THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:

NCTA
Technical
Papers

A New Economic Coax Approach to MDU and Hotel Internet Access

Jack Terry
coaXmedia, Inc.

Abstract

This paper presents a new, but DOCSIS-compliant, economic system approach for multi-dwelling unit (MDU), hotel, and university campus or hospital high-speed Internet access that operates effectively in existing in-building TV coax. The system offers plug and play end-user attachment without PC reconfiguration or installation of an Ethernet NIC card as multi-megabit Internet access is achieved through the use of the PC's existing parallel or USB port. Within-MDU ingress interference is isolated from the main hybrid fiber-coax network and the system offers improved bandwidth management and efficiency, particularly in the upstream or return direction. The coaXmedia user interface consumes less than one tenth of the power of that of a Cable Modem thus making it easy to provide for lifeline services.

BACKGROUND

The demand for high-speed Internet access is driving the telecommunications industry like few forces have in the past. While the Cable and Telephone industry position their networks for the future, ever-changing technology has previously made it both costly and risky to invest in new delivery systems.

Most current approaches for delivery of Internet services in MDUs utilize telephone wiring in "data above voice" configurations. Such approaches usually require selective identification and disconnection of each telephone pair and the insertion of a modem function at the central end of the telephone loop. Such intrusive installation is both costly and time consuming. A second modem is required at the user end of the telephone pair to connect to the user's PC or in-home network. Since MDU telephone wiring generally has worse inter-pair crosstalk performance than that of outside wiring and suffers considerable electrical ingress interference it is usual to insert the data on the telephone loop within the building to ensure adequate performance.

The high frequency loss of longer telephone loops between the central office and the MDU considerably limits their potential two-way transmission speed.

The use of low-cost wireless data transmission works well where the distances are short and spectrum is abundant. For densely populated MDUs this is not usually the case.

THE PRESENT CABLE ENVIRONMENT

Cable Modem Internet service has now penetrated well over one million residences and has become extremely popular due to its exceptional speed. However the introduction of Cable Modem service in MDUs is problematic due to the complex and irregular topology of the TV coax wiring and the sharing of limited available upstream bandwidth. In addition, points of ingress interference in MDU coax distribution and home wiring are very difficult to locate and particularly difficult to isolate. Such ingress interference can cause failure of two-way services to all users in an MDU and potentially other users upstream of the MDU on the Hybrid Fiber-Coax (HFC) network.

Both Cable Modem and Telephone loop data modems are usually interfaced to the PC using an Ethernet 10baseT connection. This requires that a Network Interface Card (NIC) be installed in each PC and the PC network software configured. Since the average PC users are not usually technically skilled, this installation and/or configuration is frequently performed by the Cable or Telephone network provider. In this way the network provider becomes potentially liable for problems in the PC, often when the trouble is not related to the network provider's work. While this issue can be alleviated in some cases by use of USB ports, a large proportion of PCs are not so equipped. In hotel/motel situations, users do not generally require networking between themselves and are rarely adept or willing to reconfigure their PCs each time they rent a room or return to their home or office.

MDU and hotel coax distribution systems, which can be served by Cable, Satellite or Broadcast network operators, are usually configured as passive "tree and branch" systems using splitters and/or relatively long coax runs with taps or couplers arranged to serve the apartments or rooms. Such passive distribution arrangements frequently serve from 30 to 100 rooms or apartments and are arranged such that the TV signal levels fed to each apartment or hotel room are typically within a 10 dB range. These coax distribution systems typically have losses in the range of 15 dB to 25 dB and are usually fed from a centralized one-way broadband TV channel amplifier to ensure adequate signal levels for the users. Larger high-rise MDUs and hotels usually have a number of centralized amplifiers each feeding a passive coax distribution sub-system serving separate areas or floors of the building.

THE OPPORTUNITY

The spectrum of the MDU TV services usually lies below 750 MHz, whereas the coax cable can handle frequencies beyond 1 GHz. The passive splitters or couplers, although usually only rated for use in the TV bands, usually perform adequately in terms of loss and/or port isolation when carrying more robust digital signals of up to 1 GHz. Furthermore, the loss per unit length of the in-building coax wiring, rather than being a problem, helps attenuate echoes at these higher frequencies and thus permits much simpler equalization in digital receivers.

Clearly there is an opportunity to utilize the higher frequency spectrum of in-building coax for high-speed Internet access services using robust digital modulation techniques. Ingress interference is much less at frequencies above those of TV channels and, being contained by the one-way characteristic of the central TV channel amplifiers -- at least at the TV downstream channel frequencies and higher, any ingress interference is prevented from exiting the MDU and interfering with the HFC Cable network.

The available above-TV-channel spectrum in in-building coax can be arbitrarily divided up to offer high-speed data in both directions. Due to the relatively high field-strength radiation of portable cellular handsets it is prudent to operate at frequencies of 900 MHz and above. Using presently installed splitters and couplers it is also better to keep to frequencies of 1 GHz and below. This available 100 MHz of available spectrum is plenty to serve the statistical two-way Internet access needs of 50 to 100 users or client modems. If higher capacity is needed, additional downstream spectra can be allocated in bands between 1 GHz and about 1.6 GHz, provided that higher frequency specified splitters are substituted. Such higher uni-directional capacity can provide for additional digital video-on-demand (VOD) services, in either Internet Protocol (IP) format or in native MPEG2 format. In all cases the spectrum between 900 MHz and 930 MHz can be utilized for upstream transmission. The use of this single upstream spectrum provides adequate traffic capacity and simplifies control.

AN ALTERNATIVE APPROACH

An alternative system approach, named coaXmedia, has been devised which takes advantage of the topology and performance of in-building coax distribution to provide high-speed Internet services.

This system architecture is DOCSIS-compliant at a network level, consistent with existing Cable Modem operation and service practices and yet offers plug and play end-user attachment without PC reconfiguration or installation of an Ethernet NIC card. At the same time the approach isolates within-MDU ingress interference from the main hybrid fiber-coax network and provides bandwidth management and efficiency, particularly in the upstream or return direction.

The per-MDU common equipment installation is extremely simple and there is no need for a truck-roll or appointment to provide service to each customer. Indeed, the customer interface can be drop-shipped to the consumer and is easier to hook-up than a VCR. Multi-megabit Internet access is achieved through the use of the PC's existing parallel or USB port using a simple "enabler" which can be optionally loaded from the MDU central hub modem, via the PC's existing serial connector -- no floppy disks or CDs.

The primary purpose of this "enabler" is to place a "connection" icon on the user's desktop for ease of access to the service. There is never the need to perform another enabler load when moving the PC between coaXmedia client modems, such as when moving between hotel rooms or returning home, as the "enabler" does not need to contain any addressing or configuration information.

The coaXmedia client modem is extremely simple since it does not require a tuner or even a microprocessor for its operation. Other simplifications result in a complexity of around a quarter of that of a conventional Cable modem. The client modem is thus very low in cost and this cost will continue to track at significantly less than half of the cost of technology-evolving conventional cable modems. Additionally, the coaXmedia user interface consumes less than one tenth of the power of that of a Cable Modem. Installation costs are minimal and marketing of the service by the Cable MSO is simplified as service may be offered on a same-day trial basis.

The coaXmedia client modem can be packaged on a single printed circuit board housed in a plastic case of approximately the size of a small cellular phone. This case may be included as a pod inserted in a piece of coax cord connected to a coax wall receptacle. This pod will also have a thin data cord with a multi-faceted connector that may be inserted into the parallel, serial or USB connector on a PC or laptop. Future connectors may include an infrared transceiver for communication with similarly equipped PCs or PDAs. Power is provided using a low-cost, single AC voltage, UL/CSA approved, transformer cube.

THE COAXMEDIA ARCHITECTURE

A diagram illustrating the overall coaXmedia architecture is shown in Figure 1.

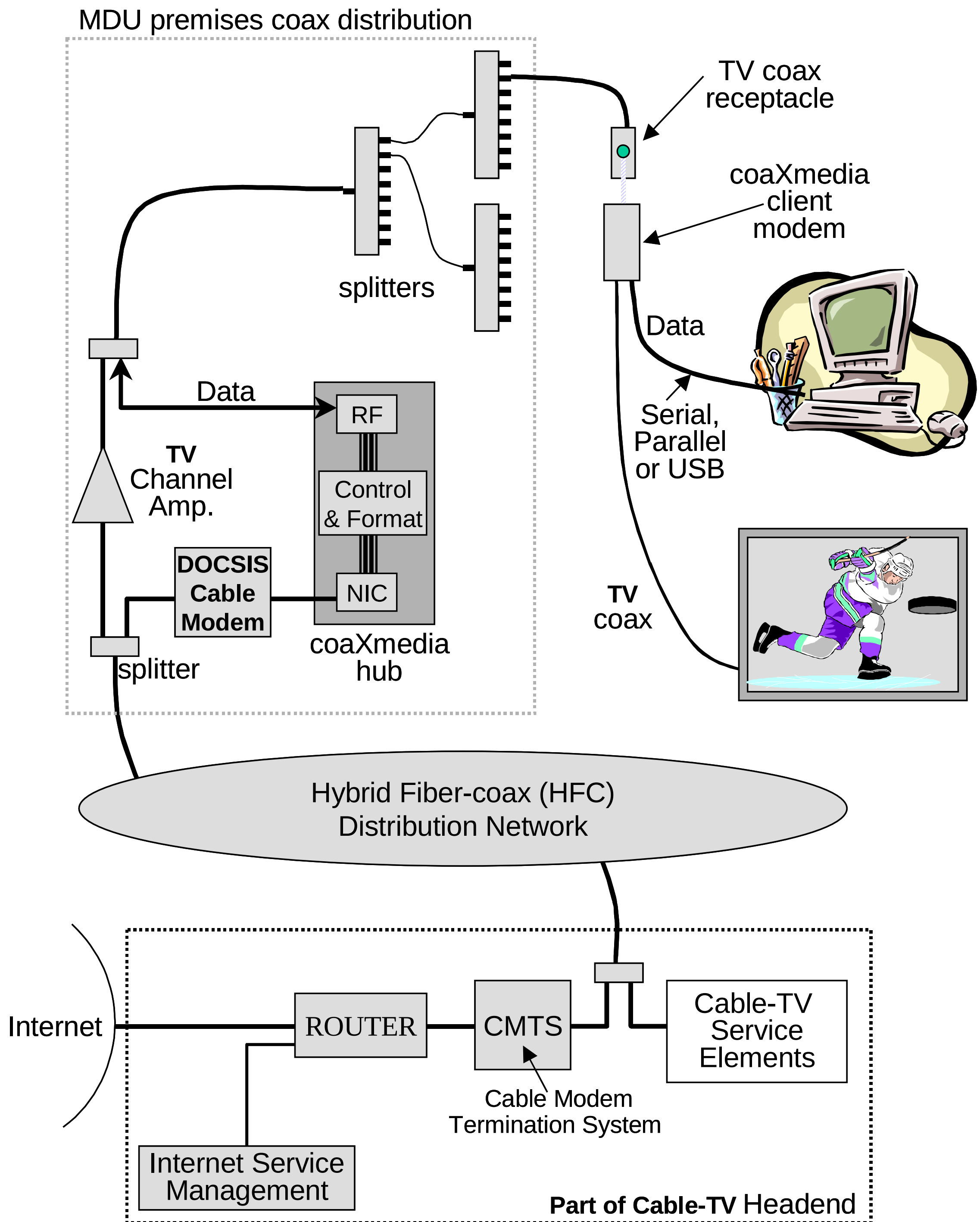


Figure 1. The coaXmedia Architecture

In this arrangement a single DOCSIS-compliant off-shelf Cable Modem is used to serve the statistical data needs of multiple users connected via a passive in-building coax distribution system.

At the user or client ends of the system a very simple modem interface is used to interface to the user's PC via its existing serial, parallel or USB port. In this way no NIC card or network configuration is required in the users

PC. Point-to-Point Protocol (PPP) is carried on RF channels on the in-building coax distribution to a central RF modem.

A protocol converter is provided between this central RF modem and the shared DOCSIS-compliant Cable Modem. This converter translates the data format between the Point-to-Point Protocol used by the PC and the 10baseT used by the DOCSIS Cable modem. Thus any IP protocol, such as TCP/IP, UDP/IP, etc., is carried transparently to and from the Internet. Special prioritization is available for low-latency requirement traffic, such as IP voice or multimedia, in both directions of transmission.

The protocol converter also acts as a proxy server in order to connect the many coaXmedia client modems and their PCs to one or a few DOCSIS-compliant Cable modems. This involves providing IP addresses to the PCs in response to PPP connection requests. The protocol converter translates single or multiple socket addresses that uniquely identify multiple sessions or windows running within each PC, in order to present unique socket addresses to servers that exist on the IP network.

The field-trial version of the protocol converter is supported by a PC motherboard and is packaged, together with the coaXmedia central modem RF board, in a PC rack-mount, pizza box sized case, for wall mounting. This PC motherboard, upon booting, makes a DHCP request via its Cable modem to a server in the headend and receives a leased IP address – just like a user-PC provided with regular Cable modem service. If the protocol converter has multiple Cable modem connections to the headend then this action is repeated for each Cable modem.

The many client-PC's are be made to appear, from a headend service management perspective, as though they are connected via individual Cable modems. Thus a function is provided in the headend that collects associated user-PC MAC and assigned IP address information from the coaXmedia protocol converter and presents this as an interface to service management functions that also manage single-user Cable Modem services.

RF TRANSMISSION

The in-building RF system presently uses 15 Msymbol/sec BPSK or QPSK modulation in a single downstream "channel" with a center frequency of approximately 970 MHz. Higher symbol rates are planned which could offer at least 30 Mb/s net downstream data capacity.

The downstream signal is transmitted continuously and formatted in a standard MPEG2/DVB structure. The MPEG2 frames comprise a framing (47 hex) / super-framing (inverted 47 hex) byte, 187 information bytes and 16 forward error correcting (FEC) bytes – a total of 204 bytes. Certain reserved MPEG2 "Packet IDentification" (PID) codes are used to indicate that the following information bytes are data of a particular type rather than digital video or idle frames.

Conventional synchronized scrambling is employed for spectral reasons and the 16-byte FEC field is always used or reserved for error correction. These structures facilitate the use of the same industry-standard off-shelf set-top technologies in both data and digital TV applications. Frame interleaving, while available, is not used in in-building passive coax distribution as this would delay latency-sensitive traffic and is not necessary for error protection purposes.

Upstream transmission in the in-building coax uses a BPSK modulated 915 MHz RF signal carrying a 15 Mb/s digital stream. Upstream transmission is only permitted from one coaXmedia client modem at a time as specified by downstream “polling” contained in the downstream data control envelope. Thus there is no collision of upstream signals. The upstream signal comprises a preamble signal that is ramped up in level followed by a sync byte. A scrambled client modem source address, a length field and then data follow this preamble. The length of the data field is dependant on how much is requested by the central modem or the remaining amount of upstream data buffered in the client modem. As in the downstream direction, special provision is made for the needs of low-latency traffic.

COAX PATH LOSS COMPENSATION

Path losses between each client modem and the central modem will have a wide variation due to the coax distribution topology and loading variations. The system is designed to accept losses of 40 dB or more.

Loss variations in the downstream direction are compensated by an automatic gain control (AGC) function contained in each coaXmedia client modem receiver.

The upstream AGC method involves adjusting each of the client modem transmitters such that their signals, upon arrival at the upstream receiver in the central modem, are approximately equal.

Each time a data burst is sent to a client modem an extra bit is included which indicates if the previous transmitted burst from that modem was above or below the ideal level required at the central modem receiver. This bit is used by the client modem to slightly adjust, either upward or downward, the level of its next transmitted burst. Thus all signals received by the central modem from every client modem become aligned in level and cycle upward and downward by a small amount. This is an ideal situation since the upstream BPSK receiver has a much wider acceptable input signal range than the small level variations received. Control systems of this type are fast to react to changes in transmission path attenuation and are intrinsically stable.

PRIVACY

A minimal cost moderate level of data privacy is provided using individual spectral scrambling sequences and/or sequence start points for each client modem in each direction. The method of establishing such scrambling sequences is itself secure. Higher levels of encryption security, like those used in DOCSIS-compliant Cable modems, will be made available, where required, at a slightly additional cost.

TECHNOLOGIES

The systems about to go into public trial use available low-cost, commercial RF and digital technologies. The coaXmedia client modem receiver may use tuner/demodulator chipsets commonly used in satellite set-top boxes.

Near-term plans include moving most functions into a pair of custom chips; one a small RF analog chip, the other a semi-custom chip containing the digital functions. This technology evolution will result in a coaXmedia modem the size of a small cellular phone that may become part of a coax cord assembly and consume very little power.

The coaXmedia hub or proxy server is presently constructed using a normally rack-mounted diskless, low cost, PC motherboard equipped with an RF/protocol board and one or more 10baseT NIC interfaces. This may be mounted, together with one or more off-shelf cable modems, on a wall adjacent to the existing building TV distribution amplifier.

INSTALLATION

As illustrated in Figure 1, the central installation requires only the addition of two coax splitters to which are attached a conventional cable modem and the coaXmedia hub. The coaXmedia client modems are simply introduced, *by the end-user*, between the TV wall receptacle and TV set (if any). An associated transformer cube is then plugged into a convenient power receptacle and the data cord plugged into the user's PC. No network-stack configuration of the PC is required, thus offering a real plug-and-play high-speed Internet access service.

SUMMARY

The coaXmedia system presents a new, economic approach for MDU or hotel high-speed Internet access that works well over existing in-building coax.

This coaXmedia system is DOCSIS-compliant as seen from the headend networking elements, consistent with existing Cable Modem operation and service practices and yet offers plug and play end-user attachment without PC reconfiguration or installation of an Ethernet NIC card. The per-MDU common equipment installation is extremely simple and there is no need for a truck-roll or appointment to provide service to each customer. Indeed, coaXmedia modems can be mailed and are easier to hook-up than a VCR.

The approach isolates internal MDU ingress interference from the main HFC network and provides improved bandwidth management and efficiency, particularly in the upstream or return direction.

Multi-megabit Internet access is achieved via the PC's existing parallel or USB port using a simple "enabler" that places a connection icon on its desktop and activates the PC's existing PPP direct connection facility. The "enabler" can be loaded from the coaXmedia central hub via the PC's existing serial connector -- no floppy disks or CDs.

The coaXmedia system approach is, and will track at, significantly less than half of the cost of a conventional Cable modem approach. Additionally, the coaXmedia user interface consumes less than one tenth of the power of that of a Cable Modem.

Marketing of the service by the Cable MSO is simplified as whole-MDU installation may be offered on a same-day trial basis.

BIOGRAPHY

Jack Terry was with Nortel and its subsidiary, BNR from 1974 until 1999. He was responsible for technical development of Nortel's DMS-100 series of telephony switches that resulted in the universal deployment of digital switching in USA and Canada during the early 1980s. Jack joined BNR following 15 years engineering experience with Marconi Communications Systems R&D in U.K. where he was engaged in the fields of high-power TV transmitters, Video equipment and Digital Telephony PCM Transmission and Switching systems.

As a Nortel technology VP during the 1990's, Mr. Terry provided industry leadership in the field of digital architecture and services in Cable-TV networks that helped seed today's direction of Internet access (DOCSIS) Cable Modems. In April of 1999 Jack and a small team founded coaXmedia, Inc. that today is operating as a small R&D company specializing in exploitation of TV coax wiring for high-speed Internet access.

During his career Jack has filed numerous patents. He is a Fellow of the IEEE and in 1995 received the IEEE Engineering Leadership award for his pioneering contributions to digital switching and line circuit technology. Jack is also a Fellow of the International Engineering Consortium, a Senior Member of The Society of Cable Telecommunications Engineers and a member of Sigma Xi. Mr. Terry's previous leadership activities within IEEE included Chair, IEEE Communications Society / Society of Cable Telecommunications Engineers joint Technical Committee on Cable Telecommunication Systems and Applications from 1995 to 1998.

Advanced Optical and Digital Architectures for Video-On-Demand

John Trail and Dawn Emms
Harmonic Inc.

Abstract

In the future an increasing amount of video content will be provided to the cable subscriber as an on-demand service rather than a broadcast service. This advanced service can be expected to become a very significant portion of the revenue stream for cable operators.

Historically the high costs of video-on-demand (VOD) have discouraged significant deployment. The falling costs of computing power and digital hardware, combined with innovative hybrid fiber-coax (HFC) transport techniques such as dense wavelength division multiplexing (DWDM), are now making the business case much more appealing.

In this paper we will review how one can centralize the video server hardware in the headend and use DWDM to transport the digital video streams in 'on channel' QAM256 format out to the hubs. Once at the hubs, the VOD channels can be easily combined with the broadcast content and then routed out to the subscriber. The paper will review the technical issues, performance, and approximate costs of a 16 and 32 wavelength DWDM system for VOD transport.

We will also review the latest technical developments in the area of digital stream manipulation required for successful, low-cost VOD deployment: PID filtering and program assignment, DVB/Simulcrypt scrambling, conditional access provisioning and QAM modulation. Particular emphasis will be placed on how this is implemented in an open-standard configuration, taking into account OpenCAS and other standards-based initiatives.

Several such systems are in deployment in trials worldwide. In the last section of the paper we will review the deployment currently underway by Telewest in Britain.

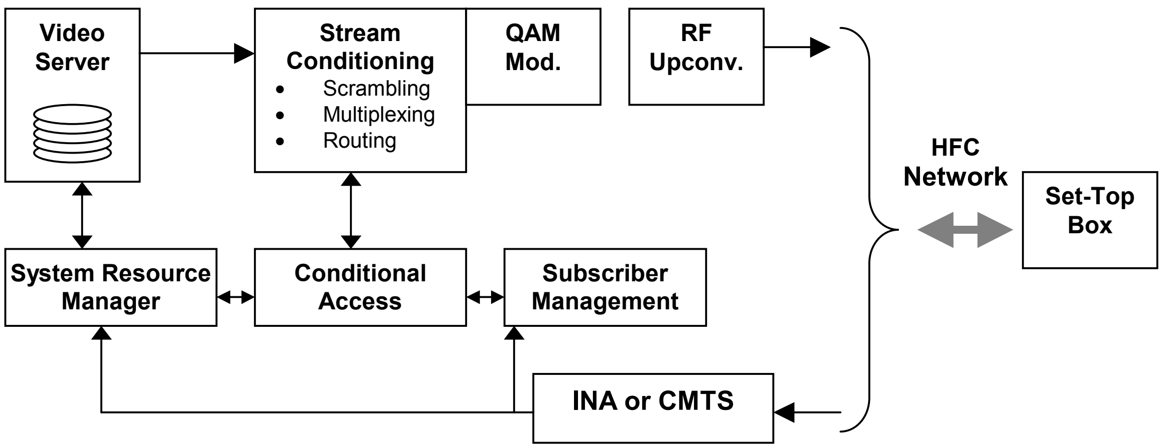
1. A VOD System: The Building Blocks

Before delving into the detail lets first review the overall structure of a VOD deployment. A VOD deployment over an HFC network consists of several functional blocks. Figure 1 shows the functional blocks in a graphical fashion.

From the perspective of the operator it is naturally desirable to implement these functions in the lowest cost and most reliable fashion. One key aspect of a cost-effective system is the use of open standard interfaces between the functional blocks. In this way it is possible to implement best-of-class components, in particular such items as the servers, the interactive network adapter (INA), and the conditional access.

Also, by grouping certain functions onto a single circuit board it is possible to dramatically reduce cost and size. In particular an attractive option is to group the scrambling, multiplexing, and QAM

Figure 1: The building blocks for a VOD system.



modulation together in a single device. This is illustrated in figure 1 by the combined box titled 'stream conditioning' and the smaller adjacent box for QAM modulation. The stream conditioning is covered in section 8.

The transport function is covered in section 4, 'Server Location and Stream Transportation Options'.

2. Cost Goals and the Financial Model

VOD is a service with great strategic competitive advantage for HFC operators. Custom video requires a large amount of narrowcast bandwidth capacity – bandwidth that HFC networks can provide significantly more cost-effectively than competing technologies such as direct broadcast satellite, or twisted pair telephone line. It is also a service with well-known demand – at least to the extent that one can extrapolate from consumer behavior at the local video rental store. It is no surprise that VOD is high on the list of new services that the leading US cable operators are looking to deploy.

However, for VOD to have wide deployment it must provide a return on investment that is compelling when compared to alternative investment options available to the HFC system operator. The key parameter to determine in any discussion of VOD architectures is the delivery cost per stream of video. The 1998 paper "Video-On-Demand, The SeaChange Business Model" [1] outlines a useful financial model for VOD delivery. In that model the target price for the various building blocks includes \$350 per stream for the server, \$300 per stream for the combined functionality of scrambling, multiplexing, QAM modulation and RF upconversion, and \$200 per stream for the transport. Given these costs the Seachange business model shows a return on investment time of less than 18 months.

As we will see in the following sections of the paper the use of open standard conditioning combined with DWDM

---

[1] "Video-On-Demand, The SeaChange Business Model", 1998, Yvette Gordon, SeaChange International, www.schange.com

transport enables these cost targets to be met, and in fact exceeded.

3. Server Location and Stream Transportation Options

The HFC network operator has two options for server location – the hub or the headend. . Although both options are valid, and both are used by various operators around the world, we believe that there are several advantages to centralizing the equipment in the headend. The three main options for location of the VOD equipment are shown in figure 2.

Option 1:

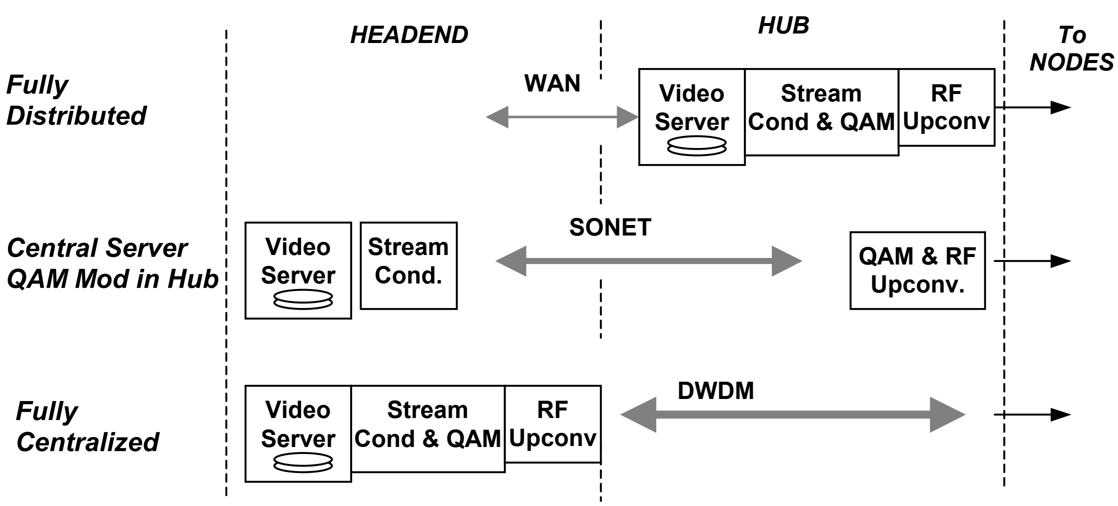
Locate the servers and conditioning equipment in the hub. Use a low or moderate speed Wide Area Network (WAN) to update the servers with the latest content. This minimizes transport costs but has the disadvantage of requiring significant hub space and also requires at least two servers per hub for redundancy.

Option 2:

Locate the servers in the headend, transport the video streams from the headend to the hub using a high quality of service (QOS) baseband digital transport such as SONET or SDH, and place the QAM modulation and RF upconversion in the hub. By centralizing the servers one can minimize the server capacity required – from both the aspect of coping with peak demand rates, and from the aspect of providing a fully redundant system.

Option 3: Locate the servers, QAM modulation, and RF upconversion functions in the headend and move the video streams to the hub, or through the hub, via DWDM transport. In addition to maximizing server efficiency by centralizing the server, this approach also transports the video in set-top ready format from the headend and therefore requires minimal or no processing at the hub.

Figure 2: Options for Server Location and Transport



4. DWDM Transport: Technical Issues

DWDM provides cost effective, high capacity transport in QAM format from a headend to a hub with minimal fiber use. The general architecture is shown in Figure 3 below. Figure 3 illustrates the structure in a ring architecture where one drops off a couple of wavelengths at a hub location and sends the remaining wavelength on the subsequent hubs on the same single fiber. It is straightforward to modify this for a star headend-hub architecture as the optical drop now becomes a terminal optical demux.

The diagram shows 10 QAM modulated channels upconverted and RF combined into a block of channels which then drive one of the ITU grid DWDM transmitters. For consistency throughout the paper the diagram shows 10 RF channels allocated for VOD service.

Once at the hub there are two main techniques for continuing the transport to the node.

1) Receive the optical DWDM signal at the hub and electrically combine with the broadcast signal at the hub. This approach could be used if the operator has an existing 1310 distribution network out of the hub.

2) Unbundle the desired DWDM wavelengths at the hub and optically combine the signal with the distribution network. This arrangement is shown in Figure 5. This architecture is very similar to the general DWDM architecture widely deployed by AT&T BIS for their advanced internet and telephony service.

DWDM has been extensively tested and deployed for the transport of QAM modulated signals. 8 wavelength systems have seen significant recent use in the US and now in Europe. Crosstalk can be kept to a minimum by keeping fiber launch powers at a moderate level plus proper attention to transmitter and filter properties.

16 wavelength systems are also available for deployment. We have tested a 16 wavelength DWDM system over a 100 Km fiber link carrying QAM256. The system is shown in Figure 6. With received power per wavelength of greater than –10 dBm we observed SNR of greater than 40 dB and BER on the QAM256 of better than 10^{-8} without forward error correction.

Figure 3: DWDM transport architecture

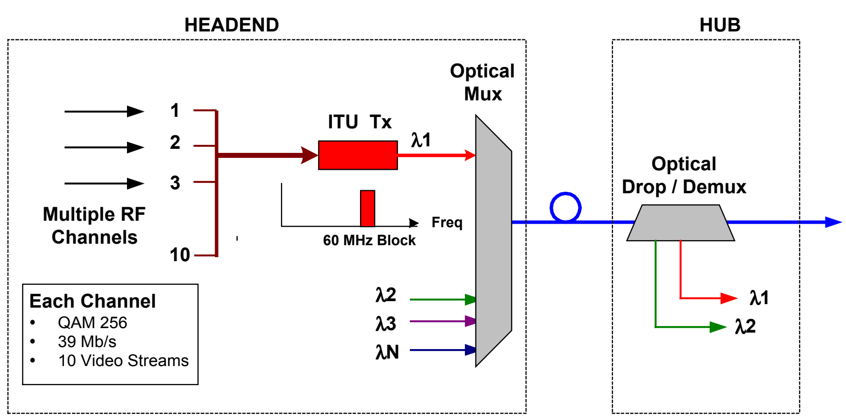


Figure 4 Drop at the hub and RF combine with the broadcast signal

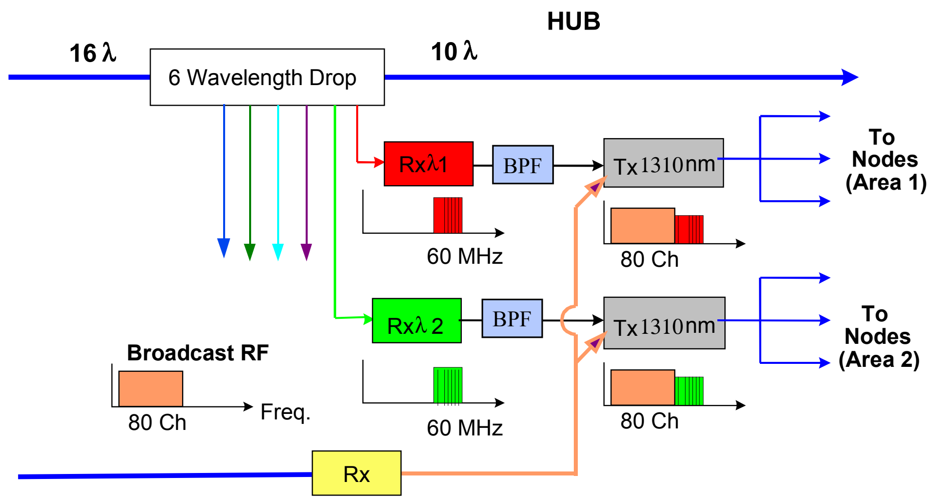


Figure 5 Drop at the hub and optically combine with the broadcast signal

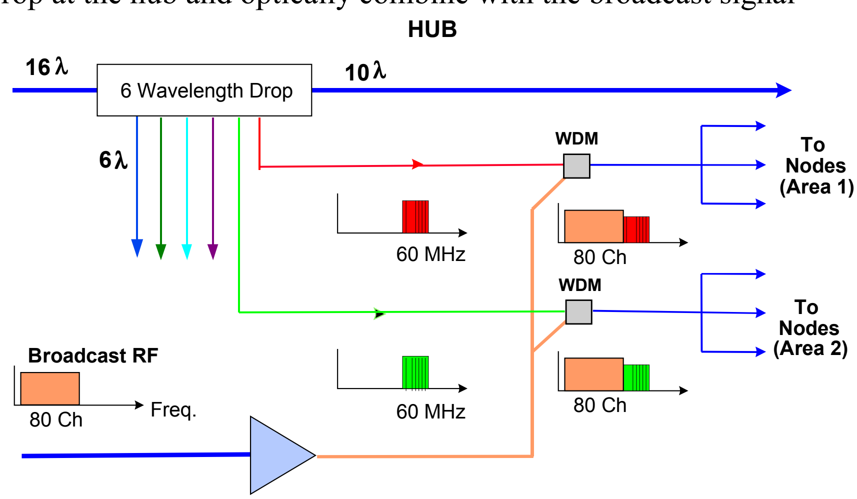
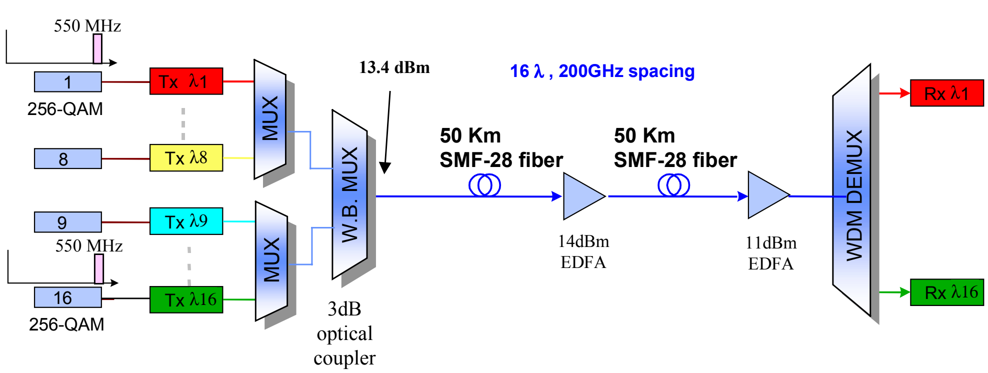


Figure 6 Experimental set up for 16 Wavelength DWDM test over 100 Km with QAM256



5. <u>DWDM Transport: Costs</u>

Optical component costs are falling rapidly in this competitive and rising volume world of optical transport. DWDM costs have been falling at faster than 20% per year and we have no reason to expect this trend to change in the near future. The costs given below are typical at the time of writing but for operators considering this approach we recommend that you contact preferred vendors for latest pricing.

The 'per stream' cost of DWDM transport is easily calculated. First determine the cost of a point to point transport for a single wavelength. A DWDM directly modulated transmitter costs around $5000 including platform and power supply. The average cost for the single port of an 8 or 16 wavelength mux or demux is $800, i.e. $1600 for a Mux and demux pair; and receivers are in the range of $1800. Total point-to-point cost per wavelength is therefore ~ $8400.

The transport cost per stream is a very simple function of QAM format and the number of channels allocated per wavelength. For QAM256 format we assume that, given a typical per stream bit rate of 3.6 Mb/s, one can carry 10 streams per RF channel. With 10 channels allocated to VOD each wavelength would therefore carry 100 streams. The cost per stream is then $8400/100 or <u>$84 per stream</u>. Table 1 below shows the scaling in cost per stream for several other likely combinations.

It is worthwhile comparing these costs with a typical SONET or other high quality of service baseband transport –as would be used in option #2 described in section 4. A single OC12 terminal currently costs of the order of $40,000 to $60,000 with the lower prices coming from newer companies in the market. Therefore even in the best case a point to point baseband link with 622 Mb/s capacity will require two terminals and will cost in the range of $80,000. Using the same typical per stream bit rate as above this capacity corresponds to 170 video streams, or $470 per stream. This SONET cost is significantly higher than the equivalent DWDM cost.

6. <u>The Benefits of a Standards-Based VOD Solution</u>

One key aspect of a cost-effective system is the use of open standard interfaces between the functional blocks shown in Figure 1. In this way it is possible to implement the most appropriate product for the specific cable operator, in particular such items as the servers, the interactive network adapter (INA), and the conditional access.

Table 1: DWDM Transport Cost Per Stream vs Channel Allocation and QAM Format (NTSC)

	Number of Channels Allocated for VOD				
DWDM Cost Per Stream*	**2**	**4**	**6**	**8**	**10**
QAM 64 (7 streams per Ch)	$ 600	$ 300	$ 200	$ 150	$ 120
QAM 256 (10 streams per Ch)	$ 420	$ 210	$ 140	$ 105	$ 84

**Assume cost per wavelength of $8400*

Historically most of the early VOD solutions have been closed systems with perhaps some limited choice of video servers. However if we look at the functions outlined in Figure 1 we note that for many of the functions there are several different vendors offering products with different cost-performance characteristics. Due to the disparity of the core technologies, different vendors are best qualified to provide the individual components. Working with partners who are experts in these core-competencies ensures that the VOD operator is taking advantage of the latest technology advances to drive down the cost.

The key providers of these VOD system components are not only committed to providing the optimum technological solution at the lowest cost, they are also committed to working together to provide an open-standard seamless solution to the network operator. These leading vendors are working together both in the field with real deployments (as we shall see in the example below) and on standards bodies to ensure that in future deployments interoperability is maintained. For example, OpenCAS is a working group established to define standards such that vendors can achieve interoperability between conditional access and content collection subsystems (scheduling, encoding, multiplexing, scrambling) to speed system integration and reduce risk. All the leading vendors are key contributors to this forum and at the present time their recommendations for the interface standards have been submitted to the SCTE.

7. Stream Conditioning

By stream conditioning we are referring to the combined functions of scrambling, PID filtering, and multiplexing. In a broadcast environment the number of unique video streams is low therefore the cost of stream conditioning is not excessive. Historically, the functions – multiplexing and routing, scrambling, QAM modulation and RF upconversion for a VOD deployment, have been handled using equipment tailored for a broadcast system. This is inefficient both in cost and space allocation. A system such as this can quickly grow to occupy most of the space in a headend.

Today, equipment has been designed specifically for the stream conditioning of narrowcast services. This technology dynamically selects program services from the high-speed asynchronous serial interface (ASI) from the media server, automatically routing the services to a QAM channel and the associated RF upconverter. Every time the media server adds or drops a service, the associated service information and program information are updated. The content needs to be scrambled prior to transmission. This functionality has been incorporated in the stream conditioning.

Today the cost-per-stream for the stream conditioning plus QAM modulation and RF upconversion is approximately $450. In the very near future the cost-per-stream for this combined functionality will be driven to under $300, meeting the target set in the SeaChange model. This can be accomplished by combining the stream conditioning, QAM modulation, and RF upconversion for several RF channels on to a single circuit board. The resulting cost and space savings will continue to strengthen the business case for VOD.

8. Current Deployments

This VOD architecture we have described above has recently been deployed by

Telewest in Britain and is about to go online at the time of writing in mid March 2000. The Telewest deployment makes use of the open standard stream conditioning approach we show in Figure 1 and also uses the DWDM transport as shown in Figure 4.

Here are some key technical details. The initial deployment covers 337,000 homes passed. Each serving area contains about 8,400 homes passed and there are 5 channels allocated for VOD service in the delivered spectrum. As this is PAL format, each of the five channels is carrying 9 video streams at 3.6 Mb/s, in 8 MHz of spectrum using QAM 64 modulation. Therefore, the system delivers 45 unique VOD streams per 8,400 homes passed serving area. This capacity is sufficient for the present customer and digital box take rates, however it is expected that the capacity will be scaled up as demand rises..

In keeping with an open-standard approach many of the key functions are provided by different vendors: the media servers are provided by both SeaChange and nCUBE; the scrambling, multiplexing, and QAM conversion are provided using the VSG card from Harmonic; conditional access is provided by NagraVision and the set-top boxes from Pace.

NTL, the other major operator in Britain is also deploying a DWDM architecture. For NTL the primary driver was to enable transport of the out-of-band signaling to the specific serving areas, however going forward NTL intend to make use of this narrowcast capability and offer VOD and other interactive services. NTL in some franchise areas have a 1550 to the node architecture and are using the optical combining method shown in Figure 5. It is worth noting that this optical combining is performed in the streetside cabinets rather than at the hubs.

British cable operators have historically faces intense competition from satellite service providers. By offering a significantly differentiable product in the form of this true VOD service, Telewest expect to significantly improve their competitive position, increase their customer take rate, and increase their revenues per subscriber.

9. Summary

Video-On-Demand is service with a key competitive advantage for HFC operators. New technologies with lower costs enable operators to deploy systems today that provide this competitive edge with an immediate return-on-investment. The business case is there.

In this paper we have reviewed some of the new directions that are enabling this business model.

- Combining several key video stream conditioning functions into a single low cost unit
- Using an open-standard approach which enables the operator to choose the latest and most cost-effective products from a competitive environment
- Using DWDM technology to provide low cost and flexible transport of VOD streams to the serving area

Finally, we reviewed the current deployment by Telewest of a VOD system which has incorporated all of these features discussed above. Stay tuned!

John Trail and Dawn Emms are both with Harmonic Inc. www.harmonicinc.com

Advances In Content Management And Protection

Wim Mooij
Mindport

Abstract

With the accelerating adoption of digital formats in multiple application domains, it becomes possible to transfer content from one domain to another. Especially with the emerging home entertainment environment with digital storage devices, content can flow between a large variety of electronic devices. Protecting the interests of all value chain participants is the challenge of a new set of technologies called Content Management and Protection. This includes security solution from multiple application domains. Content Management and Protection is critical to the industry growth in the digital content and information era.

INTRODUCTION

Digital Broadcasting is now supported by a wide range of globally adopted specifications for consumer and professional Television equipment. With the migration of broadcast signals to the digital domain, the traditional distinction between various media applications disappears. After all, who can tell the difference between an audio, video and data bits? With the advent of internet, digital television and digital telephony many networks now carry only digital data streams. It then becomes possible to transport the same bits over new distribution infrastructures creating new potential value propositions for network operators, consumers and intellectual property creators. The digital intellectual property in the digital content may represent significant (future) commercial value to participants in the value chain. Content management and protection system aim to enforce the copyrights and/or other intellectual property contained in the content. Such systems need to operate in a rapidly changing environment, and no longer can ignore the move of broadcasting into the "e-Environment".

COMMERCIAL ANALYSIS

Content in digital form can be transported over a wide variety of distribution and communication media. This facility also is available to the consumer. The same content thus may appear in a variety of infrastructures and consumer equipment. This great flexibility allows greater freedom, but with that freedom also comes a vast increase in unintended commercial side effects. Especially the ability to make perfect copies of digital content in any of the domains and the capability to introduce such copies into another player infrastructure completely changes the underlying assumptions of many distribution regimes. Often such implications are only discovered after the fact. It thus is necessary in many cases to re-visit the complete application and determine the basic commercial rules that make it a viable business proposition to all value chain participants. With these basic rules established, it becomes possible to devise a suitable content management and protection regime.

A key observation is that any content management and protection regime takes away some freedom of use from the consumer. In many cases this freedom would allow the consumer to seriously violate the commercial rules driving the application. The capability of copying and distributing amongst friends without some form of compensation to the intellectual property creators is one such limitation. In all cases the chosen content management and protection system should consider the capabilities of equipment in other domains. So, it is a fallacy to assume that copying is not possible because all copying devices in one application domain, have some build in copy restriction mechanism. If such a mechanism is not implemented in another application domain, it may be possible to make unrestricted copies.

Content in the digital domain may have a substantial value associated with it. The easy copying and distribution of digital information requires that the content itself is protected in a persistent manner. Copies of protected content then are not a problem, as the intellectual property is safe in the protected form. The consumer can play the protected content on any rendering device as long as that device is capable of dealing with the protection layer for the protected content. As the rendering quality of content continuously improves, it becomes increasingly relevant to extend the content protection into the “analog” domain. This aims to prevent the removal of the protection layer through a re-encoding process.

In some CMP systems a removable element such as a smartcard may be used to implement some or all functions of the content management and protection (CMP) system. Other CMP systems utilize periodic on-line communications to implement some protection system functions. Although implementations of CMP systems vary widely, they all share the basic function to enforce the business rules associated with the use of the content.

In an e-Environment where all content is digital and all networks are digital, content can flow very easily between networks and devices. It is unlikely that a single CMP system will be able to deal with all existing and future applications. Hence, CMP systems have to find a way to work with other CMP systems in order to facilitate this natural flow of content.

The main challenge is to make this protection framework easy for the most frequent types of consumer usage of the content. It is a widely held belief that consumers only will accept content protection if this is realized in an interoperable fashion and permitting horizontal market structures. Consumer interests are protected by commercial interests of the industries involved and by possible regulatory involvement. As the need for content protection increases due to a rapidly growing abuse of intellectual property in content, a substantial effort is required to create and implement an open content management and protection system infrastructure. Several standards setting initiatives are now dealing with (aspects of) the content management and protection issues facing a wide range of application platforms.

TECHNICAL ANALYSIS

A content management and protection system architecture operates in an environment where multiple content players are interconnected through an in-home distribution network. An example diagram of such a network is shown in Figure 1.

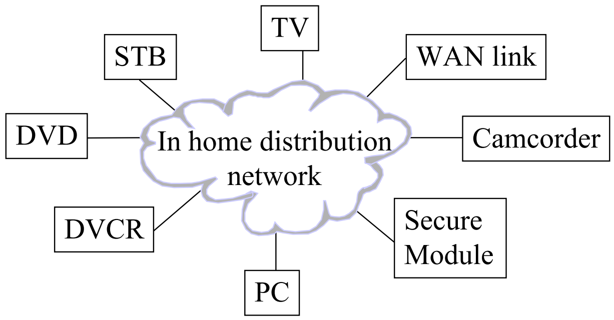


Figure 1 Diagram of the in-home distribution network.

It is desirable in such an environment that content can be shared between content players. During content move and/or copy operations, the material should remain protected. The enforcement of the business rules associated with the content is the task of the CMP System.

APPLICATIONS

There is a continuously expanding range of applications that distribute content. Most of these applications are under control of some form of CMP System. Several distribution infrastructures can be identified.

- Broadband (broadcast)
- Internet
- Wireless networks
- Pre-packaged media

In all of these networks a rich variety of media types can be identified such as real time stock price information, premium movies, archived newspaper articles. In all cases the current and future value of the content differ substantially. Hence the CMP System features also have corresponding differences. Similar differences exist in the payment structure. In the broadcast domain the following different payment structures exist.

- Free TV
- Subscription TV
- Pay per use
- Near Video on Demand
- Video on Demand
- Services that add value to those mentioned above, e.g. paid-for Electronic Program Guides and other such services

Similar differences in service types and payment structures apply to other content delivery structures, further increasing the range of options that CMP Systems need to support.

The threat model for each commercial application also varies significantly. Early releases of premium movies have an entirely different value compared to highly volatile stock price information. Hence the required security features of the CMP System can differ some orders of magnitude. Thus, it is not surprising that all attempts at defining a single CMP System for a reasonable range of applications is impossible.

INTEROPERABILITY

There is a world of proprietary application domains that each have their own governance structures defined by a set of CMP systems. An example of a such a set of CMP Systems is traditional Conditional Access Systems for broadcasting applications. Interoperability among such CMP systems creates a structure where operators and consumers benefit from new commerce structures.

An early attempts at interoperability of CMP Systems is the SimulCrypt concept developed in the European Digital Video Broadcasting project. SimulCrypt is a mechanism where multiple business rule sets are linked to the same broadcast material. Although this works within a broadcast environment, it does not facilitate interoperability with other digital content application domains.

A multiple application domain interoperability between CMP Systems can be based on the specifications developed in the Open Platform Initiative for Multimedia Access (OPIMA). OPIMA provides tools to exchange a set of authenticated identifiers, called OPIMA Credentials, to enable Protected Content to flow within and between application domains. When OPIMA peers have an on-line connection, the OPIMA Credentials can be exchanged prior to the delivery of the Protected Content. OPIMA Credentials and CMP Systems can be combined with the Protected Content when the delivery infrastructure does not support on-line connections (e.g. storage media, broadcast media). The OPIMA Credentials contain necessary information that may be used to enable Protected Content flow between application domains. For a given instance of Protected Content that is intended for consumption in two application domains, a Broadcast conditional access domain and an Internet music delivery domain, OPIMA Credentials from both domains are associated with the Protected Content.

OPIMA enables generic interoperability between different applications, devices and CMP systems belonging to different *compartments*. A compartment is a class of OPIMA enabled devices that share some common elements in their CMP interfaces and/or architectural components. For example, digital TV broadcasting can be considered as a compartment, which in turn contains other compartments defined by specific CMP system. Content does not necessarily flow between all compartments, however, OPIMA provides a schema for facilitating content flow between compartments.

The OPIMA architecture is peer-to-peer. The core OPIMA element is a peer called the *OPIMA Virtual Machine* (OVM). OPIMA provides protocols and infrastructure components that enable secure (trusted) inter-operation amongst these elements.

This section describes an in terms of. Figure 2 depicts the functional units of the OPIMA Peer.

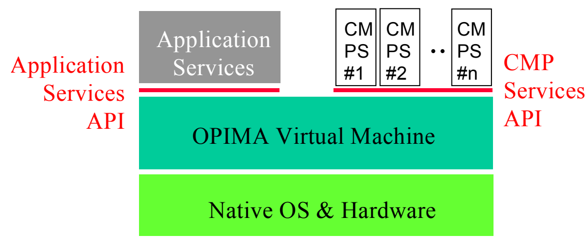


Figure 2 Diagram of the OPIMA peer

The OPIMA peer contains a group of basic functional elements that implement the backbone of trust. This is called the OVM. The basic functionality of the OVM allows for application-specific extensions. The OVM is responsible for establishing authenticated, secure channels amongst OPIMA compliant devices. The OPIMA Application Services API allows services to communicate with the OVM and the CMP system installed in the OVM. CMP Systems can access the OVM functions through the CMP Services API.

The OPIMA specification alone is not sufficient to achieve horizontal markets. This requires further specifications to be defined in a particular application domain. OPIMA solves the problem of permitting content to flow across domain boundaries and thus solves a major problem in the digital content environment.

RENEWABILITY

One of the fundamental assumptions for any CMP System is that eventually someone will compromise the system. This follows directly from the observation that it is possible to copy any system implementation given sufficient resources (time, money and skills). If the eventual breach of a CMP System is an integral part of its design and implementation, it will handle such situations more efficiently. In most cases, the CMP System has some flexible response strategy to deal with such security breach eventualities. This strategy allows the renewal of part(s) of the CMP System with minimal impact to operations and the consumer. This section analyses various implementation options for CMP Systems and describes how these impact the renewability of the system.

The general structure of a Content Management and Protection (CMP) System is assumed to use classic cryptography. Other techniques such as watermarking possess similar security threats and implementation principles. For the sake of simplicity this analysis only deals with the case of cryptographic technology CMP Systems. The key components of such a system are shown in the following diagram.

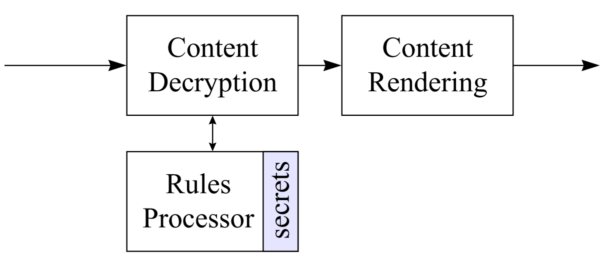


Figure 3 Diagram of a CMP System

Before the content can be rendered (decompressed, D/A converted etc.), the CMP System needs to remove the protective encryption layer around the content. This is based upon some decryption function. The keys to the decryption process are provided by the rules processing engine, which decodes the business rules associated with the content. These business rules can be enforced by the rules processing engine as it uses secrets to decrypt the keys that will unlock the content. Obviously, these secrets need to be protected using tamper resistance techniques. The key issue is the implementation of the CMP System. There are three main trends:

1. Integrate all three elements onto a single silicon device.
2. Integrate content decryption and content rendering onto a single device and use a separate rules processor.
3. As option 1, but make provision for a separate rules processor

The arguments in favor of the first implementation is that this makes it the most difficult (and expensive) for an attacker to find the relevant secrets to compromise the player infrastructure. And even when an attacker manages to find the secrets, it is not possible to introduce a breach into the system without breaking into the integrated device at the chip level. Oddly, the complexity of distribution of a breach introduces a weakness in this implementation. As soon as the secret has been obtained, it becomes necessary to build a modified player that bypasses all security features to distribute the security breach As technology progresses at such high speed, such solutions become increasingly feasible over time. It is thus inevitable that such devices will become available. The pricing of illegal devices can be low as no licenses are paid. The main attraction is that these illegal devices play legitimate content. Thus, the CMP system implementation forces the security breach to be distributed using a alternate (illegal) device infrastructure. As these new devices are outside the scope of the legitimate content distributors, only legal counter measures are possible. Note that counter measures such as software downloads and revocation don't work when the security system is completely known. Market dynamics at some point will cause the illegal players to grow very rapidly, encouraging manufacturers to also cut corners where security measures are concerned. Over time, the complete device infrastructure is lost to legitimate content distributors. Note that this process accelerates where software receivers are possible.

The second CMP system implementation method offers a slightly lower cost to the attacker as only the replaceable rules processor needs to be reverse broken into. And even worse, the implementation provides an easy way for a security breach to be distributed to the legitimate player infrastructure. It would seem that this is a worse situation than the first implementation. Again, appearances deceive. The distributor of legitimate content now can upgrade the rules processor, driving out the security breached version as this cannot process new content. This will quickly reduce the level of illegal rules processors. The attackers then are forced to break the legitimate rules processor and the cycle repeats itself. Obviously, there is a cost involved in the replacement of rules processors. The important effect of this implementation is that the infrastructure of players always remains legitimate. Only the rules processor can become illegal from time to time, but also can be recovered as a legal content player. This way the security breaches can be contained and controlled without losing the entire device infrastructure.

The third implementation form is a combination of the first two. It pairs a slightly higher burden for the first attack and still allows the infrastructure to recover from a security breach, if implemented through the rules processor interface. If the security breach is distributed through the method described in the first implementation method, the recovery through a rules processor upgrade no longer is possible. This depends upon the discovery of existence of this interface by the attackers.

CONCLUSION

This article describes the challenges placed on CMP Systems in the converging world of digital content and networks. It describes the environment were perfect copies are just one mouse click or a remote control button push away. Supporting the consumer reasonable demand for novel uses of the new network and content possibilities, places strong demands on CMP Systems. They now also need to consider how they can work with competing systems in other application domains. They need to consider how they can deal with security infringements. CMP Systems also need to be flexible to support new business options and new commerce structures. The article describes these problems and shows some recent advances in the CMP Systems that address these important issues for a bright digital content future.

REFERENCES

1. ISO/IEC 13818 part 1, MPEG-2 Systems.
2. IPMP FAQ, MPEG-4 IPMP
3. http://www.cselt.it/ufv/leonardo/opima
4. http://www.dvb.org

ACKNOWLEDGEMENTS

The author would like to thank his colleagues for their contributions to this work. Many of the described concepts are from discussions in standardization activities. This work is supported by beautiful girls in Amsterdam.

Bandwidth Monitoring Parameters for Capacity Management

Dennis Cleary
High Speed Access Corp.

ABSTRACT

Operators need to effectively plan and manage capacity when designing and building a broadband network. These networks need to be capable of handling services available today. They must have the flexibility to grow and meet future demands. The success rate of a cable operator's system, from a service and an economic stand point can be directly effected by circuit capacity management. An excess of bandwidth can be economically inefficient when on the other hand, not enough bandwidth will cause delays resulting in possible lost revenue. We will discuss bandwidth monitoring parameters to assist the cable operator with a general understanding of Bandwidth issues for their systems. The basic parameters follow the 70/30 rule. The following is the equivalent of a brief white paper with some of the issues that impact the 70/30 rule.

DEFINITIONS

The 70/30 Rule: This means that at 70% and above, contention occurs, and multiple users are vying for access. This 70 % and higher would generally be considered peak periods. When peak utilization spikes routinely exceed 70%, additional bandwidth is required to avoid contention-based delays. Conversely, when capacity is in excess of 30% for average utilization (that is, during normal use, and continuous utilization runs 30% or more), bandwidth allocation needs reviewing.

Additional bandwidth is required to avoid contention-based delays, where average utilization exceeds 30%, and peak utilization hits 70% or more, consistently. The degree of accuracy with respect to these capacity percentages is directly related to the timeframe of the monitor (i.e., it is more accurate to monitor every 15 seconds than to monitor every hour). The monitor process, itself, uses the network. Thus, it is important to consider this when establishing these monitoring timeframes. The more frequent the monitor, the more the network is utilized.

Capacity: For the purposes of this discussion the definition of "capacity" is the information carrying ability of a telecommunications circuit.

Bandwidth: Bandwidth, in a broad sense is directly proportional to the amount of data transmitted or received, per unit of time. In a qualitative sense, bandwidth is proportional to the complexity of the data for a given level of performance. Obviously, it requires more bandwidth to download a graph or picture than to download a page of text in the same unit of time. Computer programs, audio files, and animated videos require the use of more bandwidth in the same unit of time. The consumer is still pushing hard for the delivery, of streaming media, today. Oh, by the way, let's not forget gaming with its utilization requirements.

All this looks good on paper, but how do we apply these theories, manage them, and survive the high cost of bandwidth today? A good place to start, is the 70/30 rule, combined with many other elements that make up the capacity management formula. We will touch on other supporting tools such as:

Capacity Management Tools
Caching Technologies
Capacity Planning Model
Circuit Monitoring Software
Organization Issues

With respect to bandwidth, it is important to understand that the traditional mode of operation for many companies has been less than "managerial". Basically, when contingency occurred, a circuit was ordered. This put the fire out, but did it solve the problem? More often than not, this solution did nothing more than create more billing, and an over abundance of circuits.

The idea with this discussion is to increase management's awareness in order to utilize bandwidth in a network more efficiently and economically. By doing so, companies will order circuits less often and more effectively utilize the existing bandwidth for a longer period of time Done correctly, this will result in improved customer service and additional savings.

The point is, the cost of circuits may make or break the core business plan and may, quite possibly, exceed any other company expense.

<u>CAPACITY MANAGEMENT TOOLS</u>

Some of the software tools available today that will enhance capacity management capabilities include (but are not be limited to) products such as:

Transcend
Optivity
Centillion
Spectrum
Cascade View
Cisco Works
HP Open View
Net View
NetManager

And let's not exclude the ability today for companies to completely outsource their network monitoring and bandwidth management needs to companies owning very sophisticated Network Operations Centers ("NOCs").

A recent article in the Telephony Magazine[1] states "About 95% of the U.S. population has cable TV, proving that providers have built confidence and found a place in the homes of consumers".

For successful capacity management, a close relationship must be in place with the Marketing Department. One marketing campaign could easily affect the network. When sales increase at a particular location, in a short period of time, the network can be drastically affected. Tight coordination with Marketing allows the Capacity Management Department to ensure that there will be enough bandwidth to accommodate the projected customer acquisition rates.

With that in mind, demographics should certainly be one element to consider in a capacity management formula. Other elements to consider include:

Is the system a 2 Way System or a 1 Way Cable Modem System?

Does the system support Dial-Up customers?

What are the total subscriber counts, both for the commercial and residential markets.

How many homes passed?

What marketing campaigns are on the horizon.

A review of the sales forecasts.

Planning for the long time lines for installation of the LEC and back bone provider circuits (i.e., DS-1s or DS-3s).

Is varied bandwidth being offered (ranging from 128 kbps to 512kbps).

Is VSAT an option in suburban areas?

The cable operator might consider several philosophies and theories, when applying capacity forecast tools. Some individuals view the process as a science and others view this process as an art. In either case I am not convinced that a complicated and detailed formula for bandwidth capacity is the answer.

Generally speaking, historical records of bandwidth utilization also deserve some study. This research could produce ratios of circuits vs. subscribers. Certainly not fool proof, ratios are commonly used to help keep networks in check.

For example, history might show that in a dial-up network, a ratio of 3:1 would exist, providing services to 72 customers per DS-1 on the local side. On the back bone side, the ratio might be 200:1 or 300:1, which would support 200 or 300 customers per DS-1.

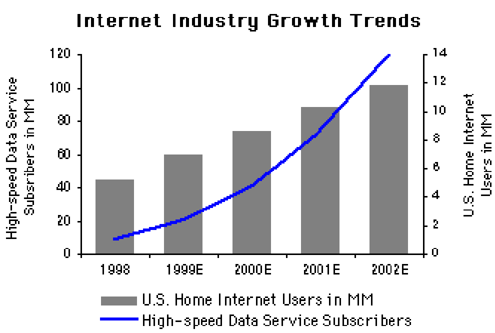
CACHING TECHNOLOGIES

Should local caching be considered as a bandwidth capacity management tool. As mentioned earlier, streaming media files are very large and bandwidth intensive. The distance between the provider and the customer is usually substantial. As more users request streaming content, it requires additional bandwidth. Multiple users viewing the same program, in the same system without caching will consume the bandwidth required on an individual basis. For instance, if twenty different users in the same system want to view the same movie, then multiply the bandwidth required by twenty.

By using caching technologies the Cable Operator can provide high quality streaming video, and increase revenue, without wasting bandwidth. The bandwidth saving will not only effect the customers viewing the streaming video but deliver a higher quality service to other subscribers on line.

Streaming video is only one example and one should consider the effects of caching for their entire system. We estimate an across the board 20 to 30 percent savings in bandwidth utilization with caching technologies.

Bandwidth management is, without a doubt, a very important issue with Cable Operators delivering Internet services. The Futurist Magazine estimates that "more than 70 percent of U.S. adults will have Internet access by the fall of this year."[2]



Broadband Access Management www.dyband.net

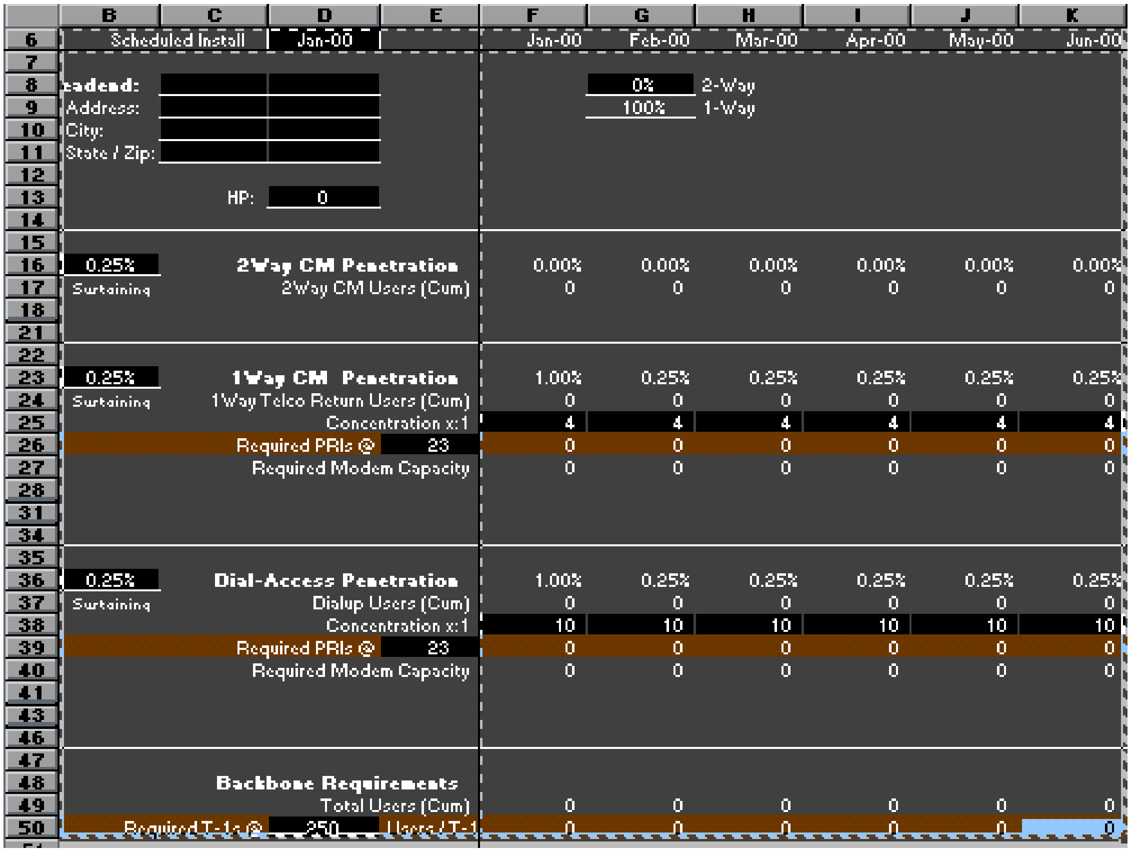
CAPACITY PLANNING MODEL

We, at High Speed Access Corporation, are continuously evaluating all processes and procedures involved in capacity management, from the onset of receiving an Authorization To Proceed (ATP), to the Site Survey through the Alpha test, into the Beta test, and finally to the ongoing monitoring of a site (once in full service).

Today, when our Sr. Project Managers receives the Authorization to Proceed (ATP) from a system, a very detailed Site Survey is completed.

Based on our capacity planning model, circuits are ordered. The formula for the circuits is written into the capacity planning model. In a cable modem application, the policy has been to apply a suitable ratio dictated by the capacity planning model. In a system where dial-up is also present, a different ratio might be used than in a system where 2-way is the only service. Although the model for a 2-way system shows that in the month that the headend is installed, we are at a 1% of homes passed. And each month thereafter, we increase the penetration by .5, we actually order whatever the model calls for in the first month that the headend is scheduled for installation and, then, in three months, we schedule for the remainder of what the model calls for (by the fourth month). For example: if the model calls for 1 DS-1 to be installed in January for the headend installation, we order one DS-1 at installation. Then, the model shows an additional DS-1 each month (Feb, Mar, and Apr) for a total of 4 DS-1s by the April date for revenue launch.

In this scenario, We would typically order 2 of the DS-1s for installation in January and an additional 2 DS-1s to be installed March 31, for the April total of 4, which the model calls for.



	B	C	D	E	F	G	H	I	J	K
6	Scheduled Install		Jan-00		Jan-00	Feb-00	Mar-00	Apr-00	May-00	Jun-00
7										
8	eadend:					0%	2-Way			
9	Address:					100%	1-Way			
10	City:									
11	State / Zip:									
12										
13		HP:	0							
14										
15										
16	0.25%		2Way CM Penetration		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
17	Sustaining		2Way CM Users (Cum)		0	0	0	0	0	0
18										
21										
22										
23	0.25%		1Way CM Penetration		1.00%	0.25%	0.25%	0.25%	0.25%	0.25%
24	Sustaining		1Way Telco Return Users (Cum)		0	0	0	0	0	0
25			Concentration x:1		4	4	4	4	4	4
26			Required PRIs @	23	0	0	0	0	0	0
27			Required Modem Capacity		0	0	0	0	0	0
28										
31										
34										
35										
36	0.25%		Dial-Access Penetration		1.00%	0.25%	0.25%	0.25%	0.25%	0.25%
37	Sustaining		Dialup Users (Cum)		0	0	0	0	0	0
38			Concentration x:1		10	10	10	10	10	10
39			Required PRIs @	23	0	0	0	0	0	0
40			Required Modem Capacity		0	0	0	0	0	0
41										
43										
46										
47										
48			Backbone Requirements							
49			Total Users (Cum)		0	0	0	0	0	0
50		Required T-1s @	250	Users / T-1	0	0	0	0	0	0

Sample Capacity Planning Model View

This practice seems to vary, depending on the number of prospects in the Operator's prospect database. If Project Managers see that there are 2000 potential customers ready to sign-up; they would escalate that order to accommodate for a more robust market. If there were only 200 potential customers, they would not order as many DS-1s as with the first example.

BROADBAND ACCESS SERVICES

According to Broadband Access Services, data traffic has been doubling approximately every 100 days, or at an average rate of 1000% for the last three years. With this kind of growth, monitoring your network is an absolute. The future of broadband access services depends on the ability to manage the distribution of bandwidth, dynamically. Without Bandwidth monitoring, heavy users can monopolize bandwidth at the expense of other users, resulting with dissatisfied customers. Here again, monitoring your network is essential. Bandwidth management is important for another reason, as well. Security. Unusual traffic patterns could easily point to a suspect and the problem stopped before damage is caused. An example of this, are people who spam using your network, or set up a DNS server—selling their own services—with your IP address.

One of the most important aspects to achieve is to educate your personnel. They need to know how to analyze bandwidth needs before ordering circuits. In many cases, the individuals who ordered circuits were simply placing the orders with little or no analysis or thought. This is the first change towards bandwidth management.

CIRCUIT MONITORING SOFTWARE

Generally speaking, circuit monitoring software may offer a built-in feature for notification. This type of feature can notify specified individuals when circuit capacities exceed pre-determined thresholds. The active circuit inventory monitored by the circuit monitoring software is usually loaded manually. Periodically, an audit of the circuit inventory is necessary to ensure accuracy. It's important to have circuit monitoring software that has the ability to access each site, and accomplish an automatic inventory of circuits and equipment. This will accommodate for changes in the field, as time goes on.

Our circuit monitoring software currently executes every 5 minutes. The results are stored and accumulated in several time frames for historical value, per site. These time intervals are adjustable. The information is blocked. First, for every 5 minutes then every 4 hours, 6 times per day, weekly, monthly and yearly.

These time intervals are set based on our preference. Your preferences may differ. It is important to note, that while monitoring is taking place, you are utilizing your network.

When new locations are added to the system, your procedures should reflect the addition of that location to your monitoring software.

ORGANIZATIONAL ISSUES

A pro-active effort must be taken at all levels within the company. This means from the President, at corporate level, to field personnel. Extremely close attention to the billing by the CFO, with a clear focus on circuit costs, generating questions and commanding answers from subordinates, can begin the process. In addition, more than one department may take part.

For instance, if several departments have access to the capacity management tools and are reviewing this data, this will create an effective check and balance process.

This awareness and focus on capacity management will, no doubt, improve the customer experience and allow you to utilize your networks in a more economically sound manner.

Accurate database accounting for each and every circuit—both carrier, and LEC providers—that indicates every cost per circuit, will help heighten awareness within the organization.

It is important to mention here that, personally, I am unimpressed with the accuracy of the billing from the circuit providers. A word of caution that an internal checks and balance process be implemented to confirm pricing, prior to payment. It's amazing that so many companies today pay the bill they receive without the means to verify bill.

I would go as far as to say that an internal billing audit might be appropriate. (Many third-party companies will audit these bills on a contingency basis. (Often times, with no cash outlay from your organization).

CONCLUSION

Throwing money at a problem such as capacity issues, without proper management, prior to the purchase of any circuits is not a recommended practice.

Under utilized networks are equally as important as over utilized networks. Good bandwidth management will adjust for both seniors.

The future will bring high-speed data, voice and video to our homes and offices. Cable Operators are in an advantageous situation. With even more exciting opportunities on the horizon.

An aggressive approach to provide the services and technologies the consumer wants through high speed access is the key to success.

---

[1] Telephony Magazine, February 21, 2000 Solution For Success: author Chrissy Moch
[2] What's Hot for the 21st Century, February 25, 2000: author Barbara Reinhold

Beyond Moore's Law

Oleh Sniezko and Xiaolin Lu
AT&T Broadband

Abstract

The computer industry has been proudly riding on the exponential curve based on Moore's law for decades. What Moore did not expect, and perhaps nobody has ever dreamt of, is that the photonic technology, especially its application in cable industry, has similar if not more profound impacts on the communication infrastructure and service delivery mechanism.

Since the advent of linear laser that introduced the optical transport technology into coaxial networks, the cable industry has embarked on extensive network upgrades that continually push fiber deeper into the network to evolve the infrastructure from a broadcast-based trunk-and-branch plant to a two-way broadband network with superior quality and reliability to deliver advanced telecommunication and entertainment services.

This paper will discuss the evolution of cable infrastructure from the point of view of a Photonic Moore's Law that is based on the continuous technology innovation in lightwave, RF, and digital processing. We will describe the new exponential curve that represents this evolution, starting from conventional Fiber-to-the-Serving-Area with fiber node segmentation, through DWDM-based Secondary Ring architecture, to the LightWire™ network with fully-passive coax plant, and to the CoraLight™ architecture based on digital and distributed processing platform with DWDM deployment to the last thousands feet.

PHOTONIC RULES

The innovation and implementation of photonic technology have been accelerating at much higher speed than what Moore projected for computer industry. This has been shown in at least three areas.

Firstly, many barriers in lightwave transmission and applications have been broken. The transmission speed and switching capacity has been increasing exponentially over the past 10 years **(Figure 1)** and reached more than terra-bits per second. 10Gbps system moved from laboratory fantasy to commercial products in 6 years, while the incoming 40Gbps system may only take half of that time.

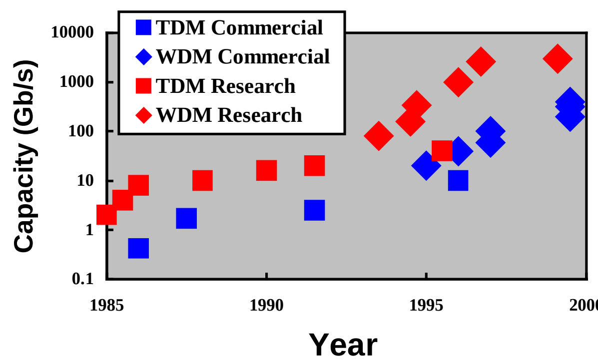


Figure 1. Single fiber capacity

Not only the speed, but also the applications of photonic technology continually broaden their horizon. The advent of linear laser introduced the optical transport technology into cable industry's coaxial networks, which enables tremendous opportunities for emerging service delivery that otherwise would be difficult over conventional twist pair based telephony networks or purely coxial RF networks.

Secondly, with aggressive deployment of lightwave technology, especially in access networks, the price of photonic components and systems has declined at a speed of 1.5 times of that in electronics and DSP **(Figure 2)**. This further motivates deeper fiber penetration and photonic technology deployment on even larger scale **(Figure 3)**.

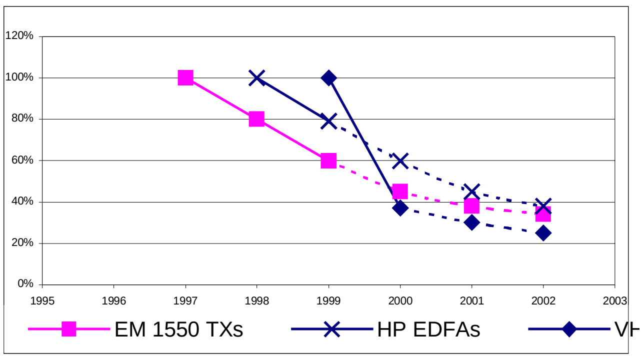


Figure 2. Price decrease of photonic devices

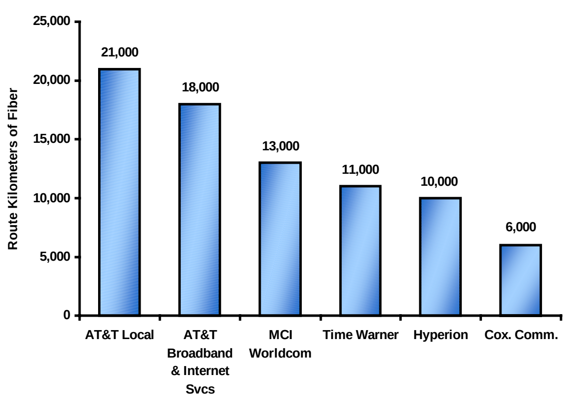


Figure 3. Top CLEC's fiber deployment

Lastly and most importantly, the photonic technology has stuck its head out of pure physical transport and moved to higher layer operation. Many network features and functions that were previously carried by electronic and DSP are now better performed by photonic components and systems. Examples include cross connect, switching, fault protection etc. This allows much more efficient and simpler operation and prevents software and DSP explosion due to the complexity of electronic processing.

All these changes enable new architecture alternatives and different ways of operating the networks with profound impacts on our business.

NETWORK EVOLUTION

Historically, cable networks were established for video broadcasting services. Ever since the advent of linear lasers, the cable industry has been embarking an extensive fiber deployment in the networks (HFC: Hybrid Fiber/Coax) to evolve the infrastructure from a one-way trunk-and-branch plant to a broadband two-way network with superior quality and reliability.

The total bandwidth in a HFC network has expanded exponentially over the past 35 years, enabled by the deployment of fiber optics and technology innovation in RF devices **(Figure 4)**. Most recently, emerging interactive services, always-on applications, and therefore substantially increased network usage increase demand for further bandwidth expansion, especially bases on bandwidth per household passed.

Further, new services, increased customer expectation, and competition motivate network evolution for simple and efficient operation, and continual cost reduction. All these lead to tremendous technology innovation in bringing a traditional broadcast cable network to a broadband two-way infrastructure and further to an end-to-end digital platform with wide implementation of the advances in lightwave and digital technologies.

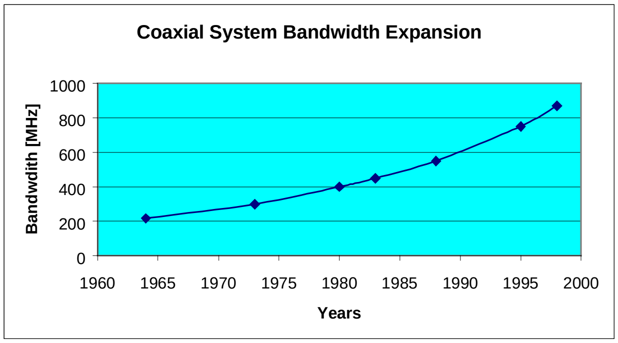


Figure 4: Coaxial Cable System Bandwidth Expansion

DWDM Ring and FN Segmentation

The purpose of fiber node (FN) segmentation is to increase the bandwidth per household passed (HHP) without incurring the cost of relocating fiber nodes and installing new fibers **(Figure 5)**. By dividing the original 1,200 HHP coax bus into four 300 HHP buses, quadrupling of capacity per HHP is realized. Using DWDM technology, end-to-end transparency can be achieved with DWDM Secondary Ring architecture with dedicated wavelength to newly created buses.

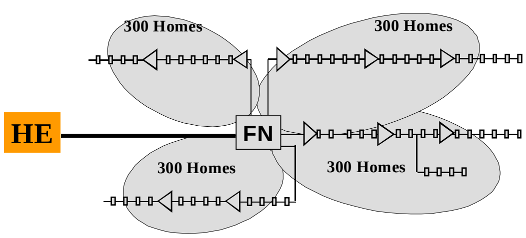


Figure 5. Fiber Node segmentation

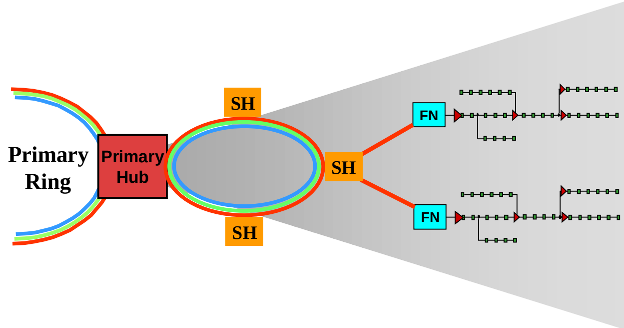


Figure 6. DWDM based secondary ring architecture

Secondary hub concept was introduced as a way of providing route diversity and lowering fiber counts in a single fiber cable route. Since then, secondary hubs have evolved into facilities for signal concentration and distribution, therefore supporting headend consolidation for advanced services **(Figure 6)**. Without this, operating emerging two-way services (telephony, data, etc.) becomes expensive due to the high fiber counts for optical cables connecting primary hub to each fiber node, or the operation cost of allocating HDT and CMTS at each remote headend. **Figure 7** shows the secondary hub architecture using DWDM technology. Combined with DWDM FN segmentation, end-to-end transparency is achieved.

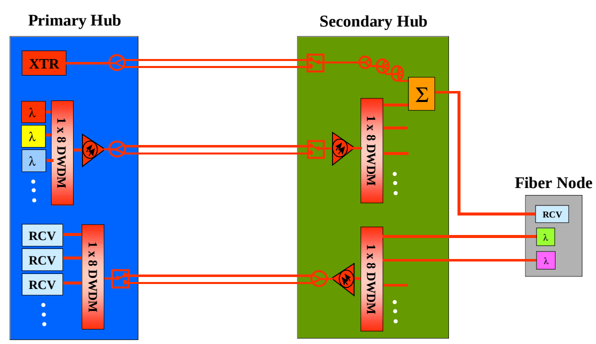


Figure 7. DWDM trunk

Fiber - To - What?

The Photonic Moore's Law enables deeper fiber penetration and deployment of advanced photonic technology in the network. The objectives are to:

1. Establish a future-proof network
2. Simplify operation and reduce operating cost
3. Significantly improve network reliability.

All these considerations lead to industry's continuous efforts in defining and re-defining architectural solutions for HFC networks to capture the ever-changing landscape of service demand and affordability (cost/benefit ratio) of new technological solutions.

LightWire™

The ultimate evolution of the FN based upgrades leads to *LightWire*™ architecture. In this architecture, mini fiber nodes (mFNs) eliminate all the coax amplifiers **(Figure 8)**, and carry both current and new services over passive coax plant.

Fibers connecting multiple mFNs are terminated at the MuxNode that resides either at the original fiber node location or at a location that "consolidates" multiple FNs. As its name implies, the MuxNode performs certain concentration and distribution functions. It "multiplexes" the upstream signals and sends them to the primary hub through the secondary hubs. It also "demultiplexes" the downstream signals received from the PH-SH fiber trunks and distributes them to mFNs.

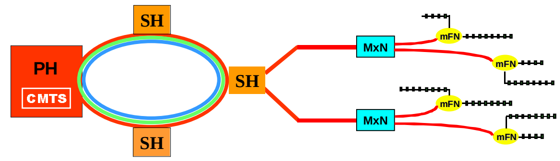


Figure 8. *LightWire*™

One of the interesting features of this architecture is that it maintains the characteristics of conventional HFC networks of being transparent to different signal formats and protocols, therefore fully supporting the existing operation for current services.

CoraLight™

One challenge the industry is facing is to provide flexibility for future expansion and growth provisioning while minimizing the incremental cost and service interruption. To resolve this issue, we proposed a so-called *CoraLight*™ architecture **(Figure 9)**. Based on their geographic locations, mFNs are daisy-chained together with three fibers: one fiber carrying downstream broadcast signals, one carrying the remaining downstream narrowcast signals, and one fiber carrying upstream signals. This therefore implements an optical bus (physical) while the previous *LightWire*™ architecture is a physical star. The advantages are the reduced fiber handling and the cost associated with it, and the flexibility for future expansion and growth (the optical bus can be further expanded to cover more areas). Our preliminary study indicated that, especially in green field situation, the cost of *CoraLight*™ is of parity with that of a traditional HFC network.

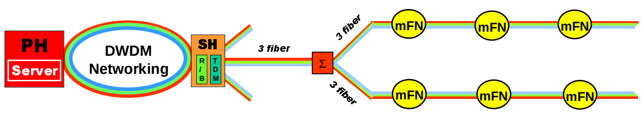


Figure 9. *CoraLight*™

Based on the physical bus, logical star or bus operation can be implemented. Using upstream transmission as an example, each mFN could perform repeating function to realize a bus operation **(Figure 10)**. On the other hand, utilizing WDM technology, a logical star can be implemented with the upstream traffic from each mFN being carried by different wavelengths.

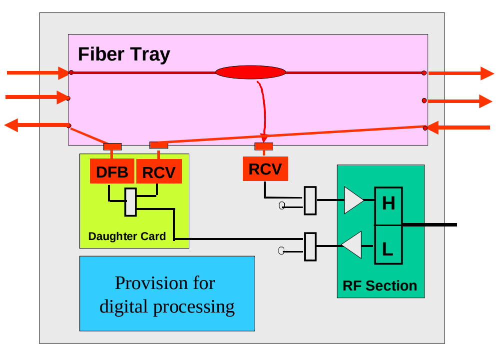


Figure 10. mFN platform with provisioning for the future

Analog to Digital

The linear lightwave technology enabled the implementation of RF subcarrier link over the hybrid fiber/coax infrastructure. This end-to-end format transparent link provides us with many different service delivery mechanisms and new service opportunities. On the other hand, with fiber penetrating deeper into HFC networks, the proliferation of RF technology, together with the power of DSP, motivates distributing certain control functions into the networks to simplify operation and improve network scalability. Terminating RF subcarrier link at the fiber/metallic transition point therefore makes more sense, if not just from cost reduction point of view.

In addition to digitizing the upstream RF bands, the unique position of each mFN enables a considerable simplification in defining media access control (MAC) protocols. Each mFN can do local policing, and resolve upstream contention within its serving area without involving other parts of the networks. The typical headend equipment can therefore be distributed into the network. For example, the RF interface (modulator and demodulator) and MAC functions can be placed at mFNs and the multiplexing/demultiplexing functions can be placed at the secondary hub. By doing this, the current RF optical transmitters and receivers can be replaced by digital ones, and passive optical network (PON) technology, such as spectrum slicing technique with high power LED and DWDM add/drop at each mFN, can be utilized. This further simplifies the transport network, reduces its cost, improves service performance, and enhances network flexibility and scalability.

More On Photonic Rules

The amount of fiber in HFC network has been increasing continuously over the past 12 years from 5% in fiber backbone architecture to 30% in the *LightWire*™ architecture. Different from Moore's prediction, the photonic technology not only contributes to the speed of the transmission, but also adds more features and values to the communication network. DWDM technology, together with the power of DSP and RF technology, provides us with more flexibility in network operation and service provisioning. It is our belief that this trend will continue **(Figure 11)**, and it is certainly to our benefit to take advantage of these opportunities.

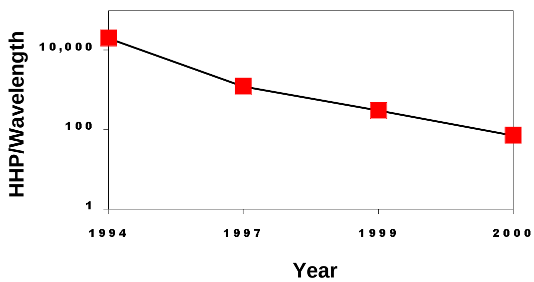


Figure 11. DWDM deployment: HHP per wavelength

BIBLIOGRAPHY

1. Oleh J. Sniezko & Tony E. Werner, Invisible Hub or End-to-End Transparency, 1998 NCTA Technical Papers.
2. Oleh J. Sniezko, Reverse Path for Advanced Services — Architecture and Technology, 1999 NCTA Technical Papers.
3. Oleh Sniezko, Tony Werner, Doug Combs, Esteban Sandino, Xiaolin Lu, Ted Darcie, Alan Gnauck, Sheryl Woodward, Bhavesh Desai, HFC Architecture in the Making, 1999 NCTA Technical Papers.
4. Tony G. Werner & Oleh J. Sniezko, Simplifying the HFC Transport for High Capacity Voice and Data Services, Montreux Symposium '99.
5. Donald Sipes & Bob Loveless, Deep Fiber Networks: New, Ready-to-Deploy Architectures Yield Technical and Economic Benefits, 1999 NCTA Technical Papers.
6. Venkatesh Mutalik, DWDM: Matching Technology Advancements with Business Requirements, 1999 NCTA Technical Papers.
7. Dr. Robert L. Howald, Advancing Return Technology..........Bit by Bit, 1999 NCTA Technical Papers.
8. Oleh Sniezko and Xiaolin Lu, How Much ”F” and “C” in HFC, 2000 Conference on Emerging Technology

Bringing Home the Bandwidth: Optimal HFC Access Architectures for New Builds

Chris Bonang
Harmonic, Inc.

Abstract

As consumer demand for services such as high-speed internet access, multiple voice lines, and video on demand continues to grow, HFC networks are increasingly being recognized worldwide as the only single, proven, residential access network that can deliver the enormous bandwidth necessary to supply these services. Any network built today must be "future-proof", capable of scaling to whatever amount of bandwidth that will be necessary to support any future services and applications that may appear in the next 10-15 years. Even in existing HFC systems with extensive legacy equipment, it is possible to apply solutions developed for new builds when planning system upgrades.

In this paper, optimal HFC designs for such "greenfield" builds are reviewed. The expected growth in bandwidth demand from cable subscribers over the next several years is first reviewed. Deep fiber optical node segmentation schemes which will allow highly scalable bandwidth delivery, and at the same time minimize or eliminate RF actives from the coax plant are then discussed. It is then shown how DWDM can provide significant cost savings and deployment convenience by reducing costly hub real estate and minimizing or eliminating expensive SONET transmission systems. Cost-effective implementation of a combination of digital transmission and DWDM in the return path of such deep fiber architectures is examined. Finally, the discussion of digital return is extended to include the possibility of demodulation, and even reduced CMTS functionality, in the node.

BANDWDITH DEMAND

Over the past few years, the exploding popularity of the internet has revealed a bottleneck in the local access network. Uptake of internet services has been delayed by the frustratingly slow speed of 14 kbps dial-up modems. Recently, new technologies such as cable modems and xDSL have improved peak available downstream residential bandwidth to a bearable 500 kbps. Although the associated customer premises equipment (CPE) is capable of much higher bandwidths, the shared access networks themselves typically do not yet support them. Network operators in the process of deploying new systems do not want to face any such service bandwidth limitations, and so are typically choosing hybrid fiber-coax (HFC) architectures, due to their unparalleled bandwidth delivering capabilities – 800 MHz or 5 Gbps per laser transmitter. The key to making this enormous bandwidth available to subscribers is dedicating narrowcasting services to transmitters and optical nodes serving small service areas in a scalable fashion. Before discussing the details of this process, it is first reasonable to characterize the services whose demand is giving incentive to cable operators to build new HFC networks or upgrade their existing ones.

Peak data rates for current typical internet usage is more than sufficient at 1 Mbps. The increasing popularity of applications like streaming video probably add another 1 Mbps to the potential bandwidth demand. The expanding prevalence of digital cameras, both for still pictures and video, will increase both up- and downstream bandwidth demand, as individuals exchange such material over the internet. This bandwidth forecast is probably conservative, since internet backbone traffic is forecasted to increase by x10 per year, and this increase must also then be reflected in the access network. In addition, the improving economics of video-on-demand (VOD) delivery systems will likely soon result in 1 or 2 movies per household at 4 Mbps per movie commonly being purchased. Based on this analysis, the peak rate to a single subscriber could be as much as 10 Mbps in the next 1-2 years. The services mentioned above are particularly downstream intensive, so a traffice asymmetry of 10:1 downstream to upstream is assumed. Peak upstream bandwidth might be on the order of 1 Mbps in this case. New services such as video conferencing, interactive gaming, and future services that have yet to be conceived must also be taken into account. Consumers will likely use as much bandwidth as is made available to them. Therefore, network operators must choose architectures that will scale to support this almost unlimited bandwidth demand.

NODE SEGMENTATION

The number of subscribers served by an optical node for a typical HFC system has steadily decreased over the last decade. Formerly, 1000-2000 home nodes were the norm, but today new builds and upgrades are more likely to be in the 100-500 homes per node range. Pushing fiber deeper into the network results in improved bandwidth, performance and reliability, particularly due to the reduction RF actives from the coax plant. Completely eliminating RF amplifiers from the system is an excellent goal, since the system would have a 33-50% less active components resulting in greater reliability and reduced power consumption (~40%). Completely passive coax networks are possible for cases of very high subscriber density – multiple dwelling units (MDU's) with +200 subscribers per mile. However, for more typical densities of 100 subs/mile, it is necessary to segment down to approximately 35 homes per optical node in order to create a completely passive coax network. This assumes a four output node with 51-53 dBmV outputs. Unfortunately, for most cable plant, it is difficult or impossible to effectively use all four outputs from the node, and there is not a tremendous cost difference between nodes utilizing 2, 3 or 4 ports. Although it is generally better to push fiber as deep as possible, 35 homes per node is probably not a cost-effective alternative at this time. A practical trade-off is to employ a four-output node passing 100 subscribers and add one line extender to each node output. This provides plenty of RF level at the home, and potentially anywhere from 20 to 50 Mbps of dedicated downstream bandwidth. With only a single line extender in each path, high performance based primarily on the optical link(s) is easily obtained, and higher order modulation formats like 256-QAM and even 1024-QAM can be supported.

DEEP FIBER ARCHITECTURES

Traditional Double-hop HFC Architecture

A traditional HFC architecture is shown in Figure 1. Broadcast analog video is transmitted by an externally modulated 1550 nm transmitter from the head end to the hub. A redundant transmitter and path are included for signal protection. Narrowcasting services (data, voice, and possibly digital video) are typically transmitted via a SONET link, although other options include ATM, IP or proprietary baseband digital transport are also available. At the hub, the digital narrowcast signals are processed for transmission over the HFC network by the appropriate interface unit. A cable modem termination system (CMTS) converts the downstream data signal to QAM, and QPSK demodulates the upstream signal, as well as supplying media access control. It is assumed that any CMTS/Cable modems deployed in the future will be based on the DOCSIS standard. The host digital terminal (HDT) performs the same tasks for cable telephony. Video servers are shown at the head end, but could also be located in the hub to provide digital video and VOD service. After processing and QAM modulation, the 550-860 MHz narrowcast signals are combined with the 50-550 MHz broadcast analog signal and fed to 1310 nm transmitters, which in turn feed multiple or individual optical nodes. Each optical node could serve anywhere from 100-1000 subscribers; for the deep fiber architecture, we will assume 100 subs/node. Since each hub serves on the order of 50,000 subs, the hubs must be relatively large buildings in order to house all of the equipment and fiber connections. The cost of the building and real estate in metropolitan areas can be as much as $2M, assuming that a suitable site can be located.

Several drawbacks exist for operators deploying traditional double-hop architectures in order to serve today's bandwidth-hungry subscribers. Aside from the difficulty and expense of locating and building a large hub site, there are the functional problems of redundancy and sheer numbers of fibers necessary to feed ~500 nodes from each hub. For services such as telephony, most operators require both equipment and fiber path redundancy down to at least the 500-1000 home level. This is difficult to achieve in the double-hop architecture unless there are redundant receivers in each node fed by redundant fiber rings and transmitters. This arrangement provides redundancy down to the 100 home level, which might be considered overkill, and is definitely expensive. A star architecture from the hub to the nodes would mitigate the redundancy problem, but would be prohibitively expensive in terms of fiber cable, since there would be a unique fiber cable to each node.

An additional disadvantage of the double-hop architecture is the high cost of SONET or proprietary digital transport as bandwidth demand increases beyond 1 Mbps per home. For a hub serving 50,000 subscribers, providing just 2 Mbps to each requires 100 Gbps from the head end to the hub. This corresponds to 10 OC-192 terminals in both the head end and the hub, a great expense in both equipment and space. Also, much of the SONET bandwidth is

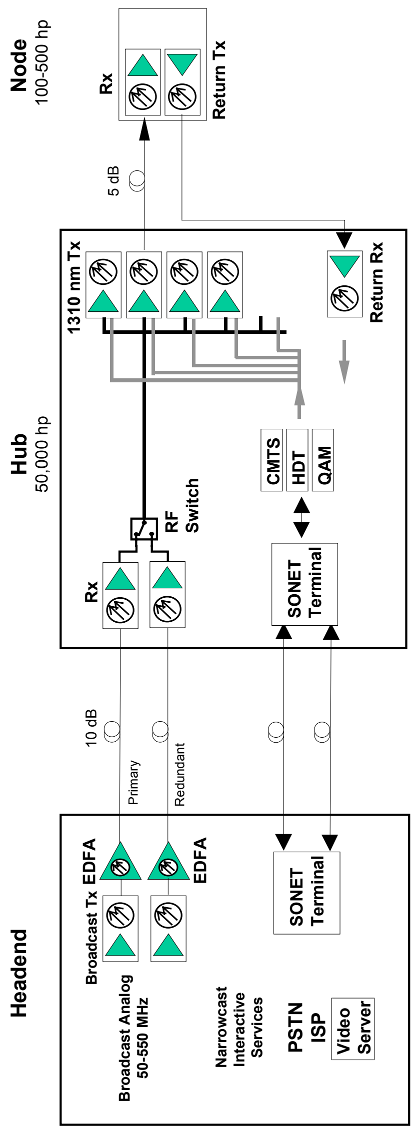


Figure 1: Traditional HFC Architecture

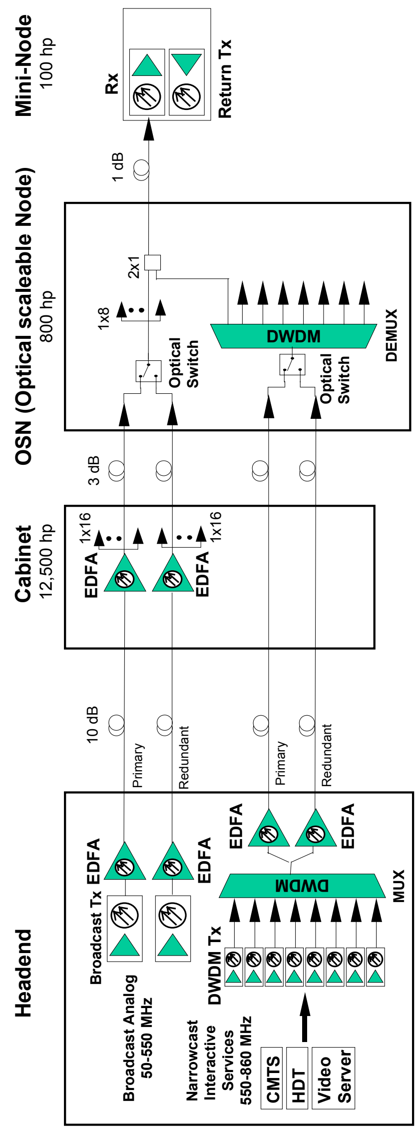


Figure 2: DWDM Deep Fiber HFC Architecture - Forward Path

wasted, due to the 10:1 asymmetry of the downstream and upstream traffic.

DWDM Deep Fiber Architecture

A Dense Wavelength Division Multiplexing (DWDM) deep fiber architecture is shown in Figure 2. As before, the broadcast analog video is redundantly transmitted via an externally modulated 1550 nm transmitter from the head end to a hub. At the hub, the broadcast signal is amplified and split to feed optically scaleable nodes (OSN's). The signal at the OSN, which may be either strand- or cabinet-mounted, is further split to feed 100 home mini-nodes. The narrowcast signals are transmitted via directly modulated DWDM transmitters, whose wavelengths correspond to those of the ITU grid. The wavelengths are multiplexed together at the head end, optically amplified by an erbium doped fiber amplifier (EDFA), pass through the hub site, and are demultiplexed at the OSN. In this ultimate configuration, each wavelength serves a single mini-node, providing 300 MHz of narrowcasting QAM channels (~2 Gbps or 20 Mbps per subscriber). However, the system can be scaled such that initially each wavelength is shared amongst eight mini-nodes. A simple optical splitter is deployed in the OSN instead of a DWDM demultiplexer. As bandwidth demand increases, additional DWDM transmitters are added at the head end, and DWDM demultiplexers replace the optical splitters in the OSN.

In the configuration shown in Figure 2, each OSN serves eight mini-nodes. That limitation presently exists primarily due to the temperature-dependence of DWDM demultiplexers. Wavelength spacing of 200 GHz is necessary for the extreme temperature conditions of a strand-mounted environment. However, next-generation temperature-hardened DWDM couplers will soon be available which allow 100 GHz spacing, and therefore 16 wavelengths can be transmitted to each OSN. Each OSN can then serve 16 mini-nodes, making the system somewhat more cost-effective.

The broadcast and demultiplexed narrowcast signals are combined at the OSN and transmitted over the same fiber to a single receiver in the mini-node. The signals may be combined via a simple 2x1 optical coupler, or by a 2x1 DWDM multiplexer, depending on the available loss budget. If a DWDM mux is necessary due to link budget considerations, then another option is to leave the broadcast and narrowcast signals separate, and transmit them over individual fibers to separate receivers in the mini-node. The cost of a 2x1 DWDM mux and 10 km of fiber is greater than the cost of an additional receiver. In addition, avoiding the 2x1 combiner more easily enables the equipment associated with a 1x16 split to be packed into the OSN, as opposed to only serving 8 mini-nodes from each OSN. An additional advantage of this approach is that it eliminates potential problems associated with the CSO and CTB from the narrowcast signals interfering with the broadcast signal when using a single receiver.

Due to the fact that the forward path signals are passively transmitted through the hub, it is easy to replace the large hub site with a small, inexpensive cabinet or vault. Such cabinets are much easier for which to find locations and only cost on the order of $20K. In

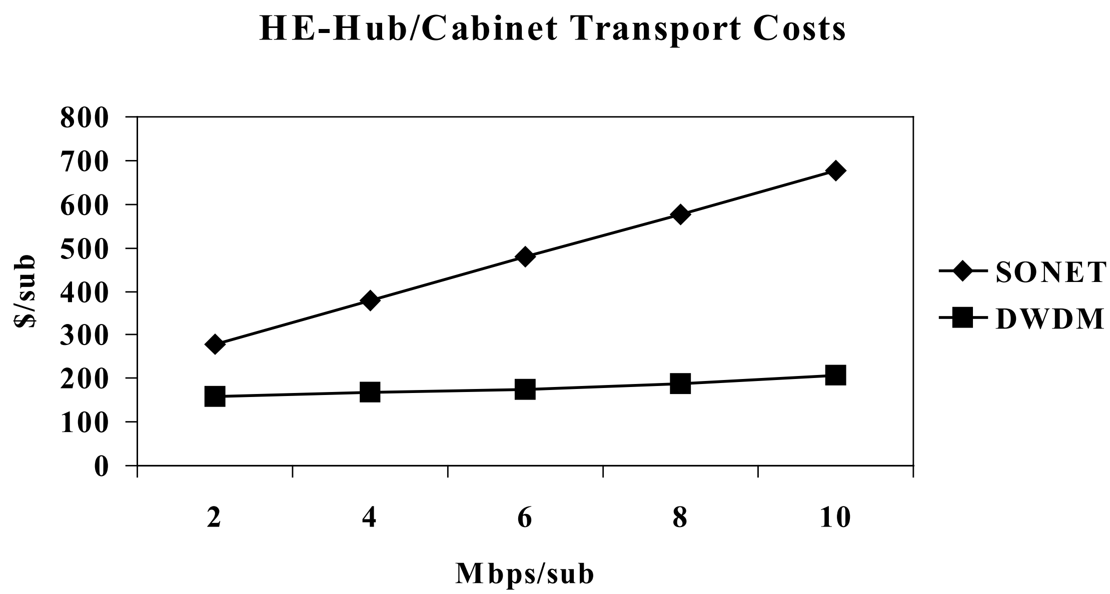


Figure 2, the 50,000 subs formerly served by a single large hub are served by four cabinets. Placing all of the equipment at the head end has the additional advantages of lowering operational costs and permitting less total equipment to be deployed in the initial stages. In the traditional architecture, it is necessary to locate at least one CMTS, HDT and possibly a video server at each hub, regardless of how limited the demand. With the CMTS and HDT pulled back to the head end, this equipment can be shared amongst multiple cabinets/hubs when demand is very low.

The DWDM deep fiber system shown in Figure 2 does not require expensive SONET or other digital transport systems. As shown in Figure 3, when combined with the use of cabinets rather than expensive Hubs, this results in a very large savings as narrowcast bandwidth demand to the home increases into the several Mbps range. The traditional architecture cost increases rapidly with increasing bandwidth demand, but the DWDM architecture only requires additional, and relatively inexpensive, DWDM transmitters. Similar to SONET, the DWDM architecture also provides path and equipment protection via redundant optical amplifiers and optical switches. However, for certain high-priority services, such as telephony, many operators are more comfortable with the extensive protection and monitoring capabilities of SONET. For these operators, a hybrid approach is possible. The low bandwidth, high-priority services like telephony can be transmitted to the hub/cabinet via SONET, while the high bandwidth, non-lifeline services such as VOD can be transmitted via DWDM.

Return Path

Traditionally, multiple return paths are simply combined such that the upstream signal from thousands of subscribers is fed to a single CMTS or HDT. This results in very high noise levels, and limits the number of subscribers who can access the service. As service penetration increases, operators must be able to segment the upstream to serve much smaller numbers of subscribers. Several methods exist for segmenting the return path, thus providing dedicated upstream bandwidth to customers. Figure 4 illustrates the pure DWDM option, which is basically the mirror image of the DWDM downstream. ITU return path transmitters in the mini-node transmit back to the OSN, where the signal is DWDM muxed with the signals from the other mini-nodes served by the OSN. Since the signals are 5-40 MHz analog, it is necessary to amplify with an EDFA before transmitting back to the head end in order to maintain acceptable performance. At the head end, the signals are demultiplexed and fed to individual return path receivers. The DWDM upstream option provides excellent segmentation. But it is not scalable, since the required ITU lasers in every mini-node, and the cabinet EDFA's, combine to make the initial system deployment relatively expensive. The system does provide excellent return path bandwidth of up to approximately 100 Mbps, assuming 16-QAM modulation. This corresponds to 1 Mbps peak rate per subscriber, which may be more than necessary in the early stages of deployment.

A more scalable and less expensive return path option is to combine digital transmission with DWDM. As shown in Figure 5, the 5-40 MHz upstream signal is transmitted by a 1310 nm laser from the mini-node to the OSN. The laser could be either a relatively low-inexpensive uncooled distributed feedback (DFB) laser or a very low-cost Fabry-Perot (FP) laser. The choice between the two depends on how much combining the operator plans to do. FP lasers are more noisy, particularly when no signal is driving them. DFB lasers therefore may be necessary when combining many return path segments, and when high priority services like telephony are offered.

At the OSN, the signal is received and combined with three other upstream signals. The combined 5-40 MHz signals are then digitized by a 10 bit sampling A/D converter. This results in a baseband digital signal of approximately 1 Gbps. This signal is then time division multiplexed (TDM) with the digitized signal from four other combined receivers and transmitted back to the cabinet via a 2.5Gbps ITU transmitter. At the cabinet, the signals are DWDM muxed with other return path wavelengths and transmitted to the head end. Because the signals are in digital format, an EDFA is not necessary. At the head end, the signals are demuxed, converted back to analog format, and fed to the appropriate CMTS, HDT or VOD controller. This system uses fewer ITU lasers than the pure DWDM return option, and no EDFA's, so it is more cost-effective. However, in the initial deployment shown, it only permits segmentation to the 400 home level, which corresponds

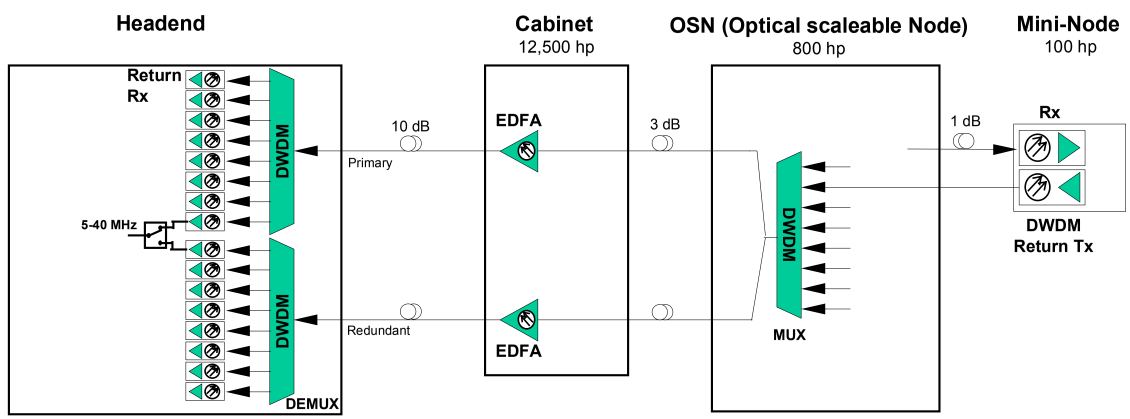


Figure 4: Deep Fiber HFC Architecture - Pure DWDM Return Path

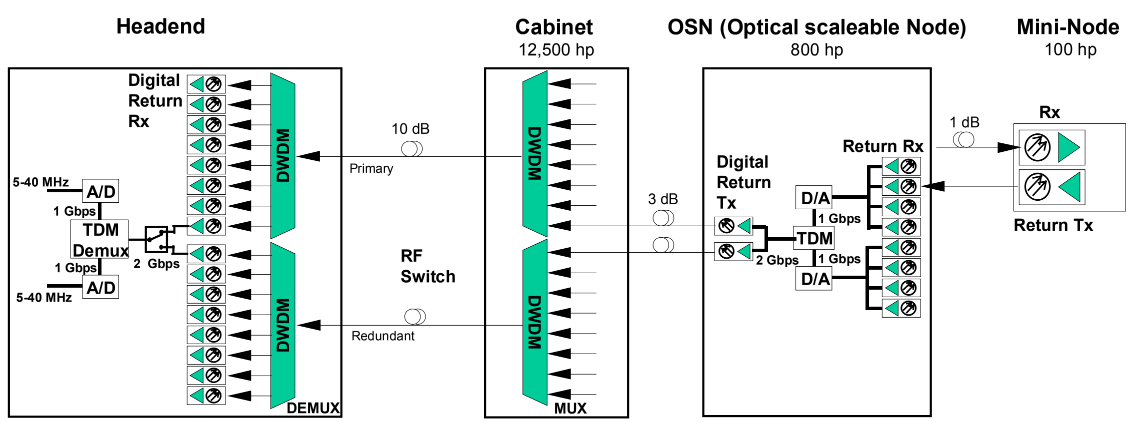


Figure 5: Deep Fiber HFC Architecture - Digital/DWDM Return Path

to a peak upstream rate of 250 Kbps per sub. In order to provide the same segmentation and bandwidth of the pure DWDM option, every return path segment must be digitized separately without combining. Either additional digital 2.5 Gbps ITU transmitters will be necessary, or the eight digitized return segments must all be multiplexed together before transmission. However, this would require a 10 Gbps transmitter, which must be externally modulated, and is probably cost prohibitive compared to a common, directly modulated 2.5Gbps transmitter.

An final return path option is to employ a more efficient form of digitization of the QPSK and 16-QAM signals in the 5-40 MHz return band. Digital sampling of the return path signal is somewhat effective, but very inefficient. The 5-40 MHz waveform is digitized to produce a 1 Gbps signal, despite the fact the maximum useful information carried by the signal, assuming 16-QAM modulation, is only 100 Mbps. A possible solution to the digital upstream efficiency problem under development is to remotely demodulate the DOCSIS QPSK or 16-QAM upstream signal by moving some of the functionality of the CMTS from the head end to the OSN or mini-node. Utilizing such a technique makes the return path more scalable. A further possible step is to locate an entire reduced-functionality CMTS in the mini-node. The device consists of the PHY portion of a regular CMTS (QAM modulator, upconverter, QPSK demodulator) and a rudimentary MAC layer. A two-way ethernet switching fabric transmits the baseband digital signal from the head end to the mini-node.

SUMMARY

Utilizing a Dense Wavelength Division Multiplexing (DWDM) deep fiber architecture overcomes many of the drawbacks associated with traditional HFC architectures. The system is capable of providing enormous amounts of dedicated bandwidth. The architecture is also completely scalable, and cost-effective when compared with traditional dual-hop HFC architectures. The return path uses a combination of digitization and DWDM to provide segmentation and scalability. Next generation technologies may distribute demodulation presently associated with the CMTS out into the node.

Broadband Readiness

Bruce F. Bahlmann
MediaOne

Abstract

The most basic requirement of service activation is still connectivity. Without connectivity it does not matter how quickly new services can be provisioned. Broadband Readiness proposes a way to achieve significant efficiencies in wring customers for new services.

Today, a majority of the homes passed by our cable plant have facilities to support cable television (CATV). These facilities include:

- A drop connecting that home to the cable plant
- Some type of grounding device to protect the home from lightning strikes on the cable plant
- Some type of splitter network device that allows the single drop to provide cable signal to multiple devices
- Cabling that runs from the splitter to one (or more) rooms in the house

It is these facilities that allow various cable television information services to be "relatively" quickly installed and activated in customers' homes. These facilities are based on an installation model developed by the Society of Cable Telecommunications Engineers, Inc. (SCTE) that is 20+ years old. However, new services like telephony, high-speed data (HSD), and digital video require enhancements to this model before they can work properly. These enhancements range from replacing out dated splitters to pulling new cable to the desired room(s). It is these very modifications that represent Broadband's single largest obstacle in terms of streamlining the installation of new services. Time studies of cabling enhancements range from 5 minutes to 2+ hours depending on the size and complexity of the job. Due to this wide range of time spent cabling a home (which is referred to as the radio frequency (RF) portion of the install), the growth in installation rates for new services has been slow because of its high dependency on labor.

While efforts proceed towards streamlining the remaining service specific portions of installs (i.e. performing installation/configuration tasks unique to a specific information service – like configuring a PC for Internet access), each installation still requires completion of the RF portion of the install to activate the customer. Therefore, no matter who/what completes the service specific portion or how quickly it can be completed, the RF portion will remain the single most time consuming task associated with the install. Installation of these new services must follow stringent hardware and cabling requirements to achieve the manufacturers' recommended operational signal strengths. As a result, if cabling is done improperly, its likely to result in a service call to that customer to correct the problem or worst, the improper cabling could negatively affect other customers' service or performance.

The need to complete installations correctly along with the need to grow the installed base for new services gives rise to the conclusion that the current method or model of installing cable (for all services) is suspect. I propose a new model or method of installing services that I call "Broadband

Ready." Instituting Broadband Ready will further elevate MediaOne as a leader in providing broadband services. Broadband Ready is as much a strategy as it is an installation model. Basically, the idea is to expand the task of standardizing the wiring of customer residences to "every" field organization in the company. Make it the mission of every member of the MediaOne field staff to convert every customer residence to a Broadband Ready residence. Field staff in this case would include installers of all types (video, HSD, telephony), service technicians, even field sales staff. If everyone focused on wiring every residence to make them Broadband Ready, all services would benefit from faster installation times, better cable plant performance, and reduced customer initiated wiring. Along with this initiative, it is beneficial to encourage standardization and stomp out the competition. The following action items would limit competition and significantly ease our development and deployment of new broadband services.

Action Item #1: Don't wire a house with category 5 (cat5) twisted pair wire when you can instead wire it with RG6 (MediaOne's chosen standard for coax in the home). By wiring the house with RG6 you prevent competitors from easily replacing our products with their own and then using our wiring. If we wire with cat5, we open the home to any Internet provider that is able to convert to this widely accepted wiring standard.

Action Item #2: Standardize all homes to a similar wiring model. This allows MediaOne to more easily expand the number of services in each home without completely rewiring it with each new addition. Implementing a standard wiring plan for all homes passed increases the market value of each home and simplifies the job of engineering new services to work in that environment.

Action Item #3: Establish regional boards that work with local building professionals, city planners, and state certification agencies to uphold recommended wiring practices and material requirements for expanding only the "approved" use of broadband. Closely track the progress of these regional boards in changing public policy, codes, and new building practices to use the approved standards. Make these boards responsible for reporting on their progress and set goals for the assimilation of all involved parties to the new standards.

By the Numbers:

MediaOne currently performs three different types of field activity: installation, service, and direct sales for each service type (video, HSD, telephony) it provides. The numbers in Table 1 represent a single region's annual occurrences of this activity for an area with 583,707 homes passed and 317,819 active customers. Table 2 attempts to translate these numbers across all regions by basing these numbers on per homes passed and per active subscriber which was obtained from Table 1. The objective of Tables 1 and 2 is to provide reasonably good numbers that represent how often we visit a customer's residence. What this means for the Broadband Ready initiative, is that each visit to a customer residence represents an opportunity to perform the needed work to make that residence Broadband Ready.

	Per Home Passed	Per Subscriber
Number of Subs	583,707	317,819
Installation Visits	150,780	150,780
Service Visits	204,180	204,180
Direct Sales Visits	1,712	1,172

Table 1 -- Annual field activity for MN region

	Per Home Passed	Per Subscriber
Installation Visits	0.2583	0.4744
Service Visits	0.3498	0.6424
Direct Sales Visits	0.002933	0.005387
Combined Visits	**0.611033**	**1.122187**

Table 2 -- Estimated number of annual visits

In between the lines of these numbers represent "reality" in terms of actual non Broadband Ready residence (or unique) customer visits – where "unique" visits represent those to customers that are not already upgraded to the Broadband Ready. For example, each year MediaOne does not perform installations to 47.44 percent unique subscriber residences or service 64.24 percent of unique subscriber residences. There are repeat service calls to the same residence and repeated installations to the same subscriber residence (i.e. apartments). Therefore some fraction of these numbers hold true for year-to-year unique subscriber residence visits. Conservatively, I estimate that ¼ of each of these types of visits are actually to a unique subscriber residence. While the initial years of implementing the Broadband Ready initiative would have a high degree of unique subscriber visit opportunities, the years that follow would steadily decrease. Also note that during the first year there may be higher than estimated unique subscriber visits. However, not all customers will want or be willing to allow the upgrade to take place because of personal preferences, time constraints, or other inconveniences that could reduce the number of Broadband Ready upgrades actually performed. Any lower than average response during the first two to three years, would merely result in a slower accumulation of Broadband Ready homes passed in the years to come.

Assuming the number of homes passed remains constant, its predicted in Table 3 that over first 3 years of implementation approximately 45 percent of the homes passed would become Broadband Ready. Since, as a company, we typically penetrate close to 60 percent of the homes passed, the 45 percent of homes passed that are Broadband Ready would represent 75 percent of our current subscribers as being Broadband Ready. Any increase in the homes passed (i.e. system trades) would only extend the initial accumulation phase, however, note that after the initial accumulation phase the percent of Broadband Ready homes passed will grow more slowly over time.

	1st Year	2nd Year	3rd Year
Installation visits	6.48 %	6.48 %	6.48 %
Service visits	8.75 %	8.75 %	8.75 %
Direct sales visits	0.0733 %	0.0733 %	0.0733 %
Accumulated % Broadband Ready homes passed	**15.30 %**	**30.60 %**	**45.90 %**

Table 3 -- Accumulation of Broadband Ready homes passed

<u>Projected Payback:</u>

The payback of the Broadband Ready initiative is significant. It is also likely to yield additional installation, marketing, and engineering efficiencies than what is explained in this document across all MediaOne services -- not to mention the under the cover increases that the model would bring to plant signal quality and improved video. The most obvious efficiency that the Broadband Ready initiative will yield is to HSD and digital video installations (which are the focus of this document). Activating the Broadband Ready would mean that approximately 75 percent of all HSD installations to current subscribers would not require any RF installation time to complete. Like wise, each year that the Broadband Ready is active there would be further increases in the percentage of HSD installs that would not require any additional RF wiring. If the Broadband Ready initiative was taken seriously and adopted as a long term company goal it is realistic to anticipate a time where only trivial RF wiring (i.e. custom jumpers) is required to install services. Table 4 displays the projected return on investment in terms of increasing the installation capacity of the existing HSD install team without increasing the number of its full time employees (FTE)s.

	1st Year	2nd Year	3rd Year
Accumulated % Broadband Ready homes passed	15.30 %	30.60 %	45.90 %
% HSD Installs in Broadband Ready residences	25 %	50 %	75 %
Installer time savings on Broadband Ready installs	½	½	½
Increase in overall number of HSD installations due to encountering Broadband Ready residences	+12.5 %	+25 %	+37.5 %

Table 4 -- Payback on investment for Broadband Ready

Basically, each of the first three years that the Broadband Ready initiative is active, MediaOne should expect an approximate 12.5 percent increase in installation capacity* of its existing HSD install team. This increase would be the result of being able to do twice the number of "normal" HSD installations 25 percent of the time the first year, 50 percent of the time the second year, and 75 percent of the time the third year.

**Note this does not take into account the improvements in streamlining the PC portion of the HSD installation. When improvements to the PC portion can be realized, the impact (when combined with Broadband Ready efforts) would yield significant reductions in the overall time spent to install HSD services. In fact, its likely that we'd see average HSD installations times in the order of 30 minutes at that point – most of which would be time taken for customer education. Also note that installation capacity is dependent on sales/marketing and cannot be directly attributed to actual increases in subscriber count.*

The other payback is the inside sales data that is accumulated as a result of the customer having a Broadband Ready residence. For example, if a customer is flagged as having a PC at the time they are upgraded to Broadband Ready, this information could be used by MediaOne telemarketing to offer them HSD services at a later date. Alternatively (once ready), MediaOne could send them a self-installation CD so they would be able to subscribe to HSD services at their convenience (without any truck roll)**.

***Caution: It is my belief that MediaOne provides a "personal" service to the customer by going to their residence and helping them get access to information services (be it video, high-speed Internet, telephony, digital video, pay-per-view, etc.). Additionally, MediaOne helps the customer use this service by teaching them how to use it as well as answering questions one-on-one. The fact that we come to the customers' residence is what makes MediaOne a superior information/technology provider.*

Therefore, it is NOT in our best interest to provide service without any customer contact. It's the contact (sales, installation/instruction, and service) with the customer that differentiates us from our competition.

Projected Costs:

To proceed with the Broadband Ready initiative, MediaOne will need additional resources (FTE)s and contractors to help carry out the mission of making every subscriber residence Broadband Ready. A majority of these resources already exist within the current video side of the business and merely need to include additional duties with each unique subscriber visit.

	1st Year	2nd Year	3rd Year
Accumulated % Broadband Ready homes passed	15.30 %	30.60 %	45.90 %
% of video installs and service calls requiring upgrade to Broadband Ready	15.30 %	15.30 %	15.30 %
Time required to upgrade customer to Broadband Ready	½ - 2 hours	½ - 2 hours	½ - 2 hours

Table 5 -- Projected time requirements of implementing Broadband Ready project

Table 5 represents the additional time required to complete Broadband Ready upgrades on new or existing unique customer residences. Once a customer is upgraded to Broadband Ready, a flag in the billing system would be set for that residence upon check-in process. This flag wound then provide information to those routing technicians to installations and service calls as to the anticipated amount of work required during each customer visit. This information could also be used by the billing system to correctly associate the number of points with Broadband Ready customer visits as opposed to non-broadband ready customer visits. Based on the requirement of up to a 2 hour extension of 15.30 percent of the customer visits during each of the first three years, field fulfillment organizations should plan on an increase of 15-20 percent in the number of FTEs currently supporting video installations and service calls. Excess and peek demand for Broadband Ready upgrades could be handled by hiring of contractors to fill the gap. Note that in the event of direct sales initiated upgrade, additional wiring could be provided or scheduled with a contractor that day.

In addition to the resources needed to carry out the Broadband Ready, another dollar figure must come into play for the extended time and materials video service or installation personnel would require to complete an Broadband Ready upgrades. This approximate dollar figure (see Formula 1) is based on the following assumption:

The technician is already on site (i.e. the Broadband Ready upgrade does not incur the full costs of a truck roll because the technician was already required to be at that customers' residence).

Cost of Broadband Ready upgrade ($27.37) = Cost of Additional Outlet labor/parts (~$19.95) + HSD Required Parts ($7.42)

Formula 1 -- Cost of each Broadband Ready upgrade

To determine the cost of implementing Broadband Ready from a time and materials stand point (less the cost of actually employing the required additional FTEs mentioned earlier) the following formula calculates this out by using the annual combined visits per homes passed calculated earlier (see Table 2):

0.09349 unique visits per hm passed = 0.611033 Combined annual visits per hm passed x 15.30% Broadband Ready residences growth

Finally, the number of unique visits per home passed allows us to determine the cost of Broadband Ready per home passed. This is calculated as follows:

$2.56 dollars per home passed = 0.09349 unique visits per hm passed x $27.37 cost of Broadband Ready upgrade

Based on these calculations, one can approximate the cost of time and materials required of implementing the Broadband Ready program for any MediaOne region (again minus the cost of actually employing additional FTEs mentioned earlier). To determine this cost multiply the $2.56 figure times the total number of homes passed. For example, in the case of the MN region where it currently has 583,707 homes passed, the annual cost of time and materials for implementing the Broadband Ready would be approximately $1,494,289.92

<u>*Review of Costs and Paybacks:*</u>

As with any long-term project of this magnitude the costs up-front are significant! However, the long-term payback for this out weighs the up front investment. For example the number of FTEs required by a region to increase HSD installation rate by 12.5 percent alone may be worth the investment.

Installers	100
Maximum number of installs per day	40
Maximum number of installs per year counting 2 weeks vacation and no other absences or turnover	96,000
12.5 % increase would amount to this many additional installs	12,000
Equivalent number of FTEs to produce same output: 120 / (40x24)	12.5 FTE

A 12.5 percent increase in HSD installs would require a similar 12.5 percent increase in the number of FTEs dedicated to HSD to achieve a similar increase in the number of install capacity. Any increase in installation personnel is costly as it requires paying the FTE's salary, benefits, vehicle, and all the supplies needed to outfit the installer to accomplish his/her job. Using the numbers from above and assuming the costs of each installer is $70,000 ($20,000 vehicle, $40,000 FTE/benefits, $10,000 installation test equipment and materials) the savings of 12.5 percent* FTEs would save the company $875,000*.

After the third year the 37.5 percent increase in the number of HSD installs and reduction in FTEs would save the company $2,625,000*. MediaOne could also capitalize the opportunity cost created by the addition of approximate 36,000* additional HSD customers it would not have realized without the Broadband Ready. This amounts to approximately $2,876,400 in annual revenue. These savings would continue to rise as the HSD installation rate continue to increase (due to the time savings of the Broadband Ready) while the costs of running the Broadband Ready decline (due to reducing number of non Broadband Ready visits).

**Based on the example above using 100 installers*

Based on these numbers, and after 3 years, MediaOne could let go or promote the expanded FTEs required to implement the initial labor oriented phase of Broadband Ready initiative and fall back to today's field fulfillment FTE count yet improve this groups overall installation capacity 37.5 percent. Additional benefits of improved video performance, reduction in the numbers of service calls, and the reduction of time required to complete video or telephony installations are difficult to quantify in terms of cost savings. However, it is projected that as the number of Broadband Ready homes passed approaches 50 percent (a 5 year projection) MediaOne will have paid back the investment required to initiate the Broadband Ready and begin seeing enormous reductions in its operating costs and average installation times.

Compatibility Issues in Block Conversion for Return Path Concentration

James O. Farmer
Mark A. Linford
ANTEC Corporation

ABSTRACT

A few months ago we received reports of problems using block conversion with DOCSIS modems. We attempted to duplicate the problems in the lab, but were unable to do so. We are not able to show reasons for the failures, but we are able to discuss possible sources of the problem. Our tests indicate good safety margins with the equipment we used. Block diagrams of practical laboratory tests are shown, which can help identify any possible problems before system deployment.

SUMMARY

Block conversion is a technique used to improve the utilization efficiency of the return path. Return path spectra from multiple nodes are converted to different bands, then the bands are combined to modulate one laser. This provides a convenient, economical way of using one fiber to transport up to 18 return spectra at one wavelength.

A few months ago we heard of field problems involving block conversion and DOCSIS modems. The report was that the modems didn't work over a block conversion system. We set up a simulation in the lab to understand the problem. Commercial concerns precluded our obtaining the same equipment used in the failed field test, so we had to use available equipment. For better or worse, we were not able to induce failures similar to those observed in the field. In fact, we found excellent margins to any failure modes. Thus, we are unable to report the source of the field failure. However, we can speculate on some of the possible reasons for the failure, and show test systems that will allow simulation of field conditions, allowing you to do your own testing.

BACKGROUND

Figure 1[i] illustrates a basic block conversion system. Multiple return path spectra are converted to different frequencies before being modulated onto a single optical transmitter. As illustrated in Figure 1, block conversion is used at a node, where up to four return paths are placed in the spectrum from 5 MHz to about 210 MHz, and modulated onto one laser for upstream transmission. This allows one fiber to be used to transport four individual return paths. In the simple system illustrated, single conversion is used, so that the three bands that are frequency translated are also inverted. At the headend the bands are block down converted, with the spectrum being returned to its normal relation.

An alternate form of block conversion is employed to allow return signal transport from a hub to the headend. In this application all blocks are converted to fit the spectrum from 50 MHz to 860 MHz. In the North American channel plan, up to 18 blocks can be accommodated on one optical path. The composite spectrum is applied to an optical transmitter of the type normally used for downstream transmission. In this configuration, double conversion may be used to eliminate the spectrum inversion of figure 1. Of course, double conversion adds

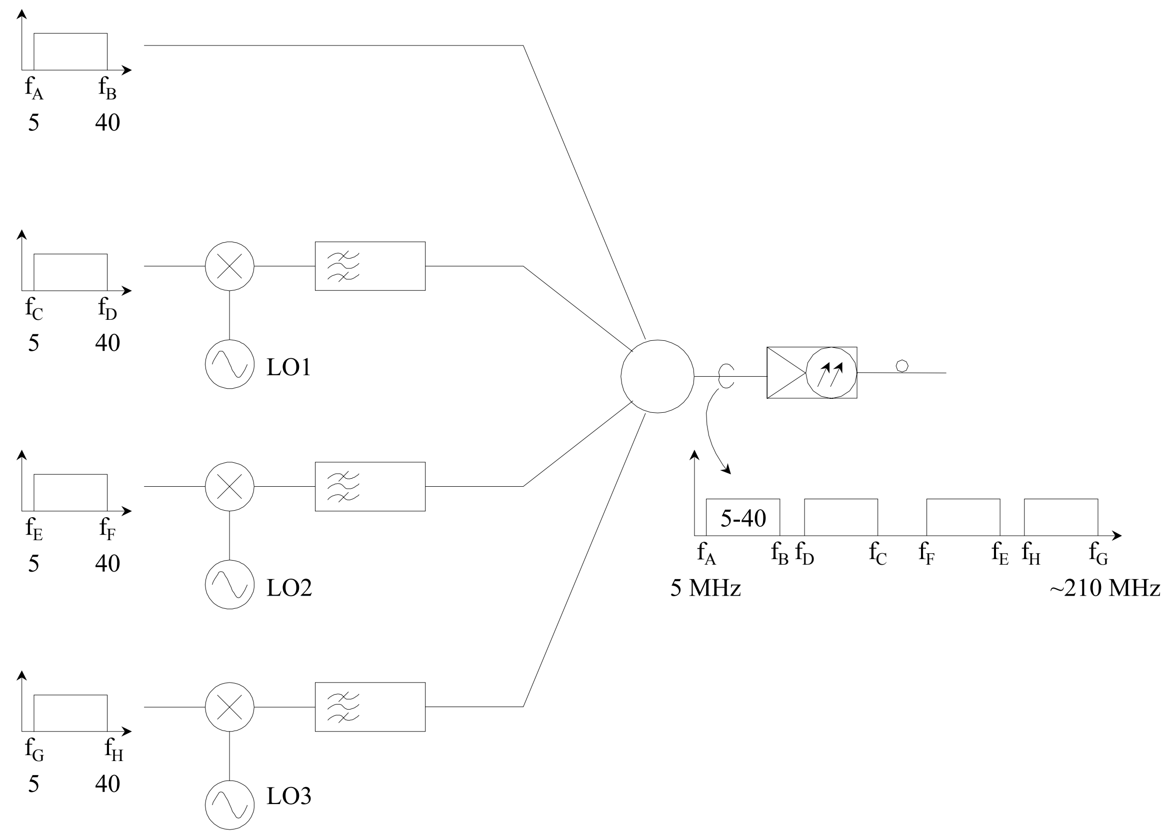


Figure 1. Basic Block Conversion

complexity and potentially more phase noise, but you can show that overall, it provides a more practical block conversion system for the application. It is quite possible to further multiplex different wavelengths, providing up to about 288 return blocks (10.7 GHz) on one fiber.[ii]

SYSTEM SIMULATION

When we first became aware of problems in the field, we set up a simulated system in the laboratory in order to try to duplicate the problems. Unfortunately we were not able to gain access to the system that had shown problems in the field, nor were we able to obtain that equipment for the lab test, so we set up a similar system using different equipment. We were not able to duplicate the problems - the lab system worked quite well - but we did gain insight into what might have happened in the field.

Figure 2 illustrates the system set-up in the laboratory used to simulate the system tested by others in the field. The "headend" on the left supplies downstream signals to a node. The downstream signals consisted of the incoming feed of our local cable system to 550 MHz, combined with the output of an Arris CMTS1000 cable modem termination system operating in the 256QAM mode. This signal was supplied through a typical length of fiber to the node, which in turn supplied signals to two RCA brand cable modems purchased at retail. Each modem was connected to a computer, and the task against which we judged performance was the transfer of a large file from one computer to the other. This exercised both the downstream signaling, which was not under test, and the upstream signaling, which was under

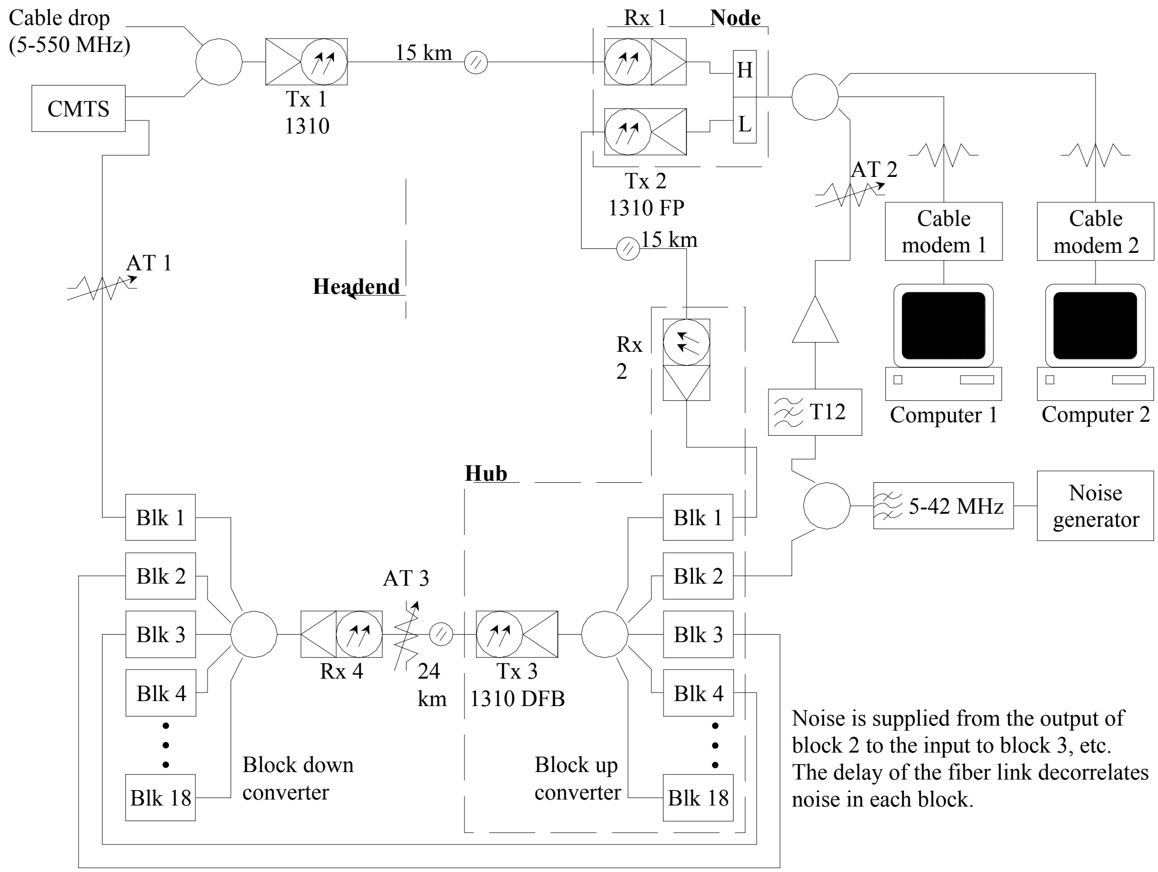


Figure 2. Simulation of Field System

test. The upstream communication was operated in the 16QAM 10.24 Mb/s mode, the highest upstream speed currently defined for DOCSIS-compliant modems.

To simulate a real return path, we used a Fabry-Perot (F-P) return laser in the node, which supplied signals to a "hub," which comprised an 18 band block converter and a DFB return optical transmitter. The return path was loaded using a noise generator with a 5-42 MHz bandpass filter. This signal was split, with a portion of the signal being supplied to the second block converter of the 18. At the "headend," the output of block 2, which contained the noise generator signal, was looped back to the "hub" through a short cable, and supplied to block 3, whose output was supplied to block 4 and so on. By doing this, we de-correlated the noise supplied to each block. This is necessary to make the resultant signal have about the same peak to average characteristics as would real return signals. The roughly 130 μs of delay in the return fiber ensured that peaks in the random noise did not occur simultaneously in each block. For practical reasons, we didn't loop all 18 channels as shown, but we did loop them in three groups, which yielded about the same results.

Result of System Simulation

We were able to transfer files between the two computers through the CMTS with no measurable errors. The operational dynamic range of the system could be tested by adjusting AT1. This attenuator changed the return path signal level received by the

CMTS. Long loop automatic level control resulted in the output level of the two modems varying as AT1 was adjusted. Simultaneously AT2 was changed to keep the noise at a constant level relative to the modem outputs. When the dynamic range was measured at channel T12, we had an error-free range of 27 dB. At the high-level end of the range, we experienced a transition from error-free transmission to complete failure with a 1 dB change in level. This was likely due to clipping in the F-P laser. At the low-level end a 1 dB change produced a transition from error-free transmission to errors reported by the CMTS, and an increase in file transfer time from 223 to 275 seconds. We did not experience a failure to communicate, however.

When the test was repeated using a modem return center frequency of 9 MHz, the high-level limit was similar, but we experienced a worse low-level limit and a dynamic range of 14 dB. This is believed to be due primarily to noise from TX2, which was loaded very lightly. Many modern return transmitters employ dither techniques to improve the dynamic range at low levels, but the transmitter chosen for this test did not have a dither circuit. The problem may have been exacerbated by group delay at the low end of the spectrum.

The tests were repeated using blocks 9 and 18 with substantially identical results.

MARGIN TEST

Finding no problems with the system set-up, we investigated the amount of additional degradation that the system could tolerate. The block conversion system was modified to introduce errors that might cause failure of the return path, and we investigated how much additional error could be tolerated before we encountered system problems.

The block conversion system used in testing has a pilot carrier that is transmitted from each block to the headend. The pilot is used to stabilize the gain of the return path against changes due to temperature and optical path changes. The pilot is also used to force zero frequency error in the block conversion process. Frequency conversion error is not necessarily a problem for all return systems. However, there are some return applications that demand zero error. For example, a few years ago the industry was looking seriously at using cable to link PCS (personal communications service) mini-cells. This is a cellular-like telephone system that uses small base stations. It is common for a phone to be simultaneously in contact with multiple base stations, which are linked back to a master controller by the cable plant. The master controller can work with signals being received simultaneously by multiple base stations, but only if there is no frequency error between the received signals. If the several base stations communicate upstream through different cable nodes, there must be no frequency translation error.

The system tested by the MSO may not have phase locked the up- and down-conversion processes, so we experimented with breaking the loop and introducing intentional frequency error. Another concern is with phase noise in the frequency conversion process. Any phase noise in the local oscillators will be transferred to the signal, and if enough phase noise is added, demodulation of the return signal can fail.[iii]

Figure 3 illustrates the configuration used to test these and other hypotheses regarding what could go wrong. The CMTS was connected to our internal network to provide access to the internet, and was con-

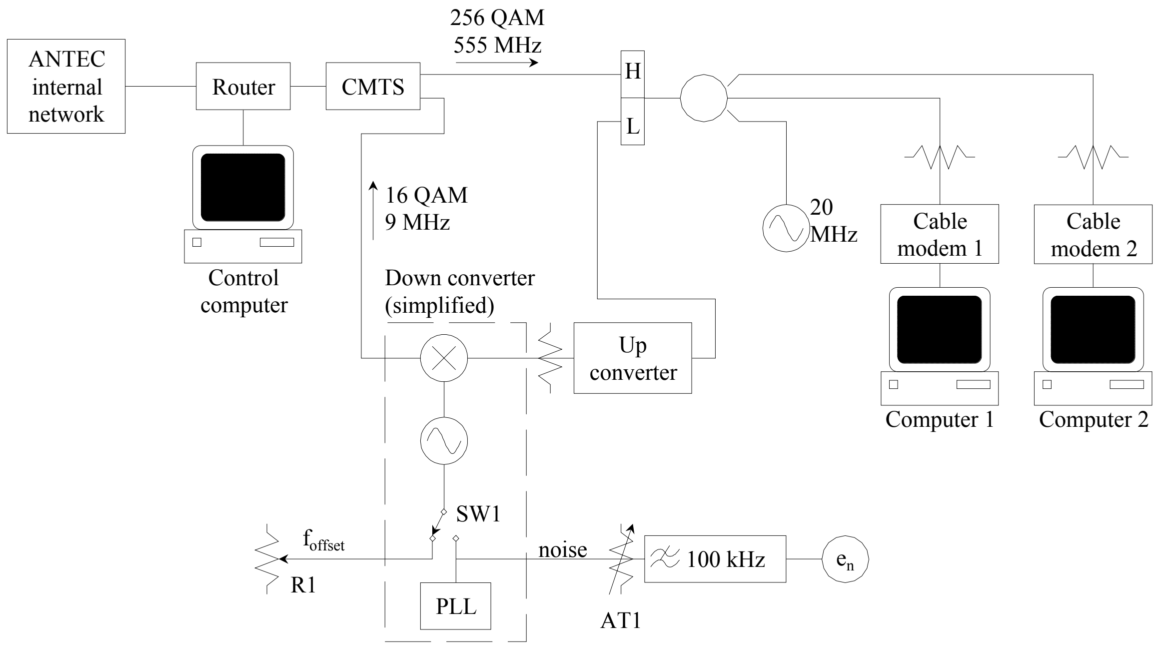


Figure 3. Block Converter Margin Test

nected to the two modems and computers as described above. In this case we did not add the optical network, since we were concerned only with what might happen in the block converter. The forward path was transmitted directly from the CMTS to the modems. A diplex filter routed the return signals through a single up converter and down converter, of the type used in Figure 2. The down converter was modified to allow us to break the phase locked loop in order to introduce frequency errors. We also added the ability to introduce phase noise into the closed loop. The block down converter is actually a dual conversion device as described above, but we have shown only one conversion here, because we did nothing with the other conversion.

The 20 MHz generator was used to give us a return path signal that we could measure in order to determine frequency offset and phase noise. By observing the 20 MHz signal on a spectrum analyzer at the "headend" we could tell how much frequency error or phase noise we introduced.

To introduce phase noise, we used a pseudorandom noise generator followed by a low pass filter to limit the noise to about 100 kHz. Attenuator AT1 allowed us to adjust the amount of added phase noise until we encountered errors. We monitored the control computer for errors reported in the return transmission, and defined failure as occurring when we saw any reported errors or delays in file transfer. We used the same file transfer test reported above. We also tried using a web radio broadcast as the source, but there is little return data required to keep the broadcast alive, so it didn't really stress the return path at all. Also, we found so many internet problems that it was hard to tell when the return path was at fault.

Frequency Offset

We were able to offset the frequency +30 kHz and -35 kHz by changing the control voltage to the oscillator. Over this range no change in operation of the return path was noted. This corresponds to a worst case

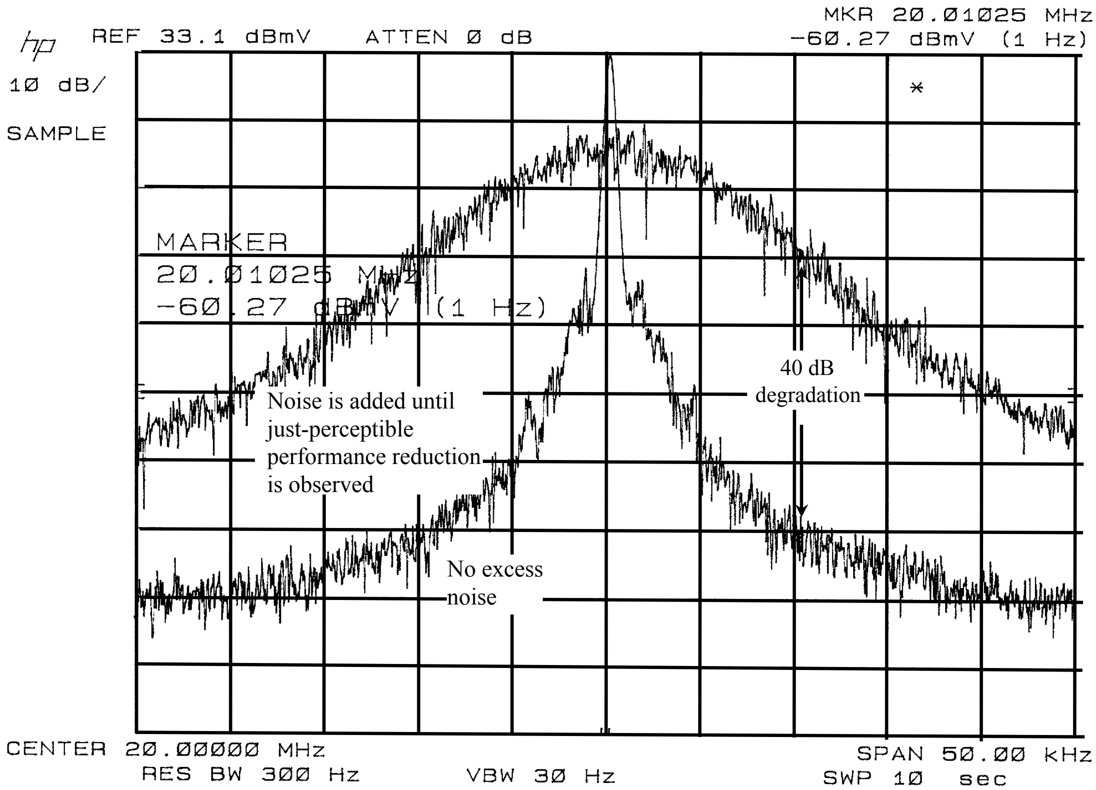


Figure 4. Normal Phase Noise and Enough Phase Noise to Stress System to the Threshold of Error Introduction

frequency translation error of about 37.5 ppm, which is not a difficult number to achieve today. Since we saw no problems with this much error, we concluded that even a block conversion system that was not phase locked should work, at least with this combination of modems and CMTS.

Phase Noise

It is known that if enough phase noise is introduced into a transmission path, digital transmission will fail. Figure 4 shows the phase noise on the 20 MHz carrier without any added noise, and also shows the phase noise with enough additional noise to induce errors into the 16QAM transmission. Often phase noise is measured as so many dB down from the carrier at a 10 kHz offset, measured in a specified bandwidth. For this test, we were not as much interested in the exact number of dB the phase noise was down, but we were interested in knowing how far we were from problems.

We found that we had to add enough noise to bring the noise sidebands up 40 dB (at 10 kHz offset) in order to induce errors. Note that this was so much noise that the carrier peak is not discernable at the resolution bandwidth used. Comparison of the no-excess noise plot with archival records dating back to initial product approval showed that we were measuring more than 10 dB more noise here than in the approval measurements. We suspect that this is due to noise on the 20 MHz oscillator, which was a medium-quality variable signal source. Thus, we probably had more than 50 dB margin in phase noise.

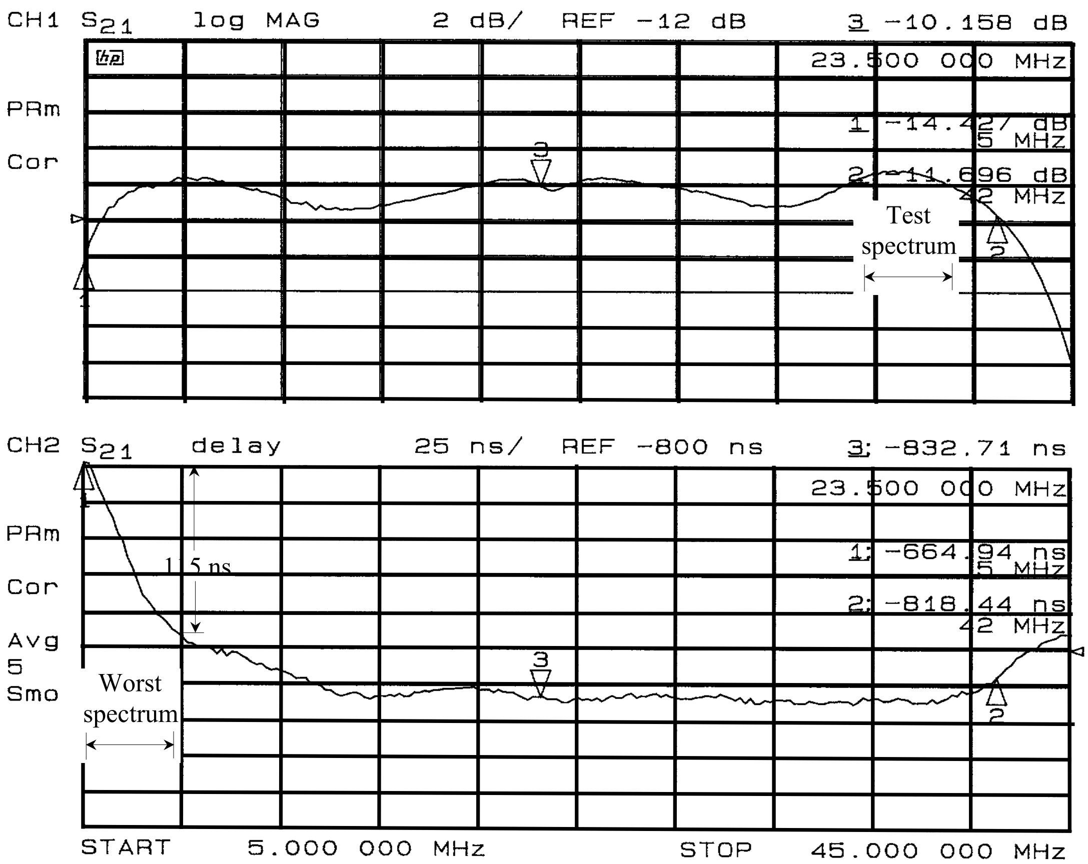


Figure 5. Amplitude and Delay Response of Block Conversion System

WHAT ELSE CAN GO WRONG?

Having not found any significant problems that would explain the field failures, we searched for other issues that may have been limiting factors in the field performance of the other equipment. It certainly is possible that errors in frequency or delay response of the block conversion system could cause premature failure to communicate. In the present DOCSIS specification, there is no adaptive equalizer in the return path. Inclusion of adaptive equalization is difficult because the amount of equalization required is different for every modem, depending on the equipment between the modem and the headend. Also, lower order modulation methods, such as the QPSK and 16QAM used in DOCSIS return paths, are not that susceptible to errors. (In future generations of DOCSIS specification, where higher levels of return path modulation are used, the specification may provide for pre-distortion in the modem transmitters, based on headend measurements relayed back to the modem.)

Figure 5 is archival test data showing the amplitude and delay response of a typical block conversion system, including a 15 km optical path. In the S_{21} log MAG (amplitude response) path at the top, we show the test spectrum where the return path signal was operated for this test. The spectrum is 3.2 MHz wide, the widest bandwidth currently specified by DOCSIS.[iv] The peak-to-peak amplitude error is negligible over this bandwidth. Placing the return carrier at other frequencies would have made little difference.

Of possibly greater interest is the group delay shown at the bottom. The block converter uses ceramic resonator filters that do exhibit group delay, but delay equalization is provided. Note that the residual group delay is worst at the very low end of the spectrum. Over the bandwidth occupied by the DOCSIS signal, the delay is about 115 ns. The maximum symbol rate specified in DOCSIS for the return path is 2.56 Ms/s (mega-symbols per second), so the period of one symbol is the reciprocal, or 391 ns. This is still longer than the delay, so we might suppose that group delay alone would not be fatal to a data signal carried in this spectrum. However, other delay issues and, more importantly, noise considerations, would preclude use of a return path this low in frequency. If any signals are placed this low, they should only be extremely robust BPSK or FSK signals at very low data rates, with the ability to accept errors.

Marker 2 is at 42 MHz, the highest frequency specified for the block converter system under test. Note that the group delay is starting to rise slightly, but group delay at this frequency will be dominated by that of the diplexers in the node and amplifiers. For this reason it is not recommended to use frequencies above about 40 MHz with a 42 MHz return cutoff.[v]

CONCLUSION

We received reports from an MSO that he had tried a block conversion system with DOCSIS modems, and experienced unsatisfactory performance, which was attributed to block conversion. We were unable to obtain the equipment used in his test for commercial reasons, but we did set up several laboratory experiments to try to duplicate the results, using equipment available to us. We were not able to experience the same failures, but we were able to define some parameters that an operator may want to look at to ensure satisfactory performance. Our test confirmed that DOCSIS modems can work very well with block conversion.

CREDITS

We appreciate the help of David Hodgdon of Arris Interactive, in getting the CMTS up and running in our laboratory quickly. John Kenny directed much of the testing.

James O. Farmer, Chief Technical Officer
Antec Network Technologies
11450 Technology Circle
Duluth, GA 30097
Tel: 678.473.8005
Fax: 678.473.8470
jim.farmer@antec.com

Mark A. Linford, Staff Engineer
Antec Network Technologies
11450 Technology Circle
Duluth, GA 30097
Tel: 678.473.8037
Fax: 678.473.8040
mark.linford@antec.com

---

[i] Ciciora, Walter et. al., *Modern Cable Television Technology: Video, Voice and Data Communications*, San Francisco: Morgan Kaufmann, 1999, p 579
[ii] Kenny, J., *DWDM Block Converted Return Transmission Performance*, companion paper presented at the 2000 NCTA Technical Program
[iii] Ciciora op. cit., pp 178, 179
[iv] ibid., p 209
[v] ibid., p 579. Another reason for not using frequencies above 40 or 41 MHz is the leakage of return path signals into the IF section of TVs connected directly to the cable.

Designing Multi-Service Cable Networks

Marty Glapa, Chia-Chang Li, Amit Mukhopadhyay, Cezary Purzynski, & Bill Stenger
Lucent Technologies

Abstract

How to manage access bandwidth to accommodate multiple services with diverse bandwidth requirements, while meeting performance objectives and maximizing revenues/profits is an area of great interest in the cable industry. We have developed a network engineering tool for efficient and effective multi-service cable network design. In addition, the tool can be used to effectively evaluate the impact of service definitions and growth scenarios on bandwidth utilization and performance. In this paper, we discuss the principles of multi-service network design, review some performance considerations, and present the capabilities of the network engineering tool with example results.

1. Introduction

Today's cable services are primarily entertainment based but there is a rapid expansion underway into high-speed data and telephony. The traditional, and still the primary business of the cable industry is one-way broadcast television service delivering up to about 80 channels with more on the way. Pay channel subscriptions that are broadcast over the network have been successfully selling monthly subscription services for about the past 25 years. Pay Per View works similar to the pay channel subscriptions except there is a billing reporting mechanism – often done by polling the subscriber's set top box by telephone or a store and forward RF system using two–way compatible plant. Modern addressable set top boxes have allowed cable operators to scramble analog channels and offer them as a tiered service. MPEG-2 (Moving Picture Experts Group) compression technology and QAM (Quadrature Amplitude Modulation) have enabled the cable operators to offer broadcast digital channel with tiered services, which too are gaining in popularity. A true departure from traditional entertainment video is high-speed data with well over 1 million subscribers in service. It allows a subscriber with a cable modem to receive and transmit high-speed data at rates far beyond those of traditional analog modems.

There are many competitive pressures on cable companies today that can effect the set of core services that are offered. To help combat the competitive pressures, cable operators are looking to new future services to retain and attract customers. These services include a mixture of video services, telephony, and multimedia services. Narrowcast ad insertion could be targeted to specific subscribers based on an individual customer's profile. Circuit based telephony, while being deployed today in some markets, will be quickly replaced by IP based telephony. IP telephony leverages a common logical layer infrastructure (the IP layer) for both voice and data services – resulting in lower capital and operations costs. Video on demand is surfacing as a narrowcast service that allows individual subscribers to order, and watch a movie or event. Work-at-home, and ideal service for an IP based infrastructure, extends LAN and PBX functionality to a workstation in an employee's home. Home security monitoring and energy monitoring/management services can be rather easily deployed using an HFC infrastructure. And, streaming video, video conferencing, video telephony and interactive gaming can open new opportunities for services delivered using an IP data infrastructure.

Each new service will have its own set of bandwidth requirements. These new services will compete for bandwidth in both up and downstream directions on cable plant. Managing this bandwidth, especially those services enabled by a common logical IP layer, to maximize efficiency and minimize cost is of great interest to the industry and is the topic of the remainder of this paper.

2. Service and Traffic Scenarios

As indicated in the previous section, cable networks today support a limited set of services, i.e., broadcast video, Pay Per View, Video On Demand, data services, and, in some cases, circuit voice services. These services typically occupy different part of the spectrum. They coexist on the same HFC plant but are engineered and provisioned separately.

The introduction of DOCSIS [1] brought about an Internet Protocol platform on cable networks that enables multiple service offerings. Internet access services have been widely available in the many areas in the United States and many other countries. Voice over IP (VoIP) services have been in trials while the technology and PacketCable specifications are being developed and implemented. The first service scenario we consider includes VoIP and basic Internet access services.

In addition to voice and basic Internet access services, the potential of new services for the cable operators is enormous. Cable operators can provide different grades of service to create tiers of services. Different tiers of Internet service could be associated with different bandwidth guarantees, maximum delays, and maximum packet loss probabilities. Massive bandwidth available on cable will open up the opportunities of many new services, e.g., streaming video, interactive gaming, and VPN connectivity for work at home. Table 1 shows the typical set of traffic parameters associated with some of the new services.

3. Multi-Service Network Design Principles

The key to designing any efficient network is a good model of bandwidth requirements. Multi-service networks provide the additional challenge of estimating bandwidth requirements for several services. These services may have to satisfy multiple performance criteria and they can also have different busy hours – overlapping or non-overlapping.

Let us take the simple example of designing a residential voice and data network over the cable infrastructure. An important performance parameter for voice service is probability of blocking while for data service a key criterion is throughput. First, let us

Services	Bandwidth Requirement	On-Line Ratio	Active Ratio	Busy Hour Erlang
G.711 Voice with PHS - 2 way	121.6 kbps	N/A	N/A	0.346
Basic Data Service		60%	40%	N/A
Downstream	400 kbps			N/A
Upstream	160 kbps			N/A
Video Telephony - 2 way	400 kbps	30%	100%	N/A
Streaming Video				N/A
Downstream	1.5 mbps	25%	100%	N/A
Upstream	0			N/A
Interactive Gaming		25%	80%	N/A
Downstream	500 kbps			N/A
Upstream	100 kbps			N/A
SOHO - 2 way	1.5 mbps	100%	80%	N/A

Table 1: Example Application Bandwidth Requirements

assume that there is a common busy hour for both the services – 8:00-9:00 p.m. The network is loaded to its limits for both signaling as well as bearer traffic during this hour. Proper network engineering will require ensuring the desired blocking probability for voice traffic and providing the required throughput for data traffic. Secondly, let us now assume that the busy hour for voice is 7:30-8:30 p.m. and the same for data is 8:30-9:30 p.m. In this scenario, the design needs to meet the blocking criterion during the voice busy hour and the throughput criterion during the data busy hour. Lastly, let us assume that voice busy hour is from 7:30-8:30 p.m. and data busy hour from 8:00-9:00 p.m. In this case, the design needs to meet blocking criteria 7:30-8:00 p.m., blocking and throughput criteria 8:00-8:30 p.m. and throughput criteria 8:30-9:00 p.m.

The above is probably the simplest example of a multi-service network – only two services with one performance criterion each – and yet it poses a formidable challenge already to the network designer. In the very near future, cable operators will provide a combination of voice, data and video services to a mixture of residential, telecommuting, business, SOHO and other types of customers over a common platform. Each of the services may also be offered in different tiers, such as silver, gold and platinum classes. It is obvious that the solution space is multi-dimensional and the possible number of dimensions could grow out of control if it is not properly constrained by certain network planning and design framework.

In order to properly bound the problem, we first define as the building blocks several basic services, and, for every basic service, we assign certain traffic parameters based on a collection of expert opinions and market data. Note that these traffic parameters are part of the modeling tool inputs and can be varied to fit different service assumptions. Several basic services then can be grouped into a service package. Growth or service penetration rates then are associated with the service packages. Finally, we compute the weighted averages of the parameter values across all services for the network. This approach can be used in parts of a network (e.g., the region served by one Fiber Node) or to the entire network (e.g., a region served by one or more Head Ends). Figure 1 shows sample upstream and downstream bandwidth requirement at a Head End with respect to varying market sizes and service penetration rates.

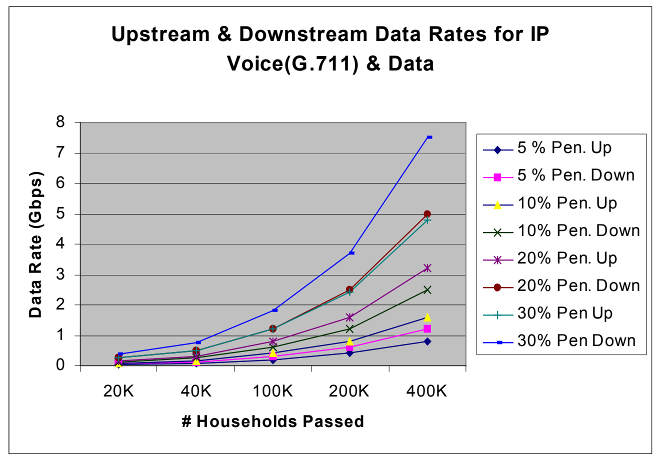


Figure 1: Voice and Data Bandwidth Demand

The basic concept of busy hour is still valid, but it has to be interpreted in the proper perspective. We no longer look into busy-hour traffic data for individual services; what we need, is to establish the busy hours for the aggregate bandwidth demand. Today, we have extensive data on voice bandwidth demand and have some measurements of Internet data traffic demand. As other types of data usage increases and as multimedia traffic grows, there is an increased need for keeping track of subscriber bandwidth requirements.

Once the basic traffic requirements are established, the next step is to set the proper performance objectives. These performance criteria will typically be different for various types and classes of service. The challenge is to engineer a network that provides the required bandwidth while satisfying the required performance criteria.

4. Performance Considerations

In order to capitalize on the promise of converged networks, it is necessary to develop models and network design techniques that allow the network designer to engineer a single infrastructure capable of supporting multiple services. These models and techniques have to be flexible enough to give services a degree of isolation from one another in terms of network characteristics to meet service-specific requirements. The designer takes into account all of these requirements in the form of delay, delay jitter, bandwidth requirements, BER, packet loss rates etc. By evaluating the performance impact of various engineering the parameters of the MAC and IP layers as well as choice of QoS schemes and proper network sizing attempts to satisfy the requirements of all services. The first step is to understand the acceptable levels of performance required by each service and identifying the major factors impacting performance. Until very recently, the Internet access over cable was the predominant, non-traditional service offered by the MSOs. The service requirements were very loosely defined. For both delay and packet loss, the general understanding was that less is better, but no specific targets have been established. This very loose definition of service levels has led to under-engineered networks and unacceptable performance levels as the subscriber population grew. Planned introduction of voice services changes that picture entirely. Voice services, and primary line services in particular, are characterized by very strict delay, delay jitter and packet loss requirements. Legacy services, such as fax and voice band data, place even stricter requirements on those performance measures. Below we discuss the major factors affecting voice and data performance in networks with HFC access. Figure 2 shows the major architectural components of such a network.

4.1. Delay

As mentioned above, voice services place very stringent delay requirements. These follow from the desire to maintain PSTN levels of voice quality in the VoIP services in the cable environment. The general requirement is that the bearer channel delay should be less then 300 ms round trip. It is widely supported by perception studies and standards. Primary line service is also characterized by strict signaling delay requirements, driven by two factors. The first is driven by customer perception of what constitutes good service and second, maybe even more important, is the desire to maintain all of the optional features (call waiting, caller ID etc.), some of which have strict timing requirements. In the following we identify the end-to-end delay components, with particular emphasis on bearer traffic delays.

4.1.1. Delay in Multimedia Terminal Adapter/Cable Modem (MTA/CM)

MTA performs A/D conversion and packetization of the speech samples. The delay introduced by both is driven by the choice of codec. All codec's have an associated block size i.e. the number of speech samples that need to be accumulated before the codec can process them. The block size ranges from the low of 1 sample (i.e. a delay of 125 microsecond for 8000 sample/s) to a high of 240 samples (or 30 ms) for G.723 codec. Some codec's also require a look-ahead time, which can be as high as 7.5 ms for G.723. The packetization time has to be an integer multiple of block size and therefore its theoretical lower limit is a single block size for the codec of choice. However, one has to consider the issue of bandwidth efficiency and the longer the packet in terms of bytes, the smaller is the overhead associated with packet headers and physical overheads. In some sense we have to strike a balance between the delay and bandwidth efficiency.

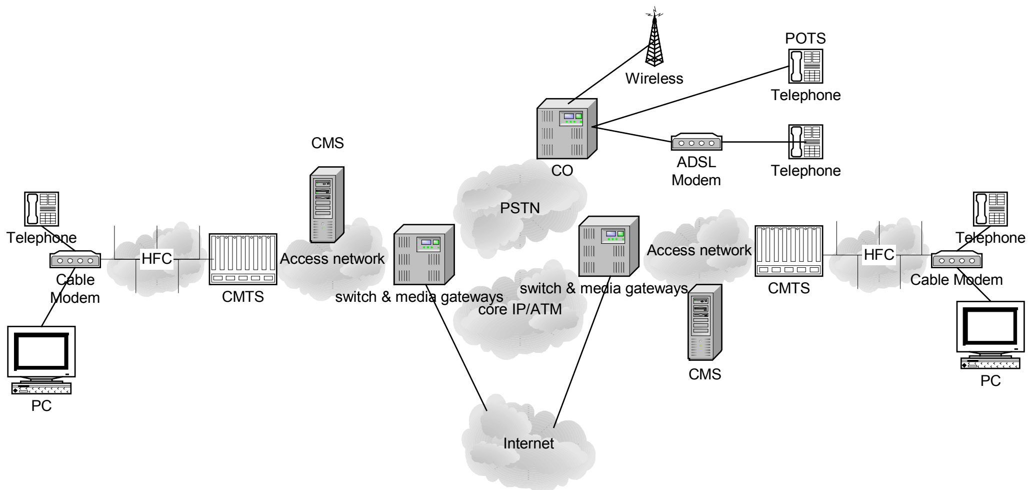


Figure 2: A converged network with ingress cable access and various possible egress networks

The additional factor to consider is that the more efficient the codec i.e. it generates smaller number of bytes per unit time of speech, the higher (in percentage terms) is the overhead, for a fixed packetization time. We will come back to that subject later. It suffices to say here that the "optimal" packetization time ranges someplace between 5 and 20 ms with 10 ms being commonly used. Additional delays are incurred at the MTA for packet processing (i.e. header generation etc.).

For data traffic the estimate is much simpler. As to data traffic, the CM is just a bridge/router with its well understood delay sources.

4.1.2. Delay in HFC Plant

The HFC plant delay includes not only the propagation delay (which is negligible) but also the Medium Access scheme. The switch from data oriented DOCSIS 1.0 to integrated services DOCSIS 1.1 has become the enabler of real time services. For real time services, Unsolicited Grant Service (UGS) is the preferred MAC access method. UGS service grants periodic, fixed size transmit opportunities to established CBR flows. In order to minimize the access delay an MTA can synchronize its packet generation time with the expected grant arrival time and hence not incur any delay (other then a small safety margin). Should the MTA choose not to implement grant synchronization, the MAC access delay becomes a random variable (uniformly distributed) in the ranging from 0 to 1 packetization time, as the packet might be generated at any time between successive transmit opportunities. That delay is random only from call to call and should remain constant for the duration of a particular call. The picture becomes much more complex, when multi-line MTA is considered, where the processor has to perform a juggling act between encoding and decoding operations on multiple connections and still try to keep these operations synchronized with the transmit opportunities.

4.1.3. Delay in CMTS and IP Interconnecting Network

Once the voice packet reaches the CMTS successfully, it will be treated preferentially, as voice packets carry very high ToS or DS marking. The delay therefore is negligible, as the only traffic a voice packet competes against is that of other voice packets. A simple estimate, based on the relative sizes of the ingress and egress interfaces and the packet size, shows that delay in the worst case is no more than 0.5 millisecond. Same conclusion remains true (and actually improves) throughout the IP interconnecting network, as the speed mismatch between ingress and egress is the worst at the CMTS. One should however emphasize that this delay is highly dependent on the interconnecting network architecture.

4.1.4. Delay in Gateways and Core Networks.

Once the traffic emerges in the primary headend, it can take two distinct paths, depending on the called number location and more importantly, transport network that a service provider has at its disposition. As of today, there are few QoS enabled (a.k.a. managed) IP backbones in existence as the technology to build them is not yet mature enough. Therefore, for the sake of this analysis, one should primarily consider two choices. A direct ATM backbone connection and PSTN hop-off. The delay analysis of the ATM backbones is relatively easy, as they are connection oriented (i.e. allow for traffic engineering), QoS enabled with call admission control. Under such a set of conditions, one can easily predict (and guarantee) the delays, which for a backbone have two major components: switch delays and propagation delays. The picture is much more complex, when a hop-off to PSTN is needed. Hopping off to the PSTN entails restoring original packet spacing, commonly referred to as dejittering, and playing the voice packets out. Playing out involves restoring packets into a TDM structure, an operation that ranges in complexity from simple recovery of individual bytes from a packet for G.711 to a full decoding operation for compression codec's. The dejitter buffer will delay packets by a fixed amount that is programmed based upon the amount of delay variation experienced by packets form the same flow on their way from the MTA to the PSTN gateway. The less jitter, the smaller the dejittering delay. This is where proper network design in terms of sizing and QoS will have a major effect.

The propagation delay through the PSTN (or any network, for that matter) is impacted by the facilities length, a factor related but distinct from the distance between the end points. It takes into account the fact that facilities do not necessarily follow the shortest path, as well as the fact that for reasons ranging from traffic engineering to protection switching the even the shortest facilities path may not be used. The correction factor of 1.3 to 1.8 is commonly used. Once the packets arrive at the egress network attached to a cable plant, all the processes the packet went through on ingress, i.e., packetization at the gateway, dejitter, and play-out at the terminating MTA will be repeated. The only difference is that now we send bits downstream on the cable plant. The transit delay is negligible, with the biggest component being the interleaving delay.

4.2. *Delay Jitter*

The delay jitters have several components. The first occurs right at the MAC access where the transmission opportunities can be scheduled within a certain window. This picture becomes much more complex if VAD is used, which employs UGS with poling. This access method works similarly to the UGS service, except that when silence is detected the flow of grants is stopped and resumed again when speech is detected. There is a delay associated with restarting the flow of grants and that additional delay will impact the first packet of the talk spurt. Additionally, the CMTS upstream MAC scheduler might not be able to schedule the transmission opportunities for the flow at the same position as previously. This will add to jitter. Furthermore, each router, including the CMTS, can experience varying levels of congestion and hence packets of the same flow waiting for an access to an outgoing link will experience variable delay through each router. Similarly, the egress (CMTS downstream) link buffers and scheduling algorithm at the CMTS should be engineered carefully to avoid potentially excessive jitter.

Since the packets have to be played out evenly spaced, dejitter buffers are used at end points to properly re-space the packets. The amount of time packets is placed in that buffer before they are played out should be slightly larger than the maximum expected jitter, otherwise packet loss will occur. A buffer that is too large leads to an unnecessary additional delay. The upper bounds on jitter can be estimated by adding all jitter elements mentioned above. Jitter in routers can be estimated by calculating the maximum number of packets from competing voice connections. Finally, it should be pointed out that jitter can become a component of constant delay if the first packet (its time of arrival establishes the play-out reference point) incurs maximum jitter.

Generally, the requirements for data are not as strict as they are for voice. It, of course, depends on type of data. Interactive gaming has actually stricter delay requirements than voice. Other types of data, such as e-mail, are practically insensitive to delays. But most of data, as represented by Internet traffic or NCS signaling traffic, do not have to be delivered to the destination with accuracy of several (tens) of milliseconds and virtually no jitter. However, one should take into account that many of these types of traffic are riding on the TCP protocol, which can be profoundly effected by long delays, excessive jitter and packet loss. All of these can throw the TCP behavior out of balance and limit its throughput disproportionately to the incremental impairment.

4.3. <u>Bandwidth Efficiency</u>

It is well known that IP is extremely bandwidth inefficient for short packets, such as voice packets. As discussed previously there exists a trade-off between delay and bandwidth efficiency. The headers associated with bearer traffic are RTP/UDP/IP headers, Ethernet MAC and DOCSIS MAC. Additional overhead is incurred by FEC and mapping of packets into minislots. Finally, there is physical layer overhead associated with guard time and preamble. As some of these overheads are unavoidable, in particular given the harsh conditions of the HFC plant, others are constant form packet to packet and can therefore be suppressed. The mechanism of the PHS can be used to suppress the packet header constant fields in particular the entire UDP, IP headers and nearly entire Ethernet header in the upstream direction can be suppressed. It is accomplished through a mechanism PHS mask (byte wise mask of suppressed fields) and telling the CMTS (or the CM) the values of the constant fields so it can restore them. The header suppression technique is not in general available in the IP backbone network. The overhead becomes more pronounced as the number of bytes in the packet gets smaller, as is the case when one shortens the packetization time and/or deploys a more efficient coder/decoder (codec). PHS is not used for data packets.

4.4. <u>Packet Loss and FEC</u>

In order to prevent excessive packet losses (voice packets are particularly sensitive, as they cannot be retransmitted), one can protect the packets with an error correcting code. The more extra bytes are added, the more errors can be corrected. However, it also results in lower bandwidth efficiency. The issue of FEC depth is of particular importance in the upstream direction where interleaving cannot be used. The FEC depth required is related to the noise characteristics in the HFC plant. In the high ingress levels will contribute to lower bandwidth efficiency, higher bit error rates and packet loss rates, which for voice services will lead to deteriorating voice quality. In general, losses in the range of 0.1% - 1% are service effecting, with the exact number highly dependent on the codec type used. In order to mask the impact of packet losses on voice quality, an error concealment method should be implemented in the voice decoder. For voice-band data (VBD) and fax, the requirements are even more stringent (0.01%

packet loss), and failure to achieve them might lead to unusable service. It should also be pointed out that fax and VBD services usually require G.711 coding, which is less bandwidth efficient, as compression codec's do not reliably transfer the analog signaling. Reliability of DTFM tone transmissions in compression codec's is of concern as well.

5. Lucent Cable Network Engineering Tool (LuCNET)

Lucent's Cable Network Engineering Tool (LuCNET) has been developed to minimize equipment cost while still satisfying necessary performance criteria. The input to the tool comprises traffic demand, network topology and service definitions (service area characteristics) and performance objectives.

The user can specify Head End (HE) and Distribution Hub (DH) location on a map and input traffic parameters on an interactive input screen. The tool generates Customer Premises Equipment (CPE) to DH bandwidth estimate. It then uses various in-built engineering rules and performance objectives to determine various equipment configuration. Once the tool is run, the user can view graphically a schematic representation of the network, sizing of inter-connecting links, and equipment sizing at each node. Numerous sensitivity analyses charts can also be generated. Network costs for various configurations can be determined interactively, when the cost data is available. The cost elements can be fed into a business modeling tool in conjunction with revenue and other relevant data to generate various financial reports. A representative process flow chart for the tool is presented in Figure 3.

6. Example LuCNET Results

LuCNET is highly flexible to generate different types of results that may be of interest to a network designer. The basic output comprises a schematic network topography connecting the Distribution Hubs and Headends in a single or hierarchical dual ring structure. The tree structure connecting the Distribution Hubs to the Fiber Nodes and to the homes is also captured in network diagrams. Links are shown towards the PSTN, SS7 and Data networks. By clicking on one of these links, one can see the type of facility (e.g., DS3, OC3…), the number of facilities (e.g., 1, 2…) and the loads of these links (e.g., 65%, 80%…).

Clicking on the nodes can show details of the nodes on the diagram. They will show what equipment and how many of them are there (e.g., Cable Modem Termination System, Call Management System, etc.) By further clicking on the equipment, one can also see the detailed equipment configuration (e.g., number of chassis, plug-in cards, etc.)

Some of the most interesting outputs from the tool are the various sensitivity analyses charts. Figure 4 shows the sensitivity of cost per subscriber with respect to market size and take rate. The x-axis shows different market sizes (with respect to the number of households passed) and the y-axis represents normalized cost per subscriber. Each curve corresponds to a particular take rate. As expected, cost per subscriber goes down as the market size and penetration increases.

Figure 5 shows the sensitivity of the total network cost (normalized) with respect to different market sizes and penetration. On the x-axis, each bar corresponds to a market size (HHP – Households Passed) and a penetration value. The y-axis represents normalized cost for the entire network. The top part of each bar represents the cost contribution from the CMTS while the bottom part represents the cost contribution from the CMS.

Presented in this paper are only a few examples of sensitivity analysis charts of common interest. The tool can create numerous such figures based on individual needs of the users.

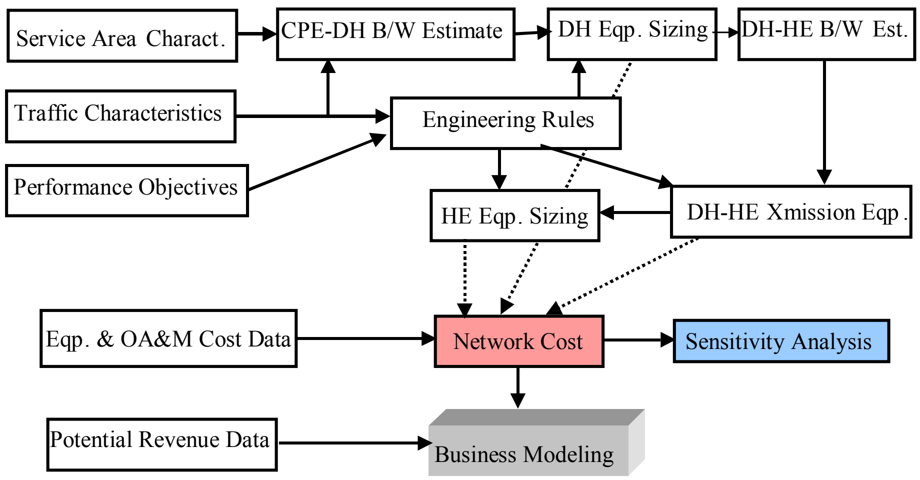


Figure 3: Network Modeling Process Flow

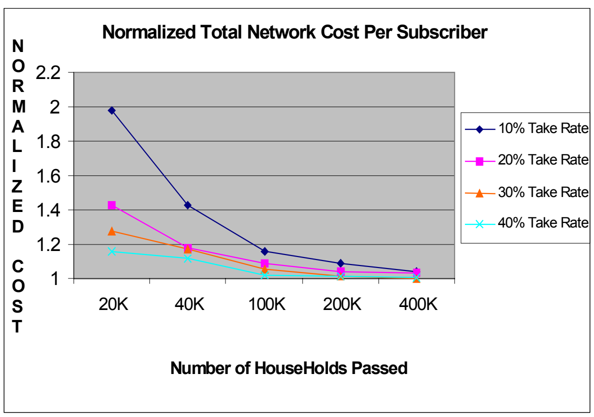


Figure 4: Cost Per Subscriber Sensitivity vs. Network Size and Service Penetration Rate

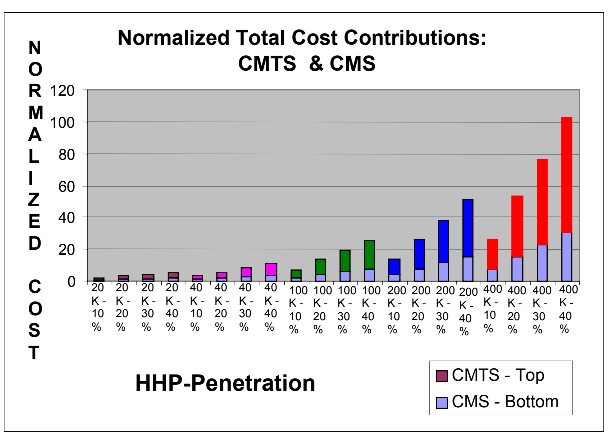


Figure 5: Normalized Cost Contribution by CMS/CMTS

7. Concluding Remarks

Cable access networks present a special challenge to network planning and design. Sharing of the network resources extend from the core to the access, which traditionally is dedicated to a single user. With the enormous amount of bandwidth on cable access, many new services will become available. These new services will further complicate the tasks of designing an efficient access network while meeting QoS objectives of all the services. In this paper, it is shown that, with appropriate design tools and good understanding of service characteristics, a cable access network can be built efficiently and effectively to offer multiple services on an IP platform.

8. References

[1] Cable Television Laboratories, Inc., Data-Over-Cable Service Interface Specifications.

Contact Author Name: Chia-Chang Li
Mailing Address: 4K430, 101 Crawfords Corner Rd., Holmdel, NJ 07733
Telephone Number: (732)949-3101
Facsimile Number: (732)949-7425
Email Address: chiali@lucent.com

DWDM Block Converted Return Transmission Performance

John J. Kenny
Mark A. Linford
ANTEC Corporation

ABSTRACT

Block conversion and dense wavelength division multiplexing (DWDM) have each been used to concentrate multiple independent return path signals onto a small number of fibers. The combination of these technologies results in very fiber-efficient return path transmission.

Performance evaluation of a return path system using these combined technologies could require a prohibitively large number of independent signal sources to load the system. Therefore, a loopback technique has been developed to minimize the amount of test equipment required for transmission testing.

This technique was used to characterize a DWDM block converted return path link carrying 72 independent 5 to 42 MHz return signals.

INTRODUCTION

The number of return bands requiring independent transport from a hub back to the headend can be quite staggering. A hub may serve 50 or more nodes and often these node returns must be subdivided to support high take rates for services. This leads to a situation in which conventional return techniques require a large number of fibers, thereby requiring technologies to reduce fiber count.

Reducing the hub to headend fiber count has several advantages: It lowers cable costs in new installations; it can avoid additional cable installation; it can free up fibers to allow redundant fiber path diversity; and it reduces the time to fusion splice fibers after a cable cut.

Block conversion and DWDM have each been used independently to concentrate returns from the hub to headend. Block conversion by itself allows up to eighteen, 5 to 42 MHz, or twelve, 5 to 65 MHz, returns to be carried in the traditional forward band frequency range of 45 to 870 MHz[1]. Initially, the DWDM return path architecture used eight wavelengths[2], while early installations generally equipped no more than four. Recently reported work[3], along with experimental verification, derived optical crosstalk design rules for analog DWDM systems with up to 32 optical channels. Using both of these technologies together further economizes fiber usage. We will eventually be able to carry at least 18 x 32 = 576 independent 5 to 42 MHz return path signals on a single fiber, using externally modulated DWDM transmitters.

TEST CONFIGURATIONS

The testing performed on DWDM with block conversion consisted of noise power ratio (NPR) measurements and dynamic range for "error free" 16 QAM transmission from a cable modem to a cable modem termination system (CMTS).

Based on availability of hardware, this testing was limited to 8 wavelengths and we used directly modulated transmitters. Since directly modulated optical transmitters carried by standard single mode fiber create a significant amount of second order distortion, the range of block converted signals was restricted to less than an octave, i.e. each transmitter was loaded with 9 blocks instead of 18. Thus, this verification testing was done for 72 independent returns.

Noise Power Ratio

During early block conversion NPR testing, we learned the importance of having independent signal sources to characterize dynamic range. Initially, we simply split the output from a noise generator to drive multiple upconverters. The measured NPR dynamic range was far worse than predicted.

have 72 independent signal sources, 72 up and down converters, 8 transmitters and receivers, etc. This leads to a very complex test configuration. We have found two effective means for generating decorrelated test signals from a single noise generator. One is to put the output of a 1 GHz noise generator into a bank of downconverters. Each downconverter samples a different portion of the noise generator's output spectrum and frequency shifts it down to 5 to 42 MHz. The second technique is to introduce time shifts between the upconverter input signals. A combination of these methods is used in the test configuration block diagram in Fig. 1. To conserve the number of up and

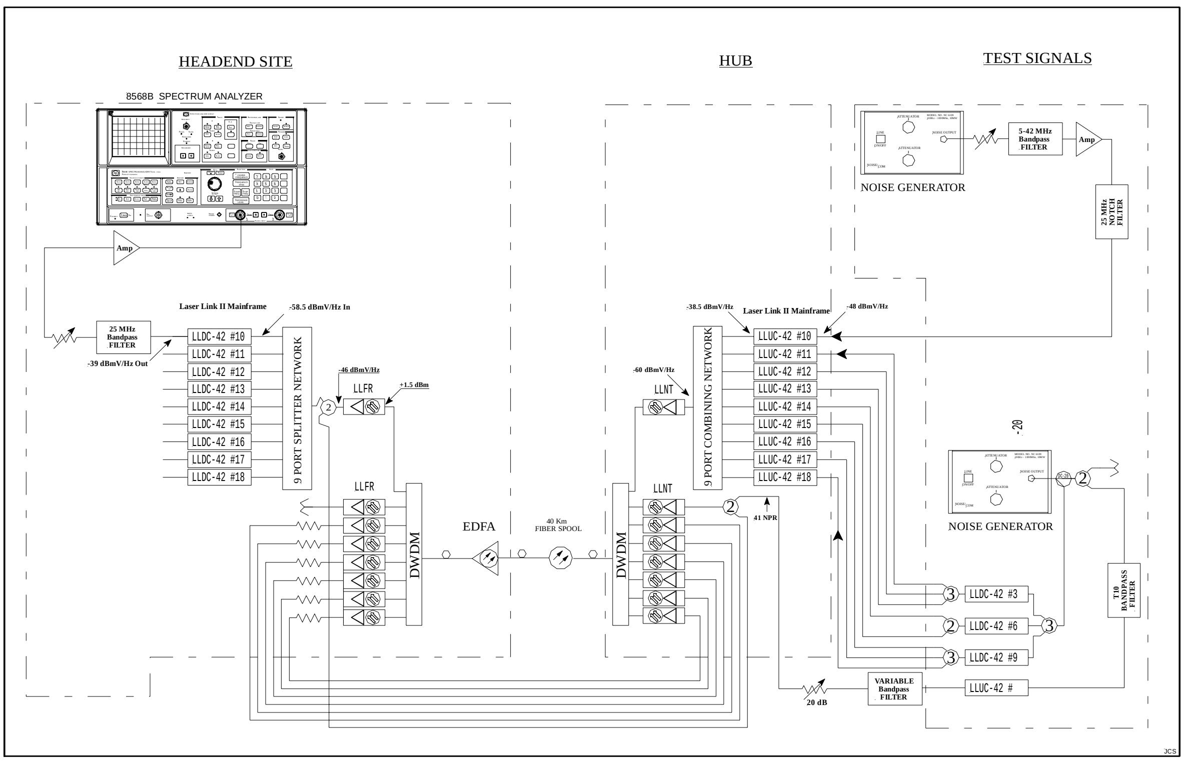


Figure 1 Test Configuration for NPR Testing of DWDM with Block Conversion

The correlation of the input signals was the culprit. Once we decorrelated the test signal sources applied to each upconverter, the measurements agreed with expectations. Therefore, to test this 9 block, 8-wavelength return transport system, one would like to

down converters, we recycled a receiver output from one path to the input of a DWDM transmitter. Again, the 40-km of fiber provides more than enough delay to decorrelate the signals loading the International Telecommunications Union (ITU) grid DWDM directly modulated transmitters.

Figure 1 shows the path under test as the top upconverter/downconverter pair and top optical transmitter/receiver pair. The signal for the path under test comes from an automated NPR test set. The inputs to the other upconverters come from the split outputs of three down converters. More downconverters could have been used so that all input signals are uncorrelated, so the test results may be somewhat penalized by taking this short cut. Looping back a receiver output to the next transmitter's input derived the RF inputs for the other seven DWDM transmitters. There is just one difficulty with doing this: Unless something intentional is done, there will be no crosstalk contribution to the measurement. At the notch frequency the inputs to all transmitters will be low in level and hence there will be no measurable crosstalk. To correct for this problem, a narrow band of noise was inserted into the noise slot for the second and subsequent optical paths.

DOCSIS Return Path

The second area of testing used two RCA DOCSIS compliant cable modems purchased off the shelf at a retail store and a model 1000 CMTS from Arris Interactive. We tested the dynamic range of the transmission path to demonstrate the amount of headroom relative to level changes.

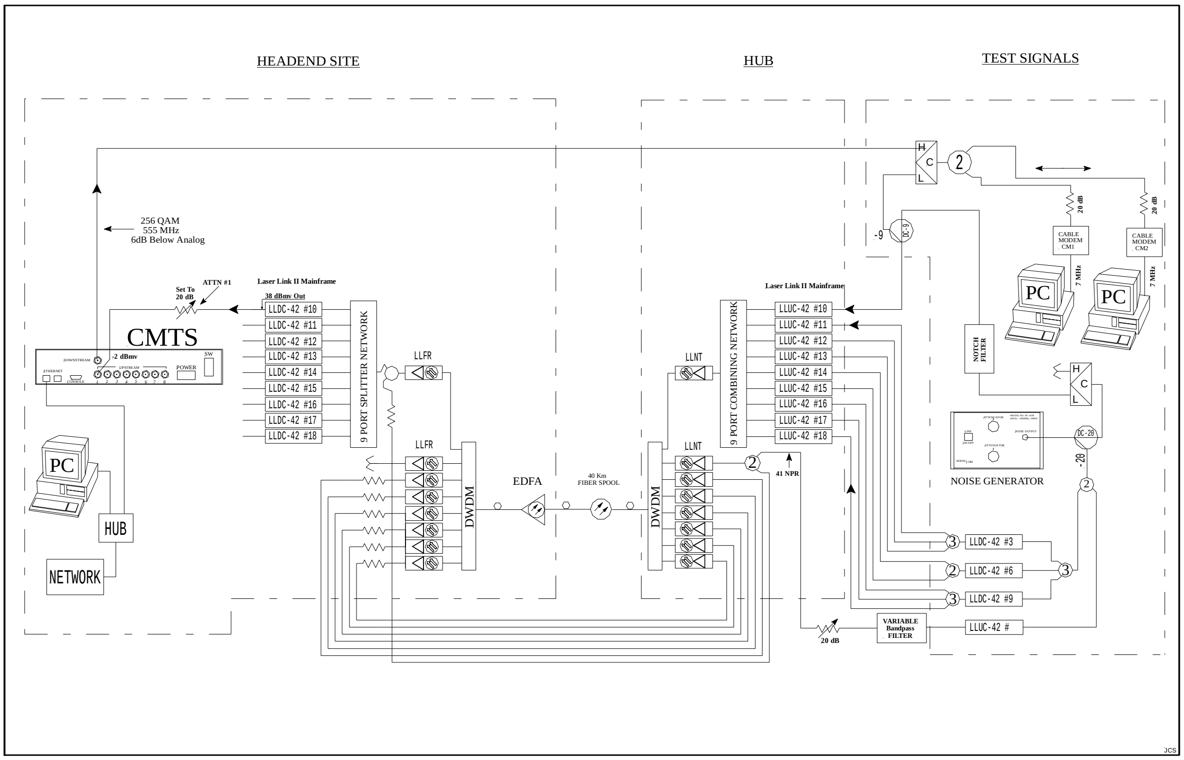


Figure 2 Test Configuration for Cable modem Dynamic Range

The block diagram of this test configuration is shown in Fig. 2. The CMTS generates a 6 MHz wide 256 QAM signal at 555 MHz. This signal is "dropped" to two cable modems through a diplexer and an RF splitter. The 3.2 MHz wide, 9 MHz, 16 QAM return signals from the cable modems are directed to the return path transmission equipment by the low pass section of a diplexer. A 5 to 40 MHz band of noise is notched at 9 MHz, combined with the cable modem signals and then it is inserted into the upconverter of the path under test. See Fig. 3 for a plot of the upconverter's input signal. Eight other block converters are noise loaded, and then the set of nine block converters is combined to drive a 1550 nm directly modulated

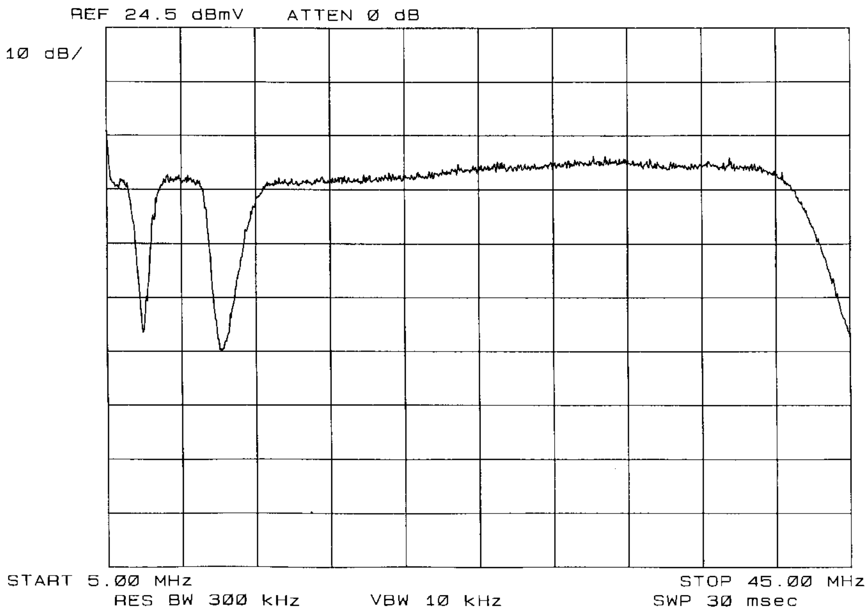


Figure 3 RF Input to Upconverter with Cable Modem Signal and Noise Load

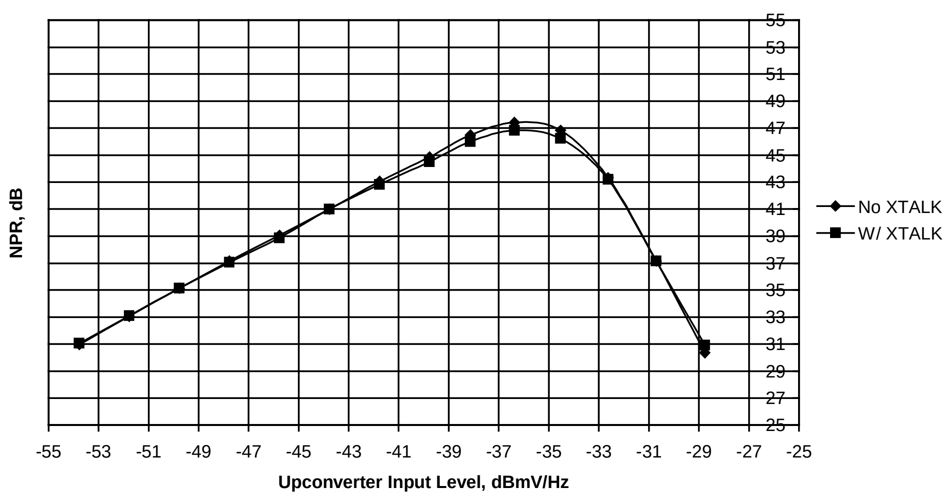


Figure 4 Noise Power Ratio, with and without Crosstalk

transmitter. Eight transmitters are optically multiplexed and sent over 40 km of standard single mode fiber. At the headend, they are optically amplified before being separated into individual optical signals in a DWDM. The optical receiver output with 9 return bands is split 9 ways. Then the bands are downconverted to 5 to 40 MHz and connected to the CMTS return input.

TEST RESULTS

NPR

The NPR tests were conducted in three blocks: 10, 15 and 18. The measured NPR dynamic ranges for several values of NPR are shown in Table 1:

NPR, dB	Block 10	Block 15	Block 18
25	29.3	31.3	30.8
30	24.5	26.5	26.2
35	19.3	20.9	19.9
38	16.3	17.1	15.6
40	14.3	14.5	13.1

Table 1 NPR Dynamic Range for DWDM with Block Conversion

A previous estimate of the NPR performance for an 8-wavelength system using directly modulated DWDM transmitters was 15-dB dynamic range at a noise power ratio of 38 dB[4]. The impact of crosstalk interference is seen in Fig. 4 by comparing the NPR curves with and without modulation on the other 7 DWDM optical transmitters. It can be seen in that figure that the influence of crosstalk is negligible for NPR values less than about 41 dB.

Cable Modem Dynamic Range

The block converter input signal, composed of the cable modem return signal and the notched noise, was varied in level to determine how sensitive the end to end DWDM with block conversion system is to input level. As shown in Fig. 2, the CMTS is connected to a PC to display data packet errors. The dynamic range as used here is based on the limits of signal level providing error free transmission of the 10 Mb/s, 16 QAM signal.[5] The test time is based on sending a 50-MB file multiple times with zero correctable bit errors, so the bit error rate is about 10^{-9}. As a worst case scenario, we used block 10 for this test (see Table 1). The nominal level at the input to the upconverter was set to be the noise density that produces approximately 41 dB NPR. We were able to raise the cable modem signal and band of noise by 16 dB in 1-dB steps or lower them by 13 dB in 1-dB steps without encountering bit errors during a file transfer. Thus we observed at least a 29-dB total dynamic range.

Comparison of NPR and Cable Modem Dynamic Range Measurements

Theoretically, 16 QAM reaches a bit error rate (BER) of 10^{-9} at a signal to noise ratio of about 22.5 dB. Practically, a modem will produce that BER at a slightly higher signal to noise ratio. Table 1 indicates that the dynamic range for an NPR of 25 dB is about 29 dB. This is show good agreement between NPR and BER dynamic range performance.

CONCLUSIONS

The measurements confirmed that the impact of optical crosstalk is negligible in block converted DWDM return band transmission. Furthermore, the performance correlations among predicted, measured and actual QAM bit error rate are very good.

ACKNOWLEDGEMENTS

The authors thank Melissa Catlin and James Street for their assistance in performing measurements, collecting data and preparing figures for this paper.

Author Information

John J. Kenny, Ph.D.
Principal Engineer
ANTEC Corporation
11450 Technology Circle
Duluth, GA 30097
Tel: 678.473.8127
Fax: 678.473.8040
john.kenny@antec.com

Mark A. Linford
Staff Engineer
ANTEC Corporation
11450 Technology Circle
Duluth, GA 30097
Tel: 678.473.8037
Fax: 678.473.8040
mark.linford@antec.com

---

[1] Kenny, J. J., "High Dynamic Range Multi-Band Block Conversion for Return Band CATV Transmission," ECC`99, http://www.antec.com/white_papers.html.
[2] Sniezko, O. J. & Werner, T. E., "Invisible Hub or End-to-End Transparency," 1998 NCTA Technical Papers, pp. 247-257.
[3] Atlas, D. A., "Nonlinear Optical Crosstalk in WDM CATV Systems," RF Photonics for CATV and HFC Systems, LEOS STM, San Diego, CA, July 1999, pp. 23-24.
[4] Ghuman, H. J., Kenny, J. J., and Vella, E, "Getting the Most out of Your Return Paht," CED, December 1999, pp. 86-90.
[5] Farmer, J. O. and Linford, M., "Comptibil-ity Issues in Block Conversion for Return Path Concentration," Cable 2000 NCTA Technical Papers.

Enhanced Broadcast Services with Complimentary Delivery Over IP

Jan van der Meer
Philips Consumer Electronics

Abstract

The increase of available bandwidth over IP, the decrease of costs of storage devices and the availability of new technology for coding and streaming of audio and video is creating more and more opportunities for new enhanced broadcast services. The paper discusses technologies that enable those services and examples of applications.

INTRODUCTION

One of the technologies that is expected to play an important role in enhanced broadcast services with complimentary delivery over IP is MPEG-4, an object oriented multimedia standard that offers a variety of tools. In addition to providing tools for coding of individual video, audio and graphical objects, MPEG-4 provides also facilities to compose an MPEG-4 scene from a set of such objects. The composition may vary in time and space, which is particularly powerful in the case of multimedia applications with a behavior that is dynamic and yet unknown at application launch, for example covering a life sports event.

To enhance an MPEG-2 based broadcast service, the MPEG-4 objects can be transported over the same MPEG-2 Transport Stream as the MPEG-2 audiovisual broadcast content. Typically, this will be done for objects that need to be transported to all viewers of the broadcast service. However, transport over IP is possible too, which may in particular be useful when transport is needed to one or more individual users only, for example in case of a personalized service.

The MPEG-4 content can be played back immediately, but it is also possible to store one or more of the objects on a storage device in the Set Top Box for playback at a later time, for example upon a command of the user. In summary, delivery and playback options can be exploited as deemed appropriate by the application.

This paper first provides an introduction on important MPEG-4 features for enhancement of MPEG-2 programs with MPEG-4 content. Next, it is discussed how to incorporate MPEG-4 features in MPEG-2 broadcast services. Finally, some example applications are presented.

MPEG-4 FEATURES

MPEG-4 is an ISO/IEC standard, defined by MPEG, the Committee that earlier developed the Emmy Award winning standards known as MPEG-1 and MPEG-2, which enabled a large variety of digital video applications for Compact Discs, Broadcast, and DVD. MPEG-4, their most recent achievement, is an object based multimedia standard, offering scalability and flexibility, in combination with a high coding efficiency over a large range of bandwidth.

Though there is some functional overlap between MPEG-4 and MPEG-2, MPEG-4 will not replace MPEG-2. At the bitrates and resolutions commonly utilized by MPEG-2 applications there is no improvement in coding efficiency and hence no justification to replace MPEG-2 by MPEG-4 tools. Instead, the MPEG-4 specification is designed to offer a set of new features that can be exploited to add value to existing MPEG-2 applications. In particular the object oriented approach of MPEG-4 enables the design of sophisticated multimedia applications. An overview of MPEG-4 features for video, graphics, audio and composition of scenes is provided below.

Video

The video tools are capable of coding natural textures, images and video. In addition, arbitrary shaped objects and transparency of video objects are supported. MPEG-4 video offers scalability and flexibility in combination with a high coding efficiency over a large range of bandwidth, from as low as 5kbit/s up to about 1 Mbit/s. This makes MPEG-4 especially suitable for video streaming applications in environments where the available bandwidth may vary.

Coding of natural images and video is achieved in a way similar to conventional MPEG-1 and MPEG-2 coding, using motion compensated predictive DCT technology, with a high level of flexibility with respect to input formats, frame rates, pixel depth and bitrates. When compared with MPEG-1 and MPEG-2, motion compensation technologies are enhanced to support very low bitrates. As a result, MPEG-4 is extremely efficient at very low bitrates, making the MPEG-4 standard the obvious choice for streaming media over the Internet.

In addition to the MPEG-1 and MPEG-2 capabilities, MPEG-4 supports arbitrary shaped objects and transparency of video objects. MPEG-4 also includes a dedicated coding scheme for textures and still images based on a very efficient and scalable zerotree wavelet algorithm. Furthermore the MPEG-4 standard supports some 2D and 3D modeling techniques, as well as synthetic objects, in particular human face and body animation.

Graphics

MPEG-4 defines a rich set of 2D and 3D graphical functions, largely based on VRML. For applications requiring low-complexity graphics, a 2D Graphics profile is defined. For transport of graphical data an efficient binary format has been specified that compresses VRML 2.0 data with a factor of typically 8 - 15. This format also allows for very low bitrate animations, typically at the bitrate of a few kbit/second.

Audio

The MPEG-4 audio tools are capable of coding speech and music over a wide range of bit rates and sampling frequencies. Three different types of codecs are defined. The lowest bitrate range is covered by parametric coding techniques; 2 - 4 kbit/s for speech with a 8 kHz sampling frequency and 4 - 16 kbit/s for music with a sampling frequency of 8 or 16 kHz. Speech coding at the medium bitrates between about 6 - 24 kbit/s uses Code Excited Linear Predictive (CELP) coding techniques. In this region, two sampling rates, 8 and 16 kHz, are used to support both narrowband and wideband speech, respectively. For bitrates starting below 16 kbit/s, and typically up to 128 kbit/s for stereo, Time to Frequency (T/F) coding techniques are applied, with sampling frequencies such as 8, 16, 24, 32 and 48 kHz.

To support audio streaming over Internet with a range of available bandwidth and a variable Quality of Service (QoS), MPEG-4 defined the Scalable Audio Profile. Using this profile, it is possible to increase the audio quality seamlessly when more bandwidth comes available and to decrease the audio quality gracefully when the available bandwidth decreases.

Scenes

MPEG-4 is an object oriented multimedia standard. In addition to providing support for the coding of individual video, audio and graphical objects, MPEG-4 also provides facilities to compose an MPEG-4 scene from a set of such objects. The composition may vary in time and space, which is particularly powerful in the case of multimedia applications with a dynamic behavior. Examples thereof are applications to enhance the broadcast of life sport events. During such life broadcasts, several situations may happen that are worthwhile to be reflected in an immediate change of the scene composition of the enhanced broadcast application. MPEG-4 allows for example to add and remove on the fly video objects simultaneously overlayed on the screen.

The necessary composition information forms the scene description that is constructed using the so-called BIFS (Binary Format for Scene description). Composition information is coded and transmitted together with the media objects. BIFS provides MPEG-4 with a rich set of scene construction operators, including the VRML graphics primitives that can be used to construct sophisticated scenes.

Delivery of MPEG-4 content over MPEG-2 Systems and over IP

MPEG-4 is an abstract standard that does not define transport mechanisms; for interoperable services additional specifications are needed to define transport of MPEG-4 data. The MPEG Committee defined how to carry MPEG-4 content over MPEG-2 System streams in Amendment 7 to the MPEG-2 System specification, while the IETF defines carriage of MPEG-4 over IP in a joint effort with the MPEG Committee.

ENHANCED MPEG-2 PROGRAMS WITH MPEG-4 CONTENT

Within MPEG-2 Transport Streams, the Program Map Table, PMT, is used to define which elementary streams form a program. Within the PMT also reference can be made to MPEG-4 content. The MPEG-4 content may represent an individual MPEG-4 elementary stream, or an MPEG-4 scene with one or more MPEG-4 objects.

Individual MPEG-4 elementary streams

An individual MPEG-4 stream may represent for example an MPEG-4 encoded complementary speech channel or a low frame rate video of a small size for overlay on full screen MPEG-2 video. Each individual MPEG-4 audio and visual stream is carried in PES packets. In the PES header, PTSs are encoded in the same way as for MPEG-2 elementary streams, based on the MPEG-2 System Time Clock. In this way the decoding and presentation of the individual MPEG-4 elementary stream is defined directly in terms of the MPEG-2 time base.

MPEG-4 scenes and objects

To identify the MPEG-4 objects that are to be composed into an MPEG-4 scene, a unique ID, the ES_ID, is assigned to each such MPEG-4 object. An MPEG-4 object may represent audio, video, text, graphics, or other content. An MPEG-4 object is not required to be MPEG-4 encoded. For example, also an MPEG-2 video or audio stream can be an MPEG-4 object.

MPEG-4 System streams

MPEG-4 audio and visual elementary streams can be carried directly in PES as individual MPEG-4 elementary stream. Next to this method, also MPEG-4 System tools can be used to carry MPEG-4 content over MPEG-2, in particular SL-packetized streams and FlexMux streams.

In MPEG-4 Systems, SL packets are the basic entity for carriage of access units. Each SL packet carries exactly one access unit or a part thereof. The header of the SL packet contains time stamps and other data for the contained access unit. A sequence of SL packets with data from the same elementary stream is called a SL-packetized stream.

The MPEG-4 FlexMux tool is capable of multiplexing SL-packetized streams into a FlexMux stream. A FlexMux stream consists of a sequence of FlexMux packets. Each SL-packetized stream in a FlexMux has its own FlexMux channel, identified by the FlexMux channel number coded in the header of the FlexMux packet.

MPEG-4 time base(s)

In principle, each MPEG-4 object has its own time base. Elementary streams carried in PES without the use of MPEG-4 System streams are locked to the MPEG-2 System Time Clock, STC, in the same way as any MPEG-2 audio or video stream that is part of the same program. MPEG-4 elementary streams that are carried using MPEG-4 System streams have a time base with the following characteristics :

- The object time base is locked to the MPEG-2 STC;
- There is a fixed time offset between the object time base and the MPEG-2 STC.

In this case the object time base is either carried by the SL-packet header or by a specific FlexMux channel. The time offset between the object time base is defined through the use of MPEG-2 and MPEG-4 time stamps. See amendment 7 to the MPEG-2 System specification.

Complementary delivery over IP

Once developed, an MPEG-4 application requires transport of the MPEG-4 content. If the streams are intended for broadcast to many clients, transport over MPEG-2 may be most suitable, but content that is delivered to a single client may be delivered more efficiently over IP. In any case both delivery methods are available as complementary options which can be exploited as appropriate.

Delivery of SL-packetized streams and FlexMux streams over IP is specified by IETF, in a joint effort with the MPEG Committee. MPEG-4 allows the content to be authored independently of the delivery method. The application requires that the content has been authored under the same conditions as for delivery over MPEG-2, and therefore the format of constructed SL-packetized and FlexMux streams is fully transparent to transport over MPEG-2 or over IP. However, applications should take into account that the end-to-end delay for delivery over IP may be significantly larger.

EXAMPLES OF APPLICATIONS

MPEG-4 Scene composition is a powerful tool to design multimedia applications. It specifies how video, audio and graphics and other objects relate in time and space. In this section four examples are given of multimedia applications designed to enhance digital broadcast services.

Audience attractor

Assume a Pay TV broadcast of a tennis event. Instead of the usual way of reporting from a tennis match with a broadcast of a single full screen video stream and a single audio stream, the broadcast is now composed of multiple objects :

- two tennis players as two foreground video objects;
- the tennis stadium as a background video object;
- an audio object with a commentary voice;
- the ambient sound in the stadium as a background audio object.

The PayTV coverage of the event is broadcasting the background objects, the tennis stadium with its ambient sound for free, to attract the target audience. Upon payment, the tennis players and the commentary voice are added to the background.

Personalized advertizing

Assume further to the previous example some advertisement boards as texture objects within the background object The advertisements on the boards depend on the geographic location and profile of the consumer, to reach the target audience for the advertisements.

Immersive sport coverage

Assume a major sport event such as the Olympics, World Football Championship, Tour de France Cycling or Car Racing. The event is covered with one or more digital broadcasts. Each broadcast consists of the usual full screen video and associated audio streams, but in addition other streams are provided too, such as one or more of the following small size pictures that can be overlayed on the full screen video. These overlayed pictures may or may not have associated audio, and may represent for example one or more of the following :

- Coverage of a simultaneously ongoing game
- Status of the race at different positions
- Performance of a specific racer
- Highlights of the game or race until now
- Summary of a previous race or game
- Replay of something that happened during the game or race.

The additional content may be available in many ways. It may be broadcast or stored locally or at a server. In case the additional content is intended for a broad audience, it is likely delivered over the broadcast channel, but in case of delivery to a single user or a small group of users, the content may be delivered more efficiently over IP.

The objective is to optionally offer users an enhanced coverage of the event. If the user wishes, after tuning in he can ask for the highlights until now, he can keep an eye on the progress of simultaneous games, have a look to the performance of his favorite racer and have a replay of what happened earlier. The user may be able to choose between watching the event in the usual lean back position or to have an exciting multimedia experience in a little bit more lean forward position or anything in between, as desired.

Enhanced EPG

Assume a bouquet of digital broadcast services offered over a medium such as cable. For user convenience and promotional purposes an enhanced EPG is offered; the EPG can be overlayed on top of any of the selected broadcast programs, exploiting the transparency and arbitrary shape features of MPEG-4. The EPG shows a vertical axis that lists all available services and a horizontal axis that represents time. The user can very conveniently select a service and a point in time and will get presented in a small window information on what is broadcast by the selected service at the selected time. The information can be broadcast, stored locally or provided over IP. The user can select programs of a certain type and when available, such programs are recommended to the user, either to watch or to store for later use.

CONCLUSION

The MPEG-4 standard provides a truly open specification with a large variety of features enabling powerful multimedia applications. MPEG-4 provides solutions for delivery over Internet with its largely varying delivery rates, but MPEG-4 also provides a solution for high QoS services such as provided by digital broadcasts. Using the same MPEG-4 standard, applications can exploit the strengths of both delivery methods. Complete new ways of broadcast become possible.

Jan van der Meer
Technology Manager
Philips Consumer Electronics
Email : jan.vandermeer@philips.com

Evolving To The IP Solution –
IP Access To Embedded Circuit Switched Systems

J.C. Proano
Jane Gambill
Lucent Technologies

ABSTRACT

Many MSOs (multi-system operators) are providing telephony service and Internet access as well as entertainment video. These are usually provided in today's market as three separate networks – a video network, based on a hybrid fiber-coax infrastructure, a data network, built on the HFC infrastructure using CableLabs' DOCSIS specifications, and telephony network, also built on the HFC infrastructure. Some MSOs are considering combining the data and telephony networks, to reduce the cost of network implementation and maintenance.

Two approaches are considered: (1) Providing an IP telephony network on top of the DOCSIS network, using the CableLabs' PacketCable specifications, and (2) Providing IP telephony access to existing telephony equipment in the headend. The latter approach allows existing circuit switches deployed by an MSO or other service provider to provide telephony service to cable users on an IP based cable access plant. This scenario allows a migration from circuit based telephony to packet based telephony. Normalized costs are evaluated. From an end-user perspective and from an access network perspective – the two approaches are transparent. However from a service provider and network architecture perspective, the two approaches are vastly different.

INTRODUCTION

Today's cable service providers are exploring adding revenues to their income by adding new services beyond traditional entertainment video. Internet access and residential telephony service are two good examples – built on the existing video distribution network, these subscriber services offer new revenue streams that leverage the existing service provider investment.

Internet access is the service that allows subscribers to use their personal computers (PCs) to access an Internet Service Provider, or ISP. *Roadrunner* and *@home* are two ISPs that partner with MSOs to provide Internet service. The MSO works with the subscribers to provide cable modems in the home, and provides a CMTS in the hub or headend that routes the data packets to the ISP. Subscribers usually pay a monthly fee for the high-speed access a cable modem network provides to the Internet, in addition to the fee for cable television services. MSOs leverage the installed HFC network to minimize the cost of deploying this additional revenue-generating service.

Residential telephone service is another revenue stream being offered by MSOs. In this model, the MSO works with the subscriber to provide a box on the side of the house – a network interface unit (NIU). The NIU has telephone jacks (RJ11) on one side, and a coax connection on the other side, for the cable drop. Each telephone jack is connected to the one twisted pair in the house telephone wiring, providing a new phone line for the house. The NIUs on the side of the house vary from vendor to vendor, providing one or multiple new phone lines for each house. Subscribers typically pay a monthly fee for each phone line activated.

Depending on the technology selected by the NIU vendor, different headend equipment and different network bandwidth is required to provide telephone service. Two technologies are currently being explored for phone service: circuit-switched technology and IP-based technology. This paper compares circuit-switched and IP-based technology for telephony, and proposes a hybrid solution providing IP access to circuit-switched headend telephony equipment. Normalized cost is used to compare the different technical solutions.

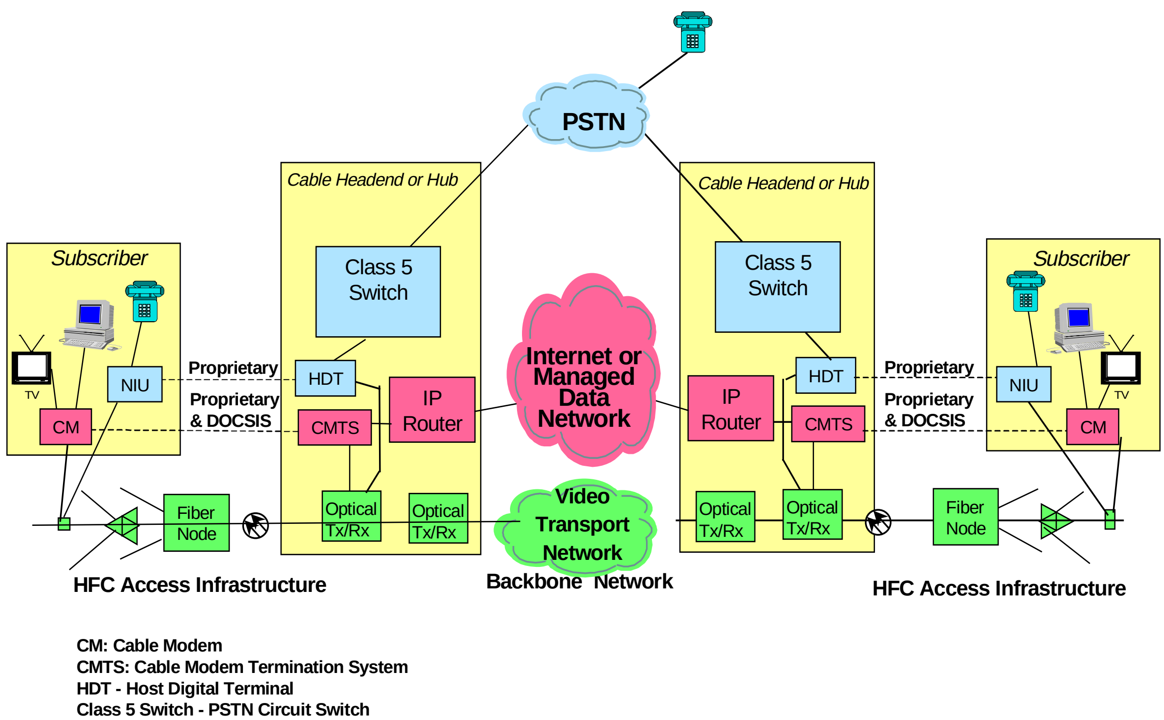


Figure 1: Overlay Networks

DEPLOYING CIRCUIT-SWITCHED TELEPHONY SERVICE

Figure 1 shows an architecture that provides telephony to the cable network subscribers. In this architecture, the cable operator chose to deploy an overlay telephony network at the service level (i.e., it uses the same coax cable to the home but the operator has separated voice and data bandwidth over the cable). Note that the only integration of voice and high-speed data in this architecture is the sharing of the HFC physical facilities.

An example of how the bandwidth on the cable spectrum might be allocated to

Figure 2: Spectrum Allocation for Overlay Networks

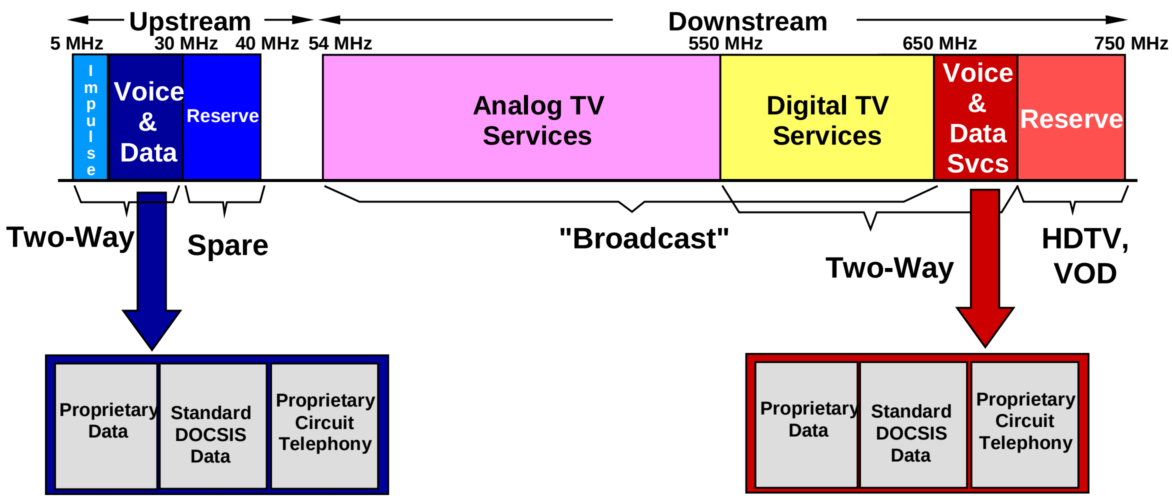


Figure 2

support these overlay networks is shown in Figure 2.

The bulk of the spectrum is reserved for television services, and the new revenue streams are allocated 5-40 MHz and 650-700 MHz. This bandwidth is divided among the standard and proprietary data protocols, and the separate circuit telephony protocol. When telephone service and Internet access are both deployed over a common IP network, then the underlying physical network, the IP network, and the common spectrum can all be shared.

This type of architecture has a few advantages:

- Provides immediate implementation of telephony service by using two-way cable networks and circuit-based backbone transmission and switching equipment.
- Quality of service is very similar to the old twisted pair.
- Telephony services are transparent to user.
- Performance and reliability is very high and is only limited by reliability of the cable plant.
- Fairly secure. In the last few years, this implementation has been supplemented with efficient security mechanism for privacy and other security attacks.

Among the disadvantages we can mention the following:

- It is an overlay network incompatible with packet switching technology.
- Evolution to packet-based telecommunication is very limited or null (forklift-upgrades are needed).
- Costly solution. Uses technology that is currently in decline (or at least not aggressively increasing)
- Current mature circuit-based over cable technology cannot take advantage of the IP backbone network without expensive multiplexer and conversion systems.

DEPLOYING IP ACCESS TO EXISTING CIRCUIT-SWITCHED TELEPHONE EQUIPMENT

A technology emerging in telephony service is IP Telephony – the ability to deploy telephone service over IP networks. IP Telephony offers efficiency in the use of cable spectrum, and in the use of cable network infrastructure, particularly in networks that include IP access to Internet Service Providers. Many, but not all, of the features used in residential telephony service are available in IP Telephony implementations. For service providers who have Class 5 circuit switches available, it's possible to upgrade the access to those switches from an overlay circuit switch network to an integrated IP telephony access network, as an interim step to providing full IP telephony.

This Hybrid Solution is characterized by efficiently providing high quality telephony and the full set of features (the same ones that the subscribers currently enjoy via the old copper twisted-pair technology), via IP access to circuit equipment.

Figure 3 shows one possible implementation of the Hybrid Solution.

This solution is appealing because it does not disrupt the video entertainment nor the packet data services infrastructure. This Hybrid solution takes advantage of the deployed IP access and backbone networks to carry the voice traffic (in packetized form) to the packet-circuit gateway (NCSG) at the edge of the network.

From an engineering point-of-view this Hybrid Solution has the following characteristics:

1) provides significant capital cost savings,
2) minimizes the development of new operations systems while improving the performance and cost effectiveness of MSO's cable network flow-through provisioning, and
3) leverages the continued use of (potential existing switches) Class 5 (telephony switches) operations support systems infrastructure.

The Hybrid solution represents a

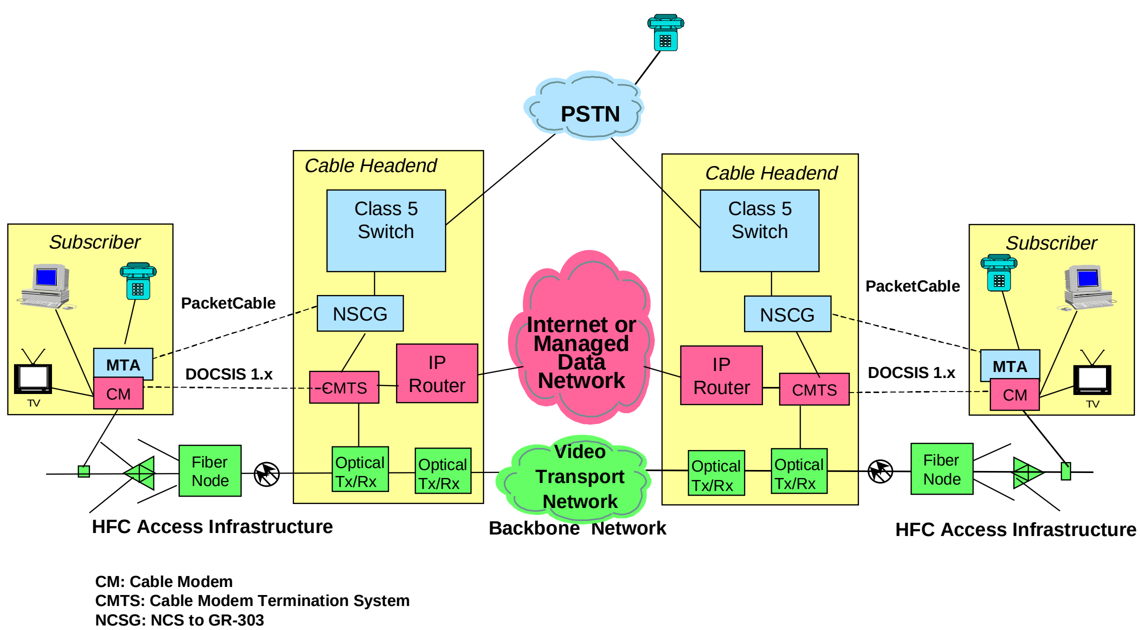


Figure 3: Hybrid Telephony Solution

major step towards the realization of most MSO's strategy for an any-distance, any-service, converged consumer franchise and the continuous reinvention of the MSO's broadband business. As the PacketCable standards mature and new technologies emerge, the cable network will become ready to go from a hybrid (IP and circuit) to the full end-to-end IP telephony solution. As the cable network evolves to a full packet-based network, the Class 5 switch evolves functionally to a full packet-based switch. This transition provides key technologies and capabilities necessary to evolve today's circuit switched networks into the cost efficient voice/packet network of tomorrow, while protecting the cable operator's investment in their embedded base of network elements and revenue generating services.

The NCSG (gateway) is supported by OAM&P systems that leverage current OSS systems capabilities and interfaces. In this solution, all network provisioning is done at the Class 5 switch through existing interfaces, and billing information flows are done over the current paths.

As the gateway between the cable network and the Class 5 switch, the NCSG translates between MGCP/NCS signaling and RTP bearer path on the cable network side to a GR-303 compliant interface that terminates on the Class 5 switch. Call processing and any associated functions in the cable network's core network, are performed in the circuit switched environment using the MSO's embedded base of Class 5 switches. The NCSG interfaces with the BTI (MTA/CM) at the subscriber premises via the CMTS over the cable plant (i.e., the Hybrid-Fiber Coax network).

The major components in the solution are:

- NCS Gateway (NCSG)
- Cable Modem Termination System (CMTS)
- Broadband Telephony Interface (BTI) consisting of Multimedia Terminal Adapter (MTA) and Cable Modem (CM)
- Data Server software platform (DNS/DHCP/TOD/TFTP Servers)
- IP Configuration Manager software platform
- IP Fault Manager software platform
- Element Management Systems (NCSG/CMTS/BTI EMSs)

A brief description of each component follows:

<u>NCS Gateway (NCSG)</u> is deployed at the edge of the circuit-switched network. On its feeder side, it provides DS1 interfaces (or STS1(E) or DS3 interfaces) upstream to an Class 5 switch, a local digital switch. On the distribution side, the NCSG provides access to IP networks utilizing Voice over IP (VoIP) packet technology. From the perspective of the serving digital switch, the NCSG is one or more GR-303-compliant remote digital terminals; from the perspective of the Broadband Telephony Interface (BTI), the NCSG appears as a PacketCable Media Gateway and Call Management Server.

<u>Cable Modem Termination System (CMTS)</u> – terminates the HFC Network at the cable system's head-end. The CMTS includes an IP router that essentially transfers IP packets between the cable distribution network (HFC, BTIs) and the NCSG while performing the appropriate physical and data-link layer conversions (i.e., RF cable to 100BaseT). The CMTS interfaces to the IP network and to the HFC.

Additionally, the CMTS provides the following capabilities:

- Data service support, auto discovery and auto provisioning, IP multicast, IP filtering (according to port number and source/destination addresses).

- Bandwidth allocation controls, RF channel operations, channel assignments, registration and management authentication, encryption key management, and quality of service (QOS) processing, and event message generation.
- SNMP Agent for CMTS management.

Multiple CMTSs may be provided depending on the capacity needs of the service provider.

Broadband Telephony Interface (BTI) – consists of a Multimedia Terminal Adapter (MTA) and a Cable Modem (CM) combined as a single unit:

Multimedia Terminal Adapter (MTA) - is a hardware device that interfaces standard analog phones to an IP network providing analog voice, analog fax, and telephony modem communication over an IP communications network. On the user side, the MTA provides multiple telephone line ports. Each telephone line port interfaces to standard analog touch-tone devices. The MTA originates or terminates voice/FAX telephone calls at its telephone line interface ports. The MTA communicates with the NCSG to provide subscribers with the telephony features set that reside in the Class 5 switch.

For telephony call management, the MTAs communicate with the NSCG by transmitting various control messages using the Network-based Call Signaling (NCS) protocol, which is carried over UDP/IP through the CM and CMTS and terminating on the NCSG. These messages also include the control messages to setup calls, transmit control information such as DTMF digits, and release calls. The MTA is initialized and registered with the NCSG, using NCS Protocol. The voice/fax data for calls to/from a MTA is transmitted over the HFC plant to the destination MTA via Real Time Protocol (RTP) over UDP/IP. The NCSG converts the voice packet to a DS0 stream for the Class 5 switch.

Cable Modem (CM) – acts as a transparent bridge to interface the MTA or terminal equipment attached to a local Ethernet LAN to the HFC Network. The CM transmits IP packets generated by the MTA or attached terminal equipment to the CMTS and forwards IP packets received from the CMTS to the MTA or attached terminals as appropriate. In particular, a CM interfaces the MTA and a PC to the HFC Network according to the DOCSIS CMTS-RFI Specification[i] .

The CM also provides the following functions:

- IP packet filtering : IP filters may be applied to restrict the types of services accessible to a CPE including e-mail server and web server. The IP filters may be defined either through a configuration file or through SNMP.
- SNMP Agent: the CM provides full management access supporting the MIB II and the DOCSIS MIBs (RF, Cable Device, BPI).
- Provides monitoring information to the CMTS.

Data Server software platform (DNS/DHCP/TOD/TFTP Servers) – is a suite of software solutions for the integrated IP services management that are required to support IP Telephony over cable networks. The DHCP and DNS Servers are fully integrated systems to synchronize updates in real time and run autonomously. This provides maximum flexibility in configuring and deploying services across the network. In addition, the data server platform solution facilitates building redundancy and fail safe uninterruptability into the network infrastructure. Data Servers support

10/100BaseT Ethernet interfaces. A brief description of each Server follows.

The DHCP Server is used to make dynamic IP address assignments to the BTIs.

to perform operations such as retrieving configuration information out of a user profile database and sending them to a CMTS or other network device. TFTP could also be used to update firmware in BTIs.

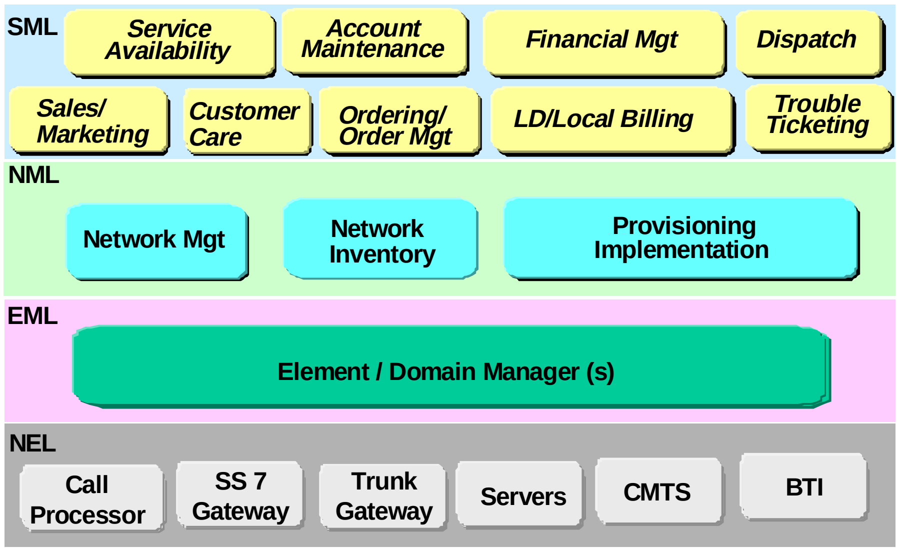


Figure 4: Telecommunications Management Network Model

The DHCP Server supports the definition of vendor classes, configuration of address pools, and the specification of lease parameters for IP Address pools.

The DNS Server provides the NCSG the data necessary to maintain a full cache of Fully Qualified Domain Names (FQDNs) to BTI IP addresses. It supports policies to check and handle if the client's requested hostname are a duplicate within a domain or across the HFC enterprise. It also supports secondary server updates.

The TOD Server is used to support the synchronization of IP devices supporting the IP telephony service.

The TFTP Server is used to support software downloads to IP devices supporting the IP telephony service. TFTP may be used

Configuration Manager software platform (IPCM) - is the customer provisioning system in an overall operations architecture based on the Telecommunications Management Network ("TMN") model. TMN is a layered model that divides the functionality needed to manage network elements and the services provided by Business Management, Service Management, Network Management, and Element Management. Figure 4 is an illustration of that model. The IPCM is part of the network management layer and is the heart of the solution's flow-through customer provisioning process. The IPCM accepts customer service provisioning requests from the Network Inventory.

Once network and Class 5 switch provisioning is complete, the IPCM provides customer service NCSG provisioning, and full BTI provisioning when the BTI is powered up at the customer site. The solution supports both data and voice provisioning and utilizes the PacketCable MTA Provisioning Flow. The IPCM supports northbound standard CORBA/IDL interfaces and interfaces to the BTI EMS and NCSG. The northbound CORBA interface supports create, modify and disconnect changes to the NCSG.

The OSS architecture for the NCSG solution, presented here, is focussed on two areas: customer service provisioning and fault management.

IP Fault Manager (IPFM) - receives network element alarms from the NCSG, CMTS, BTI and the Data Servers. The IPFM performs standard fault maintenance capabilities (filtering, thresholding, throttling, correlation, etc.).

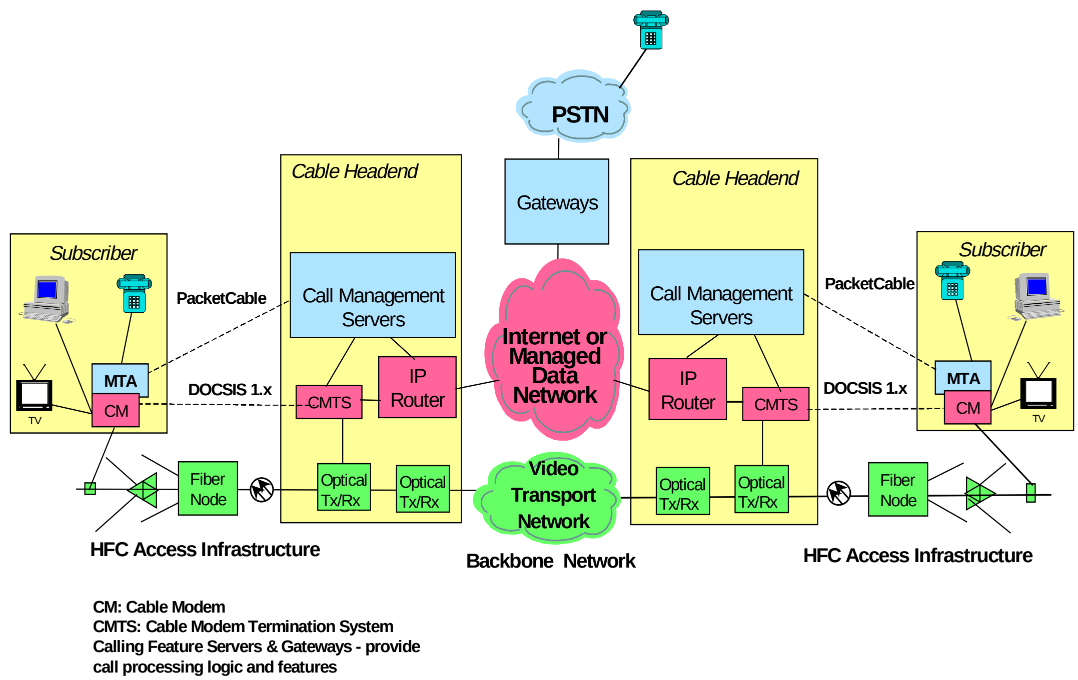
Element Management Systems (NCSG/CMTS/BTI EMSs) - BTI-EMS, CMTS EMS and the NCSG Element Manager provide element management layer support for the IP access network.

The proposed Hybrid solution assumes that an appropriate infrastructure made of data routers, optical systems, data servers, etc., are already in place in the MSO's network.

DEPLOYING IP ACCESS TO IP TELEPHONY EQUIPMENT

Figure 5 shows the architecture for the end-to-end IP telephony over cable. This architecture has no circuit-based equipment and the transition to full packet-based telephony (IP telephony) has taken effect. Notice that this transition includes an evolution of the NCSG system to a call management server and gateways' implementation. The functionality of the different boxes, included in the description of the NCSG solution, applies here as well. Call management servers and gateways shown in the diagram above are PacketCable compliant; basically, they translate the functionality of the Class 5 switch into the end-to-end IP network environment. The overall effect of providing IP access to the telephony system on the Internet access network is to raise the level of the network shared among the services to the IP level. This allows the maximum re-use of deployed equipment and applications.

Figure 5: IP Telephony



As an additional point of comparison, the telephony and high-speed data integration even at HFC level is more effective in the NCSG/Full-IP solution. Figure 6 shows the bandwidth shared by the HSD and telephony services (in addition to the video entertainment services).

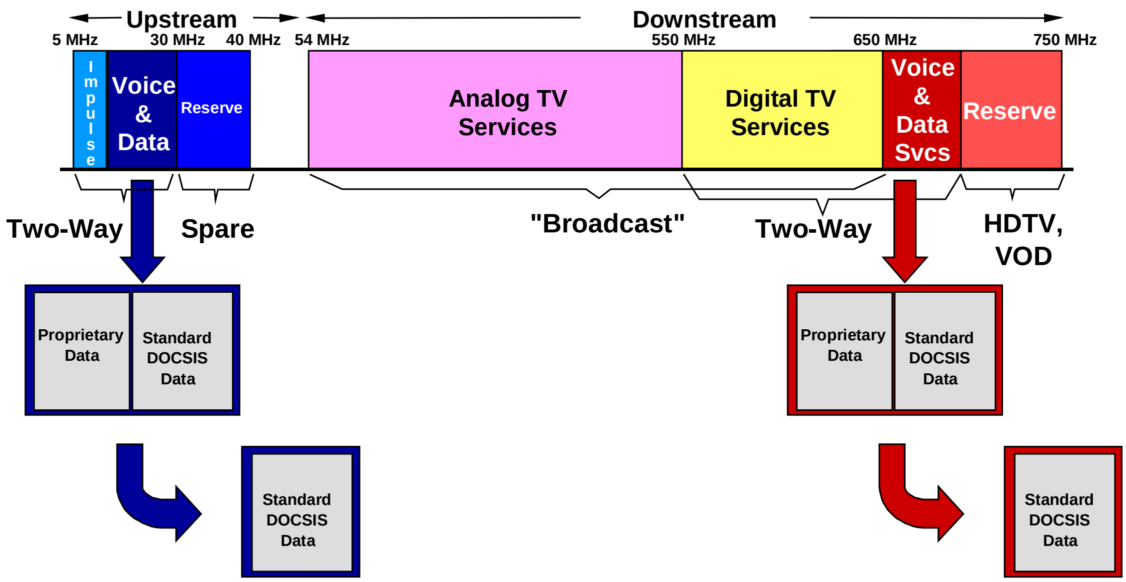


Figure 6: Spectrum Allocation in Full IP Solution

COST COMPARISON

Figure 7 shows a comparison between the circuit-based telephony overlay network architecture, the Hybrid solution using NCSG gateway and the full end-to-end converged IP telephony network solution, using normalized price-per-subscriber.

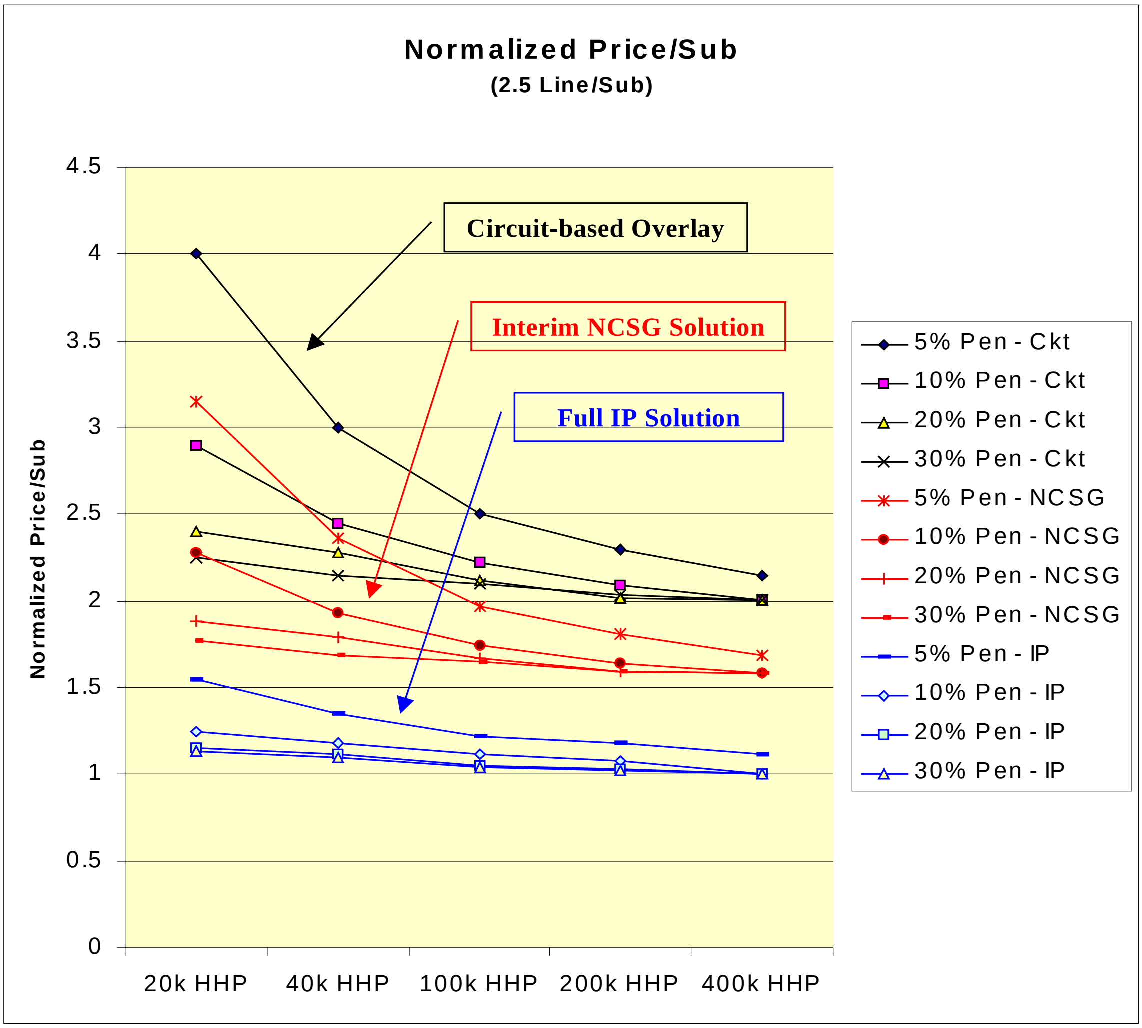


Figure 7: Normalized Price / Sub

Notice that even for large deployments (optimal scenarios for the overlay solution) the circuit-based telephony overlay network is twice as expensive as the full IP solution. The reason behind this is the duplication, in the case of circuit-based network overlay, of equipment throughout the network. The Hybrid solution appears to be significantly less expensive than the circuit-based overlay solution.

Up to now the comparison has been done on equipment alone. If we consider that the circuit-based overlay network solution has two full networks to operate, maintain, provision, etc., the cost of the network increases significantly. In this architecture there are two independent and complete networks, this means that it requires two sets of specialized personnel, training, procedures, etc. The operation cost is significantly higher in the circuit-based solution than the other two architectures under discussion.

CONCLUSION

The IP technology for telephone service provides enhanced use of Internet access networks, so an MSO expecting to provide both Internet access and telephone service should consider the use of IP telephony equipment for telephone service. If the MSO has invested in headend telephone equipment as well as Internet access network, an Hybrid solution like the one described here can provide an excellent method for leveraging existing investments as well as evolving the network to a converged multi-service network in the future.

ABOUT THE AUTHORS

J.C. Proano and Jane Gambill work in Lucent Technologies Cable Communications Architecture, providing network solutions to drive the growth of the cable industry, and the Bell Labs development of Cable Solutions.

For the last 10 years, Jay has been a member of Bell Laboratories in AT&T and Lucent Technologies. He has a broad experience in domestic and international telecommunications systems development. During his work at Bell Labs he has held positions in systems engineering and architecture development for: SONET transmission systems, ring technologies, HFC networks, CATV analog video systems, SDH broadband digital switching systems, and Cable networks. Jay is currently responsible for IP telephony architecture. Jay holds Ph.D., MS, and Diploma Engineering degrees, in Electrical Engineering and Applied Mathematics.

Jane has been with Bell Labs since 1982, designing and developing IP telephony equipment since 1992. Before that, she designed and developed business circuit telephony equipment and real time operating systems. She received an MS in computer science from the University of North Carolina, Chapel Hill, in 1982, and a BA, with majors in mathematics and computer science, from the University of Tennessee, Knoxville in 1980.

Contact information:

J.C. Proano
Lucent Technologies
Cable Communications Architecture
Member of Technical Staff
11900 N. Pecos St., DR 30j36
Westminster, CO 80234
303/538-1663 fax 303/538-3907
proano@lucent.com
www.lucent.com/cableconnect

Jane Gambill
Lucent Technologies
Cable Communications Architecture
Senior Manager
11900 N. Pecos St., DR 30j23
Westminster, CO 80234
303/538-3894 fax 303/538-3907
jgambill@lucent.com
www.lucent.com/cableconnect

---

[i] CableLabs Data-Over-Cable Service Interface Specifications, *Radio Frequency Interface Specification*, Version 1.1.

Extension of Cable-based Services to Home Networks

Terry D. Shaw
Cable Television Laboratories, Inc.

Abstract

CableLabs has been investigating technologies that could be used for provisioning and delivering cable-based services throughout the home environment. This paper provides a discussion of several issues surrounding provisioning cable-based services through home networks.

Introduction

CableLab's Home Networking Project has been investigating technologies that could be used for provisioning and delivering cable-based services throughout the home environment. At the outset of this project, two key strategic principles were established for this investigation:

- Different cable services have significantly different requirements to transparently pass through to home networks and must be addressed, in some degree, separately.
  - Cable modems represent a first-generation IP gateway to interconnect to home networks for high-speed data (HSD) services.
  - Video-centric networks are evolving to support home-based client-server convergence entertainment services.
  - Any home networking solution should not degrade existing consumer telephony experience as IP-based telephony products are deployed.
  - At this time, technologies are being developed that embed a cable modem directly in devices that provide each of these service categories. While the number of cable modems embedded in household devices may grow to be significant, the industry would like to optimize the number of DOCSIS cable modems per home that actively communicate with the network in order to more efficiently use bandwidth and simplify network operations.
- The audio/visual entertainment network is seen as largely an island unto itself with limited interactions with other networks present in the home.
  - While communication with other networks is desirable, the set-top box is not seen as a focal point for voice and data delivery.
  - A/V efforts on home networks should concentrate on the issue of video redistribution of SDTV, HDTV, and, possibly, digitized analog video.

In the fall of 1999, CableLabs and its member companies surveyed home networking vendors on the questions of in-home network transport and access system interfaces. The primary lesson learned from this vendor survey is that if no action is taken, current market trends will likely create three separate home networks over which cable services will be delivered:

- A video-centric entertainment network based on 1394 technology.
- An IP data network for distribution of products slated for consumption by PCs and Internet appliances.

- A telephony network that uses a network interface unit (NIU) with an embedded cable modem on which "primary line" voice calls are distributed on the home's twisted-pair infrastructure.

It is important to stress that the above home networking scenarios are likely to evolve absent any proactive development efforts by the cable industry. That is, services connected to these home networks are unlikely to be customized to extend cable services over the home network segment to the benefit of customers. Moreover, this outcome limits the ability of cable operators to deliver services across the different networks. For example: In order to provide a call waiting message on the television, a message must be sent from the NIU serviced by one cable modem to the CMTS in the headend to the cable modem embedded in the set-top box, where it can be inserted into the video data for delivery to the television.

A number of factors, which set bounds on home network architecture, will allow cable operators to effectively extend and create new services over home networks, or permit the integration of a single home networking solution for all services and home locations:

- **Implementation costs**. Use of legacy wiring, installation costs, and component costs should be low enough to ensure a wide utilization of the architecture.

- **Varying service requirements**. The overall architecture should have components that address issues such as the transfer of high volumes of MPEG-2 video data and the provision of power to telephones as part of a primary line service.

- **Hardware divergence in an era of service convergence**. As IP and Internet-based products mature, a huge number of different hardware platforms are being developed in order to consume these services. In order to support network-based services, control of the service demarcation point is of strategic importance. The architecture should be constructed so that these natural market forces can be leveraged to the best advantage of the cable industry.

- **Divergence of business and technology strategies both inside and outside the cable industry**.

Given these strategic confines, the remainder of this document describes several technical issues concerning the extension of cable-based services.

Home Network Technical Issues

As a point of philosophy, consumers will only purchase and use services that provide value. In general, a consumer is inclined to purchase any service for which the perceived value exceeds the cost (this applies also to "free" services). If the performance of the home network degrades the quality of service to a perceived value that is less than the cost, then the customer will not purchase (or use) the service.

The primary considerations for ensuring quality service delivery over cable networks are:

- Support of varying levels of QoS.

- Network performance: Data rate, latency, jitter, and packet loss characteristics.

- Support of cable network management functions: Registration, Administration, Security (RAS), copy protection, billing.

- Support of cable network operations: Installation, configuration management, performance management, and fault management.

Overall, this issue involves a number of closely entwined matters including home network management systems, transport technologies, and protocol issues.

Home Network Management Systems

The widespread use of home networks, as part of the consumption system for cable-based services, will require "bullet proof" home network management systems. It is expected that if a customer consuming cable-based services experiences a problem, then the customer will no longer desire the service even if the source of the problems lies entirely within the functionality of the home network. The ability of a home network management system to appropriately distinguish between an access network-related problem and a home network-related problem will greatly enhance the value of the network to the consumer.

Transport Technologies

Four different types of transport media have been proposed for home network applications:

- Phoneline
- Wireless
- Powerline
- Special wiring

The choice of the technology best capable of supporting cable-based services is a complex issue and depends largely on the specific scenario of implementation. While there is no one technology that can service all applications, there is substantial evidence that phoneline and wireless technologies can serve a majority of scenarios. At this time, 1394 is a rapidly maturing technology for media and IP transport (Sony, NEC, etc.) and has been selected for OpenCable™ interface. Powerline and special wiring systems are in development that show some promise for the delivery of cable-based services. In order to effectively use the capabilities of any of these technologies, the requisite interfaces to the DOCSIS protocols must be developed. Vendors of all of the transport media home networking equipment have high data rate equipment with QoS hooks in development, which use protocols similar to the Ethernet-based technologies of phoneline and wireless technologies. These capabilities and the associated technology are, in general, not completely defined and must be adapted for use with cable-based services. Characteristics of wired transport media and wireless transport systems are summarized in Tables 1 and 2, respectively.

Table 1. Characteristics of wired transport media for home networks.

Wired LAN				
	HomePNA 2.0	MediaWire	1394	AC Wiring
RF Band	7 MHz–14 MHz	0 MHz–25 MHz	100 MHz–800 MHz	0 MHz–10 MHz
Digital Bandwidth	10 Mbps	100 Mbps	400 Mbps	10 Mbps
Wiring	Co-exists w/POTS	Uses POTS	New 6 conductor	Power wiring
Shipping Products	Yes	No	Yes (lower rates)	Yes (lower rates)

Table 2. Characteristics of wireless transport media for home networks.

Wireless LAN			
	HomeRF	802.11b	Bluetooth
RF Band	2.4 GHz	2.4 GHz	2.4 GHz
Digital Bandwidth	1.6 Mbps	11 Mbps	1 Mbps
Distance	Whole house	Whole house	Room
Shipping Products	Soon	Yes	No

Protocol Issues

Several key home networking protocol issues must be resolved in order to ensure the delivery of high quality cable-based products. These issues touch on the three primary CableLabs projects—DOCSIS, OpenCable, and

PacketCable™—and in some cases significantly overlap planned, in process, or already developed specifications of these projects. These issues include:

- Discovery of other devices and applications on the network
- Download Control
- Security (authentication, copy protection, firewalls)
- System Interfaces
  - Quality of service
  - Transport
  - Streaming protocol
- Network Management: Diagnostics and statistics

In some cases, different protocol approaches are used to solve similar issues (e.g., the security approaches on PacketCable and OpenCable). These differences are, in large part, driven by the nature of the specific products being delivered.

Conclusion

The extension of the fundamental bandwidth advantage of cable throughout the home promises to create tremendous value for both consumers and cable operators. The creation of seamless network interfaces will be a complex task involving network management systems, the performance of transport technologies, and network protocol issues. The solution for the delivery of multiple services must be closely coordinated with the CableLabs DOCSIS, OpenCable, and PacketCable projects to ensure that the consumer can enjoy a quality experience, regardless of the device of consumption.

HPNA 2.0 10Mbps Home Phoneline Networking

Edward H. Frank and Jack Holloway
Broadcom Corporation

Abstract

Computers are now in more than 50% of American homes, and about 20 million of these homes have at least two computers. The multiple PC home, together with the anticipated growth of Internet appliances, has created the need for a low-cost high-performance home networking technology. One approach uses the same pair of wires as the existing analog telephone service (POTS). Standardized under the auspices of the Home Phoneline Networking Association (HPNA), this technology is already in its second generation, operating at speeds up to 16 Mbps. Phoneline networking, unlike traditional Ethernet, must work robustly over a widely disparate range of transmission channels that have significant dynamic impairments.

1. Introduction

We live in an age of ever-accelerating technological change. The signal event at the end of the Second Millenium was almost certainly the explosion of the Internet. In 1995 there were 20 million users on the Internet, by 1998 there were 160 million. It is estimated that by 2003 there will be 500 million users worldwide, and over 14 countries will have more than 40% of their population on-line – countries that represent more than half of the world's GDP. Internet-based commerce has grown from essentially zero in 1995, to $50B in 1998 and is projected to grow to $1300B by 2003. There is an unprecedented level of investment in Internet-related business ventures – a direct consequence of the appreciation that the "new world order", built on a wired information network, will profoundly effect the way we work and live.

What is less well appreciated is that the electronic dendrites of this network will extend beyond the PC to every electronic device within the home – connecting literally billions of devices (See Figure 1).

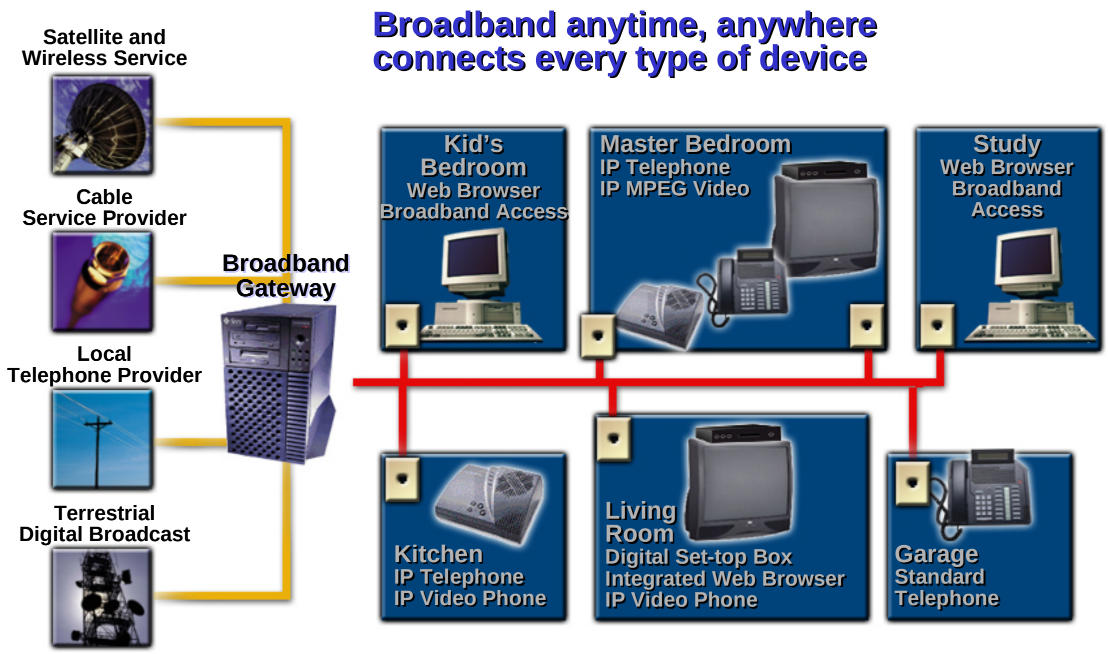


Figure 1: Connectivity in a networked home.

Traditional consumer electronics – television, stereo audio, telephones -- are already in the process of being redefined to use digital technology. In the new era, these devices will be designed with "the Network" built-in as a standard component, mirroring the absorption of the embedded microprocessor that occurred in the previous era. Network-connected devices will be smarter, easier to use, easier to maintain. The very nature of television, radio and the telephone will be transformed.

If every consumer electronic product will have an "Internet Inside" sticker, what connector will be used? How quickly will network-enabled products be adopted if consumers have to install network wiring and learn how to setup and administer a network? Will the home have to be "network-enabled" before these products can be used? The hard reality is that consumers don't want to buy networks – but will be motivated to buy smart network-connected devices that entertain, inform, educate, connect and increase convenience and choice. To initiate rapid market adoption, these devices will need to plug-in as simply as a telephone, with "no new wires".

In the home, there are basically three existing wiring infrastructures that can be exploited: phone wiring, wireless and AC power wiring. It appears that all three will be used, with phoneline networks deployed first.

In 1998 the computer and semiconductor industries created an Alliance to select, promote and standardize technologies for Home Phoneline Networking (see HPNA 2.0 system [1] and http://www.homepna.org). This group has introduced a first generation 1 Mbps technology (based on a system developed by Tut Systems), and a second generation 10 Mbps technology based on a proposal from Epigram, Inc. (now part of Broadcom Corporation). Home Phoneline Networking is well suited for the interconnection of broadband voice, video and data within the home. Industry reports estimate shipments of 1 million HPNA compatible interfaces by the end of 1999, and somewhere between 5-10 million interfaces by the end of 2000.

Networking over the existing home phoneline infrastructure suffers from many impairments (as do all no-new-wires physical media), namely high attenuation, reflections, impulse noise, crosstalk and RFI ingress/egress. These challenges must be overcome by a successful technology.

2. Requirements for Home Networking

It is our belief that for a home networking technology to be successful, it must properly address the following major issues:

1. Leverage existing wiring infrastructure and be easy to install.
2. Leverage existing standards and interwork with common operating systems and software platforms.
3. Implement a quality of service (QOS) mechanism that provides low latency for telephony and other voice applications; implement guaranteed bandwidth for streaming audio and video applications.
4. Be very robust and provide connectivity in essentially every home.
5. Support data rates in excess 10BASE-T Ethernet, and scale to 100 Mbps in a way that remains compatible with installed earlier generations.
6. Provide reasonable privacy at the physical layer. (Wireless and powerline require some level of encryption to achieve wired equivalent privacy.)
7. Be future safe, employing designs that are scalable and extensible so that users do not have to do “fork-lift” replacements when upgrading their networks in the future.
8. Be implementable with sufficiently low cost to allow inclusion as standard in a wide variety of products.

Table 1 summarizes how well the principal choices for home networking technology meet these criteria.

Parameter	HPNA 2.0	Wireless	Powerwire	Ethernet (Cat 5)
1. Leverage existing infrastructure	Good	Good	Good	Poor
2. Leverage Standards	Good (802.3 compatible)	Medium (too many standards[1])	Poor (no standards)	Excellent
3. QOS Support	Good	Good to Poor (some standards have no QoS provision)	Unknown	Medium (Simple hubs don't support QoS. More expensive switches may.)
4. Robustness	Good	Medium	Unknown (Highly impaired channel)	Good
5. Performance	>10 Mbps, 100 Mbps next generation	1 to 11 Mbps Up to 50 Mbps at 5 GHz	Unknown (Highly variable channel capacity)	10, 100, 1000 Mbps
6. Privacy of Physical Medium	Good	Poor	Poor	Good
7. Future-safe	Good	Poor (too many standards, potential for interference)	Unknown	Good
8. Cost	Good	Medium (RF circuitry is harder to integrate)	Unknown (but should be comparable to HPNA)	Medium (low hardware cost, but higher cost if one considers installation of new wiring)

Table 1: Comparison of networking technologies.

---

[1] Several competing systems are under development and proposed for the unlicensed 2.4 GHz band (Bluetooth, HomeRF, 802.11b). The 2.4 GHz band has multiple sources of interference such as DECT phones and microwave ovens. The 5 GHz NII spectrum may also be used for home networking, using 802.11a or some other standard. Other standards and frequencies are proposed for systems to be used in Europe and Japan.

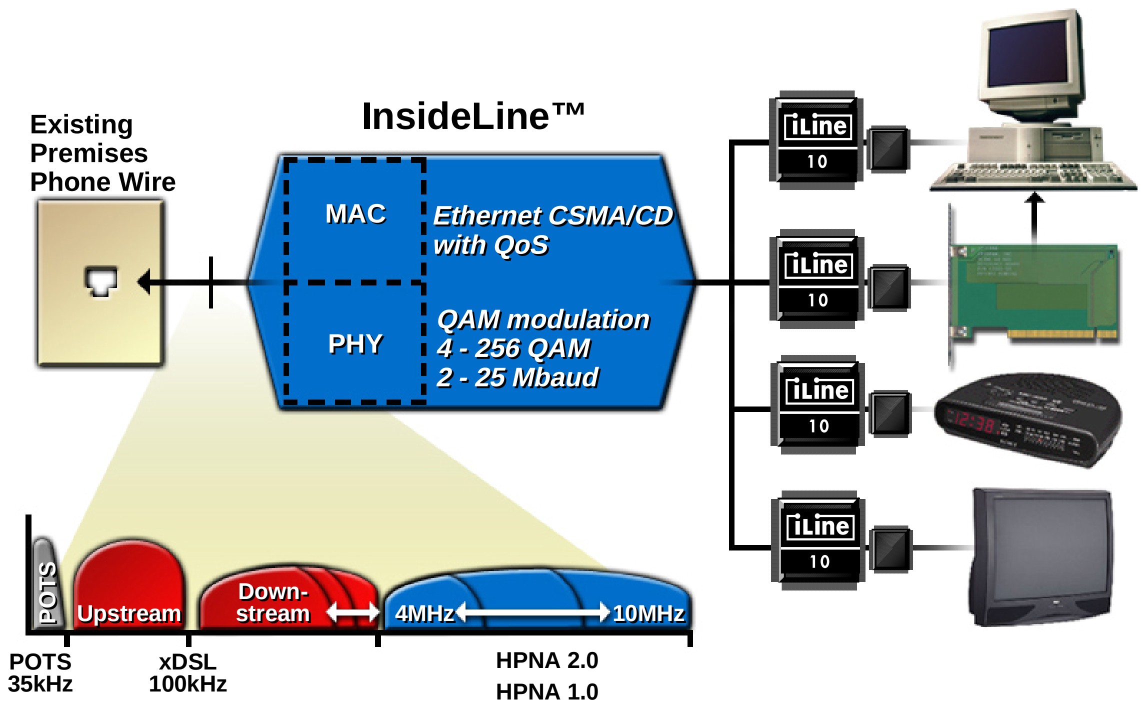


Figure 2: A view of the HPNA stack and spectrum

3. The HPNA 2.0 System

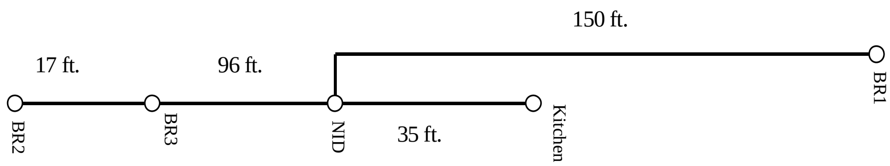
Figure 2 is an illustration of the HPNA 2.0 system from the point of view of network stack and frequency spectrum. The HPNA 2.0 system is a multi-point CSMA/CD packet network that supports unicast, multicast, and broadcast. As will be discussed in this section, and illustrated in Figure 7, it has the look and feel of Ethernet. However, it differs from 10BASE-2 and 10BASE-T in a number of respects. First and foremost, HPNA 2.0 places no restrictions on wiring type, wiring topology, or termination. Moreover, like 10BASE-2, but unlike 10BASE-T, HPNA 2.0 uses a shared physical medium with no need for a switch or hub. 10BASE-T on the other hand requires dedicated point-to-point CAT-3 or CAT-5 wires.

Physical Layer

At the Physical layer, the system is frequency division multiplexed on the same wire as standard analog phone service (POTS), as well as other splitterless ADSL[2]. Analog telephony uses the low part of the spectrum below 35 kHz. ADSL (both G.Lite and G.Heavy) use spectrum up to 1.1 MHz.

HPNA selected the 4 to 10 MHz band for several reasons. The lower limit of 4 MHz was chosen to make it feasible to implement the filters needed to reduce out-of-band interference between HPNA and splitterless ADSL. After modeling several thousand representative networks with capacitive telephones and common wire lengths, it was determined that the spectrum above 10 MHz was much more likely to have wider and deeper nulls caused by reflections [3]. Crosstalk between phonelines increases with frequency, and the analog front end is harder to implement at higher frequencies. The particular choice of 4 to 10 MHz only overlaps a single Amateur Radio band (40 meters), which simplifies ingress and egress filtering.

The no-new-wires media that are available for networking within homes have the problem that the communications channel can be severely impaired. The nature of the impairments is illustrated in Figure 3 and Figure 4, which shows the kind of channel response one might find on a typical phone wire loop inside of a house.



4

Figure 3: Topology of a simple home network

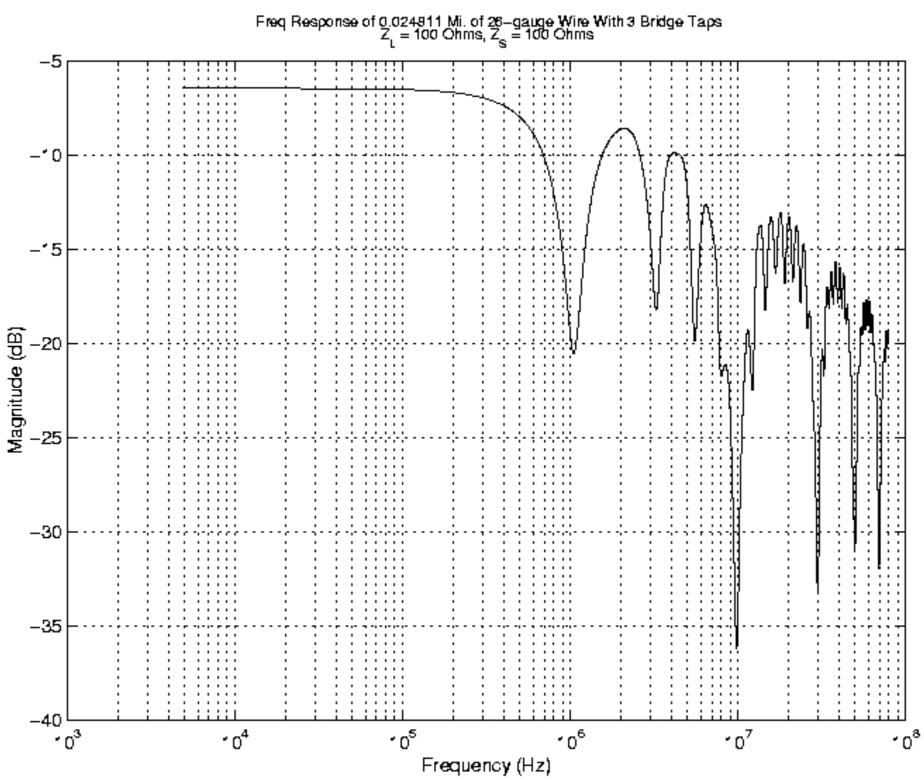


Figure 4: The channel response of a simple home phonewire network.

The wiring topology found in homes is ad-hoc resulting in reflections and frequency dependent channel transfer functions; the transmission parameters of the wire used is uncharacterized and highly variable, especially at higher frequencies; telephone instruments on the same wiring present a wide range of frequency-dependent impedances; While HPNA 1.0 uses Pulse Position Modulation as mentioned above, HPNA 2.0 uses quadrature amplitude modulation, both to get more throughput in the same bandwidth, as well as to achieve greater robustness. However, due to fact that the channels may have very deep nulls, and multiple nulls in band, two techniques are used to improve robustness. The first technique is to be rate adaptive. Instead of having a fixed number of bits per symbol, a transmitter may, on a packet by packet basis, vary the packet encoding from 2 to 8 bits per symbol. The second technique is to use spectral diversity as discussed below

Frequency Diverse QAM

Unfortunately, the nature of channel nulls can be such that even rate adapting down to 2 bits per symbol is not sufficient to guarantee that the packet can be received.

In a traditional QAM system, if there is an extreme null (i.e. one with which the equalizer can't cope) in the band then the system will fail to operate. At its 2 Mbaud rate, HPNA 2.0 implements a modified version of QAM invented by one of our colleagues Eric Ojard, called Frequency Diverse QAM (FDQAM) [5]. While a full discussion of FDQAM is beyond the scope of this paper, Figure 5 through Figure 8 illustrate the basic concept.

In a traditional QAM system, a single copy of the baseband signal is sent and received. Because in FDQAM the baud rate is less than half the width of the filter, the output signal has two redundant copies of the baseband signal, as shown in Figure 5. Thus, the signal is frequency-diverse, motivating the name FDQAM.

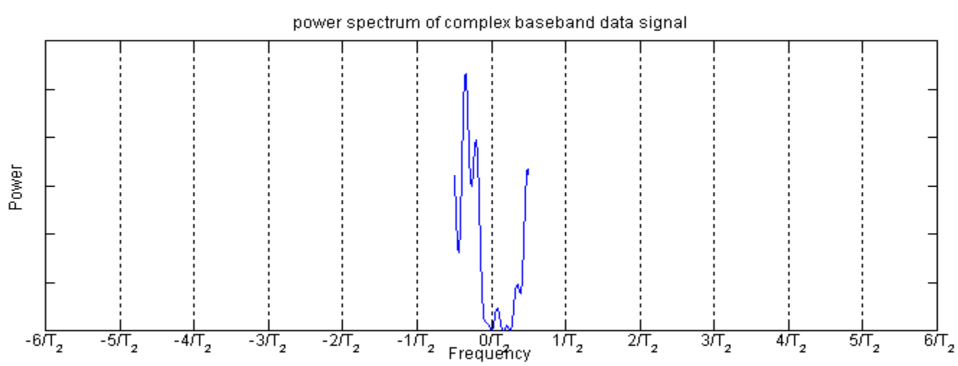


Figure 5: Spectrum of complex base-band signal.

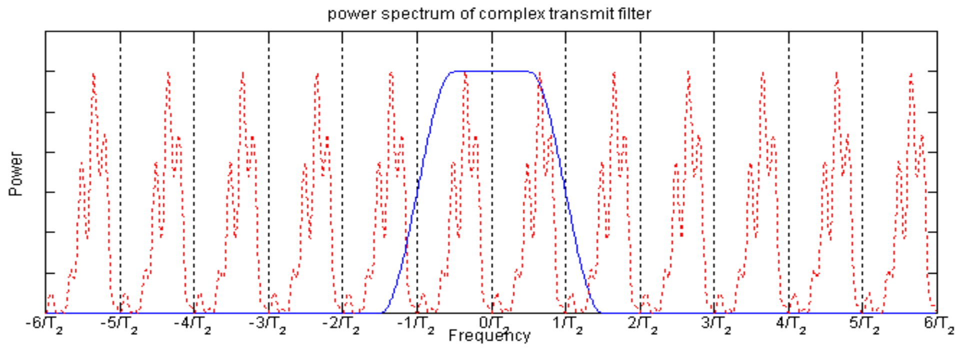


Figure 6: Spectrum of upsampled complex baseband signal, overlayed with low-pass filter.

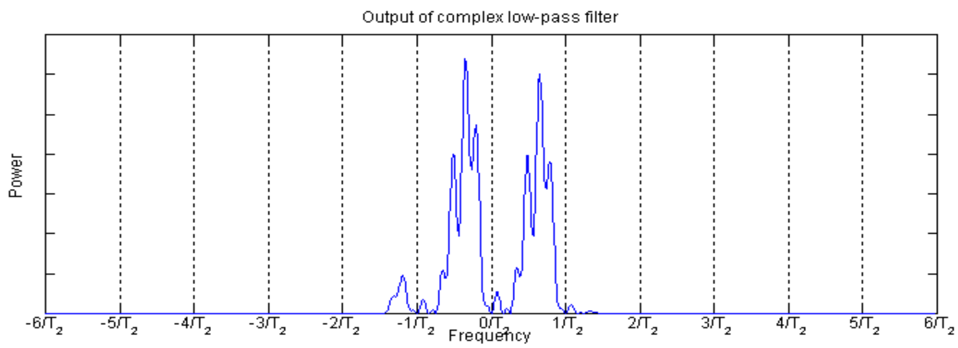


Figure 7: Ouput of frequency-diverse low-pass filter.

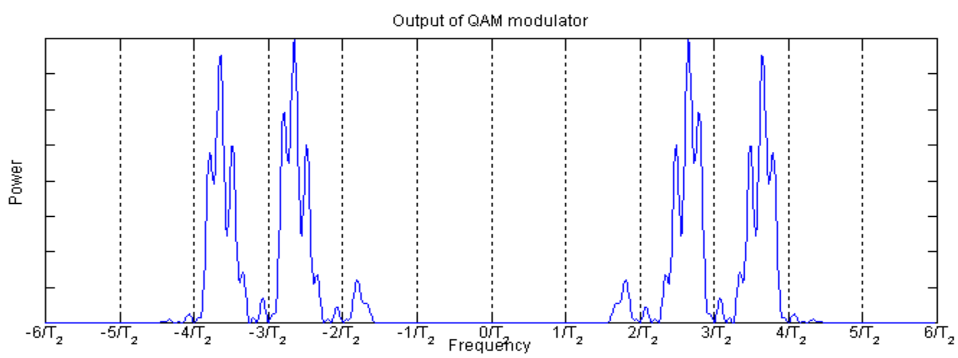


Figure 8: Output of FDQAM modulator.

Intuitively, it's easy to see that on channels where half of the spectrum is nulled out, one copy of the signal will still make it through. Quantifying the performance of FDQAM versus QAM on arbitrary channels is more complicated, and this analysis is not included here. It can be shown, however, that on channels with low SNR where a large part of the spectrum is severely attenuated, FDQAM works robustly in many cases where uncoded QAM modulation would fail. Such channels are common on home phonelines. Unlike most other methods of handling severe channels, FDQAM requires no knowledge of the channel characteristics by the transmitter, simplifying the protocol and enabling robust performance over time-varying channels.

In cases where the channel nulls are not particularly deep, HPNA 2.0 allows for a higher performance 4 Mbaud mode, which achieves peak data rates up to 32 Mbps, and throughput above 20Mbps.

Frame Format

Figure 9 shows the frame format on the wire. The frame begins with a known 64 symbol preamble. The preamble is used for several purposes:

- Robust Carrier Sensing and Collision Detection
- Equalizer Training
- Timing Recovery
- Gain Adjustment

Following the preamble is a Frame Control field. The first part of the FC is an 8-bit frame-type. Frame-type=0 is shown, where other codes can be assigned for frame formats used by future systems. Following the frame-type is an 8-bit field that specifies the modulation format (bits per symbol for example). There are miscellaneous other control fields in Frame Control including an 8 bit header CRC. The remainder of the packet is exactly an 802.3 Ethernet frame followed by CRC16, padding and EOF sequence. The CRC16 covers the header and payload, and reduces the undetected error rate for severely impaired networks.

Key to operation is that the first 120 bits of the frame are sent at the most robust 2 Mbaud, 2 bits per symbol rate. The reason for this is that if it is possible for any station to be able to demodulate a packet, it will be at this encoding. Thus, even if the payload is encoded at a rate or bits per symbol that the receiver can't demodulate, it will be possible to demodulate the header. In this situation, the receiver sends a Rate Request Control Frame to the sender asking it to reduce the number of bits per symbol or the symbol rate.

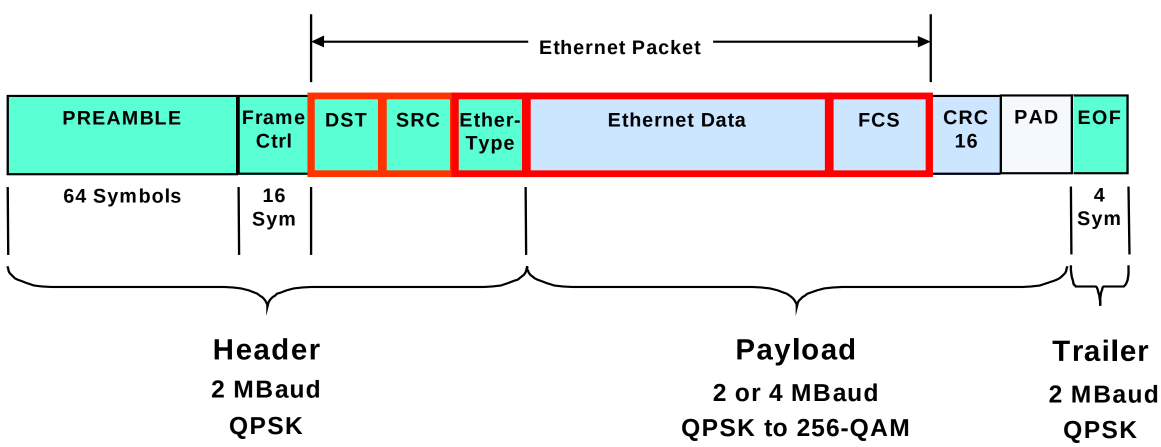


Figure 9: HPNA 2.0 Frame Format.

Media Access Control

As mentioned above, HPNA 2.0 is a Carrier Sense Multiple Access (CSMA) with Collision Detection (CD) system, just like standard IEEE 802.3 Ethernet. HPNA 2.0 introduces eight levels of priority and uses a new collision resolution algorithm called Distributed Fair Priority Queuing (DFPQ).

Voice telephony requires a low-latency network service, and streaming audio or video applications require a guaranteed bandwidth service. With the MAC in Ethernet, there are no real service guarantees, as shown in Figure 10.

In this example, three nodes (N1, N2, and N3) are contending for access to the network. Node N2 is transmitting a voice-over-IP (VoIP) packet. Initially, N0 accesses the wire transmits a frame (TX), and during this transmission N2 has packetized a voice sample and is ready to transmit, but must defer to N0. At the end of the first transmission, N0 has a second packet ready to send, and when N2 and N0 contend for access (resulting in a collision) N2 by chance chooses a longer backoff interval than N0. N0 gains access again and transmits. During this time, another station N1 becomes active, and starts deferring waiting for N0 to finish. Now, when N2 attempts to transmit it collides with N1. A possible outcome is that N1 succeeds in the collision resolution and N2 further increases its backoff. In this manner, the queuing disciple can become very unfair for N2. If N0 and N1 are PC's engaged in file transfers, they can generate enough traffic loading on the network to cause errors in the VoIP service operating on N2.

One solution to this problem is to introduce different access priorities, where the VoIP station uses a higher priority than best-effort file transfer traffic. HPNA 2.0 accomplishes this by organizing the time following the inter-frame gap into an ordered series of priority slots.

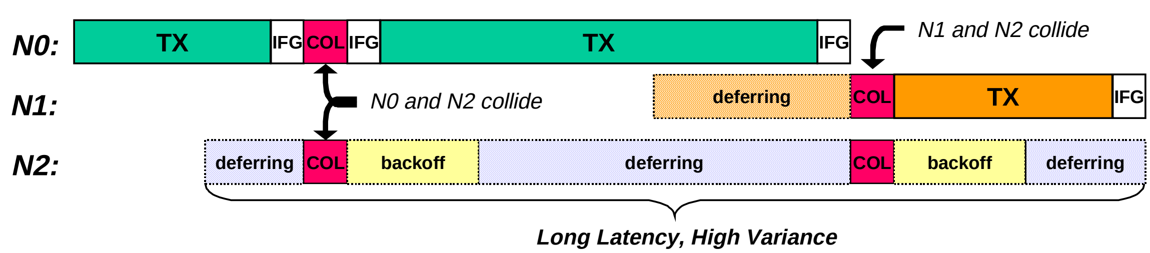


Figure 10: An example of access latency unfairness

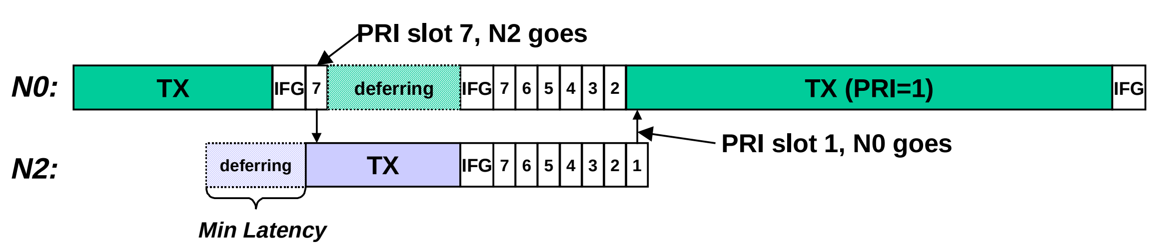


Figure 11: Priority

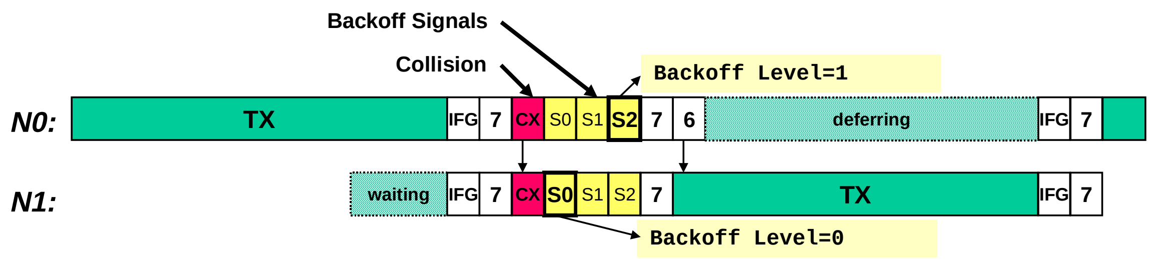


Figure 12: HPNA 2.0 Collision resolution algorithm

Now in the example shown in Figure 11, when N0 finishes transmitting, all stations on the network that have lower priority than 7 wait, while N2 begins to transmit (without collision). After N2's transmission, no stations have traffic with priority higher than 1, so N0 again gains access to the channel with its next transmission.

Access priority lets software define different service classes, such as low-latency, controlled-bandwidth, guaranteed-bandwidth, best-effort, penalty, etc., each using a different priority level.

Within a given priority level, HPNA 2.0 uses an algorithm for collision resolution where each station keeps track of a *Backoff Level* and after a collision randomly chooses to increment the backoff level by 0, 1 or 2. During a collision resolution cycle, stations incrementally establish a partial ordering –eventually only one station remains at the lowest backoff level and gains access to the channel.

In the example shown in Figure 12, N0 and N1 enter into a collision resolution cycle. N0 randomly chooses to increment its backoff level by 2, N1 by 0. To optimize the partial ordering, eliminating null levels, stations send a special signal immediately following a collision which reflects the backoff increment chosen (0 and 2 in the example shown). All stations observe these signals and perform a distributed computation to calculate the new (partial) ordering. In this case, N0 increments it's backoff level by 1 since it saw the Backoff Signal from N1 in S0 but no station indicating in S1.

In practice, even on saturated networks, HPNA 2.0 is very well behaved, and unlike traditional Ethernet does not exhibit the capture effect. The relative performance of DFPQ and Ethernet are shown below.

Figure 13 shows the distribution of access delay for Ethernet for offered loads ranging from 30% to 120% of network capacity. The simple simulation shown assumes a Poisson traffic model, 10 stations and a uniform frame size of 1500 bytes. Delay is indicated in units of frame transmission time, where a perfect queuing discipline would have a maximum delay equal to 10 frame times (the number of stations). As the offered load increases, Ethernet experiences delays of 100's of frame times for several percent of transmission attempts.

In Figure 14 the HPNA 2.0 access latency is shown. By comparison with Ethernet there is a negligible distribution tail beyond 20 frame times, even at high offered loads. It should be noted that the results shown are for contention within a single priority level. High priority traffic will see on the average less than one frame time of delay when contending with lower priority traffic.

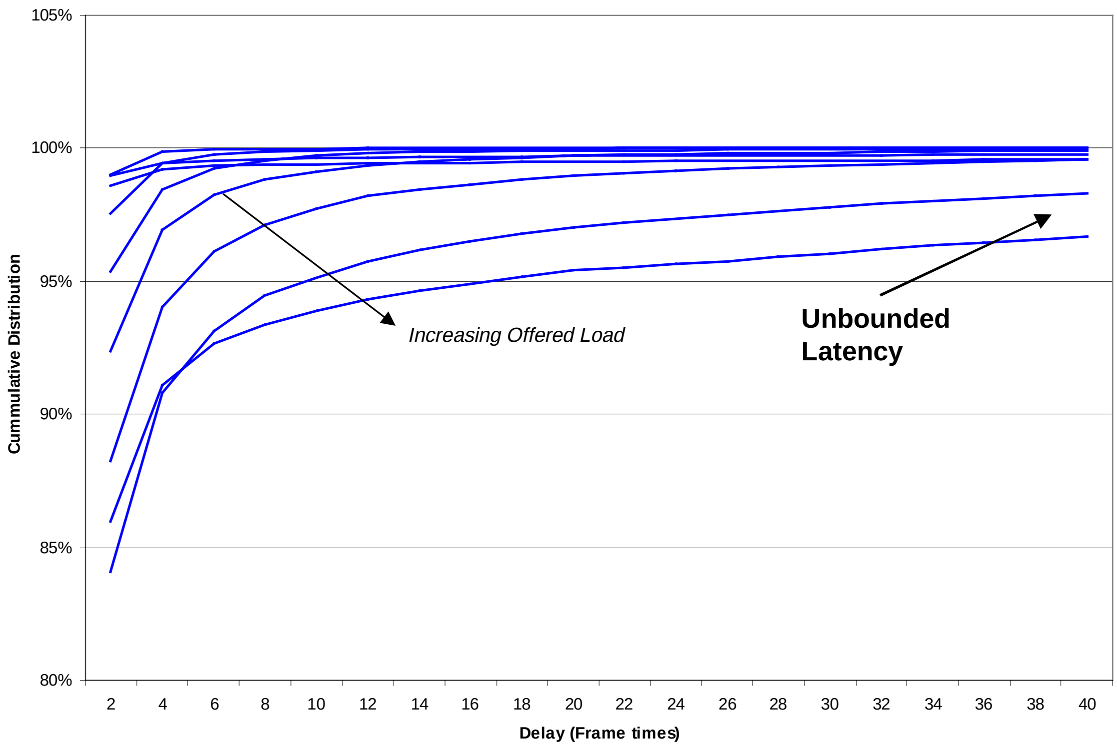


Figure 13: Ethernet Access Delay Distribution

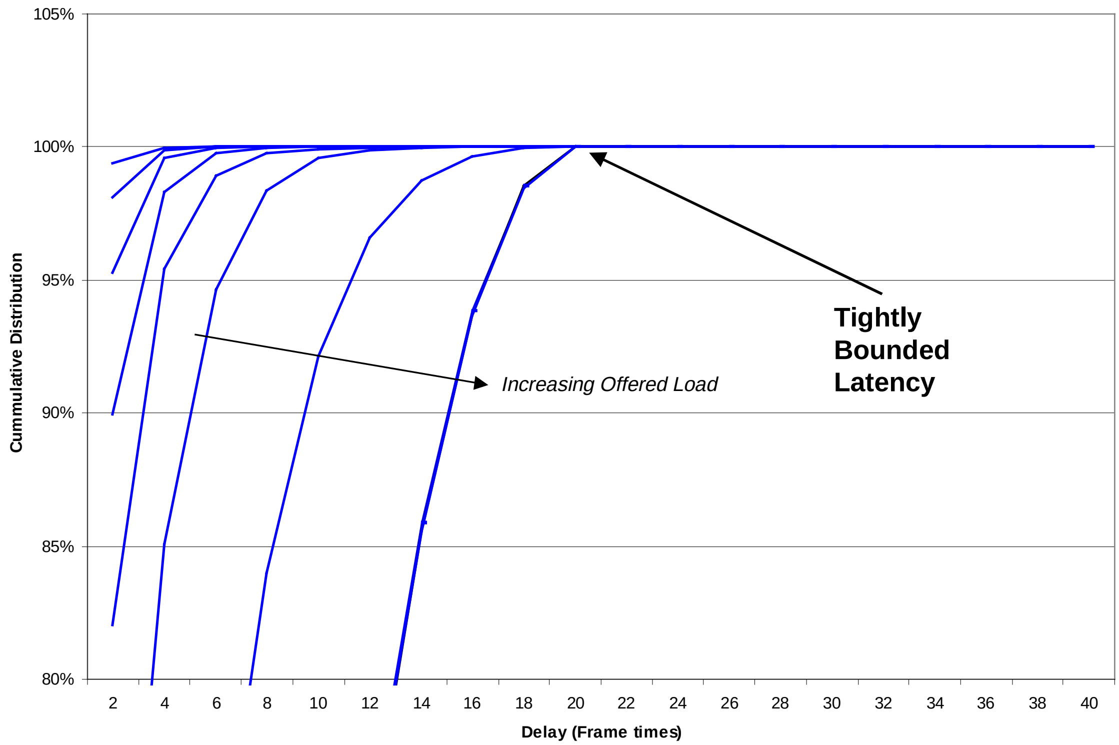


Figure 14: HPNA 2.0 Access Delay Distribution

Link Layer Protocols

One impairment not mentioned previously that is a problem for all home networks using "No new wires" is impulse noise. On phonewire impulse noise exists due to phone ringing, switch hook transitions, and noise coupled from the AC power wiring. Fortunately, the impulses tend to be short, and destroy only a single packet. While there are coding techniques that might reduce the number of packets destroyed by impulses, we have chosen to use a fast retransmission mechanism we call Limited Automatic repeat ReQuest (LARQ). Because LARQ is implemented (in software) at layer 2, and because it is only on a single segment of the network, it is very effective in hiding packet erasure from TCP/IP, as shown in Figure 15.

Finally, it is worth mentioning, the HPNA 2.0 implements a link integrity mechanism, which can be implemented either in hardware or at low levels of a software driver. The virtue of link integrity is that it provides a quick and easy way for the end user to determine if the network has basic connectivity. Link integrity frames are sent once per second, unless there traffic on the wire, in which case the number of frames sent may be reduced.

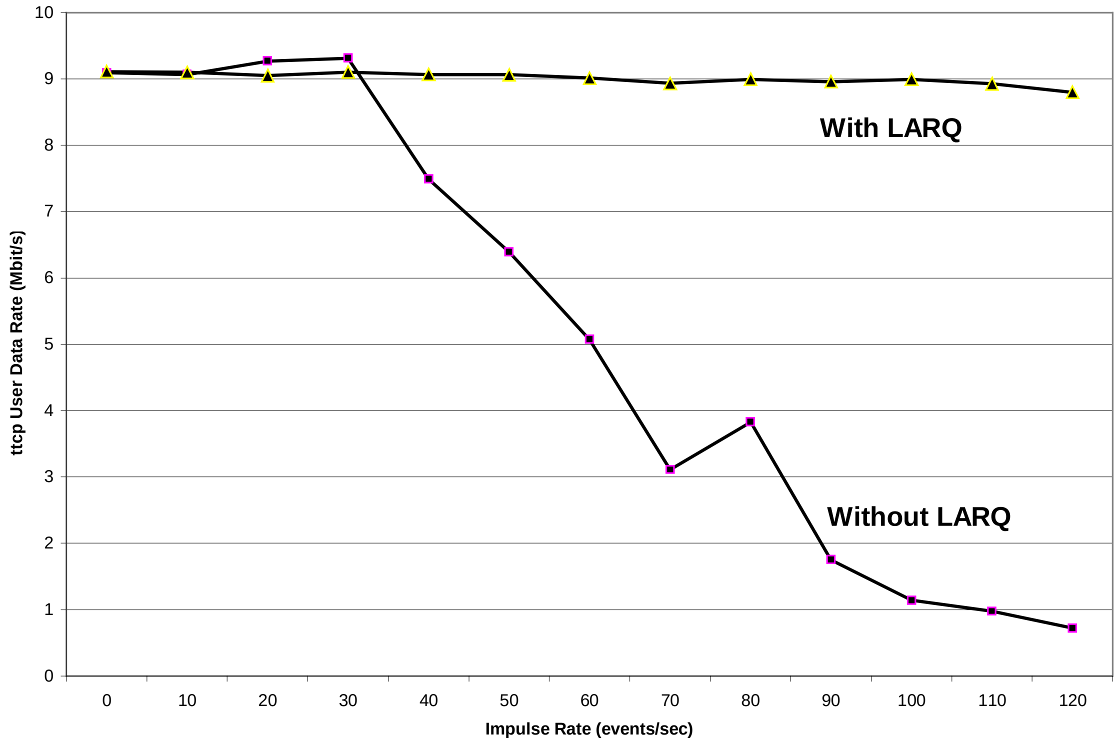


Figure 15: User level throughput vs. Impulse noise events/sec

4. Example Implementation

Figure 16 shows the high integration at the chip-level and the low complexity at the system-level that can be achieved with HPNA 2.0, in this case using Broadcom's iLine-10 chipset, which combines the MAC and PHY for both HPNA 1.0 and HPNA 2.0. Chipsets from other vendors should have similar characteristics. The major components are:

A. The BCM4210 MAC/PHY chip (the larger iLine chip)

B. The BCM4100 Analog Front End chip (the smaller iLine chip)

C. Magnetics module for phoneline isolation/ protection (large black module)

D. Serial Prom with MAC address and other configuration

E. Crystal

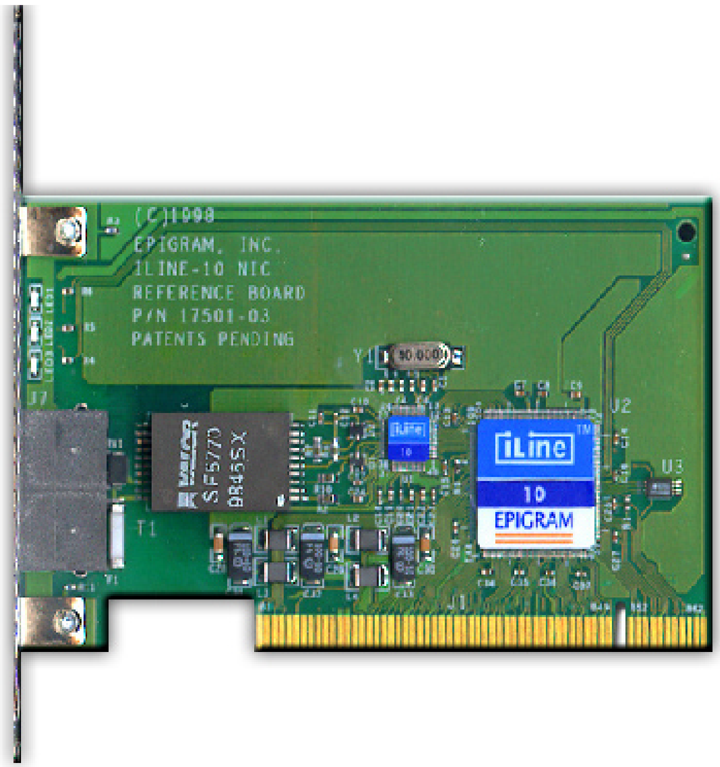


Figure16: Broadcom's iLine10 HPNA2.0 NIC reference board.

Conclusion

Home networking is now a reality. In 1999 estimates are that over 1 million home networking nodes were shipped. Estimates range as high as 10 million home networking nodes will ship in 2000. Using advanced signal processing techniques and high density CMOS it is possible to transmit data over existing media, such as in-place phone wire, at rates once considered impossible. Equally important, the cost of these solutions is such that the chips can be built into a wide variety of computers and internet appliances. Just as the Microprocessor has become an essential component of every digital device, we anticipate that a communications element, which we call the Internet-Chip or I-Chip, for short, will be come an equally essential element in every digital system over the next several years. Consumers will come to expect that the devices they buy have an I-chip in them. By the year 2005, if not sooner, consumer devices that can't communicate in the Home LAN will be obsolete.

Acknowledgements

The work described in this paper is the result of the efforts of many people. In particular we would like to recognize Ruben Alva, Neal Castagnoli, Alan Corry, Mike Dove, Greg Efland, Nino Ferrario, Hariprasad Garlapati, Ray Hayes, Mark Kobayashi, Harrison Kuo, Jim Laudon, Joe Lauer, Guillermo Loyola, Tracy Mallory, Rich McCauley, Subrat Mohapatra, Tushar Moorti, Walter Morton, Neal Nuckolls, Eric Ojard, Jay Pattin, Kevin Peterson, Henry

Ptasinski, Tim Robinson, Dane Snow, Bill Stafford, Jason Trachewsky, Chris Warth, Larry Yamano, and Chris Young.

Biographies

Edward H. Frank (ehf@broadcom.com) was co-founder and Executive Vice President of Epigram, now the Home Networking Division of Broadcom Corporation. Prior to Epigram, Dr. Frank was Vice President of Engineering for NeTpower, Inc., and was a Distinguished Engineer at Sun Microsystems, where he worked on the Green Project, which created Oak, the precursor to Java, and where he was co-architect of several generations of SPARCstation workstations. Dr. Frank has a PhD from Carnegie Mellon University, and BSEE and MSEE degrees from Stanford University. He holds over 15 issued patents.

Jack Holloway (h@broadcom.com) was co-founder and Chief Technology Officer of Epigram, now the Home Networking Division of Broadcom Corporation. Prior to Epigram, Mr. Holloway was Vice President of Broadband at MicroUnity, a closely held mediaprocessor developer. Mr. Holloway was a Vice President, Broadband Technology for Bolt Beranek and Newman, where he started an ATM product division called Lightstream, which was eventually acquired by Cisco Systems. Mr. Holloway was a cofounder of Symbolics, Inc, and for several years was a Principal Research Scientist at M.I.T. and co-Director of the M.I.T VLSI Laboratory.

[1] *HPNA 2.0 Specification*, October, 1999, available to members of Home PNA, see http://www.homepna.org

[2] Chen, W., *DSL: Simulation Techniques and Standards Development for Digital Subscriber Line Systems*, Macmillan, 1998.

[3] "Spectrum Selection for Home Phoneline Networking", ITU-T Study Group 15, NG-101, Nuremberg Germany, 2 Aug 1999.

[5] "Frequency Diverse Quadrature Amplitude Modulation", by Eric Ojard, in preparation.

HSD Traffic Behavior in HFC Networks and RDC Interconnects

David Brown, Wayne Ebel, Esteban Sandino and Oleh Sniezko
AT&T Broadband

Abstract

The high-speed data access provided by HFC networks has been a great success story. In many systems and municipalities, the penetration within the first year exceeded all expectations. Moreover, the early adopters were Internet and computer networking savvy and generated high traffic in both downstream and upstream directions of the HFC access network. The same high capacity utilization was experienced in the interconnects between CMTSs, proxy server locations and regional data centers (RDCs).

This paper analyzes capacity utilization of several major components of the high-speed data access in HFC networks:

- *access side of CMTSs (downstream and upstream channels),*
- *network side of CMTSs,*
- *interconnects between CMTS locations and proxy server locations,*
- *RDC LANs, and*
- *interconnects to global Internet.*

This analysis accounts for the number of customers served and for the user behavior. The purpose of this exercise is to develop simple tools for initial network design and capacity engineering for different levels of penetration and for different behavior of the users. The data for this analysis has been collected over a period of several months. This data by itself is interesting and representative of diurnal and weekly traffic patterns, and can be used for capacity engineering in networks shared by different user categories: residential and business.

INTRODUCTION

The explosive growth of the demand for high speed data (HSD) access services, although welcomed and partially anticipated by the HSD access network operators and HSD access service providers, introduced an element of surprise. Traffic generated by the early adopters did not follow the expected patterns. At low penetration rates on the access side of the HSD plant and unpredictable traffic patterns of the early adopters, the efficiency of caching and proxy servers in traffic containment and traffic load reduction was low. This could lead to unexpectedly high capacity utilization if this inefficiency were disregarded. This in turn could result in unexpected capacity exhaustion in an under-engineered HSD network segment.

As the penetration increases, the more predictable user behavior and traffic patterns prevail, and engineering of the interconnect capacity based on the average user behavior becomes well grounded. Moreover, it is also expected that the server and caching efficiency in traffic containment will increase as the number of customers served by the server location increases.

Other factors such as BER, rate shaping and service tiering, customer traffic patterns based on the customer type, and aggressiveness of the IP protocols must be considered and monitored beside using historical data for traffic and utilization prediction and capacity engineering. Therefore, traffic and capacity utilization monitoring as well as development of the demand-extrapolation tools must be continuous in nature.

The authors present several metrics and statistics that may be useful as simple predictive tools. However, these proposals are very preliminary. Although the tools must be simple and intuitively interpretable, they may become quite more accurate and adaptive with the use of today's data analysis and processing engines.

HSD NETWORK ARCHITECTURE EXAMPLES

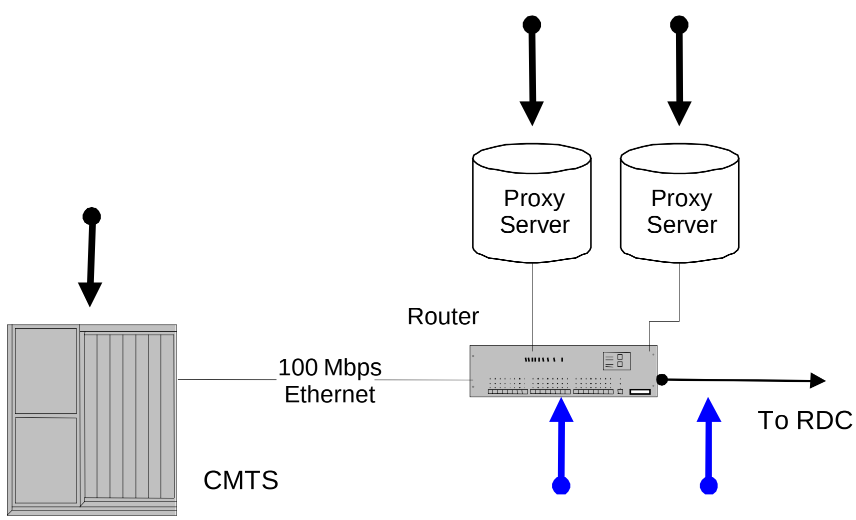
Three systems were selected for the analysis presented in this paper. The systems/markets were selected to represent three different sizes.

Single-CMTS Architecture

The first system (see Figure 1) is based on a single CMTS with collocated proxy servers. The interconnect with the RDC was initially engineered for a capacity of four T-1s and has been upgraded to 22 Mbps capacity during the period of collecting data for this paper.

An HSD NOC reports *In Data Rate* and *Out Data Rate* for the CMTS on a weekly basis. The *In Data Rate* statistics represent the traffic collected from either the larger Internet or from any local or proxy servers within the logical data network segment, and forwarded to customer cable modems (CMs). This is equivalent to "downstream traffic". The *Out Data Rate* statistics represent the traffic collected from customer cable modems within the HFC service areas and destined to either the global Internet or to any local or proxy servers within the logical data network segment. This is equivalent to "upstream traffic". The NOC also provides statistics for each downstream data transmitter and upstream data receiver on the access side of the HSD HFC plant.

Figure 1: HSD Architecture for Single-CMTS Network

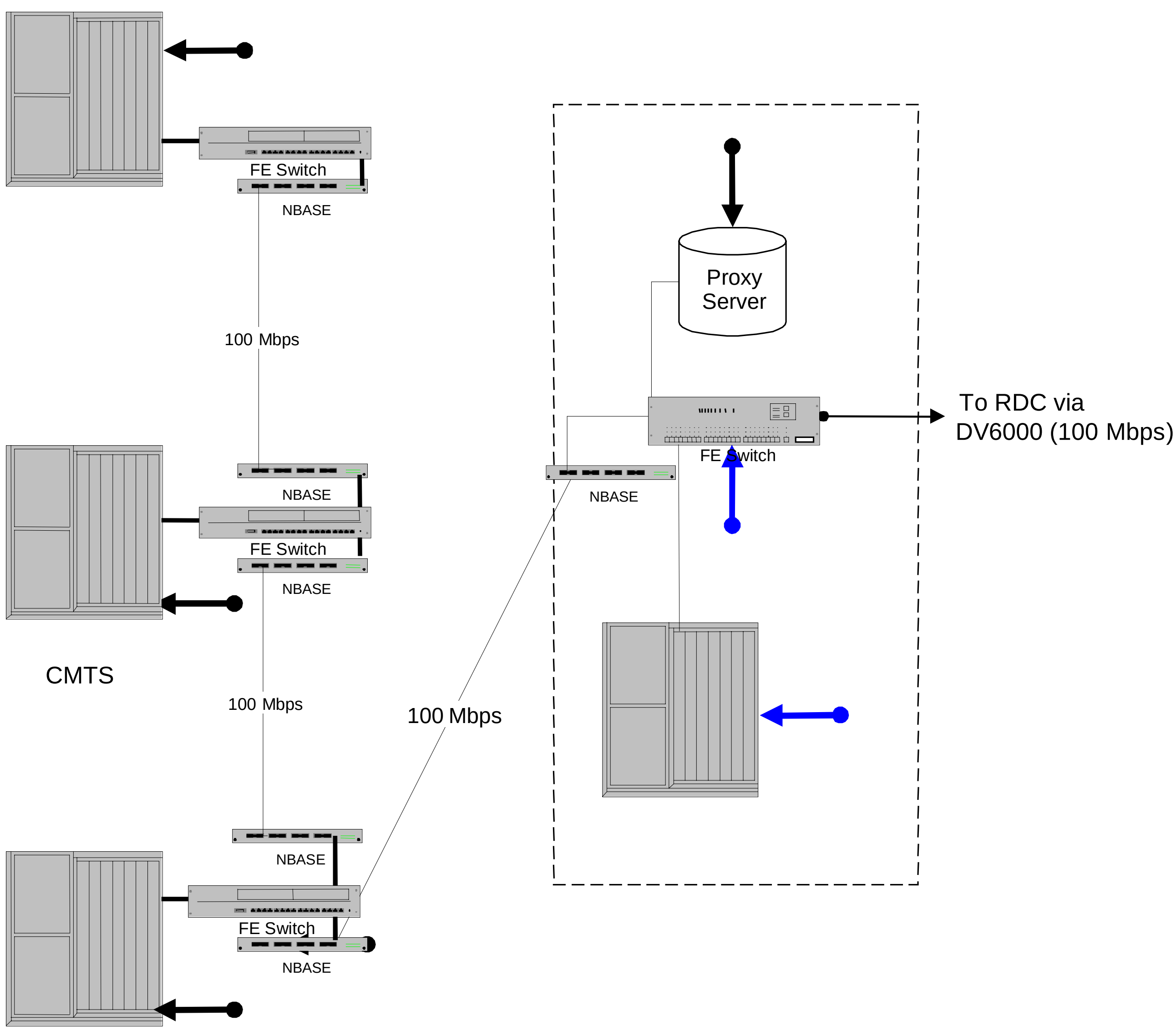


Medium Size CMTS Configuration

The second network (see Figure 2) consists of three CMTSs connected via a Fast Ethernet link with a proxy server location. This location is in turn interconnected with an RDC via a 100 Mbps EtherRing.

As described above, the HSD NOC reports *In Data Rate* and *Out Data Rate* for each CMTS 100BaseT input/output interface on a weekly basis.

Figure 2: Medium-Size CMTS Configuration



Large-Size HSD Network

An example of the large-size network consists of 12 headends and 32 CMTS units. However, traffic data for this network was available only for 20 of the CMTS units. Traffic statistics were also available for the Gigabit Ethernet internal RDC LAN at the master headend. The internal RDC LAN aggregates traffic for the entire market data network. This Gigabit RDC LAN also provides access to the global Internet for all customers in this market.

The distribution data network consists of a single 100BaseT Fast Ethernet backbone traversing from the northwest headend through all west headends and hubs (CMTS locations) to the master headend, and finally heading through east headends and hubs (CMTS locations) into the hub in the northeast service area. Additional 100BaseT spurs along this main route feed into the main backbone. There is a single proxy server at one headend in the West and two proxy servers at another headend in the East.

Logically, the main Fast Ethernet backbone is split into two main segments: West and East. The West Fast Ethernet segment services CMTSs in several headends and hubs in the West. All Internet content requests generated from customers within the West Ethernet backbone service area are first directed to the proxy server in one of these headends. At this location, a determination is made on whether to forward those requests to the RDC and the larger Internet cloud.

The East Fast Ethernet segment services CMTSs in several headends and hubs in the East. All Internet content requests generated from customers within the East Ethernet backbone service area are first directed to the two proxy servers in one of these headends. As is the case with the West portion of the Fast Ethernet backbone, the two proxy servers determine whether to forward content requests to the RDC and the larger Internet cloud.

As already described, for each CMTS 100BaseT input/output interface, the HSD NOC reports *In Data Rate* and *Out Data Rate* on a weekly basis.

The NOC also reports *In Data Rate* and *Out Data Rate* for the Gigabit Ethernet internal RDC LAN in the master headend. In this case, the *Out Data Rate* represents incoming traffic from the global Internet that is forwarded to both the West and East Fast Ethernet backbone segments for distribution to all customers in the market data network, i.e., “downstream traffic”. The *In Data Rate* represents incoming traffic originating from all customer data terminals arriving at the RDC from both the West and East Fast Ethernet backbone segments and destined to the global Internet, i.e., “upstream traffic”.

CMTS TRAFFIC ANALYSIS

Due to the concerns of the industry outsiders, the access network capacity engineering has been the focus of the HFC engineering effort, despite repeated reports on the results of simulation and traffic modeling that showed significant capacity margins in the access plant. These results can be found in several publications (two are referenced in this paper).

The traffic statistics presented in the next several subheadings will support the earlier results and will also allow to draw some preliminary conclusions that may be useful during capacity engineering effort and in capacity exhaustion predictions.

Single-CMTS System

Channel Utilization — Downstream

The CMTS in this market served a maximum of 1313 active modems during the week for which the traffic statistics are presented in Figure 3. The transmitter s6.p6 served a maximum of 820 active modems. As can be seen from Figure 3, downstream traffic did not reach at any point 50% of the capacity utilization (downstream capacity equal to approximately 27 Mbps).

After traffic aggregation on the network side of the CMTS, the maximum traffic levels were much below the capacity of the CMTS.

An interesting conclusion can be derived from the analysis of the downstream data rate per CM. It seems that the log-linear plot of this statistic (see Figure 5) fits very well a power trendline. However, the data represent too few observations to draw binding conclusions.

Channel Utilization — Upstream

The transmitter s6.p6 served a maximum of 820 active modems. As can be seen from Figure 4, upstream traffic did not significantly exceed 50% of the capacity utilization for any receiver at any point in time (upstream capacity equal to approximately 2.5 Mbps for each of the six upstream receivers).

The upstream data rate per CM does not follow the same pattern as data rate in the downstream direction. However, the upstream packet rate per CM (see Figure 4 for raw data plots) for this particular example follows similar pattern (see Figure 6) to the pattern for the data rate per CM in the downstream direction.

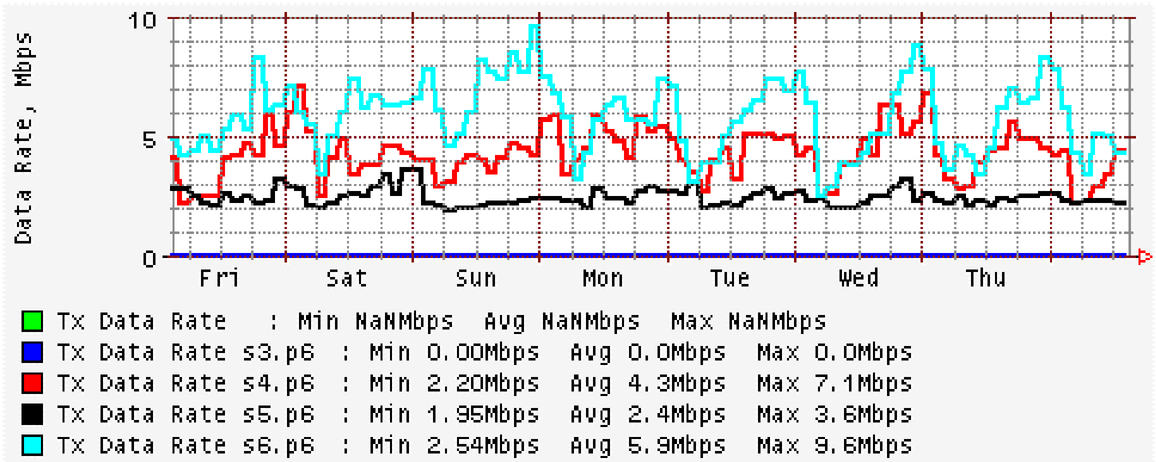
Diurnal and Weekly Traffic Patterns

The downstream traffic reaches the highest data rates at midnight, tapers down fast to reach a minimum at 6 a.m. and then starts increasing at a slower or faster rate to close the daily cycle at midnight. This pattern is discernable in both downstream traffic data on the access side and downstream aggregated traffic data on the network side of the CMTS as long as the number of users is sufficient to generate sizeable traffic and perform statistical traffic aggregation. This pattern is not clearly visible on the access side for the upstream traffic. However, the aggregated data rates for the upstream traffic on the network side as well as the packet rates for the upstream traffic on both sides of the CMTS show similar patterns. The difference for the upstream data rate diurnal behavior on the access side can be explained by statistically smaller group of users and by higher variability of the packet sizes in the upstream direction. In most cases, the downstream and upstream data and packet rate traffic reaches its peak during weekends. Moreover during weekends, the traffic reaches high levels already at noon.

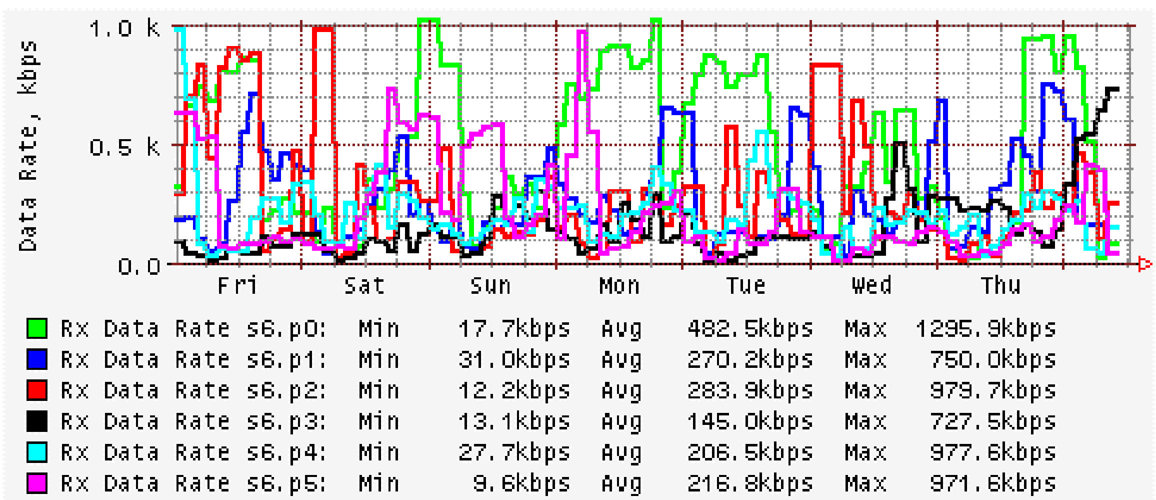
To engineer the access side of the HSD plant, one must account for the peak values of the data rates unless traffic shaping and throttling schemes are implemented either for downstream or upstream traffic or in both directions.

Figure 3: **Single-CMTS System Data Rate Statistics**

a) **Access Side of CMTS — Downstream**



b) **Access Side of CMTS — Upstream Receivers Corresponding to TXs6.p6**



c) **Network Side of CMTS**

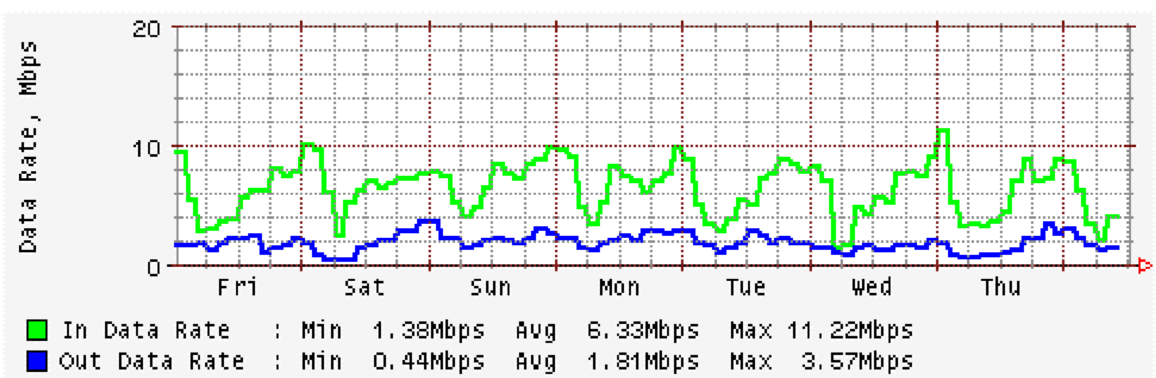
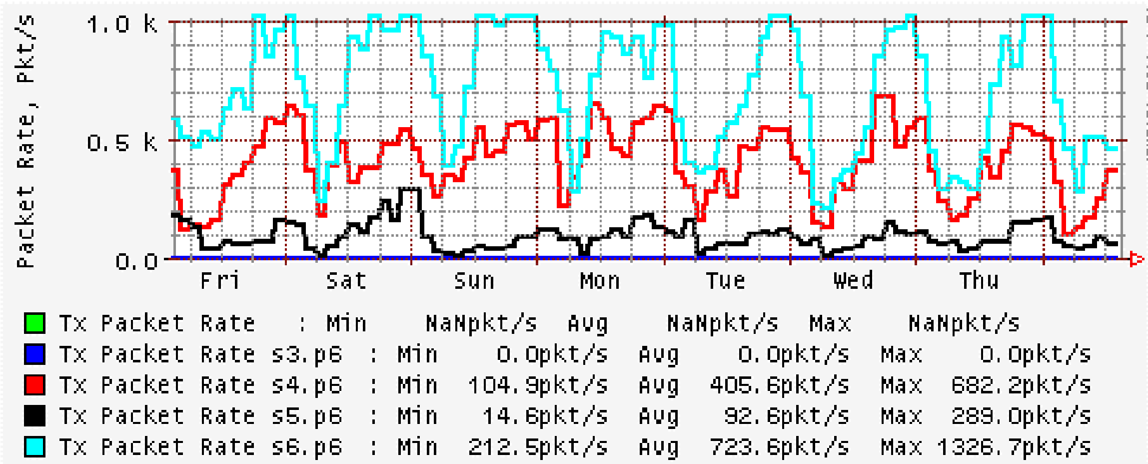
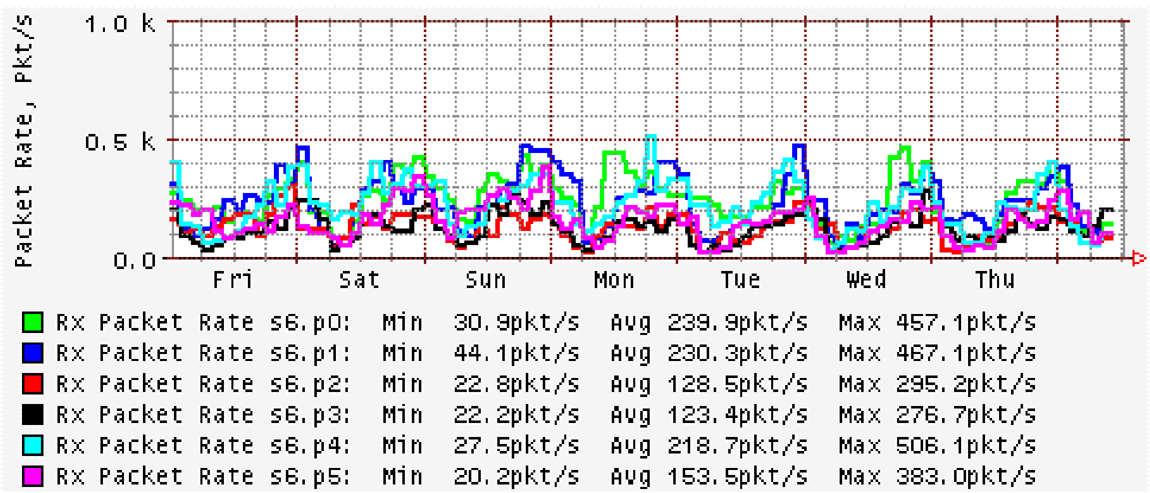


Figure 4: **Single-CMTS System Packet Rate Statistics**

a) **Access Side of CMTS — Downstream**



b) **Access Side of CMTS — Upstream Receivers Corresponding to TXs6.p6**



c) **Network Side of CMTS**

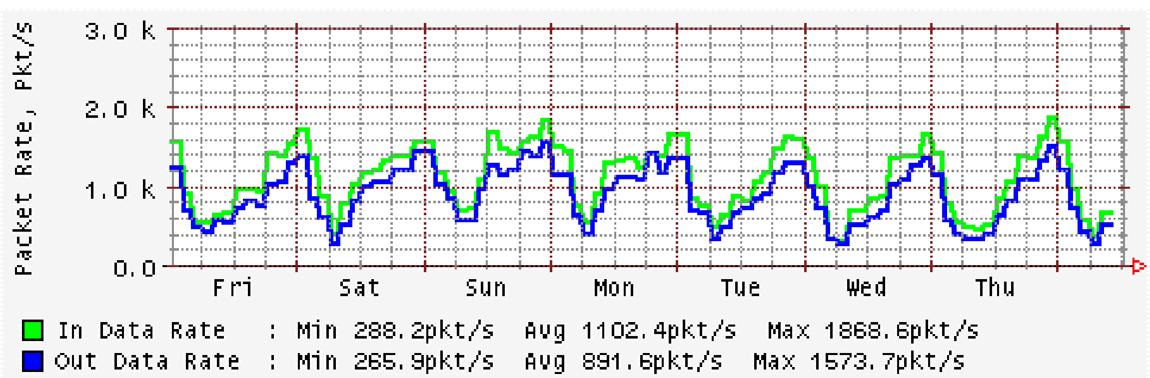


Figure 5: Downstream Data Rate/CM Behavior

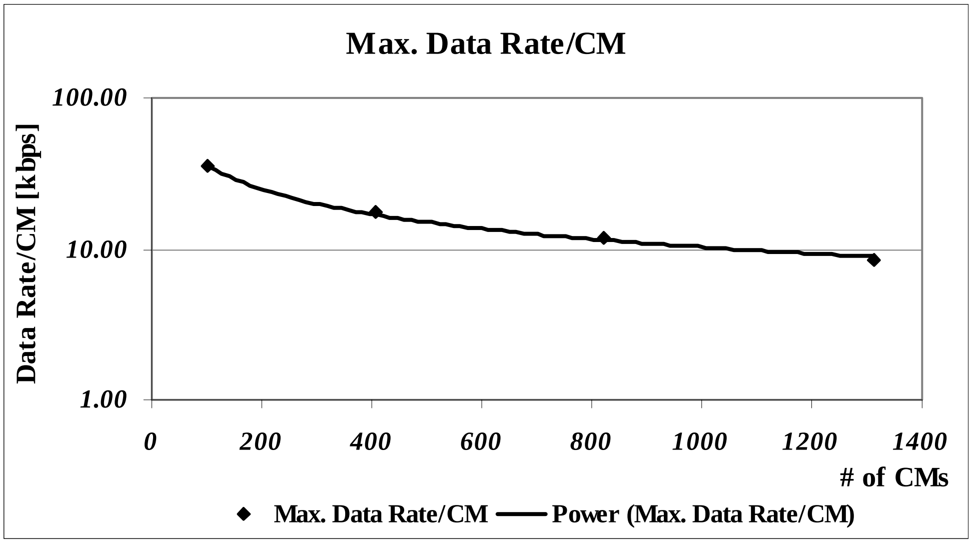
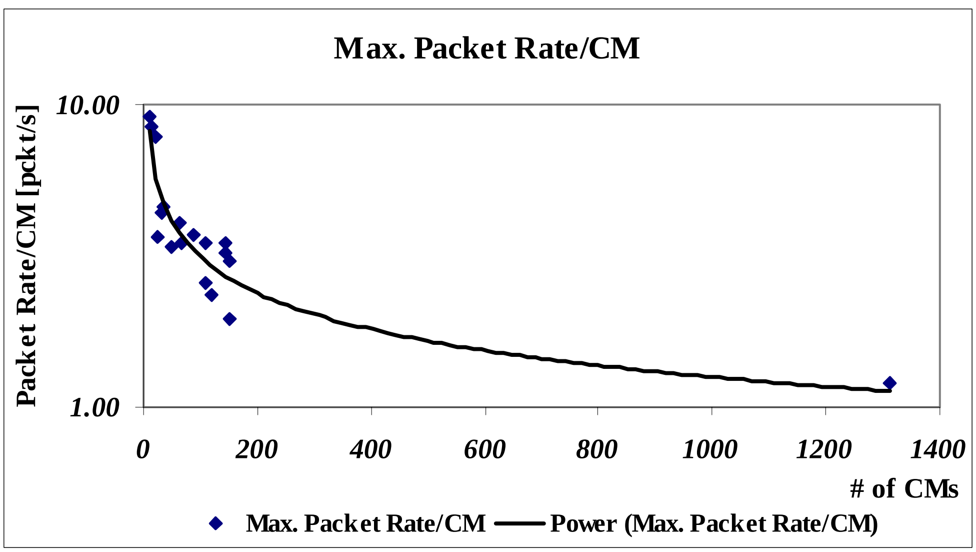


Figure 6: Upstream Packet/CM Behavior



Medium-Size HSD Configuration

The traffic patterns from each CMTS have similar characteristics as the traffic characteristics for the CMTS in the single-CMTS system. There are some small differences in the upstream traffic patterns that will be analyzed later in this paper.

Of interest is the data presenting statistics on the traffic in the link between the RDC and the proxy server location for this system. These traffic records show that the diurnal and weekly patterns described above for a single CMTS are even more pronounced at higher levels of traffic aggregation. Moreover, the aggregate traffic reached above 20 Mbps (peak value of 5-minute averages) over the period of six months since the service launch. The number of active modems reached 4,000 over that period of time.

Downstream Data Rate/CM Behavior

The results of the regression analysis show that there is a reasonably good fit between the collected data points and power trendline described by the following equation:

$$y = cx^b$$

where *c* and *b* are constants.

In the case depicted in Figure 7, the data rate per CM drops approximately by a factor of two for each quadrupling of the number of the active CMs (total data rates increase twice for each quadrupling of the CM number). For 1,000 of CMs, total data rate reaches 10 Mbps (10 kbps/CM); for 4,000 CMs, total data rate reaches approximately 20 Mbps (5 kbps/CM). These statistics should be continuously verified as the customer behavior may change as well as new, more bandwidth demanding applications may start dominating the downstream traffic. However, they can be used today for capacity engineering of the access and network side of the CMTSs.

Figure 7: Downstream Data Rate/CM Behavior for Large Sample of Observations

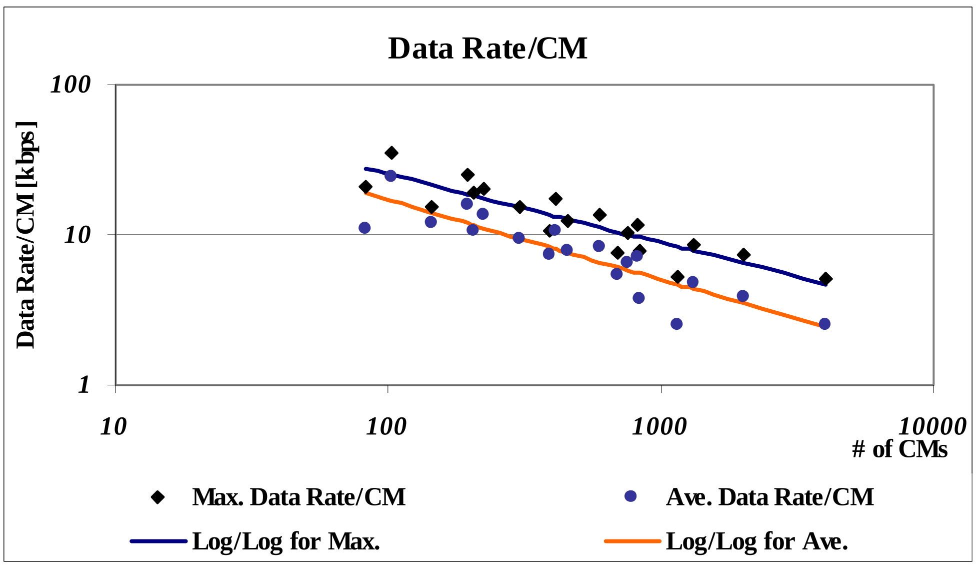


Table 1: Linear Regression Analysis Results for Max. Downstream Data Rates (Log-Log Scales)

Regression Statistics						
Multiple R	0.8915					
R Square	0.7947					
Adjusted R Square	0.7827					
Standard Error	0.1064					
Observations	19					
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	0.7457	0.7457	65.81875	3.01446E-07	
Residual	17	0.1926	0.0113			
Total	18	0.9383				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.3264	0.1537	15.1401	2.67E-11	2.0022	2.6506
X Variable 1	-0.4602	0.0567	-8.1129	3.01E-07	-0.5799	-0.3405

Upstream Packet/CM Behavior

The results of the regression analysis for the upstream traffic are less conclusive. There is no good fit for data rates as a function of the active CM number. There is only acceptable fit between the packet rate and the number of active CMs (see Figure8). The only convincing conclusion based on the data collected is that, for CM numbers exceeding 500, the packet rate is approximately 1 pkt/s/CM and the data rate does not usually exceed 3 kbps/CM. These numbers can be used for capacity engineering, especially when supported by the previous conclusion that for the number of CMs approaching 1,000 units, upstream channel capacity utilization rarely exceeds 50%. With four to six channels per single downstream channel of 27 Mbps (this channel can serve approximately 6,000 active modems), there is sufficient capacity of 10 to 15 Mbps in the upstream direction. Based on the numbers listed above, this capacity could serve 8,000 to 12,000 active modems.

Figure 8: Upstream Packet Rate/CM Behavior for Large Sample of Observations

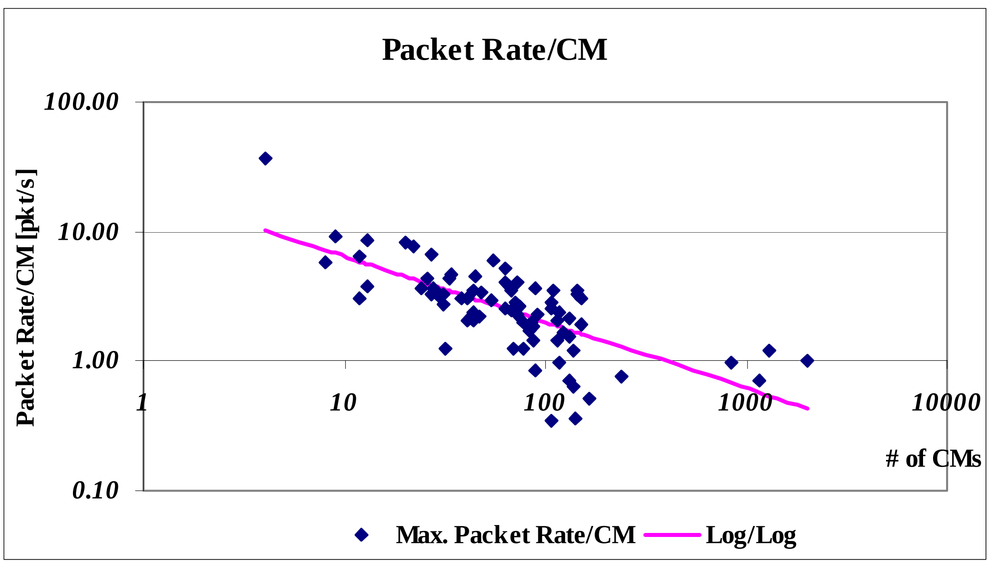


Table 2: Linear Regression Analysis Results for Max. Upstream Packet Rates (Log/Log Scales)

Regression Statistics	
Multiple R	0.7052
R Square	0.4973
Adjusted R Square	0.4906
Standard Error	0.2375
Observations	78

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4.2412	4.2412	75.1696	5.72298E-13
Residual	76	4.2881	0.0564		
Total	77	8.5293			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.3189	0.1101	11.9792	3.51E-19	1.0996	1.5382
X Variable 1	-0.5096	0.0588	-8.6700	5.72E-13	-0.6266	-0.3925

Assumptions and Definitions

The downstream channel capacity is estimated for 6 MHz 64 QAM DOCSIS channel, the upstream capacity is estimated for 1.6 MHz QPSK DOCSIS channel.

All the conclusions for the downstream and active traffic and customer behavior are based on historical data and are may be dependent on other factors not included in the analysis.

TRAFFIC AGGREGATION BY PROXY SERVERS IN LARGE-SIZE MARKET

Available Data

Data for the large-size market has been collected from individual 100BaseT interfaces at each of the CMTSs polled. As such, this data reflects the raw traffic before the local headend data routers direct it to either the proxy servers (if available) or into the appropriate 100BaseT Fast Ethernet backbone segment. The 100BaseT Fast Ethernet backbone segments carry all traffic into the master headend and RDC location. The RDC location is the main gateway to access the global Internet. Access to the Internet cloud is via two OC-3 packet-over-Sonet (POS) interfaces.

The traffic statistics for the Gigabit Ethernet internal RDC LAN that aggregates all incoming data from the 100BaseT Fast Ethernet backbone are also available. Unfortunately there was no meaningful data available from the proxy servers in this market at the time the data was collected.

Data Analysis Results

From among 32 CMTSs in this market, data was available for 20 units only. The resulting estimated traffic statistics database allows for aggregation of the specific traffic sources within the data network.

Traffic aggregation has been done separately for the sources feeding the West Fast Ethernet and the East Fast Ethernet logical backbone segments. This allowed for an estimate of the current utilization for each of these two segments. Traffic aggregation for all combined sources has also been done. This latter aggregation allowed for direct comparison against the reported traffic statistics available for the internal RDC Gigabit LAN. It also allowed for an assessment of how effectively the proxy servers at both West and East proxy server locations perform traffic containment and traffic load reduction on each of the two Fast Ethernet logical backbone segments. The following sections summarize the findings to date.

Utilization of West and East Fast Ethernet Backbone Segments

Figures 9 and 12 for the West and East Fast Ethernet backbone segments respectively illustrate the Internet traffic transported from the RDC Gigabit LAN to each CMTS. The West Fast Ethernet backbone peaks at 90% utilization (90 Mbps). This peak utilization starts at around 6:00 p.m. everyday and reaches its maximum level of 90 Mbps at just before midnight. Utilization tapers off rapidly after that to its minimum level at around 6:00 a.m. everyday. The East Fast Ethernet backbone has a worst-case peak of 100% utilization (100 Mbps). A similar pattern of traffic utilization is present here.

The charts in Figures 9 and 12 indicate that both Fast Ethernet backbone segments are heavily utilized. Figures 10 and 13 for the two Fast Ethernet backbone segments illustrate the traffic transported from each CMTS location to the RDC Gigabit LAN and destined to the global Internet (upstream traffic). This upstream traffic utilization is very uniform. Maximum utilization for this case is less than 20 Mbps for the West Fast Ethernet backbone and less than 30 Mbps for the East Fast Ethernet backbone. Figures 11 and 14 illustrate the ratio of downstream to upstream traffic for the West and East Fast Ethernet backbone segments respectively and have been included here for information purposes only. The observed ratios fall between 3:1 and 6:1. The higher ratios appear to track the volume of traffic and occur at approximately the same time as the peak utilization for each Fast Ethernet backbone segment.

Figure 9: Aggregate Downstream Traffic for West Backbone

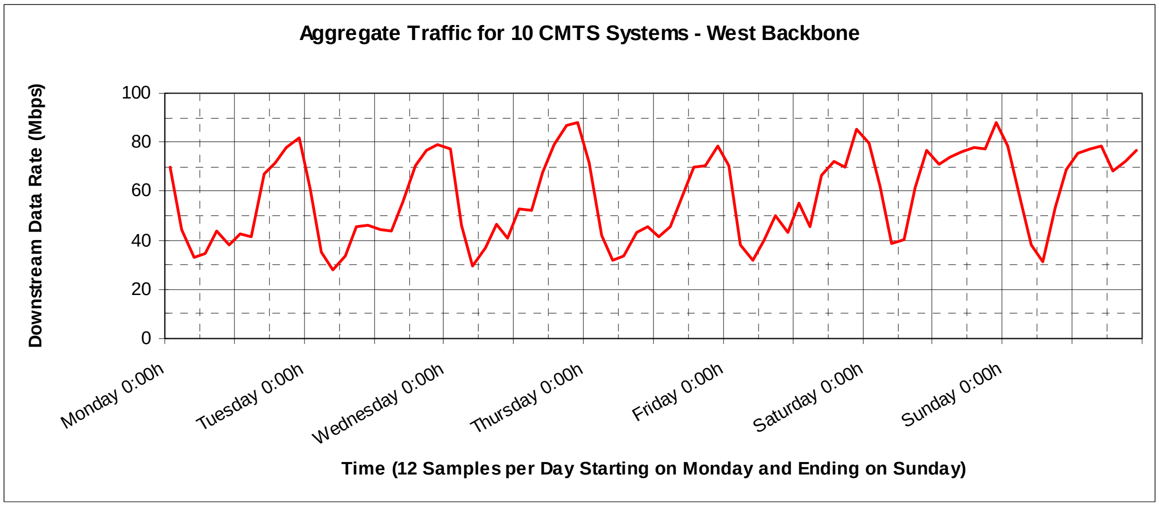


Figure 10: Aggregate Upstream Traffic for West Backbone

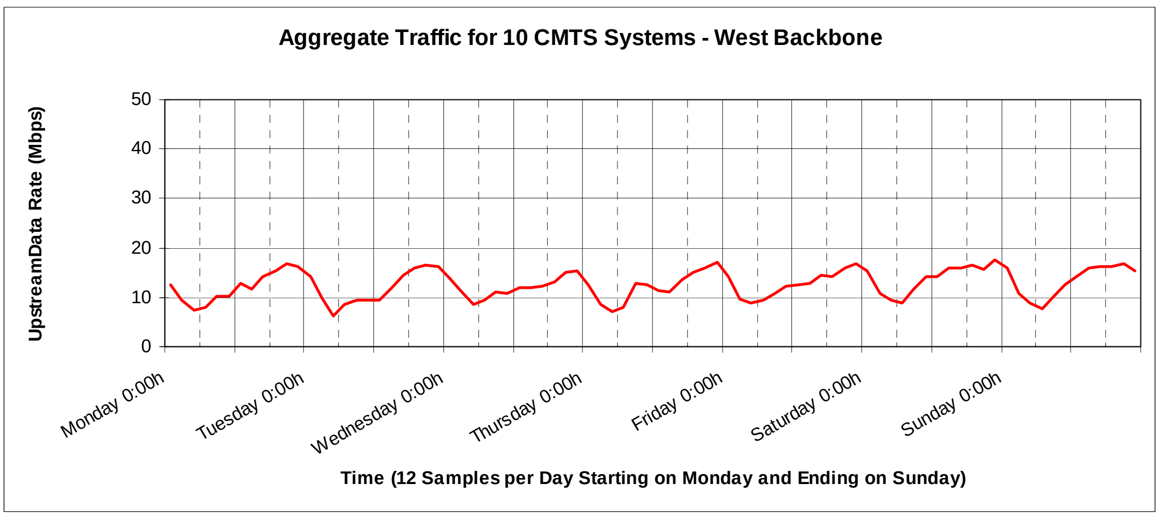


Figure 11: Downstream-to-Upstream Data Rate Ratios – West Backbone

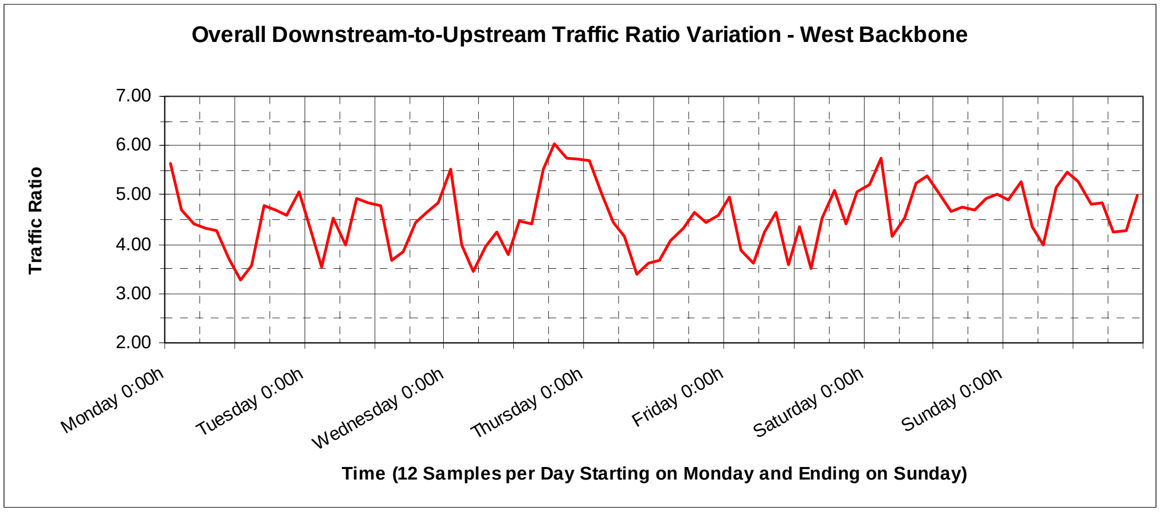


Figure 12: Aggregate Downstream Traffic for Eat Backbone

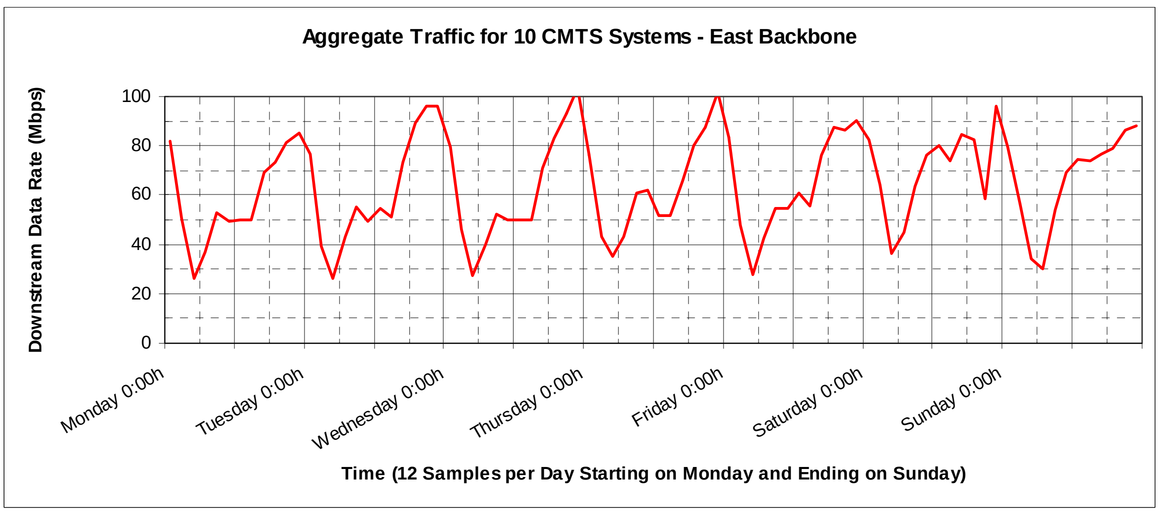


Figure 13: Aggregate Upstream Traffic for East Backbone

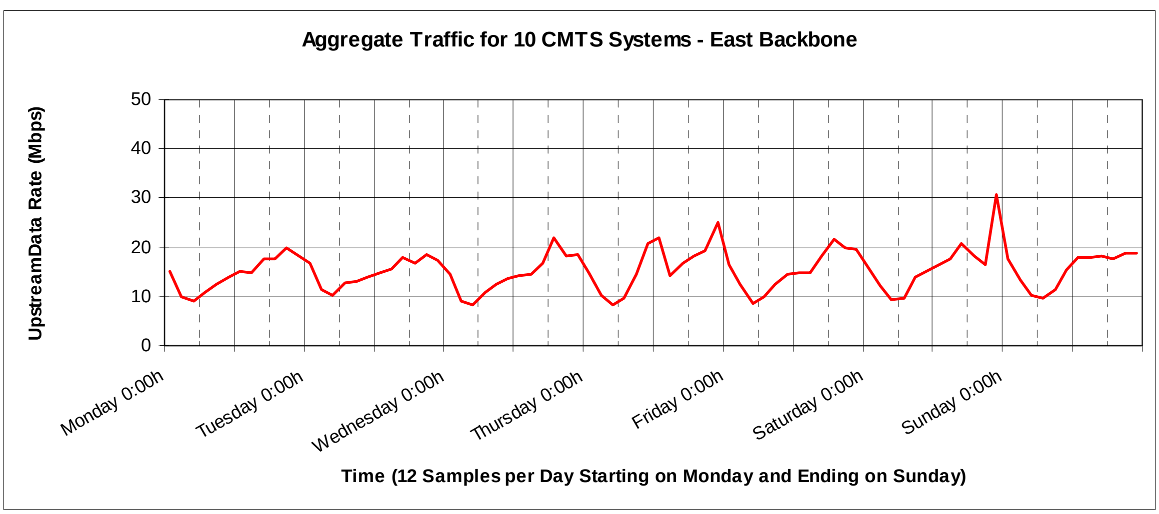
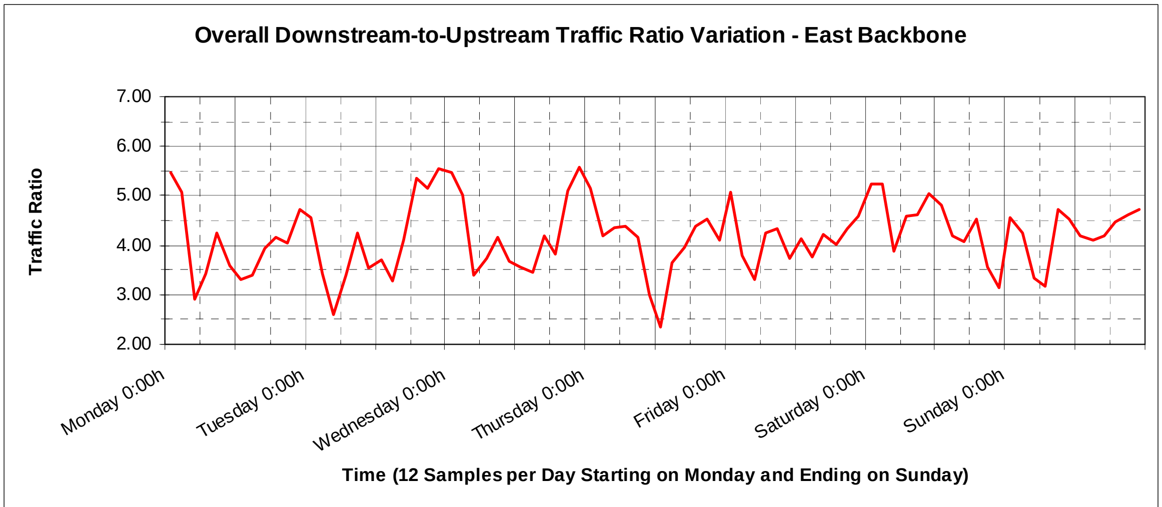


Figure 14: Downstream-to-Upstream Data Rate Ratios – East Backbone



T

Utilization of Internal RDC Gigabit LAN

Figure 15 shows the aggregation of all the traffic entering each of the monitored 100BaseT interfaces for all 20 CMTS units under study, i.e., downstream traffic into the subscriber modems. This is the aggregation of all the reported *In Data Rate* traffic in the charts for the CMTS units.

Figure 16 shows the aggregation of all the traffic leaving each of the monitored 100BaseT interfaces for all 20 CMTS units under study, i.e., upstream traffic generated from all the subscriber modems. This is the aggregation of all the reported *Out Data Rate* traffic in the charts for the CMTS units.

Figure 17 shows a representative weekly traffic chart tracking utilization for the Gigabit Ethernet internal RDC LAN. The Out traffic trace should represent all incoming Internet data traffic leaving the RDC LAN for distribution into each of the CMTS units via the 100BaseT Fast Ethernet backbone segments. The In traffic trace should represent aggregation of all data traffic entering the RDC LAN from the 100BaseT Fast Ethernet backbone segments and destined for the global Internet.

The proxy servers in the network are intended to contain Internet-related traffic (web page traffic) within specific sub-network domains and away from the 100BaseT Fast Ethernet backbone segments. The objective is to minimize the amount of duplicate packets on each of the backbone segments and eliminate duplicate user requests to access the same information over a period of time. The proxy servers in this scenario would store the most requested Web pages and Internet content locally. Local storage closer to the end user should also result in faster access to the requested information, i.e., reduced delays.

Under the current scenario, we would expect to find the utilization of the Gigabit Ethernet internal RDC LAN for both the *Out* and the *In* traces to be somewhat less than what we would find if straight traffic aggregation from all CMTS sources took place. The proxy servers would be expected to limit traffic on each of the Fast Ethernet backbone segments feeding into the RDC LAN. The data collected so far does not support this expectation. A quick visual check of Figures 15 through 17 shows that almost all In and Out traffic for the Gigabit RDC LAN is aggregated in a straightforward way. However, the above analysis is by no means complete. We are missing some of the traffic contributions for the additional CMTS sources. The current analysis represents only 20 out of the 32 CMTS unit universe in the distribution network.

Fast Ethernet Backbone Utilization

Another interesting finding is that, although a Gigabit Ethernet LAN is implemented at the RDC, the two Fast Ethernet backbone segments connecting all CMTS traffic aggregation points are still point-to-point Fast Ethernet links. From the RDC perspective, there are two separate 100BaseT links feeding traffic into the Gigabit Ethernet LAN. However, these two Fast Ethernet links are approaching the limit of their capacity.

Figure 15: Aggregate Downstream Traffic for 20 CMTSs

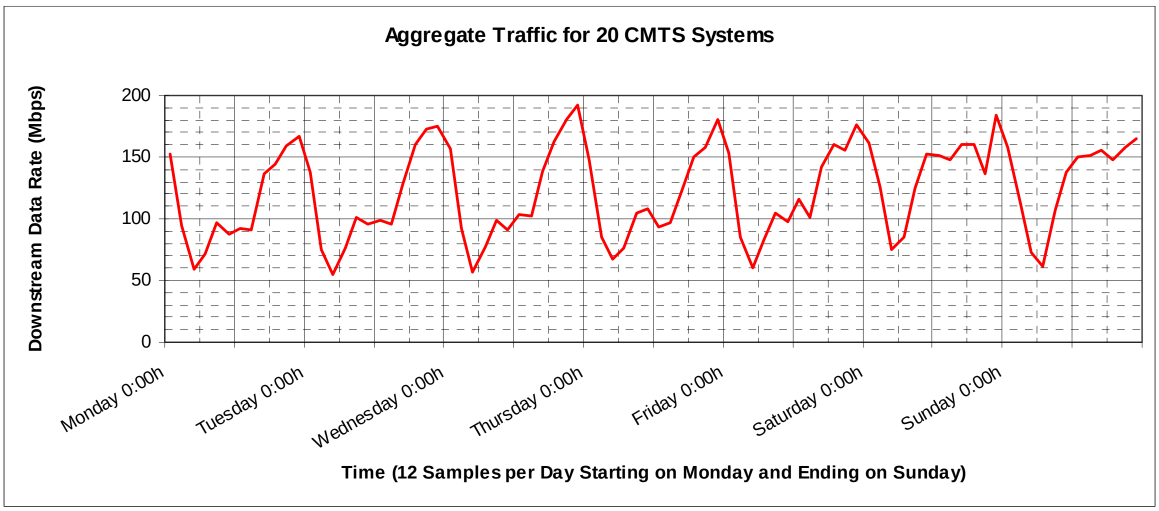


Figure 16: Aggregate Upstream Traffic for 20 CMTSs

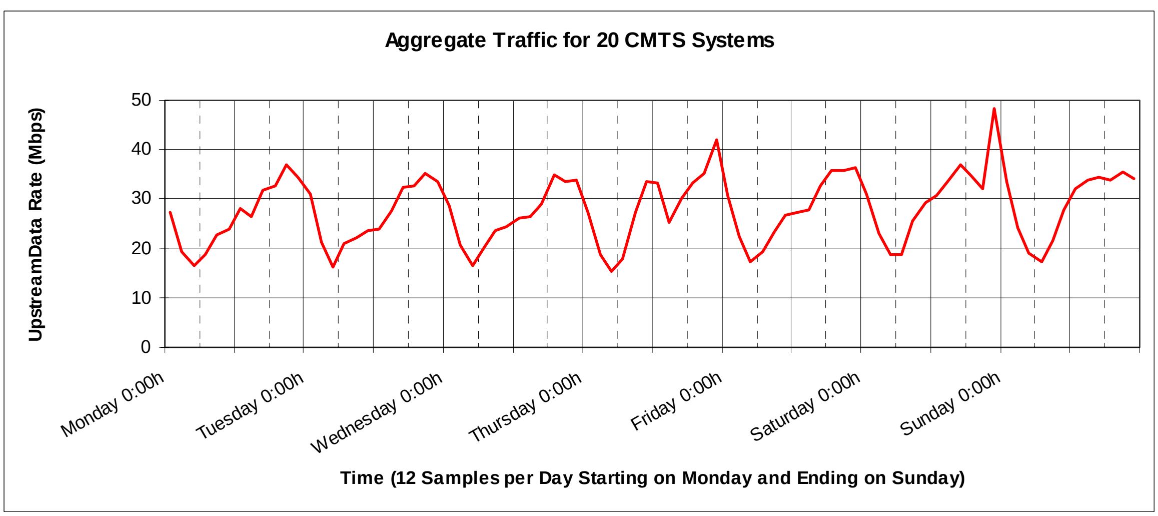
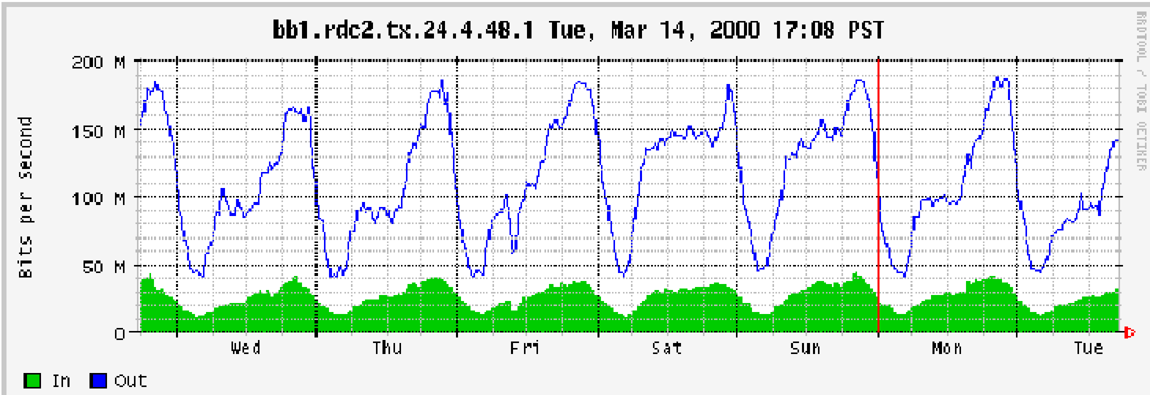


Figure 17: Traffic Activity in Gigabit Ethernet RDC LAN



TRAFFIC STATISTICS ANALYSIS

Table 3 below contains some data rate and packet rate statistics that can be useful in capacity engineering.

Table 3: Simple Traffic Statistics

Statistic	Average Value	Standard Deviation
Downstream/Upstream Packet Rate Ratio	1.20	0.0571
Downstream/Upstream Data Rate Ratio	3.20	0.70
Downstream Packet Size (Network Side)	746 bytes	42 bytes
Downstream Packet Size (Access Side)	1295 bytes	277 bytes
Traffic Aggregation Gain for Downstream Packet Rates	121%	8%
Traffic Aggregation Gain for Downstream Data Rates	224%	55%
Upstream Packet Size (Network Side)	305 bytes	60 bytes
Traffic Aggregation Gain for Upstream Packet Rates	224%	91%
Traffic Aggregation Gain for Upstream Data Rates	360%	57%

The data shows that the upstream traffic aggregation gains (also called the multiplexing gains) are much higher than those for downstream traffic are. This fact indicates that upstream packet rate and data rate peaks from different customer groups do not coincide in time as well as the downstream rates do. Moreover, these gains in both direction are much lower than previously expected (usually for traditional LAN applications, the access to trunk capacity ratios are much higher). This may be caused by the fact that traffic peaks from different customers coincide in time.

Also, traffic rate asymmetry is much lower than expected. For peak and average traffic, the asymmetry for packet rates is 1.2:1 and for data rates is 3.2:1 downstream to upstream.

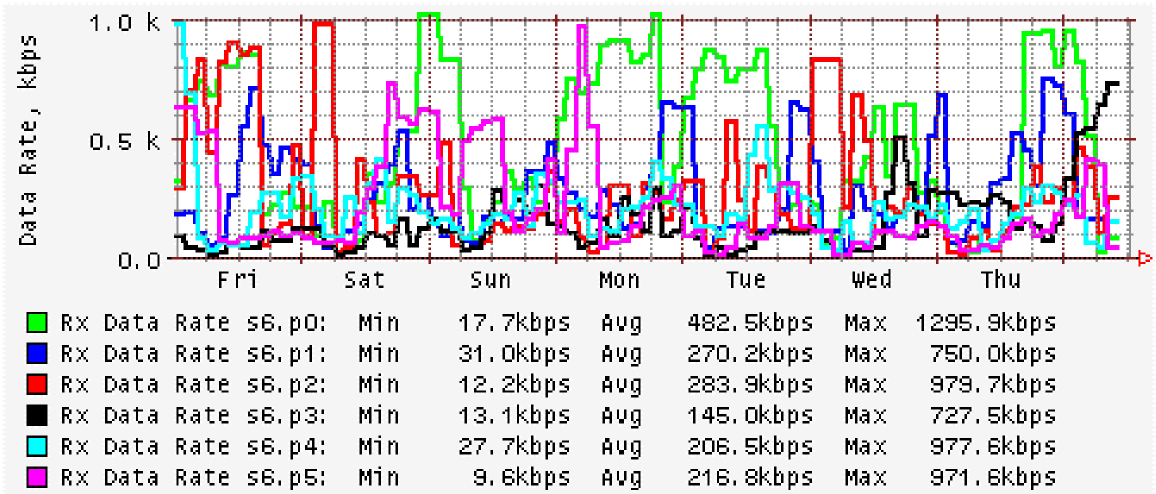
OTHER FACTORS TO CONSIDER

BER Impact

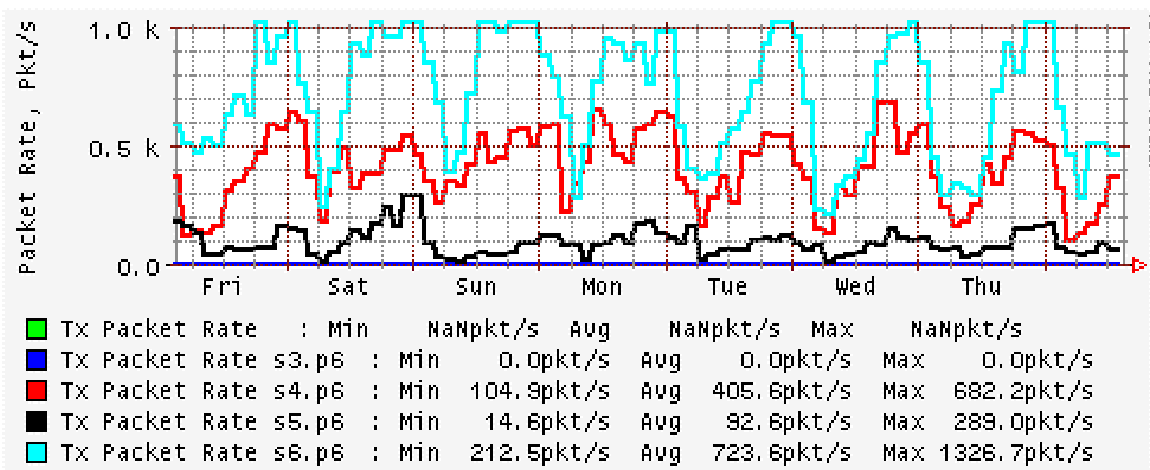
Figure 18 shows an example of the access network with very low BER and FEC activity. Figure 9, on the other hand, shows an example of the access network with very high BER and FEC activity. The impact on the upstream traffic pattern and behavior is visible. Some impact can be also observed in the downstream traffic (not shown in Figures 18 and 19).

Figure 18: Low BER and FEC Activity Access Plant

a) Data Rate



b) Packet Rate



c) BER and FEC Activity

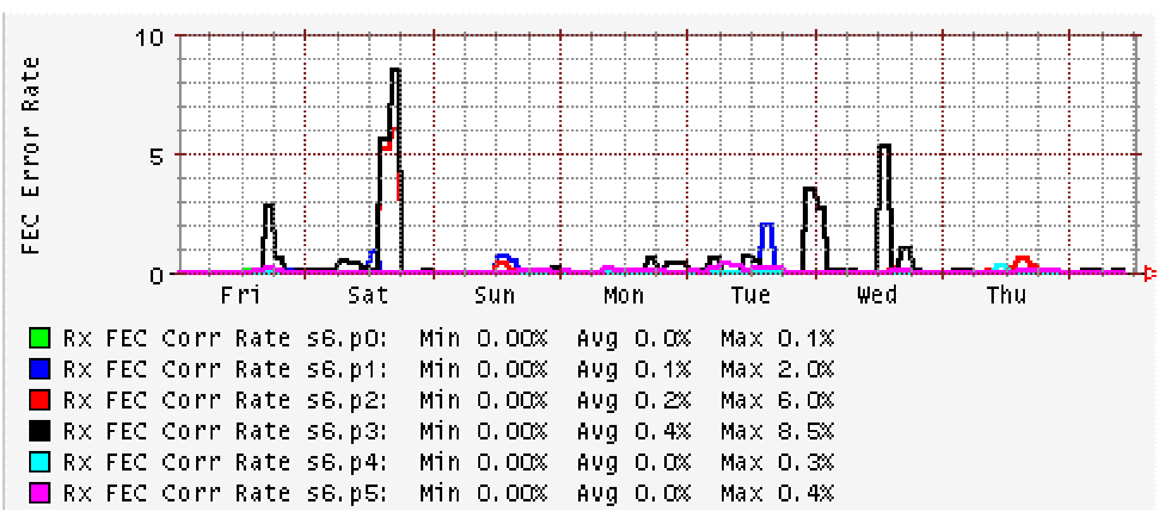
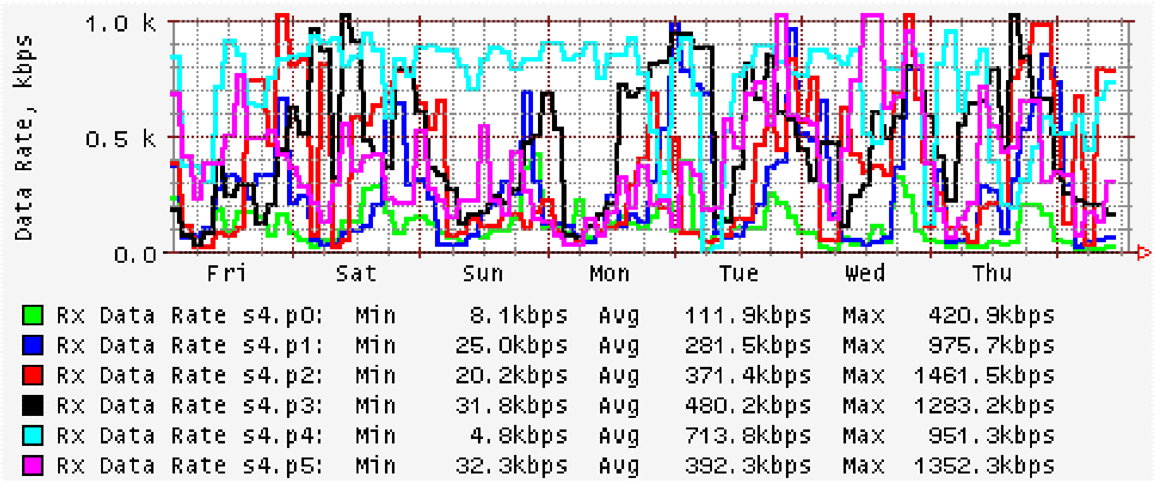
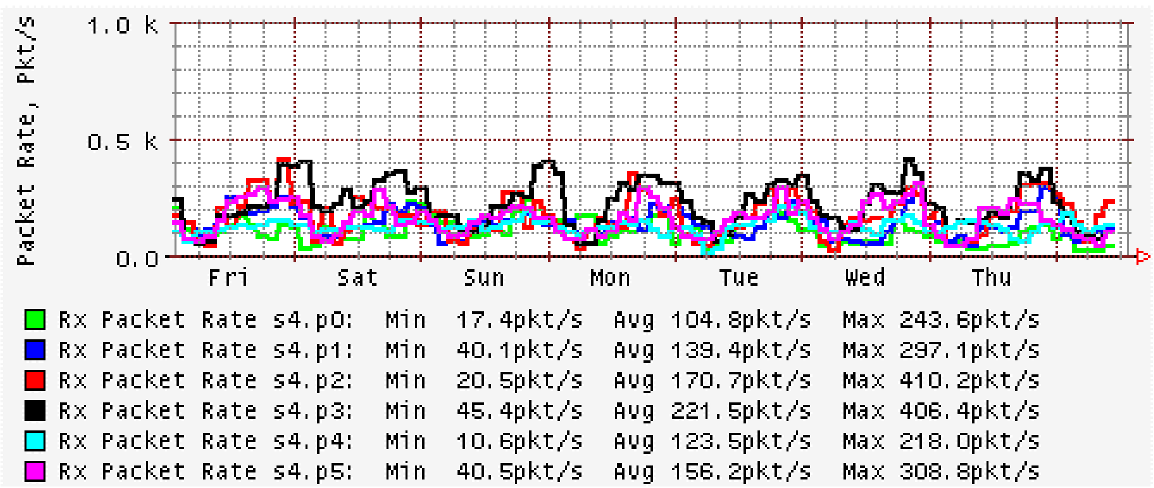


Figure 19: High BER and FEC Activity Access Plant

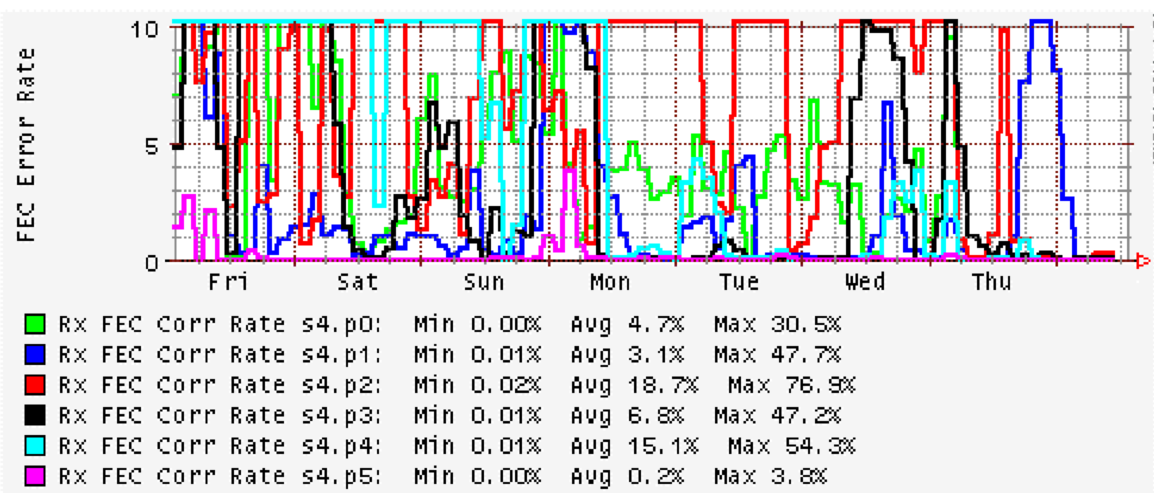
a) Data Rate



b) Packet Rate



c) BER and FEC Activity



Capacity Exhaustion

Figure 20 shows data rate on the network interface of the CMTS connected to the RDC via a link with 100% utilization for most of the time. Its impact on traffic is visible in flattening the diurnal patterns. After the link upgrade, the typical diurnal patterns were restored.

Applications

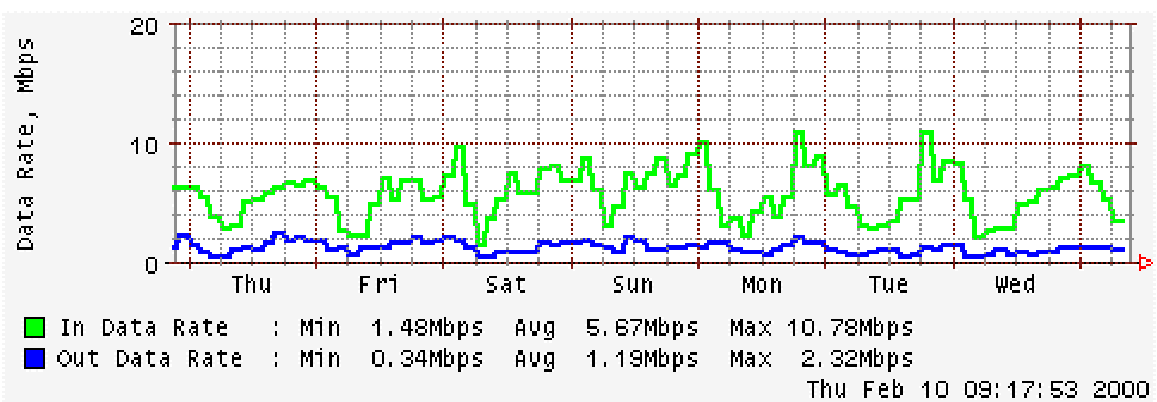
Downstream traffic data rate peaks and packet rate peaks coincide with each other almost perfectly. The downstream packet sizes are more uniform (see Table 3). This is not the case in the upstream traffic where data rate peaks do not coincide with packet rate peaks (compare Figures 3 and 4). This may be caused by different packet composition in the upstream direction as well as by high BER and packet retransmission rates.

Miscellaneous

Other factors such as traffic shaping and service tiering as well as data rate throttling may soon start playing a dominant role in influencing traffic rate patterns and behavior. However, they were not included in the analysis presented in this paper.

New applications and more aggressive protocols (for media streaming) may also affect the traffic patterns and behavior. For example, if the media streaming will dominate in the downstream (as opposed to upstream), the asymmetry of data rates and especially packet rates may increase significantly. Other traffic statistics may also be affected (for example, packet sizes and data/packet rates per CM).

Figure 20: Capacity Exhaustion Influence on Diurnal Traffic Patterns



CONCLUSION

The first step to understanding all the elements of the HSD networks in the metro areas is to monitor the traffic behavior in its:

1. CMTS access and network interfaces,
2. Interconnects between CMTS locations and proxy server locations,
3. RDC LANs, and
4. Interconnects to global Internet.

Based on traffic behavior and pattern databases, a simple set of data rate, packet rate and traffic statistics and metrics can be developed for capacity engineering of all the HSD network elements listed above. Moreover, trend monitoring will allow for early adjustments of the statistics and discovery of unusual traffic patterns and behaviors. As an example, this monitoring can and should be used in elimination of high BER.

This paper presented preliminary results of the traffic database analysis from several HSD systems of different sizes. Although further analysis is still required, the paper presented some analysis tools as well as the analysis results. The most useful are:

1. Downstream data rate/CM of 5 kbps for the number of CMs higher than 1,000;
2. Ratio of downstream-to-upstream data rates of 3.2:1 for weakly peak traffic values (not necessarily coincidental in time);
3. Significant multiplexing gains (albeit lower than expected) for upstream traffic after the traffic integration from the access side to the network side of the CMTSs (in excess of 300%) and some gains for the downstream traffic (in excess of 200%).

These results are based on historical data and their application must always be accompanied by verification whether the traffic patterns and behavior assumptions remain the same. The data collected proved that the capacity of the downstream and upstream DOCSIS HFC channels can support up to 5,000 active modems. On the other hand, the results of the analysis on effectiveness of the proxy servers and information caching in traffic containment and traffic load reduction are inconclusive and further analysis on richer traffic parameter databases is warranted.

REFERENCES

1. Dan Pike and Tony E. Werner, "On Plant Renewal Strategies," 1998 NCTA Technical Papers.
2. Harold S. Fluss, "Effective Performance of Shared Bandwidth Data Channels in Hybrid Fiber/Coax Networks," 1999 Emerging Technologies Conference.

THE AUTHORS

David Brown

David Brown, M.TEL., University of Denver, B.S.E.E., University of Colorado at Denver, is a Director for Broadband Engineering at AT&T Broadband (formally TCI Communications). He has over 15 years of experience in distributed communications architecture and engineering. He has held engineering and telecommunications positions at Martin Marietta Corporation, now Lockheed Martin Corporation. He is also a member of IEEE and ACTE.

Wayne Ebel

Wayne T. Ebel is the Director of Network Implementation of High Speed Data at AT&T Broadband (formerly TCI Communications). He has 20 years of telecommunications experience in HFC operations, engineering, and advanced services. Prior to joining TCI, he worked at Pacific Bell where he was involved in the deployment of advanced services for HFC and MMDS. He is actively involved in SCTE as former Chapter Secretary and DOCSIS speaker.

Esteban Sandino

Esteban Sandino, Bachelor in Engineering Physics, McMaster University in Hamilton, Ontario, is a Director of HFC Engineering at AT&T Broadband (formally TCI Communications). He joined Rogers Cablesystems in Toronto, Canada in 1989 and was later involved with CableLabs in the initial effort to characterize two-way cable plant for high-speed data applications. Esteban is also a chairman of the Hybrid Management Sublayer (HMS) subcommittee under the SCTE Engineering committee for the integration and management of HFC outside plant network elements.

Oleh J. Sniezko

Oleh Sniezko, MBA, York University in North York, Canada, B.S.E.E, Wroclaw Polytechnic

University, Poland, is the Vice President of Engineering for AT&T Broadband (formerly TCI Communications). Prior to joining AT&T, Oleh worked 8 years at Rogers Engineering in Toronto. For his achievements in implementing fiber optic technology, Oleh received the 1997 Polaris Award sponsored by SCTE, Corning, and CED. Oleh is a licensed Professional Engineer in Ontario and a long-term member of IEEE and SCTE. He has been active in a series of CableLabs and NCTA projects, and represents the cable TV industry in NEC Panel 16.

Improve Plant Performance with Laser Clipping Suppression

Robert L. Howald and Timothy J. Brophy
Motorola Broadband Communications Sector

Introduction

Forward path digital loading has become common in HFC signal multiplexes over the past several years. The added bandwidth on HFC systems requires not just improved frequency response from the equipment in the distribution chain, but it also requires hardware with better dynamic range. As the forward path signal load increases with heavy digital loading, the total output power that must be achieved without degrading distortion is increased. Depending on plant layout, the total power could also be shared by backing down the analog load as a trade-off with the increasing digital load's effects. However, with slightly lower analog levels, CNR would degrade unless a lower noise floor exists. Thus, ideally, devices with improved dynamic range yield the most flexibility.

Towards this end, linear optics has improved significantly as the plant engineering focus has shifted to more sophisticated architectures spanning large geographical regions. Product families have been continually growing to support longer link lengths with higher power lasers and Erbium-doped fiber amplifiers (EDFA's). Lower noise lasers and ITU grid wavelengths also count as important achievements. Broader RF bandwidths and improved RF noise and distortion characteristics have also occurred in this time frame. This paper discusses a technique which crosses into the realm of both RF and optical to achieve its goals. This technique, laser clipping suppression, uses RF sensing to provide dynamic laser bias adjustments. This approach mitigates the laser's dominant impairment relative to digital signal carriage – laser clipping events – by shifting the laser's bias point at the onset of large RF envelope variations, which might otherwise introduce clipping.

Forward Path Operating Point and the Effect of Clipping on Analog and Digital Signals

Linear optics carries a composite signal multiplex for advanced HFC systems. The multiplex consists of traditional analog video channels, typically below 550 MHz. Above 550 MHz, digital signals are carried. Among these are digital video broadcasts, targeted services, cable modem traffic, and possibly telephony or other telecommunications services.

In all views of what HFC will be carrying in the future, heavier loading is a constant. Adding digital channels increases the total load on the transmitters, the significance of which depends on relative digital levels. It works out that the magic "10 dB" number, the rule-of-thumb threshold by which decibel addition of contributors means a negligible contribution, becomes violated for 300 MHz of digital loading at –6 dB relative to analog. The relative digital load at that point becomes about –8 dB of the total.

The need for high analog CNR results in the need for a particular per-channel optical modulation index (OMI). The above loading increase exacerbates this need, as the total power must be shared. The OMI level for high CNR is such as to generate clipping events that tend to drive the uncorrected BER performance of the QAM in the optical portion of the link. Thus, the RF drive to the laser is a careful trade-off between generating a high enough rms OMI per channel for the CNR needs of the analog video, and the clipping impairment associated with higher OMI. It also involves assuring that digital-on-analog composite intermodulation noise (CIN) does not degrade analog CNR, although this often becomes a bigger issue deeper in the RF plant where high output levels are required.

From the outset then, digital signaling in the forward path is confronted with clipping limitations of the linear optics, and relies on FEC to deliver the end-of-line performance. Of course, reliance on FEC for one impairment means less available FEC "budget" for other impairments in the forward path (microreflections, CTB, CSO, phase noise, frequency response variations). While FEC can be proven to mitigate clipping-induced errors, the movement to 256-QAM in the plant will place more constraints on the quality that the plant must deliver relative to CNR (6 dB worse than 64-QAM) and distortions [1]. The coding gain for 256-QAM is slightly weaker than the 64-QAM. Thus, it may be desirable to build the additional margin into the plant design to be prepared for the additional difficulty in transporting 256-QAM.

The average clipping repetition rate, ν, can be estimated [2], and is in units of 1/sec or Hertz. It is related exponentially to the rms OMI. The quantity νT, for QAM symbols of period T, is thus unitless, and represents the average number of clipping pulses per QAM symbol T. This term is valuable for a few of reasons. First, it points out the regularity with which a QAM symbol will be imposed upon by a clipping event – which does not automatically mean that an error will result. Second, the stress on the forward error correction (FEC) can be recognized, assuming it will be responsible for cleaning up some of these impulsive errors. Third, this parameter is part of the key to developing BER analysis for clipping in the forward path, and also plays a role in performing this same analysis for the case of clipping suppression.

Because of the short duration nature of clipping events, their effect on signal quality for analog and digital signals is different. The spectral nature of short duration impulses is, intuitively, wideband in nature. As clipping is gradually increased by increasing RF drive and optical modulation index (OMI), a noise floor is created that on a spectrum analyzer looks like additional white noise. The spectrum analyzer is an averaging device, or, correspondingly, a lowpass filtering device. The human eye also has filtering characteristics. Aside from the raw ability to overlook rapid changes, perception is influenced by what surrounds a picture element and the experienced brain's expected observations. The impairment impact of clipping on analog signals can be correlated to the CNR degradation due to the carrier-to-nonlinear-distortion (C/NLD) associated with the laser clipping phenomenon.

For digital signals, this is no longer the case. The digital receiver, not the human eye, is the processing mechanism faced directly with the impulsive clipping effects. The receiver measures energy in each symbol (this is actually a simplified view of the processing in the demodulation stages). In any case, the digital receiver relies on symbol energy measurements over the small intervals of time corresponding to the symbol rate of the transmission. Because of this, impairments that the eye may not detect are imposed upon a transmission format (high rate digital symbols) that make them detectable, in this case ultimately manifesting itself via the familiar "tiling" associated with digital picture impairment.

<u>Early Linear Optical Transmitter Development History</u>

With the advent of linear fiber optics in the late 1980's, system architectures began to take on new and more reliable forms. The signals that were carried then were strictly analog channels, most often 42 or 60 channels of broadcast video. This lighter load relative to today's 110-channel analog or combined analog and digital systems reduced the design constraints on the laser transmitters. The baseline performance of the fiber optic portion of a network requires about a 50 dB CNR, and distortion levels better than about -65 dBc. These levels were achievable with state-of-the-art 42-channel laser transmitters in 1988, and have remained so with the increased channel count and improvements in transmitter design and linearization since that time.

The earliest fiber optic transmitters employed only second (even) order linearization circuits. Those linearization schemes often employed predistortion of the RF drive signal such that the laser's non-linearities were compensated by the distorted RF drive. In predistortion schemes in general, two features are characteristic: the non-linear element of the predistorter that generates the error signal used for correction, and the topology into which the non-linear element is configured. In this instance, the non-linear elements are pairs of amplifiers in a push-pull configuration, which are connected in a path parallel to the main signal path. A small portion of the main signal is split out of the main path, sent through the non-linear element, amplified, and phase adjusted before being added back into the main signal path. In a similar fashion, third (odd) order predistortion was added in an additional parallel path. This combination of multiple parallel paths and even and odd predistortion constituted the logical progression of the early optical transmitters.

Advances in predistortion technology have led to combined RF and optical subassemblies that previously were separately handled. The non-linear element is now diode pairs, and the topology has been simplified to one in which the elements are in the main signal path. This has allowed tuning of the predistorter in a more predictable and controllable manner, as the RF and optical subassemblies were combined onto a single board, enabling manufacturing cost and process savings.

The Anti-Clipping Concept

In the most recent advancement of the transmitter circuitry, anti-clipping technology has been added. Incoming RF signals can have large envelope variation, periodically becoming large enough to create distortion due to laser clipping. It is a somewhat similar phenomenon to amplifier clipping, but in that case the input-output transfer function is a soft distortion mechanism as the amplifier gradually loses dB-for-dB behavior. In the laser clipping case, the distortion mechanism is a hard clip, much like a perfect limiter, but occurs only in a single-sided manner for directly modulated lasers, when the instantaneous RF drive exceeds bias level. If the large amplitude RF drive can be avoided, or the laser biased to a high enough point, clipping does not occur. In the forward channel, the periodicity of the RF signal's peaking events are such that reasonable predictions can be made about its amplitude and repetition characteristics. Because of this, an "anti-clipping" circuit can be designed based on statistical knowledge of the events. The circuit samples the RF and predicts the duration of a high amplitude RF envelope occurrence. A drive circuit is used to partially control the DC bias point of the laser, and when a high amplitude event is recognized, the drive circuit temporarily biases the laser to a higher level to prevent the onset of clipping. The duration of the events and the mechanism of changing the laser bias point have not been detrimental to the overall operation or performance of the laser transmitter.

A simple way to understand the clipping mitigation approach is to think of it as a threshold detection mechanism followed by a laser bias adjustment, the period of which is determined by an RC time constant, or equivalent lowpass function. The input to a forward path laser is a composite multiplex of traditional analog video and digital services, typically riding above the analog spectrally. As mentioned above, the resulting signal has random qualities, creating a high peak-to-average ratio that looks much like noise. However, for the forward path, there is an inherent component of peaks of that occur, not coincidentally, with a 6 MHz repetition rate and predictable duration with which to set the time constant. The amplitude characteristics of this burst period depend upon the relative phases of the analog carriers. Because of the large number of (typically) independent carriers (not phase-locked), the randomness qualities vary slowly with the nature of crystal oscillator drift.

There has been a significant amount of study to determine the impact of laser clipping in composite analog/digital multiplexes on both the analog CNR and on the QAM BER. Analytical expressions have been derived and measurements take in support of these results to help system designers optimally align laser loads for both analog and digital performance. A recent contribution [3] provides simplified BER expressions under the practical assumption that, during a clipping event, the AWGN contribution is negligible compared to the clipping. The is particularly applicable to HFC systems, where clipping can still be considered a relatively rare event in terms of number of events on a per symbol basis, and where the CNR on any 6 MHz channel is relatively high.

Even for digital channels that are run backed off from the analog video by up to 10 dB, CNR's in the low 30's will exist as a minimum. For the optical link only, we assume that the only impairments of significance to deal with are AWGN (RIN, optical RX, interferometric intensity noise) and the clipping effect. The expression for BER under the above assumption can be represented as

BER(ν) = Prob(no clipping events in a symbol) $\bullet$ Prob(error/no clipping events) + Prob(clipping event in a symbol) $\bullet$ Prob(error/clipping event in a symbol),

or

$$\text{BER}(\nu) = [1-\nu T] \bullet \text{Prob(error/AWGN)} + [\nu T\ ] \bullet \text{Prob(error/clipping event in a symbol)} \quad (1)$$

The first term in (1) is the traditional demodulator case – symbol detection in the face of the AWGN channel. The second term recognizes that a different error rate expression must be calculated when clipping events occur. This is because the impairment PDF has changed from Gaussian to something different.

This expression is a good starting point for predicting the effect of clipping mitigation, by relying on the prior work regarding the characteristics of the clipping impulses. It is described in [2] how clipping can be modeled as a Poisson process, with a parabolic pulse shape, the duration of which is Rayleigh distributed. The mean duration is related to the frequency content of the modeled spectrum and the rms OMI driving the laser.

From this statistical information, the effect of the matched filter on the clipping event can be evaluated. In particular, the impulses are so short, on the order of nanoseconds, relative to the symbol periods (on the order of hundreds of nanoseconds) in real systems, that the parabolas look nearly like impulses to the input of a matched filter, which allows simplification of the filter output. Thus, the impulse response of the filter plays a key role in evaluating the effect of the short clipping pulse. Effectively what occurs is that the matched filter spreads the clipping energy out into the order of a symbol time, and the analysis simplifies to evaluating the impulse response of a matched filter, but with a noise term of a statistically varying output amplitude related to the clipping level. Since the matched filter integrates over a symbol period to determine an energy for symbol slicing (detection), the varying output amplitude is, in fact, a function of the Rayleigh statistics of the clipping impulse duration. The actual statistics of the output amplitude is a complicated expression [2] because the PDF of the impulse duration is modified by the matched filter operation. Refer to the probability density function (PDF) of this matched filter output amplitude due to a clipping impulse as $p_c(C)$.

How does this all relate to the analysis for the clipping mitigation case? Consider (1):

$$\text{BER}(\nu) = [1-\nu T]\ \text{Prob(error/AWGN)} + [\nu T\ ] \bullet \text{Prob(error/clipping event in a symbol)}$$

What changes now to the predict performance of the equipment using this approach is that clipping mitigation that has been introduced. This effects the coefficient terms in this expression. In the analogy given above for the mitigation circuitry – an RC filter – when the onset of clipping is sensed, a bias change is introduced for a time period related to the duration of a burst of events. The period is governed by an RC time constant. That is, the nature of the forward path multiplex is that when there are clipping impulses, they typically come several or many over a period of time for which this time constant represents the average. The term νT would still represent the average rate of clipping occurring, the description here merely says that they occurrence is not necessarily evenly distributed over time.

Now, when clipping suppression kicks in, the further assumption is made that the error rate again reduces to the AWGN-only case. The new expression for BER can be written by noting that this occurs when the amplitude of the clipping event at the matched filter output exceeds some threshold. The threshold that must be exceeded is, of course, at the transmitter on the other end of the link. However, for a clipping impairment that dominates the AWGN contribution when both exist – an assumption from the beginning for this simplified analysis - it is equivalent to fixing a threshold at the matched filter output also. Call this threshold boundary at the matched filter b, then the BER expression becomes

$$\mathrm{BER}(\nu) = \left[1-\nu T+ \int_b^{\infty} p_c(C)\,dC\right] \bullet \mathrm{Prob(error/AWGN)} + \left[\nu T - \int_0^{b} p_c(C)\,dC\right] \bullet \mathrm{Prob(error/clipping\ event\ in\ a\ symbol,\ modified).} \qquad (2)$$

The expression here shows how the mitigation of clipping events is taken into account to express the new likelihood’s of the two BER’s that make up the composite expression.

Physically, the interpretation of this expression is that, for the percentage of amplitudes that would exist above the threshold at which the laser bias is adjusted, it is assumed that clipping mitigation is employed and successful. Doing so reduces such a situation to the AWGN detection problem. Thus, the first term in the expression above corresponds to the case where no clipping events occur, naturally, combined with the case where a clipping event would have occurred, but was ameliorated by the mitigation effect of the technology. The second term corresponds to the case where clipping exists, but the amplitude of the impulsive event is not enough to trigger bias reduction. Thus, the detection problem is still one of detecting the signal in the face of a clipping event. The expression in the second term for probability of error of the remaining clipping impaired symbol, however, is not the same as that used in (1). The expression for the error probability given a clipping event is an integral expression. The region over which the integral is evaluated now is only over the range up to the threshold boundary, b. Thus, the probability of error given a clipping event in a symbol is slightly different in (1) and (2). An intuitive way to look art this is that in (2), this expression only applies for a subset of the cases of equation (1), the clipping events which do not exceed the threshold. Thus, the error probability given a clipping event in a symbol must be smaller.

This analysis can be used to predict performance as a function of the threshold of declaring a clipping region, channel CNR, and rms OMI.

Measured Results

A block diagram of the test setup is shown in Figure 1. Note that Figures 1 through 3 are all located at the end of the paper.

A Matrix generator was used to place 77 channels of analog video in the 52-550 MHz passband. This CW comb was RF combined with the simulated digital load, which consisted of a bandpass filtered wideband noise source, with the filter delivering a flat spectrum across 553-860 MHz. The combined signal was passed through a channel deletion filter at the location of the desired 256-QAM test channel, 555.25 MHz. The QAM test signal, an uncorrected 256-QAM RF carrier, is upconverted and combined at the proper relative power level to complete the signal load. This load is split three ways to drive three lasers. The three lasers represent three classes of performance from the Motorola product family. There is a standard performance product, the LM-9, a premium performance product, ALM-9, and finally the clipping suppression product, the ALM II-9. The –9 in each product represents the maximum optical link length the transmitter is designed for (essentially, this defines laser output power).

The transmitters are alternately run through an optical link of 20 km, where the signal is detected by a forward path optical receiver inside a node. The node in this case is a Motorola SG2000. Thus, it was assured that the performance were relative to the same link and optical receiver. The node's RF output feeds a bandpass filter for the QAM channel, through a subsequent downconversion to get to the test modem IF frequency. The demodulator interfaces with a BERT and a PC. This completes the test setup. All of the BER measurements are with no error correction.

The testing was broken down into a focus in three areas of study – digital bandwidth, relative digital levels, and increased analog levels. The test results are shown in Table 1, Table 2, and Table 3 corresponding to the three areas of focus.

In these tables, the term SNR has been switched to to represent the digital channels, while standard term CNR has been used to represent analog measurements. Note that the effective SNR improvement (the column labeled "BER SNR Delta") of the 256-QAM performance is based on converting from BER to SNR (or Eb/No in this case, the relationship is still one-for-one). Figure 2 shows a BER plot of 256-QAM in AWGN, relative to the BER curve for 64-QAM, 16-QAM, and QPSK. Note that the horizontal axis is based on Eb/No, and not SNR. To put in terms of SNR requires augmenting with the bits per symbol relationship. For 256-QAM, for example, the SNR is $10\bullet\log(8) = 9$ dB higher than Eb/No, since there are eight bits per 256-QAM symbol. This leads to the useful rule-of-thumb that a BER of 1e-8 requires (25+9) = 34 dB for uncorrected 256-QAM. This is useful because, in terms of SNR, the significant modulations are all about 6 dB apart. In other words, 64-QAM requires about 28 dB SNR for 1e-8 and 16-QAM requires about 22 dB SNR.

Varying Digital Bandwidth

In Table 1 below, the performance of the lasers against increasing the digital bandwidth is measured, for analog levels at nominal per-channel settings and a fixed relative to analog digital per-channel level of –6 dBc.

Table 1 – Varying Digital Bandwidth

Test Conditions w/9dB Link			256 QAM BER			BER SNR Delta
TEST #1						ALM II-9 vs. ALM-9/LM-9
			LM-9	ALM-9	ALM II-9	
550MHz Analog (nom per/ch) Single 256 QAM ch @-6dBc			6.30E-08	3.70E-08	1.00E-10	1.25 dB/1.50 dB
	547.25MHz	CNR	51.8	53.9	53.9	
550MHz Analog (nom per/ch) 200MHz digital @-6dBc w/555MHz			2.40E-07	1.40E-07	2.90E-09	1.00 dB/1.25 dB
	547.25MHz	CNR	50.9	53.0	53.6	
550MHz Analog (nom per/ch) 300MHz digital @-6dBc w/555MHz			8.70E-07	2.90E-07	4.90E-08	.50 dB/1.00 dB
	547.25MHz	CNR	51.4	53.4	54.0	

Evaluating Table 1, the following points are of note:

- The anti-clipping technique in the ALM II results in virtually error free performance when a single digital channel augments the analog multiplex; this verifies the anti-clipping mechanism as the standard models show error accumulation even for this favorable condition.
- There is about two orders of magnitude BER improvement for the ALM II for the first case and the 750 MHz system with 200 MHz of digital load.
- There is about an order of magnitude BER improvement for the heaviest loaded case of 300 MHz of digital; in this case the total digital power represents –8 dB of the total power, so it is no longer considered a negligible power load.
- There is 1-1.25 dB of effective SNR gain on the digital channels for 200 MHz of digital loading, and .5 dB-1 dB of gain on the 300 MHz of digital load.

The measurements verify the effect of the anti-clipping circuitry. The point of the CNR numbers taken on the last analog channel was to recognize that the analog performance was essentially unchanged around a high quality level of CNR performance. In addition to proving anti-clipping effects, this test shows the significant link gain available for the 200 MHz of loading. While the 300 MHz loading case shows less gain, this case also crosses the threshold at which the digital loading can be ignored as a percent of the total. In other words, the laser in this case is being driven above its nominal rms OMI because of the 100 MHz of additional digital loading. The following tests explore other options that may more efficiently use the performance gain.

Varying Relative Digital Levels

The emphasis in Table 2 below is on the level of the digital channels relative to the analog, for a fixed 300 MHz of digital bandwidth, and fixed analog levels at nominal per-channel settings.

Table 2 – Varying Relative Digital Levels

Test Conditions w/9dB Link			256 QAM BER			BER SNR Delta
TEST #2						ALM II-9 vs. ALM-9/LM-9
			LM-9	ALM-9	ALM II-9	
550MHz Analog (nom per/ch) 300MHz digital @-6dBc w/555MHz			8.70E-07	2.90E-07	4.90E-08	.50 dB/1.00 dB
	547.25MHz	CNR	51.4	53.4	54.0	
550MHz Analog (nom per/ch) 300MHz digital @-8dBc w/555MHz			1.50E-07	5.75E-08	2.10E-09	1.00 dB/1.25 dB
	547.25MHz	CNR	51.4	53.6	54.3	
550MHz Analog (nom per/ch) 300MHz digital @-10dBc w/555MHz			5.80E-07	3.46E-07	1.00E-10	2.00 dB/2.25 dB
	547.25MHz	CNR	51.6	53.7	54.4	
550MHz Analog (nom per/ch) 300MHz digital @-12dBc w/555MHz			2.40E-06	2.00E-06	4.00E-08	1.25 dB/1.25 dB
	547.25MHz	CNR	51.7	53.9	54.6	

Evaluating Table 2, the following points are of note:

- The performance of the ALM II actually improved as relative digital levels dropped, to a point. This is attributable to what was implied by the Table 1 measurements of 300 MHz of digital bandwidth – the significance of the power load. As the power load becomes once again negligible (-10 dBc relative levels), the laser experiences a typical level of clipping. The circuit is designed to handle typically found clipping statistics well, and the link becomes "less" clipping limited, instead of completely clipping limited, like the LM and ALM. This can be recognized by noting that BER does begin to degrade at the -12 dB relative level for the ALM II, because of lower QAM levels relative to the remaining impairments.

- The traditional transmitter families show predictable BER degradation as signal level is dropped. These models are still impacted by clipping as the dominant impairment, and lowering the channel power makes the C/(clipping),the dominant impairment parameter, lower.

- For digital channels run –10 dBc, there is a 2-2.25 dB effective SNR gain in BER due to the anti-clipping design; the link runs virtually error free, versus E-7's for the traditional models.

- A major advantage of the lower digital levels - not recognizable in this link test – is the effect of the lower digital levels on the end-of-line plant distortion performance. As RF output levels increase to eliminate actives in fiber deep architectures, the imposition of digital CIN onto the analog CNR can become a major issue. This is particularly an issue at –6 dBc levels and above because of the additional power load mentioned above. The relative digital impact represents about an additional .6-.7 dB of total power, which is doubled for determining third order distortion impact – a noticeable result.

Increasing Analog Levels

In Table 3 below, the focus changes to the increase in analog levels for a fixed digital level, and fixed digital bandwidth. Please note that the digital levels relative to analog are absolutes. In other words, in Table 2 above, the first test was run with digital channel levels at –6 dB relative to analog levels. In the first test below, the analog levels have been increased by 1 dB. The digital level was not changed, therefore the digital level relative to the analog level now becomes –7 dB. In the last test of Table 3, it is noted that the analog levels are increased by 2 dB. The digital levels are noted as –12 dBc. This means that if the analog levels were at their nominal per-channel settings, then the digital levels would be –10 dBc. Thus, the digital level is set the same as it was in the third test in Table 2.

Table 3 – Increasing Analog Levels

Test Conditions w/9dB Link			256 QAM BER			BER SNR Delta
TEST #3						ALM II-9 vs. ALM-9/LM-9
			LM-9	ALM-9	ALM II-9	
550MHz Analog (nom per/ch + 1 dB)						
300MHz digital @-7dBc w/555MHz			7.10E-06	3.20E-06	5.70E-08	1.50 dB/1.75 dB
	547.25MHz	CNR	52.4	54.3	55.1	
550MHz Analog (nom per/ch + 2 dB)						
300MHz digital @-8dBc w/555MHz			1.60E-04	9.40E-05	3.80E-06	1.75 dB/2.00 dB
	547.25MHz	CNR	52.8	54.4	55.8	
550MHz Analog (nom per/ch + 2 dB)						
300MHz digital @-10dBc w/555MHz			2.10E-04	1.90E-04	1.53E-06	2.75 dB/2.75 dB
	547.25MHz	CNR	53.3	55.4	56.2	
550MHz Analog (nom per/ch + 2 dB)						
300MHz digital @-12dBc w/555MHz			5.30E-04	4.40E-04	4.30E-06	3.00 dB/3.00 dB
	547.25MHz	CNR	53.6	55.7	56.2	

Evaluating Table 3, the following points are of note:

- The additional 1 dB of analog level degraded the BER on the two traditional models, while for the case of the ALM II it remained virtually unchanged.
- For a 2 dB increase in analog levels, the traditional transmitters are creeping towards crashing with BER's in the E-4 range. The ALM II is still comfortably in the E-6 range.
- The ALM II continues to show CNR increase with the additional 1 dB of analog level, while the traditional designs show the increase for a 1 dB increase in level, but not for a 2 dB

increase due to higher distortion and clipping contributions to CIN.

- Dropping digital levels accompanying the 2 dB analog increase causes further BER degradation from the additional relative CIN with low QAM on the traditional models, and the increase in clipping events caused by the analog levels. The ALM II again shows some slight improvement to a point, as it is not as clipping dominated. It again shows slight degradation as the composite of contributions to BER begins to appear only for the lowest digital levels.

- Effective SNR improvement due to the ALM II achieves values as high as 3 dB for the high analog/lowest QAM case shown. Even in this case, the ALM II shows BERs in the E-6 range, two orders of magnitude better than the traditional models.

- The same discussion of digital levels and CIN effects can be applied in this case. In fact, the increased analog levels have a greater impact on the total load than the increasing digital. It helps that the digital levels are not increasing in this case, although the dominant CIN effect for analog/digital is two analogs mixed with one digital.

- It is worthwhile to mention that FEC performance is generally considered capable of correcting BER's in the E-4 range, although it is not desirable to have all of it allocated to the optical link [1].

It is important to note that this paper discusses only the optical portion of the link. Thus, there is virtually no discussion of distortion impacts such as CTB, which is driven by performance of the RF plant. As it turns out, CTB can effect QAM BER under certain conditions [1]. In the sense of effecting the channel performance on the end-to-end system, the optical portion of the link has a major impact on CNR delivered and on the BER of the digital signals because of the clipping issue. Other BER impairments specific to the RF plant must also be well understood to understand the complete performance.

The results above point out that there are various ways to enhance plant performance with the anti-clipping approach. The various ways to spend the performance advantage are

- Lower uncorrected BER with typical analog and digital loading
- Wider digital bandwidth with the same analog loading
- Lower digital levels with the same analog loading
- Higher analog levels with a constant digital loading

Translated to SNR, BER improvements can be as high as 3 dB. The benefits to the total plant of lower digital levels also helps CTB performance under the heaviest digital loading at today's highest amplifier RF output levels. Because of the dB-for-dB digital contribution effect (two analog by one digital) of third order composite analog plus digital distortion, it can be roughly estimated that CIN drops a dB for every drop of digital level that can be implemented. Another scenario is systems squeezed for analog CNR margin for long links or multiple hops can afford an increase in level at no BER cost. Finally, upgrading the plant to add more digital can be done pain free because of the margin provided by the clipping mitigation to the loading effects.

Return Path Application

Linear optics are also used in the return path. It is thus natural to consider this technology for return path applications. While this is under investigation, there are two reasons that this technique is of less value in the return path. First, unlike to forward path, the return path is a bursty, generally time domain (and frequency domain) accessed channel. It is thus very rare, under normal operating and alignment conditions, for the reverse path band to be loaded heavily enough for clipping to have consistently expected occurrences the way the forward path does. For example, consider Figure 3, which represents the output Carrier-to-Noise-plus-distortion [C/(N+NLD)] versus total input level. The plot is an example of a Noise Power Ratio (NPR) curve, and is valuable in determining how best to align laser drive as a compromise of performance and robustness in the return path. Assume that the setup point of the laser is backed off from the optimum point of C/(N+NLD) by 3 dB. The set point is based on a fully loaded band of return signals, typically allocated on a power-per-Hz basis, although this condition is very unlikely in practice. Backing off of the peak performance is common to guard against signal level variations, equipment variations, and other unexpected events. The goal is to have enough room from the peak such that the slippery slope of the NLD part of the curve is avoided.

For the return path, the NPR back-off is often set greater than the forward, to account for outdoor plant and level variations, as well as protect against noise and interference transients (or constants) that can overload the reverse path. Assume that it is set only 3 dB off of the peak NPR. Further assume that the maximum simultaneous usage of the return path is 25%, such that the maximum loading ever present to the laser is yet 6 dB lighter. This is a total of 9 dB back off from the NPR peak. This region of the NPR curve is clearly noise dominated, as opposed to noise and NLD, and the probability of clipping even for a Gaussian-like noise signal is extremely small. The number can be quantified easily by analyzing the Gaussian probability distribution function (PDF) and integrating the curve to provide the likelihood of exceeding the 9 dB peak to average in one direction only. However, suffice to say without the math, that to attempt to mitigate the events of very low probability will have very little impact on typical performance.

Secondly, even if the return were fully loaded, the statistical nature of the clipping events would differ significantly. It has been described how, in the forward path, the approach taken is a rather simple one - recognizing the onset of a clipping event, and using the a-priori information we know about these events. Their duration and statistics of their duration, as well as their shape, has been analyzed in detail [2]. It has also been discovered, not surprisingly, that there is a strong 6 MHz periodicity of peaking to the composite forward path load. Obviously, this relies on analog carriers, and if the load were to be switched over to all digital channels, there would be some complexities. By knowing the nature of the periodicity of the forward path time domain waveform, the timeframe with which the laser bias must be adjusted can be “predicted” on average, and mitigated against. The known characteristics of the forward path’s waveform and its 6 MHz-related characteristics do not apply in the return path. There are no fixed channel assignments (or few, none of significance to loading), no fixed service bandwidths, and no clear-cut relationships to leverage for bias changes enabling efficient clipping suppression as there is in the forward path.

Digital reverse paths make the return clipping suppression idea seem of less value. In principle, however, it is still very applicable, as an A/D converter has NPR characteristics similar to a laser because of its tendency to clip. For the A/D, however, clipping is two-sided in nature, and this reveals itself as the steeper slope on the distortion side of the curve. In fact, digital technology offers the opportunity to manipulate the bits to deal with sensed clipping events from an A/D’s overload flag, as opposed to in the analog transmitter, providing more flexibility in processing techniques to consider. However, the reasons above apply still in terms of use of the concept in the context of return paths.

Conclusion

Linear optics has made the development of the HFC plant possible. Continuing advances have allowed HFC to evolve into a variety of architectural scenarios. The deployment of digital signaling on HFC created a need to have a deeper understanding of the laser clipping issue, as the clipping phenomenon manifests itself as BER impairment on the digital channels. This paper describes the first significant product attempt to specifically address laser performance from the standpoint of clipping mitigation. It has been shown that by addressing the clipping issue in the design of the transmitter directly, using some known stochastic properties of the clipping events, significant performance gain is achievable. How this performance gain is used is up to the designer. However, the advantage can be translated into simple a lower digital BER. Or, it can be spent as wider digital bandwidth with the same analog CNR and BER, higher analog levels and analog CNR with the same BER, or lower digital levels of the same BER and thereby higher CNR due to lower CIN.

References

[1] Howald, R. et al., “Distortion Beat Characterization and the Impact on QAM BER Performance,” 1999 NCTA Technical Papers.

[2] Pan, Qi and R. Green, “Amplitude Density of Infrequent Clipping Impulse Noise and Bit-Error Rate Impairment in AM-VSB/M-QAM Hybrid Lightwave Systems,” IEEE Transactions on Communications, October 1996.

[3] Pan, Qi and R. Green, “Simplified Analysis of QAM BER Impairment in Hybrid AM/QAM Lightwave Systems,” IEEE Transactions on Communications, April 1999.

Acknowledgement

The author would like to recognize the efforts of Mike Short, Tech Specialist in Systems Engineering in Motorola BCS' Transmission Networks business unit.

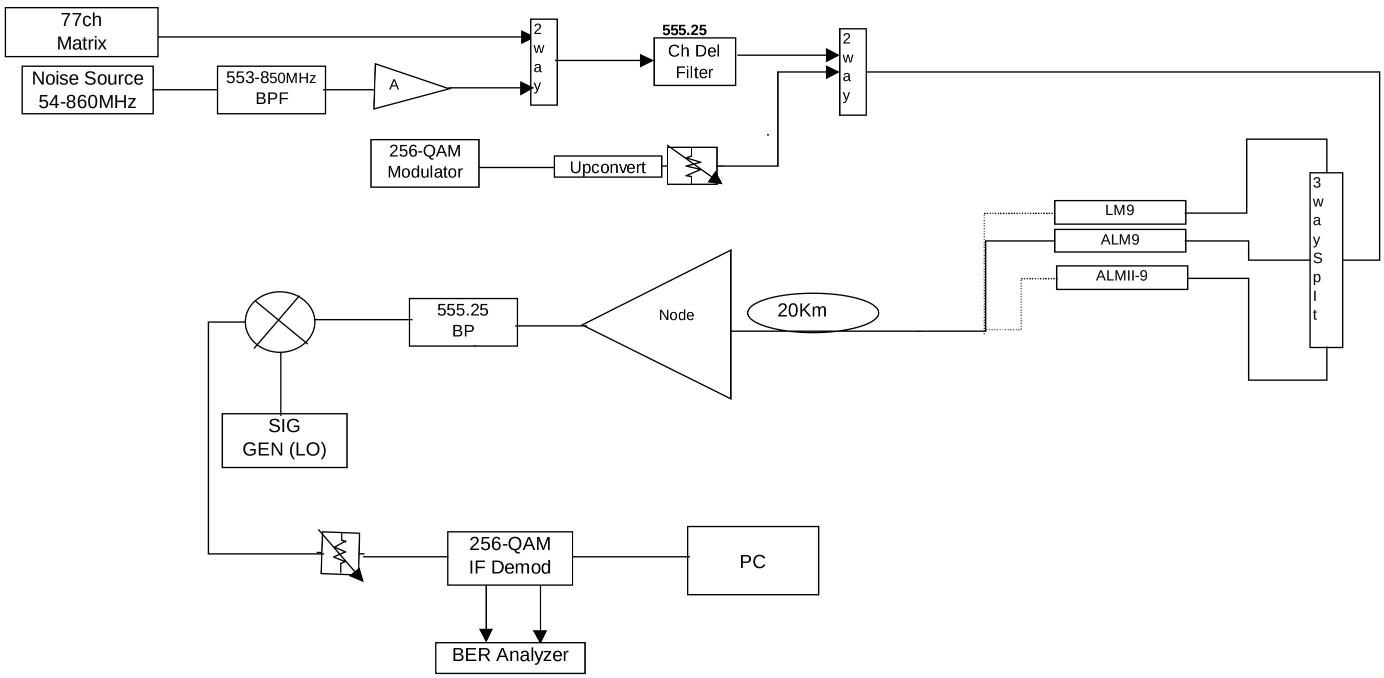


Figure 1 – Test Setup Block Diagram

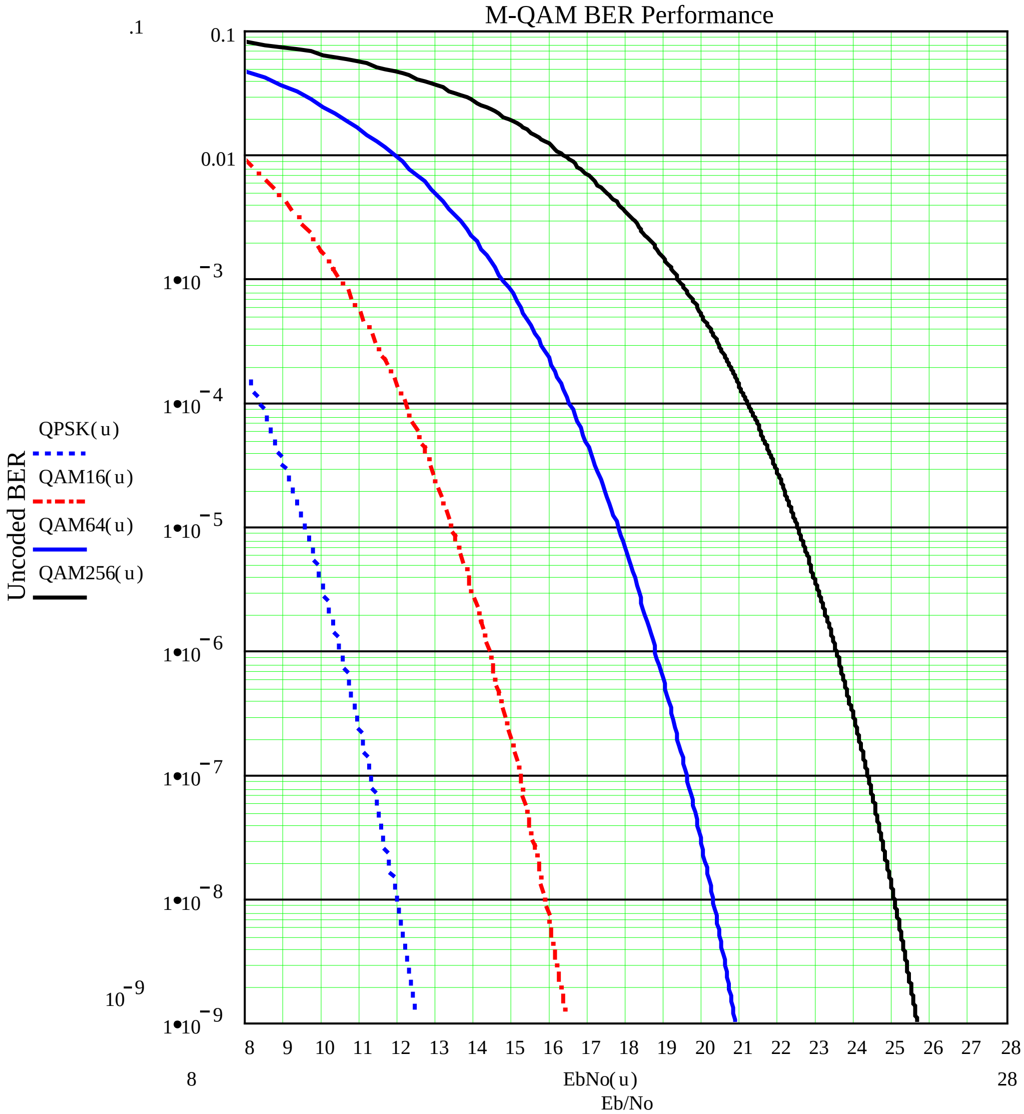


Figure 2 – Theoretical BER of 4/16/64/256-QAM vs. SNR per Bit

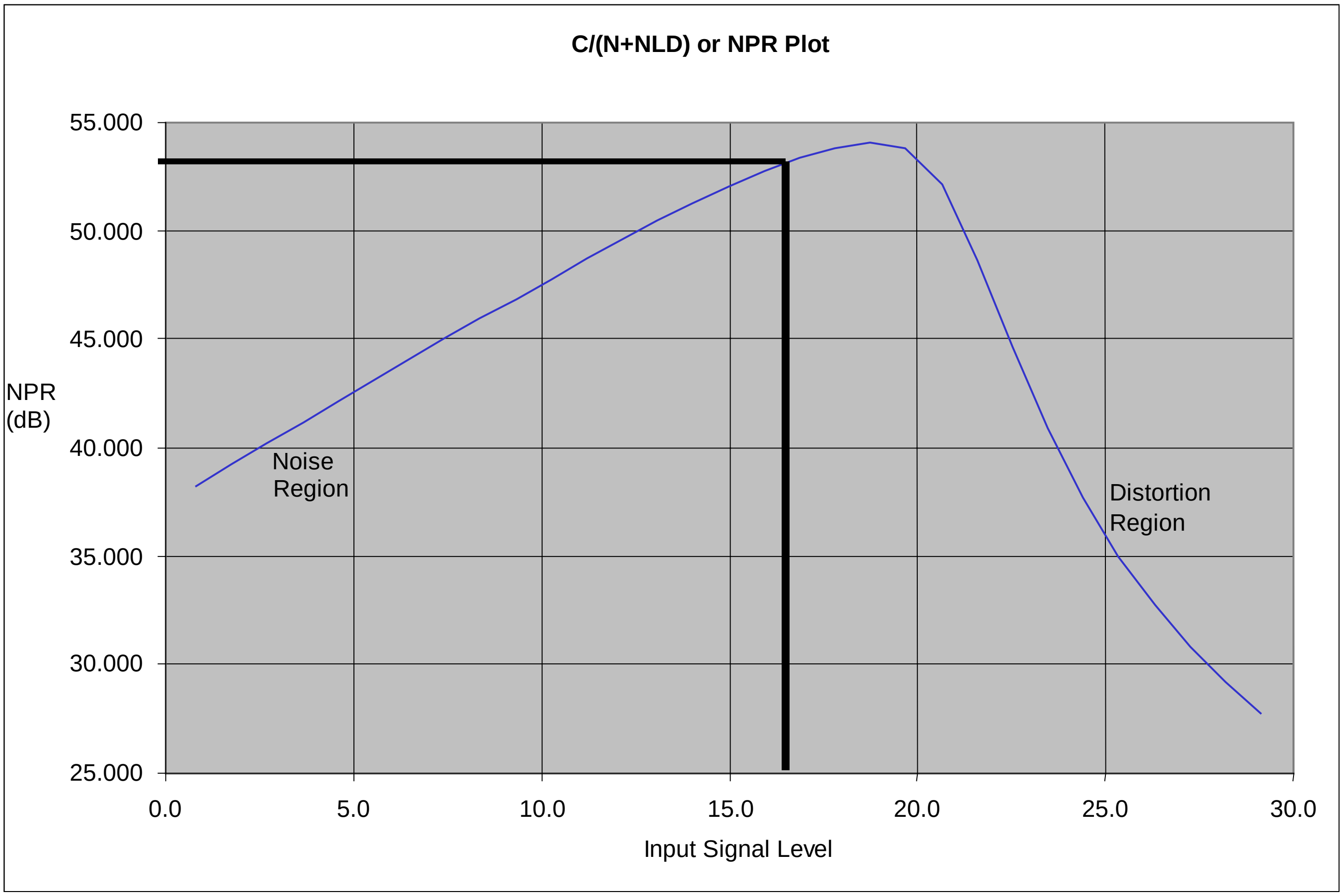


Figure 3 – NPR for Return Operating Point with Headroom

Integration Of Home Networking Technologies With The HFC Residential Gateway

Jed Johnson
Motorola Broadband Communications Sector

Abstract

This paper analyses the currently available and developing home networking technologies from a Hybrid Fiber-Coax (HFC) based services perspective. The following home networking technologies are assessed for suitability to support these services; Home Phoneline Networking Alliance (HPNA) 2.0, powerline networking, wireless networking and cluster networking. Each technology's basic characteristics, advantages and disadvantages are described. Finally, technology gaps for implementing a robust end-to-end service model are identified.

INTRODUCTION

Home Networking has become a hot topic over the past year and it is an issue that must be addressed by cable modem providers. One of the pressing issues is how to deploy Home Networks given the cost to provide the wiring infrastructure for traditional methods like running Ethernet over 10 base T. In this paper I will address some of the technologies that can be deployed to provide Home Networking capabilities without adding new wires in the home. I will discuss the popular technologies for wire line and wire less networking in the home both in terms of where they are and where they are headed. I will conclude with some brief comments about what capabilities these "no new wires" technologies need to provide to support the advanced applications that service providers may want to deploy. Applications like telephony and streaming media.

HOME NETWORKING DRIVERS

Before I dive into the technologies I will cover some background on what is driving the need for Home Networking. There are two near term and one general long-term driver. In the near term the need for Home Networking is being driven by the growth in the number of homes with multiple PCs and the number of homes with broadband access. Longer term the need for Home Networking will be driven by advanced services delivered over the broadband network.

Multiple PC Homes

The growth in the number of homes with multiple PCs is growing faster than the number of homes with any PC at all. The chart below from a Yankee group report supports this finding. In addition the data from the Yankee Group shows that 78% of the homes with multiple PCs have a printer. Once there are two PCs in a home with a printer there is a strong need for a home network unless all the devices are located in the same room.

Broadband Access

Broadband access is growing very rapidly and all the analysts' predictions are for continued rapid growth. Our data indicates that over half of the households with broadband access have multiple PCs. Once broadband hits the home there is a strong desire to share the bandwidth among all the PCs in the home.

Advanced Services

Over time, as broadband access becomes pervasive, new services that use the broadband access will appear. In addition more and more devices in the home will have Internet access capability and will want to share the broadband connection. The long-term outlook is for broad networking capability in the home with bandwidth and quality of service demands increasing.

No New Wires Technologies

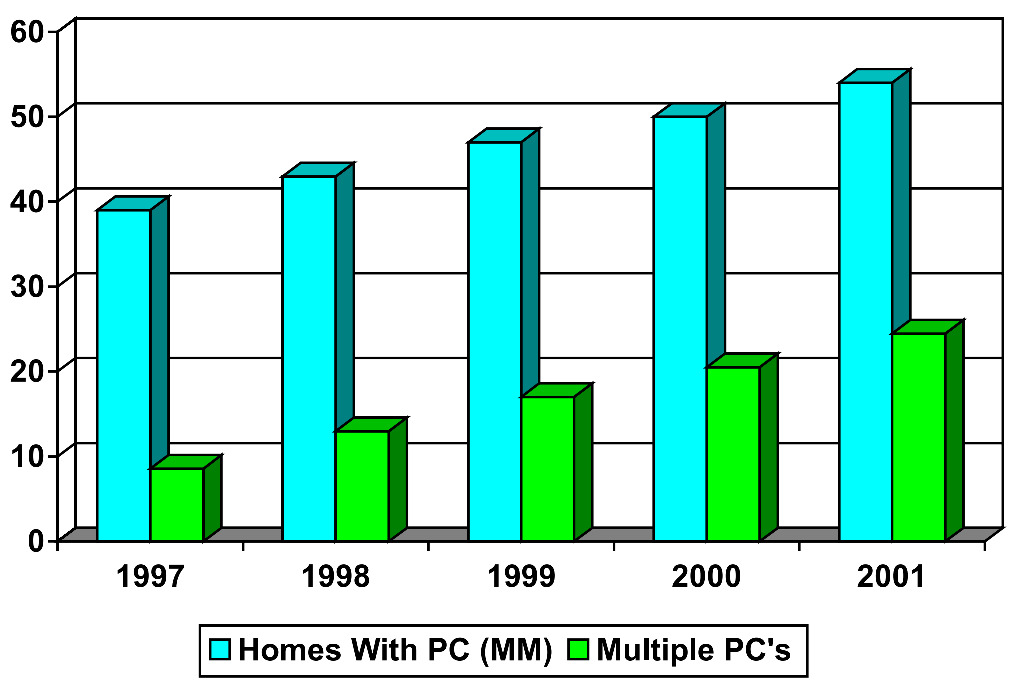
The rest of the paper will look at technologies that provide Home Networking capabilities without having to wire the home. This means either using the existing wiring infrastructure or use wireless techniques.

Existing wire

The two existing wiring infrastructures in nearly every home that has a PC are phone lines and power lines. Both are viable alternatives for Home Networking.

Phone Lines

An industry group to standardize the use of phone lines for Home Networking called HomePNA was formed almost two years ago and the second generation of products are now available. HomePNA is currently the defined standard for phone line Home Networking.



Power Line

An industry group to standardize the use of power lines for Home Networking called Home Plug was formed in early March this year. The intent of Home Plug is to create a single standard from the range of power line technologies evolving for Home Networking.

Wireless

In the wireless area there are two main technologies gaining acceptance for Home Networking. They are the IEEE 802.11 standard called 802.11b and the standard developed by the HomeRF group call SWAP. Both of these standards use the 2.4 GHz band. I do not include Bluetooth in this discussion because Bluetooth is not a networking technology

For each of these technologies I will discuss the current status and future trends.

HomePNA Today

As I mentioned above HomePNA is an industry organizational that has released a specification for Home Networking on phone lines. This technology coexists with POTS and ADSL by occupying a different frequency spectrum.

Version 2.0 of the specification was released by HomePNA late last year and many products that support HomePNA2 are on the market. The significant advance in HomePNA 2.0 is an increase in the data rate from 1mbps to 10 mbps.

There are however two other interesting changes in the 2.0 spec. The PHY layer is capable of supporting 32 mbps. While there are no publicly announced products that take advantage of this to go beyond 10 mbps, the technology will support it.

The 2.0 spec also adds a feature called Differential Fair Priority Queuing, also called DFPQ. DFPQ allows for upper layers and applications to signal the MAC layer to provide priority access to the media. This means that HPNA2.0 could be extended to support telephony and streaming applications.

HPNA Future Trends

The demand for higher speed continues to grow and will likely result in higher data rate standards from HomePNA. Even thought the current PHY spec supports 32 mbps it is quite possible that there are PHY extensions that could allow phone line networking to be extended even further.

To support telephony there will need to be agreement on the signaling protocol to allow telephony terminals from multiple vendors to interoperate with each other in the home and to gain access to the voice network which might be provided by another vendor. The point is that for voice interoperability it is required that higher layer protocols, beyond the MAC and PHY layers, are agreed to.

It should also be noted that while DFPQ looks very promising there have not yet been any telephony or streaming products developed using DFPQ. We need to see products using DFPQ successfully to fully determine its viability.

Power Line Today

An industry group called Home Plug has been recently formed to develop and endorse standards for using the power line infrastructure in the home for Home Networking applications. There have been a number of companies developing technologies for power line home networking but to this point none have proven to be effective either because data rate is low or because error rates are high or both.

This technology continues to evolve however and shows promise for being able to achieve the data rates and Quality of Service required for Home Networking applications.

Power Line Future Trends

The notion of using the power lines in the home for Home Networking is compelling because power outlets tend to be ubiquitous in the home. In addition society is used to the idea of a power plug on a device and if networking could be achieved without another cable it would be easy to promulgate.

The current baseline for home networking is the ability to support 10 mbps with error rates in the neighborhood of 5% or below. Power line technologies need to be at this level of technology and competitively priced today to become accepted.

Over time the expectations of the capabilities of the Home Networking will increase and all technologies will be required to keep pace or fall by the way side.

Because of the ubiquity of power outlets in the home, power line networking will always be welcomed in the door if the technology can compete.

Wireless

Wireless LAN products are becoming very popular since some proprietary products hit consumer price levels last year. Recently products based on HomeRF and 802.11 standards have reached the market. Both HomeRF and 802.11 operate at 2.4 GHz and each has its own advantages and disadvantages.

Wireless is generally regarded as the first choice by consumers because of the mobility that is offered by the technology.

HomeRF Today

HomeRF operates at 1.6 mbps and products based on HomeRF have just recently been released. The current spec also provides for 4 voice channels and at least one company has stated an intention to develop voice products.

The obvious problem with HomeRF is the data rate. An FCC ruling to allow HomeRF is operate at 10mbps is expected to be approved by mid year to allow vendors to address this issue.

HomeRF Future Trends

HomeRF will need to get to higher data rates. The immediate opportunity is to take advantage of the FCC ruling and extend the 2.4 GHz technology to 10 mbps.

It may also be possible that HomeRF will pursue 5.5 GHz technology to obtain even greater bandwidth capability.

Because HomeRF has a TDM channel and an a greed to signaling protocol it is quite possible that voice products based on HomeRF will have an impact on the market.

802.11b Today

802.11b is also a 2.4 GHz standard that provides 11 mbps data rate. There are multiple vendors providing NICs and access points based on this standard. The current prices are slightly higher than HomeRF but many consumers are willing to pay extra for the increased data rate.

One problem frequently mentioned with 802.11b is that lack of QoS capability in such products appear on the market.

On the cost front the conventional wisdom is that volume will drive down cost to nearly the same levels as HomeRF. Only time will tell.

5 GHz

One observation is that there are two overlapping wireless alternatives and it is not clear that one is going to win out over the other.

Another observation is that the demand for greater bandwidth will continue to rise and that HomeRF and 802.11b will not be able to fulfill that demand.

A growing likelihood is that the 5 GHz band will become attractive for technology suppliers that want to provide greater throughput for applications like streaming video. This creates an opportunity for the industry to the MAC layer to add support for telephony applications. Some also feel that the security provided by 802.11b is not very strong.

802.11 Future Trends

Extensions to the MAC layer to provide QoS support and to improved security have been proposed. A Task Group called 802.11e has been assigned to evaluate these proposals against a set of requirements that were developed last year. The goal is to have a draft proposal for the 802.11 committee by November this year.

The QoS extensions will provide support for applications that require bounded latency, however to support telephony applications a signaling protocol will also have to be standardized. Currently there is no project within 802.11 to create such a standard. It is however possible that a defacto standard could be created by one of more companies that are able to develop and deploy telephony products based on 802.11. Either way it is very likely to be well into 2001 before adopt a single standard.

There are currently three standards that I know of in this range, HyperLan1, HyperLan2 and 802.11a. HyperLan2 and 802.11a have very similar PHY layer specifications and HyperLan1 is perceived to be a lower cost solution. This unfortunately sounds familiar. Currently there are no 5 GHz products widely deployed for Home Networking applications so there is still a chance for a unified standard for Home Networking in this band.

Summary

While the wireless technologies are attractive and are frequently the consumers' first choice the additional cost will very likely mean that a wire line alternative will always exist. Also working against wireless is that it doesn't work sometimes because of RF dead spots in the home or because of interference.

The current technology of choice for wire line is HomePNA that as mentioned earlier is a technology that operates using the telephone wiring in the home. At the same time consumers would prefer a power line solution because of the ubiquity of power connections in the home. It is also perceived to be easier to use because there would not be the need for an additional network cable.

The power line technologies still have a long way to go however and as long as HomePNA has an advantage in data rate and QoS it is likely to remain the technology of choice for wire line Home Networking.

At the same time wireless technologies will evolve and will always be more attractive to consumers because of the mobility provided. This mobility however brings with it a greater need for security. The wireless standards groups will need to develop standards that provide effective encryption and authentication.

The encryption available today is considered weak and authentication for 802.11b as a standard does not exist yet. Authentication currently requires a single vendor solution but the whole point of standards is to proliferate solutions from multiple vendors. As I mentioned there are plans to deal with these issues.

The two applications that require more capability from the MAC and PHY layers are telephony and streaming.

HomePNA has developed DFPQ and HomeRF has developed voice support using TDMA channels with DECT for signaling. At this point only HomeRF is prepared to support telephony and it looks like some products will be available late this year. For HomePNA to support voice the vendor community will need to agree to a common signaling protocol. Power Line and 802.11a are further in developing QoS standards.

To support streaming media the MAC layer needs to be able to provide a bounded latency and either maintain a very low error rate << 5% or employ a mechanism for retransmission or error correction (without contributing materially to latency).

HomePNA has an advantage here with DFPQ and because the error rate over phone lines in a vast majority of cases will be low enough so that lost frames will not be noticeable.

HomeRF and 802.11b need to develop extensions to the MAC layer to support streaming.

The other summary point is that demand for greater bandwidth will continue. HomePNA might have some capabilities to substantially increase the data rate on phone lines but for wireless it will mean moving to the 5 GHz spectrum.

End to End Management: The Next Frontier

Once Home Networking technologies are develop to support advanced applications like telephony and streaming media that require QoS from the MAC layer the need for the service provider to manage this QoS will appear. Service providers will have a need to manage the network that is delivering services to end users to maintain and ensure customer satisfaction with the service. Home Networking technologies will need to provide the following to be effectively managed:

1. An open and well-defined interface for configuration and status monitoring.
2. Protocols and open interfaces to request QoS and to resolve conflicts when over subscribed
3. Protocols and open interfaces to support event reporting and fault isolation.

These capabilities along with a Management System solution will allow delivery and management of advanced services to the home over a packet based network without deploying new wires in the home.

Glossary

ADSL – Asymmetric Digital Subscriber Loop

MAC Layer – Media Access Control Layer.

PHY Layer – Physical Layer

POTS – Plain Old telephone Service

QoS – Quality of Service.

Intel's Vision Of Sports Immersion

Guy Blair and Rajeeb Hazra
Intel Corporation - Intel Architecture Labs

Abstract

Intel is exploring the technologies required to make sports immersion a reality within the next few years. Our vision is to develop a new application category that delivers highly interactive and personalized sports content to the client. This valuable content will drive a range of novel uses, ranging from Webcasting personalized sports highlights to integration with reality-based games.

Our goal is to develop the "ultimate" fantasy sports experience. To accomplish this, we will combine compelling live sports event content, sophisticated semantic video processing and 3-D rendering.

A VISION OF SPORTS IMMERSION

Intel's vision of personal sports immersion is full motion, 3-D, real time, trans-port independent, interactive sports entertainment based on live athletic events. This technology ultimately will allow the client (end user) to access virtually unlimited depth of information and viewing options while a sports event is being broadcast from anywhere in the world.

We call this "pull entertainment" because the clients will pull from the event the unique data that they desire. These data include video and query-based textual information about the event.

For example, what if you could…

- Replay of a goal being scored, but see it from any angle of view that you select — and see it in 3-D.
- See a wide range of game statistics on the performance of your team or favorite player, in real-time using simple query language.
- Keep track of players' movements to review how a goal score play was set up.
- Select a player on the live video and continuously highlight him or her throughout the game and obtain detailed player profile information on-demand.
- Assemble personalized highlights of one or more sports events, based on the specific types of events that you want to see.
- Eventually, play a 3-D, interactive simulation or game based on actual players using performances from real games, called reality-based gaming.

Currently, video broadcasts and Webcasts of sports events offer limited interactivity for the client. The levels and options of interactivity are predetermined by the vendor.

Sports immersion will transcend these limitations. It will become a new type of entertainment experience that features highly personalized interactivity. A client will enjoy the "what if" examples mentioned previously, as well as options not yet conceived.

First generation applications

There are examples of early glimpses into sports immersion technology. For example, Orad Hi-Tec Systems Ltd., based in Israel, has developed a first generation technology called "virtual sports Webcasting," based on their VirtuaLive* video system. They demonstrated this technology during the 1998 World Cup soccer events and with Premier League soccer matches that same year.

Sports.com deployed Orad's video clips on their Website. Soccer enthusiasts around the world could view the clips using Orad's VirtuaLive Console™ player, which has controls similar to that of a VCR remote.

Figure 1 depicts a generic type of controller. Using this controller, the client can search for a particular scoring event on a menu. Then, the event can be replayed. As this technology evolves, the search and viewing options of the player will reflect the increased amount of data available to access.

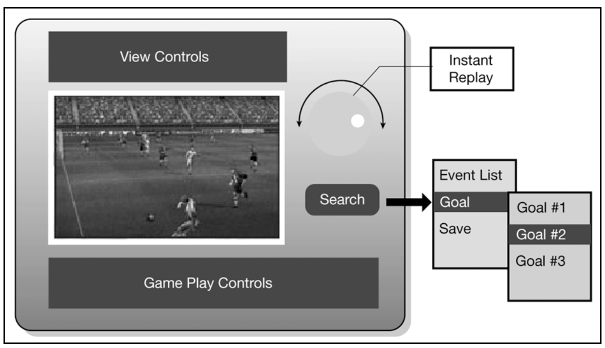


Figure 1

INTEL'S MODELING ARCHITECTURE

Development of this modeling architecture is in its early stages, so an in-depth exploration of the elements and processes is not available. However, the following covers the basics from video to production to server/broadcast to client participation.

Live video data from the sports event are generated from cameras positioned around the stadium. A tracking module processes the input video to generate 3-D coordinates for the players, ball and officials. The 3-D coordinates place each object with respect to real-world coordinates in the field of play and stadium, such as the field boundaries, goals and stands. The data are streamed to an analysis engine, where event detection and event modeling occur.

Figure 2 shows a system diagram that illustrates the sequence of processes that occur as live data are generated. The data flows from the live event through analysis and production levels to the end user device, such as a computer, television or game console.

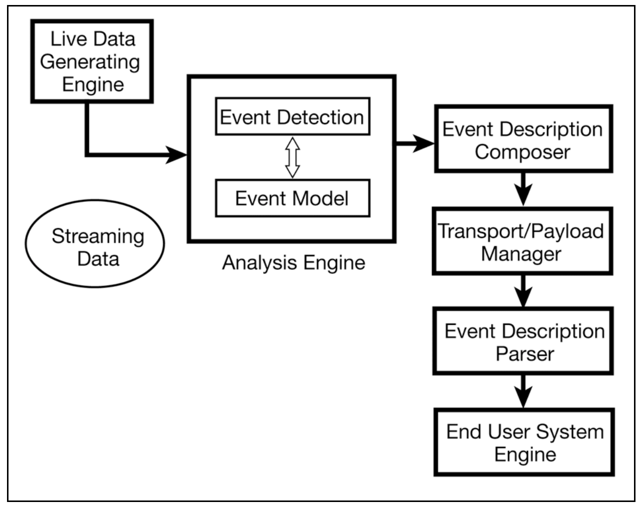


Figure 2

The modeling architecture is key to this technology. Some actions by the players can be interpreted automatically by the event detection engine. This includes player and ball movement within the field boundaries, as well as basic entity relationships.

However, certain actions cannot yet be reliably interpreted automatically by the engine. This requires human intervention. An operator must view the video and annotate it with assigned actions. At this point the model becomes "instantiated" with that additional data. For example, the operator would see a kick on the video and add the detail of it being a kick by the player's right foot. Automatic detection only would indicate ball contact with this player and ball movement away from the player, but not the details of the action.

Eventually, automatic event detection and modeling architecture will improve so that an operator no longer is needed to add semantic events, such as types of kicks, fouls and other details. This advanced level of detection is necessary to bring the immersive sports experience to live event broadcast.

Soccer has been considered a viable sport to test the early generations of immersive sports technology. But all other major sports also are candidates for this technology. For each sport, a hierarchical model must be developed.

To gain an understanding of how this technology is applied to a sport, consider the example event model shown in Figure 3. It represents a hierarchical event model for American football. The general breakdown of the sport, on the left, identifies key categories of information. The example model on the right is a specific application of this approach to a particular football game.

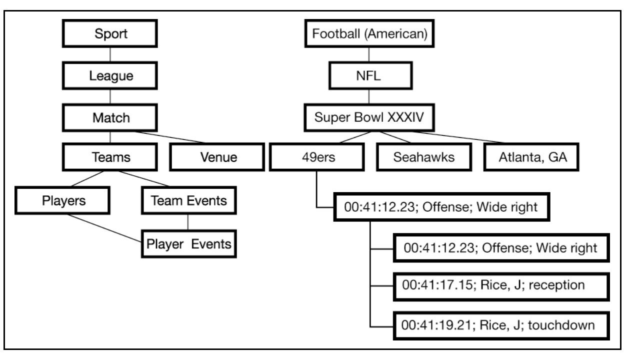


Figure 3

Adding value to content

An instantiated model of a sports event is an entirely new type of content and a valuable property. Once a sports event is recorded this way, it can be repurposed in any number of ways on a wide range of platforms.

The content could be used by networks to provide enriched live television broadcasts. And because the content includes a database of statistics and event information, game developers could create new types of reality-based entertainment.

Sports immersion content also offers flexibility with respect to the delivery plat-form. Content could be delivered via television, cable, DVD, CD-ROM, the Web or a game cartridge.

CONCLUSION

Our vision is a new way for sports enthusiasts to experience a broad range of sports and compliment broadcast television coverage of sports events. It's a way for everyone to get become a participant, not just a passive viewer.

With Intel's vision of sports immersion, everyone will be able to get into the game!

CONTACT INFORMATION

Guy Blair, Ph.D., Marketing Manager
Intel Corporation - JF1
2111 NE 25th Ave.
Hillsboro, OR 97124
(503) 264-8453
guy.blair@intel.com

Intel Architecture Labs
http://www.intel.com/ial/home/digentertain

Orad Hi-Tec Systems Ltd.
A leading developer of video and real-time image processing technologies for the TV

broadcast, post production and advertising markets.
http://www.virtualive.com
http://www.orad-ny.com

Introducing Intermodulation - Its Role in Cable Modems And Reverse Path Operation

Martin Lee & Keith Mothersdale
Channell Corporation

Abstract:

With the advent of two way communications over Hybrid Fiber Coax (HFC) networks, new technical demands are being made of conventional element of the coax drop. Products originally engineered and designed to accommodate forward path transmissions only are increasingly being used in two way systems. In these networks high level reverse path transmissions hold the key to future revenue growth and proliferation of services offered by cable operators.
This paper examines a specific problem, Intermodulation, as it relates to transmission of high level reverse path signals through RF passive products. The factors contributing to Intermodulation Distortion occurring are explained in detail and a variety of solutions currently being pursued by passive product manufacturers are presented. Isolation performance of RF passives is also discussed as it relates to optimizing reverse path stability and managing intermodulation. In conclusion, passive product design criteria is presented which should be considered carefully in the selection of next generation passives.

A History of CATV Coaxial Network Design:

Cable Television systems are designed for optimum forward path operation (in one direction only), allowing 6MHz analog video channels to move from the Headend to the home over a Hybrid Fiber/Coax medium. The products used in these networks, from amplifiers to passives, were designed with forward transmission as the key requirement. As RF travels from the cable system Headend to the home, a number of things happen. For the purpose of understanding Intermodulation and why it occurs, the key thing to remember about RF signals passing through any medium is that they attenuate or lose strength. Cable systems are engineered to ensure that the RF signal present at the entry to every home is optimized to guarantee a minimum video level at the highest frequency of operation. To accomplish this, passive products are designed to attenuate RF signals that pass through them by varying degrees. Depending on how far away a particular home is from the last amplifier (distribution and drop cable length), the dB value of a tap (coupling value) is chosen to ensure that the dB level at all homes fed from that particular amplifier is constant. Homes fed from different taps at different distances from the amplifier should all receive the same signal level in order for the set-top box to work correctly (typically 7dB at the input to the home splitter). Consider the following example (See Figure 1).
In this next example, let's look first at the forward direction. The output level of the amplifier is 50dBmV. This is attenuated by the following losses as it moves towards the home:

- 1.0dBmV Insertion Loss of the first Tap installed on the Amplifier output.
- 5.0dBmV loss of the distribution cable between Amplifier and Tap #2.
- 5.0dBmV loss of the drop cable between Tap #2 and Home B.

If a 7dBmV signal level is desired at the home entry point, a 32dB Tap must be selected in order to properly budget the available RF. Amplifier output (50dB), minus Insertion Loss of first Tap (-1dB) minus cable loss (-5dB) minus Tap dB attenuation (-32dB), minus

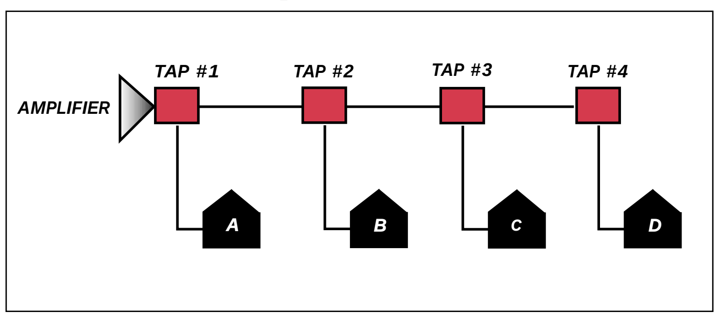


Figure 1. The RF signal level at the entry point to each home must be approximately 7dBmV. The coupling values of Taps 1 through 4 are chosen to ensure this consistent RF level at each home entry port.

drop cable attenuation (-5dB), equals signal available at the home. [50dB minus 1dB, minus 5dB, minus 32dB, minus 5dB] equals 7dB. This RF budgeting has been carried out with only the forward path considered.

CATV Coaxial Network Demands Today:

Cable systems today are beginning to utilize their reverse paths extensively. The best example of this is the rapid deployment of cable modems, which allow customers high speed Internet access through their cable television networks. In order for cable modems to operate properly, they rely on being able to receive and transmit data from the home to the Headend over the cable network. As cable systems deploy cable modems, they are learning more about the limitations of passive products in this area. The limitations have primarily to do with the passive product's ability to pass high level RF signals in the reverse direction. From a fundamental design standpoint, RF passes freely through any passive product in either direction. The technical issues have more to do with what happens to the signal as it passes. More specifically, the Intermodulation issue relates to and results from the high RF signal level at which a cable modem must transmit a signal in order to ensure that it reaches the headend accurately and within an acceptable Bit Error Rate range. Just as a RF signal is attenuated through a passive devise in the forward direction, it is also attenuated by the same degree as it passes through that devise in the reverse direction. Consider the following example (See Figure 2):

Reverse Path Signal Level Budgeting:

Let's now consider a signal originating at the cable modem and the obstacles it must overcome in order to reach the amplifier, at which point it can be amplified and continue its journey back towards the Headend. Consider the following example (See Figure 3):
In this case, we are looking at a signal originating at the cable modem output. This signal must be transmitted at a level sufficient to overcome interference due to ingress (noise funneling from home appliances etc.), and the losses of passive devices and cable between the modem and the first amplifier upstream. These losses are the same as in Figure 2, with the

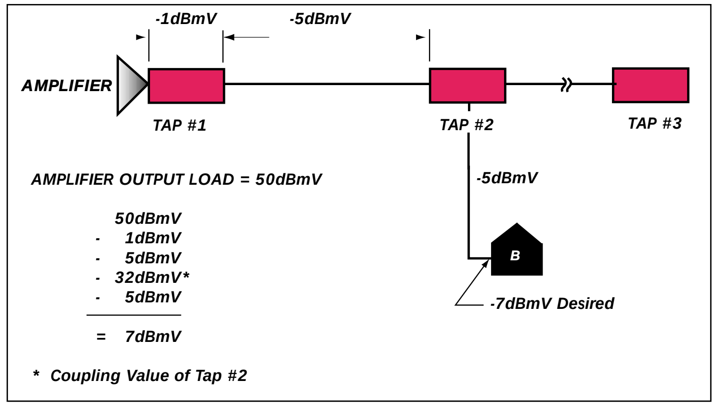


Figure 2. This example shows typical attenuation losses, in the forward direction, of taps and drop cable for a desired 7dBmV at the home.

addition of a loss through a 2-Way Splitter inside the home which was added to allow the cable modem to be connected, and the added cable between the entry point to the home and the Indoor Splitter. Beginning at the cable modem, the transmitted signal must overcome the following losses:

- 3.5dBmV loss of the 2-Way Splitter
- 5.5dBmV loss of the drop cable
- 32.0dBmV loss of the distribution tap
- 5.0dBmV loss of the distribution cable between tap and amplifier.
- 1.0dBmV insertion loss of the first tap on the amplifier output

TOTAL LOSS = 47.0dBmV

Let's assume that the reverse amplifier requires a minimum input signal level of 13dBmV in order to optimize the carrier to noise and dynamic range. The total output level of the cable modem must now be 60dBmV (47dBmV + 13dBmV) in order to overcome all losses along the way and still reach the amplifier with an acceptable signal strength.

This extremely high cable modem output level of 60dBmV is the root cause of the intermodulation problem we will now discuss.

What Is Intermodulation?:

CATV Passive products are constructed using transformers with ferrite cores and various resistive/capacitive tuned circuits. These circuits function together to ensure that the RF signal is minimally compromised as it passes through the devise in order to redirect, split, reduce it in power by a predetermined factor, or combine it with other signals.
The following circuit diagram depicts a typical 3-Way

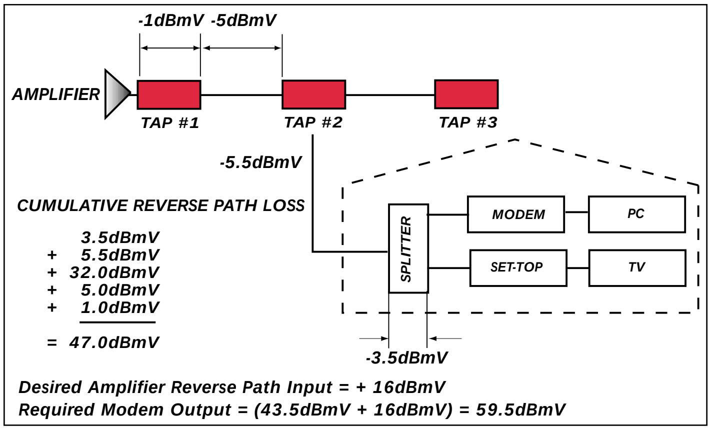


Figure 3. This example shows added attenuation losses of 3.5dBmV due to 2-Way Splitter addition for connection of cable modem.

splitter circuit:

The ferrite core material is particularly sensitive in terms of its magnetic characteristics (linear magnetization curve) and how it allows a broadband (5-1000 MHz) RF signal to pass through it with minimum interference. It is used because of its ability to man age such a broadband signal while continuing to operate linearly or within a predictable and flat range across the entire 5 to 1000 MHz spectrum. The linear operation of a ferrite core is dependent on its magnetic properties remaining within a stable range (linear region). The stable magnetic range of a ferrite can be represented by what is known as a hysteresis curve. This curve depicts the stable operating range of the ferrites magnetic property which occurs between the upper and lower hysteresis limits (linear region).

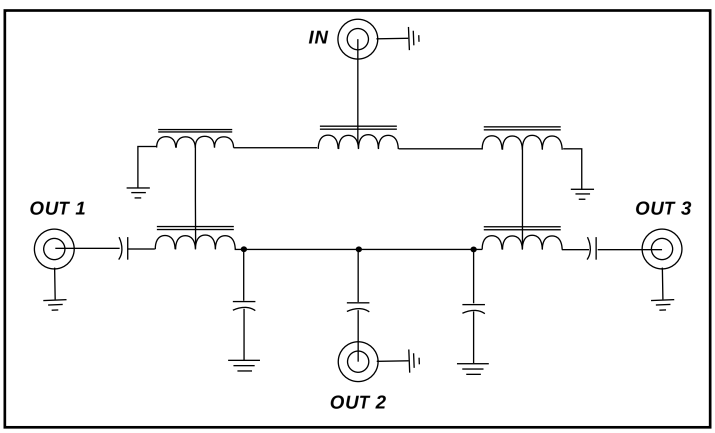


Figure 4. A typical passive device circuit is designed to redirect, split, and/or reduce signal power by a predetermined factor, or to combine it with other signals.

The Problem:

As we have demonstrated earlier, a cable modem must output signal levels over a wide dynamic range (25 to 60 dBmV typical) in order to overcome the losses it encounters on its way back to the amplifier. In fact, cable modems are designed to automatically increase their output level until they can verify the signal they are transmitting has been received at the headend within an acceptable bit error rate range. This adds an element of unpredictability to the actual output level of any given modem at any given time.

Assuming the modem output level is a realistic 55 dBmV, the first passive device encountered is the 2-Way Splitter used to connect the cable modem to the CATV network. Depending on where the ferrite material is sitting about the hysteresis curve, the ferrite cores may no longer be capable of operating within their linear range when hit by such a high signal level. The high RF level changes the magnetic properties of the core, and results in operation of that core outside of its stable hysteresis range. This is referred to as “saturation of the core” (non-linear operation).

When saturation occurs, undesired beats or signal sources are generated at harmonic's of the original frequency (fundamental frequency). For example, if the return path signal is being transmitted at 40 MHz and the ferrite saturates due to the high level being transmitted, a second harmonic beat will appear at 80 MHz (twice the fundamental frequency). This second harmonic beat now falls in the forward video band; its level is unpredictable and can distort the video channel being transmitted at 80MHz (causing severe interference on the TV picture).

Other lower level beats may appear at higher harmonic frequencies depending on the level of the original beat. Detection of the problem will only occur when the cable modem is operating and a television set close to the modem is tuned to a channel which falls on one of the harmonic frequencies. This makes the interference itself intermittent and difficult to troubleshoot if not understood by the service technician. The following graph depicts this interaction (See Figure 5).

Common Approaches to Controlling Intermodulation in Passive Products:

Manufacturers have for some time known about this problem and have made some attempts to develop and implement a solution. Initial solutions for the most part fell short in one fundamental manner. They did not address the problem at the root cause, which is the physical design, material composition and linearity of the ferrite component itself. Other contributory solutions include making the ferrite less susceptible to saturation by merely increasing its size, improving the "as produced" magnetic properties of the ferrite by ensuring it is completely demagnetized on leaving the

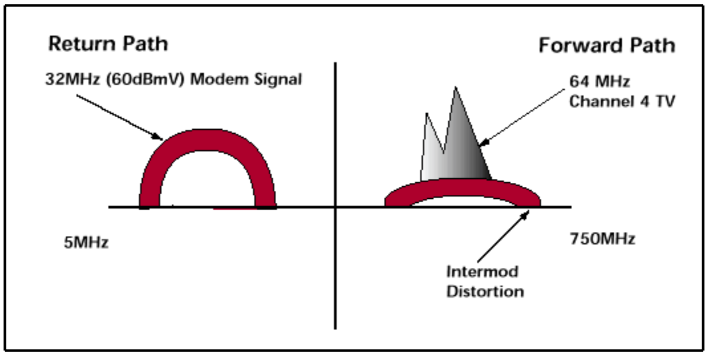


Figure 5. This example shows a return path signal that is being transmitted at 32MHz. As the ferrite saturates, due to the high level signal being transmitted, a second harmonic beat will appear at 64MHz (twice the fundamental frequency).

factory, or adding components which prevent or limit other sources of interference such as impulse noise or other forms of voltage spikes. A closer examination of these different solutions reveals some of their inherent drawbacks if relied upon entirely.

Ferrite Size: By simply increasing the size of the ferrite, it can become more capable of handling higher signal levels before saturating. However, it is not only physical size that matters, the material composition also plays an important role. Merely increasing ferrite size has a direct negative impact on the broadband performance of the ferrite. In other words, as the size increases it becomes increasingly difficult to maintain desired return loss and frequency response characteristics across the 1000MHz band and the lower and upper portions of the band are most severely affected.

Demagnetizing Ferrites in Production: This process is known as "Degaussing" and involves passing the final product through a degaussing chamber just before final packaging at the factory. The resulting ferrite is demagnetized and operating at optimum linearity on leaving the factory. The problem with this solution is that various things can happen in a standard network to re-magnetize the ferrite rendering the product susceptible once again to saturation. The most common re-magnetizing force is a voltage spike caused by impulse noise or lightening. It actually requires rather low voltage levels to change the magnetic properties of the ferrite once again rendering its intermodulation performance unpredictable. Changes in temperature can also vary the magnetization characteristic.

Adding Components to Prevent Re-Magnetization of the Ferrite: Many manufacturers are now adding blocking capacitors with various voltage ratings to all ports of an indoor splitter. These blocking capacitors essentially reduce the level of the voltage spike reaching the ferrite, thereby preventing the ferrite from being re-magnetized. This method does help reduce the effects of low voltage spikes, but high voltage spikes can still induce enough magnetic change to cause the ferrite to operate non-linearly. It is therefore at best only a partial solution or defensive measure against the problem.

Isolation and It's Role in Cable Modem Operation:

Isolation refers to the ability of a product to prevent a RF signal from coupling onto adjacent ports and interfering with other subscribers or other sets connected off the same passive product. For example, the signal transmitted by the cable modem, in an ideal situation, travels directly back to the headend and the isolation performance of the passives through which it must pass is sufficient to ensure that it does not interfere with adjacent ports.

Unfortunately, once again the high output level of a cable modem places unusual isolation demands on particularly the first passive it encounters, since its signal level is highest at this point. This again will be the indoor splitter used to connect the modem into the CATV network. Because of this, it is not only necessary to prevent intermodulation in the design of the splitter, but also to increase the port-to-port isolation performance of the splitter to protect adjacent ports from the high level cable modem output signal.

Passive Products - Design Solutions:

The critical consideration in controlling or eliminating intermodulation in passive devises lies in maintaining the broadband linearity of the ferrite component itself. A multi-stage solution involving careful selection of ferrite materials and fine-tuning of ferrite manufacturing processes has been shown to predictably provide a level of linear broadband performance greater than ever achieved in the past. Advances in circuit designs have also been shown to achieve a guaranteed higher level of broadband RF performance. In addition to these measures, some of the performance insurance measures presented above, such as optimizing the physical size of the ferrite and adding blocking capacitors, should be incorporated to provide redundant protection.; The selection of PCB material, tolerances of individual components and production tuning methods have also been attributed with Isolation performance improvements in the order of 10dB over conventional passives. This is particularly necessary over the important reverse path frequency band where high level signals are present.

Biography:

Martin Lee is the Managing Director for Channell, RF Products Division in Charlotte, North Carolina. Presently he is actively involved with the Society of Telecommunications Engineers (SCTE) and serves as Director of Working Group II on the SCTE Interface Practices Subcommittee. This Working Group develops industry standards for coaxial products used in broadband networks.

IP Data Over Satellite To Cable Headends And A New Operation Model With Digital Store And Forward Multi-Media Systems

Paul Harr
Wegener Communications

ABSTRACT

The success of the Internet has proved it will be an important service component for cable operators. The advent of high-speed IP access has opened the door for rich multi-media content to be played out on consumer PC's.

High-speed cable modems with data rates up to 5Mbps allow the savvy PC user to not only use his PC for surfing the Internet, but also for downloading streaming video and audio. Increasingly, the PC is being used to receive multi-media content while its user is also punching out a little work .

However, delivery of rich media content requires high performance from high-speed IP access systems. Today, performance has not been a critical issue as high-speed Internet users believe they are receiving improved performance over analog dial-up. But, this is changing and fast. As more users come on-line, the infrastructure to access the huge amounts of content will become the bottleneck.

This paper will analyze the growing problem IP Data delivery to high-speed access systems. And, will offer technical solutions that include satellite IP Data backbone to overcome the limitations of the terrestrial infrastructure.

INTRODUCTION

The Internet explosion has created a new and lucrative revenue source for the cable operator. With modem technologies standardized using the Data Over Cable Service Interface Specification (DOCSIS) 1.0 standard, subscribers can realize access rates up to approximately 5 Mbps. Although, few if any, cable operators are actually delivering that performance today. Most are rate-limiting service rates to about 1 Mbps.

Subscribers are signing up for cable modems with the belief they will have the fastest access performance available. And, in a market with relatively few subscribers, performance limitations go unnoticed. However, that is about to change.

Three factors will soon uncover the limitations of cable's Internet access infrastructure and the need for an alternative solution. First, the number of subscribers signing up for cable modems is outpacing the build-out of infrastructure. Currently, there are approximately 1.8 Million cable modems installed in North America. However, by 2003, that number is forecast to exceed 15 Million.

Second, the cumulative demand of high-speed cable modem users will exceed the economic feasibility of the terrestrial infrastructure to deliver content at the same high performance.

Third, Internet users are signing up for high-speed access because they want fast access to rich media content. This does not mean simply surfing the web, but downloading rich multi-media files such as streaming audio and video multicast services. A growing number of Internet Service Providers are providing streaming multi-cast services.

Background

Cable's infrastructure has evolved over the last 10 years from being a distributed network of headends to a more centralized architecture. The old distributed model required many headends serving far fewer customers than today. The reason for this distributed architecture is that the distribution or transport technology was all coaxial cable. Coaxial cable transport is very lossy and requires distances from headends to be relatively short so that picture quality is maintained.

Today's new cable architecture model has moved to a fiber-based model that is more centralized serving much larger subscriber bases. This model has great appeal because it allows consolidation of headends within a given market while increasing bandwidth and maintaining constant quality of video service (QoS). Fiber to the serving area (FSA), as it is called, has in large part created the massive cable system consolidation and horse-trading of cable systems among multiple system operators (MSO's).

By using FSA architectures and centralizing cable headends, operators achieve excellent economies of scale for video services. Since traditional video services are broadcast only, reducing the number of headends also reduces the expense of operating multiple headends and improves service quality.

The Problem

The problem with the centralized model of FSA is that it was never intended to deliver all the advanced interactive services being placed upon it. In addition to traditional broadcast cable services, the network is delivering video on demand (VOD), Internet access and IP telephony. The very nature of interactivity requires unique session with each subscriber and distributed access.

And, unless the network has been designed to provide 100 per cent availability, there is the high likelihood subscribers will not gain access to some services. In particular, high-speed Internet access will be seen as taking a step backward with very slow performance.

As a result, a centralized architecture will not economically scale in bandwidth performance.

Subscriber Growth

One of the factors that will stress the current IP infrastructure is cable modem subscriber growth. Cable subscriber growth estimates for high-speed Internet access varies widely among forecasters. Figure 1 below offers a range of projections. Low estimates are for only six million cable modem subscribers by 2003. High estimates have the penetration exceeding sixteen million over the same period.

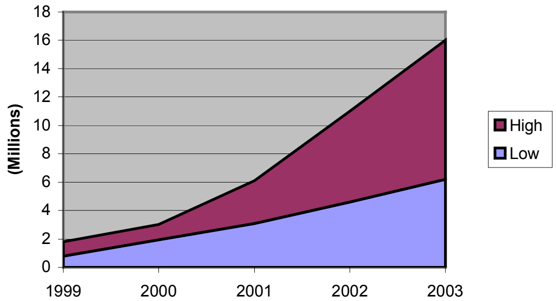


Fig. 1 Cable Modem Shipment Estimates

If the conservative estimate of six million by 2003 is used, cable modem penetration of almost 10 per cent would be achieved. In a cable system of 250,000 subscribers, 25,000 would be using cable modems for high-speed access. If the aggressive forecast is used, over 57,000 users would subscribe to cable modem access. While its unlikely that all subscribers are on-line simultaneously, cable Internet access is being billed as always on service. The following table illustrates the potential demand from the system.

	1.5 Mbps	5 Mbps
Low Estimate (25,000)	37.5 Gbps	125 Gbps
High Estimate (57,000)	85.5 Gbps	285 Gbps

Figure 2 Maximum Bandwidth Demand

Running out of Bandwidth

Cable Operators are deploying the latest fiber technology available. Using SONET and ATM, gigabit transport within the cable fiber network is possible. See Figure 3, Cable's Fiber Architecture. Connected to these internal pipes are incoming terrestrial transports up to OC-48 rates. These huge pipes virtually guarantee plenty of bandwidth for IP Data.

However, Internet access over terrestrial infrastructure creates a session or connection with the backbone for every subscriber. As subscriber penetration increases and the demand for more rich multi-media content increases, the terrestrial connection to the Internet will become the weakest link in the transmission chain. Even though software caching systems can be used to distribute Internet content, the cost of having high bandwidth terrestrial pipes up to OC-48 in every cable system becomes prohibitively expensive for an MSO.

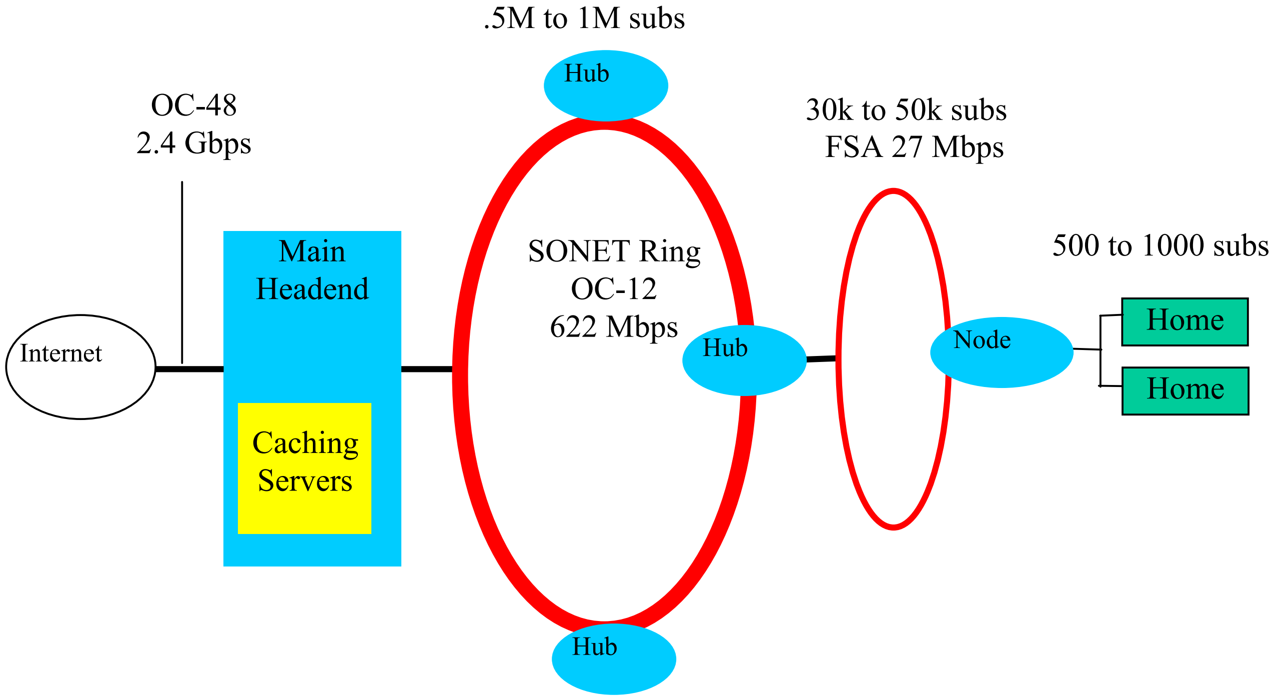


Figure 3. Cable's Fiber Architecture

Rich Multi-media Content

The final factor that will affect high-speed Internet access performance is the demand for rich multi-media content. Internet users today are already downloading multi-media content. In addition to large file download, audio files are in high demand. Many web sites have sprung up offering MP3 audio files for download. And while the copyright issue for downloading MP3 files is still hotly contested, the technology has proved the Internet can deliver CD quality audio. Its only a matter of time before the business enterprise infrastructure supports the commerce of MP3 distribution.

Streaming video is also one of the fastest growing content areas on the Internet. Users want instant and immediate access to news and information without waiting for television to broadcast the story. In addition, many Internet users want live events and previously recorded content for viewing. Users are willing to accept a lower quality video on their PC in order to get immediate access to information. However, again the technology is making it possible for video delivery over the Internet. MPEG2 and MPEG4 encoding and decoding provide a good quality real-time video, if the network can deliver the bandwidth needed for bitrate performance. Acceptable performance for streaming video at rates of 500 kbps provides good quality on small PC viewer.

However, as more subscribers use high-speed Internet access for streaming video and audio, the incoming terrestrial infrastructure becomes more of a bottleneck.

IP Data over Satellite Alternative

As high-speed Internet access over cable continues to grow, cable operators will have to look at alternatives for accessing rich media content. Satellite provides an excellent alternative to high-bandwidth terrestrial pipes. Terrestrial Internet access alone does not provide enough bandwidth to satisfy real-time demand. Nor, does the cost for high-bandwidth terrestrial access compare with satellite transmission costs.

In order for the high-speed Internet access experience to remain enjoyable as penetration grows, a distributed approach to Internet rich media content must be considered. In figure 4 it can be seen that using satellite backbone infrastructure along with a content caching system, high demand content is pushed to the edge of the network where it is cached locally on IP Data servers at the cable headend. The distributed model can be further extended to the edge by locating IP Data receivers at fiber nodes.

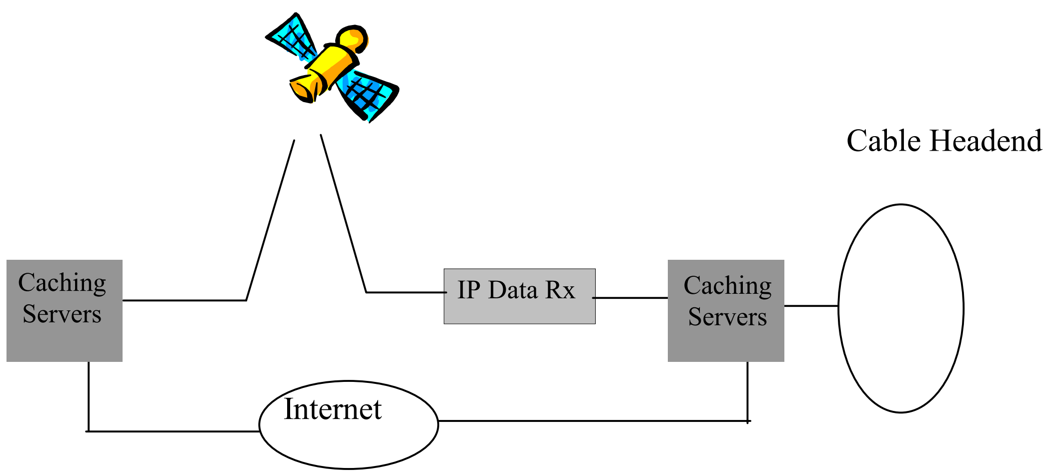


Figure 4. Satellite IP Data Backbone

Intelligent software manages the content and transmits only the information that has changed. This relieves the burden of repeat transmissions of the same content in terrestrial networks. It also eliminates multiple sessions of the same content to different users.

Satellite transmission augments the terrestrial infrastructure with data rates up to 70 Mbps. In a 24 hour period 756

GigaBytes of information can be replicated at thousands of downlink locations at a fraction of the cost. By augmenting terrestrial with satellite backbone the Internet Service Provider can provide a high QoS and rationalize the cost of Internet access.

Conclusions

Today's fiber rich cable networks are designed as centralized access points to traditional media content. As services such as high-speed Internet access delivering rich multi-media content emerge, the centralized model will no longer work. ISP's and cable operators will need to look at alternatives to provide high bandwidth desirable content to the network's edge.

The best technology to do this is a satellite IP Data backbone infrastructure. Terrestrial networks are expensive fat pipe access points which are required for every cable headend or ISP. Satellite overlay provides a store and forward content distributed model which eliminates congestion caused by repeat transmission in the terrestrial backbone. Internet access costs can be significantly reduces using satellite for cached content and terrestrial for occasional real-time access.

However, both infrastructures are still required for effective high-speed Internet access. While satellite provides an excellent method to distribute content economically to the network's edge, real-time performance is superior with terrestrial networks.

High-speed Internet access infrastructures will be challenged as rich multi-media content develops and personal computing power increases. The interactive experience of the PC as a multi-media tool is limited only by the infrastructure where it is connected. High-speed Internet access providers will ultimately differentiate themselves by the quality of service delivered by their networks and the technology used to provide high-speed access.

Paul can be contacted at:

Paul Harr
Vice President, Marketing
Wegener Communications
11350 Technology Circle
Duluth, GA 30097

Tel: 770-814-4048
Email: pharr@wegener.com

Local Commercial Insertion in the Digital Headend

Mukta L. Kar, Cable Television Laboratories, Inc.
Sam Narasimhan, Motorola Broadband Communication
Richard S. Prodan, Terayon Communication Systems

Abstract

Existing ad insertion systems enable cable headends and broadcast affiliates to insert locally generated commercials and short programs in a channel seamlessly before delivery to the home. The revenue generated by local ads and short programs is very significant. Current ad insertion systems are hybrid systems, where commercials are stored in the ad server in digitally compressed form, while splicing is accomplished using analog techniques. The compressed commercial is decompressed and converted to analog format before insertion into an analog channel. The complete digital technology for delivery of compressed digital audio, video, and data is superior to existing analog methods. However, dealing with compressed video poses several challenges. One such challenge is splicing into a compressed digital bitstream compliant to the MPEG-2 standard seamlessly, that is to say, without adversely affecting the display due to decoding discontinuities at the insertion point. These applications involve insertion of locally generated compressed digital commercials and short programs into a digital channel containing previously multiplexed and digitally compressed programs.

The Society of Motion Picture and Television Engineers (SMPTE) has defined standard SMPTE 312M "Splice Points and Cue Messages for MPEG-2 Transport Streams" to promote interoperability of compressed digital video splicing equipment. SMPTE 312M is aimed at studio broadcast and requires a constrained (a synonym for pre-conditioned) MPEG-2 transport stream. For use in deployed digital cable systems, this will require significant changes to encoding methodologies to enforce predetermined conditions at the splice points.

The Society of Cable Telecommunications Engineers (SCTE) Digital Video Subcommittee (DVS) has established a splicing standard DVS 253 "Digital Program Insertion Cueing Message for Cable" to provide cue messaging and splicing in a more digital headend friendly manner, which does not require restrictions or constraints on MPEG-2-compliant transport streams. This paper will present a brief overview of existing analog/hybrid program insertion systems. The limitations of SMPTE 312M will be discussed. The solutions employed in DVS 253 will be described including digital headend friendly features. The implementation of DVS 253 to insert compressed commercials at the headend, including the issue of invisibility from commercial killers, will be addressed. As cost is related to complexity and flexibility of implementation, complexity issues and network operational constraints will also be examined for various implementations of digital program insertion systems and constituent components.

Introduction

Traditionally, the cable industry has been a popular conduit for the delivery of premium analog television content to the home. The access to the broadcast premium content was controlled using an analog set top box at the consumer's premises. The signal scrambled at the headend could only be descrambled using proprietary hardware embedded in the set top box. The advanced audio/video compression technology and subsequent standardization along with advances in digital modulation has made delivery of digital audio/video to the consumer's home a reality. The digital technology is very bandwidth efficient and provides superior video quality compared to its analog counterpart. In the last few years the cable industry has been deploying delivery of digital signals using a hybrid analog/digital set top box mainly for premium and pay-per-view (PPV) services.

Over the years, cable systems have grown significantly to become a primary conduit for the delivery of television contents to the majority of the homes. The industry also started inserting commercials both at national and local levels in some channels to enhance their revenues. Currently, the revenue generated from local ads and short programs contribute significantly to the total revenue of the cable industry. The revenue generated from local ads and short programs has been growing significantly.

Typically when a content is created, a few avails are included to insert commercials. A certain percentage of the avails are used for national ads and the rest for the insertion of local commercials. However, the networks (such as CNN, TBS, etc.) fill all the avails with the national ads before broadcasting to the headends using satellite. The headends can replace a percentage of these national ads with locally inserted commercials using the network cue signals. Basically the cue signal has three components, the Pre-roll (setup time), the Start (beginning of an avail) and the Stop (end of the avail). The Pre-roll is sent in advance to allow downstream devices at the headend (such as the ad-server and the splicer) time to prepare for the insertion. When the Start signal is received, the splicer switches from the network program to the commercial. After receiving the Stop signal the splicer leaves the commercial and returns to the network program.

Until a few years ago most of the ad-insertion systems were analog where the ad-insertion equipment is analog and analog storage devices were used to store the ads. However, the majority of the ad-insertion systems have been replaced by a hybrid one. In a hybrid system, the ads are stored in digital compressed format and the ad insertion equipment is analog in nature. In the hybrid system, the digitally compressed commercial is being decoded and converted to analog before insertion.

Analog Ad Insertion

In the analog domain, splicing between two NTSC video sources or TV programs is relatively easy. A switch between two synchronized video sources at any frame boundary (i.e. the vertical blanking interval) can be accomplished, since the frames are equal in size and time duration and are synchronized to one another. The resultant video, with the switched video source at the insertion or splice point, will appear seamless when displayed.

Figure 1 shows the block diagram of a typical commercial insertion system in cable headends and broadcast stations. In existing analog systems, networks distribute their content to cable headends and broadcast affiliates via satellite modulating a baseband bandwidth close to 10 MHz. As shown in Figure 2, the lower portion of the transmitted signal spectrum is used for NTSC video. The upper portion of the spectrum above 4.2 MHz contains a few FM subcarriers, which are used to transmit mono, stereo, multi-language audio, and data. One of the subcarriers is used to transmit cue-tone signals. The subcarrier is modulated with a cue-tone signal using frequency modulation (FM) or frequency shift keying (FSK). Some networks send cue-tones using a pair of FSK tones (e.g., 25 Hz and 35 Hz) within one of the audio channels. The integrated receiver-decoder (IRD) receives the RF-modulated signal, demodulates the cue-tone subcarrier signal, and provides an output signal. Either a dual-tone multiple frequency (DTMF) audio cue-tone sequence and/or a relay contact closure signal are provided to the insertion equipment. The networks send their schedules and spot availability (avail) in advance for local advertising insertion in a program. The precise location of the spot is signaled by a cue-tone during program broadcasts. The cue-tone consists of a sequence of numbers indicating the start and the end of an insertion opportunity. For example, the Discovery Channel uses a DTMF local avail cue-tone with 826* indicating the start, and 826# indicating the end of the spot.

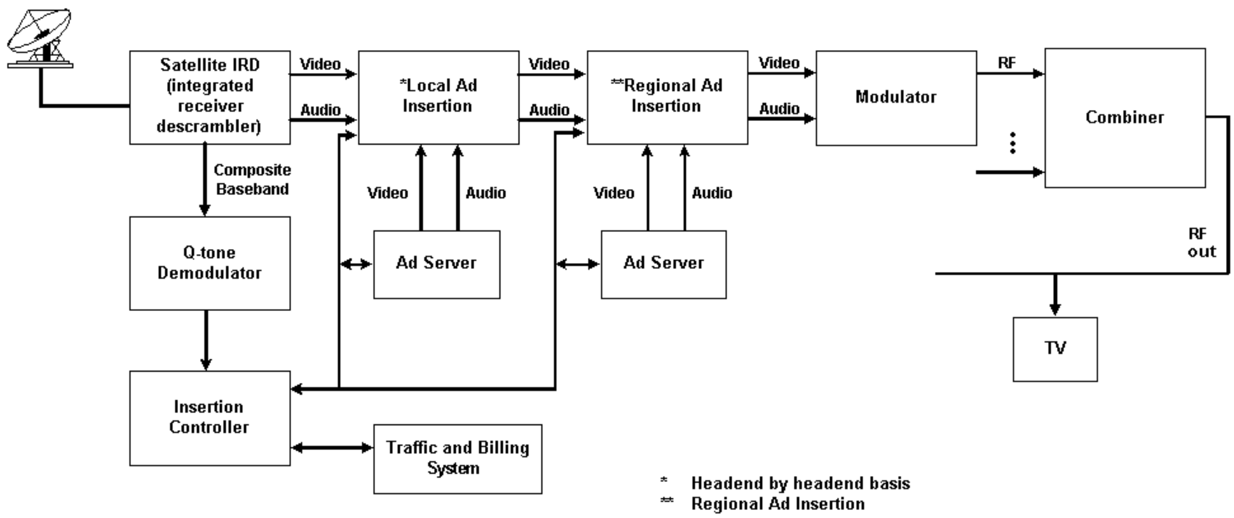


Figure 1. Typical Commercial Insertion System

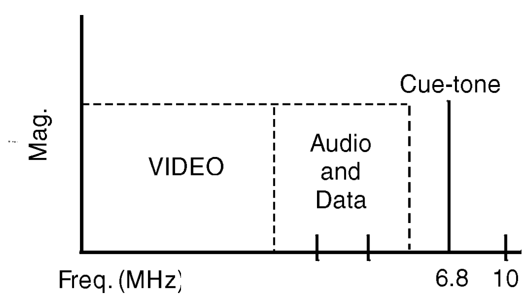


Figure 2. Analog Baseband Used by Networks

<u>Digital Program Insertion</u>

At the time of standardization of the MPEG-2 compression and system/transport layers [1], it had been expected that switching and splicing of compressed elementary streams and transport streams will be needed for editing commercial insertions, etc. Subsequently some hooks were provided for this purpose. After MPEG-2 standards were completed, the Society of Motion Picture and Television Engineers (SMPTE) decided to create a standard for splicing compressed digital streams. SMPTE completed the splicing standard in 1998 and is designated as SMPTE 312M [2]. SMPTE 312M is more suited to post production studios. Some of the major difficulties with SMPTE 312M are:

The splicer requires a constrained input bitstream.

Cue Signaling is not flexible enough. Cue signaling can only be sent to perform splicing on all the component streams of a program but not on a single component if needed.

No encryption mechanism for cue signaling was provided.

The impact of the requirement of constrained bitstreams for splicing alone is significant. For example, existing MPEG-2 compliant TV contents have to be pre-conditioned before any local commercials can be inserted into them. Also the installed encoders either have to be modified or replaced by the new ones. In essence, the standard is not very cable friendly for insertion of local compressed commercial in a cable centric statistical multiplexing environment [3]. The Society of Cable Tele-communications Engineers (SCTE) Digital Video Subcommittee (DVS) reviewed the SMPTE standard in detail and decided to work out a separate, more cable-friendly standard. SCTE established the digital program insertion ad-hoc group to work out such a standard in 1998. After working for more than a year, the ad-hoc completed their work, which has become DVS 253 titled "Digital Program Insertion Cueing Message for Cable" [5].

In creating such a standard no precondition or constraint for the MPEG-2 transport stream is assumed. The standard does not control the behavior of the splicing device in any way. The cue message carries timing information, which the splicing device may use to perform the splice. In other words, it may be treated as an opportunity for splicing but not a command. Cue messaging, if required, may be passed on to the downstream equipment such as a pass-through via a remultiplexer to a set-top box. The timing correction needed after remultiplexing may also be transmitted to maintain timing accuracy of the cue signal.

A cue message has basically two components – the first schedules the insertion points far in advance (such as the previous week). The second defines frame accurate insertion points. These messages are carried in an MPEG-2 private section syntax and called a splice_information section. A splice schedule defines the insertion time using coordinated universal time (GPS UTC) accurate to one second, while the splice insert uses MPEG PTS time, which enables frame accurate insertion.

Another important feature added to DVS 253 is the capability to encrypt cue messages. This will enable selective authorization of splicing devices in multiple headends. If a headend is not authorized, its splicing device will not be able to identify the insertion spot. The headends, which are authorized, will get the correct information about an avail.

In addition to these two message components, each splice command enables splicing of complete programs or individual components of a program (such as Video or Audio or Data). Today's systems use only the program level splicing where all the components of a program are replaced at the splice point. In the future, we will see programs and advertisements that are 'enhanced' using data broadcast as well as programs that offer interactivity. These systems will use the component level splicing where only selected components of a program are replaced at the splice point. The component level splicing enables pre-loading of data streams that are part of the same program by splicing the data component ahead of video/audio content. This may be done to load and run the data enhancements in the receiver's application engine. In the future, we will also see part of the bandwidth being used for 'opportunistic data' carriage and these components of the program are not spliced over during the insertion.

The components of a program are identified using the component_tag defined in the program map table using a new descriptor called the stream_id_descriptor. In this command mode, each component of the program can signal a different PTS time for insertion and the components that are not signaled are continued in the stream. When the inserted streams (such as enhanced commercials) have more components than the network program, the program map table will typically include the definition of these additional components even though they are not present in the original program. At the splice point, these components are enabled or disabled. Typical enhanced commercials could include delivery of discount coupons, prize drawings, or free software.

The use of the splice event_id and equipment interfaces such as between the splicer and ad_server are not defined. The DPI group has started working on the definition of APIs between the splicer and compressed commercial servers. DVS 253 along with standardized APIs will allow splicers and commercial servers to be interoperable.

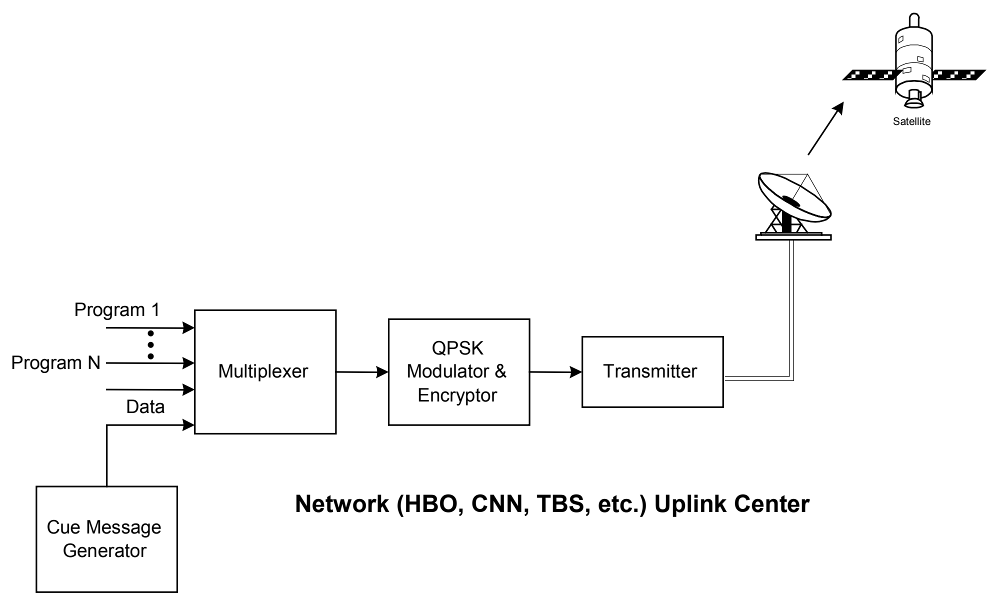


Figure 3. Digital Cue Messaging System (per DVS 253)

Figure 3 shows a typical block diagram depicting how cue messages will be multiplexed with a program before transmission to the headend via satellite. A Cue Message Generator generates a cue-signal per DVS 253. The digital cue message gets multiplexed with the other elementary streams to form an MPEG-2 compliant transport stream (TS). This baseband TS then can digitally modulate a RF carrier and be distributed to headends.

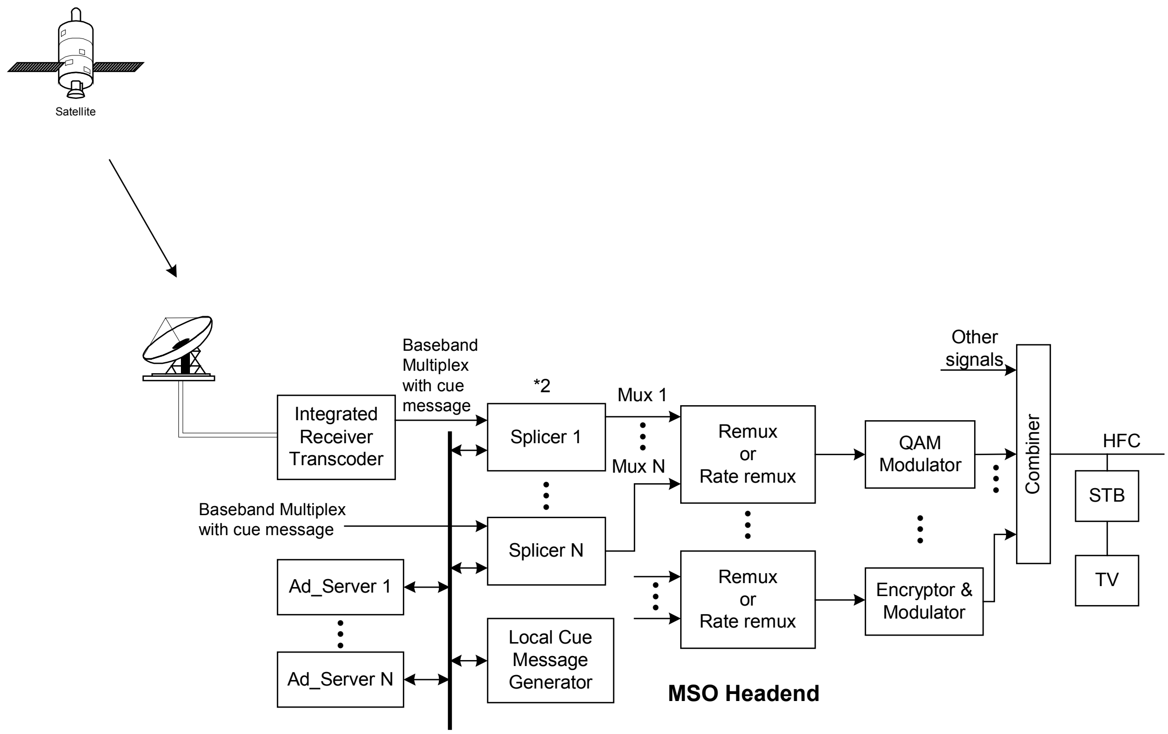


Figure 4. Typical Commercial Insertion System per DVS 253

Figure 4 shows the block diagram of a digital headend where local commercials are inserted into a program utilizing the cue message multiplexed in the transmitted program. The satellite integrated receiver demodulates the RF signal to its baseband MPEG-2 transport stream and decrypts it. The cue message, multiplexed in the transport stream is detected and decoded by the splicer. If it is a "pre-roll" it communicates with the ad_server to prepare. When cue message with a "splice_insert" is detected, the splicer switches from the network program to the local ad. At the end of the avail, the splicer switches back to the network program. A headend may have multiple splicers and ad_servers interconnected among them so that the same commercial need not be stored in more than one server. If for any reason a network cue signal is not detected or decoded properly, a local cue signal generator may be used instead. The local cue signal generator may be used as a backup system.

The cue message standard DVS 253 does not specify how to splice between two bitstreams. The techniques and resultant constraints for splicing compressed streams have been left to the innovation of the splicing equipment manufacturer. Splicing may be designed to splice seamlessly, near-seamlessly, or in a non-seamless way. Also, the compressed bitrate of the inserted ad may need to be reduced to the original program compressed bitrate to provide a seamless transition (i.e. the spliced stream cannot violate the MPEG-2 TSTD buffer model). Figure 5 shows the typical trajectory of a decoder buffer fullness. The constant transmission rate of the compressed video bitstream fills the buffer and the decoder removes variable size compressed frames (such as I, P or B) every display period. For the MPEG-2 main profile at main level (MP@ML), the minimum size of the decoder buffer must be 1.8 Mbits. Decoder buffer overflow or underflow should be avoided so that no audible or visual artifact is created.

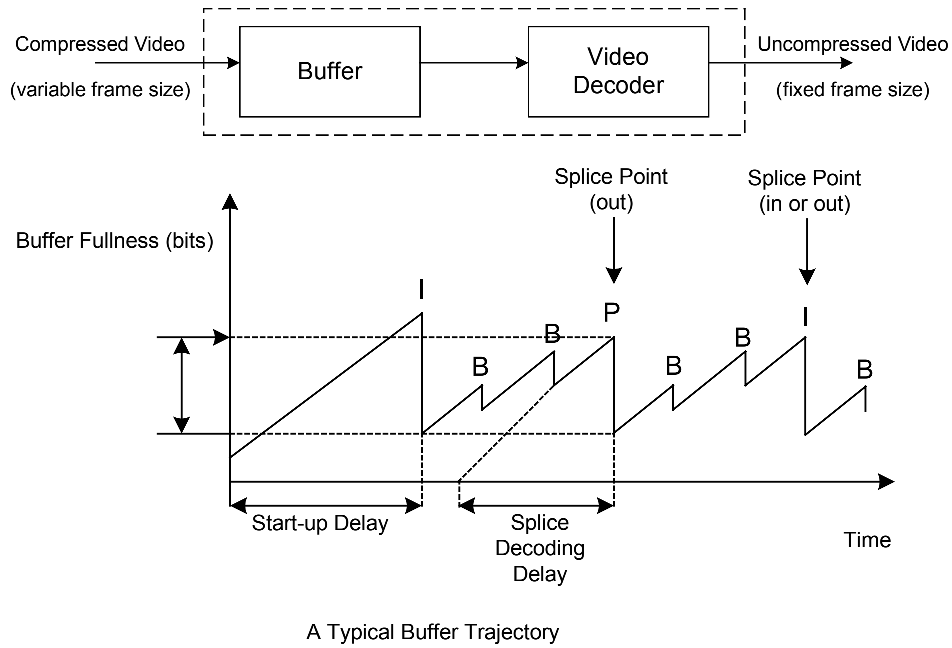


Figure 5. Simplified Representation of a Video Decoder

Splicing for Unconstrained, Seamless Insertion

Statistical multiplexing of compressed video streams helps to utilize digital channels better than constant bitrate encoded streams, assuming that the peak demand for bits from all video encoders does not coincide. The important constraint here is that all the videos need to be encoded while multiplexing. A previously compressed stream typically had to be decompressed and re-encoded before it can be inserted in a statistical multiplex. An alternative to such an approach to splicing compressed video streams is to use remultiplexing or "grooming" so that new services such as digital ad insertion may be added more efficiently without compromising visual quality due to cascaded decompression and recompression.

A *remultiplexer* receives one or more multiplexed streams as input and creates a new output multiplexed stream from local operator selected programs such as local ads. A few of the important functions performed by a remultiplexer are:

Demultiplexes the input multiplexed streams (unbundles the individual programs);

Creates a new multiplex out of the operator-selected programs and includes PSI for the new multiplex;

Maintains the bitrate constraint such that sum of all elementary stream bitrates, including PSI does not exceed the transmission channel rate;

Removes jitter in Program Clock Reference (PCR) time stamp values and maintains A/V synchronization within the programs;

Provides perceptually seamless switching capability from one program to the other without any audible or visual artifacts.

Nominally, a remultiplexer does not alter the bitrates while constructing a new multiplex out of the input multiplexed streams. The technology that deals with the multiplexing of compressed video streams along with other streams (audio and data), and trims the resulting multiplex to match an assigned constant total transmission channel bitrate, is known as *rate-remultiplexing* [4]. Rate-remultiplexing meets the latter constraint by transcoding individual video stream within the output multiplex. Transcoding is the technique by which a compressed video stream is translated to a lower bitrate strictly within the compressed domain. Thus without cascaded compression, degradation in picture quality is not noticeable with occasional or moderate reductions in the average bitrate of individual video streams.

The above advantages of rate-remultiplexing make it a very attractive choice over standard remultiplexing. Rate-remultiplexing combines existing remultiplexing technology with a new capability known as transcoding. Transcoding can reduce the bitrate of MPEG-2 compressed video without fully decoding and re-encoding a bitstream with the attendant loss of picture quality inherent in such cascaded compression.

In creating compressed commercials, content producers can produce one high-quality version and store it in a server. But, based on the availability of digital bandwidth in the multiplex at various times and at various systems, the commercial's bitrate will have to be reduced during insertion. It is also possible to create different versions of commercials with different bitrates. However, storing different versions of the same commercial could be redundant if rate-remultiplexing is employed. Rate-remultiplexing technology provides the capability to insert compressed digital commercials into digital channels at the headend and removes the need to match the bitrate of the local compressed commercial with that of the remotely transmitted program, or creating and storing different bitrate versions of the commercials in the ad server.

Related Standards Development

Currently two work items are in progress in the Digital Program Insertion Ad-Hoc Group of the SCTE DVS Subcommittee – development of a standard splicing API and a splicing Guide for the use of DVS 253.

The goal of the API is to create a common interface for communication between the Ad Servers and Splicers for inserting ads into any multiplex. The API will be flexible enough to support multiple Ad Servers attached to one Splicer and one Ad Server attached to multiple Splicers.

The Guide will explain briefly the features of DVS 253. This will help not only to minimize the chance of misunderstanding of the features of the standard but will also reduce the miscommunication between the creators of DVS 253 messages (the networks) and the ad insertion equipment manufacturers.

ATSC specialist group T3/S8 has just completed the specifications for 'Directed Channel Change' (DCC). This is an extension to PSIP standard and specifies the signaling to automatically switch a 'virtual channel' based on the user preference at appropriate times. The specification also signals the end of this event so that normal programming can resume. This feature is expected to be used for applications such as targeted advertisements as well as enhanced advertisements. The time reference in this specification is based on UTC/GPS time and is only accurate to 1 second and the specification is very similar to the splice schedule command. Acquisition of the new 'virtual channel' as well seamless transition to the new channel requires splicer architectures that are very similar to what is presented in this paper. In this model, the streams for advertisement are expected to be within the same transport multiplex identified by a different virtual channel number so that the channel change can occur at the receiver. In the cable headend model the server is resident at the headend and the switching occurs in the headend instead of at the end user equipment.

Conclusion

In summary, the problems and solutions for flexible, unconstrained insertion of local digital compressed programming such as local ads into remotely downlinked digital video multiplexes has been described. The evolution of technology from previous analog and hybrid ad insertion systems to this new digital insertion/remultiplexing technology has been described. The features of the new digital program insertion cueing message standard for cable were summarized. Splicing technology using rate-remultiplexing for unconstrained, seamless local insertion and grooming of remotely compressed digital video programming was described. Related standards developments were noted.

References

1. CODING OF MOVING PICTURES AND ASSOCIATED AUDIO, ISO/IEC 13818-1 (Systems), ISO/IEC 13818-2 (Video).

2. *Splice Points for MPEG-2 Transport Streams,* SMPTE Standard 312M, December 1997.

3. *Digital Program Insertion for Local Advertising,* by Mukta Kar, Majid Chelehmal and Richard S. Prodan, NCTA Technical Papers, 1998.

4. *Rate-Remultiplexing: An Optimum Bandwidth Utilization Technology,* by Richard S. Prodan, Mukta Kar, and Majid Chelehmal, NCTA Technical Papers, 1999.

5. *Digital Program Insertion Cueing Message for Cable,* SCTE Standard DVS 253, December, 1999.

Acknowledgements

The authors would like to thank Carol Derr of AT&T, Joe Davis of Atlantek Electronic Development Corp. and other members of the Digital Program Insertion (DPI) Adhoc group of SCTE for their participation in the various digital program insertion technology development and standardization activities.

Low Complexity Real-time Video Encoding for Soft Set-Top Box Platforms

Krasimir Kolarov, Feei Chung, William Lynch
Interval Research Corporation

Abstract

This paper presents a very low-complexity wavelet-based video codec, the Wavelet Z-Codec (WZD), implemented in real-time on a programmable media chip, the Media Accelerated Processor (MAP-1000).

The WZD algorithm is carefully designed to require only adds and shifts. Temporal wavelet transforms are used in place of motion estimation and compensation. With block processing and intermediate coding, WZD is 15 to 20 times less complex than MPEG2. An entire encoder requires only 50K gates and 128KB of memory and can be implemented for a couple of dollars as a chip macro. Experimental results have shown that it achieves similar performance to MPEG2 for typical bitrates and content.

1. INTRODUCTION

The fast development of broadband technology brings video capabilities to everyday consumers. The next generation set-top boxes (STBs) will allow us to receive, store and manipulate large amounts of video and audio information. Those boxes will have a hard drive and editing capabilities. Because digital video is very resource consuming (an uncompressed 2 hour video can take as much as 216 GB of memory), there is a strong need for encoding capabilities in the box. MPEG, the current standard approach, is very expensive for consumer set-top box encoding applications and does not allow for easy editing. Having been explicitly designed with a broadcasting model in mind, MPEG is asymmetric, with the encoding significantly more complex and costly than decoding.

Having in mind low cost, easy to time-shift and edit, soft set-top box applications, we have designed a wavelet-based technology WZD (Wavelet Z-coDec) for quality compression of full size (D1) video. Our approach uses wavelet transformations (as opposed to the discrete cosine transform DCT in MPEG) in both the spatial and temporal direction. Thus we avoid the complex and expensive process of motion estimation and motion compensation. The overall algorithm is 15 to 20 times less complex than the MPEG approach and achieves similar or better performance for all important bitrates.

WZD is a symmetric codec with very low Mips overhead and low latency encoding (< 6 frames). We require very low power operation and allow cheap color conversion and direct compression of the composite NTSC (our compression technology also decodes chroma from the carrier – see www.interval.com/wzd).

The main implementation benefits are: simplified processing, reduced hardware & software overhead and saved memory and memory bandwidth. They result from our wavelet compression, multiplierless coefficients, block by block processing and intermediate compression. Our solution is ideally suited for real-time processing of analog video sources for a variety of embedded applications - set-top boxes, PVRs (personal video recorders), home networks, cameras and camcorders, security applications.

We will present the WZD algorithm in the next section, followed by a section on the MAP-1000 processor. Experimental results will be shown next, and finally we will conclude with a summary.

2. WAVELET Z-CODEC

An image transform codec (compression/ decompression algorithm) generally consists of three basic steps in the forward (compression) direction:

1) forward transform,
2) quantization,
3) entropy coding.

The decompression direction consists of an entropy decoding stage followed by the reverse transform. The forward and inverse transforms are exact inverses of each other and are thus lossless. They are often linear, and are performed to decorrelate pixel values so that the resulting coefficients can be better compressed. The quantization stage is lossy, where less important visual information is discarded. The entropy coding stage is again lossless, where the quantized coefficients are encoded into compressed bitstreams. A lossless compression is achieved by omitting the quantization stage.

The Wavelet Z-codec follows these same basic steps, using a three-dimensional wavelet transform in the transform stage. Optional intermediate quantization and coding stages may be added to greatly reduce RAM. The basic process flow in the forward direction is:

1) forward transform:
   - 1-a) spatial transform, with TS (2-6) wavelet filters,
   - 1-b) intermediate quantization (optional),
   - 1-c) intermediate entropy encode (optional),
   - 1-d) intermediate decode (optional),
   - 1-e) temporal transform, with Haar-Haar wavelet filters,
2) final quantization,
3) final entropy encode.

Steps 1-b, c and d are optional, depending on implementation platform. If used, the memory required is significantly reduced and the entire coder including memory can be implemented on a fraction of a chip.

As is typical of DCT transform codecs, the processing steps are not applied to entire video frames. Instead, each video field is divided into blocks, where each block is processed separately. The definition of a block as well as details of each of the non-optional processing steps are described in the sections below.

2.1 Processing 'Units' – GOPs, Stripes, Blocks, Sticks and Stones

A WZD video stream is broken into groups of pictures (*GOP*) of a fixed number of frames (two), each consisting of two (2) fields. Each GOP is processed independently of each other GOP. We have found that for WZD this GOP length makes a suitable tradeoff between picture quality, bit rate, and complexity. In addition, this short GOP length is well suited for editing and searching. At any point in the compressed stream you are within one of the desired frame.

By contrast, the GOP for MPEG codecs is typically fifteen (15) frames. However different GOP lengths and structures can be used depending on scene content. In any case, motion estimation and compensation are performed independently on each GOP. Even such a simple editing operation as "cutting" from one stream to another is a complex operation for an MPEG stream.

For WZD, each field in a GOP is further broken into 8-line *stripes*, with each stripe processed independently of each other stripe.

Each stripe is then broken into 8-line by 32-column blocks of pixels, each block processed independently of each other block.

During the temporal transform four corresponding stripes, one from each field of the GOP, are processed together as a *stick*. Within each stick, four corresponding blocks, one from each field of the GOP, are processed together as a *stone*. These processing units are illustrated in Figure 1 and

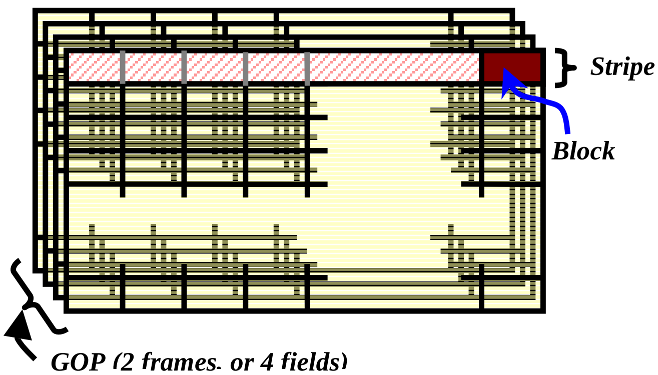


Figure 1 – A GOP, a Stripe and a Block

The main advantage of using a small processing unit of an 8 x 32 block, or an 8 x 32 x 4 stone, is that it reduces the memory / data cache size and bandwidth requirements. It also reduces occurrences of costly data cache misses in soft implementations and reduces silicon area requirements in hardware implementations. Block processing does present the potential problem of blocking artifacts due to transform discontinuities across block boundaries. This problem is resolved in WZD by the incorporation of edge filters, as described in the next section.

2.2 Processing Step 1-a, Spatial Transform

For the spatial transform, the WZD algorithm uses the TS (Two-Six, or 2-6) wavelet filters, which are quadratically lifted Haar wavelet. The basic equations that transform an input sequence x_i , $0 \leq i < 2N$, into half-length low-pass and high-pass sequences f_i and h_i respectively, $0 \leq i < N$, are:

2-6 Forward Transform:

$$f_i = x_{2i} + x_{2i+1} \qquad (1)$$

$$g_i = x_{2i} - x_{2i+1} \qquad (2)$$

$$h_i = g_i - \text{floor}\{(f_{i-1} - f_{i+1}) / 8\} \qquad (3)$$

2-6 Inverse Transform:

$$g_i = h_i + \text{floor}\{(f_{i-1} - f_{i+1}) / 8\} \qquad (4)$$

$$x_{2i} = (f_i + g_i) / 2 \qquad (5)$$

$$x_{2i+1} = (f_i - g_i) / 2 \qquad (6)$$

Equations 1, 2 (and 5, 6 in the inverse direction) form the simple Haar transform, and Equations 3 and 4 perform the "lifting" method [3], which is in essence a quadratic interpolation.

Equations 3 and 4 cannot be used at the two edges of the sequence and different filters are required. To minimize blocking artifacts, we have designed special edge filters that ensure polynomial continuity across block boundaries. These filters effectively result in constant, linear or quadratic interpolation for input sequences of length 2 ($N = 1$), 4 ($N = 2$) or ≥ 6 ($N \geq 3$) respectively. Quadratic interpolants typically account for 98% of the variation at a point.

The equations for the case $N = 1$ (replacing Equations 3 and 4) are, for $i = 0$:

$$\textit{Forward:}\ h_0 = g_0\ . \qquad (3b)$$

$$\textit{Inverse:}\ g_0 = h_0\ . \qquad (4b)$$

For $N = 2$, the equations are, for $i = 0, 1$:

$$\textit{Forward:}\ h_i = g_i - \text{floor}\{ (f_0 - f_1) / 4\}\ .$$

$$\textit{Inverse:}\ g_i = h_i + \text{floor}\{ (f_0 - f_1) / 4\}.$$

For $N \geq 3$, the equations are:

Left Edge (i = 0):

$$h_0 = g_0 - \text{floor}\{ (3f_0 - 4f_1 + f_2) / 8 \}$$

$$g_0 = h_0 + \text{floor}\{ (3f_0 - 4f_1 + f_2) / 8 \}\ .$$

Right Edge (i = N-1):

$$h_{N-1} = g_{N-1} - \text{floor}\{ (f_{N-3} - 4f_{N-2} + 3f_{N-1}) / 8 \}$$

$$g_{N-1} = h_{N-1} + \text{floor}\{ (f_{N-3} - 4f_{N-2} + 3f_{N-1}) / 8 \}\ .$$

Note that all filters above map integers to integers and are reversible. They are linear except for the **floor** operations. They are also short, and use dyadic rational coefficients (i.e., they are integers divided by powers of two) with small numerators. Thus, the entire WZD transform can be implemented with only adds and fixed shifts; no multiplies or variable shifts are required.

Note also that all filters above are one-dimensional, but they can be applied to the horizontal or vertical direction. In the horizontal direction each application partitions its input sequence into *left* (L) and *right* (R) '*subbands*', where the left subband holds the low-pass (or average) result sequence $\boldsymbol{f}$ and the right subband holds the high-pass (or vertical edge details) result sequence $\boldsymbol{h}$. In the vertical direction each application partitions its input sequence into *top* (T) and *bottom* (B) *subbands*, where the top subband holds the low-pass (again, average) results $\boldsymbol{f}$ and the bottom subband holds the high-pass (or horizontal edge details) results $\boldsymbol{h}$. For each 32-column by 8-line block of input image pixel luma component data, the wavelet filters are applied in both horizontal and vertical directions several times successively to form the final wavelet pyramid in Figure 2.

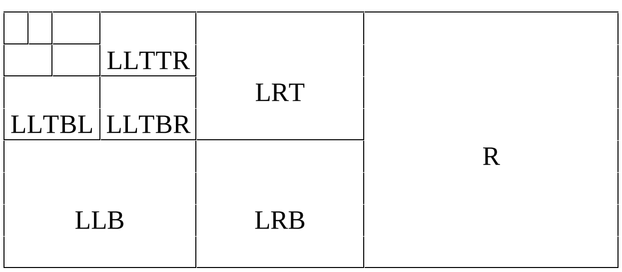


Figure 2 – Wavelet pyramid for luma;
0 = LLTTLLTL, 1 = LLTTLLTR, 2 = LLTTLRT, 3 = LLTTLLB, 4 = LLTTLRB

Each block of the chroma components in component video is half as wide as luma (16 columns by 8 lines) because the data is sub-sampled in the horizontal direction. Thus, one fewer horizontal transform is performed. Figure 3 shows the resulting wavelet pyramid.

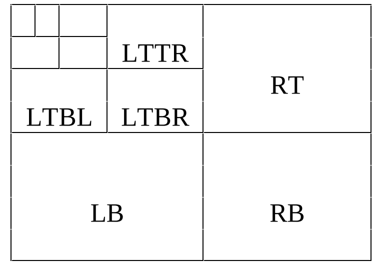


Figure 3 – Wavelet Pyramid for chroma;
0 = LTTLLTL, 1 = LTTLLTR, 2 = LTTLRT, 3 = LTTLLB, 4 = LTTLRB

Note that at the end of the spatial transform, subband 0 contains the average value of all pixels in the block. Thus, subband 0 from all blocks form a 'thumbnail' image of the original. Other subbands contain more edge sharpness information, and we expect their coefficients to be small or zero except at very sharp edges. In fact, due to the smoothness nature of typical video data, we expect the R, RT, LB and RB subbands to be mostly zero. Furthermore, many small values in the LRT, LLB and LRB subbands can be quantized to zero without much visual loss.

2.3 Processing Step 1-e, Temporal Transform

A Haar-Haar wavelet transform is used in the temporal domain for increased compression. The equations used, i.e., the Haar filters, are shown below. They are simply Equations 1, 2 and 3b in the forward direction, and Equations 4b, 5 and 6 in the inverse direction.

Forward Temporal Transform:

$$\boldsymbol{f}_i = \boldsymbol{x}_{2i} + \boldsymbol{x}_{2i+1} \qquad (1)$$

$$\boldsymbol{h}_i = \boldsymbol{g}_i = \boldsymbol{x}_{2i} - \boldsymbol{x}_{2i+1} \qquad (2), (3)$$

Inverse Temporal Transform:

$$\boldsymbol{x}_{2i} = (\boldsymbol{f}_i + \boldsymbol{h}_i) / 2 \qquad (4b), (5)$$

$$\boldsymbol{x}_{2i+1} = (\boldsymbol{f}_i - \boldsymbol{h}_i) / 2\,. \qquad (4b), (6)$$

The filters are applied to every pixel in each stone, taking as input the four values from the four fields at the same pixel location, as illustrated in Figure 4.

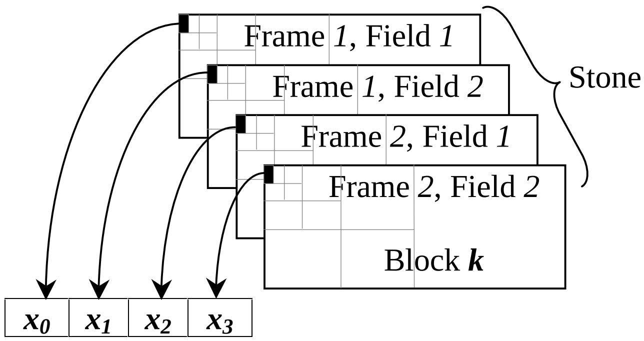


Figure 4 - Temporal transform input sequence construction example

For each pixel, i.e., for each input sequence, the filters are applied twice (hence Haar-Haar), yielding the subbands *Slow-Slow*, *Slow-Fast*, *Fast-Slow* and *Fast-Fast*. As with the spatial transform filters, the temporal transform filters can be implemented with adds and fixed shifts only.

The temporal transform is much lower in complexity than the motion estimation and motion compensation done in MPEG. Furthermore, it operates on a GOP of two frames, as compared to a typical GOP of fifteen frames for MPEG, thus making it easier to navigate compressed video streams, facilitating video manipulations such as editing and searching. The short GOP does imply that when there is significant motion in a video sequence such as a sports footage, we expect our WZD algorithm to perform worse than MPEG. However, for most sequences including movies, WZD yields results comparable to MPEG.

2.4 Processing Step 2, Quantization

The WZD quantization strategy is based on the well-known idea of discarding information that the human visual system (HVS) is unable to perceive under expected viewing conditions. This imperceptible (and unusable) information primarily consists of contrast resolution at high spatial frequencies. Such information is imperceptible due to optical losses (diffraction and chromatic aberration) in the HVS.

Quantization is traditionally performed by dividing a coefficient by a value Q, Q dependent on the spatial frequency of the coefficient. WZD simplifies this operation by restricting Q to be a power of two ($Q = 2^q$) so that quantization is implemented simply by means of discarding the rightmost bits of those coefficients, i.e., shifting the coefficients by q bits.

Processing step 1-b, if performed, would be done in a similar manner so that only shifting would be used rather than division.

2.5 Processing Step 3, Entropy Coding

The wavelet pyramid, after being quantized, will have substantial runs of zeros as well as substantial runs of non-zeros, which makes it natural to code the coefficients into two streams, one being a stream of significance bits, and the other being a stream of non-zero coefficients. There is one significance-bit for each coefficient, a one if the quantized coefficient is non-zero and a zero if the quantized coefficient is zero. In addition to having substantial runs of zeros, the significance bits are also highly correlated, where each significance bit is most likely the same as the preceding significance bit.

Arithmetic coders are known to compress streams such as the significance stream very well. We use a modification of the low complexity, binary arithmetic coder known as the Z-coder. We have found that, for video, the Z-coder calculations can be done with ***8*** bits of precision and that only ***3*** bits of context (***3*** previous significance bits) give a good prediction of the next bit. These three bits of context imply a 2^3 entry table of 8-bit increments. Details of the Z-coder can be found in [2].

The non-zero coefficients are encoded using a Huffman coder. For each coefficient, the codeword is obtained for its absolute value and merged into the code stream, followed by the sign bit. The Huffman coder has been thoroughly studied in academics and industry, and will not be described here. For details, on the Huffman and arithmetic coders refer to [6].

As with the transform filters, we designed the Z-coder and the Huffman coder to require only adds and shifts to maintain low-complexity in implementation.

3. THE MAP 1000 PROCESSOR AND ARCHITECTURE

Equator Technologies' MAP (Media Accelerated Processor) 1000 is a low-cost programmable processor designed to handle the large computational demands of digital media processing, including image and video codec, 2D and 3D graphics, as well as digital audio codec, synthesis and spatialization. It provides a flexible, high-performance environment for developing real-time multi-media applications, many of which have rapidly evolving functional requirements due to the recent growing need for multi-function digital TV and STBs.

3.1 The Map Architecture

The MAP 1000 has a high computation rate because it is based on an on-chip parallelism technique known as Very Long Instruction Word (VLIW) architecture. VLIW is one of two well-established schools of thoughts on how a high degree of instruction-level parallelism can be achieved, the other being the super-scalar architecture. Today's popular microprocessors such as the Pentium II are based on the super-scalar approach, which require special on-chip hardware to look through the instruction stream and find independent operations which can be executed simultaneously by multiple execution units available so that parallelism can be maximized. This special hardware typically takes up a significant portion of the chip's die area. Each instruction in superscalar processors codes for only one single operation, and the grouping and scheduling of instructions for execution is done at execution time in hardware.

On the other hand, for VLIW processors the difficult task of finding parallelism is moved to the compiler. The compiler searches for eligible operations, checks for data dependencies, controls resource conflicts, and packages these operations into VLIW instruction words. Thus, each instruction word explicitly describes the parallelism by specifying which operation is to be performed by each execution unit during each cycle. Because VLIW processors do not require special on-chip hardware, they can be much cheaper than superscalar processors, but they require sophisticated compilers to perform the instruction-level parallelism properly.

3.2 The MAP compiler

The Equator MAP compiler is critical in ensuring that the computational power and architectural features of the MAP 1000 are fully utilized. It supports a wide range of optimizations, including software pipelining, preconditioning, as well as trace scheduling which can search a whole routine for eligible operations. It can also explore across natural code boundaries such as branches for opportunities of increased parallelism far beyond the limited search window seen in existing super-scalar architectures.

The MAP core processor is programmable in C, and the MAP C compiler is generally compliant with ANSI standard C. It supports a media intrinsic API with a large suite of 32-bit and 64-bit integer and floating-point scalar operations, partitioned operations with 64-bit registers, plus operations with 128-bit partitioned long constant/variable that facilitate vector arithmetic.

With extensive use of partitioned media intrinsic, the implementation of the WZD wavelet transforms on the MAP core processor requires only three (3) cycles per coefficient.

3.3 MAP Coprocessors

The MAP-1000 has several functional units in addition to the VLIW media core processor. The Data Streamer intelligently handles inter- and intra-chip data transfers as well as hides the penalty due to data cache misses. It can be programmed to transfer data of various shapes and forms. The Data Streamer executes commands in a separate thread of control with minimal CPU support. Thus, conventional double-buffering schemes can be used to fully overlap data movement and VLIW core or Variable Length Encoder/Decoder (VLx) processing.

Another functional unit is the Fixed Function Block, which includes graphics accelerators, a video scalar and the VLx. The VLx is actually programmable, currently in assembly. It is a RISC coprocessor which can be used to perform bit-sequential tasks that are not well suited for the VLIW core. The VLx contains special-purpose hardware, called the GetBits engine, for bitstream processing and hardware accelerated MPEG2 table lookup. More details on the MAP platform can be found in [1] as well as at ETI's web page http://www.equator.com .

4. WZD on MAP

For the implementation of the forward direction of the WZD algorithm on the MAP-1000, the Data Streamer is programmed to transfer pixel values from frame buffers to contiguous memory locations on the on-chip SDRAM one stone at a time. Double-buffering is done so data from the next stone can be transferred in while the current stone is processed.

Data in the frame buffer is in field-interlaced YuYv 4:2:2 format, thus the Y (luma), U (chroma 1) and V (chroma 2) components are first extracted out into separate buffers in the core. The wavelet transforms for each component are then performed for the stone, with resulting wavelet coefficients placed into contiguous memory locations in SDRAM. Again double-buffering of the coefficients buffer is done so that the coefficients from the current stone can be transferred to the VLx while those from the next stone are calculated.

The Data Streamer transfers the wavelet coefficients from the SDRAM, stored in its natural order of scan-order, into memory accessible by the VLx called CM1 in subband-order. Coefficients grouped in subband-order are expected to have longer runs of zeros or non-zeros than in scan-order and are thus expected to yield a smaller Z-coder codestream.

The final quantization and final coding stages are implemented in the VLx core. The data streamer and the GetBits engine are used to transfer the resulting codestream bits back to SDRAM. The core then writes the codestreams out to file.

The inverse direction of the WZD is implemented in a similar manner.

5. EXPERIMENT RESULTS

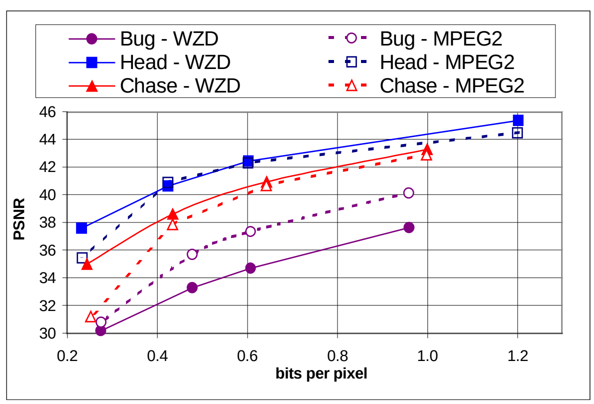


Figure 5 - Rate-Distortion curve comparison of WZD software implementation vs. MPEG2

Figure 5 plots the quantitative results of a software implementation of the WZD codec, compared to those of a standard software implementation of MPEG2, Test Module 5 (TM5). 3 footages were processed, each 300 frames long. "Bug" is a footage from an animated movie, "Head" is a talking-head footage from a motion movie and "Chase" is a motorcycle chase footage from the same movie.

The curves are known as rate-distortion curves. Here the distortion is represented by the Peak Signal to Noise Ratio (PSNR) in dB, actually the opposite of distortion, and the rate is represented by the bit rate of the compressed bitsteam in bits per pixel (bpp). For full-size frames, 1 bpp corresponds to roughly 10 Mbps (million bits per second). Although PSNR may not always be an accurate indication of the visual quality of a moving sequence, typically the higher the PSNR, the better the quality. Figure 5 shows that WZD yielded better quality than MPEG2 for two of the three sequences. Furthermore, WZD quality degrades smoothly as the bitrate is reduced, where as MPEG2 quality drops off sharply below 0.3 bpp.

Figures 6 and 7 show sample results of the WZD algorithm implemented on the Equator MAP-1000, intermixed with results of TM5, the standard software MPEG2 implementation. The image quality can be seen to degrade as the bitrate is reduced, and WZD results are comparable to TM5 results at most bitrates. At low bitrates WZD results are generally better. For example, in Figure 7, WZD at 1.8 Mb/s is visibly better than TM5 at similar rate.

6. SUMMARY

In this paper we have described a soft real-time implementation of the Wavelet Z-coDec or WZD, a very-low complexity wavelet-based video codec, on the MAP-1000. WZD was designed specifically with low cost, ease of compressed video stream manipulation and soft set-top box applications in mind. The MAP-1000, with its powerful VLIW architecture, compiler and supporting coprocessors, enables this soft implementation of the WZD to be in real-time. Combining the WZD with MAP-1000 allows timely response to rapid changes in market demand and industry standards.

WZD is far less complex than MPEG-2, and experiment results show that it yields comparable quality to MPEG-2 in general.

In addition to the implementation of the WZD technology on the Equator MAP-1000, we have also built a software implementation on the Pentium PC as well as another popular media processor – the Philips TriMedia. Some of the advantages of our approach come to life in a hardware implementation, i.e. an ASIC or part of an ASIC (system-on-a-chip). Because the entire algorithm can be implemented with shifts and adds only, we anticipate to be able to design a full encoder with 50 – 60,000 gates, significantly less than MPEG-2 encoders with comparable quality. For more details on such implementations, please refer to [4] and [5], which together with several other papers and descriptions can be found at http://www.interval.com/wzd.

Acknowledgements

Many thanks to Equator Technologies, Inc., which has provided wonderful technical support, and to Bill Arrighi, who has made important contributions towards implementing the WZD on the MAP-1000 as well.

References

[1] C. Basoglu, K. Zhao, K. Kojima and A. Kawaguchi, The MAP-CA VLIW-based Media Processor from Equator Technologies inc. and Hitachi Ltd., *White Paper ETI*, January 2000.

[2] L. Bottou, P. G. Howard, and Y. Bengio, The Z-Coder Adaptive Coder, *Proceedings of the Data Compression Conference, pp. 13-22,* Snowbird, Utah, March 1998.

[3] K. Kolarov and W. Lynch, Compression of Functions Defined on Surfaces of 3D Objects, *In J. Storer and M. Cohn, editors, Proc. Of Data Compression Conference, IEEE Computer Society Press*, 1997.

[4] K. Kolarov and W. Lynch, "Very Low Cost Video Wavelet Codec," *SPIE Conference on Applications of Digital Image Processing*, Vol. 3808, Denver, July 1999.

[5] W. Lynch, K. Kolarov and B. Arrighi, "Low Cost Video Compression Using Fast, Modified Z-coding of Wavelet Pyramids," *Proc. of the International Conference on Image Processing ICIP'99*, Kobe, October 1999.

[6] I. Whitten, R. Neal, and J. Cleary, Arithmetic Coding for Data Compression. *Communications of the ACM 30, 6 (June) 1987, pp. 520-541*

(a) Original

(b) MPEG2 @ 6 Mb/s

(c) WZD @ 4.5 Mb/s

(d) MPEG2 @ 3.9 Mb/s

(e) WZD @ 2.6 Mb/s

(f) MPEG2 @ 2 Mb/s

Figure 6 – Processing results of MPEG2 and WZD codec algorithms applied to a motorcycle chase sequence. (a) through (f) are ordered in decreasing bitrate.


(a) Original


(b) MPEG2 @ 3.7 Mb/s


(c) WZD @ 3.1 Mb/s


(d) MPEG2 @ 1.8 Mb/s


(e) WZD @ 1.8 Mb/s


(f) WZD @ 1.2 Mb/s

Figure 7 – Processing results of MPEG2 and WZD codec algorithms applied to a talking head sequence. (a) through (f) are ordered in decreasing bitrate.

Low-Cost Mass-Storage
"The Last Piece of the Home Server Puzzle"

Mitch Webster
Media4u a Mindport Company

Abstract

Future business models for interactive television and media-based e-commerce in the broadband environment are spawning the need for a new type of media enterprise: the Media Service Provider (MSP), a combination of an ISP and traditional broadcasting company. Since subscriber delivery of interactive, instantly accessible video and audio media is the key revenue driver in this concept, there will be need for low-cost, reliable mass data storage in the home.

This paper surveys the next generation of multimedia home storage and suggests how the home media server might merge with the traditional television set-top box, the current entertainment "gateway" to most living rooms. The future model is neither a fat nor thin client, but an entirely new model: a storage client.

Introduction

Recent trends indicate that interactive digital television, driven by the economic engine of media-based e-commerce, will offer compelling new business opportunities in the broadband era. This merging of the traditional Internet web site with entertainment quality audio and video is a strong candidate to become the "killer app" of media distribution in the next decade.

With this new media distribution model will come the need to store vast amounts of quickly accessible multimedia content in the home. The ability to provide consumers with instant access to large media files and video exceeds the capabilities of the current generation of conventional personal computers. Even as the capacity of the PC expands, its primary role as a general purpose, highly flexible-computing device makes it inappropriate to function simultaneously in the secondary role as a media server. A new type of device is required to meet the needs of home entertainment storage and distribution.

Since the television set-top box is now the primary control terminal for entertainment services to most homes, the functions of a home media server comfortably integrate -- perhaps fully merge -- with the set-top box to create a new, far more powerful home entertainment hub. In addition to accessing and serving traditional television and audio entertainment programming, this device would handle a wide range of new of interactive media services and could connect to a broad universe of home appliances. In some instances -- remote areas, for example -- the home server could act as a cache for a large number of Internet web pages, removing the "wait" from the WorldWide Wait.

Enabling Technologies: An Overview

Several technologies are "coming of age" that will enable the home server and its associated network of appliances to come to market at an affordable price. These technologies include:

a) RF/Wired home communications (Home RF, Blue Tooth, Home PNA (using existing home phone wiring for networking) to seamlessly link together components in a home network.

b) High-performance, low-cost microprocessors (300 MIPS for <US$10) dramatically boosts computing power at increasingly lower cost.

c) Low-cost mass storage (> 17 GBytes, < US$80) is here. Rapidly falling prices will soon allow 100 gigabytes of storage at a cost of less $1 per GB.

d) Digital broadband transmission technology, via cable, DSL and satellite, is being rapidly deployed throughout the United States.

While not precluded from functioning as a home server, the generic personal computer is a less likely candidate due to the following reasons:

a) *Reliability and Security*. Multipurpose devices are also inherently less reliable and more prone to crash than dedicated, application-specific appliances.

b) *Availability*. PCs employed with traditional tasks may not be available for entertainment functions at desired times.

b) *Cost.* General-purpose computers cost more than dedicated computing appliances.

c) *Control*. The PC storage device is controlled by the end user, not the media service provider. Storage may not always be available for service provider downloads, thus a guaranteed quality of service is not assured.

d) *Security*. A generic PC is inherently less secure than special purpose server device.

e) *Complexity*. A general purpose PC is more complex than an appliance designed for a specific application.

The key issues in development of the home entertainment server involve:

a) Protection of high-value content.

b) Provision of media services to the home.

c) Distributed data management.

d) Content rights management.

Applications

The home entertainment server concept allows cable, satellite and broadcast operators to become full service media providers through a multi-function entertainment gateway into the subscribers' home. In addition to real time services, the technology offers a new business model that allows all kinds of content to be downloaded to subscribers during the hours of off-peak bandwidth. Because the hub can connect to range of information devices, the service possibilities extend far beyond traditional audio/visual programming and web services.

Through USB, Ethernet and IEEE 1394 "Firewire" ports, the media server hub connects not just to television sets, but with personal computers, digital cameras, printers, cell phones, Palm-sized computing devices, e-Book readers, CD burners, MP3 music players, and telephone equipment. A vast range of media types, from music files to electronic books and newspapers, can be delivered to audiences during the low traffic night time hours.

Promising new business opportunities for media service providers include true video-on-demand with timed viewing, music-on-demand (home jukebox), enhanced personal video recording (based on viewer preferences), games, online banking, subscription data and reference services (online phone book, education material, etc.), media-enriched home shopping, digital photography, multimedia e-mail, video conferencing and multi-featured telephony. In the future, new applications for the network will appear that we don't yet even imagine.

The Internet web surfing experience can also be significantly enhanced when coupled with a home media server. When tied to the server, several home PCs or Internet appliances can simultaneously access locally cached web pages and associated multimedia files that were previously downloaded to the home storage device. This arrangement allows significantly faster access to web content and lessens the delay that users encounter even on the fastest broadband systems. The connection of home information devices to the server is enabled via the home network, either a through RF or Home PNA.

Specially formatted television-centric supplementary web content (HTML, Shockwave, etc.) could also be accessed instantly from the home server while viewing programs. The depth of interactivity between web content and television programming is dramatically increased when a local server is used to store data. For example, the night before the Super Bowl the MSP can download hundreds of Super Bowl-specific web pages that could be accessed from the server during the game. When the game ends, the pages can be automatically deleted from the storage device and replaced with updated material for future programs.

Typical Application Storage Requirements

Application	**#**	**GBytes**	**Total**
VOD (@4 Mb/s, 1.5 hour) [i]	5 titles / day	3	15
MOD (@ 150kb/s, 3 minutes)	100 titles	0.003	3
WWW hosting (100 Kbytes / page)	10k pages	0.00001	1
PVR (8 hours, 4 Mb/s)	8 hours	2	15
Infomercials (3 minutes, 4 Mb/s)	20 slots	0.09	1.8
Subscription Data Services (?)	5	0.2	1
		TOTAL	36.8

The Customer as Programmer

The home media server is a powerful successor to the Personal Video Recorder (PVR), a hard disk-enabled device that enables television viewers to take control of their viewing experience. Unlike the current generation of Personal Video Recorder, the home media server is not a consumer appliance, but an extension of the media service provider into the customer's home. The server -- just as today's set-top box -- is maintained and controlled by the MSP, yet operated locally by subscribers in the home. Empowered by the server, the customer becomes the programmer, enjoying instant access to a wide range of highly personalized media from a vast array of sources.

Neither Fat nor Thin, but a Storage Client Model

A new computing model is created – neither fat nor thin – but rather a storage client. The personal computer, dependent on increasing amounts of processing power, is considered a fat client model. Information appliances, dependent on a network for its power, are categorized as thin client devices. On a storage client, tasks are segregated between the media server and the network operator. Local CPU intensive processing is not needed, yet extensive local manipulation capability is needed locally for handling large video and audio files. The storage client is a logical extension of the "Edge Server" model that is currently used in Internet applications.

New Technologies Required

In order that a home media server system – as described here – can be brought to market, several existing technologies need to be further developed and applied to the specific application. They include:

a) *Profiling*. An area under intense development for related applications, the profiling functions needed in a home media server environment need further refinement. Essential to the success of the media server model is intelligent software that can determine the preferences of the home viewer and determine which household member is watching at any given time. Needed are improvements in learning preferences, demographic identification, and a more sophisticated method of tagging and categorizing incoming content. Since the profiling used in this home-based system stays on the server within the confines of the subscriber's home; the privacy of personal viewer information is fully protected.

Streaming media, now an essential shortcut for delivering multimedia over bandwidth-

b) *Content protection.* The protection of content stored on a server is different from the protection of content delivered in real time. For example, content may reside on the home server a full year before it's used. Much content protection today involves key cycling, but when media is stored long-term on a server, keys can easily lose sync. Server-specific content protection must be implemented.

c) *Intelligent storage management.* Because the home server does not have an infinite storage capacity, there is a need to select and manage the content it stores. As media changes and ages, conflicts can emerge. For example, based on a personal profile, the server's software may face a choice between replacing older, unviewed material with a more current program that might be of interest of the recipient. The system must be intelligent enough to make the correct choice.

d) *Home Scheduling.* Depending on the household members watching at a given time -- male, female or both -- the system needs to determine preferences for the viewers from its stored database. It's here the local server makes decisions previously reserved for the network programmer. The software must engage in dynamic decision-making based on what it knows about the viewers.

The New Entertainment Hub

The media server will co-exist with the home PC. The PC will remain the dominant "lean forward" information device for tasks, while we see the media server as the dominant "lean back" device for entertainment uses. New devices, such as Internet radios, telephone and other networked appliances will emerge over time and complement the server.

starved networks to the PC will find new applications with the home media server. Media may be streamed from the server to Internet appliances with low bandwidth connections, and could overcome the limitations of home networks. For example, the MSP might download 5,000 song titles to the server overnight. Using audio streaming, anyone in the home could access any single title on a wide choice of listening devices. The role of streaming media to relieve bandwidth on the Internet shifts to relieving bandwidth on the home network.

Proprietary Networks Must Adopt Architectures Complimentary to the Home Server

Today, proprietary networks dominate home entertainment delivery. Cable and satellite operators each use incompatible closed systems. The home server model can work only if these networks are adapted to accept and control it. Such a full-scale integration of technologies is, of course, no small task. Only recently, and after years of difficult negotiations, has the television industry made progress on the seemingly simple issue of interconnecting home electronics components.

As we consider the benefits of the home server model, a key paradigm shift must occur: control of home media storage must shift from the consumer to the media service provider. Consumer ownership of media storage has a long history, dating back to the early days of audio and video recording. With today's consumer electronics model for PVR products, the owner of the product controls the hard disk storage on the recording device. However, for the service provider to insure the needed data storage space and maintain a high quality of service, consumers must forfeit primary control of the storage device, much as with today's television set-top boxes. This will require a creative marketing effort designed to demonstrate the clear benefits of shifting this task to a service to an outside vendor.

Conclusion

The engine of the home media server concept is entertainment. Interactivity, personalization and massive storage, combined with a high-speed gateway to a vast array of on-demand entertainment services, offers significant new business opportunities to pay media broadcasters and huge benefits to consumers. This new model, the client server, deserves serious consideration as a cost-effective method for efficiently delivering interactive multimedia content to consumers over broadband networks.

---

[i] This will depend on the compression method used. The figures here assume a worst case of real-time, MPEG2 encoding at 4 Mbits/s.

Monitoring Client Experience

Bruce F. Bahlmann
MediaOne

Abstract

The customer demand for quality Internet access is prompting a change in the way Internet information services (or high-speed Internet Service – HSD) will be engineered, maintained, and marketed in the future. As a result, traditional measurement applications of Internet service will give rise to more sophisticated applications which focus on customer experience and quality.

Best Effort Quality

Today's Internet service is not very sophisticated when it comes to quality control. Almost magically, from a customer's perspective, Internet service continues to work and "seems" fairly reliable. Behind the scenes it's a constant flurry of activity where projects to upgrade circuits, equipment, and software abound. This constant change creates many challenges operationally to achieve standardization and yet get the most out of the equipment in service. As a result, quality control & maximizing return on investment are in constant conflict with the deployment of new services and stabilization of the network build out.

Providers of Internet service in this space reside in a very competitive market that requires them to constantly seek new affiliates. They use terms like "redundancy" and "high availability" to lure affiliates into signing service contracts with them. However, the language of many of these service contracts all but define the level of service or overall quality of service individual customers can expect. The service contract merely binds the affiliate to the Internet provider in exchange for Internet provider making its best effort at providing a quality Internet experience for the affiliates' customers. Sometimes the service agreement is entirely void of service quality related measurables. In this case, an additional document called a service level agreement (or SLA) is necessary to define the level of service the affiliate can expect for signing on with a particular Internet provider. Ideally, this document represents a compromise between the type of service the Internet provider is willing to supply and what the affiliate is willing to accept. For reasons that will be explained later, it is often difficult (if not impossible) to reach a compromise on the SLA. As a result many SLAs go unsigned – which means the Internet provider is under no agreement to provide a particular level/quality of service to the affiliate.

Today's Quality "Guarantee"

In cases where a service level agreement is signed the Internet provider is "bound" to the level of service expected in return by the document as well as the affiliate. Within a typical SLA are several points of interest. Notably, the "Key Performance Indicators" and the "Network Services Conformance" sections provide the operational parameters that the Internet provider has committed to supplying. Key performance indicators are focused on response to outages or escalations where network services conformance is concerned with availability. The rest of this document will focus on the subject of availability.

Availability

One of the commonly used terms with regard to providing Internet service is "Availability". Availability is defined as capable of being obtained and/or accessible for use. Internet providers use the word availability to signify the amount of reliability they intend to provide with respect to various services they supply. Availability is typically defined in terms of percent (%) with higher percents equating to higher reliability.

The availability projections within the SLA are usually based on the Internet provider's "best effort" to measure the accessibility of the services they provide. One of most common tools in use today to measure availability is ping. The ping application communicates with Internet hosts to determine their operational status. For example if a host is operational (or "up") it is reported as "alive" by the ping application. If the host is not operational (or "down"), it reports "no response" or "request time out" by the ping application. Although the ping application is a useful operational tool on the Internet, it is not a very reliable means of measuring availability. For example, the host may be up but the application (or service) supplied by the host could be down. In this case the availability is reported incorrectly. As a result, there is a difference between application availability (measured via the application's client) and host availability (measured via ping).

Unfortunately, availability is not mentioned with respect to service windows used by the Internet provider. A service window is when Internet providers perform necessary installations, changes, and upgrades. The service agreement dictates certain times and days as potential service windows that Internet providers may use to maintain service. These days and times are usually coordinated with affiliates so the affiliate and its customers understand any resulting down time during the window. The ease at which service windows can be scheduled and the fact that service windows rarely count against service reliability affords the Internet provider a means of constantly changing the system it uses to provide Internet service. The resulting constant change along with the lack of affiliate-initiated acceptance and/or safeguards prevents quality control from being achieved.

Surprisingly, the Internet provider often does the only monitoring of availability levels to measure its compliance established in the SLA. The Internet provider supplies this because the affiliate does not always have the means to do this on their own. However, SLA's typically do not stipulate the type of monitoring (application availability or host availability) they require. In absence of any specific request for monitoring method, host availability is likely reported as the default as it's the easiest to obtain. As a result the monitoring data reported by the Internet provider often does not reflect the actual availability seen from a typical affiliates' customer perspective. This allows the Internet provider to maintain compliance with the SLA while providing a level of service and quality of service that is actually lower. Unless the SLA is re-written to dictate the monitoring method used by the Internet provider the SLA will not represent any guarantee for level/quality of service.

Future Quality Guarantees

Fortunately, customers are beginning to sense some differences among providers of Internet service. Although price is still the biggest factor, level and quality of service are moving up fast on the importance scale. The result of customer initiated preference for quality and reliability is changing what the affiliates' need to satisfy their customers. These changes will include such things as incentives and penalties enforced on Internet providers as they strive to meet minimum service levels dictated by the SLA.

Incentives & Penalties

To address this affiliates are realizing that availability of Internet service impacts their bottom line. The impact that availability has on things like call volume, truck rolls, and higher sales is not known at this time. However, the ability to compare availability with call volume, look for trends, and establish some relationships between the two is gaining interest from affiliates. At the time of this writing, it seems reasonable to expect that there is a relationship between call volume and availability. It is projected that further analysis could potentially derive a cost factor per customer that is absorbed by the affiliate as a result of lowered availability. Additionally, the cost calculated could in turn be used to establish minimum availability levels an affiliate will accept. Thus having a tool that could provide affiliates with up to the minute calculations on availability could help them better understand the relationships between availability and support costs and reduce the burden that lower availability has on affiliates' bottom line.

Providing motivations to Internet providers is the key to establishing realistic minimum application service levels. Obtaining the history of an Internet provider's performance, one can establish the average service availability level provided. This average availability level could then be used to drive the affiliate's required service availability levels. Combine this with impact studies above could result in the affiliate providing incentives for the Internet provider to perform above their required service availability – such has a kick back premium per customer. Like-wise, service availability levels below the required levels would result in service discounts per customer (to enable the affiliate to recover the added support costs that were the result of lower availability levels caused by the Internet provider). Providing these kinds of incentives and penalties would allow availability to be treated equally with other methods of evaluating an Internet provider's performance. The reality of the matter is that Internet providers need this level of information to make informed decisions of service upgrades and network build out.

Informed Decisions

Having the application availability information provides affiliates with the means to make informed decisions regarding escalation of calls to the Internet provider's tier two services, scheduling of service calls, and acceptance of system upgrades. In fact, this information could actually drive affiliate requests for specific application performance upgrades in some cases. Making informed decisions is the key to cost savings and reduction in outages caused by unnecessary upgrades.

Informed decisions also enable significant cost savings to Internet providers. By using availability information systems could "truly" be scaled in concert with demand thus eliminating costs of over engineering solutions for under populated areas. This would enable targeting of capital expenditures to areas of need (a type of scratch where it itches approach towards network upgrades) would permit Internet providers with a means of controlling costs and increased operational efficiency.

The ability to make informed decisions would also enable both parties to collect historical information needed for capacity planning.

Historical Data

Historical data enables the most efficient use of resources to solve problems. Establishing such things as “baselines” and “peaks” allows vendors to build products that handle the kind of beating that an Internet service in this space demands. Today vendors cannot fully understand the dynamics of the operational environment to build products that can withstand the punishment of taking on all the Internet can dish out. As a result, new products are forced to burn-in while in production mode rather than in less service-impacting mode. The ability to collect application-based data is the single largest factor impacting the collection of historical data.

Application Monitoring

An application called a client experience monitor (CEM) has proven potential to provide affiliates with the information they need to quantify the level of service they receive from Internet providers and guide future agreements for continued service. A working prototype of the CEM is explained as well as a snapshot of the data that has been collected.

Client Experience Monitor Prototype

The goal of the CEM is to regularly perform “client-like” tasks. The CEM is responsible for storing application response results along-side “traditional” availability tests (pings - which are performed in parallel) – see Performance Figure below for sample data collected. This data will enable separate CEM tools to produce periodic reports to summarize compliance with service level agreement, and produce a client experience rating based on the responsiveness of the applications supplied by the Internet provider

Design Goals & Hypotheses

It is projected that a delta exists between up time (from a client’s perspective) and application availability reported by the Internet provider. The delta will be the result

Performance:

Last Polled: *@09:48:04 Tue Sep 14 1999*

			Ave Ping Response Time				Ave Application Response Time:			
Service:	**IP:**	**Status:**	**Day:**	**Month:**	**Year:**	**Overall:**	**Day:**	**Month:**	**Year:**	**Overall:**
dns1	209.32.160.10	UP	0.078	0.064	0.063	0.063	0.147	0.142	0.139	0.139
dns2	209.32.160.11	UP	0.067	0.059	0.060	0.060	0.145	0.139	0.140	0.140
dns3	24.31.3.8	UP	0.053	0.048	0.048	0.048	0.134	0.123	0.122	0.122
dns4	24.31.3.9	UP	0.057	0.069	0.068	0.068	0.093	0.095	0.095	0.095
news	24.31.3.15	UP	0.082	0.083	0.083	0.083	0.263	0.317	0.316	0.316
ntp1	24.31.3.8	UP	0.070	0.076	0.076	0.076	0.397	0.401	0.401	0.401
pop3	24.128.1.94	UP	0.158	0.162	0.162	0.162	0.862	0.784	0.783	0.783
tftp	24.31.3.8	UP	0.048	0.052	0.051	0.051	0.276	0.284	0.284	0.284
web2	24.31.3.10	UP	0.041	0.036	0.036	0.036	0.141	0.144	0.144	0.144

All response times are listed in seconds...

of degradation in application performance to a point where it is unacceptable to the client (or noticeably impacts its ability to use the service). During these periods of degradation the application availability will remain unchanged when in actuality, the application is “effectively down” from a client’s perspective.

It is also projected that a relationship between call volume and application availability exits. The increase in call volume as a result of a decrease in availability would provide evidence of an additional metric that must be considered with respect to the SLA as its currently absorbed by the affiliate. The prototype actually generated data that allowed availability to be plotted but call volume data was not yet available at the time this document was written.

Additionally, it is projected that during application outages the availability of these applications will fail to depict the actual accessibility of resources provided by the Internet provider due to the resulting increase in load. Instead, the application is “effectively down” much longer from the client’s perspective – see Availability figure below for sample data.

Availability:

	Ping Avaiability				Application Availability:			
Service:	Day:	Month:	Year:	Overall:	Day:	Month:	Year:	Overall:
dns1	100%	100%	100%	100%	99.14%	99.31%	99.33%	99.33%
dns2	100%	100%	100%	100%	100%	99.86%	99.86%	99.86%
dns3	100%	99.99%	99.99%	99.99%	100%	99.92%	99.92%	99.92%
dns4	100%	99.96%	99.97%	99.97%	100%	99.97%	99.98%	99.98%
news	100%	99.99%	99.99%	99.99%	100%	99.94%	99.93%	99.93%
ntp1	100%	100%	100%	100%	100%	100%	100%	100%
pop3	99.83%	99.18%	99.17%	99.17%	97.96%	98.62%	98.62%	98.62%
tftp	100%	99.99%	99.99%	99.99%	100%	100%	100%	100%
web2	100%	100%	100%	100%	100%	100%	100%	100%

The CEM and its data will seek to provide affiliates with a reliable means to monitor the Internet provider’s compliance with the SLA. Monitoring of client experience will strive to eliminate potential bottlenecks or single points of failure to provide the most accurate measurement possible. The CME will also seek to establish a range of “acceptable” client experience ratings. This range is expected to raise the bar on the Internet provider’s application performance to account for quantifiable demands by the affiliate for higher service quality and capacity.

The design of the CEM is based purely on a “proof of concept”. The goal of building the prototype is to demonstrate a working CEM and collect sample data for analysis and hypothesis confirmation. The prototype will also provide direction for follow-on work and serve as an example for future efforts and/or spin-off projects.

Results

After running the prototype for six weeks several data points were realized. Most importantly all hypothesis were confirmed.

1. *Degradation of performance*
   Several instances of performance degradation occurred during the CEM prototype trial. They included DNS round trip times exceeding one second, POP3 connects exceeding 30 seconds, and NEWS requiring more than one minute to download a single article. These minimums where established arbitrarily and not based on actual tests with customers to determine their acceptance.
2. *Increase in call volume during reduced availability*
   Increased call volumes were confirmed by phoning call center during perceived outages. Each time the call center confirmed that call volume had increased during the duration of the perceived outage. No further effort was spent to determine the actual increase.
3. *Difference between application and host availability*
   Any type of host outage rarely accompanied these periods of degradation. Based on the calculations of the prototype tool the "effective outage" was more nearly three times that of any perceived host outage.

Learned Results

Interestingly, ping outages seemed to follow periods of application outages. It was like the host became overwhelmed with the application demands and went down from a ping perspective. Then a short time later came back up however was bombarded by requests from clients thus resulting in another application outage.

The tool also exposed several configuration errors made by the Internet provider. For example, several DNS servers lay idle while one DNS server seemed to be handling a majority of the requests by clients. This could be simply fixed on the Internet provider's DHCP server if only they had detailed performance information on each application and they were attentive at optimizing the use of every component in their system.

Further Study

It is likely that call volume may vary by application and the extent of the outage. Meaning, some applications may cause more calls than others, just as some outages are more extensive than others. Further study is needed to establish specific relationships between various applications and their various outage tendencies before any kind of penalty can be established for such an outage.

Forward

Since the affiliate is ultimately responsible for providing the service (or seen in the eyes of the customer as responsible for sustaining reliable Internet service), it must seek ways to provide the highest quality service possible. One of the best ways to provide reliable service would be to pass along these requirements to their Internet provider. The following suggests some ways to accomplish this:

- Establish some means of confirming the quality and reliability of the service supplied by the Internet provider.

- Establish motivations for the Internet provider to seek the highest availability possible.
- Dictate terms such that the affiliate will conduct its own service level verification and share this with the Internet provider as a means to allow it to maintain the terms of the contract
- Provide customers with access to current status of various applications, scheduled outage windows, etc.
- Provide the data needed to make more informative decisions regarding handling customer trouble calls and coordinating requested upgrades by the Internet provider.
- Strive to negotiate every service window rather than opening the gates for constant change.

Providing reliable Internet service helps the affiliate in the following ways:

- Increased availability (higher reliability) means lower trouble calls and potentially fewer truck rolls. Every call answered that is trouble related is potentially one less sales call answered.
- Increased availability means higher customer confidence in providing Internet service via cable TV lines and thus opens doors for sales in new markets.
- Increased availability also means more satisfied customers which translates into greater demand.

Consideration of client performance as a driving factor for application availability levels has not yet reached the main stream and "quality" features such as availability and reliability play a limited role in today's customer selection of an Internet information service. However, as customer's choices of Internet access become more equal in terms of speed, capability, price, and flexibility, "quality" will be what differentiates one Internet Information service from another.

As the market for Internet service shifts gears to begin focusing on quality, affiliates need to be ready to quantify the service levels they want to provide. Work at home customers will be one of the first to demand the highest possible levels of service and will likely compare various options before buying. Having access to up to the minute service levels will enable marketing to go after these highly demanding customers. Thus there is a need for such a tool or system to drive up service availability levels and empower affiliates continued growth in the future.

MSO Profit Target: The Business Internet User
How Advanced Internet Caching Appliances Can Assist The MSO In Capturing This Business

Jim Royle
CacheFlow Inc.

OVERVIEW

The goal of this paper is to identify a new profit center for the MSO: The Business Internet User. By applying general core competency already existing inside of an MSO operation, it is now possible, through the recent advancement of sophisticated Internet caching systems, to vigorously attack this mature market with what can be the premier last-mile product in the industry.

There are several obstacles in the path of this business decision, however. And, as you will see, only caching can provide a ready-now solution to these issues, and allow for quick entry into this market.

HISTORY

This paper will not go into detail about the history of the Internet, only to say that a majority number of business users still use the same archaic type of circuit they first installed five years ago. Traffic surges have dictated a major network upgrade.

The birth of the Internet happened in the early 70's with the early users being the academia and government employees. By the arrival of the 80's and the birth of the PC, all elements were starting to fit into place to allow this new communication medium to explode. This did not happen until the early-90's, when fiber networks were installed across the country by new Inter-Exchange Carriers (IXC's). The fiber networks provided not only a pure digital signal, but also a cost-effective transport medium to allow these new business players (ISP's) to get into the game. And when they did join the game, they started the process of building these massive router-to-router-to-router based networks, spanning the country with a myriad of connections.

Peering relationships between the larger carrier based ISP's started to take place in the mid-90's. Not only did this function create faster throughput for the end user, (peering helped deliver the user directly to the network that owned the origin server), but it also saved the carrier significant access fees to the common Internet cloud.

These peering points are the key to how today's Internet really works. The vast majority of transactions today connect over a series of router-based carrier networks that are tied together at various private peering points around the country. The original Internet network still exists, but is primarily used to resolve the delivery process of any domain address that is not currently housed with one of these a major players.

This transport process is *critical* to the MSO in its plans to attack the business market, as these router-based carrier networks are now congested to the point of failure, with no quick solution in sight from a transport perspective. In today's model, the average browser-to-origin server journey for any common www.xxx.com request creates an average of 12-15 routers hops each way. If one closely examined the congestion possibilities on each one of these hops that occur during the millions of connections and handshakes that are required to pass along internet data, they would soon realize why things on the net seem so slow.

So, the issue at hand is how to address latency in the Internet, not how much bandwidth the user has between his equipment room and the ISP's POP.

This solution plays right into the High-Speed Data (HSD) cable modem feature set.

Latency becomes even a bigger issue when these legacy-type networks attempt to offer any new and exciting applications, such as streaming, distributed content, secure socket layers transmissions, and e-commerce support. In some cases, these applications require up to 100 times the downstream bandwidth of a static web page session. Again, this is a perfect fit for a HSD cable modem service.

The only solution to the latency issue is to cache this data as close to the end user as possible, completely removing the latency problem. Only the MSO has the perfect kind of deployment architecture that can cost-effectively support these needs in the business sector.

TARGET MARKET

The specific business user that the MSO should be looking for should have the following common traits:

- Reasonably close to the existing coax plant
- Subscribes to a high-speed data offering from a legacy ISP
- Uses the RBOC or a local xDSL provider for their last-mile connectivity to the ISP's POP
- Has an on-site router connected to the current uplink circuit
- Looking to upgrade to faster speeds from his provider
- Unhappy with latency

It is estimated by the Department of Energy that there are over 21,000 skyscrapers, 26,000 midsize buildings, and 61,000 small buildings in the United States. Using a blue-sky factor that 90% of all domestic businesses will be directly connected to the Internet by the end of year 2000, this leaves a target market of an estimated 3,000,000+ users of this service. And perhaps the most important statistic is that the MSO's current nationwide coax plant passes right by more than 80% of these profitable customers.

Let's examine this average user just a little closer. Over the last six years, when an organization wanted a connection to the net, they simply called one of the popular ISP's to set up service. In the early days, if this location had less than 20 users, a dial-up connection was the common choice. If there were more than 20 users, a Local Area Network (LAN) device was installed, along with a router.

This network then connected the router to a new technology called "frame relay" which offered uplink speeds from 56KB up to 1.544 MB. These circuits required a dedicated last-mile DS-1 circuit from the end user to the ISP's frame switch for any speeds greater, or equal to, 128KB. This frame switch then provided the uplink to the carrier's ISP backbone and they charged the end user up to $2K a month for this service.

The MSO can win big in this market by simply providing standard HSD services that have sophisticated caching devices installed in their network. This design will improve ***delivery of content*** by over 10-20x, verses the old ISP model.

Coupling the current features of HSD, with caching, will simply terminate the competition and remove the fear, uncertainty, and doubt (FUD) that have kept the MSO out of this game in the past.

COMPETITION

In the late 90's, we saw xDSL architecture develop. As long as to many flavors of xDSL are not placed in the same 50-pair copper binder, or to many high-speed xDSL circuits with legacy DS-1 circuits are grouped in the same copper binder, performance will be satisfactory. And as long as the RBOC can provide a "dark" copper pair completely free of any bridge tabs or digital loop carrier (DLC) devices, the same positive outcome will be achieved. If all of the above conditions are met as planned, xDSL service works extremely well.

A new application we are starting to see in the industry, which should be of strong concern to the MSO, is the deployment of small xDSL boxes into so-called "lit" buildings. This kind of service is designed to be full-flavor with voice, video, and data offerings bundled into one xDSL stream. Creating this model requires an expensive uplink dedicated circuit from the "lit" building to the vendors POP, to allow transport of these multiple services to take place. As a result of this costly design, the MSO can successfully compete with, at a minimum, the data side of this offering. Perhaps there's even a new business partnership with these players that should be explored in more detail by the local MSO.

xDSL players can, and in some cases, are, deploying caching systems in their models today. But, they still cannot match either the speed or the low delivery costs that a MSO can offer with a cached cable modem design.

We come back to the end user who currently has a 256KB dedicated Internet connection with an incumbent that is consistently getting slower and slower due to increased latency. The only option offered from the incumbent is to increase the speed of the last-mile to, for example, a 512KB circuit, which almost doubles the user's monthly cost.

The downside in this commonly offered solution is that, in the end, the improvement to overall service is less that 5% primarily because: ***the latency is in the backbone and not the last-mile!***

MSO OBSTACLES

"If it was easy everyone would be doing it," someone once wrote. Well, attacking this market is not as difficult as it seems, but will require some dedication, focus, and commitment on the MSO's part.

There are five major obstacles the MSO must overcome to enter into this business market. The technical issues can be solved today through the implementation of caching, with the remaining obstacles being a simple business development and/or partnering issue.

These obstacles are:

- Footprint
- HFC Uplink Issues
- Origin Servers behind Cable Modems
- Sales Force (or lack of)
- Layer 2 and wiring management and installation

Footprint

An easy argument can be made that the MSO has historically never wired any business locations with HFC. They have only wired homes, and this business model is not cost-effective.

Everyone then jumps to the conclusion that running fiber is the best means to success.

HFC Uplink

There has always been an underlying FUD issue about the limitation of the uplink capability of a particular HFC community. Although this issue does exist in a number of cases due to local cabling architecture coupled with the number of cable modem end users, it can be addressed in a number of ways. These ways include, but are not limited to caching, splitting off communities into separate frequencies and extending HFC fiber.

What needs to be addressed in this model, however, is the increased load placed on the network by a business-community user verses a one-work station user. Bandwidth savings on the HFC uplink will become a priority issue down the road as more and more businesses join the pool.

Origin Servers Behind Cable Modems

Everyone has a web page these days. The MSO can no longer simply turn this kind of client down, as they soon will run out of potential clients to sell to. So, the question is: How can a MSO survive if his HFC plant starts taking a gazillion hits going to one cable modem location?

To resolve this issue is rather simple with caching.

Sales Force

The MSO has two basic options here.

The first option is to proactively create an internal sales organization to market their own in-house labeled ISP service.

The second option is to support Open Access thereby creating an instant wholesale sales force model. By allowing every ISP the right to tap into your network to use this HSD cached last-mile connectivity will allow you to create a near zero cost sales model, increasing your profits significantly. We all know (and maybe fear) that when this business model finally does get deployed, these hungry players will blanket your current audience with strong advertising pitches to “buy now!”

In either of these sales models, the MSO can become very profitable using this sophisticated last-mile delivery system, allowing them to generate huge profits by simply applying their existing core competencies.

Layer 2 and Wiring Issues

In past models, the ISP simply told the RBOC where to install the circuit to allow the customer to plug and play. As long as your cable installation team can place the cable modem in the same spot as the telecom connection, there are no issues.

But reality indicates there could be a problem, as buildings are already internally wired for telecom connectivity and not necessarily cable. So, risers, equipment rooms and other labor-intensive connectivity problems could become deal-breaking issues if not resolved.

To solve this, the MSO must enter into a business relationship with either a local, or national expert in this field who can easily handle this task.

WHAT IS CACHING?

We discussed earlier how the latency in the Internet is caused. A computer user clicks onto his web browser, types in a web site connection and hits the button. The request goes out across the carrier’s network, in-and-out of numerous routers, until it finally reaches the origin server on the other side of the net. Then, a series of transactions take place to move objects from the server to the desktop to allow them to be painted onto the screen.

As we can see in our own daily use of the Internet, objects take between 2 seconds and forever to arrive for viewing. This delay is nearly always caused by the inherent latency in today’s connectivity architecture.

A cache intercepts this request before it goes out to the Internet path. It then attempts to instantaneously deliver fresh data from its storage medium to the user. How each cache product accomplishes this task is completely different, but the goal remains the same: deliver data fast, without getting caught in the latency trap.

Let’s remember that a router was built to move bytes from and to, and a cache was made to do just the opposite and keep the bytes close to home.

There are various types of caching systems available today. They range from full-scale platforms that perform a variety of sophisticated tasks, including caching, to custom-built devices using a freeware called Squid. The available caching systems also include new plug-in applications that connect to legacy devices, such as filers, and more recently simple to use caching appliances that just plug and play.

Some key factors to study when choosing a cache are:

- Is caching the primary function of the box?

- How hard is this product to install and maintain?
- How does this device react when it sees a web page request for the first time?
- How does the product go out and refresh and store its data?
- Can it support future applications such as streaming, distributed content, secure socket layer transactions, and address security and filtering issues?

The process of selecting a cache should carefully take each of these issues into consideration, as they directly relate to both MSO manpower needs and overall performance objectives.

THE TECHNICAL SOLUTION

Footprint and HFC Uplink Issues

By installing basic coax based services and not expensive fiber connections into these business sites, the MSO can now use its core competency to provide cost-effective HSD service.

What is still missing in this design, however, is how to overcome the FUD (some real/some perceived) already laid down by the competitors on this type of offering. The answer is to offer an HSD ***content delivery service*** that is cached at different locations in the network verses a plain bandwidth pipe to nowhere.

The MSO can now turn the FUD against the legacy provider. Caching devices should be placed at the main Head Ends, out into the network next to the CMTS, and even in the equipment room of multi-tenant buildings, driving the most frequently asked for content right to the edge of the network. Not only will this design significantly reduce the load off the HFC plant, but it will also create an unbelievable user experience that has never been matched by their old ISP.

This enhanced last-mile cached network design can then also be offered for a premium fee to the Open Access ISP when, and if, they arrive on the scene. Everyone will buy the cached version, as they will not want to be the slowest one on the block.

Origin Server Issue

A different set of problems exists here.

To allow the MSO to provide service to any end user who has a web site which is currently active on an in-house origin server, they must move the content of that web page as close as possible to the main uplink Internet connection to the carrier’s network. Then, all traffic coming into the web site will be satisfied by the primary uplink circuit, and as a result, will never burden the HFC plant.

One fast way to accomplish this task is to outsource this activity to a web-hosting specialist. Not only will you lose total control of the customer, but this option is also a timely and expensive experience for all parties involved. So, instead, let’s build one of these sites and keep all of the profit internally. With no track record of revenue, a very time consuming and expensive process, coupled with it being totally outside of the typical MSO core competency, we would be hard pressed to find a divisional MSO President who would fund this.

Instead, solve this issue by caching. By employing a simple DNS redirection process, one can point the DNS of the origin server directly over to the cache device installed at your Head End. When any Internet user requests this web site, the path will now go directly to the cache from your ISP uplink connection, and be delivered instantaneously back to the user.

The origin server can then easily contact the cache when it wants to update data on a particular object, in order to maintain freshness.

One additional feature of this design is that you will automatically create a firewall-type security system to keep hackers from corrupting the client's origin server. This is a billable item. (Not all caches can perform this feature, so please check before you try and deploy this model.)

The initial cost of this cached web-hosting model can be as low as $10K per city, making it both cost effective and easy to manage by the MSO team.

RISK/REWARD

The MSO should ask at this point "What's in it for me?"

The answer is: revenue and lots of it!

The current billing to an end user for a 1.544 MB DS-1 connection to the Internet can average well over $1200.00 per month, with a 512KB circuit averaging over $600.00 per month. The cost the MSO can now bill this business user is, at a minimum, 18:1 over what they can bill a residential user.

The great news is that this service will use the same gear, the same plant, the same billing systems, and the same CMTS devices they already have in place to service the residential community. The only additional architecture costs to deploy this business service will be the cost of adding the caching equipment.

The associated risk if the MSO ignores this market segment is that they will miss out on a tremendous market, which is begging for change. Furthermore, without a distributed cached network, the current HSD architecture will never be able to support all of the bandwidth glut type applications that are just now starting to create tremendous revenue streams for other players.

SUMMARY

Malcolm Forbes once stated that "for a company to succeed it must do one thing well".

The MSO has already proven that this strategy works. The challenge for the industry is to again apply its core competencies to create additional revenues from the surging Internet market, without disturbing the existing business model.

Jim Royle can be contacted at jim.royle@cacheflow.com

Network Interface Unit, A Broadband Gateway Into The Home

Robert S. Burroughs
Lucent Technologies, Cable Communications

Abstract

Cable has been promising two-way services since the early 1970s, but only recently have significant numbers of Cable plants upgraded to full operational two-way.

Requirements are presented for Network Interface Units (NIU) in two-way HFC networks that are deploying or plan to deploy blended services[i]. Numerous attempts to deploy NIUs, mostly for broadband Conditional Access, have traditionally not been very successful. An argument is put forth which encourages Cable to use NIUs to effectively deploy telephony and to use them to secure a beachhead against competition.

The What, Why, and How of deploying NIUs is discussed as well as several of the benefits, options, and problems to overcome.

INTRODUCTION

Competition is the raison d'etre for NIUs. Competition for the established Cable companies comes from overbuilders, the Local Exchange Carriers (LEC), and the Competitive Local Exchange Carriers (CLEC)[ii].

Overbuilders are differentiating themselves from the established Cable companies, by being more aggressive and deploying more of the new services, such as digital TV, Cable Modem, and cable telephony quicker. The LECs are capitalizing on their Carrier Class networks, their anticipated higher security, the world wide ubiquity of their lines, their dominance of the commercial telephony market, their deep pockets, and their ability to bundle telephony and high speed data in a single line.

Market analysts are projecting revenues for high speed data services, from Cable Modem, to be from $11.9 billion [1] to $26 billion [2], by 2003[iii]. Cable telephony could reach 9 million[iv] subscribers by 2003, at an average of $35 per month, representing additional revenue of approximately $4 billion. The projection for Cable's video revenues for 2003, is about $38 billion. Clearly, cable data and telephony represent considerable incentives to attract competition and venture capital.

Most current Cable telephony deployments are for switched rather than packetized telephony. This is due, in part, because the PacketCable specification is not complete. But, analysis demonstrates that even at the highest scales of economy, the total price per subscriber for the IP solution is about half of that for the circuit switched solution. The price gap is even larger when there are only a few subscribers on the network. This suggests that IP telephony would be the natural tool for CLECs and overbuilders to compete against the local telephone companies, and eventually against Cable's high-speed data and telephony offerings.

WHAT IS AN NIU?

Definition : A Network Interface Unit (NIU) is a physical enclosure, located at the service endpoint, which connects between the

HFC network and local residential wiring. It establishes a point of demarcation between the HFC network and the subscriber premise and may include a variety of functional elements.

The NIU demarcates, or clearly separates the HFC network from the subscriber's network, and consequently demarcates the Cable operator's area of responsibility from the subscriber's. Although the concept of an NIU can be used in traditional non-HFC networks, the NIU is more suitable to the evolving HFC network architectures. HFC was invented to exploit the broadband nature of Cable. It has done very well for traditional one-way services for the past decade, but its real purpose was to encompass two-way services. Because two-way services were not widely available when HFC was deployed, (the Internet was just beginning to emerge, telephony wasn't even legal in most states, interactive TV was faltering), NIUs were not seriously considered as an necessary part of HFC networks. But in today's competitive environment, NIUs and network powering need to be an integral part of new HFC network architectures.

To be an effective network element the NIU needs to be more than just an RF interface device. It can easily play a roll as part of the distributed intelligence of modern HFC networks.

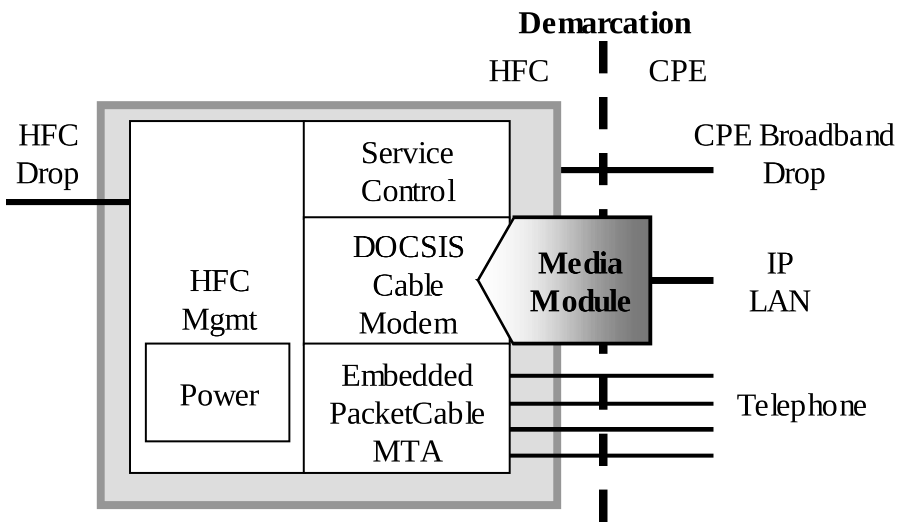


Figure 1: NIU Major Components

NIU Components

- HFC Management: terminates the coaxial cable. It separates the RF signal into three components:

  – 50MHz to 860+ MHz (downstream)
  – 5MHz to 42 MHz. (upstream)
  – DC to 1 MHz. (power plus status)

  Additional Management Functions:

  – status monitoring of;

  - NIU modules,
  - power usage,
  - the subscriber's upstream RF, and
  - other non-time-critical parameters
- control of upstream RF, from the subscriber, and downstream RF to the subscriber
- control of average power usage
- other low speed data functions, such as meter reading, load shedding, etc.

- Service Control provides demarcation of the RF, going to and coming from the subscriber. It functionally:

  - controls upstream impairments from subscriber's network (e.g. RF drop), with an On/Off switch or filters
  - turns downstream services On and Off
  - allows a variety of conditional access control modules for analog and digital services, e.g. de-scrambling, traps, jammers, and whatever clever contraptions the vendors can conceive.

Digital television and Open Cable present the same interfacing problem Cable has faced since the introduction of the first converter, Cable has more functionality than standard TV receiving devices provide. The QAM is a prime example of this trend. The 64/256 QAM, chosen by Cable, is not compatible with either current TV receiving devices, or future digital TV receiving devices, since 8VSB has been chosen as the North American broadcast standard for digital transmission. Even if digital TV receiving devices were equipped with 64/256 QAM receivers, the legacy NTSC and non-QAM receiving devices will predominate well into the future. In the near term, it appears that Set Top Boxes (STB) will be with us for some time, at least in the initial phases of digital TV. One possible solution is to add a Gateway, in the home, which would work together with the NIU. The Gateway would control all of the Home Networks and do all of the necessary protocol and signal conversions between Cable's digital TV signals and those required by digital and/or analog TV receiving devices.

Some of the ways that an NIU can alleviate the Cable network/ CPE device interface dilemma are that the NIU:

  - controls service access to the Gateway, for all TV channels (both digital and analog)
  - shares circuits, (such as the Out Of Band modem), when there are multiple TVs in the home
  - does analog conditional access at the demarcation point, so that this feature is not required in STBs or gateways (this is best done in the NIU so that subscribers don't have direct access to conditional access).

- DOCSIS Cable Modem, as an optional embedded element of the NIU, has key system advantages:

  - Cable Modems have an Ethernet bus, which can support multiple PCs in the home. All that is needed is a twisted pair from the NIU (or a wireless LAN) to each device that uses a data connection. Multiple Cable Modems can be added for MDUs and enterprises.
  - it is always active, because of the lifeline feature, and is used for monitoring intrusions and hack attacks
  - it can also be used in non-DOCSIS modes to assist in HFC network performance monitoring and control.

- PacketCable MTA is the CableLabs interface, from Cable Modem to standard analog telephones. It is specified to have four twisted pair analog lines. If a Cable Modem is included in the NIU, it is logical to include the embedded MTA to save a truck roll when the subscriber wants cable telephony.

  – the MTA was created to attach to the Ethernet bus of a Cable Modem located near a PC. But there is rational to embed the MTA in the NIU, where twisted pairs to phones terminate
  – each of the four lines can be provisioned remotely from the OSS, no truck rolls.
  – one of the lines can be provisioned to carry HPNA for other data services.

- Media Module (MM)

  – The MM takes the IP output from the Cable Modem and converts it to other standard interfaces, e.g. 802.11 and Bluetooth wireless, HPNA, etc.
  – a standard interface, such as PCMCIA is used to allow a variety of options offered by several vendors.
    – the PCMCIA interface may make more sense in a Gateway[v] in the subscriber's residence, so that the subscriber has easy access
    – in this case, the Ethernet, IP line, can be brought to the Gateway
  – it only makes sense to add the MM to an NIU if it is more cost effective than adding a Gateway.

- Power is one of the critical elements, and problems, of NIUs. Operators are requesting 5 watts or less of average power consumption (3 watts in some cases). Their concern is the cost of power, which is significant when there are thousands of NIUs deployed.

NIU power:

– Lifeline service[vi] is required for primary line telephony

– For lifeline service, network power is

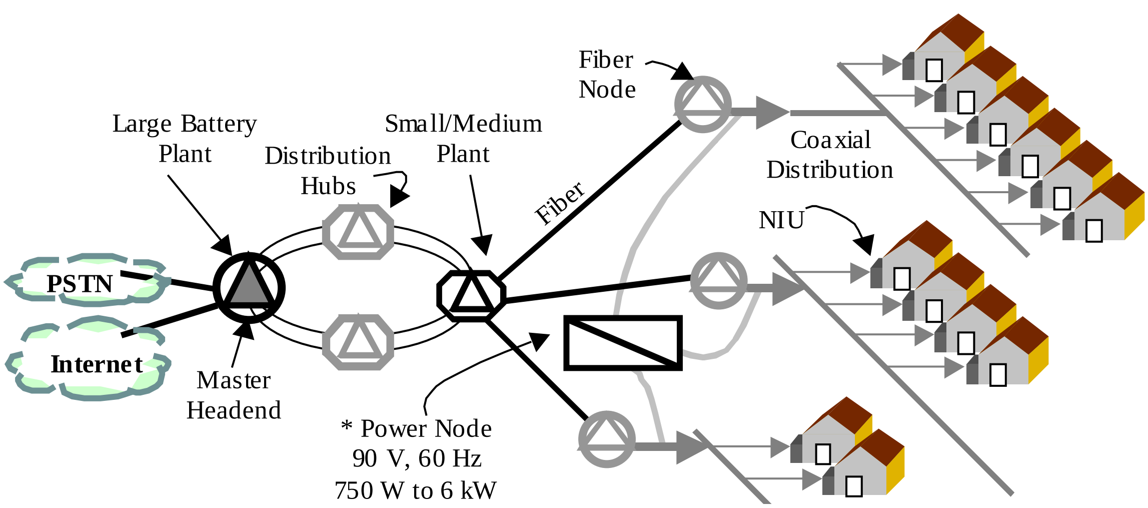


Figure 2: Gratuitous Cable Network Power Architecture (Centralized Network Power)

more cost effective than batteries at the NIU, when penetration of telephony exceeds about 10%.
- NIUs can be normally powered from the home and seamlessly switched to the network power when home power is lost
- New power distribution architectures are evolving for Cable (Figure 2)
  - these new architectures will have status monitoring and control capabilities, which could lessen the burden of CMTSs to support similar functions

Operations and Maintenance (O&M)

A means is provided to provision an NIU, even if there is not an embedded Cable Modem or the Cable Modem has not yet been initialized. An Out Of Band (OOB) modem is used either to work, as the Open Cable OOB channel, or over the cable network power system, using power line transmission standards.

Provisioning includes functions such as:

- Initial installation of NIU,
  - NIU arrives fully tested and configured with all security mechanisms ready
  - NIU is installed in a convenient location for service and if possible, close to telephony and power.
  - before installation
    - subscriber has prepared all networks and wiring
    - Cable operator could use a separate group that does wiring installations
    -
- Changing and ordering services (basic, pay, data, telephone)
- Service and maintenance of NIU and subscriber's account

Securing the Service

The NIU's primary security feature, it is out of reach and not easily tampered with by the subscriber. It has tamper alarms that indicate attempted tampering, either with the HFC drop cable or the NIU. If there is an intrusion attempt, an alarm is sent to the Operations Support System (OSS) and the appropriate action is taken locally within the NIU, (e.g. turn off all services, turn off specific services, inhibit the upstream), to be followed up by the Cable operator's action.

Key security functions:

- NIU is always active
  - tamper alarms
  - continuous monitoring of NIU status and HFC network
  - lifeline functionality for telephone
- Control of subscriber's access to HFC:
  - upstream RF
  - downstream RF
- Modular for easily upgrading or changing services and features by service technicians
- Common X.509 certificate and secret encryption keys, embedded by manufacturer,
  - difficult to clone, because Cable operator knows credentials of all NIUs
  - retail Cable Modems, MTAs, and STBs, still represent a threat, but if there is any suspected attacks, the NIU can block service.

WHY USE AN NIU?

Competition. By installing NIUs, the competition can roll out high speed data and telephony services faster than retail distribution or in-home boxes. And the NIUs provide a more secure environment and a well-behaved cable plant. NIUs are a key network element for competing in the high-speed data and telephony businesses.

- High speed data and telephony are profitable businesses in their own right and are a good way to gain new revenues (about $ 20 billion by 2003).
- DOCSIS and Packet Cable standards offer an attractive opportunity for CLECs and Cable overbuilders
- End-To-End IP telephony is substantially less expensive to deploy than the standard switched telephony approach, making it easier to compete with established telephone companies using switched technologies.
- The battle for the subscriber's data business could well hinge on the telephony offering.
- Since PacketCable telephony requires a DOCSIS Cable Modem, it makes sense to deploy NIUs with these devices installed
- LECs hold the edge in telephony, with carrier class networks, business experience, and worldwide presence. When they get their high speed data act together, they will be a formidable competitor.

Even though the Cable Modem and MTA are planned to be available through retail and supported by several MSOs, there are good system reasons for including these devices in the NIU. An NIU provides three important HFC network functions, essential for telephony and effective high-speed data:

- Network integrity,
- Operations and Maintenance,
- Service access control.

Network integrity

- Access and throughputs for Cable Modem upstream traffic are increased because impairments generated in the home do not get into the HFC upstream. These can come from:
  - RF leakage; from TVs, VCRs, and STBs.
  - motors and cordless phones,
  - computers and other digital devices
  - disconnected coaxial drops dragged across the radiator

- Hackers don't have direct access to the HFC network
- High peak levels of upstream transmitter bursts, (from Cable Modems or STBs in the home), will cause clipping in the upstream laser. Harmonics are generated which can affect the full 5 to 42 MHz spectrum. This can be detected and controlled by the NIU.

Operations and Maintenance (O&M)

The key to O&M is to reduce truck rolls:

- is conveniently located for service access.
  - typically external to the residence
- has status monitoring and control functions,
- is modular for easy upgrading ,
- has tamper alarms

Service Access Control

Cable now can offer a broad array of services to the subscriber. All of these services require access control, not only for the distribution of downstream services, but also controlled use of the upstream, where bandwidth is scarce and the spectrum needs protection.

It is more difficult to secure services in the subscriber's home, without an NIU:

- the devices can be powered off, so that the devices can not be continuously monitored for tampering, as can be done in the NIU
- encryption keys and certificates are not in a trusted environment.
- cloning threats, which can be reduced for in-home units as described in [3], still extorts a considerable burden on the CMTS

There are several attacks a hacker can launch against data and telephony services:

- <u>Disruptive Attack</u>s - inject impairments into the network to disrupt service,

- <u>Mild Attacks</u>
  - sniff the downstream to learn when other Cable Modems are scheduled to transmit, and jam their bursts with the intention of gaining more bandwidth
  - get your own map information, but jam other modems form getting their information in the downstream

- <u>Sophisticated Attacks</u>
  - clone Cable Modems, MTAs, and STBs
  - use authorized Cable Modems to steal service:
    - jam the downstream so that Cable Modems begin searching for another downstream frequency.
    - the hacker then inserts their own downstream channel, and begins to initialize other Cable Modems, bypassing security.
    - once initialized the rouge CMTS can then download new software to the unwary Cable Modems .
      - with the new software, the hacked Cable Modems can now be directed to send data to the hacker's CMTS, or to process any requests the hacker has for data.

<u>HOW TO IMPLEMENT AN NIU</u>

How to provision an NIU, when it is initially installed, is a matter of strategy. Not every subscriber needs an NIU initially. It is only needed in those nodes that have telephony or Cable Modem subscribers. What to put in the NIU, is also a matter of strategy. The basic elements are all required for network integrity, but different flavors of each element can be implemented. If the NIU and its interfaces are standardized by the SCTE, a variety of vendors will be available to supply modules.

Installing the Cable Modem and MTA in every NIU, is primarily a cost and powering issue. If the Cable Modem is deployed in the NIU, the MTA might as well be added too, because it is a modest cost-up. The cost will likely be offset by a simplification of provisioning, (e.g. reduced truck rolls), when the subscriber eventually decides to try Cable telephone service.

Semiconductor vendors are continually integrating more functions into less chips with smaller geometry's. Soon a single chip that does a complete Cable Modem and MTA will be available. A logical argument could be made to include the Cable Modem and the MTA in all NIUs.

- HFC Network Interface includes:
  - diplexer for separating the spectrum:
    - 40 to 90 VAC power
    - power line modem (status and control)
    - downstream (50 - 860+ MHz.)
    - upstream (5-42 MHz.)
  - embedded Cable Modem (5 to 860+ MHz)
  - Low Noise Amplifier (LNA), *optional*, to improve the noise figure of the downstream RF.

- Power supply
  - operates with network power (40 to 90 V, 60 Hz)
  - or 110 VAC home power
  - primary power is software selectable from either network or the home.
    - when primary power is from the home. Network power is seamlessly activated when home power is lost (Lifeline capability)

- Status, monitor and control
  - powerline modem is used (e.g. CEBus spread-spectrum modem)
    - alternatively Open Cable Out Of Band (OOB) modem, or Cable Modem could be used
  - monitored parameters
    - power usage and status
    - subscriber upstream power and impairments
    - NIU functional status (all modules)
    - downstream status (requires Cable Modem or status monitoring plug-in for downstream processing module
    - tamper alarm
  - controls
    - downstream service control (On/Off, conditional access)
  - upstream control (On/Off, filter adjust)
  - Cable Modem disconnect/connect

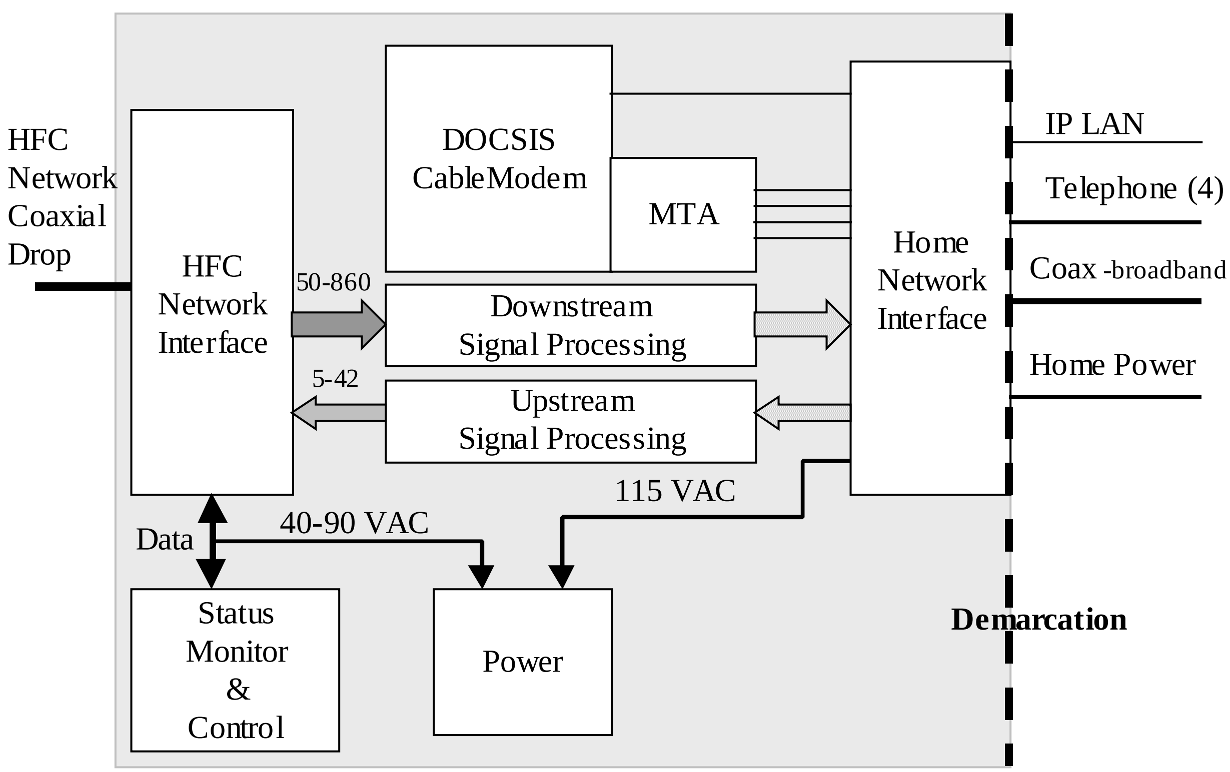


Figure 3: NIU Architecture - Typical Implementation

– NIU master reset

- Downstream processing

A simple embodiment includes:

– service On/Off switch
– Broadband amplifier, with digitally controlled output level
– diplexers

Provisions are made to incorporate additional functionality by using plug-in modules:

– filters and traps
– tuners
– analog descrambler
– other conditional access technologies

Downstream processing could also contain QAM demods for digital TV

- Upstream processing

Basic functionality includes an On/Off switch and upstream power level detector

Other plug-in modules could:

– filter specific upstream channels

- Embedded Cable Modem and MTA

There are several advantages to including the DOCSIS Cable Modem and the Packet Cable components in the NIU:

– the data and telephony modules will be isolated from any impairments generated in the subscriber's RF network
– a higher level of security is achievable because, the Cable Modem can always be guaranteed to be active.
– cloning attacks described in [3] are not a serious problem, because the subscriber doesn't have access to either the embedded keys or certificates. And with tamper monitoring in the NIU, any attempt to gain illegal access to the NIU will result in an alarm being transmitted to the OSS and a possible disconnection from service.

- Home Network Interface (+Media Adapter)

There are four primary interfaces into the home, from the NIU:

– coaxial drop, feeds any TV receiving devices or Cable Modem devices
– IP Ethernet, either 10BaseT or 100BaseT is available
– four twisted pair standard analog telephone lines
– home power (optional)

The Media Adapter allows other interfaces into the home:

– Home PNA, Cable Modem data is multiplexed with one of the analog lines
– PCMCIA, allows Cable Modem data to interface to several transport interfaces such as IEEE 802.11 wireless.

<u>Enclosure:</u>

There are several challenges for housing the NIU functions, including:

- Environment
- Heat dissipation
- Component access
- Security

High tech plastic has been effectively used to deal with typical environments that the NIU will encounter.

The use of multilayer polymer structures, for the enclosure allows taylor-made properties meeting a wide range of requirements for the NIU. Plastics can be made to meet the stringent requirements of an NIU, but are much lower in cost than similar metal enclosures. For example, an enclosure consisting of a multilayer sandwich structure can be made particularly suitable for NIUs used outside. Structures can be made to exhibit excellent long term weatherability and UV stability, good impact properties, (even at temperatures below -40 degrees centigrade), and good chemical resistance.

Heat dissipation is a major problem when designing low cost electronic enclosures for outdoor environments, particularly those Arizona summers. A common approach is to put a large heat sink on the back of the NIU to dissipate the heat. This approach can be expensive and makes for a heavy NIU. New tools for meeting the challenges of outdoor environments are becoming available. Semiconductor technologies, such as silicon-on-insulator, are currently used to make Cable tuners covering the Cable spectrum to 860+ MHz. The typical operating range exceeds 85 degrees centigrade. Other high tech means to absorb or dissipate the heat from NIU components, such a heat absorbing gels and forced ventilation techniques, are being tested. The problem is still waiting on an elegant solution. Fortunately there are many solutions currently being developed, not only for Cable, but for NIUs deployed for telephony and other equipment such as cellular base stations.

Component access provides security, determines the ease of provisioning and servicing, and the flexibility to upgrade. It would be a major advantage to the Cable industry, if it were to turn its standardization efforts toward standards for NIUs and NIU plug-in modules.

Two types of standardized modules are needed. Analog RF modules, which plug into an RF bus and digital modules which use a standard data bus, such as the PCI bus. All modules should be plug-and-play at the cable network level so different configurations can be easily changed or upgraded by service personnel working in conjunction with the OSS.

The downstream and the upstream signal processing are separated in the NIU and each is processed independently. The upstream processing has the capability to turn off the upstream or filter impairments from the Home Network. The downstream has a wide band amplifier to compensate for losses incurred by the Home Network. Additionally, conditional access technology for the analog and/or digital channels can be added, depending on the Cable operator's preferred implementation. Various types of conditional access technologies are available in the market, and could easily be adapted to the NIU if standardized interfaces are defined by the industry.

Status Monitor

A key NIU element is the status monitor and control modem. This is used for provisioning (activating / deactivating conditional access, supporting plug-and-play capability), monitoring the status of the upstream, downstream, and key elements of the NIU.

Power

And finally a means of maintaining power in the NIU, even if the power is lost in the home. This only critical when cable data and telephony services are deployed, to be able to provide the lifeline capabilities provided in primary line telephony, and may not be needed for NIUs that only provide an RF demarcation. The NIU can be powered from the home, but when there is a power outage in the home the HFC network should provide the power, seamlessly, with no interruption of service. This power could be, and is currently, provided by batteries located at the home, but when the service penetration is greater than 10%, network powering becomes less expensive than the cost of the batteries, not to mention the ease and cost of maintenance.

CONCLUSION

Ten years from now, 2010, virtually all HFC networks will have NIUs. That is a pretty strong statement, but operators deploy HFC for their ability to provide blended two-way services. Telephony could be the trump card that captures the high-speed data subscriber. This could be a likely scenario because, Cable has three strong competitors, LECs, CLECs, and overbuilders[vii] who will push for combined telephony and high-speed data services. As competition intensifies, over the next few years, NIUs will be required to be able to manage the HFC networks efficiently and to enable the deployment of new services and innovations rapidly.

REFERENCES

[1] Research group Kinetic Strategies, (11/99)
[2] IDC report (12/99)
[3] Medvinsky, Sasha; Anderson, Steven, “DOCSIS Security vs. Cable Modem Cloning”, Conference on Emerging Technologies, January 11-13, 2000, Proceedings Manual.
[4]Jones, Doug, “PacketCable Security Overview”, Conference on Emerging Technologies, January 11-13, 2000, Proceedings Manual.
[5] Burroughs, Robert, “A Point of Entry Interface for 2-Way Broadband Information Delivery”, June 7, 1993, NCTA Cable ’93 Proceedings Manual.

ACRONYMS

CLEC	Competitive Local Exchange Carriers
CMTS	Cable Modem Termination System
CPE	Consumer Premise Equipment
DOCSIS	Data Over Cable System Interface Specification
HPNA	Home Phoneline Networking Alliance
LEC	Local Exchange Carrier
LNA	Low Noise Amplifier
MM	Media Module
MTA	Multimedia Terminal Adapter
O & M	Operations and Maintenance
OOB	Out Of Band
OSS	Operations Support System
PCMCIA	Personal Computer Media Communication Interface Adapter
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency (> 5 MHz)
SCTE	Society of Cable Telecommunications Engineers
VSB	Vestigial Side Band

NOTES

[i] Blended service, is a term that is currently in vogue to describe the new services that will be introduced when the PacketCable standard is implemented. It includes, high-speed IP data, IP telephony, streaming video, digital and interactive TV. With PacketCable, all of the services tend to blend together.

ii CLEC, sounds like a tribe out of Star Wars. But, they are entrepreneurial companies that are competing with local telephone companies for the local telephone business. They're typically very aggressive and looking for every avenue to compete.

[iii] Cautionary note: Kagan, Q1 99, projected Cable's high-speed data revenues to be only $3.8 billion by 2004. But, with Cable modems reaching more than 2 million units by 2/2000, this number appears too low, if you assume an average 1 month rental of $40.

[iv] Typical number, at high end, from several sources: Forester 2/99- 9.1 million, Goldman Sachs 7/99 - 9 million, Kagan 6/99 - 5.8 million

v The term Gateway usually refers to a device that passes traffic between different protocols, such as between Ethernet and Bluetooth protocols. The term gateway is used in its generic context, in the title of this paper, to mean the door between the subscriber's Home Network and the HFC network.

vi Lifeline power: power is available to make telephone calls even if there is no power in the residence.

[vii] since the Internet represents a huge opportunity, there are more than the four major players looking to gain market share: satellite, wireless - PCS and broadband.

Network Support Infrastructure For Pod-Based Systems

Mark DePietro
Motorola Broadband Communications Sector

Abstract

The FCC Report and Order of June 1998 called for the availability of retail navigation devices and PODs in the July 2000 timeframe. In order to enable a set of services in a POD-based set top or digital TV, it is necessary to consider corresponding changes that must be made to the supporting digital network infrastructure. This paper examines those changes in the context of the feature set that has been defined by the OpenCable process for July of 2000. In addition, other system level considerations that affect the success of OpenCable systems are discussed.

BACKGROUND

In order to be able to declare that PODs and Hosts have been successfully deployed, it is necessary to establish criteria by which to measure success. In the OpenCable process, success has been defined as the ability of the entire system, consisting of Headend, POD, and Host, to exhibit a set of behaviors identified as the J2K Feature set. In this context, J2K stands for "July 2000."

J2K Feature Set

The following elements comprise the feature set associated with the J2K rollout of PODs. This feature set was derived as a result of OpenCable discussions that took place in the summer and fall of 1999.

- Clear Analog Services
- Clear Digital Services
- Subscription Digital Services
- Call Ahead Pay Per View
- Impulse Pay Per View
- Support for HDTV Passthrough via 1394
- Copy Protection on the 1394 Interface
- Copy Protection on the POD-Host Interface
- System Information in Accordance with DVS-234 Profile 1
- Emergency Alerts Based on DVS-208
- Closed Captions Based on DVS-157 and DVS-053
- Parental Control Supported by Content Advisories in the PMT
- Electronic Program Guides

Multiple Deployment Scenarios

Three distinct POD/Host deployment scenarios have been identified to date. In the ***Native Network*** scenario, the Host manufacturer and the POD manufacturer coincide. In the ***Foreign Network*** scenario, the Host manufacturer and the POD manufacturer differ. Finally, in the ***POD-less Host*** scenario, a Host is deployed without a corresponding POD.

Each scenario enables completion of a different range of features in the overall feature set. The POD-less Host scenario enables the smallest set of features for obvious reasons. Without a POD, the Host has no access to the out-of-band data feed, nor does it have access to encrypted channels. As a result, a PODless host can not access anything other than clear analog and clear digital services.

Code Download and Foreign Network Constraints

In the foreign network scenario, the POD and Host are from different manufacturers. Because code download was not part of the J2K feature list, Hosts are generally not able to participate in code download operations in foreign networks. Since hosts are sold at retail, they can not be assumed to possess an electronic program guide that is suitable for use in the subscriber's cable system. This is true because retailers typically serve consumers whose dwelling locations span multiple MSO service areas, and it is not practical for a manufacturer to pre-load every conceivable EPG into a host. If suitable business relationships between the Host manufacturer, retailer, and MSO have been established, it is possible to envision a solution to this problem wherein the retailer performs the service of downloading an EPG into the Host on behalf of the other two parties.

Alternatively, the manufacturer may establish a private mechanism to obtain EPG downloads and data feeds that bypasses the MSO (e.g., through a manufacturer-specific telephone dialup service). Because these solutions are beyond the scope of OpenCable, they are not considered further. Other features that rely on the existence of either a download mechanism, either directly or indirectly, are not available in the foreign network scenario for J2K.

In the remainder of this paper, all features being described should be assumed as required in both the native network and foreign network scenarios, unless explicitly stated otherwise.

SYSTEM LEVEL IMPACT ANALYSIS

There are 5 specific features in the J2K Feature set that drive a need for systemwide changes. these features, and the required system changes to support them, are described below.

COPY PROTECTION

The design of the copy protection system for OpenCable has had significant systemwide development impacts. This section describes the reason why copy protection is needed in the system, and discusses the related issues of certificates, certificate revocation lists, content tagging, and ultimate limitations of any copy protection solution.

The Need for Copy Protection

One side effect of partitioning a set top into Host and POD is the creation of an easily accessible interface over which all inband content flows. In the absence of a copy protection scheme, this provides an easy point of attack for pirates wishing to make unauthorized digital copies. To remove this deficiency, it is necessary to encrypt the interface between the POD and the Host. This requires the selection of:

(1) An encryption algorithm, and
(2) A key management scheme

If both of these selections are completely known in the public domain, the resulting solution strength is ultimately equivalent to having a clear interface. To avoid this weakness, it is useful to introduce a set of secrets that are known a-priori only to legitimate Hosts and PODs. This introduces the need for a secret access-granting authority. CableLabs plays this role with respect to POD-Host interface encryption system parameters. Finally, even with the existence of such an authority, it is not wise to assume that these secrets will be maintained (as evidenced by the various DVD breaches that have emerged). Thus, it is also desirable to have in place an intellectual property barrier whose violation can be used as the basis for litigation against an interface attacker. The use of DFAST in the key generation process plays this role in the POD-Host interface, and CableLabs administers access to the DFAST intellectual property.

The Role of Certificates

By putting encryption and key management in place across the POD Host interface, attacks mounted by third parties against the interface are effectively thwarted. However, this level of protection was not considered adequate during the OpenCable requirements analysis phase. It was also desirable to know ahead of time that Hosts would not make inappropriate use of content once it traverses the interface from the POD. To receive assurances in this area, the Host is required to possess a certificate that ensures it will behave as a "good citizen" with respect to handling of high value content. Furthermore, the POD is required to: (a) validate the authenticity of the Host certificate, and (b) report the results of the authenticity check (in a non-spoofable fashion) back to an entity in the control system. The control system receives the authenticity verification from the POD, and compares the Host ID against a list of known bad hosts - the so-called ***Certificate Revocation List*** (CRL). If the germane Host ID is not present on the CRL, the Host is considered to be well behaved with respect to protection of high value content. As a result, the control system sends a "Host Validation" message to the POD. The Host Validation message is protected by a digital signature to prevent spoofing attacks. Until the POD receives a Host validation message from the control system, it will not decrypt any content that has been tagged as "high value."

Tagging Content in the Control System

To complete the copy protection solution, it is necessary to have a facility in the system that provides a distinction between low value content and high value content. In the OpenCable solution, a data element known as Copy Control Information (CCI) plays this role. Four values of CCI have been defined to date. ***Copy Freely*** content requires no protection. The other three values, ***Copy Once***, ***Copy Never***, and ***No Further Copying***, all require protection in the form of POD-Host interface encryption. The CCI value "No Further Copying Permitted" is typically a source of confusion. This particular CCI value is only used as a designation on second generation material where the ancestral content was tagged as "Copy Once".

In addition to tagging content, cable operators participating in the OpenCable process wanted to put in place an infrastructure that enabled price differentiated purchase opportunities for a single item of content. As an example, suppose an item was marked "Copy Freely". The OpenCable participants wanted to have a way to offer the content for purchase at two different prices. At the lower price, the content would appear to have a more restrictive "Copy Never" tag from the perspective of the Host. At the higher price, the content would appear to have the less restrictive "Copy Freely" tag, again from the perspective of the Host. In order to provide this capability in the system, each individual POD needs to have the capability to present a different tag to the Host based on criteria associated with the mode of purchase. As a result, the access control system was required to be involved in the CCI delivery chain. This caused a ripple effect back to the Billing systems, because these systems act as the point of entry for pay per view schedule data into cable access control systems. The ripple effect also extends to the Electronic Program Guide (EPG) providers, who will need to develop user interface screens that are capable of presenting the purchase options to the subscriber.

Advanced Tagging

For J2K, the four values of CCI defined above will be implemented. However, during OpenCable discussions, more esoteric forms of CCI were also considered and ultimately rejected for the J2K timeframe. Given the recent popularity of Personal Video Recorder systems, it is possible to envision a need to tag content with respect to even ephemeral copying that is required to support functions such as program pause and time-shifted viewing. This class of tagging has been considered in ATSC, and there are proposals on the table that extend CCI to include values such as "Pause No More", "Copy with a 15 Minute Lifetime", etc. These extended forms of CCI may be required in a later version of OpenCable, and are likely to have systemwide requirements ripples similar in nature to those induced by the current set of defined CCI values.

Choosing Content to Protect

It should be noted that the current copy protection system provides protection primarily for the content owner. During the course of the development of the OpenCable copy protection specification, there was a fair amount of discussion surrounding the choice of content that should be encrypted over the POD-Host interface. To protect against a re-distribution attack on the conditional access system, it would be necessary to encrypt content on the POD-Host interface whenever the content is encrypted as it traverses the HFC system. In addition, if a conditional access system employs a very long crypto-period, it is possible to use the clear content on the POD-Host interface as a point for mounting a plaintext/ciphertext attack on the CA system. Both re-distribution attacks and plaintext/ciphertext attacks were considered to be beyond the scope of the envisioned threat model. As a result, content is only encrypted over the POD-Host interface when it has been tagged as "high value", meaning that the CCI value is not set to "Copy Freely."

Other Copy Protection Considerations

It should be noted that the above described copy protection system provides protection against a pirate who wishes to make pristine first-generation digital copies of content. Digital copies of good quality can still be obtained by making a camera-based copy of a television image, by taking an analog output and digitizing it, etc. It should also be noted that in order for "bad hosts" to be shut down, it is necessary to place the hosts on the CRL. The current copy protection system provides no watermarking feature. As a result, there is no defined facility in the copy protection system that could be used to ascertain the point of origin of content that is discovered to be pirated. Thus, the creation and maintenance of the CRL will require the existence of an administrative process that is based on an approximate audit trail.

Summary of Copy Protection System Impacts

As a result of the considerations noted above, it is necessary to modify existing conditional access systems and interfaces, billing systems, and electronic program guides to support the OpenCable copy protection solution. In addition, it is necessary to implement the administrative processes that will be used to govern the content of Certificate Revocation Lists.

SYSTEM INFORMATION

One of the key elements required to support Hosts from a variety of manufacturers is a uniform facility for communicating channel lineups that are active in the system. In today's digital systems, channel lineups are typically communicated out-of-band in a data structure known as the ***Virtual Channel Map*** (VCM). The VCM is a logical view of the channel lineup that allows a subscriber to maintain a constant interpretation of channel map even when there is variation in the underlying physical channel lineup.

In digital systems from different vendors, these VCMs are often transported in non-uniform ways. For example, some systems use MPEG transport on the out-of-band, while others use ATM transport. In addition, some systems support multiple channel maps within the context of a single hub while others employ one channel map per hub. In OpenCable, it was necessary to hide these intersystem differences from the Hosts in order to avoid unnecessary complexity in Host out-of-band processing firmware. The POD plays a crucial role in hiding these inter-system differences from the Host. It strips off the transport headers that are used in a particular system, and accumulates sections of the VCM for delivery to Host. It also filters out any VCMs that are not relevant to the Host, giving the Host the illusion that it is operating in an single VCM environment. So, in addition to being the security element in the system, the POD also plays a key role in supporting the delivery of System Information in a uniform fashion.

EMERGENCY ALERTS

In currently deployed digital systems, different methods are used by different vendors to denote the occurrence of an emergency condition and to provide the cable subscriber with needed emergency information. Often, the solution involves the generation of a proprietary message on the OOB that redirects the set top's tuner to an analog station that is carrying the emergency information. When the emergency condition terminates, the set top is returned to the previously tuned channel. In OpenCable systems, the use of private mechanisms to communicate emergency conditions is not acceptable. As a result, it was necessary to develop a common emergency alert mechanism that relied on public domain messages. SCTE DVS-208 was developed to fulfill this role. For OpenCable Systems, it will be necessary to carry DVS-208 messages in addition to any existing private messages that might currently be in use. This requires an upgrade to each headend that normally takes the form of a firmware download to one or more headend components.

CLOSED CAPTIONS

In today's digital systems, two closely related but different methods are used to carry closed caption data in digital streams. One method is based on DVS-157; the other is based on DVS-053. In some deployed hosts, there are hardware constraints that require DVS-157 to be present in order for closed captions to operate. As a result, in OpenCable, it will be necessary to dual carry DVS-157 and DVS-053 forms of closed captions, and it will be necessary for OpenCable Host devices to be prepared to process both forms. The Hosts are required to support both forms in order to facilitate any transition periods that may exist during which only one form of CC data is present in the system. To support dual carriage of DVS-157 and DVS-053, currently fielded encoder systems will need to be upgraded.

RATINGS AND PARENTAL CONTROL

In current systems, program ratings information is typically carried in the Electronic Program Guide data feed using proprietary data formats. In addition, parental control functions used to block access to undesirable programming are facilitated via the EPG. In OpenCable, it could not be assumed that any given EPG data feed would be present on all systems, nor was it desirable to mandate the presence of a single EPG. At the same time, it was considered essential to provide a uniform, public ratings conveyance mechanism, in part to achieve compliance with FCC rules related to V-Chip functionality in the digital domain.

In the OpenCable process, it was decided that ratings information will be carried with the content itself, and will be captured at the point of content encoding. The ratings information will be carried in the MPEG program map table (the PMT), using a ratings descriptor that is congruent with the one defined in EIA-766. OpenCable Hosts are required to monitor incoming PMTs, and must take appropriate actions (e.g., blank video and mute audio) whenever incoming content has ratings that violate established parental control constraints. Hosts must take these actions within 200 milliseconds.

PMT Version Change Considerations

As a result of these requirements, MPEG PMTs will undergo a version change at every program boundary to signify the presence of a new ratings descriptor. These PMTs are typically transmitted very frequently, on the order of 10 times per second in each multiplex, in order to facilitate rapid channel acquisition. Once a channel has been acquired, it is *usually* not necessary to examine the PMT again until a channel change operation occurs, since PMTs tend to be relatively static objects. Nevertheless, PMTs can change and when they do in current systems, they typically signify the occurrence of a disruptive change to the underlying PID structure of the transport multiplex. Thus, in most systems, when the PMT changes, set top firmware initiates a program re-acquisition cycle. In the future, when ratings information is carried in the PMT, a version change in the PMT will now most likely not signify a disruptive change to the PID structure.

As a result, it will be mandatory to rethink the logic used to respond to PMT changes, and put in place an ***Advanced PMT Monitor*** (APM). The APM will be responsible for analyzing the details of changes to an incoming PMT, and will need to avoid a reacquisition cycle when the PMT is only used to identify a ratings transition point.

Headend Remultiplexing Considerations

Set top firmware and encoder systems are not the only elements that are affected by a requirement to carry ratings information in the PMT. Every headend is also affected in some way, particularly when the headend supports remultiplexing operations. Whenever an MPEG remultiplexing operation occurs, it is necessary to tear down any incoming PMTs and reconstruct the new PMT to ensure consistency with the newly formed multiplex. In current systems, this normally means that pointers to video, audio, and data PIDs must be changed to complement any PID remapping that is occurring in the multiplex. However, with the addition of ratings information to the PMT, pointer preservation will not be adequate. It will also be necessary to preserve and re-insert incoming ratings descriptors. Thus, every deployed digital cable headend will need to be modified to support PMT ratings descriptor preservation functions.

ELECTRONIC PROGRAM GUIDES

For J2K, there will not be a single EPG that spans all systems. In addition, there is no OpenCable code download mechanism defined for J2K. As a result, EPG capabilities will exist only in the native network scenarios. Over the longer term, the cable industry and CEA have agreed that a certain base level of EPG data will be carried in the system in the form of inband PSIP. This, along with a future code download mechanism, will allow for future portability of EPGs.

OTHER SYSTEM LEVEL CONSIDERATIONS

In addition to the J2K feature list, it is worthwhile to consider some factors that will play a part in defining the evolution of OpenCable systems. Four relevant factors are briefly considered in the following sections.

Heterogeneity Of Security Features

Different conditional access systems have different features that present different experiences and system capabilities to the user. Examples of areas where CA systems differ include, but are not limited to:

(1) The quantity and nature of entitlements differ from system to system. In some systems, entitlements are conferred via service keys, in others the tier concept is used. Many other mechanisms are possible, and likely to exist.
(2) The method used to implement the concept of money differs from system to system. In some systems , there is no monetary concept. In other systems, a credit mechanism is used. In other systems, time is the element that equates to money.
(3) Some systems implement concepts such as free previews, program packages, and blackout areas, while others do not.

Given the existence of these differences, the following alternatives existed with respect to the interface between the host module and the security module:

(1) Account for the lowest common denominator features in the Host to POD functional interface.
(2) Define the interface to include all capabilities that could be encountered in any CA system, and use a profiling mechanism to identify differing levels of support.
(3) Define a very low-level host module to security module interface, and then require the existence of security module specific "driver firmware" to exist on the host device.

In OpenCable, a combination of alternatives (1) and (2) was chosen. As a result, applications developed for OpenCable will need to be aware of the existence of various security profiles, and will need to be designed "defensively" so that the lack of a particular feature in a given system can be recovered from in a graceful manner.

Firmware Distribution and Host Configuration Identification

In current digital systems, there is a primary vendor that provides both headend and set top equipment. Firmware destined for a particular population of set tops is distributed directly by the manufacturer to the MSO's site, often through the use of a field engineering organization. This direct distribution model does not scale well in an multi-host vendor to multi-MSO environment. Ultimately, there will be a need for a multi-hop distribution channel in which an intermediate business entity facilitates the many-to-many distribution operations.

In addition, each delivered firmware object will be relevant to only a selected subset of Hosts in a given system. This is true even in the presence of a defined OpenCable middleware solution for the following reasons:

(1) The middleware engine implementation will be host specific.
(2) The middleware solution itself will be an object that is subject to evolution. Eventually, capability profiling or discrete forms of a middleware application will need to exist to distinguish between capability sets resident in different classes of Hosts.

As a result, there will ultimately be a need to have a Host configuration identification mechanism that spans multiple vendors. A two-tiered identification scheme, based on an organizationally unique identifier (OUI), makes the most sense. Some entity such as CableLabs will need to emerge as the OUI administrator.

Compatibility Matrix Generation and Maintenance

In current systems, features are added to deployed systems by making synchronized complementary changes to the headend and set top populations. The configuration management discipline of ***Compatibility Matrix*** generation is used to record information that defines version sets of various system elements that are known to operate properly together. In OpenCable systems, this discipline will become more important, since the number of different system elements that must be coordinated will increase. Some economic model must be developed to support large multi-vendor compatibility matrix generation efforts.

System Level Troubleshooting

In single vendor systems, or systems wherein a small number of vendors are involved, the locus of responsibility for performing system level troubleshooting is narrowly defined. In OpenCable systems, where multiple vendors supply PODs and Hosts, and where subscribers have a business relationship with both a retailer and an MSO, it is not at all clear what system level troubleshooting model or models will exist in order to ensure subscriber satisfaction. The emergence of clear, explicit models must occur in order to avoid customer confusion and dissatisfaction that would likely result otherwise.

CONCLUSIONS

This paper has examined the system level changes that are needed to support POD based systems. Significant progress has been made by the OpenCable community to ensure that systems will be ready to support PODs in the July 2000 timeframe. Additional work will be required to support portability of a compelling feature set.

Noise Power Ratio the Analytical Way

Robert L. Howald
Motorola Broadband Communications Sector

Michael Aviles
Motorola Broadband Communications Sector

Introduction

Noise power ratio (NPR) testing is a valuable tool for characterizing return path link performance capability. It provides a quick snapshot of a link's noise and distortion performance with one easy to understand performance curve. Additionally, it is a relatively simple test to perform, particularly with the development of automated NPR test equipment. The development of HFC into a high performance two-way interactive communications medium has meant the need to assure a high quality link. NPR testing is a valuable characterization tool towards meeting this goal.

For the period of time that NPR has been associated with CATV (it has long been associated with other applications, such as satellite communications), the performance has generally been evaluated using manual means. The test is relatively easy, but manual operation can be time consuming. More recently, test equipment has been developed to support NPR testing using an automated measurement run from a PC. This has simplified the test and evaluation process. However, it is always the case from the standpoint of system design to have tools that allow accurate prediction of system performance. With such tools, it is feasible to explore all of the possibilities in a design via modeling and simulation before committing significant resources to equipment design, integration, and test. Pure computer simulation techniques can be applied, but simulation approaches can have potential for inaccuracy in applications involving nonlinearity, particularly when the nonlinear process cannot itself be simplified with convenient assumptions.

In this paper, we describe an approach that develops a mathematical model that accurately captures the multiple nonlinear impairment sources, providing a complete representation of the composite distortion generation process. Such a model has been used in product development programs to optimize designs, analyze link performance, and predict performance of various HFC architectures and link cascades. Predications are made using a library of equipment models developed consisting of the various link components.

NPR Basics

The idea of NPR can be considered a very straightforward extension of the common two-tone distortion concept. In that case, a device's nonlinear performance is evaluated by driving it with two CW carriers in its passband, and looking at the device output for new frequency components that did not exist in the input signal. Nonlinear components typically expected are harmonic tones associated with either of the two inputs, as well as mixing products of the two input tones. Observing these undesired components allows determination of the device's intercept points (IPx), the fictional point representing the crossing of the distortion input-output curves with the fundamental linear gain curve. In turn, these intercept points are very useful for cascaded system analysis, as the distortion behavior of entire chains can be predicted if the intercept points of the components are known.

The IPx concept has generally been associated with narrowband systems. CATV has typically used composite second order (CSO) and composite trip beat (CTB) to characterize nonlinear distortions. These parameters account for the fact that the distortion generation process with a large carrier multiplex creates multiple contributors to second order and third order impairment

Under a reasonable set of assumptions, these composite distortions can be related in a straightforward fashion to the intercept points. For example, in the case of CTB, we can write [9]

$$CTB\ (dB) = 2\ [Pc - IP3] + 10\ \mathrm{Log}\ (3N^2/8) + 6\ dB$$

In the above equations, IP2 and IP3 represent the second and third order intercept points, respectively. The number of analog (CW) channels is given by N. Pc is the power per carrier. For this third order case, the equation is the mid-band CTB, which is the worst case under the assumptions. One such assumption for this to be the case is flat IPx performance versus frequency. This is not necessarily a practical assumption for a multi-octave device, and, depending on magnitude, can be important to the accuracy with which performance can be predicted. It is typically handled today with empirical approaches in the forward path. In the reverse path, analysis and measurement as part of the studies undertaken in this paper have shown it to be less significant.

Now, it is often the case that literature or analysis concentrates on third-order distortion, and point out that the third-order intercept point (IP3) is the key parameter in evaluating distortion performance. The reason for this bias is that third-order distortion contributions can fall within the signal band even for narrowband systems, the majority of cases for RF systems. By contrast, second order components tend to fall outside of the signal band of interest. However, in HFC return systems, the North American split of 5-40 MHz is three octaves wide. Thus, the second order components can greatly contribute, and this analysis includes those effects.

Extending the concept of two-tone distortion, consider that the actual signal carriage on the return path is generally modulated data-carrying RF signals, applied to the return path at a predetermined power allocation. The reverse path multiplex means that the set of signals applied will each generate distortion components themselves and with one another. Because of the broad passband, the distortion components can fall anywhere in the return band. From the input signal mix of frequencies, it is predictable where the intermodulation distortion (IMD) components will fall exactly, and how many will land in one frequency location or another using common "beat-mapper" programs.

RF and Laser Distortion Contributions

The ability to estimate the NPR of a system requires the analysis of both its noise and distortion performance. A CATV HFC return link will consist of both RF and optical components. It is convenient to define the optical portion of the link parameters in terms of equivalent RF parameters. This facilitates the mathematical cascading of the link in order to generate the overall intercept points and noise figure, which are then be utilized in the NPR modeling effort. In addition, for optical links, the laser diode clipping characteristics must be known.

The RF link contributes to the cascaded noise figure of the network, as well as to the distortion performance through its 2^{nd} and 3^{rd} order intercept points (IP2 and IP3). How much of the contribution to the overall link performance is due to the RF section is determined by its relationship to these same parameters in the optical portion of the cascade. In general, the optical link has been the limiting factor in system performance, both in terms of its noise and distortion performance. With improvements in high performance and lower cost optics technology, and the development of more sophisticated RF processing, such as return RF collector architectures and block conversion, the RF section's performance may take on a larger role in determining the overall characteristics of a link.

Lasers contribute to NPR limitations in two ways. The first is due to relative intensity noise (RIN), which can be converted into an equivalent noise figure and cascaded with the rest of the system. The second way is through its distortion characteristics, of which there are two main types. One is the familiar RF-like distortion generated due to the non-linearity of its transfer function, which can be characterized by the standard two-tone distortion parameters.

The other key distortion contributor is due to the clipping phenomena that can occur because of the peak-to-average variation of a noise-like return load. This clipping effect can be mathematically evaluated as a non-linear distortion (NLD) parameter [7]. The C/NLD due to clipping usually sets the upper limit of the NPR curve by defining the point at which the carrier to noise plus distortion (C/(N+NLD)) becomes unacceptable, regardless of how good the non-clipping related distortions may be. This curve is also useful, without the RF part of the link, for estimating optimal alignment of optical levels, as can be seen from Figure 1 (note that all figures are located at the end of the text).

The Concept of Multitone Distortion

Estimating the NPR performance of module and cascades using two-tone distortion information requires modification of the resultant intercept points to an approach based on an (ideally) infinite number of infinitesimally spaced tones. Multitone analysis is based on the adjustment of the distortion products as the number of tones increases, while restricting the total output power level to that of the two-tone level.

In terms of the 3rd order products, the resultant distortion level asymptotically approaches a level that is approximately 8 dB higher than that of the two-tone case. This level may be considered the worse case, occurring in the middle of a contiguous band of multitones.

An alternative, but valuable and more computational approach would be to use a beat mapping program in combination with the measured intercept point to calculate the discrete 3rd order distortion product levels as they fall over the frequency range of interest. An artificially high number of tones across the band simulates a typical noise density, with the more accurate results occurring with the highest number of tones used. This approach also allows for the estimation of approximate distortion levels over any range of frequencies, instead of just the middle of the signal band, at the expense of computational power. In other words, NPR versus frequency can be analyzed, and this often turns out to be quite useful. For equally spaced tones at equivalent levels, the analysis would show a 3rd order product buildup with a shape similar to that of a bell curve centered at the mid-point of the multitone frequencies [8].

For optical systems especially, the 2nd order distortion performances may also be a major contributor to the final achievable NPR levels. Using the above logic, it may be deducted that the 2nd order level based on an infinite number of tones approaches a point 3 dB higher than the two-tone levels. Care must be taken, however, in evaluating 2nd order distortion for NPR evaluation, due to the frequency dependence of its distortion mapping. The 3 dB point would fall at the location of the maximum 2nd order product buildup. Unlike the 3rd order situation, this location would depend on the bandwidth, center frequency and spacing of the tones.

Second order distortion patterns consist of two bands of products, each shaped as a triangle with the peaks above and below the frequencies of the original multitones due to the singular addition or subtraction of the tones [8]. The maximum number of products corresponding to the 3 dB point mentioned earlier occurs at the peak of the lower triangle, whose frequency is a function of the spacing of the multitones. This peak occurs at DC. It is necessary to identify the frequency of interest in the triangle to find the actual level of the 2nd order distortion contribution. For a 5-40 MHz return path bandwidth, the mid-band multitone 2nd order distortion level is approximately the level of the initial two-tone distortions and is a function of both the lower *and* upper product triangles. Thus, for analysis that targets midband 5-40 MHz performance, it is sufficient to use the two-tone distortion level for this portion of the NPR contribution.

NPR Mathematical Modeling

The procedure for generating NPR estimates involves evaluating the noise and distortion performance of the system. The NPR curve itself is a measure of carrier-to-noise-plus-nonlinear distortion ratio, C/(N+NLD), where NLD in this cases considers both that due to the optics and the intermodulation distortion (IMD) of the RF nonlinearity. For NPR, the desired signal, C, is represented by a noise source.

The overall C/N must be found to characterize the link noise performance. Noise parameters may be specified or measured on the individual components of a link or on an entire link. Components of the link may have a guaranteed noise performance. However, depending upon the component, the noise parameter used to characterize the device may vary. The optical/RF transducers are the obvious examples of such components. But an equivalent NF can always be generated, and can be determined directly from the circuit design if the design is known. Following this conversion to NF, cascaded NF analysis common to system analysis can be used. The relationship between C/N of a component or link and its equivalent NF is simply

$$C/N(dB) = Ps(dBW) - \{[10Log(k \cdot 290 \cdot BW)] + NF(dB) + Gain(dB), \quad (1)$$

for signal power, Ps, Boltzmann constant k, and system bandwidth, BW.

For the distortion characterization required to develop NPR, the single second order (SSO) intermodulation distortion (IMD2) is found in order to calculate C/IMD2. Single third order (STO) distortion is also found, to calculate C/IMD3. A key step is the adjustment of SSO and STO values to account for noise signals versus tone analysis by extrapolating two-tone theory to noise-like signals using multitone theory previously mentioned [5]. The resultant C/IMD2 and C/IMD3 are added power-wise to get a composite C/IMD for the RF distortions.

It has been found through modeling and verification that RF distortion products up to third order only is necessary to develop accurate NPR models. As with the noise case, distortion can be measured or specified at the component or link level to use in predicting NPR performance. Cascade analysis applies when information is known at the component level, using well-known cascaded intercept point relationships. Individual intercept points can be found from the following well known two-tone distortion relationships:

$$SSO = IP2 - Ptone \quad (2)$$

$$STO = 2\,[IP3 - Ptone], \quad (3)$$

for per-tone power Ptone. Note that all intercept point references and analysis in this paper refer to the output intercept point of the component or the cascade.

For adjustment between tone distortion and NPR, modeled via multitone analysis, the midband C/IMD3 becomes about 8 dB worse as previously described (actually 7.7 dB). By contrast, for this case, the second order midband response does not change. However, this is not the case across the entire, multi-octave return. This can be seen in modeling and measuring NPR versus frequency.

When there is an optical link, as in the HFC case, the clipping C/NLD and C/N_{iin} (Carrier-to-noise due to interferometric intensity noise), if necessary, are included. The C/NLD contribution can be calculated [2][6][7]. The NLD due to clipping does account for non-negligible higher order distortion products that this mechanism can generate. The total NLD due to clipping is a function of the rms optical modulation index (OMI) used to drive the laser. C/NLD can be approximated as

$$C/NLD = 10\mathrm{Log}\{[(.4U^3)/(1+6U^2)]\exp(-1/2U^2)]\}, \quad (4)$$

where

$$U = (\text{OMI per ch})\cdot \sqrt{(N/2)}.$$

The OMI is, of course, tied to the RF drive level at the point in the cascade where the link converts the RF to linear optics. Thus, as with typical distortion contributors, C/NLD varies as a function of RF drive level just as C/IMD2 and C/IMD3 do.

For optical networks that use DFB lasers and go through appreciable link lengths (>10 dB loss, for example), it may be necessary to include the interferometric intensity noise generated by the laser/fiber interface. The procedures and equations necessary to calculate this effect, and the theory behind them, is also available [1][4].

The RF and optical distortions are combined to generate the total distortion power. The noise and distortion contributions create a composite C/(N+NLD). This process is repeated over a range of RF levels at the front of the cascade to sweep the complete NPR curve.

A library of component models and their various noise and distortion parameters allows prediction of NPR performance across a variety of architectures.

Bandwidth and Frequency Dependence

This paper concerns itself primarily with the estimation of NPR at the mid-band of a 5 to 40 MHz return band. However, NPR modeling estimations are in fact frequency dependent. This is largely due to the product mapping discussed previously of the 2nd and 3rd order distortions. The frequency dependence is secondarily due to the possible frequency variations associated with the numerous noise contributors.

The 3rd order product mapping results in a bell-shaped response with the peak occurring at mid-band of the tone generating frequencies. The distortion level roll off at the edges of the generating band has been investigated previously [8][9] and has been shown to be approximately 1.76 dB lower than that of the mid-band response.

As shown in [8], the 2nd order products resemble two triangle patterns that may or may not fall partially in the band of interest. The location of the additive products for the NPR models will begin at twice the starting frequency and end at twice the stop frequency with the maximum density occurring at the start plus stop frequencies. The subtractive products are found from DC (the triangle is completed in the negative frequency realm) to the stop frequency minus the start frequency. As can be seen for the 5-40 MHz return, because of the multi-octave nature of the passband, some resultant 2nd order products do fall within the generation band and should be included in any estimation of NPR.

There may be instances when the products fall outside the band of interest, such as a forward path with a link dedicated to carrying digitally modulated RF only between 550-870 MHz. Under such circumstances the NPR may be estimated without the contributions due to 2nd orders. It must be kept in mind that, while this may indeed be true in evaluating the particular digital channels, the out-of-band 2nd order products can cause interference problems with standard forward analog channels. Unless filtering is applied, this may be the case when the analog channels are located below the digital at a point where the transport link becomes a composite signal load of analog and digital.

For optical networks there would also be secondary frequency dependent effects to NPR that are due to the laser and fiber phenomena themselves as well as the interactions between the two. Both interferometric intensity noise (iin) and phase to intensity noise are frequency dependent and would have to be evaluated as such in any widely swept frequency analysis of NPR. Typically, phase to intensity noise is a negligible effect and can be ignored for modeling purposes. Finally, practical cascaded systems can also incur a thermal noise floor variation over frequency.

Analysis vs. Measurement

Using the modeling approach discussed, we can compare estimated NPR performance of various return path systems with measured results.

The first system consists of an unisolated Fabry-Perot (FP) 0.4 mW laser transmitter operating through 25 km of fiber into an optical receiver with a 9.7 pA/√Hz noise current. Optical loss associated with the 25 km of fiber at 1310 nm was measured at 9 dB, close to the expected .35 dB/km. The transmitter is a Motorola model SG1-FPT, one of a family of node transmitters available for return path applications. The link terminates into a Motorola return path optical receiver, model AM-OMNI-RPR/2C. The signal loading for the transmitter consists of 35 MHz of noise (5-40 MHz) at total power levels from 7 dBmV to +30 dBmV. The recommended operating total signal level for the transmitter is +20 dBmV.

Measurements of the FP diodes used in the transmitter show a typical relative intensity noise (RIN) of -132 dB/Hz. This is an equivalent RF NF of 42.8 dB for an efficiency of 0.1 W/A. When modeled with the RF portion of the transmitter, this resulted in an overall noise figure of 26.6 dB. Two tone distortion data for the diodes showed average 2nd order levels of –53.4 dBc and 3rd order levels of -64.6 dBc. After cascading this with the RF portion of the transmitter, the resultant overall levels dropped to -51.7 dBc and -64.5 dBc respectively. Based on these numbers, the calculated intercept points were +65.1 dBmV for 2nd order (IP2) and +45.7 dBmV for 3rd order (IP3).

The optical receiver RPR/2C uses a photodiode with a typical responsivity of 0.85 A/W. The internal gain control was set for the mid-gain of it's 20 dB range, resulting in the measured EINC of 9.7 pA/√Hz. This corresponds to an equivalent noise figure of 4.5 dB at an optical input of -13.0 dBm. Typical measured 2nd and 3rd order levels of the receiver at 0 dBm optical input and 20% OMI per tone at the transmitter are

-65 dBc and -75 dBc respectively. This results in an IP2 of 117.0 dBmV and an IP3 of 89.5 dBmV.

The transmitter and receiver were cascaded through the 25 km fiber (RF loss of 18 dB) and the overall electrical parameters for the chain were generated for use in the NPR estimations. The link is shown in Figure 2.

Choosing a single point of operation, for an input to the transmitter of 20 dBmV, the resultant C/N, distortions, and final NPR were found as follows:

C/N = 40.5 dB
IP3 = 64.3 dBmV; C/IMD3 = 56.8 dB
IP2 = 83.5 dBmV; C/IMD2 = 51.5 dB
C/NLD = 70.4 dB
Total NPR @ 20 dBmV = 40.1 dB.

Now, this being the recommended operating point, it is expected that linear operation would be the situation here, as can be seen by the dominance of C/N to the resultant NPR. Of course, for a complete NPR curve, the above calculation is performed at every drive level. The modeled mid-band NPR curve based on the cascaded NF and intercept points added to a carrier-to-nonlinear-distortion C/NLD curve of the clipping generates the overall modeled NPR curve shown in Figure 3. The example link was then measured for NPR using the Noise-Com automated NPR Test Station. These results are also shown in Figure 3 for comparison. The modeled versus measured data show good correlation.

A second return system utilizing a 1 mW distributed feedback (DFB) laser diode was also evaluated. The transmitter used was a 1 mW DFB, Motorola model SG2-DFBT. The RPR/2C was again used as the receiver. By using the same procedures as outlined for the SG1-FPT, the overall cascaded electrical results for this system are shown in Figure 4. The resultant model and measured NPR curves are shown in Figure 5. Again, the plots show good correlation. In addition to the curves generated by the equivalent RF electrical parameters and the diode clipping, the overall model for the DFB link also includes the interferometric intensity noise (iin) contribution due to the scattering effects of the laser/fiber interface.

As the return path is slowly transitioned from analog to digital, the ability to estimate the NPR performances of systems utilizing analog-to-digital converters (A/D's) becomes important. This technique can also be used to incorporate the A/D's quantization and saturation noise performance into an NPR response for a complete digital link. Figure 6 shows the modeled and measured performances of a 10-bit A/D digital return system. A 10-bit system has roughly the same performance as linear optics with DFB lasers. A full discussion of the theory behind evaluating digital links and their analogy to analog transmission types may be found in [3].

Conclusion

NPR has become a valuable too for characterizing return path performance capability. It provides, in one easy snapshot, both noise and distortion characteristics of a complete link. In addition, it is a simple test to setup and make measurements on. The drawback with NPR since it has come into the realm of CATV has been that system designers have been constrained to some extent by having to rely on measurements. Measurement requires time and resources each time a system design verification is desired. Simulations put the designer at the mercy of the programming code and the potential for inaccuracy in dealing with nonlinearity. In this paper, we have presented a mathematical approach extrapolated from the well-understood distortion analysis of both RF and optical systems. In this approach, the two subsystems can be integrated in a single spreadsheet analysis that can accurately predict NPR. This provides system designers the tool they need to characterize new designs and new architectures, develop performance specifications to the component level, and yield deeper insight into the mechanism of noise power distortion effects.

References

[1] Darcie, T., Bodeep, G., Saleh A., "Fiber-Reflection Induced Impairments in Lightwave AM-VSB CATV Systems," IEEE Journal of Lightwave Technology, 1991 9(8).

[2] Frigo, N. and G. Bodeep, "Clipping Distortion in AM-VSB CATV Subcarrier Multiplexed Lightwave Systems," IEEE Photonics Technology Letters, July 1992, vol. 4 no. 7.

[3] Howald, R., Advancing Return Path Technology….Bit by Bit, 1999 NCTA Technical Papers.

[4] Judy, A.,"Intensity Noise from Fiber Rayleigh Backscatter and Mechanical Splices," Proc. 15th European Conference on Optical Communications, 9/89.

[5] Leffel, M., "Intermodulation Distortion in a Multi-Signal Environment," RF Design, June 1985.

[6] Phillips, M. and T. Darcie, "Numerical Simulation of Clipping-Induced Distortion in Analog Lightwave Systems," IEEE Photonics Technology Letters, Dec. 1991, vol. 3 no. 12.

[7] Saleh, A., "Fundamental Limit on Number of Channels in Subcarrier-Multiplexed Lightwave CATV System," Electronics Letters, June 1989, vol. 25 No. 12.

[8] Some Notes on Composite Second Order and Third Order Intermodulation Distortions, Matrix Test Equipment, Inc., Application Note MTN-108.

[9] The Relationship of Intercept Points and Composite Distortions, Matrix Test Equipment, Inc., Application Note MTN-109.

Acknowledgements

The paper presented here is the result of the extensive efforts of several key Motorola BCS employees, including Dean Stoneback, Dave Ciaffa, Vipul Rathod, Mike Short, and Ricardo Guevera.

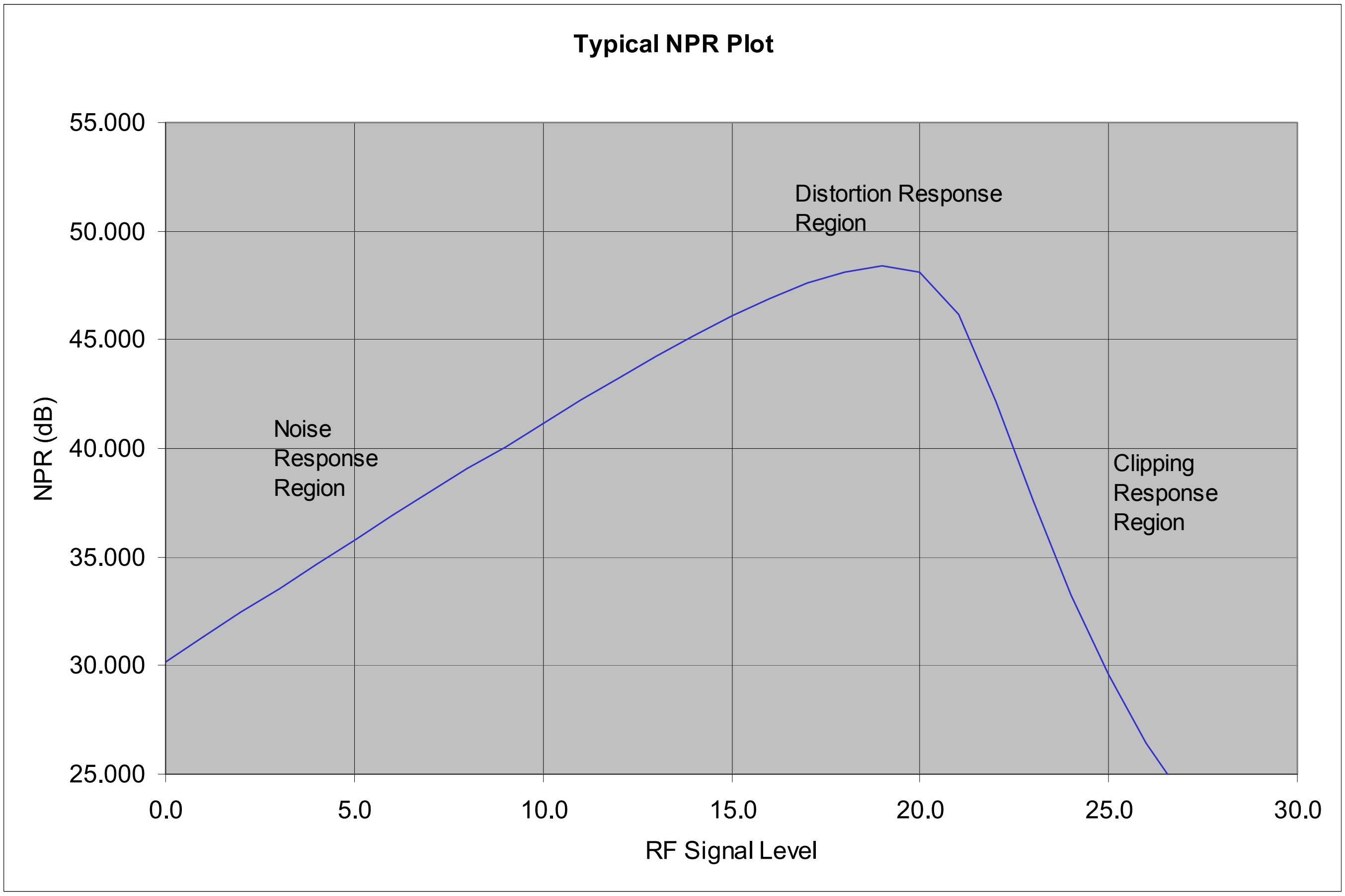
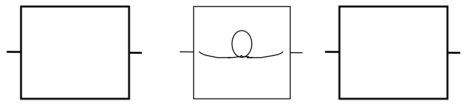


Figure 1 – Typical NPR Plot



	SG1-FPT (FO TX)	Fiber Loss (RF)	RPR/2C (FO RX)	Cascade Total
Gain (dB)	-3.59	-18	36.59	**15.00**
NF (dB)	26.55	18	4.47	**29.31**
Output 1 dB Comp dBmV	300	300	300	**299.94**
Output IP3 (dBmV)	45.68	300	89.5	**64.26**
Output IP2 (dBmV)*	65.06	300	117	**83.47**

Figure 2 – Cascade Analysis for a 9 dB Fabry-Perot (FP) Return Link

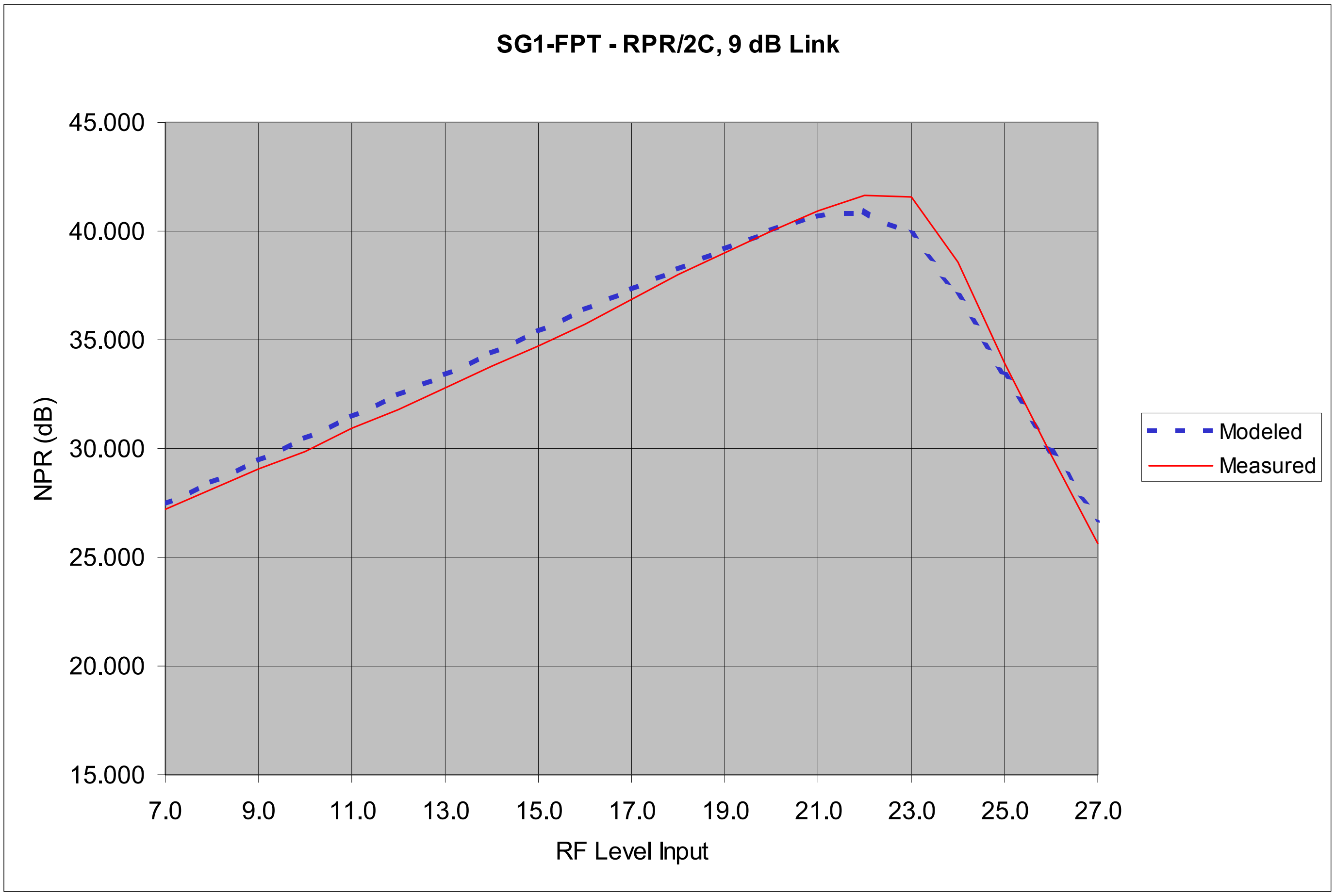
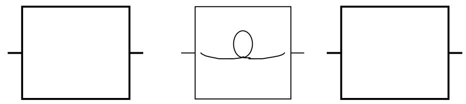


Figure 3 – NPR Plot for a 9 dB Fabry-Perot (FP) Return Link



	SG2-DFBT	Fiber Loss (RF)	AM-OMNI-RPR/2C	**Cascade Total**
Gain (dB)	5.51	-18	36.59	**24.10**
NF (dB)	6.26	18	5.23	**18.00**
Output 1 dB Comp dBmV	300	300	300	**299.94**
Output IP3 (dBmV)	55.51	300	89.5	**73.98**
Output IP2 (dBmV)*	81.16	300	117	**98.63**

Figure 4 – Cascade Analysis for a 9 dB Distributed Feedback (DFB) Return Link

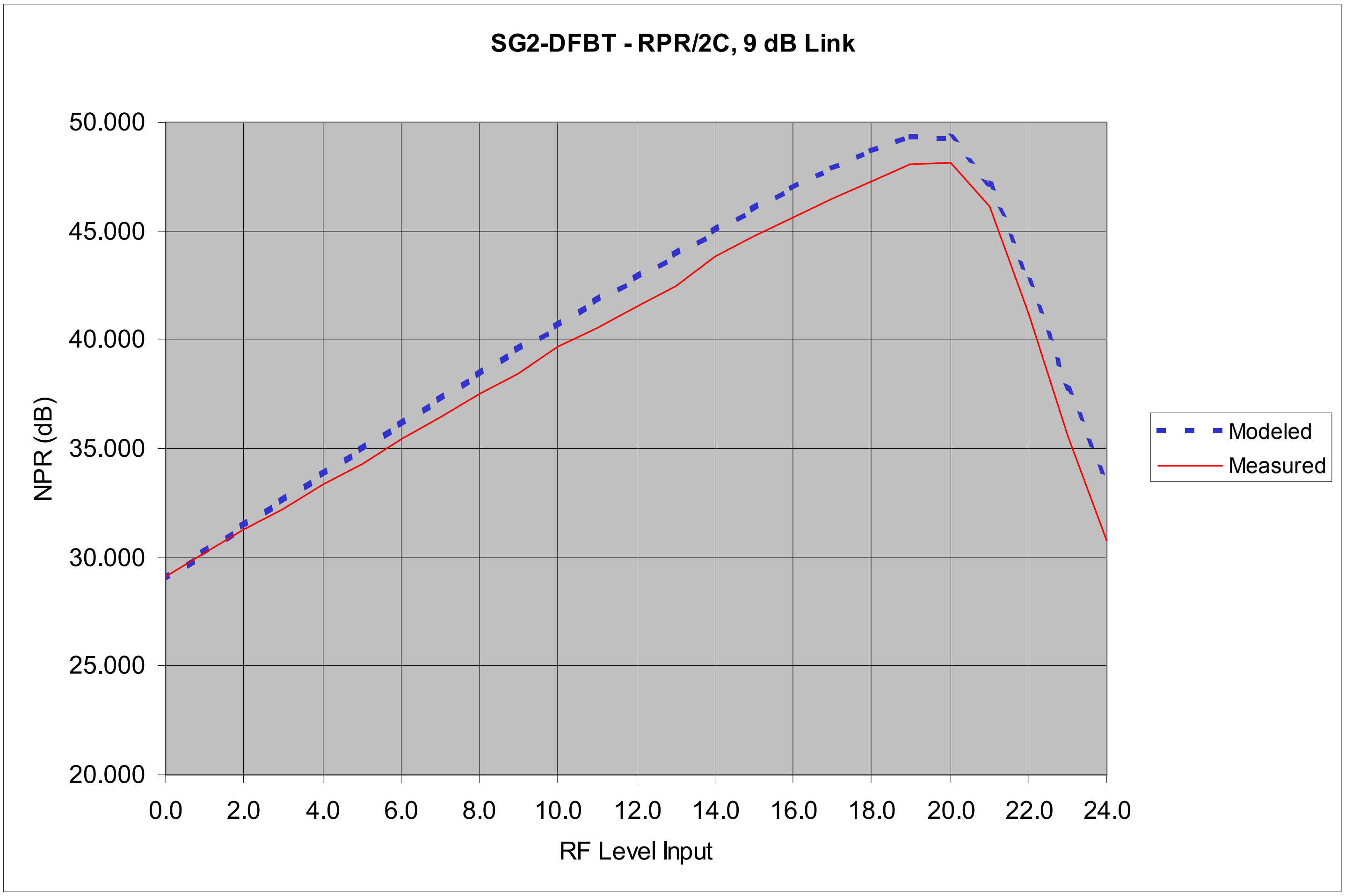


Figure 5 – NPR Plot for a 9 dB Distributed Feedback (DFB) Return Link

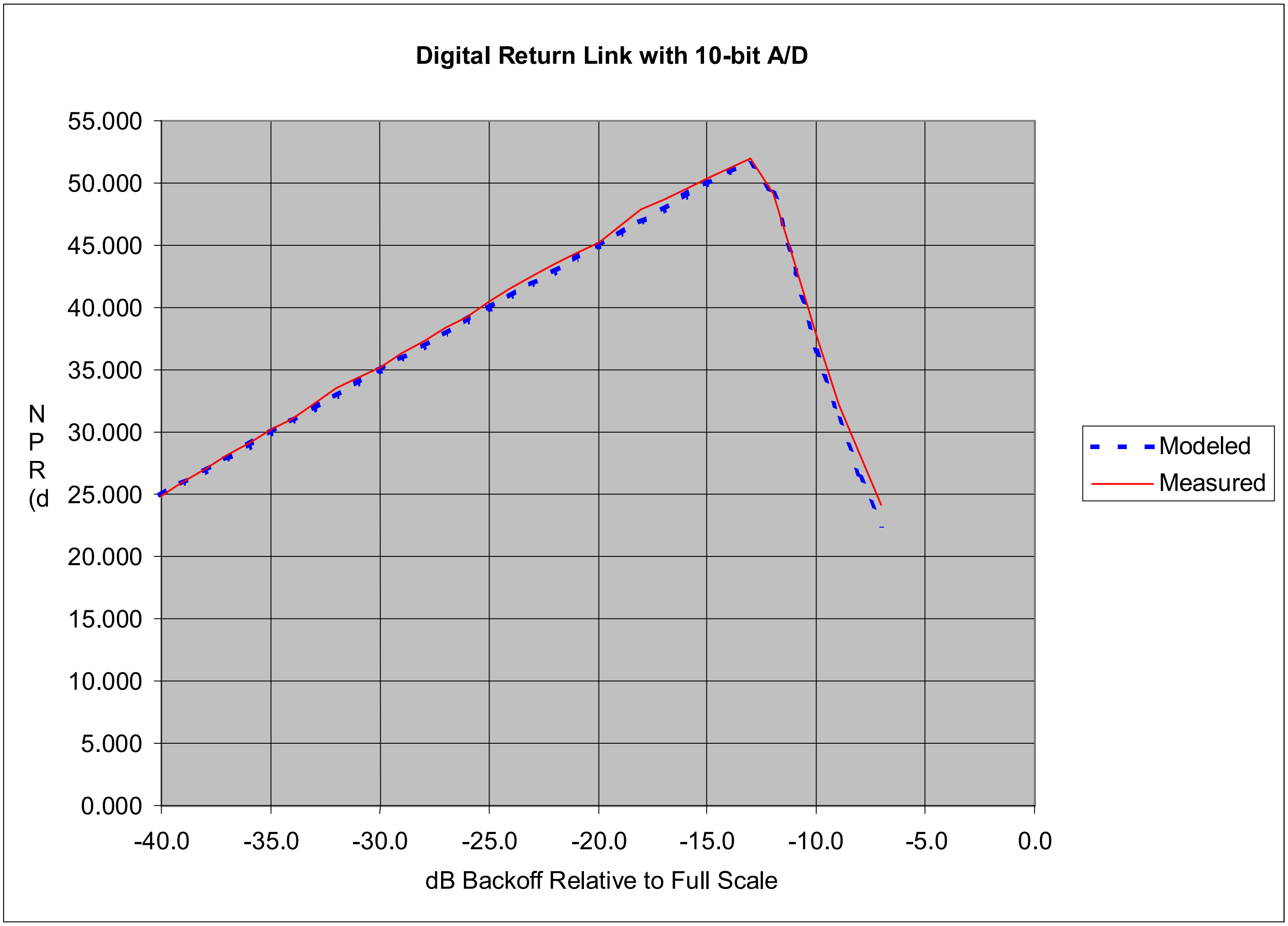


Figure 6 – NPR Plot for a 10-bit Digital Return Link

Open Middleware and the OpenCable POD Module: Versatile Solutions for Portable, Secure Digital TV

Anthony J. Wasilewski
Scientific-Atlanta, Inc

Abstract

Two aspects of a secure, portable environment for interactive DTV services are discussed. The OpenCable POD and an open middleware approach can be useful tools on "the road to retail".

The OpenCable Point of Deployment (POD) module is an important component of the open standards specifications for digital TV. The POD supports the total separation of the conditional access system from the host terminal while still supporting a wide array of features and applications and providing high performance video/audio/data services. By allowing host terminals to become more generic, the POD may play a critical role in any transition to retail availability of set-tops.

This paper will cover the following aspects of the POD:

- Brief History/Origins
- Regulatory Issues and Timelines
- Relevant Standards & Documents
- Architecture and Features
- Interface Descriptions
- Copy Protection

To complete the support of an open platform for hosting of a rich set of portable services, the application software environment of the host also needs to be standardized. The goal is to provide an environment in which a large measure of freedom to craft applications with varied feature sets and powerful graphics exists, while also fostering a high degree of portability of those applications to different hardware platforms.

To this end, this paper describes an open set of middleware that includes:

- HTML
- Javascript (ECMAScript)
- MIME
- Personal Java
- XML
- HTTP
- SSL
- DOM
- XHTML
- ATVEF

These middleware components can be and/or are deployed on existing digital set-tops in currently launched systems.

<u>POD History/Origins</u>

The concept of using a removable device to encapsulate all security and conditional access processing has origins in both the DVB (Digital Video Broadcasting) Common Interface process in Europe and the NRSS (National Renewable Security Standard) effort carried out as a joint engineering committee of CEA and the NCTA under EIA sanction in North America. Both groups eventually adopted a PC Card (PCMCIA) form factor although the NRSS specification also includes an extended ISO 7816 smart card format as an additional choice.

The POD extends both the DVB and NRSS standards by adding:

- explicit handling of out-of band (OOB) data channels
- a copy protection mechanism
- an application interface
- extensions for cable-ready applications

Regulatory Issues and Timelines

From a regulatory viewpoint, the major influences on the existence of the POD have been:

•The Telecommunications act of 1996 requires that cable subscribers be given the option of owning the equipment required to receive cable services

•The FCC's Report and Order (63 Fed. Reg. 38095) requires that cable operators make separable security modules available by July 1, 2000 in order to facilitate commercial sale of navigational devices

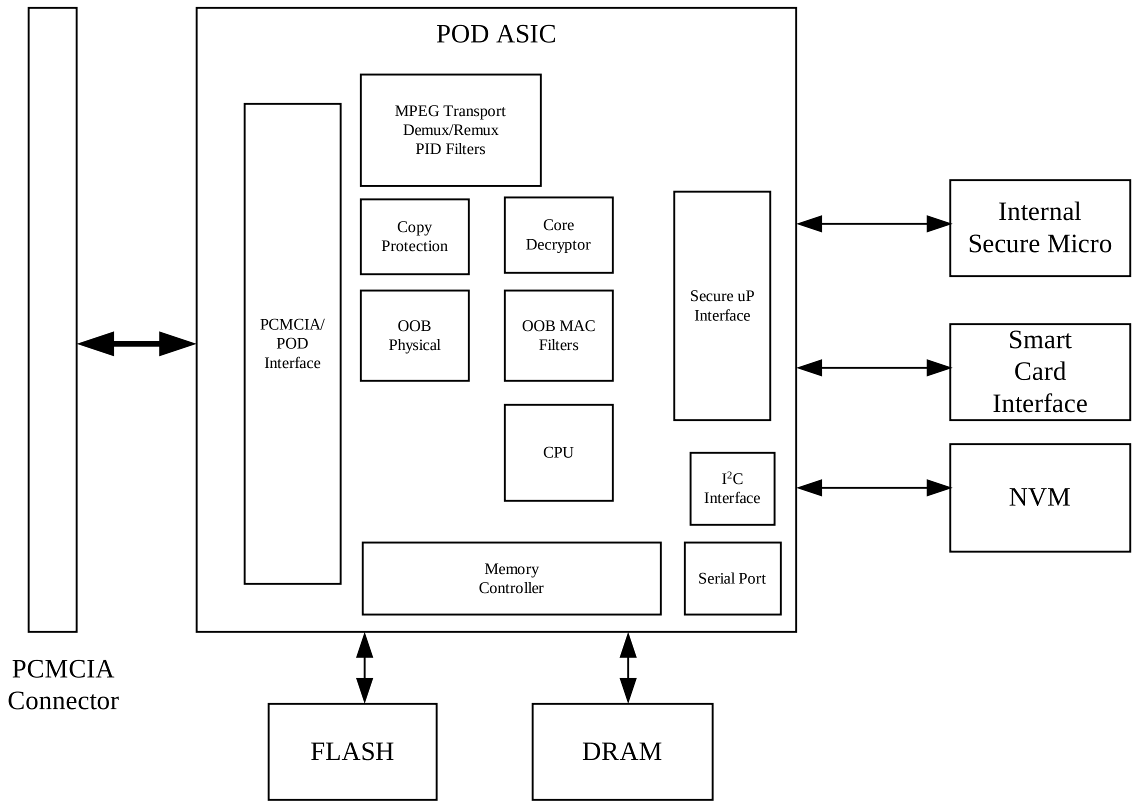
The industry responded by including in the CableLabs® OpenCable® process, a working group to define the functionality for a Point-of-Deployment or POD module and its interfaces and also to specify a suitable copy protection method that would be acceptable to content owners. This work has progressed well and manufacturers are responding to the challenge in a manner that should result in suitable product being available in the FCC-mandated time frame.

POD Architecture and Features

The POD handles both in-band MPEG Transport and Out-of-band (OOB) control channels on behalf of the host device. One of its primary responsibilities for MPEG processing is the conditional access-level decryption of subscriber-selected content for which the subscriber is authorized. The POD accomplishes this in conjunction with the CATV headend through the conditional access system that is implanted within it. For the OOB channel, it interprets the format of the control bit stream that has been sent to it from OOB RF receivers in the host.

The POD must support five different interfaces: PCMCIA, MPEG Transport, an out-of-band channel (either DAVIC or Motorola (formerly GI)), the POD data channel and the POD extended channel. It must support traditional applications such as digital broadcast and IPPV but must also be capable of other OpenCable services such as VOD. It must also be able to support the man-machine interface (MMI) of host applications utilizing a graphical interface based on HTML 3.2.

A basic block diagram for a POD module is shown below:



To fit the amount of functionality the POD specifications require into a small form factor such as Type II PCMCIA, considerable integration in silicon is called for. Thus, much of the major components in the POD block diagram above are found on an ASIC. This chip, of course, has interfaces to memory and other chips such as secure microprocessors to complete its mission.

Interface Descriptions

The following are the classes of interfaces supported by an OpenCable POD module:

- PHY
- Extended Channel
- Link Interface
- Application Interface

The *PHY or physical interface* is compliant to the 68-pin PCMCIA interface. This supports parallel, full-duplex transmission of MPEG-2 Transport streams at the bit rates typically deployed in North American CATV systems. There is also support for signaling and CPU-to-CPU communication. Upon power-up, the POD performs the standard 16-bit PC Card Memory Only initialization, after which the POD and Host activate the "POD Module Custom Interface" which has a registered interface ID of hexadecimal 341. The POD also follows PC Card power management standards.

The *Extended Channel* provides a data path between the POD and Host for information flows outside the MPEG-2 Transport that are not terminated by the POD. Thus, for example, it supports the flows of IP packets or MPEG sections that have arrived for the Host via out-of-band (OOB) pathways from the headend. Some data, such as Entitlement Management Messages (EMMs) from the CA system, arrive over the OOB pathways and are not forwarded by the POD to the Host.

The POD *Link Interface* is compliant with the Command Interface of the NRSS Part B specification. This implements a set of protocols that establish communication about and to *resources* relating to conditional access, host control, the man-machine interface (MMI), copy protection, generic IPPV support and the Extended Channel. The link interface takes care of identification of flows and Protocol Data Unit (PDU) fragmentation.

The *Application Interface* is used to support "cable-ready" applications that use the data channel that are not defined or adequately covered by NRSS Part B. Such functions include Host Generic Feature Control, POD Emergency Recovery and specific application support. This interface defines the POD/Host MMI, additions to the low-speed communication interface, additions to the host control resource, additions to extended channel support, modifications to the generic IPPV resource and specific application support for hosts that have software download capability.

Copy Protection

Because the POD processes (decrypts) and passes digital content streams to the host, suitable copy protection is required. If these content streams were sent "in the clear" over the PCMCIA interface to the host, it would be relatively straightforward to make exact digital copies of this content. Thus, to protect the bit streams as they flow from POD to host, the POD must apply additional encryption. While this is fairly simple to do, the POD is also required to authenticate the host to verify that it has not been previously identified as an illegal device. This is accomplished through the use of digital certificates and signatures. In this manner, a highly secure form of key exchange may also be practiced.

Once the POD and Host have agreed on keys to be used, the POD encrypts the content of the MPEG-2 Transport packets that require copy protection as signaled in CCI (copy control information) bits which are sent in authenticated form via the conditional access system.

Open Software Environment

The POD module only provides part of the solution for a portable application and content environment. There must also be a method that provides content and application interoperability.

An open software environment relies on published standards that enable portability between platforms. If we begin with that assumption that the TCP/IP protocol suite will be the foundation for messaging, client-server

TCP/IP protocol suite will be the foundation for messaging, client-server applications and data access and that MPEG video and transport and Dolby AC-3 audio are the foundations for transmission of entertainment content, then portability concerns need to be focused on other aspects of the system. There are at least three other major areas of support needed to enable the desired portability:

a) Content rendering
b) Application code execution
c) Network protocols

These areas can be covered by an architecture that includes standardized elements for the following functions:

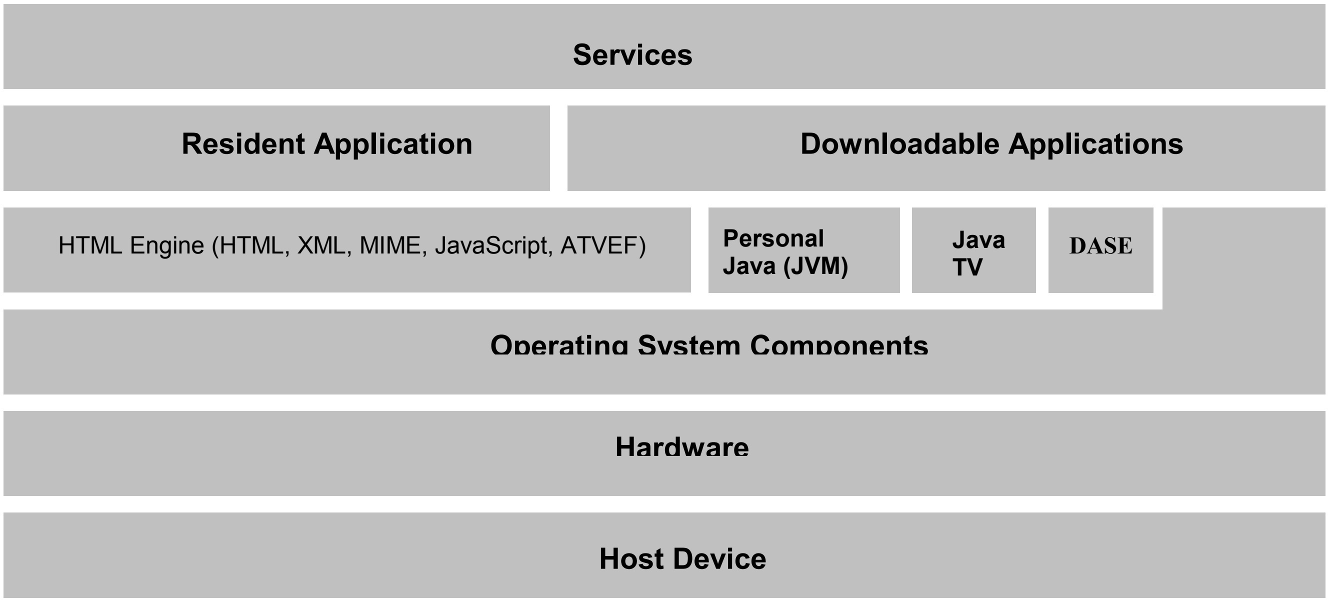
- Presentation Engine
- Application Engine

In addition, the platform would require elements that provide the following functionality on behalf of applications:

- Network services
- Platform services

Ideally, these network and platform services would be supplied via an *operating system* that, in conjunction with the middleware, completely abstracts the details of the underlying hardware platform and network and permits applications to migrate between platforms with little or no modification. The operating system may or may not be a standardized element, however, it is likely that, within any one MSO system, only one operating system would be deployed.

A layered model of an interoperable software environment for a Host device is shown in the figure below. In the figure, the *middleware* components are located between the applications and the operating system (i.e., in the "middle", hence the term "middleware"). The foundations of the middleware, are the *presentation engine* components: HTML, MIME, ECMAScript, ATVEF, DOM) and the *application engine:* PersonalJava. Examples of *extensions* to the middleware are the Java.TV classes and the ATSC Digital TV Application Software Environment (DASE) which has been defined by the S17 Group of ATSC.



The *multi-purpose Internet mail extensions (MIME)* provide useful standards for content rendering such as JPEG and *Portable Network Graphics (PNG)* and audio standards, such as WAV and AIFF.

The *Document Object Model or DOM* is a platform and language-neutral model that allows programs and scripts to dynamically access and update content. The DOM provides a map of a document's structure and style and supports generic access to its parts. Combining DOM0 with ECMAScript is equivalent to Javascript 1.1.

The Advanced Television Enhancement Forum or ATVEF is a cross-industry alliance of companies that have defined protocols for HTML that can be used to deliver enhanced programming over many transports to a range of intelligent receivers. These enhancements include announcement protocols, trigger handling for real-time events and a local identifier URL scheme.

XML or *extensible markup language* is a very promising addition to the markup language approach started in HTML. XML supports the separation of the definition of the data from the description of how it should be displayed. This promotes more dynamic content as well as providing strong portability since the rendering can be specified for TV-based graphics, print-oriented graphics, speech synthesis or even *Braille* without changing the content coding at all. It also supports domain-specific data definitions that can be used to formulate standardized formats for specific types of data, so that all applications can use and interchange the *same* data.

Further capitalizing on this separation of content definition and rendering is XHTML (*Extensible Hypertext Markup Language).* XHTML is a reformulation of HTML 4.0 as an application of XML 1.0. It has the advantage of being easily extensible, which allows applications to be updated with relatively little effort and it is designed to be highly portable, so that content can be transferred to many diverse platforms (such as PCs, cell phones, PDAs, TVs, etc.) and be acceptably displayed without modification.

The *Application Engine* is a complete execution environment within itself. One example is PersonalJava® or pJava, one of the Java® application environments developed by the Javasoft division of Sun Microsystems, Inc. Java is implemented on a Host device as a virtual machine. This is a software program that executes byte codes, which are standardized instructions for the machine. As long as a Java Virtual Machine (JVM) is available for a platform, applications written in the Java language can be readily ported to that platform. The JVM also provides a security framework to ensure that "renegade" applications do not wreak havoc on the host platform.

The *operating system components* supply the interface (and abstraction of) the *hardware components* of the Host device. Thus, applications and the middleware layer need not concern themselves with the details of different types of tuners, on-screen display graphics drivers, conditional access elements, and other vendor-specific components. Also, the operating system must provide a robust event model and the facilities to handle the event-driven aspects and requirements of applications. These include display focus, inter-process messaging, timers, semaphores, memory management and the like.

Finally, the operating system must also support important networking protocols such as TCP/IP, HTTP and SSL. Hypertext Transfer Protocol is one of the foundations of the World-wide Web and the Secure Sockets Layer has rapidly become the *de facto* Web protocol for securing communications between clients and servers.

Conclusion

Critical parts of the support of the portability required to support retail availability of Host devices in CATV systems are supplied by the OpenCable POD module and by open approaches using standard middleware. Combined with a robust operating system that abstracts the details of the platform hardware and provides system services to the middleware and applications, the basic foundation of interoperability can be formed. The POD module supplies complete separation of conditional access functions from the Host device and copy protection functions that are acceptable to the content industry. An open middleware approach provides additional standardization and abstraction for content rendering and the application execution environment.

Using this foundation, application developers can produce support for new services, confident that their efforts will be applicable to a wide range of Host platforms.

References and Standards

The following documents are some useful references that provide additional reading and information regarding the OpenCable POD and open middleware environments and standards:

1	Document markup language HTML 4.0: http://www.w3.org/TR/REC-html40/
2	Document scripting language ECMAScript: http://www.ecma.ch/stand/ecma-262.htm
3	Document Object Model DOM Level 0: http://www.w3.org/DOM
4	Hypertext Transfer Protocol (HTTP) 1.1 (RFC 2068): ftp://ftp.isi.edu/in-notes/rfc2068.txt
5	Aggregation & encoding of multiple resources into a single resource for delivery: MIME multipart/related: http://info.internet.isi.edu/in-notes/rfc/files/rfc2387.txt MIME HTML (rfc2110): ftp://ftp.isi.edu/in-notes/rfc2110.txt
6	Extensible Markup Language (XML) 1.0 Specification, http://www.w3.org/TR/REC-xml
7	OpenCable Host-POD Interface Specification, IS-POD-131-INT01-991027
8	NRSS Part B Specification, EIA-679-A, Part B
9	OpenCable POD Copy Protection Specification: IS-POD-CP-INT01- 000107
10	Java Specification: http://www.javasoft.com/aboutJava/communityprocess/maintenance/JLS/index.html
11	ATVEF Specification: http://www.atvef.com/library/spec1_1a.html

Author's Contact Info:
Tony Wasilewski
5030 Sugarloaf Parkway
P.O. Box 465447
Lawrenceville, GA 30042
Phone: 770-236-5004
Fax: 770-236-3080
E-mail: tony.wasilewski@sciatl.com

Opencable – Opening Soon

Don Dulchinos
Cable Television Laboratories

Abstract

An overview of the OpenCable project is presented, from its inception to the current status. The OpenCable process is described, along with its success in enabling the participation of multiple MSOs, traditional cable vendor companies, and non-traditional cable vendors from the consumer electronics industry. Recent interoperability testing is described which suggests that several vendor companies are likely to be certified OpenCable compliant in the next few months. Finally, some thoughts are offered on future developments.

OPENCABLE OBJECTIVES

OpenCable was conceived and launched in the summer and fall of 1997 in response to a number of industry and technology trends that were then working both for and against the interests of cable television system operators.

At the time, cable companies had placed orders for digital set-top boxes but were still waiting on delivery of boxes at price points that made economic sense. Part of the delay was deemed the result of a lack of sufficient competition among suppliers.

It was further noted at the time that the computer industry was riding extremely steep declining cost curves into an era of new growth and increased innovation. Further, the advent of the World Wide Web pointed to real consumer acceptance of interactive services.

A series of meetings coordinated by CableLabs between cable CEOs and Silicon Valley CEOs led to a belief that the convergence of the industries held the key to cable's future. At the same time, cable executives were wary of handing over the keys to this new engine of growth to another limited set of vendors. Thus, OpenCable was born.

One additional OpenCable objective is worth noting. Due to a clause in the Telecommunications Act of 1996, a requirement was placed on cable companies to enable retail distribution of "navigation devices". OpenCable was shaped in order to meet this requirement, not only for regulatory purposes, but also to enable cable companies to shift equipment investment off their books. Although the financial need to do so receded as cable company finances improved (in part due to investments in MSOs by Silicon Valley companies), the appearance and growth of direct broadcast satellite competition led to a need for cable companies to facilitate a competitive response in retail consumer electronics outlets.

OPENCABLE PROCESS

The cable industry announced an RFP for a design for next generation digital set-top boxes and other devices. The end result of the RFP process was not a selection of a single vendor solution, but a realization that there was enough commonality in the approaches for the cable industry to seek to drive the creation of an open specification.

The OpenCable process is modeled on the highly successful DOCSIS, Data Over Cable

System Interface Specification, which has to date certified cable modems from 17 different companies. Some key aspects of the DOCSIS process are directly relevant to OpenCable:

- A commitment to an open, collaborative process.
- Inclusion of vendors as specification authors and true partners.
- A neutral venue for development work.
- A feedback loop between equipment development and refinement of specs.
- And building a strong consensus within the cable industry on cable's requirements.

The first step was the development of a specification outline, as illustrated in Figure 1.

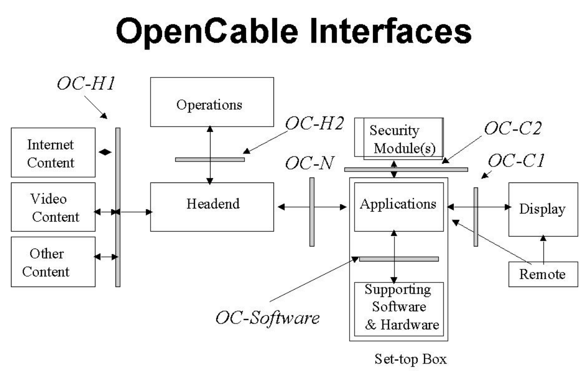


Fig. 1

The next step was to solicit vendor authors, who ultimately came from such diverse companies such as Sony, SCM Microsystems and Time Warner Cable. Authors prepared draft documents, and then submitted them for review by cable MSO representatives. Once a cable industry consensus was reached, documents were put out for further review by all interested vendor companies who signed a non-disclosure agreement (promising not to publish draft documents.) There are currently over 400 vendor companies that are involved in this process. Finished specifications are then submitted to SCTE and other standards organizations for due-process standardization.

1394 INTERFACE WITH 5C COPY PROTECTION

The interface between the set-top box and the consumer device, such as a television, turned out to be a point of much controversy. Hollywood studios expressed to cable companies their concern that a digital interface between a digital set-top and a digital television meant the appearance of a digital video stream in the clear, a tempting target for pirates who would then have a pristine copy for further duplication.

Cable, Hollywood and consumer electronics companies settled on a 1394 interface with so-called 5C copy protection as a viable solution. This technology was written into the OCI-C1 specification, also known as the Home Digital Network Interface, shown in Figure 2.

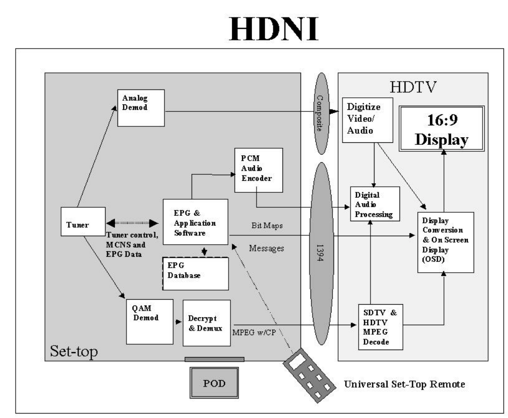


Fig. 2

POD MODULE

A second area of the OpenCable specification that occupies a central role is the point-of-deployment security module, or POD module.

During FCC deliberations on how to implement the “navigation device” provision of the 1996 Telecommunications Act, the concept of a removeable security module gained favor and was eventually written into the FCC's order implementing the provision.

The POD module was modeled on the NRSS-B standard for renewable security. The form factor is a PCMCIA card, chosen for renewability as well as being considered incrementally more secure than smart cards that were the basis of security in Europe. (See Figure 3.) PCMCIA cards also have the processing capability to handle complex out of band signaling protocols.

Fig. 3

INTEROPERABILITY TESTING

Following the successful process pioneered in the DOCSIS program, CableLabs began holding interoperability events as a way to assist developers in building products that could meet the OpenCable specification and eventually be certified.

The first interop event was held in July, 1999, and the second in December 1999. The December interop was then transported onto the exhibit floor at CableNET '99 at the Western Cable Show to offer industry observers the opportunity to see the progress of the effort. A third interop event was held at CableLabs in March, 2000. The companies that have participated in interops to date are shown in Figure 4.

<u>Headend</u>

DiviCom
Motorola
Scientific-Atlanta

<u>POD Modules</u>

Mindport
Motorola
NAGRA
NDS
SCM Microsystems (test tool)
Scientific-Atlanta

<u>Host Devices</u>

LG Electronics
MARGI Systems (test tool)
Microsoft/SCM
Motorola
Panasonic
Philips
Samsung
Scientific-Atlanta
Sony
Thomson
Zenith

Fig. 4

The most recent interop covered the following functions:

- Initialization Between POD module & Host
- Display A/V Decrypted by POD module
- Display of Closed Captions from MPEG
- Channel Change
- OOB Messaging between the Headend and POD module via the Host

- SI processed from the Extended Channel
- POD-Host Interface Encryption

These elements are a subset of the full range of features that will be part of the Acceptance Test Plan that will form the core of the OpenCable certification process. The first OpenCable certification wave, which will encompass further interoperability testing as well as certification testing, begins on April 10 and runs through May 19.

Based on the significant progress observed in the successful performance of many of the above items since the last interop in December, it appears that some POD module and host vendors may be certified in the very near future.

FUTURE DEVELOPMENTS

Vendors currently participating in OpenCable interoperability testing at CableLabs are mostly developing set-top boxes. However, one company did work on a television receiver integrating set-top functionality and the POD interface, while another developed an architecture that may allow a personal computer to function as an OpenCable host device. These two products illustrate the flexibility that the OpenCable platform holds for development of innovative products to which cable's broadband delivery system can deliver exciting new services. Figure 5 suggests the variety of devices that might eventually become certified as OpenCable-compliant and provide an array of new services to cable customers.

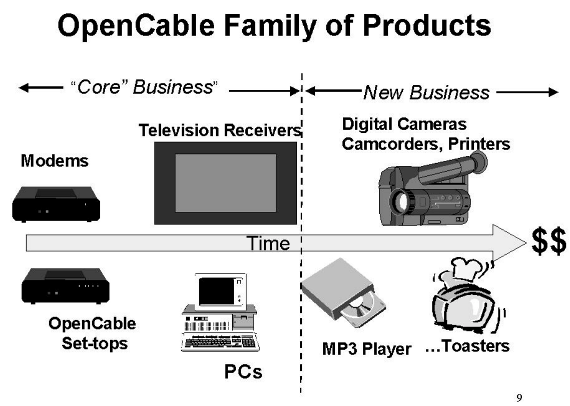


Fig. 5

OpenCable was conceived from the outset as a family of products. In addition to meeting its objectives of increased competition among suppliers and enabling innovation, OpenCable also makes possible a viable retail distribution strategy. Equally important to a retail product offering is the ability to offer not only hardware that is portable between different cable systems, but also services that are portable. To this end, the next phase of OpenCable has begun to tackle specifying a common software environment to which applications and interactive services can be written. This environment will extend the flexibility of the OpenCable design to enable innovation in the realm of interactive services and applications, just as it has done in the hardware space.

Opportunities For Application Service Providers In Cable Networks

Pablo L. Martinez
Lucent Technologies

Abstract

Application rental services take advantage of the "always on" broadband access provided by cable networks. In this service model end users remotely invoke application features from simple thin clients. Applications run on network-centric server clusters.

This paper addresses the technical applicability and revenue opportunities of the Application Service Provider model in cable networks. It considers service models and both network and application architectures. A business case built upon these models is presented.

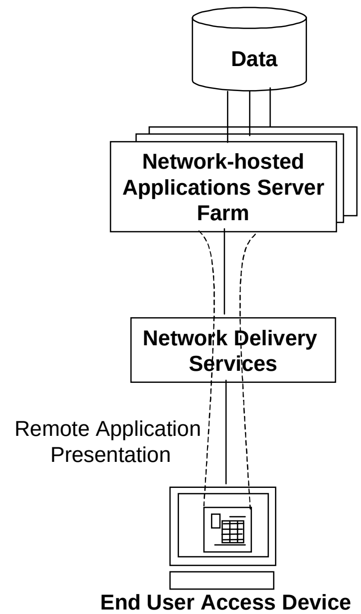


Figure 1. Delivering Applications as a Service

INTRODUCTION

Internet access for e-mail and web browsing is currently driving initiatives to upgrade cable plants to IP-centric platforms. However, the flexibility and ubiquity of IP technologies give Cable Operators the opportunity to offer new innovative services not only to residential customers but also to telecommuters, Small Office Home Office (SOHO), and medium business customers. One example is application rental services, where end users remotely invoke features from applications running on network-centric server clusters. This is illustrated in Figure 1.

This network-centric service model simplifies end user system requirements and maintenance. It provides Cable Operators the opportunity to offer application services that take advantage of the "always on" broadband access that cable networks offer. In this role, Cable Operators become Application Service Providers (ASPs).

This paper provides an overview of the ASP concept and addresses its technical applicability to cable networks. First, the ASP service model is introduced. Second, a generic ASP architecture is described and then mapped to cable networks. The resulting architecture is used to run a high-level business case analysis that shows the revenue potential for Cable Operators. The last part of this paper presents the conclusions.

ASP SERVICE MODEL

The ASP service model is "one-to-many." Applications run in a network-hosted environment to serve a dispersed customer base. Initially, the ASP service model targets the small and mid-market business customer segments. However, the enterprise (i.e., large business) segment is also showing interest. And in the lower part of the spectrum, income-generating home offices (i.e., home-based businesses) may quickly become one of the key segments benefiting from this model.

There are many reasons why the ASP service model is quite compelling. Below is a partial list of these benefits.

1. Access to enterprise-grade applications and IT resources at a lower price
2. Shifts large, unpredictable, up front capital costs to smaller, predictable, recurring monthly expenses
3. Lower operational costs: smaller IT staff focused on business core competencies, longer equipment life (8 years for thin clients vs. 3 years for PCs), reduced system downtime costs, better overall utilization of specialized applications
4. Software rental for short-term projects
5. Always access latest version of applications
6. (Global) access from anywhere, anytime, on any device
7. Have multiple "desktops" for both personal and business purposes
8. Reduced risk of virus propagation and other security threats
9. Quickly add new end users

ASP Value Chain

The ASP value chain goes from co-location services to full-managed services. Co-location services are for customers wanting to have total control of their applications and servers while leveraging the ASP infrastructure. Full-managed services are for customers looking for the ASP to manage their applications and infrastructure. In all these service arrangements, customers may want to monitor security status, and application/network performance. Predictable performance translates into the ability to offer Quality of Service (QoS) guarantees that in turn drives Service Level Agreements (SLAs). It is important that compliance of SLA metrics (e.g., network availability, application response times) be proactively reported to end users as part of service offerings.

Partnering to Add Value

Given the nature of the ASP concept, where applications become services that are delivered over networks, the service model is structured into the "layered" framework shown in Figure 2. The lowest layer provides infrastructure such as data center facilities, physical connectivity (network transport and access), and data networking equipment including routers and firewalls. This layer is responsible for maintaining expected levels of network performance, reliability and security. The next layer provides the ASP platform and includes application-specific infrastructure, computing resources such as servers and operating systems, data storage resources, and application management. Next is the applications layer where application services reside. A professional services layer provides application planning, consulting, and integration services. The Operations Support Systems (OSS) layer provides fault and configuration management, accounting, application and network performance monitoring, and security functions. It also supports subscriber management and customer care functions.

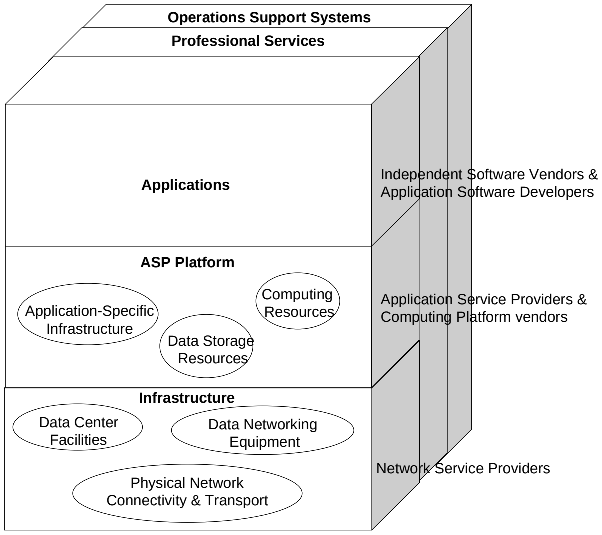


Figure 2. ASP Components

Service providers may partner with others to support all these layers. For instance, a Network Service Provider (NSP) partner provides the networking infrastructure. An Independent Software Vendor (ISV) partner provides applications and tier 2/tier 3 application-related customer care. An ASP partner provides tier 1 customer care and overall service management. A Professional Services partner provides consulting, planning and integration services. In some cases, an ASP may be cross selling services from other ASP partners. And in other cases, an ASP may have a presence in an ISP portal thus allowing the ISP to offer ASP services.

Target Applications

As mentioned earlier, the ASP service model initially targets business segments. In particular, this model is attractive to small and mid-tier business customers because it provides access to enterprise-class applications at a lower price.

At this time, business applications that do not require extensive integration or software-code customization are the most suitable to offer via the one-to-many ASP service model. Good examples are desktop productivity suites, E-mail hosting, Web hosting, calendaring, data warehousing and storage, virus control, and unified messaging. However, applications that require a higher level of customization, such as E-commerce hosting, electronic customer care (including Customer Relationship Management), sales force automation, and back-office applications (e.g., human resources, payroll, supply chain management, electronic payment) are what end users are demanding the most from ASPs.

Initial service offerings are targeting traditional, data-centric enterprise applications. A next step is to offer voice-enabled data applications that take advantage of the convergence capabilities of new programmable communications platforms. There are also opportunities to offer application services to residential end users. Some example applications include managed home networking services, (content) media streaming, and network hosted games.

The ASP service model bundles network access, managed network services, and network-hosted applications as one service. That gets complemented with ancillary services such as end user authentication, application usage reporting, and application monitoring. The advantage of providing these value-added service bundles is that it significantly reduces end user churn. This model takes service bundles to the next step in the value chain.

ASP ARCHITECTURE

The sections below describe the ASP architecture at the service, network, data center, and application levels.

Service Delivery Architecture

The ASP service model dictates a service delivery architecture. This architecture is client-server in nature, although the functionality is distributed differently when compared to traditional "fat client" architectures. In a "fat client" architecture the client performs some of the processing and relies on remote servers, if needed, to provide additional data and/or processing functions. In the ASP service delivery architecture, clients do not perform any processing functions other than presentation functions to locally display or "publish" remotely executed applications. In other words, clients only perform functions to display the user interface of invoked applications. A specific protocol is used between the client and the server to carry keystrokes, mouse clicks, and screen updates across the network.

As shown in Figure 3, there are two variants of the ASP service delivery architecture. One is Web-based, where end users remotely run applications on a web server and associated backend servers. In this case the HyperText Transport Protocol (HTTP) is used between the browser client and the web server. The other architecture is thin client-based. In this case the end user relies on a Citrix®[1] client to access network-hosted applications running on Citrix-enabled application servers. The server runs a multi-user operating system. The remote presentation services protocol used between the Citrix-based client (browser) and the application server is the Citrix Independent Computing Architecture (ICA®[1]).

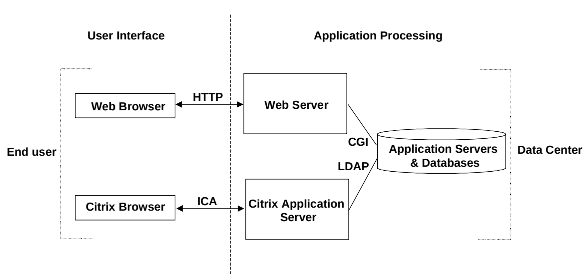


Figure 3. ASP Service Delivery Architecture

[1] Citrix and ICA are registered trademarks of Citrix Systems, Inc.

Network Functional Architecture

Figure 4 shows a high-level, generic network architecture that supports ASP service offerings. Applications are hosted in server farms running in data centers. Multiple instances of these centers are dispersed over a geographic area to increase service availability, and improve application response times via load balancing. Backbone network interconnectivity may be leased from a Network Service Provider. As shown in the figure, a Network Access Point (NAP) provides public peering to connect to the Internet. For improved performance, multiple private peering connections from data centers to major Internet backbone network providers can be coordinated to bypass congested Internet NAPs.

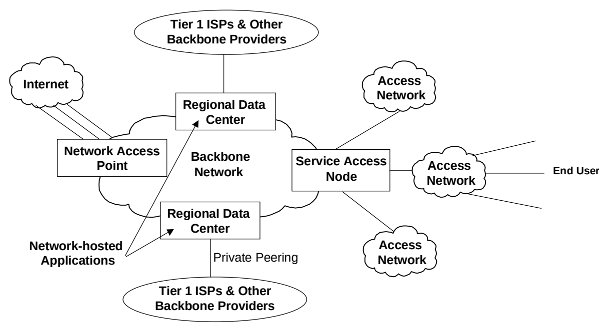


Figure 4. High-level, Generic Network Functional Architecture

Given that the ASP service model initially offers business applications, security is an important consideration. A layered security scheme that covers host, network, application, and end user (authentication) security should be adopted. Another important consideration is application performance, in particular response times and packet loss. Resource management plays a key role in supporting this. This includes traffic management to evenly distribute the traffic load over the network, and traffic shaping to enforce SLAs. Quality of Service (QoS) treatments are another part of this. These QoS treatments are based on application-specific and/or end user-specific policies.

Virtual Private Networks that provide end user access to network hosted applications may offer the necessary security, resource management and QoS treatments in a coordinated way. These are “ASP value-added VPNs” where VPNs complement ASP service delivery offerings. A special case of this is the offering of managed Extranet services.

Data Centers

The key goal of data center facility design is the optimization of application service availability. That means that data center facilities need to be highly secure and disaster resistant. That includes physical security (escorted access, card-key access, surveillance cameras and intrusion sensors), equipment redundancy, heat/smoke detectors, fire suppression systems, water sensors, air filtration systems, power/surge protection, dual utility feeders, backup power sources, and temperature control systems. However, no physical facility is completely disaster resistant. Business continuity plans are required to manage multiple redundant or backup data centers. Backbone network connectivity should be redundant and follow diverse paths. These connections should be engineered with spare capacity built-in (50% or more).

Data centers consist of 4 basic elements: application processing, data networking, transport, and operations. Figure 5 shows a functional decomposition of these elements.

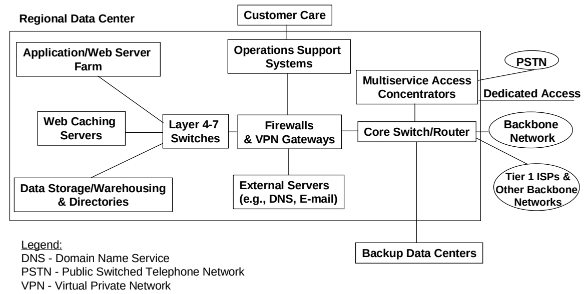


Figure 5. Generic Data Center Functional Architecture

Application and Web server farms run on shared or dedicated servers. Redundant servers may be located at separate data centers and accessed via multiple network connections. This can be coupled with Layer 4-7 (web) switches to provide local and distributed load balancing among servers. These web switches monitor server and application response times, and network utilization. Web caching servers complement load-balancing functions to offer web access acceleration services. Highly redundant database clustering or Storage Area Networks (SANs) provide data storage management services with fail-over capabilities. Firewalls, VPN gateways and intrusion detection systems provide secure access to applications, including those running on external servers. This may include secure ID token-based end user authentication, host-based security, and router access control lists.

A core switch/router interconnects components in data centers and provides connectivity to the backbone network. Multiservice access concentrators may support dial-up and dedicated access.

The Operations Support System performs application monitoring, management, and billing. A customer care gateway, or Customer Network Management (CNM) system, allows end users to manage their services on-line. With CNM, an end user can monitor network and server availability/performance, and enter and monitor the status of trouble tickets among other features.

Application Architecture

The ASP service model is multi-user (i.e., one-to-many) and subscription-based. That has an impact on application design, as network-hosted applications need to be scalable and customizable while keeping operational costs down. End users demand customized applications at a cost lower than owning a fat PC client. This gets more critical when ASPs want to target business customers with premium applications while adopting a commodity service delivery model. One way to deal with this is to design applications end users can easily configure via templates or wizards. Alternatively, ASPs could build libraries of frequent customized versions of the applications, although that requires keeping larger inventories.

ASP ON CABLE

Major efforts are underway to upgrade existing cable access plants to support two-way data communications. In addition, Cable Operators are looking at other access network types, such as xDSL, fiber, and broadband wireless to expand their access network portfolio and footprint. That gives Cable Operators the proper conduit to reach end users as an ASP. What is still required is for Cable Operators to extend their IP infrastructure to include data centers. A generic network diagram is shown in Figure 6. In the figure, the data center also includes NAP functions to access the Internet. Private peering to other backbone network providers is also possible. In the case of smaller Cable Operators who may want to become ASPs, smaller scale data centers could be co-located with Head End facilities (e.g., Master Head End).

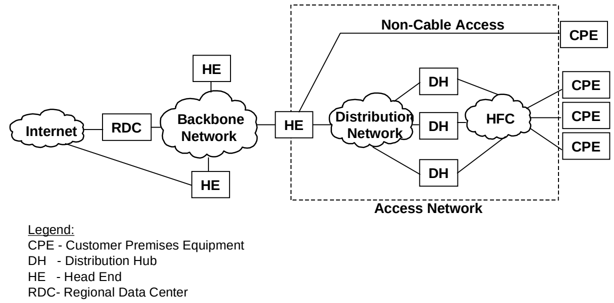


Figure 6. Regional Data Centers and Cable Networks

In a way, the ASP service delivery model may help overcome some of the bandwidth limitations of the HFC plant in the upstream direction. The ASP service delivery model keeps application processing in network data centers and relays presentation functions to end user terminals. This not only reduces bandwidth requirements but also keeps it at a predictable rate, independently of the applications used. For instance, Citrix ICA reduces required bandwidth to approximately 13 kilobytes-per-second.

Another benefit of adopting the ASP service model on cable networks is that it improves end-to-end security. Applications are exclusively run in the network, where the Cable Operator has total control. This could be complemented with Cable Operator-managed VPNs for secured access. These VPNs also help in providing end-to-end QoS over the cable plant. One interesting observation is that with the ASP service model, QoS is better served in the sense that the networked data centers are properly designed and managed to guarantee the right levels of QoS to applications. The backbone network is traffic engineered with enough bandwidth to support such operations. In the access network, only application presentation streams are transmitted with more predictable traffic profiles and with much lower bandwidth requirements. That does not preclude applying traffic prioritization in the access network and in fact, network equipment is becoming aware of the remote presentation services protocols used to support these transactions.

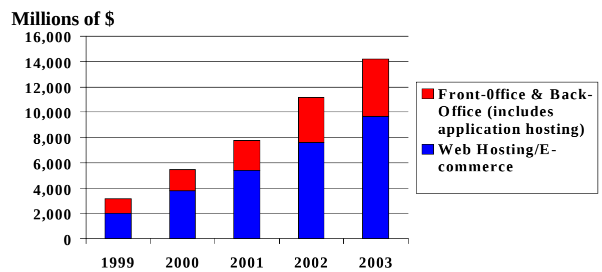
Another key feature of the ASP service model that fits well in Cable Operator offerings is service bundling. This feature is becoming critical in a time when technological advances are making possible the convergence of voice and video communications with network-hosted data applications. Cable Operators are in a good position to bring all these services together as network-hosted applications. Adopting ASP may be a good complement to current cable telephony and digital video initiatives. The "always-on" broadband access experience Cable Operators provide to end users may be fertile ground for new application and service opportunities that may not be possible for other service providers. End users connect to IP networks and the Internet for three basic reasons: communications, content, and commerce. Cable Operators are in a good position to provide services along those three dimensions. The ASP service concept may be the catalyst that may allow Cable Operators to offer these as compelling, valued added service bundles.

The ASP service model simplifies end user equipment to "multimedia" user interface devices. This is aligned with the Cable Operator's goal of reducing truck rolls to install and configure Customer Premise Equipment (CPE). The emergence of ASP service offerings is driving retail sales of "Internet appliances" that are not only easy to install and configure, but also easy to maintain and has a longer life when compared to PCs. Another trend is to support wireless access and the ASP service model may position Cable Operators to serve those markets as well by adding wireless Web gateways/portals to their infrastructure.

From an ISP perspective, the ASP service model is a next step in the value-added chain. Cable Operators could adopt the ASP model as a way to differentiate themselves from traditional ISPs. And they could do so without extensive partnering to provide full service network-hosted applications.

BUSINESS ANALYSIS

According to the Yankee Group, the ASP market will grow from $3.1 Billion in 1999 to $14.2 Billion in 2003. This is shown in Figure 7. According to this forecast, web hosting and E-commerce are the key revenue generators.



Source: Yankee Group, 8/99

Figure 7. ASP Market Forecast

A market segment of interest is income generating home offices. By the end of 2002, IDC[AG99] expects over 30 million US home office households with someone running a business. About 8.2 million US households will be equipped with cable modems, out of which 6.2 million are expected to be home offices. This represents over 75% of the cable modem customer base.

One of the elements that drives home-based businesses to access and have a presence in the Internet is that it serves as a low-cost conduit for revenue-generating opportunities (e.g., E-Commerce). Also, the Internet is quickly becoming a strategic portal for business information and research, in particular for small businesses which tend to have a higher percentage of knowledge workers. This type of customer is very cost sensitive, prefers to deal with local service providers, and expects high-quality customer service. The ASP service model may help Cable Operators satisfy those needs.

A business case built around a simple scenario is presented below. A cable operator, who currently provides traditional Internet access services to residential end users via cable modem, wants to become an ASP. In this scenario, the cable operator built a data center capable of supporting a total of 50,000 end users (20,000 end users subscribed to ASP services and the remaining 30,000 end users subscribed to regular Internet access services). Another option would be for the Cable Operator to have a third party service provider host the data center. This option reduces up front capital outlays and allows faster entry into the market. But for the purposes of the business case presented here, it is assumed that the Cable Operator builds its own data center. The rest of the assumptions used to build the business case are presented in the table below.

Item	Assumption
Market Size Assumptions	
Initial footprint (i.e., first year)	150,000 end users
Growth rates	30% (years 1-3) 25% (years 4-6) 20% (years 7-9)
General Assumptions	
Weighted Average Cost of Capital (WACC)	12%
Terminal rate	4%
Tax rate	36%
ASP Service Penetration	
Initial penetration	1%
Annual increase	2%
Maximum	6%
Churn Rates	
Initial churn	12%
Incremental churn	0% [Economically stable service area]
ASP Service Pricing	
Service revenue	$150/month/subs
Annual increase	-$5
Partner share	10%
Equipment Expense	
CMTS (incremental to support ASP subs)	$150/sub
ASP equipment (1 data center)	$1.7 Million ($700K for software, $300K for servers, $400K for data networking, $300K for data storage)
Data Center Expenses	
Recurring expenses per year	$500,000 + 5% of gross revenue
Engineering & design	$200,000 (first year only)
Billing and OSS Expenses	
Recurring expenses per year	$250,000 + 3% of gross revenue
Customer Service & Support	
Recurring expenses per year	$30 * average number of subscribers
Sales and Marketing Costs	
Recurring expenses per year	$100K + (150 * number of new subs)
General & Administrative	
Recurring expenses per year	$500,000 + 3% of gross revenue
Installation Costs	
Installer salary & benefits per year	$100K
Number of installations per technician per year	7,500

Table 1. Cable-ASP Business Case Assumptions

The business case assumes that the Cable Operator starts offering a relatively simple set of ASP applications such as basic human resource applications, financial management, collaborative computing, sales automation, groupware and e-mail, all offered as a service bundle. In addition, a simple pricing plan is assumed based on flat user/seat/month fees. More sophisticated billing schemes are needed to support usage-based pricing either at the application level or at the transaction level. Determining the right level of billing granularity not only depends on current technological capabilities but also on the ASP business/service arrangements upon which application frameworks are implemented.

Figure 8 [JT2000] shows how ASP pricing models may be evolving over time. The diagram shows a shift towards subscription-based and transaction-based models, and implies that eventually pricing models may rely less on traditional software licensing. The ASP service model offers applications to thousands of users on a monthly subscription basis. This requires adapting application-licensing schemes to fit a dynamic recurring monthly fee model. One example is software licensing utilities that enable ASPs to provision applications for rental without incurring up-front license fees. The software licensing utility measures concurrent usage of application software and the ASP makes monthly payments to Application Software Vendors accordingly. This utility may also apply tiered discounts as the ASP's customer base grows.

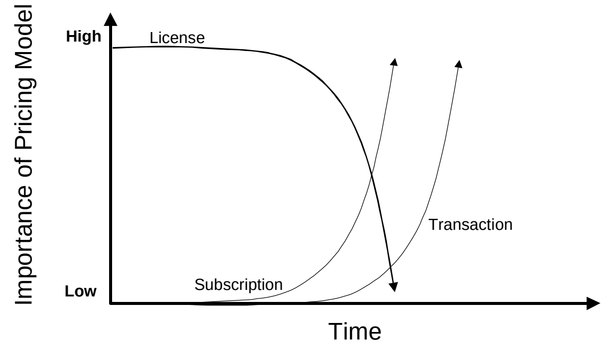


Figure 8. Evolution of ASP Pricing Models

The flat rate pricing plan used in this analysis is based upon application types. The pricing criteria takes into consideration application value and application configuration time. There are other pricing plans in use today. An example is charging according to server type and configuration (e.g., shared vs. dedicated servers). This type of pricing plan can be broken further into hardware and maintenance fees. Another example is charging according to end user access rights to application data (e.g., read-only vs. editing privileges).

Independently of the pricing plan used, the ASP service model rests on a pricing structure that generates monthly recurring revenues. ASPs have many opportunities to increase these revenues. For instance, many existing applications are being "ported" to network-hosted environments. In addition, new network-hosted applications are emerging. This creates opportunities not only to expand service portfolios, but to offer professional services as well. In fact, the ASP service model allows ASPs to cross-sell solutions from other ASPs. The service model simplifies adding new end users as well.

Figure 9 below shows the gross revenue results of the example business case analysis considered here. After the fifth year, gross revenue starts decreasing. This is due to the assumption that there were no plans to expand beyond the capacity of the single data center deployed initially, and that service pricing decreases as technology matures. This means that after five years of steady customer base growth, either new data center facilities need to be deployed or the capacity of the existing should be expanded. This of course depends on the growth rate profile assumed in the analysis.

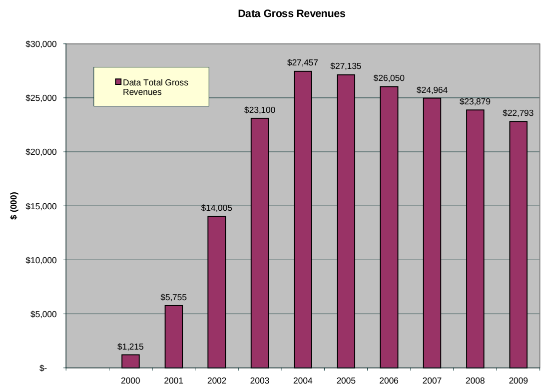


Figure 9. ASP Business Case - Gross Revenues

In terms of expenses, there are several elements that need to be considered. One element is the cost of implementing data center facilities and improved IP infrastructure. Storage costs are of particular importance. Another element is the cost of customizing application software. As mentioned before, customized software does not fit well with the one-to-many ASP service model. The time that an ASP spends customizing an application for a customer is time that cannot be applied to serving the needs of other customers. Also, application customization may increase the time it takes to complete application software upgrades. Applications should be designed in ways that optimize its customization capabilities or at least in ways that expedite the creation of libraries of pre-customized application templates.

Other elements include application delivery costs, application service trial costs, best practice implementation costs, the cost of integrating new applications into existing service bundles, and IT staff costs (e.g., hiring and training). In terms of IT staff costs, this cost is spread over a growing customer base, thus providing economies of scale benefits. IT utilization is "bursty" in nature when dedicated to one company. Once IT resources are shared among multiple customers, their utilization increases and stay at a more stable rate.

Going back to the business case, Figure 10 below shows the total expense results. Again, after the fifth year the expense growth rate slows down considerably. This corresponds to the fact that the data center assumed in the analysis has reached its maximum capacity at that time.

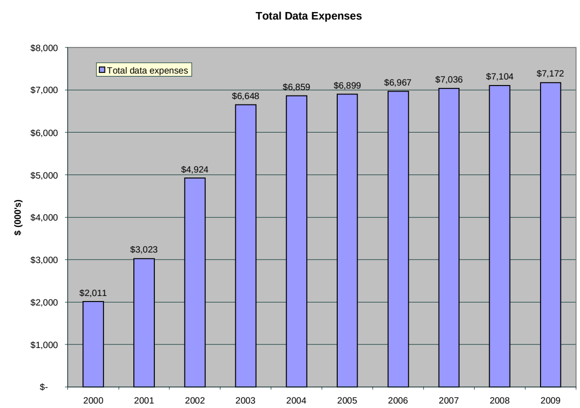


Figure 10. ASP Business Case - Total Expenses

Figure 11 below shows the results of the free cash flow analysis of the business case. Again, after the fifth year cash flow starts decreasing given the assumption that, at that point, maximum capacity is reached in the data center and there are no plans for additional growth. At the same time, annual service revenues keep decreasing. Of course, no additional investments are made as well. In a more realistic scenario, the Cable Operator may plan for growth, both of the customer base and the service portfolio. Also, there will be technological advances that will increase infrastructure capacity and enable more profitable emerging applications.

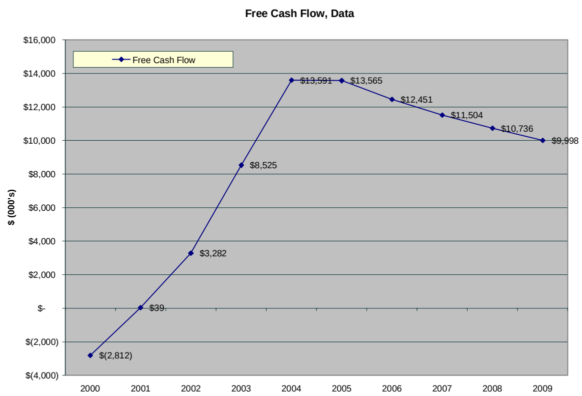


Figure 11. ASP Business Case – Free Cash Flow

Figure 12 below shows CDCF results of the business case. With an initial investment of about $2 Million, a relatively simple portfolio of application offerings and limited growth planned, the business case predicts over $41 Million in 10 years with the break-even point reached in less than 3 years.

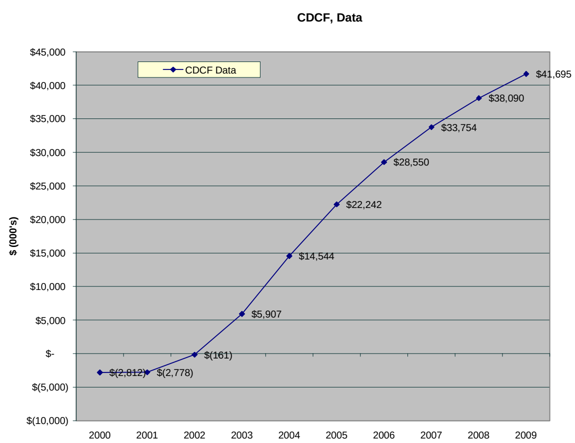


Figure 12 ASP Business Case – Cumulative Discounted Cash Flow

CONCLUSIONS

The ASP service model and architecture provide Cable Operators the opportunity to differentiate from traditional ISPs and exploit their strengths in offering converged service bundles. This model moves application processing, security, and QoS to the network and relays presentation functions to end user terminals. Security and QoS treatments are still needed to guarantee proper delivery and application response times, but these can be provided in a more efficient and simpler way over the cable access network. End user equipment gets simplified, meaning a reduction in truck rolls and overall maintenance support.

There are some challenges, however, that need to be addressed. Depending on the situation, the ASP service model may require partnering with other service providers to provide certain components. That means proper measures need to be put in place to guarantee the combined security and QoS that satisfy end-to-end SLAs. Another challenge is the definition of best practices. Critical areas include: data center operations, network operations, client-server operations, application management and monitoring, and CPE management.

Other challenges include: designing applications that run on distributed network computing environments under a service subscription model, evolving data center computing platforms to support converged voice/video/data applications, and developing schemes that will allow Cable Operators to guarantee the required application and network performance levels as dictated by SLAs.

ACKNOWLEDGEMENTS

The author would like to acknowledge the contribution of Sabet Elias in providing the business modeling tools and Laurie Schoonover in providing some of the ASP market data used in this paper. A special acknowledgement goes to Charles Lee for his outstanding contribution to the business modeling analysis.

REFERENCES

1. [DP99] Pappalardo, Denise, John Cox, “Customizing ASP apps no easy chore,” Network World, Volume 16, Number 45, November 8, 1999, p.1
2. [JM99] Makris, Joanna, “Hosting Services, Now Accepting Applications,” Data Communications, March 21, 1999
3. [BQ99] Quinton, Brian, “Kicking ASP,” Telephony, November 1, 1999, p.38
4. [AG99] Gilroy, Amy et.al., “Home Office on the Internet,” IDC Report #18331, March 1999
5. [JT2000] Traynor, Jim, “The Future for ISVs and the Integration Channels,” Great Plains Software, slide presentation, ASP Summit 2000, San Jose, CA, February 14-15, 2000

CONTACT INFORMATION

Pablo L. Martinez
Lucent Technologies
Crawfords Corner Road, Holmdel NJ 07733
(732) 949-1733
pablo@lucent.com

Requirements for Reliable Communications in HFC Based Broadband Data Networking

Paul Nikolich
Broadband Access Systems, Inc.

Abstract

This paper will briefly address the reliability performance requirements of the Broadband Data Networking transmission system cable modem termination systems (CMTS), and HFC plant). In the context of this larger perspective, it will specifically illustrate the critical importance of a maintaining a high availability CMTS. For example, failure of a single CMTS downstream port will affect hundreds of subscribers. The service objective is to keep downtimes below 53 minutes a year (99.99% available.)

This paper will then focus on techniques that can be applied to maximize the availability of the Cable Modem Termination System. Mechanisms that increase availability are redundancy, high mean time to failure and low mean time to restore. A fully distributed, no single point of failure architecture also enhances reliability.

This paper will be of interest to all cable operators interested in understanding the impact of deploying high availability advanced services and how that translates into infrastructure equipment requirements.

OUTLINE:

1) Introduction
2) Service Availability 101
3) An HFC network Failure Model
4) Analysis of a the HFC Failure Model
5) Conclusions

1. Introduction:

Broadband Data Networking has tremendous potential to become a key source of revenue for cable operators by enabling the delivery of not only Best Effort Data services, but of Packet Voice and Packet Video services as well. However, along with the revenue potential comes the increased expectations by the consumers for high availability of these new data, packet voice and packet video services. Hence operators must ensure their networks are 'high availability' with performance similar to what the telephone company objectives are today: 99.99% uptime.

2. Service Availability 101:

Service Availability is defined as the percentage of time that service is available. This value is a function of the number of network elements between the service 'source' and the service 'sink', how these elements are interconnected and reliable these elements are. Each network element can be assigned a mean time between failure (MBTF) and a mean time to restore (MTTR). It is assumed these parameters are assigned during the 'steady state' portion of their life cycle, and not early on when there is a high 'infant mortality' or late in life when 'wearout' becomes a factor.

Availability, A, of one specific network element is defined as

$$A = MTBF/(MTBF+MTTR)$$

Obviously, for each network element, it is desirable to make the MTBF as high as possible and the MTTR as low as possible to maximize the availability.

Conversely Unavailability, U, is defined as

$$U = 1 - A$$

$$U = MTTR/(MTBF+MTTR)$$

When network elements are connected in series, any one-network element failure results in service unavailability. In this case, the total availability (A_{total}) of the network segment comprised of series elements is equal to the product of the individual availabilities.

For example, if network element one (NE1) with availability A1 is in series with network element two (NE2) with availability A2, the total availability is

$$A_{total} = A1*A2$$

If A1 = A2 = 0.99 (99%) the resulting A_{total} is 0.98 (98%) and the total unavailability is

$$\begin{aligned} U_{total} &= 1 - A_{total} \\ &= 1 - 0.98 \\ &= 0.02 \text{ (or 2\%)} \end{aligned}$$

When network elements are connected in a parallel redundant fashion, any one-network element failure will not result in a service disruption. In this case, the total unavailability (U_{total}) of the parallel network elements is equal to the product of the individual unavailabilities. Assuming NE1 and NE2 above are in parallel, the resulting

$$U_{total} = U1*U2$$

If U1 = U2 = 0.01 (or 1%, note this is equivalent to availabilities A1 and A1 above), the total unavailability is

$$\begin{aligned} U_{total} &= 0.01*0.01 \\ &= .0001 \text{ or } .01\% \end{aligned}$$

$$\begin{aligned} A_{total} &= 1 - U_{total} \\ &= 1 - .0001 \\ &= 0.9999 \text{ or } 99.99\% \end{aligned}$$

In summary, implementing redundant network architectures along with maintaining high MTBF and low MTTR in the network elements will help cable operators achieve the very high availability targets (for example 99.99% or 53 minutes of downtime per year) for their networks.

3. An HFC network Failure Model:

Armed with an understanding of how to calculate service availability for a network, a model of the HFC network for failure analysis can now be constructed. Figure 1 illustrates a typical HFC network. The Cable Modem Termination System (CMTS) is connected to the Internet via a conventional 100Mbps Ethernet port. The CMTS connects to the HFC network via its downstream and upstream RF interfaces. The downstream interface is connected to an fiber transmitter via the combiner, the optical signals sent down 10 miles of fiber to the Fiber Node, where the signals are converted to RF and transmitted into the neighborhood via a series of trunk and line amplifiers over 75 ohm coaxial cable. The signals are tapped off the coax and brought through a drop cable into the home where they are routed to the television and cable modem subsystems in the home. Signals sent upstream from the cable modem to the CMTS follow the reverse path illustrated in Figure 1 through an independent fiber return.

Even though there are differences between Hybrid Fiber Coax Network shown in Figure 1 and the telephone local exchange carrier (LEC) architecture that is used to deploy fiber to the curb, for the purposes of developing a failure model they can be viewed similarly[1]. The LEC criterions are defined in a document called Generic Requirements for Fiber-in-the-Loop Systems[2] (TA-909). TA-909 defines the maximum annual outage objective to be less than 53 minutes (99.99% availability). Note that this is an objective, and is not necessarily what the LECs are achieving.

TA-909 makes the assumption that certain failures are not included in their model:

- failures of utility power
- failures of switching equipment (in the cable operator's case this is head end equipment)
- failures in cabling from the curb into the home (in the cable operator's case this includes the drop cable, splitters, set top box, cable modems, televisions, PCs, etc.)

In this paper, we will look at the contribution of the head end equipment to the overall service availability, but we will not address power failures or failures that occur from the tap into the home.

4. Analysis of the HFC Failure Model:

All of the MTBF numbers shown in this paper assume that the equipment is operating in 'mid-life steady state'. It is during mid-life that the MTBF values are the highest; infant mortality has ended and wearout has not begun.

Note that dominant source of downtime in the network are the DOCSIS CMTS (120 minutes) interface. The chassis controller and 100BaseT Ethernet interfaces also have significant contributions (24 minutes). Also shown in Table 1 is what happens to the downtime when the CMTS, Ethernet interfaces and chassis controllers are run in 1+1 redundant modes--the downtime drop to less than 0.01 minutes per year per module. The overall chassis downtime reduces from 171 minutes per year to just over 3 minutes per year--a significant reduction. Clearly it will be a requirement for cable operators that must deploy highly reliable services to utilize equipment that is capable of running in a redundant mode.

Each of the network elements depicted in Figure 1 has a specific MTBF and MTTR as shown in Table 2. All of these values are taken from the ADL study[1] except for the values used for the CMTS subsystem. The CMTS subsystem values are based on estimates of equipment with similar complexity.

Figure 2 is a graphical depiction of the downtime on a per network element basis. Figure 2 also illustrates the downtime of a set of network elements together; the CMTS system both in redundant (3 minutes) and non-redundant (171 minutes) modes. The downstream fiber (15 minutes), upstream fiber (16 minutes), trunk coax/amplifiers (22 minutes) and feeder coax/amplifiers (19 minutes) are shown. (Note the y-axis is on a log basis.) The grand total of for the entire HFC network, assuming the CMTS is running in redundant mode is 76 minutes per year--clearly higher than the stated goal of 53 minutes per year.

One method of reducing the downtime to less than 53 minutes a year will be to place redundant paths in for the downstream (15 minutes) and upstream (16 minutes) fiber runs. This has the potential to reduce the downtimes to fewer than 1 second per year, bringing the total system downtime to approximately 46 minutes per year--well within our goal.

Of course the benefits of redundancy come at a price, it roughly requires a doubling of investment in the CMTS and Fiber infrastructure portions of the network.

5. Conclusions

It is possible to provide highly available HFC network systems, with downtimes of less than 53 minutes per year--better than what the local exchange carriers have as an objective--but it will require the deployment of redundant systems both in the head end (CMTS) and in the fiber plant.

Also remember that the effect powering, drop cables, in home cabling and equipment were not included in this analysis--these factors must be considered when caclculating the service objectives the from the end user perspective.

[1] Failure Modes and Availability Statistics of HFC networks, Stu Lippoff, Arthur D. Little, HFC '96: High Integrity HFC Networks, SCTE and IEEE Comm Soc.

[1] Generic Requirements of Fiber-in-the-Loop Systems, Telcordia TA-NWT-000909, Issue 2, December 1993.

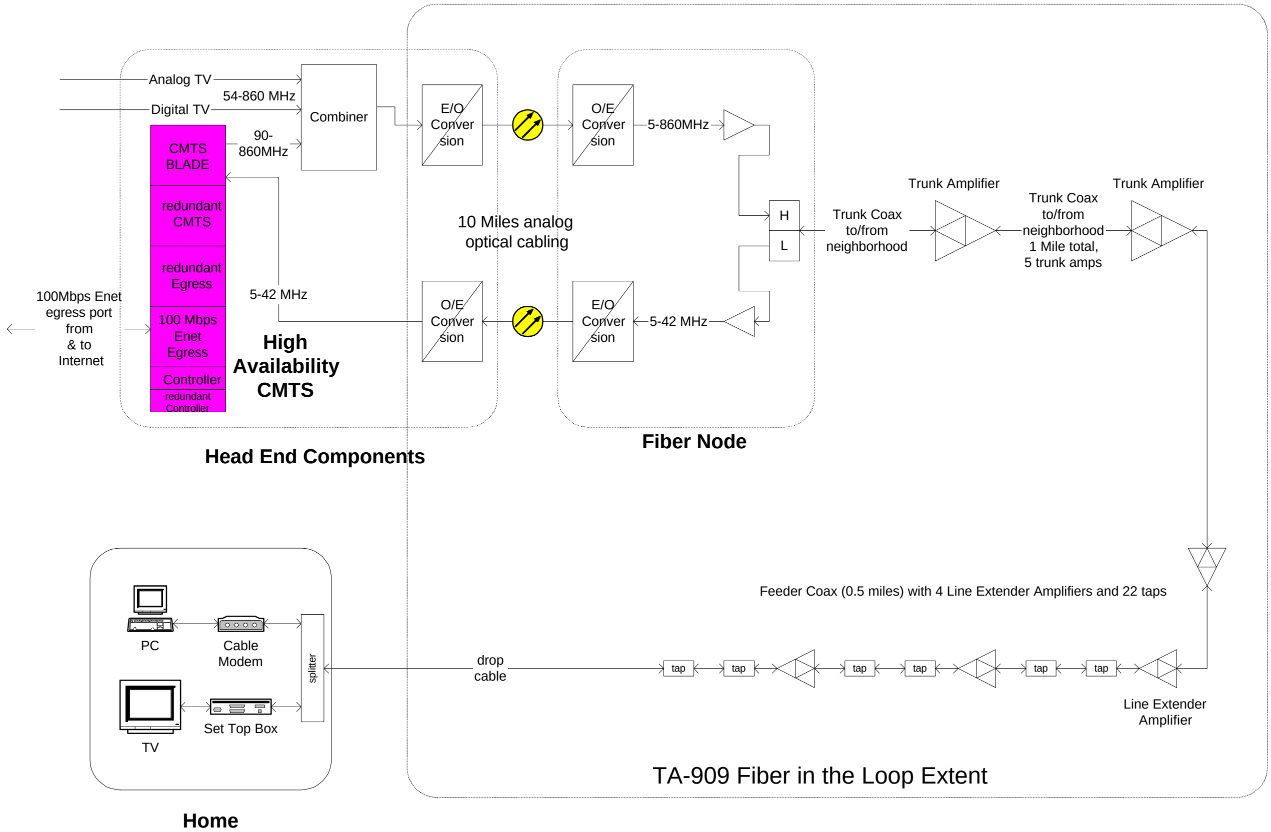


Figure 1 HFC Network Failure Model

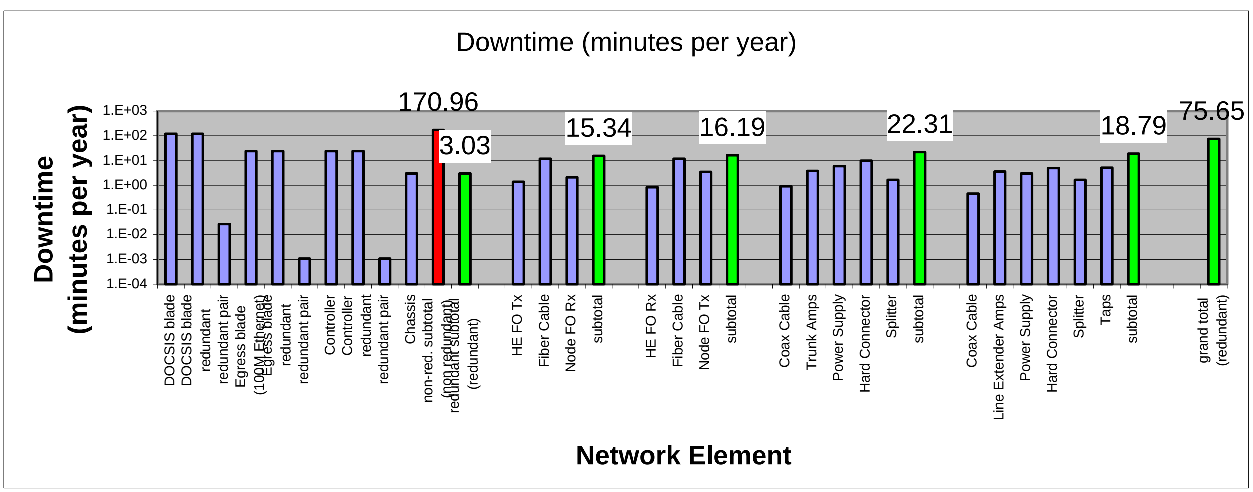


Figure 2 HFC Network Element Downtime in Minutes per year

Table 1 HFC Network Failure Mode Data

Network Segment	Network Element	Failure Rate (%/yr)	Number	MTBF (years)	MTTR (hours)	Avail	Unavail	Down-time (min/year)
CMTS	DOCSIS blade	100	1	1.0	2	0.999771742	0.000228258	119.97
	DOCSIS blade redundant	100	1	1.0	2	0.999771742	0.000228258	119.97
	redundant pair					0.9999999479	5.21019E-08	0.0274
	Egress blade (100M Ethernet)	20	1	5.0	2	0.99995434	4.566E-05	24.00
	Egress blade redundant	20	1	5.0	2	0.99995434	4.566E-05	24.00
	redundant pair					0.9999999979	2.08484E-09	0.0011
	Controller	20	1	5.0	2	0.99995434	4.566E-05	24.00
	Controller redundant	20	1	5.0	2	0.99995434	4.566E-05	24.00
	redundant pair					0.9999999979	2.08484E-09	0.0011
	Chassis	1	1	100.0	5	0.999994292	5.70773E-06	3.00
	non-red. subtotal (non redundant)					0.999674739	0.000325261	170.96
	redundant subtotal (redundant)					0.999994236	5.764E-06	3.03
Downstream link	HE FO Tx	2.33	1	42.9	1	0.99999734	2.65981E-06	1.40
	Fiber Cable	0.439	10	22.8	4.5	0.999977449	2.25509E-05	11.85
	Node FO Rx	1.396	1	71.6	2.5	0.999996016	3.984E-06	2.09
	subtotal					0.999970805	2.91945E-05	15.34
Upstream link	HE FO Rx	1.396	1	71.6	1	0.999998406	1.5936E-06	0.84
	Fiber Cable	0.439	10	22.8	4.5	0.999977449	2.25509E-05	11.85
	Node FO Tx	2.33	1	42.9	2.5	0.999993351	6.6495E-06	3.49
	subtotal					0.999969206	3.07938E-05	16.19
Trunk Coax	Coax Cable	0.439	1	227.8	3.5	0.999998246	1.75399E-06	0.92
	Trunk Amps	0.514	5	38.9	2.5	0.999992666	7.33442E-06	3.85
	Power Supply	2	2	25.0	2.5	0.999988585	1.14154E-05	6.00
	Hard Connector	0.28	16	22.3	3.68	0.99998118	1.88197E-05	9.89
	Splitter	0.13	7	109.9	3	0.999996884	3.11643E-06	1.64
	subtotal					0.999957561	4.24393E-05	22.31
Feeder Coax	Coax Cable	0.439	0.5	455.6	3.5	0.999999123	8.76997E-07	0.46
	Line Extender Amps	0.599	4	41.7	2.5	0.999993162	6.83785E-06	3.59
	Power Supply	2	1	50.0	2.5	0.999994292	5.70773E-06	3.00
	Hard Connector	0.28	8	44.6	3.68	0.99999059	9.40996E-06	4.95
	Splitter	0.13	7	109.9	3	0.999996884	3.11643E-06	1.64
	Taps	0.13	22	35.0	3	0.999990206	9.79442E-06	5.15
	subtotal					0.999964257	3.57429E-05	18.79
	grand total (redundant)						total	75.65

Return Path Linear Distortion and Its Effect on Data Transmissions

Tom Williams
Holtzman Inc.

Abstract

Much has been written on the operational challenges for cable systems caused by upstream noise, especially burst noise, radio frequency ingress, and common path distortion. There is also an upstream problem with linear distortions caused by discrete echoes, amplifier tilt, group delay and micro-reflections. Diplex and lightning filters in amplifiers, missing terminators, damaged components, design problems, and construction mistakes create these linear distortions. The effect of uncorrected linear distortions is to increase the bit error rate in the presence of random noise or other additive impairments. Uncorrected linear distortion also limits an operator's ability to go to higher-order modulation schemes. In the return band a micro-reflection problem is aggravated by the low loss of the coaxial cable. This paper presents a burst reference signal measurement technique for measuring linear distortions on working plant as well as test data from active usptream cable plant. Simulation plots and curves show increases in intersymbol interference from linear distortions.

BACKGROUND

Linear distortions can severely impair the performance of digitally modulated carriers or render them useless. The problem that linear distortion creates is an inter-symbol interference (ISI) between the symbols comprising the data stream. This impairment is also referred to as dispersion [1]. The bit error rate will be increased in the presence of random noise if ISI is small. If the ISI is severe, the data will be useless.

Fortunately, a solution to the linear distortion problem exists: adaptive equalizers that cancel linear distortion. Adaptive equalizers are typically digitally implemented filters that are programmed with the inverse of the channel's frequency response, causing them to "undo" linear distortion. For bursty upstream cable traffic, using adaptive equalizers is complicated by the reality that each subscriber's modem has its own unique frequency response determined by its upstream path characteristics, including the inside wiring of the individual house. Thus the adaptive equalizer's programming solution is unique for every active upstream modem. Multiple solutions are currently being discussed. A first solution is placing the adaptive equalizer in the modem and pre-distorting the signal so it arrives in the headend corrected. This solution is the one that is currently being planned for DOCSIS 1.1. Programming coefficients for the adaptive equalizer must be transmitted over the downstream portion of the cable plant for each active modem in the system. A second solution is to store the programming for the adaptive equalizer in the headend for each active modem. The second method is complicated when the headend controller does not know in advance which modem's transmission will be received next, such as when modems are operating in a contention mode. A third solution is to accompany an upstream burst transmission with a reference signal. The reference signal may be used to remove linear impairments from the upstream transmissions.

Each of these approaches has their own drawbacks and work-arounds. Pre-distorting the signals works well, but as the number of modems attached to a headend controller increases, additional traffic will be generated if there is a sudden change in the echo environment of each modem's signal path. For example, a changing echo on a main line will cause any modems using that same line to re-adapt when their common signal path changes. Likewise, if the adaptive equalizer predistortion coefficients in the modems are not stored in non-volatile memory, and there is widespread power outage, the recovery sign-on storm will create additional traffic and delay. This is especially true if each modem's coefficients are not stored in the headend controller.

Placing the adaptive equalizer at the headend also can work, but this places a computational load on the headend controller. The problem with contention mode can be alleviated by using a robust modulation for contention (e.g. QPSK) and a higher rate modulation such as 16-QAM for reservation packets.

Accompanying the modem's burst transmission with a reference signal is another possible solution. Using a reference signal for just the contention mode packets is another possible variant.

At this time, however, adaptive equalizers are not required by the DOCSIS 1.0 specification and are not widely deployed. The net effect is that the cable operators typically place digitally modulated carriers in the middle portion of the return band where the linearity is best, avoid higher order modulations (such as 16-QAM) in all but the best of circumstances, and attempt to maintain their plant with a flat frequency response.

If a conventional sweep test is performed on an upstream cable plant, the results will give an operator an indication of the flatness of the 5-42 MHz band, but this is only half of the answer. The other half is the phase response of the plant. That is, a frequency response is a set of complex numbers with both magnitude and phase components. The derivative of the phase component with respect to frequency is known as group delay. If the channel has group delay variation, ISI is created. Upstream group delay typically has a "bathtub" shaped response in the 5-42 MHz band with the group delay variation being worse between 5 and 10 MHz and between 35 and 42 MHz. Most of the group delay variation above 35 MHz is caused by the low pass portion of the diplex filter, and most of the group delay variation below 10 MHz is caused by the "lightning filter" which is a 5 MHz high pass filter. Unfortunately, while the amplifier designers have generally done an excellent job of maintaining a flat magnitude frequency response over the entire return band, the group delay response on commercially available amplifiers can cause serious problems on relatively short cascades even using a modulation as robust as QPSK.

Group Delay

Both amplitude non-flatness and non-linear phase can cause ISI. If the signal path can be classified as a minimum phase network, any deviation in the phase linearity will be accompanied by amplitude non-flatness. Echoes (multipath distortion) are examples of responses that are minimum phase. Group delay is a measure of the slope of the phase vs. frequency curve. Group delay variation occurs when the

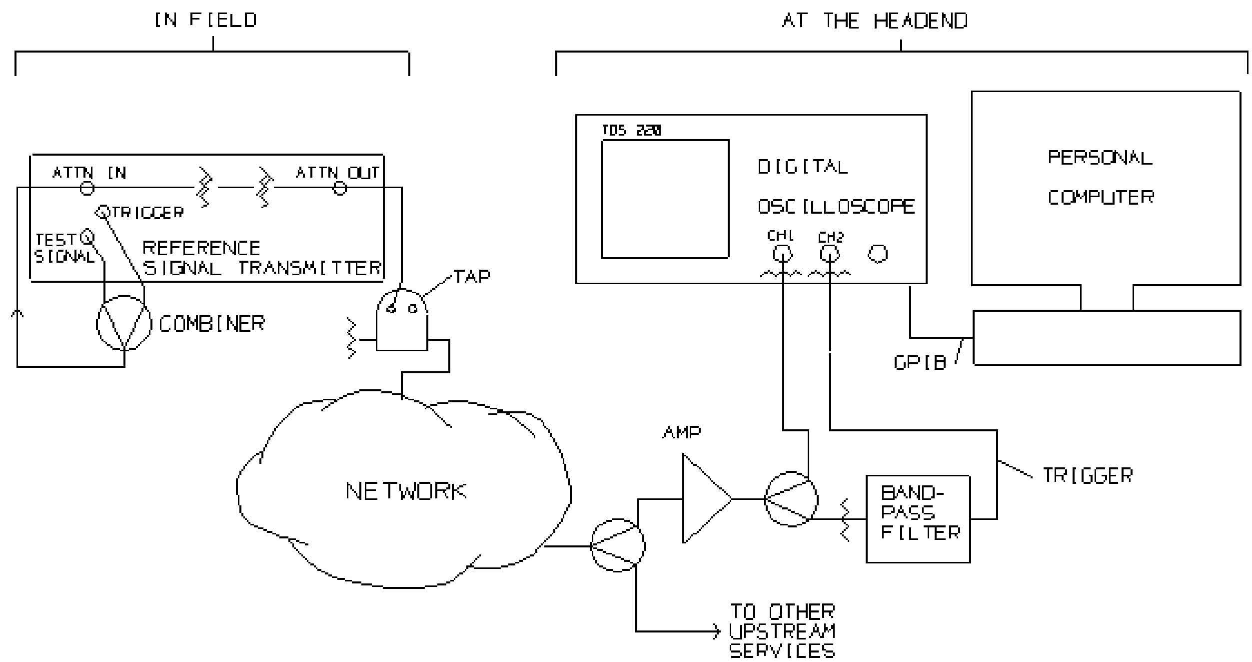


Figure 1 Wiring Diagram for Measuring Upstream Frequency Response

group delay changes with respect to frequency. Thus, if ω is frequency, ϕ is phase, gd is group delay, gdv is group delay variation, ω_1 is the bottom of the channel and ω_2 is the top of the channel:

$$gd(\omega) = -\frac{d\phi}{d\omega} \qquad (1)$$

$$gdv = gd(\omega_2) - gd(\omega_1) \qquad (2)$$

While amplitude non-flatness is an easy-to-understand impairment, group delay variation causes confusion to some cable operators. A tilt in the magnitude portion of the response means that the signals at one frequency are attenuated relative to signals at another frequency. Group delay variation means that signals at one frequency can make it through a network faster than signals at another frequency. For example, the effect of severe group delay variation on a Channel 2 downstream analog TV picture would be a misregistration between the luminance and the chrominance signals because the chrominance subcarrier, which is 3.58 MHz higher in frequency than the luminance signal, arrives at the subscribers's TV sets a fraction of a second before or after the luminance signal. Misregistration means that an actress's red lip color would be shifted to the left or right of the outline of her mouth. Reference [2] provides a definition of group delay.

In fiber optic cable, group delay variation has been renamed chromatic dispersion. The problem with fiber optic cable is that different colors (frequencies or wavelengths) can traverse through the glass fiber at different rates, distorting fast pulses, which contain energy over a wide range of frequencies [3].

Measurement of Complex Frequency Response

There is equipment that can measure complex frequency response. For example, a conventional network analyzer can make this measurement, but the input and output of the network must be in the same location. This works for test beds and temperature chambers, but is not applicable for fieldwork where the ends of the signal path are separated by many kilometers. One way to measure complex frequency response with the cable plant in the field is to use equipment that was originally designed for microwave links. These microwave link analyzers typically employ a sweep carrier that is frequency modulated (FM) as it is swept across the band. Unfortunately, this equipment requires the plant to be taken out of service to make the measurement.

MEASUREMENTS USING BURST REFERENCE SIGNALS

A technique that can be used on plant that is in service is to transmit a 5-42 MHz burst reference signal that is so brief that it does not cause interference. That is, the dwell time of the test signal in a channel is so short that any symbols that are damaged can be easily corrected by the FEC (forward error correction) circuits in the receiver. Equipment that uses this test method is commercialized under the name Cable Scope™. The brief broadband test signal is received on a digital acquisition unit (an A-D converter with memory) and is processed with an unimpaired copy of the signal to give the complex frequency response. If a subscriber service is using an upstream channel the instant that the reference signal is transmitted, the complex frequency response results will be contaminated over only the occupied band of the subscriber service.

The process works as follows. First a technician connects a reference signal transmitter directly into the receiver at the headend (or hub site) and captures an unimpaired reference signal. This unimpaired reference signal is stored directly onto a computer's hard drive. Next, the technician takes the reference signal transmitter to the field and connects its output to a tap port. Figure 1 is a block diagram that illustrates this step. At the tap location the technician injects a burst reference signal into the network on command. The reference signal is preceded by a 5 microsecond sine-wave burst that triggers the capture of the reference signal. The signal propagates back through a coaxial and a fiber portion of the network to the hub site or headend where the trigger burst passes through a bandpass filter and triggers a digital storage oscilloscope (DSO). The DSO captures the impaired reference signal and downloads it to a PC over a GPIB cable. The PC processes the impaired reference signal with a stored version of the unimpaired reference signal and displays the results to the operator in the hub site or headend.

Figure 2 (top) is a time domain trace of the burst trigger signal followed by an unimpaired reference signal that lasts about 18 microseconds. (Figures 2-6 are located at the end of the paper.) The grayed-out parts of the signal are displayed but not used by the process. Figure 2 (bottom) is a spectral (frequency domain) amplitude plot of the same trace that was obtained by performing a FFT on the time domain series. Figure 3 (top) is a time domain trace of a received impaired reference signal after it has passed through a number of amplifiers. Figure 3 (bottom) is a spectral amplitude plot of the received signal. At this point an operator in the hub site or headend instructs the technician in the field to increase or decrease the transmit level and send another burst signal. If the signal was transmitted too strong the network may be briefly clipped, distorting the test results. If the launched signal was too weak, any background noise on the network may degrade the quality of the reference signal, producing a poor quality frequency response result.

The PC processes both signals in the frequency domain to produce the complex frequency response. A complex frequency response plot is illustrated in Figure 4 with a magnitude component, a phase component, and a group delay component. The grayed-out areas below 2 MHz and above 42 MHz are the regions where the reference signal has too-little energy to give reliable results. The magnitude vs. frequency plot is displayed with a linear vertical scale. The phase vs. frequency plot actually "wraps-around" where the top of the plot (+180 deg.) and the bottom of the plot (-180 deg.) are the same phase. The group delay vs. frequency is derived as the slope of the phase vs. frequency plot. There is an adjustable smoothing function on the plots. Smoothing uses a running average of 16, 8 or 1 frequency points. Smoothing reduces the effects of noise at the expense of detail. At the receiver in the hub site an analog-to-digital converter running at 100 megasamples per second captures 2048 points in 20.48 microseconds. Frequency domain samples occur every 48.8 kHz.

The complex frequency response is also processed with an IFFT (inverse fast Fourier transform) to produce an impulse response, which is the top plot in Figure 4. An impulse response is nothing more than a voltage vs. time plot of how the upstream signal path would respond to an impulse (narrow voltage spike) injected at the tap.

While a single strong echo shows up as a ripple in magnitude and group delay plots, it shows up on the impulse responses as smaller pulse, which is delayed relative to the main pulse.

Reference Signals

There are many possible waveforms that can be used as reference signals. The characteristics that make a good reference (or training) signal are flat energy spectrum, a low crest factor (crest factor is the peak to average energy), ease of generation, and a good autocorrelation function. When a signal is convolved with itself, the autocorrelation function is found. A good autocorrelation has a single large impulse at t=0 and low energy at all other time samples.

Examples of well-known reference signals are:
1. A sin(x)/x waveform has a flat energy spectrum, but the crest factor of the waveform is high.
2. Pseudonoise (PN) sequences are a classic favorite. They have a low crest factor and are cheap and easy to generate. Unfortunately, the PN sequences do not produce uniform energy in the frequency domain; the spectral plots normally have frequencies with low energy.
3. The Koo training signal, which is used on line 19 in the vertical interval of some US analog television broadcasts, makes a good reference signal since it has uniform energy in the frequency domain, as well as high energy.
4. Chirps also work well. A chirp is a well-known waveform that has a frequency that increases (or decreases) linearly with time.

Stepped Frequency Reference Signal

Holtzman used a stepped-frequency reference signal to test two of the three systems. Figure 2 illustrates this signal. This stepped-frequency reference signal has characteristics that are similar to a chirp, but its instantaneous frequency increments in steps. This signal holds a frequency for a short period of time before stepping to a new frequency in a continuous-phase manner. Its frequency vs. time plot looks like a staircase. The advantage of this signal is that it can be easily generated by rapidly reprogramming a numerically controlled oscillator (NCO) integrated circuit. NCOs are also known as direct digital synthesis (DDS) chips.

The stepped frequency reference signal generates less interference with traffic relative to a PN sequence reference signal because the time that the stepped frequency signal spends in any frequency band is much less than the 18 microseconds needed to traverse the entire 5-42 MHz return band.

Perhaps the biggest advantage of the stepped-frequency reference signal is that a DDS circuit function is already inside most cable modems. Thus cable modems can, for very little extra cost, act as test signal generators at the end points of the network. This diagnostic capability should be invaluable to cable operators when homeowners start self-installing modems.

One of the three systems tested was characterized with a PN sequence. The stepped frequency signal produces a smoother frequency response plot relative to the PN sequence.

TEST RESULTS

As mentioned previously, three systems were tested. All systems were working two-way systems carrying modem and other traffic. Systems A and B are relatively recent rebuilds with 5-40 MHz return bands. Systems A and B used hybrid amplifiers. System C is an older system with a mix of upstream amplifiers. Some of system C's reverse amplifiers were capable of 30 MHz and some were capable of 40 MHz. The amplifiers in System C were the discrete transistor type. Systems A and B used upstream FP (Fabry Perot) lasers, while system C used DFB (distributed feedback) lasers.

A typical plot from System B is presented in Figure 4 and a typical plot from system A is presented in Figure 5. Figure 6 is a typical plot from System C. Some general observations can be made on the data. The first is that the group delay has a "bathtub" shaped response. This can also be observed as a "bending" or slope change of the phase response. Group delay variation tends to be very bad near the low end of the band and the high end of the band. Another observation is that the magnitude and group delay plots have a rough appearance. This roughness is caused, in the author's opinion, by micro-reflections. Many small echoes created by signals reflecting off taps, amplifiers, power inserters, cable kinks, and drop wiring anomalies can cause micro-reflections. If the equipment is within specifications, no single reflection is large, but all reflections summed together can produce large excursions in magnitude and phase. Micro-reflections were first predicted for the downstream plant, but were not strongly observed outside the home wiring. The probable reason for this is the reflections were damped by the high loss of the cable in the forward frequency bands. In

Node Number	Number of Actives	∆Group Delay 7.5-10 MHz	∆ Group Delay 35-37.5 MHz	∆ Group Delay 37.5-40 MHz
A1	na	-261 ns.	100 ns.	278 ns.
A2	na	-175	92	394
A3	na	-160	119	408
A4	3	-100	145	301
A5	4	-124	67	278
A6	4	-134	85	305

Table 1 Group Delay Variation Summary for System A

Node Number	Number of Actives	∆ Group Delay 7.5-10 MHz	∆ Group Delay 35-37.5 MHz	∆ Group Delay 37.5-40 MHz	∆ Group Delay 40-42 MHz
B1	6	-201 ns.	94 ns.	158 ns.	160 ns.
B2	5	-174	79	115	201
B3	4	-106	48	79	174
B4	3	-71	35	71	124
B5	3	-125	31	84	130
B6	3	-87	64	79	149
B7	3	-150	44	94	180
B8	4	-93	56	84	153
B9	4	-130	66	125	228

Table 2 Group Delay Variation Summary for System B

Node Number	Number of Actives	∆ Group Delay 7.5-10 MHz	∆ Group Delay 25-27.5 MHz	∆ Group Delay 27.5-30 MHz
C1	6	-114 ns.	26 ns.	76 ns.
C2	6	-154	65	175
C3	6	-160	13.5	77
C4	6	-172	41	48
C5	5	-106	23	64
C6	4	-127	33	135
C7	6	-132	52	96
C8	5	-130	22	25

Table 3 Group Delay Variation Summary for System C

the return band cable loss is exceedingly small. The periodic ripple on the magnitude plot of Figure 5 show a distinct echo.

Test Signals and System Clip

The levels at which networks clip can sometimes be determined by transmitting the reference signal at a high level and observing distortion products or compression in the received signal. This technique did not work reliably in system B because the distribution equipment (presumably the fiber optic transmitter) shuts the signal path down quickly when it senses a high level signal. This feature was probably installed to protect the laser diode.

SIMULATION RESULTS

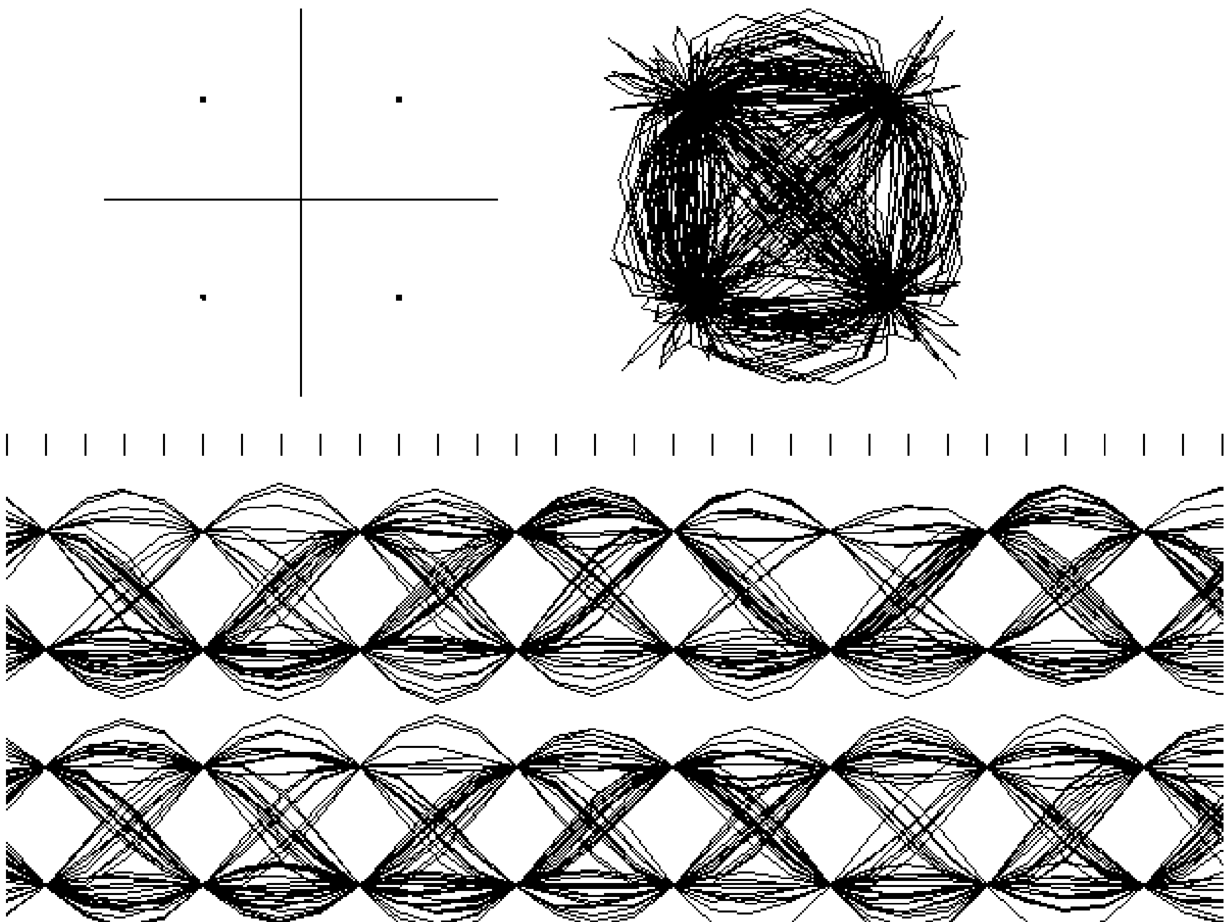


Figure 7 A QPSK Constellation Plot and Eye Diagram of an Unimpaired QPSK Signal

The next step is to take the measured results of amplitude non-flatness and group delay variation and evaluate their effect on signal integrity. To accomplish this task a simulation program was written in the C programming language to create a hypothetical test QPSK packet, distort it with group delay variation and amplitude non-flatness similar to the measured values, and demodulate the distorted signal. After the demodulation takes place, the impairment is characterized by MER (modulation error ratio). MER is a signal-to-noise related measure of the desired signal relative to the undesired noise and distortion components. MER is given by:

$$MER = 10 \cdot \log_{10}\left(\frac{\sum_{j=1}^{N}(I_j^2 + Q_j^2)}{\sum_{j=1}^{N}(\delta I_j^2 + \delta Q_j^2)}\right) dB \qquad (3)$$

where I and Q are ideal coordinates and δI and δQ are the errors in the received data

Modulation Type	QPSK
Packet Duration	179.2 microseconds
Symbol Rate	2.5 M Symbols/sec.
Number of Symbols/packet	448
Baseband Nyquist Frequency	1.25 MHz
Roll-Off Factor	0.25
Occupied Bandwidth	3.125 MHz
FFT Size	2048 points
Number of radians/sec per freq. sample	30.68E3 radians/sec.
Number of Time Samples/Symbol	4

Table 4 Characteristics of the Test Packet

points created by impairments. j is the symbol number.

Since no random noise was added in the simulation, the distortion energy comprises the undesired terms. Table 4 lists the characteristics of the test packet. This packet is similar to the modulation used for DOCSIS at its highest upstream rate.

The program works by creating a QPSK burst in the time domain comprising rectangular pulses and converting the burst into the frequency domain. In the frequency domain the burst is filtered to produce an inverse sin(x)/(x) pulse shape correction followed by raised cosine rolloff filtering using a rolloff factor of 0.25. Keep in mind that a multiplication in the frequency domain is equivalent to a convolution in the time domain.

Figure 7 (bottom) is a resulting set of I and Q eye diagrams showing an unimpaired channel with no group delay variation and a flat magnitude response. Note that the eye diagram is fully open. Figure 7 (upper left) is a constellation plot of four dots showing no constellation point spread due to linear distortion. Figure 7 (upper right) is a vector plot showing the trajectory of the instantaneous magnitude and phase over the duration of the packet.

In the frequency domain the phase is assumed to be distorted by the formula:

$$phase(n) = gdf \cdot n^2 + dly \cdot n + phase_O \tag{4}$$

where phase(n) is the phase angle as a function of frequency index, n. In this example n is equal to 30,680 radians/sec. “gdf” is a group delay factor, which produces a quadratic phase curve vs. frequency, dly is a time delay offset, and $phase_O$ is a static phase offset. The center of the channel is the reference point. The assumption that the group delay variation distortion creates a quadratic phase vs. frequency term will be less valid as the occupied band of the test signal increases.

The variation of group delay is the derivative of (4) with respect to frequency, evaluated over the channel. If n_2 is the top of the channel, and n_1 is the bottom of the channel, and gdf holds constant over the channel:

$$\begin{aligned} gd &= (2 \cdot gdf \cdot n_2 + dly) - (2 \cdot gdf \cdot n_1 + dly) \\ &= 2 \cdot gdf\,(n_2 - n_1) \end{aligned} \tag{5}$$

Figure 8 (bottom) is a resulting set of I and Q eye diagrams showing an impaired channel with a group delay variation of 333 ns. and a flat magnitude response. Note that the eye is partially closed. Figure 8 (upper left) is a constellation plot showing constellation point spread due to group delay variation. Figure 7 (upper right) is a vector plot showing the trajectory of the instantaneous magnitude and phase over the duration of the packet. If this level of distortion is placed on a 16-QAM modulated signal, the eye diagrams will be completely closed.

Another approach to modeling would be to use the measured magnitude and phase data from the plots to predict ISI. Interpolation would be needed to fill-in the vacant frequencies between sample points. This approach was not taken because any noise on the frequency response data would increase the ISI of the resulting demodulated packet.

Linear distortion caused by tilt on the magnitude vs. frequency plots is modeled by:

$$A(n) = 1 + m \cdot n \qquad (5)$$

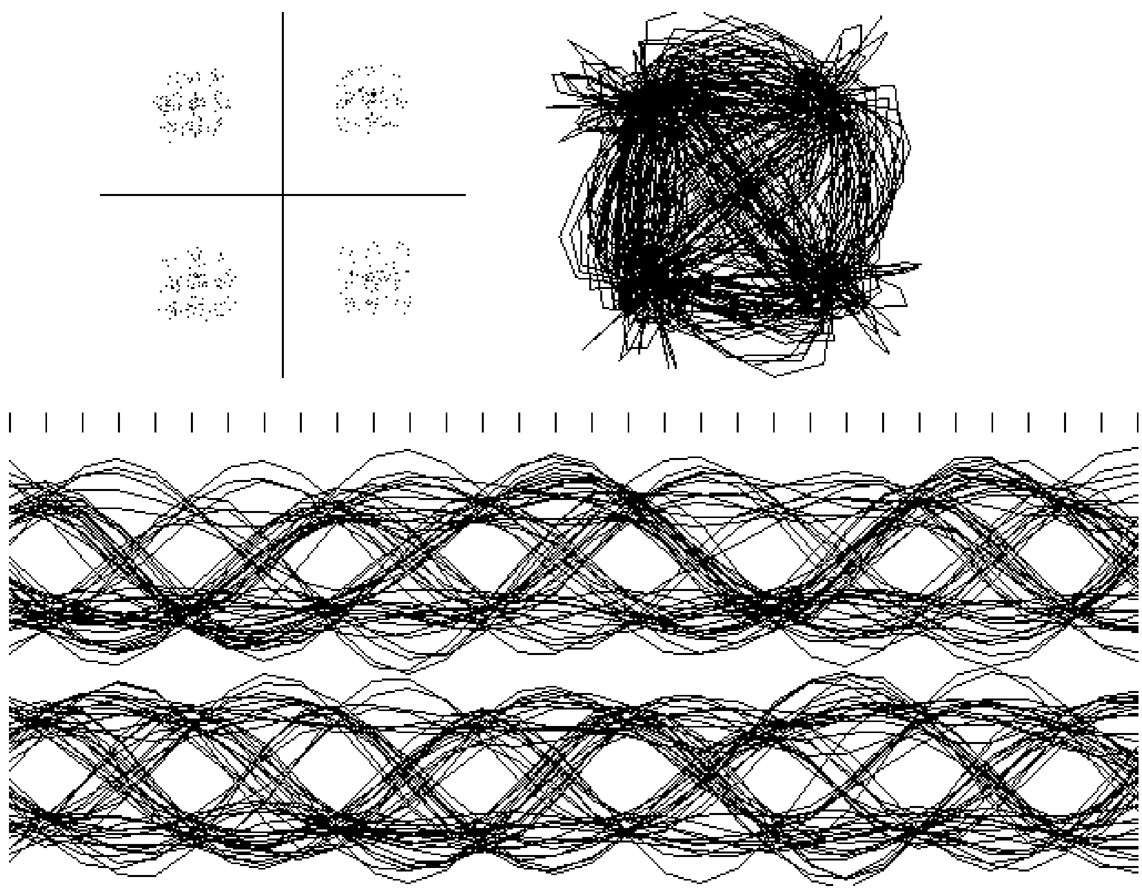


Figure 8 A QPSK Constellation Plot and Eye Diagram of a Signals Distorted with 333 ns. of Group Delay Variation over 2.5 MHz. , MER = 13.82 dB

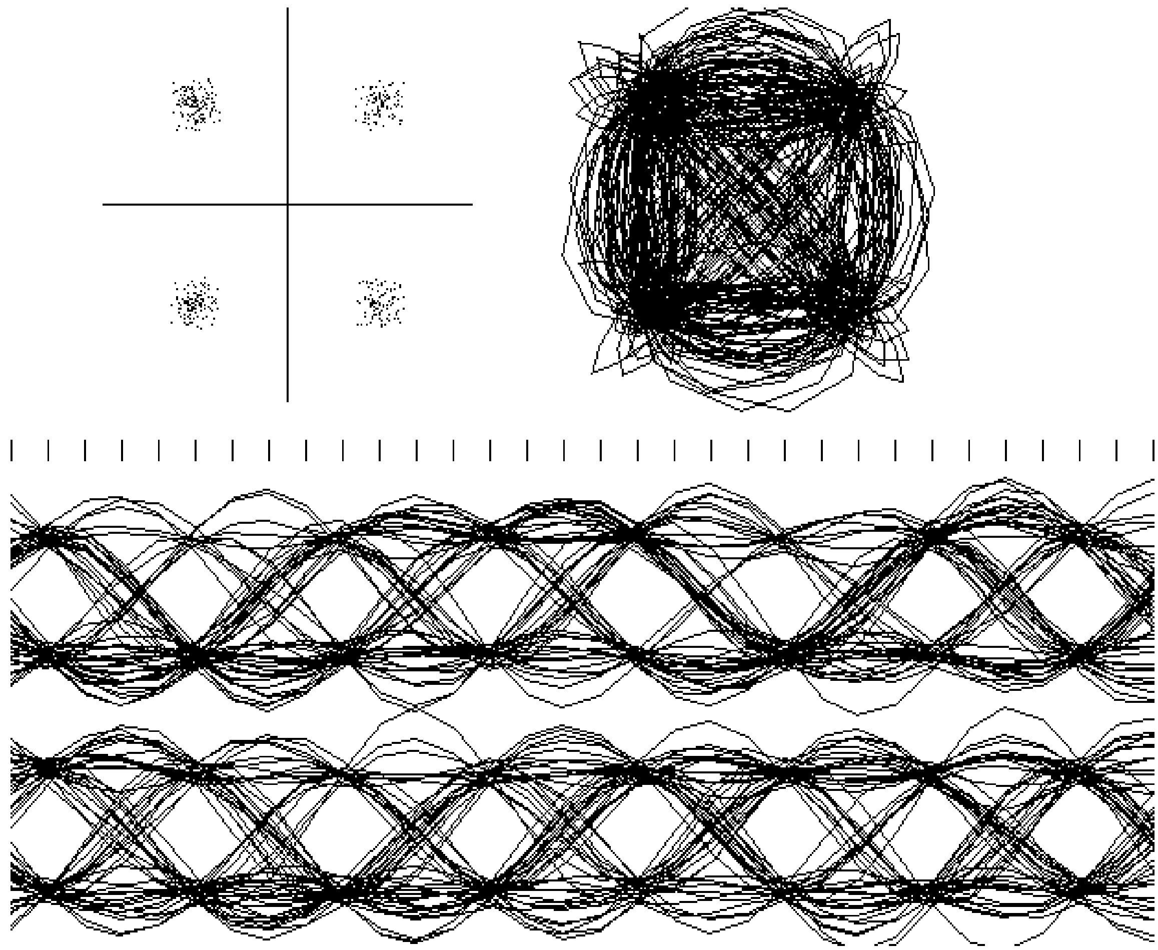


Figure 9 A QPSK Constellation Plot and Eye Diagram of a Signals Distorted with 4.4 dB of Amplitude Tilt over 2.5 MHz. , MER = 18.5 dB

where A(n) is the magnitude of the channel response vs. the frequency index, n. m is the slope factor. It is assumed that the center of the channel is the reference frequency. In decibels the channel tilt is expressed as:

$$tilt = 20 \cdot \log\left(\frac{A(n_2)}{A(n_1)}\right) \quad (6)$$

Figure 9 is a QPSK constellation plot showing the results of a 4.4 dB tilt over 2.5 MHz. with no group delay variation.

Figure 10 is a plot of MER vs. group delay variation showing the increasing corruption of a signal with increasing group delay variation. Figure 11 is a plot of MER vs. magnitude tilt showing that a channel with large amplitude tilt also has large intersymbol interference.

Patents are pending on this technology.

Observations

The data show that un-equalized 16-QAM is not going to work well over some parts of the 5-42 MHz band on some systems because of excessive group delay distortion. In particular, group delay variation below 10 MHz and above 35 MHz is too great in some cases, even for the relatively short cascades of new equipment tested. System A's upstream links have higher group delay variation than system B's links in the 37.5-40 MHz band. Furthermore, QPSK carriers will have degraded performance in the presence of random noise because of linear distortion. In evaluating the effects of amplitude tilt or group delay variation, the occupied bandwidth of the signal should be considered. If you reduce the occupied bandwidth of a signal by one half the level of amplitude tilt or group delay variation may be roughly reduced by one half.

Adaptive equalizers on the upstream path should fix problems with linear distortion

Any proposed adaptive equalization system should be thoroughly field tested prior to deployment to ensure that the system can work with real-world impairments, such as intermittent echoes.

CONCLUSION

This paper has presented field measurements on linear distortion in upstream cable plants. A simulation tool was created to quantify the effects of data measured in the field. The data show that there is a microreflection phenomenon visible on the upstream plant. The simulation shows that in many amplifier cascades the group delay variation at the high and low ends of the return band will stop 16-QAM without adaptive equalization and will degrade the performance of QPSK in the presence of noise.

Acknowledgments

The author wishes to thank the system operators, the technicians, the engineers, and technical reviewers for their support in gathering this data and reviewing this report.

References

[1] Digital Communication by E. Lee and D. Messerschmitt, p. 126, 2nd , Kulwer Academic Publishers

[2] Discrete Time Signal Processing by A. Oppenheim and R. Schafer, pp. 202-205, Prentice Hall

[3] Digital Communication by E. Lee and D. Messerschmitt, pp. 128-135 2nd ed, Kulwer Academic Publishers

Contact Information

Tom Williams
President, Holtzman Inc.
6423 Fairways Dr.
Longmont, CO 80503
303-444-6140
e-mail:tom@holtzmaninc.com

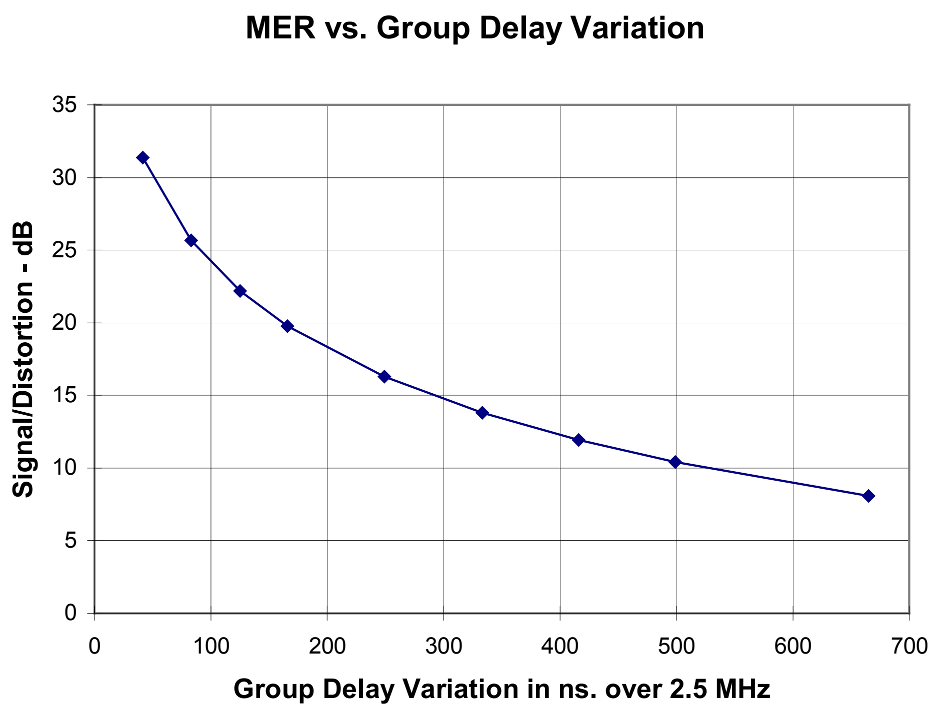


Figure 10 Eye Closure vs. Group Delay Variation over 2.5 MHz.

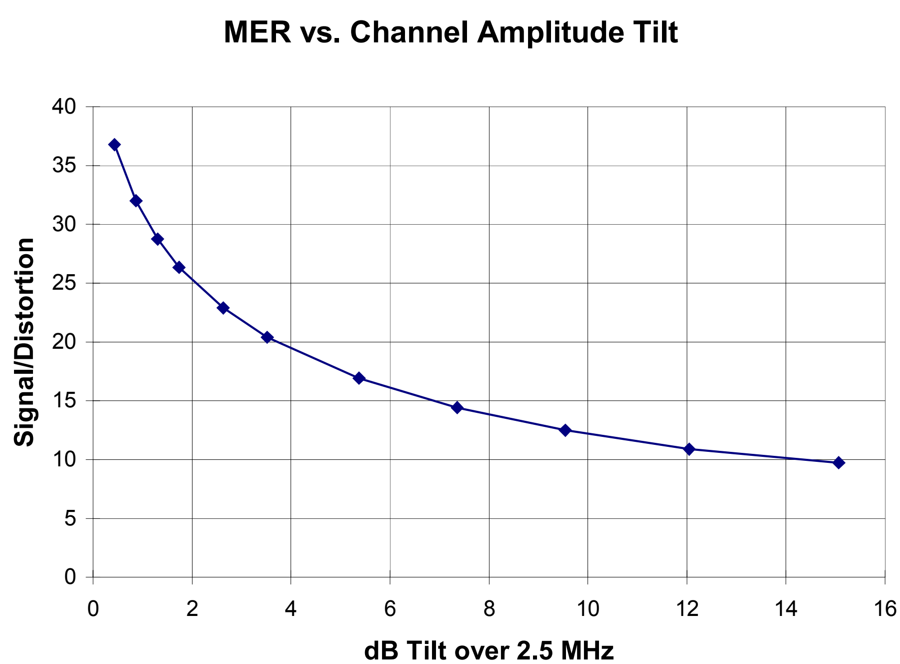


Figure 11 Eye Closure vs. Amplitude Tilt

Holtzman CABLE SCOPE

Captued:
On Tue, Nov 30, 1999 at 14:11:00

Unimpaired Reference Time Domain

0.050 Volts/Division

Power:41.25dBmv

Unimpaired Reference Frequency Domain

10 dB/div
Ref. Line: 60 dB

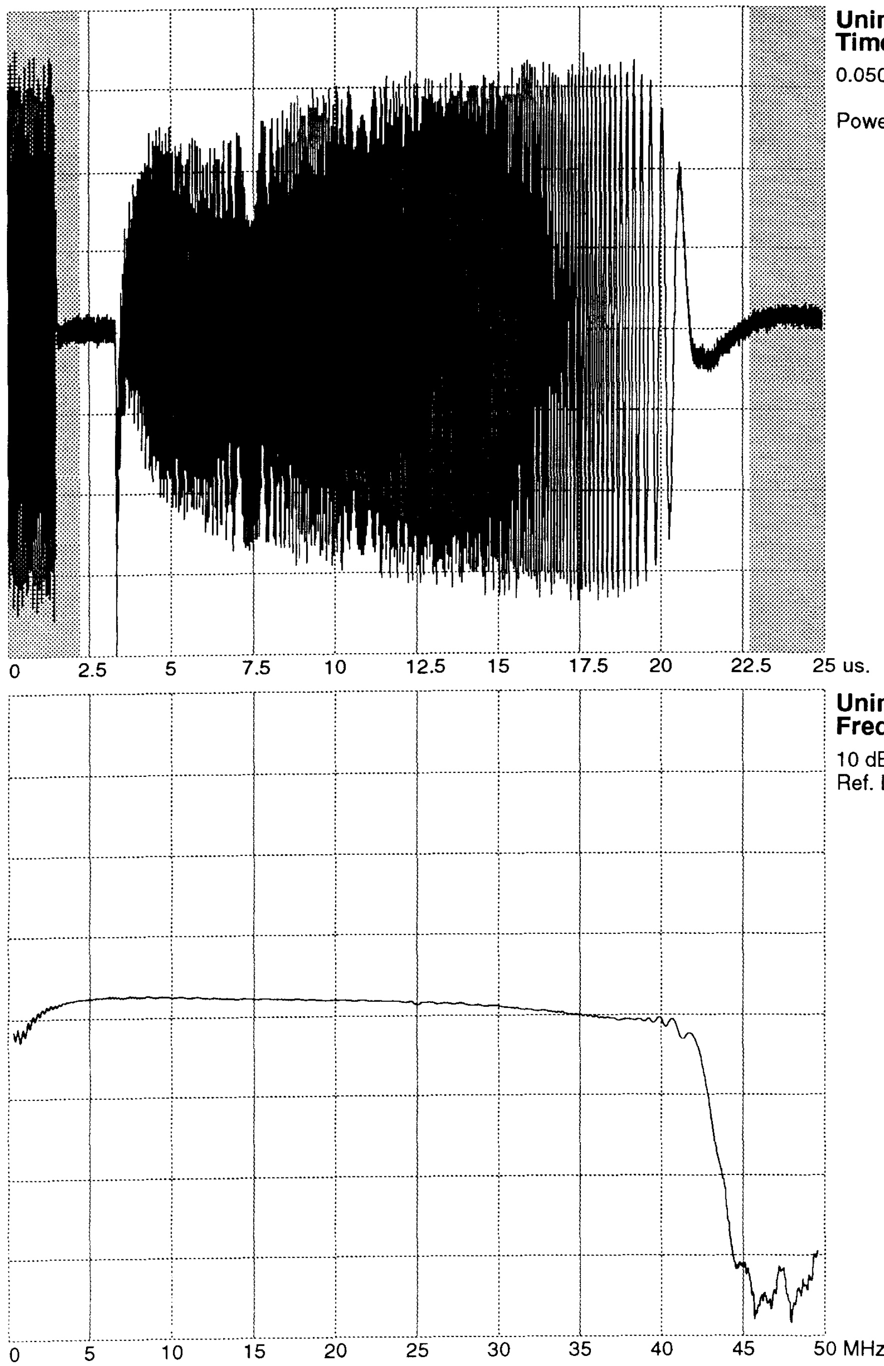


Fig.2 Unimpaired Reference Signal

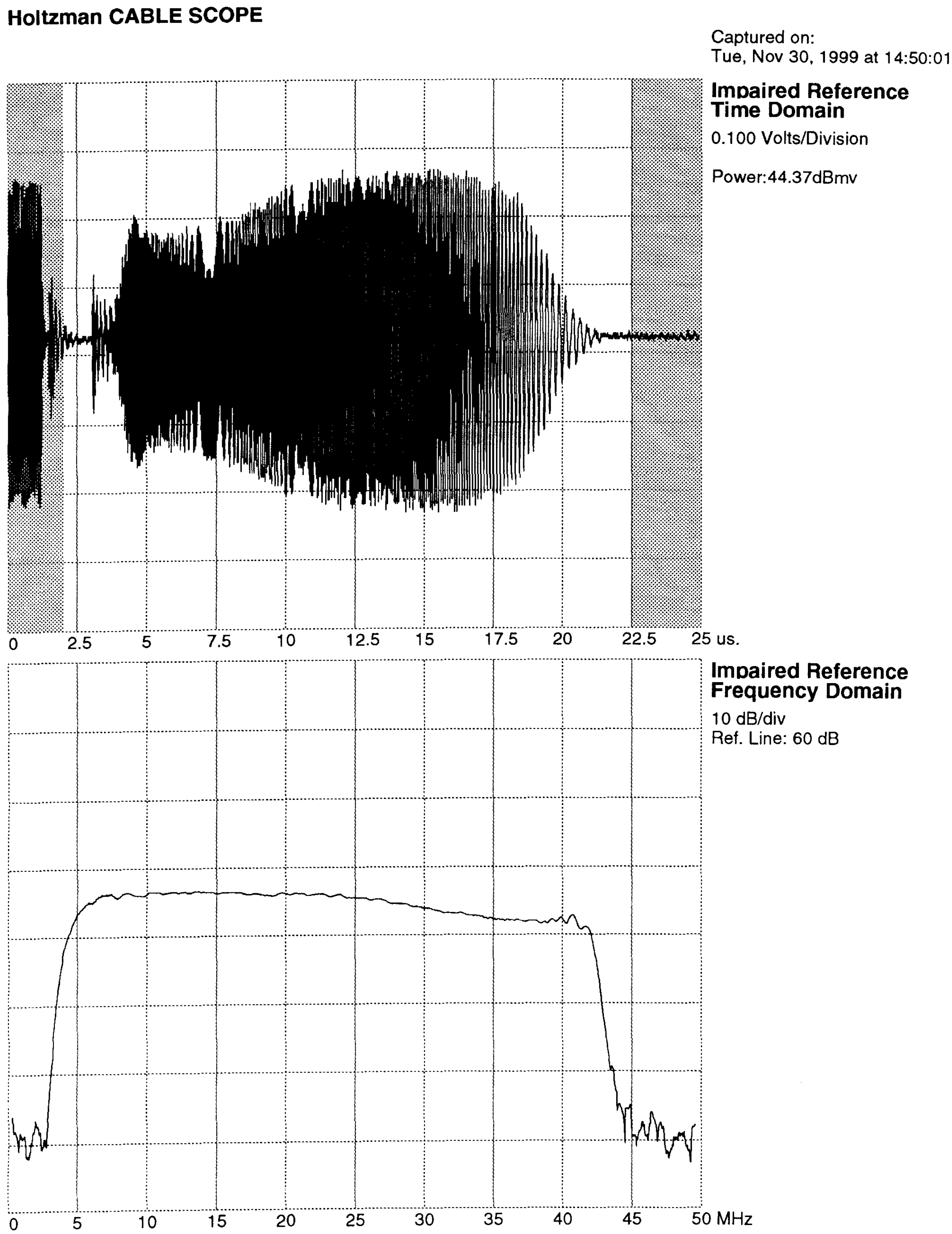


Fig. 3 An Impaired Reference Signal

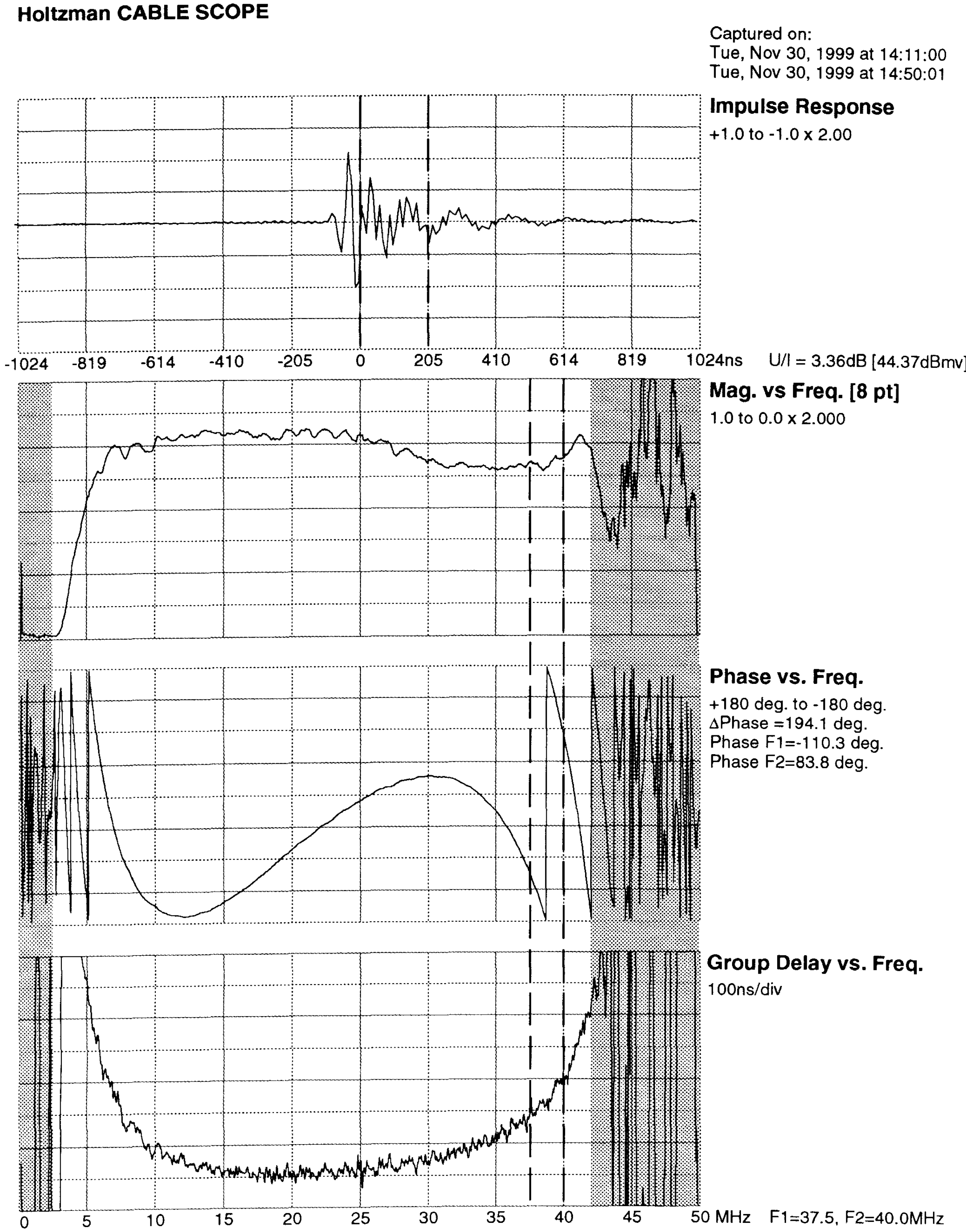


Fig. 4 Frequency Response System B

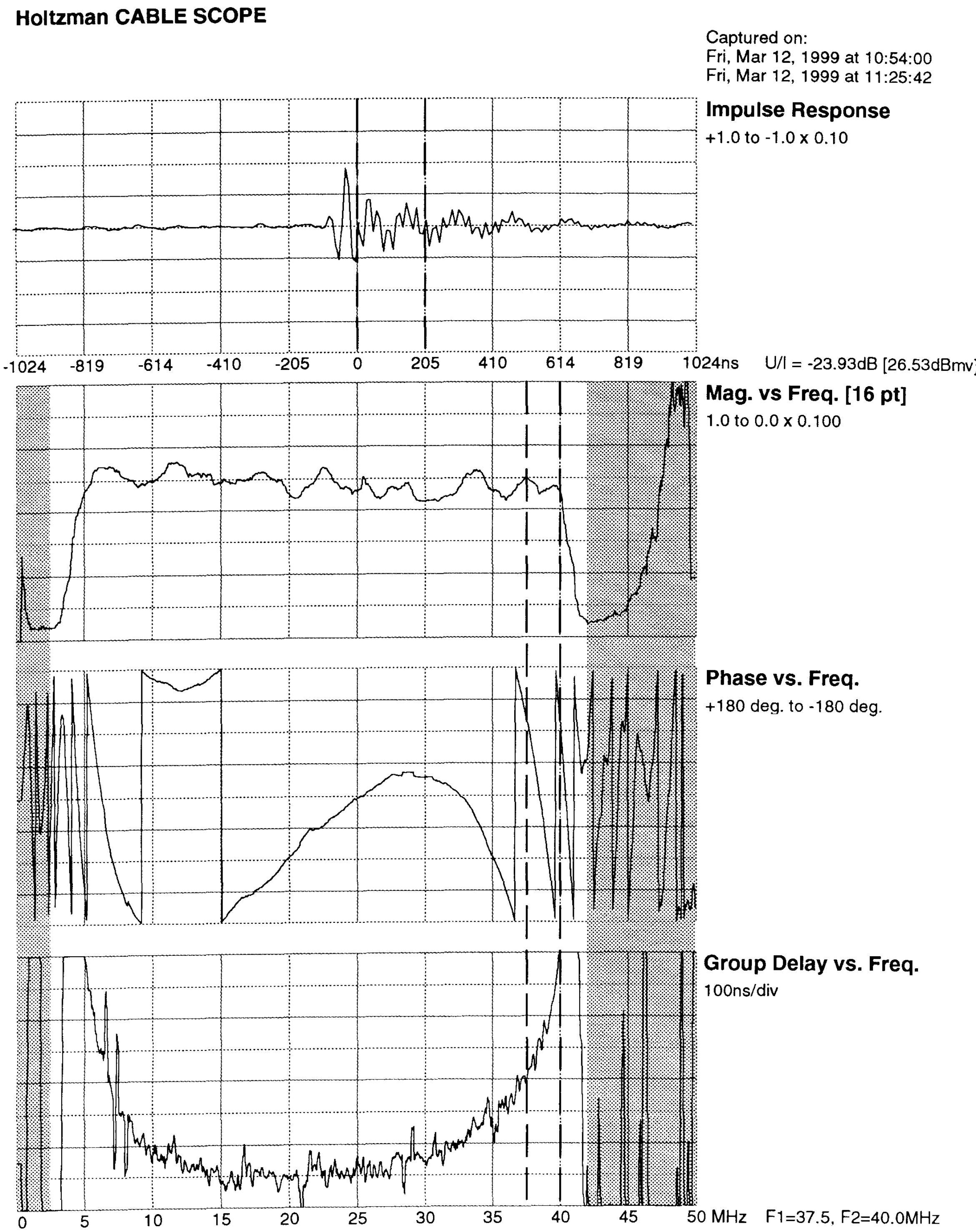


Fig. 5 Frequency Response System A

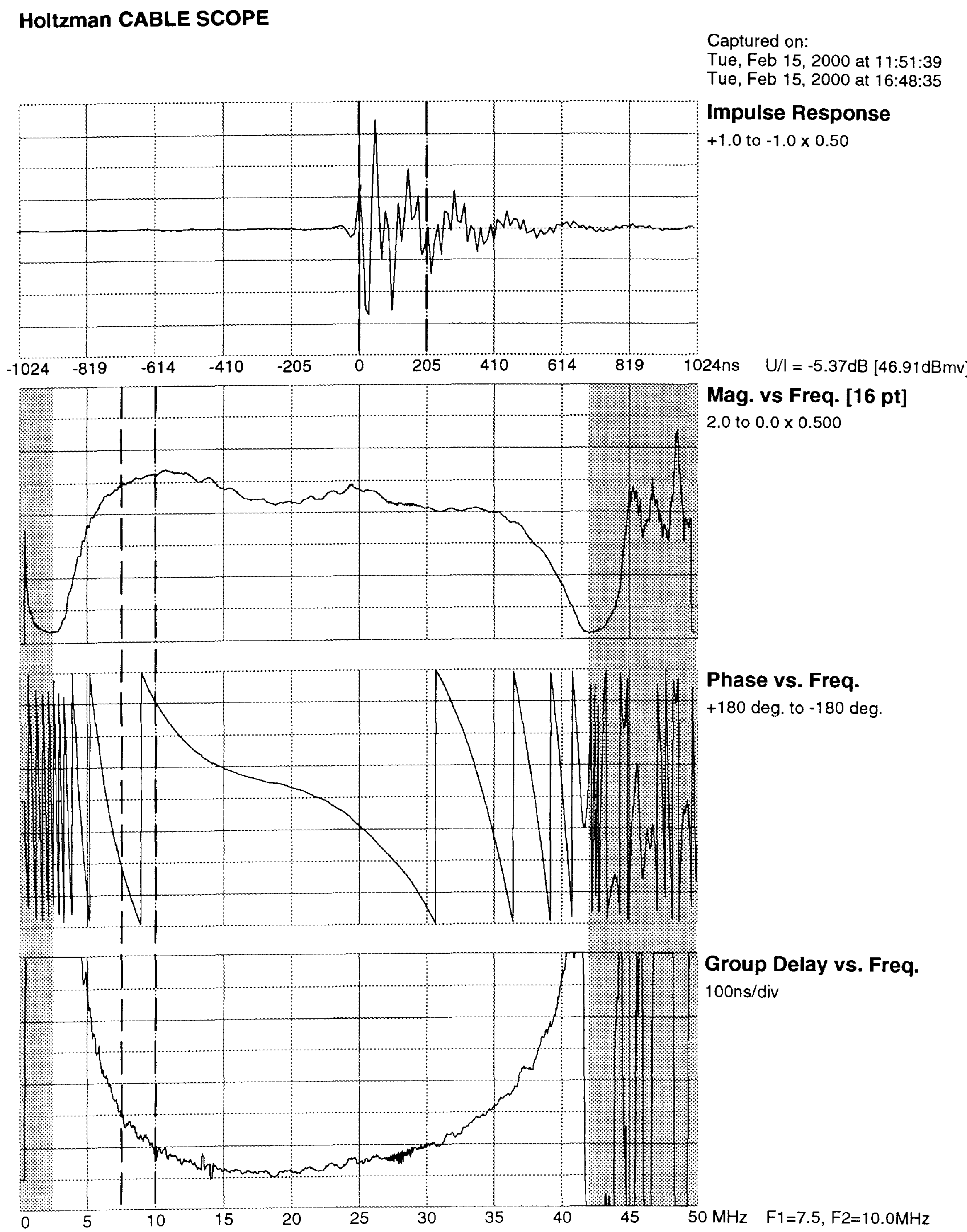


Fig. 6 Frequency Response System C

Streaming IP Media And Its Impact On The Digital Set-Top

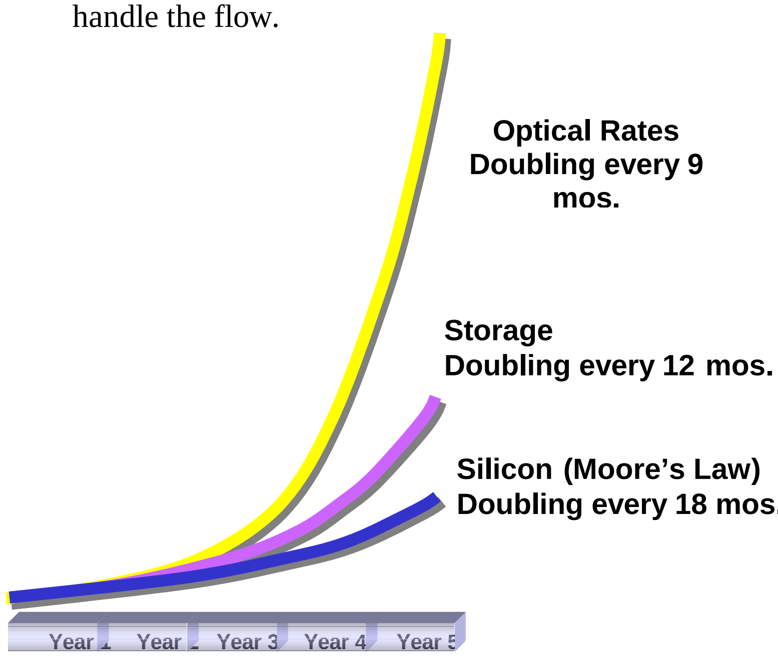
William E. Wall
Scientific-Atlanta Subscriber Sector

It's now clear that packet-based communications — using the global MPEG transport standard for downstream video transport and IP for data transport and interactive communications — will be the fundamental enabler of multiple communications media (video, data and voice) by a single broadband delivery network. This combination of bandwidth and open technology gives cable operators an unmatched competitive advantage, creating a network which is, in essence, a giant "hard drive" loaded with executable applications, all ready to be called at any time by any connected set top.

Bandwidth Explosion

What has been the catalyst behind packet's rapid emergence? For 20 years, the industry has been part of a rapidly escalating technology revolution, culminating in a "bandwidth explosion" that has altered forever the types of services we can now offer to our customers. Just consider the reductions in size and cost of computer chips that have occurred; with the corresponding exponential improvements in memory, processing and graphics power.

Together, these innovations have caused a huge "pull" from client side because all this power is creating demand for new bandwidth-intensive applications and services. So too has there been a "push" from the backbone side. Network servers have become much more affordable and powerful. Fiber is now deployed throughout the network, and it too has increased dramatically in capacity. All in all, the broadband pipeline today is truly primed to handle the flow.



Source: Lucent

Up to this point the bottleneck has occurred within the access portion of network — the last mile and the set-top at the home. However, with the lower cost of silicon, added memory, advent of hard disk drives, network capabilities, etc, that too is changing. Today it is not only possible, but essential for the set-top to be well connected to the IP network

Streaming Media Applications

We are really just beginning to develop a large category of applications that are going to proliferate and require streaming IP to the set top:

***-Internet TV*:** E-mail, Chat, Web browsing, Personalized Services (e.g. AOL TV), Walled Garden, Watch n' Surf, E-Commerce

-Home Gateway: MP3, PDA, Printer, Digital Camera, Web Pad, PC, Home Networks

-Communications: Voice, Videophone, Messaging

-Games: Interactive, Multi-player, Integrated with video

IP Delivery Mechanisms

Three basic mechanisms for delivering IP data to the set top have been developed and will be available either as standard or optional within the next generation digital set-tops — Out-of-Band IP, DOCSIS and In-Band IP. These technologies are not mutually exclusive, but in fact are very complementary and can be employed in the same network, running on the same set- tops. Each one however, does possess unique advantages and disadvantages, making it better suited for certain operators, running certain applications at certain times.

(See Figure 1)

Out of Band IP Protocol (OOB IP)

The oldest of the technologies, OOB IP is available on all set tops utilizing the DAVIC protocol. Early on a need was determined for a permanent connection available for signaling, conditional access, and other system related usage, regardless of whether the tuner was tuned to analog, digital, or data channels. Using the DAVIC channel, OOB IP is a relatively low-bandwidth channel (compared to 27 Mbps) delivering a payload of 1.2 Mbps. System usage consumes About 300 - 400 kbps, leaving about 800 -900 kbps of IP bandwidth available for other applications. That may not seem like a lot, but keep in mind that until now, most people connect to the Internet at rates below 56kbps. Thus, OOB IP falls into the niche of very good for relatively low-bandwidth applications that need to be always connected, like e-mail, limited web browsing, and instant messaging services.

DOCSIS

Originally developed as a standard for cable modems, it's incorporation into digital the set top has been the most recent. The advantage of DOCSIS is high speed — with 27 or 38 Mbps bandwidth available downstream, and up to 10 Mbps upstream, depending on network capacity and quality of the digital plant it is installed in. Typical data rates used in the field are 4 mbps upstream.

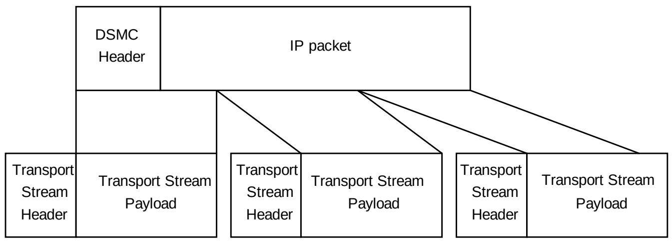
DOCSIS enjoys enhanced bandwidth over OOB IP but in many ways emulates the same sort of functioning of an always-on, always-connected service. DOCSIS can be used by operators who want constant high-speed connectivity to the Internet and for transporting telephony services, video conferencing, downloading MP3 audio or video clips. One disadvantage of DOCSIS is that is does require an additional 6 MHz tuner and additional hardware within the set-top to support the DOCSIS channel. The second disadvantage, which is also true of OOB IP, is that they are fixed services on a single channel that all set-tops in that portion of the network are connected to. And once that channel's bandwidth is all used up, operators will have to add equipment or subdivide the network to obtain more capacity.

In-Band IP (IP Gateway)

A relatively new implementation developed as a way to extend the bandwidth limitations of OOB IP, In-Band IP uses standard multiprotocol encapsulation of IP data into MPEG transport along with video and other types of services. As such, a portion of or an entire 6 MHz channel, or more can be dedicated to IP services. And it does this while making use of the existing tuner in the set-top eliminating the additional tuner cost that DOCSIS requires.

In-Band IP delivery is based on international signaling and encapsulation standards that are being implemented in networks today. Protocols enabling the setup of client/server data sessions over the QAM channels plus the use of broadcast data carousels in the network are established according to standards under Digital Storage Media-Command and Control (DSM-CC) MPEG Part 6 (ISO/IEC 13818-6).

International Standards
Encapsulation



DSM-CC Header contains destination MAC

Digital video and audio services are broadcast directly in MPEG transport. In hybrid video and data networks, IP may be encapsulated in MPEG transport packets using the protocols standardized in DSM-CC for routing over the HFC network. These multiprotocol encapsulation standards also have been adopted by ATSC and DVB, and are in use by various satellite and cable data delivery systems worldwide.

Frequency agility is a big advantage of In-Band IP. This allows an operator to dynamically switch IP connections from one frequency to another and reassign bandwidth virtually on the fly. As such, In-Band IP offers true on-demand functionality, with the ability to match bandwidth devoted to IP traffic to actual demand at the moment (e.g. traffic heavier at night) without having to re-engineer the plant.

Another potential advantage of frequency agility is with Open Access issues, where multiple ISPs are going to be allowed onto the network. Under this scenario, it will be advantageous to be able to assign certain amount of bandwidth on demand to a particular ISP for a duration or time. In-Band IP's User to Network DSM-CC signaling allows you to assign blocks of bandwidth to ISP's for short or long periods of time. An operator can know how much bandwidth an ISP is using, as well as being able to guarantee the service will not be interrupted by bursty traffic. With OOB IP and DOCSIS, ISPs can be allowed in to share bandwidth, but there are not control mechanisms to determine how much bandwidth they are utilizing.

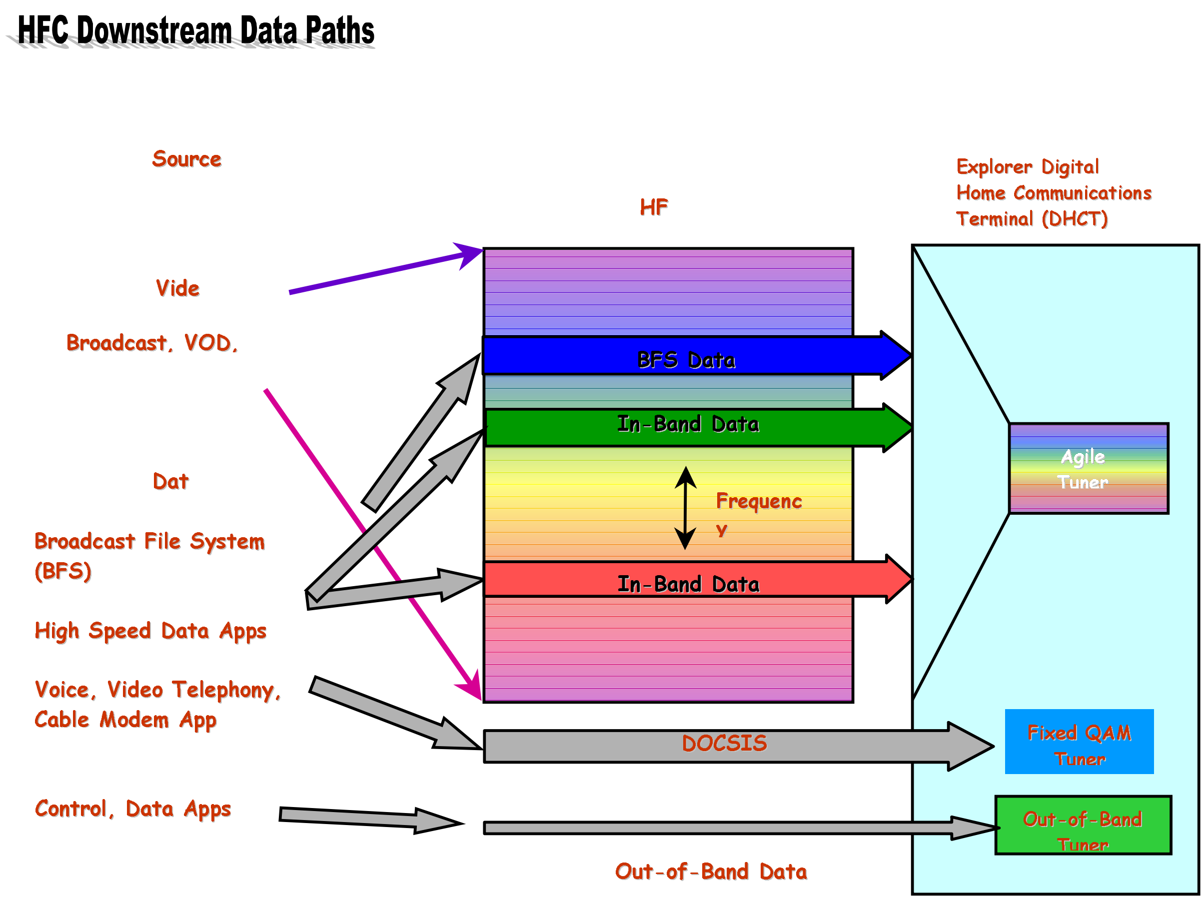
In-Band IP not only allows you to share bandwidth, but also share common equipment as well. QAM modulators, for example can be used for either VOD or IP services. Depending on the instantaneous demand on the system, the system can dynamically switch the content on a QAM modulator, (x percentage of IP, y percentage of VOD) to share and balance peak loads.

SUMMARY

In Band delivery of IP data to set-tops tops provides an attractive alternative to DOCSIS and complements OOB IP delivery mechanisms. By utilizing multiprotocol encapsulation and DSM-CC signaling, In-Band

IP provides operators with an inherently guaranteed quality of service, while allowing them to dynamically reassign bandwidth on the fly to match system demands, and to share bandwidth with other targeted services, making for high equipment efficiency. No extensive system reconfiguration or re-engineering will be necessary to implement In-Band IP, as it works through software download to the existing base of interactive digital set-tops which are already deployed throughout networks today.

Figure 1



HFC Downstream Data Paths
Source
Vide
Broadcast, VOD,
Dat
Broadcast File System (BFS)
High Speed Data Apps
Voice, Video Telephony, Cable Modem App
Control, Data Apps
HF
BFS Data
In-Band Data
Frequenc y
In-Band Data
DOCSIS
Out-of-Band Data
Explorer Digital Home Communications Terminal (DHCT)
Agile Tuner
Fixed QAM Tuner
Out-of-Band Tuner

Streaming Over HFC—MPEG-2 or IP or Both?

Mukta L. Kar and Bill Kostka
Cable Television Laboratories, Inc.
Majid Chelehmal
Terayon Communication Systems
Munsi Haque
Philips Semiconductor

Abstract

With the rapid progress in science and technology, more and more applications involving audio and video streams are emerging. The MPEG-2 compression standard, coupled with advances in digital modulation, has already made an impact in the broadcast industry by increasing the number of channels and viewing quality. However, consumers have limited choice—purchasing or renting CDs and DVDs from a limited stock or selecting from limited movie and music offerings.

Digital technology has advanced in a number of fronts, including the transport of data over public or private networks. High-speed data can be transported in a variety of ways—unicast, multicast, or even broadcast using various transport protocols. Digital information, such as audio, video, text, graphics, etc., differs only in the size of data. For example, digitized video is much larger than digitized audio. Advances in fiber technology have increased the capacity of both public and private networks to the point that audio and video streaming is becoming a reality. CableLabs, and other reseach institutes, have found that there are a few obstacles to overcome before broadcast-quality video streaming is possible over public networks. When those issues are resolved, electronic search engines will be able to find the desired content in archives located anywhere in the world. Consumers will be able to enjoy their content choices with a click of a button.

One major issue is the transport mechanism (protocol) for real-time, uninterrupted flow of digital audio, video, graphics, etc. MPEG-2 transport protocol and IP-based Real-time Transport Protocol (RTP) are currently the two leading protocols used for the delivery of digital content in real-time. This paper will analyze these two protocols in detail, and will present comparative studies for streaming technology. This technology is not mature; e.g., streaming in IP networks is implemented as part of transmission application layer protocols, where the unreliable user datagram protocol (UDP) is used mostly at the transport layer. To alleviate packet loss, RTP with quality-of-service (QOS) routing is considered for improved services in IP networks.

Issues and obstacles for streaming audio-visual content, particularly over the public networks, will be addressed in this paper. Research activities addressing some of the problems also will be discussed. Finally, content streaming based on MPEG-4, the recently completed multimedia standard, also will be discussed.

Introduction

Until a few years ago, cable systems were one-way networks used to deliver premium analog television content. Equipment used in cable plants was proprietary in nature. As a result, the majority of the equipment was not interoperable and, therefore, not portable across cable systems. Advances in digital technology have brought about a revolution in the computer and communications industries, including cable networks. Digital television signal compression technology and MPEG-2 standardization, coupled with digital modulation, have ushered in a new era in television broadcasting. One of the most important benefits is the bandwidth efficiency in spectrum utilization compared to analog broadcasting. An existing 6-MHz analog channel can be used to send multiple digital channels with equal or better picture and sound quality. This indicates that the existing limited number of analog channels can be transformed into a larger number of viewing or logical channels. These logical channels can be used for the delivery of audio, video, and high-speed data services. Digital technology has transformed the analog cable network into a broadband multimedia delivery system.

Prior to the addition of digital technology to existing analog systems, the cable industry had researched transport protocols and digital modulation systems for transmission of compressed digital audio/video over the cable networks. About the same time, the MPEG-2 standard [1] was completed by the moving picture expert group (MPEG) of the International Organization for Standardization (ISO). Unlike MPEG-1 Systems [2], which deal with a single program, MPEG-2 systems can handle multiple programs and have an added transport layer optimized to broadcast digital audio and video synchronously. Because cable systems primarily deliver premium television content, the MPEG-2 transport system became cable's primary choice. Also, hybrid fiber/coaxial (HFC) cable networks are much less susceptible to atmospheric noise; a better signal-to-noise ratio is available compared to other broadcasting systems. To take advantage of this, the cable industry chose to use a higher order digital modulation system, 64-QAM or 256-QAM (quadrature amplitude modulation), which transforms a 6-MHz physical spectrum into a larger digital bandwidth (27 Mbps for 64 QAM and 38 Mbps for 256 QAM) described in ITU standard J.83-B [3]. Digital technology vastly increases the channel capacity of the existing cable plant. This increased capacity will allow the delivery of an increased number of TV channels and other digital services as well.

In adding digital capability, cable networks are equipped for MPEG-2 transport at the baseband. Figure 1 shows a simplified diagram for delivery of digital TV signals using baseband MPEG-2 transport. The multiplexer's output is an MPEG-2 systems-compliant transport stream consisting of MPEG-2 transport packets. At the modulator, the transport packet payload, excluding the packets carrying system information, may be encrypted (optional). Forward error correction (FEC) is applied to protect against noise in the transmission channel. The resulting bitstream then digitally modulates a 6-MHz carrier. Addition of FEC and digital modulation is described in ITU Standard J.83-B. The modulated carrier is upconverted to a desired cable channel before combining with other channels for transmission over the cable network. In the upstream (return channel), baseband transport is similar to downstream transport, except that lower order modulation (e.g., QPSK) is used. Upstream transport is used for management messages and user interactivity, and is shared by many set-top boxes. QPSK provides more robust modulation so that packet loss due to noise will be minimal.

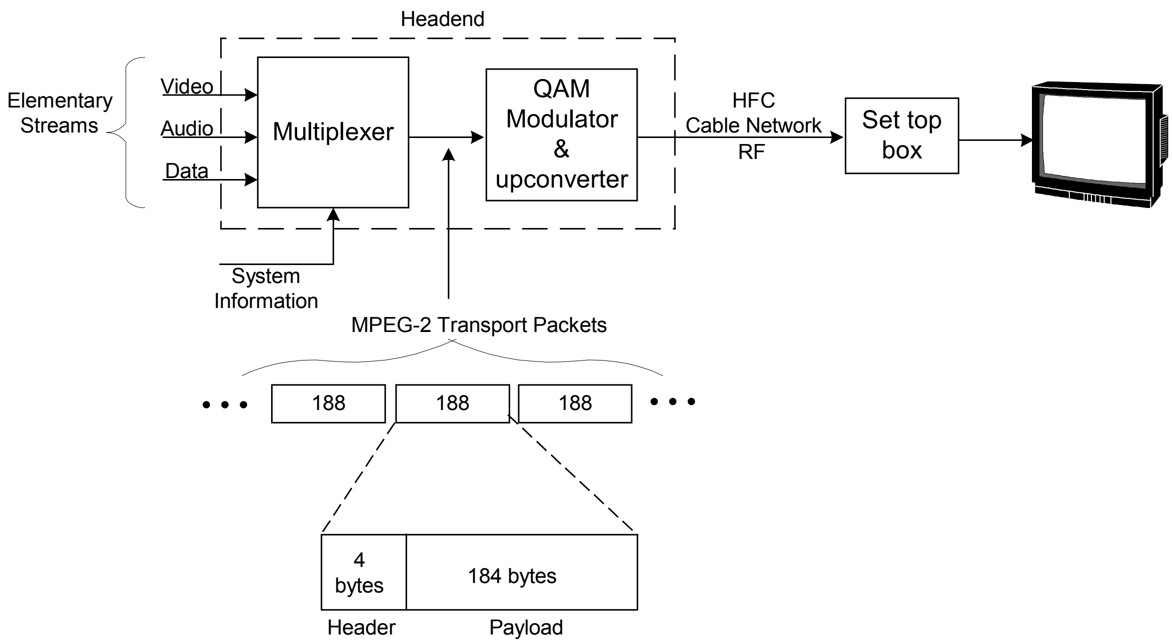


Figure 1. Broadcast of MPEG-2 audio/video over HFC cable network (downstream)

In recent years, high-speed Internet connectivity has become a popular service. Internet connectivity allows use of digital services available over the Internet, such as web-browsing, web-casting, email, home shopping, chat, etc. The cable industry decided to add Internet connectivity to cable networks, but services over the Internet require two-way networks. In adding digital technology, many cable systems have been upgraded to provide two-way connectivity. The Internet is primarily a high-speed data network and uses Internet protocol (IP) transport [11], not MPEG-2, as the baseband transport. Since the early days of cable, video delivery has been a primary source of revenue and, even today, the major part of cable systems' revenue comes from video delivery. Figure 2 shows a simple diagram of the interface between a HFC network and the Internet. CableLabs, in collaboration with its members and the vendor community, have developed an interface standard known as Data Over Cable Service Interface Specification (DOCSIS) [12]. The main objective of the DOCSIS specification is to enable many vendors to design interface equipment in a competitive market place. Equipment designed based on a standard will be interoperable, portable, and available in retail markets. Per the DOCSIS standard, IP packets are encapsulated into MPEG-2 transport packets before transmission over the network in the downstream direction. If the payload size is larger than 184 bytes, the packet is broken into smaller ones before being sent as MPEG-2 transport packet payloads. Conversion of IP packets to MPEG-2 packets is performed per following rule.

If $\mathrm{mod}(L, 184) = 0;\ N = L/184$
$\mathrm{mod}(L, 184) \neq 0;\ N = (L/184) + 1$

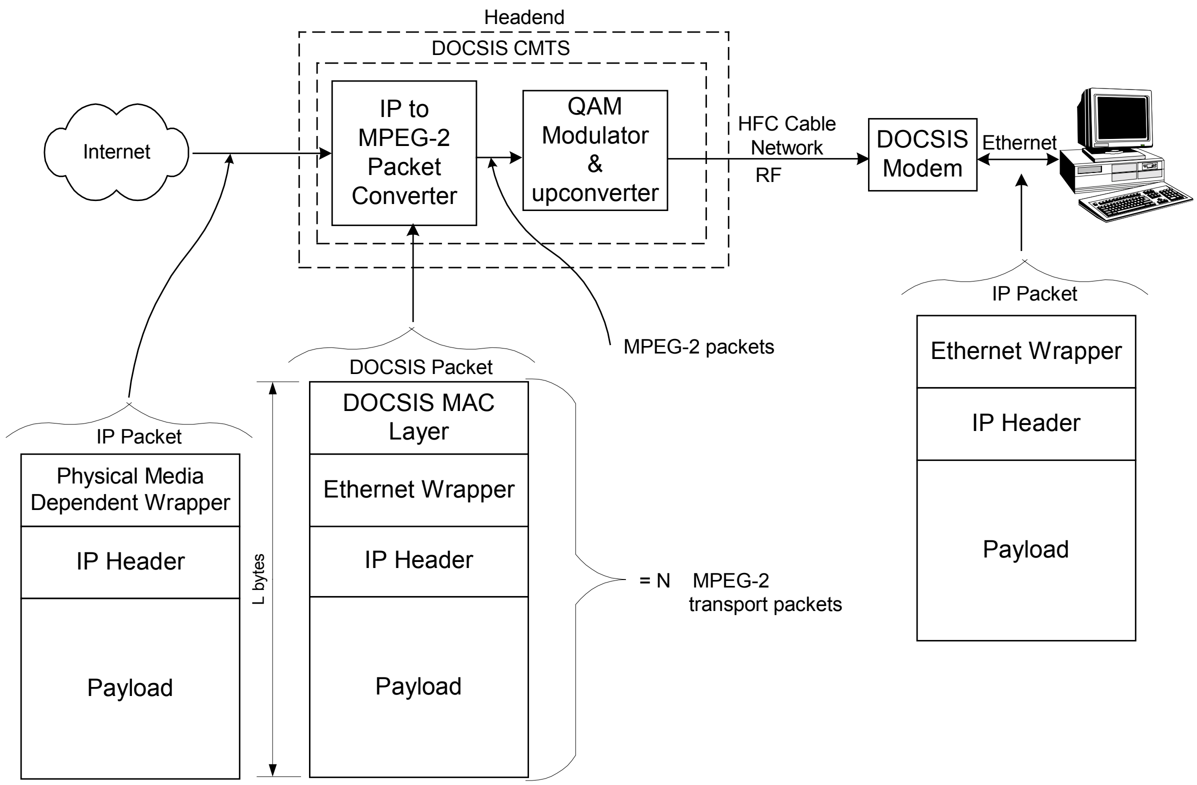


Figure 2. DOCSIS Transport (downstream)

Before delivery to the end user, IP packets are reconstructed by a DOCSIS modem (also known as a cable modem). The cable modem termination system (CMTS) at the headend, and the cable modem at the subscriber's end, make the cable network transparent to the user and make the user feel as if he/she is using a local area network (LAN). This also will enable multiplexing audio/video packets and IP packets in the same physical channel, if necessary.

In the MPEG-2 systems syntax and semantics are defined for multiplexing multiple programs in a single bitstream; the transport layer defines the semantics (or mechanism) for real-time delivery of multiple programs over error-prone channels. MPEG-2 Systems include the following functions:

- Timing and synchronization. The transmission of timing information in transport packets allows the receiver to synchronize with the encoder clock, which in turn helps audio and video to synchronize, avoiding lip-synch errors, etc.

- Packetization. The segmentation and encapsulation of elementary data streams into 188-byte transport packets. Included with each packet is the 4-byte packet header, which allows easy identification of each packet at the output of the transport demultiplexer.

- Multiplexing. The mechanisms to interspace transport packets of various elementary streams and program specific information (PSI) into a serial bitstream that complies with the MPEG-2 T-STD (transport stream decoder) model. This means timed delivery of the audio/video packets. The PSI information in the bitstream is used by the demultiplexer to demultiplex elementary streams uniquely at the decoder.

- Conditional Access. Provision for inclusion of access control information in the transport multiplex.

In the system/transport stream layer, bitstreams are split into 188-byte packets (a 184-byte payload and a 4-byte header) as shown in Figure 1. The header carries various information fields as shown in Figure 3. The PID (packet identifier), the most important field, has a length of 13 bits. The PID is a unique integer number associated with an elementary stream in a single or multi-program transport stream. The packets can carry various video and audio channels and other information, such as synchronization and timing, encryption, program information, access control, etc. The PID numbers help sort the packets into the specific streams to which they belong.

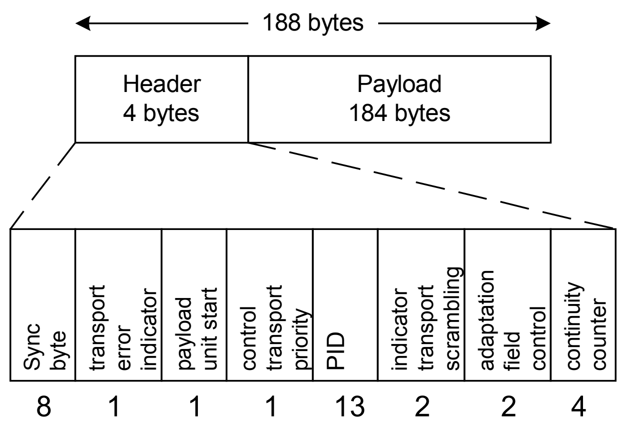
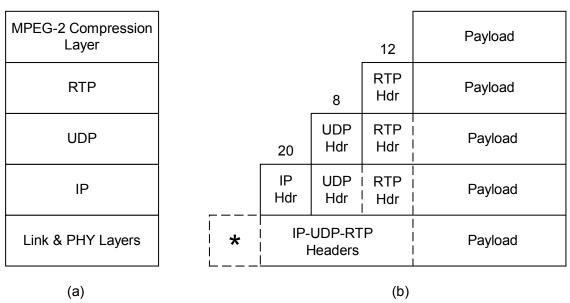


Figure 3. MPEG-2 Transport Packet

RTP—A Transport Protocol for Real-time Applications

The Moving Picture Expert Group (MPEG) has been active in creating international standards for compressed digital audio and video, and for the delivery mechanism (MPEG-2 transport)—a parallel effort was started by the Internet Engineering Task Force (IETF) for transport of data across various heterogeneous networks for non-real-time delivery. The transport standard created by IETF is known as Internet Protocol (IP). This protocol is widely used today for routing data across wide area networks (WAN), which may encompass several heterogeneous networks covering the entire world.

Various application-specific protocols were developed to take advantage of the IP protocol. TCP and UDP are the two dominant ones that sit on top of the IP layer. TCP is a connection-oriented protocol and, therefore, has additional, specific messages—a protocol for applications to request distant connections and a means for destinations to identify that they are ready to receive incoming connection requests. UDP provides a connectionless, unreliable service [10]. TCP was developed for guaranteed delivery of packets, whereas UDP does not. TCP is not well suited for real-time delivery of data because re-transmission conflicts with timed delivery of data. Timed delivery is an important requirement for real-time transport of audio and video. For that reason, UDP is the preferable mode of transmission for real-time applications. Below the IP and UDP headers, data-specific information is needed to convey payload-related (i.e., video, audio, etc) information. This additional information adds to the amount of the payload overhead. The protocol developed on top of actual payload of video, audio and data is known as real-time transport protocol (RTP) [4–6]. Figure 4a shows the generic packet format for RTP. Figure 4b provides an estimate of overhead from various layers such as IP-UDP-RTP. The bottom row of Figure 4b shows the payload and total overhead for a typical RTP packet.



*A media and protocol dependent header. For Ethernet the header size is 14 bytes.

Figure 4. RTP/UDP/IP Packet Layers

RTP may be used for unicast or multicast network services. Services provided by RTP include payload type identification, sequence numbering, time stamping and delivery monitoring. RTP does not provide a mechanism to ensure timely delivery or provide other quality-of-service guarantees, but relies on lower-layer services to do so. It does not guarantee delivery or prevent out-of-order delivery, nor does it assume that the underlying network is reliable and delivers packets in sequence.

RTP consists of two distinct, closely linked parts:

- The real-time transport protocol (RTP), to carry data that has real-time properties.
- The RTP control protocol (RTCP), to monitor quality-of-service and to convey information about the participant in an on-going session. Basically, it is a feedback mechanism from the receiver to the sender, related to the number of packets lost, etc. For example, loss of packets may be proportional to traffic congestion of a network segment; the server may take appropriate action to minimize packet loss.

The fixed header size of the RTP packet [4], [6] is 12 bytes, as shown in Figure 5. No header compression is assumed here. To reduce overhead, the size of the payload in a RTP packet can be increased, as range of payload size is 64 bytes to 1518 bytes. But if a packet with a large payload of compressed audio or video is lost, it may create respective audio/visual artifacts. Instead, video and audio of same time duration may be bundled in the same packet to keep the packet to an optimum size so that overhead is tolerable. This type of packaging is known as bundling or generic multiplexing of elementary streams in the same packet [9]. An extra four bytes in the header are required to indicate an offset of the audio payload and other audio-related information.

RTP Packet Header

The first eight bytes of the RTP header are:

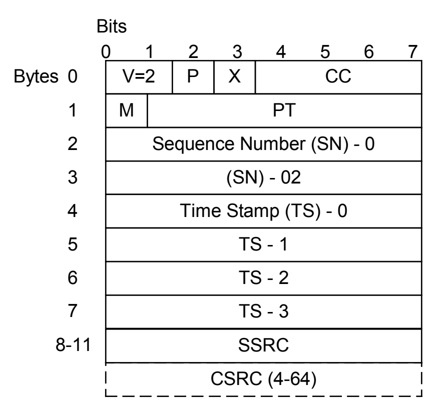


Figure 5. RTP Packet Header

The next four bytes comprise the synchronization source (SSRC) identifier and the next four to 64 bytes may comprise the contributing source (CSRC) identifiers.

RTP Payload

The payload size depends upon the audio-video codec type. For most standardized codecs, there are also payload headers [5], [8], as mentioned below, that immediately follow the fixed RTP header. As an example, the payload header for JPEG (RFC 2435, October 1998, RFC 2035, September 1996) streams consist of a 8-byte-long header, called the "main JPEG header," followed by other related headers, such as, Restart Marker header, Quantization Table header, etc. Similarly there are payload headers for H.263+ (RFC 2429, October 1998), H.263 (RFC 2190, September 1997), H.261 (RFC 2032, October 1996), MPEG-1 / MPEG-2 Video /Audio (RFC 2250, January 1998), etc.

The RTP payload size depends upon the frame-sizes of the access units (compressed frames). RFC1889, January 1996 (RTP: A Transport Protocol for Real-Time Applications) defines the default packetizing interval for various audio streams as follows:

- Packetizing interval is 20 ms for frame-type codecs with a framing interval of 20 ms or less
- It shall be the framing interval, when the framing interval exceeds 20 ms. For G.723.1, the framing interval is 30 ms, so 30 ms shall be the packetizing interval.
- For G.729 codecs, the framing interval is 10 ms, so an RTP packet has two frames.
- For non-frame-type codecs, such as G.711 or G.726, the packetizing interval is 20 ms, etc.

UDP Header

The datagram checksum is two bytes long and validates the message contents [10], [6] as shown in Figure 6.

4	Source Port Number	Destination Port Number
4	Datagram Length	Datagram Checksum
8 bytes		

Figure 6. UDP Header

IP Datagram Header

The IP Header shown below consists of 20 bytes. There may be optionally another 0 to 40 bytes may be present in the header. Details can be found in [6] and [11].

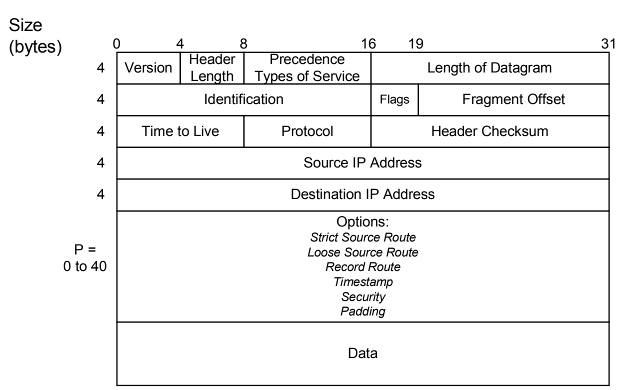


Figure 7. IP Datagram Header

Compression of Header Data

The method for compression of header data is known as payload header suppression (PHS). PHS is a process in which a portion of repeated MAC header information is suppressed. This provides a mechanism for avoiding redundant payload header data transmission. PHS is a form of data compression where efficiency of packet transmission increases; efficiency is gained by replacing a significant portion of the MAC packet header with few bytes.

A compression scheme [7] for the IP/UDP/RTP Headers (RFC 2508, February 1999) is initially targeted for applications sending audio and video over 14.4 and 28.8 dialup modems that provide full-duplex communication. This also may be used with reduced performance on simplex links. This compression scheme performs best on local links with low round-trip time.

The 12-byte long RTP packet header can be compressed to two to four bytes on an end-to-end basis. On a link-by-link basis, the combined IP+UDP+RTP header of 40 bytes can be compressed to two bytes for packets with no UDP Checksums, and to four bytes with Checksum. On a simplex link, or links with high delay, periodic refreshes with an uncompressed packet header are needed to restore the compression state in case of error. The link layer also must be able to provide an indication of switching between uncompressed and compressed header formats.

Almost half of the bytes in the IP and UDP headers may remain constant over the life of the connection. After sending the uncompressed header once, these fields may be removed from the compressed headers that follow. Differential coding on the changing fields in the remaining headers also reduces data size. For several fields that change in every packet, the difference from packet to packet is often constant, and its second-order difference is zero. When the uncompressed header and the first-order differences in the session state are shared between the compressor and decompressor, an indication of zero for second-order differences is sufficient for the decompressor to reconstruct the original header without any loss of information. This can be accomplished simply by adding first-order differences to the saved uncompressed header while each compressed packet is received.

While header suppression is an attractive technique to reduce overhead, it also has disadvantages. If a packet with an uncompressed header (reference packet) is lost, the follow-on packets with header suppression will be lost as well. To minimize reference packet loss, extra care needs to be taken, such as assigning a higher priority level or increasing robustness against bit errors in the header by adding FEC, etc.

IP Transport of Audio/Video over HFC Network

IP transport provides Internet connectivity for cable subscribers and is going to stay. Question arises if IP (RTP/UDP/IP) can be used as baseband transport for delivery of DTV signals. This way the entire HFC network will have one homogeneous transport and may provide some advantages such as provisioning, management, billing, etc. Before we make any such decision lets analyze such scenario. Figure 8 shows a typical block diagram of such an implementation. A packet framer packs MPEG-2 compressed audio/video in a RTP packet of desired size (e.g., L bytes). Actual payload will be L bytes minus the overhead (the headers). L bytes are broken into N MPEG-2 packets before input to the QAM modulator. For MPEG-2 encapsulation, the overhead is increased by N*4 bytes. No header compression is assumed here.

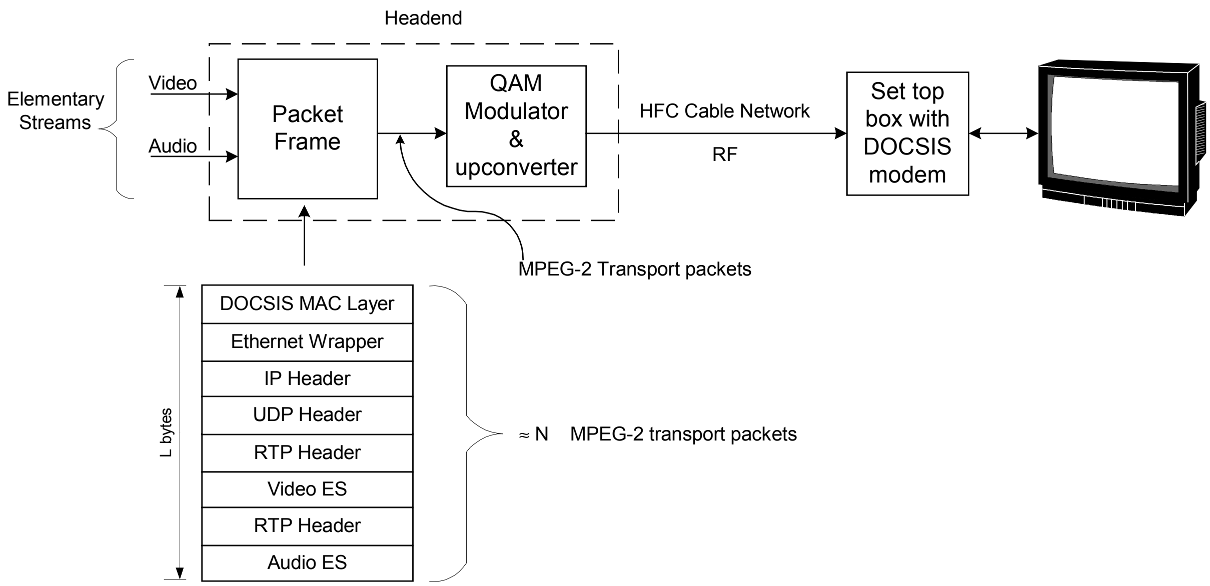


Figure 8. Typical scenerio for broadcast of DTV signal over DOCSIS channel (downstream)

Comparison of MPEG-2 and IP Transports

It may be observed from the above discussion that MPEG-2 transport is a link layer protocol. It has been created by MPEG for delivery of compressed digital audio/video in an isochronous (timely delivery of packet) mode. This protocol is particularly suited to broadcast delivery of multiple digital television programs in a very efficient way while providing synchronization between audio and video (no lipsync artifact). On the other hand, IP protocol has been created by IETF for end-to-end delivery of data in non-real-time over heterogeneous networks. IP is a higher layer protocol than MPEG-2 and addresses delivery of data on one-to-one or one-to-a-group basis, etc. It implies that IP packets are routable over WAN (wide area network) whereas MPEG-2 packets are not. For delivery of MPEG-2 compressed audio and video in real-time on the end-to-end basis IP protocol may be used to encapsulate audio and video data. To facilitate this, additional layers have been added over IP and the combined protocol is known as real time protocol (RTP). RTP is designed on top of UDP and IP, which is discussed before. Adding more layers provides flexibility but causes increase in overhead and efficiency of the transport protocol goes down. Efficiency may be defined as the ratio of actual payload divided by payload plus overhead. A brief comparison of two protocols is given in the Table 1.

The cable network spectrum is divided into a number of 6-MHz physical channels. When digitally modulated, a 6-MHz channel provides a digital bandwidth of 27 Mbps for 64 QAM and 38 Mbps for 256 QAM. For various bitrates of MPEG-2 compressed video, computations have been performed to find the number of logical channels that can be delivered over a 6-MHz physical channel using the two protocols.

		MPEG-2 Transport	RTP / UDP / IP Transport
1.	Ref. to OSI Layers	MPEG-2 is a link layer protocol. Link layer is the layer 2 of OSI.	RTP is a higher layer protocol. RTP is created by adding 3 layers above the link layer. IP is layer 3 of OSI and RTP resides over UDP/IP.
2.	Packet Size	Fixed 188 bytes.	Variable packet size (64 to 1518 bytes)
3.	Overhead	4 bytes for 184 byte payload.	40 bytes for 64 –1518 bytes. Overhead is much higher for comparable MPEG-2 packet size. Large packet size is avoided as packet loss can cause artifacts. Overhead may be reduced by bundling audio and video in the same packet
4.	Delivery	Optimized for broadcast delivery of compressed audio/video content. By adding MPEG-2 system layer conditional access mechanism delivery to one or a selected number of receivers is possible in an intranet like HFC cable network.	IP is optimized for one-to-one ore one-to-many non-real-time data delivery. RTP is an extension of IP for real-time use over Intranet. Not as efficient as MPEG-2 in a broadcast like application.
5.	Isochronous Delivery	MPEG-2 T-STD buffer model is designed for such a delivery.	No such buffer model is specified for RTP. Buffer model depends on application.
6.	Audio/Video Synchronization	Keeps tight synchronization between Audio and Video. Proven through large-scale implementation in broadcast applications.	A few issues to be solved for synchronized delivery of A/V over internet. A QOS capable network may be of some help.
7.	Routability	Not designed for routing over heterogeneous networks.	Designed for routing over internet.
8.	Efficiency	97.7%	< 85%

Table 1. Comparison of MPEG-2 and IP Protocols

Table 2 presents the computational result for a 64-QAM modulated channel while Table 3, for 256 QAM no header compression is assumed in putting the number of channels for IP. It may be observed from Tables 2 & 3, that MPEG-2 transport is preferable to IP as the former provides more channels. More channels mean more revenue. The next question comes about streaming video and audio. Streaming from the headend (such as VOD) to any cable subscriber may be implemented using either MPEG-2 transport or IP transport. MPEG-2 transport will be preferable over IP from the reason of overhead. When a cable subscriber wants to reach a source located outside the cable network for a streaming video, IP (RTP/UDP/IP) will prevail as packets constituting the stream need routing over the internet. Internet is an unpredictable network where loss of packet or variable packet delay may occur. Near isochronous delivery may be achieved using some form of QOS or a private backbone.

Encoding Resolution	Bit Rate (Mbps)	MPEG-2 Transport Logical Channels	Video Over IP Logical Channels
352x240 (CIF) (Movies)	1.5	17	14
352x480 (Half) (Movies)	2	13	11
540x480 (3/4) (Movies)	3	8	7
704x480 (Full) (Movies)	4	6	5
540x480 (3/4) (Sports)	5	5	4
704x480 (Full) (Sports)	6	4	3
HDTV 1080x1920 (Full)	19	1	1

Table 2. Logical channels for MPEG-2 Transport and Video over IP (64 QAM 27 Mbps)

Encoding Resolution	Bit Rate (Mbps)	MPEG-2 Transport Logical Channels	Video Over IP Logical Channels
352x240 (CIF) (Movies)	1.5	25	21
352x480 (Half) (Movies)	2	19	15
540x480 (3/4) (Movies)	3	12	10
704x480 (Full) (Movies)	4	9	7
540x480 (3/4) (Sports)	5	7	6
704x480 (Full) (Sports)	6	6	5
HDTV 1080x1920 (Full)	19	2	1

Table 3. Logical channels for MPEG-2 Transport and Video over IP (256 QAM 38 Mbps)

The MPEG-4 Standard and the Related Content

MPEG-4 is a compression standard and no specific transport protocol has been created to deliver MPEG-4 compressed elementary streams. MPEG-2 Systems standard has been amended to carry MPEG-4 content both at elementary stream level and as multiplexed one. Also MPEG and IETF (Audio Video Transport group) have been working together to create a protocol to deliver MPEG-4 content over RTP (RTP/UDP/IP). In this regard a few RFC have been submitted for consideration.

Conclusion

Baseband transport of digital content over the HFC cable network has been analyzed. The two leading transport protocols, MPEG-2 and IP, have been studied. MPEG-2 transport is very efficient for broadcast delivery of MPEG-2 compressed audio/video and private data. IP has a larger overhead, but provides connectivity to the internet, a world outside of the cable intra-network. It is expected that MPEG content delivery and internet connectivity will be two important parts of the cable business. Broadcast of MPEG content from the headend is very efficient using MPEG-2 transport, while providing internet connectivity via HFC cable network has become a necessity. It may be concluded that both transport protocols are going to coexist in the HFC cable network for a while before one takes over the other.

References

[1] ISO/IEC IS 13818-1, “Information Technology—Generic coding of Moving Pictures and Associated Audio Information: Systems.”

[2] ISO/IEC 11172-1, “Coding of Moving Pictures and Associated Audio for Storage Media at up to about 1.5 Mbit/sec.” Part I; Systems.

[3] Annex B to ITU-T Recommendation J.83 (10/95), Digital multi-programme systems for television sound and data services for cable distribution.

[4] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” RFC 1889, January 1996.

[5] M. Reha Civanlar, Glenn L. Cash, Barry G. Haskell, “RTP Payload Format for Bundled MPEG,” draft-civanlar-bmpeg-00.txt, August 1996.

[6] Douglas E. Comer, “Computer Networks and Internets,” Prentice Hall, ISBN 0-13-083617-6.

[7] S. Casner, V. Jacobson, “Compressing IP/UDP/RTP Headers for Low-Speed Serial Links,” draft-ietf-avt-crtp-05.txt, July 27 1998.

[8] M. Handley, “Guidelines for writers of RTP payload format specifications,” Internet Draft.

[9] M. Handley, “GeRM: Generic RTP Multiplexing,” draft-ietf-avt-germ-00.txt, November 11, 1998.

[10] J. Postel, “User Datagram Protocol,” RFC 0768, August 1980.

[11] Defense Advanced Research Projects Agency (DARPA), “INTERNET PROTOCOL, DARPA INTERNET PROGRAM, PROTOCOL SPECIFICATION”, RFC 0791, September 1981.

[12] Data Over Cable Service Interface Specification, Radio Frequency Interface Specification, DOCSIS 1.0 and DOCSIS 1.1.

Acknowledgements

The authors would like to thank the management of CableLabs and Philips Semiconductor for their support in performing this research. Special thanks are due to Mr. Rouzbeh Yassini of YAS Corporation and other members of the team.

The Challenge of a Standard Software API

Mark Eyer
Sony Electronics

Cable Multiple-System Operators (MSOs) would like to supply not only digital audio, video, and data services, but also software applications that can run on customer-owned equipment. With this goal in mind, CableLabs® issued a Request for Proposal in 1999 to help identify and standardize a software Application Program Interface (API) for OpenCable compliant retail boxes. This paper explores the challenges involved in this effort, and identifies some of the pitfalls and obstacles that must be overcome. These include issues of platform independence, the cost and complexity of the platform, the challenge to support an evolving digital world, and the need on the part of consumer electronics manufacturers to differentiate their products in the marketplace. Suggestions for resolution of some of these dilemmas are presented for consideration.

THE MSO'S VISION

The eventual availability in the retail market of digital cable-compatible consumer devices offers the cable MSO a number of significant benefits. Whenever a customer buys a retail cable-ready device, the operator's capital expense is reduced. Due to competitive market pressures, the retail devices will be able to offer the latest technologies, including faster CPU speeds, ever-speedier graphics, and interfaces to the newest audio/video peripherals. And happier customers can result: many are more content without the need for the bulk and clutter of the set-top box, as set-top box functions are integrated with the digital television.

This picture is quite clear for services including standard- and high-definition audio/video offered on subscription and impulse-pay-per-view (IPPV) basis. The standardization of the interface to the removable security module, the network (physical cable) interface, and system and service information (and agreements to deliver it) has enabled consumer electronics manufacturers to start designing digital cable-compatible devices for retail sale, starting with digital TVs (DTV).

But what about other services, such as Electronic Program Guides (EPGs), video on demand (VOD), voice over IP (VOIP), or streaming audio and video in formats other than MPEG-2 or Dolby Digital? And what about services not yet conceived? Set-top boxes supplied by the MSO can be built to offer advanced services. How can a device available for retail sale be enabled to do so?

A simplistic view of the world, from the point of view of the cable operator, is that the primary purpose of any cable-ready device to be available at retail should be to generate revenue for that operator. To that end, the retail device would be 100% controlled by the cable operator in terms of everything that is presented for viewing—its "look and feel."

An MSO's dream, therefore, might be that a DTV or other retail cable-ready device, after being brought home and installed by the consumer, would be downloaded with code supplied by the local cable operator. At that point, any access the consumer would attempt to make of any services offered on the cable would be managed through a navigation application supplied by the cable operator.

If a special offer or preview were available, the navigator could make sure to present that information to the user. If new services were offered, the navigator could be set up to notify the user of their existence, and to guide the user towards their access. As an additional source of revenue, advertising or links to commercial sites could be included in the navigator.

Services such as VOD could be offered, because the navigator could support whatever proprietary form of access and control was required by that operator's plant and equipment. New forms of services could be offered when they became available, even if the details of presentation and decoding are yet unknown. That's because an updated navigator could be provided when the details of the new service are worked out.

An example of such a new service is a data broadcasting service. Multicast data synchronized to video is offered today, but the standardized techniques for transport and the content coding formats are not yet totally settled. When industry acceptance is widespread and the particular cable operator implements the new standard, the navigator can be upgraded to allow all subscribers to have access to the new data enhancement.

The EPG can adapt the presentation based on what services are authorized for viewing in this particular device. For example, if the user has not subscribed to *MovieMax*, the navigator can direct that user to the *MovieMax* preview channel, can notify the user of special sign-up offers, or allow him or her to sign up online (self provisioning).

With a downloaded navigator, the cable operator has direct control over the look and feel of the EPG. They can organize the guide in such a way that the services with the highest profit margin (IPPV perhaps) are given prominence. They can put effort into human factors design to help ensure that the consumer's experience is productive and pleasurable.

FACING REALITY

What's wrong with this dream? A troubling aspect of this code download scenario is the implied notion that every retail device, regardless of manufacturer, make, or model, would behave in exactly same way. In the following sections we first explore this product differentiation problem, and then discuss further difficulties, including:

- Problems with the overall philosophy of cable- operator-supplied downloaded code
- Technical challenges, such as reliability and the difficulty of porting the API
- Challenges related to the magnitude and complexity of the problem

The paper goes on to suggest some resolutions to these problems, and describe example products we would like to be able to manufacture once the software download system design and specifications are complete. We then suggest a way forward in the near-term, bridging between technologies available today and that which we will develop and refine in the next several years.

PRODUCT DIFFERENTIATION

From the point of view of the consumer electronics manufacturer, the idea that every model of every manufacturer's product ultimately runs the same cable operator-supplied code causes real marketing problems.

Competition in the marketplace

Consider a high-end product from manufacturer A compared with a high-end product from manufacturer B: when compared side-by-side on the sales floor, both products will appear to be *identical* once downloaded with the local cable operator's application suite.

Low, middle, and high-end

Commonly in consumer electronics marketing, a manufacturer offers low-end, middle, and high-end products. The middle of the road product offers some features not found on the low-end model, and the high-end product offers bells and whistles not found on the level below.

Perhaps the low-end product doesn't support software download at all. But let's say the middle and high-end products do support the OpenCable Middleware Solution. Once downloaded with the cable operator's application, when either of these boxes accesses a cable service, the user experience is the same (aside from factors like CRT display size).

In this world, a manufacturer can no longer differentiate one product from another based on a software-related feature.

PHILOSOPHICAL PROBLEMS

Native applications

A native application is one written in or compiled to the machine code of the retail device's

CPU. A cable operator may want to maintain control over native applications, for example by downloading a "Master Application" capable of authenticating them, launching them, and determining their privileges and resource usage. Is this practical, possible, or even reasonable? We think not.

Given that each model of each manufacturer's product would typically have a different set of native code, it is entirely unclear how the cable operator's downloaded application could be afforded such control.

Consider that a cable-ready device offered for retail sale must have *some* level of functionality even before an operator-supplied code download occurs. For example, it will likely provide access to analog and free services on cable, assuming that the user has a basic cable service. It would provide some form of user setup and/or diagnostic functions even without a basic cable service.

This native application cannot and should not be under the control of the Middleware Solution or the cable operator's downloaded application.

In fact, the native application needs to take priority over anything that might be downloaded, for example to allow it flush memory and re-initialize the unit in case of trouble. Or, in case the user moves it to a new city and/or a new cable system.

It is not only impractical but also unwise to say that an application provided by the cable operator should control the native applications in a retail device.

The native application on the right in Figure 1 accesses OS functions in the device directly. As shown, a companion *resident* application is also present. The resident app is written in a platform independent way by the manufacturer, and takes advantage of the middleware layer implementation.

As shown, a cable-operator supplied Application Suite is present, including a "master application." We feel that this master application should be the "master" of the elements of the application suite (EPG, VOD, web browser, as shown), but it cannot and should not be involved with the control of the resident or native applications.

Extensibility

Let's say an API is eventually agreed upon, and some number of compliant platforms are fielded. In a year or two, as history tells us, typical CPU speeds will be doubled, memory prices will be halved, and graphics capabilities will increase by a large factor. In two years it may be cost effective to include video hard disks in most devices.

If the MSO or cable operator does not create a new application suite tailored to the 2nd generation platform, the power of any new available hardware cannot be fully exploited. The size and capabilities of the application are limited by the least common denominator platform.

If the solution is to provide a new application suite to be run on the next-generation boxes, where does this progression stop? The process of defining and standardizing new platforms and API extensions is never-ending as the operator's configuration control and management problem becomes exponentially more difficult.

How many new downloadable applications might any MSO develop over a five-year period? We think the answer is something like one or two at most, given the enormous complexity of the task. A software release for a large cable plant must be rigorously and thoroughly lab-tested before large-scale deployment. Testing such an application would be an unprecedented challenge because of the large (and growing) number of target platforms upon which the application must be validated.

Control over *all* cable-delivered services

A cable operator may want to use the downloaded application to control the look and feel of *all* cable-delivered services, including the DOCSIS cable modem. Clearly, the operator grants or denies access via the cable modem to the Internet. Access to cable modem services is based on whether or not the consumer has paid for a subscription to the cable modem service.

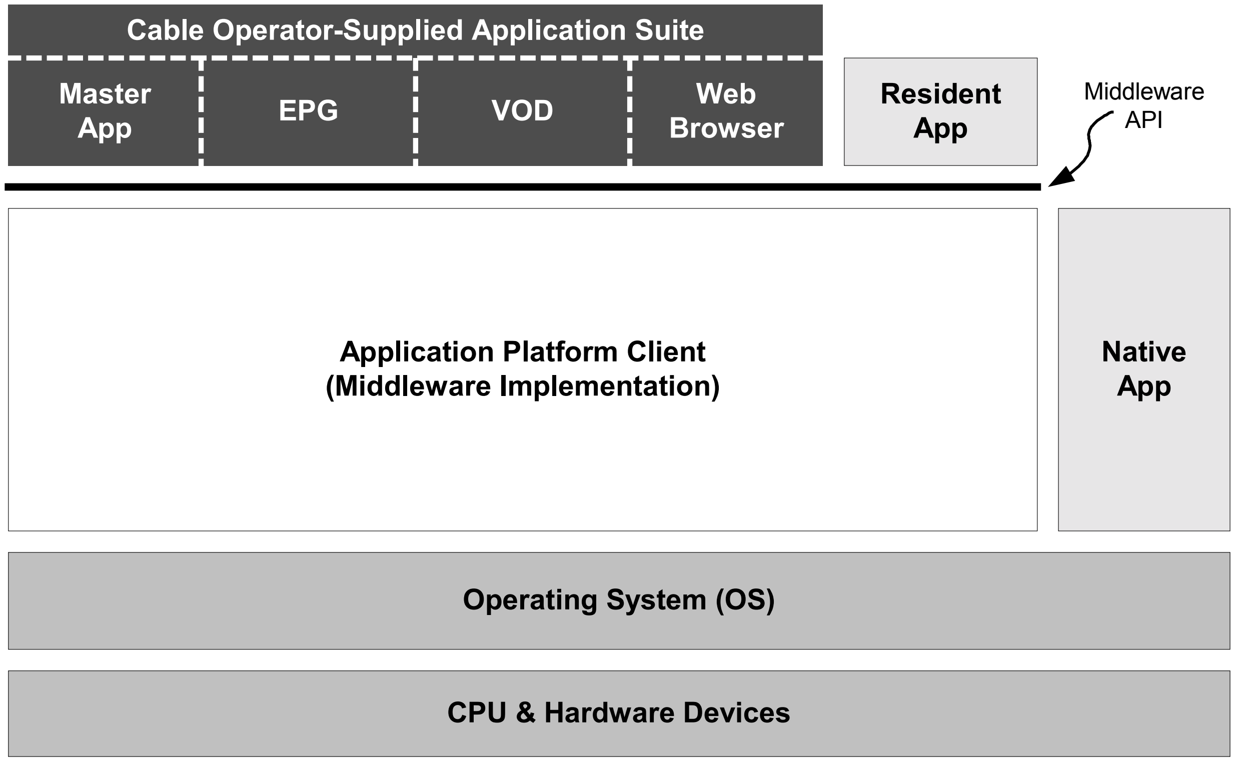


Figure 1. Software Architecture of a Representative OpenCable Retail Box

But if a cable modem service has been paid for, what control over the "look and feel" of it can a cable operator hope to impose? It will be the authors of HTML content on a website, for example, that will determine the appearance (and look and feel) of presentations based on that content.

Perhaps the cable modem in the set-top or DTV is connected by Ethernet to a PC. How could the cable operator have any control or impact of the on-screen look and feel on the PC screen?

One application for *all* OpenCable devices

A cable MSO may want to create a single application for deployment across the full range of OpenCable host devices available at retail. That suggests that the same application that runs in a retail DTV also runs in a D-VHS recorder, retail set-top box, or cable-connected Personal Video Recorder (PVR). Upon reflection, it will be clear that the one-application-for-all-devices goal is unreasonable.

For example, the PVR may be a low-cost device that happens to have a cable tuner but no user interface of its own. It can act as a slave storage peripheral and program source for other devices on a home audio/video bus, but it wouldn't have a use for (let alone the resources to support) a cable-operator-supplied navigator application.

Even if the PVR could accept an application, a PVR is a PVR and not a DTV. The application must account for the functional aspects of the type of host platform it runs on.

Any cable-operator supplied application *must* take into account aspects of the platform upon which it is expected to run. An application downloaded to a retail set-top box must be optimized for a set-top box type of product, just as an application for a DTV must be adapted to the DTV.

Control over *all* resources

A cable MSO may wish to use a downloaded application to control *all* resources available on the host device. We see a difficulty here.

Let's say a new model of consumer device offers wireless connectivity to other devices. Can the cable operator's downloaded application have access to the resources offered by the wireless port? Such access will not be practical until the

Middleware Solution itself is extended to support wireless access. How can the application take advantage of wireless access anyway, given that the capabilities of this particular host device are unknown to it?

Furthermore, as an unending succession of new types of resources are invented and popularized, will the Middleware Solution be continually extended to include each new one? This seems impractical as a road forward.

TECHNICAL CHALLENGES

100% reliability

The reliability of the cable-ready device is very important. Clearly, we cannot allow tomorrow's digital television or set-top box to move anywhere near today's notion of a PC, where program crashes and the need to re-boot the machine are frequent and commonplace.

On the other hand, 100% reliable crash-proof code is impractical with today's technology, especially given the size and complexity of the code involved, and the environment in which it is intended to run.

Even with the Java programming language and its elimination of pointers, some types of programming errors can cause available memory to diminish to zero over time. The program may not "crash" but it can become unusable, non-responsive to RCU keys, and require some form of re-booting or reset.

Another form of crash can occur as two separate software processes, or "threads," both try to access the same data structure in memory. Unintended results can occur, or in some cases a "deadly embrace" can result where both processes are blocked from further operation.

Also, of course, software testing isn't foolproof. By some reports for example, the first release of Windows 2000 included 63,000 known "issues." The count didn't include those bugs not yet discovered of course.

It is possible to write a very stable application that will function reliably for long periods of time, given a stable and well-understood environment in which it can execute. The challenge we face in the OpenCable approach is that the environment is not always well known. There will be many platforms and many implementations.

Lack of standards for code download and authentication

Compared to the issues above, this one is rather easy. Security functions in an OpenCable compatible device are handled within the POD module. Standards are needed to allow a Host device to make use of API calls to the POD module to determine if a particular piece of executable code has come from a reliable source, with no tampering or corruption. Some early proposals have been floated to some standards committees addressing this need.

Porting a complex API definition

When a manufacturer wishes to support the OpenCable Middleware Solution in a product, it may be necessary to port it to the Operating System chosen for that product. If the OS is a common one, it is possible that an implementation of the Middleware Solution standard for that OS can be purchased. Even so, there will always be a significant amount of work involved—the device-level support aspects must be home-grown in almost all cases. These include the aspects of tuning, demodulation, MPEG-2 decoding, SI section filtering and parsing, graphics and display control, audio control, front panel and RCU interface, signal switching and routing, and communications protocol stacks.

Cost and return on investment to support standard API

Once the Middleware Solution is standardized, manufacturers of cable-compatible devices will be asked to provide support for it. The cost to add support for this API will be nontrivial. Many megabytes of RAM and ROM will be required. A significant portion of the cost to develop such a product will be involved with testing the implementation.

The manufacturer is likely to ask: "where is the return on my investment to include support for this API? Will my customer recognize the value of it to the extent that it increases the cost of the product?"

After all, it is the cable operator that will reap the monetary benefits. For manufacturers to embrace the standard API concept, it appears that some form of business arrangement will need to be made with cable operators as an incentive.

THE CERTIFICATION CHALLENGE

Many compliance and interoperability questions are raised by the notion of a standard software API to be used by retail consumer devices. How will a manufacturer certify that a certain implementation of the Middleware Solution is fully compliant? Given the large number of manufacturers involved, and the need to individually test each different product from each manufacturer, the number of products needing testing in a given year is very large.

While CableLabs has undertaken to certify standalone DOCSIS modem implementations, that organization does not appear to be capable of supporting, on its own, certification of all these new consumer devices.

The Middleware Solution specification document is likely to be extremely large and complex. As an example, consider the candidate specifications promoted by ATSC by the DTV Application Software Environment (DASE) group and by Sun Microsystems in the JavaTV effort. The 922-page DASE API specification (version 1.08.01) includes over 350 Java packages, classes, and interfaces.

One might ask, "What's new here—haven't there been other similarly challenging compliance problems?" I think the answer is "no." Let's look for some examples.

In the realm of Microsoft-Intel PC platforms, certainly there are a large number of manufacturers and implementations, but the compliance problems are much simpler. Most importantly, all platforms use the same type of CPU, so all code is "native." Compliance testing has been achieved and many vendors have created clean-room implementations of the Basic I/O System (BIOS). If the BIOS is compliant, any Operating System or application riding on it will work. Unlike the OpenCable Middleware Solution, which will be *very* large and complex, the IBM BIOS is small and very well defined.

In the world of satellite TV, different vendors can license the technology needed to build satellite IRDs compatible with a certain operator's signal. Typically, the number of manufacturers is quite small (three or so). Each manufacturer is free to use a proprietary hardware platform, APIs and native applications. Nevertheless, the compliance testing conducted by a typical operator is quite complex and time consuming.

The essential difference in the case of retail cable-compatible devices is that 1) the API is very large and complex, and 2) there are likely to be a very large number of equipment manufacturers, each with multiple products.

SUMMARY OF ISSUES

In summary, the following issues and difficulties with the Middleware Solution concept proposed for OpenCable have been raised here:

- No provision for or acknowledgment for the need for product differentiation is made
- Authenticating and controlling execution of native applications is impractical and unnecessary
- There is a clear need for a well-understood evolutionary path, or of an approach to realizing extensibility in a practical way
- The concept of "one application for all retail devices" is impractical
- "Crash-proof" code is practically impossible
- The cost of implementation and testing the API is an issue to the CE manufacturer
- Certification and compliance testing will be an unprecedented challenge

PROPOSALS FOR ACHIEVING OUR GOALS

The following sections outline proposals for an expanded Middleware Solution concept that we think can work for consumer electronics manufacturers as well as cable operators and MSOs. The most fundamental proposed change is that the Middleware Solution must support resident and manufacturer-supplied applications alongside the cable- operator-supplied downloaded code.

We first argue why this change is necessary, and then describe some of the new system requirements that result.

PRESERVATION OF COMPETITIVE URGE

The competitive urge must be preserved. If a Middleware Solution were designed such that product differentiation was not supported, the only challenge for a manufacturer would be to make the most cost-effective implementation. Enhancing the product line with more and better features could not occur. Innovation would be stifled.

Clearly, this scenario is unacceptable. Product differentiation ultimately benefits both the consumer electronics industry and the cable MSO. If we agree that product differentiation must be possible, then the following conclusions result:

- The Middleware Solution must support unrestricted execution of native and resident applications, under control of the consumer (not the cable operator). This was discussed above.
- The Middleware Solution must support native or manufacturer-specific extensions and "hooks."
- The Middleware Solution must allow any specific implementation to fully take advantage of (on its own) hardware and interface features that might be present.
- Communications resources such as the cable modem must be freely available to the native and resident applications (given the proper basic authorization for cable modem service, of course).

Support for API "hooks"

To support certain desirable product features, the resident or native application needs to be aware of certain user actions. In some cases in fact, the resident app needs to be able to take control of the user interface.

The manufacturer's native application needs to be able to "register" with the Middleware Solution so that it can be notified about certain events. For example, it may want to be notified whenever events of the following types occur:

- Any tuning-related API is called
- The user targets a program scheduled for future viewing (for example to set a timer)
- A timer activates or expires
- An Emergency Alert notification is received
- Access is made to web content related to the television program being viewed

The API must support such access to these events, as well as any others deemed useful by the implementation.

Support for hardware features

A Middleware Solution may be implemented on various devices, each with a different set of included hardware and interface features. These might include 3D graphics acceleration, special content decoding formats (MPEG-1, DirecTV transport, etc.), support for specific peripheral devices on the IEEE-1394 serial bus, camera and video input ports, game controller ports, USB, storage peripherals, DVD player/recorders, personal video recording capability, format conversion (camcorder to MPEG, etc.), wireless communications ports and protocols, or countless others.

If any of these hardware-related features are present in the cable-compatible device, the manufacturer-supplied resident and native applications must be afforded access to them. Furthermore, to the extent possible via API hooks and extensions, the native code should be able to integrate them with the cable operator-supplied application suite.

For this philosophy to work, some fundamental aspects of the Middleware Solution API may need to change.

EXAMPLE PRODUCTS

The following sections explore some possible cable-compatible products to show how the concepts we have presented apply.

Example: Video game console in the box

This product combines a state-of-the-art video game machine with a cable-compatible digital television. A DOCSIS cable modem is included so that interactive games may be played via the Internet.

By way of review, here is a brief list of the issues:

- If the video game platform uses DOCSIS (a cable service), and the Middleware Solution must rule the look-and-feel, how can that work on top of the video game? Clearly each individual game played on the game console has its own look and feel.
- How could the Middleware Solution authenticate each game before it is played? Clearly that's not possible nor is it needed.
- To save cost, eliminate redundancy, and provide proper sharing of resources, the video game application may want to take advantage of API calls within the Middleware Solution. Such sharing of the API appears not to be allowed in the current middleware concept.
- The cable operator's downloaded application is supposed to be offered the ability to interface to any native extensions (in this case, the video game hardware and software support library). This can't work because these extensions are very likely to be proprietary. Even if they were open, the "one application" likely couldn't take advantage of resources available only in a small fraction of implementations.

Example: High-end retail cable-ready box

Consider a high-end cable-compatible device to be offered for retail sale. The manufacturer of this device is aggressively leading-edge and wishes to offer the following product features:

1. Support for a sophisticated file system for personal video recording on a built-in hard disk
2. Support capture of video clips via a CCD camera port or images from a digital camera for attachment to e-mail
3. Support an enhanced navigator function, combining web-hosted databases with data derived from expressed user preferences and observed viewing habits
4. Support a marketing product tie-in service, allowing the user easy access to products and services associated with broadcast television programs
5. Support a feature where the device acts as a home audio/video master control center, capable of routing signals between a/v equipment throughout the home
6. Support a "home gateway" function including a private personal website, through which the homeowner may login from anywhere in the world to access guide data, check on and set recording timers, check home security and access other equipment controlled by the device.

The standard Middleware Solution may or many not include APIs that would support this set of features (it's not likely). Even if it did, while a cable-operator supplied application suite *could* possibly use the standard APIs to include these features, it is extremely improbable.

Even if the operator-supplied application suite did do a few of these features, the user should be able to choose which he or she wishes to use.

CO-EXISTING APPLICATIONS

In an environment that includes a cable-operator supplied application suite alongside the manufacturer's native code, some new challenges arise.

Memory management issues

The cable- operator-supplied application will require or request certain memory resources (RAM, Flash-ROM, even hard disk file space), as will the resident or native application. The resident and downloaded applications may both request, through API or OS calls, “all the remaining memory.”

Once a system of cooperative sharing of memory resources is worked out, both these competing entities can peacefully co-exist.

Another problem relates to management of the persistent storage. A Flash memory file system is an example of a persistent storage system. The file system is often used by applications to store preferences, passwords, “cookie” type data, viewing history, forms data, or any other data that has some kind of lasting value.

The problem arises as the file system inevitably overflows, and no more space is available to create new files or to expand the size of existing ones. Some kind of file space reclamation process must be run. That process must decide which files are good candidates for permanent deletion.

With personal computers, we don’t allow such a process to make decisions about what files to delete. We do it manually. Writing a file space reclamation routine will be a challenge, because unexpected results can easily occur when files expected by a certain application turn up missing.

Security policy issues

In a typical Middleware Solution, a *security policy* is an inherent part of the API. Through an enforced security policy, certain applications can be granted or denied access to specific API calls. A Java applet, for example, is not allowed to perform file I/O.

As discussed above, we do not feel it is workable to allow the cable operator’s *master application* to dictate and establish all security policies in effect within a cable-compatible device. So the question is open: who will allow whom to do what? Agreements on management of security policies will need to be worked out and documented.

Resource sharing issues

In the scenario we are discussing now, an application suite provided by the cable operator is present alongside a manufacturer supplied resident or native application. If we wish to allow both of these to have actively executing threads, contention for resources can result.

For example, both applications may wish to make use of the tuning API. If one has control of tuning, the other will need to deal gracefully with the unavailability of that resource. This simple approach may result in some undesirable behavior, where even a high-importance request can go ungranted. Such problems suggest that a priority scheme should be adopted for resource management.

Clearly, a new level of complexity is introduced. Such priority mechanisms have been discussed in some groups, but the problem has so far not been brought to a clean solution.

PROPOSAL FOR STEPWISE EVOLUTION

Given the scope and magnitude of the unresolved issues, we feel that we are perhaps two to three years away from a workable solution to general-purpose code download. So, in light of these difficulties we present here a proposal for a practical way forward that can benefit both the cable operator and retail device manufacturer.

We start with the definition of a cable-compatible device that can be built today, using currently published SCTE and OpenCable standards. The stepwise evolution therefore starts exactly where we are today.

Level 0: the “Watch TV” box

The features of the level 0 box include the following:

- Performs the basic “watch TV” function.
- Conforms to the OCI-N network interface, tunes/demodulates 64- and 256-QAM, etc.
- Supports all video formats defined in EIA-818 and OpenCable, either by down-converting to NTSC, or passing compressed HD video via 1394 to DTV.
- Navigation based on SI/EPG tables as standardized by SCTE DVS.

- Adheres to OpenCable requirements for diagnostics, audio and video performance.
- Hosts OpenCable POD module for access to premium services, pay-per-view.
- Provides copy protection on all analog or digital outputs.
- If an upstream transmitter is provided, supports impulse pay-per-view (IPPV).

We expect level 0 DTVs to reach the market sometime in late 2001. No further standards work is needed aside from finalizing some of the details of the POD interface and POD copy protection (this work is underway now).

Level 1: the "Web Browser" box

We now define a cable-compatible device with features and capabilities going beyond those offered by the Level 0 box. The Level 1 cable-compatible device provides all the features of Level 0, plus it:

1. Supports a DOCSIS cable modem
2. Supports a TCP/IP protocol stack plus HTTP, SNMP
3. Is capable of interpreting and displaying HTML-based content (for example, HTML 4.0, ECMA Script, DOM1, CSS1)
4. Supports HTML extensions for a standardized TV-based URI scheme so that HTML pages can include links to TV channels (per ATVEF, for example)
5. Supports a scheme whereby the retail box knows the "home page" offered by the cable operator. By this means, the cable operator can operate a *portal* and offer links to services and promotional offerings.
6. Supports access to world-wide web as a pay service (otherwise, box can only access the operator's intranet).
7. For set-top devices: supports OpenCable HDNI specification for pass-through of compressed HD video to DTV.

Except for one detail (#5), we have standards available to allow us to formally define the level 1 device today.

We need to make sure all the standards are in place to allow a cable-ready device purchased at retail to be brought home by the consumer, plugged into the cable, and become operational without the need for the cable operator to roll a truck.

For this discussion we can assume that before even buying the new device, a subscription to basic or premium cable service has already been established. When the new box is first plugged in, it will be able to access unscrambled analog or digital services.

The customer then calls the cable operator to indicate a desire for cable modem and premium movie services, and requests that they send an access card. A POD module arrives in a day or two. The customer plugs it in.

The presence of the POD module triggers an initialization sequence in which the device communicates with the cable headend, registers itself in the network, and is provisioned for access to cable modem and the requested premium channels.

At this time, with a suitable new protocol we need to standardize, the POD module can give the Host a URL to be used to access the cable operator's local home page. In Denver, for an example cable operator named *XYZ Cable*, this URL might be http://stb.xyz-cable.den.com. Now, an RCU button labeled **HOME** could trigger opening a browser window using this URL, and *XYZ Cable*'s portal would pop up.

From the *XYZ Cable* home page, the user can access the following types of functions, at the discretion of this cable operator:

- Links to information on the channel lineup
- Links to Electronic Program Guide data presented in HTML format. A search function can be supported; listings can be by time or genre, etc.
- Hyperlinks to account information pertinent to this particular customer
- Hyperlinks to allow self provisioning (for example, to allow one to sign up for new premium services, special events, etc.)
- Search functions related to program offerings or FAQs

- Telephone numbers to call for service problems or account information
- Hyperlinks to local businesses. The portal can of course include advertising banners.

It should be clear that the level 1 device is a very powerful and flexible platform, offering the cable operator a "look and feel" presence in the retail box and a rich set of operator-supplied features. Importantly, this can be done with existing standards (with the small exception noted).

Level 2: Web browser plus Java

The Level 2 device provides all the features of Level 1 plus:

- Adds a Java Virtual Machine to support Java applets associated with web pages.
- A specified set of Java class libraries is resident to support the desired applet capabilities (for example, it may be a subset of Personal Java and JavaTV).
- Minimum RAM requirements are specified for downloaded applets, cache.
- Minimum graphics resolution and performance requirements are specified.
- Support for a prescribed set of content and mime types is specified (graphics formats, audio formats, streaming audio/video formats and plug-ins, etc.)

The level 2 device can respond to web pages enhanced to include applet-based applications. The level 2 retail device can be built with today's technology, with a few exceptions. At this writing the Java components are not yet finalized. Completion of these is expected sometime in 2000.

Also, to be effective, some form of memory management would have to be standardized so that important applets would be cached in RAM for quick retrieval when needed. It wouldn't be practical to have to wait for an applet download whenever a channel was changed, for example.

With the addition of this basic level of Java support, the cable operator can now offer fully-featured Electronic Programming Guides with improved presentation and better look and feel, Video-On-Demand, enhanced broadcasting, enhanced e-commerce applications, and networked games.

Since only one application is assumed to be executing at any given time (the one associated with the web page in current view), many of the complexities associated with resource sharing and application lifetime are avoided. In spite of this limitation, an unlimited array of new services, applications, and features can be supported.

Level 3: Cooperative downloaded applications

At level 3, which we think is realistically three years away, the technical challenges identified in this paper have been successfully met. This allows the co-existence in one retail cable-compatible device of:

- A resident application suite present in ROM or Flash at the time the product is purchased
- An application suite downloaded upon consumer installation of the device

Furthermore, either or both of these two basic applications can be upgraded (or replaced entirely) via a download update mechanism.

CONCLUSION

This paper began by describing the MSO's vision of the capabilities and benefits of the Middleware Solution for support of software downloads to retail devices. It then explored those aspects that are felt to be impractical or unreasonable.

The paper discussed some of the technical challenges that must be met before a general solution to the code download problem can be reached. The argument was made that the preservation of product differentiation in the marketplace is essential.

The paper concluded with a proposal for "stepwise evolution, where HTML and applet-based approaches would be used in the interim, until the challenges of the downloaded application approach can be fully addressed.

ACRONYMS

API	Application Program Interface
ATSC	Advanced Television Systems Committee
ATVEF	Advanced Television Enhancement Forum
BIOS	Basic Input/Output System
CPU	Central Processing Unit
CRT	Cathode Ray Tube
DASE	DTV Application Software Environment
DOCSIS	Data over Cable Service Interface Specification
DTV	Digital Television
D-VHS	Digital VHS
DVD	Digital Versatile Disk
EIA	Electronic Industry Association
EPG	Electronic Program Guide
FAQ	Frequently Asked Question
HD	High Definition
HTML	Hypertext Markup Language
HTTP	Hypertext Transport Protocol
IP	Internet Protocol
IPPV	Impulse Pay-per-View
IRD	Integrated Receiver-Decoder
MPEG	Motion Picture Experts Group
MSO	Multiple System Operator
OS	Operating System
PC	Personal Computer
POD	Point of Deployment
PVR	Personal Video Recorder
QAM	Quadrature Amplitude Modulation
RAM	Random Access Memory
RCU	Remote Control Unit
RFP	Request for Proposals
SCTE	Society of Cable Telecommunications Engineers
SI	Service Information
VOD	Video on Demand
VOIP	Voice over Internet Protocol

* * *

About the author

Mark Eyer has worked in the field of satellite and digital television systems for over eighteen years. He has contributed to various standards-making activities in the ATSC, EIA, and SCTE. He is currently employed as the Director of Systems for Sony Electronics, Digital Media of America in San Diego.

Mark Eyer
Director, Systems
Digital Media of America
Sony Electronics
San Diego, CA 92127
(858) 942-7130
mark.eyer@am.sony.com

The Effect Of Digital Signal Levels On Analog Channels In A Mixed Signal Multiplex

Marc Ryba and Joseph B. Waltrich
Motorola Broadband Communications Sector

ABSTRACT

In a cable system carrying both analog and digital signals, the digital signals are typically run at –10 dB relative to the analog channels. This is done in order to minimize effects of digital third order distortion on the analog signals. Since distortions produced by digital signals are noise-like, their effect appears as a decrease in Carrier/Noise ratios in adjacent analog channels.

Selection of digital power levels is a compromise between minimizing effects of analog distortions (i.e.- burst errors) on the digital channels and maintaining an acceptable noise floor in the analog channels. This is of particular concern as the industry moves toward higher density modulation formats such as 256QAM. This paper will discuss the results of testing to determine optimum levels of digital signals added to a 77 channel analog multiplex.

BACKGROUND

In a typical cable system, analog channels are located in the lower portion of the spectrum and digital channels are appended to the high frequency end. In such a system, theory has shown that digital third order distortion produces a noise-like spectrum that extends into the lower adjacent analog channels. This phenomenon is known as Composite Intermodulation Noise (CIN) and has been described in earlier publications [1], [2]. The result is a decrease in the C/N ratio of the channels at the upper end of the analog spectrum. The decrease in C/N performance is proportional to the digital signal power as well as to the number of digital channels.

Although operating the digital signals at reduced power levels minimizes analog C/N effects, it makes the digital signals more susceptible to burst errors resulting from analog CSO/CTB peaking [3]. Therefore a series of laboratory tests were conducted to determine optimum digital/analog ratios for operation of a mixed signal multiplex.

TEST SYSTEM DESCRIPTION

Test Setup

The system configuration for the tests consisted of a 77 channel analog multiplex in the 55-550 MHz range combined with digital channels above 550 MHz. Tests were conducted with a total of 32, 16 and 8 digital channels. Digital power levels were varied from zero to –10 dB relative to peak analog carriers and C/(N+CIN) levels were measured throughout the analog spectrum.

A block diagram of the test system is shown in Fig. 1. A 77 channel analog headend was used to generate the analog signal multiplex. The analog channels were independently modulated by separate video test signals. The digital multiplex consisted of a maximum of 32 digital signals, received from various satellites and transcoded to 64QAM via a bank of Motorola IRT1000 Integrated Receiver/Transcoders. The IRT outputs were up-converted to RF using Motorola C6U up-converters and inserted into the combined signal multiplex starting at 552 MHz. The average digital power levels were adjusted in 2 dB increments from 0 to –10 dB relative to peak of analog sync. Regardless of the number of digital channels, the digital signals were always contiguous throughout the course of the testing.

reliable secondary indication of AGC effects on laser drive level and modulation. As the

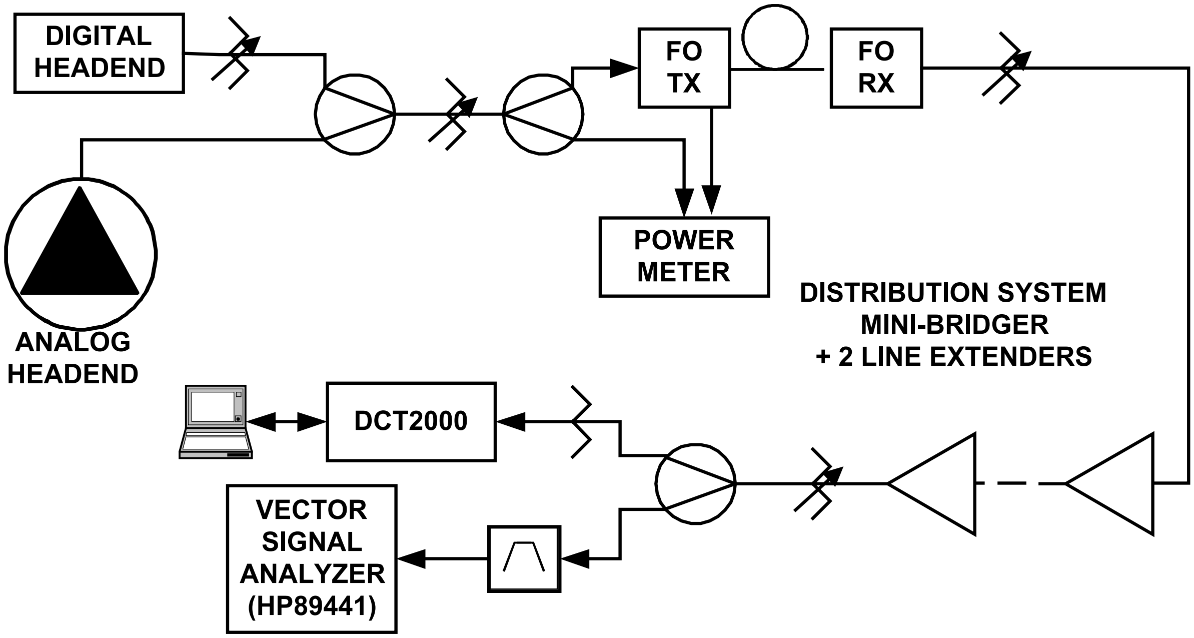


FIG. 1 - BLOCK DIAGRAM OF TEST SETUP

The distribution system used for the tests consisted of a 1310 nm laser transmitter driving a 20 Km fiber optic link with a 1.8 dB in-line optical attenuator followed by an RF system consisting of a Motorola MB-86SH/G mini-bridger, two Motorola BLE-86S line extenders and 21 taps. The RF distribution system was designed with 10 dB of tilt for 750 MHz spacing. The amplifiers were set up to be driven at their specified output levels for 77 NTSC carriers plus 200 MHz of compressed data (37/44/37 dBmV for 750/550/50 MHz, respectively).

The level of the combined analog/digital multiplex was adjusted to provide an input signal level that is near optimal for the fiber optic transmitter. With full channel loading, the fiber optic link was monitored for clipping effects and found to be several dB away from clipping according to Motorola BCS's laser clipping test procedure. A calibrated RF test point is mounted on the front panel of the laser transmitter module. This point was monitored via a power meter to provide a channel loading varies, the transmitter's microprocessor controlled AGC maintains a constant drive level to the laser and adjusts the total input power accordingly.

Analog C/N ratios were measured using a HP89441 Vector Signal Analyzer. The analog test signal was monitored after the second line extender throughout the test. This test point provided a signal that was greater than +25 dBmV per channel for accurate C/N measurements. A DCT2000 set top terminal was used to monitor digital signal quality during the course of the testing. The monitored digital channel was always located at the center of the digital spectrum.

Test Procedure

Unless otherwise specified, all measurements were performed in accordance with NCTA recommended practices and procedures.

Analog and digital signal levels were adjusted to provide a flat spectrum at the input to the distribution system. The peak analog power levels were set at +15 dBmV per channel. The average digital power levels were adjusted from 0 to –10 dBmV per channel relative to the analog carriers. The total average input power to the laser, after AGC, was approximately -17 dBm.

All C/(N+CIN) measurements were made at the first tap after the second line extender. The HP89441 Vector Signal Analyzer was used to measure RF analog C/N ratios and digital Es/No ratios. The peak level for the channel under test was maintained at a constant +25 dBmV at the input to the HP89441. A tunable bandpass filter, centered on the test channel, was used for all measurements to eliminate saturation and distortions at the input to the test equipment. In-band ripple of the tunable channel filters did not exceed 0.5 dB in a 6 MHz bandwidth.

A Motorola DCT2000, modified to run Broadcom QAMLink™ software, was used to monitor Bit Error Rate (BER) and Modulation Error Ratio (MER) of the digital channel under observation. The digital test channel was always located in the center of the digital spectrum.

TEST RESULTS

Tests with 32 Digital Channels

Test data are shown in Table 1. Table 2 presents the contributions of CIN to the analog noise floor. Plots of C/(N+CIN) and C/CIN are shown in Figs. 2 and 3, respectively. CIN was calculated by treating it as an additional noise source and subtracting the effect of Additive White Gaussian Noise (AWGN) (i.e. – the system noise measured with no digital channels present) as shown by Equation (1).

$$CIN=10*Log(10^{0.1*(AWGN+CIN)} - 10^{0.1*AWGN}) \quad (1)$$

Where CIN is expressed in dBc and AWGN is the system noise floor (dBc) with no digital channels present.

From Figs. 2 and 3 it is seen that the analog channels adjacent to the digital multiplex are affected most severely by the increase in the analog noise floor due to digital third order distortion. This is in accordance with theory [1].

Although FCC requirements [4] specify a minimum system C/N of 43 dB, most cable systems are designed to meet more stringent requirements – typically 49 dB at the end of the system. When digital channels are added to an analog multiplex, the effect of CIN must be treated as an additional noise source to be added to the system AWGN. Therefore system C/N design goals must include the effect of CIN as well as AWGN. Table 3 shows the system C/AWGN requirements for a 49 dB C/(N+CIN) design goal. The system C/AWGN values are calculated as follows:

$$C/AWGN = -10*Log(10^{-4.9} - 10^{0.1*CIN}) \quad (2)$$

Where C/AWGN is the system Carrier/AWGN in dB that must be maintained to meet the required C/(N+CIN) (49 dB in this case) and CIN is expressed in dBc.

The blank entry in Table 3 for channel 78 at 0 dB is indicative of a value for which the CIN is too large to meet a 49 dB design goal. That is, equation (2) yields the log of a negative number. The blank entries in Table 2 at the lowest frequencies are indicative of negligible CIN values (i.e. – no measurable difference between CIN and the system noise floor).

In practice, system C/AWGN values greater than the low 50's are difficult to achieve. An inspection of the data in Table 3 shows that in a cable system with a C/(N+CIN) target of 49 dB, the digital channels would have to be operated at power levels of –6 dB or lower relative to the analog channels in order to maintain desired C/(N+CIN) performance.

Tests With 16 Digital Channels

Test results are shown in Table 4. Table 5 presents the contributions of CIN to the analog noise floor. Plots of CIN plus system noise vs. frequency are shown in Fig. 4. Fig. 5 presents a plot of CIN vs. frequency. The blank entry in Table 5 indicates a negligible CIN value. Table 6 shows the C/AWGN values required to meet a 49 dB system C/(N+CIN) design goal.

As is the case with 32 digital channels, a 49 dB system C/(N+CIN) target would require that the digital signals be operated at –6 dB or lower relative to the analog carriers. It should be noted, however, that the CIN generated by 16 digital channels at equal power to the analog signals decreases more rapidly than the CIN for 32 channels in most of the channels immediately adjacent to the digital spectrum. This corresponds to theoretical predictions for an octave band decrease in the number of digital channels [1].

Tests With 8 Digital Channels

Test data and CIN contributions are presented in Tables 7 and 8, respectively and shown graphically in Figs. 6 and 7. Table 9 shows the system C/AWGN requirements needed to meet a design goal of 49 dB for C/(N+CIN).

From Table 9 it is seen that, in a typical cable system with a C/(N+CIN) target of 49 dB, the digital channels could be operated at a power level of –6 dB or lower relative to the analog channels. This would depend on how well the system AWGN could be controlled.

Since the analog channel most affected by CIN is the highest channel (i.e. – the lower adjacent channel to the digital spectrum), maintaining a desired C/(N+CIN) design goal in this channel should assure that the design goal would be met in the remainder of the analog spectrum. Fig. 8 presents plots of system C/AWGN vs. A/D ratio required to achieve a 49 dB C/(N+CIN) ratio in the worst case analog channel (EIA channel 78). From Fig. 8 it is seen that an A/D ratio of 6 dB or greater would be required to operate with a realistic C/AWGN level. Lower A/D ratios may be achievable, depending on the number of digital channels and the extent to which a system operator can control system AWGN.

CONCLUSIONS

Test results have shown that digital third order distortion may be regarded as an additional noise source that adds to system AWGN to produce an increase in the noise floor of the adjacent analog channels. Test data show that this effect is worst in the highest analog channel and decreases with decreasing analog channel frequency. Since digital signals are affected by CSO and CTB peaks, setting digital signal levels is a compromise between generation of digital distortion in the analog multiplex and optimizing digital signal robustness. To date, it has been a practice to set levels for 64QAM signals at –10 dB relative to analog carriers. As 256QAM is deployed, these levels may have to be increased. Test data show that a level of –6 dB relative to analog signals can be attained without generating objectionable CIN levels. The choice of an acceptable operating level is dependent on the number of digital channels in the system, the current C/AWGN performance and the extent to which the operator is able to control system AWGN.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to Bill Otto, Chris Lynch and Lenton Jones for their assistance in conducting the tests.

REFERENCES

[1] J. Waltrich, “Distortion Produced by Digital Transmission in a Mixed Analog and Digital System”, Communications Technology, April, 1993

[2] J. Hamilton, D. Stoneback, “The Effect of Digital Carriers on Analog CATV Distribution Systems’, 1993 NCTA Technical Papers

[3] O. Sniezko, D. Stoneback, R. Howald, “Distortion Beat Characteristics and the Impact on QAM BER Performance”, 1999 NCTA Technical Papers

[4] FCC Regulations, Part 76

EIA Channel	2	16	7	12	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	**57**	**135**	**177**	**207**	**273**	**327**	**393**	**459**	**513**	**519**	**525**	**531**	**537**	**543**	**549**
A/D Ratio (dB)	**System Additive White Gaussian Noise + CIN (dBc)**														
0	-53.5	-52.8	-52.5	-51.9	-53.1	-52.1	-52.0	-52.5	-49.9	-49.6	-49.7	-50.3	-50.4	-49.8	-48.1
2	-54.1	-53.7	-53.5	-53.4	-54.6	-54.0	-54.0	-54.7	-53.3	-53.0	-53.1	-53.5	-53.7	-53.2	-50.9
4	-54.6	-54.0	-53.7	-54.0	-55.3	-55.0	-55.1	-56.2	-55.2	-54.7	-54.8	-55.4	-55.5	-55.2	-52.7
6	-54.7	-54.3	-54.1	-54.4	-55.7	-55.7	-55.7	-56.7	-56.2	-55.5	-55.7	-56.3	-56.4	-56.2	-54.5
8	-55.1	-54.4	-54.3	-54.5	-55.9	-56.0	-55.9	-56.9	-56.6	-56.0	-56.1	-56.6	-56.9	-56.7	-55.1
10	-55.2	-54.6	-54.4	-54.5	-56.0	-56.0	-56.0	-57.3	-56.8	-56.3	-56.5	-56.9	-57.1	-56.9	-56.2
No Digital	-55.2	-54.5	-54.3	-54.6	-56.3	-56.2	-56.4	-57.4	-57.3	-56.8	-56.9	-57.3	-57.6	-57.7	-57.3

Table 1 – System AWGN + CIN for 32 Digital Channels

EIA Channel	2	16	7	12	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	**57**	**135**	**177**	**207**	**273**	**327**	**393**	**459**	**513**	**519**	**525**	**531**	**537**	**543**	**549**
A/D Ratio (dB)	**CIN (dBc)**														
0	-58.4	-57.7	-57.2	-55.3	-55.9	-54.2	-54.0	-54.2	-50.8	-50.5	-50.6	-51.3	-51.3	-50.6	-48.7
2	-60.6	-61.4	-61.2	-59.7	-59.5	-58.0	-57.7	-58.0	-55.5	-55.3	-55.4	-55.8	-56.0	-55.1	-52.0
4	-63.5	-63.6	-62.6	-62.5	-62.2	-61.2	-61.0	-62.4	-59.4	-58.9	-59.0	-59.9	-59.7	-58.8	-54.5
6	-64.3	-67.8	-67.6	-68.3	-64.6	-65.3	-64.0	-65.0	-62.7	-61.4	-61.9	-63.2	-62.6	-61.5	-57.7
8	-71.5	-70.8		-71.2	-66.5	-69.5	-65.5	-66.5	-64.9	-63.7	-63.8	-64.9	-65.2	-63.6	-59.1
10				-71.2	-67.8	-69.5	-66.6	-73.7	-66.4	-65.9	-67.1	-67.5	-66.7	-64.6	-62.7

Table 2 – CIN Contribution to Analog Noise Floor for 32 Digital Channels

EIA Channel	2	16	7	12	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	**57**	**135**	**177**	**207**	**273**	**327**	**393**	**459**	**513**	**519**	**525**	**531**	**537**	**543**	**549**
A/D Ratio (dB)	**System C/AWGN (dB)**														
0	49.5	49.6	49.7	50.2	50.0	50.5	50.7	50.6	53.7	54.3	54.1	52.9	52.8	54.2	
2	49.3	49.3	49.3	49.4	49.4	49.6	49.6	49.6	50.1	50.1	50.1	50.0	50.0	50.2	52.0
4	49.2	49.2	49.2	49.2	49.2	49.3	49.3	49.2	49.4	49.5	49.5	49.4	49.4	49.5	50.4
6	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.2	49.3	49.2	49.2	49.2	49.2	49.6
8	49.0	49.0	49.0	49.0	49.1	49.0	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.2	49.4
10	49.0	49.0	49.0	49.0	49.1	49.0	49.1	49.0	49.1	49.1	49.1	49.1	49.1	49.1	49.2

Table 3 – System C/AWGN Required to Meet a 49 dB C/(N+CIN) Design Goal in a System with 32 Digital Channels

Fig. 2 - System AWGN + CIN for 32 Digital Channels

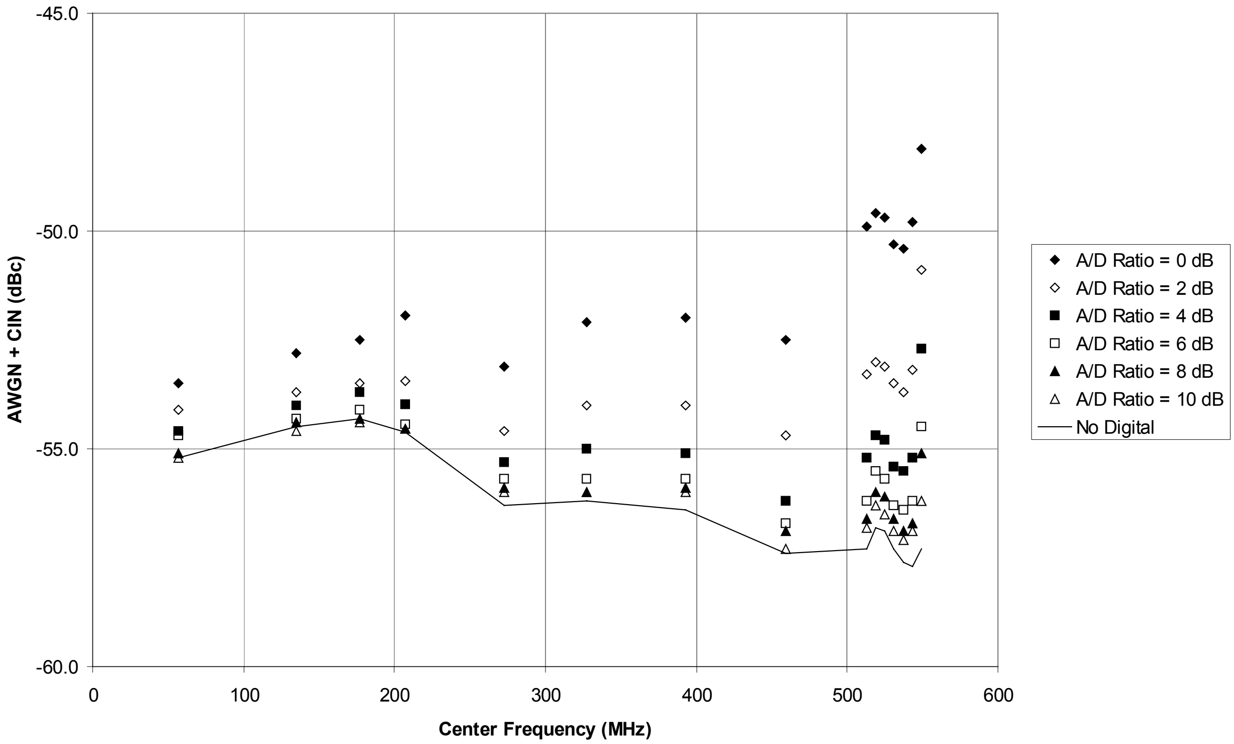
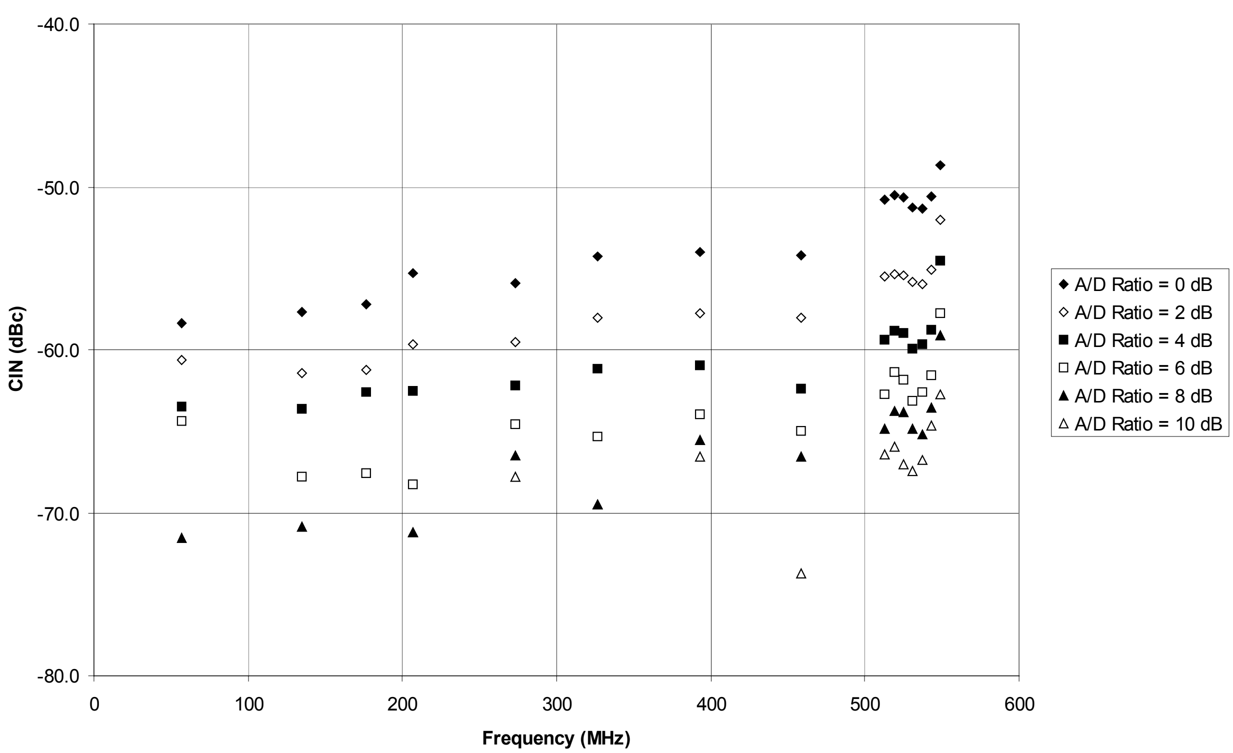


Fig. 3 - CIN Contribution to Analog Noise Floor for 32 Digital Channels



EIA Channel	2	16	7	12	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	57	135	177	207	273	327	393	459	513	519	525	531	537	543	549
A/D Ratio (dB)	System Additive White Gaussian Noise + CIN (dBc)														
0	-54.0	-53.3	-53.5	-54.0	-54.0	-54.7	-54.7	-55.0	-54.0	-53.0	-53.7	-53.9	-54.0	-53.5	-49.6
2	-54.3	-53.6	-54.0	-55.0	-54.8	-55.4	-55.4	-56.0	-55.0	-55.0	-55.2	-55.5	-56.0	-55.2	-51.7
4	-54.5	-54.0	-54.2	-55.0	-55.4	-56.0	-56.0	-56.6	-56.0	-56.0	-56.0	-56.0	-56.0	-56.1	-53.2
6	-54.8	-54.2	-54.3	-55.0	-55.7	-56.4	-56.4	-56.9	-56.0	-56.0	-56.4	-56.5	-57.0	-56.6	-54.6
8	-54.9	-54.2	-54.6	-56.0	-55.7	-56.5	-56.5	-57.0	-57.0	-56.0	-56.6	-56.8	-57.0	-56.9	-55.4
10	-54.9	-54.4	-54.6	-56.0	-55.7	-56.6	-56.6	-57.2	-57.0	-56.0	-57.1	-56.9	-57.0	-57.1	-56.4
No Digital	-55.2	-54.4	-54.7	-56.0	-55.8	-56.8	-56.8	-57.4	-57.0	-57.0	-57.2	-57.2	-58.0	-57.6	-58.3

Table 4 – System AWGN + CIN for 16 Digital Channels

EIA Channel	2	16	7	12	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	57	135	177	207	273	327	393	459	513	519	525	531	537	543	549
A/D Ratio (dB)	CIN (dBc)														
0	-60.2	-59.8	-59.7	-58.7	-58.7	-58.9	-58.9	-58.7	-56.3	-55.7	-56.3	-56.6	-56.8	-55.6	-50.2
2	-61.6	-61.3	-62.3	-61.7	-61.7	-61.0	-61.0	-61.6	-59.9	-58.9	-59.5	-60.4	-59.8	-58.9	-52.8
4	-62.8	-64.6	-63.8	-63.4	-66.0	-63.7	-63.7	-64.3	-62.4	-61.4	-62.2	-62.2	-62.5	-61.4	-54.8
6	-65.4	-67.7	-64.9	-64.1	-72.1	-67.0	-67.0	-66.5	-64.6	-63.7	-64.1	-64.8	-65.1	-63.5	-57.0
8	-66.7	-67.7	-71.0	-69.1	-72.1	-68.3	-68.3	-67.6	-67.2	-65.9	-65.5	-67.4	-66.6	-65.2	-58.5
10	-66.7		-71.0	-69.1	-72.1	-70.1	-70.1	-70.7	-70.3	-67.0	-73.5	-68.7	-69.0	-66.7	-60.9

Table 5 – CIN Contribution to Analog Noise Floor for 16 Digital Channels

EIA Channel	2	16	7	12	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	57	135	177	207	273	327	393	459	513	519	525	531	537	543	549
A/D Ratio (dB)	System C/AWGN (dB)														
0	49.3	49.4	49.4	49.5	49.5	49.5	49.5	49.5	49.9	50.0	49.9	49.8	49.8	50.1	55.1
2	49.2	49.3	49.2	49.2	49.2	49.3	49.3	49.2	49.4	49.5	49.4	49.3	49.4	49.5	51.4
4	49.2	49.1	49.1	49.2	49.1	49.1	49.1	49.1	49.2	49.3	49.2	49.2	49.2	49.3	50.3
6	49.1	49.1	49.1	49.1	49.0	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.2	49.7
8	49.1	49.1	49.0	49.0	49.0	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.1	49.5
10	49.1	49.0	49.0	49.0	49.0	49.0	49.0	49.0	49.0	49.1	49.0	49.0	49.0	49.1	49.3

Table 6 – System C/AWGN Required to Meet a 49 dB C/(N+CIN) Design Goal in a System with 16 Digital Channels

Fig. 4 - System AWGN + CIN for 16 Digital Channels

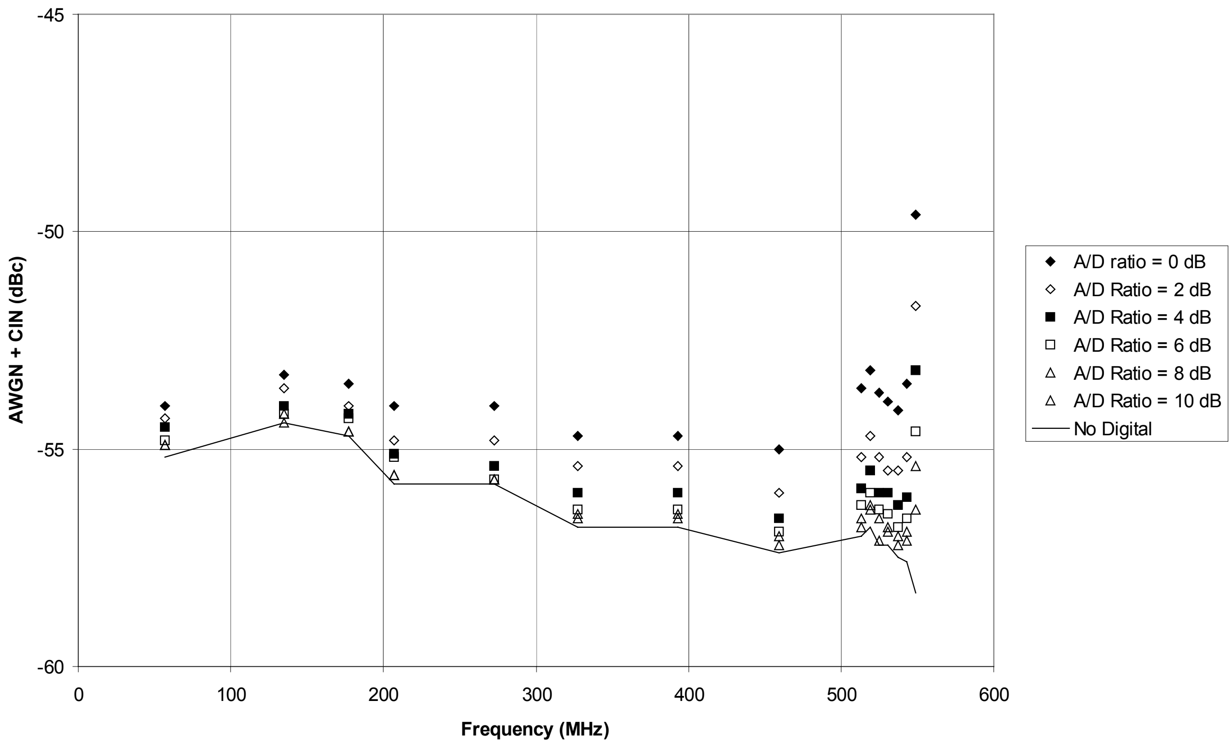
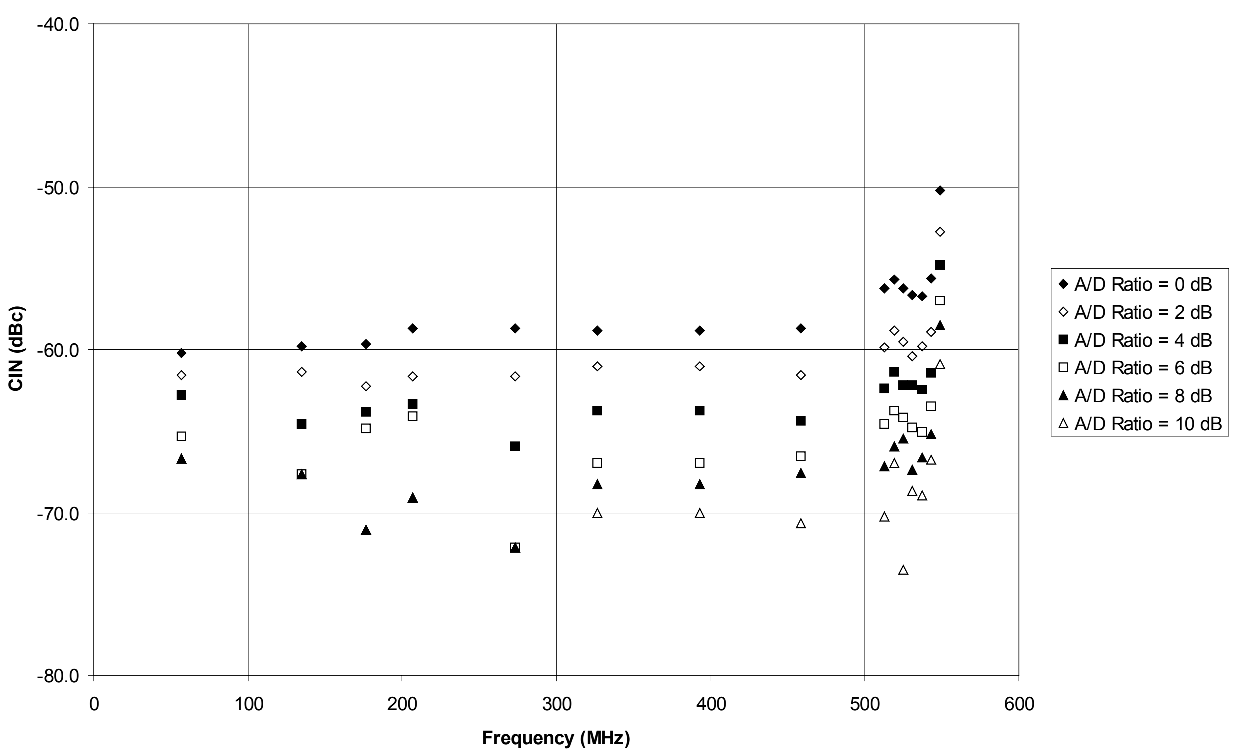


Fig. 5 - CIN Contribution to Analog Noise Floor for 16 Digital Channels



EIA Channel	2	16	7	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	**57**	**135**	**177**	**273**	**327**	**393**	**459**	**513**	**519**	**525**	**531**	**537**	**543**	**549**
A/D Ratio (dB)	**System Additive White Gaussian Noise + CIN (dBc)**													
0	-54.3	-54.6	-53.9	-55.1	-54.4	-55.5	-56.7	-55.8	-55.3	-55.9	-56.2	-56.4	-55.7	-50.5
2	-54.6	-54.8	-54.2	-55.4	-54.7	-55.9	-57.2	-56.5	-55.8	-56.4	-56.9	-57.2	-56.6	-51.9
4	-55.0	-55.1	-54.4	-55.6	-55.0	-56.2	-57.5	-56.8	-56.3	-56.8	-57.2	-57.5	-57.2	-53.5
6	-55.1	-55.3	-54.5	-55.8	-55.1	-56.5	-57.5	-57.1	-56.6	-57.0	-57.5	-57.7	-57.5	-54.7
8	-55.1	-55.3	-54.7	-55.9	-55.2	-56.6	-57.7	-57.2	-56.7	-57.1	-57.7	-57.8	-57.8	-55.8
10	-55.2	-55.3	-54.7	-56.0	-55.4	-56.6	-57.8	-57.5	-56.7	-57.3	-57.7	-58.0	-57.9	-56.5
No Digital	-55.4	-55.4	-54.7	-56.1	-55.6	-56.8	-57.8	-57.6	-57.1	-57.5	-58.0	-58.2	-58.4	-58.3

Table 7 – System AWGN + CIN for 8 Digital Channels

EIA Channel	2	16	7	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	**57**	**135**	**177**	**273**	**327**	**393**	**459**	**513**	**519**	**525**	**531**	**537**	**543**	**549**
A/D Ratio (dB)	**CIN (dBc)**													
0	-60.8	-62.3	-61.6	-62.0	-60.6	-61.4	-63.2	-60.5	-60.0	-61.0	-60.9	-61.1	-59.0	-51.3
2	-62.3	-63.7	-63.8	-63.7	-62.0	-63.2	-66.1	-63.0	-61.7	-62.9	-63.4	-64.1	-61.3	-53.0
4	-65.6	-66.9	-66.2	-65.2	-63.9	-65.1	-69.3	-64.5	-64.0	-65.1	-64.9	-65.8	-63.4	-55.2
6	-66.9	-71.7	-68.0	-67.6	-64.7	-68.3	-69.3	-66.7	-66.2	-66.6	-67.1	-67.3	-64.8	-57.2
8	-66.9	-71.7		-69.4	-65.8	-70.1	-74.1	-67.8	-67.3	-67.7	-69.5	-68.4	-66.7	-59.4
10	-68.7	-71.7		-72.4	-68.9	-70.1		-73.9	-67.3	-70.8	-69.5	-71.5	-67.5	-61.2

Table 8 – CIN Contribution to Analog Noise Floor for 8 Digital Channels

EIA Channel	2	16	7	32	41	52	63	72	73	74	75	76	77	78
Center Frequency (MHz)	**57**	**135**	**177**	**273**	**327**	**393**	**459**	**513**	**519**	**525**	**531**	**537**	**543**	**549**
A/D Ratio (dB)	**System C/AWGN (dB)**													
0	49.3	49.2	49.2	49.2	49.3	49.3	49.2	49.3	49.4	49.3	49.3	49.3	49.5	52.9
2	49.2	49.1	49.1	49.2	49.2	49.2	49.1	49.2	49.2	49.2	49.2	49.1	49.3	51.2
4	49.1	49.1	49.1	49.1	49.1	49.1	49.0	49.1	49.1	49.1	49.1	49.1	49.2	50.2
6	49.1	49.0	49.1	49.1	49.1	49.1	49.0	49.1	49.1	49.1	49.1	49.1	49.1	49.7
8	49.1	49.0	49.0	49.0	49.1	49.0	49.0	49.1	49.1	49.1	49.0	49.1	49.1	49.4
10	49.0	49.0	49.0	49.0	49.0	49.0	49.0	49.0	49.1	49.0	49.0	49.0	49.1	49.3

Table 9 – System C/AWGN Required to Meet a 49 dB C/(N+CIN) Design Goal in a System with 8 Digital Channels

Fig. 6 - System AWGN + CIN for 8 Digital Channels

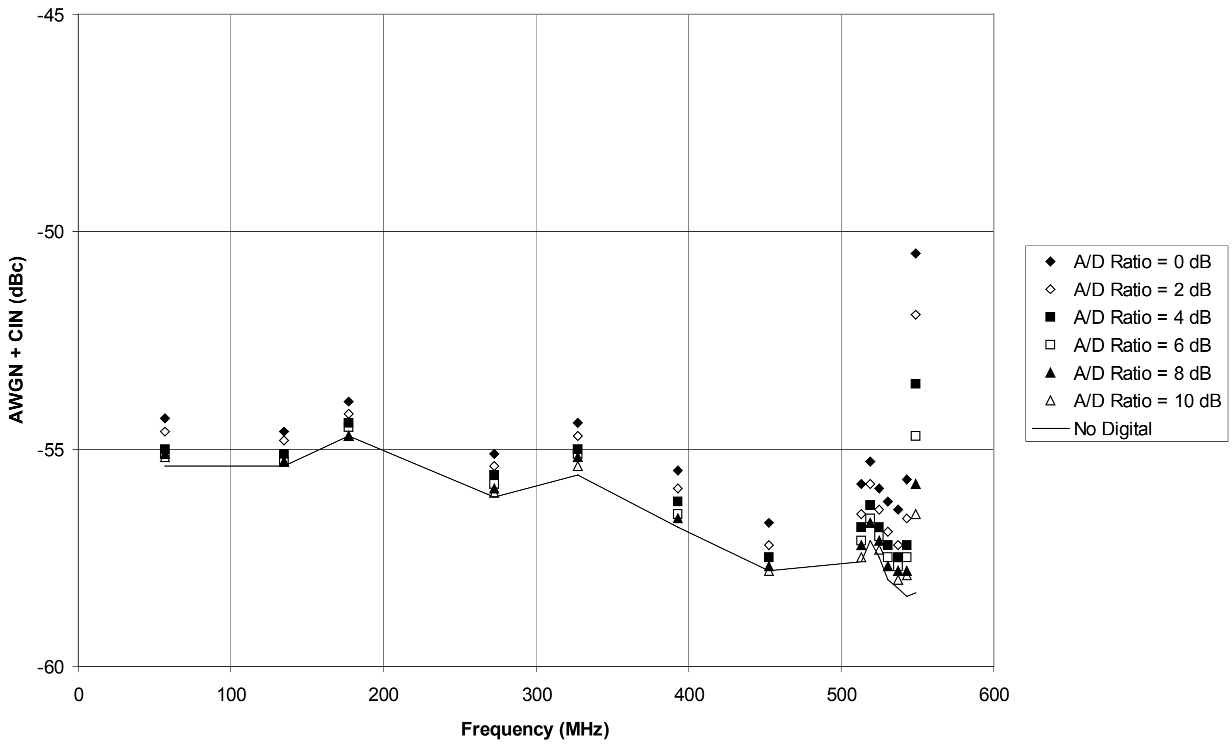


Fig. 7 - CIN Contribution to Analog Noise Floor for 8 Digital Channels

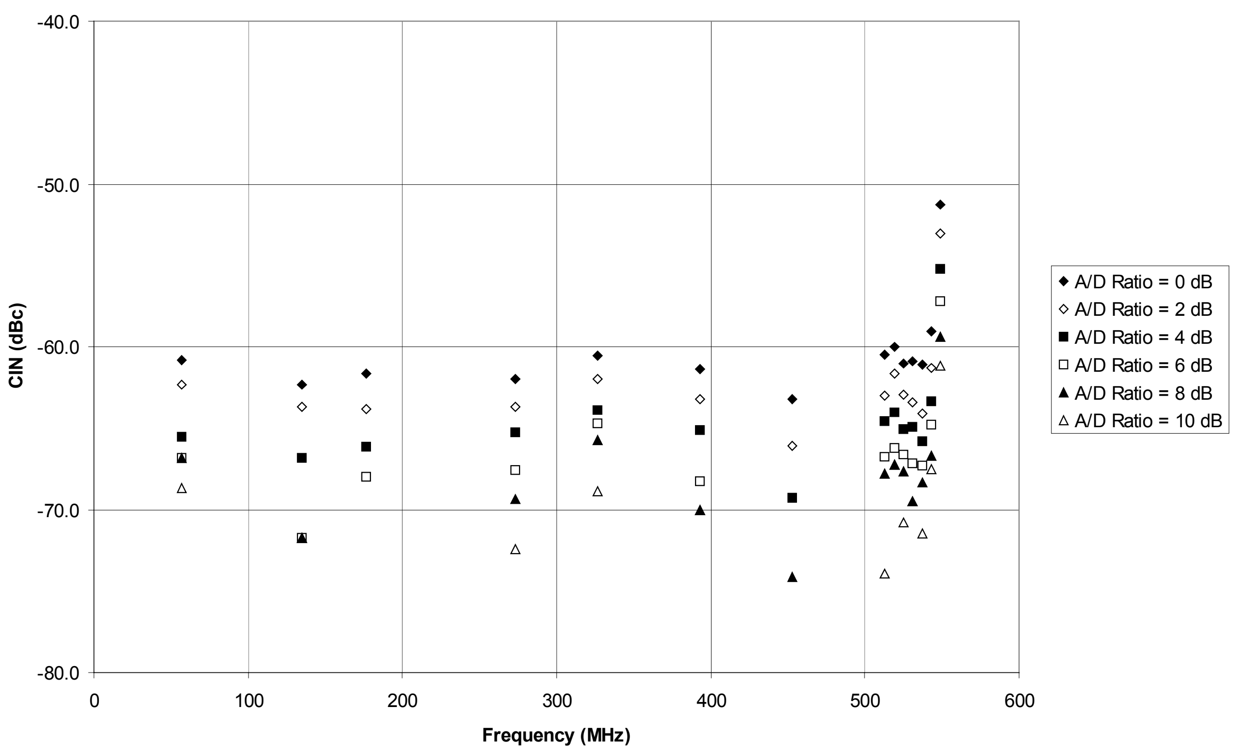
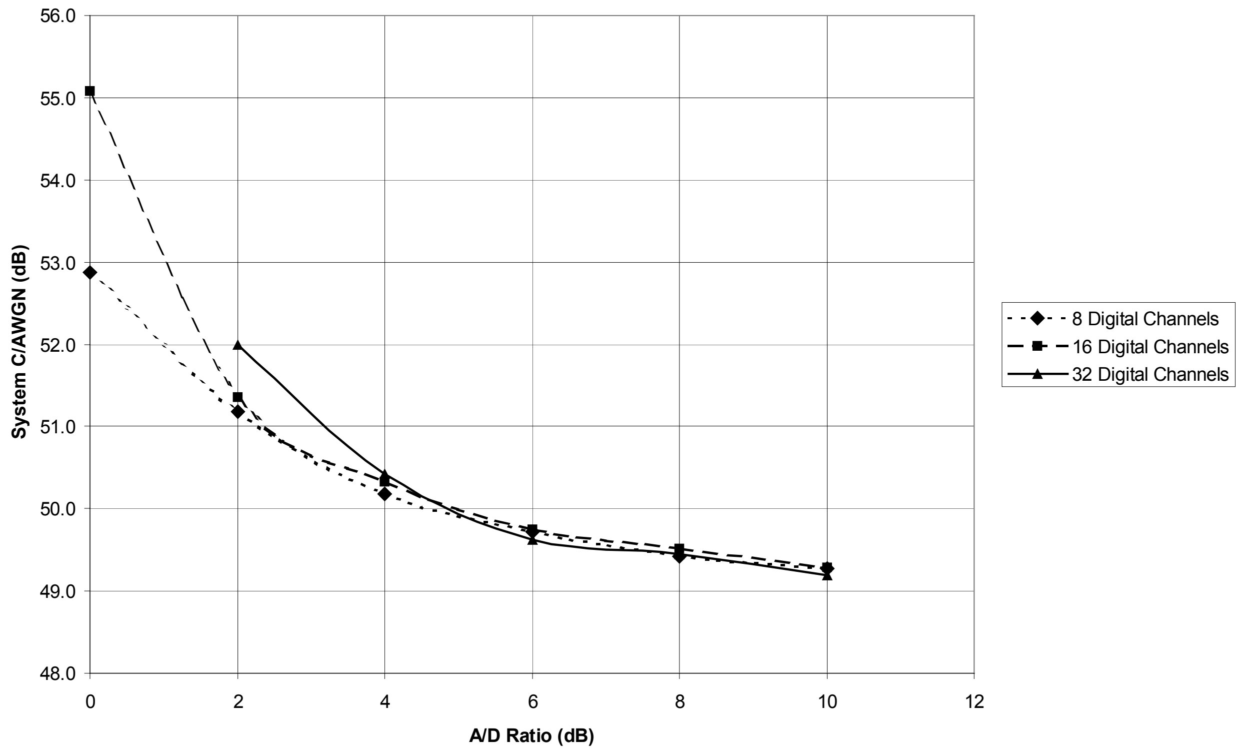


Fig. 8 - System C/AWGN Required to Meet 49 dB C/(N+CIN) in Channel 78



Traffic Engineering for Video-On-Demand Systems

Bill Wall
Scientific-Atlanta

Abstract

Traffic engineering techniques borrowed from the telephony world are used predict the performance of VOD systems. This paper provides a brief overview of traffic theory and how it can be applied to VOD systems. Erlang's B formula is used to calculate service blockage as function of buy rate, system capacity, and service group size. The implications of these calculations are discussed.

THE PROBLEM

Video-On-Demand (VOD) Systems are now a commercial reality and are operational in a number of cable systems. These systems offer the cable operator the potential for significant revenue increase, but also represent a significant capital investment. The objective, of course, is to maximize the revenue for the minimum investment. In VOD systems, individual video *streams* are created for each active user of the system. The incremental capital cost of adding additional stream capacity is relatively linear, and the cost of VOD systems is often measured in *cost per stream*. If a cable system is under provisioned, that is the peak demand for VOD outstrips the system capacity, then customer dissatisfaction with not being able to get on-demand services on demand may actually cause a loss of revenue. On the other hand, over provisioning by having a higher stream capacity than peak demand costs more than is necessary for delivering the service. Traditional wisdom, based on trial experience, has suggested that VOD system stream capacity should be 10% of the potential VOD customer base. For example, if a cable system has 40,000 digital subscribers capable of receiving VOD, then the VOD system should support 4000 individual video streams. One objective of this paper is to examine that premise and its assumptions. The problem then is to determine what is the optimum capacity needed to support user demand, and what is the optimum way to deploy that capacity.

CONCEPT OF SERVICE GROUPS

In the previous example, 4000 video streams were required for the 40,000 digital subscribers. Assume for the moment that these numbers represent a cable system of 100,000 homes passed, 80% cable penetration, and 50% of those subscribers take digital. Also for the moment assume that the 10% peak VOD usage represents the correct VOD deployment for the system. Typical encoding rates for good quality VOD are about 3.5 Mbps. Using 256 QAM in the plant, with a payload of 38 Mbps, each QAM modulator can carry 10 video streams. The capacity for 4000 streams then requires 400 QAM modulators to support the VOD service. If the cable system is built with 500 homes passed/fiber node, then we could distribute two QAM modulators to each of the 200 fiber nodes, using 12 MHz of plant bandwidth. The same two six MHz channels would be used for VOD applications in all nodes. Alternatively we could feed identical signals to two fiber nodes forming a *service group* of 1000 homes passed that are now logically one group. Here we could feed four QAM modulators to each service group and achieve the same 10% stream capacity, but now using 24 MHz of spectrum. In a similar manner we could combine 4 nodes to form service groups of 2000 homes passed and use eight QAM modulators, occupying 48 MHz of bandwidth, and so forth… Clearly the first case used the least spectrum, but are there advantages of larger service groups? Intuitively, larger groups should have some advantage by “averaging” over a larger population of users during peak usage times. Does this advantage exist and can it be quantified? Similar questions have been dealt with for generations in telephony systems using traffic engineering.

FUNDAMENTALS OF TRAFFIC ENGINEERING

Methods for determining how much capacity is required as a function of expected demand are called *traffic engineering methods.* These methods, a branch of applied probability theory, have been developed over years to accurately predict the performance of telephone and other telecommunications networks.

The demand on a traffic system is called the *offered load* and is the product of the average rate of customer requests (*average arrival rate, r*) and the average time they require service (*average hold time, t*), or offered load, *a*, is given by

$$a = rt$$

The value *a* is dimensionless and expressed in *erlangs*, named after the founder of traffic theory, A. K. Erlang. If the instantaneous demand on the system exceeds the capacity of the systems then the call is *blocked.* Two classes of systems are typically used in the telecommunications world. In the first class, when a call is blocked, it is dropped, and the user must retry at a later time. These systems are called *blocked-calls-cleared* (BCC) systems. In the second class, when a call is blocked, it is put in queue to be serviced at a later time; these are called *blocked-call-delayed* (BCD) systems. Hybrids of the two classes are also popular, where a fixed-length queue is used, but when the queue is filled, calls are dropped.

Telephone traffic cannot be predicted exactly, but may be viewed as statistical processes. A common assumption is that the probability of a call arrival during an interval T is proportional only to the length of the interval, and the constant of proportionality is the average arrival rate *r*. This assumption leads to the fact that the probability that k calls arrive in an interval T is described by a *Poisson distribution*, and any process following this distribution is a *Poisson process.* This process gives an accurate description of telephone call arrivals. Calls exiting the network are assumed to follow a similar process, in an interval T, each call will terminate with probability T/*t* where *t* is the average hold time. This leads to a *negative exponential distribution* H(T) denoting the probability of a given call lasting for a duration of T.

Assuming a BCC system that has a capacity of supporting c calls, and a random load *a* is offered, then Erlang showed that the probability B of an arriving call being blocked is given by the formula

$$B(c, a) = \frac{a^c/c!}{\sum_{k=0}^{c} a^k/k!}$$

This equation is often referred to as the Erlang Loss Formula, or Erlang B Formula and is central in the planning of telecommunications systems. Similar formula can be derived for BCD systems, as well as a calculation for average delay in the queue. These results can be found in most traffic engineering texts.

APPLICATIONS TO VOD SYSTEMS

In many ways VOD systems are analogous to telephony systems. A single VOD session connects a client to a server, much like a telephone call connects two users. VOD sessions and telephone calls both use a similar connection procedure; in fact the DSM-CC session set up procedure was loosely based on Q.931 call setup procedure. Both are initiated randomly by a user. Both last some finite time and terminate. The most straightforward way to model a VOD system is as a BCC system, where when the system is busy, service is denied and a user must try again later. BCD systems could be implemented, but with the relatively long average hold time of a VOD movie, queuing delay could be unacceptably long. This is perhaps a topic for further investigation. Service request arrival statistics should be similar to that of a telephony system, and a Poisson distribution should accurately model this process. However call hold time in a VOD system is less likely to be as random as a telephone call due to the deterministic nature of the fixed length of a video program. However users will terminate early, invoke pause and rewind functions that will extend the length of the program, and programs will vary in length. Perhaps a Gaussian or Raleigh distribution would more closely match VOD systems hold time statistics than the negative exponential distribution. Here we have very little data from fielded systems. This area is clearly one where more work needs to be done. With these caveats, we press forward and use the Erlang B formula to calculate blocking probabilities in a service group and explore the blocking percentage as a function of service group size and buy rates. One remaining key issue is how to relate buy rates to peak offered loads.

ESTIMATION OF PEAK BUSY HOUR

Traffic engineering theory generally assumes that the processes are stationary, which means the parameters describing the process (average arrival time and average hold time) are constant or vary slowly compared to the actual call rate. Traffic loads do vary with the time of day and day of week, as well as season, but in general in telephony systems these variations are slow enough that traffic theory works well. In order to provide high reliability and a high level of customer satisfaction, telephone systems are engineered based on the busiest hour of the day in the busiest season. It would seem appropriate to engineer VOD systems to similar criteria. To date, there is limited public data detailing buy rates for VOD systems versus time of day and time of week. For the calculations shown as example in this paper, some assumptions must be made.

The buy rate model used in the following analysis makes two simplistic assumptions, first that all VOD buys occur in a six hour time segment each evening, and second that Saturday night buy rates are double the buy rates of other nights. We know the first assumption overstates the buy rates during primetime, but the second assumption most likely understates the popularity of weekend primetime. In some manner this may come closer to actual peak rates on weekends. This model would predict that 15% of all buys occur during a three hour primetime period on Saturday night. Cable operators are used to thinking of Pay-Per-View in terms of buys per month. Using the above model we relate peak busy hour average arrival time to buys per month. Based on a four week month, and six hour per day buy period yields a 168 hour/month buy opportunity. The average buy rate per hour for b buys per month per sub would be b/168 and the peak buy rate would be 2b/168 or 0.012b. This value times the number of subs per service group yields the average arrival rate for peak busy hour for that service group.

CALCULATION OF BLOCKAGE RATES

The first case examined looks at the sensitivity to service group size for a fixed percentage rate of VOD deployment. The baseline assumption is a plant design of 500hp per fiber node, 80% subscriber penetration, and 20% digital penetration. This first case looks at a fixed buy rate of four per month. Table 1 shows the relevant parameters and the calculated blockage rates. Average hold time used was two hours, to correspond with average movie length.

Number of Nodes	Service Group Size	Number of Digital Subscribers	Number of QAMs	Number of Streams	Offered Load	Blockage Probability
1	500	80	1	10	7.6	10.4%
2	1000	160	2	20	15.2	4.9%
3	1500	240	3	30	22.8	2.7%
4	2000	320	4	40	30.4	1.7%
6	3000	480	6	60	45.6	0.7%

Table 1 - Probability of Blocking vs. Service Group Size

Note first that the offered load using this model is just slightly below the suggested 10% VOD provisioning number, which implies that small statistical variations above the offered load would block at that deployment level. The actual deployment level in this example was 12.5% of digital subscribers in order to match up with QAM granularity. Note also the strong dependence of blocking probability with service group size. This result validates our earlier thought that larger service groups would provide better "averaging" of the load. With this level of digital deployment and this buy rate, a service group size of 2000 homes passed has a blocking probability of under 2%
.

Buys/ Month	Offered Load	Blocking Probability		
		30 Streams	40 Streams	50 Streams
3	22.8	2.7%	.032%	3E-5%
4	30.4	13.9%	1.7%	.03%
5	38.4	27.5%	9.5%	1.2%
6	45.6	37.5%	19.3%	6.0%
7	53.3	45.8%	29.0%	14.2%
8	60.8	52.1%	36.8%	22.5%

Table 2 – Blocking Probability

The second case examined looks at blocking probability as a function of buy rate and level of VOD capacity. The parameters examined would correspond to digital deployment of either 20% in a 2000 homes passed service group or 40% digital deployment in a 1000 homes passed service group. Table 2 lists the relevant parameters and the blocking probabilities.

Note the sensitivity to buy rates, which suggests that it would be impractical to provision a system where blocking does not occur. These results can give an operator a feel for the issues involved in planning a VOD system, however because of the assumptions made in determining average arrival rate for peak busy hour, a better model is needed. Before engineering a VOD system based on these methods, real data needs to be collected and used to determine peak busy hour, and predicted blocking rates need to be verified against real data.

CONCLUSION

A method has been described that can be used in helping to engineer the deployment of VOD systems. Before it can reliably be used, the method needs to be verified against field data. Once verified, this method can help operators design VOD systems and make the business tradeoffs in terms of capital expended versus the probability that a customer is denied service when he attempts to use the system. This method can also be used to aid engineering the tradeoff between service group size and spectrum used for VOD services. One result shows that it will be likely that deployed systems will have occasional denial of service during peak periods, and marketing techniques will need to be developed to cope with this fact. Systems with queuing of requests when capacity is full need to be explored as well. This preliminary analysis also suggests that the 10% capacity rule may be on the low side, but more data is needed. Finally, traffic engineering has been used successfully in telephone systems for decades, and should provide an important tool for the cable industry.

Unified Headend Technical Management Of Digital Services

Ran Oz and Amir Bassan-Eskenazi
BigBand Networks, Inc.
David Large
Media Connections Group

Abstract

Historically, cable system designs were based on a one-way broadcast model whereby all signals were sent throughout the network. Even though operators are now introducing digital services and subscriber-specific information streams, they are forced by this broadcast legacy to make fixed allocations of expensive equipment and bandwidth to each new service. This inefficient methodology fails to leverage the inherent advantages of cable's two-way hybrid fiber/coaxial (HFC) networks relative to the networks of competitive broadband service providers.

The authors argue that cable's multiple system operators (MSOs) face an imperative to convert their service delivery architecture from its hard-wired legacy to a truly switched and routed fabric in which resources and bandwidth are dynamically assigned to sessions, of video, data, voice, or any multimedia form, in response to real-time subscriber demand.

The new headend element required to achieve this is a new class of product, the Broadband Multimedia Router, through which all digital services will be controlled.

THE COMPETITIVE EVOLUTION TOWARDS MERGED APPLICATIONS

Homes and businesses access many services from networked connections: telephone, video, Internet access, etc. Traditionally, those services have been delivered by separate suppliers over largely-separate networks.

It is increasingly clear that these networks are converging to carry the full variety of services. For subscribers, this has advantages such as streamlining service provider relationships and the potential availability of new hybrid services that combine various media forms.

Network operators welcome the opportunity to maximize revenues by offering multiple services and to achieve cost efficiencies by combining services into singular, well-managed conduits. In cable there is clear optimism about this future, as indicated by the soaring values of cable systems.

Cable television has been moving aggressively by offering both innovative new video services and non-video services, such as high-speed data and telephony. Video-on-demand (VOD) trials are being widely conducted with a new sense of optimism; millions of people use cable modems for broadband Internet access; and MSOs who offer telephony are achieving very meaningful penetration levels.

The advantages of convergence have not been overlooked by cable's competitors. Direct broadcast satellite (DBS), traditional telcos and fixed wireless operators (multichannel and local multipoint distribution service, or MMDS and LMDS) are all positioning themselves to leverage their network infrastructures to deliver non-traditional services.

The major DBS operators now offer enhanced services including integrated Web access, email and digital video recorder (DVR). While their capabilities do not match cable's, they can provide some VOD-like features, Internet-on-TV display and direct links between television ads and related Web sites.

Major incumbent telephone companies are investing billions of dollars in their plants to provide digital subscriber line (DSL) service at speeds of several megabits per second, capable of carrying high-quality video.[1] Competitive DSL data carriers are now starting to announce video initiatives of their own. These companies want to be full-service suppliers to consumers, and strive to present their offerings as being functionally equivalent or superior to those offered by cable systems.

Finally, MMDS/LMDS companies, which have been only marginally successful in delivery of analog video, are taking advantage of technological developments from some of the most prominent global technology companies in their plans to offer high-speed data.[2] While fixed wireless networks are particularly challenged, from engineering and cost perspectives, in their efforts to achieve broadband speeds, they do hold the enticing promise of enabling mobility in access to converged services.

Clearly, this is a highly competitive and changing marketplace. Each type of network brings its own inherent advantages and disadvantages to this competition. It is incumbent upon MSOs to leverage their HFC advantages in order to compete and prosper.

CABLE'S LEGACY ARCHITECTURE

Historically, cable system architecture was almost exclusively coaxial tree-and-branch, whereby all subscribers were fed from one or a few trunk lines leaving the headend. All services were transmitted in common throughout the network, with various access control methods utilized to limit optional service reception to selected customers.

This network was optimized for the one-way delivery of analog video channels and was constructed as economically as possible for that purpose. For many years the primary innovations in the network were limited to RF bandwidth expansion, reach extension, and access control methods.

While the network was a model of cost-effective engineering and performed its task well as the cable market grew, it was not suited to certain types of new services. In particular:

- There was no effective way to subdivide the subscribers into smaller groups (required to support any significant level of subscriber-specific communications).
- Nearly the entire spectrum was consumed by bandwidth-intensive analog video signals.
- The reliability was limited by the long equipment cascades.
- Two-way communications were limited by both cascade lengths and the large number of network branches.

UPGRADE TO HFC

At its most basic level, a hybrid fiber/coaxial network is one in which optical fibers carry the signals part of the way from a central location to subscribers' terminal equipment. At the local distribution level, it usually means that a portion of the former coaxial trunk network is replaced by a functionally-equivalent linear fiber link, leaving some residual coaxial distribution equipment between each fiber node and subscribers. Depending on the implementation, there may also be higher optical network layers which may carry either analog or digital signals between a headend and intermediate hubs.

The introduction of HFC technology enabled cost-effective bandwidth upgrades and improved reliability. Additionally, it provided the theoretical ability to feed different signals to each of the various segments of the network by re-use of channels, a technique sometimes known as space division multiplexing (SDM).

HFC architectures also make the operation of high-speed, two-way communications practical through reduced sizes, in terms of both cascade depth and number of subscribers, of the individual coaxial distribution sub-networks.

INTRODUCTION OF DIGITAL SERVICES AND STANDARDIZATION OF DIGITAL TRANSPORT

The latest round of downstream spectrum expansion has differed from previous upgrades in that the upper portion of the spectrum is usually reserved exclusively for carriage of digital signals. MSOs are making business commitments to use these assets to tap new revenue streams through introduction of advanced services. This is the first step in the eventual, inevitable, analog-to-digital transition for all cable-delivered signals.

Digital carriage over cable is made possible by several fundamental advances. MPEG compression enables 8-12 video streams to share a 6 MHz channel formerly occupied by a single analog television signal. Cable modem equipment utilizing 256QAM modulation can achieve up to 38 Mb/s gross data rates in a single channel, with the usable data rate divided among customers in each group, where groups may evolve from several combined nodes to parts of a node, as demand increases. Standards such as CableLabs' PacketCable and DOCSIS1.1 facilitate voice-over-IP (VoIP) telephony in which voice traffic can efficiently share bandwidth with data services.

To the credit of the various standards bodies, digital video, data and telephony all use the same MPEG transport protocol, the same RF bandwidth and the same QAM modulation. Nothing precludes these services from being intermixed in the same channels. For example, the DOCSIS interface specification includes a "Downstream Transmission Convergence Sub-layer" which allows multiplexing of video and data MPEG packets. Header information identifies to receiving devices the content of each packet to facilitate multiplexed streams.

Cable operators have developed systems which have a potentially unbeatable combination of characteristics, if properly utilized:

- A theoretical downstream information capacity of several Gb/s
- A two-way wide-band communications path between headends and customers
- The ability to divide customers into small groups
- A common transport protocol for all digital services

REMAINING CHALLENGES

Unfortunately, cable systems currently fail to take full advantage of their potential. They contend with inefficient utilization of precious bandwidth and high capital cost for each service, which is often a barrier to introduction of new services.

The bandwidth and associated equipment, such as modulators and upconverters, for each digital service are typically allocated based on the number of subscribers in a sharing group, the expected service penetration, and the simultaneous usage rates among subscribers. These calculations are typically made for each service, with the assigned bandwidth based on predicted busiest-hour loadings. The bandwidth and resources may be highly utilized, or even insufficient, during periods of high demand, but are grossly under-utilized when demand is low.

Traditional approaches are inefficient on a number of grounds. Peak usage times for various video and non-video services do not occur simultaneously, and thus resource scarcity in one part of the spectrum is likely to coincide with overabundance on another. For example, watching movies is an activity that tends to be more popular on weekends, whereas accessing the Internet tends to be more popular during the work week.

Projecting usage levels for untried services involves, at best, an educated guess. When a popular movie becomes available on VOD, or a highly-anticipated download becomes available via Internet, the demand projection maybe rendered useless. Inexact demand projection leads to excess bandwidth being set aside for penetration increases or peak loading that may only be realized in the future, if ever, compounding the unequal spectrum utilization. More speculative or less popular services may not be offered at all because of the corresponding need to fully dedicate channels and assets.

Finally, these methods ignore synergies possible because video, data and voice all use the same transport protocol.

Figure 1 shows the degree of complexity and functional duplication that results.

Today's methodology means that a new service must somehow present a business case that justifies the use of some integer number, and thus at least one, of RF channels and also justifies the purchase and deployment of all the equipment required for signal processing from signal generation or acquisition through delivery of modulated carriers to combiners. The result is such a high barrier that many promising new revenue streams may never be tapped because of capital and bandwidth constraints and the uncertainty of demand projection. Thus, the "Wonderful new interactive digital service" of Figure 1 may never get launched, even though the cumulative capacity that goes unutilized by other services is more than sufficient to carry the signals associated with the new service.

Assigning dedicated spectrum to each service means that the available spectrum in even a 750 or 860 MHz cable system is quickly being allocated. Furthermore, analog broadcast television will be with us for many years to come while expected must-carry requirements for digital over-air broadcasters will require additional bandwidth, as will new demands coming out of retransmission consent negotiations. Innovations like high definition television (HDTV) may also tax spectrum allocations. Furthermore, many cable operators are constantly under pressure to add new basic service cable networks for both political and competitive reasons.

Operators often find themselves unable to meet external demands for new programming and unable to take advantage of new revenue opportunities, while significant portions of their spectrum are underutilized. Put another way, they are in the ironic position of having the highest per-subscriber bandwidth delivery system of any party in the broadband convergence competition, yet are held back by network sophistication limitations in trying to offer new services.

The authors wish to propose here a more integrated and efficient signal and bandwidth management for switched digital services by elevating the concept of routing in HFC networks.

THE OUTLINES OF A SOLUTION: THE BROADBAND MULTIMEDIA ROUTER

We suggest that dramatic improvements can be made in headend configuration and bandwidth management, in large part through the installation of a single new type of component: a Broadband Multimedia Router (BMR). The BMR is foundation to a service-independent headend management system for all digital services. Its function is open switching and routing of MPEG transport packets of all types, from any advanced digital service, to any node or other grouping of subscribers.

Instead of the headend incorporating a lot of redundant, but under-utilized, equipment, it becomes a switched and routed environment, simplified and streamlined to the configuration in Figure 2. All digital sources feed the multimedia MPEG transport router, while the router feeds a bank of modulators and upconverters which provide the modulated RF signals to the combiner for each node. The BMR would need certain DOCSIS capabilities to support cable modem deployment methodologies.

Any downstream 6 MHz RF channel can then be loaded with whatever mix of MPEG sessions is required to support services requested by customers in a given node, be they VoIP, Internet access, VOD or some other service. The intermixing of different services in shared data streams allows instantaneous spectral sharing and thus the maximum gain from statistical frequency division multiplexing (FDM).

It furthermore means that the mix of communications types may be different in different nodes. Thus, we also gain from statistical space division multiplexing (SDM).

The implications of incorporating the BMR into the headend are extensive, including:

- Elimination of the concept of fixed "channels" and frequency allocations dedicated to particular digital services and sources. The total information capacity that lies within the bounds of each channel can simultaneously carry different services, while packets from any given service can be transported over any channel with sufficient capacity. Dynamic frequency assignment is already provided for in the terminal equipment: video set-top boxes learn about the locations of programs via informational data packets, while cable modems are assigned to frequencies by the CMTS.

- Open and dynamic allocation and re-allocation of spectrum in each node on the basis of real-time demand by customers in that node. Any portion of spectrum can carry any digital service at any time.

- Enhancement of dynamic allocation through phasing out of proprietary equipment and fully embracing such standards as DOCSIS, OpenCable and PacketCable standards that have underlying MPEG transport compliance. This will also facilitate meeting the FCC's retail availability requirements.

Because the BMR is driven by routing tables, it supports new application of intelligence to how the cable plant is utilized. Different communications streams can be assigned appropriate priorities that allow each RF channel to be better used. For instance, high QoS priority sessions (e.g., voice or video) can be distributed over a sufficient number of channels to guarantee a lack of conflict, with the variable remaining space in each channel assigned to traffic whose timing constraints are not as tight.

Using the BMR as the heart of the headend digital management system allows both bandwidth and required equipment to scale with actual customer demand for total utilization of the cable plant, rather than as a function of offered services, each with a projected penetration. Any given service need only provide a standards-compliant transport stream to the BMR input to effect connection with every node in the system. The BMR will sort the packets (including duplication, if required) to feed the appropriate nodes. New transport equipment is only required in response to cumulative plant usage, and not in response to expanding the variety of offered services.

With all digital services now delivered in a switched/routed format, networks can be engineered to provide a total per-subscriber capacity, regardless of what services are offered, which of those services the customer chooses to access, and which services are simultaneously being accessed by others. New services can be added without requiring duplicate equipment and without allocating dedicated bandwidth. It is simply a case of connecting the service to the switching and routing fabric achieved by the BMR.

To the extent that the new services are functional alternates to existing services (for instance, an alternate source of video which is unlikely to be used simultaneously with an existing service) or which have different peak usage times, the total capacity requirements might not change at all and, if they do, it presumably means that customers are more heavily using the network and, therefore, that additional income streams have been tapped. Scaling now becomes a function of total network utilization and the need to guess at the penetration of any given service is greatly reduced.

THE BUSINESS ADVANTAGES OF THE BMR APPROACH

A network configured as described above “looks” unlike a conventional cable television system from a management standpoint. In effect, it changes character constantly: it may become a predominantly a VOD network on Friday nights, a predominantly Internet access network on Monday morning and, perhaps, a predominantly e-commerce shopping network during the Christmas buying period. There can even be node-to-node differences in the “look” of the network at the same time.

Battles over whether some new service is carried no longer happen. Only the signal acquisition or generation equipment is required to add it and, if the service is not popular, it can as easily be discontinued. No significant network engineering is required. Just as easily, a successful service can be scaled up to increase capacity by simply installing more servers or increasing the bandwidth of the connection to the service’s source.

The available spectrum is utilized as efficiently as possible at all times. Therefore, upgrades to increase bandwidth or to split nodes can be delayed, leading to major savings in rebuild capital.

While the BMR can delay node size reduction for some MSOs, others who intend to achieve small sizes, as is being attempted in mini-fiber-node trials, can also benefit from the introduction of switching and routing into the cable plant. Smaller node sizes increase the imperative to distribute service originations with some concentrated upstream and some residing closer to subscribers, according to issues such as service penetration and scalability. Under this scenario, it may be beneficial to distribute the functionality of the BMR between the headend and some lower network level. Increasing control over communications and supporting a variety of protocols, such as SONET, would enable the BMR to facility signal distribution in this heterogeneous environment.

At initial installation, the BMR can simply be added between existing MPEG sources and existing QAM modulators/upconverters. At this level, the significant efficiencies of shared data streams will be realized, while each application can keep its proprietary customer interface. Thus, no significant retraining of headend personnel will be required and the change will be transparent to customers.

Because transmission equipment is now shared across applications, the failure of any individual modulator or upconverter will have limited effect on overall performance. Automatic redundancy switching will simply transfer the load to alternative equipment. Thus, the overall reliability off all services will increase.

In summary, the BMR enables intelligent and dynamic allocation of bandwidth and resources among all digital services. This allows cable to leverage its core strengths to create the greatest effective per-customer information delivery capacity of any multimedia, multi-service provider. Cable, further has the ability to add new services and scale existing services in response to market demand, at the least possible incremental cost. A single new type of device, the BMR, elevates cable to the most sophisticated network for carriage of broadband multimedia convergence.

LONG-TERM BENEFITS AND OPPORTUNITIES

Beyond its immediate implications, the BMR can serve as a foundation for further dramatic innovations to be easily incorporated into cable plant.

Extension of the routing table concept can provide a software management level above automatic routing and resource management. Through this layer, operators can set policies for handling various situations. For instance:

- During times of high capacity demand, the operator may wish to determine a priority by which certain less profitable sessions may be denied, delayed or slowed down.
- Selected video streams (for instance an old movie, but not a live sporting event or high-priced advertisement) may be more highly compressed, with a determined quality reduction.
- Certain customers or customer types may be given access or speed priority.
- During low demand periods, non-real-time sessions (e.g., streaming video to customer premises local caches such as digital video recorders) may be opportunistically initiated.

By monitoring customer response to various network loading situations, these policies can be reviewed and changed as required.

Since not all nodes will require the same total information capacity, nor will the peak demand occur at the same time in every node, a future extension of the proposed scheme could allow the number of modulators and upconverters assigned to a node to vary, based on the total communications needs at any one time. For example, a more commercial part of the system may require a greater total data rate to support small office cable modem service during business hours, but a residential area may more heavily utilize the system for Web browsing and television viewing during the evening hours. By dynamically assigning hardware, as well as spectrum, an operator will need not only less spectrum, but fewer expensive signal processing elements (or looked at in the inverse, provide more services using the same amount of equipment).

New merged services may be enabled by channels carrying a multiplex of several types of communications. For instance:

- Customers may be able to jump between a streamed broadcast video and a related Web site and to optionally initiate a two-way voice conversation with a manufacturer's customer service representative. The "live" program may continue in a window on the screen or even be spooled, so that the customer picks up where he left off when he returns from the Web site. The BMR can smoothly effect the transition between video and data streams and back.

- Ads in broadcast programming may be tailored to the demographics of individual groups of customers, or even to the characteristics of individual subscribers. E-commerce purchasing opportunities may be seamlessly available from any ad. The BMR can simultaneously insert different ads in digital programming being delivered to various nodes.

- Play-along quiz shows that combine interactivity through the Internet with broadcast video may become common. Several precursors to this have emerged, such as the Disney and NFL conducting trials of a parallel Internet game with Monday Night Football in 1999. The BMR can assure coordination in the delivery of streams of different media types to the same subscriber to assure synchronization.

The obituary for analog has already been written by the FCC. While it is uncertain how much longer the patient will live, it is clear that NTSC video will eventually be discontinued. When that happens, operators may well choose to carry all programming on an effectively on-demand digital basis (this can be completely transparent to users, but will result in significantly more effective data capacity, as programs will only be delivered to any given node if somebody served from that node desires to view it). At that time, the full spectrum, capable of delivering multiple Gb/s, can be fully available to deliver services to each grouping of subscribers.

The initial vision for the BMR leverages the ubiquity of MPEG transport protcols. However, with constancy of change a certainty in cable networks, it is important to realize that current standards and protocols may, at some point, be combined with others or displaced. By introducing increased intelligence and communications control, the BMR can extend its benefits to any changes effected in the network. For example, if IP transport is selected by some MSOs for certain services, the BMR would be able to ease this introduction.

If and when desired, services can be offered which are not tied to traditional 6 MHz channelization. This means that the peak data rates available to certain heavily immersive, and as yet unimagined, services or simply higher access speeds for rapid downloads and communications could increase significantly by use of wider RF channels. For instance, a 100 Mb/s data stream would require about 15 MHz of bandwidth.

The BMR would dramatically improve the positioning of MSOs to provide services in the age of broadband multimedia convergence, with both significant near-term implications due to the device's fundamental concept, and far-reaching eventual implications as a foundation for future innovations.

SUMMARY

In summary, cable has moved from its all-analog broadcast roots through the construction of HFC, two-way networks and through the offering of digital services, many of which involve delivering information streams to individual customers.

Allocating a fixed bandwidth and dedicated signal processing equipment to each service based on anticipated peak demand, however, is inefficient of both capital and signal processing equipment.

The authors propose that operators take advantage of the common digital services transport protocol by use of a broadband multi-media MPEG router which will allow dynamic allocation of both spectrum and equipment across many services and nodes. The use of a common headend platform for digital services provides greater spectral efficiency and also allows services to be added and scaled without re-engineering the system. The BMR enables operators to avoid distribution system upgrades that would otherwise be required to increase RF bandwidth or split nodes.

END NOTES

[1] "SBC Orders from High-Fiber Menu," Multichannel News, October 25, 1999.

[1] "Cisco, Broadcom Wireless Gear Boost MCI, Sprint MMDS Plans," Multichannel News, November, 1999.

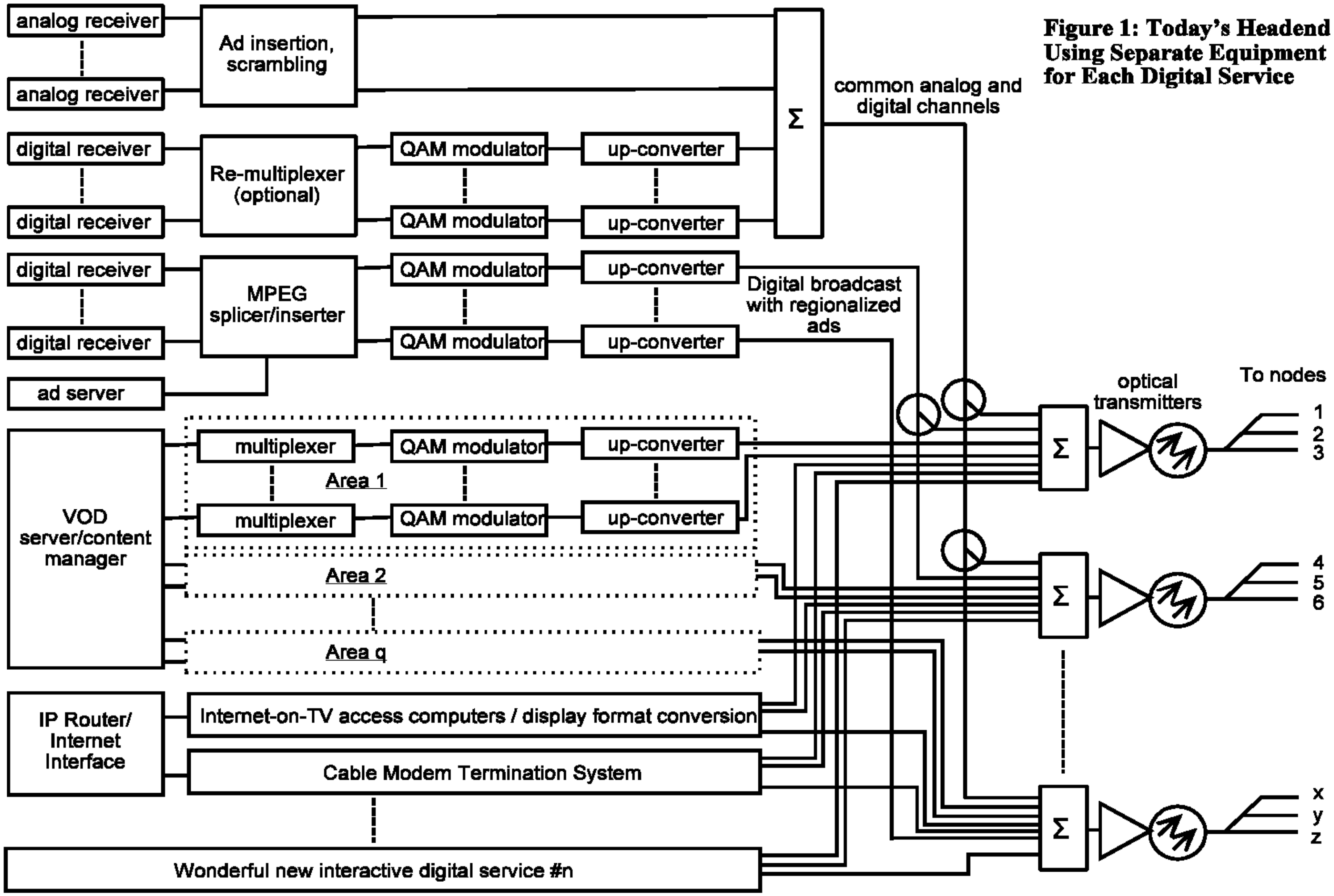


Figure 1: Today's Headend Using Separate Equipment for Each Digital Service

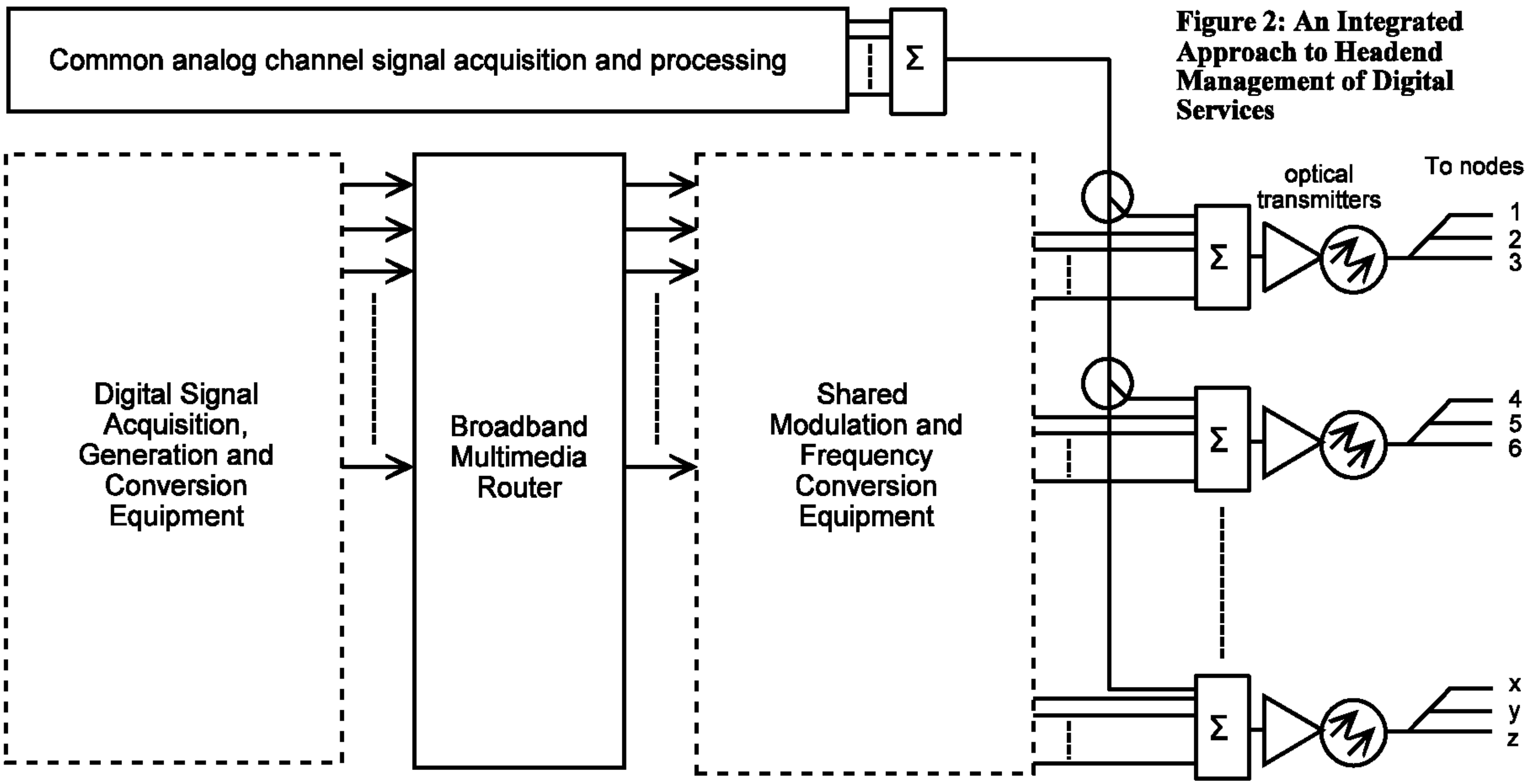


Figure 2: An Integrated Approach to Headend Management of Digital Services

Video On Demand – More Than Movies

Yvette M. Gordon
SeaChange International

Abstract

As real-time 2-way networks are becoming common place, the industry has begun to shift its focus to the advanced services that can now be offered in our digital network environments. Applications such as Internet-on-TV and Video on Demand are in initial deployment and trial phases worldwide. Video on Demand, however, tends to be associated mostly with "Movies on Demand" – allowing subscribers to watch a movie whenever they wish and having VCR-like capabilities. There are, however, many different applications of Video on Demand, each bringing unique revenue opportunities and technical challenges.

As we begin to move past the initial deployments and become comfortable with basic Video on Demand services, such as Movies on Demand, we can begin to explore other applications that utilize on demand video streaming, and thus leveraging our investments. This paper will review various Video on Demand applications, their technical infrastructure, business model overviews, and technical challenges ahead.

WHAT IS VIDEO ON DEMAND

As costs have decreased and digital platforms have been deployed, Video on Demand has become a topic of renewed interest within the cable industry. It is surprising, however, that as most MSOs are preparing for initial VOD deployments or trials, the business of what Video on Demand will be used for remains somewhat undefined. There are many different businesses that can utilize on demand video streams including Pay-Per-View movies on demand, Subscription-based movies on demand, after-broadcast programming on demand, interactive advertising, news on demand, walled garden streaming, and much more. As MSOs are evaluating these potential applications, several key questions are being asked –

- What is the business model (cost payback, revenue modeling, cash flow, etc.) for these services?
- What is the technical infrastructure?
- How deployment-ready is the application and what are the remaining technical hurdles?

Certainly, maintaining a competitive edge is a major factor in launching advanced interactive services, but the above issues are also serious considerations in selecting one service over another. Let us first look at the infrastructure of a Video on Demand network and then evaluate what services there are in addition to movies.

THE VIDEO ON DEMAND INFRASTRUCTURE

A typical Video on Demand system is broken into seven major categories, as depicted in Figure 1. Depending on the digital network, these components communicate differently. Some of these components were built with broadcast television in mind, and others are new to the digital infrastructure. We will attempt to address the general Video on

Demand issues for each of these components.

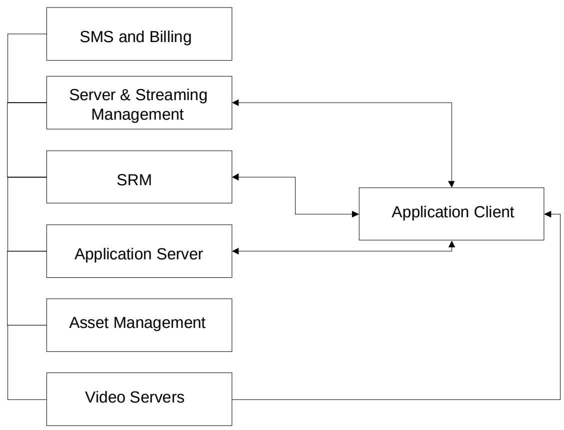


Figure 1. Typical VOD Components

<u>SMS and Billing</u>

Subscriber Management Systems (SMS) and Billing interfaces have traditionally been built for broadcast systems and are also closely linked with encryption processes. Typically, a pre-scheduled program is entered into the system and the encryption and views (billable events) reference this scheduled event.

With the addition of Video on Demand, billing systems need to support flexible schedules – i.e. having a program play at any time as opposed to a scheduled time. Encryption likewise needs to be supported on a flexible basis (either the content needs to be pre-encrypted, or the content needs to be encrypted on an arbitrary schedule as it is played out of video servers). Since billing systems are also the Subscriber Management Systems, they need to support real time management of subscriber information and non-movie purchases.

Many of these VOD requirements are new to billing systems and have been worked around by making VOD events look like Pay-Per-View (PPV) events and by having separate SMS platforms. With the addition of non-movie services, however, the need for advanced billing and subscriber management systems is increased.

<u>Server and Streaming Management</u>

The management of VOD servers and the constant establishment and teardown of interactive streams is typically done in software developed by VOD server vendors. The key to the success of these components is their scalability and fault tolerance, which differs from one vendor to another.

<u>System Resource Manager (SRM)</u>

A SRM manages the topology of the network and RF spectrum as it relates to VOD streaming resources. In some networks the SRM is a part of the standard digital platform; in others it is the responsibility of the VOD server vendor to provide a SRM. Having a successful SRM depends on the ability to support key VOD requirements such as narrowcasting.

<u>Application Server</u>

The application server is the headend or hub based component that is the counterpart of the set-top based Application Client. There can be many application servers in a network, one for each application such as Movies on Demand, Sports on Demand, News on Demand, etc. In a truly open network, each application server can even be developed by different software development companies, allowing various VOD applications to become a new business opportunity. The technical hurdles and status of various applications will be discussed later in this paper.

<u>Asset Manager</u>

The Asset Manager is a critical VOD

component. It manages the content, content groups, and content information for each asset (movie, artwork, fonts, etc.) on the system. The function and capabilities of the Asset Manager are key to the operational ease of running a Video on Demand system. This component is typically a key software platform developed by video server vendors.

Video Servers

The video servers are obviously the heart of any Video on Demand system. Much development spending in the last 5 years has resulted in improvements in efficiency and cost reduction. Video servers that used to cost thousands of dollars per stream are now hundreds of dollars per stream and a fraction of the size. Some of the critical factors in selecting a good video server are its scalability, fault tolerance, and streaming efficiency.

Video server providers have also taken various approaches towards on demand conditional access and encryption, from building in QAMs and pre-encrypting content to supporting unique session-based encryption in external components, and some without any encryption at all.

Application Clients

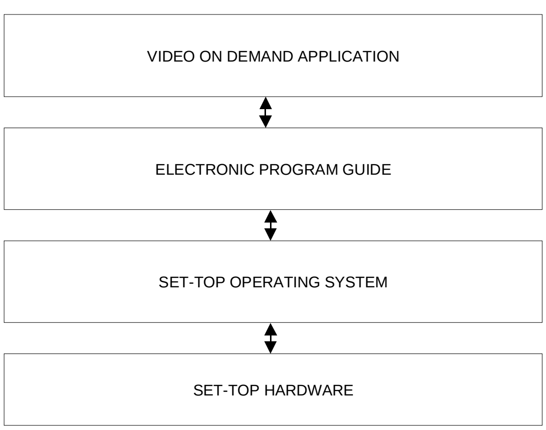


Figure 2. VOD Set-top Layers Today

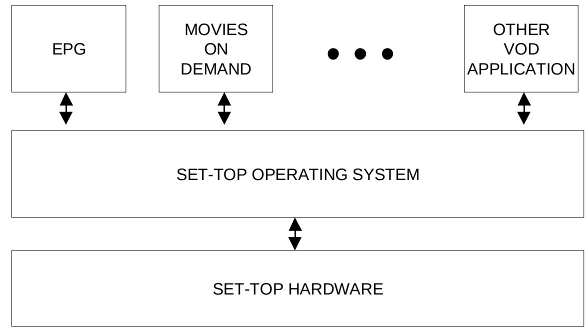


Figure 3. VOD Set-top Layers In Progress

In order to deploy multiple applications, the set-top box has to be able to support switching between these applications. In a traditional digital broadcast environment, the Electronic Program Guide controls all set-top activity. These were built without “application sharing” in mind. In addition, there is no industry standard for application interfaces. These two points mean that each application developer today must build the application specific to the set-top box and with special integration efforts with the EPG vendor, as depicted in Figure 2.

It has become obvious that moving between various interactive applications, and offering a variety of new services, is a requirement in todays networks. There are two methods in progress to ease application sharing. In the first, shown in Figure 3, the operating system, such as PowerTV or WindowsCE, manages the various applications. Since it is likely that some applications will require sharing information (such as parental control), the Operating System must also enable such application data exchange.

The second option, for ease of application sharing that is in progress, is adding a middle-ware layer on top of the Operating System, as depicted in Figure 4. This allows the concept of application porting from one set-top platform to another in that the operating system is masked to the application software. The EPG could also perform such a middle-ware function. In many cases, this middle-ware is also being designed to support common functionality, such as HTML support and Video on Demand streaming session setup and control. This allows applications to use common functions in a consistent manner and enables new applications to deploy easier.

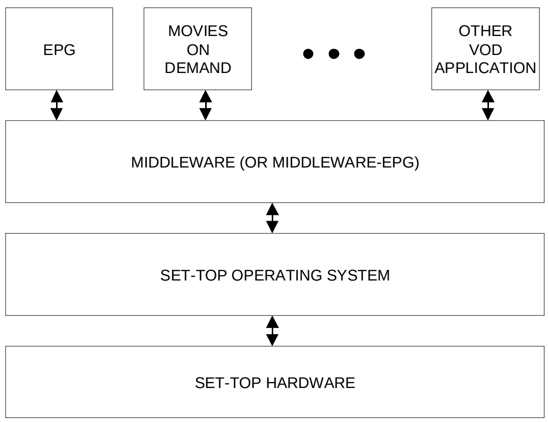


Figure 4. Set-top Layers with Middle-ware

There is much work to be done to enable set-top application development to become as common place as PC software development is, but the focus on application sharing and common middle-ware is an important step in that direction. Having common set-top software standards is also a key part of enabling open set-top hardware.

Component Summary

Of course there are several other parts to a VOD system such as content acquisition and general distribution that are not discussed in this paper. To summarize site-based technical components, current VOD deployments that support only Movies on Demand, have worked out all of the requirements to properly fit into a typical digital network. These platforms are readily deployable today. Much of the key functionality has been developed, either by the digital network platform, or the video server vendor.

In the future, however, as other on demand services become commonplace, subscriber management and billing interfaces will need to become more flexible in order to avoid requiring duplicate databases of subscriber information. In addition, the ability to switch seamlessly and efficiently between applications is the key towards allowing new applications (and businesses) to be deployed on our networks. Platforms that will enable this application switching in a cost-effective manner will become critical in the interactive era.

VIDEO ON DEMAND APPLICATIONS AND SERVICES

Now let us look at various applications that fall into the “Video” on Demand category. As we will see, each of these applications has its own business model and technical accomplishments and pending hurdles.

Movies on Demand

Movies on Demand – having a list of movies to select from and being able to view it whenever one wishes. Once rented, the movie is typically available for viewing within 24 hours and can be put “in progress”. The comparison to traditional PPV is the ability to have a dedicated movie copy with VCR-like controls.

Movies on Demand (MOD) is clearly the most common application of Video on Demand. Many times the term VOD is used to describe what really is MOD. There has been much analysis of the Movies on Demand business model. Taking some general assumptions, a high-level business model is summarized below.

Total Homes Passed:	200,000	
Total Subscribers:	140,000	(70%)
Digital Subscribers:	56,000	(40%)

Buys Per Month	3
Weighted Avg Movie Cost:	$4.25

(Averages adult and non-adult pricing)

Per Subscriber Revenue:	$ 12.75
Monthly Revenue:	$ 714,000
Cost Per Subscriber:	$ 80.00

As we can see, there is a significant revenue potential with Movies on Demand. The service easily reaches all digital subscribers, and adds a cost of $80 per subscriber (a relatively small amount in comparison to the recent investment of hundreds of dollars per digital subscriber for the set-top box), and can bring in approximately $12.75 per subscriber per month.

Movies on Demand Accomplishments:

- The building block for VOD as a whole has been established through Movies on Demand deployment planning.
- Integration into various digital networks has taken place.
- Movies on Demand is deployed today.
- Substantial business models are being developed from early deployments.

Movies on Demand Hurdles:

- Encryption techniques and costs must still be worked out.
- Flexible, two-way billing and subscriber management must still be enhanced and developed.
- Content acquisition and distribution details are still in progress, including the determination of what the cable MOD licensing window will be.
- Application switching and integration into set-tops and EPGs must continue to be enhanced.

Subscription-Based Movies (SVOD)

SVOD – The ability to watch certain movies from a package at any time, or to join on a movie that is in progress from the beginning. Currently all movie packages are scheduled events, SVOD will allow subscribers to have on demand access to certain movies within the package for a fixed monthly fee.

Since SVOD is a new service, there is no proven business model. Taking the same system-wide subscriber numbers that were used in the Movies on Demand example, and assuming a 40% digital take rate of a SVOD service, a high level revenue model is listed below.

Total Homes Passed:	200,000	
Total Subscribers:	140,000	(70%)
Digital Subscribers:	56,000	(40%)
SVOD Subscribers:	22,400	(40%)

Per Subscriber Revenue:	$ 15.00
Monthly Revenue:	$ 336,000

SVOD Accomplishments:

- SVOD can easily piggyback on a MOD launch. It uses the same streaming infrastructure. Therefore MOD sites are likely candidates for SVOD services.
- The revenue per subscriber is fixed and therefore not subject to changing viewing trends.
- Billing is easy for SVOD in that it is a subscription package, as we have today.
- Studio rights have been secured by some providers.

SVOD Hurdles:

- SVOD is not yet deployed and take rates are not yet well defined.

Bandwidth implications for a pay-once, view-often service are not yet understood and may require more video streams and RF bandwidth to support the viewing tendencies.

Interactive Advertising (IAD)

Interactive Advertising has been used for years for PC based applications. Since local and national broadcast advertising revenue is a major part of programming today, the use of advertising in on demand and interactive television services are of obvious interest.

There are many approaches to interactive and targeted advertising, as well as how these interact with broadcast programming. Many of these approaches remain vendor or service specific today, but the bottom line is that the consumer impact, implementation, and revenue modeling for advertising in an interactive environment is not yet understood. It may be that each advertiser continues to pay for broadcast 'spots' while paying a per-use fee for interactive spots. Targeted spots may command a higher fee. Much of the success of advertising will depend on the interaction and delivery capabilities, which are being agressively developed by companies like SeaChange International.

Since no firm revenue model yet exists, we will instead look at some of the parameters that will be used in building such a model:

1. Broadcast advertisement cost per 1000 views (used per number of cable subscribers)
2. Target advertisement cost
3. Number of potential interactive views (used per number of digital cable subscribers)
4. Number of actual interactive advertisement users per ad
5. Advertiser payment per minute of interactive advertisement use
6. Number of participating advertisers and programmers

There are many other factors that can be included in this model; however, most of the additional inputs are unique to various solutions that are being developed. Although the business models are still being tested as solutions are enhanced and built, it is clear that interactive advertising will be a part of digital networks, both through standard interactive applications, and on demand video applications.

Interactive Advertising Accomplishments:

- Basic Interactive Advertising can be supported through existing VOD infrastructures, using a unique IAD application service.
- Helps justify cost of VOD system for additional revenue in the successful local advertising space.

Interactive Advertising Hurdles:

- IAD Presentation, traffic and billing parameters, and set-top based handoff between interactive, target, and broadcast advertisements are still being worked.
- Interactive IAD client/servers need to be developed, along with supporting traffic and billing systems (that support all of the new parameters for traffic management on VOD servers and advertiser payments).
- Programming rights issues need to be addressed.

After-Broadcast Programming On-Demand

After-broadcast programming (ABP) uses the VOD servers to also support real time recording of broadcast programs. These programs would then be available for viewing on demand after their scheduled broadcast time. This service is sometimes confused with a Time Delay product, which only postpones the broadcast play-out in a broadcast mode.

There have been several attempts at broadcast replay of programming. Most recently, Tivo and Replay are also offering personal video recorders (PVR) that offer similar services to subscribers that pre-plan their broadcast program recording. There is likely to be a market for after broadcast programming both in the subscriber set-top or television directly as well as from a video server, or in some combination of the two.

Since a ABP video server based service does not yet exist, we will make some assumptions for our revenue model. Since eventually, subscribers will have both PVR and video server based ABP access, we will assume that an average household will use the server based after broadcast program access once weekly. We will also assume a $1.99 price per replay.

Total Homes Passed:	200,000	
Total Subscribers:	140,000	(70%)
Digital Subscribers:	56,000	(40%)

Buys Per Month:	4
Average Cost Per Program:	$ 1.99

Monthly Revenue: $ 445,760

ABP Accomplishments:

- Enables a new method for watching TV – truly what you want when you want it as it relates to available programming
- Enhances the PVR market to offer subscribers the ability to plan program recordings as well as watch them for a per program fee if recording was not planned.
- Billing interface is similar to Movies on Demand billing.

ABP Hurdles:

- Server based ABP buy rates are still relatively unknown.
- Network integration work, including MPEG compatibility of digital signals from various satellite and encoder sources is a significant issue. Some MPEG sources do not include critical data like I-frames for performing fast forward and rewind functions.
- Real-time encoding and MPEG import capabilities are in their infancy in servers. Mostly, asset management and scheduling integration efforts are required.
- Licensing rights for both programming and advertisements are undefined and undeveloped for server based ABP solutions.
- Bandwidth impacts of this service are not modeled.
- Storing all possible content on video servers will take a significant amount of automation and disk storage. The trade off between storage & management costs and revenue potential will need to be weighed.

News or Weather on Demand

In addition to the functionality of the ABP service, there is also opportunity for pay-per-view or subscription based services for special interest programming. It is likely that we will see Video on Demand (or non-video based interactive applications) services for target audiences, such as News on Demand or Weather on Demand. With such a service, a subscriber could watch previously aired news services on demand and potentially tailor their news delivery order towards their interests. Likewise, for a Weather on Demand service, a subscriber could search for weather conditions and receive updates and recent videos at the click of a remote.

Modeling for such a service is somewhat challenging. Since these services could be offered in many ways, we will assume a fixed subscription fee with a 20% take rate.

Total Homes Passed:	200,000	
Total Subscribers:	140,000	(70%)
Digital Subscribers:	56,000	(40%)
Svc. Subscribers:	11,200	(20%)

Per Subscriber Revenue:	$ 5.00
Monthly Revenue:	$ 56,000

NOD, WOD Accomplishments:

- As with SVOD, subscription based services are easy to implement for billing purposes.
- The revenue per subscriber is fixed and therefore not subject to changing viewing trends.
- Television-based weather updates is a highly desired service. Providing this convenience with interactivity can be a very compelling application.

NOD, WOD Hurdles:

- Business model is fairly unproven for TV based services, although much research has been done (and is likely to continue).
- Content updates can be an issue, especially for a broadcast video service like News on Demand. Management of the constantly changing content can be impacting.
- The revenue for these types of services may not warrant heavy video (bandwidth) use. A service such as Weather on Demand may be very successful as a graphics-only application that requires little bandwidth impact and can use existing internet site caching for content updates.

E-Commerce, Internet Applications, Food Purchase Applications

These applications are commonly non-video based and can easily utilize the same underlying delivery technology. Once an internet-on-TV application exists, it becomes fairly easy to launch HTML or XML applications (depending on the browser support) that are managed locally.

Many Internet-on-Cable applications are also targeting the set-top layering and application interoperability problem described in the first section of this paper. This makes the thought of an internet browser middle-ware product appealing to operators. Issues remain, however, as to the real standardization of set-top interfaces to avoid the potential of a proprietary solution dominating the middle-ware marketplace.

Total Homes Passed:	200,000	
Total Subscribers:	140,000	(70%)
Digital Subscribers:	56,000	(40%)
IOC Subscribers:	16,800	(30%)

Per Subscriber Revenue:	$ 20.00
Monthly Revenue:	$ 336,000

The business model for Internet-on-Cable applications is complex. As with Video on Demand, a middle-ware product enables many applications, and therefore one is justified to spread the cost across many services. However, as with Movies on Demand, one application must exist to kick start the deployment and have a viable standalone business model.

A monthly internet access fee of approximately $20 per subscriber is likely for Internet-on-Cable. Other add-on applications are likely charged externally. For example, a Pizza store would pay on-going fees to the operator to have an on-line, marketed application. Fullfillment is done outside of the cable billing system in that the payment is made upon delivery, directly to the store employee. Likewise, fullfillment for most e-commerce applications is handled directly with the vendor in question. This leaves a reliable, constant revenue stream with the operator without any subscriber billing requirements.

Internet-on-Cable Accomplishments and Benefits:

- Many times, Internet middle-ware platforms also add application interoperability software, addressing one of the major issues within set-top box software.
- Internet access easily enables electronic commerce applications.
- Internet middle-ware easily enables local applications such as Pizza buying or local news bulletins.
- Interactive Offering (also called walled garden) applications can use an existing Video on Demand system to do local on demand streaming integrated with HTML applications.
- Having cable modem and Internet-on-Cable back-end interoperability would be very positive.

Internet-on-Cable Hurdles:

- Cost vs. revenue factors are still being determined.
- Affect on the infrastructure is being tested and is not yet well understood.
- It is not known if EPG vendors will be willing to be a 'subset' application as opposed to the middle-ware itself (this would affect non-EPG vendors only).
- Video on Demand interoperability work is still in progress.
- Having internet or walled garden sites specifically tailored towards television viewing remains a general issue.
- Subscriber-specific applications require billing workarounds since most billing systems are designed for Pay Per View billing only.

IN SUMMARY

Revenue Summary

First, let's look at the various revenue models, both on a revenue-per-subscriber and on a revenue-per-system basis. It is easy to note that the key factors are the anticipated take rates and digital penetration numbers. *The purpose of this paper is not to anticipate these, but instead to point out the revenue possibilities and to offer some suggestions as to what factors go into the various calculations.* These hypothetical numbers and listings of accomplishments and outstanding issues will hopefully enable discussions of what is ahead and where we are.

Service	Per-Subscriber Monthly Revenue Est.	Per-System Monthly Revenue Est.
MOD	$ 12.75	$ 714,000
SVOD	$ 15.00	$ 336,000
ABP	$ 7.96	$ 445,760
WOD	$ 5.00	$ 56,000
IOC	$ 20.00	$ 336,000

Figure 5. Potential Revenue Estimates for Interactive and On-Demand Services

Another issue that arises in looking at these new revenue potentials, is what the maximum actually is of what a typical subscriber will buy. This all the more justifies applications like interactive and target advertising, in that the revenue comes from non-subscriber sources.

Technical Summary

We have certainly come a long way in the field of Video on Demand. The technology has gone from the perception of having only expensive trials to being one of todays most viable new businesses. This has been accomplished through the deployment of standard digital infrastructures, and the completion of Video on Demand platforms into these infrastructures. The primary VOD components (video servers, SRM, server management, streaming management, asset management) are complete, low component costs justify the VOD business, and early deployments are under way.

There is no doubt that Movies on Demand business has been the jump-start for Video on Demand. Enhancements and technical hurdles remain mainly in the areas of advanced billing/subscriber management, encryption techniques, and application interoperability. At the same time, we continue to see improvements to component performance, density, fault tolerance, and cost. Various issues exits for new VOD applications, including the impacts on back-end services, bandwidth, and storage. Such services are likely to begin in trial phases to quantify these impacts.

As we begin to see more wide scaled deployments of the basic Video on Demand infrastructure, new services will certainly emerge as add-on applications since their costs will be lessened by the initial investment and as we continue to strive to maximize our digital network investments.

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