to identify individuals that are responsible for abuse, either by embedding recipient information in the media or ownership information that indicates copy restriction.

One way of marking media is done by adding information to the digital media file that is ignored during normal playback, but can be extracted and used by more specialized tools. Apple iTunes is widely assumed to be using this method to tag DRM-free file distribution (1). This kind of tagging enables identification of the unmodified file, but are easily removed and destroyed when the file is re-recorded or converted to another format.

For more robust marking, visible text or dots have been used to carry identifying information in video, and while this information does survive re-recording, it can easily be identified and removed in order to disable the ability to track the content. In order to maintain robustness, these markings are also visible, a compromise that degrades the consumer experience.

Digital watermarking is another marking approach that has been suggested in many variations. Common digital watermarking schemes embed information, by introducing manipulations at a certain positions in space or time. These watermarks are detected by specialist software (2). In other words, it is a process for modifying media content that embeds a machine-readable code into the data content.

When interpreting the manipulations during machine processing of the content, processing can be greatly reduced through pre-knowledge of the insertion positioning. When the positions are modified, i.e. misplaced or weakened, the readout may become difficult or impossible. Such modifications do routinely occur during simple media treatment such as cropping, rotation or conversion to another file format, including lossy compression, during which perceptually insignificant information is eliminated to reduce the size of a digital media file.

Critical transformations that occur during piracy re-recording include:

- Camcorder capture
- Digital-to-Analogy (DA) conversions to S-video or VHS
- Re-compression to low bitrates using popular formats like H.264, DivX, MPEG2
- Geometric distortions like heavy scaling, rotation, AR change, random bending, cropping
- Color conversion, e.g. grayscale
- Filtering like blur, sharpen, contrast modification, noise reduction and de-flicker
- Frame rate conversion

Relative misplacement of the mark and underlying content can also be created intentionally by imperceptible, slight or combined modifications, such as shifts, rations and time jitter. Publicly-available tools (3, 4) apply these modifications, also called attacks, in an automated fashion. Since current image processing algorithms are not very successful in recognizing misplacements in distorted content (a process also called registration) these modifications render these types of machine-readable digital watermarks ineffective.

Digital still images have been the early focus of watermarking research and commercial exploitation (5). Video watermark approaches often have been based on the application of the image watermark applied to video frames. While this is the natural progression and allows the embedding of much data, this approach does not efficiently use the time domain for gathering embedded information because detection is only successful if some information in individual frames can be recovered. This approach fails when none of the watermarks can be read at least in part due to a failure in registration or destruction of relevant areas.

Digital watermarking typically involves a complex transformation of the original image and of the message to be embedded in order to allow invisible embedding. In the forensic, user specific application, watermarks are used to embed information about an individual playback device to allow tracing of the last legal recipient. It this case, it acts as a serial number or license plate and requires execution of the embedding in a playback device, which typically has little processing power to spare for additional functionality like digital watermarking.

HUMAN-READABLE WATERMARKING

To overcome the challenges outlined above, a new approach is required that takes the specific requirements of forensic watermarking into account. The approach starts with the realization that the recognition of distorted content is a task that humans can still perform better than machine. This fact is used for example in so called CAPTCHA images (6) that blocks Web site robots and identify human users. So instead of embedding machine-readable information, a human recognizable image is invisible embedded in the video, distributed over time. During extraction, that information is aggregated in order to derive a human-readable image containing the embedded information.

The result is a robust mark that exists with the actual media and survives transformations thereof, unlike bitstream manipulation and encryption. Unlike visible marking it is unnoticeably hidden in the media. The difference to previous watermarking approaches is the possibility of using the human perceptual system for actual recognition, which is superior to current machine reading capabilities, in particular for degraded, noisy and transformed content.

The message is emphasized by a computer, but the actual interpretation of the mark can be performed by a human, making it possible to detect the mark in degraded content and independent of registration. While it is very difficult for a watermark technology to interpret a misplaced mark, a message can be easily read by a human even if it is rotated, made smaller, stretched and on a noisy background.

During the embedding process, only some areas are modified, while all video data is used for detection. The areas that are modified are chosen by a perceptual model that takes into account where in the content the modification is able to stay just below any noticeable difference and thereby make the modification invisible to the consumer. Another component of user-specific forensic watermarking is that it ensures security by distributing the embedding locations in a random fashion. Consequently, the precise locations of the alterations can not be observed or removed, even if the mark can be recognized.

The mark is spread over the entire frame area and combined with the media using the basic content, such that the embedding portion can not be generally removed without destroying the video.

Additionally the mark is spread over time, and while it generally can not be recovered from a single frame, each frame contributes to the detection result. The detection process accumulates results from several frames over time, contributing to invisibility and robustness of the mark.

The basic principle of combining frames in time is that during the process the actual movie content of the frames will average out to be a more or less smooth surface with pixel values close to the mean –with a faster convergence for high-motion content. The information embedded in the content in contrast is at constant locations and therefore will increase in relative signal strength as the frame content diminishes. A simple averaging though is not sufficient to expose a readable mark. To actually archive robust detection, additional filtering is applied to enhance the mark in every frame. The filter is aimed to reduce the effect of underlying content to be significant in the combined image while at the same time improve the effect of individual frames. Over the last years of researching this approach, we have evaluated a variety of different filters and often found surprising, counter intuitive results on their individual effectiveness. The best achievement so far to highlight the remaining bits of information is a combination of filters applied in sequence.

Our initial research however already revealed that the method of combining frames over time shows remarkable robustness, even in the presence of modifications that would suggest a removal of the minute, unnoticeable variations. These degradations include compression of the color domain or analog transformation. The reason for the approach being effective in these scenarios is that a fraction of the modification will survive any transformation that leaves the content in a reasonable quality.

The specifics of the technology do not allow for automatic measurement of the actual extracted signal strength as it is part of the design that the information is read out by a human. While the mark can be measured for verification by correlation to the known embedded information or OCR applied for reading, this will not be adequate to determine the level of human readability. The detection is therefore somewhat asymmetric. For one, the readout is done by a human and the embedding by a machine. In addition the extraction filter is mostly independent of the embedding and does not follow the specifics used during embedding.

For instance, the embedding locations are spread in a pseudo random fashion that does not have to be known during extraction. Both levels of asymmetric embedding and reading increase the security against attacks that aim to understand the embedding algorithm and to invert them. It also provides robustness against the so called oracle attack that degrades the watermarked content in an automated loop to find the least amount of degradation that causes the detector to fail to recognize the mark.

Unlike machine-readable digital watermarking, the detection process displays an obvious and unambiguous human understandable outcome, e.g. a serial number. When uncovering the mark, the embedded graphic slowly appears from the marked video, showing that the mark is derived from the content. The extraction can be interpreted or read by a layman. The result is easy to understand and can be used as persuasive evidence of wrongdoing during an investigation into content misuse.

HOW TO USE ROBUST WATERMARKING

User-specific digital watermarking is ideally suited for forensic applications that aim to identify the source of piracy. It acts as a deterrent because it registers the individual video to its owner. The embedding task is performed in each set-top box (STB) or DVR, allowing individually-marked copies, without the requirement to process individual videos for each receiver at the head-end. The extraction is performed on copies that are in public distribution and originate from a single piracy source. To identify the embedded information, the copies are read in a central location avoiding the distribution of the extraction service preventing a possible attacker to verify the success of modification applied to the video with the intention to remove the mark.

CONCLUSION

With the obvious need for better protection and infrastructure in the growing business of digital video distribution, a layered protection approach that incorporates digital watermarking can help alleviate the revenue loss content owners are facing through current piracy. Digital video delivered through an STB provides a unique environment for content protection using user-specific marking, potentially providing a superior distribution channel and a preferred form of distribution by movie studios that are looking to secure their sensitive early release windows. As a business stimulus, such early release windows translate into more demand for pay-TV operators and they are a crucial component to increase the momentum and customer base of this distribution channel.

Transparency of consumption enabled by watermarking may be an additional important factor for the bottom line - maintaining the consumer motivation to actually *pay*-per-view for premium content compared to trawling for "free" peer to peer downloads. Effective content protection can make the difference between few subscribers that distribute movies to a large Internet community and a growing consumer base that values premium content.

References

1. http://www.theinquirer.net/inquirer/news/357/1050357/drm-free-itunes-files-have-your-number (retrieved on 1/29/09)

2. Jian Zhao, "Look, It's Not There", in: BYTE Magazine, January, 1997

3. Fabien A. P. Petitcolas, "Watermarking schemes evaluation", I.E.E.E. Signal Processing, vol. 17, no. 5, pp. 58–64, September 2000

4. Fabien A. P. Petitcolas, Ross J. Anderson, Markus G. Kuhn, "Attacks on copyright marking systems," in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH'98,, Proceedings, LNCS 1525, April 1998

5. Saraju P. Mohanty, "Digital Watermarking: A Tutorial Review," Dept. of Comp Science and Eng, University of South Florida, Tampa, 1999

6. Luis von Ahn, Manuel Blum and John Langford, "Telling Humans and Computers Apart Automatically," In Communications of the ACM, vol. 47, pp. 56 – 60, February 2004

Niels Thorwirth
Director Advanced Technology Implementations
Verimatrix, Inc.
6825 Flanders Drive
San Diego, CA 92121
+1 (858) 677-7800 x3003
+1 (858) 677-7804 Fax
+1 (858) 357-1529 Cell
nthorwirth@verimatrix.com

AN OVERVIEW OF TV WHITE SPACE FOR THE CABLE INDUSTRY

Tom du Breuil, Dave Gurney, Mark Kolber
Motorola, Inc.

Abstract

After years of study, public comment and testing, the FCC has adopted rules for unlicensed operation of TV Band Devices (TVBDs). The TV spectrum's desirable propagation properties and the promise of significant amounts of usable spectrum make this band attractive. TVBDs are expected to support a wide range of applications, from providing broadband wireless internet connectivity in rural areas to consumer-based whole house automation and networking solutions. This paper provides an overview of the TVBD rules and their impact, including some potential TVBD applications and key cable operator concerns.

INTRODUCTION

In the United States, 402 MHz of spectrum in the range of 54 MHz to 806 MHz has been allocated for TV broadcasting as shown in Figure 1. Coincident with the DTV transition, now scheduled for completion June 12, 2009, the uppermost 108 MHz of that spectrum, referred to as the 700 MHz band, will be vacated from all TV broadcasting and the spectrum has been reallocated for commercial and public safety systems. The auctions for this spectrum have brought almost $20B[1] to the government with one block still unsold. The remaining 294 MHz of this prime radio spectrum will remain for over-the-air (OTA) TV and other licensed operations. However, throughout the United States, portions of this 294 MHz resource remain unused.

After the DTV transition, all of the full power analog TV transmitters will go off the air, freeing up additional spectrum compared to today. This is often termed the "Digital Dividend" by regulators. The amount and exact frequencies of unused spectrum, i.e., unused channels not assigned to any TV broadcast stations, varies from location to location. These unused segments of spectrum are referred to as TV White Space (TVWS).

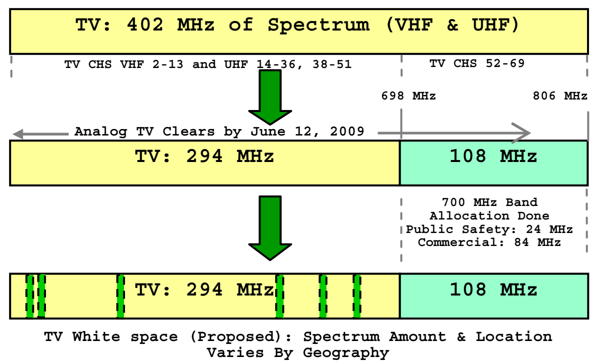


Figure 1. US TV Spectrum Before and After the DTV Transition

Since each market has different TV channel assignments, the potential "white space" in each market is different. For example, in Baltimore, Channel 25 is available for use as white space after the DTV transition while in neighboring Washington DC, it will not be available as a low power analog station, WZDC, will still be using that spectrum. Unlicensed TVBDs will be regulated under the FCC Part 15 rules and are allowed to operate at relatively low power levels.

Making this white space spectrum available is intended to address the need for more spectrum for Wireless Internet Service

Providers (WISPs), home WiFi-type wireless devices and other wireless services.

This white space spectrum should not be confused with the spectrum above 700 MHz that is going to become available after the digital TV transition. As noted, the spectrum above 700 MHz has been re-allocated to licensed commercial services and public safety services and will no longer be used for TV broadcasts. TV stations currently broadcasting in the 700 MHz band will move to lower parts of the TV band.

APPLICATIONS

Many potential TVWS applications have been identified. Wireless rural broadband service (Figure 2) is an application that has received significant attention and could potentially be a good fit in the TVWS bands since rural areas tend to have fewer TV stations, and thus have more spectrum available. In addition, these rural areas are less likely to have wired broadband options due to the low population densities.

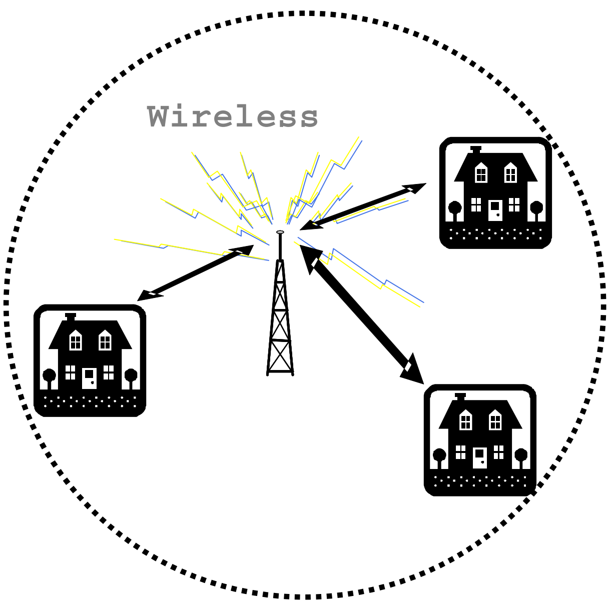


Figure 2. Wireless Consumer Broadband Service

Similarly, the favorable propagation characteristics of the TV spectrum compared to traditional WiFi spectrum potentially allow wireless Internet Service Providers (WISPs) to service rural communities with fewer transmission towers. Such wireless broadband is not limited to rural areas, however, and could also be offered in certain limited suburban areas as an alternative or supplement to other licensed mobile data services or unlicensed (WiFi) wireless data services.

TVBDs may also potentially be used for wireless home networking, both for home automation applications as well as high speed wireless HD video streaming. Thermostats, lights, and the alarm system could all be networked with TVBD technology in order to allow the homeowner seamless wireless integration which permits operation from anywhere in the home. TVBDs could further be incorporated into a smart power grid solution where home appliances communicate in real-time with the local utility company in order to more efficiently utilize both generation and distribution capacity by balancing the real-time price against the timing of when major appliances are operated in order to provide better efficiencies for the utility and lower power costs for the consumer. TVBD networking applications need not be mutually exclusive of other network technologies; a home automation application might connect through a wired broadband service permitting the homeowner to securely peer into the TVBD-enabled wireless camera at the front door. Smart grid applications would also most likely include a combination of networking technologies.

TVBDs may be used to provide the networking infrastructure for a whole home video entertainment network, and can also be used to augment existing WiFi data networks. Where there is sufficient white space available, multiple vacant TV channels can be used to concurrently and wirelessly stream multiple HD videos around the home from a digital video recorder in one room to multiple

TVs in other rooms. TVBDs could also be embedded in consumer camcorders and enable wireless streaming from an HD camcorder directly to the home entertainment system for easy wireless viewing or to a PC for wireless file transfer.

Another potential TVBD application area is data networking for enterprise and public safety applications. Such applications already use licensed spectrum and public safety will get additional spectrum in the 700 MHz spectrum, so these TVBD applications may allow supplemental services. A university or large business might use TVBDs to add high quality video conferencing between campus buildings without impacting existing LAN infrastructure.

FCC TVBD RULES OVERVIEW

In November, 2008, the FCC released the first draft of the Second Report and Order detailing Unlicensed Operation in the TV Broadcast Bands [2]. The order covers the initial operating rules for unlicensed TV Band Devices. The order provides for unlicensed operation of TVBDs using geo-location database as the primary method of incumbent protection mechanism, along with sensing as a secondary method of incumbent protection.

Geo-location database enabled TVBDs must determine their location to $\leq$ 50 m accuracy and access a geo-location database via the internet to determine open channels in that location. The geo-location database contains protection information for analog and digital OTA TV broadcast services, as well as for cable headend and TV translator receive sites, radiotelephone services, and certain (e.g., FCC Part 74) fixed wireless microphone deployments. Generally, broadcast TV stations are protected from any co-channel TVBD operations in their grade A and Grade B service areas to avoid harmful interference. A white space database coalition consisting of several major technology companies including Motorola and Google [3] has been recently formed to address the need for industry standard geo-location databases.

There are also provisions in the rules for sensing-only TVBDs, but such devices must go through an additional testing and certification process that is not yet fully defined. Therefore, one can expect that geo-location enabled TVBDs will be the first to reach the market, once third party database providers are established.

The FCC order allows for operation of two classes of unlicensed TVBDs: fixed and personal/portable devices. Fixed TVBDs may output up to 4 W EIRP, and operate on OTA channels 2, 5-36, and 38-51, while personal/portable TVBDs may output up to 100 mW EIRP, and operate on channels 21-36 and 38-51. Sensing-only TVBDs would be limited to 50 mW EIRP. Table 1 summarizes the operating restrictions for the different devices. There are two classes of personal portable TVBDs: Mode I client devices, and Mode II master devices. Mode II master portable TVBDs utilize geo-location databases to determine usable channels in their area, as do fixed TVBDs. Mode I client portable TVBDs must rely upon a master device (fixed or portable) to acquire a list of usable channels for a geographic area. Portable TVBDs are permitted to operate at locations inside of an adjacent (TV) channel service contour at a maximum power level of 40mW. Fixed devices are presently not allowed to operate on adjacent channels. Note that the FCC has left the record open regarding fixed TVBD operation on adjacent channels, and higher powered rural fixed TVBDs. The FCC prohibited TVBD use of Channels 3 and 4 because of the widespread use of home RF remodulator interconnects on these channels between set top boxes, VCRs, and TVs.

Specification	Personal/ Portable TVBDs	Fixed TVBD
Tx EIRP max	100 mW 40 mW adj ch	4 Watts
Tx TPO max	100 mW 40 mW adj ch	1 Watt
Geo-location	yes, from host or self	yes
Sensing	yes	yes
Prohibited channels	2-20, 37	3, 4, 37
TVBD database registration	no	yes
Tx ID	no	yes
Tx power control	yes	yes
FCC certification	yes	yes
Tx antenna height	n/a	≤ 30 m
Rx antenna height	n/a	≥ 10 m
Tx Spectral Mask	-55 dBr (measured in 100 kHz BW)	
Sensing threshold	-114 dBm with 0 dBi antenna (ATSC, NTSC and wireless microphones)	
Border restrictions	32 km from Canada 40 km from Mexico UHF 60 km from Mexico VHF	
Database check period	at power up and relocation	once per day

Table 1. Summary of FCC TVBD Rules

While, the database is the primary method of protecting OTA TV broadcasts, sensing must also be utilized by all TVBDs to detect the presence of incumbents. TVBDs sensing ATSC or NTSC signals above the -114 dBm threshold must report the detection to the equipment operator and give the operator the option to vacate the channel (i.e., TVBDs are not required to vacate a channel when sensing detects a TV signal if the database indicates that the channel is open). Wireless microphones must also be sensed and TVBDs must vacate the channel when a wireless microphone transmission is detected (regardless of the database). In general, TVBDs must sense a channel for a minimum of 30 seconds before using it, and sense at least every minute during use to determine the presence of incumbents. Cable system issues and protections are discussed in detail in the next section.

There are no FCC restrictions as to the types of modulation or bandwidth that TVBDs may utilize. Vendors are free to innovate in this area and some IEEE groups have expressed interest in developing standards for operation. The IEEE 802.22 group has been developing a standard for wide area networking (i.e., broadband wireless internet service) for well over two years.

POTENTIAL CABLE OPERATOR ISSUES

The deployment of TVBDs can impact cable systems in several ways. There are possible interference issues as well as possible business opportunities as cable operators may want to extend their wired networks to wirelessly support additional applications. The interference issues fall into two broad categories:

1) Interference to the reception of weak OTA TV signals at the headend , and,
2) Interference due to direct pickup (DPU) at the subscriber's residence.

Interference in the cable distribution plant between the headend and the subscriber taps of the cable system is not expected to be an issue. This is because the distribution portion operates at higher RF levels and employs well shielded hardware. Shielding weakness in this portion of the system would cause a cable operator to fail their Cumulative Leakage Index (CLI) requirements and require immediate remediation. The fiber portion of a Hybrid Fiber Coax (HFC) system is also immune to RF interference.

Cable Headend Protection

The newly adopted FCC TVBD rules afford protection for OTA cable headend receive sites. As noted previously, a TVBD device is not allowed to operate co-channel to a TV station within the TV station's Grade A and B service contour. This provides protection to cable headend receive sites and TV viewers alike. Cable headend receivers that operate outside of OTA TV protected service contours are afforded specific protection by the geo-location database, as requested of the FCC by both NCTA and Motorola [4][5]. A TVBD is allowed to operate beyond the grade B contour, but not within specified distances of a headend that is registered in the geo-location database as detailed in Table 2 and Figure 3. Note that both co-channel and adjacent channel operation is prohibited within the "keyhole zones" although the area of protection for adjacent channel is much smaller than that for co-channel as shown in Figure 3.

	Within Grade B Contour	**Beyond Grade B Contour**
Co-channel protection	TV station in data base	Headend in data base establishes "keyhole protection zone"
Adj channel protection	no protection	Headend in data base establishes "keyhole protection zone"

Table 2. Geo-location Protection Mechanisms for Headend ATSC Reception

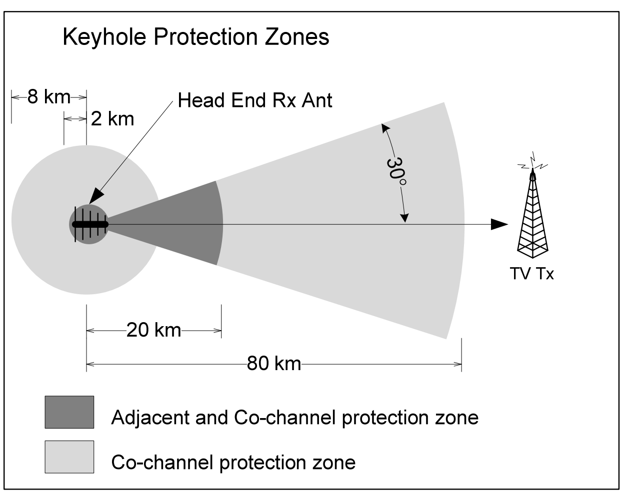


Figure 3. Distant Cable Headend Database Protection

Cable operators will need to register information about these cable headend receiver sites including their location and affected channel(s) with the TV white space database provider.

There is no specific protection given to cable headend receivers that operate inside of the grade B channel contours from adjacent channel TVBD interference. Recall that portable TVBDs are allowed to transmit up to 40 mW EIRP within the grade B contour on the adjacent channel of a TV station. Cable headend receivers as well as TV receivers could potentially be adversely affected by adjacent channel interference from TVBDs. Therefore, it is important that cable headend OTA receivers have excellent adjacent channel rejection, and that the OTA antennas be as high and directional as possible.

Direct Pickup (DPU) Interference at the Subscriber's Residence

As stated above, the newly adopted TVBD operating rules allow up to 100mW EIRP levels for personal/portable devices. It was shown during several FCC tests that direct pick-up (DPU) to cable systems can occur at levels well below 100 mW [6][7]. Both Motorola and the NCTA recommended limiting in-home devices to much lower power levels (i.e., 10 mW) to avoid the DPU interference issue [5][4]. However, the FCC chose to allow higher power levels in the TVWS ruling. Remediation of DPU interference will become more critical in cable systems when TVBDs are deployed.

DPU Analysis

Interference from TVBDs to the cable system headend is addressed by the geo-location database and sensing. These are the same mechanisms designed to protect OTA TV reception. However, interference can also occur at the other end of the cable system, at the subscriber's home as shown in Figure 4.

This is the result of co-channel ingress or leakage into the cable system. Ingress can enter the cable system anywhere there is a weakness in the system shield including poorly shielded or damaged coaxial cable, poor quality, loose or corroded F-connectors, open system ports on splitters and wall plates, and poor shielding in TVs, PC tuner cards, VCRs, set top boxes, cable modems, or multimedia terminal adapters (MTAs).

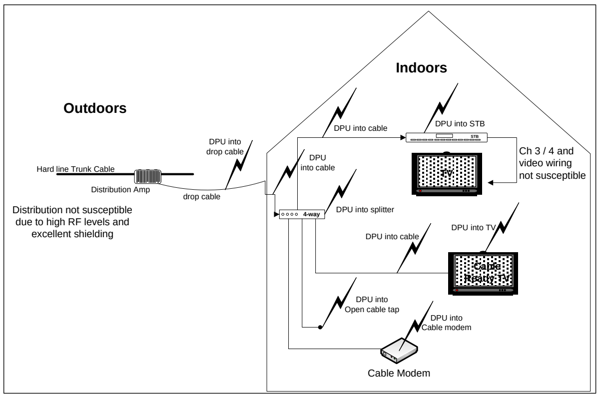


Figure 4. Home Cable System Ingress Points

The interference due to DPU from TVBDs is similar to DPU interference that cable operators have experienced in the past from over the air signals including those directly from TV stations, ham radio operators, land mobile police and fire radios etc. More recently, as some cable system operation has extended above 800 MHz, DPU interference from cell-phones has also occurred. After the digital TV transition, the 700 MHz to 800 MHz band will become occupied with high powered 4G wireless transmitters (e.g., 700 MHz base stations EIRP levels can exceed 1 kW) as well as additional mobile handsets (e.g., with typical 200 mW transmit power levels). The possibility of DPU interference due to these services will also need to be addressed through the same remediation techniques.

It is interesting to note, that after the digital transition, what was previously one of the most troublesome forms of cable system DPU, pickup from analog TV transmissions, will become less troublesome, as the full power analog TV transmitters go off the air. In the VHF range, the OTA and cable TV video carriers were often on the same channel and DPU caused severe beat interference and ghosting. After the transition when the off air analog signals are replaced with 8-VSB digital signals, the high energy video carriers will no longer be present and the visually severe DPU interference will be replaced by a lower level of snow-like DPU interference for on-channel analog cable services. Offsetting this, however, the wide-spread deployment of 700 MHz 4G mobile handsets as noted above, along with nearby lower powered TVBDs have the potential to significantly increase the instances of DPU becoming a problem. Sufficiently strong DPU can completely disrupt services on the affected channel, which may include multiple SD and HD video, high speed data, or telephony services. Some operators have chosen to stop using some cable channels in the 800 MHz band that were suffering from widespread severe DPU, caused by mobile handsets.

The fundamental interference protection mechanism used by TVBDs to protect OTA reception is the selection of an unused frequency or white space. Unused OTA frequencies, however, have little relationship to unused cable channels. There are very few, if any, white spaces on a cable plant, especially with the growing demands for more HD, on-demand and data services. This means that every TVBD has the potential to interfere with one or more services on the cable plant via DPU. The remaining interference, or DPU, protection mechanism is the "shielding effectiveness" (SE) of the cable system, which includes the entire residential cabling system including all the devices attached to it such as TVs, set tops, and modems. Once an interfering signal has entered the cable system, this undesired interferer travels throughout the home along with the desired cable signals. As such, a TVBD causing interference may be located at the opposite end of the home well removed from where the interference is observed.

The FCC has established the RF DPU standard for analog cable ready TV receivers in 47 CFR 15.118 as 0.1 V/m. The rules do not explicitly refer to digital cable ready digital TV's but we can assume the same requirement applies. The FCC acknowledged the widespread use of Channels 3 and 4 as the in-home interconnect frequency and has prohibited TVBDs from operating on those channels. Subscribers using a set-top box (STB) with a Ch3/4 connection from the STB to the TV will be protected from interference in this interconnect. Likewise systems using a baseband audio / video (e.g., composite, s-video, component, DVI, or HDMI) connection should not suffer from interference to this connection. However, for those subscribers not using a STB but instead using a cable ready TV connected directly to the cable, the TV itself may be susceptible to DPU as shown in tests conducted by the FCC [4].

Figure 5 shows the free space far field strength at various distances radiated by TVBDs operating at 40 mW, 100mW and 4 Watts. The dashed horizontal line at 0.1 V/m represents the minimum FCC specification 47 CFR 15.118 for cable ready TV set immunity. The lines cross at about 40, 60 and 400 feet respectively. Thus we can expect the possibility of DPU interference directly to a cable ready TV that just meets minimum FCC DPU specifications at these distances in free space.

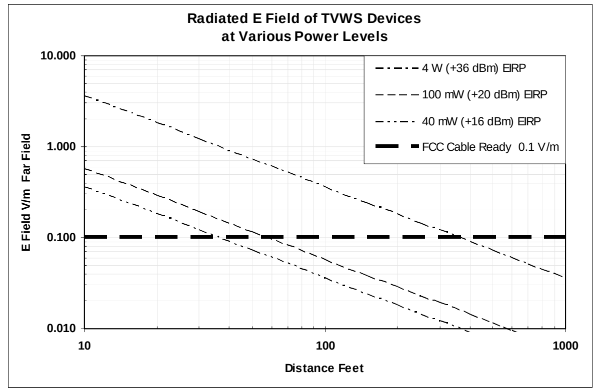


Figure 5. Susceptibility as a Function of Radiated Power and Distance

Even when a cable ready TV is not being used, the in-home cable wiring itself can be susceptible to DPU [5]. Motorola conducted tests on various home wiring components including a selection of retail purchased cables and splitters and found a wide range of shielding effectiveness. For the purposes of these tests, the shielding effectiveness (SE) of a sample of cable is the ratio between the external electric field measured in V/m and the induced voltage on a 75 Ω load connected to the cable. For example, if a cable exposed to 1V/m (60 dBmV/m) induces 0 dBmV into the load, the shielding effectiveness is 60 dB.

For comparison, the shielding effectiveness of an FCC 47 CFR 15.118 compliant TV set was calculated. FCC 47 CFR 15.118 specifies that with a desired signal level of 0 dBmV and a 0.1V/m interfering field, the DPU interference must be better than -45 dBc. Thus a 40 dBmV/m interference field must induce an equivalent interference signal below -45 dBmV. Therefore the minimum required shielding effectiveness is 85 dB.

The specified minimum tap level for 256 QAM is usually -12 dBmV. The C/I ratio for Threshold of Visibility (TOV) for 256 QAM is about -27 dBc. Therefore any interfering DPU signal above -39 dBmV will disrupt the QAM signal. Note that in the case of an analog signal, the DPU interference becomes visible above -45 dBmV but will not completely degrade the picture or sound. In the case of 256 QAM digital signals, interference above -39 dBmV will not just become visible but will completely destroy the picture and sound. This is due to the "cliff effect" of the digital video systems. This difference needs to be kept in mind when considering the ramifications of various levels of interference. Subscribers may be expected to tolerate some level of visible DPU interference to analog signals but cannot be expected to tolerate a total loss of picture and sound due to DPU interference to digital signals.

Motorola tested the shielding effectiveness (SE) of various coaxial cables and wiring components and provided this information to the FCC [5]. An example of the SE test data is shown in Figure 6.

The best cables had a SE measured around 130 dB which was the limit of the test equipment. The worst cables were in the range of 70 to 80 dB. The type of termination on the end of the cable was also found to have a large impact on SE, and hence DPU interference susceptibility.

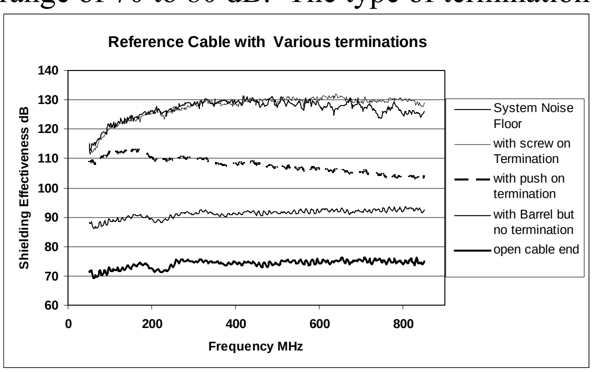


Figure 6. Shielding Effectiveness of Various Cables

Table 3 summarizes the mid-band 500 MHz value of shielding effectiveness we found for various samples of cables, terminations and splitters.

Sample	Typ SE at 500 MHz
Noise floor of test system	130 dB
High Quality Cable with barrel and termination	110 dB
High Quality Cable with barrel, NO termination	90 dB
High Quality Cable, open F connector	75 dB
Splitter with soldered back, all ports terminated	125 dB
Splitter with glued back, all ports terminated	80 dB
Splitter with soldered back, all ports open	95 dB
Retail cable, RG6 Quad Shield, Snap ‘n Seal	130 dB
Retail cable, RG59, Molded connector assemblies	85 dB

Table 3. Shielding Effectiveness of Various Cables and Terminations

Figure 7 shows the effect of the common practice of leaving an unused cable end exposed (i.e., unterminated) which causes large degradation to the SE. This is likely to be typical of pre-wired homes in rooms without TVs or other devices connected to the cable system.

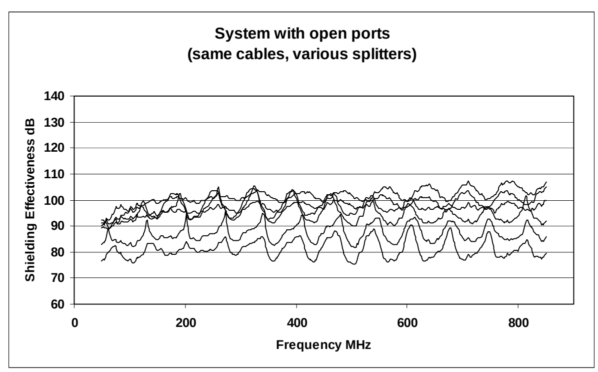


Figure 7. The Effects of Unterminated Ports (Open Wall Plates)

DPU Mitigation

There are several steps a cable operator can take to mitigate interference from TVBDs, the new 700 MHz services, and the existing 800 MHz mobile services. In the case of TVBD interference to the headend, the operator must register headends receiving OTA signals beyond the grade B contour in the TVBD geo-location database and keep these records accurate.

For all headends receiving OTA signals, the OTA antenna(s) should be highly directional, properly pointed, and mounted as high and as far as possible from likely locations for portable/handheld TVBDs. ATSC headend receivers should also have good adjacent channel rejection characteristics and quality low-loss cabling between the antennas and receivers.

A combination of remediation approaches may be needed to resolve DPU interference in subscriber homes. One of the most important techniques is to properly terminate open ports. Cable drops, splitters and terminations should all be of high quality and properly installed. TVBDs and cell phones should be located away from the cable system and TVs. Old or unused cabling/splitters in the home should be disconnected from the system. And in extreme cases, it may be necessary to disconnect poorly shielded cable ready TVs and install well shielded set top boxes in front of them. Cable operators may also want to work with local retailers to insure that cable components with high shielding effectiveness are readily available in retail outlets. And cable operators may want to prepare and distribute information to their subscribers explaining self-help remediation procedures.

There is another technical step the cable operator can consider. Increasing the signal levels on the in-home wiring directly reduces the effect of DPU interference. Signal levels delivered to the inputs of set top boxes and cable ready TVs should be set to levels that are well above minimum system and regulatory requirements. Every extra dB of signal level here provides another dB of DPU immunity. As analog signals are replaced with digital signals, the need to maintain very low levels of composite second-order (CSO) and composite triple beat (CTB) are reduced. Amplifiers can be used at the residential point-of-entry to increase the signal level on the in-home wiring without having to raise the signal levels throughout the entire distribution plant.

Figure 8 shows the E Field DPU susceptibility for 256 QAM against a range of cable shielding effectiveness for several levels of QAM operating level.

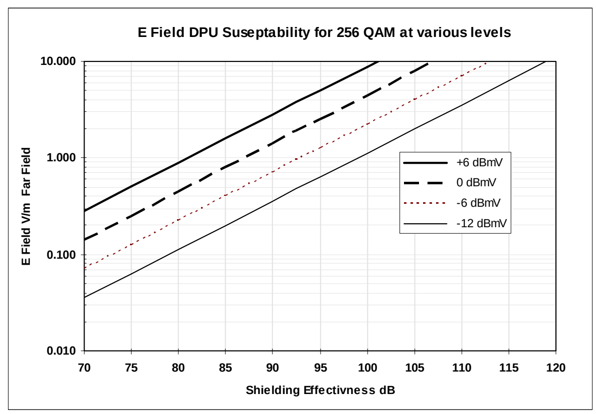


Figure 8. DPU Susceptibility as a Function of SE

For example, using a cable with 100 dB SE and a 256 QAM operating level of -12 dBmV yields a DPU susceptibility of 1 V/meter. We can use Figure 5 and Figure 8 together to examine the relationship between desired signal level, and system SE vs. TVBD power level and separation distance. Using the 1 V/meter level we obtained from Figure 8, and referring to Figure 5, we should not expect DPU problems from a 4 Watt TVBD at a distance of about 35 feet or greater free space. Note that any intervening walls or structures between the TVBD and the affected cable device would decrease the required separation distance.

CONCLUSION

TV White Space presents interesting opportunities to utilize the TV broadcast spectrum for a wide range of unlicensed applications. The FCC rules currently provide for two classes of TVWS devices: fixed and personal/portable TVBDs; each are permitted different maximum power levels. The primary protection mechanism for licensed users of the TV spectrum is the mandatory use of a geo-location database. Sensing is defined as a secondary protection mechanism. Additionally, the FCC rules allow for a class of sensing-only products at a lower power level, but such devices will have to undergo additional testing and qualification by the FCC before such devices may be sold.

For cable operators, TVWS presents both the opportunities for additional wireless applications and services as well as risks to their current cable system operations. The risks include the potential for TVBDs to cause interference in cable systems.

The FCC acknowledged the NCTA and Motorola comments on the potential of TVBD interference to headend OTA reception of distant ATSC TV signals and the FCC provided a specific protection mechanism in the TVBD rules. Such headends are permitted to register in the geo-location database and are granted a keyhole protection zone to facilitate these distant OTA reception requirements.

Cable systems today typically operate up to 750 MHz, 860 MHz, or 1GHz and often have no white space due to the high subscriber demand for more HD, on-demand and data services. As such, all TVBDs operating in broadcast TV spectrum “white space” are, by definition, transmitting co-channel to one or more cable services. As such, any ingress of a TVBD into the cable system will cause interference and possibly even loss of service on the affected channels. Such ingress, referred to as DPU, is most likely to occur in subscriber residences both due to the lower signal levels used within subscriber drops and homes versus the main cable plant as well as the wide variety of cabling, device types, quality and installation of the home wiring systems. Additionally, many of the potential TVBD applications are consumer focused, and such TVBDs are likely to be in close proximity to subscriber cable systems.

The TVBD rules that the FCC decided to adopt, allow higher in-home power levels than either NCTA or Motorola recommended in FCC comments [4][5], as well as power levels higher than the levels at which the FCC found caused cable system DPU interference in its own field tests [6]. As a result, many subscriber homes may experience DPU interference resulting from the deployment of TVBDs and remediation would be required to resolve these issues. These interference issues are likely to be similar to those that some operators have already experienced with existing mobile phones in the 800 MHz band. With the reallocation of the 700 MHz band including significant spectrum licensed for high powered 4G mobile services, cable services on those frequencies are likely to see additional interference as well. Fortunately, the same remediation techniques can resolve all these potential interference issues.

Consumer demand for bandwidth continues to grow, and this demand provides continuing growth prospects for cable operators. TVBDs provide unlicensed access to significant spectrum, thereby enabling additional cable operator wireless applications expansion. Cable operators will have to lead DPU remediation efforts with both their subscribers and local retailers in order to maintain and grow the capacity of their current cable systems, paving the way for these additional wireless applications.

REFERENCES

[1] FCC Auction 73, 700 MHz Band.
[2] FCC Report and Order, Docket 04-186, 11/14/08.
[3] White Spaces Database Group, 2/4/09, http://www.google.com/intl/en/press/pressrel/20090204_whitespaces.html
[4] Comments of NCTA, Docket 04-186, 1/31/07
[5] Ex Parte of Motorola, Docket 04-186, 12/7/07
[6] FCC OET DPU Test Report, Docket 04-186, 7/31/07.
[7] FCC OET Phase II Test Report, Docket 04-186, 10/15/08.

Cloud Computing for Cable TV: Expanding Choice, Control & Content for a New Generation of Television Viewer

John W. Callahan, Chief Technology Officer, ActiveVideo Networks

Jeremy Edmonds, Director, Product Management, ActiveVideo Networks

INTRODUCTION:

The multichannel television industry is caught between two significant realities: Growing consumer demand for increased choice and control of TV content, and an installed base of technologically fragmented, resource-limited digital set-tops and CE devices.

Consumers have become acclimated to the media richness, immediacy and dynamic nature of Internet video. It follows that consumers – particularly those under 30 years of age -- are increasingly seeking to avoid the usability limitations associated with traditional broadcast multichannel video delivery systems.

As a consequence, content owners, service providers, and advertisers alike face major challenges to traditional business models as video consumers shift to the one-to-one world of unicast, on-demand video streams that characterize VOD, DVR, and Internet video.

The notion of "just plug a broadband connection into your HDTV" – a predominant trend at the January 2009 Consumer Electronics Show -- is the latest step in the process of moving from the traditional operator business model of bundling access and content, to a world where the consumer chooses "what, when, and where" to "watch television" [as our parents called it]. The implications for cable are clear. First, this movement damages the operator's traditional revenue streams, i.e., bundling, "tiering" and 30-second ad insertion.

And second, the inexorable growth in IP access network bandwidth, exemplified by the advent of DOCSIS 3.0, serves to enable further disaggregation of the bundled model for video consumers.

As the operators' service delivery platform continues to change, however, so does the software applications platform. The availability of cable-delivered EBIF [Enhanced Binary Interchange Format], tru2way and unicast streaming applications such as NDVR [network-based digital video recording], will allow operators to more easily introduce new video services – subscription, transactional, and advertising supported. The debut of a new generation of "broadband-connected" TV sets will catalyze this as consumers will find it easier to "watch television" without subscribing to the traditional operators' services.

In the same way that the Internet is shifting toward a "cloud" model -- where computing resources are provided as a service, over the Internet -- so can multichannel video providers leverage their local, regional and national footprints to be the "TV cloud" for more and better content.

Specifically, this approach could deliver a stream of personalized content as a single, standard MPEG stream -- to any digital set-top box or web-connected CE device.

This paper provides an overview of current impediments to TV-based interactivity, and describes the high-level requirements and capabilities of the "cloud computing for the TV" architecture that could improve the outlook.

THE CASE FOR THE CLOUD

The language of the computing "cloud" is typically associated with the Internet, even though the term itself pre-dates the Internet by at least two decades. Before it was the "cloud," computer scientists recognized the need for "grid computing," "distributed processing," and other variations on the theme of sharing a processing workload over clustered computers.

The benefits of cloud computing apply to interactive television, as well. The "TV cloud," for instance, could remove workflow gaps and bridge the application and media processing requirements between head-ends and set-tops. A cloud approach also would enable developers of programming and advanced advertising to use familiar, web-based development tools that are similar to or the same as those used to provide interactivity within a web site (e.g., DHTML, JavaScript, etc.).

As in all approaches, there are pros and cons to the notion of locating computing and media processing power in the operator network, versus investing in generations of more powerful Customer Premise Equipment (e.g., set-top boxes). Noting that each user is given an individual, personalized video stream can sum up the con of the network approach. Obviously, this uses the operator's network bandwidth to carry the unicast stream. However, the pros of this approach include:

- Faster software application lifecycle and lower total cost of ownership (development, integration, deployment, content management).
- The ability to deploy rich media services (i.e., complex video, graphics, and interactivity)to all CPE, eliminating "lowest common denominator set-top issues" and increasing scale for operators, programmers and advertisers.
- Consistent presentation of content and advertising opportunities (which minimizes unpredictable nature of the user experience on different CPE and optimizes the ability to accurately place advertising elements).

There are certainly capital investment attributes to the argument above. Generally speaking, investment in server-based computing and network bandwidth leads to less expensive CPE being able to provide an array of video services (i.e., simple MPEG decoders). That said, CPE with more capabilities can be equally supported. With respect to the investment in software technologies, the greatest expense often is in targeted development, integration and regression testing across dozens of different CPE platforms -- each with its own performance characteristics, graphics display capabilities, and consequent impact on the viewer experience. While tru2way, EBIF, and other standard client software platforms are certainly an improvement, they do not eliminate the cost of integration and regression testing.

As cable operators and multichannel video providers have deployed more and more sophisticated applications, the costs of client development have increased greatly – measured both in the time to market (24 to

36 months not being at all uncommon) and direct technology expense for complex labs and software expertise.

The server-centric approach allows an application to be developed, integrated, and regression tested on a single software platform. The time to realize a new or modified application can be measured in months (if not weeks). The server-centric approach takes advantage of powerful computing resources to enable sophisticated, media rich applications that can meet the operators' collective need for a rapid and repeatable way to "change the category," by creating must-have features that are unique to the video platform.

Given the growth in IP bandwidth – both the access and backbone networks – delivering "TV Quality IP Video" is quickly becoming a viable model. The server-centric approach is well suited to deliver video services over the IP network (somewhat obvious when one considers the growing number of such providers).

Questions for operators relate to development and deployment of new video services that maximize their investments in their robust, high QoS "last mile networks" for both traditional HFC and IP connected devices. What is the appropriate application software model for video applications? In short, via "the server cloud or the CPE [set-top] client?"

Addressing the industry's set-top fragmentation problem – meaning the hundreds of permutations of differing hardware and software combinations that exist amongst the 37+ million digital cable boxes currently installed – will go far in assuaging the "critical timing factor" necessary to quickly remove applications that aren't attracting consumers, and to quickly add those that are.

In addition, the total cost of ownership for applications lifecycle is a phenomenon well understood in the enterprise and personal computing domains, but only recently becoming an issue for multichannel video providers. In this sense, "total cost of ownership" includes not only initial applications development, but also customization, integration, bug fixing, and regression testing. That is, a "fat client solution," to the exclusion of all else, is an extraordinarily difficult and expensive business proposition.

One of the features of the Web is that it is essentially a "cloud computing" model in the "software as a service" sense. It allows the application lifecycle to be managed on a small and well-behaved domain of common devices – network servers and common client devices, i.e., PCs. The fact that PCs all have enormous computing capabilities means that media-rich applications depend on client-based media processing technologies for application execution (e.g., Flash).

Multichannel video operators can share the benefit of the Web approach and utilize network-based servers to support complicated, media-rich applications -- but they cannot depend on the client to have the necessary capabilities. Unlike the ubiquity of just one or two operating systems and common APIs, the multichannel video provider has dozens, if not hundreds, of client [set-top] platforms - all differing in the details of capabilities and performance.

Standardization of the client side will take many years. By analogy, it took more than a decade for Microsoft Windows® to become a de facto standard for desktop PCs. Application deployment to PCs is a demonstrably slow and expensive proposition as compared to the "web model" of software as a service. Web applications are iterated frequently and at low cost. Cable operators all have one common "standard" in their client devices, i.e., MPEG decoders. This, combined with server-based applications in the cloud will bring a more "web like" application life cycle model to the television viewing platform.

This Web model approach would combine the concept of "cloud computing" with some aspects of the traditional set-top approach. In essence, it binds together the best features of cable, cloud, and client.

CURRENT IMPEDIMENTS TO TV-BASED INTERACTIVITY

The storied past of TV-based interactivity goes back as many as three decades, yet "interactive TV," as a category, seems always to be on the horizon. The word "interactivity" itself is part of the problem: As soon as an "interactive application" gains consumer traction, it takes on a descriptor other than "interactive." Consequently, the application exits the perceived realm of "interactivity," becomes part of the "normal" viewing experience, and ceases to exist as an example of "interactive TV." Examples include the electronic program guide (EPG) and video on demand (VOD). Both provide interactivity, yet neither is considered an "interactive application" as they have passed into the realm of the "normal" viewing experience.

The historical predominance of "destination-based" interactivity is also a factor. To access VOD, consumers were taught to "go to" a place to find and order titles. Likewise for the electronic program guide, which exists as a separate menu destination. A strong argument can be made for interactivity to occur as a natural part of the viewing experience that it enhances.

Immersive interactive video applications will bring the desired content "to the viewer," not make the viewer search to find a "destination" in an unnatural way. This "surfacing the content to the viewer" (versus "destination-based interactivity") can be found on many video streaming websites, e.g., YouTube, where the activity of viewing any given video stream is augmented by meta-data links to several other video assets (as well as non-video applets).

Transposing this experience to a full-screen video monitor, viewed at distance, and eliminating the computer mouse and keyboard bring the "VOD" and "interactive TV" models together as natural features of "watching television."

THE WORKFLOW ISSUE

Several other impediments to TV-based interactivity exist. One is an overall lack of an automated systems infrastructure to connect the "sales order process" with the "creative process" to the "content management and provisioning process" and finally to the "delivery process". The overall word to describe this is "workflow."

For the traditional multichannel video subscription business, this workflow is well established. In its simplest form, movies and TV shows are produced, licensed to an aggregator (e.g., NBCU), wholesaled to an operator (e.g., Comcast) for distribution, and then retailed to the consumer. The advertising and subscription models are well established for this process.

The important point is that there are automated systems (encoding, content protection, "billing systems," trafficking systems, royalty payment and settlement) that support this model so these businesses can scale.

With respect to interactive applications, this “workflow” does not exist in any uniform, scalable way. The current ecosystem of extant and desired interactive video applications and services relies on a patchwork of business systems and creative tools, all of which are delivered to a heterogeneous population of operators with no “billable event tracking” except by sneaker-net and swivel chair operations. Without the “back-end” tied to the “front-end” via an automated workflow that generates invoices and tracks payments and respects copyrights, it will be very hard to build a scalable business around a popular interactive application.

A specific example of the workflow conundrum is the notion of the “bound” application, meaning an application that executes synchronously with the program or advertisement within which it runs.

EBIF, the Enhanced Binary Interchange Format, is the Cable Television Laboratories specification developed to establish bound applications over two-way video plant. EBIF’s strength is its overall reach – essentially the entire installed base of digital cable set-tops. However, EBIF defines only a portion of how to execute bound [program-synchronous] and unbound applications.

Specifically, EBIF defines only the delivery chain of the “trigger” or “widget” that enables a consumer to “click” from the remote control and to engage with the TV and the program or advertisement at hand. Such definition is critical and necessary, but for EBIF-based “bound” applications to become mainstream, a necessary scaffolding of workflow must emerge.

That workflow scaffolding includes the following: A known, easy and repeatable method for creating applications and applying any QA [quality assurance] mechanisms to ensure applications behave at their best; data collection, to fulfill the application’s intent, and to feed any primary or third-party billing mechanisms; and the links to those billing systems.

Consider an advanced advertising application that allows the viewer to click on a widget associated with an ad to receive more information on the product. From a workflow perspective, gaps emerge immediately:

1) *Creative*: What should the widget look like? Who builds the creative for the campaign – and to what template, using what authoring tool(s)?
2) *Application provisioning*: Operationally, the interactive application must be provisioned on to the network – its widget assets transferred for playout, its availability parameters fed into the traffic/billing system.
3) *Stewardship*: All ad campaigns follow general and specific rule sets – competing products may not be shown within the same ad pod; time parameters to protect children from inappropriate content, etc.
4) *Data Collection*: After playout, data associated with the spot needs a method to flow into the aggregation engines feeding national and local campaigns.
5) *Billing*: Any additional revenue associated with the interactive spot needs a feed into operator billing systems.
6) *Reporting and Settlement*: automated mechanisms must be available to

operators and advertising constituencies, etc. to create reports both for advertising effectiveness and contract fulfillment purposes.

While many efforts are underway, the authors do not know of any available solutions that will connect the traditional day-to-day business of advertising sales to the operators' broadcast and unicast streaming platforms. Individually and combined, workflow gaps prevent the business from scaling and the ability for multichannel video providers to build both local and national advertising revenues.

THE CHALLENGE OF THE INSTALLED BASE

Digital cable set-tops, as a category, are approaching their 15th anniversary. Until fairly recently, they've existed as the "thin clients" that lag the Moore's Law trend of computing devices. Compared to PCs, digital set-tops have long been dismissed as not including enough processing power or memory to enable immersive, media rich applications. In short, what's thick today is thin tomorrow, and, for digital cable boxes -- compared to PCs -- it's always tomorrow.

The installed base of digital boxes presents a "lowest common denominator" problem for application development and software version control. Building applications only for high-end boxes reduces potential reach; building applications for all set-top variations reduces the application's attractiveness to the lowest common denominator of graphics chips, processing power, and memory.

Put another way, operating interactive applications solely upon the limited capabilities of the aggregate set-top base, and without the benefit of network server resources means the wealth of capabilities in the newest units is eclipsed by the care-and-feeding needs of the oldest units.

THE CHALLENGE OF THE INSTALL-*ING* BASE

Equipment fragmentation problems are not contained to the set-tops of multichannel video providers. While the "TV Widgets" so prevalent at the 2009 Consumer Electronics Show achieved high marks for "cool factor," they will ultimately face similar challenges.

Assert: Consumers will tolerate poor quality of service when using an "interactive application" on the Internet versus on the TV. Consider: Millions of dollars have been spent trying to make channel changes take half a second less time, because consumers dislike having to wait two seconds instead of one and a half to change from one video stream to another.

Consider the developer, though, seeking to get a new "Yahoo Widget" into a Samsung television set. First, the widget gets submitted to Yahoo's TV widget working group for a testing cycle (with associated costs) that may be lengthy. After approval, the widget is presented to Samsung, likely requiring a convincing business model agreement to justify association with the Samsung brand. These steps are repeated for inclusion on other "widget-ready" devices, and likely would not be possible for an independent widget developer to perform.

Another popular client environment, Adobe Flash®, typically releases an entirely new software stack on an annual basis. Each release assumes a Moore's Law-like increase in available computational power and memory in the underlying platform –

yet that assumption likely will not fit the razor-thin margins and cost justifications required in the CE manufacturing environment.

Indeed, making a Yahoo Widget or Adobe Flash engine run exactly the same, on all consumer display devices, and so that the application author is not exposed to a maze of confusing, conflicting specifications is a very complex task. The result, as has been the case with set-top boxes, is the likely pruning of platform features to the lowest common denominator.

REQUIREMENTS TO ENABLE SERVER-CENTRIC APPLICATIONS ON THE ACCESS NETWORK

The "new new thing," vis-à-vis using server-centric applications to deliver interactive video applications to thin client set-tops and/or CE client devices, is connecting the population of client devices to the server cloud with sufficient bandwidth. This seems obvious yet, in practice, it more than likely is not. Why: Multichannel video operators generally maintain copious amounts of bandwidth to connect their IP access networks to the Internet, yet the traditional headend is not likely to have more than a data connection for File Transfer Protocol (FTP) support. This is a result of the fact that the subscription video content has traditionally been served by point-to-multipoint satellite distribution, or by low bandwidth links for IPG data or other non-video applications.

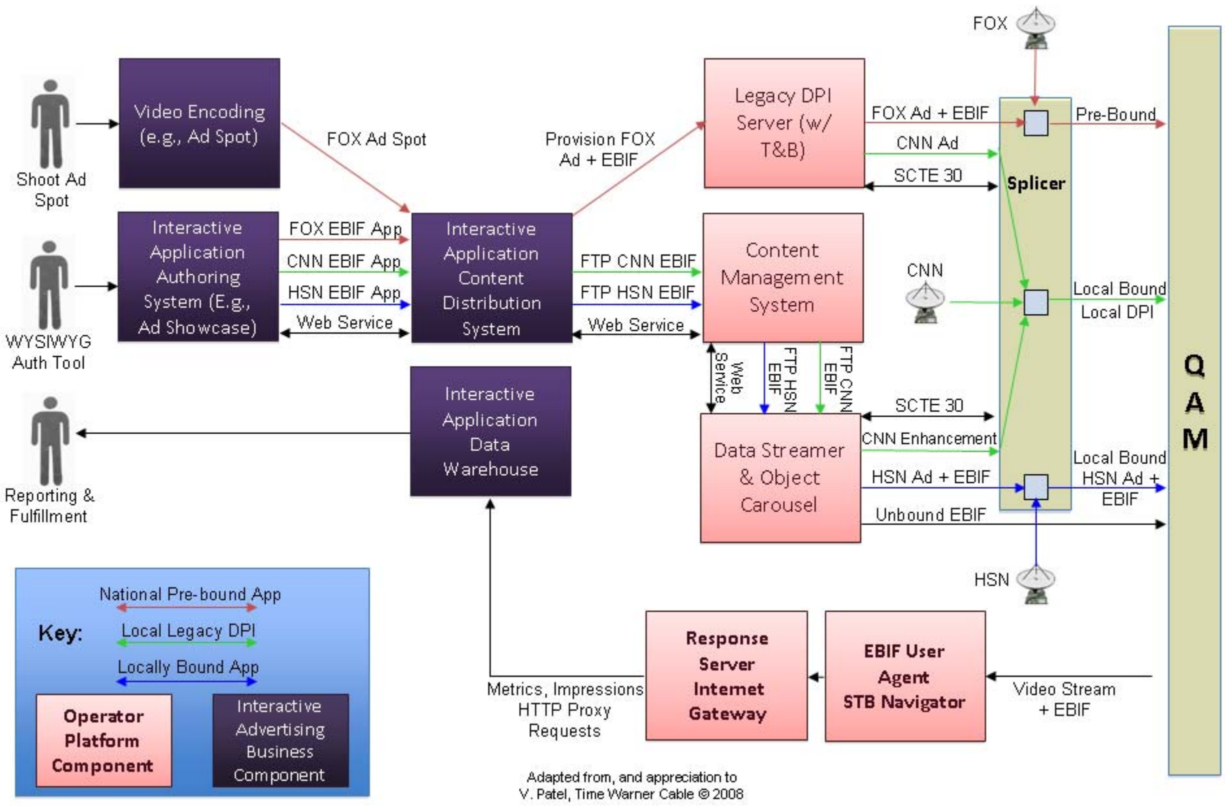
Requirements to enable server-centric applications, using content cached locally as well as pulling content from Internet sources, include:

Requirement 1: Provisioning of backbone network interconnections to 100 Mbps or greater between headend servers and relevant Content Distribution Network locations. This is obviously sensitive to the characteristics of individual applications. In effect, operators need to enable a terrestrial caching point at the edge of set-top network (as is done for IP services). The required bandwidth will be a calculated value derived from customary traffic analyses.

Requirement 2: The provisioning of adequate access network bandwidth to deliver unicast video streams to a meet a specified QoS. That is, the operator will estimate "peak simultaneous stream usage" and ensure the network Session Resource Management components are provisioned accordingly.

Requirement 3: The transcoding of content to formats supported by the client population. While there is a diversity of formats on the Internet (and a diversity of plugins to process them), the video streaming format that ALL CPE devices can process is MPEG. In addition, for cable operators, the ability to multiplex EBIF applications and/or deliver Java applets is required.

Interactive Advertising Application Creation/Delivery Work Flow



<u>Requirement 4:</u> Workflow support for integrating the application platform with "business systems" and "content management systems" back ends. This issue was outlined in the workflow discussion.

<u>Requirement 5:</u> The availability of common Web authoring. That is, tools that use DHTML as well as common scripting and video object toolsets. As applications are transcoded to the specific client capability (as above), the application developer does not need to be schooled in the various client capabilities and idiosyncrasies.

<u>Requirement 6:</u> The guarantee that applications will present the user experience per the designer's intent, consistently, on all client devices

<u>DOWNSIDES OF A "PURE CLOUD" TOPOLOGY:</u>

To be fair, an entirely cloud-based approach to interactive TV isn't the right answer, either. Doing so wouldn't take advantage of client device capabilities, like local graphics blending, and a location execution environment for "overlay applications." An easy example of this is the ability to detect embedded data "triggers" on broadcast video streams (think "EBIF Trigger"). This allows a local application to "overlay" graphics on the video stream (and execute application logic) only for those viewers who choose to do so; it creates a locally interactive individual experience with a broadcast video service. There is no need for such a mechanism with unicast video streams as the viewer is already engaged in an individual, one-to-one, experience. As the bandwidth intensive video is shared (broadcast), this approach can be more efficient than unicast streams.

Likewise, certain applications are almost entirely client-based. Chat and social networking applications are good examples. They typically require a user to log on and the network to maintain knowledge of their location, a.k.a., "presence-based" applications. These applications are typically comprised of low-bandwidth text and graphics and are not associated, or bound, to a unicast video stream. Simple text and graphics can usually be processed by client devices using graphics overlay and are not directly associated to a unicast video stream (think "chat room associated to a live sporting event").

The ideal environment for immersive, video-linked interactivity, which infuses Web-like characteristics into TV shows and ads, is a combination of cloud computing and a traditional client-server applications architecture approach. Network servers have the processing power and resources to create media-rich, video-intensive applications. Client-side application execution can provide associated overlay applications, as well as detect embedded video triggers that can then be locally processed (on the client) to further enhance the viewer's experience by offering hyperlinks to related video or other applications.

An example of this could be a TV commercial with embedded EBIF triggers that enable telescoping into an ad microsite. The user reacts to the trigger, initiating a local application overlay. The overlay application displays several options, including the ability to see special how-to videos. The user selects a video of interest and the local EBIF application signals a network server, which initiates a unicast video stream in concert with the local application. The user is now seamlessly viewing a linked video. This video stream is inherently personal (as it is unicast) and may contain further embedded triggers. As the server can process user requests for complex media types and encode them to client compatible media formats in real time, the application may bring all the depth of a typical multimedia web site application to the viewer.

Meanwhile, a scaled version of the broadcast stream can be included in every scene, enabling the viewer to continue to view the original programming channel. Given appropriate "TV production values," the net effect on the viewer is no obvious "interactive" or "VOD" application, simply viewing a video with all the richness

Example of EBIF Triggers and Telescoping to an Ad Microsite

1. TV viewer is watching Trading Spaces, a linear TV program about home remodeling on The Learning Channel, when a two minute ad break begins

2. In the last spot, a Home Depot commercial airs. An EBIF on-screen prompt then appears giving the viewer a chance to see do-it-yourself videos

3. After pressing "OK", the viewer telescopes to a Home Depot-branded microsite featuring do-it-yourself videos, product demos and other features. Meanwhile, a scaled version of the linear broadcast plays in the bottom right corner

4. The viewer could watch do-it-yourself videos while learning more relevant products such about Black & Decker tools and Benjamin Moore paints

5. Home Depot can even choose to provide dedicated multi-layered showcases for each featured brand, effectively creating "microsites within a microsite"

6. At any point, the viewer could exit the Home Depot microsite and return to the full resolution linear broadcast

of the Web and the immediacy of the classic television presentation experience.

CONCLUSION

The optimal solution to provide the best viewing experience that combines the media richness of the modern Web with the quality of service and ease of use of traditional television uses a software applications platform that maximizes both server-centric computing and media processing power with client overlay and local application logic. The Web model, if used as a guide for modernizing and realizing the promise of the cable broadband television platform, would argue to use authoring tools and publishing workflow for applications targeted at both mainstream video and advertising applications (the latter, taking advantage of the oft cited "targeting" capabilities inherent in the Web model and when using unicast video applications).

The judicious use of server-centric software applications will ameliorate (but not completely eliminate) the problem of lowest common denominator set-top fragmentation. This is a serious issue when considering "scale." The server and client hybrid – one that maximizes the inherent strengths of a high-bandwidth, real-time, two-way connection between the headend and home, and that takes advantage of simultaneous MPEG and IP transport paths -- is the optimal approach to maximize the return on investment in software delivered multimedia applications.

COMMON ADVANCED ADVERTISING SYSTEM

Arthur Orduña
Canoe Ventures LLC

Abstract

The cable industry has a long history of working together to advance our services for the benefit of consumers and our commercial customers, including the advertising community. CableLabs, National Cable Communications (NCC) and the numerous interconnects that dot the map are examples of this cooperative spirit. It is a natural evolution that cable would work together to develop a standardized national platform for advanced digital advertising. That is the prime development task for Canoe Ventures.

At the heart of Canoe's ecosystem will be the Common Advanced Advertising System and its over-arching business process management system. This implies a massive data record keeping system and associated business intelligence capable of aggregating, storing, mining and reporting on extremely high volumes of data. Advanced analytics and inventory management systems are also crucial to reaching the full potential for addressability and adaptive campaign management.

COMMON ADVANCED ADVERTISING SYSTEM

Overview

The Common Advanced Advertising System (CAAS) consists of the national Advertising Stewardship business systems and enabling infrastructure that interfaces with advertisers and advertising agencies, media owners and MSO distribution systems to deliver a range of advanced digital advertising products. The CAAS will be a centralized control center and clearinghouse, and will communicate with many different MSO delivery systems.

To operate within the CAAS, it is generally acknowledged that the Participating MSOs must adhere to common guidelines. MSOs must provide assurances that each of their participating systems will:

- Support uniformity of product and ad formats
- Support the defined business processes
- Provide common system interfaces
- Provide consistent data per the agreed upon interfaces, semantics and policies

Given the difficulty of building the CAAS to support multiple heterogeneous MSO delivery systems, the scope will initially be limited to MSO video services. However, this will be expanded to include other platforms based on market demand (e.g. mobile, HSD portals, other service providers).

Reference Architecture

The reference architecture presents the envisioned end-to-end system and highlights the portion of the system for Canoe. The underlying MSO distribution systems will use a common set of interfaces, however the specific design may vary by MSO or systems within an MSO.

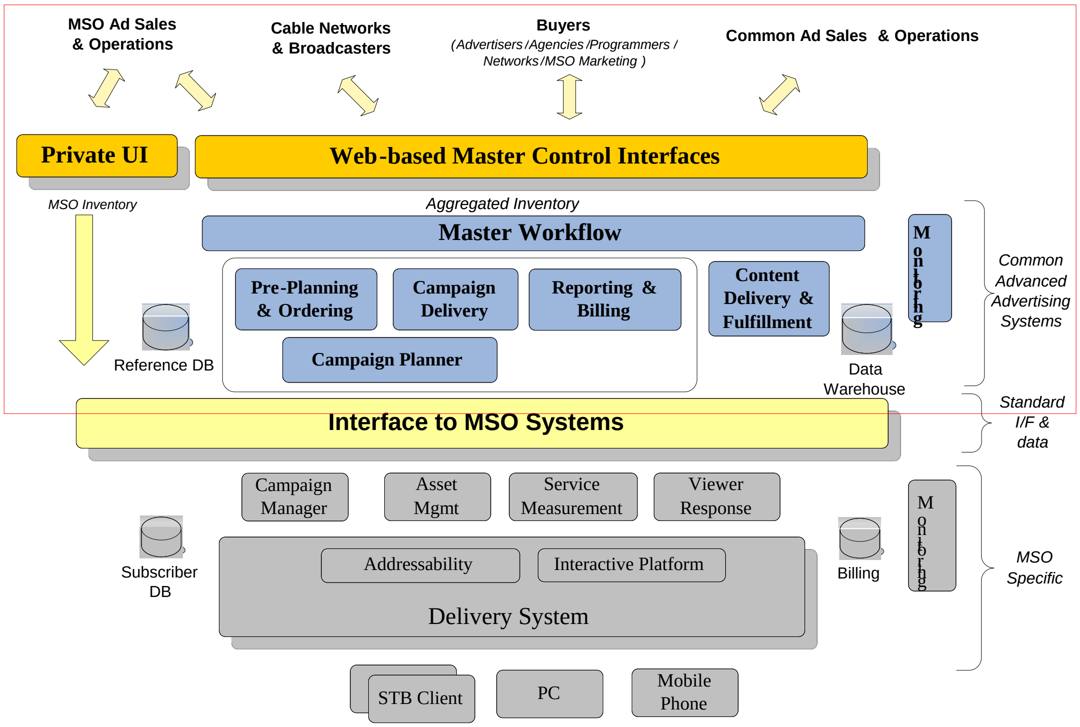


Figure 1 - Reference System Architecture

The following reference diagram deconstructs the CAAS into its primary elements and business processes.

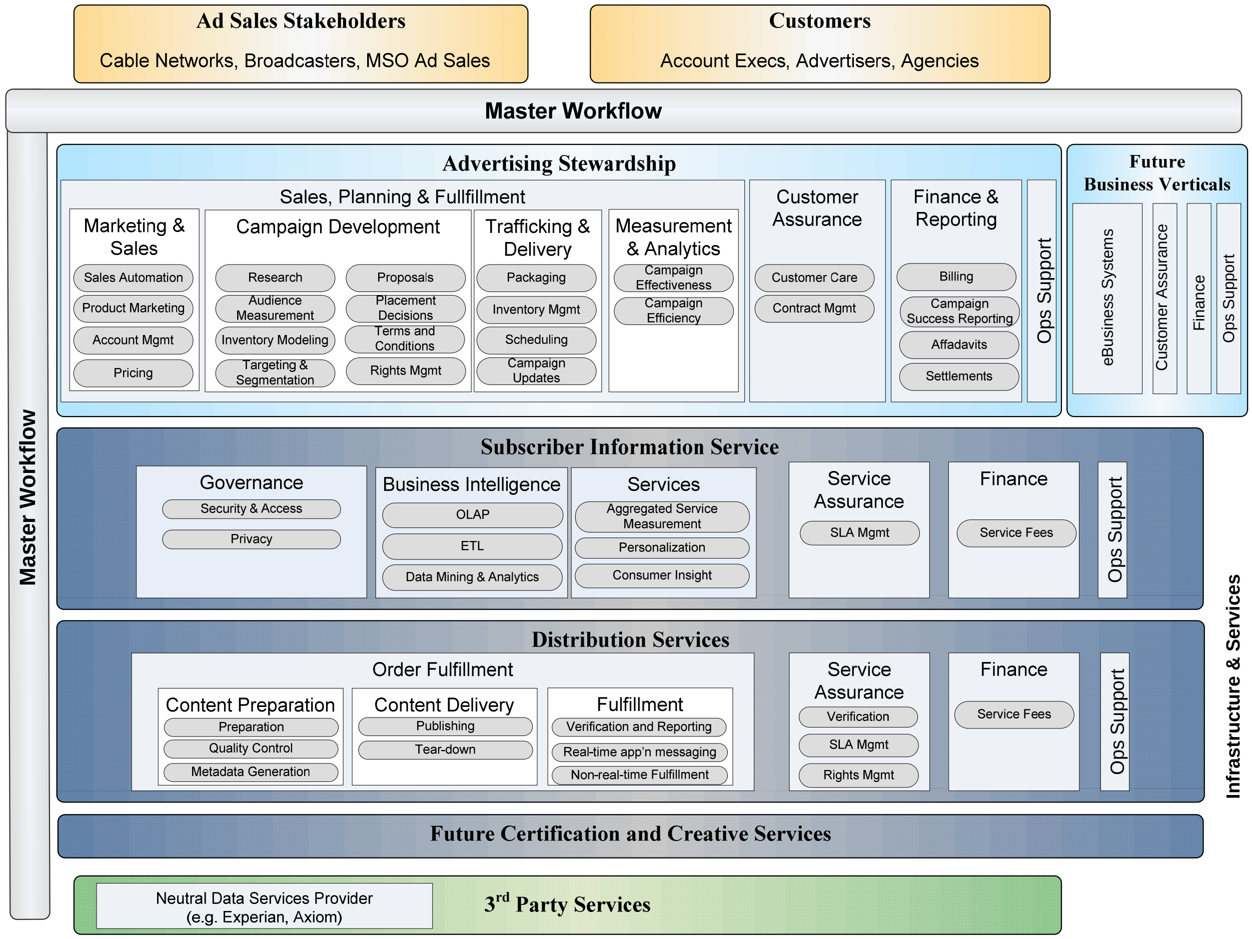


Figure 2 - CAAS Business Processes

Advertising Stewardship enables a national sales organization to manage the complete ad sales lifecycle. Web based interfaces will provide the necessary access for both ad sales stakeholders and customers for sales and campaign development, in-flight campaign management and post-flight reporting and financials. These systems will rely heavily on the underlying infrastructure services as well as interfaces to MSO systems.

The Subscriber Information Service (SIS) is the primary data warehouse aggregating audience and service measurement data from all of the distribution systems. It will supply the necessary storage, processing, and business intelligence functions to support the Ad Network, and could possibly support future business verticals. The data within the SIS will support a wide range of data mining and analytics process, one of the most important being audience segmentation to support targeting. Through the use of bonded 3rd party data services providers, blind-matching can be performed against MSO and advertiser customer lists while maintaining full consumer privacy.

The Distribution Services are intended to simplify the process of creating, delivering and fulfilling digital advertising across multiple MSOs. While customers could interface directly with the MSO interfaces, this service greatly reduces the complexity and customer overhead by providing a single organization to handle content preparation and delivery across many different systems, and potentially different target platforms such as wireless or broadband. And, while it is the long-term objective for all MSOs to use common content formats and delivery standards, in practice it will take time for all systems to reach a consistent level of deployment, and this service enables a more manageable migration path.

Tying all these systems together is the Master Workflow. Workflow automation will connect the many business processes and stakeholders with automated and traceable processes. Because of the many stakeholders and complexity of the business processes, a highly sophisticated, flexible, and secure solution based on proven technology is mandatory for successful operation of the system.

Workflow

The following diagram illustrates a high-level workflow between sellers, buyers, the CAAS and MSO systems.

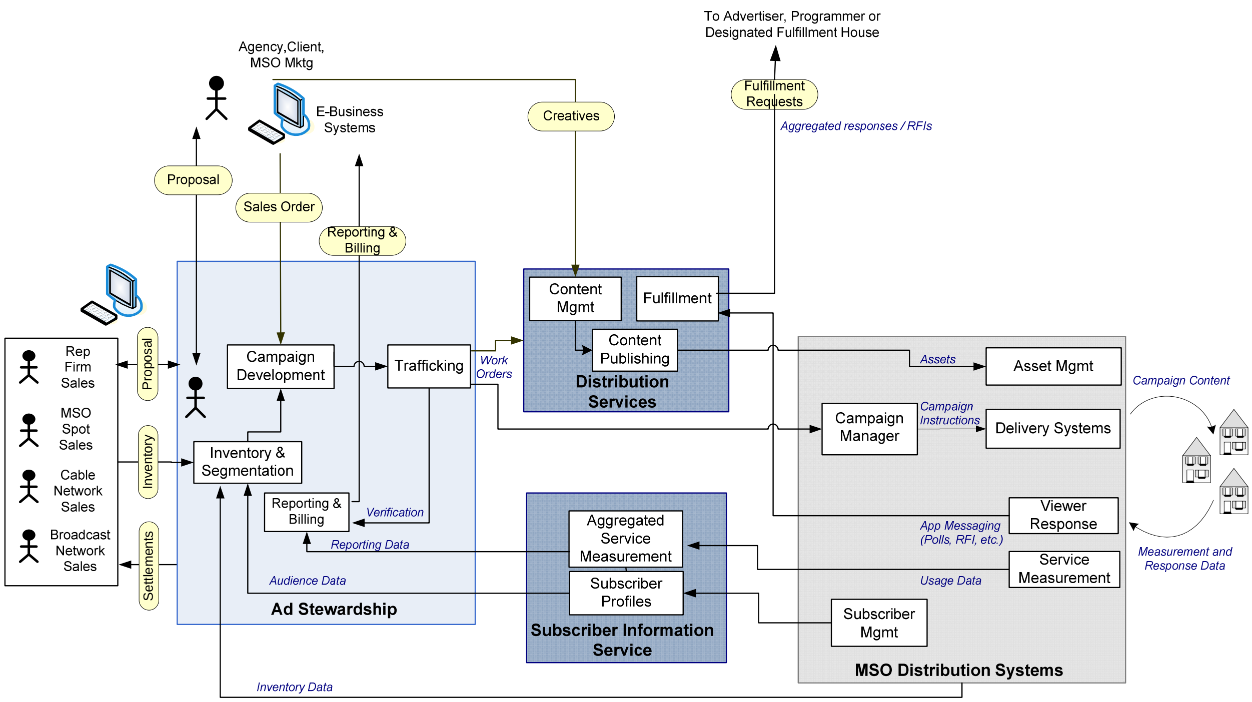


Figure 3 - Example Workflow

Design Objectives

Time to market is of paramount importance for the Common Advanced Advertising System. Therefore, it is highly desirable to minimize custom development, and maximize the use of existing or off-the-shelf systems. However, it is important that the recommended systems meet, or will meet with modifications or enhancements, the complete set of CAAS requirements.

Within the above overarching guidelines, our design approach has been to:

- Evaluate best of breed products that provide functionality in line with the CAAS requirements and long-term objectives
- Minimize unnecessary custom work
- Enable procurement flexibility and product innovation
- Limit unnecessary dependencies and integrations – balance the benefits of off-the-shelf products against vendor management complexity

The resulting design centers on a loosely coupled architecture framework with open interfaces enabling multi-vendor solutions and independent operation of distinct elements of the system. For example, the Subscriber Information Service is a complex information system that might benefit from different technologies, physical distribution and/or hosting models than the Advertising Network. Part of the design exercise weighed the different models and present trade-offs in terms of reliability, extensibility, scalability, operability, cost and time to market.

It is also crucial to define and automate the business processes using proven workflow automation technology. Because advanced digital advertising is still a nascent business, a flexible rules engine and defined roles and permissions are mandatory in order for the system to keep pace with changing business objectives. The workflow analysis should place equal emphasis on core business processes and operations monitoring and quality control functions.

<u>Design Considerations</u>

There are critical system elements that are expected to require significant new design.

<u>Information Systems</u>

A robust data infrastructure coupled with the business intelligence necessary to extract maximum value is the cornerstone of this effort. As the business objectives for the CAAS evolve, and as Internet advertising continues to experience strong growth, the data required to enable this effort has also increased in magnitude and importance. The ability to accurately target and adapt campaigns, and thus the required Information Systems to effectively manage, analyze and securely govern very large amounts of data are critical to the success of the Common Advanced Advertising System.

Table 1 presents the range of data that passes through the system. The design of the CAAS and MSO system interfaces will cover these as well as anticipated future needs.

Table 1 - Common Advanced Advertising System Data Needs

Reference & Stakeholder	Order & Execution of Order	Operational	Reporting
Agencies	**Order Details**	**Service Level Management**	**Report Types**
Networks	Products	**Capabilities**	Delivery Confirmation
System Configuration	Price	**Availability**	Affidavit
SYS Code	Rules	**Monitoring**	Operational Reports
STB	Financials	System Wide	Technical Operations Reporting
MSO Subscriber DB	Content Reference	Regional	SLA Management
HSD (Broadband)	**Expected Info**	Full Footprint	Sales Pipeline
Wireless	Reporting		Complete Financials
Historical Orders	Affidavit		HR People Reporting
	Verification		MSO Report Types
Planning	Enhancement		Inventory Utilization
Proposal data	**Approval**		Aggregate Financials
Inventory	Order Expressions		Dashboards
Dynamic	Confirmation		Historical
Capacity	**Management**		Category
Product	Revisions		Sector
Pricing	Change orders		Performance
Business Rules	Cancellations		Comparative reports
Permissions	Make Goods		Co-op affiliate reporting
Reach	RFI Fulfillment		
	Polling Results		
	Performance		
	Capture more than will report		
	Clicks associated with Views		
	Service Delivery		
	VOD		
	STB		
	Raw Data Sets		
	Impressions		

Initial Sources for service measurement from MSO subscribers includes but is not limited to:

- Linear ad insertion
- Settop box usage, including linear viewing, DVR and time-shifted viewing, and Interactive usage
- On demand usage

As a service provider, adherence to applicable consumer privacy laws and policies are always of critical importance. Any design must include strict data governance rules and auditing capabilities.

Addressability

National, Regional, and Local Advertisers want to segment customers, develop targeting profiles and identify targeting groups, deliver specific messages to each targeting group, and determine the efficacy of their messages.

The primary components of the targeting capability include:

- Definition of targeting profiles to match the advertiser's segments of interest using data from multiple sources.
- Identification of the targeting groups in an iterative process of refining the profiles to maximize the reach (e.g. how many households) – through blind matching and generating counts.
- Creation of advertising packages consisting of message, creative, and metadata that are then trafficked and delivered.

Segmentation can be applied to the subscribers' profile for household (and eventually set-top box level) level targeting.

The efficacy of targeting is measured through key metrics that are specific to the advanced advertising product (e.g. linear addressable, RFI, etc.).

The targeting capability is largely a part of the Campaign Development step in the advertising lifecycle. It is applicable to all the products being contemplated within the CAAS advanced advertising tracks (e.g. targeted overlays, linear addressable, etc.).

It is envisioned that the MSO may require a 3rd party data services provider as an infrastructure partner for meeting Personally Identifiable Information (PII) requirements, and to provide support for advertisers in understanding and applying targeting to their advertising campaigns. In most cases, a "blind" data match will be required for qualitative profiling to protect the privacy of the specific subscriber file and information. Clarifying the relationship between MSO data systems, the CAAS and 3rd party partners and enumerating the trade-offs for structuring the underlying data systems is a critical design objective.

Inventory Modeling and Placement Decisions

The inventory management requirements for the CAAS are extremely complex and dynamic given the number of distribution systems and the different types of inventory represented by the envisioned set of products. It is the objective for distribution systems and media owners to communicate inventory status to the CAAS through well-defined interfaces. The CAAS will correlate this information and use it to support inventory modeling during campaign development and to drive trafficking and placement decisions.

Invoicing and Accounting

In order for the CAAS to offer a cohesive buying and tracking mechanism for customers across the Cable footprint, it must provide customers with a single invoice across media owners and product lines. To accomplish this, the design must include billing and accounting solutions focused on a secure infrastructure that supports accounting standards and provides proven transactional integrity. The billing design will accommodate the creation of customer invoices across companies and products as well as the distribution and tracking of these invoices. Any associated accounting transactions will be tracked in concert with GAAP Accounting standards and support customer collections as well as vendor payments. Finally, the design will allow for periodic reconciliation and settlement of financial transactions between participating companies.

Workflow Automation

As illustrated in the CAAS reference diagram, all of the business processes will be orchestrated through automated workflow. The workflow engine will support flexible and extensible rules definition and will provide multiple views into the current state of any campaign based on stakeholder access rights. Due to the highly distributed nature of the problem, it is very likely for campaigns to encounter errors somewhere in the process. For example, content may not be successfully received at all sites, or rights may not have been cleared by the required date. Therefore, the solution should also provide active monitoring and alerting capabilities for defined error conditions.

Interfaces with MSO Systems

An abstraction level – called the MSO System Interfaces – will be designed to allow data and message exchange between the CAAS and the specific systems and equipment at the MSO delivery sites.

A critical activity during the design phase is to perform a detailed analysis of the data and messages that are used by the MSO delivery systems shown in Figure 1. Canoe is working with each Participating MSO in order to collect a full set of requirements to describe the data and messages that must be generated for and collected from the MSO delivery system in order to ensure proper operation within their systems. This may require analyzing the capabilities and interfaces of deployed traffic and billing, VOD, and interactive systems to fully understand the how these systems will interoperate within the end-to-end solution. Individual MSOs will make independent technology and procurement decisions regarding their delivery systems, including the portions of those systems that interface to the CAAS.

There is a strong desire that the interchange between the CAAS and the MSO delivery system be governed by standards-based implementation. This covers the current suite of relevant standards and specifications from ANSI, SCTE, CableLabs, IETF, and other involved entities. Of particular interest are SCTE-30, SCTE-35, SCTE-104, SCTE-118, SCTE-130, DVS-766, CableLabs ETV, CableLabs Metrics, and several subprojects of the CableLabs VOD Metadata and OpenCable initiatives. These are key to the development of the MSO delivery system interfaces.

Standard	Specification
SCTE 30	Digital Program Insertion Splicing API, SCTE 30, 2006
SCTE 35	Digital Program Insertion Cueing Message for Cable, SCTE 35, 2004
SCTE 104	Automation System to Compression System Communications Applications Program Interface (API), ANSI/SCTE 104 2004
SCTE 118	Program-Specific Ad Insertion - Data Field Definitions, Functional Overview and Application Guidelines, ANSI/SCTE 118-1 2006
SCTE 130	Digital Program Insertion – Advertising Systems Interfaces, SCTE 130
SCTE 138	Stream Conditioning for Switching of Addressable Content in Digital Television Receivers, SCTE 138 2009
OC-ETV-AM	OpenCable Enhanced TV Application Messaging Specification, OC-SP-ETV-AM1.0-I03-060714
OC-ETV-BIF	OpenCable ETV – Binary Interchange Format 1.0, OC-SP-ETV-BIF1.0-I03-060714
OC-SP-METRICS	OpenCable Receiver Metrics Gathering Specification, OC-SP-Metrics-I02-070416

Table 2 - Standards Reference

Interfaces with Agency and Advertiser Systems

Canoe will also define an approach for data exchange with the Advertiser and Agency community. Again, there is a strong desire that the approach between CAAS and its customers move toward a standards based approach leveraging and driving the creation of standard specifications within the Advertising community (AAAA). However, it also must be recognized that there will be an ongoing need to electronically transact business with this community within the constraints of the capabilities of legacy stewardship products in use today.

This effort will involve not only defining the system components and communication standards referenced within this document, but also integrating the new products into the document trafficking workflow, protocols and formats currently in place with Legacy Stewardship systems for various documents including proposals, orders, revisions and invoices.

This demands familiarity with existing AAAA common documents and protocols, work with MSO and National Sales, as well as with the lead software vendors in this space, in order to define an approach for document exchange that can support both current and future capabilities of these platforms.

Conclusion

This paper described at a high-level the functional architecture of Canoe's Common Advanced Advertising System. It is our intent, in publishing this paper, to educate the technical community on what we are designing and building for our various stakeholders.

It is our goal that the successful launch of the CAAS will mark the first iteration of a standardized national platform for advanced digital advertising, resulting in additional and ongoing innovation from multiple parties – not just Canoe and the MSOs - at the services and application levels.

CONNECTING CABLE'S NETWORK AND APPLICATIONS WITH ITS INDUSTRY PARTNERS WHILE MAINTAINING OUR INDIVIDUALITY

Mitch Weinraub
Thomas Engdahl
Joshua Seiden
Comcast Media Center (CMC)

Abstract

The cable industry is working aggressively to expand the range of services that meet consumers demand for personalized programming. In order to deliver these new services efficiently, the industry is changing its video delivery infrastructure towards architecture that is open and network –based, connecting headends to backbone networks and using data to drive events.

From its position at the intersection of programmers, advertisers, application developers, MSOs and hardware manufacturers, the Comcast Media Center (CMC) has been able to develop a unique perspective on the current state of systems development and the gaps that are yet to be resolved. From this intersection, it has become apparent that the cable systems serving standalone markets will want to join larger MSOs in expanding their level of connectivity, both vertically and horizontally, in order to the opportunities that are emerging succeed in today's market place.

The connections that the cable industry is making fall into two primary categories. First, the industry will want to create tighter connections to existing partners and other MSOs. Second, the industry is extending its connections to key partners in the media and entertainment industry and to advertisers. *Perhaps more importantly, cable systems will also want to expand the connectivity initiatives that involve seemingly competitive industries, such as mobile media providers and internet advertising structures*.

INTRODUCTION

Throughout the history of cable television in the United States, the business has been based on local providers providing a unique service in each market. Besides providing an acceptable mode of operation, it has also proven valuable in competing against satellite and other providers that, due to their technological differences, were forced to deliver a homogenous service to the entire country.

Even as some cable MSOs grew larger and began to align their services, a number of traditional cable aspects had remained unchanged. More recently, cable MSOs have begun to standardize their navigation, packaging and other offers across their footprint, but these standardizations often remain within a particular MSO.

However, there is less standardization between MSOs. Examples include major market DMAs served by two or more cable systems that haven't created ad interconnects. It is becoming easier to make these types of connections thanks to standardization among cable operators, as evidenced by CableLabs' recently issued specifications for Modular Headend Architecture.

Another example of movement toward greater standardization is the multi-MSO funded organization known as Canoe Ventures. With this organization in place, some major, positive steps are being taken to make sure that interactive and advanced advertising efforts are as standardized as possible. This, when successful, will allow for the widest possible footprint for new and exciting forms of advertising.

The industry's move toward greater interconnectivity will support the steps being taking by initiatives such as Canoe in order to offer highly competitive advanced advertising and other new business opportunities in today's environment. As this paper will demonstrate, one of the major challenges facing the cable industry in the near future is the ability to maintain the local and individual aspects that have defined cable throughout its history while making these new connections..

The significant and meaningful connections that are emerging include using SIP (session initiated protocol) and other key technologies to connect service platforms between MSOs, connecting to content providers, broadcasters and major studios, connecting with interactive television authors and platforms, and connecting with prominent advertisers and their agencies. These connections allow the cable industry to continue to grow and remain foremost in the minds of consumers as the primary service provider for entertainment and telecommunications services.

Things Changed When Consumers Became Connected

Some business historians may trace the beginning of a connected marketplace back to the use of the U.S. Highway system, a national telephone system, railroads, or even the Wells Fargo Wagon. Similar to these past significant events, the ability for consumers to go online has served as a major catalyst for the connected marketplace. Thanks to the Internet's World Wide Web, services now had the ability to aggregate communities and consumers in ways not possible before. In a significant way, this added a new dimension to business and created brand new opportunities. Before this change, it was possible to be successful by providing a good product or service to a limited (in both size and geographic diversity) market. Local stores flourished and providers that best serviced a focused marketplace won the day.

When consumers went online, this all began to change, but it took a while. At first, online services looked a lot like the off-line services that preceded them. Compuserve, Prodigy, and America Online became the dominant players. Each provided a suite of services that was individual, unique and catered to the specific users that they were trying to attract. And for a while this worked well. Even as the web began to emerge, these services lived on as the best way to connect to the web as well as to access additional "walled-garden" services. However, as web based services began to add features, the walled-garden approach became less and less valuable. Web based services - those that allowed access by millions of consumers regardless of their internet service provider (ISP) or means of access - quickly became more valuable than those (within the walled garden) that only served members of the private club.

The recent Internet success stories have been services that reach out to the widest possible audience. Whether you look at it from the perspective of a "tipping point" or a "network effect", the volume of users doesn't just demonstrate linear growth for these services, it is also responsible for the basic success and is the only way the service could survive. To paraphrase the famous line from Marshall McLuhan, in the new business environment, it could be said that the mass is the message.

One example of the mass of a service being critical to its success is eBay. eBay began its life as Auctionweb, a component of the online services offered by Echo Bay, a consulting firm that served the San Francisco bay area. The Auctionweb portion of the site quickly grew to become the entire Echo Bay site, which was renamed eBay and broadened its scope to serve all comers on the web.[i] Once available to the world, the volume of buyers, sellers and items available grew exponentially and the rest, as they say, is history. In this case, the mass of the offering has completely transformed the service.

Likewise, YouTube.com and social networking sites like Facebook provide excellent examples of mass or scale. With the entire world of on-line users able to participate, these services have been wildly successful (from a usage if not yet a financial perspective). For these services, the network effect or mass is they key to their success. Their features may not be extraordinary or cutting edge, but the sheer number of users and connections that can be made has been integral to their success.

A number of additional examples could be cited, but the critical point here is that these services represent a new kind of business success. The success of these kinds of services wouldn't have been possible without the mass or scale of the connections being made. These services could not survive in small or isolated environments. In fact, it could be argued that they are successful primarily because they didn't isolate themselves but opened and connected their services to a bigger market.

This is the model that the cable industry is beginning to follow with its deployment of advanced video interactive services. In today's marketplace there are multiple competitors aiming to deliver service, many in a non-traditional fashion without even owning their own networks. In response, cable operators are opening their networks to connect a wider range of services to their customers, markets and vendors, which allows the industry's new advanced services to be successful. In this environment, the whole will be greater than the sum of the parts. This requires having a (big) slice of an extremely large pie as opposed being limited to a single cupcake. The following sections of this paper outline specific areas of the business where the potential for these opportunities may be realized through building connections.

MSOs NEED TO CONNECT MORE THAN EVER BEFORE

In the past, cable was a parallel sort of service where each head-end could handle and process programming independently, without any connection or knowledge of one another. In the case of traditional linear services (like CNN and HBO) it didn't matter if there wasn't an interactive connection. Conversely, the new advanced services that the industry and its partners are striving to create are often the kinds of services that can't succeed or perhaps even survive without extensive connections. From gaming to advertising to shopping to new ways for viewing content, the cable industry is poised to play a big role in the interactive, connected and personalized service offerings of the next decade. However, many of these services fit the new business models outlined above. This means that they will require mass and scale as opposed to closed systems in order to succeed.

Canoe Ventures is already driving some of the key connections that will be required to enable advanced advertising related services. These efforts are expected to prove extremely valuable and successful for all participants. This section of this paper focuses on the connections between providers that will be necessary to support and drive the growth of services outside of those directly related to advertising, which Canoe is already addressing.

Some of the most basic interactive services being developed today are in the area of gaming. While these games are fun and entertaining on an individual level, they really become interesting when individuals can play and measure their performance against others. Although larger MSOs may be able to aggregate enough of a user base within their own footprint, this type of mass market just won't be possible MSOs serving smaller markets or single headends. In addition, game developers will have the opportunity to create truly "game changing" gaming opportunities when they can draw upon and rely upon a large and diverse user base, such the one available via the Internet.

Likewise, voting and polling applications are often sited as early opportunities for cable's new interactive platforms. Like gaming, these applications thrive only when the base of users is large enough. One of the reasons that American Idol works so well and attracts such a large audience is that anyone and everyone can cast their vote and be part of the process. In order for EBIF and tru2way™ applications to become the new standard for consumer interaction, they will need to work everywhere -on every cable system regardless of size or location.

Additionally, new cable interactive services will need to connect to other systems that may not be based on the same middleware. For instance, cell phone text messaging systems, 1-800 numbers and websites can already cast votes and aggregate polls. Applications riding on the cable industry's iTV platforms will need to connect to these existing platforms (and others yet to be invented) in order to gain the mass and scale necessary to be successful.

Like the games and voting/polling services described above, both social interaction services (living room versions of Facebook, etc.,) and content sharing services (such as YouTube) will also require the interaction of extremely large number of users in order to succeed. As a result, cable and other last-mile service providers will want to find ways to connect their offerings and share or aggregate their users and interactions across the video universe, as well as extending to online, mobile and other electronic platforms.

While applications or content services providers require open systems that support the seamless delivery of their services, it will be important for the connected ecosystem to support each cable system's or MSOs requirements for certain forms of autonomy. In addition to differences with local navigation, each system or MSO will want to maintain the look and feel that supports its brand identity as the provider of services that are delivered over the local network. Fortunately, the connections required on the back-end that allow for mass and scale to be aggregated don't need to affect the front end where user interfaces and branding opportunities are presented. Using structures defined later in this paper, we can accomplish one without impacting the other.

In the case of cable system operators and other last-mile service providers, the question is “How can these connections be made?” The initial plans for most of the services described above focus primarily on the local servers necessary to deliver applications to customers. However, the connections required to allow these services to work well and really grow are as important as the core functionality. Examples like cell phone call processing or SABRE in the airline industry represent forms of “coop-etition” where competitors in the same marketplace have come together to create common systems allowing an entire industry to grow. For example, the SABRE airline reservations tool is used by more than 90 airlines across the globe. This has enabled single interfaces and on-line booking services, but it wasn’t always the case. Likewise, it wasn’t always so easy to get your cell phone to work across different phone networks. These industries have found a way to share data and transactions. This is the model that the cable industry could follow in order to achieve the full potential for advanced services while maintaining differentiation.

For the cable industry, this involves establishing standard and trusted points of data exchange. In this scenario, each cable system operator would need to trust a centralized service that will allow all players (cable and non-cable alike) to publish high scores and aggregate votes and user generated content in a way that allows the entire pie to grow. Historically, there are examples of cable MSOs that banded together to address this need and Canoe serves as the latest example. Achieving these requirements would involve:

1. Establishing a high level committee that would agree (on behalf of the interested MSOs) on the definition of centralized services for data exchange.

2. Creating an RFP to allow interested vendors to describe the details of the systems necessary to support the exchange.

3. Establishing a timeline for the deployment of a beta system that application developers would use to test against.

Making these connections creates numerous service opportunities. The biggest opportunities typically lie in the services that will be invented in order to take advantage of the connected platform. However, some of the services that can quickly emerge include:

1. **Television- based social networking –** Whether more like Facebook, Dating on Demand or some combination of the two, meeting and sharing with others across the country and the world will undoubtedly be significant.

2. **Television program related sharing –** Examples include real-time on screen fantasy sports matchups; or what to watch next - recommendations from friends or just real time program chats with social connections. Interactive applications that directly relate to television content are also likely to proliferate.

3. **Meaningful Real-time Voting and Polling –** When most if not all viewers can be connected and

aggregated, voting and polling with immediate results becomes possible.

4. **Real-time competition on a mass scale –** ideas in this space include massively multi-player quiz shows or prediction contests. When the industry is able to match the quality and presence of television with real-time, two way connectivity on a large scale, the kinds of games customers are likely to play will evolve and expand past what can be envisioned today.

As mentioned above, the best ideas are almost always those that can't yet even be imagined. Innovations that creative developers using tru-2-way can come up with in their proverbial garages can't even be imagined until the world that provides these kinds of connections comes into existence. The successes of the internet are a great reminder that vast, new capabilities are a prerequisite for game-changing successes such as eBay and YouTube. History has also shown that replicating walled garden environments, such as Prodigy or Compuserve, on a big a screen in the living room may not be enough take to take the industry where it wants to go.

TECHNOLOGICAL REQUIREMENTS FOR THE CONNECTED ECOSYSTEM

Offering the next generation of services described above will require video services to become IP-based. This will allow for new applications to utilize the Internet to provide enhanced services as well as connect to existing Internet-based services and applications.

The primary goal of the industry is to not only increase the choice of applications offered to customers, but to also begin to leverage ubiquity to generate new forms of revenue. The current approach for video delivery in the cable industry focuses on "every MSO as its own entity". The resulting effect is that some MSOs find themselves able to create new revenue streams while others struggle to follow on the coattails. As history demonstrates, innovation occurs at a faster level when the cable industry can leverage its existing capabilities in other platforms and implement them across the majority of cable operators

As organizations such as Canoe focus on deploying interactive ads nationally, they cannot be as successful without ubiquity. Ubiquity is achieved when a single application can be distributed and function across the entire cable footprint, in every environment. The most straightforward approach for accomplishing this objective would be to construct applications generically, using generic interfaces, thus preventing the need to create new versions of applications for every single deployment in the country. This is obviously a daunting challenge, but with a few changes to how the industry does business today, it can be accomplished.

The Elements of Application Creation and Distribution

EBIF (Enhanced Television-Binary Interchange Format) and tru2way represent set-top abstraction layers. While these two abstraction layers work in different ways, they accomplish a similar task. Deploying applications that run in EBIF or tru2way, make it possible to have the applications be portable from one platform to another without modifying the application's code.

Application success is typically defined by both reach and use. It will be easier to achieve national distribution if only one version of the application exists. One of the challenges to overcome is allowing each MSO to retain its own look and feel while all MSOs are using the same application. This can be accomplished more easily if first, the application is broken up into several components, second, it relies heavily on using library-type functions, and third templates are used for allowing the application to adapt to the unique elements of a local environment without touching the software. The nature of how the software is structured can be expanded from its current functionality to achieve these goals.

As application developers innovate these new products, implementation and adoption will be easier for those that implement proper design and technique. Applications developed with industry-approved Software Development Kits (SDKs) keep developers within the proper guardrails for the EBIF client or tru2way stacks deployed in the field. In addition, applications that implement Templates will allow for quick customization for new skins, new behaviors, new look and feel, without modifying the core code.

Once well-designed applications are created, they will need to be deployed quickly and efficiently. Understanding that MSOs may deploy a wide variety of carousels and infrastructure in their networks underscores the need and benefits of using standard data formats such as CODF (OpenCable Content Definition Format). An important goal is to avoid the implementation of proprietary interfaces in the application distribution, data gathering and reporting chain. A "publish and distribute" architecture available to all content providers and application developers, and subscribed to by multiple cable operators, can help to dramatically accelerate the deployment of national applications in rapid fashion. Once this cross-MSO CDN exists, an advertiser or application developer can gain access to a nationwide footprint at a fraction of today's costs.

It is conceivable to implement intelligent features into this application delivery CDN, such as the ability to customize applications on the fly. Creating this abstraction layering capability can allow an application developer to ingest a single application into this CDN. An automated workflow engine will attach or bind the application to any number of individual templates. The results of such a process work for both the MSO and the content provider. Each MSO can receive a custom version of the application that is branded and behaves in a way that maintains the common look and feel of their network. Multiple content providers and advertisers can use the same application and achieve similar functionality, but with a look and feel that responds to the content.

Set-top Abstraction Layers

Supporting these new interactive applications on cable plants will require leveraging new open architecture technologies such as EBIF and tru2way to support widely deployed applications across a wide variety of headend networks. EBIF and tru2way provide abstraction layers to both the network and the set-top box in different ways. EBIF is a data format that allows an EBIF client on a set-top box to act as a presentation engine for EBIF applications. Tru2way, conversely, enables an entirely new platform and infrastructure. Tru2way will also allow developers and advertisers to create widgets that would work with video-based applications similar to the role of widgets with Internet-based applications.

Due to the existing footprint of well over 60 million EBIF-capable set-tops, EBIF is becoming the technology of choice for advertisers and application developers looking to immediately deploy their applications. In addition, cable MSOs are also deploying infrastructure to support tru2way applications.

Interactive Application Sources

Interactive applications will originate from a variety of sources, including interactive content authors who are new to the cable industry.. As previously outlined, the open, global and interactive characteristics of the Internet have allowed it to benefit from the development of new and innovative applications. By creating translation engines and other types of interfaces to these applications, the cable industry can accelerate the deployment of complex, cross-platform applications on its video platform.

Abstracting Cross-Platform Communication

Because EBIF applications have the ability to communicate with application servers on the Internet, those application servers can achieve a cross-platform presence. An example of this would be an EBIF application that communicates over the Internet to a Facebook application through a proxy on the cable headend network. Once the connection to the Facebook application is made, that application can leverage the cross-platform footprint of Facebook and carry the customer experience onto iPhones, Blackberrys and other mobile devices. Leveraging applications such as Facebook in this manner allows an MSO to use the cross-platform connectivity of Internet-based application as an abstraction layer. While the actual EBIF applications running on the set-top box will be a simpler application than the Internet-based applications they can be utilized to initiate this cross-platform experience and provide customers with the type of global connectivity and interactivity they are seeking.

Maintaining MSO Identity - Application Customization

As more and more applications become available, maintaining a common look and feel on the MSO's network that is also consistent with their navigation paradigm will become an important part of successfully deploying applications. Unlike applications on a PC, iTV applications are presented as a service offered by the MSO. While application vendors strive to have their own look and feel. A customer who interacts with many applications from multiple vendors may find a different style each time they launch a different application. This could be unsettling if an application loads or navigates differently from the other applications on the same cable plant.

Allowing each MSO to maintain consistency with their navigation paradigm involves building applications that can be customized to "fit" an application to the MSO's network. However, application customization will impede progress unless it is done in a way that reduces, or even eliminates, re-testing the application on each MSO's network. Rapid deployment and wide-scale adoption cannot occur when there are multiple testing and certification processes. Instead, the most successful applications are likely to be those that "white-label" their applications, allowing each MSO to "skin" the application to their standards. Options include allowing each MSO to use its own colors, logos, and messaging. Expanded functionality would include capabilities to drive customers to other services like VOD, or launch other interactive applications. For example, an MSO or content provider may deploy an application that is bound to a specific television show (i.e. voting and polling, etc.) In addition, the MSO or content provider may choose to add a button on the application that can launch VOD sessions of older episodes of the program.

News and weather applications will also require a level of customization. These apps will require the ability to utilize localized feeds as a source of data.. Traffic reports, local news and sports, and local weather may come from existing RSS feeds or other data feeds. A single application must be able to define all of these data sources, skins, and behaviors without modifying the source code of the application.

Standardized Infrastructure requirements

A key success requirement for quickly deploying applications on a cable plant is to limit the amount of required testing while ensuring that the application will execute safely and perform well. Accomplishing this will require both the industry's continued embrace of open standards and focusing on building a standardized architecture.

A variety of technologies may be used for developing this standardized architecture. First, by standardizing on a few software development kits (SDKs), application developers can be assured that they are developing their applications within a certified and tested framework. Second, the use of templates will allow for advertisers and application developers to quickly customize and deploy applications. Finally, and as previously described, by standardizing the back-office infrastructure required to support interactive applications, cable systems will be able to deploy applications developed by multiple application vendors across the same infrastructure.

SDKs

Widely used for online interactive applications software development kits (SDKs) are typically a set of development tools that allows developers to create applications for a particular software, middleware or hardware. The iPhone is a great example of how an SDK can be leveraged to successfully build a variety of applications. Application developers need to know little about the iPhone if they have the proper SDK to build applications. The iPhone SDK keeps developers confined to the capabilities and limitations of the device.

The cable industry currently has several choices when it comes to EBIF and tru2way SDKs. As the technologies mature, the choices of SDKs are likely to consolidate, thus providing a consistent and common toolkit for developing interactive applications. Common toolkits allow developers to leverage a common framework, and help to ensure that applications will exhibit similar behaviors. The benefits to this approach are that the resulting applications will tend to have a common look and feel, they will behave similarly, and they tend to be more predictable. This approach will also serve to shorten the test time.

Moreover, applications developed with an EBIF or tru2way toolkits have many advantages over applications developed by proprietary tools. SDKs provide an environment where a developer can simulate the application's behavior before loading it on a set-top box. This allows for a certain level of testing of an application and the middleware before the code is loaded on a set-top. Developer toolkits also help to keep software developers within the guardrails allowed by the EBIF client or the tru2way stack. As a result, applications are developed with "safe and tested" methods, thereby making the code less intrusive than a native application.

Templates

What if an organization wants to develop interactive ads or applications but it doesn't have the luxury to hire its own software development staff? For example, an advertiser may want to implement an interactive EBIF ad and distribute it into a national broadcast. If the advertiser doesn't have a software development staff, or doesn't want to hire a third party to develop a customized application, it could still implement the application by utilizing "templates".

The concept of templating is used throughout the Internet advertising industry today. Advertisers can purchase a template and quickly customize it for their advertising needs. Their main benefit is that the deployed ad is a fully tested application without the advertiser having to invest in software development. For example, advertisers can purchase Google AdWords templates to create Internet-based ads for their business. The templates require little software development experience and allow advertisers to instantly deploy ads that display during internet searches or navigation to topical websites.

Interactive application templates will come in two forms: content-driven templates and application-driven templates. Content-driven templates implement a single function, but they can be configured to present entirely different sets of data. Request for Information (RFI), VOD Telescoping, and Voting/Polling applications are examples of content-driven templates. An example of a content-driven template is a voting and polling application. A Voting/Polling template allows the content provider to modify the look of the application and to also define the data that it will present, such as a question on which contestant to vote for or a set of trivia questions.

The second form of template is an application-driven template. Application-driven templates tend to be more complex applications that were originally developed as stand-alone applications, like electronic program guides (EPGs) and fantasy football applications. Application-driven templates will allow MSOs to customize the application to achieve a similar look and feel to other key navigational functions and to provide a consistent experience to the consumer.

Implementing an application-driven template will allow an application developer to develop a single application and customize it for every customer quickly and safely.

Standardizing back-office requirements

SDKs and Templates simplify the work on the client-side. However, the back-office and infrastructure required to run these applications also needs to be standardized, as described earlier. CableLabs and the cable industry are currently working on ways to standardize back-office requirements. This will be essential to allowing cable video customers to experience e-commerce and other transactional applications with the same ease of use that they enjoy with online shopping or ordering an On Demand movie.

APIs

The complex testing of applications occurs today because it is necessary. Most of the software deployed on current cable plants is hardware-specific, so any changes or variations can cause problems in the network. However, implementing new services with abstraction layers and standards not only allows for portability and broader support, but also limits the risk when the new services are introduced or changed. As the industry experienced with DOCSIS and PacketCable, open standards and abstraction layers work. Hardware and software vendors in the online ecosystem do not have to regression test every iteration of hardware and software on their platforms because applications are developed using open or published (sometimes licensed) APIs to the operating system. Similarly, web browser developers are not required to test their browsers against all possible websites because the browser is acting as a presentation layer for a standardized format of data – not unlike EBIF.

Sandboxes

In addition to supporting open standards, the cable industry's connected ecosystem can benefit from faster methods of integrating new services. One of the approaches being used to shorten the development cycle is to allow developers access to a live, MSO Grade, test cable plant that is representative of a real cable network. Providing application developers access to this test network, or "sandbox" during their development process gives them the ability to learn, and identify issues that cannot be identified in a simulation environment. To encourage innovation from a wide cross-section of developers, access to a sandbox may be the only way a developer may be able to afford to test their application. To accomplish this, organizations such as CableLabs, and centralized services providers such as HITS AxIS can provide "sandboxes" to application developers to develop and test new applications on a representative cable plant. This approach can not only reduce the development time, but also accelerate the test cycle, enabling applications developers and MSOs to deploy new services much faster.

Providing these sandboxes also enables the cable industry to incubate new application developers and encourages innovation on cable networks. By reducing the cost of building an environment to innovate new applications, sandboxes also help to foster the next generation of application developers. Organizations like OEDN, which encourage college students to begin developing applications for the cable industry as they do today for iPhone and the Internet, also provide the infrastructure for students to integrate their applications. Enabling these new application developers helps to grow the ecosystem that will provide MSOs with more choices of applications to offer their customers.

On-boarding

Organizations that provide an operational environment for application developers should be prepared to offer developers an easy, accessible on-boarding process that allows for rapid introduction of new applications into the market. Mini-mobile applications and the iPhone "App Store" illustrate that the connected world is more about small, inexpensive and often temporary applications rather than complex, expensive ones. Accordingly, good on-boarding organizations help to breed innovation and competition by encouraging a wide universe of application developers to develop, test and deploy an even wider range of applications to run on cable networks.

A good on-boarding process also helps in the building of a knowledge base of best practices, known infrastructure problems and typical pitfalls that can be shared amongst all developers. In addition, it offers application developers the privacy and protection required for keeping their intellectual property safe.

A CONNECTED ECOSYSTEM FOR ADVERTISERS

By connecting MSOs and back offices, local cable systems will be able to present advertisers and content providers with the opportunity to communicate to their customers with the most powerful medium (high definition, large screen, surround sound television) while using the web-proven tactics of personalization and interactivity. However, some additional connections will be required in order for the advertising and content community to take full advantage of these advanced opportunities.

In addition to the connections discussed earlier, the capabilities of a more robust platform for seamlessly delivering and managing advertising and other forms of content will help to increase effectiveness. Moreover, these enhanced capabilities would allow content placement opportunities on the cable video platform to be more competitive with other content platforms, largely by leveraging their capabilities.

The connected ecosystem described in this paper provides advertisers with a platform that combines television's powerful ability to engage audiences with the interactive capabilities and reach offered by the Internet. Today, advertisers use on-air campaigns to drive traffic to web sites where they can provide more information and transactional opportunities. The connected cable ecosystem will allow these interactive elements to occur over the video platform itself.

Personalizing Content Through Changeable Elements

The connected ecosystem can support the advertising and content communities' interests for optimizing the value of their content by expanding the methods for reaching potential customers who are the most interested in their ads or content. Currently, one of the advantages of placing ads by node is the ability to sell that inventory to neighborhood advertisers. The connected cable ecosystem would expand the market for these hyper-local ad avails by providing the capability for national advertisers to reach specific neighborhoods or other demographics anywhere in the country.

Supporting the growing demand for segmented advertising requires an expanded approach towards managing, tracking, and delivering content. It is relatively simple to manage one of a few pieces of content that are distributed nationwide, or to a few regions of the country. It becomes exponentially harder to manage content that may have multiple versions, with different interactive applications attached to them, with different text overlays, linking to different actions, and all designed to reach different communities.

Combining demographic data concerning a neighborhood or customer with the cable system's ability to deliver content by node, opens up the possibility for personalized content. However, delivering the message that will be the most relevant to that neighborhood can involve managing as many as 20 different elements that customize the message for maximum impact. An example would be a cruise ship ad that features older adults when it appears in neighborhoods with that demographic while replacing those scenes with images of family activities in neighborhoods that skew to that demographic. Stitching those elements together requires pairing the ad server with an ad-decision engine that will ensure the ad containing the right elements appears in front of the right viewer.

Delivering this type of content to the hosting and serving devices will require the ability to automate management of the bidirectional metadata. This is to ensure that the multiple elements of the ad delivered to the serving environment are assembled the correct way for that particular node. The ability to take advantage of advanced capabilities such as these is result of both the technology platforms being available, and accessible at large-enough scale for the content to be available to a large audience. The connected cable ecosystem plays the important role of being able to mange the delivery of the content and the metadata across a national footprint, but to also aggregate engagement data across all markets and by each demo, while observing strict privacy guidelines.

Interactive advertising

In order to fully support interactive advertising – all of the content and its associated elements should be managed and tracked across multiple delivery platforms. As an illustrative example, assume an interactive ad may have links to telescope to a long-form video on VOD, and also to support a prospective consumer's ability to receive more information about a particular product or service. To achieve the desired result, the linear ad needs to be delivered to the Ad insertion system, the long-form content to the VOD server, and the information requests content and instructions to the RFI system. All of this content and the associated metadata that describes it need to be delivered to all of the devices, and verified to ensure the consumer experiences the desired result. Ads delivered via tape are physically incapable of containing the data required for interactive capabilities. The open standards architecture described earlier in this paper is intended to support interactive advertising applications by providing the interface to remote carousels or ad servers and the asset and scheduling metadata to insert an asset into a digital video stream. Advertisers interested in taking advantage of the connected ecosystem's iTV capabilities will also require the interactive application templates that were previously described.

Accelerating time to market

Another requirement of the advertising community for delivering personally relevant content that can be addressed by a connected ecosystem is accelerated delivery. Rapid, timely delivery of all of the content and its associated elements across the entire footprint is an important element of the video system operating in a comparable manner to the other (i.e. Web, mobile, etc.,) platforms. Examples of ads that are heavily dependent upon time-to-market for their effectiveness include political campaign ads where a candidate will want to create and place an ad to respond to a new attack ad from his or her opponent. Another example would be advertisements that encourage greater audience engagement by mixing images from the program with ad content, such as a carmaker's ad that features "the play of the game."

To achieve rapid delivery requirements, the ability to accelerate the collaboration and approval processes can also be improved. Advertising content is typically created in a collaborative process between one or more production houses. Ad agencies approve this content before it's distributed to the end destinations. As the number of versions of the content, and associated elements grow, the processes that track and support the approval will need to take on the same level of sophistication as the delivery processes.

Meeting the demand for accelerating time-to-market is well beyond the capabilities of tape dubbing and overnight delivery systems. These ads will need to travel at the speed of light, not FedEx. A network-based architecture appears to be the most appropriate means to support these content delivery needs. Understanding that the participants in this ecosystem will be comprised of both large and small companies located around the globe suggests that the content management and delivery system will need to overcome such issues as packet loss, corrupted files, longer transport session times, lost productivity, manual monitoring and the lack of real-time, item-detail reporting of file transfer activity. These asset delivery systems will need to provide fast, fault-tolerant, secure, and audited digital delivery that will represent a core element in the supply chain for addressable advertising. Delivering in real time leaves no time for mistakes.

Meeting the Advertiser's Quality Requirements

Ensuring quality throughout the asset delivery process is also increasingly important to the advertising community. Their ads are built and tailored for customers who will be viewing them on large, high-resolution display devices. As such quality of experience becomes an increased focal point. The ecosystem will also need to provide content providers and others stakeholders in the distribution chain with the ability to monitor and measure quality throughout the content distribution process. Incorporating these design, operational and measurement elements into the asset delivery system will improve the likelihood of adoption and probability of success for advertisers and other enterprise market stakeholders that use connected cable architecture to meet the needs of their customers and grow their businesses.

As with the connections described in previous sections of this paper, meeting the interest of the advertising and content communities will also involve extending beyond the connected cable video network. In order to provide access to the features and flexibility found on the web, connected cable networks will also need to connect to the systems and services that enable these web capabilities.

One example of the systems cable may want to connect with are advertising decision engines. These web based systems that decide which ad to provide to a particular customer, at a particular time are becoming extremely advanced and valuable. Today, these systems primarily support banner ads but they are expanding into video ads as well. These types of decision engines will be advantageous for expanding the capabilities that cable's new platform can offer to advertisers and content providers. In addition, leveraging the mass and scope of the data collected and analyzed by these web-based systems will enhance the value of the cable video platform to companies that require this information in order to optimize the effectiveness of their content delivery activities.

SUMMARY

With all of the recent advances in technology, the cable industry is quickly moving past a wired world into a connected world. In fact, consumers aren't just connected occasionally, they are always connected. They can get e-mail, connect to the internet and text or twitter anytime of day or night, regardless of location.

For cable to lead the development of advanced services from advertising to game playing to social networking, it will want to embrace the concepts of connectivity. Historically, cable systems were being limited to a single medium that provided "last mile" connectivity for video consumers. However, a connected view of the world can't end there. Limiting the view to ways to connect to the consumer would squander the opportunity for being a major player in the products and services that consumers access through those pipes.

Consumers of today and tomorrow have an expectation that a single device or application is connected to the rest of their world. Evidence that this is necessary becomes obvious in situations like looking around the plane when the flight lands. Almost everyone immediately reaches for his or her cell phone. In many cases the user doesn't immediately make a voice call, but rather spends the taxi time reconnecting to their world, checking e-mail, answering texts or checking news and scores. This is the environment in which cable will play a greater role by continuing to build upon its commitment to an open, connected ecosystem.

The decisions that all cable MSOs make over the next few years will help to realize the potential for cable's video platform to serve an increasingly larger and more dynamic environment . Realizing the fullness of that potential will require the connections that allow consumers to access the personally relevant services they want more easily.

By connecting to other wire providers and cable systems, the industry could ensure that its services can deliver the mass and scale required for success in today's world. This entails connecting technologies in ways that allow for localization and personalization without requiring custom development and extended deployment cycles. It also involves connecting advertising platforms in ways that allow advertisers to enjoy the power and presence of the high definition, large screen, surround sound television with the interactive, dynamic and personalized capabilities that have evolved on (the far less impactful) Internet platforms.

The time is right for evolving to meet the connected society of the future. Connections to both its existing base of friends and partners, as well as connections to other, somewhat competitive, electronic industries will be critical for success in a connected world. The cable industry is in a unique position to provide both consumer and back office services connections in ways that can take advantage of a world where often, the mass is the message.

---

[i] ***How did eBay start?*, A brief history of eBay, By Aron Hsiao, about.com**

Content Management Systems versus Content Delivery Networks

Michael Adams
TANDBERG Television

Abstract

In today's on-demand deployments, content management and content delivery are intertwined into a store-and-forward, multicast delivery system that relies on satellite transmission for secure asset delivery. The mode of delivery is a "push" from the content provider to the on-demand site, such that all assets must be provisioned in the video-server memory before they are offered to the customer. This same system ensures that titles are offered in a consistent and flexible way according to the operator's marketing requirements.

In contrast, cable's high-speed data services use content delivery networks (CDN) to deliver web assets, as a series of file-transfers over terrestrial networks, such as leased interconnects and backbone networks. Moreover, the mode of delivery is a "pull" from the subscriber, based on HTML requests from the website. Websites are essentially unmanaged by the operator, except for caching frequently served pages to manage backbone bandwidth.

There is considerable interest among cable operators in migration towards using a CDN approach for on-demand content. However the transition is not trivial because of cable's unique requirements related to scale, management, and security for on-demand assets.

This paper explains in detail the requirements for a Content Management System (CMS) to manage end-to-end delivery of on-demand assets in a hybrid network environment using satellite and terrestrial links. We show how a migration towards terrestrial content distribution is possible without giving up some of the fundamental advantages of today's "push" delivery model, namely that very popular titles do not generate unmanageable spikes in network utilization and that customers are never offered "hot" assets until they have been successfully provisioned at the video-server.

Nevertheless, there are excellent reasons to migrate to terrestrial networks for content distribution including the ability to turn-around assets much faster when required; an example would be a political campaign advertisement. In addition, the explosion in content choices, coupled with the increase in the number of high definition titles, makes the "push" content delivery model impractical at some point. Therefore, mechanisms to identify certain, less popular titles as "library content" become essential, as does selectively employing a "pull" content delivery model.

Finally, this paper will contrast the requirements for a CMS with those for a CDN, and show how the two technologies can be used together in certain circumstances to blend the advantages of both types of approach.

INTRODUCTION

Video-on-demand systems have now been deployed by all major cable operators across their entire footprint and these systems work extremely well, serving millions of customers on a daily basis.

Cable on-demand services are based on a QoS guaranteed, connection-oriented model from the streaming service to the subscriber. This model relies upon session resource management which allocates a guaranteed slice of bandwidth between the server and the set-top for the duration of an on-demand session. This is possible because the path from the streaming server to the set-top is over a relatively simple access network topology (typically via a Gigabit network to a QAM modulator). The only possibility of blocking is in the HFC network itself, but this well-managed by allocating constant bit-rate sessions using standardized QAM resource management. Cable operators have become adept at increasing the total bandwidth per subscriber by reducing the size of the service group as on-demand peak usage increases.

However, this means that video servers have to be placed at the edge of the network and they need to be pre-provisioned with all the assets that the customer could possibly want. This leads to some challenges to solve in deployed systems:

1) Scaling the number of assets from 10,000 to over 100,000 to support long-tail content.
2) Effectively managing a very large number of assets across multiple sites.
3) Ability to dynamically change metadata independently of content propagation.

In this paper we will examine the role of a Content Management System (CMS) and a Content Delivery Network (CDN) in solving these problems and show how they address different aspects of these challenges.

ON-DEMAND CONTENT DISTRIBUTION

Content distribution products were first developed to meet the specific needs created by cable operators as they deployed on-demand systems on a headend-by-headend basis.

In the earliest days, content files were shipped around on tape and metadata was manually entered into the VOD system at each headend. In order to scale on-demand services, the concept of a package was developed – this is essentially a collection of assets and metadata files that completely describe the on-demand title. Each package is transformed into a single file using a UNIX utility called tar. These tar files are transmitted over satellite, using multicast to greatly reduce bandwidth requirements. Thousands of hours of content are distributed to systems via satellite from multiple content providers. Although an hour of standard definition content requires about 1.7 gigabytes to be transferred, the actual transmission rate from each content provider is relatively low, in the order of 10 Mbps.

Satellite distribution provides an extremely efficient, "push" mechanism to get the required content to a very broadly scattered set of cable systems. This so-called "pitcher-catcher" approach uses a reliable multicast algorithm to make efficient use of satellite capacity to push assets to a large population of headend systems simultaneously. Also included is robust encryption to prevent unauthorized access to the content files.

Unfortunately bundling the content files with the metadata means that the metadata is also multicast, leading to a one-size-fits-all approach. This forces the operator to make metadata changes at each system, after content distribution rather than centrally. As a result, asset management has become complicated and time consuming because each system is managed separately. Worse still, there is a lack of transparency to corporate marketing and central operations groups. If delivery of title to a site is unsuccessful, a manual "re-pitch" is required, which is both costly and inefficient.

The title metadata includes fields on pricing, categorization, and availability window. A content provider distributes packages with these values set according to its business rules. The operator may need to override some of them such as different pricing on a system-wide or system-by-system basis. Some assets may need to be filtered because they are not required at the particular system, or because there is insufficient storage capacity for them. Because of the need to change the metadata for each operator, and even down to the specific site, “Asset Management Systems” were introduced. Initially asset management systems were deployed on a site-by-site basis but they have evolved to support multiple sites in a regional environment.

CONTENT MANAGEMENT SYSTEMS

On-demand systems rely on thousands of hours of content distributed via satellite from multiple content providers. With so many files in so many locations, it becomes a complex task to keep track of everything and today’s manual systems are struggling to keep up. There is often a significant cost in on-demand operations due to so-called “content propagation errors”. These occur when part of the content distribution chain fails during the provisioning process. In most cases, the harm is done because there is no mechanism for the operator to automatically check that all the necessary assets have been placed on all the appropriate servers. So the provisioning error is often discovered first by the customer, who requests a title and gets back an error code. This impacts the session success rate. Worse, it discourages the customer from using the on-demand service and reduces operator revenues.

In practice it is very difficult to make the content distribution chain completely “bullet-proof” and doing so would come at a significant cost, so it is better to automatically check the asset status on a regular basis. If a particular asset is missing from a server, it can be automatically rescheduled for delivery. Since assets are typically propagated before they are made available (before the viewing window), it is possible to repair any missing assets before they have any customer impacting effect.

The situation is similar for on-demand dynamic ad placement – in this case it is especially important to verify that the ad content is actually on the server before it is scheduled to be inserted into an on-demand stream. The ad decision is usually made at session start-up time, and at this point a play list is presented to the server.

Since operators offer the majority of on-demand titles across their entire footprint, it makes sense to manage them centrally. Since content distribution and delivery is spread geographically across a large number of systems, a distributed content management solution is required with centralized control. Since local divisions and regions also want to offer local on-demand titles, the solution must support local management of those titles.

A central control panel to manage content metadata provides a long list of advantages for the operator, including the potential to enable powerful and timely promotional campaigns along with pricing discounts, to provide targeted advertising support, to adjust the viewing window, to remedy errors, or to enrich metadata after the title has been provisioned (for example to support extended descriptions, advanced search, or recommendations).

Let’s explore some example scenarios that are problematic for operators today and I’ll explain how a content management solution can help to make these scenarios much easier to manage effectively. I’m going to focus on scenarios that are operationally intensive or impossible to implement today:

The corporate marketing and programming department believes that a title has truly underperformed and decides that a promotional campaign along with a price reduction is justified in the final week of availability. With centralized control of the metadata and the resulting ability to easily lower the price for the title, the marketing and programming department now has the ability to pull together a cross-channel advertising campaign in partnership with the content provider. Further, they can use their content processing solution to embed ETV triggers in that advertising, to direct viewers directly to the "buy" screen.

A set of titles with a common element or theme can be listed in a special category – for example, all the movies featuring "Ben Kingsley" or all the "Bond" movies. By cross-linking the metadata of titles in this way, this also creates a foundation for automatically generated recommendations for new titles.

An unpopular title can be removed from the system before the end of its availability window to free up storage space for new titles. As new titles are introduced, a set of rules is used to distribute them to each market according to pre-defined priorities so that the on-demand storage in each market is optimized for maximum revenue generation.

A set of titles can be ingested in the traditional on-demand infrastructure, and then automatically processed according to a set of rights and rules housed in a seamlessly connected solution to the appropriate content compression and metadata formats for broadband and mobile video service platforms. Depending on the rights and rules, in some cases, only the trailers are sent to these additional platforms, and in other cases, the movies can be purchased and viewed on them.

In order to do all of the above effectively, a fundamental change is required in the way that the content management solution structures and handles the relationship between a so-called "heavy" asset and its associated metadata. In short, the processing, managing and delivering of metadata needs to be separate from, but not disconnected from, the associated assets, and vice versa. The content management solution needs to maintain more intelligence, awareness and flexibility in this critical linkage between assets and metadata.

For example, for certain distribution needs, the delivery of metadata might be unbundled from the delivery of the associated heavy asset. This is the case for many operators that do not have the backbone capacity to handle distribution of their content asset. By separating the delivery of these two assets, the operator can deliver very large content assets to local systems over a very cost effective satellite network, and yet still deliver metadata via their intranet. Thus metadata and rules for the management of the metadata can be organized centrally and delivered to the local systems whenever content is distributed or updated.

Finally, operators will need to be able to provision content so that it can be played on any device. This provides the operator with a competitive response to "over-the-top" video providers by providing a seamless extension of on-demand cable services to additional devices within the home; for example, to TVs with broadband connection, to PCs, or to an ultra low cost IP set-top box. To make multi-platform services manageable and scalable, the content management solution must enable the operator to provision a single title with multiple playout options across multiple devices.

CONTENT DELIVERY NETWORKS

Content Delivery Networks have grown up to support web-based delivery of information, including arbitrary content formats for graphics, music and video. Wikipedia defines a content delivery network as "a system of computers and storage networked together across the Internet that co-operate transparently to deliver content most often for the purpose of improving performance, scalability and cost efficiency to end users".

Essentially a CDN is a cache-based system where files are moved around the network based on the pattern of usage. Considerable ingenuity has been applied to the algorithms and protocols to do this. Peer-to-peer (P2P) protocols are one example of an approach to building a distributed, collaborative network of content stores.

A typical CDN implementation is to populate files into an origin server (which itself may be a distributed entity). This server is connected to many cache servers by network links. The cache servers directly serve the clients, which are web-browsers. When the user navigates to a web-page, the embedded markup language (HTML) causes the browser to request all the necessary files for that page using HTTP requests. In the first fetch, the cache server must retrieve them from the origin server and pass them along to the client, keeping a copy in the cache. On subsequent fetches, the local cached copy can be used to save bandwidth on the network link back to the origin server and to save processor bandwidth on the origin server.

The CDN architecture allows for multiple tiers of cache servers for better scalability and allows the edge cache server to be pushed closer to the client, allowing for faster response times and higher bandwidth across a given link. However, because each client request is independent of every other, each fetch is a unicast session and considerable effort has been applied to cache management algorithms, load-balancing, and so on to make best use of each fetch operation.

Although this is a simplified overview of how a CDN works, it captures the important principles of operation. It is fair to state that CDN technology has been developed for the general case of managing a large number of small files with little regard to what function those files actually represent at the system level. As such, the model lacks a higher level abstraction such as the asset grouping and its metadata that ties certain files together. The reason for this is that the hyper-linked web pages themselves provide the framework and the navigation for the website. Essentially the "metadata" equivalent is embedded in the HTML markup of the web pages. Web technology improvements including dynamic HTML, applets, and servlets make web-sites much more dynamic and flexible than before and allow support for localization and personalization, but this principle remains unchanged.

Although the CDN automatically responds to client requests to make ensure that the most recently used files are held in the cache servers to satisfy overall demand, this may not provide the best indication of asset popularity. This is because many cached files are related to navigation requests versus asset requests. Moreover, there is no intelligence in the CDN to pre-position files in anticipation of a demand for them which may cause unpredictable spikes in bandwidth as new, popular assets are discovered by the clients. [1]

CMS VS CDN

Table 1 summarizes the functions supported by CMS and CDN technologies.

Function	**CMS**	**CDN**
Metadata management	Y	N
Transcoding control	Y	N
Workflow	Y	N
Content Distribution	N	Y
Multicast Distribution	N	N
Cache Management	N	Y
Load Balancing	N	Y

Therefore, we conclude that CMS and CDN functions are complementary to each other; a CMS allows for functions related to the semantics of the content while a CDN provides the mechanism to distribute the various files across the network of servers.

The next step in this analysis is to discuss how the two technologies can best be combined in a practical way.

CMS AND CDN

In this section we will discuss a potential migration strategy to add CMS and CDN technologies to an existing on-demand deployment. As we have already discussed, CMS can provide a powerful management layer that will work with the existing satellite content delivery model to provide operational advantages and additional control and monitoring capabilities.

Therefore, the first logical step is to add CMS to the existing on-demand deployment to reap the benefits of this technology.

As demand for long-tail content increases, the system may reach a limit on the number of assets that can be "pushed" to the on-demand systems in the available time. Ultimately this is limited by the cost of satellite bandwidth available to content providers and the cost of storage in each on-demand system. Titles towards the end of the long-tail distribution will be so rarely viewed that it does not make economic sense to distribute them and store many copies of them locally.

At this point, CDN technology provides a complementary addition to the existing "push" content distribution model. It should be noted that the latest generation of "pitcher-catcher" based solutions will operate across a mix of satellite and terrestrial facilities. Nevertheless, for popular titles that are expected to generate significantly more than one play on average per system per month, the "push" model is still superior from a cost perspective because of its use of a multicast distribution protocol.

Therefore the best solution is to use the CMS to implement distribution rules based on popularity indicators embedded in the metadata in order to select which delivery mechanism to use on an asset by asset basis.

CONCLUSIONS

In this paper we have reviewed the capabilities of CMS and CDN technologies in on-demand video deployments and conclude that:

1) They are complementary technologies and there are advantages to using the two in conjunction when the number of assets is greater than 10,000.
2) CMS technology is a useful addition to existing on-demand deployments in conjunction with todays "push" content distribution systems.
3) CDN technology may be added as an additional distribution technology for long-tail assets, essentially "pulling" these assets from a central content library only when they are required on the first customer play.
4) Extensions to support multi-platform delivery of assets will require a CMS platform to manage the additional complexities of content transcoding and multiple content files formats per asset.

---

[1] ***THE BENEFITS AND CHALLENGES OF DEPLOYING LARGE REGIONAL VOD ASSET LIBRARIES by Michael W. Pasquinilli, Sunil Nakrani, Jaya Devabhaktuni, 2008 NCTA Technical Papers™***

DEVELOPING A GRADING SYSTEM FOR DIGITAL VIDEO QUALITY

Dave Higgins
Chuck Wester
Comcast Media Center

Abstract

This paper describes an "End to End" measurement and analysis philosophy for transport of digital video signals from a centralized distribution point to the consumer. A key component of the effort has been to develop specific and repeatable metrics that describe both the integrity of the signal delivery, as well as its relative subjective quality.

Given that there are a number of approaches and technologies used to deliver linear video content to households, this paper examines the use of multiple transport options commonly deployed in the marketplace today and how metrics can be created to measure the efficacy of the end product as delivered to the consumer. Specifically, the delivery plant considered is based on MPEG 2 transport to the consumer, however the model can easily be conformed to support an MPEG 4 approach.

Central to the measurement efforts employed is the adoption of common terminology to describe known digital video impairments. Further, the ability to describe these impairments in a manner that can be easily understood, communicated, and taught is a key outcome of the effort.

EXECUTIVE SUMMARY

As multichannel video providers achieve parity in the quantity and subscriptions fees for popular television channels and programming, the quality of service (QoS) associated with the delivery platform has become a key differentiator that can significantly impact the successful acquisition of new customers as well as dramatically influence whether a customer is retained. In fact, HDTV can be seen as a clearly defined competitive battleground based almost entirely on the proposition that the recovered video "Quality" must exceed a consumer's current service. According to recent industry studies, quality issues are responsible for 40% or more of all customer churn, second only to cost.[i] In addition, industry analysts have observed that cable system operators risk losing the entire RGU (revenue generating unit) when bundled customers move to an alternative multichannel video programming provider (MVPD).[ii]

With so much at stake, cable system operators must actively participate in the video content delivery chain which includes the creation, management and distribution of HD content. This paper identifies as many as eight critical touch-points in the content distribution process that can directly impact video quality performance and affect the customers' perception of quality and satisfaction with their video services provider. In addition, it is based on the end-user HD experience to provide a more accurate depiction of the impact on quality perception from multiple factors including advanced compression techniques, the nature of source content acquisition, and satellite vs. fiber transport delivery.

While most cable operators focus their quality monitoring efforts from the output of their headend to the consumer set top, achieving optimal and competitive quality assurance requires establishing and applying a consistent system for measuring the quality of digital video and audio from the source to the display device. To achieve this objective, the use of a repeatable grading system that summarizes both the characteristics typically noted by expert or "Golden Eye" viewers as well as impairments the average consumer objects to is key. The challenges associated with creating such a program include the lack of industry accepted standards and practices or inconsistent application of complex technical concepts in an operational environment.

Implementing an end-to-end system for quality assessment requires tools such as probes for measuring the quality and reliability of HD content at each touch-point. In addition to monitoring the probes and applying the grading system to anomalies that are detected and measured, data will need to be exchanged between television programming networks, cable system operators, and others involved in the origination and distribution of HD content in order to address and correct reported shortcomings.

This paper also includes a summary of a grading system that is being used today and examples of the ways that the system is helping to minimize impairments and improve the HD customer's experience. It will conclude with recommendations for steps that the industry can take in order to facilitate the implementation of a quality grading system, such as greater automation of the quality assessment process, additional independent research, and consumer education.

BENEFITS

A recent MRG research report found that 90 percent of cable operators consider video quality monitoring was either "crucial" or "very important", and 58 percent viewed end-user quality of experience (QoE) as critical and needing to be maintained.[iii] Cable operators indicated that service quality issues are one of the main reasons for customer support calls, resulting in a significant reason for customer churn – as much as 40%. With millions of customers across the country, end-user video quality monitoring is an integral part of a cable operator's business," according to the report.

Extrapolating from the report's findings, the level of churn attributed to QoS represents millions of digital video customers, hundreds of millions in annual revenue, and billions in asset value. Given the importance associated with customer retention, several system operators are launching significant efforts to focus on QoS. MRG's analysis found that some cable MSOs are spending as much as $2 – 5 per subscriber per year for QoS.

"With competitive pressures increasing, cable operators need a comprehensive video monitoring solution to ensure they meet customer expectations, or face possible increases in churn and operational costs," the study concluded.

Developing an accepted and customer-driven quality grading system will enable cable system operators to address misperceptions that hinder the use of advanced compression techniques, such as HD3:1 using MPEG-2. If left to applying only the simple math behind 3:1, or the greater compression offered by MPEG-4, potential HD customers would assume that the quality of HD signals cannot be as good as an HD signal using compression levels of 2:1 or lower. However, consumer research conducted by the CMC (Comcast Media Center) and others has demonstrated that by using best case practices for digital video encoding, stat-muxing, and transport, HD3:1 is highly competitive with MPEG-4 via DBS or HD2:1.

This research has helped to demonstrate that HD3:1 can allow cable operators to launch more HD content using less bandwidth and without sacrificing the customer QoE. It also underscores that there is a tremendous opportunity - and need - for consumer education concerning all of the factors that affect their viewing experience.

INTRODUCTION

Existing industry standards describe video quality comparison of the video measured at any point along the delivery path to a source reference; a referenced video quality measurement which determines how much a system or process has degraded a given video service. However, they do not address a significant issue which is the true quality of the video as measured at all critical touch points in the delivery path as well as at the consumer display. Regardless of the Herculean efforts to improve the video source and delivery systems, the quality of the video delivered to the cable subscriber can be no better than the worse case video quality at any touch point.

Source providers, equipment manufacturers, and MSOs have made tremendous improvements to their video transport and delivery systems. The challenge is to find further significant video quality improvements to the end video product as delivered to the cable consumer. Given that technical advances have allowed for greater density of channels in a multiplex while achieving approximate parity with earlier efforts has resulted in a tangible increase in plant capacity. The balance associated with “Quality vs Quantity” continues to be a complex and challenging equation to solve. That said, the improvements associated with upgrading one area of the Content Distribution Chain can be marginalized by other impact points and in some cases made worse. Cumulative degradation, which occurs as a video service is created or acquired, processed or re-encoded, delivered through various networks, groomers, commercial ad insertion equipment, and finally to the customer’s set tops, is an emerging area of opportunity to improve.

Prior to digital video, the quality of video delivery systems was measured using vertical interval test signals (VITS), Vertical Interval Reference Signals (VIRS), and performance metrics outlined in documents such as ANSI/EIA/TIA-250-C. Operations engineers could use these test signals at numerous points in the distribution chain without impact to the service to perform measurement of the quality of the delivery of analog video. Today, transmission operations engineers are more likely to use MPEG analyzers to determine the quality of the delivery path. Typically these analyzers are used to measure known impairments as part of ETSI TR 101-290 compliance such as continuity counter errors, missing PIDs and Jitter performance. While this approach is effective for measuring transport stream errors such as packet loss, MPEG analyzers do not provide any ability to measure customer perceived video quality particularly as it relates to subjective components (i.e. sharpness, noise, macro-blocking, etc.)

Further, the consumer marketplace has overwhelmingly adopted larger display screens that provide enhanced noise reduction, filtering, improved contrast ratios, and materially greater resolution which result in the ability to see enhanced details in all video content. Unfortunately this ability also facilitates more critical viewing of digital impairments or “artifacts” inherent to digital video compression and transmission. The obvious result is that these impairments whether they are associated with HDTV or SDTV, are becoming much more evident to customers and there is considerable mis-information present in the marketplace. One example is channels branded as "HD" which show a majority of their content as standard definition 480i video "upconverted" to HD. Another example is larger consumer displays which are often broadly lumped together and referred to as "HD Displays" while the viewing characteristics between the different types can be dramatically different, particularly under varying lighting and "off-axis" viewing conditions.

The confusion in the consumer marketplace is understandable, given the complexity of the topic in general, as well as the technical appreciation required to acknowledge all of the possible reasons behind a given video quality impairment. Differences in the video quality displayed by various brands and types of monitors and increasingly rapid video monitor response rates add to the confusion. While video engineers evaluate video quality using defined parameters such as viewing angles and distance from the monitors, there are differences in ambient conditions of monitor locations consumers likely do not appreciate.

VIDEO QUALITY TOUCH-POINTS

The examples of video quality measurement provided in this paper are based upon analysis that is being conducted by the Quality Assurance team for the CMC, which is currently placing nearly 2Gbps of MPEG2 video onto the an IP network for delivery to cable markets. Drawing 1 is a simplified view showing a typical service path of an HD signal broadcast out to cable subscribers.

As an example, assume that there is a sporting event being produced at the *Venue* which needs to be transported back to the studio where an on-air announcer is providing commentary. The transport path, whether satellite or fiber, is very likely being compressed, particularly if the feed is native HD. This initial compression causes digital impairments that will never completely be removed from the "video" and as you will see, can become materially worse as they travel thru the distribution chain to the consumer.

The studio camera is likely the very best source of video quality we will consider here as the conditions are controlled, the lighting is optimal, and there is very little opportunity to induce impairments beyond those introduced by the camera optics. Typically, material is mixed between the *Venue*, the camera, and server playback of pre-recorded media in *Master Control.* Given the cost associated with media storage, particularly HD native content, *Video Servers* will ingest content utilizing compression and induce another set of impairments. Again, the deeper the compression the more likely video impairments will result.

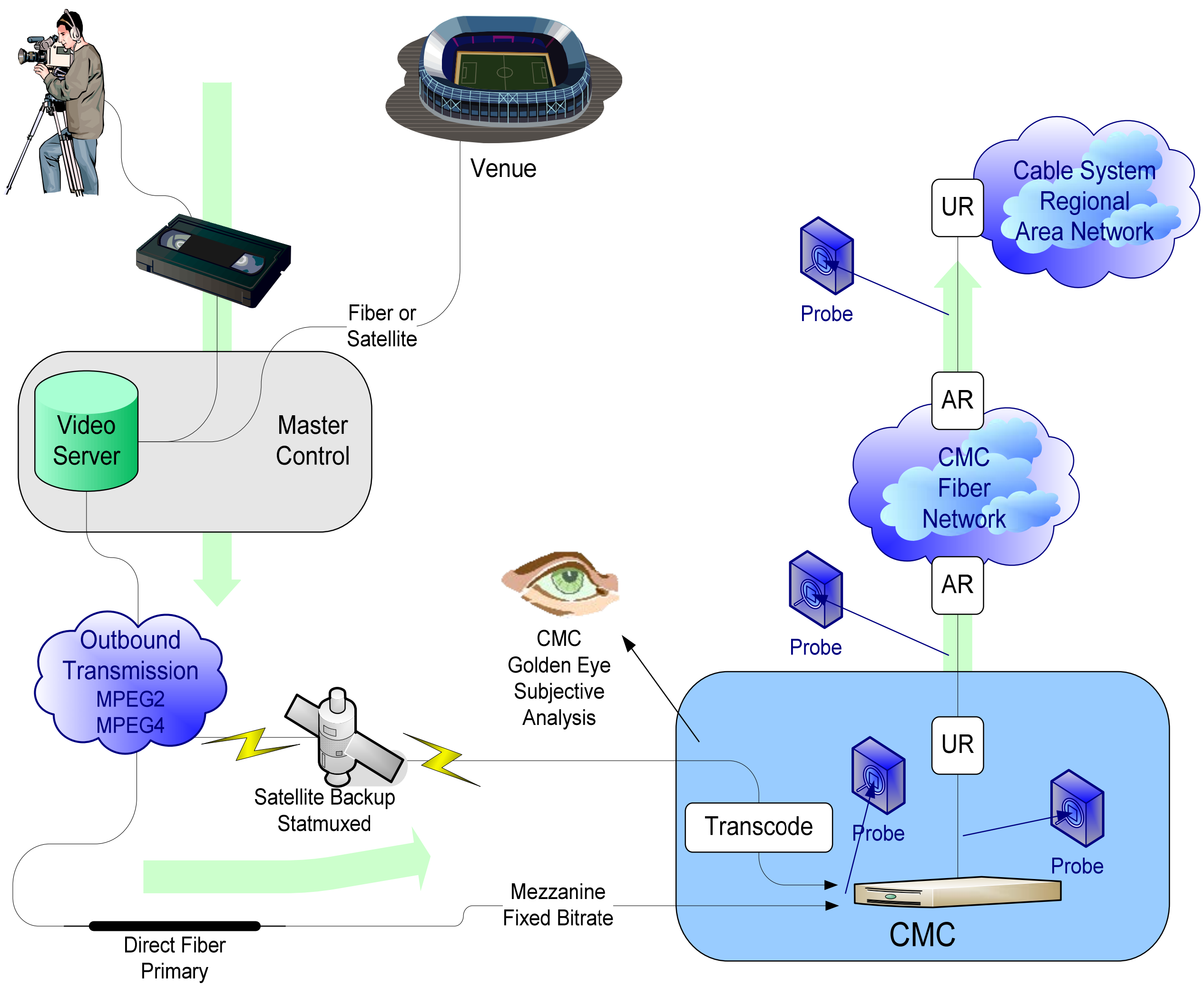
T he next step in the process is associated with transporting the video signal from Master Control to the *Outbound Transmission* system. In many cases this is where a significant impact can occur in the overall delivered video quality product. All major programming providers utilize compression, whether it is MPEG 2 or MPEG 4 based, to maximize the use of satellite or terrestrial fiber bandwidth. Traditional approaches for MPEG 2 statistically multiplexed HD bandwidth usage here are in the range of 12 to 15 Mbps per channel, but there are exceptions in both directions. Whatever the approach, the outcome can be seen at the baseband video output of the satellite receiver or receiver/transcoder and in all cases there are video impairments.

The CMC's Quality Assurance Team is also monitoring content that is distributed over *Direct Fiber*. The connections, which will serve all of the major programming providers, will operate at a higher bandwidth usage model to minimize the impairments associated with the *Outbound Transmission*.

Once the signal arrives at the CMC, it is either converted to MPEG 2 if required or presented to the *Imagine Encoder* as a Gigabit interface. The efficiency associated with removing the need to take the receiver output down to its baseband video components and then re-encoding has allowed the opportunity to integrate 3 HD signals into a single multiplex while sustaining competitive video quality.

As Drawing 1 shows, the output of the *Imagine Encoding* system is then transmitted onto the Fiber Backbone for distribution.

While not specifically addressed in this paper, downstream impairments can obviously be caused by local re-encoding, rate-shaping or grooming as well as local Ad Insertion systems. Further, the set-top box, the consumer display device, and the viewing environment all play a contributory role in recovered video quality.



Drawing 1

The working assumption is that if the video source is of superior quality, the re-packaging encoders are performing optimally, and the network is functioning as it should, then the video from the source should be delivered to the consumer with little degradation and the availability of the service should not be an issue. However, it is critical that as a service provider, the CMC tracks not only the video quality but the performance of the delivery platforms in order to determine the quality of each service, as best as possible.

To measure the uptime performance of its HD video delivery over the fiber network, CMC has installed a system of MPEG transport stream probes placed strategically across its network to monitor the availability of each of the video services transported. It is important to note that only service availability (i.e. Packet Loss as an example) is monitored through automated means. CMC has been using the MPEG probe system since January, 2008 to measure its delivery performance, optimize the network, and to accumulate fault data for statistical and alerting purposes. Table 1 shows an example of several weeks of performance data collected on a per program basis facilitating analysis and review; note the anomaly that occurred at Site 11, Red, yellow and green color coding indicates the severity of the impact.

Programs by Imparied Seconds	Source (in)	CMC (out)	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12	Site 13	Site 14
Channel A	2	16	11	10	11	11	11	14	11	18	11	11	168	11	11	11
Channel B	1	1	0	0	0	0	0	0	0	6	0	0	160	0	0	0
Channel C	9	2	1	1	1	1	1	1	1	5	1	1	159	1	1	1
Channel D	0	12	11	10	11	11	11	11	11	15	8	11	168	11	7	10
Channel E	1	7	2	2	2	2	2	2	2	6	2	2	160	2	2	2
Channel F	9	19	9	9	9	9	10	9	9	13	9	9	167	9	8	9

Programs by Imparied Seconds	Site 15	Site 16	Site 17	Site 18	Site 19	Site 20	Site 21	Site 22	Site 23	Site 24	Site 25	Site 26	Site 27	Site 28	Site 29
Channel A	12	11	20	11	11	10	14	11	11	14	16	11	11	11	11
Channel B	0	0	4	3	2	2	2	1	1	2	7	1	0	2	0
Channel C	1	1	5	2	1	1	2	1	1	2	6	1	1	1	1
Channel D	10	10	15	8	10	10	12	7	11	12	16	11	11	11	10
Channel E	3	2	6	2	2	2	2	3	3	6	5	3	2	2	2
Channel F	9	8	14	8	9	8	9	14	14	17	12	15	8	10	9

Table 1

SUBJECTIVE ANALYSIS

Because the probe system only measures the transport stream delivery quality data, CMC developed and implemented a *Golden Eye* program to ascertain a subjective video quality rating of the sources of services it delivers. This data is accumulated over time and used to determine average video quality of each service placed on the backbone.

This program is a subjective video quality assessment method devised by the CMC using trained observers to perform subjective quality measurements of the source video, which is then processed and placed onto an IP network. Each service is viewed for 10 minutes on a predetermined schedule, which ensures random quality measurements. There are considerable differences noted between content that is aired (i.e. SD infomercials on the Overnight block, legacy or older material that is upconverted during Day-part, and first run Native HD material in Primetime.) It is important to note that these channels are marketed as “High Definition”, “HD” or “High Def” and the consumer expectation is not a variable.

The Golden Eye observer assigns a level of quality to the channel under test based on their observations of any impairment during the test cycle. A sample of the Golden Eye observer subjective test results is shown in Table 2 and is based on the test criteria which are listed in Table 3.

At the heart of video service quality grading is the use of common industry terms, definitions, and tolerances to known digital video impairments such as contouring, haloing, macroblocking, noise, smearing, and pumping. While these terms are commonly used for describing artifacts that impair the customer's QoE, it is very important that Golden Eye analysis apply them in a consistent manner. Table 3 was developed in order to assure a common language to describe the nature of the digital artifact presence in the video being analyzed.

Understanding there will always be some level of each of these impairments in every digital video service, and that some are more irritating to the cable subscribers than others, weighting of each impairment measurement type is required. After each impairment type is graded, the impairments that are more irritating to the consumer must be assigned a greater weighting. The total of the six impairment types combine with these weightings to determine an overall score. This provides a view of program picture quality and the probable offending impairment types contributing to lower quality scores.

Date	Time	Channel	Program	CONTOUR	HALO	MACRO	NOISE	SMEAR	PUMP	Channel Entry Total	Channel Average
12/15/2008	12:15	A	Program 1	2	2	2	2	3	3	44	44
12/19/2008	23:40:00	A	Program 2	4	4	4	4	5	4	82	82
12/24/2008	04:32:00	A	Program 3	3	4	5	4	4	5	86	86
12/15/2008	12:20	B	Program 1	2	2	3	2	3	2	48	48
12/19/2008	23:50:00	B	Program 2	3	4	3	3	4	3	65	65
12/24/2008	04:39:00	B	Program 3	3	3	4	4	3	3	71	71
12/15/2008	12:25	C	Program 1	2	2	2	2	4	4	48	48
12/19/2008	23:56:00	C	Program 2	2	3	3	3	3	3	58	58
12/24/2008	04:45:00	C	Program 3	3	4	3	3	2	4	63	63
12/16/2008	11:20:00	D	Program 1	2	3	2	5	5	2	64	64
12/19/2008	24:05:00	D	Program 2	5	4	5	4	5	4	90	90
12/24/2008	05:00:00	D	Program 3	3	4	2	3	3	4	59	59
12/16/2008	11:25:00	E	Program 1	2	2	4	4	2	4	66	66
12/24/2008	05:52:00	E	Program 2	5	3	4	4	5	4	81	81
12/16/2008	11:30:00	F	Program 1	2	2	2	4	3	4	56	56
12/24/2008	06:01:00	F	Program 2	4	3	4	3	2	4	68	68

Table 2

Compression Artifact Definitions		Compression Artifact Ratings			
CONTOURING	**1**	**2**	**3**	**4**	**5**
Contouring is a defect where abrupt changes between shades of the same color create color bands instead of a gradual change. This can be seen during scenes with a large amount of smooth color or in scenes where this is a color contrast change like in wide-angle sunsets, dawns, or clear blue skies.	Present in all scenes.	Present in most scenes.	Present, but not distracting to viewing.	noticeable, but not consistently seen.	no contouring.
HALOING	**1**	**2**	**3**	**4**	**5**
Haloing is typically seen around areas of high contrast, such as sharp lines, text edges, and graphics. On close inspection, part of the graphic appears to extend into the background. Halling can manifest as smaller details in graphics appearing to soften or loose edge resolution resulting in apparent blocking along graphic edges.	Heavy blocking that occurs on the edges of all objects on the screen.	Blocking along the edges of most objects.	Localized blocking noticeable along the edges of objects; always present at the edges of all text and graphics.	Blocking occurring at the edges of logos only.	no haloing.
MACROBLOCKING	**1**	**2**	**3**	**4**	**5**
Macroblocking is a defect where the edges of blocks or rows of blocks, are typically seen as a grid-like-pattern. This defect often occurs during dissolves from one scene to another or during action scenes involving a great deal of complex movement. Another way this artifact is presented is with small to large pixels and/or blocks containing corrupted or green pictures. This is caused by transmission or transport anomalies.	The whole screen Blocks up, regardless of scene content. **Note: This is a very rare event.**	10 - 15 seconds of blocking covering half of the of screen. e.g. Consistent blocking occurs during every scene transition, fade/dissolves, action scenes, etc.	5 - 10 seconds of blocking covering at least half the screen. e. g. Blocks occur at the focus of the screen; often during motion.	3 - 5 seconds of blocking around a small portion of screen. e.g. Scene transitions / dissolves / short action scenes (explosions).	no blocks.
NOISE	**1**	**2**	**3**	**4**	**5**
Noise appears as random speckles on an otherwise smooth surface and can significantly degrade video quality. Although noise often detracts from an image, it is sometimes desirable since it can add a grainy look that is reminiscent of film. Noise can also increase the apparent sharpness of an image.	Black speckle clusters that manifest clearly defined block edges. Closely resembles traditional Macroblocking, though not associated with motion.	Scenes presenting a raining effect of black speckles that cluster and move into the foreground. e.g. Noise that manifests itself and is a distraction within the scene.	Scenes with black speckles that appear to be moving on static backgrounds. e.g. walls, curtains, or sky that appear to have movement in the background.	Black dots that randomly pop into any portion of the scene.	No noise.
SMEARING	**1**	**2**	**3**	**4**	**5**
Smearing is a defect where part of the image remains fixed in space while the adjacent parts of the image moves leaving a trail. Smearing may also be observed in faces or across large areas of a similar type that have fine detail (e.g. grass fields). Smearing commonly affects facial color tones causing video to take on an unnatural look.	Regular loss of object detail that manifests as localized blocking on a face or a material.	Losing object detail on faces or materials that are the focus of scene.	Intermittent loss of object detail on faces or materials.	Individual or unrelated occurrences of object detail loss	no smearing.
PUMPING	**1**	**2**	**3**	**4**	**5**
Pumping is a defect where the video or parts of the video appear to pulse at a regular interval. This is typically seen in areas of smooth neutral colors.	All scenes with static backgrounds begin to block and bleed to the foreground.	Noticeable movement on static backgrounds during both motion and still scenes. e.g. Noticeable on an overhead view of a golf course or a wall of wood paneling.	Scenes that have minor regular movement on static backgrounds.	Low motion scenes that have intermittent but not constant movement on static backgrounds.	No pumping.

Table 3

SOURCE PROCESSING QUALITY

In addition to the Probe data and the Subjective Video Quality Analysis, the CMC also relies on Imagine's ICE-Q® technique to measure its HD delivery performance. How the human visual system perceives various video characteristics is built into the ICE-Q® measurement system, such as sensitivity to analog and digital noise, spatial and temporal frequency, and factors such as luminance, color, texture and edges. Imagine's research has found that the accuracy of an objective measurement system is primarily dependent on how closely its results correlate with subjective test results from a pool of expert and ordinary viewers. In addition, since different video coding standards exhibit different types of artifacts, a good objective measurement system should be tuned and optimized for a particular video coding environment, e.g., MPEG-2 or MPEG-4 AVC.

The ICE-Q system processes every macroblock of every frame, using variable bit rate coding to achieve and maintain constant video quality. The system selects the optimal macroblocks and frames by using the objective video quality measurement system to preserve the highest video quality at the lowest possible bit rate. The ICE Broadcast System continually measures its ability to process the input signal as accurately as possible (i.e. with minimal added impairments).

The actual grading system relies on a convenient approach using a numerical scale of 1 to 100, with 100 representing the compressed source quality. For example, a score of "97" may be defined as Just Noticeable Difference (JND), in which expert viewers can rarely discern the difference between the compressed source and the re-processed signal. A score of "95" may be defined as the point of No Material Degradation (NMD), in which differences from the source can be perceived more frequently than with JND, but the quality is still excellent. Furthermore, the system can be designed such that "97" is the target average grade over time, while "95" is the target minimum quality. It is also important to design the grading system such that the numerical increments are reasonably linear with respect to subjective video results. In other words, the subjective quality difference between "97" and "95" should be similar to the subjective gap between "95" and "93."

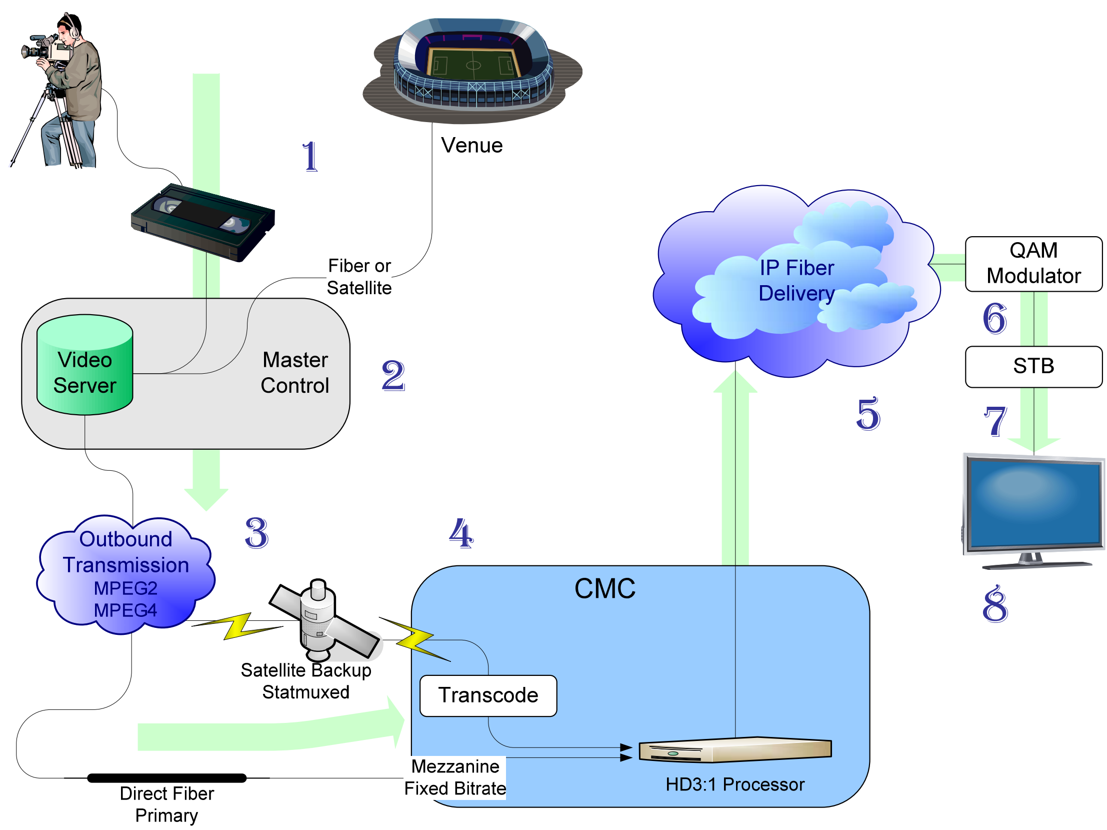
The CMC has created a summary report structure that allows the organization to monitor the incoming video quality and availability, its delivery performance, and the performance of the fiber network delivery. The level of quality is calculated from the CMC Golden Eye subjective video quality measurements, the video processor quality grading, and the delivery performance of the backbone as measured by the MPEG2 probe system. Each of these metrics is prioritized to determine actions on channels of impaired delivery quality. Channel uptime and events of packet loss are most critical followed by the Imagine quality scoring, then the CMC Golden Eye subjective scores which have the lowest priority. The results displayed in Table 4 are examples of actual video services placed on the fiber network at the CMC.

	Dec 1 - 8			Dec 8 - 15			Dec 15 - 22			Dec 22 - 29			Dec 29 - Jan 5		
Programs	Source Quality (OMC)	Average Grade	CMC Delivery	Source Quality (OMC)	Average Grade	CMC Delivery	Source Quality (OMC)	Average Grade	CMC Delivery	Source Quality (OMC)	Average Grade	CMC Delivery	Source Quality (OMC)	Average Grade	CMC Delivery
Channel A	83	98.16	5	83	98.10	80	69	98.05	35	71	97.90	16	71	97.83	8
Channel B	72	97.51	4	72	97.59	14	65	97.56	6	65	97.61	13	70	97.60	9
Channel C	81	97.90	4	80	97.98	4	77	97.95	14	67	97.99	5	67	97.99	16
Channel D	74	95.28	5	73	95.12	3	70	95.28	3	76	95.27	2	76	94.92	1
Channel E	92	95.12	3	92	95.17	80	63	94.95	8	63	94.87	4	70	95.01	7
Channel F	79	96.14	3	86	96.19	7	75	95.99	9	73	95.91	4	73	96.05	7

Table 4

QUALITY TOUCH POINTS

To maximize video quality across the delivery environment there are eight key "Touch Points" that begin at source acquisition and end with the consumer display. Video quality measurements at each touch point should be made using a non-referenced approach though a referenced approach is acceptable where a reference source is available. A referenced measurement is a comparison of the video under test to the source video providing a measurement of difference between the two; a level of degradation. A non-referenced video quality measurement is made without a comparison to the source video, which is not available at most touch points, and is the more telling measurement of the two. Drawing 2 shows the video delivery path with the Touch Points indicating critical measurement points and opportunities to improve video quality.



Drawing 2

Touch Point 1 is a measurement point at the source, such as a live event venue or a dubbing house. This is a critical point requiring non-referenced video quality measurements. Video artifacts are caused by analog to digital conversion and tape to digital storage conversion requiring compression of the video.

Touch Point 2 at the Master Control facility requires non-referenced measurements because there are many video sources required to create a video service. Service degradation causes include varying levels of video quality of programs and commercials and failures of automation resulting in loss of video or the wrong video played out. This is also the point where the CMC will be placing mezzanine encoders for direct fiber delivery.

Touch Point 3 can be a referenced measurement assuming the service provider is also the owner of the MPEG encoders and multiplexors and that a reference, uncompressed source is available. Service challenges include poor video encoding and multiplexor bit starvation causing the six video impairments described in Table 1.

Touch Point 4 is the point of entry into the CMC facilities and requires a non-referenced measurement. The service is left intact as MPEG encapsulated onto IP. The affects to video are typically due to packet loss on terrestrial delivery systems and packet loss on satellite delivery systems that are due to RF conditions. However, because this is the first point at which CMC has access to the video service, effectively making it the point of demarcation, the video service must be validated.

Touch Point 5 is interface to and from the CMC IP delivery system and requires referenced measurements assuming reference from the delivery source is available. This is the first point in the service path where the affects to the video service caused by processing can be measured. Degradations of the video service at this point are often caused by over processing and bit starvation of the multiplexor.

Touch Point 6 is a measurement point from a monitor tap of the QAM devices. This point requires non-referenced measurements because there is no reference source available. Video degradation is caused by local conditions of delivery, grooming, and commercial add insertion.

Touch Point 7 is a measurement point within the set top box in the customer's home. This requires non-referenced measurements, since there is no reference source available. Video degradations can be caused by HFC delivery and conditions within the customer's home.

Touch Point 8 is a non-referenced measurement point within consumer video displays measuring the performance on the very end of the delivery path. Clearly, this is the point where all upstream events and conditions affect the perceived video quality as well as transmission availability as presented to the cable customer.

CONCLUSION

With the increasing importance of Quality of Service as a key marketplace advantage, the ability to measure, report on, and improve video quality delivered to the consumer is an essential tool for improving competitive strategy. Replicating both ideal and imperfect viewing conditions within the software algorithms of Picture Quality Analysis tools is critical to automating non-referenced video quality measurements.

In addition, accumulated video quality scores provide MSOs reliable and repeatable metrics. This data can be used for engaging content providers (and all parties involved in content management and distribution) in a collaborative effort to achieve the best possible viewing experience for the customer. Additionally, industry accepted video quality measurements and grading enables MSOs to perform regular proof of performance maintenance of cable systems.

In the long term, digital video quality grading must take place in all devices in the path, from source to consumer, which can impact the quality of video. Further, instantaneous scoring is necessary for best determining devices and systems in the video service path that are contributing to the degradation of video quality and accumulated scores will enable trending and isolation of issues that lead to quality impacts and impairments.

It is the authors' hope that this paper helps to stimulate further discussion and movement toward a common system and automated processes for measuring and scoring digital video quality from an unreferenced source based on industry accepted tolerances.

Summary of Video Quality Measurement "Lessons Learned"

- Linear video is subject to varying video quality from one channel to another as well as from one program to another. It is also evident that video quality varies from one source provider to the next due to source acquisition and delivery approaches.

- Time of day has a direct impact on type of content being aired and whether it is high quality native HD or legacy SD upconverted material.

- MPEG 4 transcodes to MPEG 2 cause new "mini" macro-block impairments.

- Older film product and marginal quality associated with film to video transfers also have a significant impact on QoE.

- Live sports that is over compressed (whether it occurs on the path from the *Venue*, the *Outbound Transmission* or as part of a downstream re-encode process) can and will cause significant and visible impairments.

- Display viewing distance and "off-axis" viewing can impact perceived video quality.

- Display size, type, native resolution capability, contrast ratio, and pixel size and shape can have impacts on recovered video.

- Deeper compression approaches such as HD 3:1 prevent the use of downstream "Rate Shaping" or "Grooming" as the impacts to the recovered QoE are dramatic.

Acknowledgements

Video Layer Quality of Service: Unprecedented Control and the Best Video Quality at any Given Bit Rate, NCTA Technical Papers, May 18, 2008, The Cable Show '08, Ron Gutman and Marc Tayer, Imagine Communications, Inc.

---

[i] March 31, 2008, *"Cable Operator Video Quality Study,"* Multimedia Research Group, Inc. (MRG, Inc.),
[ii] January 1, 2008, *"HD Monitoring - Raising the Quality Bar,"* Craig Kuhl, Communications Technology magazine
[iii] IBID, MRG's *"Cable Operator Video Quality Study"*

DVB-C2 – THE SECOND GENERATION TRANSMISSION TECHNOLOGY FOR BROADBAND CABLE

Philipp Hasse, Dirk Jaeger, Joerg Robert
Institut fuer Nachrichtentechnik at Technische Universitaet Braunschweig

Abstract

Competition in the media and communication sector is increasing. MSOs have to align their business models with new emerging requirements on a permanent basis. Although requirements vary in different national markets in Europe and worldwide, the cost per transmitted bit is an important factor for staying competitive. Thus one of the key requirements for modern communication technologies defined by European MSOs is to increase the spectral efficiency of downstream transmissions by moving the efficiency as close as possible to the Shannon limit – the theoretical optimum.

The second generation of the DVB system for cable – called DVB-C2 – is an innovative approach making use of state of the art communication technologies which have never been implemented in broadband cable networks before. An OFDM (Orthogonal Frequency Division Multiplex) based modulation scheme in combination with a two-dimensional interleaving (in both frequency and time) and an LDPC (Low Parity Density Code) error protection mechanism provide a spectral efficiency fractions of a dB below the Shannon limit. Will there ever be any reason to develop a new PHY for cable after DVB-C2 is widely implemented?

INTRODUCTION

The Digital Video Broadcasting (DVB) transmission system for cable (DVB-C) [1] defines physical layer and low-layer signaling techniques for the transmission of digital broadcasting services via cable networks. The technology was also incorporated in the European version of DOCSIS [2]. Since its development more than 10 years ago, DVB-C has been commercially implemented in millions of products such as (Edge)QAMs, set tops, and cable modems all over the world.

During the recent 10 years, research in communications techniques has progressed. State of the art algorithms have improved significantly in terms of performance and flexibility and Moore's law has continued to happen. These trends have facilitated the commercialization of advanced systems for wireless and wire-line communication.

Forced by the increasing competition in the media and communications sector, European MSOs considered utilizing the benefits provided by such modern communication technologies in broadband cable networks. In 2006, the DVB [3] Project was approached to launch a work item aiming at the development of a second generation DVB system for cable called DVB-C2.

Figure 1: Logos of DVB and ReDeSign

This article describes the DVB-internal development process of the DVB-C2 specification. It focuses on the work carried out by the DVB Technical Module and its TM-C2 technical experts group launched to prepare the technical specification. The DVB work is supported by the research project ReDeSign. The Institut fuer Nachrichtentechnik (IfN) of Technische Universitaet Braunschweig has taken key positions in both organizations, DVB and ReDeSign. Furthermore IfN staff has been in charge of DVB work on dedicated DVB-C2 techniques such as signaling and synchronization and has made important contributions incorporated in the final specification.

DVB AND REDESIGN

The development of DVB-C2 has involved some 20 companies including MSOs, IC manufacturers and equipment vendors. The chairmanship is provided by Kabel Deutschland, one of the major European MSOs. Based on a liaison agreement, the work is supported by the European research project ReDeSign [4] which has allocated considerable resources to the work. One of the tasks of the ReDeSign team is for instance to support the standardization work by means of computer simulations.

An initial Technology Study Mission executed among DVB member companies by the Technical Module confirmed the potential improvements achievable with latest communication technologies in cable. Prior to the technical work, use cases were studied and business requirements were defined by DVB's Commercial Module. Major requirements are summarized further down in the article. The technical work was kicked off by means of a Call for Technologies (CfT). The responses created the basis of the subsequent development of the DVB-C2 specification.

An important tool for the development of DVB-C2 has been the software simulation platform. The simulations focused on two goals: First to compile the final DVB-C2 system from individual elements of the responses to the CfT and secondly to execute performance simulations of the system. A cable channel model was adopted for this purpose [5]. The model was developed by the ReDeSign consortium. To a large extent, the channel model takes account of the results derived during the work of IEEE 802.14 [6]. The non-linear behavior of digital signals, however, was researched in ReDeSign. Software models were developed incorporating the results of this work. As explained later, intensive investigations were necessary to generate the knowledge required for the creation of software tools modeling the statistical behavior of second and third order intermodulation products generated by multiple QAM signals.

COMMERCIAL REQUIREMENTS

The DVB Commercial Requirements [7] constitute the fundament for the development of the technical solution. The major requirements are summarized as follows. DVB-C2 needs to provide:

- A toolkit solution for optimal implementation under all conceivable transmission conditions of various cable networks
- Increased spectral efficiency by at least 30 % compared with the highest efficiency provided by DVB-C
- Advantageous operational flexibility
- Support of removal of fixed cable channel rasters
- A low latency mode
- Adoption of state of the art technologies
- An integral solution of the DVB Family of Standards approach: reuse of elements of existing DVB standards (e.g. of DVB-S2 and DVB-T2), wherever appropriate
- Support for possible incorporation in DOCSIS
- Cost efficient production of cable devices such as set-tops but also of trans-modulators used in headends of SMATVs

It is worth mentioning that the issue of backwards compatibility with DVB-C was sacrificed to the requirement of optimizing the spectral efficiency and thus the throughput of the new technology. 26 requirements and 5 exemplary use cases were defined in total.

THE DVB-C2 TECHNOLOGY

The DVB-C2 specification (still in final draft stage) defines the signal processing at the transmitting end. The high-level block diagram is given in figure 2. Basically the signal processing comprises an input processing block, the multiple stages FEC unit and the OFDM generator.

Physical Layer Pipe (PLP)

The DVB-C2 system is intended to provide a transparent data link for all kinds of digital information. Data formats such as MPEG Transport Stream (TS) [8] and DVB Generic Stream Encapsulation (GSE) [9] specially designed to support the transport of IP data can be interfaced to the system. For enabling the transparent transmission, a PLP concept was defined. A PLP takes the role of a basic container utilized for the transport of the data. A PLP container can for instance comprise a multiple program Transport Stream, a single program, a single application, or any kind of IP based data. The data being inserted in a PLP container are converted into the DVB-C2 internal frame structure by the Data Input Processing Unit and subsequently processed by the FEC (Forward Error Correction) encoder. The FEC is composed of a Bose-Chaudhuri-Hocquenghem (BCH) outer Code [10] and a Low-Density-Parity-Check (LDPC) inner code [11]. The LDPC is a very powerful tool for correcting transmission errors whereas the BCH's task is to eliminate the error floor which may be produced by the LDPC decoder in the receiver under certain transmission conditions. The BCH code rates add less than 1 % redundancy. The LDPC code rates available range from 2/3 to 9/10. The FEC encoded data are mapped onto QAM constellations which eventually will be assigned to the individual sub-carriers of the OFDM symbol. A number of QAM constellations are defined ranging from 16-QAM up to 4096-QAM. For each QAM constellation a dedicated set of LDPC code rates is selected. A combination of the modulation and coding parameters is called ModCod. The ModCods defined for DVB-C2 are listed in table 1. Adding a PLP header completes the creation of the PLP container.

	16-QAM	*64-QAM*	*256-QAM*	*1k-QAM*	*4k-QAM*
9/10	X	X	X	X	X
5/6	X	X	X	X	X
3/4	X	X	X	X	
2/3		X			

Table 1: ModCods of DVB-C2

Data Slice

Several PLPs may be combined to form a data slice whereas the protection power of the FEC assigned to each single container can be optimized individually by using an appropriate ModCod referred to above. The Data Slicer has the task to move the PLP container and group of containers, respectively, to a defined location within the bit stream in such a way that they will be transmitted by dedicated sub-carriers of the OFDM symbol and thus to appear at dedicated frequency subbands of the spectrum occupied by the OFDM symbol.

A two-dimensional interleaving (in time and frequency domain) is applied to each individual data slice in order to enable the receiver to eliminate impacts of burst impairments as well as of frequency selective interference such as single frequency ingress or possibly even by fading.

OFDM frame builder

The OFDM frame builder combines several data slices together with auxiliary information and additional pilot sub-carriers. The pilot structure chosen is composed of continuous and scattered pilots. The continuous pilots are allocated at dedicated sub-carrier locations in each OFDM symbol, which are constant and do not vary between the symbols. The amplitudes of the pilots are boosted by a factor of 7/3 for improved robustness. A DVB-C2 receiver makes use of the 30 continuous pilots allocated per 6 MHz bandwidth to perform fine synchronization of the signal in frequency and time direction. By an averaging calculation of the continuous pilots, the common phase error caused by phase noise (induced e.g. in the RF frontend of the receiver) can be detected and compensated. After the synchronization is completed at the receiving end, the scattered pilots are used for channel equalization purposes. The inverse channel impulse response can easily be determined from the differing values between the received scattered pilots and their equivalent values transmitted known to the receiver. Also the equalization process is easy for an OFDM system since a simple multiplication of each sub-carrier with an almost time-invariant factor can be applied rather than a complex convolution process needed in case of a single carrier solution. The number of scattered pilots is aligned with the length of the guard interval. 96 and 48 parts of all sub-carriers are used as scattered pilots in case the relative guard interval length equals to 1/128 and 1/64, respectively, of the symbol duration.

The auxiliary data referred to above contain mainly the so-called L1 (Layer 1) signaling information which is put in front of each OFDM frame in terms of a preamble. The preamble uses the complete set of sub-carriers of an OFDM symbol and provides the receiver with means required to access the PLPs. For instance, information about the start positions of the data slices within the OFDM frame is transmitted as well as the ModCod information of the PLP headers. Due to their importance, the PLP headers themselves are highly protected by means of robust modulation (e.g. QPSK or 16-QAM).

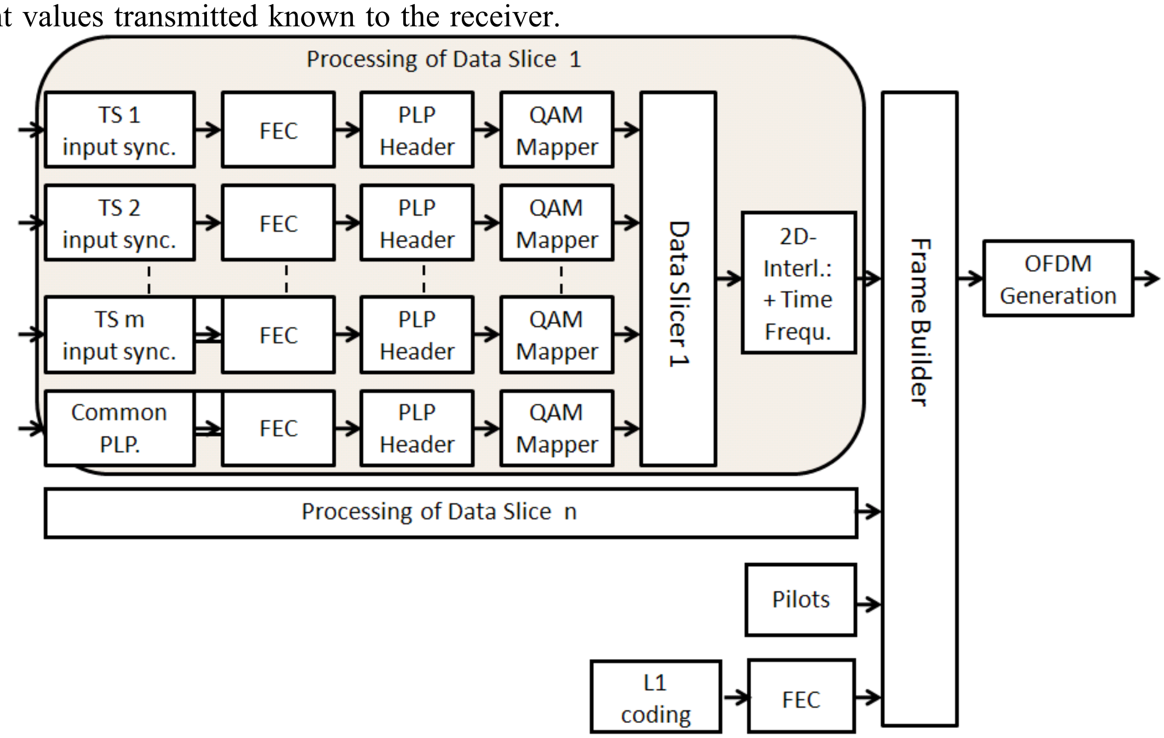


Figure 2: High-level block diagram of the DVB-C2 encoder

OFDM generator

The OFDM symbols are generated by mean of an inverse Fast Fourier Transformation (IFFT). A 4K-IFFT algorithm is applied generating a total of 4096 sub-carriers, 3409 of which are actively used for the transmission of data and pilots within a frequency band of 6 MHz. These 3409 active sub-carriers occupy some 5.7 MHz. The sub-carrier spacing calculates to 1,672 Hz which corresponds with an OFDM symbol duration of 598 µs. For markets employing 8 MHz channels such as the one in Europe, the entire system can easily be scaled up by a factor of 4/3 which results in a sub-carrier spacing of 2,232 Hz and a symbol duration of 448 µs. The OFDM signal injected in an 8 MHz channel occupies a bandwidth of 7.6 MHz. The guard interval used between the OFDM symbols has a relative length of either 1/128 or 1/64 compared to the symbol length itself and thus durations of 4.7 µs in the US respective 3.5 µs in Europe.

DVB-C2 FEATURES

The OFDM based concept implemented in DVB-C2 provides a couple of benefits in terms of flexibility as well as of efficiency advantages against single carrier solutions. Some of these benefits are introduced in the following sections.

Channel assembly at physical layer

The DVB-C2 system supports the feature of flexible and dynamic bandwidth allocation to individual data slices as well as to the entire signal. It is for instance feasible to combine various adjacent channels at physical layer in a very effective manner. Note that the term channel bonding was deliberately not used in this context in order to prevent confusion with the DOCSIS 3.0 feature of channel bonding at MAC layer. In contrast, DVB-C2 is capable of combining various adjacent channels to a single wide-band channel at physical layer. For combining 4 channels, the 4K-IFFT unit of the OFDM generator at the transmitting end needs to be replaced by a 16K-IFFT algorithm. In case 8 channels are to be combined, a 32K-IFFT units needs to be installed in the headend equipment. It is noted that higher order IFFTs can be utilized for backwards compatible transmissions in 6 MHz channels, for instance, requiring a 4K-IFFT algorithm. The feature to combine channels at physical layer has some advantages compared to the conventional bonding approach at MAC layer. For example, the method allows making use of the frequency bands located at the edges of the individual 6 MHz cable channels traditionally occupied by filter slopes of the DVB-C and DOCSIS signals (see figure 3). This feature further enhances the efficiency of DVB-C2 since, for instance, a combination of 4 cable channels results in an increase of the total bit rate which is higher than this factor 4.

Support of 6 MHz receiver bandwidth

Another example of a benefit is the possibility to map individual data slices to dedicated OFDM sub-carriers and thus to dedicated frequency sub-bands of the OFDM spectrum. In addition, DVB-C2 ensures that the L1 signaling has a periodic structure with a repetition period equal to 6 MHz (in the US – 8 MHz in Europe). As the wanted data slice has an equal or smaller bandwidth, it is possible for a receiver devices operating at a bandwidth of 6 MHz to receive a 6 MHz frequency window of the DVB-C2 spectrum although the entire bandwidth of the DVB-C2 signal is 18 MHz, for instance. In fact, the receiver decodes only those sub-carriers transmitted within this 6 MHz window, as illustrated by figure 3.

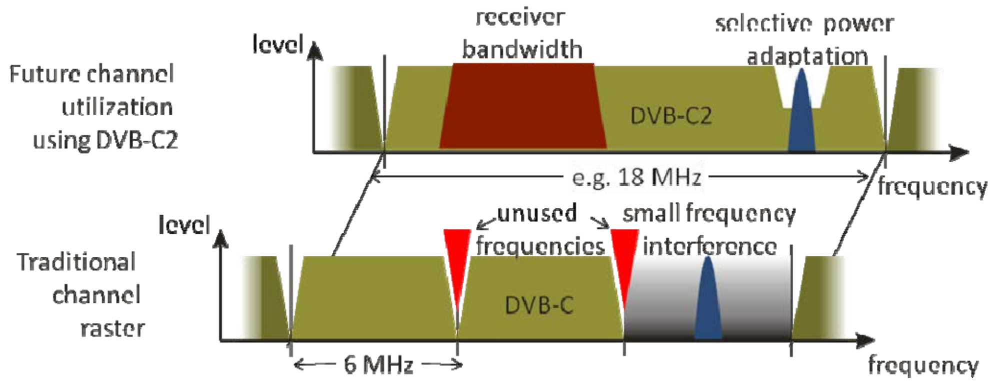


Figure 3: Illustration of features of DVB-C2s OFDM based transmission techniques

Prevention of narrow-band interference

Figure 3 also shows a further useful feature of DVB-C2. The signal power of subcarriers can be reduced to selectively adapt the power density distribution of the signal for prevention purposes of co-channel interference. This feature can be helpful in case a safety radio service transmitted at a co-channel frequency needs to be protected by regulation. While occupying the remaining part of the channel, DVB-C2 reduces only the power of those subcarriers transmitted at the critical frequencies.

PERFORMANCE FIGURES

This section provides information about the performance of DVB-C2. In many technical articles, the bit-error rate (BER) as a function of signal-to-noise ratio (SNR) is used to demonstrate the performance of a transmission system. This article chooses the spectral efficiency as an appropriate performance indicator. Digital transmission systems equipped with powerful error correction mechanisms produce BER curves with very steep slopes. In practice this means that we can refer to a single SNR value above which perfect reception is achieved whereas below that value no reception is possible (see related marks for DVB-C and DVB-C2 in figure 4). The spectral efficiency adds information about the available capacity in terms of bit-rate per Hz bandwidth.

Simulation parameters

The spectral efficiency figures presented in this article were obtained by means of computer simulations carried out at Institut fuer Nachrichtentechnik at Technische Universitaet Braunschweig for the research project ReDeSign. The results represent a quasi error-free DVB-C2 transmission in a Gaussian noise environment.

For the simulation, 4 channels, each of a bandwidth of 6 MHz, are combined to form a wide-band channel of 24 MHz. A guard interval length of 1/128 of the symbol duration was chosen as well as the corresponding pilot density of 1/96. Each of the 2 spectral guard bands at the wide-band channel edges has a width of 200 kHz.

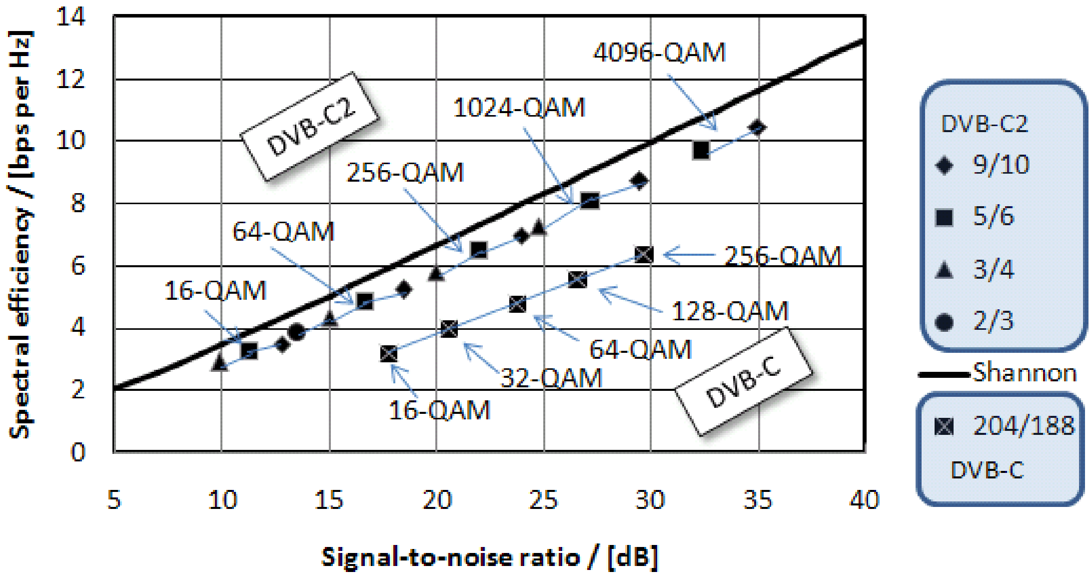


Figure 4: Spectral efficiency as a function of SNR for DVB-C2 and DVB-C

For determining the general performance of DVB-C2, an additive white Gaussian noise (AWGN) channel model was implemented. This rather simple channel model allows a better comparison with equivalent performance figures of DVB-C and DOCSIS which both represent digital cable transmission systems in operation today. In addition, the results can be better compared with the Shannon limit [12], [13] defining the theoretical channel capacity limit for the transmission of digital information through channels impaired by Gaussian noise. A more sophisticated cable channel model was implemented in the further simulations explained later.

Increased robustness

The diagram in figure 4 clearly shows the advancement of DVB-C2 compared to DVB-C (and DOCSIS, respectively) in terms of both spectral efficiency and reduction of SNR required for reception of a signal with a dedicated spectral efficiency. In the following, the 64-QAM mode is used by means of an example. This modulation scheme provides a spectral efficiency of almost 4.8 bps per Hz when applied in both DVB-C and DVB-C2, the latter using a code rate of 5/6. It is observed that the robustness of DVB-C2 increases by some 7 dB compared to DVB-C. The SNR required to receive the corresponding DVB-C signal equals approximately 24 dB whereas the simulation of the equivalent DVB-C2 signal resulted in a required SNR of some 17 dB.

Increased spectral efficiency

Comparing the spectral efficiencies of both technologies at an SNR of 24 dB(e.g. 64-QAM for DVB-C and 1k-QAM for DVB-C2), it is observed that the efficiency increases from 4.8 bps per Hz to 7 bps per Hz which is an increase of almost 50 %.

First estimations of efficiency gains based on an introduction of DVB-C2 in combination with further technologies such as statistical multiplexing and advance audiovisual coding (e.g. MPEG-4) resulted in a total efficiency increase of more than 100 %. This figure corresponds with a bandwidths saving achievable in cable networks of more than 50 %.

FURTHER SIMULATIONS

Further system simulations were carried out using a more sophisticated channel model. The model was developed within ReDeSign. It compiles information generated in earlier research work (e.g. by IEEE 802.14 [7], DigiSMATV [14]) as well as by ReDeSign internal studies and measurement campaigns. Different parameter sets of the channel model were defined to address the varying transmission conditions of real networks as well as to reflect short-term and longer-term network scenarios. For the short-term scenario, a mixed transmission of analogue and digital signals was considered whereas the longer-term scenarios reflected the situation after analogue switch-off.

The complex channel model includes typical impairments which a signal suffers when travelling through a cable network such as phase noise, echoes, impulse noise, etc. While the statistical or deterministic nature and quantity of several impairments is well known to the cable industry, one of the most challenging investigations for completing the channel model was the study of effects created by interference which are generated by non-linear transfer functions of active cable components. In particular, the contribution from digital signals in both a mixed analogue-digital and a digital-only environment was analyzed. Measurement campaigns were organized at premises of equipment manufacturers Kathrein and VECTOR (the latter also being a member of the ReDeSign consortium). At the labs of the two companies, a cable network emulator was set up compiling a chain of broadband amplifiers and an optical link as well as in-home network components. The network emulator was fed with 94 DVB-C signals. Alternatively, a mixed signal scenario with adequate signal numbers was established. Intermodulation products were captured into a large memory for statistical investigations performed afterwards. These investigations are not completed yet. First preliminary conclusions show that intermodulation products of digital signals have an impulsive noise-like behavior with a varying statistical occurrence probability of their amplitudes depending on the network load and on the signal levels respectively. Related investigations performed mainly by the Dutch research institution TNO (member of the ReDeSign consortium) are ongoing. [5] It is planned to submit the results to CENELEC [14] and IEC [15].

SUMMARY

Launched by a request from European cable operators, the DVB Project has developed a second generation downstream transmission technique for cable networks called DVB-C2. The specification defines a toolbox providing a set of parameters utilizable to adapt the system to varying transmission conditions of the networks in Europe and worldwide. The spectral efficiency compared to DVB-C and DOCSIS has been increased considerably and allows transporting more than 60 Mbps through a regular cable channel of 6 MHz. Many advantages are supported by means of the flexible channel bandwidth agility of the system. During its development the DVB TM-C2 experts have taken account of the requirement for a low latency mode allowing to incorporate the DVB-C2 physical layer technology in the existing DOCSIS system. DVB-C2 therefore has the potential to become the physical layer and lower layer signalling technology for the future generations of cable based downstream transmission systems.

ACKNOWLEDGEMENT

The authors appreciate the European Commission's funding and support given to the ReDeSign project. We also thank the member companies of the ReDeSign consortium, namely Alcatel-Lucent, ANGA, BLANKOM Digital, Telenet, TNO, VECTOR, ZON TV Cabo who all have contributed to the findings presented in this article. Without the important technical contributions to the development of the DVB-C2 specification made by the members of the DVB Technical Module and its experts group TM-C2 represented by its chairman Christoph Schaaf of Kabel Deutschland it would have not been possible to write this article. We thank all of them for the devotion to the DVB-C2 development. In particular we thank the representatives of Sony for the excellent cooperation in creating the DVB-C2 encoder and decoder modules of the software platform. Also the opportunity provided by Kathrein to analyze the effects of intermodulation products generated by digital cable signals at the labs at Kathrein premises is appreciated by the authors. Evenually, the cooperation with CENELEC made possible through a liaison with ReDeSign is greatly acknowledged.

REFERENCES

[1] Reimers, U.: DVB – The Family of international standards for Digital Video Broadcasting, 2nd Ed. Springer, Berlin Heidelberg New York, 2005

[2] Data-Over-Cable Service Interface Specifications – Downstream RF Interface Specification. CableLabs, September 2008

[3] Digital Video Broadcasting project, DVB, www.dvb.org (status 28.01.2009)

[4] ReDeSign project, www.ict-redesign.eu (status 28.01.2008)

[5] HFC Channel Model. ReDeSign Deliverable 08, December 2008, , www.ict-redesign.eu (status 28.01.2009)

[6] Kolze, T.: HFC Channel Model Submission, IEEE Project 802.14, Cable TV Protocol working Group, 26 May 1998

[7] Advanced Coding and Modulation Schemes for CATV systems (DVB-C2) – Use Cases and Commercial Requirements. DVB internal document no CM-C20018R4, January 2008

[8] Information technology – Generic coding of moving pictures and associated audio information: Systems. International Standard ISO/IEC 13818-1, December 2002

[9] Digital Video Broadcasting (DVB); Generic Stream Encapsulation (GSE) Protocol. ETSI TS 102 606 V1.1.1, October 2007

[10] Lin, S.; Castello, D. J. Jr.: Error Control Coding. 2nd Ed. Pearson Prentice Hall, New Jersey, 2004

[11] MacKay, D. J. C.: Good Error-Correcting Codes Based on Very Sparse Matrices. IEEE Transactions on Information Theory, Vol. 45, No. 2, March 1999

[12] Shannon, C. E.: A mathematical theory of communication, Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948

[13] Shannon, C. E.: Communication in the Presence of Noise. Proceedings of the IEEE, VOL. 86, NO. 2, February 1998 (reprint)

[14] http://www.cenelec.eu/Cenelec/European Committee for Electrotechnical Standardization, CENELEC, www.cenelec.eu (status 28.01.2009)

[15] International Electrotechnical Commission, IEC, www.iec.ch (status 28.01.2009)

Getting to 4G via Wi-Fi in the Real World

David Park
Belair Networks

Abstract

As the economy shrinks and the competition for profitable revenue generating subscribers increases, operators increasingly use bundled packages of services to capture and retain customers. One key tool for that is the inclusion of wireless services. This paper describes how wireless services can be effectively deployed on an MSO's existing HFC infrastructure. It describes both licensed and unlicensed technologies and how they compare. It also describes how mobility can be deployed on HFC and the difference between macrocellular and microcellular systems and their performance.

INTRODUCTION

Traditionally, cable operators have looked on wireless options associated with licensed and unlicensed spectrum as either/or propositions. But, with the emergence of fully mobile cable-optimized basestations, Wi-Fi can be used as a complement or an alternative to next-gen mobile standards that rely on licensed spectrum. Multi-radio platforms can operate in both licensed and unlicensed bands and connect directly to the HFC (Hybrid Fiber Coax) network.

One of the key business strategies for acquiring and retaining customers of the past few years has been the use of bundled service packages. In this MSOs have been uniquely positioned as the HFC offers a very cost effective platform to deliver TV, data and voice services to a customer premise. In contrast traditional Telcos has been forced to put in place extensive upgrades to their system, going to the extent of putting fiber optic connections to individual properties, in order to compete with MSOs. The next frontier in bundling is the addition of wireless services to the package.

Mobile wireless systems offer particular challenges to MSOs, notably;

- Spectrum
- Mobility

In this paper I describe the difference between macrocellular and microcellular techniques for providing wireless coverage and capacity. I also describe how the HFC can be leveraged as a mounting location for wireless basestations and some of the technologies that are required to deliver this. In conclusion I show that MSOs can deploy Wi-Fi today and achieve capacities and user experiences that are better than all of today's 3G cellular standards and the future 4G standards.

Figure 1 shows that this approach is reality today, with numerous MSO deployments of HFC based microcells in progress.

Figure 1 Wi-Fi Microcell being installed on HFC

MACROCELLS, MICROCELLS & CAPACITY

Macrocells

The traditional method of deploying wireless services used by cellular carriers has been the use of roof top macrocells. These types of cells were quite suitable for deploying services that were primarily voice orientated. A macrocell deployment has the following characteristics.

- Prone to obstructions by terrain and buildings
    - Cold spots in coverage common
    - Building shadowing causes dead zones
- Signal levels are constrained at distance by diffraction losses over buildings and into streets
    - Lower signal strengths = lower throughputs
- Frequency and spectrum re-use limits capacity at central point
    - Can't keep adding channels
- Pluses
    - Provides overlay coverage
    - Large cells mean fewer mobility events
    - Provides in-building coverage for tall buildings (3D)
- Challenges
    - Capacity and user experience

Figure 2 Macrocell approach

Microcells

The alternative approach has been to deploy microcells. These smaller cells offer the following characteristics;

- Many times the capacity of macrocells
- Go around obstructions by terrain and buildings
    - No cold/dead spots
- Signal levels are higher due to small cell sizes
    - Higher signal strengths = higher throughputs
- Frequency re-use simple
    - Reuse channels frequently
- Pluses
    - Capacity and user experience
- Challenges
    - Need more cells
    - more mobility events
    - More challenging to provide in-building coverage for tall buildings (3D)
    - Backhaul

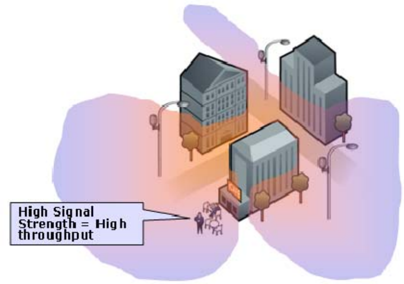


Figure 3 Microcell approach

Microcells have been used infrequently by cellular operators to date because of the challenge of getting wired backhaul to them. The high cost of deploying a fractional T1 connection into the streets has resulted in microcells being used for either cold spot (lack of coverage) of hot spot (capacity) fill-in rather than ubiquitous coverage. This has been the case for voice systems, however, as will be demonstrated, 4G systems need microcells to fulfill their capacity and throughput claims. Fortunately the MSO industry has the mounting assets required.

4G systems need microcells

One of the key differentiators of 3rd and 4th generation wireless systems as compared to the traditional cellular systems has been the use of higher order modulation schemes to deliver enhanced capacities and throughputs. MSOs are no strangers to this, with the DOCSIS® system using high order QAM to deliver bandwidth to their customers. However, the HFC is delivered to the end customer via coaxial cable, with amplifiers that keep the signal to noise high enough for all modems that are connected to the system to operate at the best modulations. A wireless system in contrast, propagates its signal over the air, and the signal that a user gets is highly dependent on their distance from the cell site. This is especially true once higher order modulations are used. This is illustrated in Figure 4.

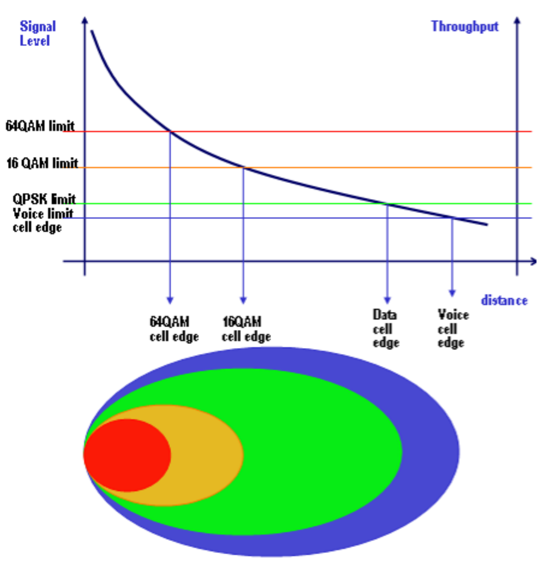


Figure 4 Modulation versus Distance

What this means to a cellular provider is that only a small percentage of the cell is covered by the highest order modulations. Given that cell site locations are driven by real estate concerns, and the customers are individuals who are mobile, it is reasonable to assume that end users are uniformly distributed throughout the cell. This means that the average throughput of a cell which also represents its capacity is degraded due to the distribution of modulation rates.

This has a profound impact on the deliverable performance from a cellular network. While most commentators describe wireless systems in terms of the peak throughput that they can deliver, or sometimes even the raw modulation rate, it is more instructive and useful to look at 2 aspects of the system in more detail.

- The peak user performance that can be expected
- The average capacity across the covered area.

To determine the average capacity of a cell we need to estimate the percentage of the cell that can operate at each modulation An example calculation is shown in Figure 5 for an LTE [1] macrocell using a 5MHz channel in a tri-sectored cell. This assumes the availability of 2 blocks of 15MHz of licensed spectrum.

Modulation	QPSK	16QAM	64QAM
Base Bits per Hz	2	4	6
Code rate	0.67	0.83	0.83
Derated bits per Hz	1.33	3.33	5
Bit Rate	6.67	16.67	25
Thoughput	5.33	13.33	20
SNR	5	15	25
Uplink Sensitivity	-97.99	-87.99	-77.99
Downlink Sensitivity	-95.99	-85.99	-75.99
BTS TX pwr	38	38	38
BTS ant gain	17	17	17
MS ant gain	-2	-2	-2
Uplink link budget			
Downlink link budget	149.0	139.0	129.0
Fade margin	9	9	9
overall DL link budget	140.0	130.0	120.0

Distance (km)		0.65	0.35	0.20
Distance (mile)		0.40	0.22	0.12
Cell Area (km^2)	1.33	1.33	0.38	0.13
Ring Area (km^2)		0.94	0.26	0.13
%		71.01%	19.53%	9.47%
Blended Data Rate (Mbps)	8.28			

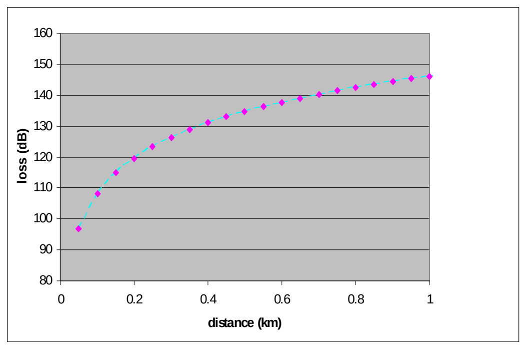
LTE Rx BW	5	MHz
subcarriers	15	kHz
subcarriers	301	
uplink resource blocks	25.083	
occupied BW	4.515	MHz
LTE BTS Noise Figure	4	dB
LTE UE Noise Figure	6	dB
LTE BTS Rx Noise Floor	-103.0	dBm
LTE MS Rx Noise Floor	-101.0	dBm

Figure 5 Example LTE link budget (macrocell)

This shows that while the 5MHz channel being used can in theory deliver a 25Mbps connection, the top modulation rate is only achieved across 9.5% of the cell. The blended capacity across the cell is 8.3Mbps. As can be seen this is far below the marketing numbers of 100Mbps download and 50Mbps upload. This capacity is shared across the entire cell/sector of the BTS. For a cell radius of 650m the circular area is 0.5 square miles. If the cell site is tri-sectored then the sector area is 0.17 square miles and the capacity per square mile is 48Mbps. This can be contrasted with the performance expected from a Wi-Fi based microcell approach where capacities of 200Mbps per square mile are deliverable today with growth to 0.5Gbps per square mile.

When similar calculations are done for today's 3G standards such as HSPA and EVDO, capacities in the order of 4Mbps per sector are achieved. Results can also be calculated for WiMAX and for use of larger amounts of spectrum. For example a blended WiMAX macrocell with 10MHz of spectrum yields 13.27 Mbps of capacity (note this is effectively less than the case shown for LTE as it is using twice as much spectrum). It should also be noted that these calculations are for the downlinks only, as that tends to drive the user experience of an end user.

The propagation model used in this case is the Empirical COST-Walfisch-Ikegami Model [2] which is representative of an "above the rooftop" macrocell type deployment. The loss versus distance and parameters used for this model is shown in Figure 6.



Frequency	f	1990	MHz
Base Station Height	h_1	27	m
Mobile station Height	h_2	1	m
Distance	d	0.7	km
Height of Building Roofs	hr	20	m
Width of Road	w	20	m
Building Separation	b	40	m
Road Orientation	p	45	degrees
City Type		0	1 = medium/suburban or 0= large metro

Figure 6 Example LTE link loss (macrocell)

MSOs have generally prided themselves on delivering the fastest user experiences, in contrast to the typical wireless system. In order to improve theses capacities, a microcellular approach is required.

Wi-Fi microcells as a complement

In the preceding sections it has been demonstrated that macrocells deliver relatively low capacities and user throughputs, but MSOs have another tool that can be used – the microcell. MSOs also have the ability to use the unlicensed frequency bands and the 802.11 Wi-Fi standard. Taking the same technical approach to a Wi-Fi based system and assessing the percentage of each cell that operates at the various modulation rates we get an average capacity of 8.3 Mbps in a 20MHz band when considering just the current 802.11b/g standards.

Modulation	cell edge	BPSK	QPSK	16QAM	64QAM
Base Bits per Hz		1	2	4	6
raw rate		6.89	13.78	27.57	41.35
Code rate		0.50	0.75	0.75	0.75
Derated bits per Hz		0.50	1.50	3.00	4.5
Bit Rate (Mbps)		3.446	10.338	20.677	31.015
Thoughput		2.76	8.27	16.54	24.81
SNR required (dB)		0	8.5	15	21
Uplink Sensitivity (dBm)		-99.97	-91.47	-84.97	-78.97
Downlink Sensitivity (dBm)		-95.97	-87.47	-80.97	-74.97
BTS TX pwr		27	27	27	27
BTS ant gain		4	4	4	4
MS ant gain		-2	-2	-2	-2
MS Tx pwr		20	20	20	20
Uplink link budget		134.0	125.5	119.0	113.0
Downlink link budget		125.0	116.5	110.0	104.0
Fade margin		9	9	6	6
overall DL link budget		116.0	107.5	104.0	98.0

Distance (km)		0.30	0.47	0.23	0.17	0.10
Distance (mile)		0.19	0.29	0.14	0.11	0.06
Cell Area (km^2)	0.28	0.28	0.69	0.17	0.09	0.03
Ring Area (km^2)		0.12	0.53	0.08	0.06	0.03
% of whole cell		40.74%	76.05%	10.86%	8.56%	4.53%
% with cell edge limit		41.22%		26.67%	21.00%	11.11%
Blended Data Rate [whole cell] (Mbps)	8.30					

Figure 7 Example Wi-Fi link budget (microcell)

This model uses a microcell power loss that was derived from empirical measurements of basestation deployed at strand height using Wi-Fi at 2.4GHz. The best fit to the data was a propagation power law of 2.8 resulting in the loss profile shown in Figure 8.

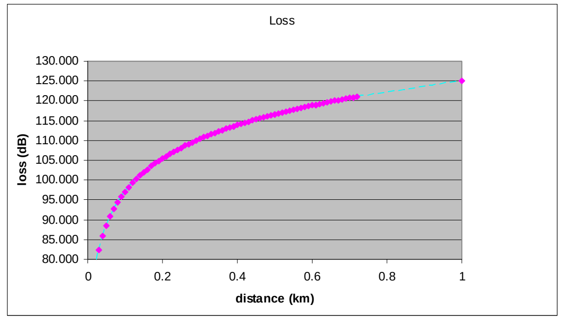


Figure 8 Example Microcell Loss

Using this number of 8.3Mbps average per Wi-Fi microcell, and using 30 microcells per square mile, we arrive at a capacity of 249Mbps per square mile, which is 5 times that capacity of the LTE macrocell. Using 802.11n, the evolution of 802.11g, capacities of in excess of 500Mbps per square mile can be achieved.

Comparison of 3G/4G/Wi-Fi systems

Figure 9 shows a comparison of the throughputs expected from various standards in the real world.

	Channel BW/Duplex	Peak user data rates (Mbps)	Typical user peak data rates (Mbps)	Wide Scale Availability (carrier)
802.11g	20MHz/TDD	22Mbps in 20MHz	4-12Mbps	Now
802.11n	20MHz/TDD	45Mbps in 20MHz 90Mbps in 20 MHz (2x2 mimo) 240Mbps indoor	10-25Mbps	2H 2009
HSDPA	5+5MHz FDD	14.4Mbps in 5MHz	1-3.6Mbps	Now
802.16e	5-10MHz TDD	27Mbps in 10MHz 12Mbps in 5MHz	expect 4-8Mbps	Early 2009
LTE	1.25MHz-20MHz	100Mbps in 20MHz DL 50Mbps in 20MHz UL	expect 6-12.5Mbps	2010+

Figure 9 Comparison of systems

A recent study by Ken Biba of Novarum [3] confirms the real world effectiveness of Wi-Fi based systems for delivering a good user experience in comparison to the 2G and 3G cellular standards that are deployed. In the study the Wi-Fi networks delivered 3 times the throughput of the WiMAX and Cellular networks with the same coverage.

A study by Nortel Networks [4] also drew similar conclusions to the above calculations. That is that the expected performance of HSPA was 3.4Mbps, LTE was 11Mbps, both on a 500m cell.

Overall this demonstrates that wireless capacities and user experiences are generally lower than that experienced by home users, but that Wi-Fi based microcells offer MSOs an important tool that can deliver superior performance.

STRAND MOUNTED WIRLESS BASE STATIONS

As mentioned earlier, MSOs have a unique asset that enables them to play a large role in the 4G wireless world. This asset, the HFC, offers the three elements necessary for deploying microcells.

- Power – Plant power, quasi square wave low voltage ac
- Backhaul – DOCSIS or fiber
- Mounting – the messenger wire

A typical strand mounted wireless BTS is shown in Figure 10.

Figure 10 Deployed Strand Mount Basestation

The basestation uses standard strand hangers for mounting and has a single connection to the HFC for both DOCSIS® and power. Various antenna options are required depending on the RF coverage that is being deployed. In order to ensure that omni directional coverage is provided, a bracket is used to offset the omni directional antennas from the basestation to avoid shadowing. Directional antennas are also possible for coverage of tall buildings or for making backhaul connections. The DOCSIS®/Power connection is made via a dc splitter and/or a power passing tap from the main coax.

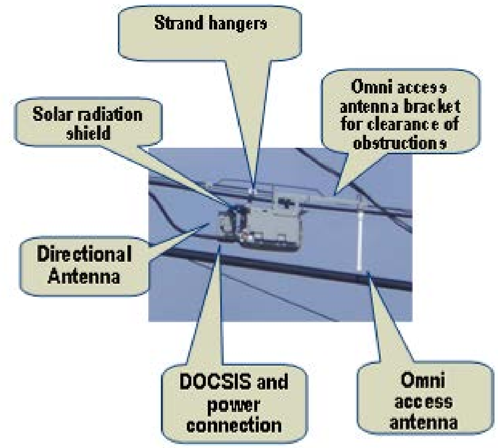


Figure 11 Strand Mount Basestation Details

The architecture of such a strand mounted basestation is shown in Figure 12. The combined DOCSIS® and power enter the unit via power protection module that splits the power and the DOCSIS® apart. This module provides filtering to prevent emissions from entering the plant and lightning and power surge protection. A power supply takes the plant power quasi square wave and converts it to usable dc voltages for the rest of the BTS.

The BTS has 2 universal radio slots, which in the example shown contain a LTE (3GPP Long Term Evolution) radio paired with an 802.11 Wi-Fi radio.

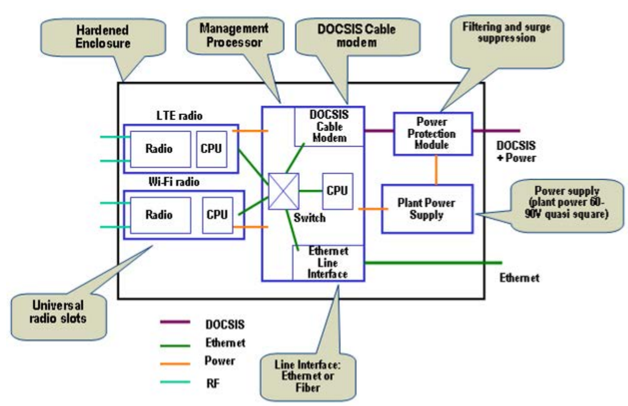


Figure 12 Inside a Strand Mount Basestation

Central to the Basestation is a control board that contains Software that enables integration into the HFC. This processor provides the packet processing to apply QoS as well as the mobility features described in the next sections.

MOBILTY IN THE HFC

One key difference between the delivery of wireless service and fixed services is the aspect of mobility. Today's HFC is designed to provide fixed subscribers with a high speed connection, and contains the assumption that the user device does not move. In most implementations there is an IP address limit per connection of 2 to 5 addresses and a MAC address limit of 8-16 devices. The network assumes that a user has a NAT (Network Address Translation) router device deployed on premise that hands out addresses to local computers. In a wireless system deployed on the HFC, the mobile devices can move between basestations at will. In fact, even a static/fixed wireless device will be mobile as variability in RF propagation and the environment means that a so called fixed device will connect to several basestations over time.

One of the main requirements is to deliver this mobility without requiring wholesale changes to the existing HFC components such as the CMTS and Fiber nodes. In order to deliver such mobility on the HFC a tunneled architecture is proposed as shown in Figure 13. The tunnels are IP constructs that can traverse a layer 2 and or layer 3 network, including a NAT device if necessary.

In order for mobility to be a seamless experience for the customer, there are some critical requirements.

- IP address of the end device shall not change.
- Customer shall not be required to login multiple times.
- Specialized Software shall not be required on the user's device.

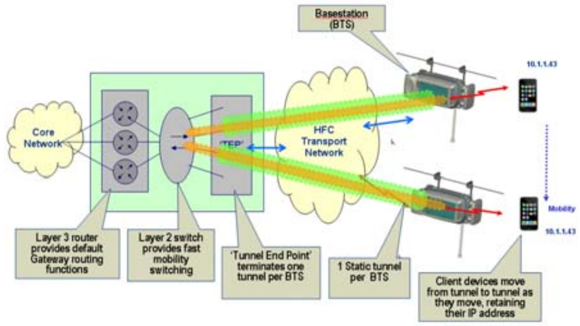


Figure 13 Tunneling approach for mobility

This level of seamless mobility can be scaled to systems in the order of 10,000 contiguous basestations. We also need to address larger scale systems and mobility between standards. To do this, a mobile IP infrastructure can be deployed as shown in Figure 14.

One might ask why not deploy a full mobile IP infrastructure to handle the mobility day one. The reason for this comes back to one of the challenges of microcells which is that there are lots of cells. This translates into a lot of mobility events in the system as the cell boundaries are frequent and users transition though cells quickly, generating a lot of mobility events. With a hybrid tunneled and mobile IP architecture, the transitions between microcells are handled at layer 2 whereas the transitions between groups of microcells and between different systems are handled at layer3.

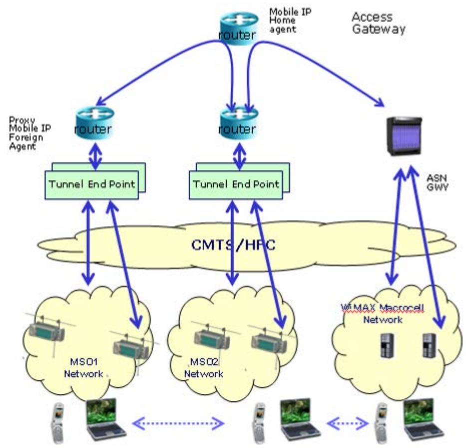


Figure 14 Mobile IP between networks

A more detailed view of the entire network can be seen in Figure 15. This shows another fundamental difference between a mobile network and a fixed network. That is the concept of centralized subscriber management.

In the classic HFC case, the modems represent the customer and the CMTS and Network can be provisioned to authorize a particular fixed device, whether that is a set top box, PVR or Cable modem. In the case of the mobile device, the device can exist anywhere on the network, and the user is attached to the device rather than a fixed location such as a home. This necessitates a centralized approach to login. This can be transparent to a user such as a SIM (Subscriber Identity Module) based authentication or MAC address based, or a web based user login.

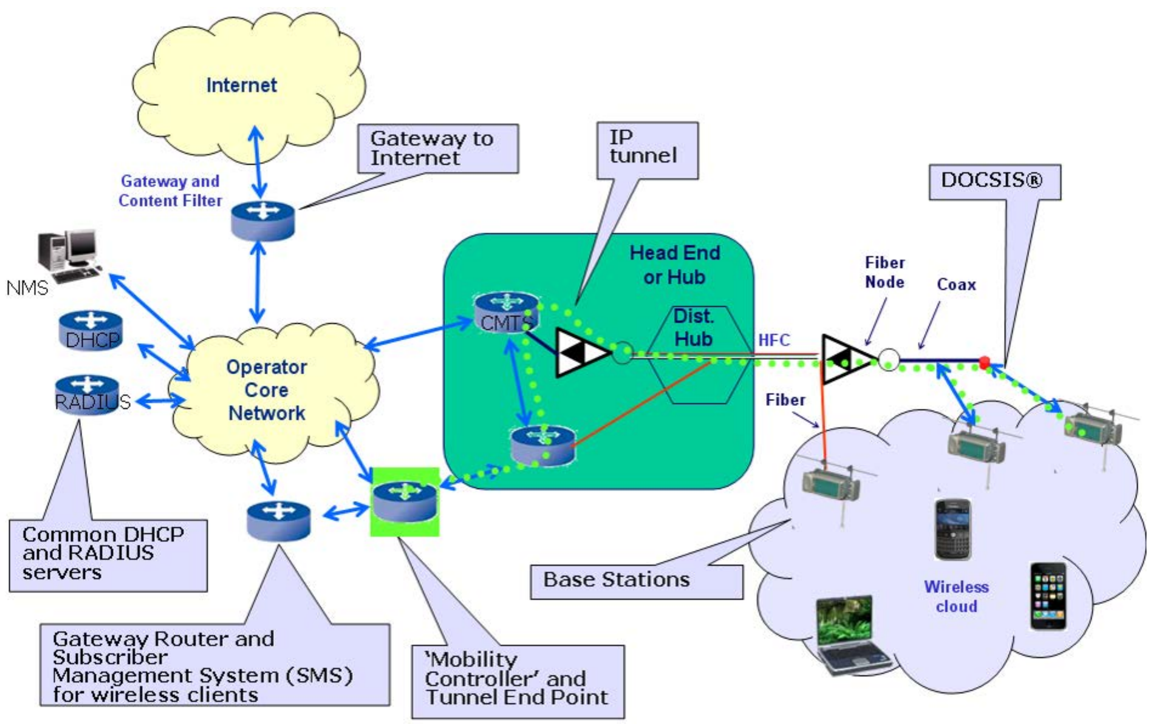


Figure 15 HFC with tunneled mobility

INDOOR WIRLESS

Indoor wireless is another part of the story that needs to be addressed. In this case, cellular technologies have struggled to deliver a consistent user experience and the experience of dropped calls within the home has been an experience that has frustrated many. In fact some cellular providers have deployed Wi-Fi already as a solution for in the home wireless using the UMA [5] technology. Others are exploring the use of Femto-cells which are miniature basestations for delivery of cellular service. This trend fully supports the argument for strand based microcells in the outdoor world, as Femtocells are becoming necessary to deliver even 3G data speeds to end users.

Fortunately the MSO/HFC friendly architectures described above can also be used to deliver wireless services into the home or small business. With the hybrid tunneled architecture seamless mobility can be provided between outdoor and indoor locations.

WHAT HAPPENS WHEN THE PLANT IS UNDERGROUND?

Of course, as ubiquitous as HFC is, not all of it is overhead strand. A reasonable percentage is underground, and is accessed from vaults and pedestals. Figure 16 shows a typical residential area pedestal with a HFC fed base station embedded.

Figure 16 Pedestal mounted BTS

SUMMARY

This paper has demonstrated that the HFC is an ideal location for hosting wireless basestations. Broadband wireless systems need microcells in order to deliver the broadband experience they promise, and microcells have traditionally been limited to locations where backhaul and power are accessible. The HFC can provide the 3 necessary ingredients (location, power and backhaul) for a microcell deployment.

Full mobility can be provided over the HFC using a tunneled architecture together with a mobile IP overlay offering a user a seamless mobile solution, while not requiring an upgrade to the HFC.

If MSOs wish to deliver the superior user experience that their customers are used to from their cable modems in the wireless world then they should consider the HFC as a key asset. Today's Wi-Fi systems offer capacities and user throughputs that are far superior to today's 3G cellular systems and to next generation 4G macrocells.

Interestingly, the 4G systems have migrated to an all IP transport system, something that Wi-Fi was designed to do from the start. This is another advantage that MSOs have, with their IP friendly networks. A Wi-Fi system deployed on the HFC can be used to deliver broadband IP wireless services to customers and as a bridge to a 4G microcellular system in an extremely cost effective manner. While not covered here, such HFC based wireless networks can also deliver a full multi-media experience, a topic for a future paper.

By using the HFC to deploy Wi-Fi today, an MSO can deploy a wireless experience that will be unmatched until 4G licensed band microcells become ubiquitous.

REFERENCES

[1] 3GPP Long Term Evolution
http://www.3gpp.org/article/lte

[2] Cost 231 Walfisch-Ikegami Model

http://www.ee.bilkent.edu.tr/~microwave/programs/wireless/prop/costWI.htm

[3] Ken Biba, Novarum
http://www.muniwireless.com/2009/02/03/muni-wifi-outperforms-cellular-and-wimax/

[4] The Business Case for LTE, Gerry Collins, Nortel
http://www.umts-forum.org/component/option,com_docman/task,doc_download/gid,1758/Itemid,12/

[5] Un-licensed Mobile Access

http://www.umatechnology.org/overview/

THE END

INCREASING MSO ADVERTISING REVENUES THROUGH MANAGEMENT OF AD SKIPPING

Dan Holden
Comcast Media Center

Abstract

New capabilities that allow customers to enhance their viewing experience, such as the ability to skip advertisements, can be disruptive to traditional business models. However, these new functions also provide an occasion for the industry to seize market opportunities that are created by shifts in consumer behavior. Rather than seeking to defeat the customer's desire to fast-forward through content that doesn't respond to his or her personal interests, it is more effective to acknowledge that desire and present alternative content. Proactively responding to the consumers actions changes the traditional spot advertising approach and benefits consumers and advertisers alike.

This paper outlines a revolutionary approach for using new technology to present advertising to consumers. The innovative solution leverages the action of ad skipping as a catalyst for allowing advertisers, cable programming networks, and cable system operators to generate additional revenue that is currently lost as a result of ad skipping.

INTRODUCTION

Many consumers actively choose to skip ads. They purchase or lease technology that helps them with this endeavor. Advancements in technology provide cable system operators with the ability to support the customer's preference to skip through ads, while presenting an alternate paid campaign that runs in the foreground while viewers are skipping the original ads or content in the background.

Industry trends and consumer sensitivities drive advertisers to be creative when placing ads. For example, digital signage has gained great penetration in retail outlets and movie theaters thus presenting consumers with campaign material regardless of their location or activity. In the same way, consumers viewing habits can be leveraged to present alternative and creative means by which linear advertising content can be presented. Specifically, time-shifted non-linear programming can be enhanced to add value to the traditional thirty-second spot. For example, it is possible to place advertising "on top" of media, much like animated bugs. This and other techniques, which we will collectively call "trick-file advertising", offer great revenue opportunities by replacing or enhancing advertisements while consumers engage trick modes such as fast forward, pause or rewind.

The main goal of trick-file advertising is to leverage consumer behavior for the benefit of consumers and advertisers alike. As consumers opt to skip advertising, they are becoming conditioned to reach for the remote at the beginning of every break. This behavior demonstrates that:

- Consumers are actively interacting with the television and the displayed programming
- They have the remote in their hand
- They are not interested in the commercial content being presented

This is truly interaction with the television platform. During the instances where viewers choose to fast forward through content, it is possible to insert alternative content in the foreground while the original content appears to be fast forwarding in the background. At this point in time, there are two clear objectives: to get more impactful content in front of the consumer and to log the event. If the alternative content is unique and compelling, it will be more valuable and relevant to the consumer. As an example, in 2005 Burger King created a viral buzz and therefore drawing increased viewership to their linear commercials as viewers flocked to see alternative content hidden in the 30-second spot.

AD SKIPPING

DVR penetration

Customers skip ads for various reasons including burnout and lack of interest in the advertised product. The potential for consumer ad skipping is increasing on a yearly basis. According to data collected in 2008 on DVR penetration, twenty-nine percent of U.S. television households are equipped with DVRs, placing the potential for ad-skipping in front of approximately 33 million households (Nielsen data)[i].

Moreover, nearly two thirds of U.S. households with $100,000-plus incomes subscribe to time shifting services or own a time-shifting device, which translates to at least one DVR in at least 15 million affluent households (Ipsos Mendelson[ii])

In the first four weeks of the Fall 2008 TV season research from IPG's Magna, indicated DVR time shifting accounted for 11 percent of all prime-time ratings on the five broadcast networks. They found DVR playbacks now equal 16 percent of all prime time viewing by consumers age 18-49. That's more than double the impact VCRs had, which was at the 90-percent penetration level. In DVR-only homes, nearly 40 percent of household prime-time viewing (and 50 percent of prime-time viewing in the 18-49 age range) is now time shifted content.[iii]

The statistics indicate that use of time-shifting equipment and functions to skip ads will continue to grow. DVR penetration is projected to increase from its current 23 percent to 37 percent of all TV households, by 2012. In addition, 25 percent of primetime content will be time-shifted by the 18-49 demographic (Magna[iv]).

Least-skipped ads

However, viewers with the capability to skip ads don't skip through all ads indiscriminately. For example, TiVo commissioned a study (Jan 2008) on advertising and determined not all advertising is skipped. The least-skipped ads include theatrical film releases, which accounted for six of the ten most-viewed commercials in DVR playback mode. In addition, eight of the ten most-watched time-shifted cable spots occurred while viewers were watching the show "Psych" on USA Network. TiVo attributed this statistic to integration of characters from the TV show into the advertising, making it difficult for viewers to distinguish between commercial and program content.

Earlier research by TiVo (covering May 2008) supported the idea that viewers skip fewer ads that they see as relevant to their circumstances and interests. TiVo's PowerWatch study specifically evaluated spots skipped by demographic. The data indicated that viewing for children's skincare products in homes with children under 12 was 37 percent greater than in homes with adults 50-plus. Advertising for toys and games had 22 percent more viewing in homes with children under 12. Contrast this to political ads, which had 15 percent more viewing in the homes with adults 50-plus. Finally, ads for hair restoration products and wigs had 10 percent more viewing in those same homes.[v]

Additionally, Fox TV has found through testing that viewer attention levels to commercials are higher for shows with shortened commercial pods. Biometric engagement research company Innerscope found that viewers watching shorter commercial pods had ad attention levels 31 percent higher; with ad engagement levels 21 percent higher. Unaided recall levels were recorded 250 percent higher and ad likeability levels 61 percent higher than viewers who watched the standard-length ad pods.[vi]

Yet another study from Innerscope used eye tracking to measure gaze location and duration and smart vests that measure arousal, respiration, heart and motion responses. During their testing, prescreened DVR users were given remote control devices that allowed them to fast-forward, pause and play at their discretion. The viewers whose programming had shorter commercial pods, and 50 percent less commercials in the program, fast-forwarded 136 percent less than viewers who watched the standard-length commercial pods. [vii]

It is possible to reduce ad skipping by disabling the fast-forward function. However, in light of the data presented this type of solution is likely to have a negative impact. An alternative approach would be to leverage the ad skipping behavior of subscribers for the benefit of advertisers and consumers. This approach can benefit customers by presenting them with an abbreviated, impactful message that would appear during the same time interval as the skipped ads.

Trick-file advertising also benefits advertisers. During fast-forward, the advertiser can be certain the viewer is engaged with the television. Captured viewers interested in the linear scheduled program but not interested in the inserted commercial is a very specific demographic. Identifying and presenting alternative content – related to the inserted 30-second spot or unrelated (alternative campaign) adds value to the inventory.

TRICK-MODES AND TRICK-FILES

Trick modes or trick play

VCR-like functions such as pause, rewind, fast-forward, replay and skip, are collectively known as ‘trick modes’ or ‘trick play.’ Trick modes are currently available on DVRs, VOD, PCs and other CE devices. A trick mode is essentially a command that controls the playback of video. Specifically, when consumers press the fast forward button, this interaction indicates their choice to skip that material. This action also represents the opportunity for inserting an alternative message rather than the traditional mapping to fast forward. Caution must be exercised when implementing this new type of functionality, as the consumer could become easily confused. The proposed solution resolves this issue by inserting alternative content that only covers a portion of the real estate. Additionally, clear delineation between ads can be created by a spatialecial separation (lines or frames between the new message)which appears in the foreground and the original content, which appears as background The ability to still see the original content in fast forward mode will allow the consumer to see that an advertising trick mode has been activated

Trick-files

A trick-file is the video a consumer experiences when they activate a trick mode, e.g. fast forward. When trick-file video is streamed, it gives the illusion of a VCR running in faster than normal mode. One method for the creation of a trick-file is to extract all I-Frames then create a new file comprised of fewer frames than the original video. For instance, a thirty second spot encoded to VOD specifications would contain exactly 900 frames of video. In order to play the spot in five seconds 750 frames of video would be removed in order to reach a total frame count of 150.

There are three general methods deployed today for VOD trick file displays: files, indexes, and dynamic video generation. The file method creates a new trick file (video file) that is played when a trick mode is activated. Indexing is a similar approach. However, rather than creating a new video file, indexes are assigned to frames of the original video. These indexes are utilized by the video pump to select the correct frames to display in order to achieve a fast forward effect. The dynamic video generation approach uses tools to calculate how many frames to skip in real-time as the video is played.

Each method outlined above may be used to create a trick file advertisement. While each method described has its own complexities associated with the creation of the trick file ad, the true complexity surrounds the successful mapping and tracking of these enhanced ads. To meet the need of the advertising community for detailed data, it is key that methods be developed to track original advertising as well as the replacement ad, while exposing the consumer to replacement spot.

Trick-file Advertising

Trick-file advertising is essentially the blending of two distinct concepts: ad-skipping and paid campaigns. Today, when customers press the fast forward button, they are indicating they wish to jump through the current programming. At this instant, it is possible to insert into the viewer's screen alternative content that is different from the advertisement or other content that had triggered the ad-skipping response. While the alternative message must be presented in a unique and compelling way, it can be from the same advertiser or content source For example a traditional 30-second car ad could be replaced by a shorter message from the local dealer containing a special offer.

Since customers will choose to view content that matches their interests, this new approach represents a win / win for the advertiser and consumer. User behavior is the key. Once the customer has the remote in his or her hand, the content presents an opportunity to turn the original commercial pod into an interactive experience. This principle also applies to other content that viewers may seek to skip. For example, a viewer that fast forwards through the first several moments of a TV series that contains clips from previous episodes could be exposed to a short promotional segment about one of the leading characters or a product endorsement.

As previously noted, trick-file advertising is consistent with the direction that consumers and advertisers are driving technology. Advertisers are open with their plans to move advertising dollars to platforms that offer greater levels of audience engagement. Meanwhile, studies such as the ones cited earlier in this paper demonstrate that viewers equipped with ad-skipping technology will choose to watch ads that interest them.

The new paradigm represented by trick-file advertising is comprised of a process that will present, when prompted by the consumer, new advertising in the foreground while the original advertising that is running in a trick mode, such as fast forward, continues to appear in the background. The customer views an alternative message during the same amount of time that they are fast-forwarding through the original content. For example, a thirty-second advertisement may be replaced with a shorter, five-second message. This type of advertising is restricted to time-shifted content (VOD, DVR, StartOver, etc), since it is not possible to fast-forward through real time content, such as a live sports cast.

In addition, trick-file advertising represents the opportunity to create new forms of personalized content. For example, another way to expand further the potential of trick file advertising would involve providing unique messages for multiple buttons on the remote, by using trick modes with each button that are mapped to a different message. This capability will allow selling more avails per placement opportunity.

To radically change the fast forward experience would not be advisable, as it is sure to confuse the subscriber. A better approach is to place the new advertising video in the foreground, and distinguishing it from the original content, which appears in the background through the use of a spatial separator. This “picture-in-picture” affect will allow the user to track the timeline of the original programming, while being presented with the alternative content ((figure 1).

Figure 1

CREATING TRICK-FILE ADVERTISING

The implementation of technology to support trick-files builds upon other advanced video developments, including VOD and the tru2way™ platform. Three approaches discussed in this paper include are file based, VOD based, and DVR based creation

File based

The first step for developing file-based trick files is to create a file, which will move the current advertising into the background and implements the fast-forward effect. For example, a normal two minute break has 3600 frames of video. To run the break in the 20 seconds that simulate a fast-forward mode, the original 3600 frames of advertising would be cut to 600 frames of video. This decrease in frames is suitable since the background video will only be utilized to track original programming. This process involves creating a file that uses only a fraction of the frames contained in the original content. Once extracted from the original ads, the 600 frames are re-constituted as a new file that creates the illusion of a background content running in fast forward mode.

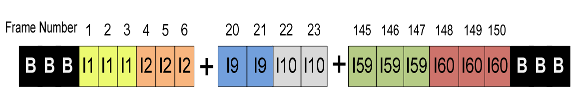
Next, the replacement content is placed in the foreground. Each and every frame of this alternative content will be squeezed and appear inside a spatial separation via a black border, so that it may be easily viewed as different content from the background video.

Once these steps are completed, the file must be indexed, tied to specific content and prepared for distribution to a playback device (VOD, DVR etc) When the new trick-file enhanced advertisement is experienced the customer views the four thirty-second spots running in the background while one twenty-second spot is running in the foreground at normal speed. This approach creates an alternative experience during trick mode but in no way modifies the experience during normal play All of the advertising being displayed is tracked via VOD server logs and can provide metrics down to the unique subscriber level.

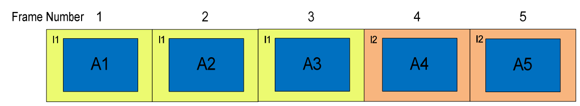
A more detailed look at the process to create trick-file advertisements begins with the following assumptions:

- One thirty second spot
- Video is encoded to VOD specification (MPEG-2 transport stream, long GOP set to fifteen, thirty frames per second)
- CMC standards for advertising (900 frames of video with bookend twelve frames of black)

Based on the specifications above, there are approx 60 I-frames of content after removing the black that appear before and after ads, resulting in approx. 60 frames of background video. In this example, the I-frames are labeled “I1” to “I60.” As all of the black frames are equivalent, they are collectively referred to as “B”. Creating a five second trick-file ad will require a total of 150 frames of video (5 seconds * 30 frames/second). A simple algorithm for reconstruction would be 3(I1) + 3(I2) + 2(I3) + … + 2(In), where n=location of the frame in the original asset. This would result in a trick-file of approximately 120 frames, thirty frames short of the 150-frame objective. Next, additional I-Frames are added at the beginning and end of the file to achieve the 150 frame objective, and to bookend the file with frames of black in order to provide smooth video transition. This process results in having the original 912 frames of video reduced to 156 frames of video that will run in five seconds. This new file will then be used to provide the background video of the original spot running in fast forward mode.



The next step in the process is to overlay the new advertising on the newly created trick-file. This process involves encoding the replacement spot to exactly 150 I-Frames. For example, these frames may be labeled “A1” through “A150”. Each of these I-Frames will have dimensions slightly smaller than the I-Frames from the trick-file background video. The following diagram is representative of the first five frames in the newly assembled trick-file advertisement.



Once the video has been correctly assembled, the next step is to add the audio track. Then, the newly created elementary stream is wrapped with an MPEG-2 transport stream component. Now it is ready for distribution to VOD pumps, DVRs, or other CE devices. This algorithm can be adapted to support a wide range of durations.

VOD based

There are two options for creating trick files as part of the VOD process. First, with a few enhancements to the video pump, it should be possible to create the picture-in-picture video effect dynamically. This would allow late binding of the advertising replacement spot. As the pump plays trick-file video, the replacement advertisement and appropriate spatial separation would be spliced into the reserved section of the trick-play video.

Alternately, it is possible to orchestrate the VOD pump, which in essence would make a copy of the trick file and provide the appropriate enhancements to it. The output of at “trick file generation device” could be attached to a feed that’s capable of splicing portions of the video.

DVR based

DVR implementation of trick-file advertising is essentially the same as VOD. The biggest obstacle to overcome is getting the trick-files on the DVR and associating them with content correctly. Within the tru2way specification, the ability to push content to and from web servers has been defined. These features have yet to be implemented within the tru2way 1.0 specification, but should be available in the future. Work is currently underway at CableLabs to enhance the metadata specifications thus allowing advertising triggers to be inserted into video and read by DVRs. Advertising engines on DVRs will then be able to dynamically insert advertising on top of trick-file playback, provided the DVR is either equipped with multiple rendering devices or splicing capabilities.

COMPARISON TO OTHER ADVANCED ADVERTISING TECHNOLOGIES

Personally relevant advertising is a key concept in the advanced advertising tool kit, and revolves around matching a campaign's message to consumer interests. Additionally, the collection of metrics to measure campaign effectiveness is sought. Trick-file advertising has the ability to achieve both of these objectives without massive upgrades to the cable plant infrastructure.

The simplest types of trick-file advertising can be implemented within the VOD infrastructure. It will not consume additional bandwidth like telescoping or linear addressable to the set top box, but will require additional VOD functionality and resources. Unlike eTV, the simpler forms of trick file advertising do not require a two-way plant. Advertising content can be pre-produced and checked for quality then staged on the VOD platform. For some VOD systems, a few changes to the pre-encryption distribution process should suffice for content delivery. In the base case, the VOD server would simply play the trick-files as it does today. VOD server logs may be used for collecting advertising metrics in a manner that upholds strict privacy guidelines.

When compared with other advanced advertising activities, trick file advertising can offer compelling opportunities for revenue generation using technology that is largely available today. These types of advanced advertising opportunities include:

Telescoping – The process of launching long form advertising from triggers that are embedded in linear or non-linear content.

Addressable – The process of placing unique advertising to match viewer interests.

Request for Information (RFI) – Is the process of gathering customer interest to create mailing lists of print, or other forms, of advertising.

Trick-File – Is the placement of adverting "on top" of standard trick play (fast forward) video.

	eTV	Additional Bandwidth	Behavior Modification
Telescoping	Yes	Yes	Yes
Addressable	No	Yes	No
RFI	Yes	No	Yes
Trick-file	No	No	No

In addition, the capabilities of trick file advertising can incorporate a number of advanced features in a two-way operating environment. For example, trick modes may be captured as state changes on the VOD platform and then mapped to specific functionality. This type of state machine will allow for the creation of a contest that is embedded within the content. One type of contest could be an Easter egg hunt, where video widgets may be placed across the span of content and accumulated into a basket, much like a shopping cart on the web. Each of these interactions would be captured and persisted. Once the consumer has interacted with all of the specified widgets they would be notified of contest results.

Other contests could span across multiple pieces of content. For example, a Monopoly-like game could be created where property widgets are hidden and embedded in multiple pieces of content for collection by consumers.

A wide variety of contests are possible. Some may be complex and last months, or they may involve simple techniques, like pressing fast forward twice followed by reverse during an advertisement in order to win a prize. The goal is to provide capabilities to both content providers and advertisers for innovative campaigns to connect with consumers. A flexible advertising architecture is sure to make advertising much more compelling and of interest to a wide segment of cable customers.

Multiple Avails

Trick-file advertising will allow for more avails per break. It is in effect a second opportunity to reach the consumer. As previously stated, different buttons on a remote can be used for creating multiple opportunities by associating different messages with each of the modes, such as Fast-forward, 2X Fast-forward, Rewind, 2X Rewind are just a few. Added value may also be realized due to the nature of the advertising, the additional opportunities and how the content that is constructed creates the option for different kinds of impressions.

CONCLUSION

Trick-file advertising offers the industry an innovative approach to harness ad skipping. It is activated at the request of the consumer, and sits squarely between ad-skipping and disabling fast forwarding. Since much of the required technology is already deployed, capital expenditure for implementation should be minimal. When deployed trick file advertising offers a dynamic way to reach specific demographic groups, and can collect advertising metrics while upholding stringent privacy guidelines.

The next step for this exciting technology is to get creative. Enlisting the talent of ad agencies will help align the new technology to consumer expectations and provide the best possible user experience. Organizations like Comcast Spotlight and Canoe may be called upon to determine how to sell the new placement opportunities and additional avails, as well as develop strategies for maximizing potential revenue.

---

[i] *"Report: More Than Half of $100K Households Time Shift,"* By Anthony Crupi, Mediaweek, January 8, 2009

[ii] Ibid

[iii] *"Study: DVR Ratings Impact Rises,"* Steve McClellan, Adweek, Nov. 6, 2008

[iv] *"MAGNA Global updates On-Demand forecasts,"* Radio Business Report, September 10, 2008

[v] *"King of Beers Defies TiVo Timeshifting,"* Television Broadcast magazine, January 2008

[vi] *"Fox Goes Sci-Fi to Measure 'Fringe' Ad Effectiveness,"* By Jon Lafayette, TVWeek, September 7, 2008

[vii] Ibid

NEW MEGABITS, SAME MEGAHERTZ: PLANT EVOLUTION DIVIDENDS

Robert L. Howald, Ph.D., Amarildo Vieira, Ph.D., Michael Aviles
Motorola Home & Networks Mobility

Abstract

The use of 256-QAM in the downstream path has nearly completely replaced 64-QAM as the modulation of choice for MSO's for two simple reasons:

> *(1) 256-QAM is more bandwidth efficient, providing a 33% increase in spectrum efficiency compared to 64-QAM*
> *(2) 256-QAM has been proven to work reliably.*

A natural question to ask, given the relatively smooth transition from 64-QAM to 256-QAM, is whether there is a convenient next step in terms of improved bandwidth efficiency. The industry has yet to make a significant move towards 1024-QAM. Prior papers, including by this author, have pointed out some of the potential hurdles. However, as the HFC network has evolved in support of new service demands, and the downstream multiplex has done the same, variables that affect the ability to reliably implement 1024-QAM are beginning to work in favor of this more bandwidth efficient approach.

It is now time to understand and quantify the practical performance and the potential limitations in order that it may be deployed properly. This paper will examine the required specification of impairments for successful transmission of 1024-QAM. The discussion will summarize the effect of HFC-specific impairments on 1024-QAM and compare them to 256-QAM and 64-QAM. Finally, we will present some conclusions, along with supporting measured data, on the proper architecture limitations and system thresholds to ensure high-performance delivery of 1024-QAM. These guidelines can be used to enable MSOs to reliably extract another 25% more bandwidth from their digital multiplex.

INTRODUCTION

Appetite for more bandwidth has continued to increase, and there is every expectation that it will continue. In particular, as MSOs migrate from broadcast-like models with a VOD component, to delivery models trending towards all-switched unicast services, the Mbps required per service group climbs dramatically. Existing tools (analog reclamation, MPEG-4, SDV, etc.) can and will continue to make the bandwidth explosion manageable. Nonetheless, continued bandwidth growth demands expansion of the HFC toolkit. The requirements are driven by a combination of high-consumption trends – personalized streams, the content itself evolving to HD and beyond, the growth in multiple and non-traditional consumption venues and devices, and the desire to continue to increase data tiers for high-speed Internet service.

The cable plant has kept up with the bandwidth consumption by adding RF bandwidth and using efficient digital modulations to mine the capacity effectively, and with robustness. What started as 64-QAM digital signals became yet more bandwidth efficient with the deployment of 256-QAM downstream, which is the dominant

QAM approach today. The ability to successfully deploy such schemes is due to the fact that the downstream channel in a cable plant has very high SNR, and a very low distortion. This is because it was designed to ensure proper conditions for supporting much less robust analog video, which had historically dominated the downstream payload. In addition to high linearity and low noise, the downstream channel has a flat frequency response on a per-channel basis, minimizing both amplitude and phase distortion, although it can be prone to reflection energy.

What has powered digital modulations historically is the ability to deliver robust link performance and services over varying degrees of link quality by boiling the required receiver function down to choosing between 1's and 0's, instead of replicating and infinitely-valued analog waveform with high fidelity. And, when channel conditions are of high quality – such as the downstream cable plant – the situation is ripe for exploiting bandwidth using very efficient digital techniques.

As a simple example of the possibilities, the theoretical capacity of a 6 MHz channel with a 40 dB SNR is approximately 80 Mbps. Yet, for 256-QAM, the transmission rate is only about 40 Mbps. When accounting for overhead, there is even less throughput. The next higher order, square-constellation, modulation is 1024-QAM. This technique achieves an efficiency of 10 bits/symbol, or another 25% efficiency over 256-QAM, and an impressive 67% improvement relative to 64-QAM. Practically speaking, three MPEG-2 HD's per QAM fit comfortably, as compared to the threat of visual artifacts of jamming a third HD channel into a 256-QAM payload. Alternatively, it would represent at least two more SD streams per QAM. By also considering statistical multiplexing efficiencies and implementing wider channels, this could possibly be increased to four HDs per QAM [2].

Of course, there is no free lunch when it comes to modulation efficiency. To support 1024-QAM, a more stringent set of specifications must be met. The goal here is to identify how pristine the plant must be, or must become, in order to migrate to this modulation profile effectively in terms of the likely candidates for disruptions to robust transmission: SNR, Beat distortion interference, and to a lesser extent, phase noise. It is worthwhile to point out that an intermediate step using 512-QAM, and its additional 12.5% efficiency, has merit in moving up the modulation complexity chain. All of the tools developed here apply to 512-QAM as well, with of course a different set of numbers drawn from them. For this analysis, however, we will focus our attention on 1024-QAM, as these modems exist for the cable space, and the downstream channel capacity favors the likelihood of adding increasingly more bandwidth efficiency to the plant.

PRIOR ANALYSIS, ESTIMATES, AND TESTING

An early analysis of 1024-QAM was presented at the 2002 SCTE Cable-Tec Expo [1]. While early modem ICs existed that could support this modulation mode [5], little attention had been given to it in the cable world by operators, and subsequently little was understood about how well it would perform in a typical plant. The paper used communication theory and what was learned during the implementation phases of 64-QAM and 256-QAM to draw conclusions about the expectations for 1024-QAM, which was thought to be coming around the corner. However, because of the potential complexities of taking on this advanced scheme and because of other more pressing priorities, this modulation mode has not been exercised, nor has there been any significant trial activity.

The conclusions of the 2002 paper covered a range of items, but ultimately focused on the major items likely to be challenging to 1024-QAM under typical HFC performance delivered: SNR, analog beat distortions, and phase noise. In each case, the conclusion was that current RF characteristics represented performance that would lead to minimal margin. And, for "below average" performing HFC plants, the result could be operating in a bit-error prone region that would add to the FEC's work of delivering error-free data. The concerns regarding these three parameters are quickly summarized below.

SNR

The difference in SNR requirements between 64-QAM and 256-QAM is 6 dB, and the difference between 256-QAM and 1024-QAM is an additional 6 dB. This is nearly exact in the non-FEC case, and is close also in the case of added coding (of the same type), although post-FEC error rate curves are much steeper. So, while 64-QAM, which delivers a 1E-8 BER with no error correction at a 28 dB SNR, zips along comfortably on a cable plant, each modulation order increase gets more difficult. At 28 dB, this allows the QAM load to be implemented with up to 10 dB of signal power back-off relative to the analog carriers and yet still support a large link margin. For example, 46 dB delivered for analog, means 36 dB delivered for digital, resulting in more than 8 dB to spare without even including FEC. This amount of back-off is important, because it allows the digital load to become much less consequential to the total power load, adding only about 1 dB to the total RF load on an 870 MHz system, assuming a 12-14 dB tilt.

Going from 64-QAM to 256-QAM means 6 dB of lost margin in the example above, which sounds perilous – 36 dB of SNR against an uncorrected 1E-8 performance threshold of 34 dB. However, the relative digital power in this case is typically -6 dB, so four of those lost dBs are recovered, at the expense of a larger impact on the total RF load of about 2.4 dB. Because that could mean 5 dB of third order distortion degradation (such as CTB), development of hybrids around digital loads had already followed suit in the expectation of larger loads. In fact, the expansion of analog bandwidth in the plant has continued to spur the development of new actives, and these are often designed with the ability to have the extended bandwidth still filled with analog channels. So, as the needs of more digital SNR for 256-QAM payloads started to matter from the RF load perspective, actives had been keeping pace so as not to increase distortion and degrade analog quality.

Now consider 1024-QAM. Figures 1 and 2 (see end of paper) show constellation diagrams of 256-QAM @ 34 dB SNR and 1024-QAM @ 40 dB SNR. These are equivalent uncorrected BER cases supporting a 1E-8 performance. The similar relative relationship of QAM symbol cloud to hard decision boundary is apparent. The congested look of the 1024-QAM diagram, emphasized by the small symbol decision regions, signals the sensitivity this scheme has to disturbance, and is illustrative of the battle ahead. It takes only small impairments to move an otherwise good symbol across a boundary. It is this 40 dB SNR and the sensitivity to analog beat distortions that led to the suggestion that 1024-QAM represents the first digital modulation choice that needs to be treated more like analog modulation than digital.

Consider what 40 dB means in terms of use on the plant. In light of the above example for digital loading, 46 dB of plant analog CNR becomes 40 dB of digital SNR. By adjusting the modulation order upward, and without any other steps taken, we have instantly evaporated virtually the entire link margin, and are now into a region of measurable bit errors, relying on FEC to finish the job under even the most benign circumstance of thermal noise only. Clearly, for plants or areas of plant that deliver poorer SNR, the situation becomes that much more challenging.

Additionally, on the STB side, there is similar margin-limited mathematics. For STB noise figures in the 10-14 dB range, and for QAM signals arriving at the STB at the low end of the power range, some simple math shows the following:

Residual Thermal Noise Floor:
-58 dBmV/5 MHz
Add STB Noise Figure (middle of range):
-46 dBmV/5 MHz
Analog Level into STB: 0 dBmV
Digital Level into STB: - 6 dBmV
(increased for 1024-QAM)
STB SNR contribution: -6 -(-46) = 40 dB

The SNR delivered *to* the STB and the SNR created *by* the STB combine to further aggravate the situation of being on the edge, in this case by 3 dB, delivering a 37 dB SNR. This led to the conclusion that existing conditions and typical deployment scenarios place the ability to ensure a smooth 1024-QAM roll-out at risk. It also reveals the necessity of the 4 dB or so of coding gain – some coding gain reduction occurs between 256-QAM and 1024-QAM for the same scheme – just to ensure link closure at a reasonable error rate.

Field tests on live plant [2] bear out the fact that running 1024-QAM means dealing with imperfections of transmission and limited margin. At the original digital transmit levels used for 256-QAM in this testing example, the authors noted errors accumulating. Increasing the digital power by 6 dB removed most, *but not all*, of these errors. It is not clearer whether the original digital level was relative to analog. A 6 dB increase of the single test channel on the multiplex would have no significant effect on loading whether the level began at -10 dB or -6 dB. However, the authors did note that the resulting SNR measured at the receiver was 36 dB, which is consistent with what is expected based on the sample calculations above. The authors additionally note that the system was tested on a relatively short cascade (N+3), which we will later see to be advantageous.

We will subsequently discuss how changing HFC variables are improving this scenario and helping change it for the better.

Distortion

Prior analysis [3], [4] had investigated the effects of analog beat distortions on 256-QAM. Furthermore, laboratory characterization had developed relationships for the comparative performance of 64-QAM and 256-QAM cable links against narrowband interference, which analog beat distortion represents. It was noted that in the modem receiver technology at the time, there was a 10-12 dB difference in susceptibility to a single, static, in-band narrowband interferer at the main CTB offset frequency of interest. It seemed logical to extend this relationship when discussing the difference between 256-QAM and 1024-QAM for this same situation.

Furthermore, analysis supported by test results indicated that the receiver would begin to count errors in a very high SNR environment for 256-QAM when C/I reached about 36 dB, then slowly degrade over the next 5-7 dB or so before becoming unacceptably error prone. This seemed to agree with field performance, as it would take poor CTB to achieve peaks of distortion of this magnitude, even given the noise-like quality to its amplitude. Minimum FCC requirements of 53 dB along with a noise-like peak-average, digital back-off, and the addition of a thermal noise component correlated with field results that generally showed good deployment results with occasional troubling installations, with a mathematical explanation as follows:

53 dB CTBmin – 6 dB (back-off) – 12 dB (pk-avg) = 35 dB

It was postulated that, relative to static CW RF interference (RFI), the interference cancellation mechanism in the receive equalizer would struggle considerably more with an interferer such as CTB because it has some finite bandwidth and a randomly varying amplitude. This expectation turned out to be the case. The concern echoed at the time for 1024-QAM was that, under the measured performance of that generation of receiver on 256-QAM, another 10-12 dB of interference sensitivity may not be tolerable in many more cases, and in cases where the CTB may be average or better.

Consider Figure 3 in the context of the above discussion about interference effects for 1024-QAM. A 1024-QAM constellation diagram with a 35 dB S/I is shown in the figure. This simulation shows clearly how such a tone, without mitigation, would cause errors with noise added in a 1024-QAM system. In fact, Figure 4 shows the 38 dB S/I case with a 40 dB SNR, and it is clear that hard decision errors are occurring that would require FEC support to correct. Analog distortion events also tend to be slow in duration relative to symbol times – a function of the frequency tolerance of the tone contributors, which results in an effective noise power bandwidth and associated time constant [3]. This then taxes the interleaver as well as the error correction mechanism, possibly requiring these receiver functions to be adapted for the increased relevance of this disturbance.

Again, we will discuss how changing variables in HFC evolution are supporting more robustness in this area as well.

Phase Noise

Untracked phase error leads to angular symbol spreading of the constellation diagram as shown in Figure 5 for 1024-QAM with .25 deg rms of Gaussian-distributed untracked phase error imposed. It was observed in [1] that this represents a reasonable limit to ensure that implementation loss due to phase noise is about 1 dB, assuming low uncorrected BER conditions, and with no practical phase noise-induced BER floor. A floor in the 1E-8 or 1E-9 region will be induced at roughly 50% more jitter, or .375 deg rms. Measurements of phase noise showed that for high RF carrier frequencies, typically associated with higher total phase noise, wideband carrier tracking still left about .33 deg rms of untracked error, enough to cause a BER floor to emerge at very high SNR.

The use of degrees rms is sometime easier communicated as a signal-to-phase noise term, and there are some simple rules of thumb to follow and make this simple, starting with 1 deg rms is equivalent to 35 dBc signal-to-phase noise. Doubling or halving entail 6 dB relationships. Thus, we have the following conversions:

4 deg rms = 23 dBc $SNR\varphi$
2 deg rms = 29 dBc $SNR\varphi$
1 deg rms = 35 dBc $SNR\varphi$
.5 deg rms = 41 dBc $SNR\varphi$
.25 deg rms = 47 dBc $SNR\varphi$

The values .33 deg rms and .375 deg rms represent 44.6 dBc and 43.5 dBc, respectively. This is instructive to compare to the SNR for the AWGN case, as it illustrates the more threatening nature of the phase noise impairment on M-QAM modulations of high M.

Recent BER measurements show that error flooring does indeed occur as measured by pre-FEC errors, suggesting that there have not been significant tuner noise improvements or carrier tracking system changes enough to mitigate this effect. However, although phase noise is a slow random process that challenges burst correcting FEC to handle, the combination of the interleaver, the Reed-Solomon encoding, and the relatively low floor, does indeed result in zero post-FEC errors. Note however, that the phase noise alone is requiring the FEC to work to clean up the output data, consuming some FEC "budget" in the process.

We will not carry forth any further analysis on phase noise, except to note that wideband, low-noise frequency synthesis is an art that has been developed in many other applications, at the expense of some cost, of course. It can be costly in particular because wideband and low noise are competing elements in frequency synthesis – high-Q oscillators characterized by low noise do not tune very far. As a result the designs often involve multiple oscillators, frequency multipliers, switches, and more complex direct digital and PLL-synthesis techniques compared to what is done today. The bottom line is if there was a will for improved phase noise, there is a way.

HFC ARCHITECTURE VARIABLES

Analog Reclamation

Analog reclamation is a "two birds with one stone" architecture variable, offering benefits to both SNR, through potential power loading adjustment, and distortion. In the case of distortion, the benefit under constant total power loading is in the quantity of analog beat components. Constant power loading allows an operator to take advantage of the opportunity to increase SNR in the digital band with newly available RF power load headroom resulting from extracting analog carriers and replacing them with lower power digital ones. Even with this constant operating point, the effect of analog reclamation is to reduce the total number of analog beats accumulating that can fall beneath a digital channel and create interference. Third-order distortions are the ones that accumulate and cascade most aggressively in the digital band, while worst case second-order distortions populate the low end of the band. Thus, we will focus on the impacts of third-order analog beat distortion, or CTB.

SNR

The use of analog reclamation to free up bandwidth for digital channels has obvious and well-understood implications for adding service value to the channel line-up. The exposed bandwidth allows for increasing HD content, more digital channels, more niche channels, and more bandwidth for data services. Also, because digital channels typically run at lower power by 6-10 dB relative to analog, replacement of analog channels with digital results in an increase in headroom that can be exploited for SNR purposes while maintaining the same total RF load on the optics or the RF actives. Table 1 shows what this headroom means in terms of the RF power load when compared to a reference power load, under different RF tilts, for the case of occupied forward path bandwidth to 870 MHz. Table 2 shows the same, but for 1 GHz of loaded bandwidth.

It would be ideal if there were 6 dB available to increase the digital levels. However, it is not critical that there is not, as the existing modulations do run with significant link margin under typical cable plant conditions. For example, 45 dB of analog CNR at end of line at a relative back-off of 6 dB it delivers a 39 dB digital SNR. That's 11 dB of margin for 64-QAM and 5 dB of margin to 256-QAM, not including coding gain. When considering coding gain, even in the 256-QAM case there is substantial SNR margin to work with. As previously described, we have the origins of the CATV network as an analog video network to thank for this good fortune, as well as the fact that its origins were as an RF-only plant. Prior to fiber optic carriage, very long amplifier cascades were required, necessitating decent noise properties from the broadband amplifiers.

Table 1 - Power Loading Effects of Analog Reclamation - 870 MHz

	Channel Uptilt @ 870 MHz					
	Flat		12 dB		14 dB	
	Delta Ref	QAM Increase	Delta Ref	QAM Increase	Delta Ref	QAM Increase
79 Analog	**Ref Load**	**---**	**Ref Load**	**---**	**Ref Load**	**---**
59 Analog	-0.7	2.5	-1.0	1.5	-0.9	1.5
39 Analog	-1.6	3.5	-1.7	2.5	-1.6	2.0
30 Analog	-2.1	4.0	-2.0	2.5	-1.9	2.5
All Digital	-4.5	4.5	-2.8	3.0	-2.5	2.5

Table 2 - Power Loading Effects of Analog Reclamation - 1000 MHz

	Channel Uptilt @ 870 MHz					
	Flat		12 dB		14 dB	
	Delta Ref	QAM Increase	Delta Ref	QAM Increase	Delta Ref	QAM Increase
79 Analog	**Ref Load**	**---**	**Ref Load**	**---**	**Ref Load**	**---**
59 Analog	-0.7	2.0	-0.7	1.0	-0.6	1.0
39 Analog	-1.5	3.0	-1.2	1.5	-1.1	1.5
30 Analog	-2.0	3.5	-1.4	1.5	-1.2	1.5
All Digital	-4.1	4.0	-1.9	2.0	-1.5	1.7

Note: For comparison of Tables 1 and 2, the delta of [1 GHz Ref Load – 870 MHz Ref Load] is as follows:

Flat: 0.25 dB
12 dB Uptilt: 1.27 dB
14 dB Uptilt: 1.56 dB

In the table, the left hand column for each case – flat load, 12 dB tilt, 14 dB tilt – represents the decrease in total RF load compared to the 79-analog channel reference. In every case, the digital carriers run at -6 dB relative level. The right column for each case represents how much *more* power could be allocated to *each* digital carrier in order to maintain very close (within 0.2 dB) to the same total RF power load. This is the added headroom available for SNR that was previously mentioned. These seemingly small available dB become important as we consider increased modulation order, as the move from 256-QAM to 1024-QAM comes with a 6 dB SNR penalty. We will also see how fractions of dBs matter in cases of SNR thresholds vs HFC conditions evaluated later.

In the tables above, the flat case represents the effect on the optical loading of the analog reclamation process. It is apparent than in a mixed multiplex from a single transmitter, the relative digital level allowable will be driven by limitations of the RF plant, where less QAM power can be allocated because of the applied tilt. However, there is nonetheless still headroom that can be exploited by increasing the total power of the analog plus digital multiplex, gaining SNR for all channels. Alternatively, Table 1 and Table 2 show that 1-3 dB of increased QAM power can be applied only to the digital part of the band of the tilted RF outputs, providing some mitigation against the 6 dB increased SNR requirement and subsequent lost margin. It will become clear later as we discuss the effect of cascade depth that these seemingly incremental dBs can have an impact on the ability of the architecture to ensure the desired minimum 1024-QAM SNR objective is met.

Consider simply increasing QAM levels to -3 dB on the 79-channel cascade. This nets a 1.5-2.0 dB total power load increase, which is enough to noticeably impact distortions if nothing else is changed (no cascade shortening). Now consider Table 3. Here, it is postulated that 1024-QAM is run at -3 dBc, with another set of digital channels (half of them) remaining at -6 dBc.

We saw in Tables 1 and 2 that the analog reclamation effort can yield possible QAM level increases for the same power load, but with this typically being less than 3 dB. Nonetheless, to win back margin lost to SNR, in Table 3 we assume that 1024-QAM is run at the -3 dBc level, and that the digital load that is split into -6 dBc and -3 dBc segments. This approach could limit the 1.5-2.0 dB additional power loading effects of having all digital channels increase to -3 dB, but also provide that extra SNR boost to a set of 1024-QAM channels. In one case, the lower half of the digital channels are set at -3 dBc, while in the right-hand columns, the upper half of the digital channels are set at -3 dBc. It is clear from this balance of QAM power that even in the worst case there is a very minor impact on total RF load of < 2 dB using split digital band loading, and essentially no net power increase to the RF load when the lower half of the digital band is used for 1024-QAM. However, it is this part of the band that sees the highest level of CTB beat accumulation, so there is an inherent trade-off between the two.

Note that SNR discussion above refers to the effect of a power increase on a fixed thermal noise floor. However, there is a distortion component referred to as composite intermodulation noise (CIN) that appears like a thermal noise floor, but is in fact a result of distortion products with a digital carrier component. As digital carriers increase, there is more digital contribution to create CIN. In system analysis, the CIN parameter combines with the thermal noise floor to create the parameter known as Composite Carrier-to-Noise, or CCN. CIN looks like thermal noise, and has effects like thermal noise, and is mathematically treated like thermal noise in the calculation of SNR. However, it aggregates as a distortion would aggregate through a cascade, dominated by third order effects. It can be easily isolated in system cascade tests, and the CIN and AWGN components of CCN identified. However, the CIN effect has not been included into the model at this point.

Table 3 - Split Loading for 1024-QAM - 1000 MHz

	14 dB Channel Uptilt @ 870 MHz			
	Lower Digital @ -3 dBc		Upper Digital @ -3 dBc	
	Delta Ref	QAM Increase	Delta Ref	QAM Increase
79 Analog	0.7	---	1.7	---
59 Analog	0.1	---	1.4	---
39 Analog	-0.4	0.5	1.0	---
30 Analog	-0.6	0.5	1.2	---
All Digital	-1.0	1.0	1.1	---

In our examples, performance analysis for nominal and increased output levels lead to the conclusion that CIN contribution is always smaller than the thermal noise contribution, and in some cases negligibly so. Reduction occurs as expected as the cascade shortens. During analog reclamation, adding digital carriers adds more potential CIN contributors as described above. However, the analog carriers drive the highest CIN3, generated as (2A+D), and the removal of the highest level analog carriers more than offsets this increase. Nonetheless, to consider the impact of a CIN effect, we evaluate the SNR threshold developed for the cascade over a range of link noise performance. From these curves, the impact of allowing for increased CIN degradation of SNR can be observed by considering the plotted SNR and adding 2-3 dB maximum degradation.

Distortion

In addition to its positive effects on digital SNR, analog reclamation offers benefits in the distortion domain as well. Some modeled relationships are shown and discussed below.

Consider a load of 79 analog, with a digital load to 870 MHz as a reference for this example. Observing an 870 MHz upper band will allow us to demonstrate the primary impact of constant loading when bandwidth extension is also considered.

Now observe Table 4, and the flatly loaded (optic) cases first. As can be seen, as a 79 channel analog load is reduced, the relative effect on worst case CTB is for it to drop. It also moves lower in frequency. However, the digital channel band extends lower as well, and thus this movement of worst case CTB does not provide much assistance to the QAM channel. The actual CTB peak extends just into the analog band, and there is about a 1 dB difference in the worst case part of the digital band – near the analog crossover. This difference increases slightly as the total number of analog channels is reduced.

The CTB beat count impact on the optical link is significant when the number of analog drops to its minimum. However, typically the optical link of an HFC cascade will dominate the SNR aspect, but not be the primary driver of distortion. That would be the RF cascade, including the RF drive from the node. The effect on third-order digital distortion, CIN3, which adds to the optical link's noise characteristics, is less significant. CIN3 is the dominant digital distortion that accumulates in mixed cable multiplexes.

Finally, note that the CIN3 improvement of analog reclamation can be nearly washed out with the addition of the digital band to 1 GHz on the optical link (flat loading). Obviously, this is just more digital spectrum to add to the nonlinear mix, at the highest amplitude levels, so this is not unexpected.

Now consider the tilted (in this case 12 dB to 870 MHz) examples. The impact of removing analog is magnified, because the analog channels being removed are the highest level channels, and thus the strongest contributors to CTB. This is quantified as up to 15 dB improved worst case CTB for minimal analog channel count, a significant gain. We can observe in Table 5 that nearly this full gain is achieved at the node output – the optical link and one RF amplification *in* the node (N+0 case).

Table 4 - 3rd Order Distortion vs Analog Channel Count

	CTB		CIN3	
Analog Channels	**Flat (Optics)**	**Tilted (RF)**	**Flat (Optics)**	**Tilted (RF)**
79	0 dB (Ref)	0 dB (Ref)	0 dB (Ref)	0 dB (Ref)
59	-2 dB	-5 dB	-1 dB	-2 dB
30	-8 dB	-15 dB	-2 dB	-9 dB
Extend Digital to 1 GHz			Add 1.5 dB	Add 3.5 dB

On the CIN3 side, we see more of the effect of removing the high analog carriers for the tilted multiplex, which drive the highest digital distortion levels as (2A + D). While the reclamation adds digital channels, it adds them at lower power than the existing channels on the uptilt, and, combined with the reduction of highest analog levels, net gains in CIN3 are achieved.

The extension of the band to 1 GHz has a magnified effect in a tilt versus flat case, obviously a result of the digital channels now being relatively higher when installed at the high end of the band.

Cascade Shortening

It is of curse not a secret that each additional amplifier placed beyond the node creates degradation. There are very well-understood rules in the RF world for cascaded degradations of equivalently performing amplifiers:

SNR → 10 Log N
CSO → 15 Log N
CTB → 20 Log N
CIN → 20 Log N

There are caveats to these rules that have to do with the mix of optical and RF distortions, noise and digital distortion, and the active technology used. It is with these rules and caveats, adjusted by offsets associated with the variables introduced in the paper – primarily channel loads and cascade depths, that we can estimate distortion effects and the likelihood it will impact 1024-QAM. While we will use particular numerical examples that represent typical characteristics, it is certainly the case that these dBs can also be subject to variation across product types and models. The intent was to deliver some practical conclusions rather than accumulate worst case assumptions that deliver a skewed result.

Table 5 shows modeled performance for a particular 1310 nm link, followed in the N+6 case by two amplifier types in a 2+4 configuration. The simulation uses mathematical models derived through hardware verification of the particular laser and receiver family, and of the individual amplifier characteristics. Of course, parameters of different lasers, receiver, amplifiers, etc can vary across product families, vendors, implementation, etc., so a “typical” arrangement using nominal levels and link lengths were chosen for a reference point. The data underscores the impact on noise and distortion of decreasing analog channel loads, and shorter RF cascades.

Again, CCN represents Composite Carrier-to-Noise – a combination of the CNR or SNR (optical link dominated) and digital distortions.

Table 5 - Noise and Distortion @ 550 MHz vs Analog Channel Count

	CCN		CTB		CSO	
Analog Channels	**N+6**	**N+0**	**N+6**	**N+0**	**N+6**	**N+0**
79	48	51	58	70	56	64
59	48	52	60	70	59	65
30	48	52	68	74	67	70

This table shows important trends in two important directions that we will further quantify to develop threshold rules for 1024-QAM. Along the rows, the advantages of going from a 6-deep RF cascade to an N+0 architecture, with the node as the last active, are on display for each parameter. This is also sometimes called Fiber-to-the-Last-Active (FTLA). Gains in both noise and distortion are clear, both of which more ably support the ability to handle higher order modulation such as 1024-QAM. Moving down columns, the benefits of doing analog reclamation becomes clear.

From the perspective of noise, the improvements available are of most significance in the shortening of the cascade, and the impact of the lack of accumulation of amplifier noise. There is 3-4 dB additional optical SNR available relative to a typical line-up and cascade depth of today. When coupled with possible loading adjustments with the larger digital tier and added distortion headroom available (from distortion improvements), there are ways to claw back close to the 6 dB of SNR – the amount of increased sensitivity of 1024-QAM compared to 256-QAM.

For both cascade shortening and minimizing the analog loading, Table 5 shows large available distortion gains. These analog beat distortions look like narrowband interferers, but with random amplitude and phase properties and a measurable bandwidth which, as previously described, makes them more difficult to cancel compared to static CW interference. Cascade shortening reduces the 20LogN accumulation of CTB distortion through the RF amplifiers, and these are the HFC elements that tend to drive link distortion. Analog reclamation reduces the amount of beats entered into the mixing process through which the CTBs and CSOs accumulate. Both are aided by changes in the directions shown in Table 5. We will discuss what these results mean more quantifiably in a later section.

TEST RESULTS

Based on field testing previously described, and lab performance characterization testing, it has been clear for some time that the output right at the end of the optical link (an N+0 link) was well-suited to 1024-QAM. The SNR and distortions at this point are as good as they are going to get n the cable, and the cascaded RF link sitting between the node output and the STB can only serve to degrade this. It is this fact that suggests that the N+0 architecture is an ideal HFC evolution supporting 1024-QAM, but also that any cascade shortening works in favor of the higher order scheme. Testing was performed to verify this prediction.

Optical Link Testing

Consider Table 6, which shows the results of typical 1310 nm optics through 20 km of fiber, with 79-channel analog loading and digital loading to 1 GHz. The QAM level is ranged over -4 dBc to -8 dBc, and the 1024-QAM channel inserted in several locations in the loaded digital band, where it's MER and pre-RS FEC BER are measured.

It is apparent, in particular given the -4 dBc data and analog CNRs in the low 50 dB range, that there is a measurement floor associated with the link, reflected in both BER and MER.

Table 6 – 1024-QAM Performance on Fully Loaded Optical Link

		1024-QAM Carrier Frequency		
		603 MHz	747 MHz	855 MHz
QAM @ -4 dB to Analog	MER	39.6	39.2	38.9
	BER	6.1E-08	1.12E-07	3.76E-07
QAM @ -6 dB to Analog	MER	39.0	38.9	38.6
	BER	1.5E-07	2.6E-07	2.5E-07
QAM @ -8 dB to Analog	MER	38.3	38.2	37.7
	BER	4.30E-07	2.02E-06	3.48E-06

The digital SNR in the -4 dBc cases should all be in the high 40 dB's, and, even with reasonable implementation losses, should be error free or nearly so. Since they are not, there is evidence of an impairment affecting the result, with possibilities in this setup being phase noise, beat distortions of 79 channels, I/Q imbalance (modem implementation loss), STB-limiting noise figure, or a combination thereof. However, in all cases in Table 4, the post-FEC BER was zero – it was error-free at the output.

Not shown was another set of data run with 1550 nm optics. The essential result of that link was that all pre-FEC BERs were in the 1e-6 order of magnitude. Based on the above data, and similar noise and analog distortion numbers, it is likely that this result stems from the effect of optical dispersion on the 1024-QAM channel when operating at 1550 nm. In this case also, however, operation post-FEC was error free.

This 1310 nm optical link performance is encouraging in that it was completely error free post-FEC, but with minor flags on taking pre-errors even at very high SNR (BER flooring). It was also completely expected that performance would be good, as the node output – essentially the N+0 case, eliminates any cascade effects. With noise performance at the node output in the low 50 dB range, and distortion outputs in the high 60's, conditions for 1024-QAM appear quite good. Of course, quantifying what distortion levels are acceptable is part of the objective of this paper. Nonetheless, given these excellent RF characteristics at the node output, coupled with previously described characteristics of the tuner performance, it points to phase noise flooring and modem implementation loss (transmit fidelity) as the most likely sources of the pre-FEC BER floor.

While this N+0 scenario presents a clear case as a sound environment for 1024-QAM, unfortunately, as the RF cascade is lengthened, noise degrades at 10LogN per amplifier to combine with this optical SNR. Furthermore, third-order analog distortion cascades as 20LogN to combine with the node output. It is these cascaded impacts that eat into margin for successful 1024-QAM.

RF Testing

To investigate thresholds of distortion interference as a function of SNR, a laboratory test bed was put together, as shown in Figure 6. While not a full cascade, power adjustment variables in the setup allowed it to accommodate a range of distortion amplitudes at different SNRs.

The RF testing performed include varying SNR under conditions of CW and *live video* CTB interference (NCTA practices define CTB with CW carriers), across multiple frequencies in the digital band (three chosen), and observing the pre-FEC and post-FEC performance. Table 7 shows the results of this testing. Note that all dB values are relative to a digital QAM signal, not an analog reference. Also, note that the "Pre-FEC Error threshold" used was 1E-7. Since the receiver bottoms out around 1E-8 (slightly below), this was selected as a point at which non-floor related errors are being taken. The "Post-FEC Error Threshold" is simply the point at which that metric, representing data after it has been through its full clean-up, is still non-zero. It may be the point at which an operator decides that CTB beyond this is, or some acceptable dB higher, is consuming enough of the FEC budget, which then becomes unavailable for other impairments. The post-FEC 1E-6 threshold is logged as a reference point. Post-FEC curves tend to be very steep, and once this level of error accumulation begins, the system is not very far from operating with unacceptable error counts or not at all. It also represents a point of visual threshold-of-visibility (TOV) in some circles. Finally, the

Table 7 - Error Threshold of Interference on Most Sensitive Channel

	CW Interference		CTB Interference			
SNR	Pre-FEC Error Threshold	Post-FEC Error Threshold	Pre-FEC Error Threshold	Post-FEC Error Threshold	Post FEC > 1E-6	Post FEC Broken
50 dB	34 dB	33 dB	55 dB	55 dB	55 dB	45 dB
45 dB	35 dB	33 dB	55 dB	60 dB	55 dB	46 dB
40 dB	36 dB	34 dB	60 dB	60 dB	55 dB	49 dB
37 dB	38 dB	37 dB	60 dB	60 dB	55 dB	50 dB

point at which the system breaks under the weight of excessive CTB is identified.

Note CTB referred to in Table 7 is as it is measured in this setup, *not* per referenced definition and typical practices. That is, rather than CW beats under an analog virtual carrier, it is live video beats underneath a QAM channel power. Beat distortion from live video beats are typically assumed 8-12 dB below "CW" CTB due to the downward analog modulation, with most references identifying 12 dB of "help" by turning on live video. And, QAM power is, of course, derated by some amount to an analog reference – typically 6-10 dB, but as we have been arguing here, this could perhaps be only 3-4 dB for 1024-QAM.

Also note that in this table, the performance noted is the *worst* performing of the three frequencies across the set of cascade measurements taken. These are important nuances, because the average measurement across the frequency set was typically 3-5 dB better than the worst measurement. And, certain cascade combinations actually showed slightly *better* CTB at N+1, for example, compared to N+0. This phenomenon has been observed before in sample test of amplifiers employing E-GaAs technology, which has different distortion-generating mechanisms than, for example, Silicon. The effect has been attributed to this reality of the device physics, and random phasing relationships. While this has encouraging benefits, it is not considered a system "rule" that these amplifiers are immune to degradation in a general sense, and system design goes according to the safe assumption that distortion is normally assumed to degrade.

Table 7 indeed proves a postulate made here and originated in the 2002 paper – that mitigation of CTB is more difficult that CW interference. Again, CTB's random properties challenge the equalizer. In addition, CTB is an averaging measurement of a noise-like waveform. Thus, it has a high peak-to-average ratio, with 12-14 dB often used for practical noise. Theoretical analysis would predict higher, while measured values tend to this range or lower. Meanwhile, the peak-to-average of a CW signal is 3 dB, and, it is of constant envelope amplitude.

Recall, the prediction that a threat could exist for 1024-QAM was based on the CW interference case for 256-QAM, coupled with these factors as follows:

CTB min (FCC) = 53 dBc
CTB pk ~ 40 dB
CTB pk, relative to QAM ~ 34 dBc

In 2002, 256-QAM showed BER sensitivity, pre and post-FEC, for S/I ratios in this mid-30's range. It is encouraging to note that in testing, now seven years hence, that the receiver evolution has led to narrowband ingress cancellation of 1024-QAM that is comparable to what once was recorded to be the limits for 256-QAM. In [5], an ingress cancellation example is shown that displays roughly 30 dB of interference mitigation of a static interferer in a noise-free environment (>50 dB SNR).

From this simple example, it is clear by comparing this 34 dBc value to the left-hand side of Table 7, that the CTB impairment's peak is achieving levels in the range that CW carrier amplitudes were seen to be of concern, but now for the more sensitive 1024-QAM case. It is this relationship that led to the conclusion that plants that were ok (but not great) could be a concern for 256-QAM at the time, and thus even average systems could create issues for 1024-QAM. Now, however, the 1024-QAM CW thresholds are much improved, comparable to prior 256-QAM thresholds. This is much better situation than had the S/I sensitivity been as previous for 256-QAM, knowing the added sensitivity created by 1024-QAM.

Table 7 reveals some possible threshold pairing of noise and distortion for a quality 1024-QAM link. Again, note that these are the most sensitive readings recorded, with averages across the board running 3-5 dB less sensitive. Nonetheless, let's point out the key result: the 1024-QAM receiver saw error degradation under some CTBs that are not inordinately high levels of distortion.

HFC PERFORMANCE THRESHOLDS

SNR

Taking a look at SNR, we can draw some conclusions about the depth of the RF cascade as a function of the expectation from the STB noise figure contribution, and the optical CCN contribution, as a function of a pre-defined overall SNR link objective. Consider Figure 7 and 8. These figures illustrate the nature of small SNR margin increments in their effect on the tolerable cascade depth over a range of optical SNRs, for two chosen link requirements: 38 dB and 40 dB. For simplicity, we have assumed equivalent RF amplifier noise and similar gain, and based these estimates on the results shared in Table 5. The effect of typical gain and noise differences do not substantially changing the qualitative conclusion to be drawn. The net noise figure impact of the gain variation of, for example, splitting the gains between trunk amplification and line extenders, is a dB or so in this N+6 case. This is small, but, as we will point out below, adjusting the RF noise contribution by a dB or two could make a difference in some regions of the curves. Thus, it is desirable *not* to be operating in such regions.

Each curve in Figure 6 and 7 represents a different value of SNR as set by the STB alone, associated with the noise figure and digital level (derated from analog) at its input. Note from the figures that there is a wide range of SNR combinations that essentially offer no practical limit to RF cascade depth as it relates to noise degradation. Clearly, tolerating a 38 dB link requirement, which would be relying very strongly on FEC for just thermal noise aspects, provides a comfortable range of noise performance of the link contributors to work with. However, the range still includes conditions that could lead to a sharp reduction in cascade acceptable. From a sensitivity analysis standpoint, such conditions hinge on small dBs and even fractions thereof. This makes it more valuable to be able to earn back SNR in the analog reclamation process by taking advantage of the load being lightened by the expanded digital, or by taking advantage of headroom in distortion offered by the dwindling beat counts. To ensure operation above the desired threshold, it is not practical to rely on the outdoor plant to maintain key noise and distortion parameters to within fractions of dBs over time and temperature, so operating in these sensitive regions should be avoided.

Again, the 40 dB SNR value represents approximately the 1E-8 point for uncorrected 1024 QAM – a noble objective, setting up the link budget to be acceptable without FEC under the otherwise ideal conditions, and reserving the FEC to deal exclusively with the set of channel issue. However, this may not be practical without lower noise contributors on the CPE side (STB).

CTB Distortion

We will make the similar simplifying assumption on the RF cascade as in the noise case – similar CTB performance and gains across the actives at nominal RF output levels. In this case, splitting the cascade as above among some typical performing amplifier types results in about a 2 dB increase (lower distortion – thus a conservative evaluation) in the third order compression point – the metric that sets CTB. Again, the point here is to recognize that around the data presented below, consider that the exact CTB may not be easily counted on to the dB level, and therefore understanding how performance behaves over a range around the anticipated CTB is important to ensure a robust actual system. The CPE end of the contribution is extracted from the analysis, because the threshold results will be based on measured data that includes that receiver contribution when BER is measured, but the recorded CTB value is that delivered to the receiver input. This leaves us with two contributors – the optical link and the RF cascade – and the variables that modify the contributions to make the assessment.

For the optical link, the beat map effect of flat loading and a fixed transmitter RF load can be relatively easily quantified mathematically and applied as such during analog reclamation. While the link distortion is dominated by the RF cascade, the optics will be included because, as the cascade shortens, it becomes more relevant, in particular with the magnified gains of the tilted channel load on CTB, compared to the flat loading gains. Thus, we will consider the optical link to be the node receiver output, where the multiplex is flat, and bundle the RF stages of the node into the RF cascade for the analysis. Mathematically, then, our N+x case will be modeled as N+x+1. We will assume actives of the same characteristics per the original assumption across the link, including the node RF, established at some nominal reference output level. Again, this is a simplifying assumption, and has single digit IP3 effects for typical ranges of parameters. However, it allows creation of a model that provides a feel for some basic relationships. Of course, the model can be extended to consider any number of individual contributors. Note that there is not a simple formulaic relationship for the RF cascade, as the tilt variation and channel maps both contributing to CTB variation.

Figure 9 shows a sample of an analog beat map for 79 analog channels on a 12 dB tilt to 870 MHz. Tools such as this are used to calculate the impact of varying channel line-ups on *relative* distortion level (tone another and versus frequency) with results as depicted in the Table 4 examples. Note that the peak distortion is third-order (CTB), and that it peaks just below the start of the digital band. As previously described, the worst digital band CTB is about 1 dB lower than this peak. Figure 10 shows the case of flat loading, such as would be carried on the optical link. In this case, the maximum CTB beat location shifts lower, and the difference relative to the worst digital band is closer to 2 dB. These results are used in lieu of an available formulaic representation to evaluate the reclamation process on the titled multiplex distortions.

We now evaluate the ability of the HFC cascade to support a CTB threshold derived from the data provided in Table 7. We will evaluate over a range of given RF amplifier CTBs specified at some typical RF output levels to which these CTB values apply, and evaluate for the varying analog channel counts used above. Two performance thresholds are used, each of which is derived by using the measured case of -60 dBc interference as the objective. The -60 dBc measurement threshold was selected because

- We care most about post-FEC performance
- Post-FEC above 1E-6 is too high, beyond a video TOV, and crashing occurs shortly thereafter on a Post-FEC curve
- It correlates with the range of the SNRs we would expect to see delivered to the home

From the -60 dBc measured value of live video underneath a QAM channel, we make the following adjustments to reference that value to a standard CTB measurement. A standard measurement would be referenced to an analog carrier level, and it would be using CW carriers as load signals. The net results of the former is a 6 dB offset (typical), and we will also evaluate a 3 dB offset to consider the case where more power is given to the QAM load to support the higher SNR requirements of 1024-QAM. When considering live video carriers are the source of the distortion beats and not CW carriers, we have another offset to account for (in the opposite direction – live video is more benign) in the range of 8-12 dB. We will evaluate the endpoints of these cases starting with a -60 dBc measurement threshold, of which we will no longer refer to as "CTB" given the assumptions that the term entails. The endpoints are:

CTBth, min = 60 + 6 – 8 = 58 dBc
CTBth, max = 60 + 3 – 12 = 51 dBc

Figures 11-14 display the results for these cases for the depth of RF cascade, across a range of RF amplifier CTBs, or net effective CTB performance, as a function of the number analog channels from a 79-channel system and going through two stages of analog reclamation.

In Figure 11, it is clear that relying on a 12 dB CTB drop with live video and boosting the QAM power by 3 dB, there is virtually no practical cascade limitations for the cases of analog reclamation employed. In the 79-channel case, only poor or malfunctioning RF amplifiers appear to be a threat to the 1024-QAM system. However, note that there is no loading impact assumed of the added QAM level. We know from prior discussion that this amount of increase in relative level, when applied to the entire digital load, can nudge the total RF level up measurably, increasing distortion. For the cases employing analog reclamation, it has been shown in prior sections that increased QAM level can be obtained with no effect on total RF load, but there is not enough headroom for a full 3 dB increase. There are also techniques such as the split loading of 256-QAM and 1024-QAM that could limit the loading impact for the existing 79-channel case.

For Figure 11, then, the assumption of maintaining the total RF load means either the nominal power load is referenced to the multiplex as described here (it includes margin headroom for this small increase). Or, only a subset of the channels employs 1024-QAM (such as the HD tier for more per-QAM efficiency, or the HSD tier). Or, analog levels are compensated downward. In any case, we will remove the constraint regarding loading in Figure 14 and recalculate this case.

Now consider Figure 12, where the CTBth,min = 58 is used as the guideline. It is clear the power of complete analog reclamation to support virtually any practical RF cascade depth in this case also. However, it also becomes clear how for 79-channel systems and 59-channel systems, some limitations are beginning to appear. This is consistent with the expectation that the march to shorter and short cascades supports the move to 1024-QAM, albeit not as powerfully as what analog reclamation creates in margin. Figure 13 removes the 30-channel analog case, where analog reclamation is complete or nearly so, thus adding granularity to the figure such that reasonable cascade depths can be better quantified as the requirement for them becomes shorter and shorter. Note that by the nature of the requirement chosen compared to the RF CTB range, the curves are unable to cross the x-axis, which represent N+0 (i.e. the combined CTB of the optics and the CTB range of RF used is better than the worst case requirement of 58 dBc).

Finally, re-consider the CTBth,max of 51 dBc under the conditions that the increase in QAM power to a -3 dB derate applies to the whole load, and that the whole load is 1 GHz of bandwidth. The total RF load then is increased about 2 dB, increasing distortions, and in particular third order distortions by 4 dB. This case is shown in Figure 14. Note the shift of a tolerable N+5 cascade to now what might be an N+3 to maintain performance against this threshold if not paying attention to the total power loading. Note also that this assumes a starting derate of QAM at -6 dBc and increasing to -3 dBc. If the starting derate is lower, than the impact on total power is smaller.

With the curves above, operators are able to evaluate the likelihood of 1024-QAM can run successfully on the plant around some basic assumptions. Furthermore, a model has been put in place that allows evaluation on a case-by-case basis for given combinations of equipment, channel plans, and alignment.

SUMMARY

The use of 256-QAM is commonplace for DTV and HSD, and operators have learned what it takes to implement it successfully on mixed multiplexes. Its performance relative to 64-QAM is well-understood, and the additional 33% bandwidth efficiency can be achieved now with minimal cost and pain. There is 25% more bandwidth efficiency available if 1024-QAM is deployed. Modem ICs exist, and this paper sets out to get a feel for the pain part of the equation, understanding that this scheme is yet another level more sensitive to impairments than 256-QAM, and rivaling in many ways the sensitivity of analog video itself.

The good news is that interference analysis conducted in 2002 based on findings extrapolated from 256-QAM performance of that era now seems to have been addressed in the receivers available today. Narrowband cancellation performance seems robust, and a measurable improvement over 2002. The performance against narrowband interference today for 1024-QAM is roughly what it once was for 256-QAM – the level of interference that used to disturb 256-QAM performs to roughly the same error rate now, but on the more advanced modulation. This is very important for analog beat distortion mitigation, which appears as noise modulated narrowband interference. The random element of it makes the receiver have to work harder, and as a result it is less capable of attenuating the interference. The improved performance makes what once looked to be a potentially troublesome CTB problem now a manageable one with some constraints on the HFC link, with performance characteristics plotted against CTB and channel line-up variations.

In terms of noise performance, not much appears to have changed, and the system display a pre-FEC floor likely associated with modem and RF imperfections, such as phase noise. The system relies on FEC to obtain error-free performance under typical plant conditions because of the 40 dB pre-FEC SNR threshold for 1E-8 performance. For 64-QAM and 256-QAM, significant link margin existed under normal conditions to set reasonable thresholds to qualify link acceptability. Because of the increased SNR requirement of 1024-QAM, it is up to the operator to make a different type of determination with respect to link margin and acceptable threshold under normal plant conditions. Acceptable margin will have to be looked at differently for 1024-QAM. The figures in the paper are meant to help operators see the trade space they are working with for SNR in that respect.

Given the results shown here, an awareness of the key drivers to link performance, and a modeling approach that can be used to assess HFC readiness to accept 1024-QAM signals, operators can start netting themselves that extra 25% of bandwidth efficiency that is currently going unused in already occupied spectrum. That is, all the tools and knowledge are in place to start getting new Megabits from the same Megahertz.

ACKNOWLEDGEMENTS

The authors wish to thank Rudy Menna and Chuck Moore for their valuable contributions to this paper.

BIBLIOGRAPHY

[1] Howald, Robert, *QAM Bulks Up Once Again: Modulation to the Power of Ten*, SCTE Cable-Tec Expo, June 5-7, 2002, San Antonio, Tx.

[2] Howard, Daniel et al, *Methods to Increase Bandwidth Utilization in DOCSIS 2.0 Systems*, NCTA Show, June 8-11, 2003, Chicago, Il.

[3] Katznelson, R., *Delivering on the 256-QAM Promise: Adopting Aggregate Degradation Measures for QAM Transmitters,* SCTE Cable-Tec Expo, June 5-7, 2002, San Antonio, Tx.

[4] Stoneback, D., Howald, R., Brophy, T. and Sniezko, O., *Distortion Beat Characterization and the Impact on QAM BER Performance*, NCTA Show, Chicago, Illinois, June 13-16, 1999.

[5] Tan, Loke Kun et al, *A 70 Mb/s Variable-Rate 1024-QAM Cable Receiver IC with Integrated 10-b ADC and FEC Decoder*, IEEE Journal of Solid-State Circuits, Dec. 12, 1998

Robert L. Howald, Ph.D.
Customer Systems Architect
Technology Office
Motorola Home & Networks Mobility
101 Tournament Drive
Horsham, Pa. 19044
407-242-9977
rob.howald@motorola.com

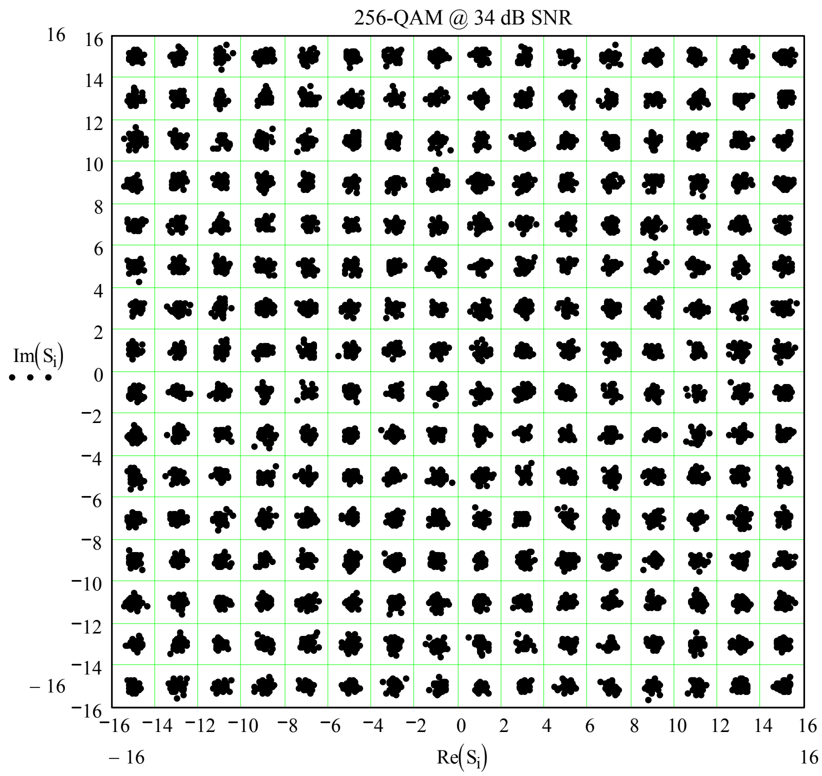


Figure 1 – 256-QAM @ 34 dB SNR

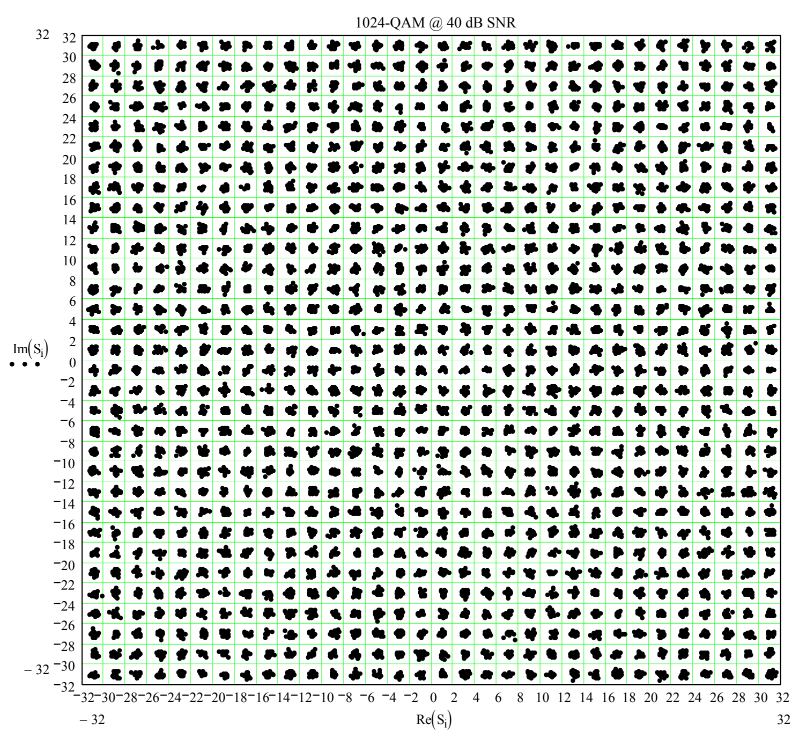


Figure 2 – 1024-QAM @ 40 dB SNR

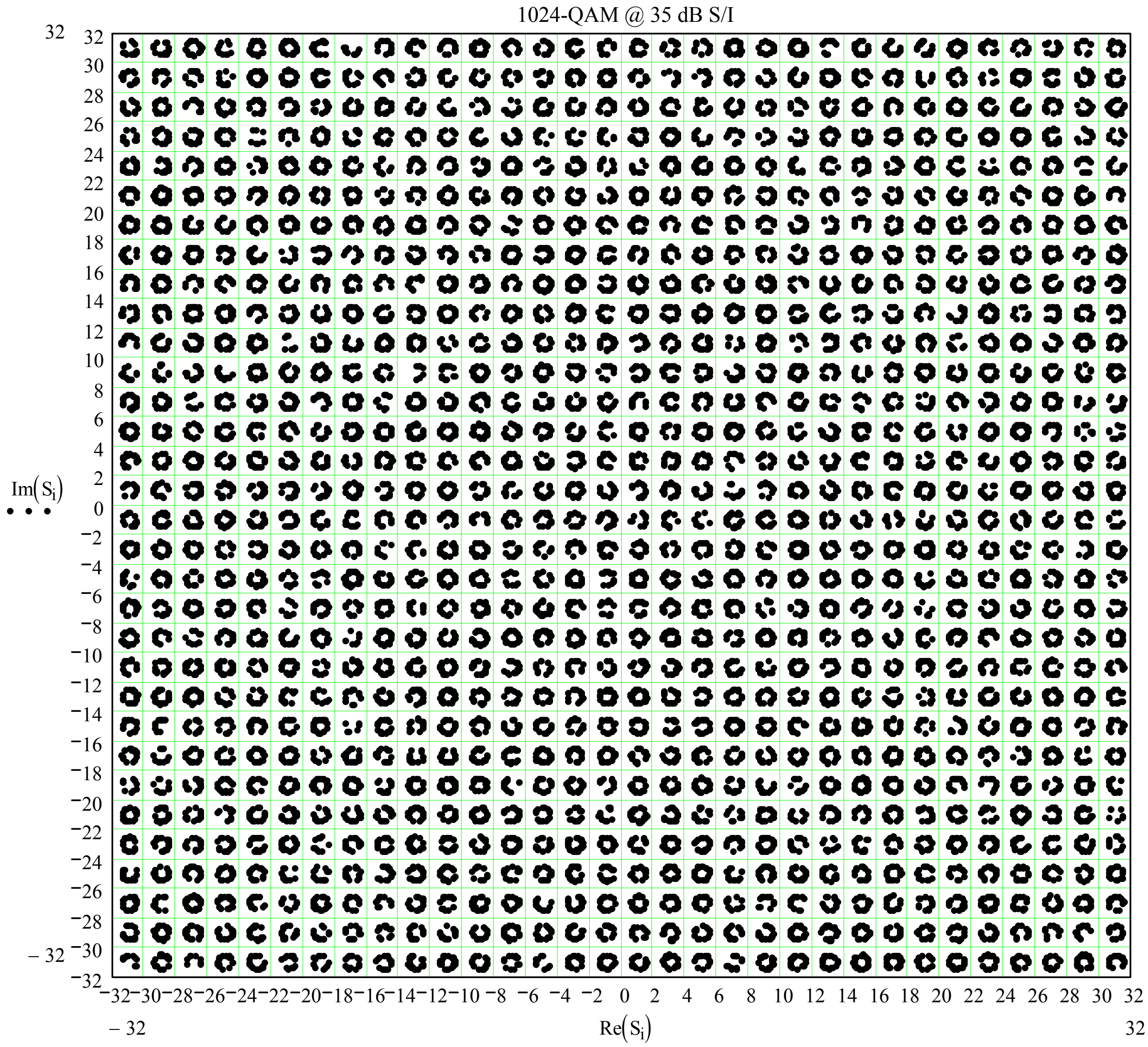


Figure 3 – 1024-QAM with 35 dB Signal-to-Interference (S/I)

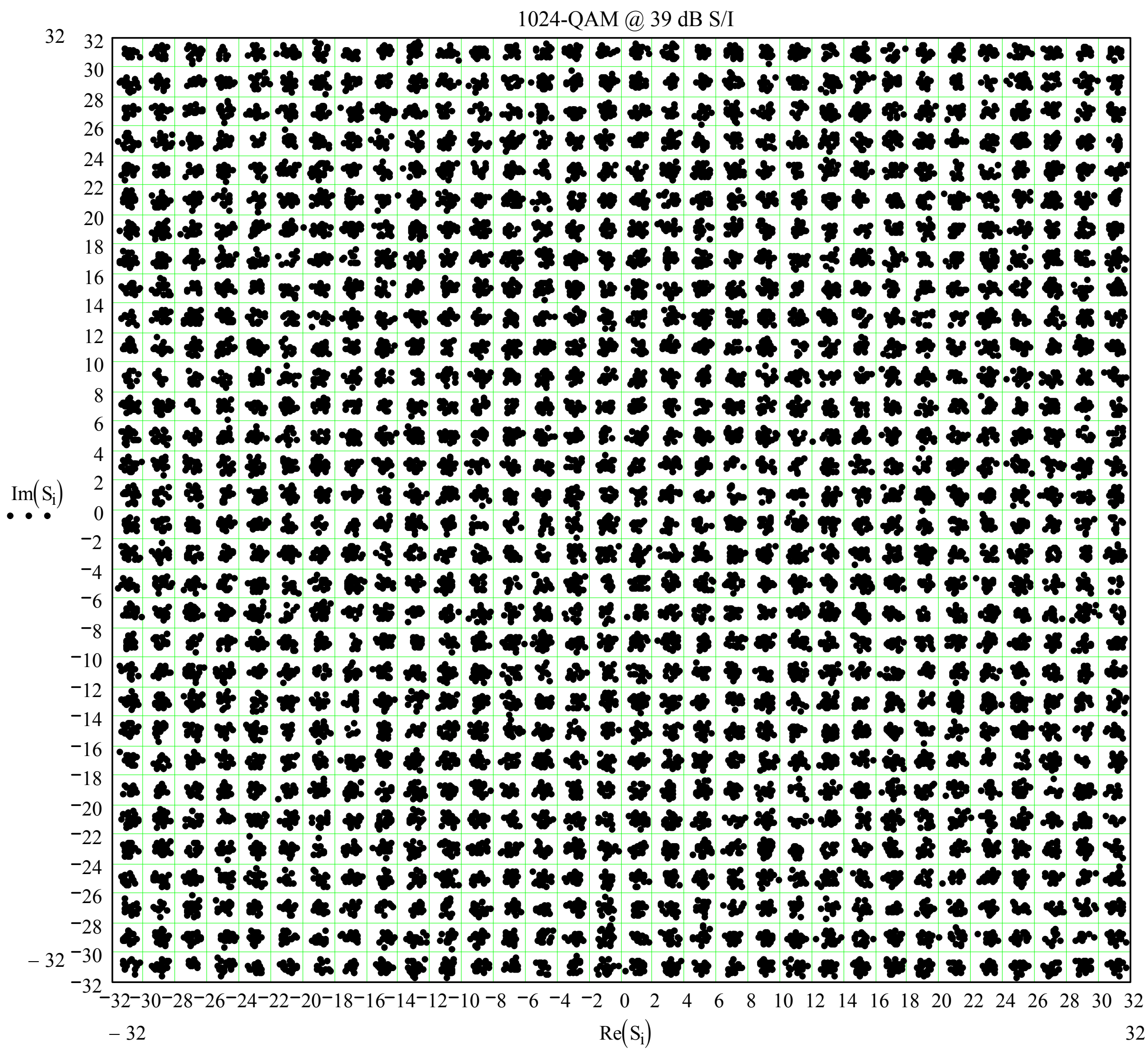


Figure 4 – 1024-QAM with 40 dB SNR and 38 dB S/I

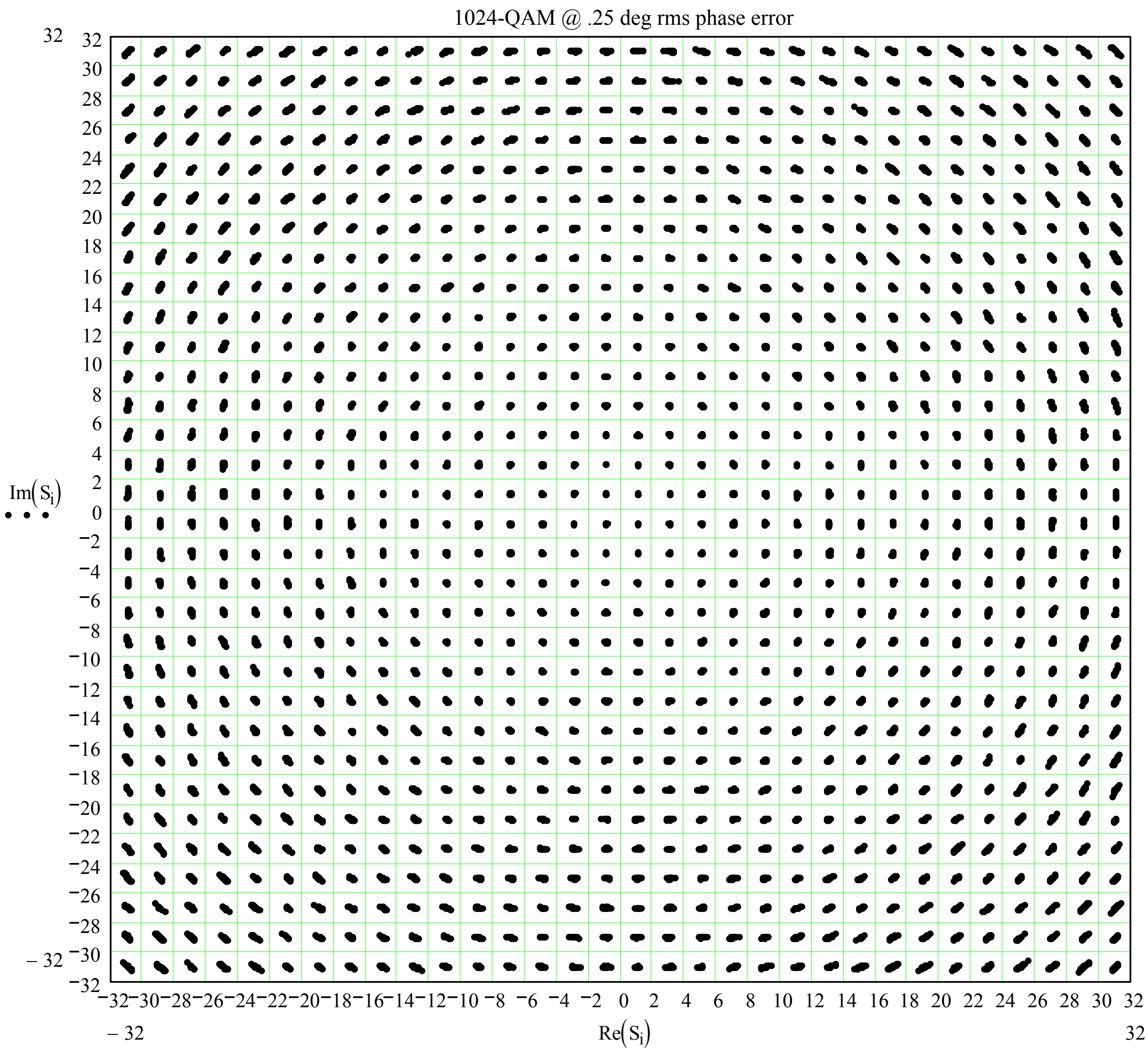


Figure 5 – 1024-QAM with .25 deg rms Gaussian Phase Noise

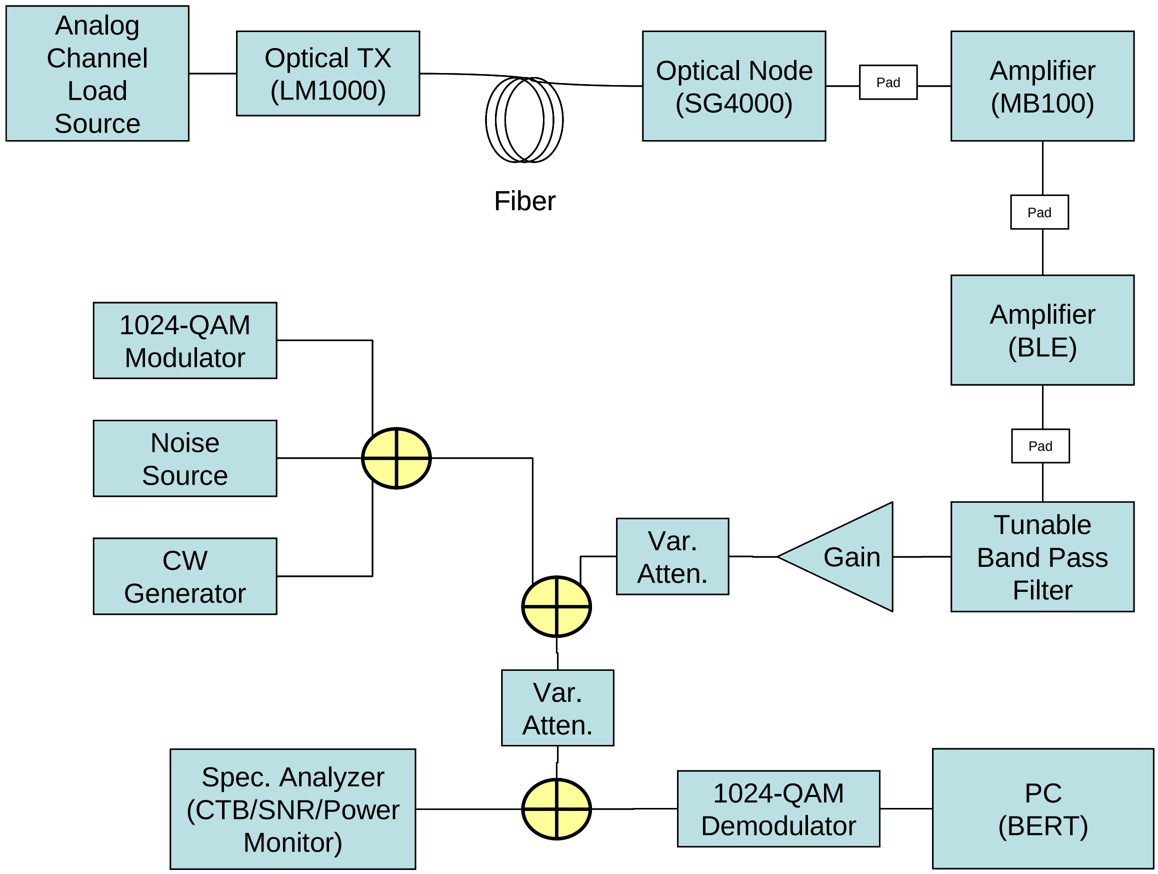


Figure 6 – Distortion Interference Test Bed

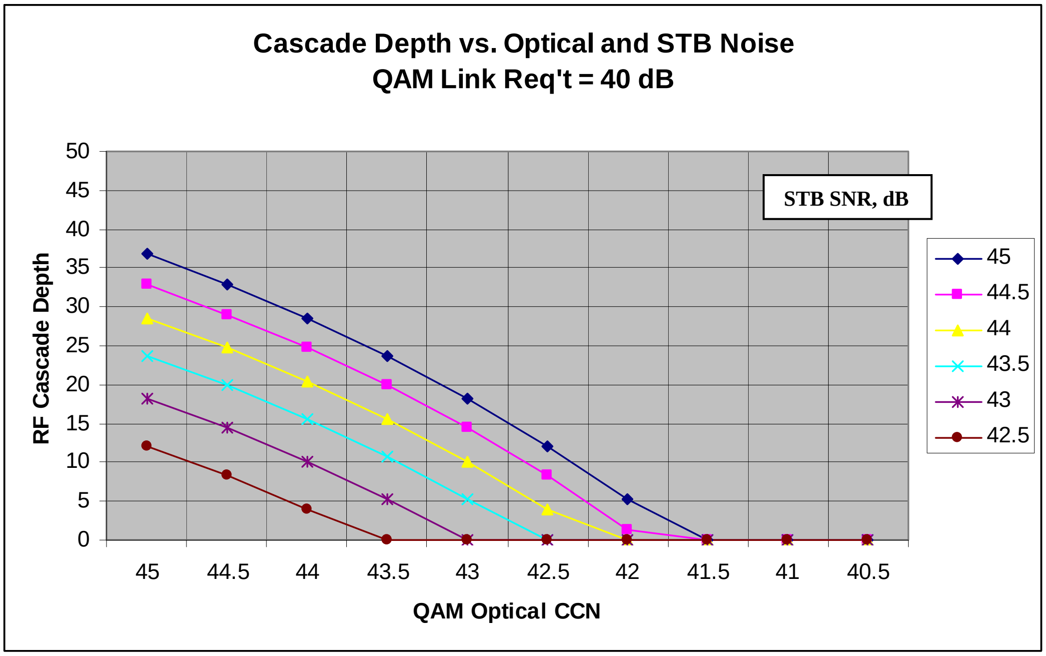


Figure 7 – Sensitivity of Cascade Depth to STB and Optical Noise Contributions for a 40 dB Link Requirement

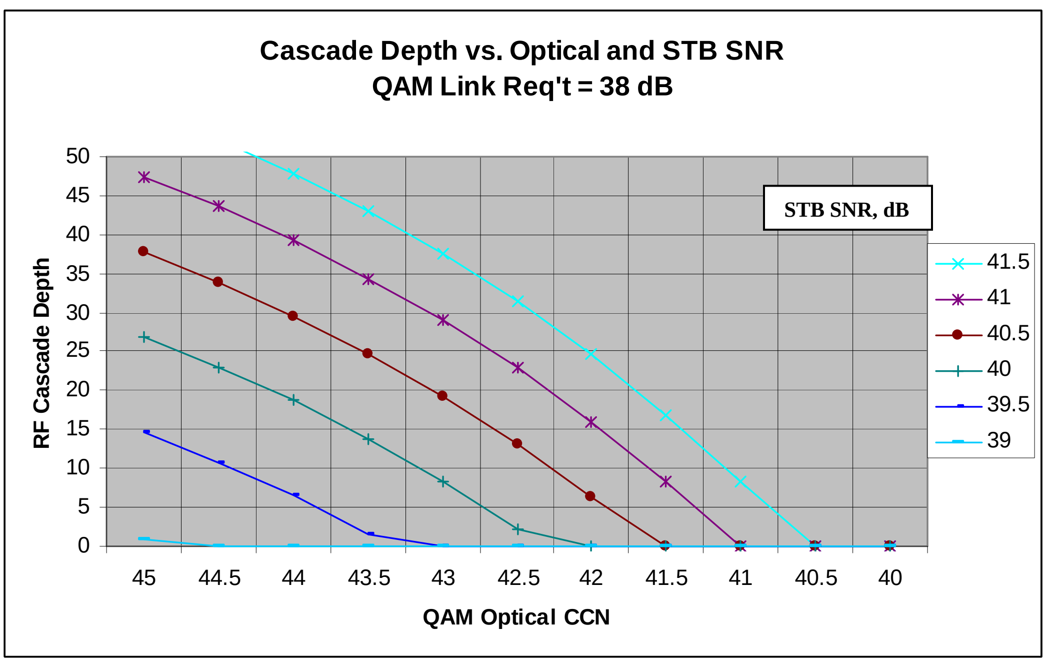


Figure 8 – Sensitivity of Cascade Depth to STB and Optical Noise Contributions for a 38 dB Link Requirement

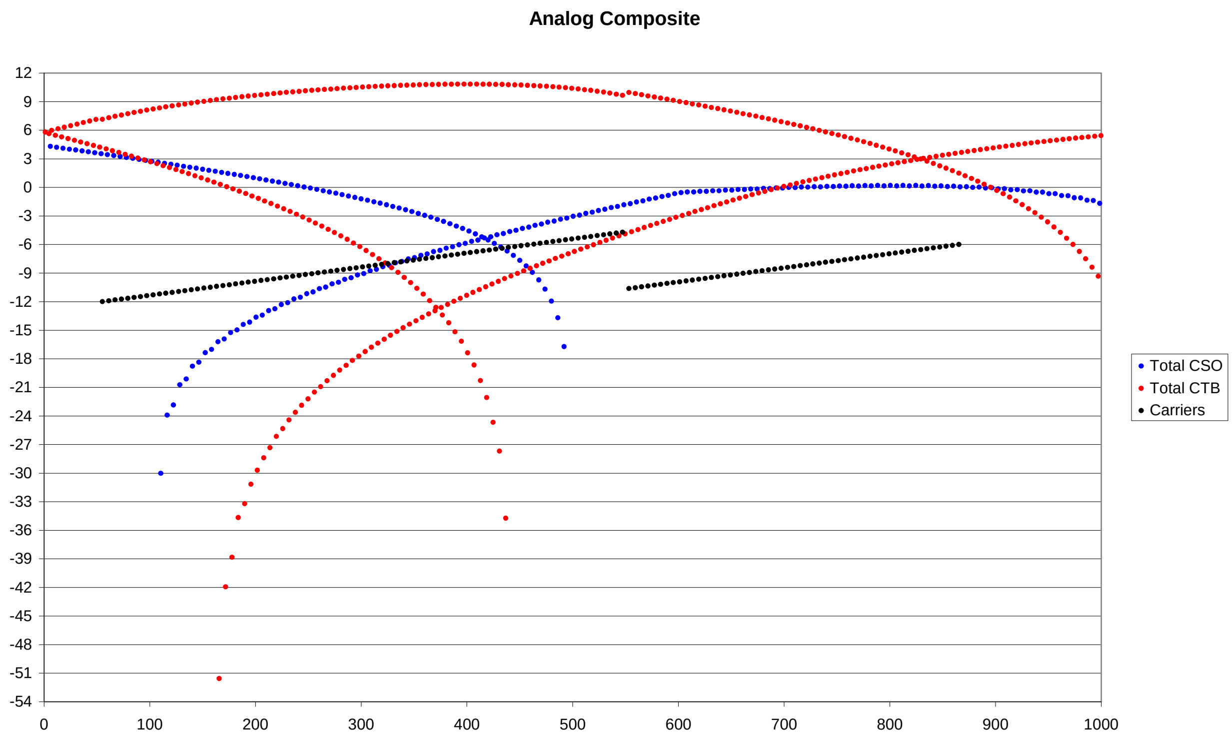


Figure 9 – Beat Distortion Map of 79 Analog Channels on 12 dB Uptilt

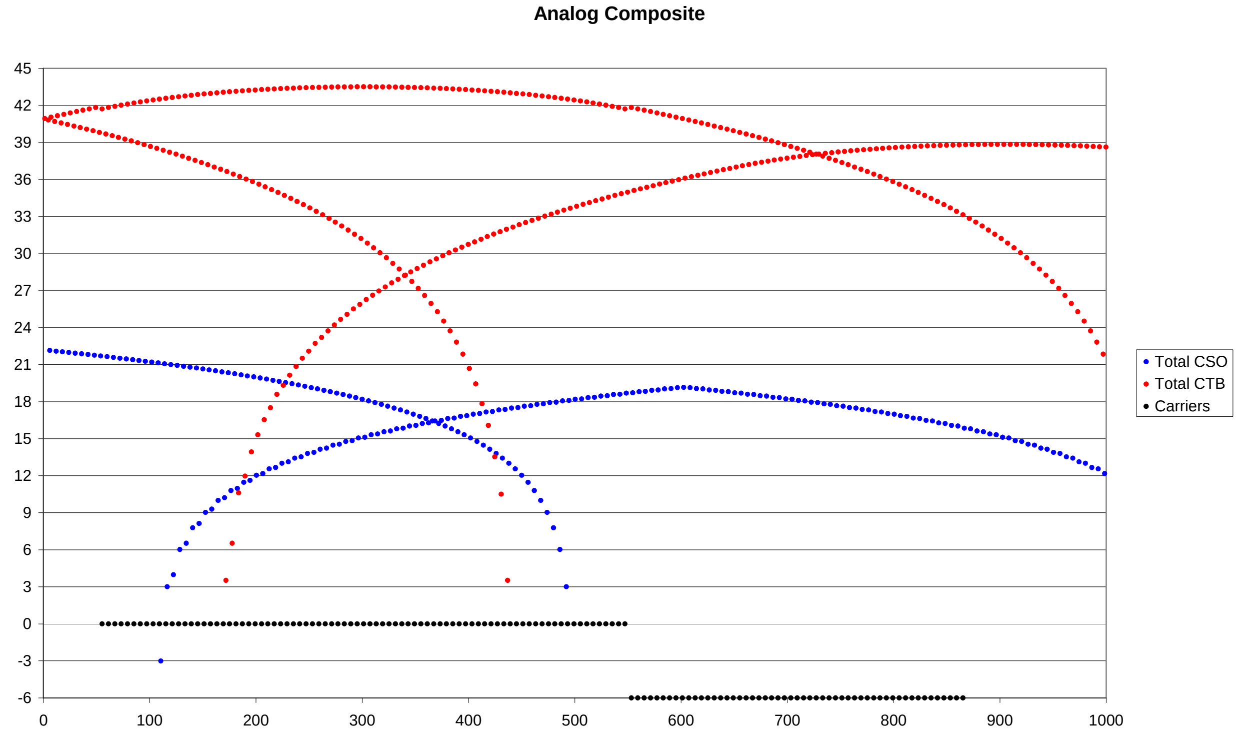


Figure 10 – Beat Distortion Map of 79 Analog Channels Flatly Loaded

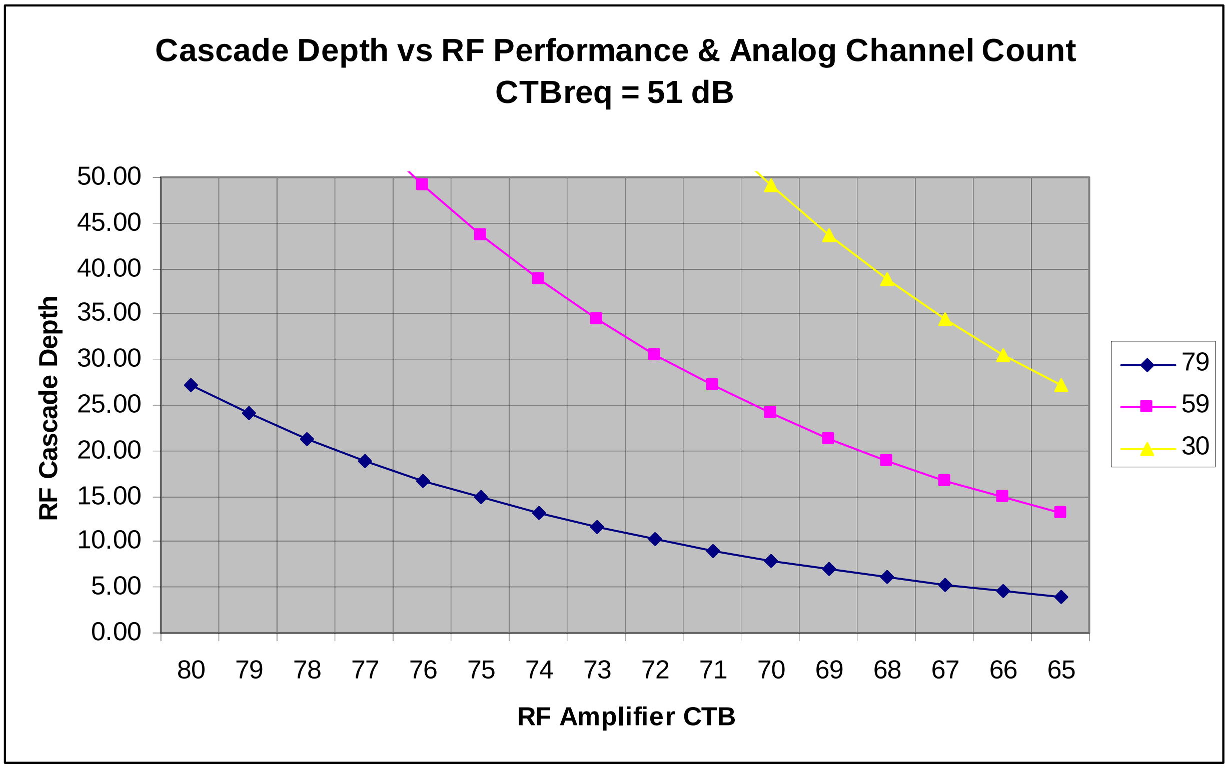


Figure 11 – Cascade Depth Thresholds vs. CTB Performance and Analog Channel Loading, CTBmax = 51 dB

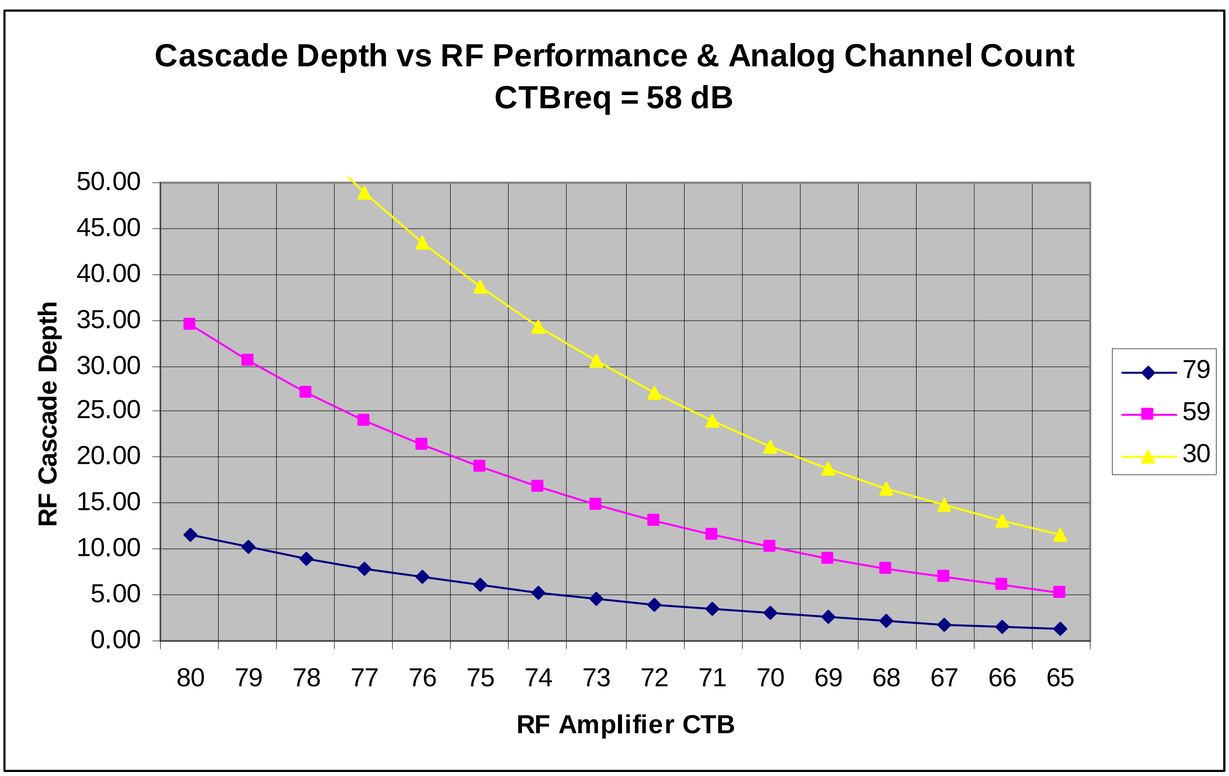


Figure 12 – Cascade Depth Thresholds vs. CTB Performance and Analog Channel Loading, CTBmax = 58 dB

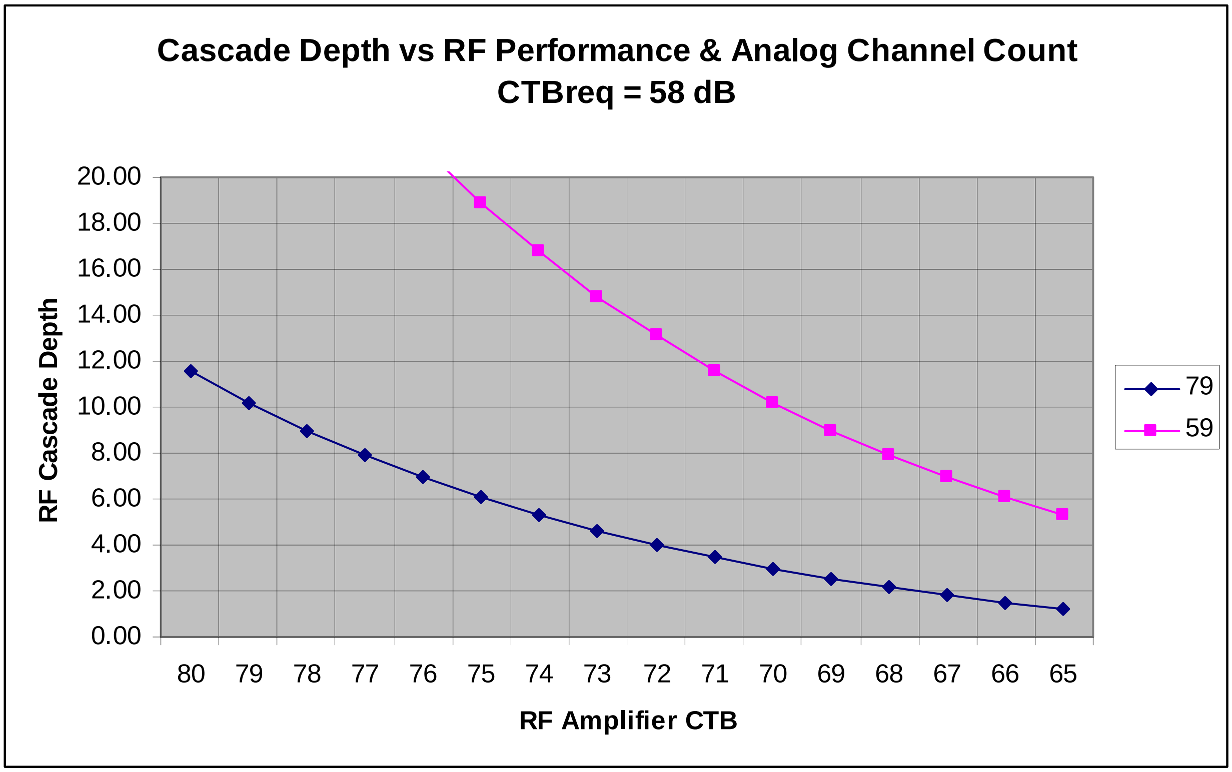


Figure 13 – Cascade Depth Thresholds vs. CTB Performance and Analog Channel Loading, CTBmax = 58 dB, Expanded for 59 & 79-channels only

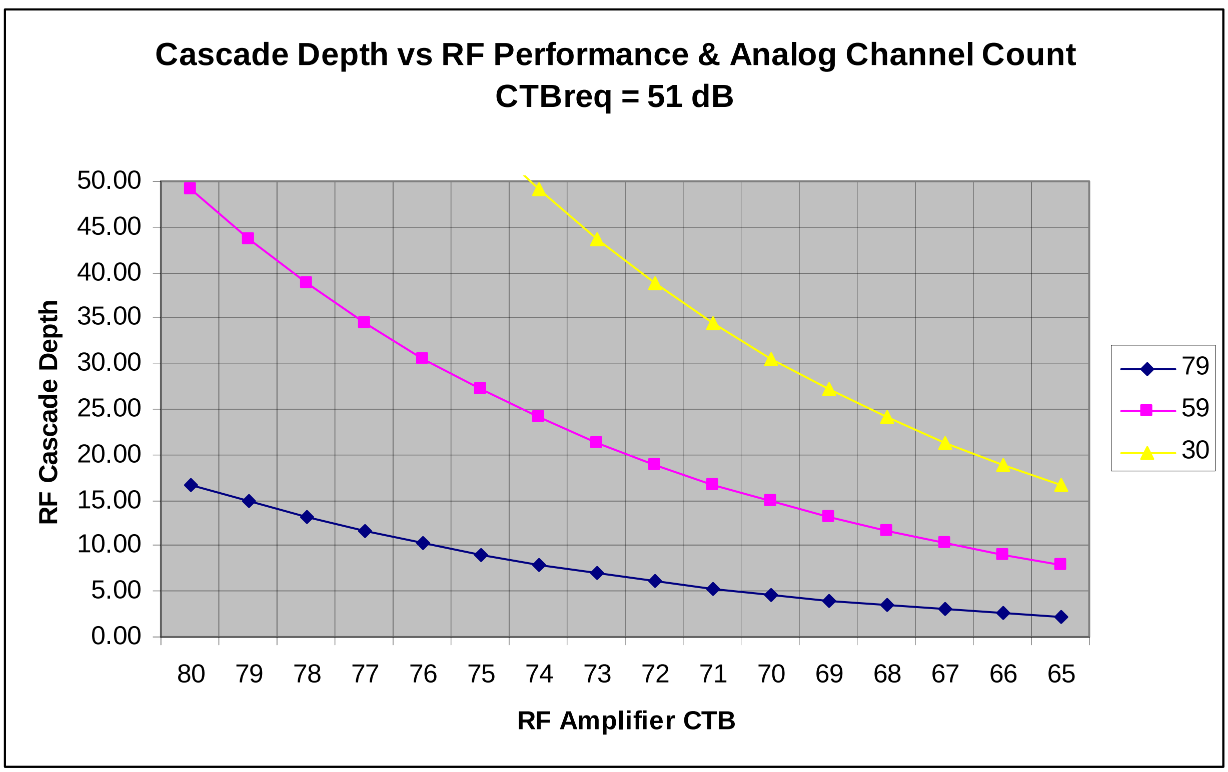


Figure 14 – Cascade Depth Thresholds vs. CTB Performance and Analog Channel Loading, CTBmax = 51 dB, QAM load @ -3 dBc

OPTIMIZING UPSTREAM THROUGHPUT USING EQUALIZATION COEFFICIENT ANALYSIS

Rob Thompson
Chuck Moore
Jack Moran
Rob Howald

Motorola Home & Networks Mobility

Abstract

The current economic climate and increasingly competitive marketplace is driving cable operators to optimize utilization and efficiency of their upstream plant. However, a problem exists when more bandwidth is utilized, as more efficient, but also more sensitive, modulation schemes are leveraged. Impairments that had previously gone unnoticed now present a significant barrier. An understanding of these problematic impairments is critical in order to properly identify, estimate, isolate, and minimize their impact to support this critical transition cable operator's now face. A goal of this paper will be to expose these key impairments that must be managed, and present a strategy for successful optimization of bandwidth utilization and efficiency.

One powerful tool that cable operators have at their disposal is the transmit pre-equalization feature (Pre-Eq) of DOCSIS. Pre-Eq is mandated by CableLabs for interoperability among vendors, and is required of all certified cable modems (CMs) and cable modem termination systems (CMTSs). Its primary function is to mitigate impairments which severely degrade upstream DOCSIS signals. It is not uncommon for the Pre-Eq feature to achieve performance improvements such that Bit-Error-Rate (BER) and Modulation-Error-Ratio (MER) support near error-free operation in an upstream environment that would otherwise be unable to support the transmission at all. Additionally, Pre-Eq has helped DOCSIS upstream signals work in environments where some test equipment will not even work.

While repairing damaged waveforms for detection is the primary Pre-Eq function, in doing so the algorithm learns important information about impairments in the plant. Simple query and analysis of the Pre-Eq coefficients can reveal the dominant impairment for which the CM-CMTS link is compensating. Conclusions may be drawn regarding the impairment contributions, including Micro-Reflections, Group Delay Variation (GDV), and Amplitude Distortion (AD). A plurality of DOCSIS terminal devices, such as CMs, may be engaged in an effort to isolate and minimize a dominant impairment.

INTRODUCTION

This paper will discuss the identification of return path impairments that can be derived by a comprehensive understanding and careful analysis of the data obtained as part of the Pre-Eq function. Conclusions can be drawn based upon coefficient analysis of the DOCSIS 2.0 Pre-Eq settings that the CM-CMTS link collaborate to create. Limits at which Pre-Eq will fail will be provided for conditions of Micro-Reflection impairment only, and a combination of Micro-Reflection, GDV, and AD impairments. These failure thresholds identify at which point the channel is too impaired to function properly, or with reasonable margin. Definition of "reasonable" margin varies among cable operators, but identifying limits will empower cable operators with the tools necessary to ensure their return plant is within their criteria.

This paper will demonstrate a mathematical simulation, supported by laboratory verification, as well as live CATV plant tests, all of which were used to establish and validate the proposed limits. A systematic approach for isolating discussed impairments within a CATV plant will be proposed. This extension of the Pre-Eq information into the realm of HFC maintenance enables cable operators to identify suspect CATV plant components that may be contributing to an impairment problem, and thus take corrective action pro-actively.

EQUALIZATION

This section introduces the fundamental digital communications receiver function of equalization, and how the operation of the equalizer generates coefficient information that can be used for diagnostic purposes. While equalization is part of virtually all modern telecommunications platforms, for cable it is instrumental in proper return operation for advanced DOCSIS systems. In order to offer higher data rates to its subscribers in the competitive world of high-speed Internet access, operators must take advantage of the throughput benefits gained from leveraging more complex digital modulation schemes, such as 32-QAM and 64-QAM. Unfortunately, these schemes are also considerably more sensitive to digital communication channel impairments than the 16-QAM channels (maximum) they are replacing in the return band.

HFC Impairments

Equalizers are very powerful tools within the digital receiver, and they can hide a lot of sins. Relevant return path impairments that can be mitigated by equalization include Micro-Reflections, AD, GDV, and Narrowband Interference. Amplitude Distortion is also referred to as Attenuation Distortion (AD). GDV, a phase-related distortion, is also commonly referred to as Group Delay Distortion (GDD) or Envelope Delay Distortion (EDD) in the telephony world.

It is important to recognize that impulse and thermal noise will reduce the equalization algorithm's ability to mitigate these aforementioned impairments. These commonplace impairments introduce random amplitude and possibly decision errors that the equalization algorithm operates on. Successful application of coefficient analysis assumes that the CATV plant meets sufficient impulse and noise requirements necessary to support a desired modulation level use. However, the fact that some level of these noise contributors exists will introduce some error into the calculated Pre-Eq solution. These issues can be overcome with modifications to the Pre-Eq update approach, but as a closed loop system, it requires careful analysis of the plant conditions to ensure stability of operation.

Digital Signal Characteristics

The digital signal characteristics used to generate all of the data presented in this paper are as shown in Table 1. The 6.4 MHz DOCSIS channel bandwidth was chosen because its widened bandwidth makes it more sensitive to impairments. Of the two formats, clearly 64-QAM is the more sensitive of the two. This actually makes the more robust wideband 16-QAM particularly valuable as a diagnostic tool in this mode.

Table 1 - Signal Characteristics

16-QAM	64-QAM
5.12 Msym/sec	
6.4 MHz	
20.48 Mbps	30.72 Mbps

RRC Matching (α = 0.25)
Square Constellation (max/min = ±1v)

Equalization Fundamentals

Equalization is the process by which a digital communications receiver (in collaboration with transmitter in the Pre-Eq case) estimates the inverse of the digital communication channel response, $H_c(f)$, and applies it to the incoming signal. The process is as illustrated by the following transfer function equation [1].

Eq 1 - Equalization Transfer Function

$$H_e(f) = \frac{1}{H_c(f)} = \frac{1}{|H_c(f)|} e^{-j\vartheta_c(f)}$$

By using the inverse function, equalization attempts restore a received digital signal to an ideal response - canceling the impairment encountered in the digital communication channel. In doing so, equalization minimizes *inter-symbol-interference* (ISI). ISI is the mechanism whereby the frequency response distortions previously noted (AD, GDV) cause in the time domain, adjacent digital symbols to leak into one another and cause interference most clearly observed on an eye diagram.

DOCSIS Equalization

Pre-Eq is required for DOCSIS compliance. The endpoints of the DOCSIS link (CM and CMTS) collaborate to converge upon an estimate of the communication channel response and bias transmission in order to, ideally, cancel any impairment that may be present. While the Pre-Eq function significantly hardens the link to impairments, as an added bonus there is much more information that can be mined to help the operator more broadly. Understanding what communication channel impairments the DOCSIS link is attempting to cancel offers the operator with extremely valuable diagnostic data. The more DOCSIS compliant devices located throughout the HFC plant, the better equipped the cable operator is to use this information for proactive maintenance, potentially eliminating a source of subscriber calls.

DOCSIS 2.0 Pre-Eq uses twenty-four, symbol-spaced coefficients, also called taps. For example, a CMTS estimates the values of these coefficients and forwards this information to a CM via station maintenance messages. These coefficients are used for amplitude and phase correction over a twenty-four symbol period time window. There are several reasons for having equalization functionality in the transmitter in addition to the receiver that are out of scope of this paper. But, in general, the more complex and diverse a channel transfer function may be, the more well-suited it is to deploying equalization resources at both ends. Cable's return path has a large range of possible frequency response signatures even within a single RF leg of an HFC plant.

Coefficient Interpretations

DOCSIS 2.0 Pre-Eq coefficients are a list of 24 complex values and may be interpreted in multiple ways, as demonstrated in Figure 1 through Figure 5. The figures presented here represent a digital communications channel with negligible levels of impairment. These can be used as a baseline to aid in the impairment identification process.

Figure 1 is the magnitude of the equalizer's impulse response, $|h_e(t)|$. A single line at Tap = 0, known as the main tap, on the x-axis is the ideal response. In that case, whatever stimuli excite the channel is perfectly replicated. Calculation of impulse response magnitude is based upon complex tap values of DOCSIS 2.0 Pre-Eq and is shown in Eq 2.

Eq 2 - Impulse Response Magnitude

$$|h_e(t)| = 20\log_{10}|h_e(t)|$$

The main tap represents the desired symbol energy, while the remaining taps represent negligible correction magnitudes of < -35 dBc. The small random magnitudes of the non-zero taps are primarily due to simulated system noise.

For coefficient analysis of multiple CMs, it is helpful to break-down impulse response measurements into regions in which dominant impairments will have the greatest impact. Numerically sorting on these impaired regions facilitates efficient organization of similarly impaired CM groups, and this can help in diagnosing issues.

Two important regions of the impulse response to focus on are the *post-tap* region and the *main tap* region. Dominant micro-reflections typically impact the post-tap region, which consists of tap 1 through tap 16. Dominant AD and GDV typically impact the main-tap region, which consists of taps adjacent to the main tap, numbers -3, -2, -1, 1, 2, and 3.

Figure 2 is the phase of the equalizer's impulse response, $\theta_e(t)$. Calculation for phase is simply the argument of the complex tap values of DOCSIS 2.0 Pre-Eq and is shown in Eq 3. The impulse response phase appears randomized between –π and π, except for the main tap whose phase correction is 0 radians. While this plot looks "noisy," recognize from the Magnitude response that the amplitude contribution of the randomly phased imperfections is negligible.

Eq 3 - Impulse Response Phase

$$\theta_e(t) = \arg(h_e(t))$$

Figure 3 is the equalizer's amplitude response, $|H_e(f)|$. Calculation for amplitude response is based upon a 1024-point, Fast Fourier Transform of the DOCSIS 2.0 Pre-Eq coefficients and is shown in Eq 4. Note that the equalizer's amplitude response is ideally constant throughout the channel's bandwidth, which is the Fourier result of the ideal impulse in Figure 1.

Eq 4 - Amplitude Response

$$|H_e(f)| = 20\log_{10}\left|h_e(t) \overset{FFT}{\leftrightarrow} H_e(f)\right|$$

Figure 4 is the equalizer's phase response, $\theta_e(f)$. Calculation for phase response is based upon a 1024-point, Fast Fourier Transform of the DOCSIS 2.0 Pre-Eq coefficients and is shown in Eq 5. Note that the equalizer's phase response is ideally linear throughout the channel's bandwidth.

Eq 5 - Phase Response

$$\theta_e(f) = \arg\left[h_e(t) \overset{FFT}{\leftrightarrow} H_e(f)\right]$$

Figure 5 is the equalizer's Group Delay (GD) response, $GD_e(f)$. Calculation of GD is based upon the phase response and is shown in Eq 6. Note that the equalizer's GDV is approximately 6 nsec across the channel's bandwidth. Group delay is anther way of describing the phase characteristics, but in a way more intuitively descriptive. Group delay represents the absolute time delay each frequency component across the band will endure. As such, it is the variation of this delay (non-flat delay) that matters most, as components of frequency arriving at different times at the opposite end of the channel result in distortion and ISI.

Eq 6 - Group Delay Response

$$GD_e(f) = -\left(\frac{\delta\theta_e(f)}{\delta f}\right)$$

The subsequent sections will show how communication channel impairments will uniquely impact the DOCSIS Pre-Eq coefficient interpretations discussed above.

MICRO-REFLECTION

This section introduces the micro-reflection impairment and how it impacts the DOCSIS Pre-Eq coefficients. As seen by a receiver, a micro-reflection is a copy of the transmitted signal, arriving late and with reduced amplitude. The result of the additional copy is the familiar image ghosting in analog video reception, but for digital communications this is ISI.

Micro-Reflection Fundamentals

Micro-Reflection sources are composed of pairs of HFC components separated by a distance of cable. What's important to understand about HFC components is that they facilitate the propagation of the signal copies in a variety of ways including return loss, isolation, mixing, and combining.

Figure 6 illustrates one of many possible micro-reflection source configurations. Two devices with poor return loss, acting as signal reflectors, are separated by a length of cable. The CM is acting as the second reflector, but any HFC component has the potential to achieve a similar result. The reflector return loss and the loss between the reflectors determine the amplitude of the micro-reflection. The delay encountered as a signal copy traverses the red path of Figure 6 will determine which equalizer tap is responsible for correction.

Note that the CM has as a design limit has a high return maximum loss value of 6 dB, meaning it may reflect up to 25% of its incident power. In the plant, design limits are typically significantly better, but over time will degrade as the plant ages and elements that contribute to good RF matching – connectors, cable, splitters, interfaces on PCBs – degrade.

QAM Signaling Impact

Micro-Reflection impairment impact may be measured on a spectrum analyzer as amplitude ripple. The peak-to-peak amplitude and frequency of the ripple, shown in Figure 7, are directly related to the micro-reflection impairment's amplitude and delay. One look at the frequency response signature quantifies the micro-reflection's parameters. In this case, the signal is impaired by a micro-reflection whose relative amplitude is -20 dBc and whose delay is 4 symbol periods.

In a QAM constellation, a micro-reflection causes the symbols to spread in a miniaturized pattern similar to the full QAM constellation itself. Additionally, phase distortion may cause the spread symbols to appear rotated. Consider first Figure 8 and Figure 9, which illustrate ideal constellation diagrams for 16-QAM and 64-QAM, respectively.

Now consider Figure 10. Figure 10 shows the effect of a micro-reflection on a16-QAM signal's constellation diagram. The micro-reflection's characteristics are those previously depicted in Figure 7. Note the spread throughout the symbol region on each 16-QAM point, and subsequently how now less additive noise has more likelihood of causing a symbol to jump a boundary and create a hard decision error than the Figure 8 case.

Figure 11 repeats the same micro-reflection scenario for 64-QAM. 16-QAM is less sensitive micro-reflections than 64-QAM because of reduced decision boundary size of 64-QAM for the same average transmit power. In other words, 16-QAM symbol size can spread more than 64-QAM. In comparing Figure 10 and Figure 11, the same level of micro-reflection impairment has spread the symbols of the 64-QAM constellation appreciably closer to the symbol boundaries than in 16-QAM constellation. The 64-QAM situation is clearly a catastrophic situation with some added noise unless some equalization processing is applied to undo the micro-reflection.

Pre-Eq Coefficient Analysis

A single dominant micro-reflection uniquely impacts the DOCSIS Pre-Eq coefficients. The post-tap region of the impulse response magnitude, illustrated in Figure 12, reveals the characteristics of the Figure 7 micro-reflection: amplitude -20 dBc relative to the main tap, and delay 4 symbol periods later than the main tap.

The impulse response phase reveals negligible phase distortion of both the desired symbol and the micro-reflection impairment.

The equalizer's amplitude response of Figure 13 shows the equalizer's amplitude response. This response derived should be compared to Figure 7 in the context of Eq 1.

The phase response shows some nonlinearity across the channel's bandwidth, especially when compared with Figure 4 phase response. The GD response of Figure 14 clarifies the additional phase distortion with appreciably higher GDV than was previously illustrated in Figure 5. Note equalizer's GDV is approximately 43 nsec across the channel's bandwidth, while the symbol duration itself is less than 200 nsec.

DOCSIS Micro-Reflection Assumptions

CableLabs has identified via the DOCSIS standards [2-5] certain assumptions regarding the nature of a single dominant echo or micro-reflection present in an HFC environment. DOCSIS compliant devices must interoperate at or below the values illustrated in the Figure 15.

AMPLITUDE DISTORTION (AD)

This section introduces the AD impairment and how it impacts the DOCSIS Pre-Eq coefficients. AD is undesirable variation in the communication channel's amplitude response resulting in distortion of the digital signal amplitude. Some common forms of AD include tilt, ripple, and roll-off.

AD Fundamentals

One common cause of AD in an HFC plant is upper return band-edge operation of digital signals, combined with long reaches of coaxial plant. Long reaches of coaxial plant accumulate multiple diplex filters from amplifiers and in-line equalizers. While individually contributing small attenuation versus frequency, when part of a deep RF amplifier cascade, the combination may contribute appreciable variation in a communication channel's magnitude frequency response. An example of amplitude roll-off has been provided in Figure 16.

QAM Signaling Impact

In a QAM constellation, amplitude roll-off causes the symbols to spread in a pattern similar in appearance to Additive White Gaussian Noise (AWGN). For reference, Figure 17 and Figure 18 are constellations for 16-QAM and 64-QAM respectively, which have been impaired by equivalent levels of AWGN.

Increasing AWGN contribution by 6 dB would show that 16-QAM is now just as close to the decision boundaries as 64-QAM was previously in Figure 18. Conversely, reducing AWGN contribution by 6 dB would show 64-QAM is now just as far away from the decision boundaries as 16-QAM was previously in Figure 17. Use of 6 dB demonstrates a well-known relationship between AWGN and modulation complexity on square constellations such as 16-QAM and 64-QAM. Each modulation order involves a 6 dB increased sensitivity to thermal noise from 16-QAM on up (for QPSK to 16-QAM it is closer to 7 dB).

The 6 dB relationship is isolated to the thermal noise impairment. Similar assumptions must *not* be made regarding the impairments discussed in this paper. Many such additional impairment relationships have been derived and discussed in the literature over time. For our case here, simulation and test is crucial for characterizing the true nature of the relationship which exists between these more complex impairments and modulation complexity, and in particular for multiple simultaneous impairments.

Figures 19 and 20 represent the 16-QAM and 64-QAM constellations that result from the amplitude roll-off illustrated in Figure 16. Note the appearance of these as compared to the AWGN cases in Figures 17 and 18.

As with prior impairments, 16-QAM is less sensitive to amplitude roll-off than 64-QAM because of reduced decision boundary size of 64-QAM. In comparing Figure 19 and Figure 20, the same level of amplitude roll-off impairment has spread the symbols of the 64-QAM constellation appreciably closer to the symbol boundaries than in 16-QAM constellation. And, again, the 64-QAM case is bordering on a catastrophic link result without some intervention.

Pre-Eq Coefficient Analysis

Amplitude roll-off uniquely impacts the DOCSIS Pre-Eq coefficients. The near main-tap region of the impulse response magnitude, illustrated in Figure 21, reveals main-tap spreading in the region of main tap +/- 3 taps. The amplitude response of Figure 22 reveals appreciable amplitude correction.

Lastly, there is linear phase and negligible GDV across the channel's bandwidth. Note equalizer's GDV is approximately 12 nsec across the channel's bandwidth.

GROUP DELAY VARIATION (GDV)

This section introduces the GDV impairment and how it impacts the DOCSIS Pre-Eq coefficients. GDV is undesirable variation in the communication channel's phase response resulting in distortion of the digital signal phase, or, as described, a variation in the propagation of frequency components of the signal across the channel.

GDV Fundamentals

As is the case for AD, one major cause of GDV in an HFC plant is upper-band-edge operation of digital signals, combined with long reaches of coaxial plant. The reasoning is the same as in the AD case. Note that filtering functions typically induce nonlinear phase responses as the band edges are approached, so the combination of AD and GDV in the same band region, understanding that diplex filtering is the cause, is perfectly expected. Different filter functions induce different GDV responses, just as different filter functions induce different stop-band characteristics. It is common that the sharper the roll-off, such as would be the case for long cascades, the worse the GDV will be.

QAM Signaling Impact

In a QAM constellation, GDV causes the symbols to spread in a pattern similar to what has already been illustrated for AWGN and AD.

As expected, 16-QAM is less sensitive to GDV than 64-QAM because of reduced decision boundary size of 64-QAM.

Pre-Eq Coefficient Analysis

GDV uniquely impacts the DOCSIS Pre-Eq coefficients. The main-tap region of the impulse response magnitude, shown in Figure 23, reveals main-tap spreading as was illustrated for the amplitude roll-off impairment. However, the amplitude response, shown in Figure 24, reveals appreciably less amplitude correction. Since the induced impairment is phase related, this makes sense.

Of course, there is an appreciable amount of phase variation in the impulse response phase and the phase response, while Figure 25 reveals appreciable GD correction over the GD correction present for the amplitude roll-off impairment scenario previously discussed. Note equalizer's GDV was approximately 30 nsec across the channel's bandwidth.

DOCSIS Group Delay Assumptions

CableLabs via the DOCSIS standards [2-5] has also made assumptions regarding the nature of GDV present in an HFC environment. DOCSIS compliant devices must interoperate under the estimated conditions illustrated in the Figure 26. The estimates shown in Figure 26 are based upon preliminary simulation and measurements of GDV and DOCSIS Pre-Eq interaction.

MAINTENANCE REQUIREMENTS

Understanding the point at which Pre-Eq will fail is the first step toward leveraging the diagnostic benefits of equalization coefficient analysis. Simulation and tests were performed of increasing single dominant micro-reflection impairment. The results of these tests reveal the highest micro-reflection impairment level that could be corrected by DOCSIS 2.0 Pre-Eq. 16-QAM and 64-QAM were both evaluated.

The test topology is illustrated in Figure 27. Seven amplifiers were cascaded with an optical link. The 6.4 MHz test signal was centrally located within a 5 – 40 MHz return path spectrum at 16 MHz center frequency, in order to minimize contributions from both the AD and GDV impairments contributed by the HFC network.

An Ethernet link was established between the subscriber side of the CM and the Wide Area Network (WAN) side of the CMTS. The Ethernet connectivity was continuously monitored as increases in micro-reflection impairment were introduced into the path between the CM and CMTS. Loss of Ethernet connectivity was assumed to be the point at which a High Speed Data (HSD) subscriber would log a service call with a cable operator.

Simulation and measurement for both 16-QAM and 64-QAM, illustrated in Figure 28 and Figure 29, reveal DOCSIS 2.0 mitigation of impairments is appreciably higher than what is assumed by DOCSIS to be present in the HFC environment. Additionally, there is a reduction of correction capability caused by a reduced decision boundary size as 16-QAM signals are migrated to 64-QAM. Note that this reduction is approximately 2 dB on average and not the 6 dB expected from QAM and AWGN impairment relationship previously discussed.

Simulation and test of increasing micro-reflection impairment were repeated with additional impairments, AD and GDV. AD and GDV contributions were increased by simply locating the test signal near the upper band edge of a 5 – 40 MHz return path spectrum, 36.8 MHz center frequency. Figure 30 and Figure 31 are measurements of the AD and GDV present at the upper band region of 5 – 40 MHz return path spectrum of the test topology illustrated in Figure 27. These results, illustrated in Figure 32 and Figure 33 for 16-QAM and 64-QAM respectively, reveal a negligible change in correction capability of the DOCSIS 2.0 Pre-Eq even with the additional impairments.

Overall, the results shown in Figures 28 through 33 are significant for the following reasons:

1. Discussed impairment levels can exceed DOCSIS HFC environmental assumptions and still be corrected by DOCSIS 2.0 Pre-Eq
2. Simulation results closely agree with laboratory measurement
3. Micro-Reflection impairment impact on modulation complexity is different from AWGN impairment impact

The HFC environment is dynamic in nature, with causes including changing loading conditions, component decay, weather, and routine maintenance practices. Allowing sufficient margin for this variation will allow the HFC environment to breath. However, exploiting the limits of acceptable performance and maintaining margin will optimize maintenance costs while also minimizing service calls.

In order to define the acceptable performance limits, simulation and measurement are necessary. However, simulation may bare the burden of exploring impairment permutations while minimizing the cost of testing resources. For example, the impact of multiple micro-reflection impairments can be studied and defined through simulation to establish acceptable performance limits which can then be verified in the laboratory and HFC environment.

Study of the single dominant micro-reflection and a combination of micro-reflection, AD, and GDV impairments has defined acceptable performance limits and behavior that is clearly different than assumptions.

Continued investigation of impairments and combinations thereof can complete the acceptable performance limit requirements of DOCSIS 2.0 Pre-Eq. Simulation can be leveraged to help manage the cost of defining these limits. And finally, an understanding of the impairment limits and relationships with modulation complexity will help cable operators define maintenance requirements and transition toward optimal upstream throughput by minimizing cost and service calls.

ISOLATION PROCESS

There may be many ways to take advantage of the wealth of information provided by Pre-Eq to isolate DOCSIS Pre-Eq related impairments. The following process has been proposed in order to isolate impairments using equalization coefficient analysis.

Step 1

Ensure that majority of DOCSIS links are supporting at least DOCSIS 2.0 with Pre-Eq enabled. The resolution of the 24-tap equalizer of DOCSIS 2.0 is better suited to identify impairments, compared to the 8-tap equalizer of DOCSIS 1.1.

Step 2

Query the DOCSIS 2.0 CM population using an SNMP query tool similar to *Modem Pre-Eq Response Tool*, illustrated in Figure 34. The Modem Pre-Eq Response Tool, which is software developed by Marc Morrissette of Motorola, has many useful features, the most important being the ability to query multiple DOCSIS terminal devices based on an IP address list. Periodic polls of coefficient values and other relevant physical layer (PHY) metrics are displayed and/or stored into a log file for post processing. This tool also provides users with a graphical view of either the impulse response or amplitude response for each CM poll.

Using tools like the Modem Pre-Eq Tool can help cable operators establish a baseline of performance, and identify problem DOCSIS links, based on CM IP addresses, within the HFC plant.

Step 3

Identify impairment problems by sorting, on increasing levels that sum the DOCSIS Pre-Eq regions previously defined. For example, determine which CMs experience the greatest amount of micro-reflection impairment by sorting on the levels which result from summing the taps located in the post-tap region.

Step 4

Understand the relevant Pre-Eq impairment problems, their impact, and how they originate in HFC plant. For example, one micro-reflection source has been discussed in the micro-reflection fundamentals section, but many possible permutations of micro-reflection sources must be understood for successful isolation.

Use results such as those in the maintenance requirements section to define what impairment levels will likely result in service calls, and consider these for potential areas to address proactively

Through repeated application of an isolation process, understand how much margin below service outages would optimize the cost/benefit ratio of proactive maintenance.

Step 5

Leverage the CM population to differentiate between CMs experiencing an impairment problem and those that are not. For example, a query of the CM population of the HFC coaxial feeder segment illustrated in Figure 35 reveals that CMs located off of amplifier 1 are reporting a micro-reflection problem, while CMs located off of amplifiers 2, 3, and T1 are not reporting a problem.

Step 6

Identify suspected HFC components using results from Step 4 and Step 5. As previously discussed, micro-reflection sources consist of pairs of HFC components separated by a distance of cable. Again referring to Figure 35, none of the CMs upstream from amplifier 1 are reporting a micro-reflection problem. Additionally, diplex filters make amplifiers located between suspected HFC components unlikely. Therefore, all of the HFC components fed from amplifier 1 are suspect.

Step 7

Inspect and repair as necessary all suspected HFC plant components resulting from Step 6. Again, referring back to the example provided in Figure 35. Inspection of the suspect HFC plant components show the micro-reflection source to be a combination of tap-to-output port isolation loss and an unterminated cable splice at the end of the amplifier 1 feeder run. This combination is what is contributing to the micro-reflections experienced by CMs 1A, 1B, and 1C. Properly terminating the splice reduces the micro-reflections to negligible amplitudes.

Step 8

Repeat CM population query multiple times and compare to baseline captured in Step 1 to ensure that the impairment problem has been eliminated and the improvements are sustainable.

CONCLUSION

The DOCSIS Pre-Eq function has enabled operators to deliver yet higher speeds of upstream service to the subscribers. The use of higher order modulation, with a strong push from DOCSIS tools such as equalization has made this a reality. However, the modulation order and bandwidth come at the expense of increased sensitivity to common HFC impairments. Not only does the Pre-Eq function ensure these higher speed links are robust, it also provides a wealth of insight into important plant characteristics. The type of characteristics is many of those that are of increased relevance as the upstream modulation complexity increases.

By fully understanding the Pre-Eq function and deploying some simple tools to perform equalization coefficient analysis on the data gathered by this function, it is possible to identify the dominant impairments for which the DOCSIS 2.0 Pre-Eq is compensating.

Single dominant micro-reflection will mostly impact the post-tap region of the impulse response magnitude, revealing the amplitude and delay characteristics of the micro-reflection source.

Dominant AD will mostly impact the main-tap region of the impulse response magnitude as well as the amplitude response.

Dominant GDV will mostly impact the main-tap region of the impulse response magnitude as well as the phase response.

Simulation and measurement are required to determine all of the points at which DOCSIS 2.0 Pre-Eq will work under various impairment combinations and levels.

Understanding the limits of DOCSIS 2.0 Pre-Eq will help cable operators establish when proactively maintaining the HFC plant will be most beneficial, leading to a refined process that helps cable operators leverage the benefits of DOCSIS Pre-Eq coefficient analysis.

REFERENCES

[1] Digital Communications; Fundamentals and Applications, 2nd Edition, ISBN 0-13-084788-7, 2001, Bernard Sklar, www.phptr.com

[2] Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification v1.0, ANSI/SCTE 22-1, 2002, Society of Cable Telecommunications Engineers, www.scte.org

[3] Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification v1.1, SP-RFIv1.1-C01-050907, September 7, 2005, Cable Television Laboratories, Inc., www.cablemodem.com

[4] Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification v2.0, CM-SP-RFIv2.0-I11-060602, June 2, 2006, Cable Television Laboratories, Inc., www.cablemodem.com

[5] Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification v3.0, CM-SP-PHYv3.0-I05-070803, August 3, 2007, Cable Television Laboratories, Inc., www.cablemodem.com

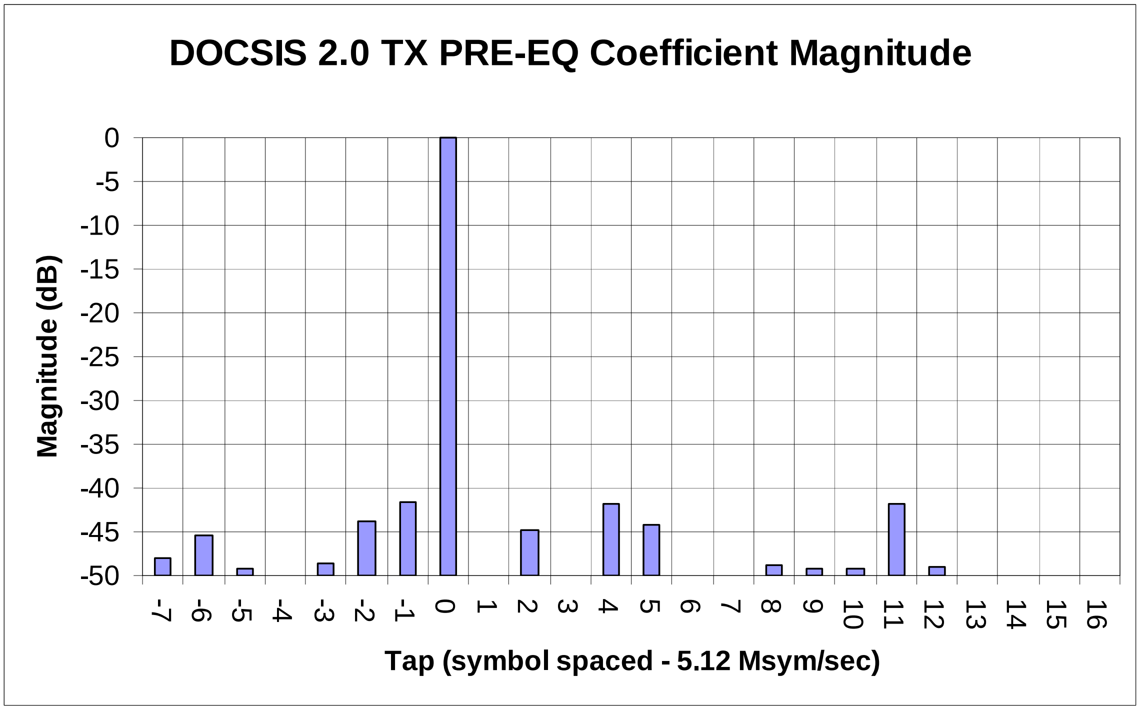


Figure 1 – Negligible Impairment – Impulse Response Magnitude

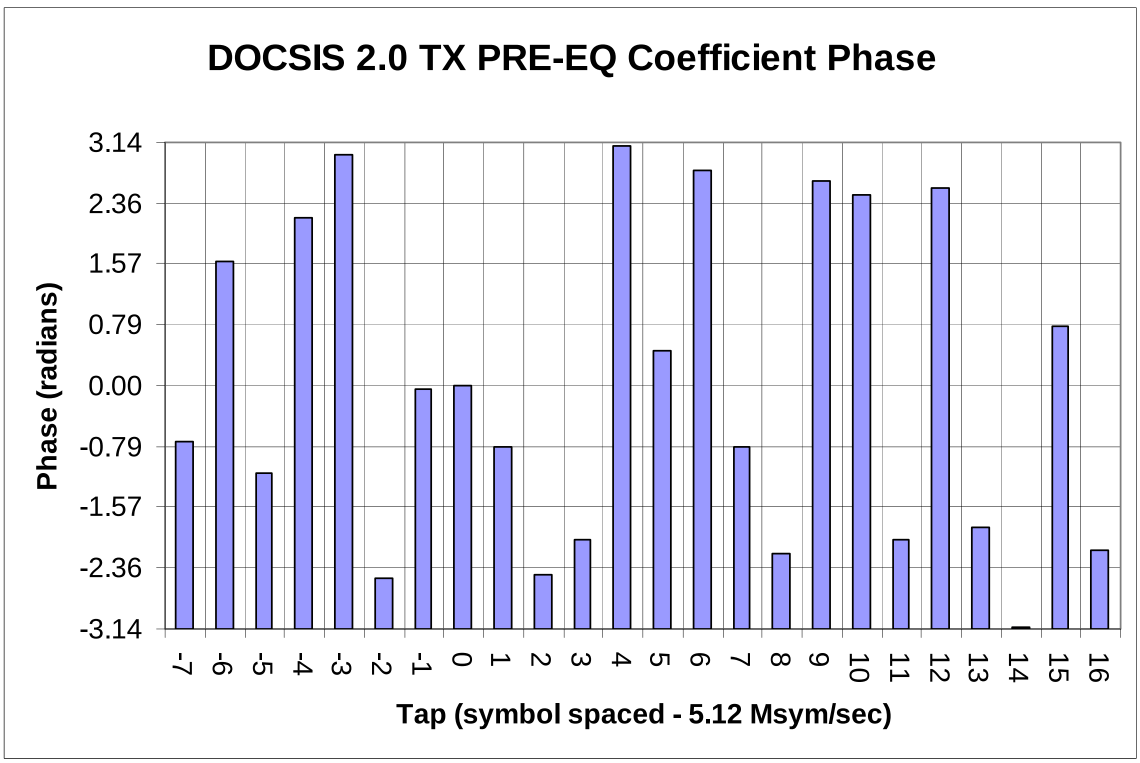


Figure 2 - Negligible Impairment – Impulse Response Phase

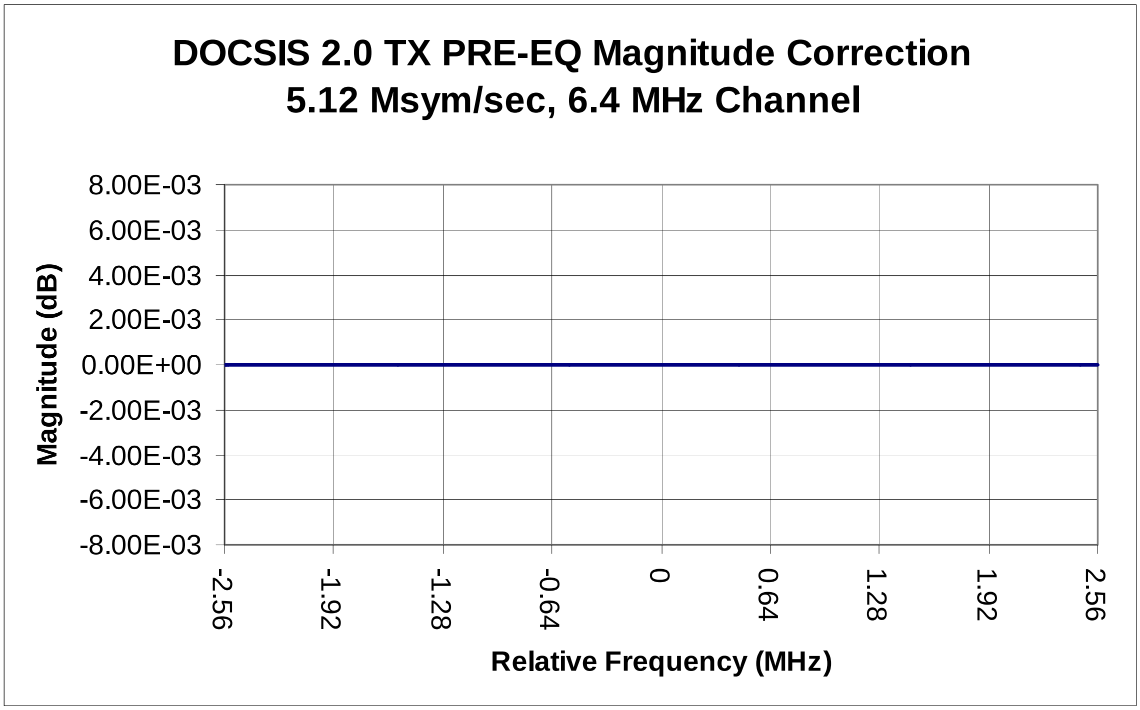


Figure 3 - Negligible Impairment – Amplitude Response

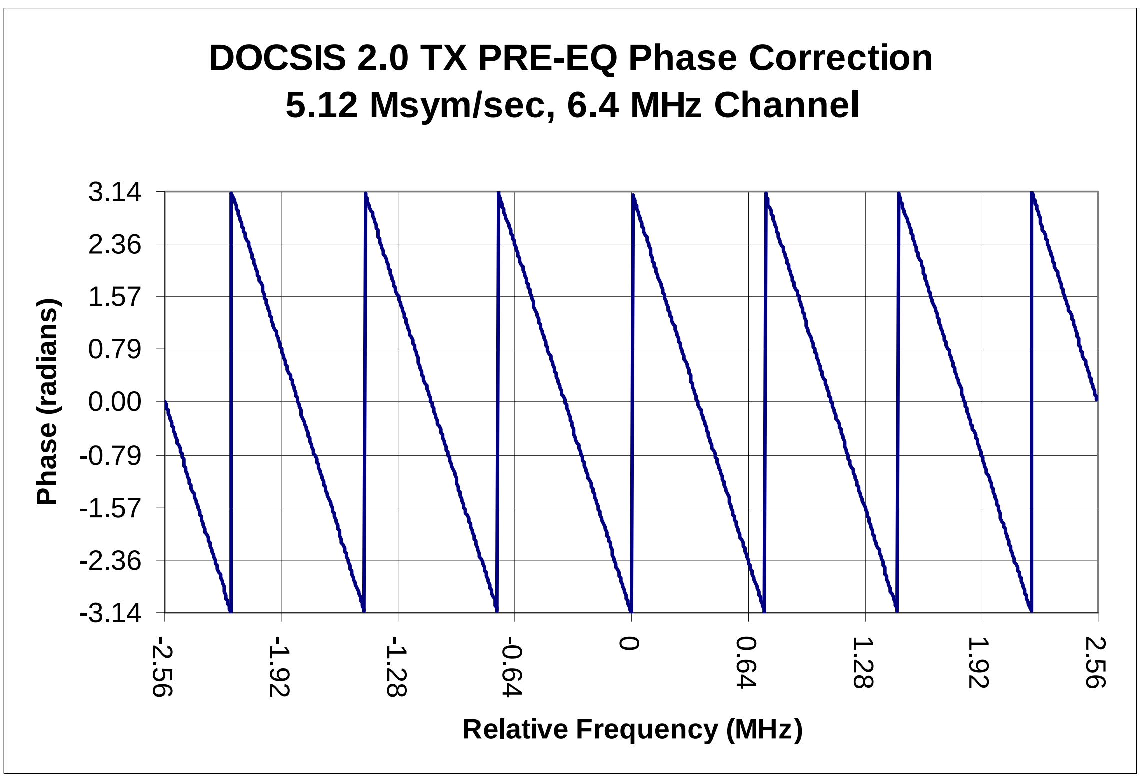


Figure 4 - Negligible Impairment – Phase Response

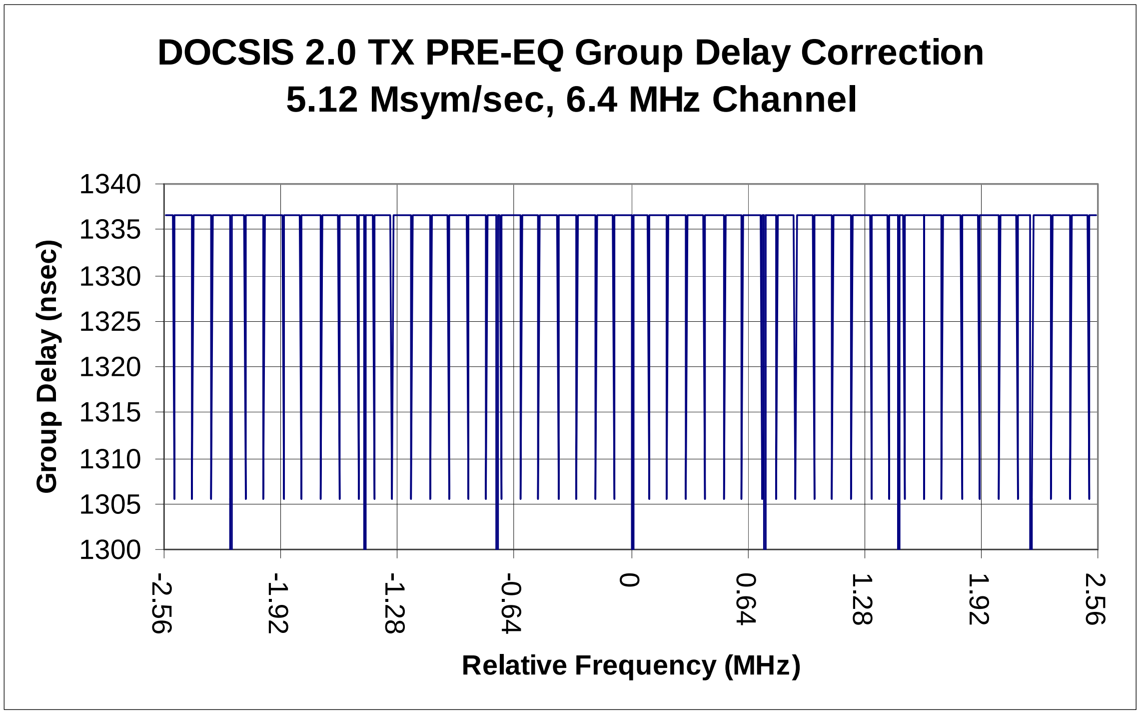


Figure 5 – Negligible Impairment – Group Delay Response

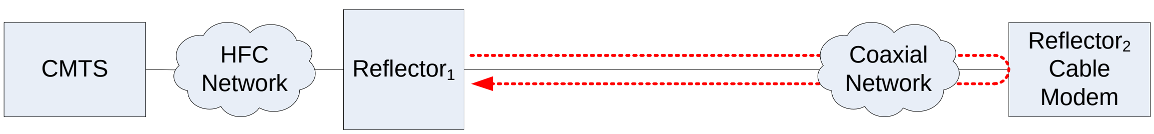


Figure 6 – Micro-Reflection Source

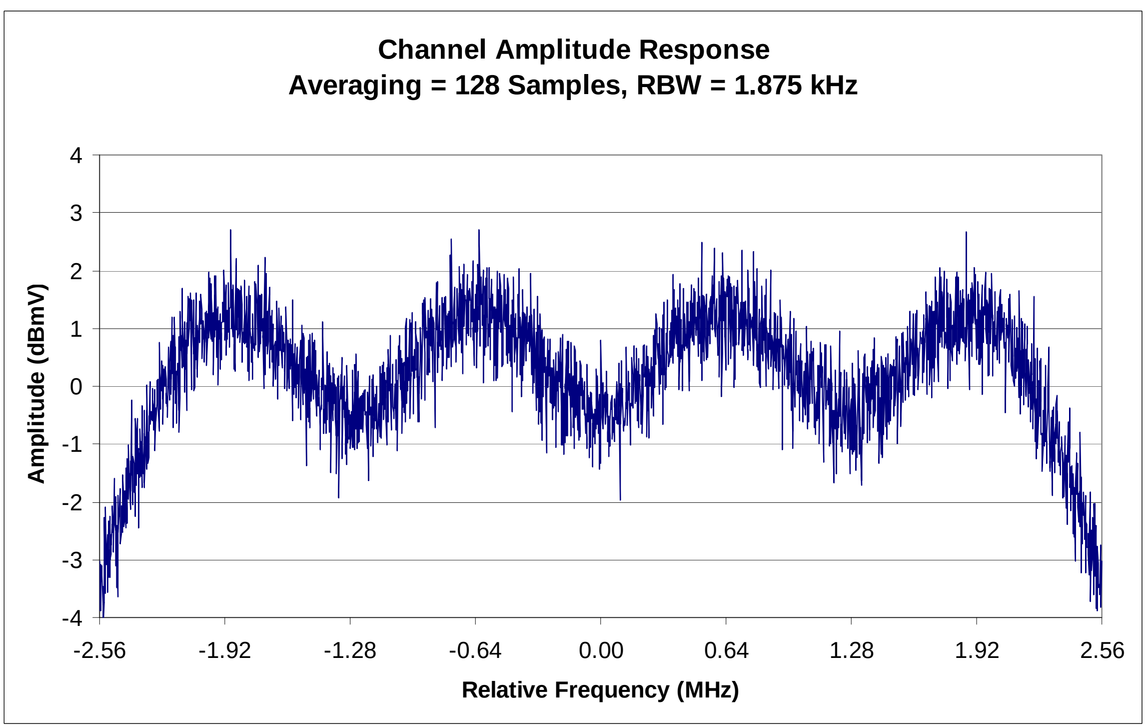


Figure 7 – Micro-Reflection Impairment – Channel Amplitude Response

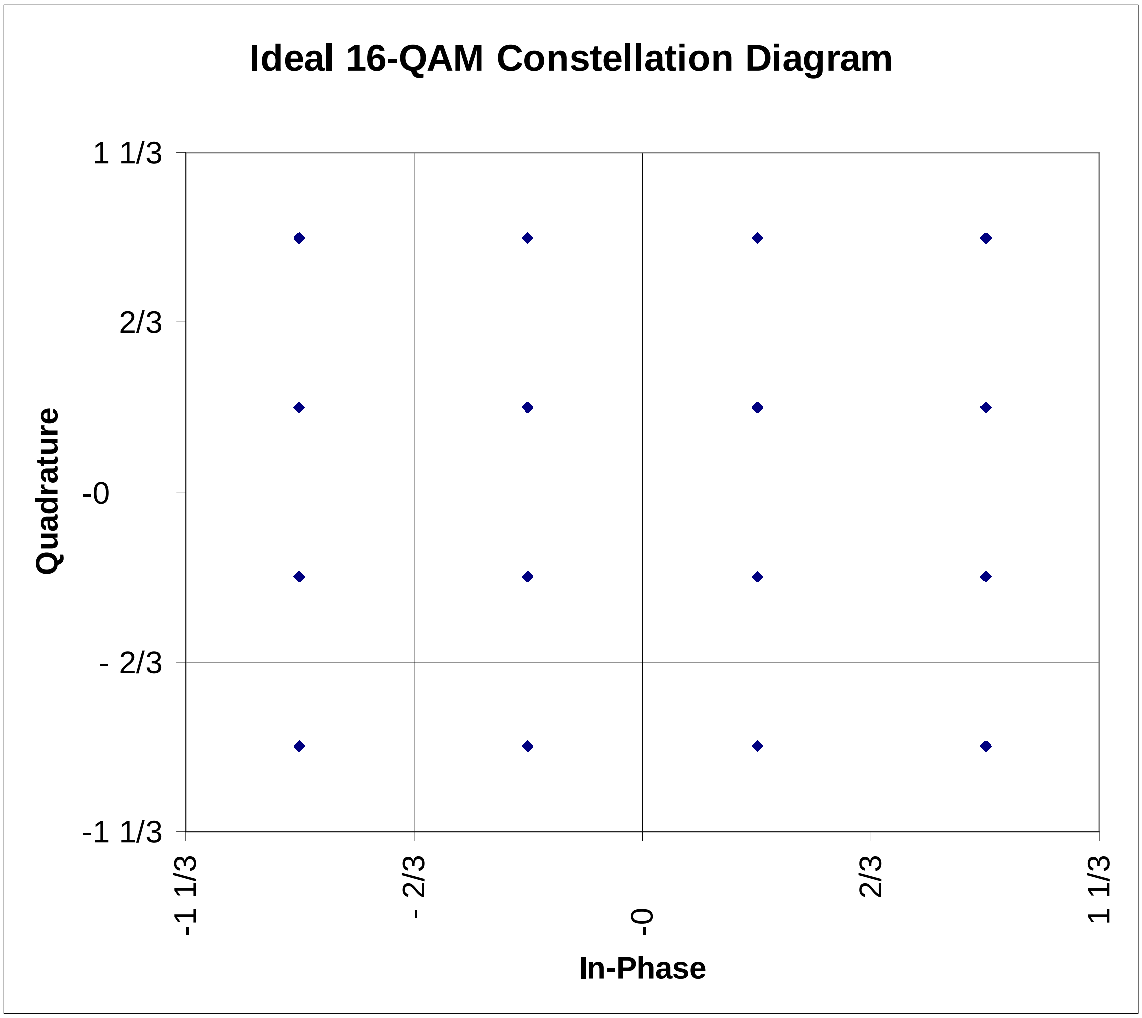


Figure 8 - Ideal 16-QAM Constellation

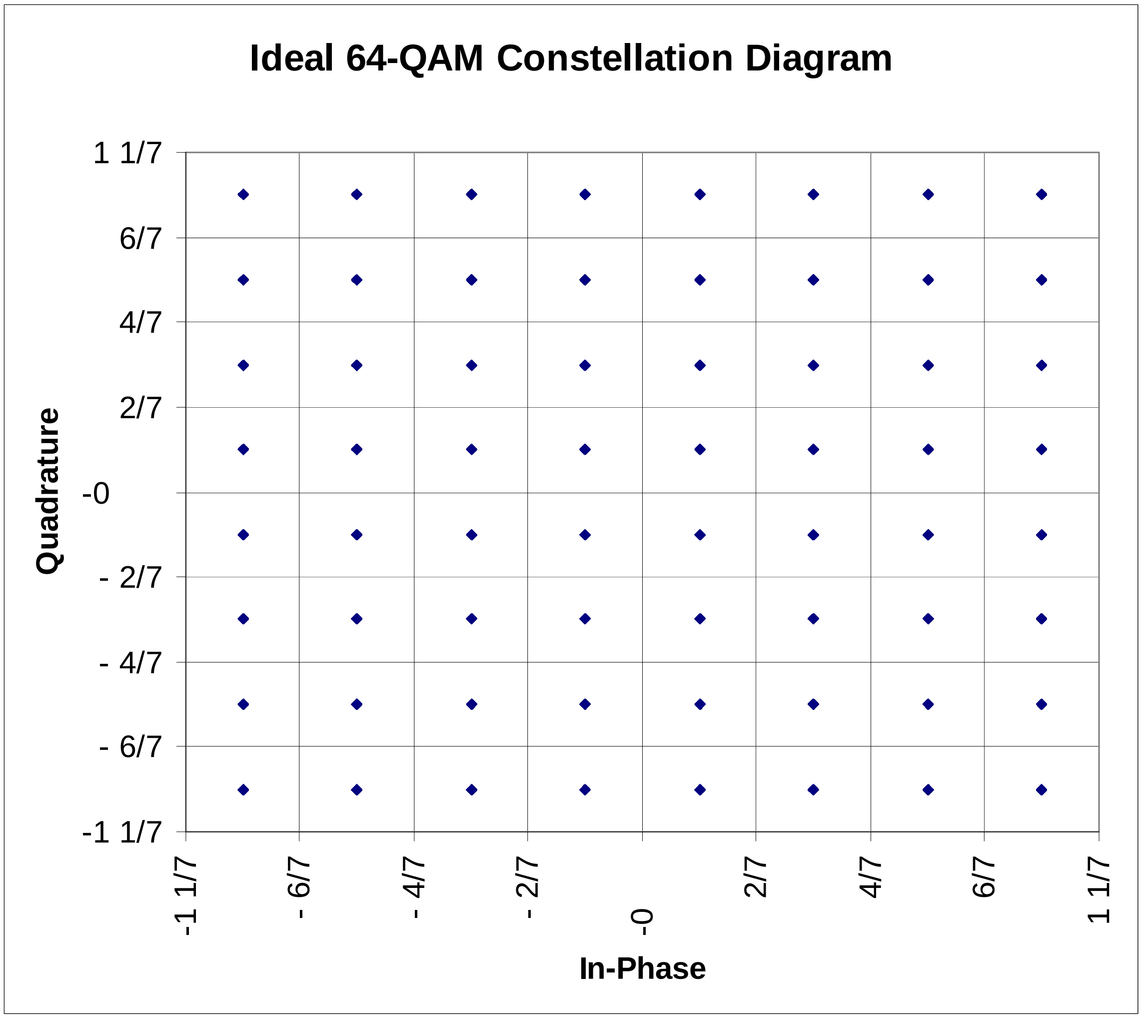


Figure 9 - Ideal 64-QAM Constellation

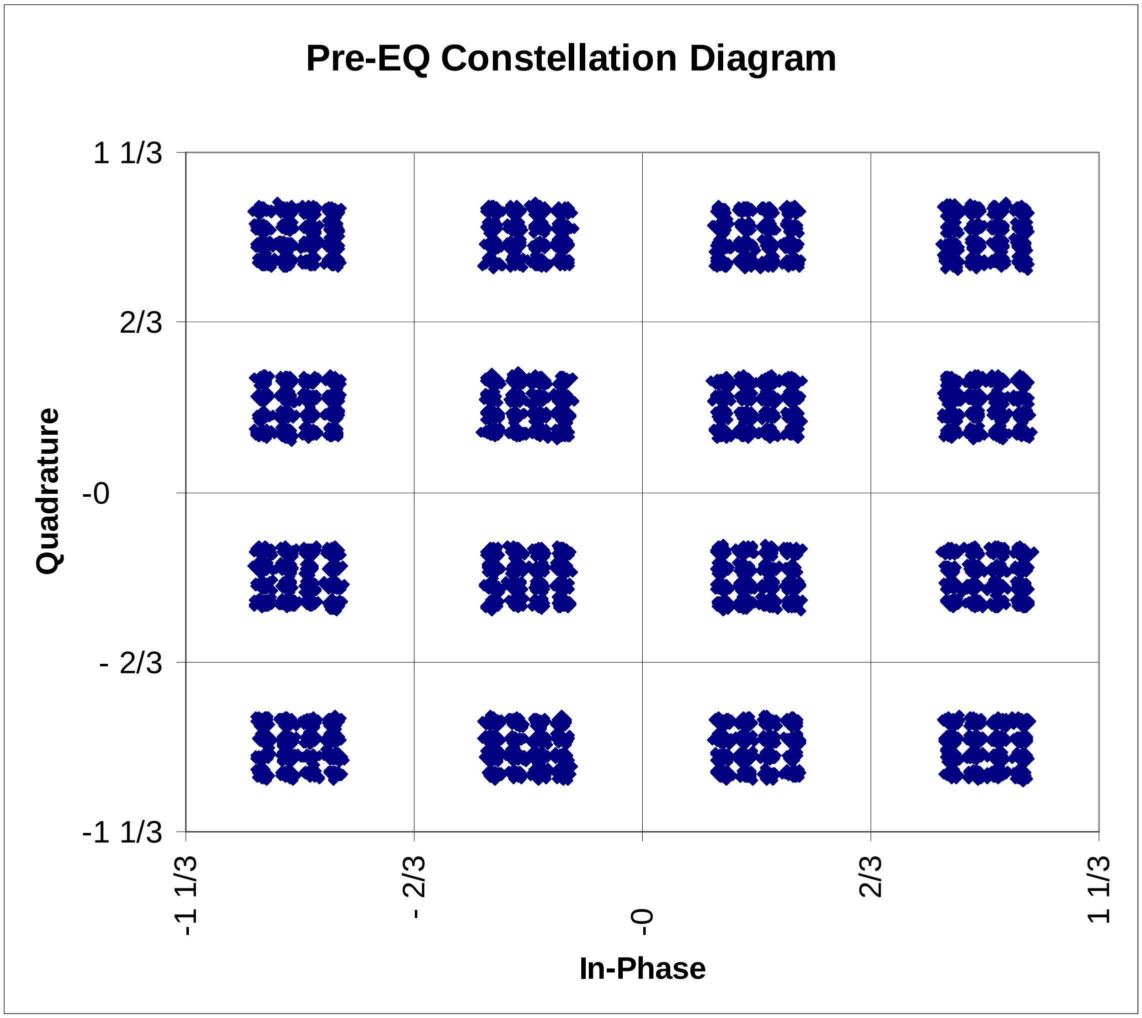


Figure 10 - Micro-Reflection Impaired 16-QAM Constellation

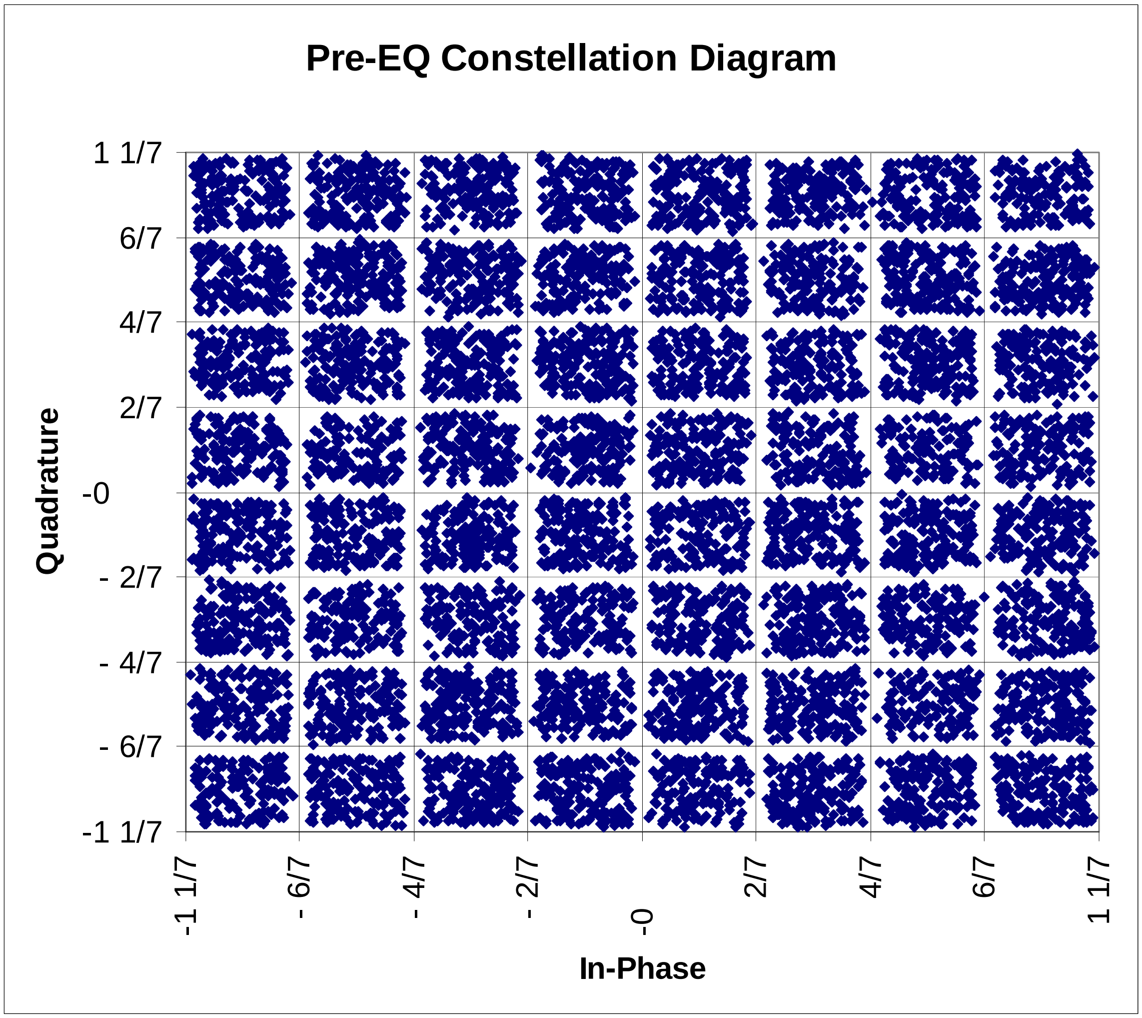


Figure 11 - Micro-Reflection Impaired 64-QAM Constellation

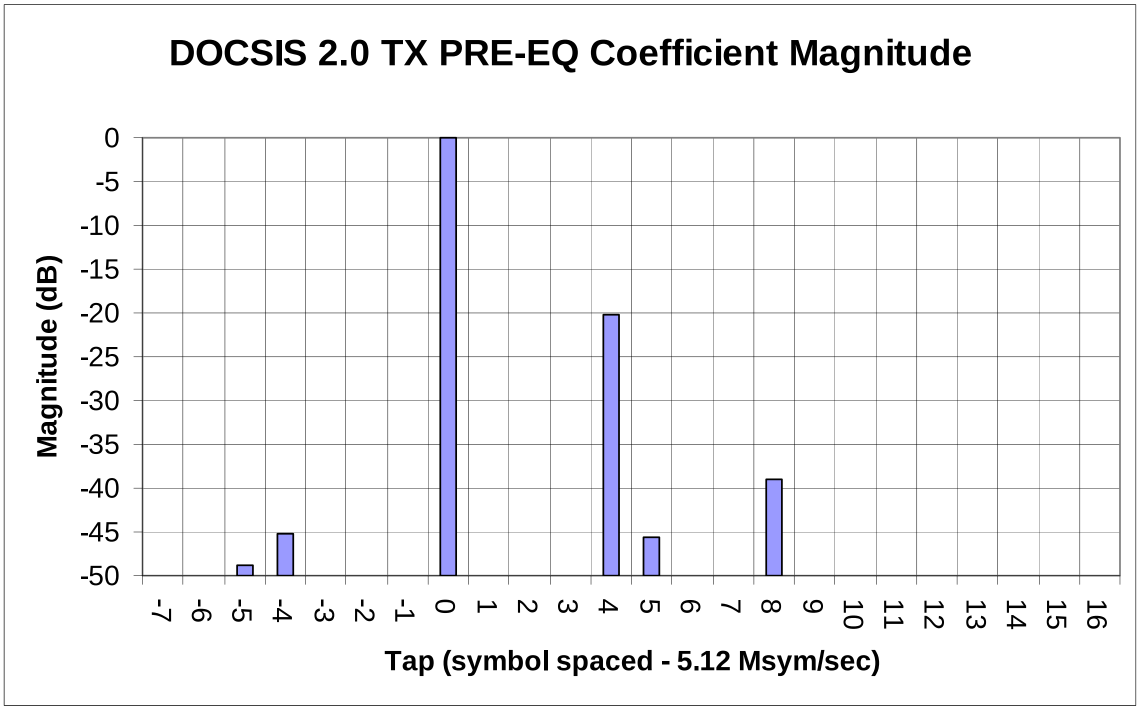


Figure 12 - Micro-Reflection Impairment –Impulse Response Magnitude

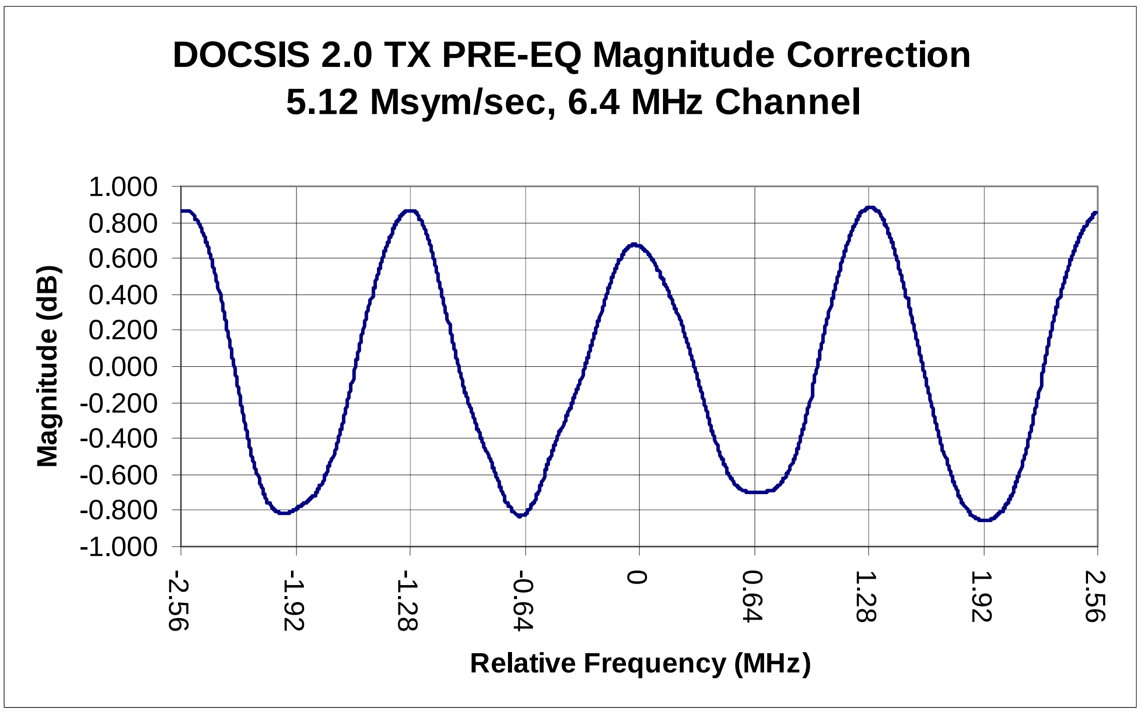


Figure 13 - Micro-Reflection Impairment – Amplitude Response

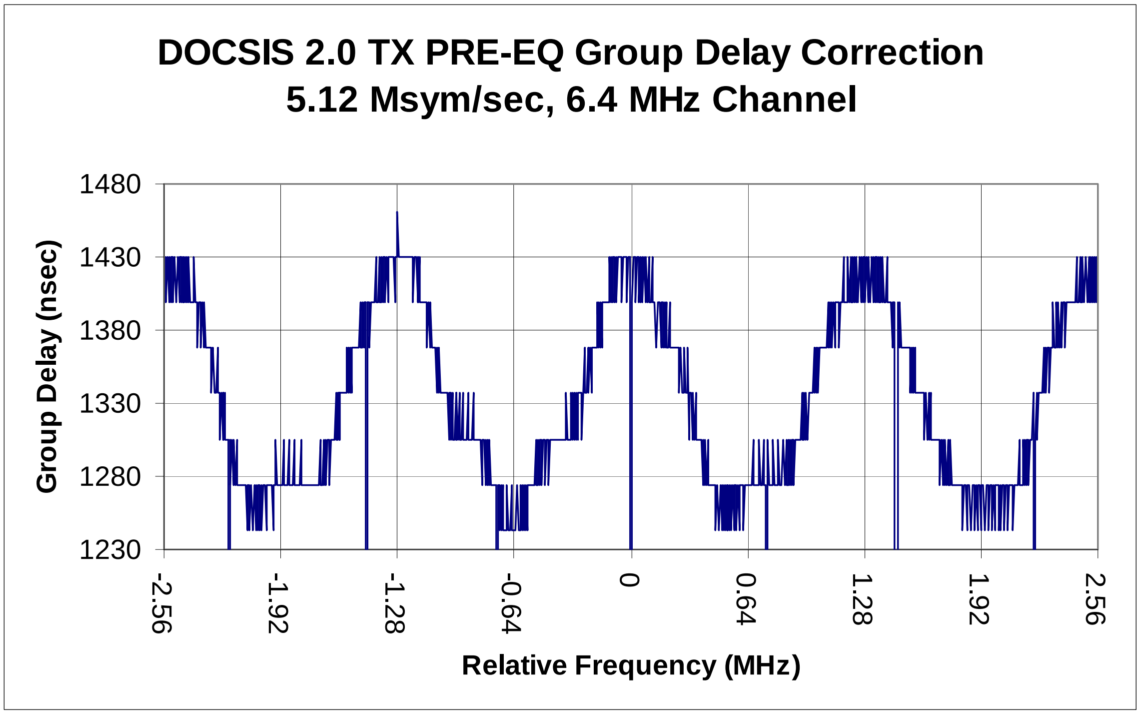


Figure 14 - Micro-Reflection Impairment – GDV Frequency Response

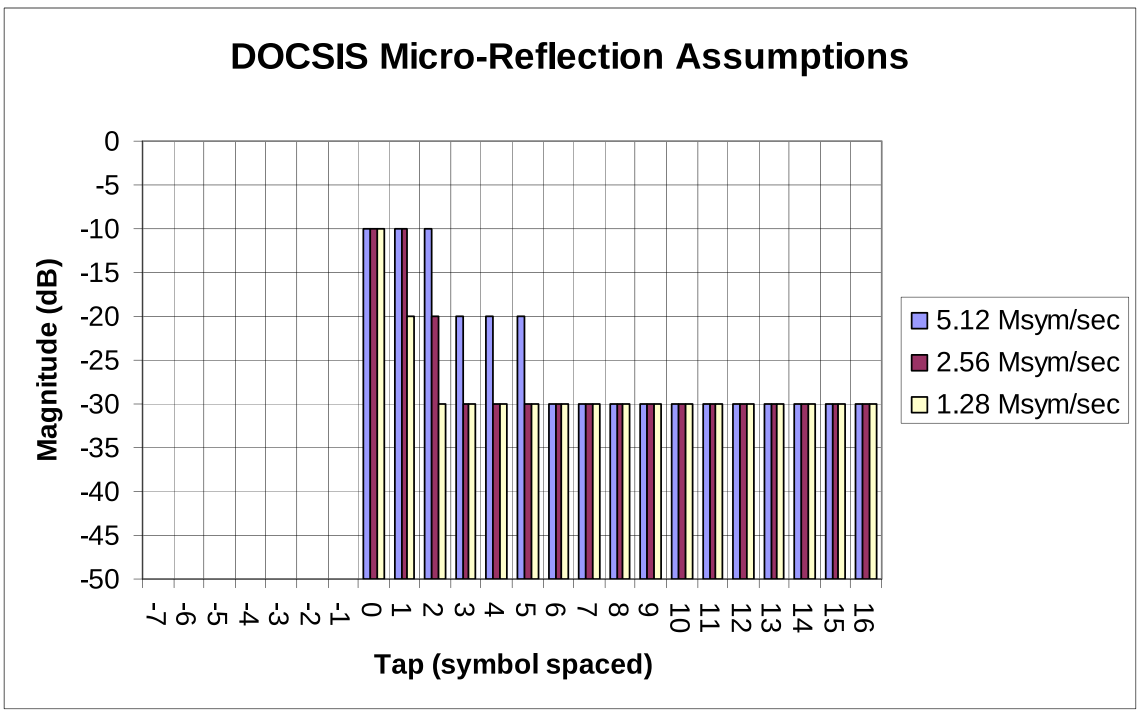


Figure 15 - DOCSIS Micro-Reflection Assumptions

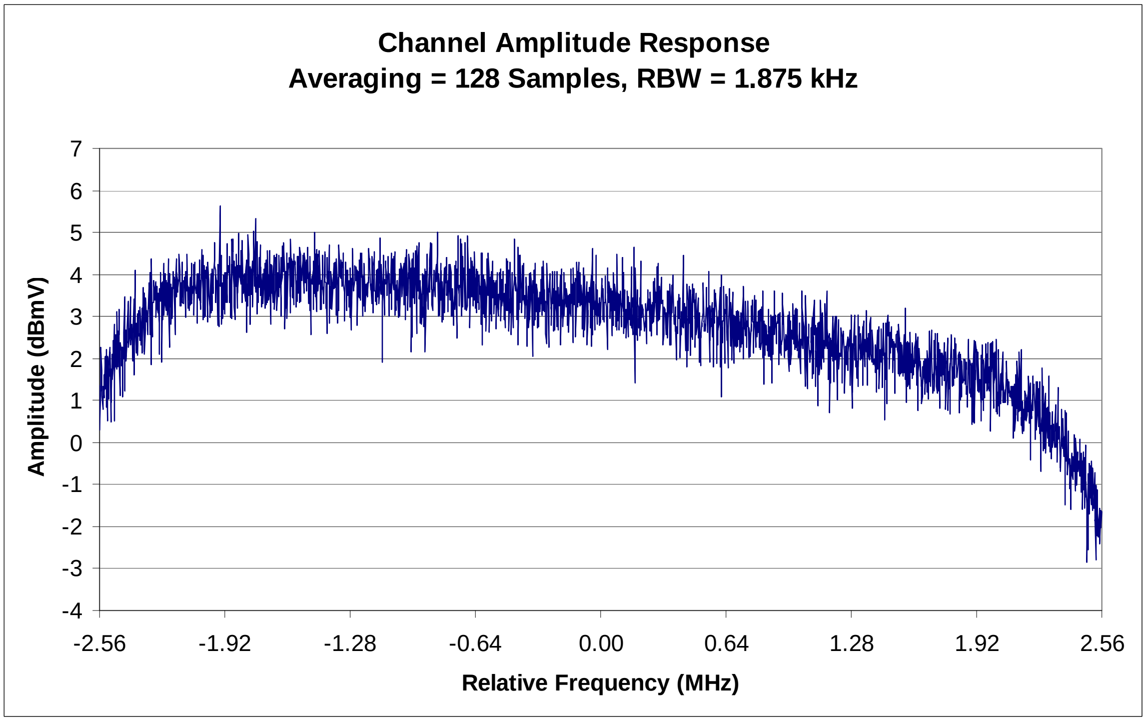


Figure 16 - Amplitude Roll-Off Impairment – Channel Amplitude Response

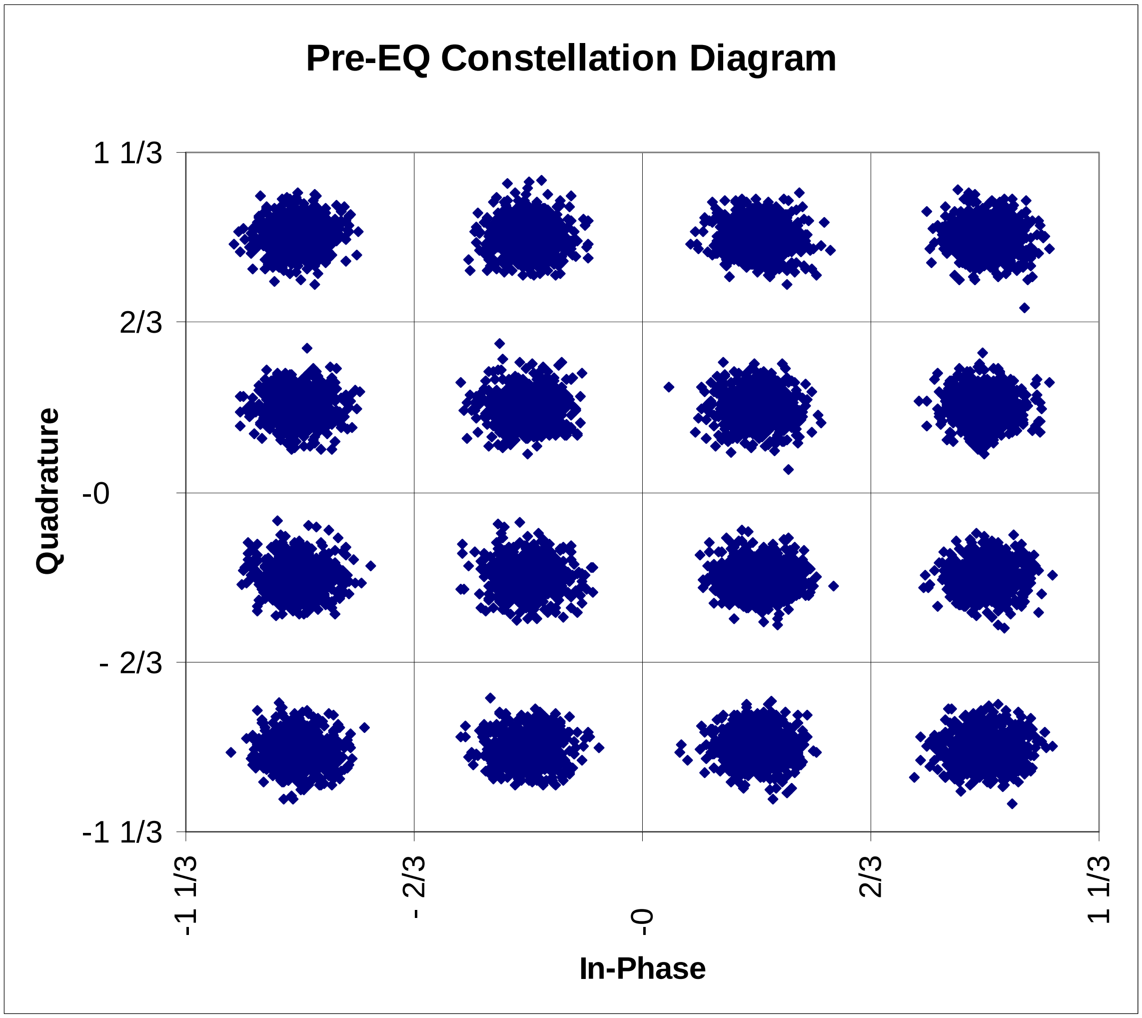


Figure 17 - AWGN Impaired 16-QAM Constellation

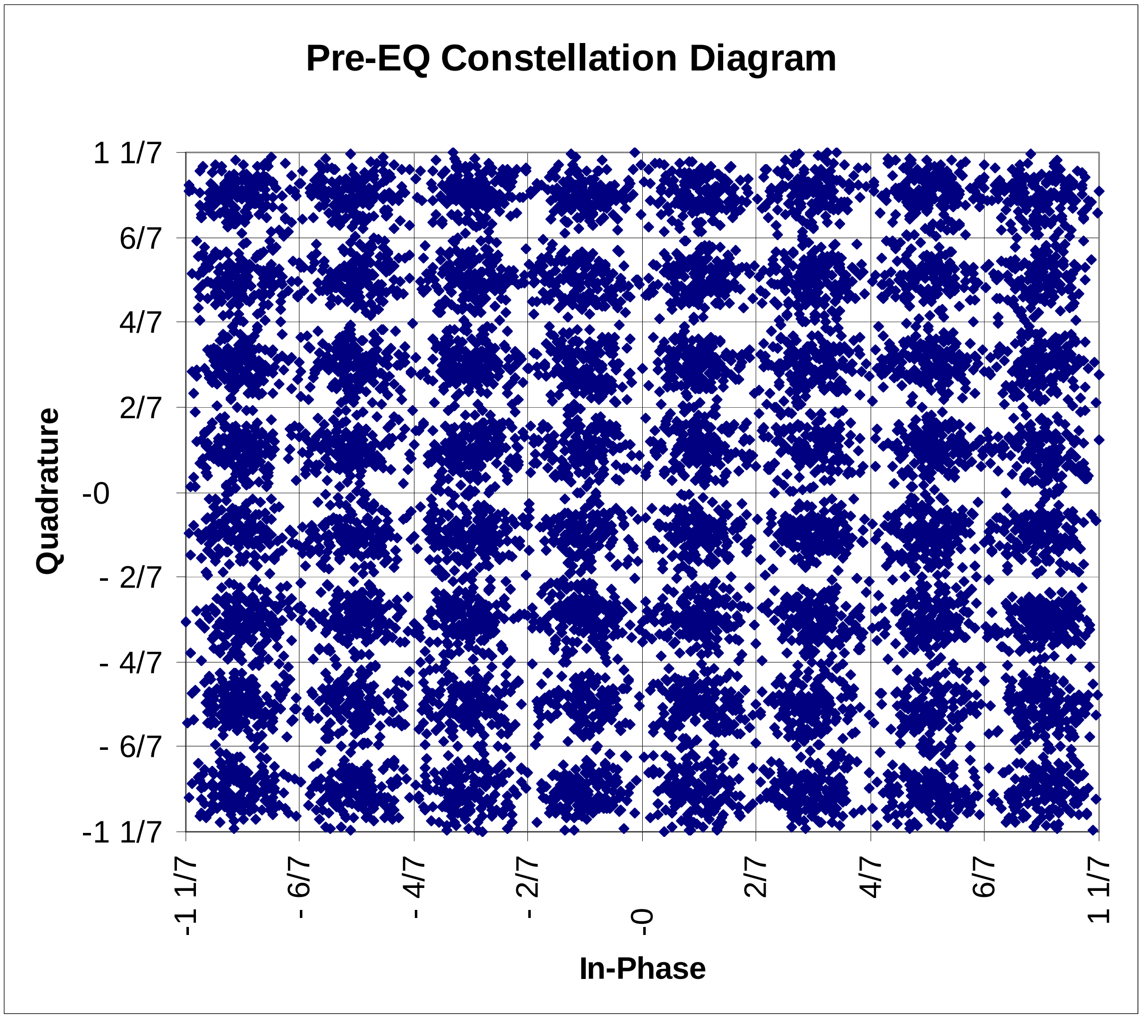


Figure 18 - AWGN Impaired 64-QAM Constellation

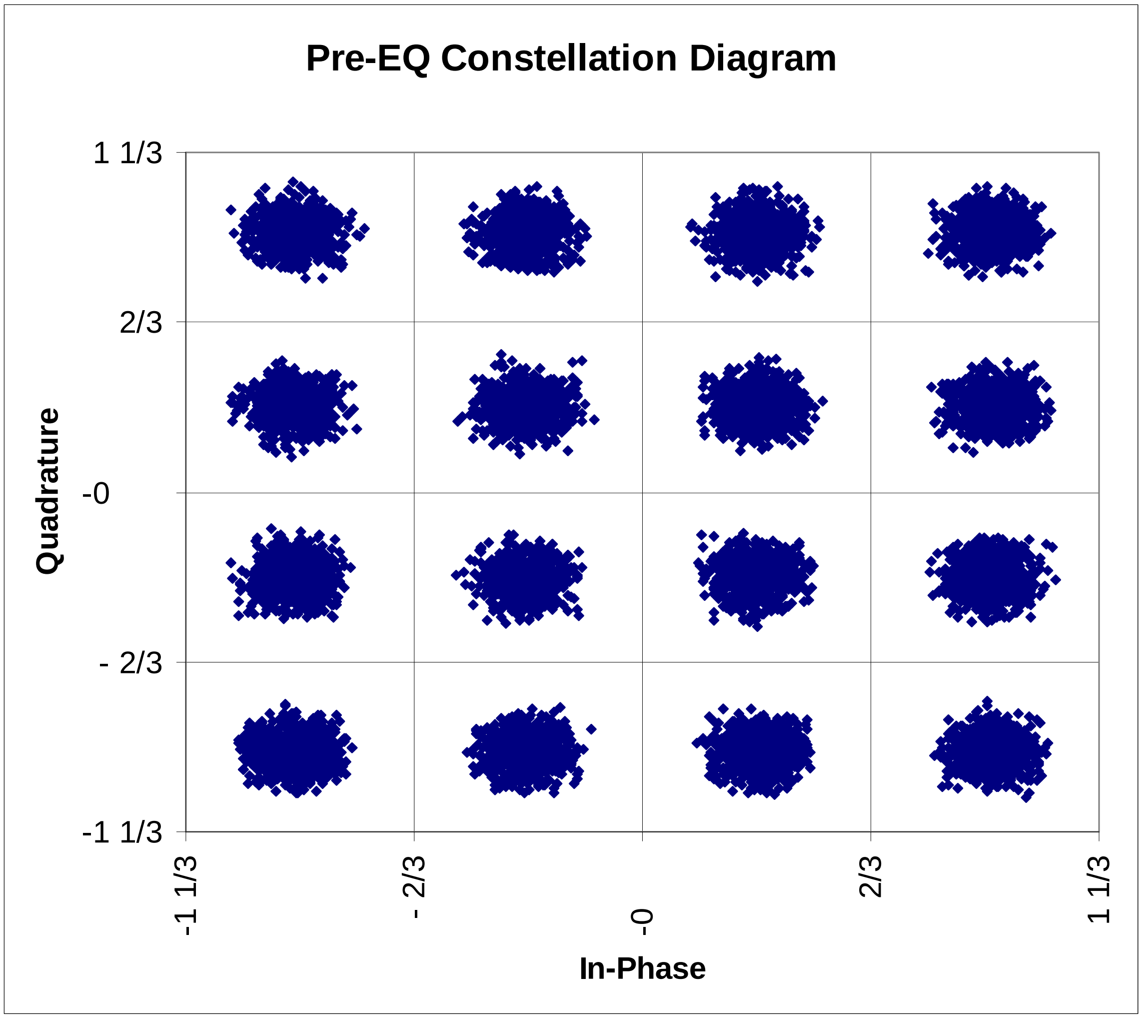


Figure 19 - Amplitude Distortion Impaired 16-QAM Constellation

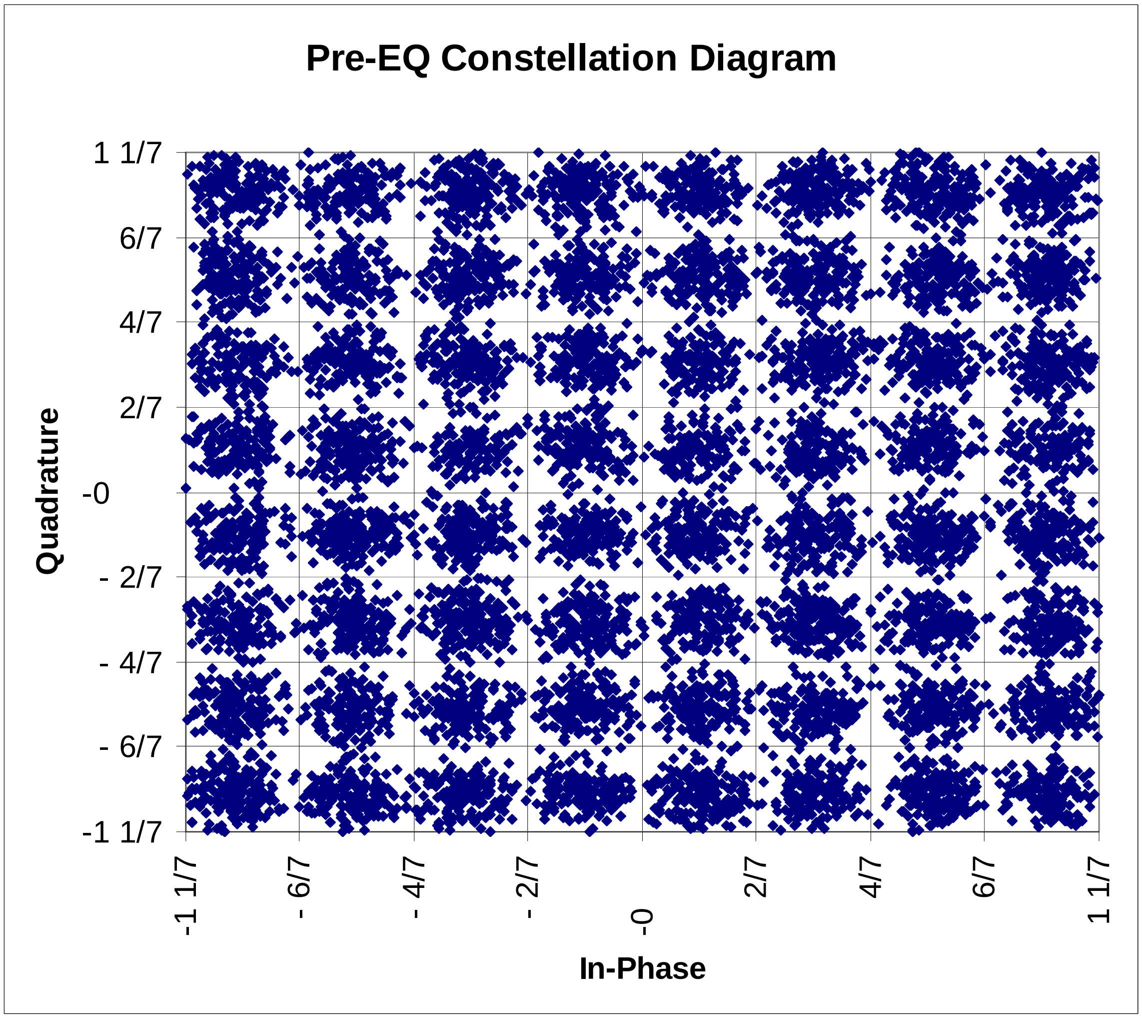


Figure 20 - Amplitude Distortion Impaired 64-QAM Constellation

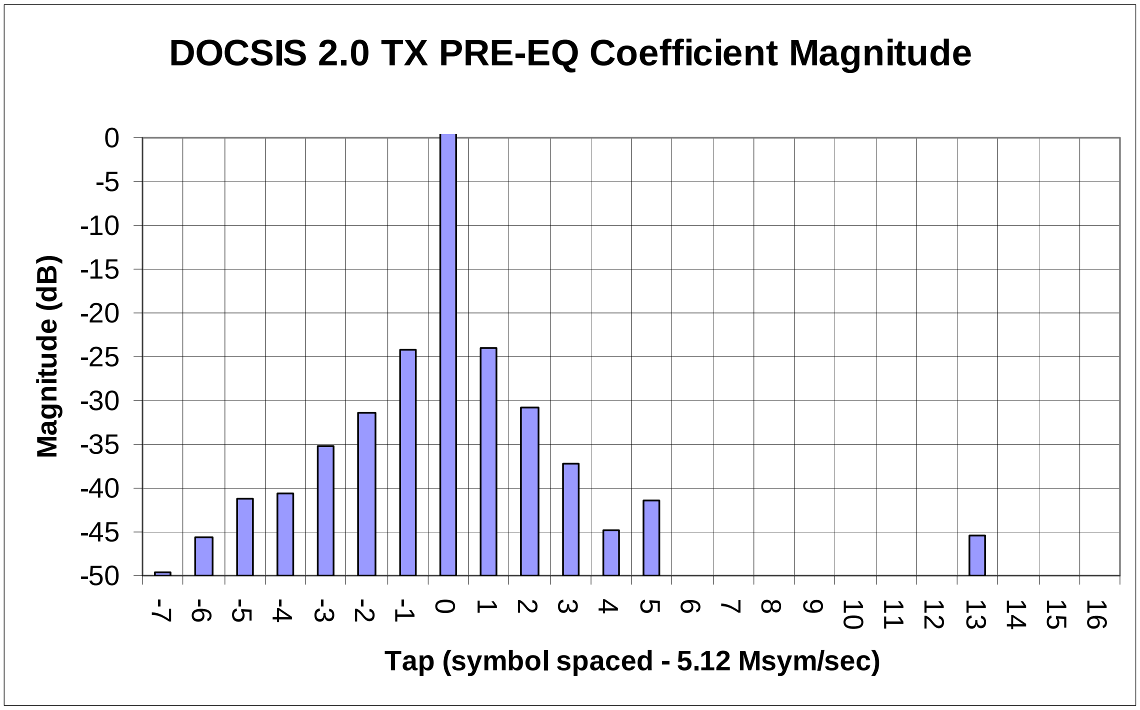


Figure 21 - Amplitude Distortion Impairment –Impulse Response Magnitude

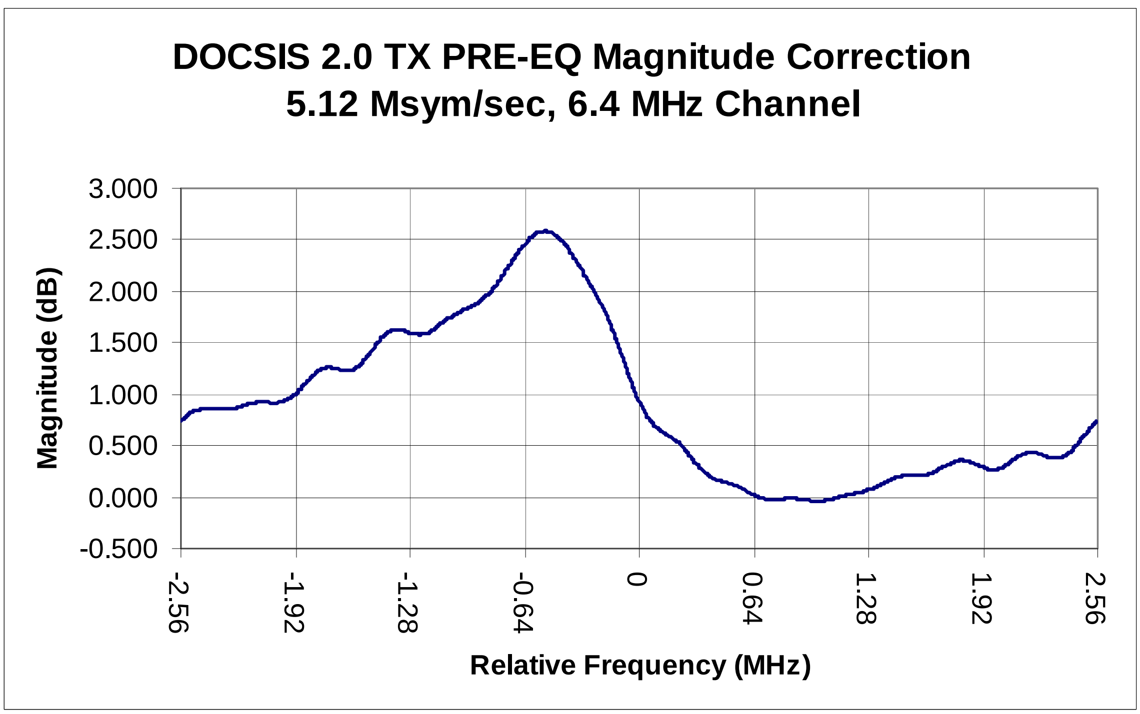


Figure 22 - Amplitude Distortion Impairment – Amplitude Response

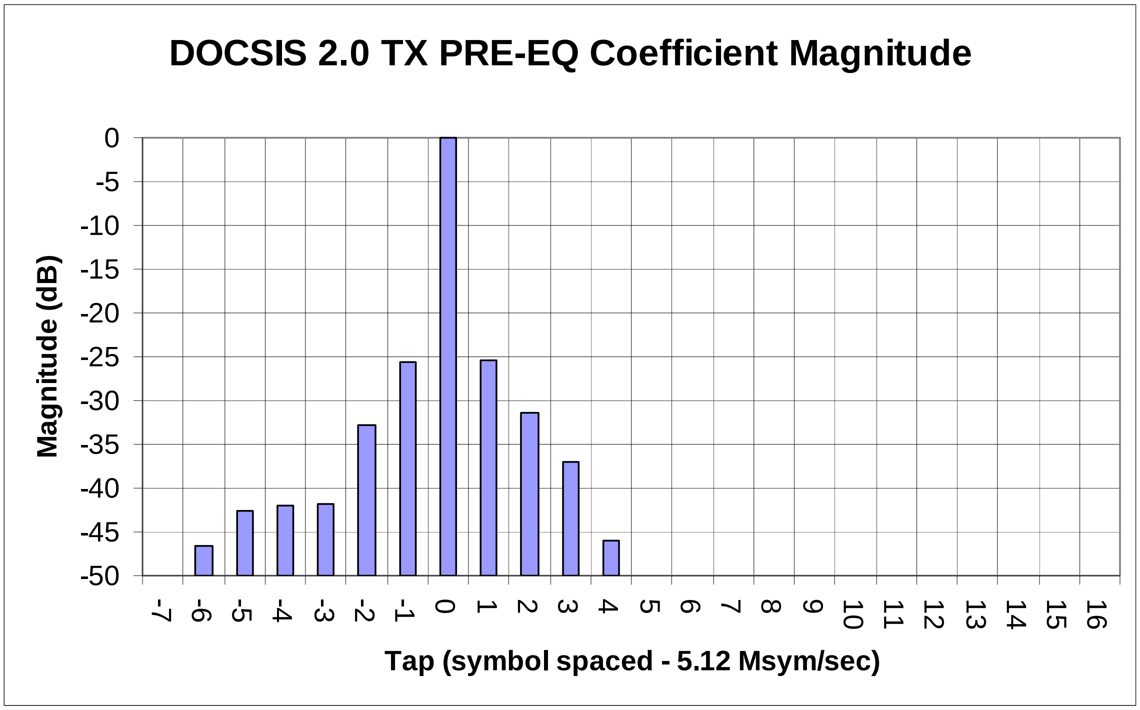


Figure 23 - Group Delay Variation Impairment - Impulse Response Magnitude

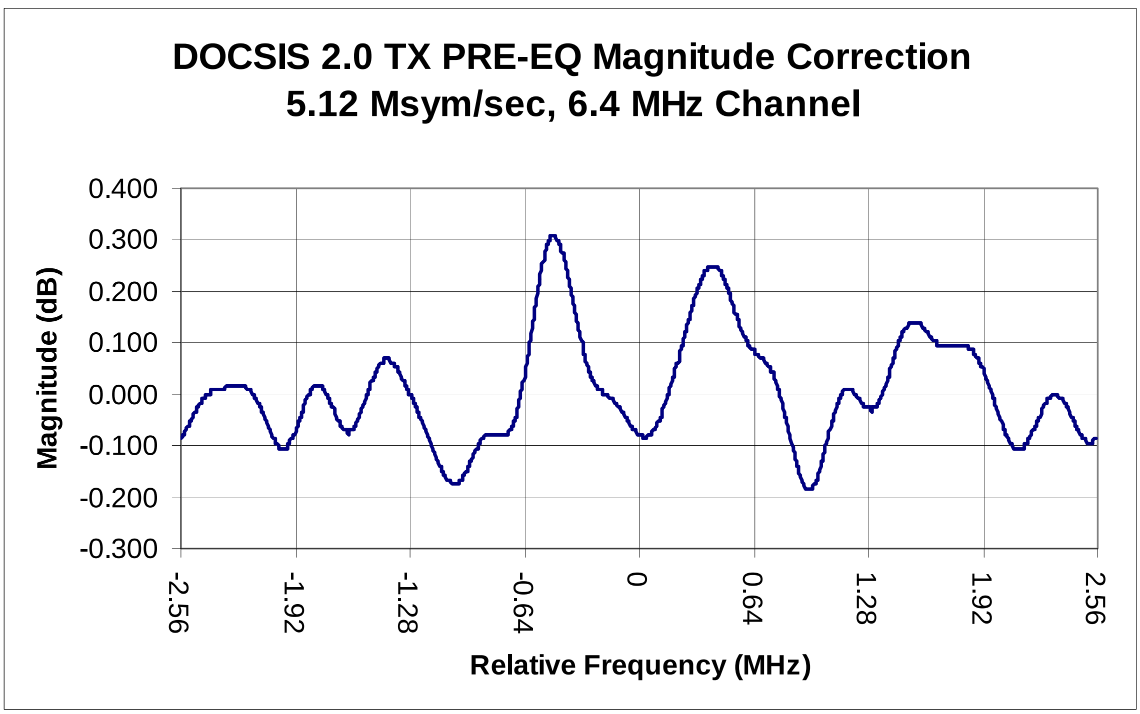


Figure 24 - Group Delay Variation Impairment - Amplitude Response

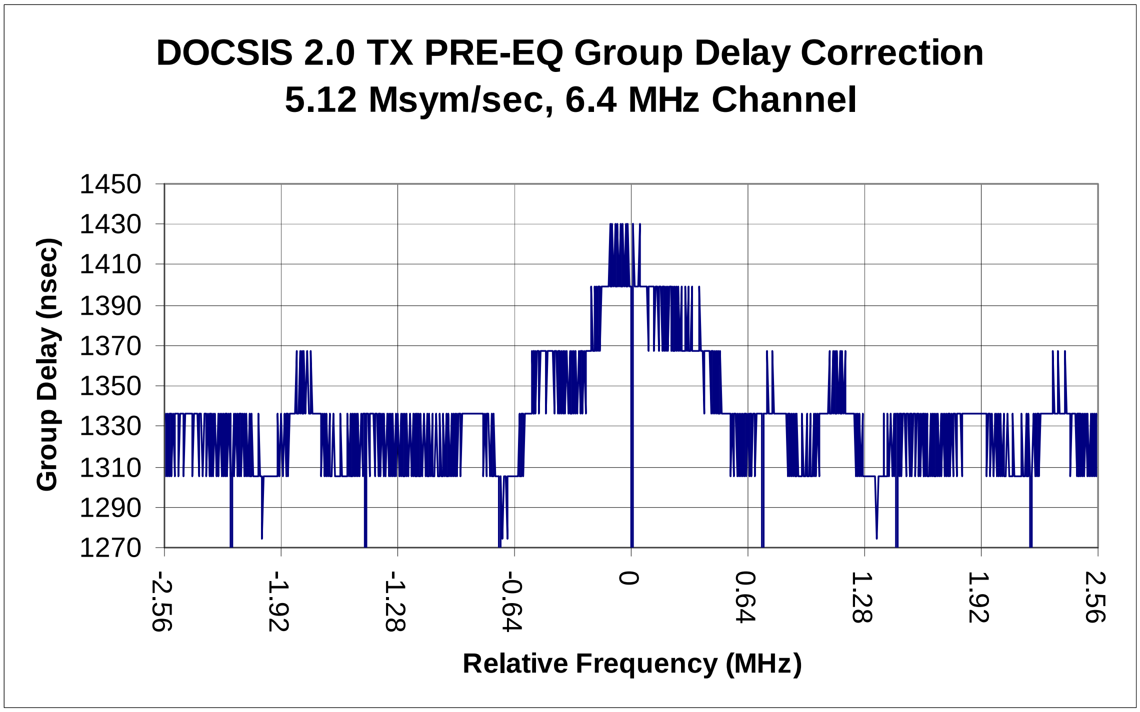


Figure 25 - Group Delay Variation Impairment - Group Delay Response

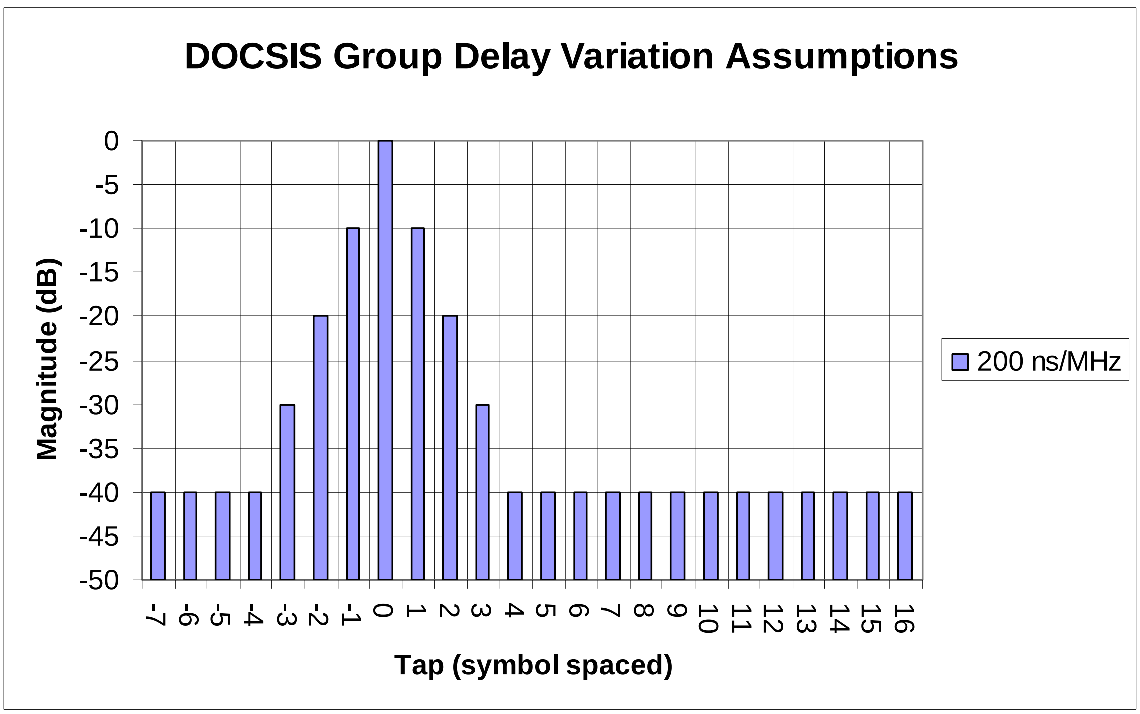


Figure 26 - DOCSIS Group Delay Variation Assumptions

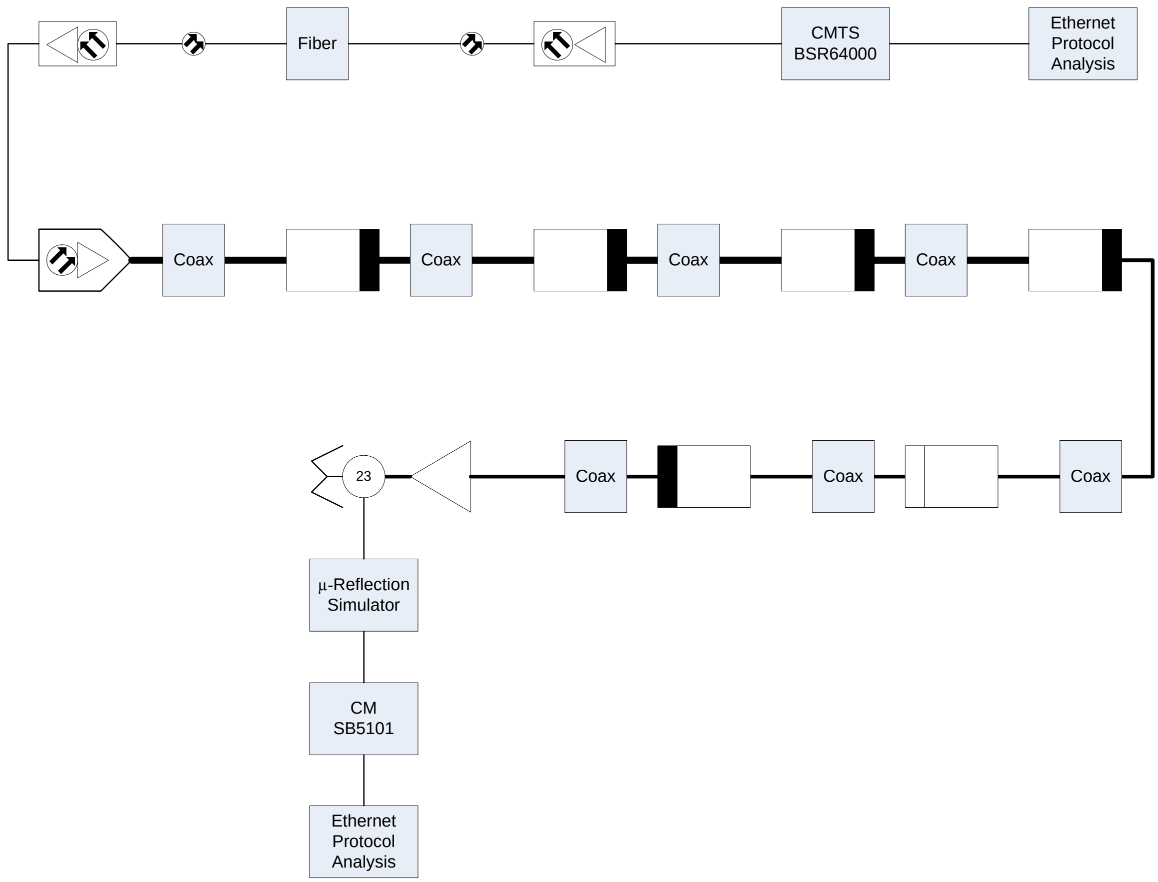


Figure 27 - Micro-Reflection Impaired Communication Channel Test Topology

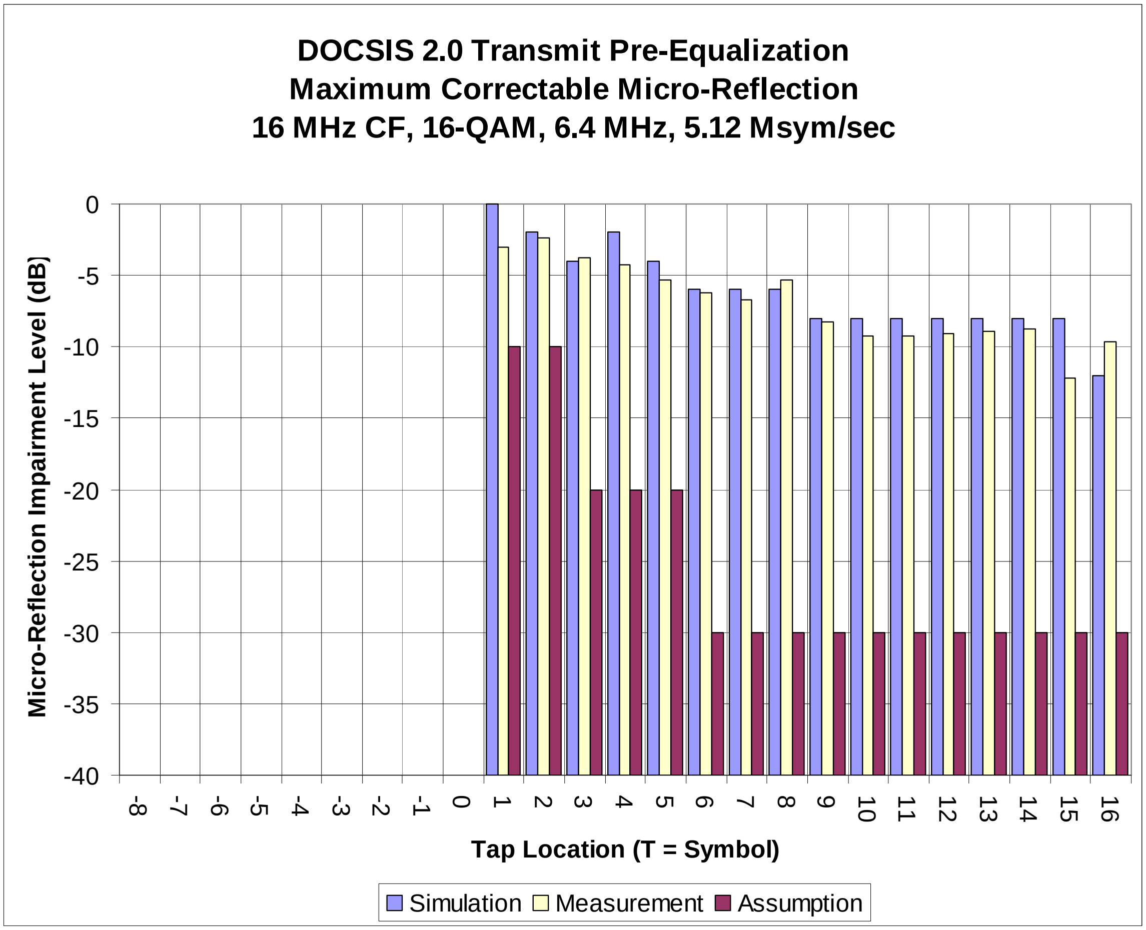


Figure 28 - Highest Correctable Micro-Reflection Using 16-QAM

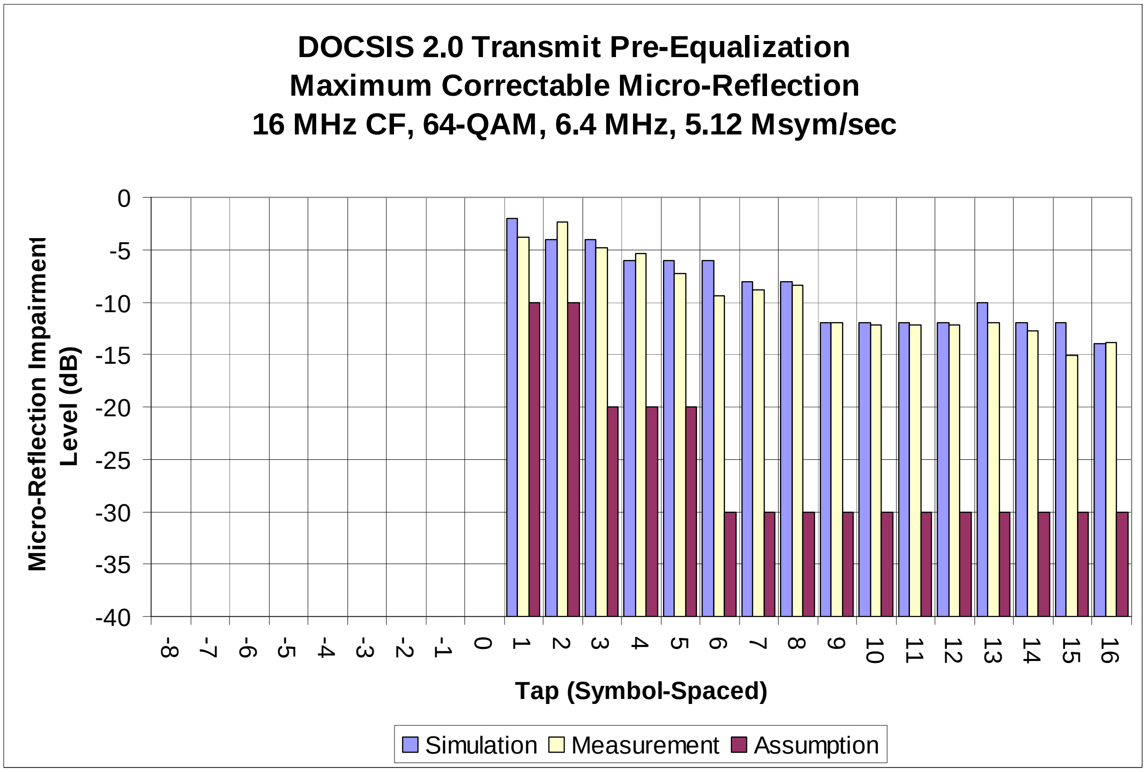


Figure 29 - Highest Correctable Micro-Reflection Using 64-QAM

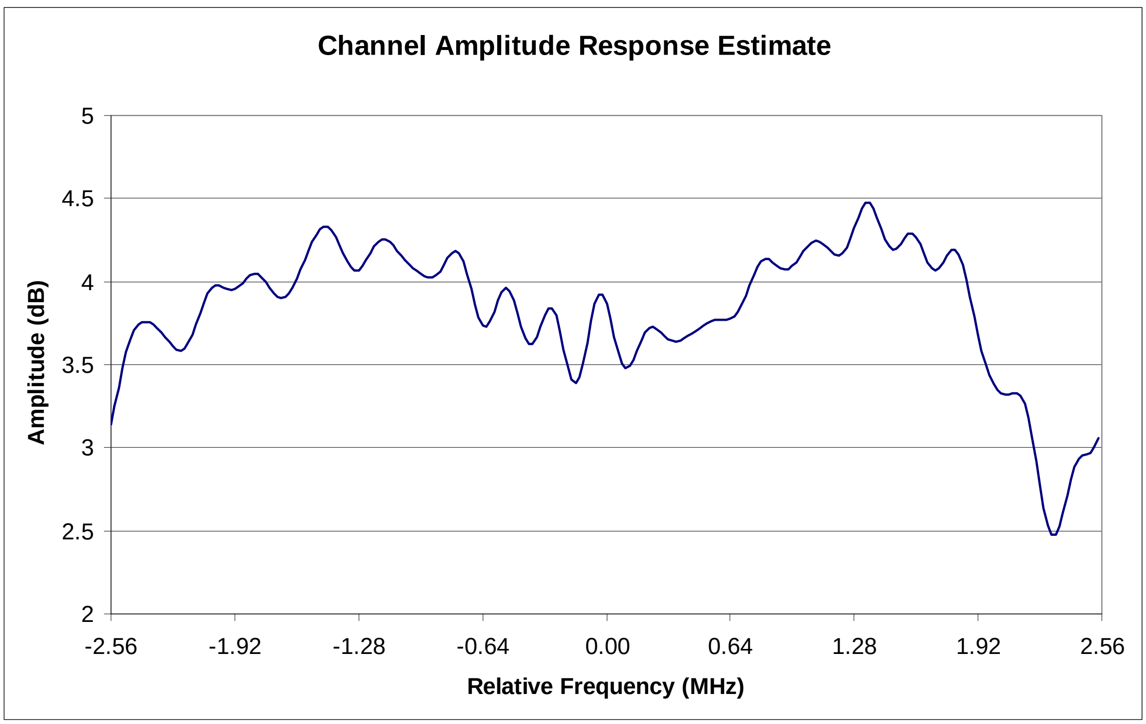


Figure 30 – Cascaded Amplitude Distortion Estimate, BW = 6.4 MHz, CF = 36.8 MHz

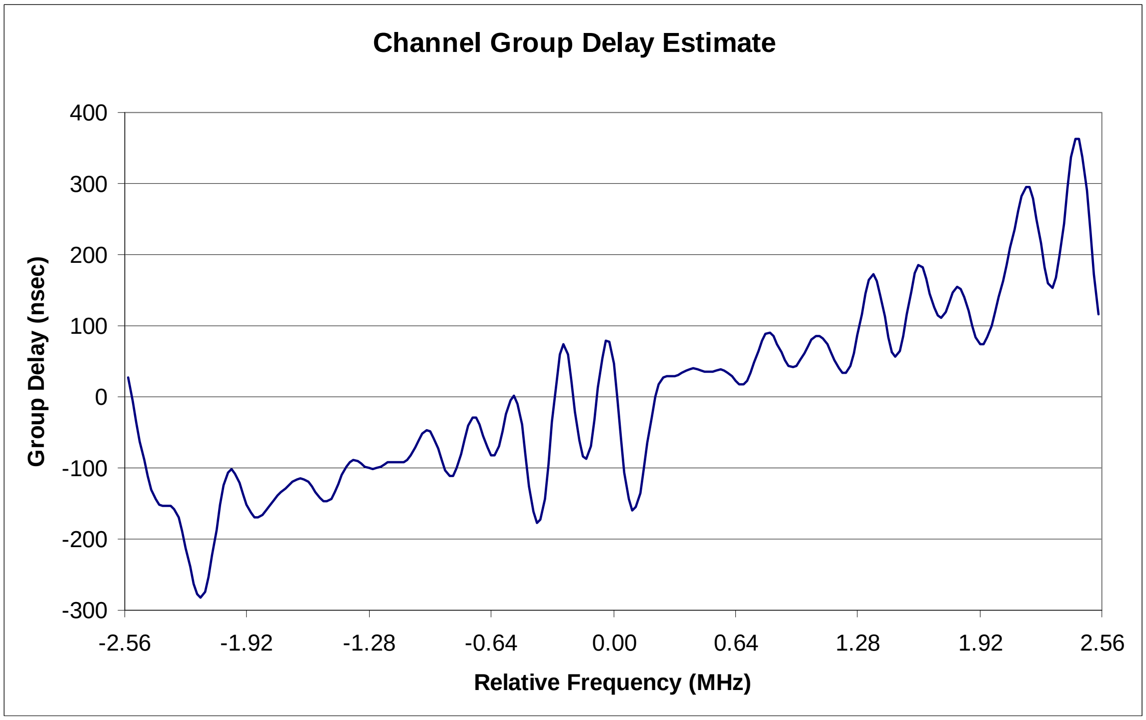


Figure 31 – Cascaded Group Delay Variation Estimate, BW = 6.4 MHz, CF = 36.8 MHz

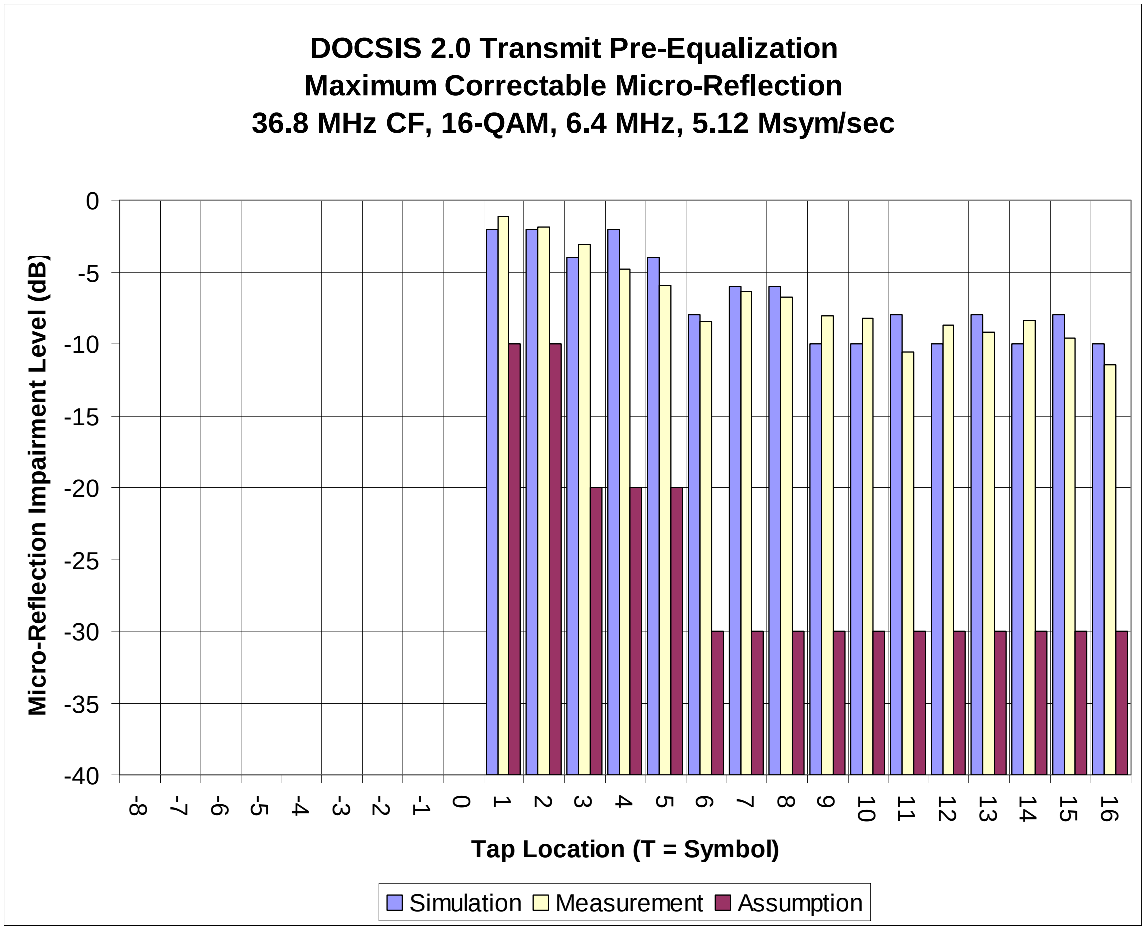


Figure 32 - Highest Correctable Micro-Reflection with Cascaded Amplitude Distortion and Group Delay Variation Using 16-QAM

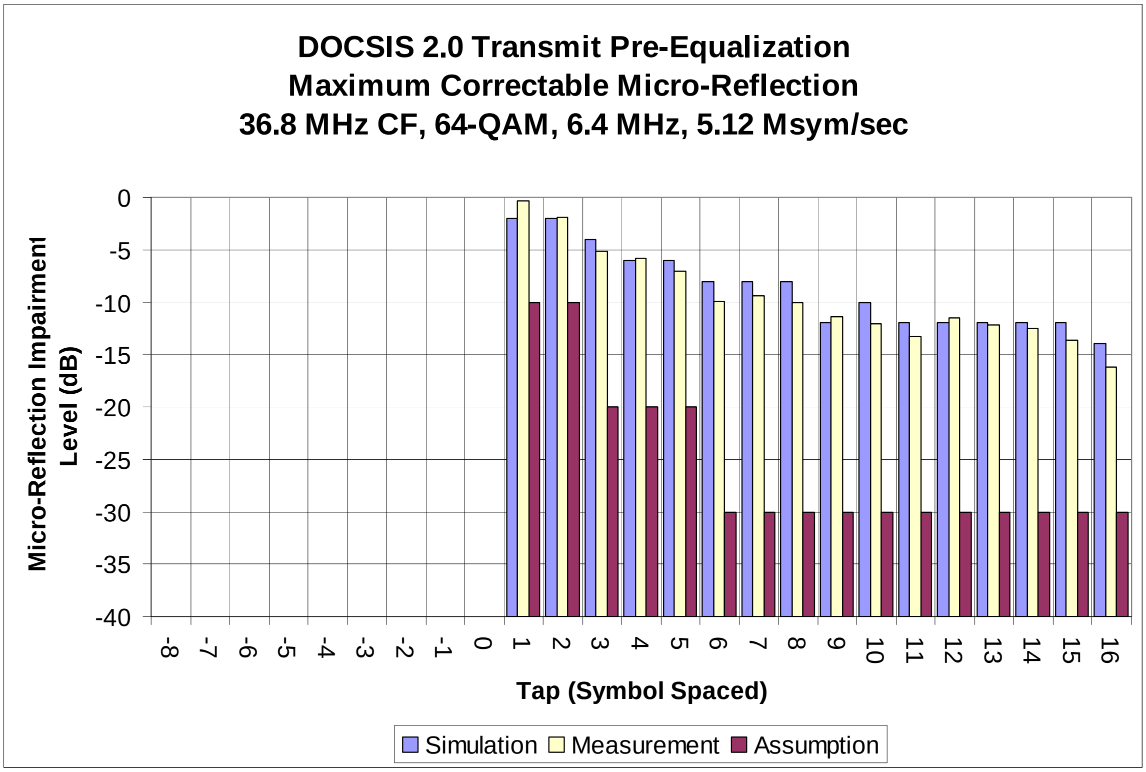


Figure 33 - Highest Correctable Micro-Reflection with Cascaded Amplitude Distortion and Group Delay Variation Using 64-QAM

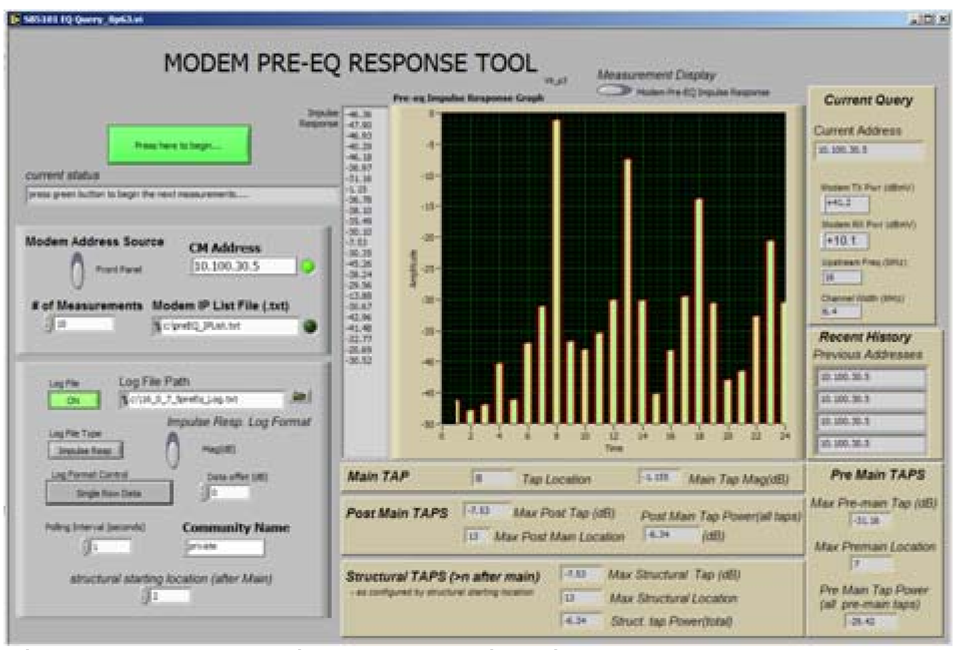


Figure 34 - Transmit Pre Equalization Query Tool

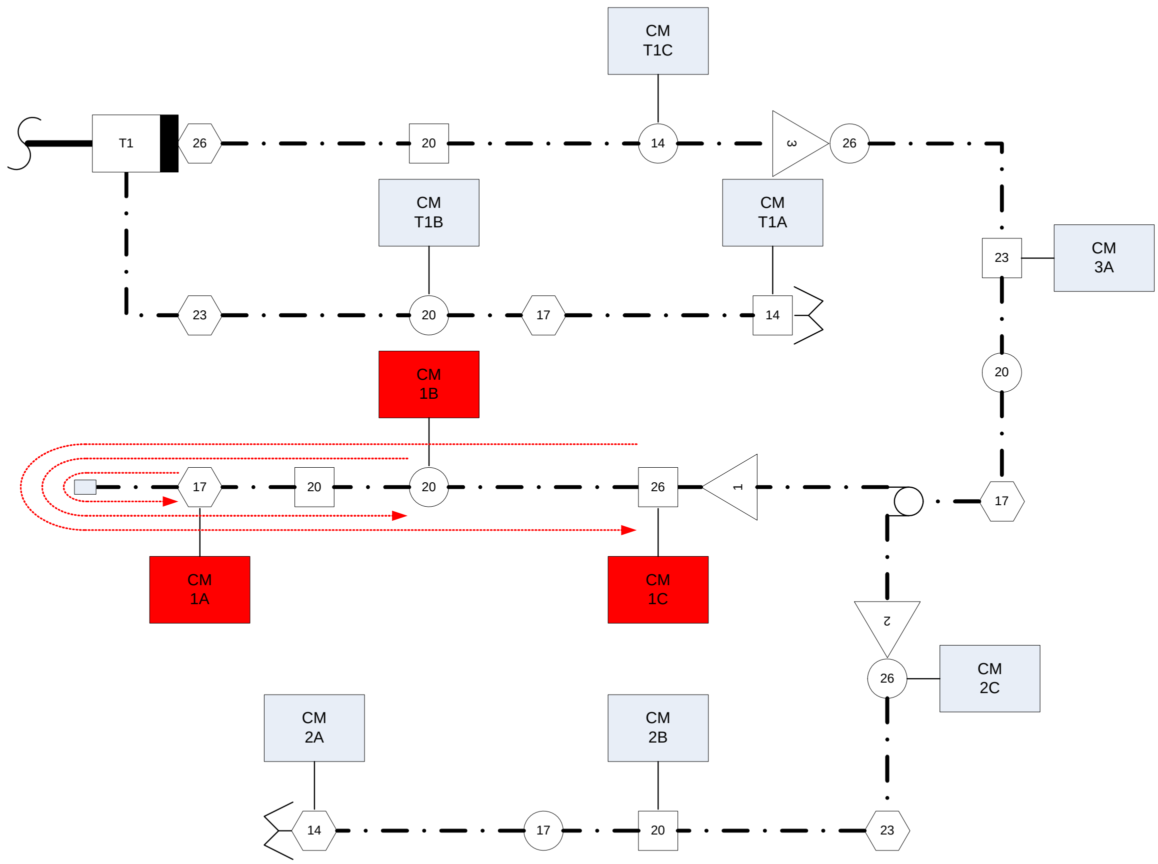


Figure 35 - Micro-Reflection Impairment Isolation Example

Separation of Video Processing and Multiplexing: The Key to Network Scaling of HD-VOD and other Bandwidth Intensive PersonalizedTV® Services

Ron Gutman, CTO
Lorenzo Bombelli, VP – Product Strategy and Management

Imagine Communications, Inc.

Abstract

Cable operators need to rapidly expand bandwidth capacity to accommodate the emerging personalized and time-shifted viewing paradigm and the accelerating availability and demand for HD content. Linear programming, traditionally broadcast to subscribers using efficient high-end closed-loop Variable Bit Rate (VBR) technology, will increasingly be delivered over on-demand networks which, using current Constant Bit Rate (CBR) technology, consume approximately 50% more bandwidth without matching the video quality of high-end broadcast VBR delivery.

Although there have been a number of proponents of on-demand VBR, architectures and solutions based on broadcast-oriented rate-shaping technology inherently rely on performing both video processing and statistical multiplexing on each on-demand stream. Such approaches result in lower video quality than high-end broadcast VBR and a cost structure that is not viable for wide scale deployment in on-demand applications.

In this paper we propose a new architecture that enables on-demand VBR by separating the video processing of each video asset from the downstream multiplexing of each stream to achieve the same video quality as broadcast VBR and eliminate the unnecessary and significant cost of re-processing the same asset multiple times. This "never cover the same ground twice" approach makes VBR viable for on-demand services, and delivers the following value to system operators:

- *Clear and sustainable competitive advantage in video quality*
- *Up to 50% gain in HFC bandwidth*
- *Reduced overall CAPEX and OPEX*
- *Uniform and controlled video quality across the entire network*
- *Improved system availability*
- *Increased service velocity*

ENABLING ON-DEMAND VBR

Limitations of Traditional VBR Technology

While VBR technology can provide significant bandwidth gains and video quality improvements, a number of important factors have impeded its adoption in on-demand applications.

During initial VOD rollouts, the challenge and complexity of developing the underlying infrastructure, combined with the relatively low initial demand for on-demand services, made optimizing video quality and bandwidth lower priority.

After initial rollouts, adoption of traditional VBR solutions was further deferred because existing VBR solutions based on broadcast-oriented rate-shaping technology could not generate enough incremental value, both in terms of video quality and bit rate reduction, to justify the incremental investment.

A key architectural limitation of these broadcast-oriented solutions is that they apply both video processing and multiplexing on each on-demand stream, which leads to two key shortcomings.

The first is that broadcast-oriented rate-shaping solutions attempt to compensate for the inherent cost associated with having to apply video processing to each stream by using bit rate as a less complex yet less accurate proxy for video quality. As a result, these solutions have no mechanisms to establish and set the optimal bit rate to achieve a desired video quality benchmark, and tend to sacrifice video quality in order to meet bit rate targets.

The second and related shortcoming is economic. Despite attempts to reduce cost per processed stream in broadcast-oriented rate-shaping solutions, it is impossible to overcome the inherent architectural limitation of applying video processing on a per-stream basis by reducing component cost. As a result, broadcast-oriented VBR solutions have been unable to meet the system cost requirements for scalable deployment on all streams.

One proposed approach to mitigate this limitation is to add an intelligent session management layer which would dynamically assign video processing resources only to over-provisioned QAM channels, relying on predictable and static user behavior patterns to allocate resources. While this approach does partially alleviate the cost and density burden inherent in a broadcast-oriented rate-shaping architecture, it does not provide the scalability required in a largely user-demand driven delivery network and is therefore a short term workaround to a sub-optimal architecture, rather than a transformative long term solution.

As a result of these shortcomings, although VBR has the potential to deliver considerable value in the rapid expansion of on-demand services, a new approach is required to make VBR viable in on-demand networks.

A New Approach for On-Demand VBR

We propose the following approach to on-demand VBR to make it viable and accelerate its adoption in cable:

1. Decouple video processing from statistical multiplexing to enable optimal technical and economic placement of each function in the network.
2. Process each asset one time upon ingest using high-end VBR technology rather than once for each stream using rate shaping technology.
3. Utilize the highest quality video processing, giving priority to video quality rather than bit rate to ensure that quality is highly controlled and not compromised.
4. Optionally, create a simple linkage between the video processor and the SRM that enables the SRM to incorporate quality in bandwidth allocation decisions.

The Case for Asset-based Video Processing

Decoupling video processing from statistical multiplexing and applying video processing on a per-asset instead of a per-stream basis enables significant improvements in both the video quality performance and the economics of on-demand VBR. Take the following example. A cable system with a half million subscribers and a deployed base of one million digital set-top boxes has a VOD system designed for 10% peak utilization, meaning that during peak usage, no more than 100,000 VOD or unicast standard definition streams can be active simultaneously on the network. The operator has taken into account user patterns across the system and has deployed an intelligent session management layer to enable over-provisioning of rate-shapers, and sufficient broadcast-oriented rate-shapers to process 50% of deployed streams. In effect, the operator has deployed 50,000 unique and distinct video processors into the network that can be assigned to user sessions as required.

Now consider a “super” video processor, with 100 times the processing power of a rate-shaper, applied on a per-asset basis at the ingest point of the on-demand library. Also suppose that at the ingest point the system requires a content refresh rate of 10,000 hours per month. Assuming real-time processing, this would require 10,000/30/24 = 14 processors.

Even assuming 20 processors to allow overhead for peaks in refresh rates, per-asset pre-processing with “super” video processors would only require 2,000 video processing units, saving an equivalent of 48,000 video processing units. These gains are further amplified if asset-based processing is applied in regional clusters of 2-3 million subscribers or national clusters of 5-20 million or more subscribers.

Video Quality Benefits

The allocation of video processing on a per-asset, rather than per-stream, basis represents a fundamental shift in the approach to on-demand VBR. With 100 times more processing power available for each asset, asset-based processing can take advantage of advanced video processing techniques that are not economically viable when applied on a per-stream basis. Look-ahead configurations can be maximized, two-pass and even multi-pass encoding can be applied, the search for more precise motion vectors can be enhanced, scene changes can be more accurately identified, and the diligence applied to noise filtering can be both infinitely granular and wholly comprehensive.

Networking and System Benefits

In addition to the quality benefits described above, shifting to asset-based video processing has other important system benefits.

The ratio of processing power required for video processing vs. multiplexing is approximately 98:2. Therefore, separating video processing from multiplexing not only unlocks the potential for higher video quality on the video processing side, but also enables significant reductions in cost, density, power consumption and system delay on the statistical multiplexing side.

Moreover, since assets are pre-processed, they can be encrypted upstream from the multiplexer (after the video pre-processing), and then distributed, multiplexed and transmitted to the subscriber without compromising the content protection or digital rights management.

Also, by consolidating video processing operations and distributing the multiplexing function, system operators can enjoy a new level of quality and operational control, ensuring that all subscribers have same high quality viewing experience regardless of the particular and immediate usage patterns in their neighborhood.

IMPLEMENTATION

Statistical Multiplexing Efficiency

The benefits of statistical multiplexing of VBR signals are well known and understood. The bandwidth efficiency benefits amounts to up to 30% gain in channel capacity. When paired with the unlocked potential of a separated video processing engine, the efficiency of the VBR statistical multiplexer can reach as high as 50%, but not without some exceptions.

Our research suggests that when unconstrained video processing is combined with constant quality VBR multiplexing, the aggregate rates of the individual programs will peak above an allocated channel rate approximately 15% of the time. The two examples given below plot instantaneous bit rates for two different 30 minute standard definition programs and provide a histogram for the number of seconds at each bit rate. In both examples, the content is coded at a constant "ICE-Q®" quality score of 96, indicating a quality level of NMD (No Material Degradation), providing a constant quality VBR signal that matches the quality of the source.

The complexity difference between the two programs is reflected in their average and maximum bit rates as well as their Effective Bit Rate (EBR). The EBR is a new metric that indicates how much bandwidth is required to maintain constant quality for any given asset. In Example A, the input CBR rate is 3.75 Mbps. The content is an HBO Sex in the City segment with an average VBR bit rate of 2.3 Mbps, standard deviation of 0.45 Mbps, a maximum bit rate of 3.32 Mbps and an EBR of 2.78 Mbps.

<u>Example A</u>

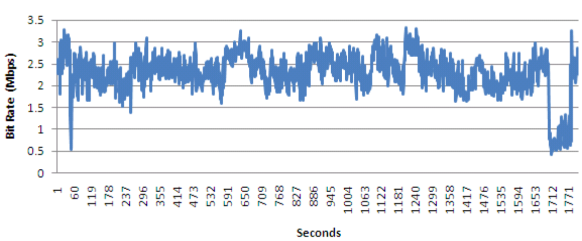


Figure 1: HBO: bit rate over time

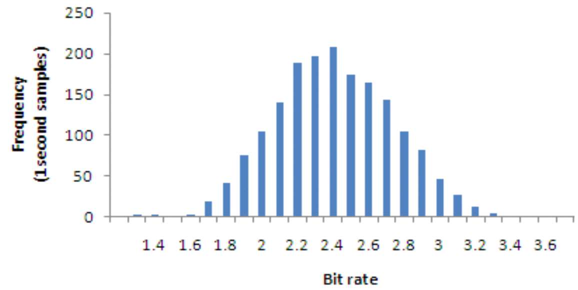


Figure 2: HBO: bit rate histogram

In Example B, the input CBR rate is also 3.75 Mbps. The content is a nature documentary. The average bit rate is 2 Mbps, the standard deviation is 0.4 Mbps, the maximum bit rate is 3.67 Mbps and the EBR is 2.46 Mbps.

<u>Example B</u>

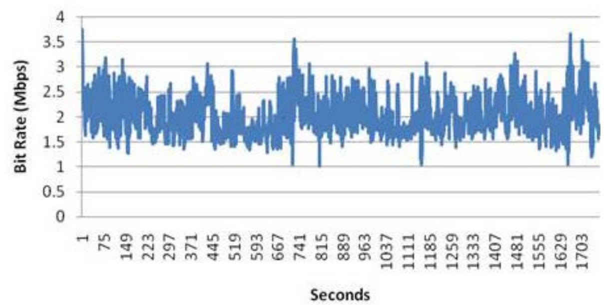


Figure 3: Africa: bit rate over time

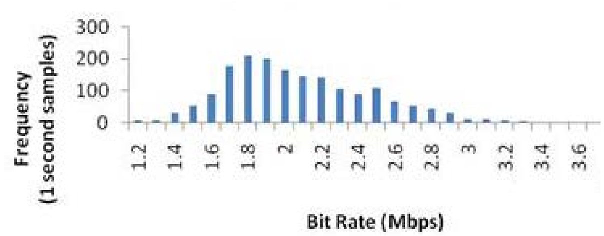


Figure 4: Africa: bit rate frequency histogram

In Example A, the percentage of seconds that are above the EBR is 12%. In Example B, the percentage is 15%. Similar analysis has been performed on a larger sample of 40 distinct HD and SD clips of varying content type and complexity. VBR bit rate distribution in this sample shows a clear statistical match to normal distribution. Overall, the SD clips have an average VBR bit rate of 2.13 Mbps, standard deviation of

0.389 and an average EBR of 2.41 Mbps. If the available QAM bit rate is 38 Mbps (allowing some reserve) the average number of streams per QAM is 15.8 according to the EBR average. Assuming optimal time-shifting and buffer management at the edge multiplexer, the probability of the aggregated average bit rate exceeding the QAM channel bit rate at 16:1 density is 0.7%. Similarly, the probabilities of streams exceeding the QAM bit rate for 3 HD plus 3 SD streams, 2 HD plus 8 SD streams, and 1 HD plus 12 SD streams are 2.4%, 2.5% and 1.3% respectively. Clearly, more streams per QAM provide larger statistical gain.

The sensitivity of video and audio communications is such that a ~2% error results in a completely unusable product. Therefore, some additional compression is required on 2% of the content.

Second-Pass Asset Processing

An additional benefit of the massive amount of video processing enabled in the asset processing stage is that all of that processing can be brought to bear in a second pass of processing whereby the variable bit rate signal can be more aggressively compressed a second time to create an alternate VBR version of the signal to the predominant constant quality "V1s." This alternate ("V2") can then be transmitted to the edge multiplexer and inserted during the 2% of the time in which it is required.

In this architecture, the statistical multiplexer performs the critical process of rate management without needing to perform rate-shaping or re-compression of the signal. And because the V2 version of the signal has been compressed using the full capability of the high-end video processor, it is of higher quality than can be provided by a broadcast-oriented rate-shaping platform, which is severely constrained by cost and density requirements.

Improved Quality and Interoperability

It is important to note that the asset-based approach for content processing can be applied to improve video quality as well as bandwidth efficiency. This is accomplished by starting with a higher input bit rate and higher quality encoded source as the input to the asset-based processing stage, for example 7 Mbps CBR for SD. The resulting EBR for the processed asset will be higher to support the higher quality level, resulting in a channel capacity that is equivalent to today's 10 streams per QAM, with each CBR at 3.75 Mbps.

These processed streams will appear to the network as typical 3.75 Mbps CBR. The video bit rate inside the 3.75 CBR transport rate, however, will be variable. In the valleys of the VBR rate, additional information can be stored that allows statistical multiplexers to recreate the higher quality program derived from the 7 Mbps source file. The file containing the extra layers of information must conform to standard VOD specifications such that servers, pumps, QAMs and asset management systems will process the asset without requiring any additional processing or format conversion. The same formula can be used for high definition content, substituting 25 Mbps input sources for the 15 Mbps CBR content prevalent today. Using this method, standard definition VOD assets can be "DVD quality," while HD-VOD assets can approach "Blu-ray quality."

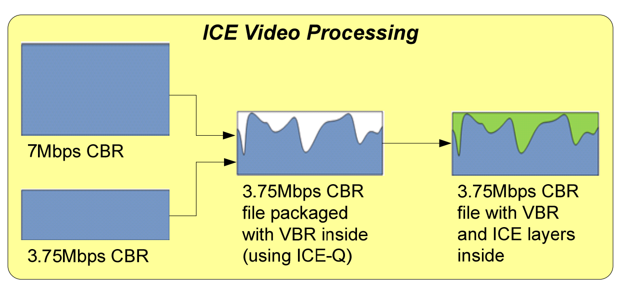


Figure 5: Improving VOD quality with ICE layers

Video Quality Aware SRM

Optionally, with minor modifications, the Session Resource Manager (SRM) can play a vital part in further optimizing bandwidth while maintaining the highest video quality in an on-demand VBR system.

From a video coding perspective, content varies in complexity. Sports content typically includes high temporal complexity (motion) while other high-complexity content is spatially detailed or includes spatial subtlety (ramp-type edges) that requires advanced processing. Other content such as static images, dark backgrounds or non-moving head-and-shoulders video is less complex and can be processed much more easily.

The fundamental problem with providing the same bit rate for all streams without taking into account content complexity is that it results in sub-optimal picture quality and bandwidth efficiency. Just as 3.75 Mbps may be inappropriate (overkill) to adequately encode a head-and-shoulders news clip, the 3.75 Mbps rate can be equally inappropriate (underkill) for encoding a complex scene. Even with virtually unlimited encoding, it can be impossible to maintain quality during dramatic scene changes, fast motion and high spatial complexity if strictly constrained to 3.75 Mbps.

For example, the processing-rich encoding in Example A resulted in an Effective Bit Rate (EBR) of 2.78 Mbps. If one tries to multiplex 16 or even 15 such streams, severe degradation in video quality is virtually guaranteed, especially if done by a processing-limited rate-shaping device. The problem is amplified in HD because the aggregated standard deviation is much higher due to the lower number of streams per QAM.

A simple solution to this problem is to provide the EBR to the SRM thereby enabling the SRM to make QAM allocation decisions based on quality with a guarantee that total channel capacity will not be exceeded.

The EBR is equal to the combination of the average VBR bit rate combined with the maximum V2 bit rate, thus guaranteeing the quality of the multiplex at any given time.

EBR = Max(Average(V1),Max(V2))

The EBR value is calculated on a per-asset basis and could be delivered to the SRM using the ADI interface making the standard SRM quality aware and enabling it to load-balance according the quality rather than bit rate. Note that this approach does not require adding more elements to the session management flow.

Furthermore, while the EBR approach described above is very simple and effective, it is possible to further optimize the solution. For example, the asset video processor could provide the SRM with the averages and standard deviations of V1 and V2 per asset:

μ – Asset average bit rate
σ – Asset standard deviation

In this case the SRM would open a session if both of the following conditions are true:

V1 – Normal_Dist(38, M_1, S_1, 1) $\geq$ 99%

V2 - Normal_Dist (38, M_2, S_2, 1) $\geq$ 99.999%

Where:
$M = \sum_n \mu(1)$
$S = \sqrt{\sum_n \sigma^2(1)}$

Therefore, enabled by the EBR (Effective Bit Rate), and using very simple methods that can be enhanced and optimized by independent SRM vendors, the Session Resource Manager (SRM) can actively load-balance the QAM sessions according to video quality rather than stream counts, thereby guaranteeing constant video quality to all sessions while further optimizing bandwidth utilization.

In other words, the SRM is transformed into a video-quality aware device that can ensure and control video quality while delivering more streams at any given system capacity.

CONCLUSION

The growing demand for bandwidth-intensive PersonalizedTV services such as HD-VOD, switched unicast, and Internet video (whether streaming or downloads) will drive substantial HFC bandwidth pressure.

While on-demand VBR can deliver significant video quality improvements and bandwidth gains, current solutions based on broadcast-oriented rate-shaping technology have significant shortcomings that challenge their viability for scale deployment.

We propose that while edge multiplexers must perform rate management to accommodate the small percentage of the time when aggregated bit rate peaks exceed the available channel rate, it does not follow that those same edge devices must also be burdened with the computational intensity of video processing platforms. By relying on edge video processing, we unnecessarily impair service quality by sub-optimally forcing video quality decisions to lower common denominator devices.

In contrast, by separating the video processing and performing it at the video server ingest point on a per-asset basis, we ensure not only a consistently higher video quality during times of non-contention, but also a dramatically improved picture during the most critical periods of aggregate peak rate channel utilization.

In addition to enabling superior video quality and bandwidth efficiency, a separated asset-based processing architecture is uniquely equipped to scale with advancements in video and audio coding, while maintaining centralized control over delivered video quality, optimizing the cost and density of the edge multiplexing solutions, and minimizing delay in the end user experience.

In summary, an on-demand VBR approach that separates advanced video pre-processing from statistical multiplexing offers a compelling solution for scale deployment that delivers the full value of on-demand VBR and provides system operators a new and powerful tool in the ongoing battle for bandwidth and the consumer.

TelePresence over DOCSIS

John T. Chapman, Cisco, jchapman@cisco.com
Harsh Parandekar, Cisco, harsh@cisco.com
Jeffrey Finkelstein, Cox Communications, jeff.finkelstein@cox.com

Abstract

Telepresence is a new technology category that delivers a unique, in person experience for virtual meetings. A telepresence solution integrates advanced visual, audio, and interactive technologies with broadband networking to bring people together from across the campus and around the world.

To date, telepresence has been deployed mostly by large enterprises to enable employees to conduct virtual meetings with fellow employees and customers at remote offices that are linked via a private corporate network.

As telepresence becomes more prevalent, businesses will want the ability to extend the system to locations beyond the corporate network. However, as a real-time, two-way service with high-definition video and spatial audio, telepresence places stringent requirements on the underlying network to provide a high-quality user experience.

This paper describes methods for extending telepresence to locations served by a DOCSIS access network, such as executives' homes. The challenges of delivering telepresence over DOCSIS networks are investigated, and potential solutions for addressing these challenges are proposed.

SERVICE PROVIDER DEPLOYMENT MODELS

Large enterprises were the early adopters of telepresence, which sought to increase productivity and reduce expenses by conducting virtual meetings with an in-person experience. These enterprises typically deployed telepresence over a private corporate network, or secured the appropriate service level agreements (SLAs) from service providers to support their enterprise telepresence systems.

Enterprises are now looking to extend their telepresence systems to more locations, including additional corporate offices, key customer and partner facilities, and executives' home offices. As enterprises continue to grow their telepresence systems, cable operators will increasingly have the opportunity, and the challenge, to support telepresence services on their DOCSIS access networks.

The stringent network requirements to support the real-time, bi-directional, high-definition video and spatial audio of telepresence will likely require cable operators to offer new levels of service to subscribers who want to access corporate telepresence systems via the DOCSIS network.

Although the enterprise will typically bear the cost of the network services required to provide remote access to the enterprise telepresence system, the cable operator will likely bill the subscriber directly as part of a residential services package.

As telepresence becomes a more prevalent medium for business-to-business communications, enterprises will want to enable their systems to interoperate with their ecosystem partners. Smaller businesses will also want to deploy telepresence, but may not have the resources or know-how to manage their own telepresence system.

Cable operators and other service providers may see an opportunity to offer a managed telepresence service to these customers as part of a commercial services package. This could entail hosting a telepresence call management server (CMS) and managing access control and billing in addition to providing the network services.

As telepresence technology evolves, one can imagine consumer electronics devices incorporating telepresence capabilities for personal use. As this evolution takes place, cable operators will increasingly be able to offer the ultimate visual networking experience to their residential subscribers.

While this paper is focused primarily on enabling telepresence over DOCSIS (TPoD) as an extension of an enterprise telepresence system, the challenges and proposed solutions described herein are also applicable to the other deployment models described above.

THE TELEPRESENCE USER EXPERIENCE

Telepresence is a technology that allows people who are in physically separate locations to communicate with each other as if they were in the same room. Telepresence combines professional video, professional audio, and networking to create a real-time in-person experience.

So what does that really mean in more technical jargon? Video is the main component. Enterprise telepresence systems, as shown in Figure 1 typically use one to three 65 inch plasma monitors operating at 1080p and 30 frames per second. The monitors are set on one side of a desk in such a way that the desk at both ends of the call combines together to create one virtual desk with everyone sitting around it.

The audio uses professional

Figure 1 – Telepresence Endpoint with Three Screens

microphones and speakers. In a larger system, multiple microphones are used, and audio streams are coordinated so that audio from an individual on the left side of the room is played back on the left speaker at the far end. This is known as spatially aware audio.

Even the lighting of the room, one of the most important considerations, borrows from the professional world. Reflective or diffused lighting is recommended as direct lighting causes shadows. Light sources are located in front of the people to light up their faces, not behind them as often happens in the webcam experience.

The lighting is chosen to have the best color temperature (4100K) to make skin tones look good. The camera is a fixed focal length 1080p camera and is calibrated to the room. There are even design guidelines for the exact size of the room, and the color and composition of the walls, floor, and ceiling.

There is a method for conveying slides from a PC. This may be done through picture-in-picture or a separate monitor or projector.

And finally, there is the user interface. Classical video conferencing systems had a complicated user interface that required lots of configuration and usually didn't work. A properly designed telepresence system uses a simple user interface such as an IP phone. For example, a call can be booked in Outlook. Automated provisioning software causes the meeting to show up on the IP phone screen. The call is established by pushing a single button.

In a properly designed telepresence system, there is nothing to adjust. All adjustments have already been done. It just works.

The quality of the experience and the careful placement of monitors, cameras, and microphones create the illusion of an in-person experience.

SYSTEM OVERVIEW

Behind the Scenes

There are other components in a full telepresence system.

In the room, there is a telepresence endpoint that contains the CODEC (Coder-Decoder) and the call management software. On one side, the chassis connects to the monitor, camera, speaker, and microphone. On the other side, the chassis connects to an Ethernet network. If more than one set of monitors is needed, then this system is duplicated and then interconnected via the Ethernet network. After all, it's all IP at this point.

The telepresence endpoint may contain an auxiliary channel that sends and receives audio and video from a personal computer. This is for sharing slides and other audio-visual material.

At a remote location, there is a call manager that manages the telepresence calls and the IP phone. For telepresence systems that use SIP (Session Initiated Protocol) for signaling, the call manager is often implemented with a SIP proxy or an IP PABX.

In order to make multi-party calls, the remote location may also have a telepresence conference bridge.

To permit interoperability between vendors, the remote location may also offer gateways that perform signaling conversion. Once example of a gateway would be to convert between older H.323 systems and newer SIP based systems.

<u>Adapting Telepresence to the Home Office</u>

The home office is not as easily configured as the enterprise environment. The easiest solution for a home office is to use an all-in-one floor standing system and put it on the far side of the desk.

The author of the paper configured a home office telepresence system with discrete components. He put a 40" LCD TV at the end of his six foot desk and sits at the other end for calls. The camera is mounted to the top of the TV. The controller was put into a cabinet with ventilation. The phone and microphone sit on the desk.

The lights in the room were changed to get the right color, and a reflective light source was added above the TV to help the camera. The windows in the room had to get blackout shades so that they would not create back lighting.

To connect back to the enterprise environment, a good quality router with build in VPN (IPsec based) is needed. As will be seen later in the paper, the existence of this router and the VPN

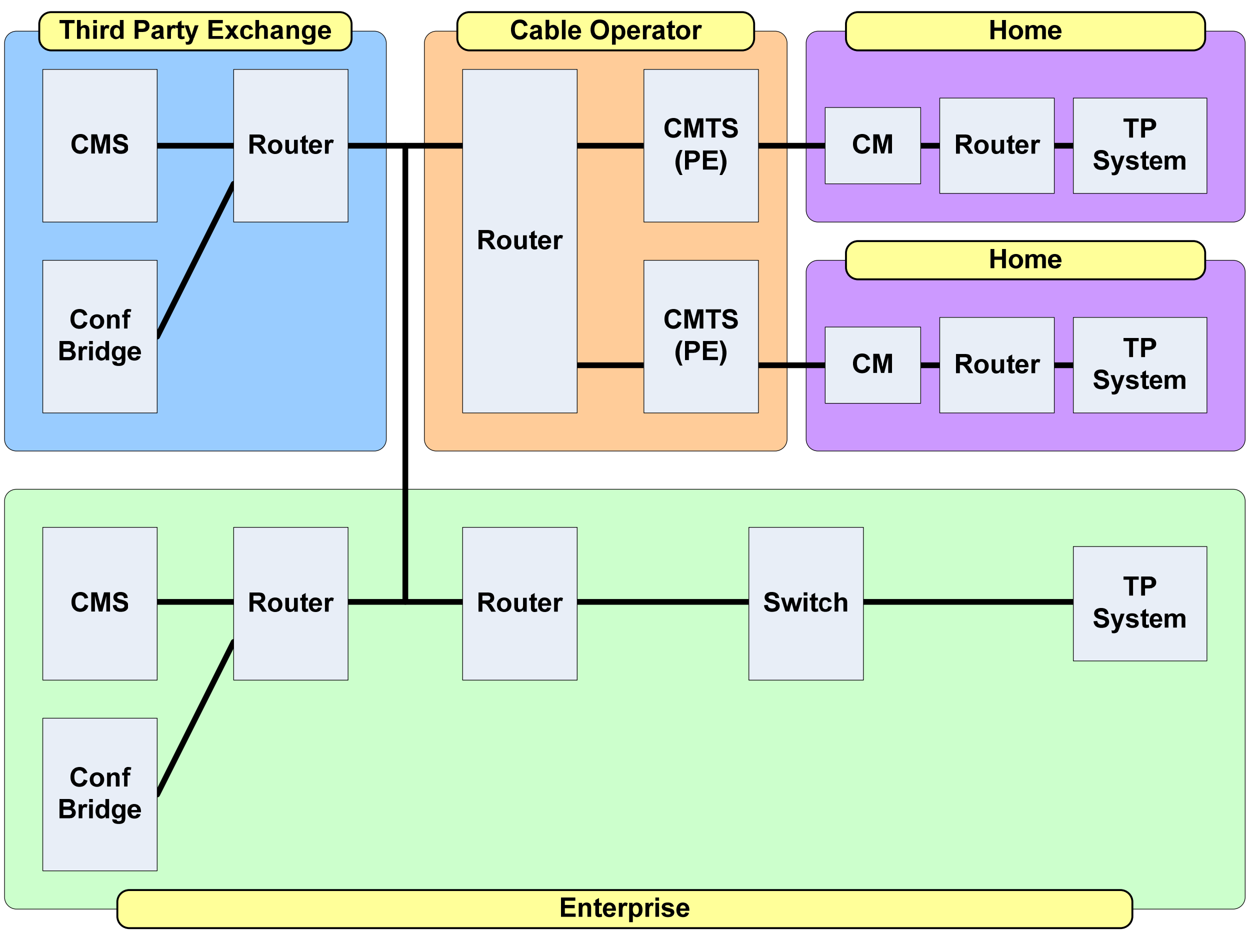


Figure 2 – TPoD End-to-End System

create one of the two major problems that have to be overcome in order to set up QoS on the DOCSIS access network.

<u>End-to-End System</u>

Figure 2 shows an entire TPoD system that spans both the enterprise and the service provider environment.

The enterprise contains one or more telepresence systems. It also contains its own IP PBX system (call manager system) that is used to connect IP phones and to connect the telepresence systems. There is a local telepresence conference bridge as well.

The home office as described earlier contains a telepresence system that sits behind a VPN router. This connects to the cable operator through a local cable modem (CM) and cable modem termination system (CMTS).

For TPoD systems that are deployed for intra-enterprise purposes, all calls are typically routed to the CMS managed by the enterprise. If the enterprise CMS is not compatible with the home office telepresence system, a third-party telepresence service may be used. Third-party service may also be used to enable inter-enterprise calls. The third-party service provider will require secure access to each enterprise. This is significant as it can impact call flow considerations.

HOME NETWORK CONSIDERATIONS

The home network consists of all the networked components in the home and the interconnectivity, including the telepresence endpoint and the DOCSIS CM. The connection between the telepresence system and the DOCSIS CM should be a wired path. Wireless connections are not recommended as they have a higher packet loss rate that can impact real-time video.

In a worst case scenario, there could be three routers in the home network.

1. A home router which aggregates all the traffic from the home network

2. A telepresence router that places telepresence traffic onto a VPN. A third-party service provider may remotely manage this router.

3. An enterprise telecommuter router to that is managed by the enterprise for data and VoIP connectivity. (Ideally, this is the same router as the telepresence router)

The main difference between these three routers is that a different individual or organization manages each router. In the simplest scenario, there is one router that manages both the home network and the telecommuter and telepresence networks.

One connectivity option is to have the home router and telepresence router connected separately to the CM. For this to work, the cable operator has to be able to configure the CM to classify and provide QoS treatment for telepresence traffic. This would require separate IP addresses from the cable operator for each router and NAT in the CM to be disabled.

If the telepresence router is connected through the home router, the home router should be configured to provide priority access to the CM for the telepresence service.

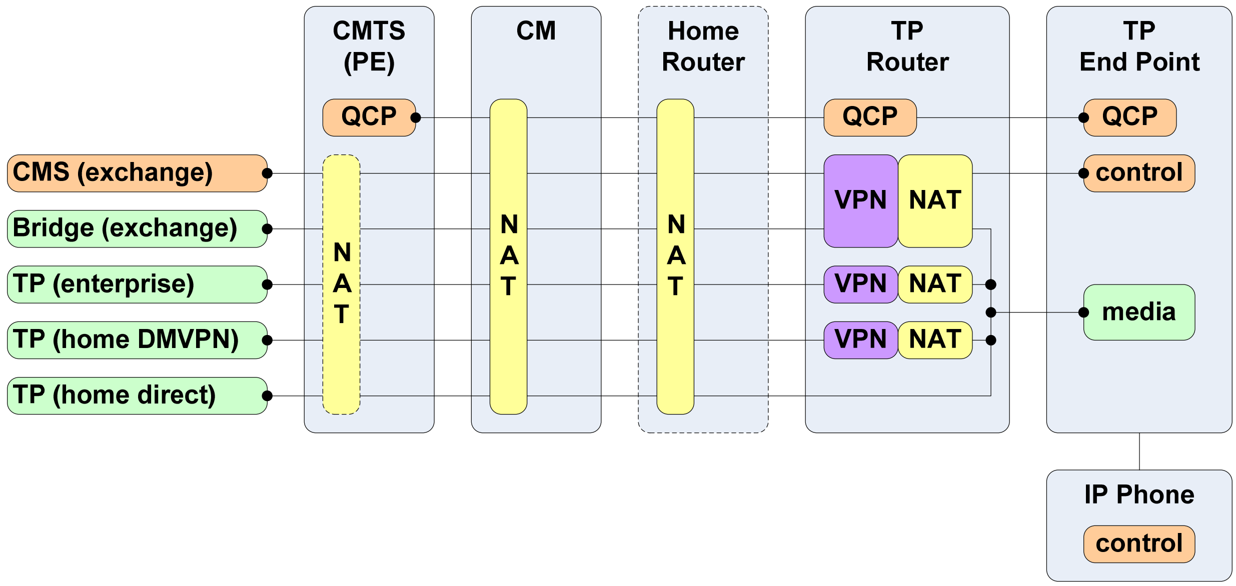


Figure 3 – VPN and NAT Scenario

For best results, the telepresence router should be directly connected to a port on the home router or the CM. This will eliminate any contentions within the home network for traffic.

ACCESS NETWORK CONSIDERATIONS

The core problem that this paper is addressing is how to deliver QoS on the DOCSIS network to support a telepresence call. To do that, the DOCSIS system needs to identify the flow of telepresence packets, separate them from the rest of the packets, and apply specific QoS algorithms.

To understand this issue better, lets look at a packet as it leaves the telepresence system and travels upstream to the CMTS.

There are two major issues with the way that telepresence is connected that prevent easy identification of telepresence packets.

- VPN
- NAT (multiple)

A VPN tunnel is built between home and enterprise routers. The VPN tunnel puts an encrypted IP packet inside another IP packet. Without the ability to read the inner encrypted packet, a SP (and the CMTS) can only use the outer IP header to route between home and the enterprise.

The home and enterprise routers terminate the VPN tunnel, remove the outer header and route only the inner packet onwards. The VPN tunnel is transparent to the telepresence system and the enterprise call manager. Since the outer IP header is only added/removed by the routers terminating the VPN tunnel, the telepresence system and the enterprise call manager only communicate using the inner IP header and have no knowledge of the outer IP address. So who is going to tell the CMTS how to build a classifier?

The situation is complicated further by NAT. Often the port on the router that is offering a VPN service also runs a local NAT. That means that the telepresence packet had its IP addresses changed prior to being encrypted and encapsulated. This is the first level NAT.

If the VPN router is connected to the home gateway router, the home gateway router often runs a NAT. Now the outside IP address of the VPN tunnel is changed. This is the second level NAT.

Most CMs today come with built in NAT. As the packet travels through the CM, the previously NATed addresses are NATed again. This is the third level NAT.

The packet is now ready to be classified by the DOCSIS side of the CM and then be sent to the CMTS.

When the packet leaves the DOCSIS MAC domain, it may be subject to yet another NAT. It is a network NAT that may get deployed after IPv4 addresses become scarce and before IPv6 can be deployed. Although it does not impact the upstream packet, by changing the packet address yet again, it will impact the packet classifier needed on the corresponding downstream of the CMTS at the far end of the connection.

In summary, the upstream telepresence packet, once it left the safety of the telepresence endpoint is grabbed by the network, NATed, encrypted, encapsulated in a tunnel, and then NATed up to three more times. This is worse treatment than Houdini used to get.

And now the CMTS is supposed to build a classifier for this packet? How is that going to work? Before we discuss our proposed solution, lets look at the second part of the QoS equation that is the traffic characteristics of the telepresence flows.

MEDIA FLOW SPECIFICATIONS

Specifications vary between the various manufacturers of telepresence systems. The following set of tables is intended to be a reference point for a typical telepresence system. Actual performance may vary.

Video

Specification	Description
Image Size	1920 x 1080 or 1280 x 720
Frame Rate - Main Video - Aux Video	 30 fps progressive 5 fps progressive
Encoding	H.264, VBR
Encapsulation	RTP, IPsec
Video Bit Rate - 1080p - 720p	 3 – 4 Mbps average 1 – 2.5 Mbps average
Max Frame Burst	65 KB in 33 ms. (12 - 16 Mbps peak)
Aux bit rate	500 kbps avg 2 Mbps peak
Max Latency	150 ms one way including all delays.
Tolerated Jitter - packet jitter - video jitter	 10 ms 50 ms
Tolerated Packet Loss	0.05%

Table 1 – TPoD Video Specifications

In Table 1, the options for high definition video are 720p or 1080p resolution at 30 frames per second (fps). 1080i does not make sense since there is no television legacy in this system. Smaller screens may be fine with 720p. Aux video from a PC can be sent at a lower frame rate.

Ironically, 720p does not necessarily help out the cable environment. While the home monitor may be okay with 720p, the enterprise monitor is likely to be larger and do better with 1080p. Further, the executive using the TPoD system at home is usually more concerned about being presented well to other participants than about how well the others look to him.

That can lead to the odd case where a 1080p signal is sent upstream from the home camera to the corporate 60" screen, while a 720p signal is send downstream from the corporate camera to the 40" home screen. This runs contrary to the cable system bandwidth that typically has more bandwidth in the downstream than the upstream.

The common codec choice for encoding of high definition (HD) video these days is H.264. Aggressive H.264 codecs can achieve a target video bit rate of 4 Mbps or better. There is a catch. People familiar with MPEG-2 encoded video on demand (VOD) and switched digital video (SDV) systems are used to 3.75 Mbps constant bitrate (CBR) for standard definition (SD) video streams. These streams are encoded such that I-frames do not cause traffic peaks that would exceed the target 3.75 Mbps encode rate.

Maintaining a constant bit rate while encoding high-definition video in a low-latency, interactive system such as telepresence requires substantial video processing capabilities in the encoder. Thus, telepresence systems typically use variable bit rate encoding, and thus streams will see a burst when an I-frame is sent. That burst can typically be 2x to 4x the average bit rate.

The peak rate in Table 1 is calculated over a 33 ms frame interval, even though the burst less if it is spread over a longer time interval if there is room in the system latency budget. In the latest telepresence systems, these bursts can be as infrequent as every 5 minutes or more.

RTP encapsulation is used so the video streams may be multiplexed into common transports. RTP also provides better management of the data path.

The maximum one-way latency goal of 150 ms is the same latency goal that is used for voice calls. The value includes all network and equipment delays.

The jitter goal is also very similar to what is allowable for voice calls since jitter tolerance is achieved through jitter buffers that in turn add latency. Video jitter does get more complicated as there are differences between what is allowable for actual packet jitter caused by packet queuing versus video frame jitter which can be caused by serialization delays on slow links.

<u>Audio</u>

Specification	Description
Encoding	AAC-LD
Raw Bit Rate	64 kbps
Packet Interval	20 ms
Encapsulation	RTP
Max Rx Channels	4
Max Tx Channels	2

Table 2 – TPoD Audio Specifications

Table 2 uses an audio codec of the quality of Advanced Audio Encoding Low Delay (AAC-LD). More compressed audio codecs do not make sense since the video already takes up so much bandwidth.

A home office telepresence system may receive up to four channels of audio from the enterprise system - three channels from the three microphones and one aux channel. Assuming a single screen home system, it could send up to two channels of audio to the enterprise system – one from the microphone and one from the aux channel.

Aggregate Traffic Profile

Specification	Description
1080p	3 – 6 Mbps
720p	2 – 4 Mbps

Table 3 – TPoD Aggregate Specifications

Table 3 sums up the audio and video from previous tables. Table values are for a home office telepresence system. The presumption is up to a three-screen system at the enterprise may be communicating with a one-screen system at the home (so three audio feeds and one video feed).

Auxiliary audio and video may or may not be present. Approximately 20% overhead is added to the raw bit rates to allow for encapsulation and for signaling.

DOCSIS Configuration

The DOCSIS transmission path easily handles bursty traffic as all traffic is rate shaped with a formula that allows for traffic bursts. The common traffic parameters used are:

Downstream and Upstream

- Traffic Priority
- Max Sustained Traffic Rate (R)
- Max Traffic Burst (B)
- Min Reserved Traffic Rate
- Assumed Min Reserved Rate Packet Size.

Downstream Only:

- Downstream Peak Traffic Rate.

DOCSIS data rates are enforced by the CMTS through rate shaping of the service in the downstream, and by metering of grants in the upstream. DOCSIS uses a token bucket based rate limit for service flows. The number of bytes forwarded is limited during any time interval T by Max(T), as described by the expression:

$$Max(T) = T * (R / 8) + B$$

Where:

T = time interval under consideration

R = maximum sustained traffic rate (bits/sec) [7 - C.2.2.5.2]

B = maximum traffic burst (bytes) [7 - C.2.2.5.3]

The DOCSIS specification includes an optional parameter for the downstream called Downstream Peak Traffic Rate:

$$Peak(T) <= T * (P / 8) + 1522$$

Note that DOCSIS does not limit the instantaneous rate of a service flow. Individual packets will always travel at the native rate of the media (~40 Mbps for an Annex B downstream, and ~10 Mbps for a 3.2 MHz, 16 QAM upstream).

The minimum value for B is 1522 (one minimum size Ethernet frame) and the default is 3044 bytes although it can be as high as 20 million byes for the DS and 3 million bytes for the upstream. This is known as “Power Boost” in the cable industry.

As a general rule, the calculation is done based upon the Ethernet frame size and does not include DOCSIS framing (except for upstream concatenated frames).

To accommodate telepresence traffic as described in these tables, the value B should be greater than 65 KB and the value R should be at least 6 Mbps. The downstream peak traffic rate should be 19 Mbps (16 Mbps plus 20% packet overhead).

Different codec configurations would result in different requirements. A system in which telepresence calls are dynamically signaled and setup could optimize the values of the service flow parameters on a per call basis.

Since actual peak traffic is infrequent, admission control can be done on the average data rate numbers.

SOLUTION OVERVIEW

This section provides an overview of three proposed solutions for providing QoS for TPoD, and discusses their pros and cons. The next sections will then respectively focus on each solution and dive into the technical details.

But first, lets define exactly what information is required in any signaling messages that are trying to convey QoS information.

Problem Definition

To implement QoS with a DOCSIS Service Flow, the CMTS must have two important sets of information.

1. Packet Classification
2. Traffic Descriptor

The packet classifier is used at the CMTS in the downstream direction and the CM in the upstream direction to direct a particular flow of packets to a particular QoS queue. A DOCSIS classifier is analogous to an RSVP filterspec. The typical items of interest are:

- Destination IP address
- Source IP address (optional)
- Destination Port address
- Packet Type

The service flow encodings need to know about the size and characteristic of the media flow. The telepresence traffic has three general flows, each with multiple sub-flows. These are:

- One or more video streams
- One of more audio streams
- Signaling to one or more end points

DOCSIS has the toolset to classify each of these flows individually, and even to provide different upstream scheduling and different QoS treatment to each type of flow. But does it make sense to do so? The video flows completely dominate the other flows in a TPoD system. Video requires the most bandwidth and the least latency and jitter. The audio and signaling are along for the ride.

Also, the job of packet classification is difficult due to the existence of the VPN and NATs. Still, the IP destination address of the IP signaling packets will be different than the IP destination address of the audio and video packets.

Each solution will have to address these issues.

Solution #1: Pre-provisioned DOCSIS

This solution tries to get the most out of the existing DOCSIS specification. The solution is focused on setting the correct parameters in the CM configuration file and on the CMTS.

The advantage of this solution is rapid time to market. It allows early TPoD systems to be deployed to see what the market interest is and how well two-way real-time video works over a DOCSIS network.

The disadvantage of this solution is the lack of visibility into when a telepresence call is set up or torn down. This complicates or eliminates the ability to do admission control. There may also be a lot of manual configuration of addresses rather than the network just figuring things out. This prevents this solution from scaling well.

Thus, this solution is targeted at field trials with limited deployment.

Solution #2: On-Path Reservation

This solution uses RSVP over UDP to communicate between the home telepresence endpoint and the CMTS. In this solution, the call manager is not involved in setting up the bandwidth reservation within the network.

This solution requires the VPN router to filter and forward the RSVP over UDP messages so that they will not be tunneled. The RSVP message has extensions that permit a bi-directional reservation.

The advantage of this approach is that by only involving the local telepresence system and the local CMTS, all DOCSIS QoS reservations can be made. This benefits the deployments in which the enterprise telepresence system is behind a firewall and unable to help out.

It also is useful for scenarios where the cable operator does not own the call manager or where the call manager cannot be easily modified for PCMM. Also, because this solution operates on-path, it can operate through NATs transparently and through a VPN gateway that performs the right signaling processing.

Because only a small number of network elements need to change, this technique can be deployed with a reasonable time to market advantage.

The main disadvantage of this system is network security. The telepresence system is an untrusted entity. Bandwidth reservations may come from that IP address that are not actual telepresence calls.

Solution #3: Off-Path Reservation

This solution looks like a classical PacketCable solution. The call manager communicates through the policy engine to the CMTS to reserve bandwidth for the call.

The advantage of this solution is that the call control is more under the control of the call management system.

The disadvantage is that a sophisticated call management system has to be in place. Today, enterprise class call managers do not support PacketCable Multimedia (PCMM) interfaces, and service provider class call managers do not support video calls. As PCMM becomes more widely deployed, this solution provides a flexible path to providing bandwidth and QoS on demand services

SOLUTION #1
PRE-PROVISIONED DOCSIS

This solution uses provisions and hooks in the existing DOCSIS protocol in order to ensure that the telepresence traffic gets the required QoS treatment over the access network.

We also need to make sure that the introduction of telepresence units in the field does not disrupt service to other modems deployed in the field. This can be achieved in multiple ways

<u>Option 1: Separate Upstream and Downstream for Telepresence</u>

This option would implement a separate upstream and downstream channel on the CMTS for telepresence service. This would guarantee that service to deployed cable modems remains unchanged and the telepresence traffic is isolated from other traffic in the field. If the home telepresence subscriber is also an existing cable modem user, they would get a second modem that is dedicated for telepresence service.

While this option has severe limitation when it comes to scaling the telepresence service, it's a safe option to start field trials and to roll out the service initially. Nailing certain cable modems to certain downstream channels can be done easily using the downstream frequency parameter in the modem configuration file.

If it's too cumbersome to add the downstream frequency in the modem configuration files for existing subscribers, a simple feature can be implemented in the CMTS to designate specific upstreams and downstreams as "telepresence-only". This could be done with a specific CM Service Type TLV.

During registration, the CMTS would look for this TLV in the cable modem configuration file and, depending on whether the TLV is present or absent, it would move the cable modem to the appropriate downstream.

Since the entire upstream and downstream channel is dedicated to a telepresence setup, the QoS profile for the telepresence modem would simply dedicate the full channel bandwidth as CIR bandwidth for that modem.

Pros

1. Very simple. TPoD should always work since resources are dedicated.

2. TPoD does not interfere with the production network. Good for early field trials.

Cons

1. Inefficient use of HFC plant bandwidth and CMTS ports.

<u>Option 2: Shared Upstream and Downstream</u>

In this option, the telepresence cable modem shares the upstream and downstream channels with other cable modems connecting to the same fiber node. The modem configuration file for the telepresence modem has to be carefully configured in order to minimize impact on the other cable modems.

For this option, if the telepresence user already has a cable modem for residential internet service, we could consider connecting the telepresence system to the same modem. The simplest way to set up the home network is to connect the telepresence endpoint to the home VPN router, and connect the VPN router directly to the cable modem (with NAT disabled) instead of connecting through a home gateway or NAT device. The VPN router would be statically provisioned with a routable IP address.

This address can be used to define the appropriate upstream and downstream classifier in the modem configuration file to ensure that other non-VPN CPEs in the telepresence user's home that are connected to the shared modem do not disrupt the QoS for the telepresence system. The CMTS would provide QoS for all the traffic generated by the VPN router.

The telepresence endpoint can also be configured to mark outbound (upstream) traffic with a DSCP code point. The VPN router could be configured to preserve these markings when tunneling traffic. The VPN router would do this by copying the DSCP values from the inner packet (telepresence packet) to the outer packet (VPN encapsulation).

The DSCP could then be used in the upstream classifier at the CM to identify telepresence call traffic. Note that once the VPN traffic hits the CMTS, the CMTS typically is configured to rewrite all DSCP values on outbound packets.

A hybrid option is also possible when we use multi-channel DOCSIS 3.0 modems. A subset of channels can be set aside for telepresence service and the remaining channels could be used other traffic to and from the non-telepresence CPEs in the home.

The CMTS would be configured in such a manner that the secondary upstream and downstream service flow created on the CMTS for the telepresence traffic would pick up the appropriate channel set. With this approach, a single modem would be deployed in the telepresence user's home but the telepresence traffic would still be kept isolated from other residential traffic in the cable upstream and downstream.

Once again, CIR reservations could be used to guarantee bandwidth. That would prove inefficient. A more reasonable approach is to use a priority service flow combined with a scheduling algorithm such as real time polling service (RTPS).

There is no dynamic admission control in this approach. Thus, the number of TPoD systems provisioned per channel should be managed carefully.

Pros

1. The pre-provisioned bandwidth functionality based on classifiers and service flows defined in the cable modem configuration file is compliant with standard DOCSIS behavior and hence is supported by a current CMTS.

2. Easy to provision and manage the telepresence service through the modem configuration file

3. If priority service flows are used for TPoD sessions, the bandwidth can be over-subscribed.

Cons

1. Telepresence service has strict QoS requirements. If static CIR provisioning is used to permanently reserve bandwidth for TPoD sessions, that bandwidth is not available to other CIR services such as VoIP or other TPoD calls. The bandwidth is, however, available to best effort flows when not in use by the TPoD session.

2. No admission control. Thus a newly added TPoD session could interfere with an existing TPoD session if there is not enough bandwidth on the channel.

SOLUTION #2
ON-PATH RESERVATION

Resource Reservation Protocol (RSVP) is the standard “on-path” bandwidth reservation protocol to ensure that the right QoS treatment is setup along the path of a media flow through the network.

The RSVP message contains a “filter spec” and a “session”. Together, these specs allows a node in the network to identify the flow – source IP, destination IP, protocol as well as source and destination port if the traffic is UDP/TCP. The RSVP message also contains a “Flow spec” that define the traffic parameters (e.g. bandwidth) of the reservation.

There are several challenges with dynamic bandwidth reservation such as NAT traversal and IPsec based VPN tunnels. These are described in detail in the next few sections.

NAT Traversal:

Bandwidth reservation is complicated when RSVP signaling has to traverse one or more NATs. First, NAT devices may not handle raw RSVP packets properly. Hence the telepresence client will use the RSVP over UDP packet format that encapsulates the RSVP message inside a UDP packet destined to a well-known UDP ports, 1698 and 1699 for RSVP encapsulation.

The telepresence client would use the destination address of the remote telepresence client for these RSVP over UDP messages. The CMTS would have to intercept these RSVP over UDP packets on the upstream by filtering the well-know RSVP encapsulation UDP ports.

Another issue with NAT traversal is that the client requesting the bandwidth would use the “inside” NAT address and port by default, while the actual traffic that traverses across the network would have the external IP address and port.

In order to avoid this problem, the CMTS would have to use the source address of the actual packet instead of the local endpoint address specified in the filter spec of the RSVP message while creating the upstream and downstream classifiers.

There are several types of NAT algorithms. The basic NAT algorithm known as one-to-one NAT replaces a packet’s source IP address and source port number with a new external IP address and external source port. There is only one internal to external mapping of the source

IP address and port for all IP destination addresses and ports.

Another NAT algorithm is known a symmetrical NAT. Symmetrical NAT changes the external mapping of the source IP address and source port for each unique instance of the destination IP address and port.

With symmetric NAT, the local port for the media flow cannot be predicted or determined easily. In order for the solution to work even with symmetric NATs, the telepresence endpoint can set a wildcard for the source port in the RSVP filter spec.

Hence, the CMTS will program the DOCSIS classifier without the local endpoint port. The upstream will match based only on the source and destination IP address, protocol type and destination port. The downstream would match based only on the source and destination IP address, protocol type and source port.

Authorization of the Telepresence Client:

As the CMTS is going to process resource reservation requests from the home, it has to be ensured that the solution is not susceptible to Denial of Service attacks from malicious cable modem users. The standard mechanism to authorize a user requesting bandwidth using RSVP is by using Pull-COPS /AAA (Authentication, Authorization and Accounting) where the CMTS would offload the authorization decision to a COPS /AAA server.

Pull-COPS /AAA authorization may not be deployed in a specific cable network. Hence another option is to add a vendor-specific TLV in the modem configuration file to indicate to the CMTS that it can accept RSVP requests from clients connected to that modem.

Teleworker VPN:

As shown in Figure 4, in most cases the telepresence endpoint would be connected to a home VPN router. In the VPN case, the CMTS will not be able to snoop packets sent from the telepresence client. This is because the VPN tunnel would extend from the home VPN router to the enterprise VPN router and the CMTS would be in the middle. By default, RSVP over UDP packets sent inside the IPsec VPN tunnel would also get encrypted along with the actual telepresence flow.

This enterprise VPN challenge can be solved by adding a new feature to the home VPN router. The VPN router will have to filter for RSVP over UDP packets and will forward the RSVP messages unencrypted outside the VPN tunnel in the upstream direction.

While this change in the VPN router would ensure that the CMTS can snoop the RSVP over UDP message, there is a further challenge in the VPN scenario. The source and destination address and port information in the RSVP message would not be useful for the CMTS, since the packets of the telepresence flow would still be encrypted within the IPsec VPN tunnel. (and therefore encapsulated with packets whose IP addresses are those of the VPN gateways).

To solve this problem the CMTS will instead have to store the source and destination IP address in the IP header of the packet carrying the RSVP over UDP message, and use those to create the classifier. When the VPN router forwards the RSVP over UDP message to the CMTS, it uses the same source and

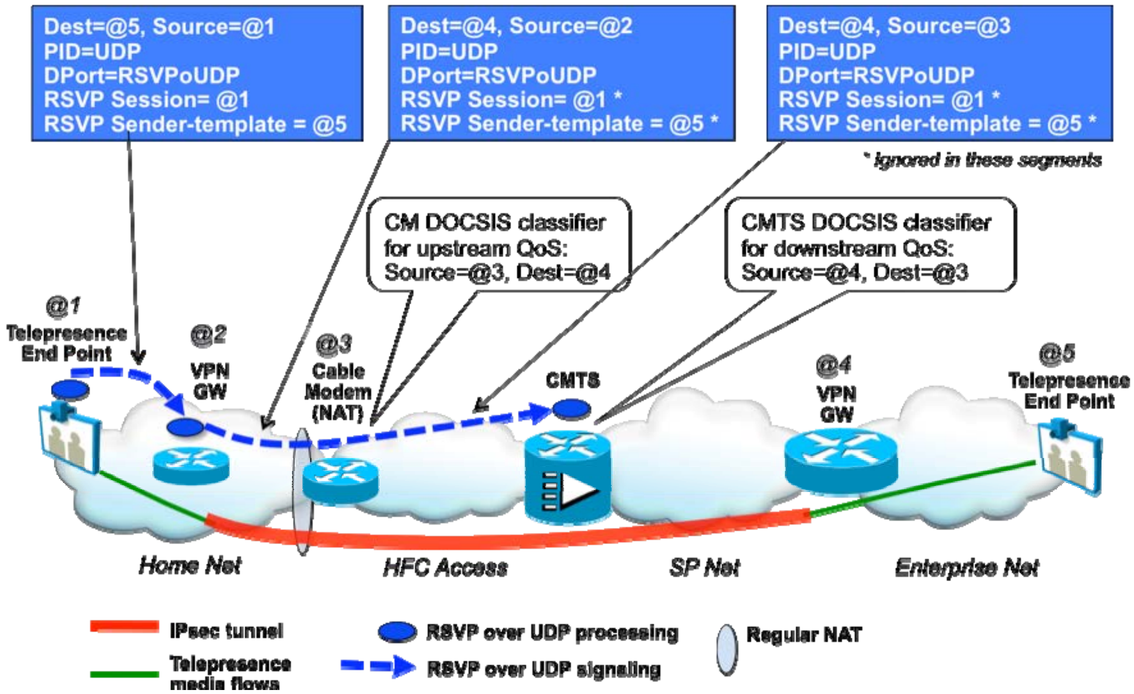


Figure 4 - RSVP over UDP Packet Flow

destination address that it will use to send the corresponding media packets within the IPsec VPN tunnel.

Hence, the CMTS can create a DOCSIS classifier that matches all the traffic within the VPN tunnel. The VPN router would have to insert a flag in the RSVP message that the CMTS would interpret as “don’t use the address and port information within the RSVP message. Instead, create the DOCSIS classifier using the source and destination address of the RSVP over UDP packet”.

In Figure 4, the telepresence endpoint sends an RSVP over UDP packet with a destination address of @5 (the far end telepresence endpoint) and a source address of @1. The VPN router intercepts this packet, sets the above mentioned flag in the RSVP message, and transmits the packet unencrypted using a destination address of @4 (the far end VPN router) and a source address of @2, which represent the near end of the VPN tunnel.

Figure 4 also shows an additional NAT operation that is performed by the cable modem which changes the source address of the packet to @3 and keeps the destination address unchanged as @4. The CMTS filters the RSVP over UDP packet and uses the destination address @4 and source address @3 to create the upstream and downstream classifiers to provide appropriate QoS to the bi-directional telepresence media.

<u>Bandwidth Reservation in the Downstream Direction:</u>

Typically RSVP messages describe unidirectional flows. If bandwidth reservation is required in both directions, it is expected that each endpoint sends RSVP requests for its own transmission in the forward direction.

Telepresence traffic is also bi-directional, but it's easier to provision the network such that the CMTS is expected to process RSVP messages only from the local telepresence endpoint. There are several issues with trying to support RSVP messages for downstream traffic from the remote telepresence endpoint

1. In case of the teleworker VPN case, traffic from the remote telepresence endpoint, including RSVP messages, would be sent encrypted inside the VPN tunnel and cannot be seen by the CMTS

2. Even for the non-VPN case, RSVP messages coming from the remote telepresence endpoint may be filtered out somewhere in the network – especially if that endpoint belongs to a different service provider.

3. The CMTS would be open to Denial of Service attacks from the Internet. It's easier to authorize the local telepresence endpoint requesting QoS than it would be to authorize a remote telepresence endpoint that the operator has no control over.

In order to avoid these issues, bidirectional RSVP can be used as defined in the PacketCable specification. This would enable the local telepresence endpoint to also request bandwidth from the CMTS for the reverse flow in the downstream direction.

<u>Step-Wise Solution</u>

1. As part of the call setup, the video resolution is negotiated. The local telepresence endpoint calculates the bandwidth to match this resolution and sends an RSVP over UDP message that describes the upstream and downstream bandwidth requirements. The source port is wildcarded in the filter spec to avoid issues with symmetric NATs.

2. If a VPN router is present, then the RSVP over UDP packet is intercepted by the VPN router, which changes the source and destination IP address of the packet to be the same as the source and destination of the IPsec tunnel. The VPN router also sets a flag in the RSVP message requesting the CMTS to "ignore the filter spec" within the RSVP over UDP message

3. When the CMTS intercepts the RSVP over UDP message, it ensures that the message is being received from a client behind an authorized modem. If authorization fails, the CMTS would drop the packet. If it succeeds it would continue to process the RSVP request.

4. If the "ignore the filter spec" flag is not set, the CMTS creates an upstream classifier based on source, destination IP address, protocol, and destination port. It also creates a downstream classifier based on source and destination IP address, protocol and source port.

   To avoid problems with symmetric NATs, in both classifiers the local telepresence endpoint UDP ports are not used for classification. If the "ignore the filter spec" flag is set, then the CMTS creates upstream and downstream classifiers based only on the source and destination IP address of the received RSVP over UDP packet.

5. The CMTS tries to admit the upstream and downstream service flows based on the TSPEC. If it fails, the CMTS

sends back an RSVP call admission control (CAC) reject notification. When the telepresence endpoint receives this notification, it can try to negotiate a lower video resolution and motion with the remote endpoint and send a new RSVP request, thereby restarting operation at step 2 above.

6. Once the local telepresence endpoint succeeds in reserving the required bandwidth, it completes the SIP call setup and telepresence media begins to flow between the two telepresence endpoints.

7. When the telepresence call is done, the local telepresence endpoint sends an RSVP teardown message which is again intercepted by the CMTS. When the CMTS receives this message, it deletes the upstream and downstream service flows created for the telepresence call.

8. RSVP uses soft state to manage the QoS reservation in the network. This soft state has to be periodically refreshed by the local telepresence endpoint by sending RSVP message periodically. If the CMTS does not receive these periodic RSVP messages, it will eventually timeout the bandwidth reservation and will tear down the service flows created for the RSVP request.

   This behavior ensures that bandwidth is reclaimed even in the case where the TP call is not gracefully torn down. An example where this is need would be if the local telepresence endpoint is powered down before it can end the telepresence call and send the RSVP teardown message.

Pros

1. This is an on-path bandwidth reservation solution that can be deployed independent of various flavors of off-path solutions (PacketCable, PCMM etc) that the operators may be using for existing voice services.

2. Complex NAT and VPN scenarios can be supported with this solution.

3. An on-path RSVP over UDP solution for TPoD can simultaneously exist with an off-path PCMM system for VoIP.

Cons

1. Even though the reference PCMM architecture does include support for RSVP clients that request QoS resource directly from the CMTS, this approach has not been fully defined in the standards specification. Hence initial implementations of this solution would be vendor specific or need to be further standardized.

SOLUTION #3 OFF-PATH RESERVATION

Nailing down bandwidth as recommended in solutions #1 and #2 will be a challenge for many cable operators. An alternative approach is to use the DQoS infrastructure as defined in PacketCable MultiMedia to dynamically create and tear down service flows as needed.

Rather than statically provisioning the QoS within the access network, the first initial SIP Invite from the Telepresence client to the call manager would trigger a secure sequence of events that would allow the CMTS to dynamically provision the QoS while still meeting the goals of solutions #1 and #2. It is envisioned that there would be no need for dedicated bandwidth and that using a shared model would provide sufficient capacity.

The call manager could be located within the Service Provider domain or be provided by a third party telepresence service. The general call flow remains the same with the addition of authentication. The use of secure protocols with authenticated is a requirement as the QoS requests will be triggered by parties external to the cable operators network.

The basic flow for requesting QOS from a hosted call manager external to the cable network is as follows.

1. Local TelePresence (CPE) issues SIP Invite to call manager.

2. Call Manager sends SIP based QOS request to the Service Edge Proxy using a secure transport protocol such as SIPS, TLS or HTTPS

3. Edge Proxy validates request and sends the request to application manager.

4. Application Manager translates SIP request to PCMM (COPS) and sends to Policy Manager

5. Policy Manager validates request, determines resources needed and sends PCMM gate-set to CMTS

6. CMTS determines resource availability, creates gate and communicates with cable modem using DSx (Dynamic Service Flow) messaging

7. Cable Modem sets up service flow

8. Telepresence media flows flow bi-directionally with the proper QoS.

Multiple checkpoints are used to maintain security:

- the 3rd party call manager must have their specific certification signed by the site they are contacting,

- the edge proxy must have details on the contacting site,

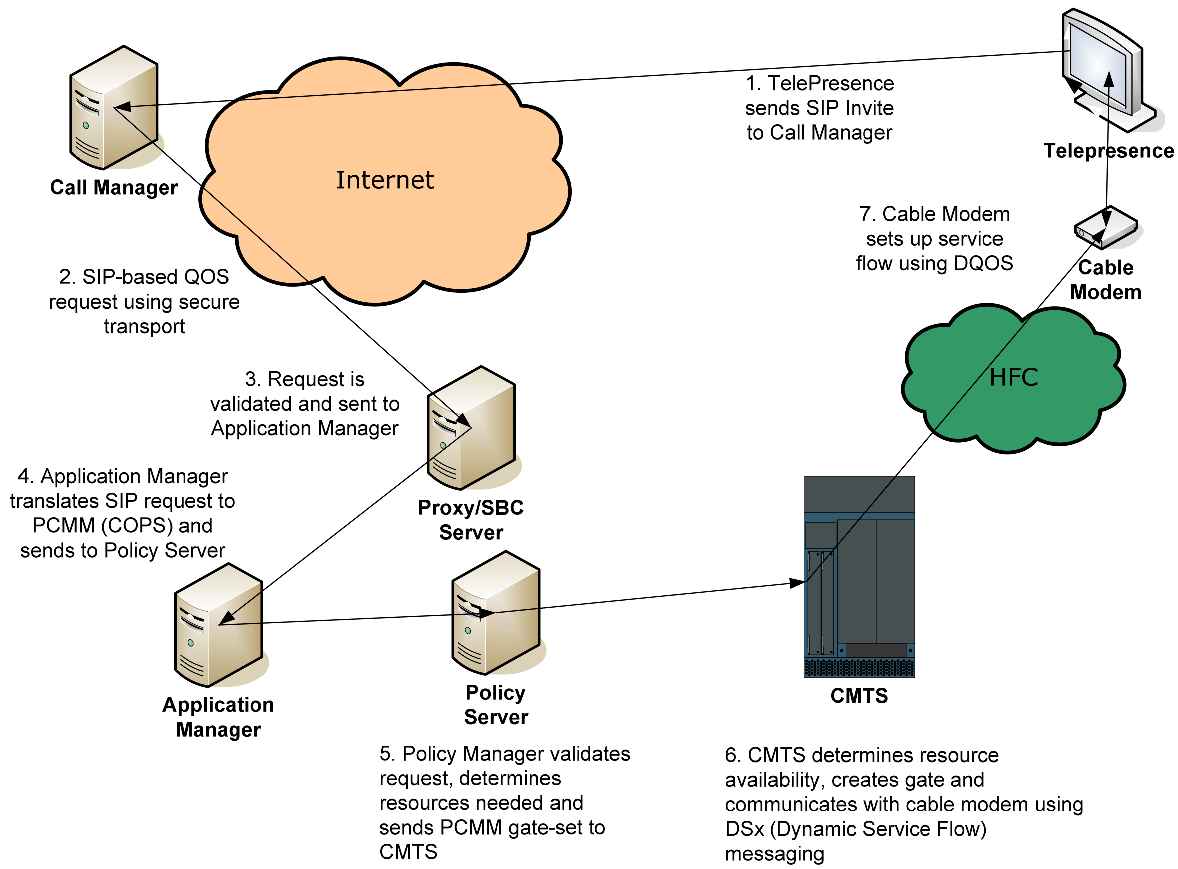


Figure 5 - TPoD PCMM System

- the application manager must have knowledge of the edge proxy and allow requests from that device,
- the policy manager must have policies that support the requests being made, and
- the CMTS must allow requests from the policy manager.

The actual flows created are equivalent to the ones created in the on-path reservation system, but rather than being static flows they are created dynamically using PCMM [4][5].

The addressing challenges from the home NAT and VPN still exist with this call setup scenario. The policy messages that arrive at the CMTS have to contain the actual IP addresses that the CMTS will see, as opposed to the IP addresses that the telepresence system uses. This is a non-trivial problem and will be the subject of additional research.

Pros

1. Uses PCMM for on demand bandwidth management. PCMM is a well-defined specification and CMTS products on the market support PCMM.

2. Does not require dedicated bandwidth

3. Will not impact the quality of voice calls

4. Secure protocols with authentication

Cons

1. Requires the operator to deploy a PCMM infrastructure.

2. Multiple points to manage increases complexity.

3. Introduces potential single points of failure.

SUMMARY

Telepresence is a new technology for conducting virtual meetings with a real-time, in-person experience. As the adoption of this new technology accelerates, users will increasingly want to communicate via telepresence in locations served by DOCSIS networks, such as executives' home offices. Cable operators will have an opportunity to offer advanced network services to support customer-owned telepresence systems, and could also offer a managed telepresence service to customers without a complete in-house telepresence system.

This paper described the challenges in supporting telepresence over DOCSIS, particularly in the areas of dynamic bandwidth reservation and quality of service. Three solutions were proposed and assessed for viability in TPoD deployments.

Pre-provisioning the DOCSIS network for telepresence services can be a suitable approach for field trials and limited deployment, but not for large-scale deployment due to the inefficient use of DOCSIS network resources.

On-path reservation enables the home office telepresence system to request DOCSIS network resources directly from the CMTS, thereby eliminating the need for integration with an external call management system. However, additional authentication mechanisms may be necessary since the TPoD end point is considered to be untrusted.

Off-path reservation enables the telepresence system to manage the reservation of end-to-end network resources, but requires further integration of DOCSIS network resource management functions with the telepresence call management system.

With further investigation and development, the on-path and off-path reservation approaches proposed can be viable solutions for large-scale TPoD systems.

ACRONYMS AND ABBREVIATIONS

AAA	Authentication, Authorization and Accounting
AAC-LD	Advanced Audio Encoding Low Delay
CAC	Call Admission Control

CIR	Committed Information Rate
CM	Cable Modem
CMS	Call Management Server
CMTS	Cable Modem Termination System
CODEC	Coder – Decoder
COPS	Common Open Policy Service
CPE	Customer Premise Equipment
DOCSIS	Data Over Cable System Interface Specification
DSCP	Differentiated Services Code Point
DSx	Dynamic Service Flow
HD	High Definition
HTTPS	Secure HTTP
IP	Internet Protocol
IPsec	IP Security
MAC	Media Access Control
NAT	Network Address Translation
PABX	Private Automated Branch Exchange
PCMM	PacketCable Multimedia
QoS	Quality of Service
RSVP	Resource Reservation Protocol
RTP	Real Time Protocol
SD	Standard Definition
SDV	Switched Digital Video
SIP	Session Initiated Protocol
SIPS	Secure SIP
SLA	Service Level Agreement
TLS	Transparent LAN Services
TLV	Type Length Variable
TPoD	Telepresence over DOCSIS
TSPEC	Transmission Specification
UDP	User Datagram Protocol
VOD	Video on Demand
VPN	Virtual Private Network

REFERENCES

1. Dierks, T. and E. Rescorla, RFC5246, The Transport Layer Security Protocol Version 1.2, http://tools.ietf.org/html/rfc5246

2. Freesoft.org, SSL V3.0 Specification, http://www.freesoft.org/CIE/Topics/ssl-draft/3-SPEC.HTM

3. Wagner, David, and Bruce Schneier, Analysis of the SSL 3.0 Protocol , The Second USENIX Workshop on Electronic Commerce Proceedings, USENIX Press, November 1996

4. Cablelabs, PacketCable™ Dynamic Quality-of-Service Specification, I03, 2007 http://www.cablelabs.com/specifications/PKT-SP-DQOS1.5-I03-070412.pdf

5. Cablelabs, PacketCable™ Network-Based Call Signaling Protocol Specification, I03, 2007, http://www.cablelabs.com/specifications/PKT-SP-NCS1.5-I03-070412.pdf

6. Cisco, Telepresence Room Requirements, 2009, http://www.cisco.com/en/US/solutions/ns669/networking_solutions_products_genericcontent0900aecd80554cb2.html

7. Cablelabs, MAC and Upper Layer Protocols Interface Specifications, DOCSIS, I09, 2009, http://www.cablelabs.com/specifications/CM-SP-MULPIv3_0-I09-090121.pdf

8. Cisco, Telepresence Network Systems 2.0 Design Guide, August 1, 2008, http://www.cisco.com/en/US/docs/solutions/Enterprise/Video/TelePresence_Network_Systems_2.0_DG.pdf

THE CHALLENGES OF STOPPING ILLEGAL PEER-TO-PEER FILE SHARING

Kevin Bauer, Dirk Grunwald, and Douglas Sicker
Department of Computer Science, University of Colorado

Abstract

Illegal file sharing within peer-to-peer networks has become a significant threat to the film and recording industries. Furthermore, such peer-to-peer protocols often use a significantly large amount of bandwidth relative to other protocols, complicating network management. In the past, copyright enforcement agencies have been hired to investigate users who appear to be sharing files illegally. In addition, broadband Internet service providers (ISPs) have actively throttled peer-to-peer traffic in an effort to reduce load on their networks. We observe that an "arms race" has begun between file traders and copyright holders/ISPs in which the file traders have started to develop techniques for hiding their involvement in the transfer of copyright-protected media files. In response, the copyright holders'/ISPs' investigative tactics are evolving to match the changing strategies. In this paper, we provide a survey of the current tactics used by file traders to hide their involvement in illegal file transfers and speculate about future strategies that may emerge on both sides of the arms race.

1. INTRODUCTION

Peer-to-peer (P2P) networks have recently grown in popularity for a variety of applications such as content distribution, streaming multimedia, and voice-over-IP. P2P networks are often built around a decentralized architecture to distribute data in a manner that offers high availability of content, inherent fault-tolerance, and efficiency. While P2P networks offer several important advantages over traditional client/server architectures, experience has shown that these networks are sometimes used to distribute copyright-protected media illegally.

P2P file sharing involving copyright protected content presents significant problems for network management and copyright enforcement. P2P networks utilize a large amount of bandwidth, particularly upstream bandwidth, complicating network management for broadband Internet service providers (ISPs), particularly during times of peak network utilization. In addition, the illegal dissemination of copyright-protected media is an obvious problem for the respective copyright holders that may result in a loss of revenue. As a consequence, there is ample incentive for both broadband ISPs and copyright holders to work to stop the proliferation of file sharing within P2P networks.

Our primary goal in this paper is to assume a proactive position toward understanding the current techniques for distributing and hiding copyright-protected content within P2P networks. We focus our discussion primarily on BitTorrent, since it is currently the most popular P2P protocol for file sharing. We observe that an arms race has already begun between file traders and copyright holders in which the file traders have started to develop techniques for hiding their involvement in the transfer of a copyright-protected media file. In response, the investigative tactics used by copyright holders are evolving to match these changing strategies. We provide a survey of the current tactics used by file traders to hide their involvement in illegal file transfers and speculate about future strategies that may emerge on both sides of the arms race.

The remainder of this paper is organized as follows: In Section 2, we provide an introduction to BitTorrent, the most common P2P network in use today. In Section 3, we describe the most common techniques that copyright holders have used to track the distribution of their copyright-protected content. These strategies often include locating individual users, issuing DMCA takedown notices, or even pursuing more serious legal actions against suspected file sharers. We also discuss the past tactics used by broadband ISPs to throttle BitTorrent traffic. In response to the copyright holders' desire to protect their content, there is now significant incentive for P2P users to shed their network identities and enjoy a certain degree of anonymity. In addition, to avoid traffic shaping, P2P users have incentive to try to hide the nature of their traffic using encryption. We discuss the current tactics used to evade ISP traffic shaping practices and copyright enforcement in Section 4. In Section 5, we describe the most common methods for achieving anonymity online and present evidence from a prior study that P2P users are beginning to use BitTorrent anonymously. We also briefly outline prior proposals to incorporate anonymity mechanisms into P2P networks themselves. We also speculate about the future tactics that may be employed to distribute copyright-protected content. Finally, we provide concluding remarks in Section 6.

2. BACKGROUND ON BITTORRENT

BitTorrent has become one of the most popular peer-to-peer protocols for file sharing. A key feature of file transfers with BitTorrent is that files are not transferred sequentially, as in protocols such as HTTP or FTP. Instead, files are broken into fixed-size *pieces* and are transferred in parallel. This enables BitTorrent to transfer data very quickly and efficiently among a large number of peers. As a result, the protocol can be particularly greedy with regard to bandwidth.

To share a file with BitTorrent, a metadata file containing the piece length, a SHA1 hash of each piece to ensure integrity, and a URL to a *tracker server* is published through an out-of-band mechanism. These metadata files are often hosted by sites such as isoHunt [4] and The Pirate Bay [8]. Once a peer obtains the metadata file for a desired file, the peer contacts the tracker server to obtain a list of other peers who are sharing the file. In the process, the peer also registers itself with the tracker. Other peer discovery mechanisms are available, including distributed trackers built upon distributed hash tables (DHTs) and gossip protocols; however, the centralized tracker server method is simple, and thus, the most commonly used. The peer finally issues requests for *blocks*, or sub-pieces (typically 16KB), from other peers. Peers who possess the complete file are called *seeders* and peers who do not are referred to as *leechers*. A file transfer using BitTorrent is illustrated in Figure 1.

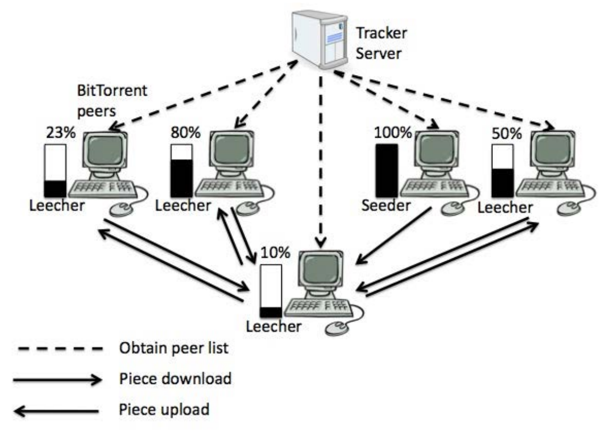


Figure 1: A file transfer with BitTorrent.

The protocol's precise sequence of messages to initiate a data transfer is provided in Figure 2. First, a leecher establishes communication with another peer by exchanging *handshake messages*. The handshake consists of a plain-text protocol identifier string, a SHA1 hash that identifies the file(s) that are being shared, and a pseudo-random peer identification string. After both peers have exchanged handshake messages, the leecher sends a *bitfield message*, which contains a bit-array data structure that concisely describes the pieces of the file that the peer has already obtained. After exchanging bitfields, the leecher knows which pieces the other peer can supply, and it proceeds by requesting specific blocks. Once a leecher has obtained a piece, it notifies other peers by sending a *have message*. More details about BitTorrent can be found in its protocol specification document [1].

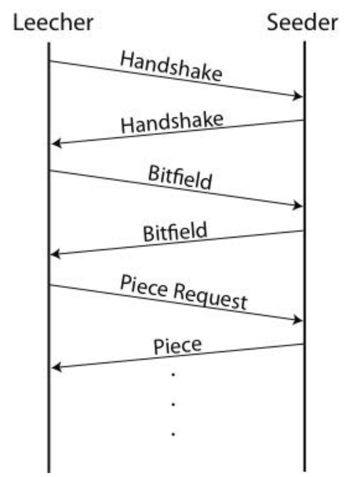


Figure 2: BitTorrent's message exchange to initiate a piece transfer.

Due to its aggressive behavior with regard to bandwidth usage (often described as *swarming* behavior) BitTorrent presents significant network management challenges for ISPs. BitTorrent is often configured to open many TCP connections simultaneously, sometimes using all bandwidth available to the user. For ISPs, this behavior may complicate network management, especially during times of peak utilization. As a result, many ISPs have actively attempted to regulate BitTorrent's bandwidth usage. In the next section, we discuss the common tactics used in the past to investigate copyright violation and to control bandwidth consumption.

3. INVESTIGATIVE TACTICS

BitTorrent is not used solely for copyright violation. There are many legitimate uses including obtaining software updates, downloading Linux ISO images, and sharing non-copyright protected movies and music. However, given the unfortunate reality that BitTorrent is being used to distribute copyright-protected movies and music, BitTorrent has caught the attention of organizations such as the Motion Picture Association of America (MPA) and the Recording Industry Association of America (RIAA). Furthermore, as a consequence of BitTorrent's aggressive network behavior, it consumes excessive amounts of bandwidth relative to other protocols, thus complicating network management for broadband ISPs. Since copyright holders and network operators *both* have incentive to curtail BitTorrent usage (though for different reasons), both have initiated campaigns aimed at slowing the proliferation of BitTorrent usage. In this section, we present an overview of the strategies employed by broadband network operators and entities representing copyright holders.

3.1 Broadband ISP Tactics

Due to the challenges that BitTorrent presents for network management, some broadband ISPs have recently adopted policies aimed at disrupting or even blocking BitTorrent traffic within their networks [3]. In particular, Comcast received extensive publicity for their use of Sandvine to specifically target BitTorrent flows with forged TCP RST (reset) packets, causing a targeted TCP connection to be prematurely and abruptly closed. This policy has been criticized by network neutrality proponents and consumer advocates in part because there was little transparency and disclosure regarding these practices. In response, researchers have produced a variety of techniques and tools [12, 21, 26] to detect this type of traffic manipulation by ISPs.

3.2 Copyright Holder Tactics

Since the tracker servers that enable illegal file transfers are often hosted in foreign countries where legal recourse against such activity is limited [11], the representatives such as the MPA and RIAA acting on behalf of copyright holders have initiated a large scale investigative effort to identify and pursue individual users participating in illegal file transfers. Such companies as Media Defender [5] and Safenet [10] have been hired to passively monitor the tracker servers for copyright infringing file transfers to obtain the list of IP addresses of the users who are participating in the file transfers. Recall that BitTorrent's primary peer discovery mechanism requires that the IP addresses of other peers participating in the file transfer be publicly advertised.

A recent study [25] found that these investigators obtain the list of IP addresses from the trackers and send an ICMP echo (ping) message to each end-host to ensure that it is alive. These investigators often target suspected file sharers with DMCA takedown notices and even have initiated more formal legal proceedings in some cases.

However, as the authors of [25] observe, this type of investigative strategy is problematic, since it is easily prone errors, especially false positive identification. False positives occur when users are wrongly accused of actively participating in the file sharing. False positives may occur as a result of normal network activity, for example, if a user obtains a DHCP lease on an IP address that had previously participated in the file transfer. However, false positives may also occur by actively polluting a particular tracker's peer list with arbitrary IP addresses. It is possible to explicitly register arbitrary IP addresses to a tracker, thus implicating any end-host in the file sharing. The authors of [25] poignantly demonstrated the shortcomings of the current investigative tactics by registering devices such as networked printers and wireless access points to tracker lists, and subsequently receiving DMCA takedown notices for these devices' alleged involvement in illegal file transfers.

Another study has found that representatives of the copyright holders actively participate in illegal BitTorrent file transfers and attempt to launch a variety of "attacks" on leechers [19]. In particular, this study identified two distinct attack strategies: *fake-block* and *unresponsive peer* attacks.

The fake-block attack occurs when peers operated by copyright enforcers deliberately reply to piece requests with invalid blocks of data. When an entire piece is obtained, the leecher verifies the piece's integrity with a SHA1 hash. However, the hash fails due to the invalid block(s). This requires the leecher to download the entire piece again (which wastes time and bandwidth), since it does not know precisely which block is corrupt.

The unresponsive peer attack occurs when a peer completes a valid BitTorrent handshake and bitfield exchange (which is the prelude to the data transfers), but the peer refuses to send any data. This attack also causes leechers to waste time and bandwidth exchanging control messages with peers that have no intention to provide pieces of the file.

The aforementioned study found that both of these attacks are relatively common. In addition, while these attacks may cause a download to take up to 50% longer, they are ineffective at stopping BitTorrent file transfers altogether.

4. RESPONSE TO ANTI-P2P CAMPAIGNS

Given the techniques used to mitigate BitTorrent usage by network operators and copyright holders, file sharing tactics have begun to evolve to incorporate mechanisms to prevent blocking by ISPs and to avoid legal sanctions by entities representing the copyright holders. In this section, we provide an overview of the next phase in the arms race between the file sharers and ISPs/copyright holders.

4.1 Concealing BitTorrent from ISPs

In an attempt to frustrate traffic shaping or blocking by ISPs, an obfuscation technique called Message Stream Encryption has been proposed as an optional extension to the

BitTorrent protocol [6]. Message Stream Encryption requires that pairs of communicating peers perform a Diffie-Hellman key exchange to agree on a shared secret and then encrypt the BitTorrent header (and optionally the payload) using the RC4 stream cipher. This feature is available in Vuze [14], μTorrent [13], and other BitTorrent clients. In order to use the encryption feature, a peer can only communicate with other peers that support the encryption feature.

However, protocol header encryption and payload encryption are relatively ineffective at obfuscating the traffic type, since the packet size characteristics remain intact. BitTorrent traffic has a distinctive signature consisting of large bidirectional data transfers, thus it would still be relatively easy to detect despite encryption. Furthermore, sophisticated techniques based on statistical or machine learning methods could be applied to detecting if an encrypted stream is BitTorrent traffic [24, 27, 28]. Encrypting BitTorrent does, however, require the ISP to develop and deploy these types of sophisticated traffic classification techniques, which may be expensive and time consuming. Furthermore, these traffic classification techniques are not perfect and may have non-negligible classification errors. This could result in a scenario in which other protocols are misclassified as BitTorrent.

In addition to encryption, it is possible that BitTorrent may adopt a UDP transport mechanism, which is rumored to be included to a future version of μTorrent [13]. The UDP transport would render the traffic shaping practices using forged TCP RST packets ineffective.

4.2 Evading Copyright Authorities

Since the large scale investigations carried out by entities representing copyright holders have resulted in DMCA takedown notices and the potential for more serious legal sanctions, counter-strategies have emerged in an attempt to frustrate these investigations. One common strategy is to intentionally introduce randomly selected IP addresses into the tracker lists (called pollution). For instance, the Pirate Bay, a popular tracker-hosting site, has implemented this policy [9]. This tactic increases the potential for false positives to a level that may not be tolerable for the investigators. For instance, wrongly accusing innocent users of sharing files illegally could have serious consequences for the copyright holders including negative public opinion or even sanctions from government regulators.

In addition, services such as PeerGuardian [7] have emerged to provide IP address blocking capabilities for P2P applications. For instance, this service could be used to block all IP addresses that are suspected of active pollution or monitoring.

More extreme techniques to evade the copyright enforcement authorities are even starting to become common. For instance, BTGuard [2] offers a pay proxy service in which subscribing users can encrypt and tunnel their BitTorrent traffic through a proxy server hosted in a foreign country. Using such a service, when a BitTorrent client registers itself with a tracker server, the tracker server knows only the proxy's IP address, and consequently, the copyright enforcers also can observe only the proxy's IP address. Provided that the proxy service does not keep records of its clients' activity, it is difficult to determine the identity of the real client. The encrypted tunnel may also frustrate ISPs' BitTorrent traffic throttling, but as described in Section 4.1, traffic analysis techniques exist that may reveal the underlying type of traffic within the encrypted flow.

5. EMERGING STRATEGIES TO HIDE P2P

The changing tactics employed by file sharers and copyright holders/ISPs can best be described as an arms race of evolving strategies and counter-strategies. In this section, we discuss the current cutting-edge and possible future strategies that P2P users may apply to obfuscate their activities from their ISPs to avoid traffic throttling and to hide from copyright enforcement authorities. Technologies that enable end-users to shed their network identities and enjoy anonymity while online are one line of emerging strategies. We present evidence to suggest that P2P users may be beginning to use anonymous networks to avoid traffic throttling by their ISPs and avoid identification and subsequent legal action by copyright enforcement authorities.

5.1 P2P and Anonymous Networks

The Internet and its fundamental protocols (*i.e.*, TCP/IP) were designed with no regard for anonymous network access. However, recent research in anonymous communications has provided the designs and implementations of anonymous overlay networks based on onion routing [22] and mix networks [17]. Anonymous networks are currently being used throughout the world for a variety of applications, often enabling freedom of speech and press within repressive countries.

Tor has become the most popular overlay network for anonymizing TCP-based applications [20]. Tor is able to provide a stronger form of anonymity than the proxy server approach (described in Section 4.2) because it is built around a decentralized design; therefore, no single entity knows both the source and the destination of an anonymous flow. Tor's system architecture (illustrated in Figure 3) consists of three components: *Tor routers*, *Tor proxies*, and *directory servers*. Tor routers forward TCP traffic on behalf of participating users by employing a layered encryption scheme similar to onion routing. A user running Tor proxy software creates a virtual circuit of precisely three Tor routers. First, the Tor proxy obtains a list of all available Tor routers from the set of trusted directory servers. Next, the Tor proxy establishes shared secret keys with each of the three Tor routers on the circuit and encrypts the user's data with each key in a layered fashion. Upon receiving a packet, the Tor router removes its layer of encryption and forwards the packet to the next router in the path. Once the final layer of encryption has been removed, the last Tor router forwards the payload to the destination server.

It is important to note that only the first Tor router on the path (called the *entry guard*) knows the true identity of the client, and only the last Tor router on the path (called the *exit router*) knows the identity of the destination server. Tor provides a strong degree of anonymity, subject to the assumption that it is difficult for a single entity to control both the first and last Tor routers on a user's virtual circuit [15]. However, an ISP or group of colluding ISPs *could* feasibly monitor the links entering and exiting the Tor network and perform traffic analysis to link the clients and destinations.

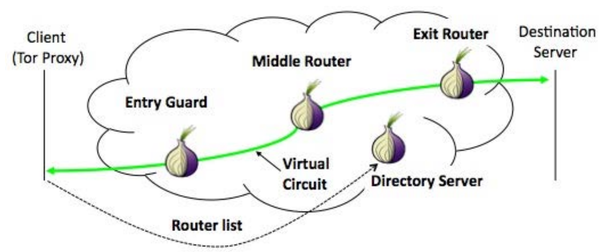


Figure 3: Tor's system architecture.

In prior work, we characterized how Tor is used in practice [23]. In particular, we analyzed the application-layer protocols that are commonly used with Tor. We discovered that individual users are starting to use Tor to conceal BitTorrent activity. While operating a Tor router for four days, we observed over 430,000 BitTorrent connections leaving the Tor network, accounting for approximately 285GB of traffic. While the number of BitTorrent connections was relatively low in comparison to other protocols such as HTTP and SSL, the amount of traffic transported over these connections was surprisingly high. However, there are plug-ins for popular BitTorrent clients (such as Vuze) that make it easy to connect the BitTorrent client to the

Tor proxy software. In addition, given the past practices of monitoring and profiling users suspected for participating in illegal BitTorrent file transfers, it is reasonable to suspect that the number of users who turn to strong anonymity mechanisms like Tor may increase in the future.

In addition to anonymizing overlay networks like Tor, it is possible that P2P users may look to other sources for anonymity. For example, the design of an anonymity layer specifically tailored for BitTorrent has been published [16]. The protocol, called BitBlender, works by introducing special peers called *relay peers* into the BitTorrent system architecture. These peers do not actively share any file(s), but merely proxy piece requests and responses on behalf of other users actively sharing the file(s).

BitBlender's primary goal is to introduce a certain degree of *plausible deniability* for peers listed by the trackers. With BitBlender, a copyright enforcement authority cannot simply examine the tracker's peer list to obtain an accurate view of the peers who are involved in the sharing. The copyright enforcer must actively participate in the file sharing and conduct sophisticated traffic analysis in order to have any chance of isolating the real active peers. However, since the relay peers exhibit many of the same protocol-level behaviors as the real peers, it may still be difficult to isolate the real peers. While BitBlender is only a proof-of-concept design (*i.e.*, there is currently no available implementation), it is possible that this relay strategy may be incorporated into popular BitTorrent clients in the future.

5.2 End Game

Until this point, we have discussed the current and emerging strategies used for hiding illegal file sharing within P2P networks. Next, we examine how the shifting strategies used to stop this type of file sharing may cause a radical shift in content hiding strategies and provide a speculative discussion of the tactics that may be used by file sharers in the future.

One potential technique for hiding content is to use a distributed and anonymous data store. Freenet [18] is a P2P network in which peers can store and retrieve files that are named by location-independent keys. To retrieve a file, a user computes a hash of the content's description - which is used as the look-up key - and forwards a retrieval request to another peer in the network. The request is forwarded through potentially many peers until the content is found, upon which, the content is sent back to the original requester through each peer that forwarded the initial request. In doing so, the replying peer does not know who actually initiated the request, and the requesting peer does not know where the data is stored. Furthermore, peers hosting files only know the hash of the file's description, so they remain agnostic regarding the content they host.

This content hiding strategy offers significant advantages over BitTorrent. The Freenet-style of content hosting and retrieval offers relatively strong deniability for both the hosts and the retrievers. Furthermore, this strategy significantly complicates investigations launched by anti-piracy agencies.

In addition to Freenet-style P2P networks, Tor offers the ability to host hidden services within the Tor network. A hidden service can be established in such a manner that the service's owner does not reveal their identity. It is difficult to shut down such a service, since its location is hidden. More details on hidden services in Tor can be found in Tor's design document [20].

Tor’s hidden services provide strong anonymity for both the service's host and those who download content, and represent perhaps the most *radical* counter-measure to anti-piracy efforts. While there is a significant performance penalty associated with using hidden services (*i.e.*, additional download time), users may be willing to cope with this limitation if there is sufficient incentive, perhaps such as avoiding prosecution. If widespread usage of Tor's hidden services for illegal file sharing becomes a popular counter-strategy, there may be little recourse for anti-piracy authorities to stop it.

6. CONCLUSION

In this paper, we presented an overview of the current strategies for identifying illegal file sharers and a survey of the counter-measures that file sharers have employed in response. We observe that a strategic "arms race" has started as the tactics for pursing illegal file sharers and hiding evolve. In addition, if this arms race continues, we speculate about the future tactics that may be used to hide illegal file sharing and conclude that strong anonymity mechanisms and location-hidden services may be the final resort of the illegal file sharing movement. Since this implies a somewhat bleak outlook for the anti-piracy authorities, we conclude that alternative strategies – including tiered bandwidth pricing models to discourage high bandwidth usage on broadband networks and lower-cost media distribution methods – should be investigated to provide individuals with more economic incentives to obtain content from legitimate sources.

REFERENCES

[1] BitTorrent protocol specification. http://wiki.theory.org/BitTorrentSpecification.

[2] BTGuard - BitTorrent Anonymously. http://btguard.com.

[3] Comcast is using Sandvine to manage P2P connections. http://www.dslreports.com/forum/r18323368-Comcast-is-using-Sandvine-to-manage-P2P-Connections.

[4] isoHunt - the BitTorrent and P2P search engine. http://isohunt.com.

[5] Media Defender - P2P anti-piracy and P2P marketing solutions. http://www.mediadefender.com.

[6] Message stream encryption. http://www.azureuswiki.com/index.php/Message_Stream_Encryption.

[7] Peer Guardian 2. http://phoenixlabs.org/pg2.

[8] The Pirate Bay. http://thepiratebay.org.

[9] Pirate Bay tricks anti-pirates with fake peers. http://torrentfreak.com/the-pirate-bay-tricks-anti-pirates-with-fake-peers-081020.

[10] Safenet Inc: The foundation for information security. http://www.safenet-inc.com.

[11] Secrets of the pirate bay. http://www.wired.com/science/discoveries/news/2006/08/71543.

[12] Switzerland network testing tool. http://www.eff.org/testyourisp/switzerland.

[13] µTorrent - the lightweight and efficient BitTorrent client. http://www.utorrent.com.

[14] Vuze HD network. http://www.vuze.com/app.

[15] K. Bauer, D. McCoy, D. Grunwald, T. Kohno, and D. Sicker. Low-resource routing attacks against Tor. In Proceedings of the Workshop on Privacy in the Electronic Society. October 2007.

[16] K. Bauer, D. McCoy, D. Grunwald, and D. Sicker. BitBlender: Light-weight anonymity for BitTorrent. In Proceedings of the Workshop on Applications of Private and Anonymous Communications. September 2008.

[17] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, February 1981.

[18] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In Proceedings of Workshop on Design Issues in Anonymity and Unobservability. July 2000.

[19] P. Dhungel, D. Wu, B. Schonhorst, and K. W. Ross. A measurement study of attacks on BitTorrent leechers. In Proceedings of the International Workshop on Peer-to-Peer Systems (IPTPS). February 2008.

[20] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In Proceedings of the 13th USENIX Security Symposium. August 2004.

[21] M. Dischinger, A. Mislove, A. Haeberlen, and K. P. Gummadi. Detecting BitTorrent blocking. In IMC '08: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement. October 2008.

[22] D. M. Goldschlag, M. G. Reed, and P. F. Syverson. Hiding routing information. In Proceedings of Information Hiding: First International Workshop. May 1996.

[23] D. McCoy, K. Bauer, D. Grunwald, T. Kohno, and D. Sicker. Shining light in dark places: Understanding the Tor network. In Proceedings of the 8th Privacy Enhancing Technologies Symposium. July 2008.

[24] A. W. Moore and D. Zuev. Internet traffic classification using Bayesian analysis techniques. In SIGMETRICS. 2005.

[25] M. Piatek, T. Kohno, and A. Krishnamurthy. Challenges and directions for monitoring P2P file sharing networks - or - Why my printer received a DMCA takedown notice. In 3rd USENIX Workshop on Hot Topics in Security, July 2008.

[26] N. Weaver, R. Sommer, and V. Paxson. Detecting forged TCP reset packets. In Proceedings of NDSS. February 2009.

[27] C. Wright, F. Monrose, and G. Masson. On inferring application protocol behaviors in encrypted network traffic. Journal of Machine Learning Research. 2006.

[28] S. Zander, T. Nguyen, and G. Armitage. Automated traffic classification and application identification using machine learning. In LCN. 2005.

The Headend Revisited: A Multi-Service Video Data Center for the Modern MSO

S.V. Vasudevan, R. Wayne Ogozaly
Cisco

Abstract

Cable MSOs are under competitive pressure to deliver an increasing number of new services to effectively compete with rival service providers, as well as emerging over-the-top players. This paper proposes a strategy which evolves the traditional cable headend into a multi-service Video Data Center which is well positioned for the challenges of the modern era.

The Video Data Center combines the best of the traditional video headend with proven data center techniques. This paper discusses the intersection of these digital delivery systems and describes a next-generation data center model. Internet data center architectures are examined and compared with the data delivery requirements of a modern digital cable system. Specific insertion strategies will describe how cable delivery networks can evolve to a multi-service Video Data Center, spanning broadcast, on-demand, and a next generation IPTV services, while integrating the best of today's regional headend with advances in data center technologies.

INTRODUCTION

Cable MSOs face considerable operational challenges as they seek to smoothly integrate innovative new video services into their existing delivery platforms. As we witness the deployment of more interactive video services, IPTV overlays, internet streaming to set top receivers, and other advanced technologies, of equal importance is the smooth integration of these new functions with existing video, voice, and data services across headend, transport, and last mile networks. These new services must coexist with and even complement existing services such as linear broadcast, switched digital video, and on-demand video services. Unlike the passive linear broadcast delivery model, these advanced services implement much more dynamic traffic flows between subscriber and headend, resulting in increasing operational considerations.

But video delivery systems are not the only area that has recently experienced great infrastructure and traffic growth. Consumption of internet web content, information which is compositionally populated with a mixture of database content and increasingly rich media components, has fueled the evolution of data centers, super-evolved versions of the "computer room" of the 1970s. Traditional internet content and service providers have deployed data center computing systems to support the 24x7 operation of web, application, and database servers, supporting the efficient delivery of interactive web content to large user populations. In order to meet the increasing performance and scale demands of a global web audience, enterprise computing, network and storage architectures have evolved into organized and tiered architectures to support high-transaction information throughput.

Most modern cable headends and hubs already use IP as a common transport for data, voice, and video traffic. Accordingly, video delivery to consumers is rising in scale, growing in interactivity, and being distributed to an increasingly diverse set of video-capable devices. The increasing adoption of interactive services such as VOD, SDV and IPTV will present a similar increasing transactional traffic load on service delivery platforms. As these changes occur and content networks expand their reach to a national and global scale, and application server compute requirements grow to respond to the increasing transactional load, the headend of old begins to resemble a data center that is serving a variety of data types, led by video, to a large subscriber population.

The service provider industry has already benefitted from the leverage of technology that was originally developed for orthogonal markets. The very use of IP switching and routing for video delivery serves as an excellent example. The switches and routers that now carry triple-play traffic were originally developed for the enterprise networking market. All of the staff-years and capital investment in ASICs, software and hardware development in the first 10 years of switch and router development went towards products that were intended for the wiring closet, and essentially funded by the global enterprise IT market. By adapting these technologies to the performance, scale, and reliability requirements of the service provider market, and through clever media processing techniques that enable the safe carriage of media streams over IP networks, IP switching and routing of entertainment grade video is now seen as the norm. The service provider industry enjoyed a drastic reduction in per-bit transport costs as a benefit of this technology transfer.

The remainder of this paper examines the evolution of the enterprise/internet data center, and provides an overview of networking and computing requirements that have driven the evolution of data centers in their efforts to scale and server their requirements. We compare and contrast between internet data center information processing requirements and cable headend information processing requirements. Finally, we propose some appropriate data center techniques and best practices that can help cable headends more effectively scale the throughput of their delivery systems.

1.0 THE DATA CENTER EVOLUTION

Introduction

Much like the cable video headend, data centers in the enterprise are under increasing performance pressures. The enterprise data center is being influenced by shifting business pressures and increasing operational limitations. The new demands of the enterprise require enhanced video services, greater collaboration, near-instantaneous access to applications and information, and compliance with ever-tightening regulatory compliance. These strains have manifested as operational issues relating to power and cooling, efficient asset utilization, element and system management and monitoring, escalating security and provisioning needs, and increasing operational expense.

A fundamental metric of data center efficiency relates to cost, power, space, computational power, and ingress and egress throughput. Any data center that is designed, trades off between these metrics, depending on the underlying requirements of the offered service. Beyond this, there is increasing attention paid to the “-ilities”: [1]

- Scalability
- Reliability
- Availability
- Serviceability
- Manageability
- Security

The data center architecture has evolved to accommodate these demands in an efficient way. With each architectural shift, the network has become an increasingly important ingredient to enable the latest round of transitions. The transformation to the newest evolution of Data Center 3.0 technologies focuses on a service-oriented design, consolidation of server farms, virtualization across disparate networked elements, and enhanced automation to manage the heavy load of requests for content, applications, and services.

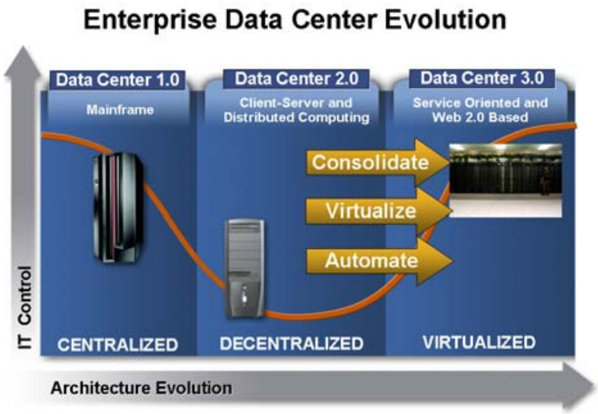


Figure 1 - The Enterprise Data Center is undergoing an extreme makeover, which is focused on server consolidation, service virtualization, and controlled automation.

The net result of this transformation is a massively scalable architecture; some data centers are built to contain upwards of 100,000 servers, stitched together as a fabric of virtualized elements.

Many design elements of the enterprise data center are applicable to the next generation cable Video Data Center. To better appreciate the similarities and differences, it is useful to present a brief overview of enterprise data center architecture.

Attributes of the Modern Data Center

Most modern data center architectures are based on a proven layered approach, which has been deployed in many of the largest data centers in the world. This layered approach provides the foundation of the contemporary data center and includes 3 major components: a ***core*** layer, ***aggregation*** layer, and ***access*** layer [2].

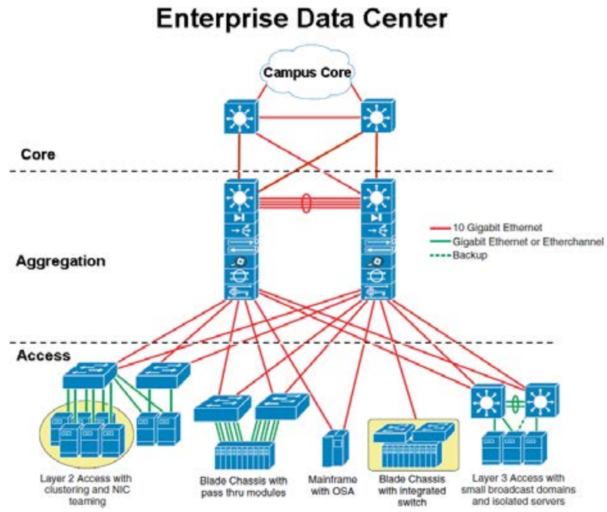


Figure 2 – The Data Center architecture implements a layered approach designed to diffuse massive loading across thousands of server elements [2].

It should be first noted that in enterprise data center terminology, the use of the descriptors *core*, *aggregation*, and *access* are used in an almost opposite manner to the way these terms would be used an interpreted by a cable headend engineer.

The core layer provides the high-speed packet switching backplane for all flows entering and leaving the data center. The core layer provides connectivity to multiple aggregation modules and provides a resilient Layer 3 fabric for this redundant system. This layer runs an interior routing protocol, like OSPF or EIGRP, and load balances traffic between the core network and the aggregation layer.

The aggregation layer includes multi-service switches which manage the interaction between servers through a Layer 2 domain. Server-to-server traffic flows through the aggregation layer and use features such as firewall and server load balancing, to optimize and secure applications.

The access layer is where servers physically attach to the network. Unlike the cable world, "access" in the data center refers to core of server design, where the network provides access to thousands of servers. Server components consist of 1RU servers, blade servers with integral switches, blade servers with pass-through cabling, clustered servers, and mainframes. The access layer network typically consists of modular switches and integral blade server switches. These switches manage Layer 2 and Layer 3 topologies which distribute flows across various server pools in different broadcast domains.

Distributing Requests Across Server Pools

Many applications run inside an enterprise data center. Each application may have one or more IP addresses associated with it, providing an identity to which Internet users send their requests. As Internet requests are received through the core and aggregation layers, specialized load balancers distribute these requests across a pool of targeted servers, using a combination Layer 2 VLANs and internal virtual IP addresses. A series of complex algorithms sift through requests, associate a virtual IP address and VLAN domain, and distribute the requests to the private addresses of physical servers [3]. These complex load balancers may confine requests for specific applications to a pre-allocated pool of servers, which are segregated into manageable clusters. In that way, server performance, redundancy, and security can be managed and scaled in a predictable way.

The multi-tier model in today's data center is dominated by HTTP-based applications. Scalable web-based services are built with multi-tier processing layers which include web, application, and database tiers of servers. Web servers field http:// requests and pass transactional information. Application servers fulfill these transactional requests by consulting one or more databases or sub-services. This information is presented back to the web server, which composes the properly formatted HTML page containing all requested information. This model enables the independent scaling of presentation, application or database resources as the service requirements dictate.

The multi-tier model may use software that runs as separate processes on the same machine using inter-process communication (IPC), or on different machines with communications over a switched network.

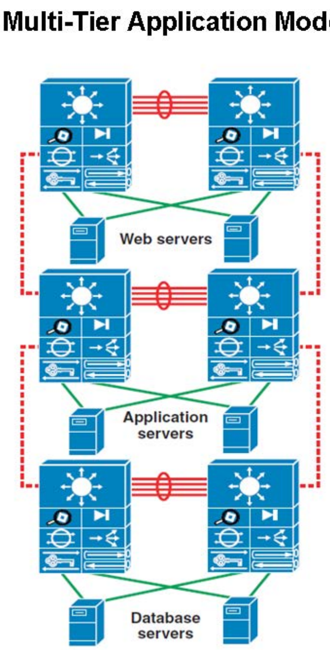


Figure 3 – For most HTTP-based applications, Web, Application, and Database servers work across clearly defined server pools using a tiered model, to process HTTP-based service requests [2].

2.0 THE VIDEO DATA CENTER

Comparing and Contrasting Enterprise and Video Data Centers

An important distinction between enterprise and video data centers can be understood by analyzing the traffic pattern of each service.

For a 3-tired web site, the amount of information offered is relatively small (an http GET method), the computational demands are moderate, and the amount of information returned is also relatively small (the average web page size is about 300kB). For a search engine web site, while the information offered is also small, the computational demands are very large (a search request may be dispatched to dozens of index servers working in concert) [4], and the information returned is also very small. For a video service provider with a service such as VOD, the information offered is relatively small, the computational demands are moderate, but the information returned is very large (a 2 hr movie in HD moves 14GB of data). Thus we can see that there is an immediate distinction in both computational as well as egress throughput requirements between the internet data canter and the video data center. Accordingly, the architecture can be optimized for the specific service requirements. But it cannot be mistaken that both types of data center have shared goals in the areas of scalability, reliability, availability, serviceability, and manageability, and security.

In this spirit, a video headend architecture implementing selected data center best practices is proposed.

Combining the Best of Data Center and Video Headend Technologies

As MSOs continue to expand the number and scale of video services, the resulting permutation of content, formats, interactive features, and end-devices has placed a burden on the traditional cable headend. This burden may be felt within the headend in a number of ways. First, services such as VOD, SDV, and IPTV require a resilient 2-way communications network, including outside plant and home connections. Upstream traffic will organically grow with the number of subscribers and applications. Correspondingly, the performance of the application servers that process these upstream requests will need to improve. In many cases the end result of the subscriber request will be the switching or streaming of video content. This information must be delivered via a resilient IP network with strict packet loss requirements. In the process of being delivered, the stream may be spliced, encrypted, or otherwise transformed by another cascading device. Additionally, sophisticated processes and systems are needed to manage the increasing rotation of licensed on-demand content, as well as ad content.

This evolving mix of new services, expansive capacity, and more interactive services, has triggered a evolution of the video headend. To accommodate this new service mix, the *Video Data Center* aims to combine the best of data center and traditional headend models into a more scalable and manageable system.

Video Data Center Design Goals

A number of key components have emerged to address the needs of the Video Data Center in the modern era:

- ***Multi-Service Support:***

  The next generation Cable Video Data Center not only needs to support core cable video services including broadcast, Switched Digital Video, and VOD, but it must also modularly accommodate new services such as IPTV, internet streaming to the TV, and video streaming to the PC or handheld device.

- ***Service Independent Scaling:***

  Each video service (broadcast, SDV, VOD, IPTV, etc) should be able to independently grow in capacity, without disrupting existing services. The IP network and multicast control plane will play a key role to both identify and independently manage each video service.

- ***Resource Modularity and Demarcation:***

  Each video resource within the Video Data Center (acquisition, grooming, encryption, etc.) is organized as a modular component with clear demarcation points. A pair of multicast addresses are used to identify the input and output points of a video stream as it traverses the delivery network. The IP network and multicast control plane are used to identify, grow, and independently manage each video resource. This methodology will prepare the way for service virtualization and improved scaling. It also enables the development of IP appliances to provide future stream processing functionality, or super-appliances that consolidate more than one stream processing function.

- ***Service Virtualization:***

  Service virtualization has many meanings, both in enterprise and service provider information processing. From the service provider perspective, it describes the implementation of a service that is logically viewed as a single client-server entity, but may physically be realized by more than one application instance. In content networking, virtualization refers to the abstraction of a video object from its physical attributes or location. For example, a movie may be stored in multiple locations, and be transcoded into multiple formats to suit multiple receiving devices, but in a virtualized implementation it need only be known as a single entity.

- ***Fault Containment and Resiliency:***

  The combination of embedded quality monitoring and enhanced redundancy techniques allow video faults and outages to be rapidly identified, circumvented, and contained to the Video Data Center. The Video Data Center provides a fully redundant system in which most failures are contained and not propagated throughout the video network.

- ***IP Early Acquisition:***

  The next generation Video Data Center will acquire large volumes of content, both real-time and asset-based, sourced from diverse locations in a wide range of formats. Much of this content will be shared across multiple video services. The "IP Early Acquisition" module ensures that this content is available to all services in an IP format, at the earliest point in the content acquisition process.

- ***Improved Operations and Management:***

  Stream visibility and quality monitoring across all services at strategic measurement points is always important. Since video streaming/processing devices and networking devices comprise the video delivery chain, it is important to have monitoring tools that can present a synthesized and unified view of the delivery system. The Video Data Center integrates video service monitoring across the delivery network.

3.0 BROADCAST SERVICES USING VIDEO DATA CENTER TECHNOLOGIES

<u>Implementing Switched Digital Video within the Video Data Center</u>

While there are some fundamental differences between the volume and type of information exchanged from a headend versus a database-driven website, there are also many similarities between each data center's design goals. Many data center architecture and networking practices can be applied in the cable headend. As an illustrative example, we

SDV Service - Design Parameters

SDV Channels	100 HD and 200 SD
High Definition Format and Rate	MPEG-2 format at 15.0 Mbps
Standard Definition Format	MPEG-2 format at 3.75 Mbps
HD Channel Bandwidth	1.5 Gbps (100 HD * 15.0 Mbps)
SD Channel Bandwidth	750 Mbps (200 SD * 3.75 Mbps)
Households Passed (HHP)	Emulate 15,000 HHP
Tuners per Household	2.0 tuners per HHP
Tuners per Service Group	500 tuners
Digital Television Penetration Rate	50% of households are digital
Total Number of Service Groups	30 Service Groups
SDV QAMs per Service Group	16 QAM channels SDV service group
Session Fundamental Bandwidth	3.75 Mbps per session

Figure 4 – The Switched Digital Video prototype service is implemented using Video Data Center techniques and includes 100 HD and 200 SD channels, mapped to 30 Service Groups, in a 15,000 household hub site.

present a hypothetical system design for a Switched Digital Video (SDV) service implementation. The SDV service described in this paper was designed and tested using the design parameters presented in Figure 4.

The advent of Switched Digital Video (SDV) technology provides a fundamental change in the way the industry delivers digital video entertainment. With SDV, service providers have the ability to offer a wider variety of programming while managing HFC network bandwidth in a sustainable way. The SDV architecture switches only selected content onto the HFC based on channel change requests from users within a service group. Thus, content that is not requested by any user in a particular service group does not occupy HFC bandwidth.

MSOs considering the deployment of SDV technology face three operational challenges. The first includes the integration of a large number of operational components required by the SDV service. Unlike the linear broadcast model, SDV implements a much more dynamic service in which SDV Servers dynamically map requested content to various QAM channels across hundreds of service groups. This increases operational complexity.

The second challenge for Cable MSOs considering SDV is how to smoothly integrate an SDV service with existing video services across the headend. The SDV service must co-exist and even complement existing services such as linear broadcast, VOD, IPTV, and other streaming techniques. The SDV system must also be managed and scaled concurrently with these adjacent video services.

A final challenge is in capacity planning for future growth. Since the SDV system has the ability to admit a great number of linear programs with only a modest increase in required HFC stream resources, operators need to rely on tools to provide visibility of system resource utilization. When the opportunity is presented to augment the amount of switched programming, the Video Data Center should be able to accommodate capacity expansions without any major disruption to the system in place.

Layered Network Approach

The Video Data Center implements a layered network approach to manage resource access, security, and scaling across a diverse set of video elements. Figure 5 highlights the Core, Aggregation, and Video Resource Access layers within the Video Data Center. Similar to the Enterprise Data Center, these layers provide a Layer 3 core connection to the Regional Network, aggregation and load balancing of video flows across various cable services, and access to a range of modular video resources. Individual video resources are linked together using the multicast control plane to create a SDV cable service.

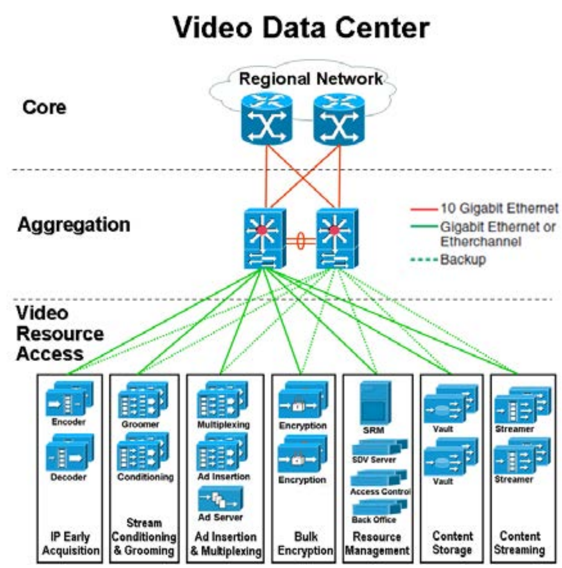


Figure 5 – Similar to the Enterprise Data Center, the Video Data Center implements a layered network approach, including Core, Aggregation and Access layers.

The Aggregation Layer includes redundant multi-service switches which provide GE access to individual resource elements. Since broadcast services within the Video Data Center operate on a much smaller scale than the traditional data center, a fewer number of multi-service switches are required. In this SDV example, two high density switches provide dual-homed GE inter-connects to all video resource elements.

Enabling Service Virtualization

Service virtualization is a critical principle upon which many of today's advanced data centers are built. The goal of service virtualization is to provide a more efficient delivery platform which is better aligned with the massive scale and complexity of today's networked environment. Although service virtualization in the data center continues to evolve, different components are directly applicable to the Video Data Center:

- ***Network Enabled Modularity***

  Resources are defined as modular components with clear network demarcation points, providing a more scalable and flexible design. Capacity can typically be added with minimal impact to existing services. (Example: SAN-based network storage which is independently managed and scaled from the server farms)

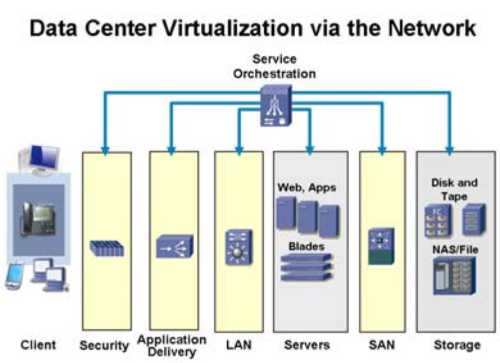


Figure 6 – Data Center storage is a modular resource, which is networked via the SAN, and can be independently managed and scaled [2].

- ***Distributed Service Models***

  Providers have the flexibility to distribute various processing stages across multiple servers, regional networks, and national backbones. Resource managers that are present for interactive services will play a principal role in the Video Data Center. These managers will optimize the allocation of resources to satisfy a number of parameters, including cost, latency, even potential revenue opportunities.

- ***Masking Complexity within Multi-Service Environments:***

  Improved resource management systems, advanced load balancing, and multi-tiered access networks hide the underlying technology from the end user. Content is available in multiple formats to a myriad of consumer devices. Exactly how this content is acquired, formatted, and delivered is hidden from the end user.

Similar techniques are implemented within the Video Data Center. As described in Figure 7, video resources operate within a modular architecture, defined by clear network demarcation points.

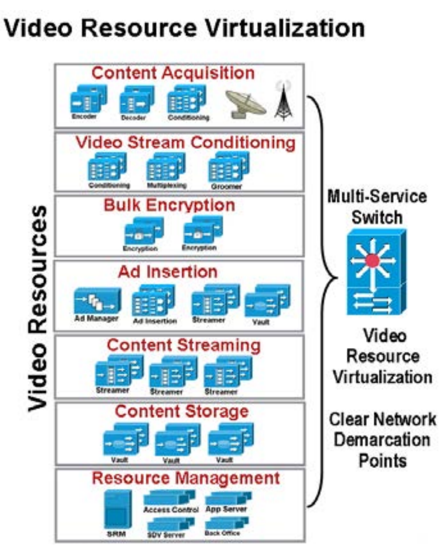


Figure 7 – Similar to the Enterprise Data Center, video resources are split into modular components and available as distributed network elements

These video resources are shared by multiple video services including linear broadcast, SDV, time-shift TV, PCTV/IPTV and other streaming services. This combination of resource modularity and network demarcation enables specific functions like Ad Insertion or VOD Streaming to be contained within a centralized Headend or distributed throughout a regional network.

SDV Service Creation

The multicast control plane plays a vital role in the Video Data Center design. Figure 8 describes how the multicast control plane strings together modular resources to create video content for the SDV service. In our example, 300 SDV SPTS streams are mapped to unique IP multicast group addresses. Devices along each processing stage issue IGMPv3 JOIN messages to draw specific multicast streams to each device over the layer 3 GE network. Redundant multi-service switches provide the routed control plane to manage the handoff of these streams between each processing stage.

Each of the major processing stages utilizes a unique multicast source address to identify primary and backup video sources, and obtain the correct content. Source-Specific Multicast (SSM), a feature of IGMPv3, also allows each stream to maintain a single multicast group address (but varying source addresses), spanning all of the video processing stages.

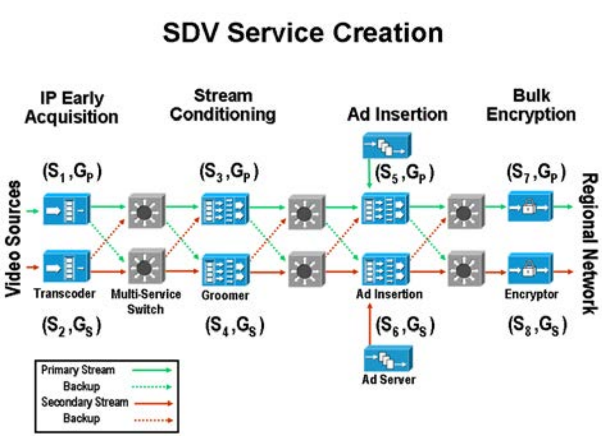


Figure 8 – The 300 channel SDV service is created by stringing together processing elements, using the SSM multicast control plane.

In this example, the same multicast group address is used for each SPTS stream across the four processing stages. There is one group address "**G_P**" mapped to the 300 primary video streams in the SDV tier, and a second group address "**G_S**" mapped to the 300 secondary or backup streams. This SSM feature greatly reduces the number of multicast groups used throughout the Video Data Center since a single group address (with varying source addresses along each stage of the delivery chain) can follow a stream throughout the Video Data Center. This technique also provides a clear network demarcation point between each stage and offers MPEG monitoring tools complete visibility into all SDV multicast streams.

Within this Video Data Center architecture, the SDV video path contains four processing stages. The IP Early Acquisition stage provides redundant access to all SDV video content. During this stage, SDV video content is acquired from redundant sources including satellite, off-air, and terrestrial links. SDV content is transcoded to an IP MPEG format at this early stage. Where necessary, video streams are converted from multi-program transport streams (MPTS) to single program transport streams (SPTS), as used by SDV.

This second stage provides Stream Conditioning of all SDV channels by redundant groomers. Since redundant copies of the SDV video streams are available, this stage performs a stream selection process in which each groomer independently selects the best available copy based on ETR-290 MPEG quality measurements. After stream selection, groomers rate cap the resultant SPTS streams from Variable Bit Rate (VBR) to Constant Bit Rate (CBR), as used by SDV.

The third stage of processing provides Ad Insertion, delivered by ad servers and standards-based SCTE 30/35/130 MPEG insertion techniques.

In the final stage of SDV processing, two independent encryption devices bulk-encrypt all SDV SPTS channels. The redundant encryption devices independently create two instantiations of the SDV channel lineup. The complete SDV channel lineup is now available for use by remote hub sites. In this fully redundant system, 300 primary streams are identified by multicast group address "**G_P**", in addition to 300 secondary video streams branded by group address "**G_S**".

Throughout this data center design, processing elements compare stream quality between the primary and backup SDV copies. In many cases, redundancy mechanisms identify MPEG level faults, and perform a cutover to the backup stream to contain the fault to the data center. These techniques prevent the propagation of video faults throughout the video network, pre-empting QAM-level cutovers to backup streams in many cases.

Source Specific Multicast (SSM) Load Balancing

A key element of the enterprise data center is the ability to load balance millions of flows across hundreds and potentially thousands of servers. These complex load balancing techniques typically use a creative combination of L2 VLANs, virtual address schemes, and server pools to evenly distribute requests across thousands of servers. This design is driven in part by the commoditization of server hardware in which enterprise-class servers are being replaced with thousands of low-cost blade servers. With the assistance of distributed computing and distributed systems management, load balancing across thousands of server elements is both efficient and reliable in the traditional data center design.

By contrast, most elements within the video headend remain specialized and far from commoditization. These specialized processors require their own sophisticated session and resource management systems. These management systems string together a number of video resources to generate each properly encoded, conditioned, ad spliced, and encrypted video stream. Fortunately, for broadcast video services, the scale of today's Video Data Center is typically much smaller than its enterprise counterpart with respect to number of elements. Given these differences, Video Data Center designs continue to employ resource managers and the multicast control plane to provide flow control and load balancing for most broadcast services.

For the SDV design, the SSM multicast address scheme provides a stable and predictable load balancing technique, which is suitable to the size and scale of the SDV service. Specific SDV streams are mapped to individual GE ports via IGMPv3 JOINs. Figure 9 illustrates how a subset of multicast group addresses are mapped to specific GE ports within the Stream Conditioning stage of the Video Data Center. In this example, 300 SDV SPTS streams are evenly distributed across 3 GE ports. The streams are divided into 3 groups of 100 streams (e.g. **$G_{1\text{-}100}$**, **$G_{101\text{-}200}$**, **$G_{201\text{-}300}$**). SSM allows the same group address to be reused across different processing stages, simplifying stream management. This processing stage employs different multicast source addresses to acquire content (from **S3)** and handoff the groomed streams (from **S5**) to the next stage.

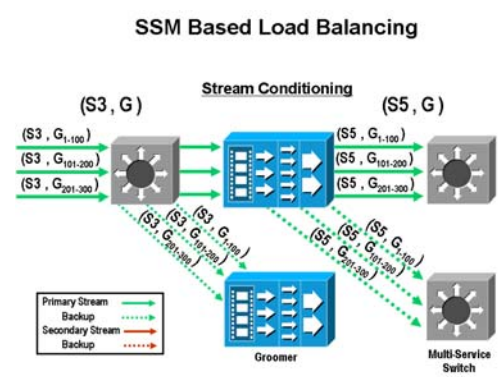


Figure 9 – SSM multicast provides a well-defined load balancing technique to distribute 300 SDV video streams across 3 GE ports.

This technique achieves a well-defined load balancing of SDV video flows. At each processing stage, multicast groups associated with video programs are joined to predefined GE ports. Each processing element is assigned a work load of multicast groups by the resource management system.

Gaining Control over Massive Scaling

The SDV system should be able to scale to a larger number of channels without impact to the current SDV deployment. Figure 9 provides the estimated bandwidth if we doubled the number of SDV channels. Based on these calculations, the total bandwidth required to deliver 600 channels of SD and HD content in the SDV tier is 4.5 Gbps.

SDV Requirements at Twice the Scale

SDV Channels	200 HD and 400 SD
Standard Definition Format and Rate	MPEG-2 format at 3.75 Mbps
High Definition Format and Rate	MPEG-2 format at 15.0 Mbps
SD Channel Bandwidth	1.5 Gbps (400 SD * 3.75 Mbps)
HD Channel Bandwidth	3.0 Gbps (200 HD * 15.0 Mbps)
Total SDV Channel Bandwidth	**4.5 Gbps**

Figure 10 - Bandwidth required if the SDV Service was doubled in capacity to include 200 HD and 400 SD programs.

To accommodate this additional capacity, extra video resources are required by the Video Data Center. Figure 10 highlights this growth in capacity, without interfering with the currently deployed SDV service. At each stage, additional multicast group addresses are assigned to the extra SDV channels. These new SDV multicast groups are drawn through the different switch ports, groomers, ad insertion, and encryption systems, ultimately providing an efficient and scalable technique to double the SDV service. Since each video processing stage is modular and utilizes clear network demarcation points, video resources within each stage can scale independently. The multicast control plane and resource management system simply directs new SDV streams to the new processing elements.

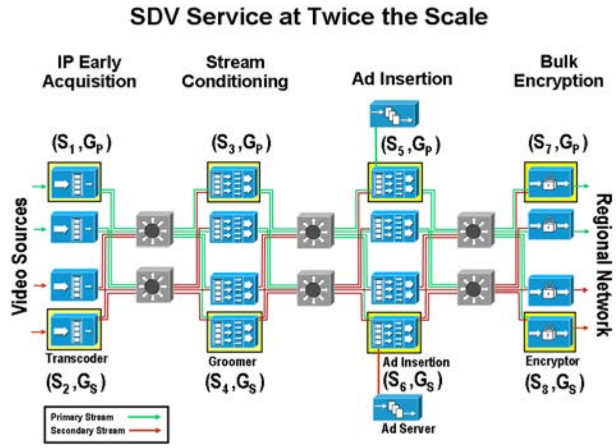


Figure 11 - Modular video resources and extra multicast groups are simply added to each processing stage to accommodate 600 SDV video channels.

4.0 IPTV INSERTION INTO THE VIDEO DATA CENTER

IPTV Insertion Strategy

The power and modularity of the Video Data Center are highlighted with the insertion of a future IPTV service. As described in Figure 12, IPTV video streams may require unique stream conditioning, ad insertion, and bulk encryption. To accomplish this unique processing, a separate set of multicast group addresses (**G_{IPTV}**) are used to draw the new IPTV streams through additional video resource elements. Similar to other cable services, a single group address for the IPTV streams will follow those streams throughout the video network. The combination of resource modularity, network based demarcations, and SSM based flow control allow this insertion technique with minimal disruption to existing video services.

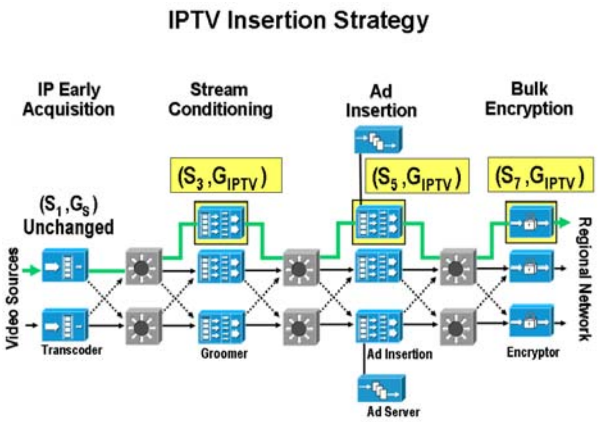


Figure 12 – The Video Data Center provides an IPTV insertion strategy with minimal disruption to already deployed services.

The IP early acquisition stage provides a common source of content for all cable services, including IPTV. As improvements to resource and element managers continue, next generation session-resource management systems can support both traditional cable services and newer IPTV systems, enabling further consolidation. This IPTV insertion strategy delivers a manageable insertion process while minimizing the impact to existing cable video services.

IP Early Acquisition

The next generation Video Data Center will acquire large amounts of video content, sourced from diverse locations in a wide range of formats. Much of this content will be shared across multiple video services. The "IP Early Acquisition" model insures that this diverse content is made available to all services in an IP encapsulated format, at the earliest point in the acquisition process.

The goal of the IP Early Acquisition stage is to convert all streams acquired from diverse sources to an IP MPEG SPTS format. Figure 13 provides a typical Headend video acquisition process in which content is acquired from satellite, off-air, and terrestrial video sources. RF video content from satellite and off-air sources is processed by a series of encoders, and groomers to convert RF, SDI, and ASI video into IP MPEG SPTS streams. Similarly, terrestrial video sources from remotes sources will be converted to an IP MPEG SPTS format.

In this example, common MPEG video streams, acquired from diverse sources are available for linear broadcast, SDV, VOD, and IPTV services.

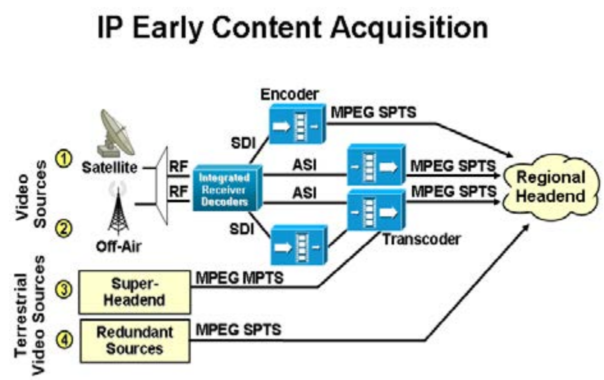


Figure 13 – IP early content acquisition collects and transcodes video content to an IP MPEG format for use by Broadcast, SDV, VOD, and IPTV services.

5.0 ON-DEMAND SERVICES USING VIDEO DATA CENTER TECHNOLOGIES

Second generation video-on-demand systems take advantage of Data Center content caching techniques and distributed video streaming to more efficiently absorb the massive scale and unpredictable loading of on-demand services. Figure 14 provides an example of an advanced on-demand network.

The intelligent IP infrastructure allows VOD content to be stored in large vault arrays at strategic locations across the network. On-demand vaults can provide a centralized and consolidated repository for large quantities of content, which are acquired from a wide range of sources to support both live and on-demand applications [5]. The ability to grow content storage independently from content streaming is a critical enhancement, and was inspired from the Web caching architectures.

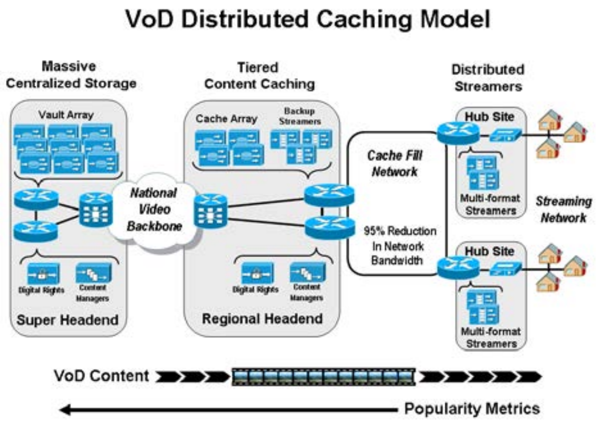


Figure 14 – The VOD content caching model is similar to Web-based cached systems in which the most popular video content is efficiently cached near the network edge, and delivered to the user via multi-format streamers.

In this distributed caching model, on-demand titles are distributed upon request using a tiered caching scheme. Cached video makes its way through the intelligent transport to video streamers located near the edge of the access network. As content streamers are moved closer to the edge, providers benefit from improved scaling, video quality, and reduced transport costs. This distributed caching model has been shown to reduce transport network bandwidth by as much as 95% since the most popular content is delivered once, cached, and reused across many requests for the same title.

Multi-format video streamers in the hub sites handle traditional MPEG video formats, but also support internet streaming formats such as Flash or Windows Media. These multi-format streaming improvements coupled with in-home gateway devices can enable a single video infrastructure to now support multiple streaming services to various in-home devices. Backup video streamers located within the regional headend can be positioned to handle unexpected peak periods, when streamers in a specific hub site are saturated with requests.

When combined with a national network, an on-demand caching scheme allows for a significant consolidation of content ingest points. Advanced caching protocols enable real-time ingest and the delivery of content throughout a national footprint within a 250 ms period. HD bit rates are also supported at a massive scale. Based on these data center design improvements, terrestrial content distribution will likely supplant the traditional on-demand "pitcher/catcher" distribution techniques of today. This networked infrastructure also provides a flexible platform to implement robust resiliency schemes, where multiple copies of content are stored and accessed from distributed locations. All of these elements work together through an intelligent network to more efficiently deliver the next massive wave of on-demand content.

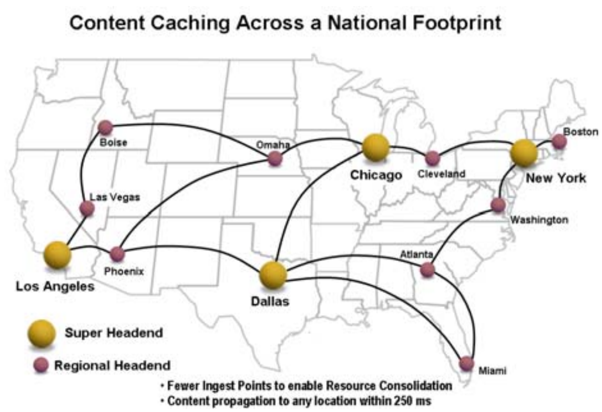


Figure 15 – Advanced content caching protocols enable real-time ingest and real-time distribution of video content across a national footprint.

SUMMARY

Enterprise computing is the midst of another evolutionary transformation, as data center architectures give way to "cloud computing" systems such as Amazon Dynamo [6]. Advances in this area are driving investments in advanced and highly-scalable networking and distributed computing architectures. The expected increase in interactive video traffic requirements is placing higher requirements on cable headends to deploy similarly scalable video delivery systems that are easily adaptable to growth and performance. From this standpoint, the evaluation and application of data center technologies and practices can greatly benefit service providers' efforts to scale and grow their video delivery systems.

REFERENCES

[1] ***"Datacenter Reference Guide",*** Sun Microsystems, July 2007.

[2] "***Cisco Data Center Infrastructure 2.5 Design Guide",*** Cisco Systems, 2007.

[3] ***"Towards a Next Generation Data Center Architecture: Scalability and Commoditization",*** A. Greenberg, P. Lahiri, D. Maltz, P. Patel, and S. Sengupta. Microsoft Research, 2008.

[4] ***"The Anatomy of a Large-Scale Hypertextual Search Engine",*** L. Page and S. Brin, 1998.

[5] ***"Climbing Mount Everest",*** S. Kotay (Comcast) and J. Pickens (Cisco). NCTA Technical Papers, 2008.

[6] ***"Dynamo – Amazon's Highly Available Key-Value Store",*** G. DeCandia et. al., ACM Symposium on Operating Systems Principles, 2007.

VIDEO QUALITY ASSURANCE ACROSS IP NETWORKS: CHALLENGES, REQUIREMENTS & WINNING STRATEGIES

New Technologies Permit Operators to Proactively Manage, Identify, And Solve Problems to Control Operational Costs and Boost Efficiency

Asha Kalyur, Marketing Manager, Cisco

Abstract

As customers demand increasing flexibility to receive video services anywhere and anytime with the best possible quality of service, cable operators of all types have deployed IP next-generation networks, which provide new levels of scalability, flexibility and fault resilience, while acting as a platform for the rapid deployment of new video services.

Operators will be challenged with monitoring and troubleshooting the growing number of video flows across these dynamic IP networks. A single lost IP packet can create visible artifacts on a large number of television screens, promoting the need for a new approach to video quality monitoring.

A new set of technologies can be used to ensure high-quality transmission of video across an IP network, using network-centric and service-centric views. This new, integrated solution – bridging the gap between "IP engineers" and "video engineers" – enables cable operators to precisely isolate and rapidly resolve video issues, while providing a common reference for diverse operational groups that will reduce the duration of impairments and consequently reduce subscriber support calls.

INTRODUCTION

The world of entertainment video is changing to support alternative methods for viewing content (e.g., on-demand) and additional devices for displaying video programs (e.g., PCs, game systems and handheld mobile devices). The industry is relatively young, but viewers still expect a flawless experience. Testimony arises from anyone who has endured a widespread video outage during a major Pay-Per-View event, Final Four tournament game or major football game. But poor quality of experience triggers more than inconvenience.

For instance, a major boxing event transmitted by a medium-size Multisystem Operator (MSO) region can generate more than $1 million in revenue. Unhappy customers will cancel subscriptions if confronted by a major outage. Further, cable television represents a lifeline service during local or national emergencies. Cable networks disseminate urgent news and critical public information, warning the public of severe weather or other impending dangers.

Service providers require a comprehensive service assurance solution in order to achieve the expectations of continuous high-quality viewing experiences. Such a solution enables service providers to proactively identify quality issues before they impact customers and leads to more efficient troubleshooting and resolution.

CHALLENGES FOR PROVIDING CONSISTENTLY GOOD QUALITY OF EXPERIENCE

The cable community understands the complexity of troubleshooting IP networks.

Adding newer video services adds to the complexity, straining network operational resources. Isolating video anomalies to a specific problem domain and allocating proper operational resources is critical to supporting high-quality video over an IP network and controlling costs.

Video relies on the network to provide customers with a high quality of experience. An effective video fault management system allows providers to detect and resolve quality issues before customers complain. Limiting help desk calls is preferred. Service calls reduce profit and increase customer churn. The average cost of a single customer case equals or exceeds the profitability of that consumer for an entire year, according to some cable providers.

In the broadcast television industry, success depends on consistently providing excellent video quality. When broadcast video services are delivered over an IP network, quality of experience is affected directly by packet loss and jitter. One IP video packet contains approximately 1,400 bytes of information, and each IP packet contains multiple MPEG encapsulated video packets. The loss of even one IP packet can lead to video impairments lasting half second or more. A single dropped packet also may cause “tiling.” Many dropped packets can lead to video freezes or loss of signal at the customer’s location.

Lengthy content delivery chains further complicate video networks. The delivery chain includes the broadcaster's head-end, the transport network consisting of multiple network elements, video manipulations such as multiplexing, ad insertions and finally through to the customer's home network and set-top box.

A video assurance solution must account for the various technologies transporting the video services and reduce the operational complexity for those charged with maintaining the quality of those services.

VIDEO ASSURANCE MANAGEMENT SOLUTION REQUIREMENTS

To assure excellent quality, service providers delivering broadcast television services over IP networks need powerful end-to-end, fault-management systems for continuous monitoring and correlation of service and network performances. These systems must meet several requirements:

- Identify faults in real-time and determine the impact to the specific video services.
- Support visibility into several operational domains (such as head-end, core, aggregation or access) for fault isolation.
- Integrate with the network infrastructure and operations support systems (OSS).
- Share common information pools and automation scripts for fault isolation, identification and resolution.
- Promote communication among personnel from different operational domains (head-end, core, aggregation, access).
- Present information in language familiar to both network operations and video operations personnel.
- Provide an architecture that supports devices from multiple vendors.
- Provide operational dashboards with intuitive drill-down navigation to detailed quality statistics per domain, per program and per stream.
- Extensible architecture to include components supporting future services.

CRITICAL FACTORS TO CONSIDER FOR VIDEO ASSURANCE

The critical factors a video assurance solution must consider include: operational costs, problem domain isolation and determination of services impacted. Related statistics are shown in Figure 1.

- **Operational Costs:** This inversely affects cable operator revenue. Four out of every five digital-program subscribers who placed service calls disconnected within the same year, according to studies. Up to one-third of all service calls requires at least one follow-up by a service technician. Increasing operational costs and decreasing subscriber quality of experience is an unsustainable model.
- **Problem Domain:** Successfully reducing the duration of a video delivery issue depends on locating the problem. Rolling out a truck to a customer site when the problem originates in the core transport network represents an unnecessary cost to the service provider. Similarly, dispatching network engineers to isolate and troubleshoot the IP network infrastructure would be unnecessary if the problem originated in the video head-end. An effective video assurance management solution must quickly identify the problem domain in the event of a service disruption, so providers don't incur wasted costs.
- **Service Impact:** Delivery of video services over an IP network represents the combination of two distinct technologies. A video assurance management solution must bridge this technology gap. Since video issues are likely to be reported as program issues, there must be an ability to correlate those services to the IP transport streams responsible for the delivery. Similarly, an operator should be able to assess the impact on a video service should IP transport issues be detected.

Figure 1. Video Assurance Statistics

83%: Problem calls related to poor video experience.
30 – 50%: Digital cable customers experience latent issues after a technician's visit.
28%: New installs require service call within one year. 80%: New digital customers with a service call in a year disconnected within that same year.
20- 25%: New installs requiring service call within one month. 5%: Digital customers disconnected within one month of install.
25 - 33%: Service calls resulting in at least 1 repeat visit

EXAMPLES OF CRITICAL PROBLEMS RELATED TO VIDEO ASSURANCE AND HOW THEY CAN BE SOLVED

1. What is the most important characteristic of an efficient video assurance management solution?

Today, there are many fragmented solutions for video assurance management. Many existing video monitoring solutions are either point solutions or rely on vendors who address only a small sub-section of overall delivery. Video assurance requires an architecture that permits interoperability between monitoring and troubleshooting. Video operations dashboards should condense the amount of information presented to the operator and provide a unified service status. Many video impairments are transient, and service technicians often fail to properly detect and isolate the video-service disruption. Without an end-to-end video assurance management strategy, technicians lack a logical trail to follow and instead work instinctively, changing cables, connectors, STBs and other equipment, trying to solve the issue. Technicians frequently are dispatched to a reporting customer site, even though the fault actually may have been miles away at the head-end, since video has dependencies within the IP network. That saddles the service provider with increased operational costs.

An efficient video assurance solution provides a complete end-to-end service assurance architecture for video services. The architecture must account for multiple domains including the video head-end, the providers transport and set-top boxes at home.

2. <u>Many operations centers fail to communicate with one another.</u>

Productivity increases when there is an ability to measure video flow as it transfers between operational responsibilities. This occurs when an anomaly is isolated to a specific problem domain in the end-to-end path, and the appropriate operational groups receive the assignment.

Most cable operators maintain separate video operations centers (VOC) and network operations centers (NOC). A video assurance management solution that alerts network operations personnel which part of the network requires attention (or what network event caused an anomaly) and simultaneously tells video operations personnel which specific programs were affected is critical to timely resolution. The ability for a video assurance solution to track the IP flows that contain a particular program (e.g., pre-ad-splice, post-ad-splice, pre-encrypt, post-encrypt) and all associated IP addresses used in the distribution of these services would allow service providers to determine both problem impact and the priority of the video services affected. Understanding the priority of the video service affected may mean the difference between troubleshooting a rerun vs. the Super Bowl. Through problem domain reduction and service impact assessment, operators can assign the personnel best-suited for a particular problem, reducing both resolution time and costs.

3. <u>How can an NOC monitor a channel across multiple multicast streams and Ad Zones?</u>

Cable operators use IP transport addressing to route packets through their networks. For broadcast video services, multicast addresses allow a single IP stream to reach large numbers of subscribers. Advertising content is inserted to video services as the multicast flows travel through the national network and regional networks to the local hubs. The insertion of advertising at these different points in the network is commonly referred to as Ad Zones.

The use of Ad Zones allow service providers to increase revenue by offering more specific advertising as the video service moves closer to the end subscribers. However, ad insertion presents a challenge to service providers. As the content of a video service is changed with new advertising, so must the multicast address of the IP transport. It is not unexpected that a video service may travel through several different multicast flows before reaching the subscriber. When a problem is reported on a particular video service, operators face the challenge of determining which multicast flows are responsible for the delivery of that complete service.

A video assurance solution must be able to associate a video service with the multiple multicast flows controlling the transport of that service through the network. The video assurance solution also must be capable of determining through which points in the network these multicast flows exist. Without this capability, operators would be unable to reduce the problem domain for an issue and would be forced to examine the entire transport network.

4. Monitoring the head-end network in a video assurance management solution.

Head-ends are the heart of a video network and can impact the largest number of customers should a serious or catastrophic outage occur.

Video content and payload quality on the ingress at the head-end is critical. Any flaws associated with the payload at this point will simply appear as poor quality on subscribers' screens. A great deal of processing takes place at the head-end that can potentially affect the video. This is where most of the video encoding, transcoding, rate shaping, program insertion, multiplexing and encryption takes place. Head-end monitoring not only guarantees the quality of the content being delivered to the subscriber, but also enables the service provider to monitor the quality of the content being supplied. The ability to identify video content errors at the acquisition source alleviates wasted truck rolls or time spent troubleshooting the IP transport, all of which reduces operational costs.

The ability to identify video content errors at the video source assures service providers with the highest quality of service for their subscribers.

5. Monitoring the last mile is critical.

Last mile and home network problems account for a majority of all quality issues reported to service providers and must be effectively addressed. While probes are effective monitoring tools, deployment to scale in the last mile is cost prohibitive.

A video assurance solution must scale to the number of subscribers in the last mile, providing error correction and quality notifications as set top boxes come on and offline. Services such as rapid channel change, forward error correction and video error repair are all components affecting video quality of experience. A monitoring solution that seeks to be end-to-end must effectively address monitoring in the last mile.

6. The role of video probes in the larger video assurance management solution.

Video probes are common tools for video monitoring. Video probes traditionally are placed at key demarcation points along the video path. Monitoring at the ingress and egress points of different domains in the transport informs the operator if the flow successfully has been transported through this domain.

Most probes collect data about each video service by examining the MPEG transport headers carried within a multicast flow. Other probes, which are generally more expensive, are capable of examining the MPEG data contents directly for errors.

The information gained from the probes, when combined with other video metrics, help isolate problems quickly in the delivery chain. Video probes are an important part of any video assurance solution. The key is tying these devices into the larger video monitoring architecture that accounts for the many other aspects of video assurance and video service delivery.

7. How important is the user interface in a video assurance management solution?

Being able to monitor multiple service silos means viewing multiple screen outputs to determine where a video service may have become impaired. This is not a simple task.

The use of a unified dashboard will prevent operators from associating diverse alerts from various screens to a common video service.

A “single pane of glass view” is critical for network operations personnel. Complexity on how the events are collected, and from what sources, should be hidden at this level. The video service status is the critical element that should be represented and monitored. An operator should be able to quickly diagnose that an anomaly exists, the domain where an anomaly is occurring and the services that are impacted.

For lower-level troubleshooting, specific engineering operators relevant to the problem domain where the anomaly is occurring can be dispatched for further additional diagnostics.

8. The challenge for cable operators to create unity among various vendor offerings.

Equipment from many vendors comprises the broadcast video network, while the IP network is primarily composed of routers and switches. Head-ends and hubs have specialized equipment that captures or encodes video signals for insertion into the network and transport to the customer premises. An effective video assurance management system can view, map and communicate with every component of this network. It must offer multivendor support and be flexible to accommodate additions, changes and upgrades to the network. For example, this flexibility would recognize the addition of last mile drop points to support new subscribers; understand software upgrades in switches and routers; and accommodate new bandwidth-management schemes over time.

SUMMARY

Consumers do not care where the problem in the network originates. They do not care whether it is in the head-end or at the set-top box or anywhere in between. The only thing that consumers care about is consistent delivery of quality video and a quality experience.

Telecom operators, broadcasters, satellite operators and cable operators are delivering more video services to end consumers today. Video is a complex, performance-sensitive service that requires a network to provide excellent quality of experience to its customers. To assure quality, cable operators who deliver video services need a powerful end-to-end video assurance solution.
We have reviewed the key drivers that can help service providers re-evaluate, redesign and build reliable, end-to-end and fault tolerant video assurance management solutions. Reduced subscriber churn and increased market share will belong to service providers who seize the opportunity to ensure customers receive the highest quality of experience through strategic and timely network management investments.

Cable operators constantly look at new service models and quality metrics that directly relate to customers' video experiences. There are myriad variables to measure and control in order to deliver an uninterrupted stream of broadcasts with a rich, clear signal and crisp sound. If service providers neglect the management aspect of this new opportunity, they may face higher costs when dealing with future operational issues. One of the determining factors in their continued success will be an overall end-to-end video assurance management solution.

Asha Kalyur is a marketing manager for Cisco. She can be contacted at 408-527-4065 or akalyur@cisco.com.

Wireless and Home Networking:
A Foundation for Service Provider Applications

Tim Burke - Liberty Global
Michael Eagles - UPC Broadband

Abstract

The explosion in the variety of Consumer Electronic (CE) devices and applications that deliver video and internet experiences has produced the need for simple and integrated home networking solutions. The cable TV community has analyzed, tested and debated the viability of offering home networking solutions for many years but competitive pressures and technology advancements has finally prompted near term action.

Although the home networking environmental conditions are reaching an inflection point, the standards, consortiums and technologies are very fragmented and the network operator economics justifying a service offering can be marginal.

This paper focuses on various wireless home networking technologies and solutions. The drivers for home networking are considered and a variety of wireless home networking configurations are discussed.

Preliminary test results from both a performance and economic basis are evaluated. Of particular importance will be to assess the user and operator experience from a set-up and maintenance perspective.

The conclusion section will contrast the technical and economic characteristics and benefits of the various wireless home networking solutions. A recommendation will be proposed that identifies areas of opportunities for wireless home networking solutions for Multiple System Operators (MSO's).

EVALUATING THE DRIVERS FOR SERVICE PROVIDER SUPPORTED HOME NETWORKING

In this section of the paper we explore the technology drivers for adoption of home networking. We consider the following drivers:

- Proliferation of wireless networked devices.
- Rise of Wireless Home Networking Standards and Ease of Use.
- Rise of Personal Web Applications, Place-shifting, and Social Networking
- Rise of Personal Digital Media and Low Cost Home Storage, and the decline of Digital Rights Management (DRM).
- Access Competition, Product Relevance and Substitution.
- Wireless Home Networking is happening today !

Proliferation of Wireless Networked Devices

Today‟s digital home includes many networking capable devices, with the range of multi-media networked devices continuing to increase. End users require home networking in order to support many of these new devices.

Multi-media Devices: Manufacturers are starting to include network connections or Wi-Fi technology in common devices like digital still cameras and MP3 players and printers. Examples include Apples Wi-Fi enabled iPod touch, or iPhone; Eye-Fi‟s Wi-Fi SD memory card; Archos personal media player; and the Hitachi Wooo camcorder or

Kodak Zx1camcorder that can stream live video using Wi-Fi to the HDTV set.

Media Extenders & Streamers: A range of IP connected media extender and streamer devices have entered the market in recent years. Examples include Apple TV, Netgear‟s recently announced ITV2000 Internet TV player.[1] In addition we see emerging streamers such as the Netflix Roku device, and networked gaming consoles such as the X-Box 360 which recently included the Netflix service. It is no surprise that today‟s TV require many HDMI ports !

Multi-room Audio: Multi-room audio solutions that leverage home networking capabilities are emerging. Examples include Sonos, Linksys by Cisco Wireless Home Audio system [2], and Apple‟s Airport-based wireless audio streaming; in addition to audio streamers such as the Logitech‟s Squeezebox, and Internet radio devices such as Tangent‟s Quattro Internet radio.

Networked TVs: TV sets with built in networking are beginning to emerge. At CES 2009 for example Sony, Samsung, LG and Toshiba were all introducing TVs with Ethernet and/or Wireless connections [3] which could be used to display Yahoo! widgets.

Rise of Wireless Home Networking Standards and Ease of Use

(a) Standards

Standards are critical in evaluating the drivers for service provider supported home networking. Support for forwards and backwards compatibility with today‟s networked devices determines the quality of experience subscribers experience and helps resolve end user issues. The key to achieving this support is the extent wireless home networking technology can be embedded in CPE devices, application layer standards for device discovery, and service provider.

(b) Ease of use

Manufacturers such as Linksys (with its LELA – Linksys EasyLink Advisor - software) have significantly improved the installation, configuration, and maintenance software for wireless networking devices.

However it is still quite complex for most mainstream consumers and causes additional call volume to ISP customer care centers.

Service providers can drive home networking further toward mass market by building a proper integrated support ecosystem. Installers and service providers could incorporate available tools to allow easy maintenance and upgrades.

Rise of Place-shifting, Media multi-tasking, Personal Web Applications and Social Networking

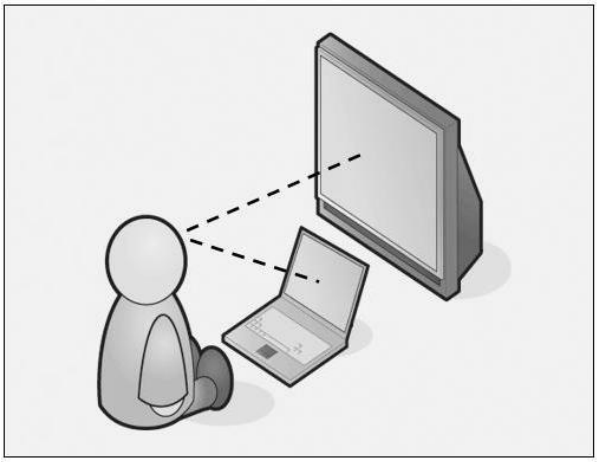
(a) Place-shifting and multi-tasking

Place shifting devices such as the Slingbox have emerged to enable services to be used in any room. Place shifting devices and services require a home network ideally based on a solution that supports location flexibility or portability.

Emerging home server solutions such as Microsoft Windows Home Server, and Linksys by Cisco Media Hub[4] are emerging that support multi-user and remote access to user generated digital content, requiring a home network.

With the proliferation of devices there is anecdotal evidence that multi-tasking is taking place in the living room, combining for example, the TV viewing experience with laptop browsing. Wireless networking is an enabler of multi-tasking.

Figure 1: Living-room Multi-tasking



(b) Personal Web Applications & Social Networking

The personalization of the web is evident in the explosion of web-based email platforms such as Gmail, Yahoo Mail and Windows Live Mail.

With the emergence of social networking applications such as MySpace and Facebook, combined with micro-blogging platforms Twitter and Friendfeed; web-based applications reinforce the personalized nature of networked communication.

The speed at which web-based personal applications have been developed for wireless networked devices, such as the Apple iPhone and Google"s Android platform emphasizes the importance of operator-supported wireless home networking.

<u>Rise of Personal Digital Media and Low Cost Home Storage, and the decline of DRM</u>

(a) Personal Digital Media

With today"s explosion of multi-media digital devices such as Camera"s, Handycam"s, and Multi-media Handheld device, users are generating more personal digital content than ever before.

As highlighted in Table 1 below, it is estimated that the typical U.S. broadband household will have almost 1 terabyte of personal digital media in the home by 2012.

Table 1: U.S. Household Digital Media Growth [5]

	2008 (GB/hh)	2010 (GB/hh)	2012 (GB/hh)
Music	11	17	24
Photos	14	47	151
Video	201	347	723

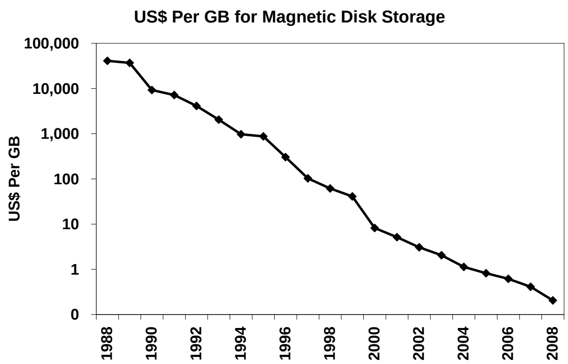
With an expanding library of personal digital content including music, photos and movies; the home network starts to play a pivotal role in allowing a user"s multiple devices to synchronize or access personal content between rooms and across devices.

Furthermore, applications are emerging to support multi-device synchronization of personal digital media such as Microsoft Live Mesh, driving the need for home networking.

(b) Low Cost Home Storage

In 1988 1 GB of storage in the home would cost about US$40,000, so storing an 11GB music library would cost US$440,000! That cost has dramatically declined so that in 2008 1 GB of storage cost of approximately US$0.20 meaning that 11GB music library could be stored for US$2.20. This dramatic shift in the economics of storage has enabled households to store large libraries of multi-media content for consumption around the home.

Figure 2: Consumer Price Per GB Declines [6]



Additionally, small form factor storage, such as SD cards, and smaller Mini and Micro SD cards are supporting multi-media from a range of new home wireless networked devices such as Apple"s iPhone and Google"s G1 Android handset. Such devices are able to both access and contribute to the personal digital media library.

(c) The demise of Digital Rights Management (DRM)

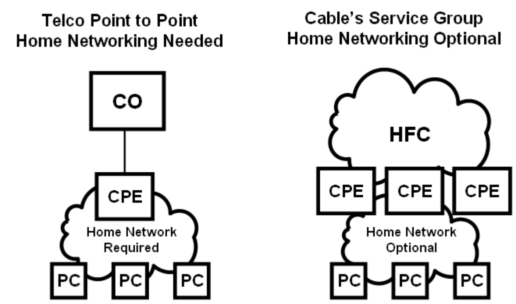
Recently Apple announced it was abandoning DRM protection for iTunes song downloads in favour of a non-DRM model[7]. The fall of DRM means that a barrier to multi-room audio has fallen and this facilitates a wider device ecology for the consumption of multi-room around the home. The key question is whether this trend will extend to the video world? In particular, the ability to move MPEG4/H.264 HD content to various devices within the home at its much lower bandwidth requirements could be a key enabler for whole home wireless networking of video, data and voice.

<u>Access Competition, Product Relevance and Substitution</u>

(a) Access Competition Driving Wireless Home Networking

Telco access architectures have been a driver of home networking. The Telco point to point architecture requires an in home solution to enable multiple devices to connect to the same service. This can be compared to Cable"s service group which can support multiple devices with a dedicated CPE if required.

Figure 3: Telco Access Architectures Depend on Home Networking



As a result Telco"s have deployed advanced residential gateways for several years by necessity.

(b) Product Relevance and Substitution Driving Service Provider Wireless Home Networking

The traditional RBOC or Incumbent Telco, has typically placed an emphasis on the residential gateway CPE device with the latest features as a way to attract subscribers, reporting the residential gateway CPE as a subscriber acquisition expense. This has provided a head start for Telco"s in residential gateway adoption and wireless home networking penetration.

A subset of Telco residential gateway deployments can be seen below in Table 2:

Table 2: Telco Residential Gateways [8]

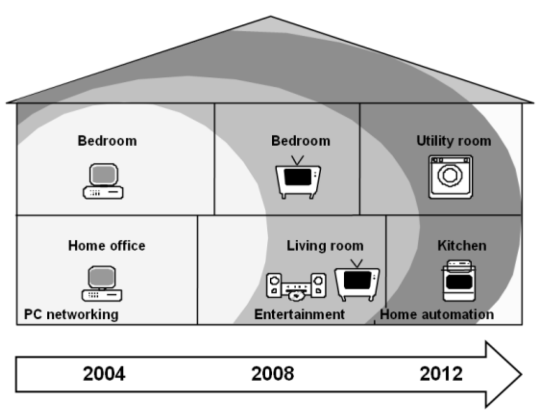
Telco	Gateway Product	Sales/Units Deployed
British Telecom	Hub (integrated Wi-Fi and DECT)	3.5m HUB boxes, about 100k per month
France Telecom	Livebox (Integrated Wi-Fi, USB DECT dongle)	7.5m sold, selling 300k per month
Deutsche Telekom	Speedport (Integrated Wi-Fi and DECT)	Selling 200k to 250k per month

Substitution challenges can also be a driver for service providers. Wireless home networking could be a unique selling point to counter wireless mobile broadband offerings in the market place.

Wirelesss home networking is happening today !!

Wireless home networking is happening today and solutions are emerging to support not only basic connectivity but also other devices such as gaming consoles, multi-media extenders and entertainment devices as outlined in the networked home in Figure 4.

Figure 4: The Networked Home [9]



Wireless is playing a larger role in home networking. As seen in the table below, outlining the shift in European online consumer home networking, there is a trend toward wireless becoming the default method of home networking.

Table 3: European Online Consumer Use of Home Networking [10]

	Q4 2006	Q4 2008
Yes – Wired	12%	8%
Yes - Wireless	13%	20%
Yes – Mixed /not sure	7%	12%
No	68%	60%

Summary of Wireless Home Networking Drivers

Proliferation in networked devices, ever-expanding user generated content libraries, the desire for „anywhere" place-shifted content access, the rise of the dynamic multi-media web, the rise of real-time social networking, and new viewing behaviors are all driving the end user adoption of wireless home networking.

Competitive access products from Telco"s already include wireless home networking, and the threat of mobile broadband substitution creates additional urgency for fixed line service providers.

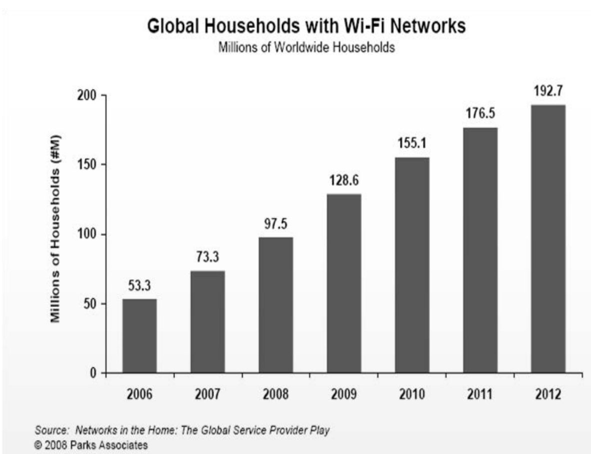
The message for the service provider: wireless home networking is happening today, service providers can either choose to participate and add service provider supported features and value for the subscriber, or watch from the sidelines at the risk of market share loss, mobile broadband substitution, and loss of relevance in the home multi-media experience.

THE POSSIBILITIES OF WIRELESS HOME NETWORKING

If a service provider, like a Cable TV Company, could utilize in home wireless technology to help solve its home networking needs then many complexities of whole house distribution could be solved. Wireless is inherently a simple, flexible, low cost and convenient medium for both the subscriber and network operator. Wireless is a particularly attractive option for the author‟s family of international companies because, unlike the United States, most homes in Western Europe, Eastern Europe, Asia and South America are smaller size, multi-dwelling units and not pre-wired with coaxial cable.

Wireless technologies (Wi-Fi, cordless and DECT phones) are currently the solution of choice for in-home data communications and phone services as illustrated by the chart below.

Figure 5: Global Households with Wi-Fi Networks[10]



The progress in viable technologies (e.g.- DLNA, UpNP, MPEG4 video compression) and applications (e.g.- Internet TV, Apple TV, Hulu, multi-room DVR…) that move video around the home has accelerated the need for a reliable in home transport medium and further raises the bar for wireless as a total home networking solution. As a total home networking tool the bandwidth, performance and quality issues of video over wireless have limited in-home wireless technology and could hamper its evolution beyond basic data and voice applications.

In this paper we will look at transmitting HD video throughout the home, replacing HDMI video cables within a room via “wireless HDMI” and traditional data/voice home networking.

Regulatory, technological and standardization advances in the wireless sector over the past few years has placed wireless home networking in a position where it could possibly meet the home networking challenge just described.

Regulatory

For instance, on the regulatory front, higher power transmission of signals has been allowed in the 5 GHz Wi-Fi spectrum in Europe. The spectrum available in Europe and the U.S. are becoming more aligned for 5 GHz Wi-Fi, UWB (6-10 GHz) and even 60 GHz. Finally, larger blocks of contiguous spectrum are being made available (~ 600 MHz at 5 GHz, 1.7 GHz at UWB, 7 GHz at 60 GHz).

Technology

Overall, wireless technology advances have made huge leaps. Orthogonal Frequency Division Multiplexing (OFDM) modulation, Multiple Input Multiple Output (MIMO) antenna schemes, large channel bandwidths (e.g.- 40 MHz channels), chip integration that puts baseband and RF functions on a single chip and high quality video compression techniques provide economical and bandwidth efficient solutions. The combination of these technological advances gives operators the appearance that wireless whole home networking is feasible even when entertainment applications are assumed.

Standards

Finally, the standards bodies, and in particular the IEEE, have made tremendous strides at continually updating and improving their wireless specifications. The 802.11 working groups have created new amendments and standards in many areas (e.g. - high throughput/802.11n, Quality of Service/ WMM, Security/WPA2, Power Save/APSD) that help improve service levels and capabilities. Although these specifications take longer than desired to become standards the progress made in the difficult political environment of the standardization process is impressive.

THE REALITY OF WIRELESS HOME NETWORKING

All the positive indicators just mentioned must be tempered with the reality of the wireless medium. The wireless channel is unstable and unpredictable in an outdoor environment but becomes extremely variable indoors as floor plan layouts, furniture, walls and living quarter sizes vary widely. The characteristics of the Radio Frequency (RF) channel can change rapidly over a time period due to fading and interference. Consequently, the capacity of the channel and signal strength (signal to noise ratio or SNR) seen by the receiver fluctuates constantly.

The evaluation of wireless whole home networking solutions must be looked at along five critical success criteria.

- High volume components & chips to meet consumer electronic (CE) device economics
- Sufficient and uniform spectrum at the right frequencies
- Transmit power levels versus interference trade-off‟s
- Quality of service capability for video applications
- Compression versus latency versus bandwidth tradeoff‟s for video applications

The interplay and relationships between the success criteria results in complex systems architectures.

Consumer Electronics (CE) Device Volumes

A large ecosystem of wireless home networking chipsets and devices are required to reach the proper economics in the C.E. world. But success in obtaining the proper device scale in the unlicensed, unregulated spectrum realm of the home network means too many wireless devices operating in the same spectrum which causes interference, quality and capacity issues. It has taken about 5 years for the 60 MHz of Wi-Fi 2.4 GHz spectrum (three 20 MHz channels) to become too congested using just voice and data applications. In some ways, the great success of Wi-Fi has bred failure for wireless as a home networking solution.

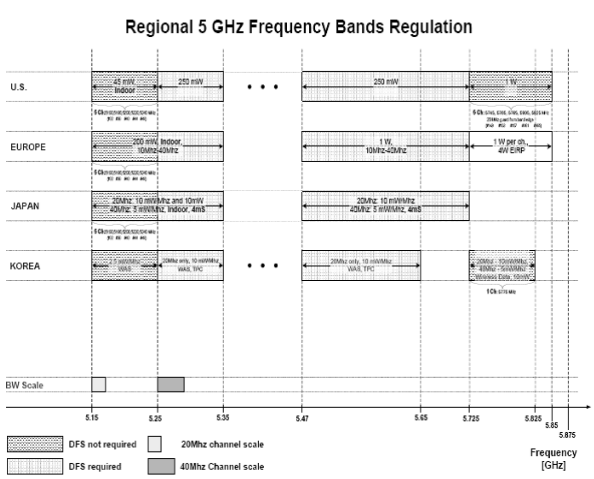
Spectrum Availability

As a positive consideration, the 5 GHz Wi-Fi frequency is relatively greenfield, contains ~500 MHz of contiguous spectrum (twenty-four 20 MHz channels) and is available almost uniformly across both the U.S. and Europe. The future success of 802.11n devices combined with the higher bandwidth requirements of video applications means a similar 2.4 GHz congestion problem could eventually occur. The capacity and quality limits at 5 GHz would occur first in the denser urban areas of apartments and multi dwelling units and in areas where public Wi-Fi networks are operating in the same frequencies. Anticipating this eventuality, standards bodies and start-up companies are venturing into the higher frequencies of UWB (6-10 GHz) and 60 GHz spectrum where much larger blocks of spectrum are available. Unfortunately, higher frequencies have the well known disadvantages of severely limiting the range the signal can go and requiring higher power and costs. For this reason UWB and 60 GHz solutions have been currently relegated to in room solutions only.

Transmit Power Levels

In general, transmit power limits in wireless devices have similar economic, cost and interference trade-off‟s. The EU has recently increased the allowable transmit power limits across most of the 5 GHz Wi-Fi frequencies to 1 Watt.[11] Although this certainly helps in getting whole home networking solutions to work in houses and apartments that are larger and made of stronger materials the RF energy from adjacent networks will raise the noise floor and interference levels. Higher power levels in devices also translates into additional costs. Transmit power levels above a certain point (~100 mW) limits the ability to integrate a power amplifier (PA) into a chip as non-linearity and peak to average power ratios cause problems in chip designs. Outboard PA‟s typically translates into higher costs.

Figure 6: Wi-Fi 802.11n 5 GHz Frequencies and Transmit Power[12]



Quality of Service (QoS)

Whole home wireless solutions that accommodate delay sensitive applications such as video and voice must at a minimum be able to prioritize various traffic types. The 802.11 standard that specifies QoS is called WMM (Wireless Multi-Media) and has the potential to ensure quality video transmissions within the home. Unfortunately, wireless home networks have not reached the point where they will guarantee and reserve bandwidth by offering parameterized QoS as is done within the cable TV and DOCSIS network of a cable operator.

Compression, Latency and Bandwidth

In the wireless home networking arena a debate rages over the effects of transporting uncompressed (e.g.- HDMI) versus a compressed video signal (e.g.- MPEG2 or MPEG4/H.264). No wireless technology transports true uncompressed HDMI as a baseband video signal. Other than the large spectrum range of the 60 GHz frequencies, all other in-home wireless technologies use some sort of real time compression. Compression (e.g.- H.264) will be needed to send video over wireless especially as video scales up with higher frame rates, deeper colors and higher resolutions.[13] Some vendors introduce the concept of lossless versus lossy compression to further differentiate their products. Traditionally the industry has regarded any compression above 4X compression ratios as lossy compression. Compression at lower compression ratios allow it to be categorized as lossless compression.[14]

Linked in with this discussion is the latency issues associated with transcoding from one codec to another when dealing with compressed video, the ability to move a compressed signal across devices while still complying with all DRM considerations and the ability to add the graphic overlay for an

EPG (Electronic Program Guide) over the compressed signal. Obviously a compressed signal has substantial advantages in moving video content around the home as the bandwidth requirements of MPEG2 HD are 15 to 20 Mb/s (7 to 10 Mb/s for MPEG4/H.264) while uncompressed HD requires > 3 Gb/s of capacity.

WHOLE HOME NETWORKING REQUIREMENTS

The CableLabs OpenCable Home Networking study group has spent a considerable amount of time defining the various use cases for moving video, data and voice applications throughout the home. As a result of this effort the minimal bandwidth, performance requirements and architectures of a home network have been proposed.

A typical deployment scenario of one Set Top Box (STB) with storage capability and two STB"s without storage depicts a requirement for four times the maximum bandwidth for a single HD stream of MPEG2 video content being transported between network elements in the home. Assuming a 20 Mb/s per HD stream requirement (MPEG2) a consistent and reliable 80 Mb/s wireless network is needed within the home that offers full Quality of Service (QoS) and prioritization (or better yet reserved capacity) of video media content. CableLabs assumes another 20 Mb/s for best efforts based data traffic and prioritized voice traffic. Therefore the MAC throughput data rate of at least 100 Mb/s is required to support all types of in home networking video, data and voice streams.[15]

OVERVIEW OF TECHNOLOGY OPTIONS FOR WIRELESS HOME NETWORKING

There is a range of wireless technology options, associated frequencies, existing or emerging standards and industry associations that could possibly meet the demanding home networking requirements of video, data and voice applications. The major alternatives investigated in this paper include:

- DECT (Digital Enhanced Cordless Telecommunications)
- High Throughput Wi-Fi (802.11n)
- Variations To High Throughput Wi-Fi (802.11n) Specification
- Optimized Video At 5 Ghz Or Proprietary Wireless HDMI.
- Ultra Wide Band (UWB) and (IEEE 802.15.3c or 802.11ad) technology

Each of these four alternatives have a number of start-up and established companies pushing their particular technology and specification. Many have established consortiums of companies and industry associations and consortiums of companies with the intent of bringing their specification to a standardization body for approval. The table below summarizes the major technology options. In all cases the standardization process is ongoing and in many cases at an early stage. Only the Wi-Fi 802.11n standard is very near completion and has a large ecosystem of chipsets and consumer electronic devices currently being built.

Table 4: Wireless In-Home Alternatives[16]

Name	Association	Spectrum	Standard Body	Major Chip Suppliers	Data Rate Claims	Indoor Range
DECT	DECT Forum	1.88–1.9GHz in Europe, 1.92–1.93GHz in the US	ETSI	DSP Group and SiTel	64Kb/s, 300Kb/s (CAT-iq)	~50m
High Throughput Wi-Fi	Wi-Fi ™Alliance [17]	2.4 GHz and 5 GHz	IEEE 802.11n	Intel, Broadcom Marvell …	300-600 Mb/s	~100m
Optimized Video at 5GHz (Proprietary Wireless HDMI)	Wireless Home Digital Interface (WHDI ™) [18]	5 GHz	IEEE 802.11aa, ac	Amimon	1 Gb/s	~50m
Ultra Wide Band (UWB)	WiMedia ™ [19]	6 to 10 GHz	Formally IEEE 802.15.3	TZero Radiospir	480 Mb/s	<10m

			a now ECMA-368	Sigma Designs, PulseLink…		
Very High Throughput 60 GHz (Proprietary Wireless HDMI)	Wireless HD ™ (WiHD) [20]	60 GHz	IEEE 802.13.3c and 802.11ad formally 802.11vht	SiBeam	3 Gb/s	<10m

DECT Digital Enhanced Cordless Telephone

DECT was developed by ETSI but has since been adopted by many countries all over the world. The original DECT frequency band (1880 MHz–1900 MHz) is used in all countries in Europe. Outside Europe, it is used in most of Asia, Australia and South America, making it well suited to the author‟s international systems.

In the United States, the Federal Communications Commission in 2005 changed channelization and licensing costs in a nearby band (1920 MHz–1930 MHz, or 1.9 GHz), known as Unlicensed Personal Communications Services (UPCS), allowing DECT devices to be sold in the U.S.

Although the DECT data rate makes it unsuitable for high speed data or video, many voice service providers consider that “DECT is the go to platform for voice calling in the home”[21]. Further, as a home wireless technology for voice, DECT has the advantage of a well established ecology of handset device manufacturers. Additionally many subscribers are familiar with DECT, the technology does not require additional subscriber education.

High Throughput Wi-Fi (802.11n)

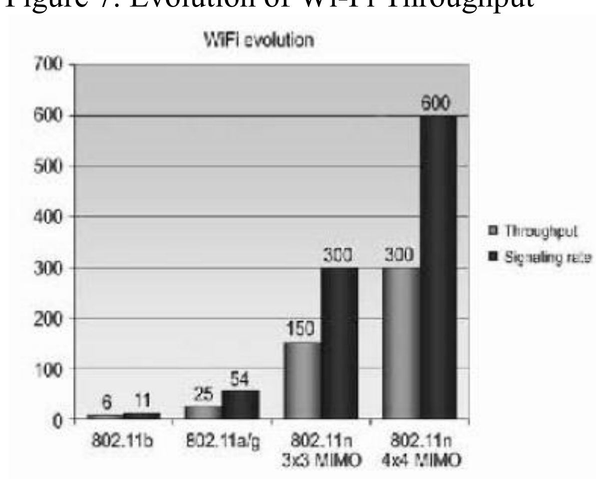
Wi-Fi is the wireless LAN technology brand developed by the Wi-Fi Alliance to certify IEEE 802.11 devices. In a little over a decade Wi-Fi has evolved from an innovative idea into an indispensible technology for consumers. The original standard has been continually updated and enhanced, which has led to the tremendous growth in the industry. The chart below outlines some of the key milestones in the 802.11 committee‟s progress.

Table 5: Key IEEE 802.11 Standards[22]

Standard	Date	Description
802.11a	1999	First 5 GHz standard. Incorporates OFDM technology with speeds up to 54 Mb/s
802.11b	1999	First standard to gain wide adoption. Operates in the 2.4 GHz using DSS CDMA technology and 11 Mb/s speeds
802.11g	2003	Operates at 2.4 GHz but employs OFDM technology at speeds of 54 Mb/s and is backwards compatible with 802.11b
802.11n	2008 for Draft 2.0	Next generation standard that uses 2.4 & 5GHz and leverages MIMO, beamforming to produce 600 MB/s speeds
802.11e	2005	Provides support for Multimedia Applications with Quality of Service, called WMM
802.11i	2007	Adds security features to improve on WEP using Advanced Encryption Standard (AES)

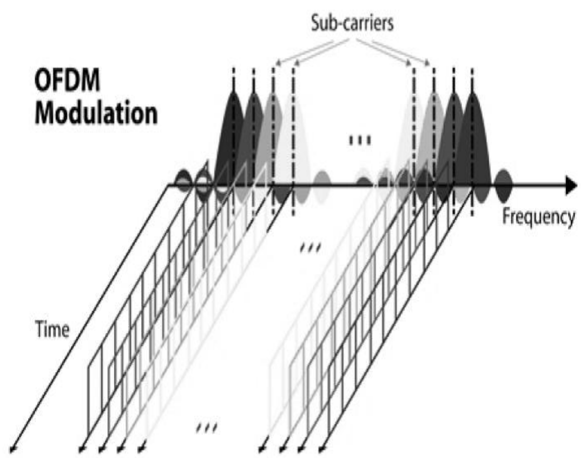
The innovative technologies employed in the 802.11n specification results in greater throughput and reliability over previous Wi-Fi solutions. Data rates as high as 600 Mb/s, 10 times greater than 802.11 a/g previous standard, are possible with 802.11n. Figure 7 illustrates the dramatic boost in network capacity and ultimately the speeds possible for in-home wireless networking applications using 802.11n technology. Contrasting this development with today‟s 802.11a/g speeds of 54 Mb/s and typical wired Ethernet 100 Mb/s throughputs begins to put this advancement in perspective.

Figure 7: Evolution of Wi-Fi Throughput [23]



The 802.11n specification is built on the cornerstones of OFDM and 802.11a and 802.11g standards. In this technique the usable bandwidth is divided into a large number of smaller bandwidths or subcarriers. These subcarriers are mathematically orthogonal or unique and can be tightly packed next to each other to gain maximum spectral efficiency. The high speed information to be transported is then divided into multiple lower speed signals and transmitted simultaneously on different frequencies (subcarriers) in parallel.[24] Both 802.11a/g and .11n specifications utilize 52 subcarriers spread across a 20 MHz channel bandwidth. Figure 8 illustrates the various subcarriers and how the data to be transported is distributed across the subcarriers in both frequency and time.

Figure 8: OFDM Technology [25]



A major advantage of OFDM and 802.11n is its ability to tolerate multipath fading by carrying small amounts of information on individual subcarriers or frequencies. If one or two frequencies/subcarriers are lost due to a fade then only a small amount of information is lost which can be compensated for via error correction coding and retransmissions.

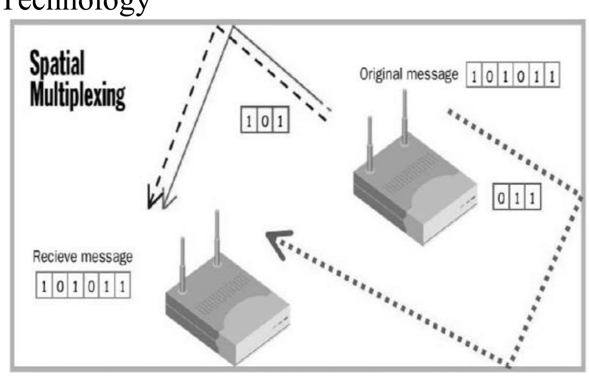
The primary innovations that allowed the 802.11n standard to make such a large leap in performance and throughput from its predecessors are:

- Multiple Input Multiple Output (MIMO) technology
- Enhancements in modulation and coding schemes
- Packet overhead improvements called packet or frame aggregation.
- Channel bonding (40 MHz channels)

(a) MIMO Technology

MIMO uses multiple radios and antennas to allow different data to be simultaneously transmitted across multiple transmit antennas. On the receive side of the wireless data link the separate unique data streams go across free space on different paths and are received by the separate receive antennas. Because of the spatial diversity of the multiple paths through the air, MIMO systems are able to transmit two unique data streams. This results in data transmissions at twice the data rate of single antenna and radio systems. This concept is called MIMO Spatial Multiplexing (SM) and is illustrated in Figure 9. In Figure 9 the original data stream or message (101011) is sent simultaneously as two different messages (101 and 011) across two different radios, antennas and spatial stream paths to the receiving device antennas. The 802.11n standard can support four spatial streams across four transmit and receive antennas.

Figure 9: MIMO Spatial Multiplexing Technology[26]



(b) Enhancements in Modulation and Coding Schemes

802.11a/g Wi-Fi networks use a single stream and eight possible modulation and coding schemes resulting in eight possible data rates. 802.11n MIMO technology can implement a concept called rate adaptation and variable modulation and coding schemes. If the RF conditions are good and the signal strength of the receiver is strong a higher modulation level (e.g. 64QAM) and weaker error correction code can employed at that instant for the data stream. This results in a higher data rate for that particular data stream. As an example of the power and complexity of the 802.11n specification there are 77 different modulation and coding schemes (MCS) possible where 8 MCS schemes are mandatory.[27] The end result of the multiple spatial streams, coding and modulation options, are the possibility for much higher data rates.

(c) Packet and Frame Aggregation

Reducing the packet overhead required for data transmission is critical for reliable high speed delivery of data operating in the 802.11n mode. In conventional wireless transmission systems the amount of overhead is fixed regardless of the size of the data packet. As the data rate increases the overhead remains the same. The 802.11n specification has fixed this problem by aggregating multiple packets of overhead into a single transmission frame.[28]

Unfortunately, aggregating multiple overhead packets can cause an increase in system latency as the radio must hang onto packets at the transmitter until it creates a transmission frame of the desired size. For that reason some real time applications (e.g.-voice) do not utilize the benefits of packet aggregation.

(d) Channel Bonding

802.11n networks gain an immediate capacity increase by implementing the concept of channel bonding. Combining two 20 MHz channels into a 40 MHz channel is a very straightforward way of increasing the bandwidth and capacity of the system. This technique is utilized more effectively in the 5 GHz range as there are twenty-four 20 MHz channels to work with versus the three 20 MHz channels in the 2.4 GHz frequency.[29]

The combination of a 40 MHz channel, two spatial streams and the packet aggregation gains of frame aggregation will provide a 300 Mb/s maximum data rate. If four spatial streams are used this peak data rate can increase to 600 Mb/s.

(e) Implementation Issues

To ensure a seamless transition to the newer 802.11n technology the specification was designed for backwards compatibility. Legacy clients (802.11a/b/g) will operate without problems in an 802.11n network. For instance, 11a clients operate in the 5 GHz spectrum and will continue to do so in a 5 GHz 802.11n network.[30]

The downside to this configuration is a reduction in performance when older devices operate in a 802.11n system. Legacy clients in an 11n network will reduce the overall throughput of the 11n system. The peak performance of an 11g client is ¼ that of 11n. So if an 11g client is operating at 10 Mb/s then the 11n capacity will only be at 40 Mb/s (1/4 the 11n system).[31]

As previously mentioned transmit power is important and can be a key implementation issue for 802.11n systems. Additional power is required for each additional MIMO antenna and to operate in the wider 40 MHz channel mode. Each MIMO antenna will require a separate PA and RF chain so power consumption and resulting cost is increased.[32] Likewise, to keep an equivalent range shown in a 40 MHz system for a 20 MHz implementation will take much more transmit power.

Unfortunately the gain of a MIMO Spatial Multiplexing (SM) system assumes the existence of multipath and uncorrelated signals for each of the spatial streams being transmitted. Radio signals need to reflect off of walls and furniture to cause the receiver to see multiple representations of the same signal arriving at different times and amplitudes. In typical residential homes this phenomena occurs but situations could arise where the intended multipath effect is not present thereby decreasing the benefit of MIMO and increases in system capacity.

In 802.11n systems it is important for all traffic to be classified as priority, best efforts or background type traffic. Implementing the Quality of Service traffic classifications called Wireless Multi-Media (WMM) will be a difficult but important process. The correct tagging and classification will be critical to properly manage the traffic in a wireless home network. Not all consumers will be up to this task and could cause performance issues that will reflect poorly on the technology.

A final implementation and design consideration in the 802.11n specification are the use of two interference control mechanisms called Dynamic Frequency Selection (DFS) and Transmit power Control (TPC). DFS is a feature that checks for the presence of military radar operating in the 5 GHz frequencies and if detected requires the 802.11n device to utilize other available frequencies. TPC requires 802.11n devices to reduce their transmit power if they are operating very close to each other.[33] Although these interference controls are a burden to the 802.11n system developer they are not hugely disruptive issues but must be taken into account when Wi-Fi systems are being implemented.

Variations To High Throughput Wi-Fi (802.11n) Specification

Because the 802.11n specification has provisions for many optional features there exists the ability of a chip designer and access point developers to offer a variety of standards compliant 802.11n implementations. The major suppliers of 802.11a,b,g chipsets and start-up companies have begun to optimize video transport within the home over the 5 GHz spectrum associated with 802.11n.

The ability to use the additional 20 and 40 MHz channel bandwidths, aggregate baseband and RF functions onto a single System on a Chip (SoC), utilize MPEG4/H.264 compression and take advantage of MIMO technologies have led to many advancements in video distribution within the home.

Some new and innovative chip design shops have creatively utilized the various core technologies listed below to great advantage while still complying with the standard.

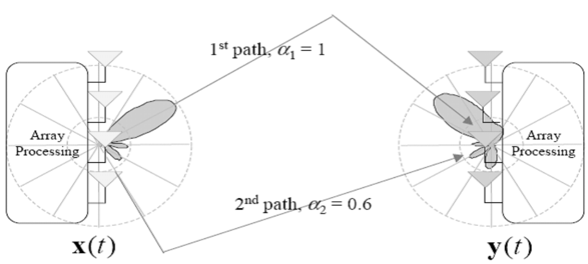
- Beamforming MIMO (simultaneous transmission of the same data streams over multiple antennas)
- Adaptive channel modulation and coding options
- Enhanced QoS (prioritizing video, data and voice services)
- Radio resource techniques (utilizing feedback information from the client devices while transmitting)
- Linking compressed video (H.264) chip and MIMO RF technologies
- Integration of advanced Power Amplifier (PA) and antenna designs

As these many variations of the 802.11n standard enter the market it will be very important for the Wi-Fi Alliance and network operators offering wireless home solutions to properly certify interoperability across access points and client devices. The beamforming variants discussed here add the most gains and consequentially the greatest complexity and interoperability issues.

(a) 802.11n Standards Compliant Implementations: Beamforming MIMO

MIMO technology has the two major design choices of either Spatial Multiplexing (SM) or beamforming. By choosing the beamforming approach the video transport can be optimized for high performance and distance. Beamforming MIMO allows for the control of the RF signal and actually steers the signal from the access point to where the client device is located. These RF pattern adjustments are made very quickly from moment to moment if either the client device moves, multiple clients are in the range of the Access Point (AP) or the RF conditions vary naturally. By choosing the beamforming MIMO option over SM MIMO the 802.11n chip or system designer typically foregoes the bandwidth gains of SM MIMO. In many cases multiple unique data streams are no longer being transmitted from each antenna simultaneously.

Figure 10: 802.11n Beamforming MIMO Architecture[34]

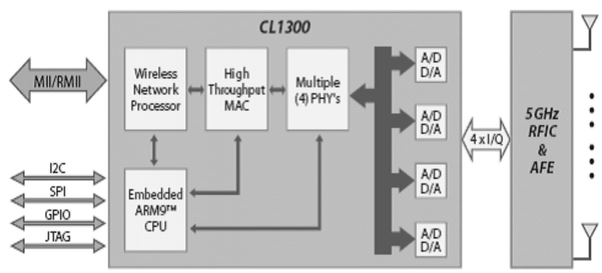


In the 802.11n specification there are chip level and system level implementations of beamforming. Chip level designs typically use mathematical and DSP based processing power to manipulate the phase of the baseband RF signal. As shown in Figure 10, these phase shifted signals are transmitted out over integrated MIMO antennas to accomplish the beam forming or optimized RF patterns.[35]

(b) Chip Level MIMO Beamforming Designs

There are many creative companies developing chip level MIMO beamforming designs. Because of the inherent bandwidth restrictions on beamforming MIMO, most chip level MIMO beamforming designs utilize an H.264 codec chip along with the baseband and RF chipset to ensure guaranteed video quality and wireless transmission of multiple HD video content.

Figure 11: 802.11n Chip Level MIMO Beamforming Architecture[36]



In one such design a 20 MHz channel carries three MPEG4/H.264 streams (1080p, 60fps) at between 8 to 10 Mb/s each. Vendor testing claims of a consistent 30 Mb/s (UDP transmission) of capacity across 25 meters and through multiple walls in a suburban home are one of many published 802.11n results.[37]

Another key beamforming MIMO chip design option involves either having the beamforming performed at one end (AP side) or both ends (AP & client). In one particular implementation the semiconductor vendor focused on delivering optimized MPEG4/H.264 HD video content where its chip is located at the transmit end of the video stream only. Therefore, it is placed in a central distribution point (access point, home gateways, multi-room DVR‟s) of HD content and interworks with standard off the shelf 802.11n chips (and MPEG4 codecs) in the client device such as a remote STB or HD TV display.[38] The reliability and distance benefit of beamforming MIMO combined with a low total cost for a whole home implementation (because of the ability to interoperate with low end 802.11n chipsets) is a very advantageous design. A second major chip vendor decided to implement their beamforming MIMO design by using their chip is used at both ends of the link. The result is probably a higher performance design but at a greater cost.

The decision to use either 20 MHz or 40 MHz (two bonded channels) channel bandwidths is likewise important. If the 802.11n implementation utilizes a single 20 MHz channel in the 5 GHz frequency to transport MPEG4 HD content then the transmit power and antenna gains can be more focused and contribute to a better range result. An implementation using two bonded 20 MHz channels (40 MHz) will need higher transmit power to get the same distance and throughputs (all things being equal). The second start-up vendors beamforming chip design utilizes 4 transmit and 4 receive antennas in their chip with a 40 MHz channel. This chip design is an excellent way to get both high capacities (40 MHz channel) and also reliable and long range performance (4x4 Beamforming MIMO).

Additionally, the transmit power and number of power amplifiers decision is crucial as it is a major contributor of cost and ability to integrate on a chip and circuit board level for vendors. Transmit power level of up to 100 mW are known to provide the best linear characteristics of the highly modulated signals (e.g.- 64QAM) of the 802.11n specification. Likewise, the number of RF chains and quantity of MIMO antennas to choose from are critical performance and cost decisions. In one illustrative beamforming design the entire video stream (comprised of 3 MPEG4 streams) is sent simultaneously over the 2 RF chains using 2 PA"s where 4 antenna choices are possible.[39] Another chip level beamforming vendor chose to utilize 4 RF chains, four PA"s and 4 antennas.[40] This design probably has more power and more beamforming patterns possible but probably at a higher cost.

The beamforming software in a particular vendors design is the "secret sauce" for their MIMO implementation. Software that chooses the best antennas and in effect creates an optimum transmit antenna pattern can make all the difference in performance. The ability to have many beamforming pattern options and continually (and quickly) adjust the pattern due to the varying conditions of the RF environment within the home cannot be underestimated. The 802.11n standard allows for information from the client side of the link to be sent back to the transmitters so that these software decisions can be made.

Unfortunately, there are multiple options and much complexity written into the standard for this critical feedback information. For instance, the implicit feedback option means the 11n client provides very limited information and the chip vendors implementation calculates how it should change its beamforming from moment to moment. The explicit feedback option in the standard spells out the continuous feedback of information the client must send. To date, no chip manufacturers of client end only chips have implemented this portion of the standard. Start-up chip vendors that have their MIMO beamforming solution on one end only are utilizing implicit feedback while those with chips on both ends use some form of explicit feedback to adjust the beam.[41]

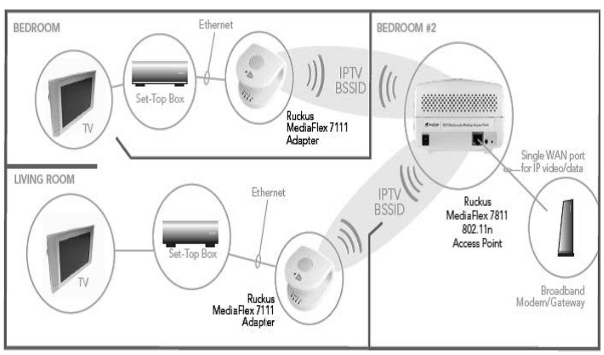
Because of the many options and complexity the interoperability testing and certification of the beamforming portion of the 802.11 standard is very much in its infancy and will be an extremely difficult area to enforce.

Suffice it to say, when beamforming is used between different vendors equipment the performance will most likely fall off considerably and firmware or even hardware changes will be require for interoperability.

(c) System Level MIMO Beamforming Designs

Companies that utilize off the shelf 802.11n chips and add their own hardware and software implementations of beamforming MIMO technologies are known as system level MIMO beamforming designs. Their area of expertise is in the design of enhanced multi antenna systems and beamforming software algorithms that continually create and direct new RF patterns.

Figure 12: 802.11n System Level MIMO Beamforming Architecture[42]



Key enablers for system level MIMO beamforming designs is the ability to create many different patterns using multiple antenna choices, directional antennas, sophisticated software that continuously adjusts and changes patterns, and efficient power amplifiers.

One such vendor‟s implementation is able to get high capacity, performance and range by utilizing directional antennas (6 in 1 of their 5 GHz implementations), beamforming technology and the higher transmit power possible with outboard power amplifiers (250mW). A standard 3x3 MIMO baseband chipset from a traditional 802.11 chip vendor is combined with 3 PA‟s and RF chains to 6 directional antennas. At any point in time an optimized beamforming pattern is sent out over the best 3 antennas using knowledge obtained from the RF channel and the far end wireless adapter. At any point in time hundreds of different antenna patterns are possible that can add gain to the system performance or even reject interference.[43]

As in many of the other creative implementations of the 802.11n standard the system level MIMO beamforming designs are only able to get the improved range and bandwidth performance claimed when their MIMO beamforming units are used at both ends of the video or data stream. Optional software programmable configurations are possible to allow the main access point to communicate with industry standard, off the shelf 802.11n adapters and extenders in laptops and other devices, but the performance will be reduced.

A nice advantage of the system based solution is that it is not optimized for video at the chip level and therefore offers a true bidirectional home networking solution where both video and data services can be transported over the data stream. QoS (802.11 WMM) is implemented so prioritized video and best efforts data transmission share the 300 Mb/s of capacity.

A key challenge for the network operators incorporating higher end implementations of the 802.11n standard in their home networking solutions is the ultimate performance these devices obtain when deployed in a mixed operating mode. The in home network performance may match claims in a perfect end to end single vendor environment but degrade severely when a variety of vendors and chipsets are used in a home network. In addition, because the beamforming implementation at their access point is a unique implementation of MIMO beamforming it may not be compatible with standard 802.11n client device implementations (such as explicit chip level feedback designs).

Optimized Video At 5 Ghz or Proprietary Wireless HDMI technology.

Targeting the large, unused, contiguous spectrum available to Wi-Fi technology in the 802.11n 5 GHz frequencies some vendors have chosen to create a unique specification targeted for video transmission and wireless HDMI applications.

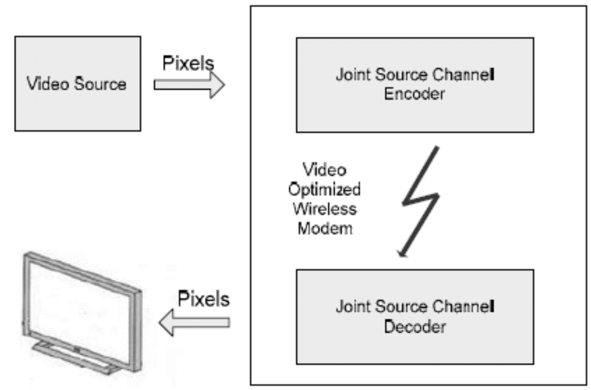
Amimon is one such start-up silicon vendor that has designed a very effective chip design optimized for video transmission at 250 to 800 Mb/s. They have created a consortium of vendors (called WHDI), that includes CE manufacturers, to create a specification and standard for wireless HDMI.[44]

Their technology also takes advantage of the multi antenna gains possible with MIMO technology (4x5 MIMO) and 40 MHz wide channel bandwidths.[25] As a result, large capacity video streams are possible with their implementation.

In addition, they claim their transmission does not use compressed video. Amimon appears to perform some form of compression but at lower compression ratios (< 4X) that allow it to be categorized as lossless compression, or for their definition, uncompressed.

The key to their uncompressed video is their ability to combine real time video processing of the video source with channel coding and modulation function present in traditional wireless transmission systems. The Amimon chip prioritizes the source video components according to their importance and only keeps the most significant bits. By tightly linking an understanding of the varying RF channel with the actual encoding and processing of the video source allows the Amimon reference design shown in Figure 13 to optimize the wireless connection for video applications.[45]

Figure 13: Optimized Video Proprietary Architecture[46]



As in all proprietary designs the key for this technology will be to obtain standards approval for the reference design in the 802.11 study groups and then build an ecosystem of supportive CE vendors.

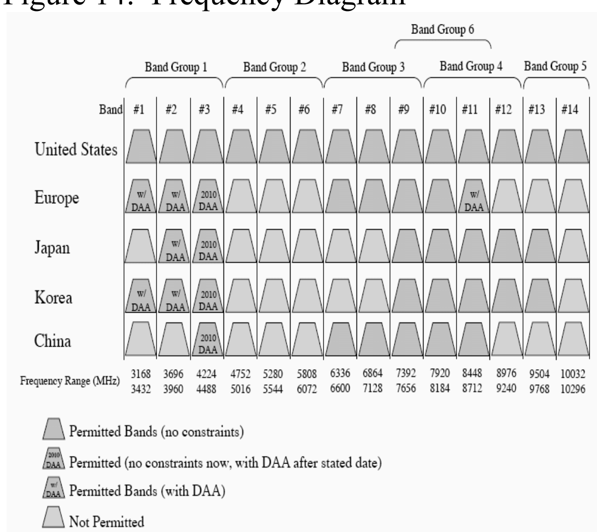
Ultra Wide Band (UWB) Technology (ECMA-368)

Ultra-Wide Band is a radio technology that uses a 528 MHz channel bandwidth (compared to the 20 or 40 MHz channels of 802.11n) within various frequencies from 3.1 to 10.3 GHz. The technology is heavily promoted by the WiMedia™ Alliance and utilizes the well known OFDM technology and is being used to transport compressed video (1080p at 30 fps) within a single room only.

WiMedia UWB can support data rates of up to 480 Mb/s and transmits at very low power levels. The FCC has placed transmit power levels (.1 mW) on these frequencies as a majority of the bands compete with other users operating in these bands. As a result of a .1 mW peak transmit power limit spread across 528 MHz the Power Spectral Density (PSD) is quite low for UWB technologies whch results in operation across very short distances (< 10 meters). Because UWB technology operates across such a wide channel bandwidth it is very resistant to frequency fading. By utilizing 128 OFDM 4 MHz subcarriers across the full 528 MHz wide channel it is able to guarantee quite high quality HD video with its built in frequency diversity.

Additionally, there is over 7 GHz of spectrum available in the various band groups from 3.1 GHz to 10.3 GHz. Only the U.S. has access to the full allocation as Europe is allowed to use only 3.250 GHz of spectrum. Also in many of the bands (e.g.- 3.5 GHz) UWB operation is required to Detect And Avoid (DAA) outside operation from WiMAX, radar operators and others before transmitting in those frequencies.

Figure 14: Frequency Diagram[47]



Although the UWB specification originally started out originally with much promise but has had much difficulty gaining traction in the marketplace and standards bodies over the years. Seven major vendors such as T-Zero, WiQuest, Radiospire, Pulselink, Focus Semiconductor, Artemi and Intel were recently forced to either shut down or merge their UWB operations leaving just two or three vendors left in the WiMedia space.[48] Most remaining vendors appear to begin offering UWB technology over coax as the WiMedia group seems to be retrenching slightly from the pure wireless applications.

One of the biggest disappointments for the UWB technology and the WiMedia Association has been their inability to obtain IEEE 802.15.3a Personal Area Networks (PAN‟s) standards approval as the specification languished for three years. Eventually the committee was disbanded and the WiMedia Association was forced to gain approval from the ECMA international standards organization[49].

It appears that the biggest issue associated with the UWB technology is that the very low power limits imposed by the FCC have limited its applications to in-room compressed HD video. This limited application is being outstripped by in-room uncompressed video at 60 GHz or whole home compressed HD video possible with the many 802.11n technologies and their variants at 5 GHz. At this writing the UWB technology appears to have become a shelved technology that has never lived up to its original hype.

Very High Throughput 60 GHz Technology (Proprietary Wireless HDMI)

The millimeter wave spectrum is very advantageous for the high bandwidth requirements of transporting HD video signals because of the large amount (7 GHz of spectrum between 57 and 64 GHz) of unused spectrum available in the 60 GHz frequencies. This technology certainly comes closest to pure wireless HDMI as full HD quality video at 1080p and 60 fps is possible over a 4 Gb/s stream including the A/V control signaling associated with HDMI. The well known benefits of transporting uncompressed video are therefore its major attraction. Additional benefits of 60 Ghz include the uniformity of this spectrum availability across the U.S., Europe, Japan, China and other major markets. The combination of large spectrum bands and wireless technology advances of MIMO makes multi gigabit rates possible.

Unfortunately, the well known downside of 60 GHz transmission is the short distance limits due to signal strength losses through free space. Although the antennas possible at these high frequencies can be made very small and compact the maximum transmit power requirements for even very short distances are quite large. SiBeam is the major chipset vendor in this area and is the leader in promoting their specification for standardization. They require a 7 Watt power amplifier to transmit one video stream 10 meters in a Non Line Of Site (NLOS) mode or 25 meters with Line Of Site (LOS).[50] Further integrated circuit advancements may reduce the power to 4 Watts and eventually 2 Watts but the large power requirements and resulting high total system costs continue to make this technology very much a niche offering.

The ability to reduce power, costs and obtain standardization will be very important for 60 GHz technology adoption. At this writing the SiBeam specifications are well positioned for being pushed through the 802.11ad Very High Throughput (formally 802.11vht) task group.[51] The WirelessHD™ (WiHD) consortium has also been effective at promoting this specification within the standards bodies. As is sometimes the case, this specification is also being promoted in the IEEE 802.15c working group which is part of the Personal Area Networks (PAN) study groups.

Because this technology and the frequency it operates in will most likely always be an in-room solution only, its applications and uses will most likely are limited and therefore be a primarily a niche technology for home networking solutions.

TEST RESULTS OF VARIOUS WIRELESS HOME NETWORKING SOLUTIONS

Preliminary Findings

Preliminary test results of unmodified 802.11n draft technology, based on a sample of standalone gateway devices that utilize 2.4GHz and Gigabit Ethernet LAN and WAN ports provided an insight into the potential challenges associated with delivering triple play services over today‟s wireless home network solutions.

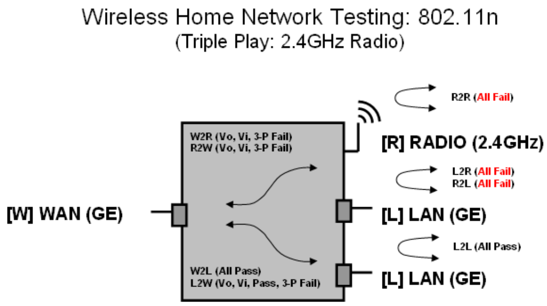
The test configuration included multiple services (Video, Voice and Data) over each wireless home networking device in specific directions as outline in Table 6.

Table 6: Triple Play Test Traffic Direction

Abbreviation	Description
L2L	LAN to LAN
W2L	WAN to LAN
L2W	LAN to WAN
W2R	WAN to Radio
R2W	Radio to WAN
L2R	LAN to Radio
R2L	Radio to LAN
R2R	Radio to Radio

Key findings indicate that when the home wireless devices were loaded with service provider traffic that Gigabit Ethernet ports including LAN to LAN performance and LAN to WAN performance met performance thresholds, where as most test that required use of the Radio (i.e. Radio to Radio and L:AN to Radio did no meet requirements.

Figure 15: 802.11n Wireless Test Flows



More detailed results are summarized in the table below in Table 7, with results in bold indicating they fell outside the threshold requirement.

Table 7: 802.11n Triple Play Performance

802.11n, 2.4GHz			
Service	Data	1-Way 12 Mbps Video	Voice G.711
Throughput (Mbps) Required > 100	L2L: 390.1 W2L: 117.6 **L2W: 41.3** **W2R: 8.9** **R2W: 15.1** **L2R: 8.6** **R2L: 7.5** **R2R: 3.2**	L2L: 11.9 W2L: 11.9 L2W: 11.9 W2R: 11.9 R2W: 11.8 L2R: 11.7 R2L: 11.8 **R2R: 10.9**	Not tested
% Bytes Lost Required: < 0.1	Not tested	L2L: 0.0 W2L: 0.0 L2W: 0.0 **W2R: 0.6** **R2W: 1.0** **L2R: 1.8** **R2L: 1.4** **R2R: 8.5**	L2L: 0.0 W2L: 0.0 L2W: 0.0 **W2R: 0.2** **R2W: 0.4** **L2R: 0.1** **R2L: 0.2** **R2R: 3.0**
Jitter (ms) Required: < 2	Not tested	Not tested	L2L: 0.0 W2L: 0.5 **L2W: 2.3** **W2R: 2.4** **R2W: 4.0** **L2R: 2.8** **R2L: 3.0** **R2R: 8.5**
1-way delay (ms) Required: <120	Not tested	Not tested	L2L: 0.8 W2L: 1.3 L2W: 3.5 W2R: 4.7 R2W: 8.0 L2R: 7.0 R2L: 6.5 R2R: 28.2
MOS Required: > 3.5	Not tested	Not tested	L2L: 4.4 W2L: 4.4 L2W: 4.3 W2R: 4.2 R2W: 4.0 L2R: 4.0 R2L: 4.1 R2R: 4.0
Key Finding	Radio speeds to not support 100 Mbps data product speeds	Radio, esp. Radio to Radio does not support <0.1 bytes lost for video.	Radio, esp. Radio to Radio does not support < 2ms jitter, or < 0.1 bytes lost for voice.

Testing in 5GHz bands indicates some significant improvement over 2.4GHz and as more devices become available in the 5GHz band this will be the subject of additional testing.

Further testing with a wider range of wireless home networking devices during 2009 will be the subject of a future paper.

Future evaluation and testing 802.11n Solutions for Home Networking

The key criteria in evaluating implementations will be:

- Capacity of multi-play capabilities, for example to determine whether the entire channel is dedicated for video transmission or other services are possible on the bandwidth simultaneously.
- Consistent and reliable performance such as voice/video quality (latency, jitter and packet loss) over range and network load.
- Effective throughput and capacity over path loss.
- Receiver sensitivity and spectral efficiency.
- Service specific quality matrix such as video startup and zapping latency.

THE SERVICE PROVIDER SUPPORTED WIRELESS HOME NETWORKING BUSINESS MODEL

What is the cost of adding wireless home networking for the service provider? Are the costs different between the various wireless technologies? We explore the economics across multiple wireless home networking solutions.

Developing the wireless home network "Pain Threshold"

We assume that the average revenue from wireless home networking is $10 per month per subscriber. Of this $10 per month we assume that 80%, or $8 is required for sales and marketing, customer care, billing and G&A per subscriber per month. This leaves $2 per month per subscriber to cover all wireless home network related technology costs.

Considering the technology costs we initially reviewed the incremental costs of wireless home networking technology we considered bit the A-end or Access Point incremental cost to add wireless and the B-end, in the case that a specialized B-end device is needed to complete the link.

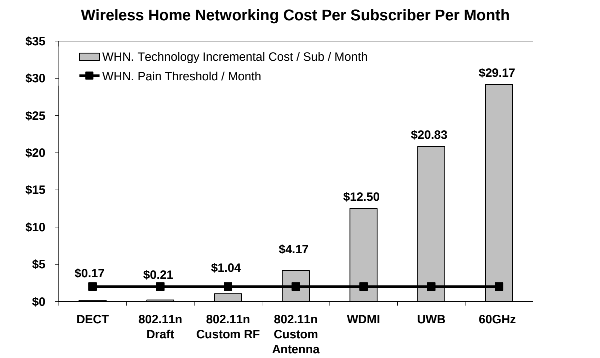
Table 8: CAPEX Cost of Adding Wireless

	Access Point Incremental Cost ($US)	Client Cost ($US)	Total Service Provider Cost
DECT	$4	Any DECT Handset	$4
Wi-Fi 802.11n (Baseline)	$5	Any Wi-Fi	$5
Wi-Fi 802.11n (Celeno) *	$25	Any Wi-Fi	$25
Wi-Fi 802.11n (Ruckus) **	$50	$50	$100
WDMI (Amimon) ***	$150	$150	$300
UWB (Tzero) ***	$250	$250	$500
60GHz (Sibeam) ***	$350	$350	$700

* Note: Celeno are targeting a chip-based solution for embedding in CPE/RG devices so no additional cost is needed for casing, packaging, cabling.
** Note: Ruckus had no plans for a CPE/RG integrated device and would only consider on a business case basis so the cost is based on an add-on access point MediaFlex 7000 at the estimated cost to the Telco rather than the list price.
*** Note: WDMI, UWB, and 60GHz solutions are primarily targeting cable replacement in 1st generation of products so the cost is including casing, packaging, cabling, power for each end of the link.

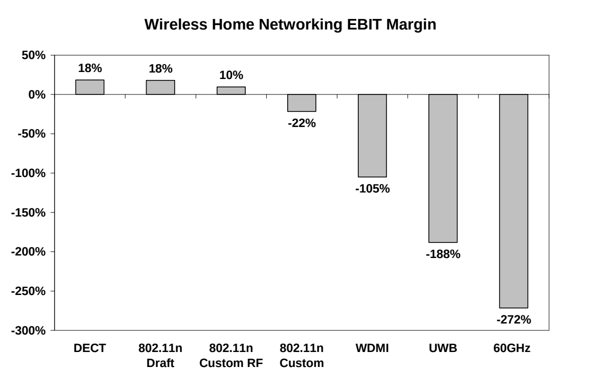
How does the pain threshold described above compare to the cost of the wireless home networking technologies? To make this comparison we assume a 24 month life span for wireless CPE accommodating the shorter life cycle associated with emerging wireless home networking technologies. The resulting cost per sub per month can be compared to the $2 pain threshold in Figure 16 below.

Figure 16: Wireless Home Networking "Pain Threshold" and Technology Costs



The analysis of EBIT margin, highlighted in Figure 17 below, indicates that unmodified 802.11n and 802.11n with minimal customizations do not add substantially to the CPE cost or require a custom or proprietary chip at both ends of the link, support a business case for wireless home networking.

Figure 17: Wireless Home Networking EBIT Margin



This clearly demonstrates the economic advantage of leveraging established device and certification ecosystems; in addition to scale and forwards and backwards compatibility in selecting a suitable wireless home networking solution.

We can summarize that 802.11n, or a variant of 802.11n, that supports a large established ecosystem of devices and also provides robustness for video, in conjunction with DECT for telephony; can potentially provide a wireless home networking foundation that is economically viable for service providers.

CONCLUSION

This paper highlights that there is not a "one-size fits all" approach for wireless home networking in support of service provider applications.

Testing of 802.11n devices against service provider triple play services indicates that it is technically challenging to utilize only a single 20 MHz radio channel for all services and service provider product requirements.

The wireless home networking technologies that are most suited to integration into the service provider CPE are those that (a) have a significant established ecosystem, (b) have volumes that support economies of scale, (c) are low cost, (d) support backwards interoperability, (e) have the flexibility to support multi-regional variations, and (f) are able to operate in well penetrated radio-congested environments.

Preliminary findings indicate that 802.11n, or a variant of 802.11n that is backwards compatible with the standards, for data and video and DECT for voice both have well established ecosystems and compelling economics and are strong candidates for service provider CPE integration.

END NOTES

[1] CES 2009, Jan 7, 2009, "Netgear ITV2000 Internet TV Player Lets You Watch BBC News Whenever You Feel Like It", http://i.gizmodo.com/5125266/netgear-itv2000-internet-tv-player-lets-you-watch-bbc-news-whenever-you-feel-like-it & http://www.netgear.com/Products/Entertainment/DigitalMediaPlayers/ITV2000.aspx

[2] CES 2009, Jan 7, "Linksys by Cisco Wireless Home Audio system unveiled", http://www.engadget.com/2009/01/07/linksys-by-cisco-wireless-home-audio-system-unveiled/

[3] CES 2009 Preview: Whole-home 1080p60 HD via Wi-Fi, Jan 6 2009, Gadgetell, http://www.gadgetell.com/tech/comment/ces-2009-preview-whole-home-hdmi-1080p60-via-wi-fi

[4] CES 2009, Jan 7, 2009 "Cisco Simplifies Digital Media Access for Consumer" & http://www.linksysbycisco.com/US/en/promo/Media-Hub

[5] U.S. Broadband Households, "Home Servers & Consumer Storage", Parks Associates, 2008, p.27

[6] Seagate magnetic storage price trends using a blend of retail pricing for magnetic Hard Disk Drives.

[7] TCM Net, Jan 6, 2009, http://blog.tmcnet.com/design-vs-functionality/2009/01/itunes_abandons_drm.html

[8] Jos van der Loop, SiTel Semiconductor BV, Jan 2009, DECT World Forum.

[9] Forrester Research, Home Networking Dynamics & Opportunities, Feb 11-12 2009, Charles S. Golvin

[10] Forrester Research, European Consumer Technology Adoption Study Q4 2006 & Q4 2008 Online Devices, Access And Telco Survey

[11] ETSI Standards, EN 301 893 V1 4.1 (2007-07), ETSI EN 302 502 V1.1.1 (2006-11), EN 301 489-17 V1.3.2 (2007-06), SIEMIC Inc., 2008

[12] ETSI Standards, EN 301 893 V1 4.1 (2007-07), ETSI EN 302 502 V1.1.1

(2006-11), EN 301 489-17 V1.3.2 (2007-06), SIEMIC Inc., 2008
[13] SiBeam Presentation, "Wireless Without Boundaries, 2009
[14] Network Zone Web Article, "Wireless Chipsets that Deliver Guaranteed Bandwith for High Speed Wireless Networking", November 3, 2008
[15] Open Cable Home Specifications Home Networking 2.0, Reserved Services Domain Technology Specification (OC-SP-RSD-TECH-101-080630), 2008 Cable Television Laboratories, Inc.
[16] "802.11n: The New Wave in WLAN Technology", Yarali & Ahsant, September 2007, Mobility 2007 conference Technical Paper
[17] Wi-Fi™ Alliance is a trade group that owns the trademark to Wi-Fi.
[18] WHDI™ is a trademark of AMIMON, Ltd.
[19] WiMedia™ is a registered trademark or service mark of the WiMedia Alliance
[20] WirelessHD™ is a trademark of the. WirelessHD consortium
[21] Kaido Kert, "Skype and DECT: Taking Skype Beyond the PC", DECT World Forum, Feb 3, 2009
[22] Comtrend, "Access and Home Networking", Andrew Morton, 2008.
[23] HP Whitepaper, 802.11n drives an Architectural Evolution, BitPipe, 2008
[24] The Beamforming Techniques in Wireless MIMO Systems, IC Access Lab, graduate Intitute of Electrical Engineering, April 2008
[25] "The future of wireless networks: Understanding 802.11n", Paulin and Rabjohn, Oct 2008, CommsDesign article
[26] WLAN MIMO Operation, Peter Cain, Agilent Technologies, October 2007.
[27] Test Strategies for 802.11n MIMO Devices, Minarsky, Octoscope, October 2006
[28] 802.11n Wireless Technology Overview, Cisco White Paper, 2009
[29] "The Practical realities of 802.11n in home networking applications", Morris, May 2008, CommsDesign Article
[30] Reliable Wireless Multimedia Streaming: A comparision of UWB and 802.11n, T-Zero White Paper, January 2009
[31] 802.11n Wireless Technology Overview, Cisco White Paper, 2009
[32] 802.11n Delivers Developing Ecosystem, McCartney, July 2008, CommsDesign Article
[33] 802.11n: Look before you Leap", Motorola White Paper, 2008
[34] IEEE Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Enhancements for Higher Throughput, IEEE P802.11n/D2.0, 2008
[35] Beamforming Webinar, Ruckus Wireless, February, 2009 http://ruckus.na3.acrobat.com/p31524561
[36] Celano product Brief, "CL1300 Wi-Fi SoC", 2009, http://www.celeno.com/Resources/pdf/
[37] Connections 2009 Conference, Blogspots, "Celano Responds to Follow-up panel questions", January 20, 2009
[38] Celano Test Report: CL1000, Celano Wireless Whitepaper, 2008
[39] Celano Test Report: CL1000, Celano Wireless Whitepaper, 2008
[40] Network Zone, "Wireless Chipsets that Deliver Guaranteed Bandwidth for High Speed Wireless Networking", November 3, 2008, EN-Genius Publishing Limited
[41] Celano Test Report: CL1000, Celano Wireless Whitepaper, 2008
[42] MediaFlex 7000 Series Product Description and Specification, 802.11n Multimedia Smart Wi-Fi Acess Point and Receiver, 2008, Ruckus Wireless
[43] MediaFlex 7000 Series Product Description and Specification, 802.11n Multimedia Smart Wi-Fi Acess Point and Receiver, 2008, Ruckus Wireless
[44] Enabling wireless uncompressed HDTV connectivity with a unique video modem approach, Meir Feder, Amimon Ltd., 2008
[45] WHDI Transmitter and Receiver datasheet, Amimon Ltd., 2008
[46] Enabling wireless uncompressed HDTV connectivity with a unique video modem approach, Meir Feder, Amimon Ltd., 2008
[47] WiMedia Alliance, "Worldwide UWB Regulatory Status", 2009

[48] In-Stat, "UWB Market sees major setbacks, but some progress", January 26,2009

[49] WiMedia Alliance White Paper, "UWB Standards", Stephan Wood, January 1, 2008

[50] Wireless HD Specification,SiBeam, October, 2007

[51] IEEE 802.11 report, Netgear Report, Jones-Petrick & Associates, January 2009

ISBN 0-940272-01-6; 0-940272-08-3; 0-940272-10-5; 0-940272-11-3; 0-940272-12-1; 0-940272-14-8; 0-940272-15-6; 0-940272-16-4; 0-940272-18-0; 0-940272-19-9; 0-940272-20-2; 0-940272-21-0; 0-940272-22-9; 0-940272-23-7; 0-940272-24-5; 0-940272-25-3; 0-940272-26-1; 0-940272-27-X; 0-940272-28-8; 0-940272-29-6; 0-940272-32-6; 0-940272-33-4; 0-940272-34-2; 0-940272-35-0; 0-940272-36-9; 0-940272-37-7; 0-940272-38-5; 0-940272-39-3; 0-940272-40-7; 0-940272-41-5; 0-940272-42-3; 0-940272-43-1; 0-940272-44-X; 0-940272-45-8; 0-940272-46-6; 0-940272-47-4; 0-940272-48-2; 0-940272-49-0; 0-940272-50-4; 0-940272-51-2; 0-940272-52-0; 0-940272-53-9; 0-940272-54-7