THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:

NCTA
Technical
Papers

A COMPREHENSIVE MANAGEMENT APPROACH FOR HFC NETWORKS

Esteban E. Sandino
AT&T Broadband and Internet Services

Abstract

Deployment of interactive data services over HFC networks demands a comprehensive approach to network management. Advanced customer terminals, such as standards-based cable modems and intelligent set top boxes, are increasingly talked about as the ideal vehicle for the collection and storage of information regarding the state of the physical HFC network. The support for standard management protocols in these new advanced terminals greatly increases the appeal of this approach: It makes the process of collecting and analyzing network data platform and vendor-independent, and also facilitates the task of integrating multiple platforms and systems.

HFC network management is thus evolving away from the more traditional approach based on the deployment of proprietary systems and protocols to poll and collect status information from active network elements. However, this does not imply the demise of this type of monitoring system. Advanced customer terminals provide operators with indicators regarding the health of the HFC network but they do not provide data on the operational status of active network elements such as power supplies, fiber nodes, and amplifiers. Thus, the data collected using advanced terminals complements rather than replaces the monitoring of active HFC devices.

This paper discusses a management framework for HFC networks predicated on the deployment of standards-based platforms. In such a framework, there is still a strong requirement for systems that monitor active HFC network elements. It is discussed how data from these systems can assist in the more detailed diagnostics and troubleshooting process following the detection of threshold alarms and early warnings generated by advanced terminal devices. To achieve this goal it is imperative that management platforms intended to monitor active HFC elements move away from proprietary protocols and support the same standard protocols as emerging management platforms for advanced terminal devices. The ultimate objective is to have open-standards based systems to facilitate not only integration with other management platforms but also to allow components from multiple vendors to operate together over the same HFC network. This paper concludes with an update on the progress of standards-development activities currently taking place within the cable industry to help achieve this objective.

INTRODUCTION

The impending deployment of advanced cable modem systems based on the DOCSIS standard as well as future deployment of OpenCable™ set top boxes has already begun to change the way HFC networks are managed.

These new advanced terminals will support the Simple Network Management Protocol (SNMP) and standard SNMP Management Information Base (MIB) definitions containing a number of variables and counters that will enable collection of system and HFC network information at various levels. These standard MIBs will support collection of data such as customer terminal RF input and output level variation, RF channel frequency response, number of transmitted and received data packets, packet transmission errors and collisions, and packet retransmission rates.

HFC network operators already realize the value derived out of SNMP MIB data analysis and the powerful insights that are gained into the performance of their physical networks and systems. A particularly powerful tool is the historical trend analysis made possible by the periodic collection of SNMP MIB counter information. Such analysis enables the continuous tracking of HFC network changes to identify non-critical chronic network problems, and to trigger preventative network maintenance activities. Early detection of service-affecting problems is also possible since periodic measurements help the operator determine when specific impairment levels exceed pre-defined thresholds.

Focus is shifting steadily out of proprietary element management systems and into standards-based systems and management platforms. In parallel with this process, there is also a migration towards adoption of a Telecommunications Network Management (TMN) or similar model for the planning, deployment and management of the evolving HFC telecommunications network. The five constituent layers in the TMN model are illustrated in Figure 1. Adoption of such models is expected to facilitate the collection, analysis and correlation of status information from multiple subsystems that are not necessarily supplied by a single vendor. These functions are performed within the three bottom layers in the TMN model: the Network Element Layer (NEL), the Element Management Layer (EML), and the Network Management Layer (NML). The available status information is then used to support higher-level network functions such as customer service provisioning, usage reporting and billing, and service expansion planning, functions that reside within the top two TMN layers: Service Management Layer (SML) and Business Management Layer (BML).

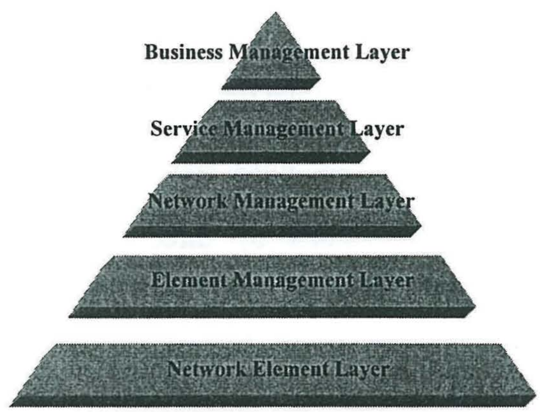


Figure 1. The TMN Model

To support these new network management models, operators must make significant investments in standard management platforms and systems. Deployment of advanced customer terminals and other network elements that support standard management protocols help justify these investments: the ability to address network elements from multiple vendors and correlate data from multiple sources is too big of an advantage to ignore.

Deployment of more traditional platforms based on proprietary protocols to poll and retrieve status information from active network elements such as fiber nodes, power supplies and amplifiers is sometimes perceived as providing only marginal benefits. Moreover, the proprietary nature of these systems makes them difficult to integrate into a standard platform without a significant development effort, which makes them a somewhat less attractive management tool. This notion is further reinforced by the belief that advanced terminals can provide all the necessary data for effective monitoring and troubleshooting of network and service problems.

Although it is true that advanced set top boxes and next-generation cable modems can act as repositories of HFC network health indicators and can provide early warning of system-wide problems, additional data is still required to assist the management systems in determining the most likely cause of those problems. This additional data can only be obtained by directly polling critical active transmission elements in the physical network. Therefore there is still significant value in providing selected active elements with the necessary intelligence to allow management platforms to retrieve data on their operational status. This supplementary data can then be correlated with data retrieved out of customer terminal devices and used to pinpoint the most likely cause of a system-wide failure or a degradation in the quality of the delivered services.

This paper describes a proposed approach for integration of various element management systems found in a typical HFC network. This approach is based on a TMN network management model. The end objective is to facilitate the gathering and correlation of status monitoring information from multiple platforms, and the effective utilization of this data to provision, deploy, monitor, and manage HFC networks. The benefits derived from cross vendor platform integration in the areas of monitoring, diagnosing and troubleshooting are discussed. The reasons for demanding compliance with open standards as a condition to allow integration of new platforms and systems into this proposed HFC network management platform are discussed. The impact this has on poll-based systems for monitoring of active network transmission elements is discussed. Finally, a brief update is presented on the current efforts to develop physical and MAC layer standards to facilitate integration of these systems with other standard-based element managers deployed in HFC networks.

HFC NETWORK MANAGEMENT PLATFORM

Figure 2 is a layered representation of a typical HFC service network and its various constituent elements. It shows a subset of the various services that are supported and the various elements involved in the provisioning, delivery and management of those services. Figure 2 attempts to depict the various network layers within the context of the TMN model. Following is a more detailed description of each of those layers and how they relate to that model.

HFC Physical Layers – TMN Network Element Layer (NEL)

These are typically the two-way RF broadband communications channels between an HFC network primary hub or headend and the end customer. In general, multiple HFC segments consisting of a set of downstream and upstream RF channels are allocated to support each of the distinct services supported by the network. Figure 2 illustrates the elements involved in supporting a subset of those services: standard broadcast as well as interactive video services, internet access services, and telephony services to name just a few.

Also included in this layer are the physical LAN and WAN that provide connectivity between various service support systems within a primary hub or headend, and the higher-level connectivity between individual headends and the central HFC management location.

HFC MAC Layers

Multiple MAC layers are supported within the HFC physical layers to implement communications protocols between the various service controllers at the primary hub or headend and the various terminal devices in the customer premises and any other active elements deployed within the various physical segments. In Figure 2, the various MAC layers arbitrate element access to the allocated HFC channel bandwidth.

The HFC physical layer is where network elements operate to handle the transmission of telecommunications data and to provide end customers access to the various network services. Collection of status monitoring information and data regarding the health of the physical network is collected within this layer. These two functions also reside within the Network Element Layer in the TMN model.

HFC Primary Hubs and Headends – TMN Element Management Layer (EML)

The primary hubs and headends are where all service processing and gateway equipment resides. This equipment performs a number of functions including: implementation of appropriate MAC and access protocols to arbitrate access to the physical HFC network, support for initial service provisioning and activation, and collection of network statistics and other health indicators for the physical plant. The latter function may involve polling of active network elements or periodic collection of stored data out of customer terminals. In general, all alarm information concerning specific services and elements is collected at this layer where some data pre-processing or data reduction may take place as well. These are the same functions that reside within the Element Management Layer in the TMN model.

Some of the service processing equipment and independent element managers within the element management layer includes the following:

Telephony Service Controllers

These are the devices used to provision and manage telephony customer network interface units (NIU). These devices may also collect data on service performance metrics such as bit error rates, dropped calls, etc. They may also provide indicators on the state of the physical RF channels allocated from the headend to the NIU.

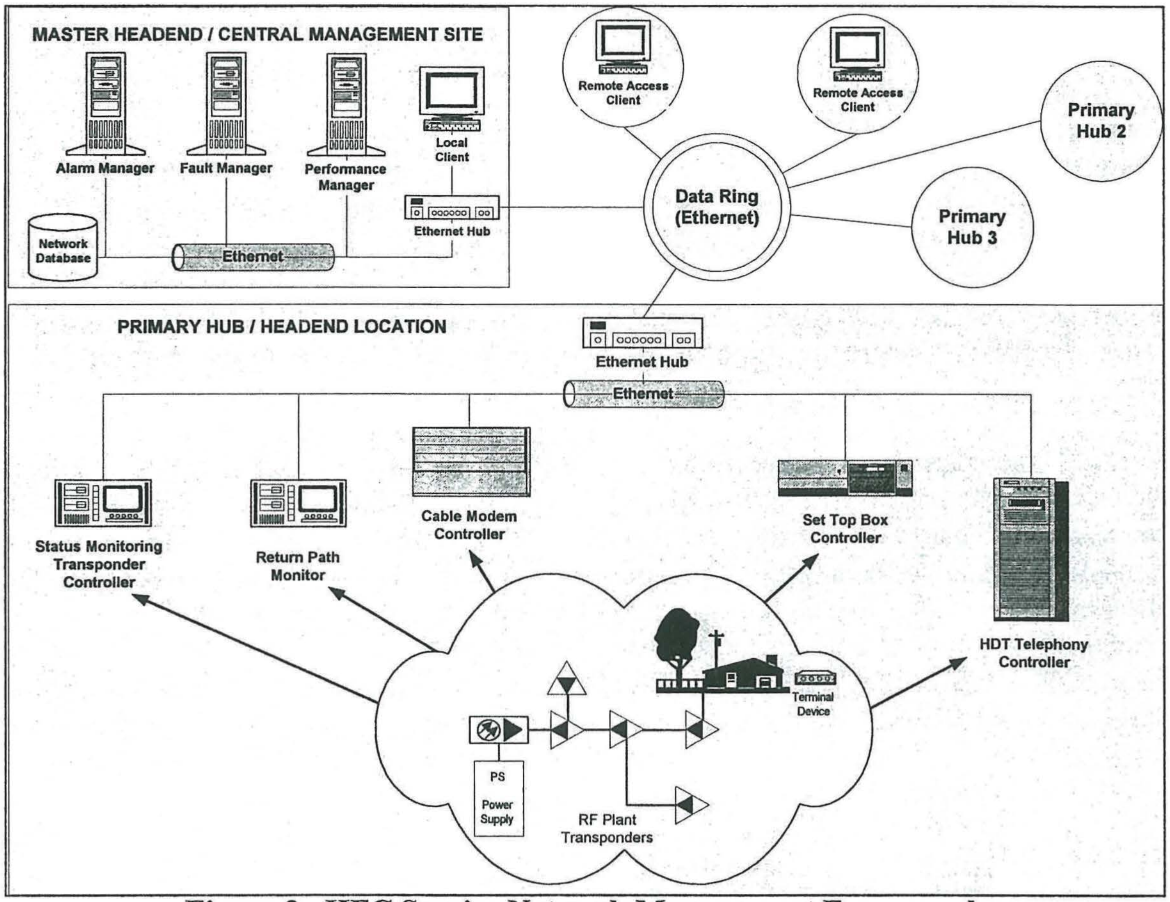


Figure 2. HFC Service Network Management Framework

Set Top Box Controllers

These devices support the provisioning and management of traditional and advanced video services through standard video set top boxes. Traditionally, set top boxes have not supported enough intelligence to support monitoring of the HFC physical transport network. However, this began to change with the deployment of advanced analog and digital set top boxes. These new terminals can provide operators with information related not only to the usage of specific services, but also related to the health of the overall transport and delivery network. The advent of OpenCable™ set top boxes will increase these capabilities while at the same time supporting standard management protocols such as SNMP.

Cable Modem Controllers

These devices manage communications with cable modems in customer premises and support the provisioning and delivery of Internet access services. Cable modems and associated control systems have supported the SNMP management protocol and related MIB definitions from early implementations. These systems allow the collection and storage of data about the status of the physical HFC network. The also provide powerful insights on parameters such as the severity of HFC RF channel impairments, the efficiency of the underlying MAC protocols that arbitrate end user access to available service bandwidth, and the efficiency of the higher IP and TCP network and transport layer protocols respectively. Cable modems and systems based on the DOCSIS standard will further advance these capabilities and will support standard MIB definitions to greatly facilitate integration of systems from different vendors.

Controllers for Status Monitoring Transponders

These allow for polling and collection of telemetry data from status monitoring transponders deployed within selected active elements in the physical distribution network. Traditionally, these systems have been based on proprietary protocol implementations that prevent efficient integration into standard management platforms as well as interoperability of products from multiple vendors. In addition, transponder deployment very often results in the gathering of a vast amount of information of limited value when analyzed in isolation. These systems are also difficult to access from higher-level management platforms.

Return Path Monitoring Systems

These are specialized monitoring systems that are used to collect information related to the severity of specific RF impairments, i.e.: impulse noise and ingress that affect the HFC return path channel availability. Impairments on any of the distribution legs in the upstream path of HFC networks will affect customers in an entire distribution area because of the funneling effect of the return paths at the system headend. Therefore, dedicated systems that can track the severity of return path RF impairments at any time become a necessity. These systems must operate independently of other systems that collect performance information for network layers above the physical HFC layer. These systems must also facilitate integration into a single management platform which makes support for standard management protocols a critical requirement.

HFC Central Management Site – TMN Network Management Layer (NML)

In Figure 2, this is the location that supports all servers and higher-layer network management platforms. The various servers support communications with service controllers distributed across multiple primary hub and headend locations. All control activities for the network originate from the central management site. This layer supports specialized servers to perform five basic network management functions that also reside within the Network Management Layer in the TMN model: performance, fault, configuration, accounting, and security management. The first two functions are particularly critical to network operators as they directly impact the quality of the services delivered over the HFC network.

Performance Management

This involves continuous communication with all element management systems deployed throughout the network to collect and store all status information about the state of the physical HFC layer as well as information on the performance at the MAC, network and transport layers.

The process of collecting performance information may rely on pre-processing of data done at the element management layer. If this is the case, care must be taken to ensure this function is not duplicated between the element management and the network management layers. If data pre-processing is not done at the element management layer, the network management layer must support data reduction functions.

Furthermore, the role of performance management also demands standard ways to access all deployed element management systems and the ability to correlate performance data from multiple platforms. Referring back to Figure 2, this is accomplished through the requirement for support of standard communications protocols and formats such as SNMP and TL 1.

Three of the main goals of performance management are:

1. Identification of underlying network problems that may not immediately result in critical service disruptions,

2. Assisting in scheduling of preventative network maintenance activities as required, and

3. Uncovering and tracking trends in resource usage. This data is handed to the higher service and business management layers where it is used to determine when to expand and add services

Fault Management

This involves collection of alarm information from all element management systems. This function must also support alarm reduction and correlation of alarms generated from multiple element management platforms. The most critical goals of fault management are:

1. Fast identification, reporting and filtering of critical and non-critical alarms,

2. Cross-platform alarm correlation and fault isolation,

3. Assist in the diagnostic and troubleshooting of network problems, and

4. Expedite problem resolution

Implementing the two network management functions just described, as well as support for the remaining three functions, requires deployment of network management platforms that can enable communication and sharing of network performance information across the various element managers and service controllers depicted in Figure 2. Some of these element managers already support standard management protocols such as SNMP for OpenCable™ set top boxes and DOCSIS cable modems and TL-1 for telephony service controllers. This greatly facilitates the migration to a TMN model for the management of the HFC network. It also enables quicker integration of equipment from multiple vendors to support additional network services.

INTEGRATION OF PHYSICAL PLANT STATUS MONITORING SYSTEMS INTO HFC NETWORK MANAGEMENT PLATFORMS

As the adoption of TMN models of network management progresses, it becomes increasingly critical to ensure that any new network elements and devices added to the HFC network management platform support open management standards. Element management systems for addressing and controlling status monitoring transponders deployed in active transmission elements in the HFC network have a very important role to fill in the areas of performance and fault management. These systems can provide valuable data on the operational status of selected active transmission elements such as power supplies, fiber nodes and distribution amplifiers. As such they complement the role of terminal devices: while the latter are able to track changes in service degradation and generate warning alarms as performance thresholds are violated, the status data on active elements assists in accurately pinpointing the exact location of a network fault.

Most transponder management systems commercially available today, however, offer limited support for standard protocols. In addition, these systems do not support standards at the physical or MAC layers which makes interoperability between platforms extremely difficult if not impossible to achieve. Furthermore, retrieving of this information by a high-level management platform requires extensive development of customized data interfaces.

The critical requirement for these systems to conform to open standards has prompted renewed efforts to define standards for both physical layer (PHY) and media access control layer (MAC) operation as well as message layer specifications. The Hybrid Management Sub-Layer (HMS) subcommittee, newly created under the SCTE Engineering Committee, is currently leading these efforts with help from both network operators and equipment vendors.

For the physical layer, the proposed standard is expected to define a minimal set of specifications for operation of headend controllers and field transponders. These include RF output transmission ranges, frequencies of operation, tuning ranges, maximum levels of spurious outputs, and data transmission rates.

At the MAC layer, the draft documents currently under discussion define the format of protocol packets, addressing functionality, MAC message lengths, synchronization procedures, initialization of network elements, algorithms to arbitrate access to the transmission channel bandwidth, and mechanisms for contention resolution.

The draft documents for the message layer specification define proposed MAC message types and their formats for communications between field transponders and headend controllers over the forward and return RF transmission channels. This also includes an effort to define standard MIBs for communications with the most commonly deployed active network elements, i.e.: power supplies, headend controllers, and fiber nodes.

To date, this standardization effort has the support of various HFC network operators as well as members of the vendor community. It is expected that throughout the remainder of 1999 consensus will be achieved among all participants to enable publication of interim specifications. Once these are published, early product implementations conforming to the proposed specifications are expected early in the year 2000.

CONCLUSIONS

This paper has described a management approach for HFC networks based on migration to a TMN model. Such migration demands the adoption of element management systems and platforms that are based on open standards. Implementation of standards-based systems has already begun with the deployment of DOCSIS cable modem systems and OpenCable™ set top boxes. However, these devices by themselves will not provide all the data required to support the monitoring, troubleshooting and management of advanced HFC networks. It is also necessary to integrate information on the status of active network transmission elements before a management platform for HFC networks can be considered complete.

The support for the monitoring of active network transmission elements, although existing today, is based on proprietary vendor implementations. This prevents proper integration with other enterprise element management systems and the sharing of network performance data. A definition of proper standards to facilitate this integration is therefore a critical task that has been undertaken by HFC network operators and the vendor community. Completion of this effort is critical to the success of network operators as new advanced services demanding ever-increasing levels of network reliability are deployed.

REFERENCES

[1] Dave Miller, "TMN and the intelligent surveillance of broadband networks", http://www.internettelephony.com/archive/featurearchive/1_6_97.html

[2] Esteban E. Sandino & Albert J. Kim, "A Structured Approach for the Collection and Analysis of HFC Service Network Statistics Using Advanced Network Terminals", SCTE 1999 Conference on Emerging Technologies, Proceedings Manual: Collected Technical Papers, January 19-21, 1999, pp. 141-148.

[3] W. Richard Stevens, TCP/IP Illustrated, Volume 1 The Protocols, Addison-Wesley Publishing Company, Inc, 1994. ISBN 0-201-63346-9 (v. 1).

ADVANCING RETURN TECHNOLOGY......BIT BY BIT

Dr. Robert L. Howald
Manager, TNS Systems Engineering
General Instrument Corporation

Abstract

Traditional HFC systems have implemented the return path, a 5-40 MHz bandpass in North America, using bi-directional RF amplifiers in the coaxial plant, and a return path laser in the fiber optic node, driving a return fiber for the optical trunking. At the hub or HE, the optical power is converted back to RF. The technology is the same analog AM-based optics approach used to transport the broadcast forward path video signals. There are numerous design and implementation issues that make this approach difficult and costly. These include analog laser specifications, laser second order response characteristics, optical link length constraints, optical receiver specification, and testing of the components. All of these contribute to the overall cost issue of developing high performance analog optics. Node laser issues are exacerbated by the fact that this component must operate in an outdoor environment, specified over a very wide temperature range.

An improved approach implements a digital transport method at the point that the RF plant terminates, and the optical trunk begins inside the node. To do this, the analog laser technology is replaced by a high-speed baseband digital technology of the type that has been used in the telecommunications industry. As a digital signal, immunity from the troublesome analog laser impairments is obtained and longer distances can be covered, potentially avoiding the need for hub repeater hardware required in analog systems, among other benefits.

This paper describes analytical and design issues associated with digitizing the return path at the node. It can be shown that the analog-to-digital (A/D) converter is a mathematical analogue to the AM modulated laser technology traditionally used. We can treat the A/D converter quantization noise as the effective "analog" optical link noise. This can be correlated with the known performance capabilities of the lasers currently used. Additionally, the distortion performance, in particular the laser clipping aspect, is replicated identically by A/D input thresholds. Conveniently, a complete technology upgrade can be achieved, while the key concepts of the mathematical analysis remain virtually unchanged. The second order and third order distortions also can be kept very low in A/D's and, furthermore, the second order distortions do not degrade to poor values as analog laser can. This can be an important issue in broadband applications. Measured performance has been taken of NPR and of BER on a loaded return, showing how the dynamic range analysis above can be applied to the digitized return as it is used in analog systems. Ultimately, the strength of the digitizing technique is

furthered by the functions and processing that can be applied once the information is represented completely as bits. Some of these concepts will be introduced.

THE DIGITAL RETURN PATH CONCEPT

The basic elements of a digitized return path are shown in Figure 1. The idea is quite straightforward. The RF piece of the CATV plant in the reverse path terminates at the node. As such, this is a logical demarcation point to terminate RF signals. Until relatively recently, the 35 MHz (North America) return path bandwidth required an RF solution, and thus employed analog laser technology for reverse transport – essentially the same technology used for broadcast analog video in the forward path. The composite return path waveform AM modulates a laser, and the light intensity varies per the RF signal applied as the modulating signal. The concept being discussed here is about transporting this reverse path signal using baseband serial optical transport – representing the signal by encoding it entirely as ones and zeroes. As shown in Figure 1, there are three main elements to this. They are

1) Converting the composite reverse path waveform to a sequence of digital words whose value represent analog signal samples

2) Arranging the word into a serial stream with appropriate synchronization information to identify the boundaries between words and to recover timing of the bits themselves

3) Converting the electrical digital signal to an optical digital signal, and transmitting the optical ones and zeroes across the fiber

4) Inverting the process at the receive side

This paper will describe the unique parts of each function, which, together, enable this concept. Performance analysis will follow, along with measured results from laboratory prototypes. Finally, the discussion will touch briefly on the next-next generation of digital implementation.

ENABLING TECHNOLOGIES

Analog-to-Digital Conversion

Analog-to-Digital converters (A/D's) have been advancing nearly as rapidly as DSP and VLSI themselves. Digitization of an analog signal for the purposes of allowing efficient transport between locations is obviously an approach with strong roots in the telecommunications industry. In that case, however, the A/D conversion can be quite pedestrian indeed. Voice bandwidths are on the order of 4 kHz, and the channel is baseband and switched in nature, so there is no aggregate build-up of analog bandwidth. In an HFC reverse path, however, both of these simplifications are thrown out. The HFC return is a shared channel, so each user sharing a node consumes bandwidth, effectively aggregating the total possible consumed bandwidth upward with more users. Rather than low frequency baseband, the HFC return is a 35 MHz wide bandpass. Additionally, individual signal bandwidths that can be placed on a cable far outstrip the stingy bandwidth capabilities of the twisted pair to the home

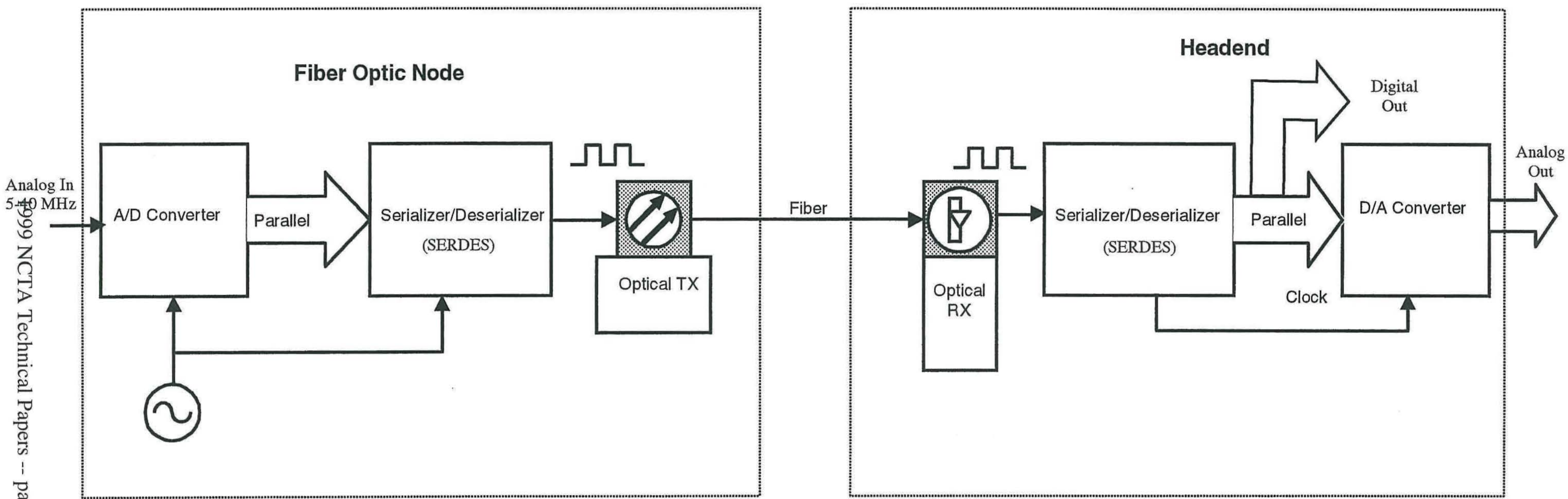


Figure 1 - Digitized Return Path - Basic Operation

originally optimized for voice. This set of differences create the requirements for a significantly different A/D beast than is used in the telco world. Manufacturer's meeting of the requirements for higher performing, robust A/D components necessary for implementation on the HFC return path has come to pass relatively recently. Some key performance items are sampling rate (speed), resolution, and environmental robustness. These first two performance parameters are discussed in more detail below.

Sampling Rate

Nyquist theory tell us that the sampling rate necessary for undistorted signal conversion or reconstruction is twice the signal bandwidth. Most textbooks begin with the most fundamental way to look at this, which is that the sampling rate must exceed twice the highest frequency of the signal. Note, however, that these introductory descriptions assume a baseband (lowpass) signal, and that such a description is really a subset of the more general sampling/ bandwidth relationship. Either view works for this case, and, if we assume that the return path to be converted is 40 MHz worth of analog bandwidth, then an A/D must be sampled at 80 MHz minimum. Now, glancing back into those old signal processing texts, recall that the digital spectrum is periodic, repeating the analog spectrum around every integer related clock component. Filtering that takes place before A/D and after the digital-to-analog converter (DAC) is required to maintain spectral purity through the processing. Practical guidelines for filter design result in the sampling rate to be reasonably higher than the minimum theoretical clock rate, such as 100 MHz clock. For this example, to implement return path digitization, words representing the analog waveform in binary form will come out of the A/D parallel outputs at the rate of 100 million words per second.

Sampling a signal at 100 MHz requires some very intricate design complexities. Solving these design issues has occurred over time. It makes great textbook reading showing nice sampled impulses of some make-believe waveform at discrete time instants. However, there are obvious complications to this picture. Any real sample instant will inherently be *not* infinitely small. Designs must be made to snatch a rapidly varying waveform and hold it steady. A/D terminology associated with these high performance sampling issues are *flash* A/D's and *sample/hold* or *track/hold* technology. Additionally, input conditioning amplifiers requiring very high linearity, but now across RF bandwidths. This is a great challenge outside of an A/D, much less as part of the device. The rapid sample taking puts pressure on the clocking and jitter capabilities, often surfacing in literature as *aperture jitter*. Also, of course, the output of the A/D is a digital word, so advances in fast logic using standard logic families pays off for this device. Lastly, the word formation has moved the digital rates and bandwidths into the realm of RF reality. This means matching, reflections, impedance games – the works. Perhaps most notable in this arena are the circuit design and layout issues. Every RF designer, particularly oscillator designers, know that RF and digital don't mix. Nice, clean RF signals get mucked up quite viciously by voltage-hungry and current flip-flopping logic components. Only now, not only are these sharp-edged waveforms on the same board containing analog inputs, but the clocks and toggling frequencies themselves, as well as the input signal, represent that same class of easily-spreading spectral components. Very careful board design and layout, with specific attention paid to analog and digital return current paths is essential.

Obviously, the evolution of the hundreds of mega-samples-per-second (MSPS) A/D's were critical to the application of the technology for HFC. It has only been within the past few years that such parts have become available at high resolution and with environmentally robust performance. There are a small number of key vendors who play in this arena.

Resolution

Any new technology development that hopes to capture interest must permit a smooth transition into its implementation. From the standpoint of hardware deployed at the node, there are several new pieces. However, because optical modules are often plug-in type packages to allow design flexibility, it is fully anticipated that the hardware aspect of the node will not be very significant. Additionally, as will be discussed below, there is fundamentally no change in alignment or return path maintenance due to the approach.

Beyond the physical implications, however, there are the performance aspects. Current return path technology – AM modulated (analog) lasers – has a certain level of well-characterized performance. It also has selectable grades of performance from low power Fabry-Perot (FP's) lasers to high power Distributed Feedback types (DFB's). There has been substantial return link characterization work going on throughout a host of companies in the industry. Because of this, SNR and distortion numbers achievable by common optical technologies available today are well known. Clearly, all of the effort that has gone into what can be effectively carried on the return path implies that equivalent performance is a logical start. However, it is worthwhile to note that there are also good arguments as to why eight bits and even six bits can effectively handle particular tiers of digital service, which opens the doors to bandwidth savings. Most notably, SNR's in the range of 40 dB have large margin above what is necessary for most anticipated services even ignoring coding gains (consider 16-QAM @1e-8 needs 22 dB uncorrected). Also, most return paths are not constrained in performance by their thermal noise component (AWGN), but instead by transient and burst noise phenomenon. Having increased SNR does not efficiently guard against these phenomenon. It is wise to assure dynamic range headroom in return performance, although overload problems are rarely ingress-related. But the amount of headroom necessary and the cost to provide it, given the link margins available, is a trade-off exercise worth evaluating.

Consider the SNR equation above. Applying a 6-bit conversion, and no oversampling, an SNR less that 40 dB is obtained. Furthermore, this is a maximum (full-scale) number that cannot be achieved in the real application because of the back-off necessary to stay within the A/D dynamic range. On the other hand, an 8-bit conversion lands in the 50 dB range. Again, there is signal back-off from peak required, but the SNR numbers now fall into the range known to be achievable and acceptable with current technology. Going through the noise and distortion numbers as we will below, it can be seen that to achieve customary performance requires the high speed A/D's to have between about 8-12 bits of resolution to be similar to the range of analog capabilities. This will be quantified in a discussion to follow.

Serializer-Deserializer (SERDES) IC's

Major efforts in serial transport technology are currently developing in two arenas – Gigabit Ethernet and Fibre Channel. The results of these standards-based developments can reap rewards, too, for the cable industry. Handled properly, these powerful, new IC's can perform all of the essential functions necessary to implement a digital return from node to HE with complete transparency. The magic in the latest advances for HFC is, again, the Gbps data rates being pushed, and the level of integration being achieved.

The basic functions achieved are fourfold.

Parallel-to-Serial Conversion (and Serial-to-Parallel Conversion)

Parallel digital words at rates around 100 Megawords per second are being delivered to the SERDES chips. The parallel words must be made into a serial stream to transport over the optical link. Thus, the SERDES must latch in the information, multiply up the clock rate, and deliver the increased serialized rate to the optics. For a 10-bit device, the result is serial rates at about 1 Gbps.

On the receive side, the inverse obviously is performed, and the output of the SERDES receiver chain is a digital word handed off to, for example, a D/A converter for analog signal reconstruction. Or, the word is handed off in digital fashion to a direct interface to the application's digital receiver.

Optical Drive

As the standards mentioned develop, compatible optics is also developing. Current technologies can often mix-and-match and be compatible, or very close to it, with just basic modifications, such as signal coupling, pull up/down, impedance transformation, etc. Part of the similarity rests solely on the fact that, at such high data rates, some form of emitter coupled logic (ECL) is unavoidable.

Timing

The parallel-to-serial scenario painted above is a simplistic one at best. The receive IC is delivered only one piece of information – a stream of bits. From this, it must achieve the timing to detect the bits (developing a synchronized clock to sample them at the optimal instant). As might be expected, each side of the link plays a role in this. Clock recovery systems and requirements have become well developed over the years of explosive digital growth. At the transmit side, the role is to encode the data such that there is guaranteed to be enough data transitions for the clock recovery circuit to acquire and hang onto. Thus, the transmitter employs encoding that provides randomization to guarantee transitions in all situations. A simple case that explains the problem would be a very lightly loaded TDMA return, such that the A/D output words are frequently all zeroes. Without transitions, the clock frequency component necessary - embedded in the data sequence through the bit transitions - cannot be tracked effectively by the receive PLL. The role of the PLL is to re-generate this timing to yield effective serial sampling. Subsequent to bit timing, the word timing information - now divided down from the serial rate - is needed at the output for regeneration of the parallel word.

Framing

While it is obviously necessary to achieve bit timing, it is also necessary to know which sets of bits belong together as one analog sample. In other words, it is necessary to make sure that the bits of some 10-bit word representing an analog sample on the transmit side get delivered together to the D/A at the other side. This is another problem with well-developed solutions in the networking world. Solutions center around sending known, predictable patterns signifying word boundaries that can be detected. For HFC, caveats exist. For example, the basic functions of the standards-based chips involves the movement of packets to and from the IC's from other digital sources. This allows, with minimal intrusion, transmission of demarcation characters to be sent between payload packets, particularly with the high sampling rates achievable creating many time slots. For HFC, however, some more intricate design functionality is necessary, as the A/D operation is delivering words constantly in real time to the SERDES function. Additional overhead must be inserted in more creative fashions.

Optics

High-speed baseband digital optics has been around for quite a while. What completes the needed requirement set for HFC is modular, small, wide bandwidth modules, encompassing the data rate needed to capture the RF bandwidths elevated via sampling and serialization. Add field robustness, dropping costs of optics, and increased data rates on the short term horizon, and the units now become both relevant to current needs, and capable of expanding as the need for bandwidth grows.

LINK PERFORMANCE: MATHEMATICAL ANALOGIES

RIN & Quantization Noise

The design complexities discussed above result in the A/D being, of course, non-ideal. There are many ways to express this. There are also ways to express how an ideal A/D *should* behave. The ideal SNR of an N-bit A/D with a full scale sinewave input (peak of sine wave at maximum input threshold limit) is

$$SNR = 6N + 10 \text{ Log } (f_s/2f_{bw}) + 1.76 \text{ dB}.$$

There are a couple of key items to mention here before moving on. Note that SNR moves as 6 dB/bit of A/D resolution, a commonly reference relationship and a good one to remember. The second term points out the benefit of oversampling. The term f_s represents sampling rate, while f_{bw} represents the bandwidth occupied by the input signal of the A/D. The noise associated with A/D quantization is basically flat across the digitized bandwidth. Higher clock rates do not change the noise power, but widen the digitized bandwidth by increasing the clock frequency, thus lowering the noise density. The result is that over the Nyquist bandwidth of interest, the noise power is lower. Finally, the last obscure-looking term actually finds its place by calculating the noise power under the assumption that it is approximated as a Uniformly distributed probability density function PDF. The result is a multiplicative factor associated with the variance of a Uniform PDF that concludes in 1.76 dB. Interestingly, this is a real case of a noise distribution having a Uniform probability density, or approximately so - not an easy thing to find in nature. The usual additive noise processes dealt with are Gaussian in fact or by assumption. Because of the source of this noise, in this discussion we will label the SNR as SQNR - Signal-to-Quantization Noise Ratio - for clarity. It is useful to separate this contribution from other thermal noise contributions in a link, such as noise funneling in the reverse path. In the HFC application, reconstruction of the signal to the analog domain also entails accounting for the noise figure contributions in the analog processing that follows. Despite the nomenclature, the SQNR is not handled any differently in system analysis.

Full Scale Input

Converter SQNR varies as a function of signal level. The reason for this is the absolute nature of quantization noise described above. The possible error of the digital encoding is still half of a quantization step size regardless of whether the signal is large or small. This may sound inconvenient, but in practice turns out not to be so for a couple of reasons. First, the SQNR is based on a CW analog measurement, in keeping with conventional practice across industries. Of course, it is not reference to a 4 MHz bandwidth, an item unique to the television industry. It is instead referenced to the Nyquist bandwidth for the conversion rate. However, this brings us to the second reason. That is, the way SQNR is applied is consistent with how specifications are being written for analog return path products, or at least how they are best written from a purely technical standpoint. That is, the SQNR over the whole bandwidth is specified. For the A/D case, it is given at maximum CW input level. This level is called the A/D *full scale* input level when the input is a sine wave with peak amplitude equal to A/D's maximum input level (in the simplest coding example, the amplitude corresponding to an of 11111111). Sometimes, the manufacturer will specify the SQNR just below full scale.

The Nyquist bandwidth is the well-known maximum signal bandwidth that can be delivered to the A/D without reconstruction problems associated with the frequency domain aspects of discrete processing. Basically, if the clocking frequency of the A/D is not at least twice as high as the A/D input bandwidth (we are treating bandwidth as a lowpass 42 MHz for purposes of this discussion, although lowpass bandwidth is not strictly the requirement for proper processing). The reconstructed output is not uniquely defined by discrete samples. Thus, if a 100 MHz clocking of the A/D is assumed, the Nyquist bandwidth is 50 MHz. In practice, there needs to be some spectral margin below the Nyquist frequency of 50 MHz for reasonable filter design, much like in diplexer design.

The Nyquist variable is not a major issue as far as performance calculations are concerned. The fact that the Nyquist bandwidth varies is accounted for in the SQNR equation given previously. However, in practice, the A/D clock rate never changes in a given system, so there is not a need for repeated manipulation to calculate SQNR. In addition, a manufacturer's specifications of SQNR apply over a range of clock rates, under the assumption that the true user SQNR is determined by the equation above once clock rates for the application are established. For high performance devices, the SQNR may, indeed, be given at several clocking rates. Very high clock rates, and especially the higher resolutions, challenge the A/D performance capabilities for noise and distortion.

Back-Off

As described, the SQNR is typically defined for a full scale sine wave input, or very close to it. Of course, this is not the look of a practical input signal. The stochastic characteristics of the input signal to the A/D are generally related to the loading in the reverse path relative to the maximum recommended design level. However, as in analog optics, under the assumption of constant power-per-Hertz loading, the link design levels and noise and distortion specifications are typically built around the assumption of a fully loaded return. This assures any loading condition will meet the requirements. For the fully loaded case or even the moderately-to-heavily loaded case, the input signal to the A/D looks noise-like for the periods of time that there are simultaneously transmissions. The light loading cases are less threatening from several standpoints in link analysis. First, there is possibility of not using uniform spectral loading if it is desired to dedicate more power to some channels and a straightforward approach to accommodate this is available. This translates to higher SNR. Second, light loading in a constant power-per-Hz approach means clipping is not an issue.

The approach to bounding the anticipated input signal dynamics from a design standpoint is to assume the composite reverse path signal at maximum suggested input looks Gaussian statistically. Single PSK and QAM channels have a noise like look on a spectrum analyzer, but not the same statistical characteristics as noise. Under the Gaussian assumption, the noise and distortion performance of the quantization becomes a tractable problem. In fact, any PDF we can assign would create a definable problem, but the Gaussian quality allows us to draw on some well-known properties to complete the analysis. The result is that it can be shown precisely what optimal input *back-off* is - the number of dB below full scale the input power is run into the A/D. Thus, the noise and distortion performance can be optimized theoretically, and even practically, in the sense that the A/D's *effective number of bits* (ENOB) – the actual resolution of the part in bits degraded by non-idealities. ENOB can be used in place of the performance of an ideal A/D in the analysis. Additionally, the noise and distortion performance can be described in what has become another very useful method of characterizing reverse path performance - noise power ratio (NPR). Elements of the analysis will be described further in a later section.

SQNR, RIN, and Analytical Models

One very convenient side benefit of the digitized return is the ability to focus all impairment analysis and design implementation at the termination of the RF plant at the A/D at the node. This assumes the eventual migration of this family of products into fully digital products in hubs and Headends, a logical and necessary step to fully reap the rewards of having all of the information being moved by transporting bits. Until that time, there still will be noise and distortions added by D/A conversion and processing. The D/A is significantly less expensive that the A/D, but contributes roughly equally to distortion. Following the D/A conversion, sophisticated application receivers now and in near future will immediately re-encode the reconstructed analog output with another A/D conversion in the receiver itself, and obviously wasteful effort. Of course, there will still be traditional boxes and services running on the reverse plant for some time, and some of these may even be analog receivers. Thus, a D/A converter will likely be a necessary option for some return applications. However, it is also the case that these applications are often simple modulation such as FSK, and not significantly effected by additional distortion contributions.

Using the A/D only model, we can conveniently relate quantization noise of the A/D to relative intensity noise (RIN) in an analog laser. In this way, we can relate the performance of an A/D to known analog optical performance. Table 1 shows this comparison using typical return path laser powers for the case of a 100 MHz A/D converter at three different resolutions - eight, ten, and twelve bits. The calculation of densities can be observed from Figure 2. The serial payload (i.e. ignoring overhead associated with any further encoding) for these cases would be 800 Mbps, 1 Gbps, and 1.2 Gbps - all well within the bounds of current optical transceiver technology.

From Table 1, some conclusions can be drawn. An 8-bit A/D provides theoretically about the same performance as the lower quality FP type laser. The 8-bit parts also will tend to provide closer to theoretical performance in terms of ENOB because of their lower resolution. A 10-bit A/D provides an in-between performance category between FP and DFB. The 12-bit device provides nearly the same performance of a return DFB laser.

Table 1 is instructive, but there are two significant items of note with this table that need to be further clarified. First, this table ignores other analog optics impairments in the link (shot noise, receiver noise noise current, iin). Second, the A/D column is based on a full scale input, when in fact the noise like signal characteristics will mean that it will be run backed off (as a laser is). Third, the table masks the distance advantage of the digital approach. Analog link performance degrades about a dB per dB, even worse, at long link lengths. At link lengths that digital optics can achieve, analog link would not be workable without repeater functions.

Table 2 shows how the numbers compare against a sample set of 9 dB optical links that include these impairments using typical receivers. The A/D numbers are adjusted by their performance at the optimum back-off setting minus 3 dB (3 dB of headroom in total power load). Note that this comparison emphasizes the point that the impairments of the digitized link are solely A/D in a complete design.

Table 2 shows that the digitized approach stands up very well to traditional analog performance, without even further considering the distance advantage. By factoring in ENOB rather than theoretical bit resolution, there is still comparable and improved performance for quality A/D converters, where good implies loss of a half a bit to a bit with today's devices.

As another advantage, a similar discussion can take place for the distortion side of the analysis with respect to the clipping impairment. That is, there is a clear relationship to comparing performance to analog optics. In fact, it is even more direct in this case - providing about an exact analytical replication of laser clipping analysis.

Analog Distortion

While the above equation captures the theoretical SNR, the fact that the A/D has an analog path leading up to the digitization means that analog distortions occur in this part. These distortions can be described in very much the same fashion as stand-alone RF or analog parts. However, by the very process of digitization, the distortion energy and its distribution across the spectrum is inherently also a function of the clock frequency. That is, the repetitive nature of the digital version of the spectrum means that traditional harmonic and/or two tone IMD components appear at frequencies related to both the input frequency and clock rate. Second order and third order distortion numbers for this class of A/D converter are, for example, in the mid-60's range for second and third order distortions, all assuming close to full-scale input levels. Like other analog parts, lower levels are generally better for distortion. However, unlike typical analog parts, only considering predictable or low order distortion components can be misleading. The aliasing back in band of intermodulation components can make any distortion component fall in some way into the desired bandwidth, and seemingly meaningless high order components that would otherwise be well out of analog band can show up because of discrete processing. This makes a composite definition for noise and distortion very useful. Examples include noise power ratio NPR, and ENOB.

An interesting interpretation of the quality of an A/D converter is given by its ENOB. Essentially, it takes into account all of the non-idealities of the converter, and expresses this analog performance obtained by the actual part in terms of what number of equivalent bit resolution is obtained. In other words, a 8-bit A/D converter, because of its non-ideal performance, may yield only effectively 7.5 bits of resolution, meaning it can be modeled as a 7.5-bit converter from the standpoint of performance. It is analogous to implementation loss as it is used in modem designs. Modems are often recognized as having some X dB implementation loss, which describes the actual performance by comparing how close it comes to theoretical BER performance. A high-speed 10-bit A/D may have an ENOB of about 9 bits, indicating loss of about one bit due to constraints of the practical implementation. Recall, this would be equivalent to a loss of about 6 dB of SQNR.

There is one other distortion component of an A/D converter, but it is not associated with the non-ideal nature of the part. Instead, it has to do with the nature of the A/D function. It is a concept quite familiar to those in HFC – clipping.

Clipping

The analog laser's clipping impairment effects have been studied quite extensively for both forward and return applications. Products are designed with requirements built around a suggested maximum input range, and this range is

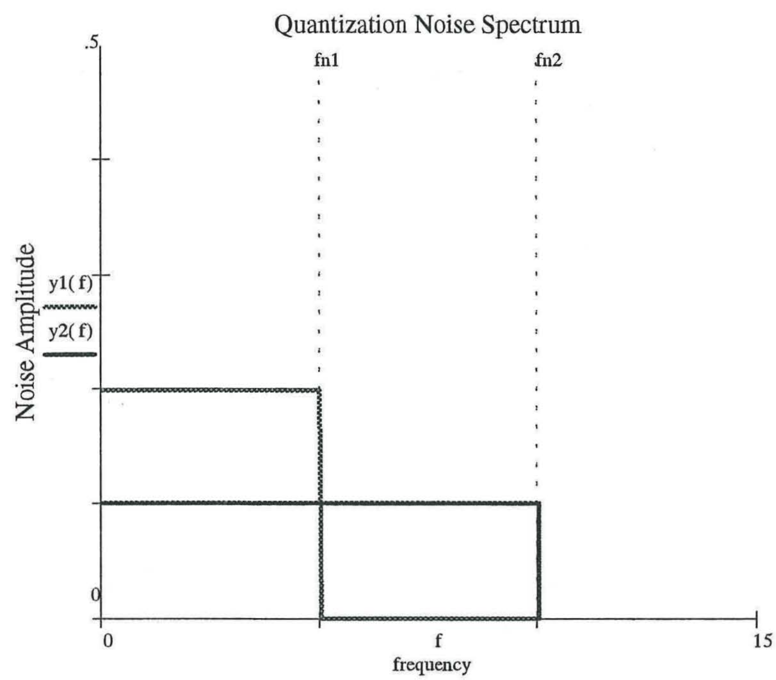


Figure 2 – Effect of Sampling Rate on Noise Spectrum (fn = Nyquist Bandwidth)

Table 1 - Noise Performance: Return Path Laser RIN vs. A/D Quantization Noise

Device	RIN	Quantization Noise (dBc/Hz @ 100 MHz)
8-bit A/D	-------------	-127 dBc/Hz
Fabry-Perot (FP)	-130 dBc/Hz	-------------
10-bit A/D	-------------	-139 dBc/Hz
12-bit A/D	-------------	-151 dBc/Hz
Distributed F'back (DFB)	-155 dBc/Hz	-------------

Table 2 - Noise Performance: Return Path Analog Optical Link (9 dB) vs. A/D

Device	FO Link Noise	Quantization Noise (dBc/Hz @ 100 MHz)
8-bit A/D	-------------	-112 dBc/Hz
Fabry-Perot (FP)	-119 dBc/Hz	-------------
DFB	-122 dBc/Hz	-------------
10-bit A/D	-------------	-123 dBc/Hz
12-bit A/D	-------------	-137 dBc/Hz

the manufacturer's recommendation for providing what they consider the best trade-off between the noise and distortion contributions. Clipping is considered a distortion contribution, although it is the case that on a spectrum analyzer, it appears as a broadband noise floor increase. This is simply due to the time domain characteristics of clipping events - they are typically very short impulsive events, corresponding to broadband events in the RF plant. At the receiver, of course, they are spread in time by the matched filter prior to detection.

Fortunately, none of the work in understanding laser clipping goes to waste when we talk about the return path digitization concept. There is just as much if not more prior analysis related to the clipping effect as it relates to an A/D converter. There is one physical difference in the mechanism of the clipping event. In the analog laser, the event occurs by negative-going amplitude peaks shutting off the laser by dropping the instantaneous current below the bias. By contrast, the A/D converter is limited by maximum and minimum peaks that exceed the all-one and all-zero words. Beyond this, all of the other well-known aspects of the clipping phenomenon apply from an analysis perspective. For a noise like load, the events where the input exceeds the threshold are short and impulsive. Some may be shorter than a sampling period, and the likelihood of sampling during an event are statistical since the signal itself is. We approach the overdrive analysis the same way. From a BER standpoint, clipping events are largely mitigated by FEC if the event is of enough energy to cause an error to begin with. The impulses are spread by a receiver's matched filter, and contribute an additive impairment. Also, the clipping impairment for A/D's creates an NPR curve that slopes steeply on the distortion side because of the hard clipping, higher order distortion components.

Dynamic Range Performance

Dynamic range, as always, describes a unit under test's performance range in a way that captures it thermal noise limitations on one end of the input signal range - its small signal performance - and its large signal performance on the other end. An amplifier, for example, is characterized by noise figure and some measure of distortion performance, such as intercept points or CTB and CSO. Use of these parameters to describe dynamic range is perhaps the most common usage of the term, although it can be generalized to other concepts.

Noise Power Ratio

An A/D converter can be described in terms of dynamic range, and it is necessary to do so to make comparisons among technologies. More importantly, its performance characteristics must be able to be expressed to allow the device to be analyzed for its contribution to link impairments, just like the rest of the pieces of the RF channel. In the return path of HFC networks, a most convenient way of doing this is through use of the noise power ratio (NPR) test.

The NPR idea is very straightforward, and the results very informative at a glance, which is one of the reasons for its widespread use. The basic idea is to use a noise load as an input signal, thus taxing the linearity of the device by delivering a signal of wide bandwidth and high peak-to-average ratio. A bandstop filter is used to create a notch in the white noise spectrum. The depth of this notch in dB is then a measure of the noise of the device only for low total input power. As the total noise power is increased, the notch will get deeper with each dB of increased signal power, until the linearity constraints of the device are reached, at which point distortion components begin to fall in the notch bandwidth, decreasing the notch depth. This curve - notch depth versus input power - is an instantaneous snapshot of the device noise an distortion capabilities.

Figure 3 shows an NPR performance curve of a typical Fabry-Perot (FP) laser through an 8 dB link, recovered through a return path receiver (RPR). The qualities shown are typical. The noise-only impairment region on the left of the peak changes dB-per-dB with input power, and the distortion region on the right side of the peak degrades much more sharply. This effect is dominated by laser clipping, which causes high order intermodulation effects as the maximum NPR is crested, creating the well-known "crash point" phenomenon. Between noise side and distortion side, the curve peaks at the maximum NPR, corresponding to optimal input power point (but with zero headroom). From Figure 3, it is apparent that an 8-bit device closely resembles the performance of the FP link. Figure 4 shows a DFB on a 9 dB link. This figure closely resembles 10-bit performance.

NPR for A/D's

The use of noise power is especially useful for return path, where multiple, random, uncorrelated, digital inputs tend towards a noise-like characteristic, and behave statistically Gaussian. As described, A/D specifications are sinusoidal-based for convenience, as most analog components. As in most of these cases, the practical signal looks much different. And, also like most cases, what happens when things change is important. In amplifiers, where distortions are generally soft limiting effects that create second and third order distortions, the multiple carriers/noise load effect is to degrade common two-tone sinusoidal intermodulation levels further. It is somewhat complex, but an analytically and numerically tractable problem. Forward amplifiers capture the effect with CTB and CSO. Reverse amps do the same for commonality, although NPR is becoming more common. Hard limiting effects like lasers and A/D's create higher order distortions, making them ideally suited to NPR.

Assuming a Gaussian input to the A/D, the nature of the discrete transfer function makes the problem tractable. An optimum back-off can be derived from the analysis, which turns out to be a function of the number of bits, as might be anticipated. Because of the A/D's discrete nature, the point at which the A/D threshold is exceeded by an input is precisely defined. The input is also defined, by assumption, to be Gaussian. The convenient part of this is that the times when we are concerned about clipping effects are primarily times that correspond with heavy loading, which is also when the input looks most noise like, making the assumption the most valid. Without developing the detailed mathematical steps, the analysis follows these steps [1]:

(1) The clipping level of the A/D can be related numerically to the average input power.

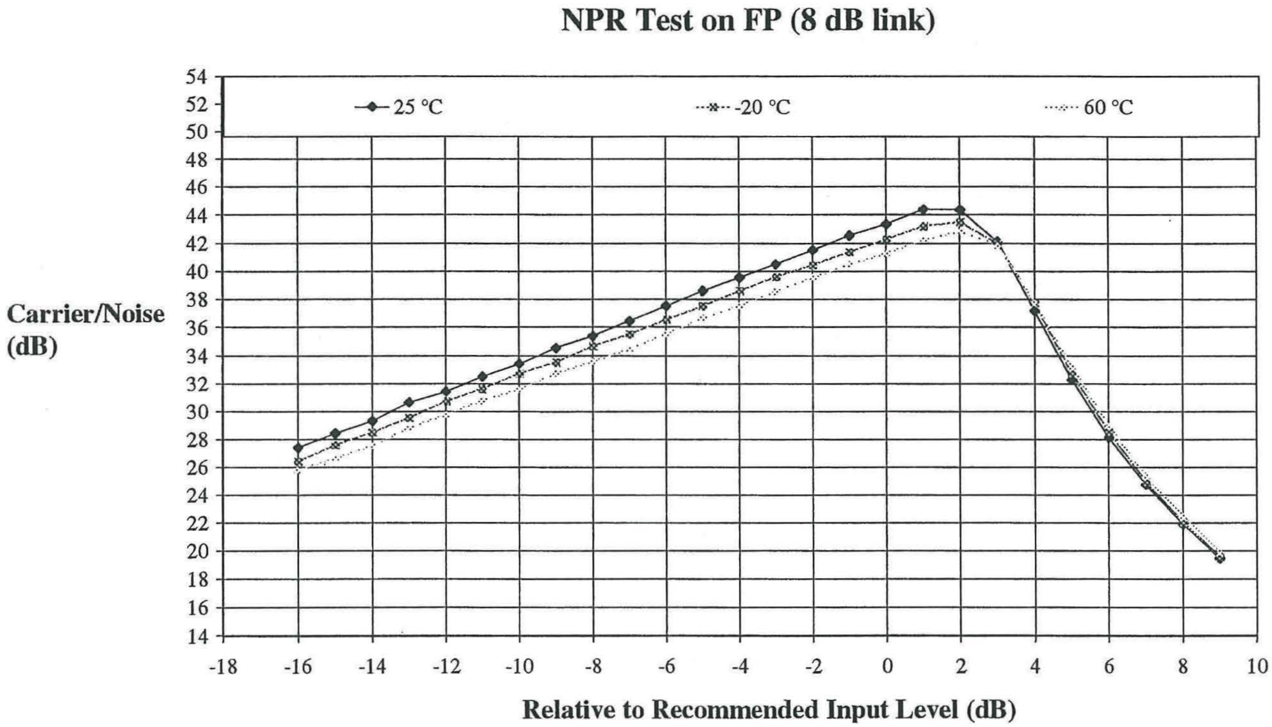


Figure 3 – NPR Performance of an FP Laser on an 8 dB Link

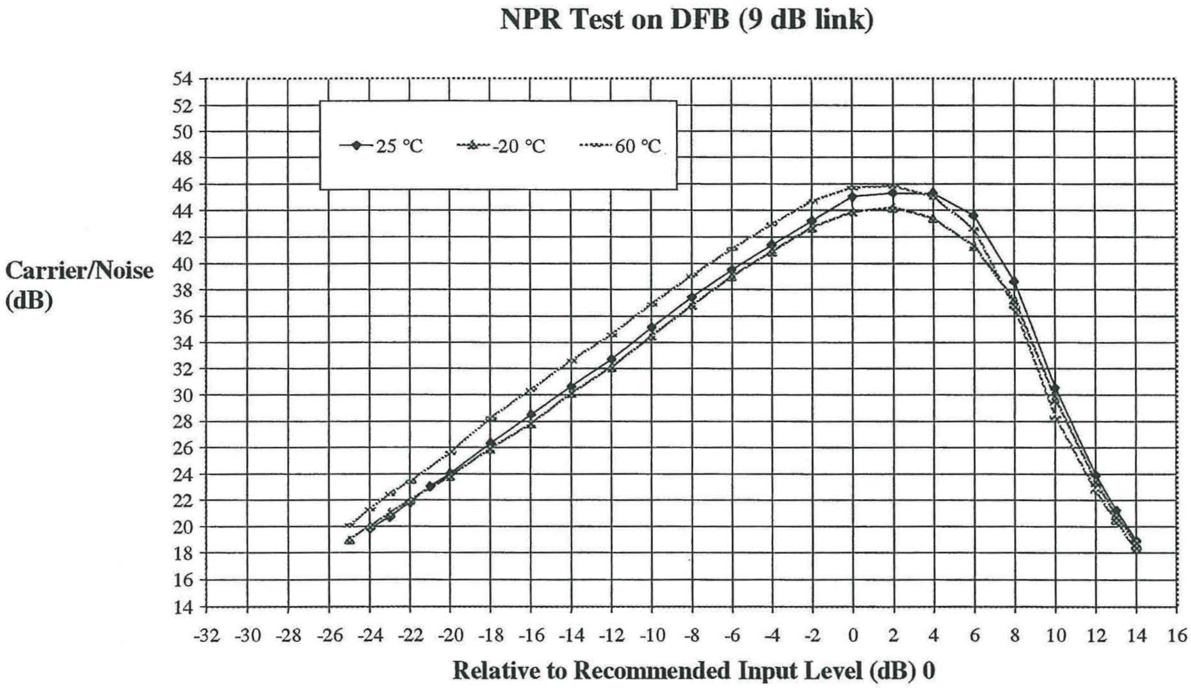


Figure 4 – NPR Performance of a DFB Laser on a 9 dB Link

(2) The input power of a zero mean (AC-coupled) Gaussian signal is its variance.
(3) The rms value of the input voltage is the square root of the variance (the standard deviation).
(4) The A/D's noise power, for a given rms/peak relationship in (1), is calculated from the error voltage squared integrated over the Gaussian likelihood of that voltage being between minimum and peak. The error voltage function of an A/D in the linear region is a sawtooth because of the quantization process.
(5) The distortion power is approached the same way. But, the calculation integrates over the likelihood of the Gaussian input exceeding the peak, and the error voltage linear increases as signal exceeds peak.
(6) The NPR is the input signal power divided by the sum of noise plus distortion power: NPR = $P_{in}/(Q_{noise} + IMD)$.

The results are shown in Figure 5 for 8-bit, 10-bit, and 12-bit A/D converters. Note the 6 dB/bit rule of thumb is easily visualized on the left (noise) side of the curve. Peak NPR's for 8-bit, 10-bit, and 12-bit devices are also about 12 dB apart. For 8-bits, the peak is about 41 dB, for 10 bits about 52 dB, and for 12 bits about 63 dB. The difference is not exactly 12 dB because at the peak of the curve, the distortion contributions are also being felt, not just the 6 dB/bit noise contribution.

Table 3 lists peak NPR, and approximate dynamic range of NPR above 25 dB, 30 dB, 35 dB, and 40 dB. For reference, 16-QAM running at a 1e-8 uncorrected BER requires 22 dB of SNR, where SNR refers to the additive Gaussian thermal noise only contribution. QPSK needs only about 15 dB of SNR. Furthermore, compared to 16-QAM, QPSK behaves more alike when embedded in a distortion floor as in AWGN, because of its robustness and lack of amplitude information in the signal structure.

It is interesting to note as a reference that a back-off of -14 dB corresponds to an OMI (peak) of 20%.

Measured NPR Performance

Figure 6 shows an example NPR measurement, capturing peak NPR performance of an 8-bit A/D (through a high resolution reconstruction D/A). Performance is close to expected, given the ENOB of the tested devices as the "theoretical" consideration. Theoretical performance of an 8-bit A/D degrades from about 41 dB to about 35 dB for a part with a 7-bit ENOB, roughly the specification for the device under test. The actual performance plot, shown in Figure 6, indicates an ENOB of about 7.4 bits. This can be calculated from the measured peak NPR of 37 dB. Thus about a half-bit is lost in the implementation of this A/D. The optimum back-off is 11.9 dB for eight bits. It becomes greater for higher resolution, because the quantization noise contribution is diminished, while the clipping portion stays the same. Thus, it makes sense to back off further into the lower noise for best performance.

THE NEXT DIGITAL WAVE

Processing

By now, most CATV folk are familiar to some extent with DOCSIS, the

Table 3 - Theoretical NPR Performance and A/D Dynamic Range

Device	Max NPR @ Backoff	DR @ 25 dB	DR @ 30 dB	DR @ 35 dB	DR @ 40 dB
8-bit A/D	41 dB @ -11.9 dB	20 dB	15 dB	8 dB	2 dB
10-bit A/D	52 dB @ -13.0 dB	32 dB	27 dB	20 dB	14 dB
12-bit A/D	63 dB @ -14.0 dB	44 dB	39 dB	32 dB	26 dB

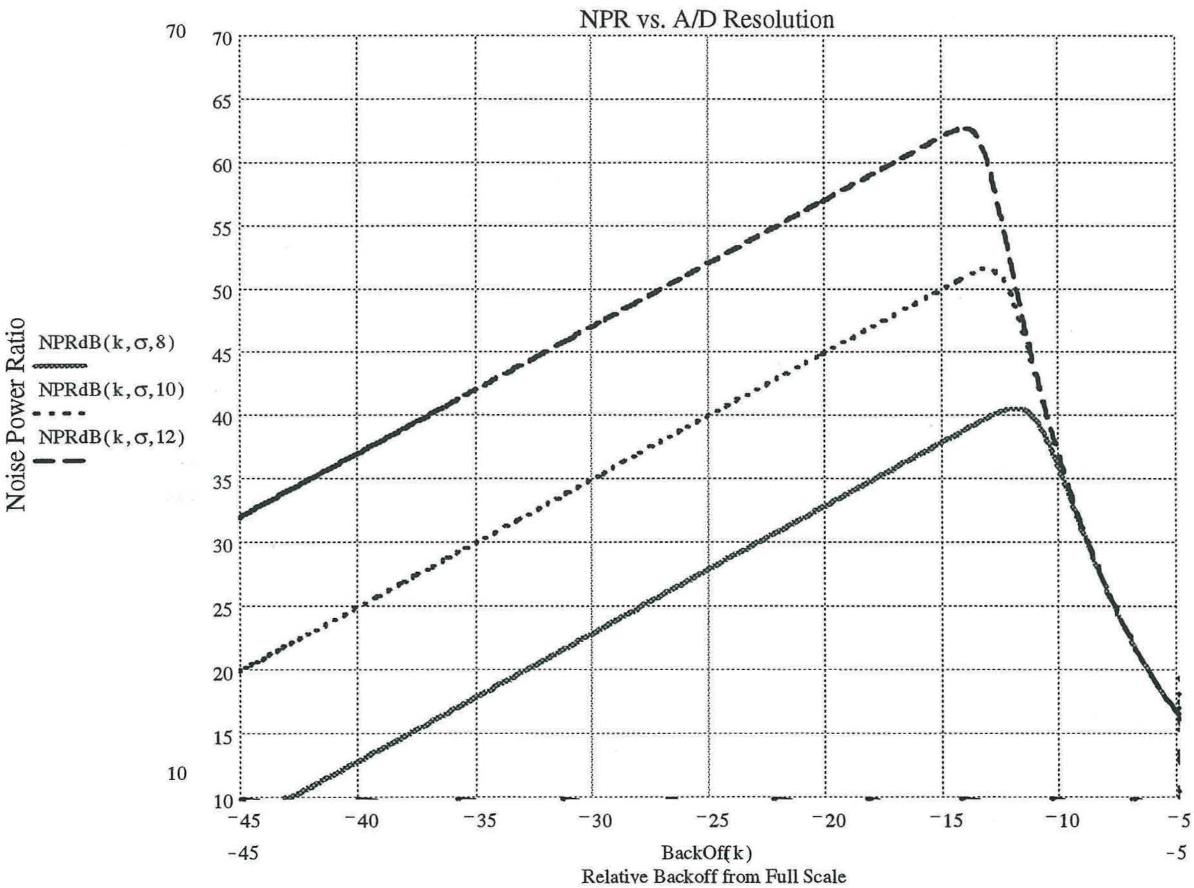


Figure 5 – A/D Dynamic Range Performance vs. Resolution

culmination of cable modem standardization efforts. A detailed read of the specification yields many impressive requirements, and representing truly a state-of-the-art communication system design in all respects. DOCSIS compatible chipsets are a very powerful breed. Similarly powerful chipsets are available in the quickly-losing ground world of ADSL, as the channel being worked with in that case also has a unique set of major impairments to deal with, as well as obvious bandlimiting characteristics. In both cases, functionality available within these chipsets opens up possibilities to network architects to bolster the system flexibility and robustness. And, obviously, they come with the built-in ability to communicate. As the return is turned into all bits, options for processing and taking advantage of processing functionality offers many possibilities for distributing functionality and intelligence throughout the network.

Multiplexers

The current short term transparency of the digital return revolves around reconstruction of the digitized waveform at the HE with a D/A of equivalent resolution (or better). The D/A function, however, introduces the same set of analog impairments as any other analog link – noise and distortion. And, in fact, the high speed products necessary for this application have some specific hardware complexities that make these devices also difficult to implement with high performance (the D/A still, however, is a significantly less costly part than the A/D). Unfortunately, the D/A output will be amplified and split, then delivered to an application receiver. This receiver will then often immediately digitize it and perform the receive, synchronize, and demodulation functions in an all-digital receiver. The wastefulness and illogic of this is obvious.

Looking past the immediate capabilities, it becomes clear that current analog HE equipment will need to take on the look of digital multiplexing products – shipping of the right bit set to the various application receivers, and speaking the right protocols. The fact that digital passage of information is involved opens up many flexibility options, as networking equipment can be controlled and programmed for maximum flexibility.

Transport

The implementation of an all bit stream at the RF-to-optical conversion point immediately brings into the equation the idea of standardized transport, a leveraging these telco-grown technologies for HFC architecture optimization. Use of SONET, SDH, or other standards-based transport simplifies network implementation, accelerates development, and brings with it a comfort level of a proven, robust technology to the HFC infrastructure. The fact that the digitization is pushed out to the node means that hub equipment can become common digital multiplexing equipment. Or, because of the distance advantage, passive equipment of removal completely of hub sites can be a possibility.

CONCLUSION

Here is a piece of non-earth shattering news: the world is going digital. And, what is already digital is becoming advanced digital. Examples abound everywhere – computers, DSP progress, advances in VLSI, even in our own backyard through the evolution of 256-QAM signaling and HDTV. This axiom will apply to the digitized return path also. The list of potential short term benefits are apparent: optical costs, total equipment costs, link distance, distortion performance, component specification, test simplicity. However, perhaps the most benefit is in the years to come, as the conversion of the return path to all bits opens up many networking possibilities associated with processing, transport, multiplexing, and terminating equipment interfaces. This paper discusses the key parameters and definitions for return HFC associated with this approach. It also shows how to relate this technology to performance parameters of interest for proper design of HFC return path systems.

References

Gray, G. and G. Zeoli, "Quantization and Saturation Noise Due to Analog-to-Digital Conversion", IEEE Transactions on Aerospace and Electronic Systems, Jan. 1971.

Acknowledgement

The author is grateful for the contributions of his colleague, Ronald Cambridge. with this approach. It also shows how to relate this technology to performance parameters of interest for proper design of HFC return path systems.

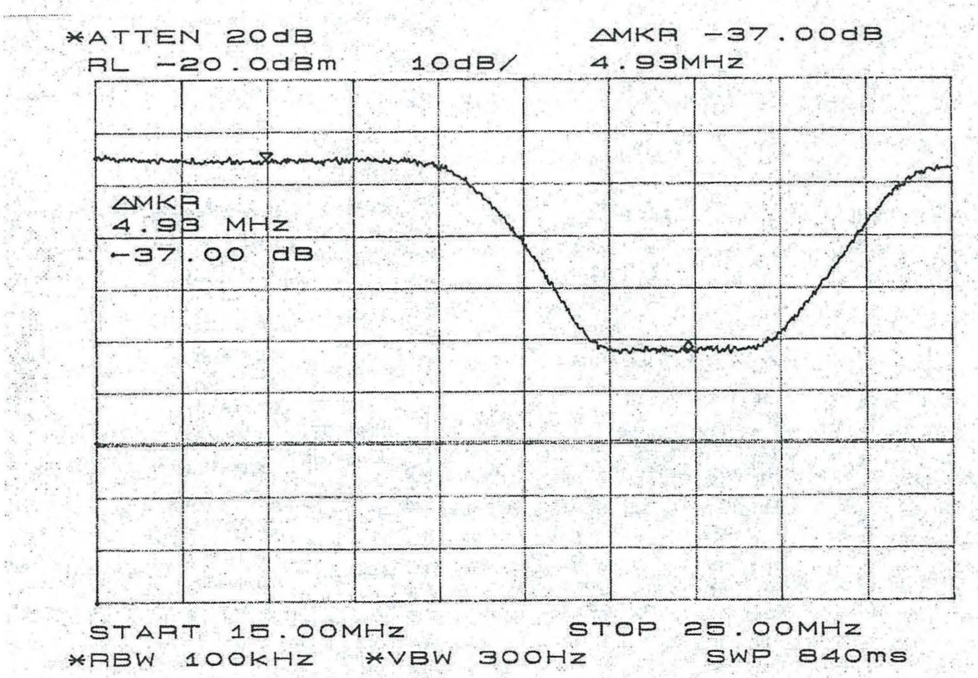


Figure 6 – NPR Performance of an 8-bit A/D

An Experimental Study of the Return-Path Performance of a DOCSIS System

Thomas K. Fong, Ph.D.
@Home Network

Bill Kostka
Cable Television Laboratories, Inc

John Chen
Cisco Systems, Inc

Abstract

This paper investigates the return-path performance of a DOCSIS system (Cisco uBR7223) over an experimental two-way cable testbed. The packet error rate (PER), as well as the downstream and upstream TCP throughputs are studied as a function of upstream noise levels. For our DOCSIS upstream burst profile, the Carrier-to-Noise Ratio (CNR) threshold for one percent packet loss was found to be 11.5dB. The corresponding TCP throughputs for a cable modem with an upstream bandwidth of 1.6MHz were 4 Mbps downstream and 1Mbps upstream, respectively. We also evaluated the effectiveness of the forward-error correction code (Reed-Solomon) in recovering corrupted bits. With the maximum value of T (T=10), the upstream CNR requirement improved by 3 to 4dB.

1. Introduction

Cable modems have become one of the highest profile communication appliances in the race to provide high-speed residential broadband services. This is a result of the exponential growth of the Internet and the development of the interoperable Data Over Cable System Interface Specification (DOCSIS) [1].

Historically, the cable network was optimized for downstream analog video broadcast. Sending data reliably in the reverse or upstream direction through the hybrid fiber coax (HFC) cable network poses a challenge to many cable service providers. The condition of the return-path varies greatly from traditional cable networks to newly upgraded HFC plant [2]. Recognizing this fact, DOCSIS allows operators to choose system parameters for their upstream channels that are suitable for their cable plants.

In this paper, we investigate the impact of upstream additive noise on the packet performance of both the upstream and downstream directions over an experimental two-way testbed. We also study the tradeoffs in the best utilization of the precious upstream bandwidth and power resources in delivering different digital services in a reliable and cost-effective manner. To assist cable service providers in facing these choices, we present an example to demonstrate the ability to tradeoff channel capacity with signal requirements. In the examples, we demonstrate that by changing the forward error encoding, the packet performance can be improved at the cost of useful capacity.

The rest of this paper is organized as follows. Section 2 presents an overview of the DOCSIS standard. In section 3, we outline the measurement techniques utilized for the experiment. Section 4 shows the performance of a DOCSIS system for a particular DOCSIS upstream burst profile. In section 5, we illustrate the tradeoff between bandwidth efficiency and impairment robustness of the upstream channel by changing forward error encoding. Section 6 contains the conclusion.

2. The DOCSIS Standard

DOCSIS is a series of interface specifications that permit the design, development and deployment of data-over-cable systems on a uniform, open and multi-vendor interoperable basis. The intended service is to provide transparent bi-directional transfer of Internet Protocol (IP) traffic, between cable system headend and subscriber locations, over an all-coaxial or HFC cable network. For two-way cable systems, the DOCSIS RF specification defines the RF communications path over the HFC cable system between the cable modem, and the cable modem termination system (CMTS) [1]. This specification includes the physical layer, link layer (MAC and LLC), and network layer aspects of the communication interfaces. It also includes specifications for power level, frequency, modulation, coding, multiplexing, and contention control.

In the upstream direction, cable modems transmit signals in a predefined portion of the upstream passband that is between 5 and 42 MHz. The modulation format and type for upstream channels are specified by the CMTS. The CMTS broadcasts periodically the parameters and frequencies for the upstream transmission to all the cable modems.

Recognizing that the quality of the upstream spectrum varies greatly in different cable plants, DOCSIS provides error correction encoding and a wide range of modulation types and formats for cable service operators to choose. An important feature in the upstream burst profile is error correction. DOCSIS has selected Reed-Solomon (RS) codes for its error detection and correction schemes.

Upstream traffic is error-protected using RS coding. Upstream packets are inserted in RS forward error correction (FEC) block codes. Block coding adds redundant parity bytes, which are used to detect and correct errors. For an RS code using GF[1](256), the number of parity bytes added is 2*T, where T is the number of bytes to be corrected [3]. The length of codewords is programmable based on shortened codes over GF(256). For DOCSIS systems, the CMTS specifies the strength of the FEC to be used by cable modems for upstream transmission. The FEC scheme is defined by the length of the FEC codeword and the value of T in the range of T=0 (no FEC) to T=10.

The expected theoretical performance of a DOCSIS cable modem is studied in detail by one of the Cablelabs' technical papers [4]. The paper illustrates the tradeoffs between bandwidth efficiency and upstream impairment robustness for DOCSIS modems and many commercially available modems. In the following sections, the experimental performance of a DOCSIS system is presented.

3. Measurement Techniques

In this section, we describe the experimental setup and measurement procedures. Section 3.1 reviews the characteristics of the upstream impairment and the performance parameters being measured while in section 3.2, the details of the experimental setup is revealed. Section 3.3 discusses the impact of error encoding on upstream performance.

3.1 Impairment and Performance Measures

For upstream channels, white Gaussian noise is generated by random thermal noise from 75-ohm termination impedances. The white noise from all terminators in all the amplifiers in the cable system is funneled into the headend in the return channel. This funneling effect exacerbates and increases the noise as it propagates through more amplifiers and connections in the cable network toward the headend. This noise is carried through each return amplifier, which adds its own noise contribution to the headend. All distribution points have their own 75-ohm terminator, and hence each leg adds its own noise contribution to the return system. This influence of additive white Gaussian noise (AWGN) is usually characterized by the Carrier-to-Noise Ratio (CNR). In this study, AWGN is simulated using a broadband programmable noise generator and the CNR is measured using a spectrum analyzer.

Since DOCSIS emphasizes delivering IP connectivity to residential subscribers, we focus our study on the performance of IP data packets (both TCP and UDP). In our experiment, we measure the performance of the following parameters:

- Packet Error Rate (PER): the number of packets dropped due to noise corruption. A packet is dropped at the receiver if the receiver cannot correctly decode all the bits in the packet.

  The upstream PER affects all traffic (TCP, UDP and DOCSIS management messages) traveling in the upstream direction. Losses of data and management frames degrade the ability of modems to transmit and receive data. Modems will drop off from the CMTS under severe loss of management messages.

- Transmission control protocol (TCP) data transfer rate, in Megabits per second, for both downstream and upstream channels. TCP throughput directly reflects on the

[1] Galois field: codes are constructed from fields with a finite number of elements. A finite field with q elements is denoted by GF(q) [3].

user's experience for applications such as HTTP and FTP [5]. For upstream channels, upstream impairments cause corruption of data packets, and hence reduce the data throughput. On the other hand, the return-path impairment degrades the downstream data transfer by causing losses of TCP acknowledgement packets and DOCSIS management messages.

TCP reacts to packet losses due to noise corruption in the same way as it does during network congestion, i.e., it slows down the transmission even though the network is not overloaded. This results in a decrease in TCP's performance. Details of the behavior of TCP over lossy links are beyond the scope of this paper. More information can be found in a paper by Cohen and Ramanathan, which analyzed the performance of TCP under different network conditions using computer simulation and proposed many methods of tuning TCP parameters [6].

3.2 Experimental Setup

Figure 1 shows the block diagram of the experimental setup. The CMTS is a Cisco uBR7223 router with an MC16 line card running Cisco IOS® software version 12.0. The CM is a Cisco uBR904 residential access router running Cisco IOS® software version 11.3(6)NA.

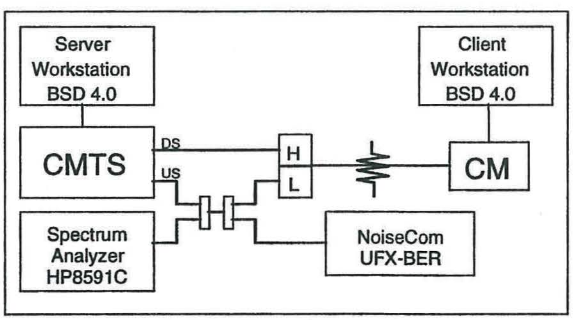


Figure 1: The Experimental Setup.

The Cisco uBR7223 series is a high-speed router with DOCSIS CMTS functionality. As a router, it has a 150K packet per second switching capacity and 600 Mbps backplane capacity. The main processor is a MIPS R4700 200-MHz CPU with 1MB of fast packet SRAM and 128 MB of DRAM. It connects to the server workstation via its 100BT interface. As a CMTS, the RF interfaces have an aggregate capacity of up to 80 Mbps in 2 downstream channels and a capacity of 120 Mbps in 12 upstream channels. The RF interfaces of a larger capacity router, the uBR7246 series, have an aggregate capacity of 160 Mbps in 4 downstream channels and 240 Mbps in 24 upstream channels. The uBR7200 series routers have passed CableLabs' CMTS qualification testing.

The Cisco uBR904 is a residential access router with a DOCSIS cable modem interface and an integrated 4-port Ethernet hub. The microcontroller is a 68EN360 33-MHz CPU with 8 MB of DRAM and 4 MB of flash memory. The flash memory is used to store remotely upgradeable firmware. The client workstation is connected to a port of its Ethernet hub.

The NoiseCom, Inc.'s programmable precision C/N generating instrument, UFX-BER acts as a broadband white noise generator. RF impairment is inserted in the upstream channel via a two-way RF combiner. CNR and power levels are measured using the spectrum analyzer following NCTA recommendations for upstream measurements [7].

In Table 1, the RF parameters for the downstream and upstream channels are shown. The center frequency for the downstream channel is 601.25MHz and 64-QAM modulation is selected. The downstream power and CNR is optimized so that the downstream RF channel is almost lossless. For the upstream channel, the center frequency is 31.184MHz. The upstream channel transmits at 1.28Mega-symbol/second and has a RF bandwidth of 1.6MHz. As shown in **Figure 1**, the downstream and upstream channels are separated with a 40/52 diplex filter.

Table 1: RF Channel Parameters

Downstream	
Center Frequency	601.25MHz
Modulation	64 QAM
Interleaving	I=32, J=4
Upstream	
Center Frequency	31.184MHz
Modulation	QPSK
Bandwidth	1.6MHz
Symbol Rate	1.28Mbps
Minislot Size	8 ticks

Two personal computers running Berkeley Software Distribution (BSD) UNIX 4.0 with 400MHz Pentium II processors and 64MB of

RAM are used as traffic source and sink. *Netperf* is used to generate and receive IP traffic. It is the de-facto standard for network performance measurement within the Internet community. *Netperf* can be used to measure various aspects of networking performance [8]. Its primary application is to measure bulk data transfer and request/response performance using either TCP or UDP.

After each test, *Netperf* reports the number of packets received and the data throughput. For UDP tests, it calculates the PER by dividing the number of packets lost by the number of packets transmitted while for TCP tests, it obtains the data transfer rate by measuring the number of bytes received over a fixed time interval.

In Table 2, the DOCSIS burst profile for this experiment is shown. It specifies the transmission formats for the different information elements (IE), including both control and data frames. Request is used by cable modems to initiate data transmission to the CMTS. Initial and Station Maintenance are management messages for maintaining communications between cable modems and the CMTS. Short and long data grants are used for carrying data packets. The distinction between them is the amount of data that can be transmitted in the grant. A short data grant uses FEC parameters that are appropriate to short packets while the FEC parameters of a long data grant takes advantage of greater FEC coding efficiency.

Table 2: Upstream Physical-Layer Burst Attributes.

Parameter	Configuration Settings	Request	Initial Maintenance	Station Maintenance	Short Data Grant	Long Data Grant
Modulation	QPSK, 16QAM	QPSK	QPSK	QPSK	QPSK	QPSK
Differential Encoding	On/Off	Off	Off	Off	Off	Off
Preamble Length	Up to 1024 bits	64	128	128	72	80
Preamble Offset	1024 config. settings	56	0	0	48	40
FEC Parity Bytes (T-bytes)	0 to 10	0	5	5	5	10
FEC Codeword Information Bytes (k)	1 to 255	16	34	34	78	235
Scrambler Seed	15 bits	338	338	338	338	338
Maximum Burst Size	0 to 255	1	0	0	7	0
Guard Time Size	5 to 255 symbols	8	48	48	8	8
Last Codeword Length	Fixed or shortened	Fixed	Fixed	Fixed	Fixed	Fixed
Scrambler On/Off	On/Off	On	On	On	On	On

Here is an explanation of the values in the burst profile:

- Modulation type is used to select between either two bits per modulation symbol (QPSK) or four bits per modulation symbol (16-QAM). QPSK carries information in the phase of the signal carrier only whereas 16-QAM uses both phase and amplitude to carry information. 16-QAM requires approximately 7dB higher CNR to achieve the same bit error rate (BER) as QPSK but it transfers information at two times the rate of QPSK. Only QPSK was used in this test.
- Differential Encoding is a technique wherein the information is transmitted by the phase change between two modulation symbols instead of by the absolute phase of a symbol. This technique makes the absolute phase of the received signal insignificant and has the effect of doubling the BER for the same CNR. Differential encoding is not required for the CMTS receiver used in this test so it is turned off.
- Preamble Length and Preamble Offset are used to define a synchronizing string of modulation symbols that is used to let the receiver find the phase and timing of the transmitted burst. A unique word in the

preamble is used to signify the start of the data portion in the burst.

- FEC Parity Bytes (T-bytes) is the number of bytes that the FEC decoder can correct within a codeword. A codeword consists of information bytes, called k-bytes and parity bytes for error correction. The number of parity bytes is equal to two times the number of correctable error (T). The size of T is dictated by channel impairments.
- FEC Codeword Information Bytes (k-bytes) for short data grant is set to 78 bytes to accommodate a minimum packet within one codeword. The minimum packet size = 64 bytes of Ethernet packet + 6 bytes of DOCSIS MAC header + up to 8 bytes of DOCSIS MAC Extended header (EHDR) typically used for baseline privacy. Information size (k-bytes) for a long data grant is set to the largest block size possible to minimize the number of FEC blocks needed to transmit a packet upstream. In more impaired upstream environments, the information size may be set to less than the largest possible block size to enable the parity bytes to act over a smaller number of information bytes and thus provide better data protection.
- Scrambler is used to create an almost random sequence of transmission symbols, which ensures an even spectral distribution of energy transmitted within the channel. The scrambler seed is an initial value that is used to start the pseudo-randomizer to "scramble" the bits. Because both the transmitter and receiver know the seed value, the scrambling can be reversed at the receiver leaving only the original data.
- Maximum Burst Size is only used to determine the breakpoint between packets that use the short data grant burst profile and packets that use the long data grant burst profile. If the required upstream time to transmit a packet is greater than this value then the long data grant burst profile is used. If the time is less than or equal to this value, then the short data grant burst profile is used.
- Guard Time is a blank time at the end of a burst transmission that exists to ensure that one burst ends before another burst starts.
- Last Codeword Length shortened enables an efficiency mode wherein all codewords except the last one are fixed in size. The last codeword may be shortened if there are not enough information bytes to fill it entirely. In a fixed operation, all codewords are the same size with the last codeword padded with nulls if there are not enough information bytes to fill it entirely. The efficiency is gained by not having to transmit the nulls that pad the last codeword.

In addition, a few MAC (media access control) layer parameters are fine-tuned so that the physical layer performance is measured accurately:

- Polling Interval: the CMTS grants a unicast Station Maintenance transmit opportunity to each cable modem, at least once every 30 seconds [1]. Each cable modem has a timer, the T4 timer, to keep track of time between station maintenance transmit opportunities. When the T4 timer expires, the cable modem is required to re-initialize its MAC layer. To reduce the effect of MAC re-initialization during a measurement, the CMTS is set to transmit a Station Maintenance message every second.
- SYNC and UCD Message Rate: Time Synchronization (SYNC) and Upstream Channel Descriptor (UCD) messages are broadcast downstream to all cable modems for initial upstream timing and frequency acquisition. We found that the rate of these messages does not affect the return-path performance measurement, and hence the vendor's default values were used for this experiment.
- Packet Transmission Rate: For the PER measurements, the source sends packets at a low transmission rate so that packet loss is caused by RF noise corruption and is not affected by other factors, such as buffer overflows and CPU workload, in the CMTS or CM.

Tables 3, 4 and 5 present the overhead in burst transmission for three frame sizes using the burst size described in Table 2. 64bytes is the minimum MAC frame size and indicates the performance of TCP acknowledgement, which is usually the dominant type of traffic in the upstream channel for a typical residential subscriber. The MAC frame sizes of 594 and 1518 bytes correspond to IP datagram sizes of

576 and 1500 bytes. These two sizes are commonly used for carrying Internet traffic.

Table 3 shows the forward error correction (FEC) overhead. For 64 byte MAC frames, a FEC T-bytes value of 5 is used. The value of T is increased to 10 for long grants to provide added protection for larger FEC blocks using 594 and 1518 byte MAC frame sizes. For short data grants, the transmission overhead is proportionally higher and the FEC scheme is less efficient for the same value of T. The Pre-FEC Bytes row includes the 6-byte DOCSIS Physical Media Dependent (PMD) header. If baseline privacy is enabled, this header increases to 11 bytes.

Table 3: Post-FEC Bytes, Upstream.

Frame Size, bytes	64	594	1518
Pre-FEC, bytes	70	600	1524
Grant Type	Short	Long	Long
No. FEC Blocks × FEC k bytes	1×78 = 78	3×235 = 704	7×235 = 1645
No. FEC Blocks ×2×FEC T bytes	1×2×5 = 10	3×2×10 = 60	7×2×10 = 140
Post-FEC, bytes	88	764	1785

Table 4 shows the overhead associated with the TDMA bandwidth allocation for a mini-slot size of 16 bytes. The overhead may be reduced by a smaller mini-slot size of 8 bytes. However, smaller mini-slots may require more processing workload for the CMTS. Each upstream transmission begins with a preamble pattern, followed by FEC blocks, and ends with a guard time interval. The last row lists the actual number of bytes occupied by an integral number of mini-slots.

Table 4: Mini-Slot Bytes, Upstream.

Preamble, bytes	9	10	10
Post-FEC, bytes	88	764	1785
Guard Time, bytes	2	2	2
Time-Slot, bytes	99	776	1797
Bytes/Mini-Slot × Mini-Slots	16×7 = 112	16×49 = 784	16×113 = 1808

Lastly, the channel overhead is computed by dividing the overhead bytes by the MAC frame size. It represents the percentage of overhead in transmission of a data packet. From Table 5, we can see that the channel overhead for transmitting short data packets in the upstream direction is very high. 112 bytes are needed for each TCP acknowledgement packet.

Table 5: Channel Overhead, Upstream.

Mini-Slot, bytes	112	784	1808
MAC Size, bytes	64	594	1518
Overhead, bytes	48	190	290
Channel Overhead	75%	32%	19%

3.3 The impact of FEC

DOCSIS provides the flexibility to modify the robustness of the upstream RF channels. Here are some techniques:

- increase the amount of error encoding (shorter codewords and larger values of T).
- decrease the upstream symbol rate to reduce the impact of burst noise. This reduces the relative duration of the burst noise.
- increase the size of the preamble to provide better bit synchronization.
- increase the amount of interpacket spacing.

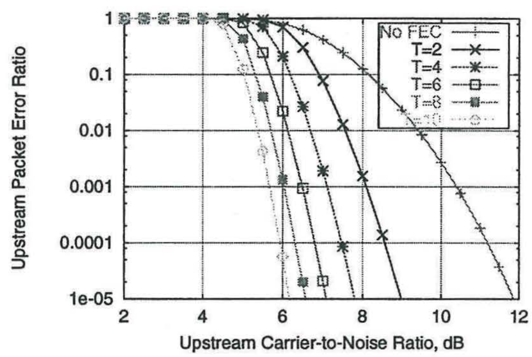


Figure 2: PER vs. CNR for DOCSIS QPSK with different values of T (packet size=46 bytes).

Figure 2 shows the theoretical PER performance for different T values. The different traces represent the range from no FEC to T=10. The maximum amount of FEC established in DOCSIS is T=10. From Figure 2, the coding gain at 1% PER is approximately 2dB for T=2 and 5dB for T=10. On the other hand, more parity bytes reduce the information rate. The traces begin to close together with higher values of T. Increasing the amount of FEC causes a rate of diminishing return. By using only T=2, about half of the possible coding gain is achieved. In Section 5, we illustrate the same relationship for a DOCSIS system in a laboratory setup.

4. Measurement Results

This section presents the measurements for the DOCSIS burst profile listed in Table 2.

4.1 Packet Error Rate

For the PER experiments, *Netperf* is configured to generate and receive UDP packets. The Packet Error Rate is obtained by dividing the number of packets lost by the total number of packets transmitted. Figure 3 shows the PER performance for packet size=46, 576 and 1500 bytes. The PER behavior for 576 byte packets and 1500 byte packets are very similar since they are both carried by the long data grant. The PER for 1500 byte packets is slightly higher than that for 576 byte packets as longer packets have higher probabilities of getting a bit error than shorter ones.

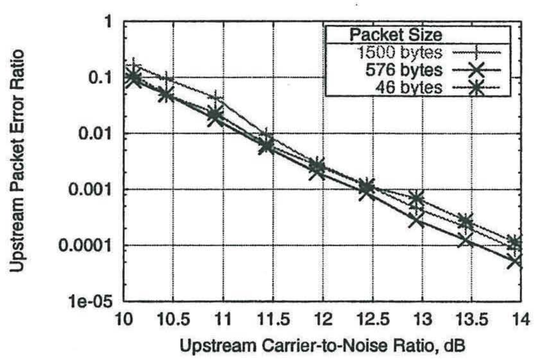


Figure 3: Packet Error Rate vs. Carrier-to-Noise Ratio for packet size=46, 576, 1500 bytes.

Note that the PER degrades very quickly as the CNR decreases (at about one decade per decibel of CNR). That means, a higher CNR margin is needed to ensure reliable data services against statistical fluctuations of CNR. For one percent packet loss, a CNR of about 11.5dB is needed.

The difference between the theoretical calculation in Figure 2 and the experimental measurement in Figure 3 is mainly due to the burst-mode characteristics of the upstream channel, the MAC layer overhead and other implementation losses. The theoretical curves are generated with the assumption that the bit synchronization for the upstream receiver is perfect and the MAC protocol does not incur any physical layer penalty.

4.2 TCP Throughput

For TCP tests, the source computer is configured to continuously send data over a fixed duration of time. The TCP throughput is obtained by dividing the number of bits received successfully at the receiver by the duration of the test. The TCP throughput is studied as a function of upstream CNR for different TCP window sizes.

4.2.1 Upstream TCP Performance: the impact of data packet loss

As seen in Figure 3, packet loss in the upstream channel increases as the noise level increases. Since TCP is a reliable data transport, the lost data packets are retransmitted using the TCP's sliding window protocol. Hence, an increase in the noise level reduces the downstream transfer rate.

Figure 4 shows the upstream TCP performance for different TCP window sizes (8, 16, 24, and 32Kbytes). We see that the TCP throughput starts to degrade when the CNR is less than 12.5dB. This means that PER of less than 0.1% does not penalize TCP performance.

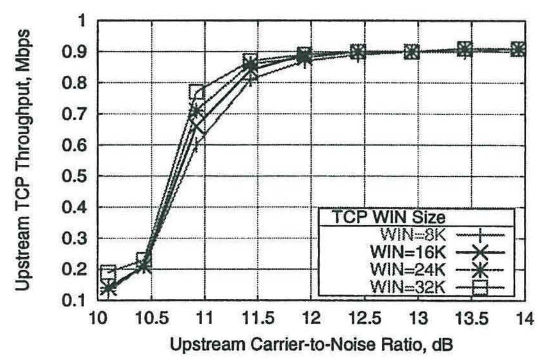


Figure 4: Upstream TCP throughput vs. US CNR for TCP window = 8, 16, 24, 32Kbytes.

As the PER increases above 0.1%, the TCP throughput degrades very quickly. At about 10% PER (CNR=10dB), the TCP throughput drops to less than 10% of its maximum value. Larger TCP window sizes do not improve the performance in a noticeable way.

4.2.2 Downstream TCP Performance: the impact of acknowledgement packet loss

For downstream TCP transfers, the upstream noise affects the performance in a completely different way as for the upstream data transfer in the previous section. Upstream noise causes loss in TCP acknowledgement packets. It reduces the downstream transfer by slowing down the growth of the TCP congestion window and causing TCP retransmission. In addition, it also affects the ability of cable modems to request bandwidth for transmission.

Figure 5 shows the downstream TCP performance for different TCP window sizes (8, 16, 32 and 64Kbytes). Since TCP uses positive acknowledgements, occasional loss of acknowledgement packets does not cause any penalty to TCP sessions with large TCP received window sizes. 3% degradation in throughput is observed for CNR=11.5dB (PER~1%) when the TCP window size is 16, 32 or 64 Kbytes. For small window size (8 Kbytes), the same loss rate causes 15% throughput degradation.

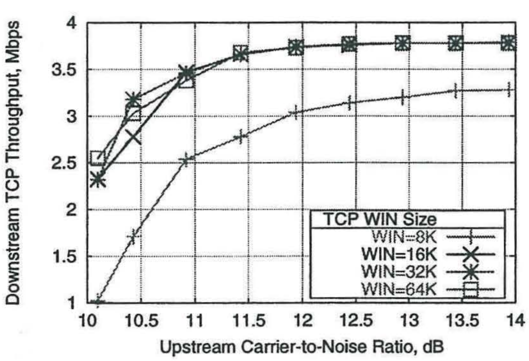


Figure 5: Downstream TCP throughput vs. US CNR for TCP window = 8, 16, 32, 64KBytes.

When the acknowledgement packet loss rate is greater than 1%, the TCP performance starts to degrade quickly. For 5% PER (CNR=10.5dB), the throughput is degraded by 50% for small TCP window sizes and 20% for large TCP window sizes.

5. Burst Profile Design Tradeoffs

This section evaluates the tradeoff between bandwidth efficiency and impairment robustness for upstream channels. The PER performance is measured for different values of T.

Figure 6 shows the PER curves for 1500 byte packets with T=0 (no FEC), T=2 and T=10. At one percent packet loss, the coding gains for T=2 and T=10 are 2dB and 4dB, respectively. As discussed in Section 3, half of the possible coding gain for DOCSIS is achieved using only T=2. Also, note that, the coding gain is higher for lower values of PER.

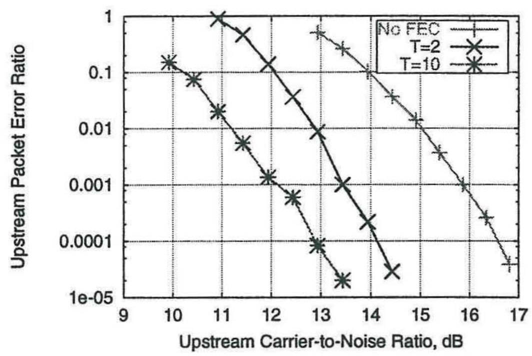


Figure 6: Packet Error Rate vs. CNR for different T values (packet size=1500bytes).

In Figure 7, the PER curves for TCP acknowledgement traffic (packet size=46bytes) is plotted for T=0 (no FEC), T=2 and T=10. The coding gains for T=2 and T=10 are 2dB and 3dB, respectively. Since short packets are less prone to packet loss than long ones, the FEC gain is also smaller.

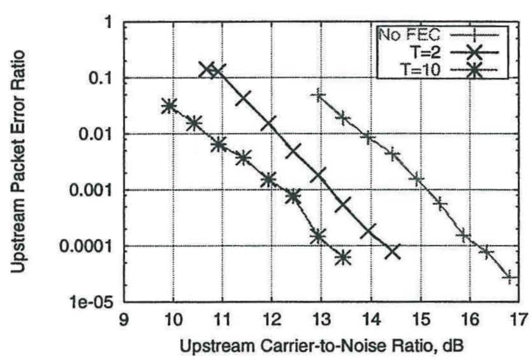


Figure 7: Packet Error Rate vs. CNR for different T values (packet size=46bytes).

Since DOCSIS provides two data grant types for carrying data packets, FEC parameters can be individually optimized for different packet sizes.

6. Conclusion

In this paper, we showed experimentally the impact of upstream additive white Gaussian noise (AWGN) on the packet error and TCP performance of a Cisco DOCSIS CMTS for different traffic types.

For the Cisco uBR7223, the upstream packet error rate (PER) decreases at approximately one decade per decibel of upstream carrier-to-noise ratio (CNR). The CNR requirement for one percent packet loss was found to be 11.5dB. It is possible to vary the signal requirement by changing the parameters in the upstream burst profile. This allows cable operators to tradeoff upstream robustness with data capacity. In this paper, we demonstrated the ability to modify the CNR performance by changing the FEC T values. For the Cisco uBR7223 using the maximum value of T specified by DOCSIS (T=10), the measured CNR improvement is about 4 dB. On the other hand, since PER degrades very quickly as CNR decreases, a higher CNR *safety* margin (at least 5dB) is needed to ensure reliable data services against fluctuations of power and noise levels.

Both the upstream and downstream TCP throughputs are dependent on the upstream CNR. For upstream TCP transfers, the throughput dropped 10% from the maximum transfer rate at CNR=11.5dB. At CNR=10dB, the upstream transfer rate fell to about 10% of the maximum value.

For downstream TCP transfers, the throughput is degraded by loss of TCP acknowledgement packets and DOCSIS management messages. Tuning the parameters of the TCP stack can reduce the impact of upstream noise on downstream TCP transfer. A larger TCP window results in a greater number of acknowledgement packets per transmission window. This reduces the probability of a timeout due to a loss of all the acknowledgement packets and thereby, increases the robustness of the TCP session to acknowledgement packet loss. In this paper, we illustrated that the throughput degradation at CNR=11.5dB is reduced from 50% (with 8Kbyte window) to 20% (with 16, 32 or 64Kbyte windows).

Using our DOCSIS upstream burst profile, the TCP throughputs at CNR=11.5dB for a computer connected to the cable modem system are 4Mbps downstream and 1Mbps upstream, respectively.

7. Acknowledgment

The authors would like to thank Jonathan Min and Bruce Currivan of Broadcom, Inc for their valuable comments on this paper.

8. References

[1] "Data-Over-Cable Service Interface Specifications: Radio Frequency Interface Specification," SP-RFI-104-980724, Interim Specification, Cable Television Laboratories, Louisville CO, July 24, 1998. http://www.cablemodem.com

[2] D. Raskin and D. Stoneback, *Broadband Return Systems for Hybrid Fiber/Coax Cable TV Networks*, Prentice Hall PTR, 1998, Chapter 2.

[3] J. Proakis, Digital Communications, McGraw-Hill, Inc, 1995, Chapter 5.

[4] CableLabs, "A Comparison of Upstream Modulation Formats," Cable Television Laboratories, Louisville CO, March 1998. http://www.cablelabs.com.

[5] W. R. Stevens, TCP/IP Illustrated, Volume I: The Protocols, Addison-Wesley, 1996, Chapter 21.

[6] R. Cohen, and S. Ramanathan, "TCP for High Performance in Hybrid Fiber Coaxial Broad-Band Access Networks," IEEE/ACM Transactions on Networking, vol. 6, no. 1, February 1998, pp. 15-29.

[7] "NCTA Recommended Practices for Measurements on Cable Television Systems", 2nd Edition (revised 1993), Supplement on Upstream Transport Issues, National Cable Television Association, Washington DC, October 1997.

[8] R. A. Jones, "Netperf: A Network Performance Benchmark, Revision 2.1," Information Networks Division, Hewlett-Packard Company, February 1996. http://www.netperf.org

[9] M. St. Johns, "Radio Frequency (RF) Interface Management Information Base for MCNS/DOCSIS compliant RF interfaces," IETF, February 1999. http://www.ietf.org

Blended Technologies
Operating in the World of HFC, Wireless, DSL, and Other Transport Technologies

Jeff Tokar
High Speed Access Corp.

ABSTRACT

Now that the standards for DOCSIS compliant cable modems have been finalized and the compliant modems are starting to emerge, cable operators are in a position to expand an early lead in deploying high-speed Internet and data services, establish a significant presence in these new markets, and begin generating revenues in anticipation of competition from telephone, satellite and wireless cable alternatives. Unfortunately a large fraction of the installed cable plant is still either unsuitable for 2-way high-speed communications, or deployed primarily in residential areas bypassing potential commercial markets where these services are highly coveted and have the potential to generate the highest revenues. By expanding upon the technologies within their arsenal, cable operators can take the lead in almost every market today, while gaining the flexibility to expand out the basic hybrid fiber/coax (HFC) technology as revenues and schedules dictate.

INTRODUCTION

Blended technologies promise to reduce deployment costs, and enable deployment beyond the traditional range of cable plants into many non-traditional markets. They can also provide a time to market advantage so that cable operators can establish a beachhead before the competition has a chance to deploy service. Not only will this improve the chances that cable will become the dominant player in a given market, it will also mean that cable operators will be able to start generating revenues, and gathering operational experience sooner.

As the initial new low-cost high-speed data provider, cable operators will be in a position to demand a premium for their services to help fund the development of the basic infrastructure. Once these costs have been paid, cable operators will be in a better position to face the competition from telcos, satellite providers, wireless cable, and others.

One of the primary limitations of many cable systems is that HFC distributions plants have been built into residential neighborhoods to deliver consumer oriented entertainment, while the business districts represent the highest potential revenues for high-speed data. Even when cable operators decide that that they will build out to business districts and industrial parks, it may prove economical to bring these businesses online and generating revenues within a few weeks, rather than waiting until the cable plant has been extended into the right location.

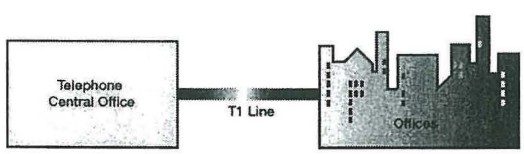
Blended architecture could also play a role in extending services into large office buildings or Multiple Dwelling

Units (MDUs), which may be difficult to wire with coax owing to costs, limited duct space, and/or restrictions on extending the cable plant onto the premises.

THE DAWN OF HIGH SPEED DATA

In the traditional high-speed access model, businesses have paid high prices for special T1 (1.5 megabit/s) or T3 (45 megabits/s) lines to be brought to their premise. These services have been offered by a combination of local telephone companies, long distance companies and competitive access providers. Prices for T1 lines can average $1000 per month including equipment rental and Internet access, although they tend to be cheaper in concentrated markets such as major cities with significant carrier competition.

Traditional High-Speed Data



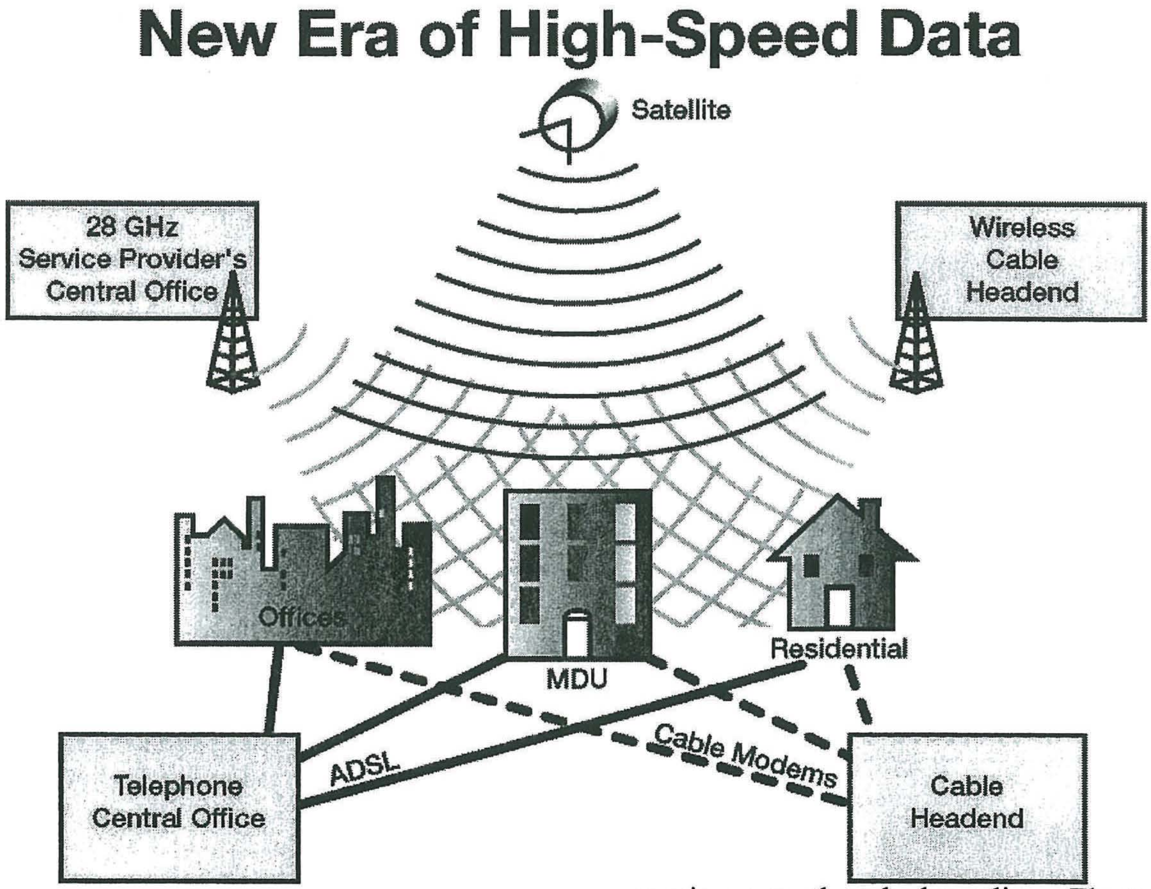
Consumers traditionally have been restricted to much lower speed alternatives that operate over more dated telephone lines. These alternatives have included modems that operate at speeds of up to 56 kilobits/s and more expensive ISDN lines that support speeds of up to 144 kilobits/s (2-B channels at 64 kilobits/s and 1-D channel at 16 kilobit/s). This has restricted consumer oriented Internet access primarily to text, video, and low-quality audio and video applications.

The last mile is in a period of rapid transition thanks to widespread interest in the Internet, and the development and deployment of new technologies for delivering data over cable TV plants, telephone networks, and wireless technologies. This is being paralleled by efforts to develop a high-speed, highly reliable backbone that will be suitable for high-quality telephony, CD-quality audio, video conferencing, entertainment television, and 3D-streaming.

The advent of these new lower cost networking approaches promises to consume the high-cost data services that exist today, and create a new market for higher-speed services which demand a higher quality-of-service. In the future consumers and businesses will regularly take advantage of the Internet for making phone and video calls, watching TV, and being immersed in streaming 3D. Streaming 3D will enable viewers to do things such as watch and replay football games from any angle they choose, or navigate through complex 3D atlases using a novel new type of interface. This technology is not just science fiction. In fact the Web 3D consortium plans to bundle the X3D standard (which is still under development) with the next generation HTML standard that is due out by the end of 1999, and to make this software freely available for incorporation into Web browsers and Internet TV appliances.

Cable operators will be delivering high speed data services over a combination of hybrid fiber/coax, wireless, and telephone wire infrastructure that will generally offer speeds up to 27 megabits/s downstream, and up to 10 megabits/s upstream, which can be shared among multiple users. For cable systems that are not yet equipped to handle bi-directional traffic, the upstream data could be carried over a regular modem. But this would operate at significantly lower speeds and would require the customer to use up a telephone line when connected to the Internet. Although this would still offer consumers high-speed access, it would mean that customer would have to pay for the extra telephone line, which would mitigate some of the added benefit of having a cable modem.

Alternatively, it may be possible to send the upstream signal over some kind of wireless signal, such as unlicensed wireless in the 900 MHz, 2.4 GHz, or 5 GHz bands, which would obviate the need for an extra telephone, and keep the telephone company out of the equation completely.



Cable companies might also run HFC plant to large buildings, such as offices or MDUs, and then use wireless, powerline, or phoneline (over lines owned by customer) networking to provide the last link to the premises.

Another group of Internet Service Providers, which may include the telephone company or others that simply lease out the telco wiring are running Asymmetrical Digital Subscriber Line services over the telephone lines. These technologies will be used by telcos to provide a new type of integrated voice, video and data service. A number of major telcos including AT&T, Sprint, and Cable & Wireless (which acquired MCI's data networking assets in the US after the merger with WorldCom) are planning integrated broadband services using such an architecture.

Wireless cable operators are also breaking into the market for high-speed data delivery by inserting a data delivery channel into an empty TV channel. They will be able to take advantage of some of the work that has taken place in the traditional cable world. For example, Hybrid Networks, which cut its teeth in the traditional cable world, has developed a technology which can send data at up to 10 megabit/s downstream and use either the telephone plant, or wireless at speeds ranging from 256 Kilobits/s to 2 Megabits/s in general and 5 megabits/s upstream for special applications. Data can be transmitted downstream over MMDS, MDS, WCS, and UHF wireless frequencies.

Meanwhile another group of 28 GHz telephone and data service providers is springing up to take advantage of the immense capacity available in this band (1-GHz). Many of these operators, such as Teligent, have already started rolling out commercial service with speeds of up to 45 Megabits/s. Faster speeds are possible due to the 28 GHz band's extremely tight radio waves, which enables the radio energy to be highly focussed, and thus does not create much interference between nearby users.

The satellite operators are also getting into the picture. The Hughes DirecTV venture has launched the DirecPC service, which is already capable of supporting speeds of up to 400 kilobit/s downstream. However, upstream service must go through a traditional analog modem. Teledesic is in the process of creating a low earth orbiting satellite that will support data rates at up to 64 megabits/s downstream and 2 megabits/s upstream for basic service, although a broadband version is planned to support 64 megabits/s in both directions.

The importance in discussing all of these different approaches is that cable operators only have a limited window of opportunity to maximize revenues, and establish new customers. Today cable operators are primarily competing against a relatively expensive technology (T1 lines at $1000 a month), while in the future they will have to compete against a range of technologies that promise to deliver the same bandwidth for as little as $40-100 a month.

It is in the cable operator's best interest to ensure that it is the first and most prominent provider of this next generation of services. If cable operators can be the first to grab major customers such as businesses and small offices, it will be easier to lock customers in -- even when competition offers cheaper services in the future. Businesses and others that grow dependant on them may be hesitant to change providers when the competition emerges due to the hidden costs and uncertainties involved in changing providers in mid-stream.

APPLICATIONS OF BLENDED TECHNOLOGIES

There are a wide variety of applications in which blended technologies can play a role in expanding the high-speed data business for cable operators. These include providing a time to market advantage, extending the wiring beyond the reach of the core business area, simplifying deployment to MDUs and business districts, reducing the cost of deploying new services, consolidating headends, and reducing the load on the core broadband network.

Time to market

In many cases, blended technologies can take advantage of existing infrastructure to speed the time to market, and the time to revenue, allowing cable operators to make an immediate return on investment while building out the basic HFC plant, and establishing a customer base.

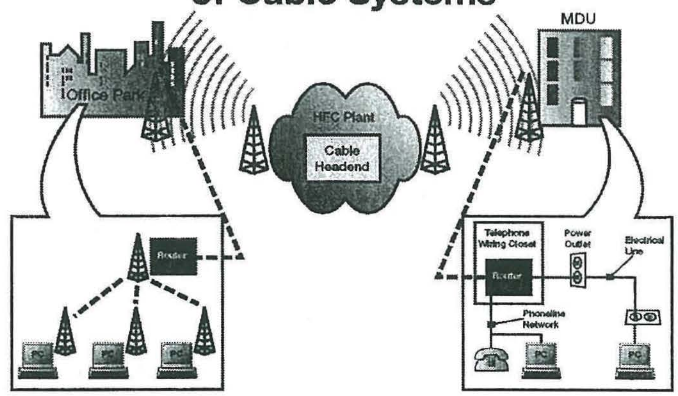
Unlicensed wireless promises to play a role in a number of scenarios. It can enable services to be offered to business parks, office buildings, and large MDUs that are up to 25 miles away from the edge of the HFC network. Short-range wireless could also simplify the deployment of services within large buildings such as offices or MDUs, without requiring any inside wiring.

An emerging new technology called phoneline networking could allow the rapid deployment of new services within a large building without requiring the participation of the local telephone company. Cable operators would simply put a router in the telephone closet, and connect to the individual customers on the customer side of the telephone line of demarcation in the building, which is the part owned by the customer.

Powerline networking can turn the power network in an office building or MDU into a giant Ethernet, which could be accessed by simply plugging into the outlet. The beauty of this approach is that it could be used in every room in the building, even those without phone wiring, or that may be shielded from radio.

ADSL could also be used to extend the reach of cable into new areas. It could be used across spare wires on a large campus without the participation of the telco. Alternatively, it might extend core service outside the range of the core cable network to a large building or office park, which could be shared among tenants. However, this approach would necessitate the leasing of lines from the telco.

Connecting Beyond Outer Reaches of Cable Systems



Connecting beyond the reach of the HFC cable system

Alternative transport technologies are cheaper to extend than traditional HFC cable plant. Several of our early wireless plant extensions are based on getting a commercial customer on line quickly while the cable operator extends the plant into the commercial area. This is a prime weapon in getting to the customer before that customer is lost to ADSL, etc.

Providing service to highly profitable business districts

In many cable systems the business districts have not been targeted because they have not had a great deal of interest in consumer oriented cable programming. However they represent the most profitable opportunity for cable operators since they are accustomed to paying more for high-speed data lines.

Businesses that have traditionally paid $1000 a month for their data connection will jump at the opportunity to have an equivalent or faster service for $200 a month. Many residential consumers are reluctant to pay more than $40-$100 unless they are a small office that can quantify and justify the value added of the expenditure.

The fact of the matter is that businesses can use the network to make money. For many potential residential consumers the Internet is perceived as primarily an entertainment phenomenon and they may be reluctant to spend too much on these services, unless they derive some concrete value like providing junior with an educational lead, or enabling dad to trade stocks with more confidence.

The other aspect of business deployments is that they make a lot more long distance telephone calls during prime-time business hours when rates are more lucrative. As cable expands into the telephony business, these added revenues are sure to play a role in the profitability of the venture. Consumers on the other hand tend to place fewer calls than businesses, and tend to make them more on nights and weekends, when rates are down.

Simplifying deployment to MDUs

Blended technologies promise to simplify the delivery of interactive services to MDUs. The MDU market has historically been tough to penetrate and hold onto. Cable operators have been required to wire up the whole facility to offer service, which has raised issues involved with management approval. This may involve issues such as expensive business arrangements that appeal to property management, or ownership of the distribution plant inside the property.

Technologies like phoneline networking would allow a cable operator to offer service by simply installing a router in the basement, and hooking phoneline modems to each line, without running a wire anywhere in the building. Wireless technology could enable cable operators to offer service without any management approval since the signal would pass through the open air, and the equipment could be installed outside of the building.

Reducing the cost of deployment of new services

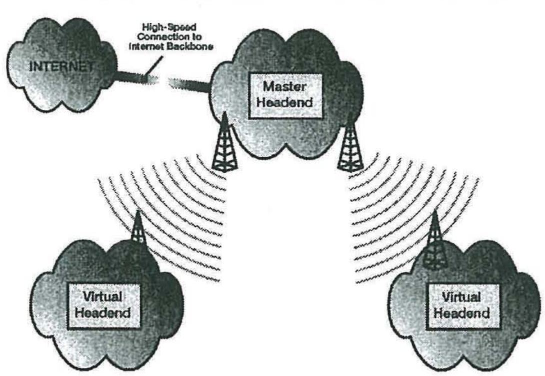
Working with existing infrastructure promises to reduce the need to build new coax plant until the revenues begin to justify the expense. This could enable cable operators to cherry pick the most profitable customers today, regardless of their location.

Blended technology could also extend the reach of cable's infrastructure into remote areas that might otherwise incur long distance tariffs for users. For example, many people are talking about using traditional analog modems to offer a return path for one way networks, or to allow workers to access their email or the web from home. A cable operator could put a virtual Point of Presence (POP) in these remote locations which connects back to the headend via wireless, thus bypassing tariffs for the customer, and hence encouraging usage and acceptance of the service.

Virtual headend consolidation

In some remote areas cable operator have established virtual headends, which simply feed satellite video into the network. Some of these may be in remote areas, which could make the prospect of deploying fiber to the area for accessing the Internet a costly proposition.

Virtual Headend Consolidation



Wireless technology could be used to link up these remote headends to a master headend that is closer to the city, and hence easier and cheaper to connect to the Internet backbone.

Different Transport Technologies

Cable operators should embrace the different technologies for extending their network and business as weapons in their arsenal. It is important to understand the basic technologies, their costs and their applications.

HFC

HFC networks provide the core of the cable network enabling the distribution of broadband data to and from the different areas that they serve. With many systems now having a total bandwidth of 450 MHz to 1 GHz, HFC promises to deliver all of the capacity that cable needs to offer a wide range of services including video, voice, and high-speed data.

However it is also one of the most expensive technologies to deploy. Costs typically range from $15,000/mile for aerial construction to $75,000/mile for underground construction. In some cases it can reach $100,000/mile if local regulations require cable operators to replace cobblestones over the cable.

Wireless

Unlicensed wireless is one of the most promising technologies for extending cable's reach into lucrative new areas beyond the reach of the core cable plant. It can be used either for point to point wireless, or over shorter ranges for enabling a single wireless station to communicate with multiple customers.

The FCC has set aside spectrum in several different bands in the 49 MHz, 900Mhz and 2.4 GHz regions for unlicensed applications that do not interfere with others, and which can adapt to interference from other devices.

Services that use unlicensed spectrum promise to be cheaper than those which do not because there is no licensing process to go through, nor spectrum to pay for. However, the growing popularity of unlicensed applications, ranging from cordless phones to Internet access, can lead to interference and impaired performance.

The FCC has also made 300 MHz available for unlicensed National Information Infrastructure devices between 5.15-5.35 GHz, and 5.725-5.825 GHz. When devices that use this spectrum do emerge, they will be able to communicate at speeds of up to 20 megabits/s over short range or several miles with large fixed antennas. The FCC rules allow devices to communicate with up to 200 mW of power between 5.15-5.25 GHz, 1 W of power between 5.25-5.35 GHz and 4 W of power between 5.725-5.825 GHz. These bands coincide with spectrum selected for the European High Performance Local Area Network (HIPERLAN), creating an international market for NII devices.

Some implementations of Europe's HIPERLAN have a throughput of up to 24 megabits/s per channel, and several such channels can be established in each geographic region. Major wireless vendors including BreezeCom, Proxim and Symbionics have developed prototype systems using NII spectrum.

For point to point applications over a long range, technologies such as Solectek's AirLAN Bridge can extend the reach of cable by up to 25 miles at speeds of up to 2 megabit/s.

We have been deploying unlicensed wireless technology from CalAmp that costs $2500 per link for a connection, or a total of $5000 to deliver a T1 equivalent to a customer. With a short range 22 degree aiming variation it has proven highly reliable, on the order of 99.99%, which exceeds AT&T's new standard of 99.95% for data services.

There are several different types of unlicensed wireless communications systems that can operate over a shorter range. For example, BreezeCom's BreezeNET, is a line of wireless radio modems that support Ethernet and can communicate at up to 3 megabits/s. The range of communications varies from 200-600 feet in an office up to 3000 feet between buildings. BreezeNET products start at a cost of under $290 per node and include access points, adapters and mini-hubs. To accommodate even greater speeds, BreezeNET can be installed in a switched Ethernet configuration that supports over 15 megabits/s of aggregate bandwidth.

Metricom has developed the Ricochet service that has already been launched in a number of markets in conjunction with power companies that support it. Ricochet has traditionally been able to support data rates of about 28 kilobits/s in the 900 MHz band. Metricom is working on a new technology, code-named Autobahn, that will support up to 128 kilobits/s. It will use spectrum in both the 900 MHz and 2.4 GHz bands.

Metricom's system is designed to plug into the existing electrical power infrastructure with little construction. It uses a network of radios connected between the solar cell controller and light at the top of many street lamps. Consequently, it is easy for a power company to set up a network. The installer merely unplugs the solar cell from the lamp, and connects the Ricochet radio between them.

WaveRider Communications has developed another wireless wide area technology that uses unlicensed radio spectrum called the Last Mile Solution. The system uses a network of base stations and repeaters, which can be spaced at a maximum of 18-km intervals. The range of the system is up to 7 km between a user and a repeater or base station. Each wireless network access point can accommodate up to 1024 simultaneous users.

The Last Mile Solution can support bi-directional data rates of 40.9 kilobits/s and can be scaled up to 10 megabit/s for high-end users. The system operates in the 902-928 MHz band, which allows unrestricted operation in the US and Canada. It is possible for the modem to operate in any band between 600 MHz and 1.2 GHz for use in other countries. It uses spread spectrum technology, which provides immunity to noise, and security from eavesdroppers.

Another option for wireless network links is to use laser beams. Laser beam communications are not licensed by the FCC because laser beams are coherent, and do not interfere with other laser beam communications. An added benefit of this is that they are substantially more secure than more traditional radio propagation technologies, making it more difficult for hackers to eavesdrop.

Systems, such as the AstroTerra's TerraLink laser can be used for communicating at speeds of up to 622 megabits/s over ranges of up to 5 miles, and 2.5 gigabits/s over ranges of up to 1.5 miles. One of the disadvantages of lasers is that they are more prone to environmental disturbances, such as heavy rain and fog than other technologies.

ADSL

Although ADSL is typically thought of as a telco only tool, it could play a valuable role in extending the range of cable delivered high-speed data services, particularly in large campuses or office building in which the wiring is owned by the customer or landlord. Typical ADSL speeds range from 1.5 -8 megabits/s downstream and significantly lower speeds upstream. It can operate over a distance of up to 18,000 feet at speeds up to 2 megabits/s and at shorter distances of up to 9000 feet can operate at speeds of 6-8 megabits/s. The downstream bandwidth is enough to provide broadcast quality digital video, which was one of the killer applications initially planned for this technology. But most telephone companies have since dropped plans for digital video and plan to concentrate on Internet access instead.

A low-end version of ADSL, called Universal ADSL or G-lite is capable of delivering speeds of up to 1.5 megabits/s downstream. One of the most appealing aspects of this approach is that it is relatively cheap to implement, and many modem manufacturers are planning on building this into their new analog modems as an optional feature that can be enabled by software. It also will operate without the need to install a splitter on the home, which would eliminate the requirement of sending out an installation technician.

By some estimates there could be as many as 6 million ADSL-lite modems installed in computers by the end of this year, although only a fraction of those will be used for service. These large volumes will make it cheap enough to be used for a variety of applications. This has implications not only for telcos, but for cable operators who may wish to use twisted pair wiring to extend the range of their services by up to 18,000 feet beyond their coaxial infrastructure, if the wiring is available.

Novel variations of ADSL promise far higher speeds. For example, VDSL (Very high-speed Digital Subscriber Line) is capable of delivering 52 megabits/s downstream and 6 megabit/s upstream over a distance of up to 1000 feet.

Phoneline networking

Phoneline networking technology is a new type of network that operates over traditional telephone wiring, without requiring any special wiring architecture. It uses frequency division multiplexing technology to enable it to share the same line voice with no interference. When installed on the consumer premise, past the telco line of demarcation, it does not require permission or coordination with the local telephone company.

It is being promoted by the Home Phoneline Networking Association (HomePNA), which includes Compaq and Intel as driving members. The current standard supports a data rate of 1 megabit/s over a range of up to 2500 feet. Epigram has developed a 10 megabit/s version, which will operate over a range of up to 1000 feet. AVIO is developing a special version that will support a rate of up to 88 megabits/s over a range of up 100 meters over category 5 wiring or 33 meters over regular wiring. Epigram has a 100 megabit/s version in the lab.

Phoneline networking modems are relatively inexpensive, costing as little as $50 per modem when bought in quantity.

Powerline Technology

Powerline technology can turn the power network within a home, MDU, or small office building into a giant LAN that can be accessed by plugging into the wall. It costs on the order of $50 per node. Intelogis has developed a version that will operate at speeds of 350 kilobit/s today over a range of up to ¼ mile, and is planning on a 2 megabit/s version by the end of the year. Enikia has developed a 10 megabit/s version, but it only has a range of 200 feet.

The longer-range technology could be good for office buildings and MDUs. However it has the limitation that everyone on a particular circuit shares the same networking segment. Another limitation is that it cannot operate through a transformer.

Caching

Caches are simply remote servers that can store data that is frequently accessed by a number of users in order to minimize the load on the core network. These caches can be used to deliver frequently accessed broadband content, such as video and streaming 3D.

Examples of Different Hybrid Architectures

Wireless used to a point miles from cable plant to individual businesses

In this scenario, the range of cable services could be extended by a distance of up to 25 miles past the edge of the cable plant. The connection could use relatively low rate, since the connection would only be with a single business. On the customer's side, the network could be connected into the existing LAN, enabling all of the computers on the network to access the Internet simultaneously.

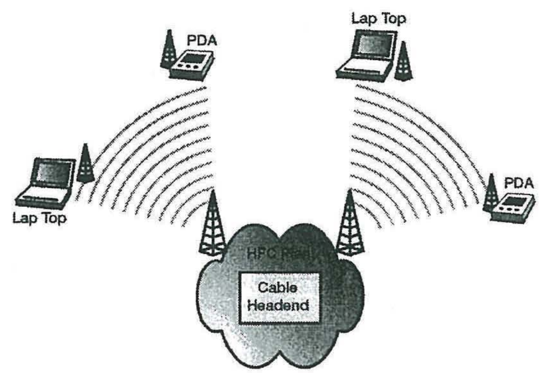
Wireless to remote Points of Presence (POPs)

A wireless link would be extended to a remote office park, office building, shopping center, MDU or small residential community. Once this connection has been established, the network could be extended to each individual business or residence via a combination of low-cost shorter-range wireless technologies, phoneline networking, powerline networking, or a traditional coaxial plant.

HFC to multipoint wireless

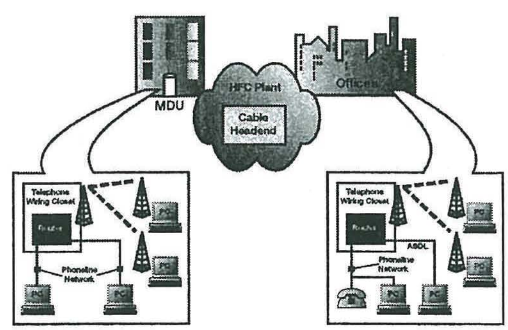
The coaxial network could be extended to a network of wireless transceivers such as Ricochet, which could facilitate communication between users roaming around campuses and office parks. Users could use this network to access the Internet, or office servers using laptops, PDAs, or special appliances such as bar code readers.

HFC to Multipoint Wireless



Although the data rates on this type of network would be slower than typical high-speed networks, the mobility factor could be important in places such as college campuses where the users may be moving around all day.

Coax to MDU or Office Building



HFC to office building or MDU

A single HFC connection is made to a large office building or MDU. Data services could be extended throughout the facility using phoneline networking, powerline networking, wireless, or ADSL technology. This would obviate the need to install any new wiring in the facility, enabling a rapid time to market for the service.

ADSL beyond range of HFC

ADSL could be used to extend the range of service up to 3 miles from the edge of the HFC network. However, it would require that the cable operator had access to a pair of wires with a direct connection to the customer premise. One scenario would be for this deployment to occur on a large campus or office park in which the management had ownership of the wires (as opposed to the telco) and an active interest in the deployment of high-speed data services.

Remote telephone modem bank

A telephone bank could be installed in a remote area where customers live, to enable them dial-up access to the network without incurring long distance charges. The telephone box could be connected to the core network via wireless technologies.

This type of architecture would make the most sense in remote neighborhoods that only have a one-way cable plant. It could also be used when a cable operator is offering service to a large company with lots of employees who work from home, and need to access the corporate network without incurring long distance charges.

Conclusion

The future promises a world where high-speed data services will be commonplace and will be offered by a number of service providers including cable, telco, wireless and satellite. Cable operators are in an advantageous position today in that the core technology behind the cable modem is well established. However, they must act aggressively in order to achieve market and mind share, build an embedded base, and to begin receiving the revenues that will enable them to build their service networks out.

Blended technologies promise to give operators the tools they need to offer service in a blitzkrieg-like, yet cost effective manner, while building out their traditional HFC plant as other business and geographical needs dictate.

Bringing Together Headend Consolidation and High-Speed Data Traffic in HFC Architecture Design

Dr. Timothy Brophy, Dr. Robert Howald, Curt Smith
General Instrument Corporation
Transmission Network Systems

The move toward headend consolidation, coupled with the evolution of the traditional CATV HFC network into a two-way interactive data communications platform, has led to an investigation of optimal architectures. The goals of such architectures are clear. They must meet current needs of multiplexed analog and digital systems and provide a straightforward and cost-effective migration. As a utopian objective, it should provide the ability to handle the maximum capacity anticipated, as two-way interactive services continue to expand in the longer term.

This latter aim, commonly referred to as "future-proofing," is much more difficult to pin down. The highly variable nature of potential service offerings, and the inability to predict the likelihood of their acceptance makes achieving this goal challenging. However, it is commonly accepted that increased use of optics and the efficient use of the RF bandwidth are the enabling network technologies to assure these goals are met. In this paper, we will discuss today's freshest technologies, DWDM and frequency stacking.

As a result of the need for more bandwidth in both upstream and downstream paths, optical and RF technologies have advanced. Today, DWDM systems are being deployed to provide segmentation and increased bandwidth. Additionally multiplexing in the RF domain is also being used in the upstream passband to increase bandwidth efficiency. From a network planning perspective, both DWDM and frequency stacking are excellent tools for "future-proofing" a network. The focus of this paper will be the combination of these technologies. Questions to be answered are:

Will it work?
How does it help my network?
How does it compare with other options?

INTRODUCTION

In a typical CATV plant today, downstream content occupies the 50-870 MHz frequency partition of the network. Return path, or upstream, signals are relegated to the 5-42 MHz frequencies. Given the asymmetrical nature of the frequency bands used it is very likely that the upstream traffic will be the first to be constrained. In this paper we will address the combining of two technologies to provide increased capacity in the upstream network. Both technologies involve multiplexing, optical multiplexing using dense wave division multiplexing (DWDM) and RF multiplexing using frequency stacking. As we will see, the combination of these technologies can increase the efficiency of the return path. But first, let's start with an introduction to DWDM and frequency stacking and their uses.

DWDM 101

CATV systems have evolved from their earliest days of delivering standard off-air channels to remote areas plagued with poor reception. In those early systems, a single headend with a large master antenna received the channels for distribution over coaxial cables and through many RF amplifiers to customers. In the late 1980s and through the 1990s, fiber optic transmission became a reality for the MSO (multi–system operators) of cable television. In a typical system, the optical fiber replaced the trunk section of the RF transmission plant and improved signal quality by virtue of avoiding the degradation of the multiple amplifier cascade. Since the system was almost exclusively designed for one-way transmission (from signal source—the CATV headend—to home), the architecture evolved into the fiber-fed, tree and branch network dominant today. [1]

Return path implementations, which were often installed due to franchise requirements, were lightly loaded and used primarily for communications with set top converters. With the advent of cable modems and the promise of IP telephony over cable, a robust return path will be required. This will be driven by higher use rates as the network evolves into much more symmetric architectures. The implication is not only for the build out of the return path, but also the ability to segment the forward signal to address individual subscribers. Further, total network traffic is kept to manageable levels by logically segmenting the signals to address individual users at the headend. The driving motivation behind the use of DWDM in CATV is the increase in bi-directional capacity without the use of additional fiber. [2] This includes delivery of independent signals to various end users, and the cost-effective collection of return signals from those users. These independent signals, or targeted services, include internet data streams, telephony, requests for and delivery of video-on-demand, near-video-on-demand, and analog channels reserved by franchise requirements for public, education, and government use.

In the CATV world today, DWDM systems are used exclusively in the 1550 nm optical window. As in the telephony world, the erbium doped fiber amplifiers (EDFA) are the enabling technology that makes this wavelength window attractive. The wavelengths that comprise the ITU grid [3] are actually a set of predefined frequencies currently spaced at 100 GHz, from which wavelengths can be derived. The wavelength spacing is approximately 0.8 nm, and the range of wavelengths covers the EDFA band, from about 1530 to 1570 nm. Not all the wavelengths need to be used in any given system and commercial products are available at 100, 200, and 400 GHz spacings, with a variety of individual product offerings. In the system to be discussed later, the spacing chosen is 200 GHz, and it is this closeness of wavelengths that make the system "dense." This is to distinguish them from some existing CATV systems, which use a combination of 1310 nm and 1550 nm wavelengths in a WDM arrangement. As described earlier, the signals traditionally sent to subscribers were analog channels, to ensure compatibility with television tuners. In CATV jargon, the signals sent over DWDM transmitters are digital. They are actually QAM signals, where a digital bit stream is encoded onto an analog subcarrier. Frequently in the literature, the terms QAM, digital, targeted services, or DWDM signals are used interchangeably.

DWDM Components

Transmitters

The laser sources for DWDM systems are housed in a headend transmitter, which provides the bias, temperature, and monitoring controls as well as the means of modulating the sources with the RF content. That RF content is either the analog broadcast television signals or the targeted services QAM signal. The modulation techniques are either external (using the modified balanced-bridge Mach-Zehnder interferometer) or direct (using the driving current control of the laser directly).

Directly modulated transmitters have the advantages of low cost and simplicity of design. Their operating wavelength (as for the sources in external modulators) is well controlled by their bias current and operational temperature through standard feedback loops. It is also possible to purchase highly linear sources that need no predistortion to yield the system performance required. Launch powers are available up to about 12 dBm. The most significant disadvantage is the chirping of the laser due to the modulating current.

Chirping (FM efficiency) converts amplitude changes into frequency changes, which broadens the linewidth in a time varying fashion. The chirp, when combined with fiber dispersion, can lead to degradation in even order distortions measured at the receiver. Recall that standard single mode fiber has comparatively large dispersion of ~17 ps/nm/km in the 1550 window. This second order distortion effect (a form of a multi-channel system non-linearity called composite second order, or CSO) is worse at high subcarrier frequencies. This limits the maximum operational frequency of the transmitter. Further, the time-varying line broadening is also detrimental due to interaction with the passive components. When the broadened line encounters a non-flat area in the multiplexer or demultiplexer, the instantaneous amplitude change is itself a source of non-linear distortion. Finally, if the linearity performance of the transmitter is limited, intermodulation distortions will occur, falling at the sum and difference frequencies of the RF content. If the loading on the transmitter is wider than 50 MHz, these IM distortions will fall into the band of the downstream analog channels.

With external modulators, the advantages are attributable to the unmodified characteristic of the laser source. The narrow linewidth is desirable so that interaction with the width or shape of the passband in the multiplexer or demultiplexer is minimized. Too narrow a linewidth can lead to fiber non-linearities, which are discussed below. Given this, a means to broaden the linewidth is required. Another desirable property is the low chirp of the source. Since the total line broadening of the link is length dependent, the low chirp characteristic leads to longer fiber links being attainable before being limited by second order distortions. On the down side, externally modulated transmitters are generally more expensive than their directly modulated counterparts, and will always need some correction to improve their intrinsically poor linearity, which results from the sinusoidal transfer function of the modulator.

Receivers

Receivers for use in DWDM systems are not distinguished from their existing analog components. Typical receivers use an InGaAs photodetector that is transformer coupled to an RF preamplifier. The photodetector has adequate energy gap so that both 1310 and 1550 nm are received for conversion into current.

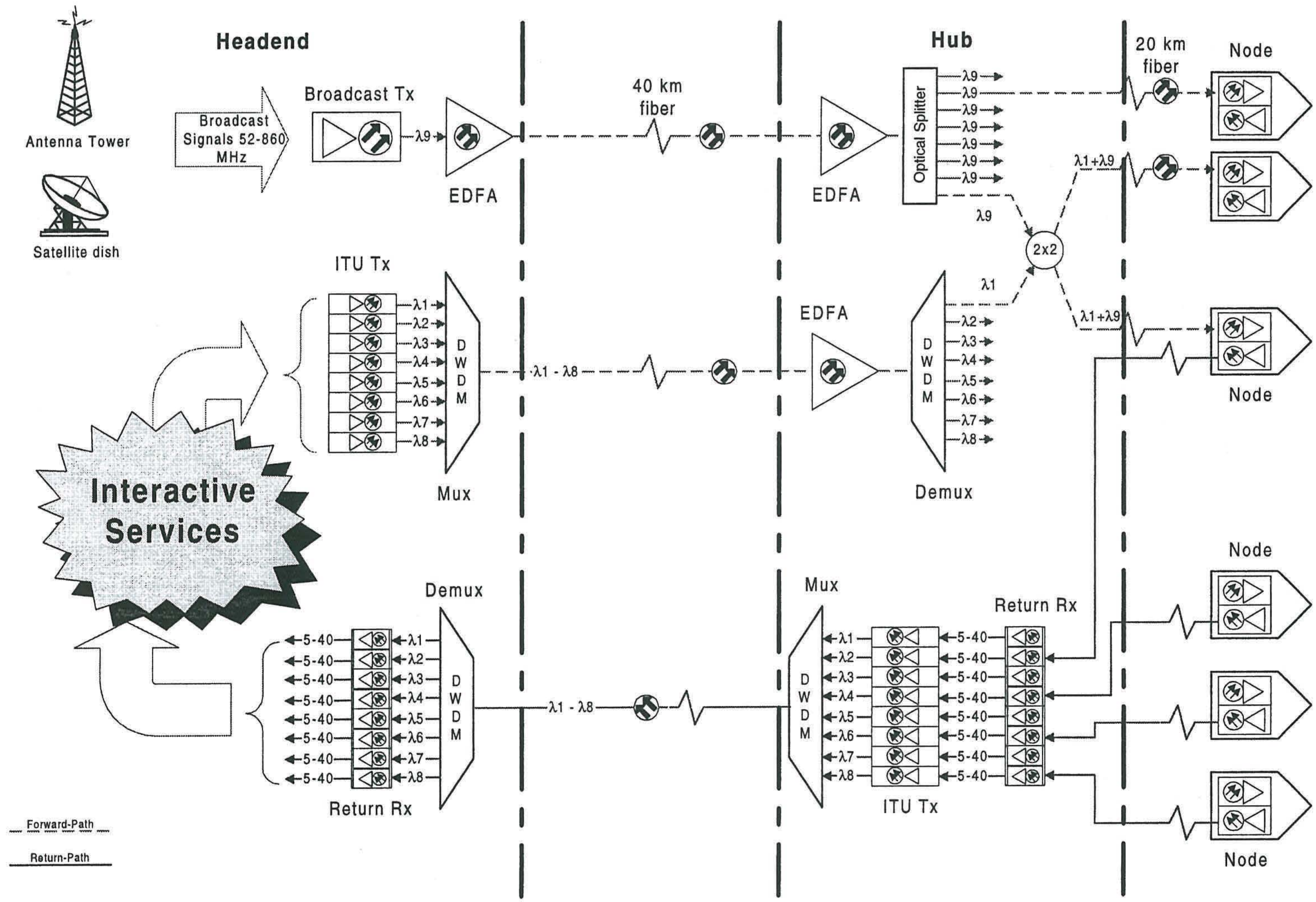


Figure 1: Generic DWDM network

Passives

Passive components in the DWDM system include optical attenuators to assure uniform power at the individual nodes and proper relative level of the targeted services delivery (TSD) wavelengths compared to the analog. In addition, there are fused fiber couplers used to combine the analog with the TSD wavelengths at the OTN location for subsequent distribution to the node. Most importantly, however are the multiplexer and demultiplexer components. These are used to combine the various ITU grid wavelengths through a low loss coupler and carry them on a single fiber and subsequently separate those wavelengths to place them onto individual fibers. There are four competing technologies that are available to create these components. They are fused biconic taper (FBT) couplers [4], fiber Bragg grating (FBG) couplers, (both of which are configured as fiber Mach-Zehnder interferometers), silicon array waveguide grating (AWG) couplers [5], and thin film interference filters (TFF).

Each of these technologies has some advantages and disadvantages relative to the others, and it is not always possible to determine which characteristics are most critical without also considering the system level interactions. Essentially, comparative evaluation comes down to cost, potential for high volume manufacture, degree of operational temperature stability (is active thermal control required?) and operational performance parameters, such as passband flatness and width, and adjacent channel isolation capability.

DWDM Applications

Downstream (Forward Path) Narrowcast Overlay With Optical Insertion

A schematic of a generic architecture for a DWDM overlay of a standard CATV distribution system is shown in figure 1. The analog transmitter is an externally modulated source, comprising a 1550 nm DFB laser coupled to a Mach-Zehnder interferometer made in a lithium niobate substrate. The output is optically amplified to a saturated level of about 17 dBm, transmitted through 40 km of standard (non-dispersion shifted) single-mode fiber (SMF) to an Optical Transition Node (OTN) location, amplified again and split into a number of outputs that matches the number of targeted-services wavelengths. After splitting, the analog signal is combined with the QAM wavelengths in an analog/digital coupler and that combination is again split to serve the number of nodes for which the given wavelength is targeted. The nodes are the optical-to-electrical conversion points for transmission of CATV signals. In our generic system, those nodes are 20 km away from the OTN and are connected using standard SMF. Note that there may be multiple nodes targeted per wavelength, especially in the early deployment stages when subscriber take rates are low corresponding to a low bandwidth requirement per node.

The DWDM sources are also externally modulated transmitters in the example system, but directly modulated sources may also be used. Eight wavelengths are shown in the figure and are combined into a single fiber in a multiplexer. The standard SMF is 40 km long, and is distinct from the fiber carrying the analog signal, but may be in the same cable. After the 40 km, at the OTN location, the combined wavelengths are amplified and then demultiplexed into separate fibers. Each targeted services

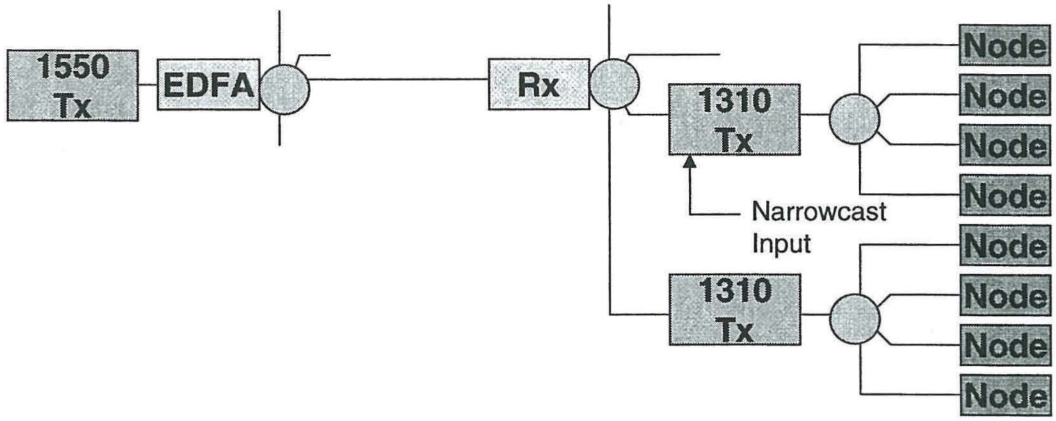


Figure 2: Dual-hop optical architecture

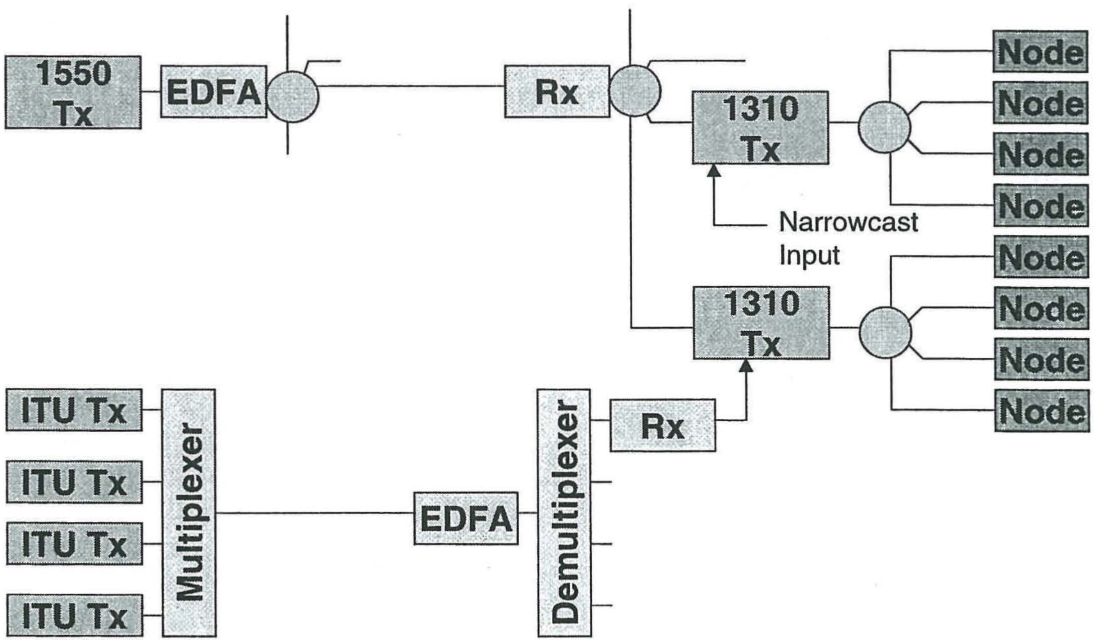


Figure 3: Narrowcast overlay with RF insertion

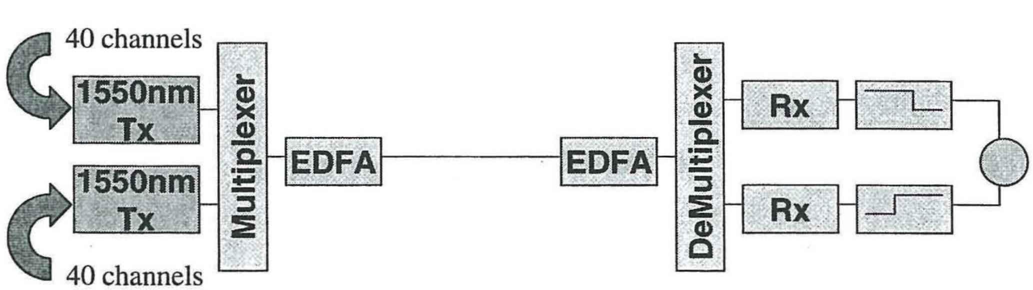


Figure 4: Split-band application using a single fiber

wavelength is combined with one of the split analog signal outputs and distributed to nodes through a single fiber carrying both the analog and digital signal. The fiber node contains a receiver that detects both the analog and QAM signals for distribution through the RF plant beyond the node.

Upstream (Return Path) Transport

In keeping with the drive towards a more symmetric network, the upstream or return path mirrors that of the downstream. The exception to this mirroring occurs not so much in the single fiber of the DWDM system, but in the portion of the return from the node to OTN. The return path is managed as a two-hop process. In the illustrated system, a temperature compensated 1310 nm laser (typically a DFB) is in the node. The RF signals from homes served by this node (1000-1200 subscribers) are collected and time division multiplexed before driving this laser. That optical output is sent over the 20 km link to the OTN, where it is detected and amplified by a receiver before directly modulating an ITU grid laser. That laser is one of several which combine the entire return path into a DWDM set for transmission over the 40 km back to the headend and for subsequent processing. Each of the DWDM wavelengths may handle the return traffic from multiple nodes using a combination of time, frequency, or code division multiplexing.

Downstream (Forward Path) Narrowcast Overlay With RF Insertion

The network solution shown in Figure 1 assumes that the optical network remains in the 1550nm window from the headend to the node. What if the existing system utilizes a re-transmission scheme at the hub/OTN. It is a goal to preserve as much of this infrastructure as possible. Fortunately DWDM can still be used to provide the narrowcast overlay.

For the purpose of discussion, we will assume that the network is configured as shown in Figure 2. This generic network uses 1550 nm technology from the headend to the hub. At the hub, the signals are returned to RF and then routed to 1310 nm lasers for distribution to the nodes. We will also assume that these 1310 nm lasers have a second RF input port for narrowcasting.

To implement the narrowcast overlay the arrangement is very similar to the approach in Figure 1. This arrangement is shown in Figure 3. Signals are placed in the proper RF spectrum and routed to the ITU grid transmitters in the headend. Outputs from the transmitters are multiplexed and transported to the hub. At the hub, the optical multiplex may need to be optically amplified to overcome demultiplexer losses. At this point the approach becomes different. Instead of combining the demultiplexed optical signal with the broadcast, the individual wavelengths are routed to receivers. Outputs of the receivers are then passband filtered, and then routed to the proper 1310 nm transmitter. Combined at RF, both analog and QAM signals drive the laser. At the node a single detector converts these signals to RF for distribution into the plant.

Downstream (Forward Path) Application to Improve Super-trunk Performance

In this application DWDM is used to reduce the quantity of optical amplifiers and fiber required while still maintaining the performance improvements associated with a split-band super-trunk configuration. Figure 4 illustrates this application.

In standard split-band super-trunk applications two fibers would be required to

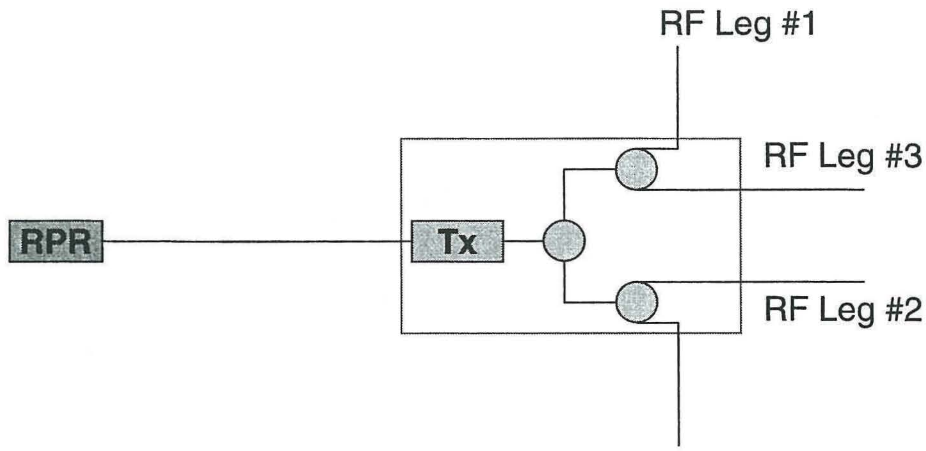


Figure 5: Typical node configuration

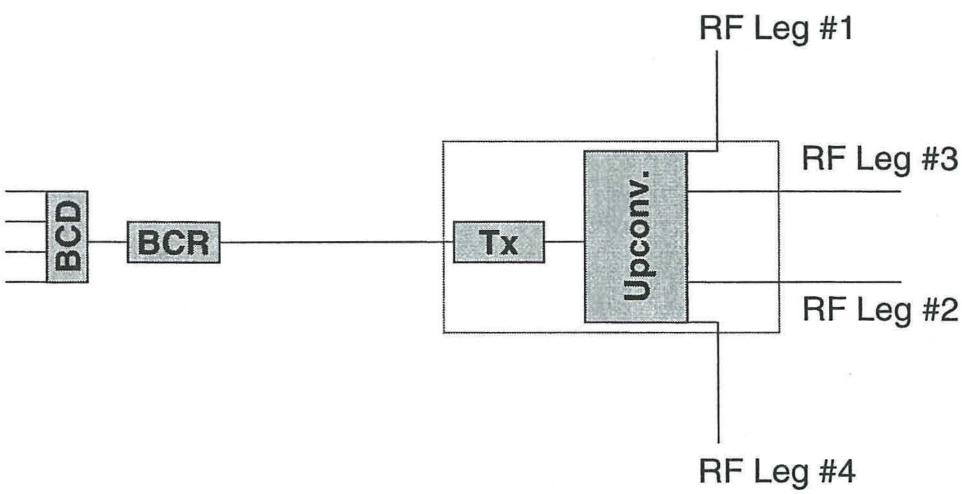


Figure 6: FSS block diagram

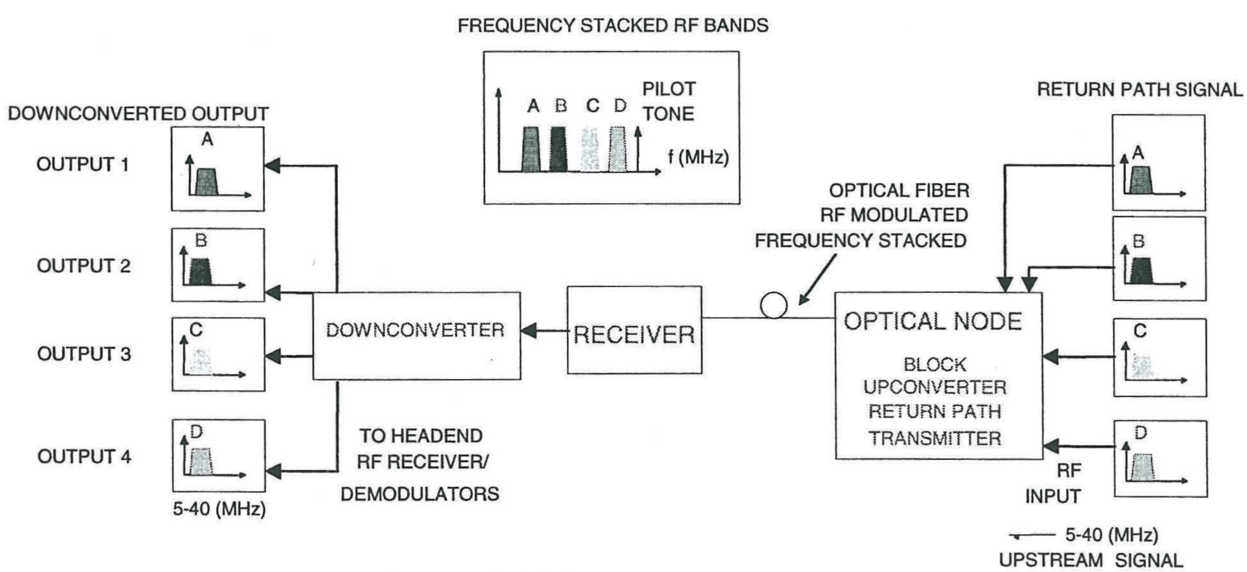


Figure 7: FSS spectrum allocation

transport the signal from the headend to the hub. In instances where there are multiple amplification points in the network it would require two identical sets of this equipment. By using DWDM to optically multiplex these signals a significant savings can be achieved. By using only one fiber the quantities of optical amplifiers is reduced to half.

FREQUENCY STACKING 101

In frequency stacking systems the 5-40 MHz return passband is block upconverted or shifted to another frequency passband. This may be done in a hub environment or, as we will discuss here, in the field located node. The main advantage of the implementation of a frequency stacking system (FSS) is the expansion of the return bandwidth per home passed. This allows for larger node sizes, which in turn reduces the overall system costs.[6]

If we look at "typical" node configuration, shown in Figure 5, all the users served by the node share the return path spectrum. If this were a 1200 home passed node, each home passed would have approximately 29 KHz of guaranteed simultaneous bandwidth. This assumes that the entire 35 MHz is available, and we can dynamically allocate the bandwidth. As Figure 5 illustrates each of the coaxial busses are RF combined into one stream.

Adding more transmitters combined with segmenting the RF paths within the node may increase bandwidth. This approach has disadvantages. Beyond adding one additional return transmitter in the node, which only doubles capacity, fiber availability issues may become the limiting factor. To achieve the same level of bandwidth per home passed as FSS three additional transmitters and fibers would be required.

An FSS approach utilizes upconversion in the node to create four passbands for the return. In this approach each leg now has its own 35 MHz of space. The four passbands are RF stacked and sent to the return laser. Figures 6 and 7 illustrate this arrangement.

FSS Components

As Figures 6 and 7 illustrate there are four major components associated with a FSS system. These components would be common in function regardless of whether the application is hub or node based. Each of these components is briefly discussed below.

Upconverter

Frequency stacking begins with the upconverter. This device, simply put, takes multiple return passbands and shifts them to other independent passbands in the spectrum. The conversion process maintains the information that resides in the original passband. In the implementation shown in Figures 6 and 7, each of the RF legs are upconverted to different passbands within the 50 – 400 MHz passband. A pilot carrier serves two key functions. First, it compensatesfor the range of link loss introduced by the optical network. Second, it is used by the downconverter to phaselock to the upconverter thus eliminating frequency offsets.

Transmitter

The transmitter used in this application is not a standard, band-limited, return path transmitter. In his implementation, a forward path transmitter designed to operate in the 50-400 MHz passband is used to

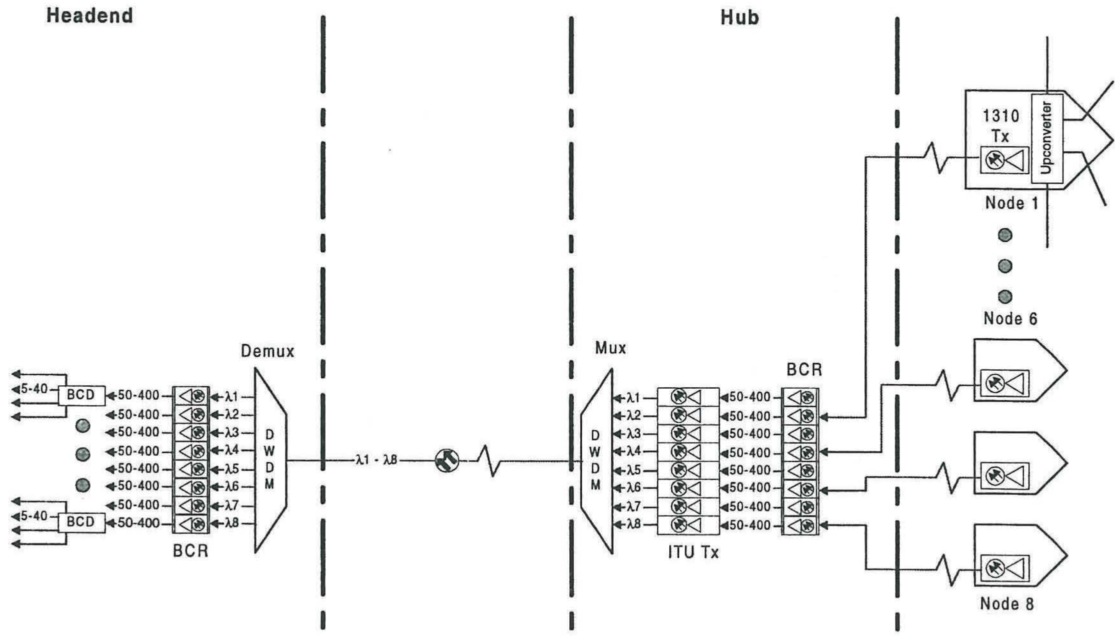


Figure 8: DWDM/FSS configuration 1

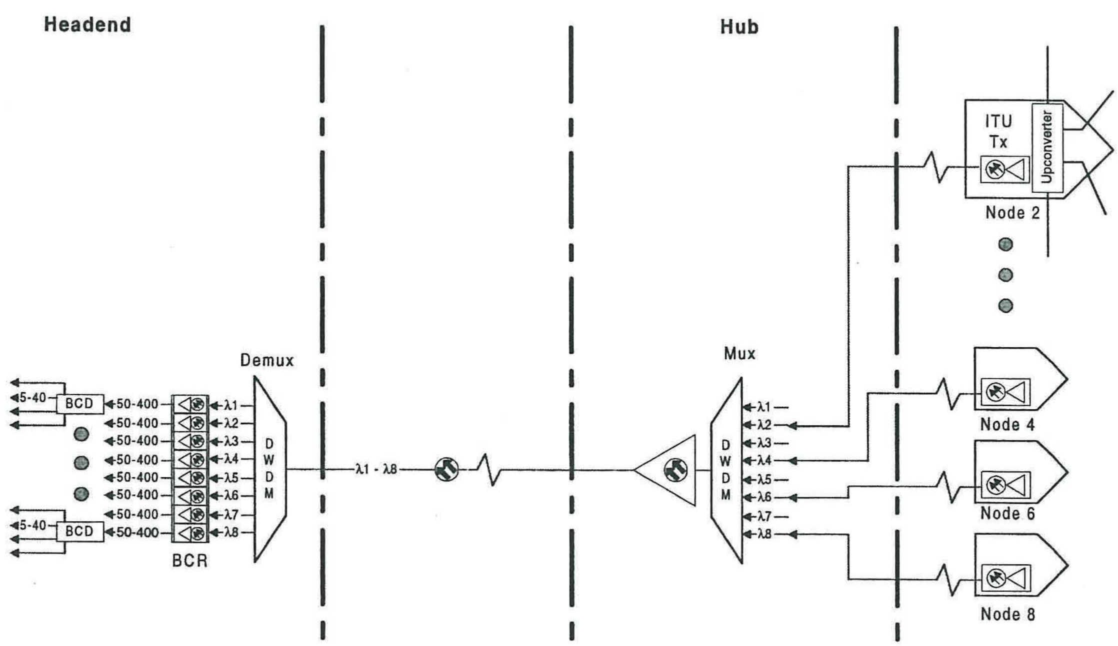


Figure 9: DWDM/FSS configuration 2

transport the upconverted signal. A DFB laser is chosen to provide optimum performance and reach.

Receiver

The FSS receiver (BCR) is also different than the normal return path receiver (RPR). Again chosen for the forward path, the receiver provides the composite RF output. Contained within this passband are the four upconverted bands along with the pilot carrier. To recover the individual bands, a downconversion process is performed.

Downconverter

The downcoverter provides the means of returning the upconverted bands to their original 5-40 MHz spectrum. Using the pilot carrier for frequency synchronization, the block downconverter (BCD) reverses the process initiated in the node. This device provides four independent 5-40MHz passbands, one for each of the upconverted bands. These outputs are then fed to the return splitting/combining network and eventually end at the individual service demodulators.

COMBINING DWDM and FREQUENCY STACKING SYSTEMS

System Description

Now that we have discussed both DWDM and FSS in some detail let's look at how these two systems complement each other. In our goal to increase bandwidth efficiency, both approaches work together to increase the efficiency of both the return distribution and return transport aspects of the network. Combining both DWDM and FSS technologies it is possible to have thirty-two 5-40 MHz return bands on a single fiber. There are two configurations currently being investigated to combine these two technologies. The difference is the location of the ITU grid transmitters. These configurations are:

1. DWDM transmitters at the hub/OTN
2. DWDM transmitters at the node

In the first configuration the ITU grid DWDM transmitters are located at the hub/OTN. As illustrated in Figure 8, this configuration upconverts the return path signals at the node location. Transmitted back to the hub/OTN via the optical distribution network they are received by the forward path block conversion receiver (BCR). At this point we begin to differ from a standard FSS network. Instead of routing the RF output from the receiver to a downconverter, it is routed to a DWDM transmitter. This transmitter has an output wavelength on the ITU grid. We now have a concentration of four discrete 5-40 MHz passbands on each of these transmitters.

Using 200 GHz spacing, we can now optically multiplex eight of these transmitters, each with its own different ITU grid wavelength, on a single fiber. **We now have 32 discrete 5-40 MHz passbands (1.12 GHz) on a single fiber.** The signals are now routed to the headend. Depending on the distances involved, and requirements such as redundancy, optical amplifiers may be required to meet the input requirements of the headend receivers.

At the headend the optical signals are demultiplexed into the eight wavelengths. Individual wavelengths are routed to receivers (one for each wavelength). These receivers are the same as those used to receive the frequency-stacked multiplex at the hub/OTN. At this point we complete the FSS system by routing the composite RF signals from the BCRs to the downconverters. The four, 5-40 MHz, RF outputs from the downconverter correspond to the four coaxial legs coming into the field node.

In the second configuration, illustrated by Figure 9, much of the same components are used. DWDM transmitters are now located in the node. Fed by the stacked RF from the upconverter the individual wavelengths are transmitted back to the hub/OTN.

At the OTN, the optical signals are routed directly to multiplexers. Since it is possible to have different optical levels, due to different node to OTN loss budgets, some level of signal equalization may be required. Output from the multiplexer is sent to the headend in the same manner as in configuration 1. The headend components are assembled as in configuration 1 as well.

The key advantage with this approach is the reduced amount of active equipment located within the hub/OTN. As figure 9 illustrates, it is no longer necessary to convert the signal back to RF. This will improve performance, but places the transmitter in a more hostile environment. This temperature stability is one of the technical issues associated with not only this technology combination, but with DWDM itself.

Technical Issues

Maintaining required performance in the return band

In the return band, the Noise Power Ratio (NPR) test replaces CNR as a discriminating criterion of robust operation. [7, 8] In this test, a marker is placed at the average level of the noise "signal" in the 5 to 40 MHz range, and the system noise floor is measured in the filter notch at 22 MHz. A conservative requirement is for a 40 dB ratio. To provide adequate headroom, the requirement of 40 dB should be met over a dynamic range of signals. In the NPR test, the RF input level is varied over power while measuring the signal level and that of the noise floor. Today's DWDM systems are able to achieve NPR's of roughly 15 dB above a NPR = 40 dB.

Additionally an FSS system can achieve well over 40 dB with a wide dynamic range. The FSS-2000 implementation shown was designed around successfully closing a 32-QAM link budget with margin, thus clearing comfortably the anticipated widespread use of 16-QAM. In the (noise+distortion) link budget sense, a required NPR of 40 dB is conservative. 16-QAM at 1e-8 BER requires 22 dB uncorrected CNR. With digital IMD looking noise-like, it behaves similarly to C/IMD. Losses are incurred from channel impairments (frequency response, interference, impulse noise, clipping, phase noise), HE combining, optical and RF link variation, and modem implementation losses. Link "gain" is achieved via error correction, as well as indirectly through equalization, in the sense that it mitigates frequency response distortions. It is also indirectly achieved in the so-called "hy-phy" approaches that will form the next round of cable modem technology (CDMA, OFDM).

When the two systems are combined, a reasonable goal is still 40 dB NPR. DWDM/FSS implementations as discussed here have about 9 dB of dynamic range for a 40 dB NPR. It is common practice in developing link budgets to ignore the benefits of error correction and frequency response equalization, harsh as that may be. Uncorrected data is an actual functional mode in DOCSIS1.0. However, completely ignoring any such benefits means that the

NPR requirement as specified assumes 18 dB is in reserve for contributions from the set of impairments described above. This provides superior technical performance, but of course must be weighed in terms of dollars for dB as to the need for this much margin. The problem is complicated by the fact that the comfortable amount of headroom necessary will vary among plants, depending on it's susceptibility to high levels of interference. We have found that commonly occurring interference is not typically large enough itself to have a major impact on the probability of clipping.[8]

Temperature stability of DWDM components

As mentioned earlier, temperature stability of the DWDM components is achievable. However this stability does have a cost. In the implementation of DWDM networks today the environment in which they are placed are somewhat controlled. Early versions of the equipment were intended for climate controlled facilities, but with the technology maturing they are being placed in less environment-friendly locations. The need for temperature compensation has occurred in two areas relative to current applications. Advances in both DWDM components and cabinet design have allowed these devices to move into more hostile environments.

However, there is still work to be done to move these components into the node. Implementations that place ITU grid transmitters in the node would require these transmitters to be very stable. In addition the amount of multiplexing will continue to increase. With 200 GHz spacing, 8 wavelengths can be multiplexed. If the industry moves to an implementation that places multiple ITU grid transmitters in the node, the need for more multiplexing will rapidly increase. Given that there are alternatives that provide the equivalent bandwidth, the networks may not reach this level. However, our crystal ball is as cloudy as everyone else's.

Comparative Analysis

Equipment usage

If we compare a hypothetical system that has a distance of 40 km between the headend and the OTN and a distance of 20 km from the OTN to node we see large differences in the amount of equipment. There will be 20 nodes fed from the OTN. The goal is to get the return path traffic for these nodes (80 individual 5-40 MHz passbands) back to the headend. The four scenarios that we will investigate are:

1. Standard 1310 return to the OTN and then retransmission utilizing DWDM back to the headend.
2. Standard FSS transport to the OTN and then retransmission utilizing DWDM back to the headend.
3. FSS transmission using ITU lasers in the node and multiplexing/optical amplification in the OTN for transport to the headend
4. Transmission from the node to the OTN using ITU lasers and then multiplexing/optical amplification in the OTN for transport to the headend.

In each of these scenarios the amount of return segmentation/efficiency will be kept equal. We will use a 2000 home passed node with the return segmented into four 500 home nodes within the node. This same segmentation can also be accomplished by implementing four 500 home passed nodes. This would require more fibers in the downstream direction.

Parameter	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Standard 1310 nm transmitters in the node	80	-	-	-
Standard FSS transmitters in the node	-	20	-	-
ITU grid FSS transmitters in the node	-	-	20	-
ITU grid transmitters in the node	-	-	-	80
Qty of return fibers from node to OTN	80	20	20	80
Return path receivers at the OTN	80	-	-	-
FSS receivers at the OTN	-	20	-	-
ITU grid transmitters at the OTN	80	20	-	-
8-wavelength muliplexers at the OTN	10	2.5	2.5	10
EDFA optical amplifiers at the OTN	-	-	3	10
Qty of return fibers from OTN to headend	10	3	3	10
8-wavelength demultiplexers at the headend	10	2.5	2.5	10
Return path receivers at the headend	80	-	-	80
FSS receivers at the headend	-	20	20	-
Downconverters at the headend	-	20	20	-

Table 1: Estimated equipment required for transport from node to headend

System costing

As we start to compare the system costs of these four approaches let us examine the fiber savings first. We will assume that the fibers will be in the same sheath. This will allow us to disregard the installation labor of the fiber cable, as it will be the same for each scenario. However, splicing costs for the different amounts of fiber in each approach will be different. We are not including these costs. To make the math easier, we will further assume that the fiber cable costs $0.01 per fiber strand per foot. So a 12-fiber cable that is 10' long would cost $1.20.

Looking at the fiber involved from the node to the OTN, Table 1 shows that the scenarios employing a FSS configuration use one-fourth as many fibers. For the 80 return passbands we want to transport, the non-FSS scenarios use 60 more fibers. At a 20 km distance from the node to OTN, this equates to a 60-fiber cable that is 20 km long. Translating this in monetary amounts with our assumed fiber price results in $39,360 in additional fiber expense in this segment. If we examine the savings in the headend to OTN link we find an additional $9,184 of fiber costs associated with the non-FSS approaches.

Today, the FSS and DWDM equipment is generally more costly than the standard DFB products used in the return path today. Will this eliminate the fiber savings? If we take scenario 1 as our baseline, how we would do it without FSS, we can compare the system costs. Providing the same level of service, a budgetary costing estimate for the equipment only is summarized in Table 2.

Parameter	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Return path equipment(includes node, OTN, & headend electronics)	$766,000	$379,060	$274,000	$560,000
Percentage difference from baseline (Scenario 1)	baseline	(50.5%)	(64.2%)	(26.9%)

Table 2: Equipment cost comparison

As Table 2 illustrates the DWDM and FSS combinations are less costly than scenarios 1 and 4. Although the FSS components are more costly, the reduced volume greatly offsets this increased unit cost. It should be noted that costs associated with DWDM components that reside in the node (scenarios 3 and 4) are estimates. The technical hurdles of placing this type of equipment in the harsh environment of the node are not fully developed, in a cost-effective solution, at this time.

CONCLUSIONS

1) These two technologies (FSS and DWDM) are complementary and not mutually exclusive in their implementation. In fact the implementation of this hybrid may reduce system costs extensively.
2) This architecture hybrid can provide adequate NPR performance for the traffic on the return path.
3) This hybrid is very well suited to implementation into existing systems that may have fiber limitations in both the headend-to-OTN and OTN-to-node optical segments.

REFERENCES

1. For a good historical perspective on the development of CATV architectures through 1990 see Walter S. Ciciora, "An Introduction to Cable Television in the United States," IEEE LCS magazine, February, 1990, pp. 19-25. Other papers in that issue are also relevant. For an up-to-date version of architectures and issues, see Walter S. Ciciora's book An Introduction to Cable Television. It is available at the CableLabs home page at www.cablelabs.com.
2. Oleh J. Sniezko, Video and Data Transmission in the Evolving HFC Network, invited talk, Optical Fiber Communications Conference, 1998.
3. International Telecommunication Union, T study group 15 (Transport, Network, Systems and Equipment); work is not yet completed.
4. Francois Gonthier, "Fused couplers increase system design options," Laser Focus World, June 1998, pp. 83-88.
5. C. Dragone, C. A. Edwards, and R. C. Kistler, "Integrated optics N x N multiplexer on silicon," IEEE Photonics Technology Letters, vol.3, pp. 896-899, 1991.
6. For more information on frequency stacking see Robert Howald, Michael Aviles, and Frank McMahon "Increasing HFC Capacity: Design and Field Test of Return Path Frequency Stacking," 1998 NCTA Technical Papers, Atlanta, GA pp 203-216.
7. Oleh J. Sniezko and T. Werner, "Return path active components test methods and performance comparison," 1997 Conference on Emerging Technologies, Nashville, TN Exton, PA: SCTE, pp. 263-294.
8. Donald Raskin and Dean Stoneback, *Broadband Return Systems for Hybrid Fiber/Coax Cable TV Networks*, Prentice Hall PTR, Upper Saddle River, NJ, 1998, p. 171.

Acknowledgements

The authors would like to thank Michael Aviles and Mike Short, TNS System Engineering staff, whose efforts were critical in the completion of this paper

Contact Information

Timothy Brophy, Ph.D.
General Instrument
Transmission Network Systems
101 Tournament Dr.
Horsham, PA. 19044

800-523-6678
tbrophy@gi.com

Robert Howald, Ph.D.
General Instrument
Transmission Network Systems
101 Tournament Dr.
Horsham, PA. 19044

800-523-6678
rhowald@gi.com

Curt Smith
General Instrument
Transmission Network Systems
101 Tournament Dr.
Horsham, PA. 19044

800-523-6678
cusmith@gi.com

Cable Headend Architecture for Delivery of Multimedia Services

Mukta Kar, Majid Chelehmal, and Richard S. Prodan
Cable Television Laboratories

Chezhian Renganathan
HCL Technologies Private India Pvt. Ltd.

Since cable headends traditionally deal with analog audio/video (AV) television signals, analog transmission service opportunities are limited. With analog technology, it is difficult to maintain interoperability and, as such, headend equipment can be proprietary in nature, as well as expensive and sometimes difficult to replace. Digital technology, combining standards-based compressed audio/video/data and digital transmissions, is more efficient and flexible. By implementing digital technology, opportunities to deliver new services over hybrid fiber coaxial (HFC) cable networks greatly increase, along with a proportional increase in revenue.

Additional services will add complexity at the headend. One of the benefits of standardized digital technology is that, through interoperability, distinct services can be integrated across system components from multiple suppliers. ISO/MPEG has provided standards for compression, storage, and digital AV transport; SMPTE has established standards for interfaces used in baseband digital AV equipment; and IETF has provided standard protocols for IP-based applications. Headend equipment may be designed with an integrated, flexible analog/digital architecture so that new standardized equipment may be added incrementally to enhance existing services or to add new services.

An integrated, flexible architecture with coexisting analog and digital services is presented in this paper. This architecture allows the enhancement of existing services or the addition of new services with minimal disruption to headend functions. Interface design issues relating to analog AV signals and multimedia signals, such as DTV, HDTV, and IP-based streaming is discussed. Additionally, information is included on digital program insertion systems for adding locally compressed content (e.g., advertising) to remotely distributed compressed programming, as well as processing of IP/Internet traffic at the headend for delivery over HFC cable networks using cable data modems.

INTRODUCTION

Until a few years ago, the broadcast industry (cable, off-air, and satellite) was limited to the broadcast of television programming. Computers were used for scientific and industrial purposes, and the telecommunication industry provided end-to-end delivery of voice and data. Emerging technologies are blurring the boundaries between the telecommunications, television/film, and computer industries. Video, sound, and communications have migrated into computers; television is incorporating the Internet; and video, audio, and interactivity are being integrated into telecommunications. All this leads to the convergence of many different applications into what is known as "multimedia." The shaded region in Figure 1 is a graphical representation of the area covered by multimedia applications.

Multimedia has become a household word. What is multimedia? Today, the media we use to transmit information is in the form of data (e.g., text), digitized video, digitized audio, still digitized images, and/or graphics. Multimedia (multiple media) refers to the combination of two or more of these media, at least one of which is discrete (e.g., text, images) and one is continuous (e.g., video, audio).

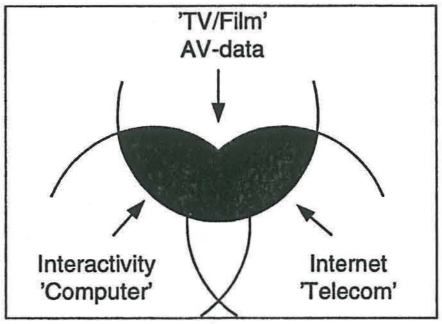


Figure 1: Multimedia Application Area (shaded)

Rapid advances in technology are producing new multimedia applications every day, opening up opportunities for traditional industries to offer new applications and services and, thus, increase revenue. The same is true for the broadcast industry. Digital technology presents a significant opportunity. Advances in video compression technology, coupled with digital modulation techniques, have opened the door to new multimedia services and applications. Using digital modulation techniques such as Quadrature Amplitude Modulation (QAM), one 6-MHz analog channel can be expanded to a digital bandwidth of 27 Mbits/sec (64 QAM) or 38 Mbits/sec (256 QAM). This QAM-modulated channel can be used to deliver either a multiplex of a few compressed TV programs or a number of other multimedia services. An analog TV program carried over a 6 MHz channel can be compressed to a 2- to 4-Mbits/sec (average) stream. Using one 6-MHz spectrum, a few analog programs can be delivered to homes with similar or better picture quality. With the size of the HFC network's total digital bandwidth, there exists the potential to deliver, in addition to analog and digital TV services, a large number of multimedia services and applications.

SIGNAL CATEGORIES

Cable systems began compressed digital programming services in 1996. Off-air broadcast of SDTV and HDTV signals began in November 1998. Per FCC guidelines, broadcasters will have to retire analog television broadcasting and completely switch to digital broadcasting by the year 2006. To be competitive, as well as to take advantage of digital bandwidth, the cable industry will add HDTV programming as well. The industry also is preparing to utilize this new potential to deliver multimedia services—IP-based services, such as e-commerce, world-wide-web, networked interactive games, video-on-demand (VOD), and others—over the HFC network. Adding all of these new services in a short period of time could be quite expensive and may not be justified economically. However, the headend may be upgraded with an integrated, yet flexible, analog-digital architecture so that new standardized equipment may be added incrementally to enhance existing services, or to add new services.

To deliver services, such as SDTV, HDTV, or any other multimedia service, the headend, the HFC cable network, and consumer premises equipment (CPE) will require different upgrade levels. For example, interactive services need a higher degree of plant upgrade than the one-way delivery of digital television programming. In particular, the integrated analog/digital headend architecture must be modular so that the enhancement of existing services, or the addition of new services, can be achieved incrementally as the resource/revenue stream permits. In this paper, we deal primarily with flexible and scalable headend architecture, which will allow incremental addition of services and, in particular, with signal interfacing and signal processing at the headend before delivery over the HFC cable network.

Signals interfaced at the headend may be categorized broadly into two types: existing analog signals and digital signals. Analog signals are already interfaced and integrated. Upgrading cable headends to interface new digital services should cause minimal or no disruption to the existing analog services. All digital signals or services can be divided into two types: MPEG-2-based digital television services and IP-based services. Some headends have been delivering digital television programming for several years. HDTV programming and the IP-based services now need to be integrated into the existing analog-digital headend. Below, we explore the tools and equipment necessary to interface these signals at the headend, as well as methods used to process them before delivery over the HFC network.

INTERFACING DTV SIGNAL SOURCES

Remultiplexers and rate-remultiplexers may be used in interfacing MPEG-2-based digital AV signals at the headend. A multiplexer is a digital device that multiplexes more than one input sequence to create a single serial digital signal (a multiplex). The multiplex bitrate will be greater than, or equal to, the sum of all the input sequences' bitrates.

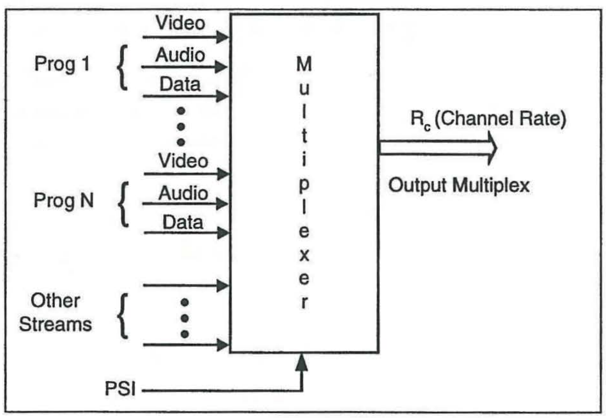


Figure 2: Multiplexer Block Diagram

Figure 2 shows a typical MPEG-2 multiplexer where MPEG-2-based elementary streams (ES) of one or more programs are packetized (188 bytes/packet) and multiplexed together to form a baseband transport multiplex compliant to ISO/IEC 13818-1 specifications. It may be noted that the multiplexer also includes PSI along with the multiplex so that the downstream equipment can demultiplex the baseband transport streams to retrieve the original constituent ES. The technology of remultiplexing is similar to that of multiplexing, but it takes one or more transport stream baseband multiplexes as input and constructs a new multiplex with the selected programs from the input multiplexes, thereby locally producing a single multiplex from multiple remote sources.

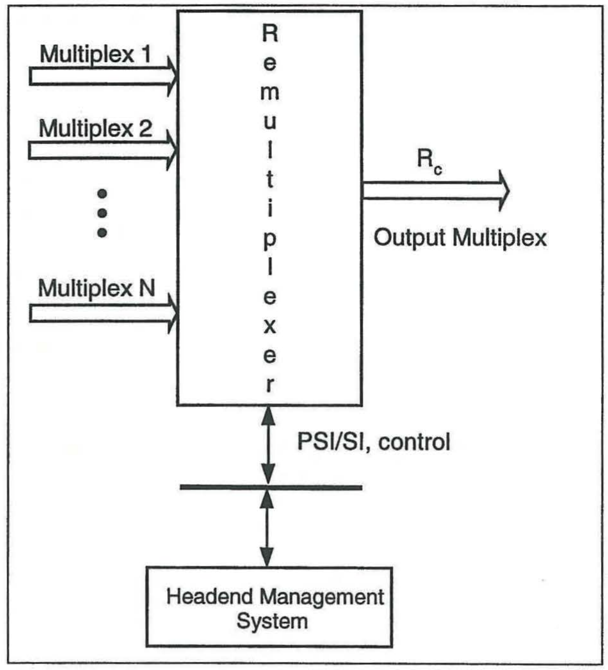


Figure 3: A Simple Remultiplexer

Figure 3 shows a typical remultiplexer. Again, the remultiplexer has the same bitrate constraints as those of a multiplexer—the sum of the ES bitrates in the selected programs cannot exceed the given bitrate of the output multiplex or channel rate. In brief, a remultiplexer does the following:

- It demultiplexes the input multiplexes and extracts the service information (SI), including program-specific information (PSI), from the incoming multiplexes;
- Constructs a new multiplex out of user-selected programs and obeys the bitrate constraint;
- It also includes appropriate PSI/SI with the output multiplex;
- Removes jitter in program clock reference (PCR) values and ensures AV synchronization within the programs;

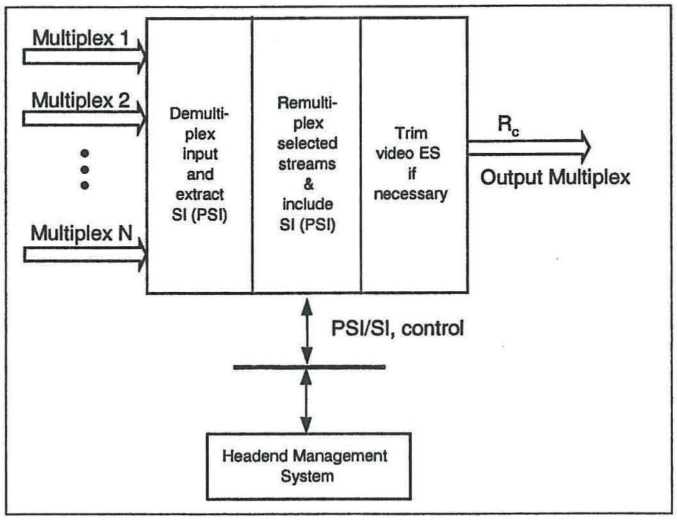


Figure 4: A Typical Rate-remultiplexer

- It provides perceptually seamless switching capability from one program to the other without any audible or visual artifacts.

An important difference between a multiplexer and a remultiplexer is that a multiplexer takes ES as input, whereas multiplexes of compressed ES are the input to a remultiplexer. In either case, ES may be of constant bitrate (CBR) or of variable bit rate (VBR).

RATE REMULTIPLEXER AND TRANSCODING TECHNOLOGY

A rate-remultiplexer (Figure 4) is a state-of-the-art piece of equipment that combines the existing remultiplexing technique with a new technology called transcoding. A standard remultiplexer does not change the bitrate of the ES, but the rate-remultiplexer can. If the sum of the bitrates of all the ES in the selected programs exceeds the given bitrate of the output multiplex, the rate-remultiplexer will reduce the bitrate of the video ES until the bitrate of the output multiplex is less than, or equal to, the available channel rate (27 Mbps for 64 QAM and 38 Mbps for 256 QAM). The rate-remultiplexer achieves this capability by implementing transcoding together with joint rate control and statmux scheduling. The latter ensures a MPEG-2 transport-compliant multiplex where video ES comply with the T-STD buffer model—no overflow or underflow occurs at the decoder buffer. Program packaging from input multiplexes for delivery over a 6-MHz channel is no longer limited by the bitrate constraint of a standard multiplexer or remultiplexer. Hence, the real advantage of a rate-remultiplexer is that it provides wider freedom in packaging a number of programs from input multiplexes into a new multiplex with a given channel rate. It is an important tool in maximizing the utilization of digital bandwidth in delivering a desired number of TV programs and, thus, may help free up some channels for delivery of other services. This rate-reducing capability provides opportunities to increase revenue with a given number of 6-MHz channels capable of carrying digital signals. The rate reducing capability is also useful in digital program insertion (DPI) and video-on-demand (VOD). The contents for DPI and VOD can be encoded with higher bitrates (e.g., 6 Mbps) than average delivery bitrates (e.g., 4 Mbps), and can be stored in the server. However, to meet the bitrate constraint, the bitrate of the stored stream will be trimmed along with other video streams when multiplexed. The advantage here is that the DPI and VOD contents do not have to be stored in different bitrate versions and it is not necessary to find a close match between the bitrate of the stored content and that of the channel [1].

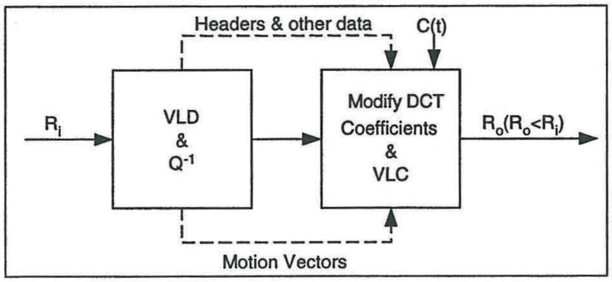


Figure 5: Open-loop Transcoder

Figure 5 is a typical open-loop rate reducer that may be implemented in the compressed domain with minimal decoding and computation. R_i is the bitrate of the input compressed video, R_o, the rate of transcoded video and c(t) may be considered as the user's input constraints. In MPEG-2 compression, the majority of the bits are used in representing the DCT coefficients of an 8 x 8 block of pixels (and there are many such blocks in a picture). The saving in bits may be achieved by reducing the number of these non-zero DCT coefficients or coarsely quantizing them. The disadvantage of the open loop scheme is that it provides little control over the picture quality of the output stream. In that respect, the closed-loop method of transcoding is significantly better where error is estimated between the picture of the input stream and that of the transcoded stream. If the error is larger than the desired threshold, the picture is re-encoded to reduce error. However, closed loop implemetation is more complicated than open loop. Irrespective of the implementation method—open loop or closed loop, pixel domain or DCT domain—the motion vectors (MV) available with the input stream should be used as much as possible in transcoding a picture (calculation of MVs is a very computationally intensive process). Transcoding of a video ES is usually associated with some degradation in picture quality. The degradation is related to the percentage in bitrate reduction. See [2] and [3] for more details on the technology of transcoding.

INTERFACING IP-BASED SERVICES

In providing digital TV services, remultiplexers (or rate-remultiplexers) are important pieces of equipment for interfacing network signals with the headend. Similarly, in providing IP-based services, a cable modem termination system (CMTS) is equally important. A CMTS and a cable modem (CM) make the entire cable network transparent to the user (Figure 6). The CMTS' and the CM's principle function is to transmit IP packets between the external IP network and the subscriber location (user), transparent of the cable network and the headend. For example, when a user uses an IP-based service via a PC (connected to a CM), the user is not aware of the cable network. The CMTS serves as the interface between the external IP network and the headend and the CM, between the cable network and the Internet appliance (e.g., a PC).

To provide IP-based services, the CMTS works as the interface between the headend and the external IP network, and may operate as a router or a bridge.

Routing and bridging are methods of moving information across an inter-network between a source and a destination. The primary difference between the two is that bridging occurs at Layer 2 (the link layer) of the open systems interconnection (OSI) reference model, while routing occurs at Layer 3 (the network layer). This distinction provides routing and bridging with different information to use in the process of moving information from source to destination. As a result, routing and bridging accomplish their tasks in different ways.

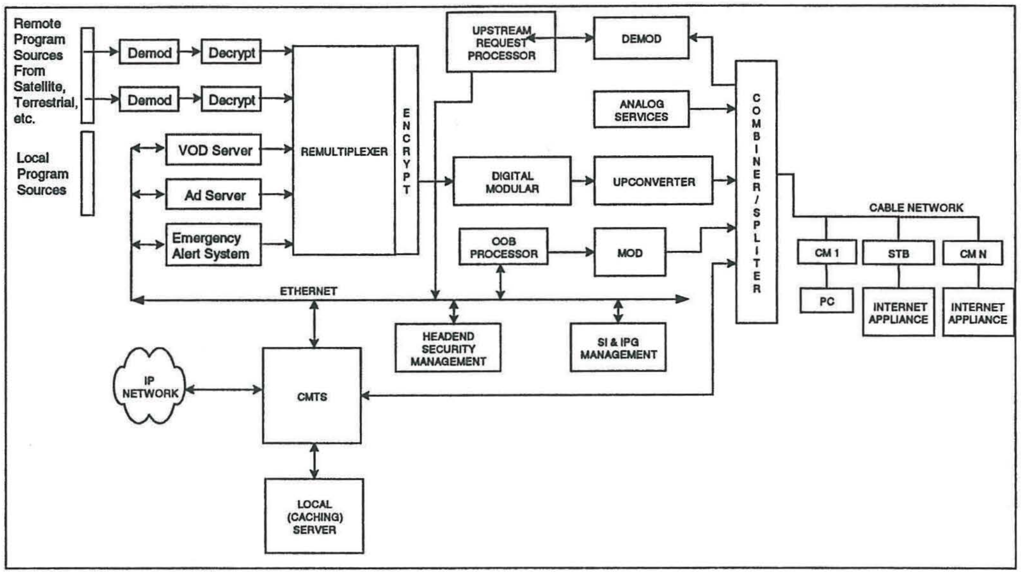


Figure 6: Headend Architecture Block Diagram

Routing involves two basic activities: determination of optimal routing paths and the transport of information groups (typically called packets) through an inter-network, the latter usually referred to as switching. Bridging and switching occur at the link layer, which controls data flow, handles transmission errors, provides physical (as opposed to logical) addressing, and manages access to the physical medium. In general, bridges and switches are not complicated devices. They analyze incoming frames, make forwarding decisions based on information contained in the frames, and forward the frames to the destination. A router looks at the logical address (e.g., the destination address in an IP packet) to decide on the routing path, while the bridge looks at the physical address (e.g., the destination MAC address) to make the decision. We consider only IP-based cable networks here.

As a bridge, the CMTS does not need to open the IP header, it simply forwards the incoming frame from the network side interface (NSI) to the radio frequency interface (RFI) on the downstream. It refers the destination media access control (MAC) address (Figure 7) of the incoming frame when deciding the outgoing interface (called a port) in the case of a CMTS with more than one interface (shown in dotted lines in Figure 9). The bridge CMTS has to maintain a MAC address versus a port mapping table to make the decision. It could build such a table by gleaning the source MAC address of frames (Figure 8) coming into a port, thus associating that port with the MAC address for future reference, a processes called "learning."

In the case of a router CMTS, packets go up to the IP layer (Figure 8). Routers in the Internet run a host of "routing protocols" which gather information about the network topology and maintain "routing tables." A routing table contains information for reaching a particular network. Typically, it could be the address of the next hop to reach a particular network. The last hop router (the router connected to the destination device) will just look at the network part of the IP address to decide on the outgoing interface. A router CMTS could just happen to be the last hop router for downstream, CPE destined packets (refer to Figure 10). Routers replace the incoming frame's destination MAC address with that of the next hop interface, while a bridge does not modify the MAC or IP addresses on transit.

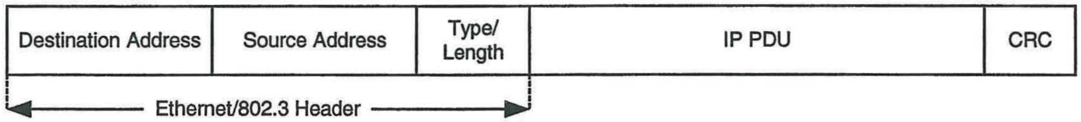


Figure 7: MAC Frame

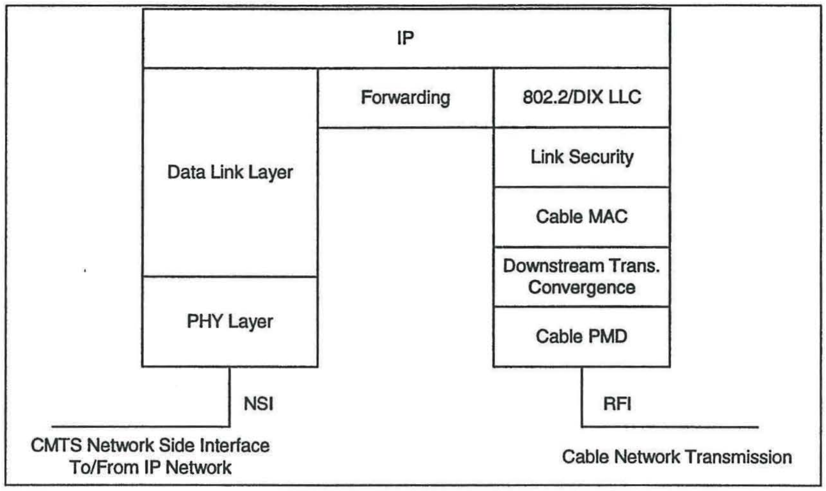


Figure 8: CMTS Protocol Stack

This decision-making process is necessary even in the case of a CMTS with two interfaces (one NSI and one RFI) if the CMTS is to stop packets not intended for the cable network from coming in from the data network (Internet).

As a router, the CMTS opens the IP packet header to check the destination IP address. Based on the network portion of the IP address, the CMTS decides on the interface in which to forward the incoming data. In such an implementation, each of the CMTS interfaces will have to be different IP networks (the network portion of the IP address will be different), while this need not be the case for the bridge CMTS.

If the CMTS finds that the network address portion of the IP destination address (present in the IP header) of the incoming data packet matches the network portion of the interface IP address of any of the CMTS interfaces, the packet is forwarded onto that interface. When the CMTS acts as a router, the incoming frame's destination address (DA) would be that of the network interface card (NIC) attached to the data network. Before forwarding on to the RFI interface, the DA address field in the MAC header has to be replaced with that of the destination's CPE (such as a PC) MAC address. For example, referring to the stack implementation in Figure 7, after stripping the incoming encapsulation, a new 802.2/DIX LLC encapsulation is performed on the RF interface.

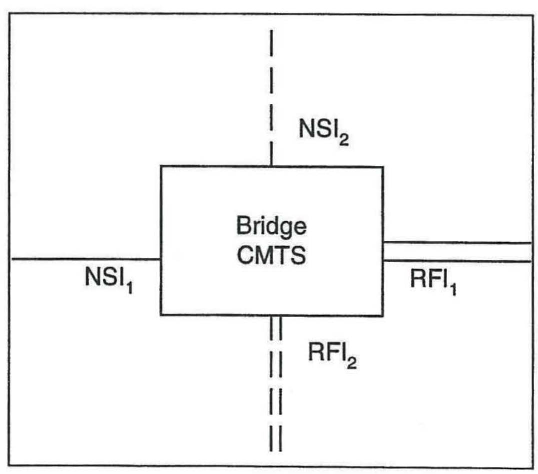


Figure 9: A Bridge CMTS

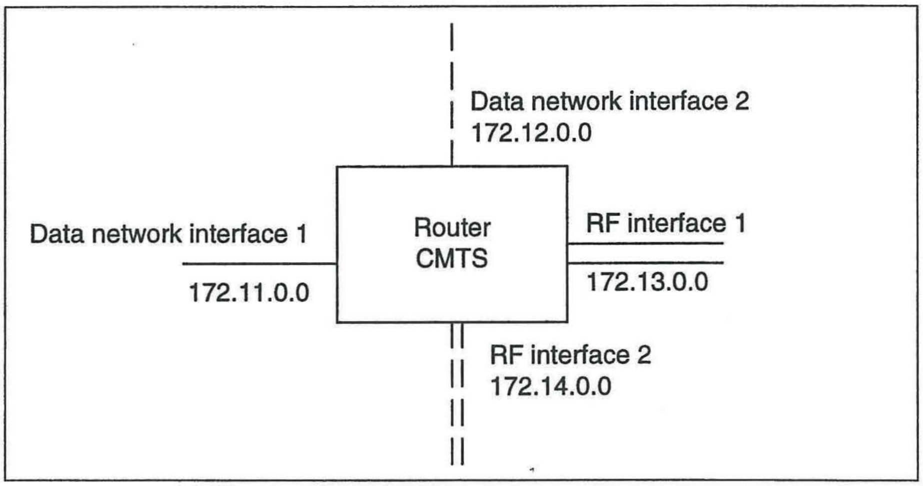


Figure 10: A Router CMTS

As a router, the CMTS could have network implementation with each of the supported interfaces attached to different networks. Therefore, handling IP packets at the CMTS entirely depends on the overall design plan of the network.

FORWARDING TO THE RFI ON THE DOWNSTREAM

Handling IP packets (router CMTS) or Ethernet/802.3 frames (bridge CMTS) on the RF interface before transmission over the cable network depends on the standards to be followed for interaction between the CMTS and the CM. Figure 8 shows the implementation recommended by the data over cable service interface specification (DOCSIS) 1.0 SP-RFI specification. In DOCSIS networks, the CPE is any device with an Ethernet NIC card. Typically, it is a PC used to connect to the IP network for services such as email, web access, etc. The functions of each layer on the RF interface are given below. Each of these layers will encapsulate incoming data.

Logical Link Control (LLC) Sublayer: The LLC conforms to Ethernet standards to provide the data link framing that the CPE's NIC needs. The following layers implement the security, modulation, spectrum management, and analog signal transmission. The functionality of these layers could be mapped as needed to other standards that define communication between the CMTS and the CM.

Link Security: This sublayer is needed to support the basic privacy needs, authorization, and authentication.

Cable MAC Layer: In this layer, the CM removes encapsulation before forwarding data packets to the CPE. It has a set of predefined messages (packets) which direct the behavior of the CMs on the upstream channel, as well as messaging from the CMs to the CMTS.

Some of the main functions of this layer include:

- CM bandwidth allocation for upstream transmission in terms of number of mini slots;
- Class-of-service support using dynamic SID allocation;
- Wide-range data rate support;
- Other features (refer to DOCSIS RFI specification).

Downstream Transmission Convergence Sublayer: This layer exists in the downstream direction only. It provides the means for delivery of additional services (e.g., digital video) over the same physical layer bitstream. This sublayer is defined as a continuous series of 188-byte MPEG transport stream packets. The PID (packet identifier), one of the header fields, identifies the type of payload it is carrying.

Cable Physical Media (PMD) Sublayer: This layer involves digitally modulated RF carriers on the analog cable network. Thus, a data packet from the Internet goes through a series of encapsulations and de-encapsulations to carry control information for each of the layers as overhead, before the IP data finally reaches the destination CPE.

For example, when a user accessing an Internet web-server from a PC (CPE) is connected to a CM by Ethernet, Hyper Text Transfer Protocol (HTTP) packets are transparently taken through the cable network. The CM encapsulates the packets inside a cable MAC header and sends them to the CMTS where they are de-encapsulated and routed out through the correct interface onto the Internet. Similarly, incoming HTTP packets are forwarded onto the cable network after cable MAC encapsulation at the CMTS RF interface. Again, the CM de-encapsulates the cable MAC header at its RF interface and bridges (the DOCSIS CM acts like a bridge) the packets to the attached CPE device. The CMTS replaces the incoming frame's destination MAC address with that of the CPE's. The CMTS could obtain the CPE's MAC address by maintaining the Address Resolution Protocol (ARP) table, which is simply a mapping of IP addresses versus NSI MAC addresses.

A CM can support more than one IP-based CPE, but will have a limit on the number of CPEs to which it can be connected. In that case, the CM should have some mechanism that should filter out packets not destined to any of its attached CPEs.

INTEGRATED ARCHITECTURE

Figure 6 shows a headend architecture where analog TV services, digital TV services, IP-based services and other interactive services have been integrated. The digital baseband signals, after appropriate processing, will be digitally modulated on a 6-MHz radio frequency (RF) carrier, then will be upconverted to desired cable channels. Finally, all the RF channels will be combined for delivery over HFC cable networks. For interactive services or IP-based services, a return path, an upstream channel or a telephone return to the headend will be required. The return signals have to be processed at the headend and user requests have to be parsed and sent to the appropriate servers or other headend equipment. The ad server is used for storing and playing local commercials; the VOD server is used for VOD services. To achieve flexibility in packaging user-selected programming and to minimize the storage requirement of compressed commercials and VOD content, a rate re-multiplexer is a better choice over a standard re-multiplexer.

FUTURE INTERFACE ISSUES

Since the MPEG-2 standard is more suited to the broadcast industry, it does not cover other application areas. To compliment MPEG-2, the MPEG-4 standard is under development and will cover wider application areas. The MPEG-4 standard will be completed in two phases; version I is already under final distribution international standard (FDIS), and version II is scheduled to be completed by the year 2000. MPEG-4 is known as the multimedia standard, and it is anticipated that new content will be created using the MPEG-4 standard in the coming years. The cable industry could deliver that content over HFC cable networks. The MPEG-2 standard has a transport protocol suitable to carry MPEG-2 compressed video, audio and data. MPEG-4 does not specify any such transport protocol. Efforts are underway to deliver MPEG-4 content using other protocols (e.g., IP, MPEG-2, etc.). The MPEG-4 standard will not be compatible with MPEG-2. An MPEG-2 set-top box will not be able to decode MPEG-4 content, but can ignore it in a backward compatible way using MPEG-2 transport encapsulation. Initially, most of the MPEG-4 content is expected to be designed for the Internet and, hence, can be delivered like any IP-based service. Consumers can decode this content using their own MPEG-4 terminals. In the future, it may be necessary to enhance MPEG-2 set-top boxes so that they can decode MPEG-4 content.

CONCLUSION

We have explored the various services and applications that may be delivered over the HFC cable network. These services include existing analog TV services, digital television services, multimedia services, and interactive applications. It was shown that by using digital technology, the HFC cable network can deliver a majority of these services and applications. These services and signals may be classified into three broad categories: analog TV signals, digital TV signals, and Internet (IP-based) signals. Analog TV signals are well understood and exist in all headends. Interfacing the other two signals at the headend should cause minimal or no disruption to analog services.

The equipment used to interface digital TV signals at the headend is known as a remultiplexer. A remultiplexer is not capable of altering the video ES bitrate of the input multiplexes, and video ES require a major portion of the multiplex' bandwidth. An enhanced remultiplexer, known as rate-remultiplexer, incorporates transcoding technology. Transcoding is a technique to reduce the bitrate of the MPEG-2 compressed video elementary streams without fully decoding and re-encoding, which is a complex and potentially perceptually degrading process.

Similar to the remultiplexer, the CMTS plays a major role in interfacing Internet signals to the headend. The CMTS and the CM make the HFC cable network user transparent. How the CMTS handles IP packets between the external IP network and the HFC cable network was explained. The protocol stack used by the CMTS as a router or as a bridge was discussed. How a CMTS may be used depends on individual plant design.

Finally, an integrated headend architecture was presented where DTV services and multimedia services coexist with analog TV services. Interface issues related to content developed using MPEG-4, the multimedia standard, also was discussed.

LIST OF REFERENCES

1. *Rate-remultiplexing: An Optimum Spectrum Utilization Technology*, by Richard S. Prodan, Mukta L. Kar, and Majid Chelehmal, presented at National Cable Television Association Conference, June 1999.

2. *Video Transcoding for Matching Network Bandwidth Constraints*, by Pedro A. A. Assuncão, a Ph.D. thesis, Department of Electrical Engineering, University of Essex, U.K., 1998.

3. Data Over Cable System Interface Specifications – SP-RFI-I04-980724 (http://www.cablemodem.com/).

4. *Internetworking with TCP/IP – Vol 1,* by Douglas E. Comer, Prentice Hall, 1995.

5. *Digital Video: An Introduction to MPEG-2*, by Barry G. Haskell, Atul Puri and Arun N. Netravali, Chapman and Hall, 1997.

6. CODING OF MOVING PICTURES AND ASSOCIATED AUDIO, ISO/IEC 13818-1 (Systems), ISO/IEC 13818-2 (Video).

Correlating Return-Band Impulsive Noise Measurements from Houses with Sheath Current Induction Test Results

Tom Williams

Holtzman Inc.
Longmont, Colorado

Abstract:
Sheath current induction is a method for finding shield breaks in coaxial cable by inducing a broadband test current onto the shield and receiving a center conductor signal when there is a shield break. This test method allows the range to a shield break to be measured, as well as a measure of the severity of the break. This paper correlates the sheath current induction test results for a house with the house's ability to generate impulsive interference with digital upstream traffic. The test method is to first create a known shield break at the ground block and characterize the break by the sheath current induction test. Next, impulsive interference is generated inside the house on the house's power wiring with a noisy 110 volt AC load. This impulsive interference is measured at the tap with a bandpass filter and a totaling counter. Finally, the known shield break is fixed and the test is redone. This test procedure was repeated on a number of different houses.

Background:
HFC (hybrid fiber coax) cable networks are widely expected to assume a major role as a transport medium for two-way high speed data including digital television, internet browsing, e-mail, cable telephony and a host of other new services. Many of the new digital services require a reliable two-way transmission capability, although the upstream bandwidth requirements are typically lower.

The downstream portion of the cable plant, which may extend from 50 to 750 MHz, is both highly evolved and well understood. The tree and branch architecture allows many high-quality copies of the composite downstream signal generated at the headend or hub site to be replicated and distributed to homes. Furthermore, the cable operator has tight control of both the signal level and the quality of the composite signal originating at the headend.

The upstream portion of the cable system, which typically extends from 5-40 MHz, is a different situation. The tree and branch architecture permits noise as well as signals from many locations to be combined into a common signal path. This well-known phenomena is known as "noise-funneling." As a net effect, a noise problem that is generated at any location affects signals from all locations that are supported by a common receiver. Typically a common receiver supports one to several nodes with 500-2000 homes passed in each node. Frequently encountered return problems are common-path distortion (CPD), broadcast ingress, and burst noise.

Burst Noise
Burst noise was found to be a prevalent return impairment in a number of studies [1]

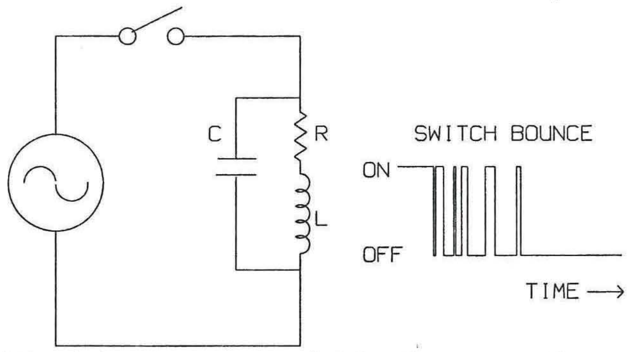


Fig. 1 A reference model for the generation of burst noise with an inductive load and switch bounce

[2]. Burst noise is frequently created when switching off and on inductive loads with mechanical contacts or with electric motors with brushes. The burst energy is typically short in duration but high in energy. Solid-state power controllers also create impulses, usually with repetition rates that are at harmonics of the power line frequency.

Received burst noise typically has the characteristic of having most of its energy constrained to the frequency band below 15 MHz. Because of the high energy associated with burst noise, the energy below 15 MHz frequently has sufficient power to clip upstream active devices, especially Fabry Perot laser diodes. Clipping upstream active devices, although brief, is disruptive to signals in the entire upstream frequency band due to a third order distortion component called cross-compression. (Cross-compression is similar to the cross-modulation distortion that is sometimes observed on downstream cable systems.)

One theory about how burst noise is getting into upstream cable plant is via shield breaks. The burst noise traveling on the AC power line finds its way onto the cable sheath and travels on the sheath until a shield break is encountered. Shield breaks are frequently caused by corroded connectors, animal chews or other mechanical damage, and consumer electronic devices with poor shielding.

Fig. 1 is a reference model of a burst noise generator created by a mechanical switch that bounces multiple times as it opens and closes. The load has an inductance and a resistance associated with the windings as well as a winding-to-winding capacitance. When a current is flowing inside an ideal inductor, the current can only go to zero instantly with an infinite voltage. In practice, the switch arcs and the capacitor rapidly charges to a high voltage. When the switch bounce causes the contacts to re-connect, the capacitor dumps its charge into the power supply lines creating a noise burst. The exact nature of the noise burst depends on the instantaneous voltage on the AC power source when the switch is tossed and the manner in which the switch bounces as it is opened or closed. The spectrum of the burst energy contains significant energy up to the VHF television frequency band. Reference [3] describes the nature of electrical interference generated by arcing contacts.

Sheath Current Induction

The most commonly used method of finding sheath breaks is with signal leakage detection equipment. Unfortunately, signal leakage detection is less than ideal for finding breaks that allow upstream-band noise into the cable plant because conventional signal leakage test equipment uses a single carrier frequency, frequently 108-120 MHz, that is outside the upstream frequency band. Furthermore, a fast Fourier transform (FFT) of burst energy captured at the headend shows a frequency selectivity that can be missed by a single frequency test signal.

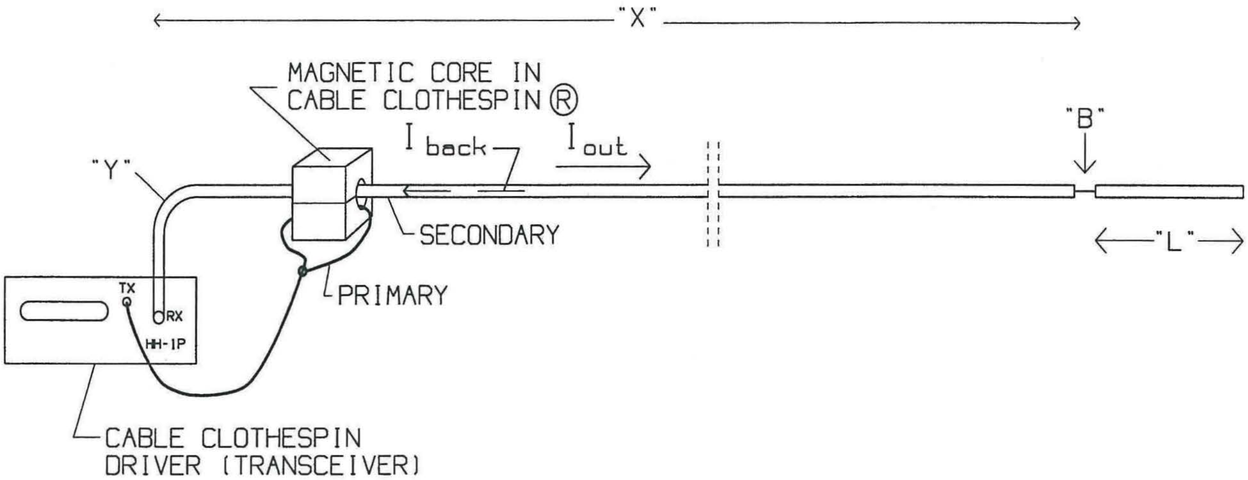


Fig. 2 The principle of testing a shield break via sheath current induction

Figure 2 is a block diagram of a sheath current test on a coaxial cable with a shield break. A technician creates a transformer by clamping a split magnetic core around a coaxial cable. Also included with the coaxial cable inside the center hole of the magnetic core is a wire connected to a reference signal transmitter which generates a broadband test signal. The technician creates a transformer with the wire forming a one-turn primary winding, and the drop sheath forming a secondary winding. The transmitter transmits a reference test signal which is measured by a receiver attached to the center conductor of the coaxial cable signal. If there is a shield break, some of the test signal will enter the inside of the coaxial cable and propagate to a receiver.

The received test signal can be analyzed to determine which frequencies are preferentially allowed into the cable, the total returned energy in the test signal, and an approximate distance to the shield break.

With sheath current induction, there are several methods for transmitting, receiving, and analyzing the test signal. The test method used for this paper was to determine the impulse response between the transmitter and receiver by transmitting a pseudonoise (PN) sequence, and performing a cross-correlation between the transmitted and received signals. A device called the Cable Clothespin® contains the magnetic core that is attached around the cable. A hand-held transceiver called a Cable Clothespin driver contains both the transmitter and receiver (transceiver). This equipment is illustrated in Photo 1. The Cable Clothespin driver may be attached to a personal computer (PC) to download the impulse response. The time delay between correlation samples is 20 ns. giving a resolution of 2.5 meters to a shield break. The PC performs a fast Fourier transform on the impulse response to show a frequency response associated with the shield break.

Photo 1.
Sheath current induction test at a tap

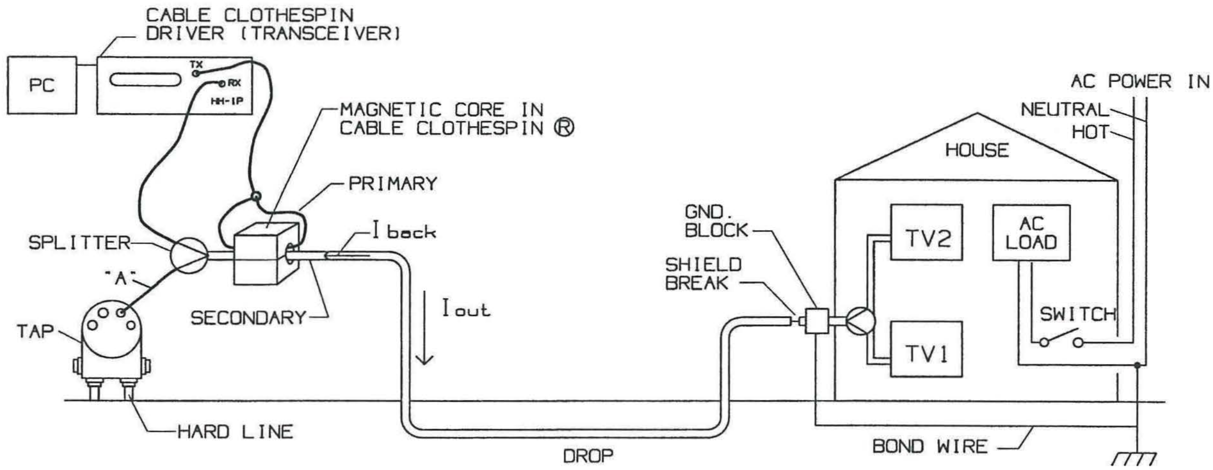


Fig. 3 Sheath current induction test for a house from a tap location

Reference [4] is an early paper written on field experiences with sheath current induction. The tester used to gather data for this earlier paper was a DSP-based complex frequency response measuring device called a Cable Scope® that employs a separate transmitter and receiver.

The Question

The question that this paper attempts to answer is: "Are the shield breaks that are typically found by the sheath current induction technique actually capable of causing packet errors in upstream data services?"

Test Method

The technical approach used to assess the capability of shield breaks to cause a disruption to upstream data was as follows. First, a known shield break near the ground block was created by inserting a 30 cm. piece of coax with a 3 mm. section of its shield removed.

Second, the shield break was characterized with the sheath current induction method using the wiring diagram shown in Fig. 3. The sheath current was injected at the tap location and the results were recorded on the laptop PC. A splitter at the tap was used to supply signals to the subscriber while the test was being performed.

Third, one operator went inside the house and created an interference on the AC power line while another operator at the tap measured the interference heading upstream. The block diagram for this process is illustrated in Fig. 4. This test utilizes a bandpass filter and a totaling counter with an accurately set threshold. Its operation will be described later in this paper. The level of the interference was measured for successively higher threshold levels.

Finally, the cable break was fixed and the test procedure was repeated.

Unfortunately, no one simple set of test conditions can be found that apply equally well to all possible upstream field situations. Significant variables are:
a. the nature of the AC power load and switching creating an interference on the power line
b. the type of upstream modulation that will be contending with the interference
c. any forward error correction that the upstream modulation will be using
d. whether the drop is aerial or buried

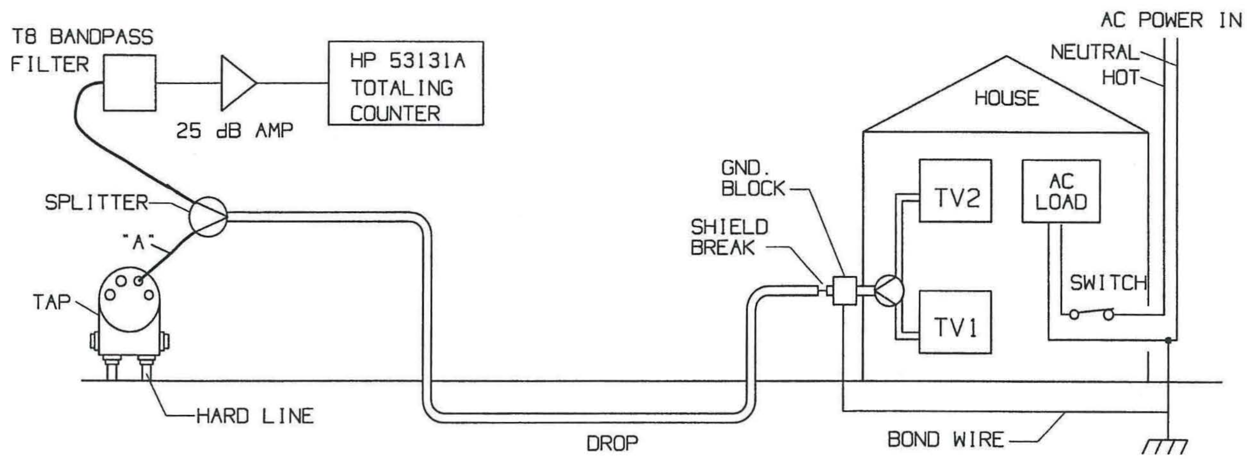


Fig. 4 Impairment test with a bandpass filter and a totaling counter

e. the frequency that the return carrier will be using
f. the bandwidth of the return carrier
g. the transmit level of a return carrier which is influenced by system design as well as the value of the tap at a given location.

For the sake of this test, a power relay with 30 amp contacts and a 110 volt AC 60 Hz coil was wired to "flutter". This was accomplished by routing the AC power supply through the relay's normally-open contact and then to the coil. As the relay energizes and starts to break the normally-open contact, the coil is temporarily de-energized. This causes the relay's spring to pull back, re-connecting the coil. The noise-generating load was plugged into a kitchen duplex power outlet in each house under test.

For the sake of this investigation it was assumed that the modulation that will be used is QPSK (quadrature phase shift keying) occupying a bandwidth of 6 MHz. QPSK is used for MCNS-DOCSIS modems and is a commonly-used modulation on upstream cable systems. The frequency chosen was T8 (11.75-17.75 MHz.) because this frequency was low enough to experience burst noise energy but high enough to be considered for use by an upstream data transmission service.

Since the transmit level of a home terminal device was unknown, a curve of symbol error rate versus the transmitter's output power was generated.

Measuring the Level of Interference
The method used to measure the level of interference heading upstream from the house is based on a bandpass filter with a known bandwidth connected to a high-speed totaling electronic counter. The sensitivity of the counter was set to increment the count value when the instantaneous input voltage level was sufficient to have caused a QPSK signal to make an error. Patents are applied for on this technology and on sheath current testing.

A qualitative description of how this technique operates is as follows. Fig. 5 has four views. View A shows four QPSK constellation points on an I-Q (in-phase, quadrature) diagram. This constellation diagram is made by demodulating a QPSK signal to DC as a baseband I signal and a baseband Q signal and sampling the signals at the correct time. An error can be made in reading a symbol if an impairment, such as the continuous wave impairment or the impulsive impairment illustrated push a constellation point across a decision threshold. The information that is desired for impairment testing is an estimate of the amount of time a signal spent over a threshold line.

Fig. 5 View B is the same as View A but the underlying QPSK signal has been removed, so the origin (I=0 volts and Q=0 volts) is the expected position without any additive impairments. A threshold region can be established at appropriate voltage, "R", and what is now of interest are any threshold crossings. Fig 5 View C shows what the impulse impairment in View B would have looked like if it had not been demodulated to baseband. The spinning rate

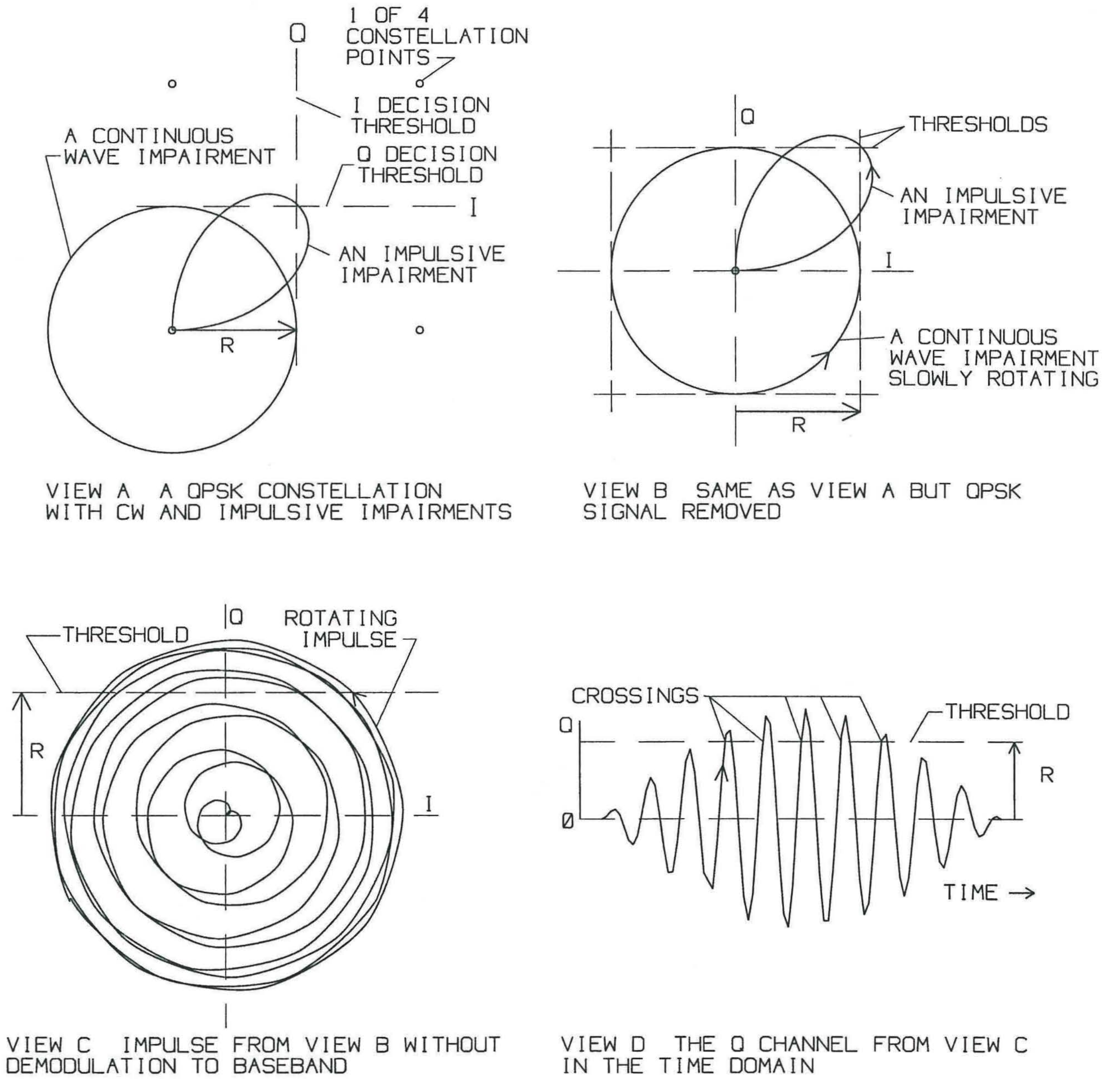


Fig. 5 I-Q Diagrams Illustrating how impairments can be characterized with a bandpass filter and a totaling counter.

of the impulse is approximately the center frequency of a bandpass filter that passed the impulsive impairment. View D is a temporal plot of the Q component of the signal in View C. By setting up a high-speed totaling counter with an accurate trigger level "R", the threshold crossings can be counted. In turn the threshold crossings can be used to estimate the amount of time an impairment spends over a decision threshold. The two variables that are important for meaningful results are the accuracy of the threshold voltage and the bandwidth of the bandpass filter that passes the impulsive energy. The bandpass filter's bandwidth should be the same as the bandwidth of the data service being tested, and the center frequency of the bandpass filter should be the center frequency of the data carrier.

This technique can also be used to test vacant downsteam channels.

Results of Sheath Current Testing

Five homes were tested for sheath current. The subdivision in which all 5 homes were located requires underground drop cable. The system was not 2-way active. House 1 was connected to the tap via a drop laying on the surface of the ground. House 2 deviated from the test plan in that the shield break was supplied by a pair of Labrador retrievers that liked to dig and chew. The homeowner did not want the damage repaired, so there is no data on how the wiring would have performed with the break fixed. Houses 3 and 5 were tested twice, once with the buried drop cable and once with a temporary aerial drop cable strung through the foliage. The dual testing was done to compare differences in sheath current results due to the presence of soil around the drop line. Table 1 lists the gross power readings for the 5 houses with and without shield breaks. The gross power was computed from the impulse response. The dB readings are relative, but absolute gross attenuation can be computed by subtracting each reading from eighty-five. Eighty-five is the effective reading obtained with a direct connection between the transmitter and receiver.

House Number	Reading Shield Broken	Reading Shield Fixed
House 1 (drop on gnd.)	48.0 dB	5.2 dB
House 2 (dog chews)	61.2	N.A.
House 3 (buried drop)	29.2	4.4
House 3a (aerial drop)	56.5	27.5
House 4 (buried drop)	37.4	-6.3
House 5 (buried drop)	47.2	11.5
House 5a (aerial drop)	60.1	23.0

Table 1 Sheath Current Gross Power Results

The impulse responses for each house with the shield broken and the shield fixed are shown in Figs. 6-12 . The vertical scale on each plot should be noted. The most striking observation is that burying a long drop produced a large reduction in the gross power readings, primarily because the high frequency portion of the burst energy is attenuated. A break at the end of a long aerial drop typically produces a lower gross power reading than a break at the end of a short drop because of skin-effect and radiation which attenuate the high frequency test signals traveling on the outside of the cable. However, this characteristic is exaggerated by burying the drop in soil, possibly because of soil conductivity. Another observation is that house 5 had non-trivial shield breaks beyond the intentional break introduced at the ground block. The homes in this neighborhood are large and typically have multiple splits feeding multiple television sets.

Unfortunately, signal leakage equipment was not available, so no correlation could be made between signal leakage and sheath current induction results.

For comparison, the results of sheath current testing in other systems is presented in Fig. 13A and Fig. 13B. These plots are histograms of gross power readings without introducing any intentional breaks on the cable sheaths. Sheath current injection was performed at the taps in Oct. 1998.

What is concluded is that all of the introduced breaks are visible on the impulse response plots, all of the gross power readings are much lower with the introduced break fixed, and a long buried drop cable attenuates the sheath current.

Results of Impulsive Noise Testing
Fig. 14 is a plot of symbol error rate versus transmitted signal level with shield breaks in place. It can be seen that for an uncorrected symbol error rate of 10^{-4} a transmit level from the houses of between 28 and 39 dBmv will be required. Fig. 15 shows what the curves looked like when the intentional break was fixed and table 2 tabulates the improvement. Note that house 5 still had a break. A symbol error rate of 10^{-9} indicates no errors.

	Improvement
house 1	>44 dB
house 3	39 dB
house 4	33 dB
house 5	22 dB

Table 2. Improvement when the intentional shield break is fixed

Observations
One observation that was made is that common path distortion was frequently observed on the hard line, and the splitter had to be disconnected from the tap at some locations to keep the return-band noise in the hard line from contaminating the test results.

Conclusions
This paper shows that the intentionally introduced sheath break at the ground block can be seen by the sheath current test method, although the sheath current test signal is diminished by a long buried drop. It also shows that the sheath break can be responsible for impairments in upstream data traffic coming from the tested home, or any other home connected to the common upstream receiver for that matter. If the home terminal transmit power levels are low, the symbol error rates will be higher.

References
[1] Two-Way Cable Television System Characterization, Cable Television Laboratories Technical Report, 1995, pp. 123-126

[2] Results of Return Plant Testing, 1997 NCTA technical Papers, by R. Prodan, M. Chelehmal, and T. Williams, pp. 142-165

[3] Noise Reduction Techniques in Electronic Systems, 2nd edition, Henry W. Ott, pub. John Wiley and Sons, Chaper 7

[3] A Method to Locate Nodes with High Return Band Ingress, by T. Williams, CED Magazine, Nov. 1997

[4] Testing Coaxial Shielding Integrity by Sheath Current Induction, by T. Williams and D. Bell, Cable-Tec Expo. Proceedings Manual, 1998

contact: tom@holtzmaninc.com
303-444-6140 (ph.)

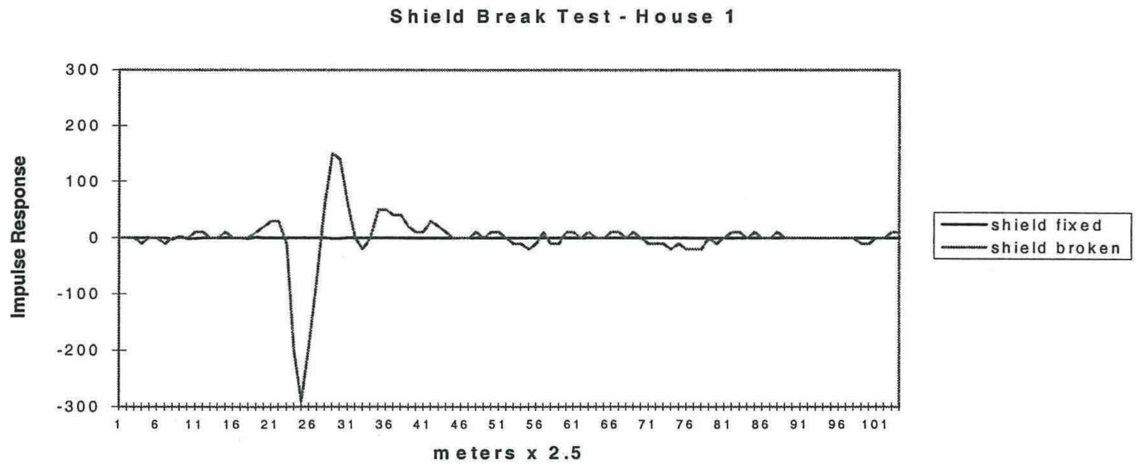


Fig. 6 Drop laying on ground. Gross power = 48 dB broken, =5.2 dB unbroken

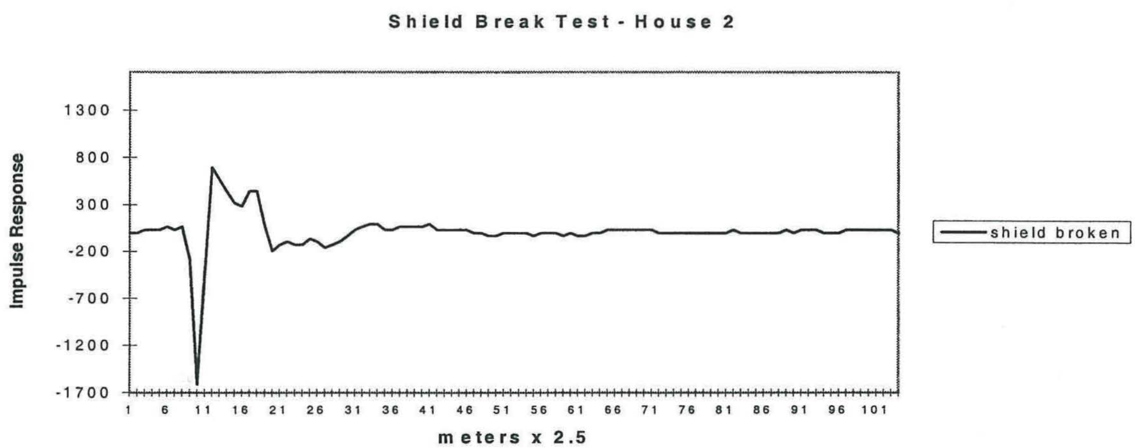


Fig. 7 Dog-chewed drop cable. Gross power =61.2 dB broken, N.A. unbroken

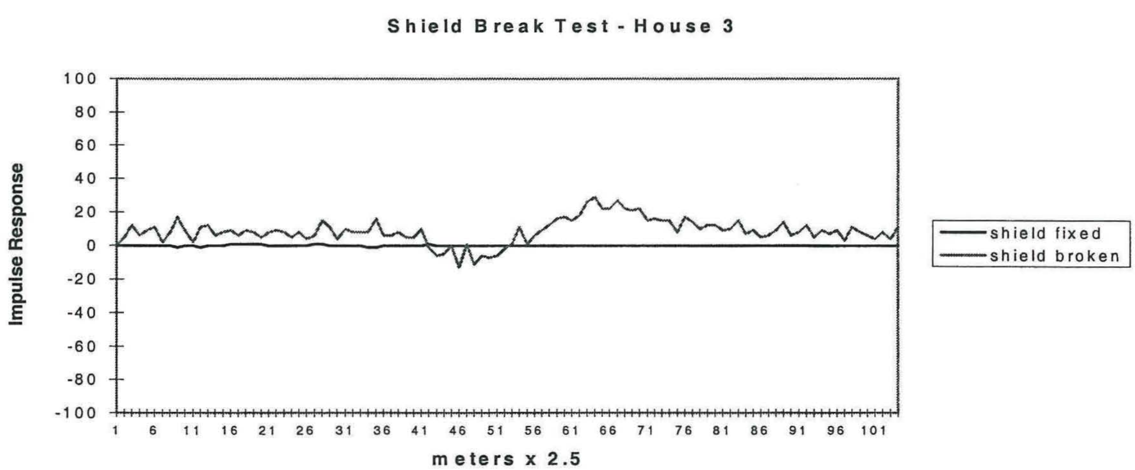


Fig. 8 Long buried drop (142 ft.). Gross power =29.2dB broken, =4.4 dB unbroken Noise on plot caused by ingress from shield break.

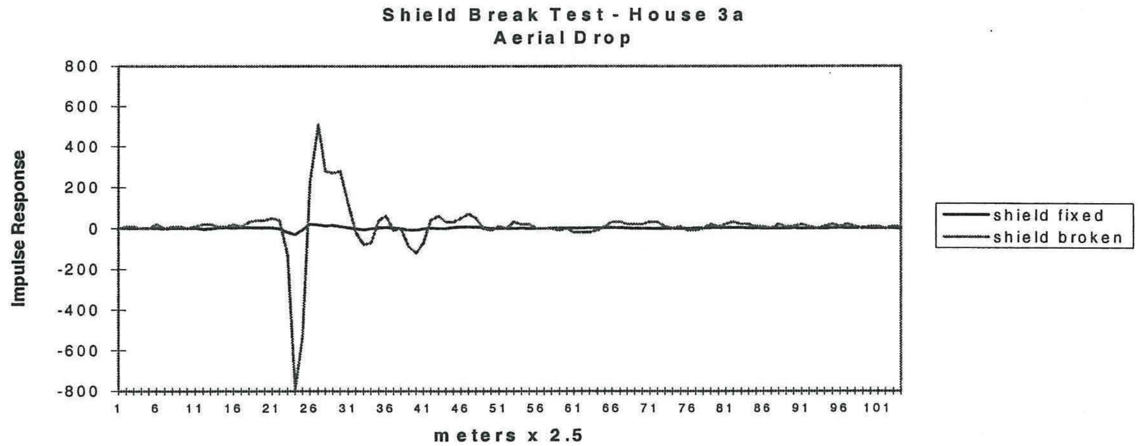


Fig. 9 Repeat of Fig. 8 with an aerial drop. Gross power =56.5 dB broken, =27.5 unbroken. Note large increase in readings over buried drop readings.

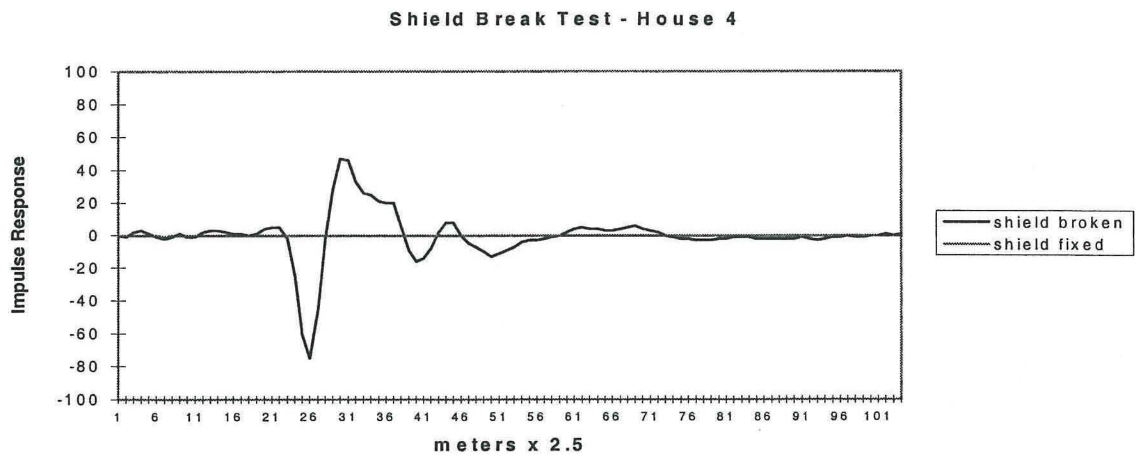


Fig. 10 Gross power =37.4 dB broken, =-6.3 dB unbroken

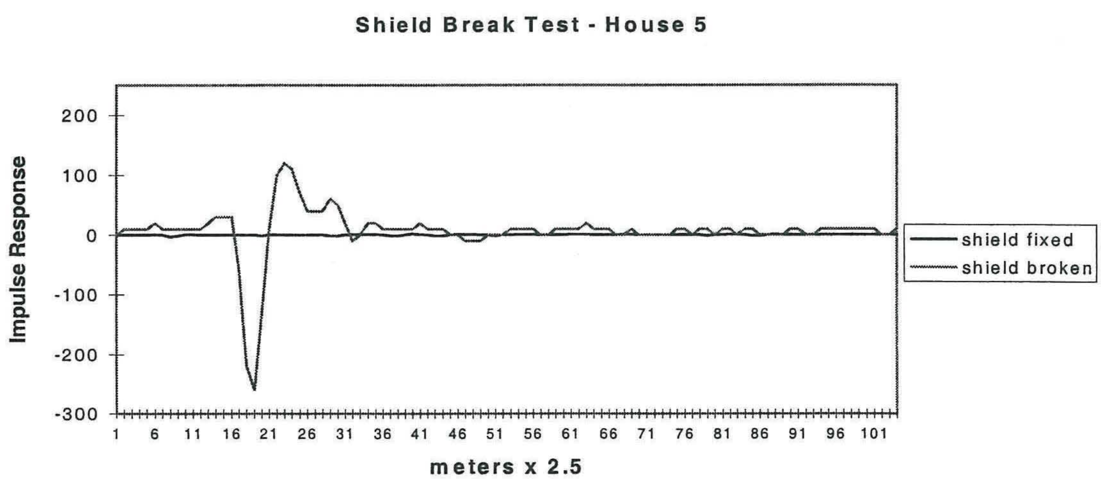


Fig. 11 Buried drop (≈50 ft.). Gross power =47.16 dB broken, =11.54 unbroken

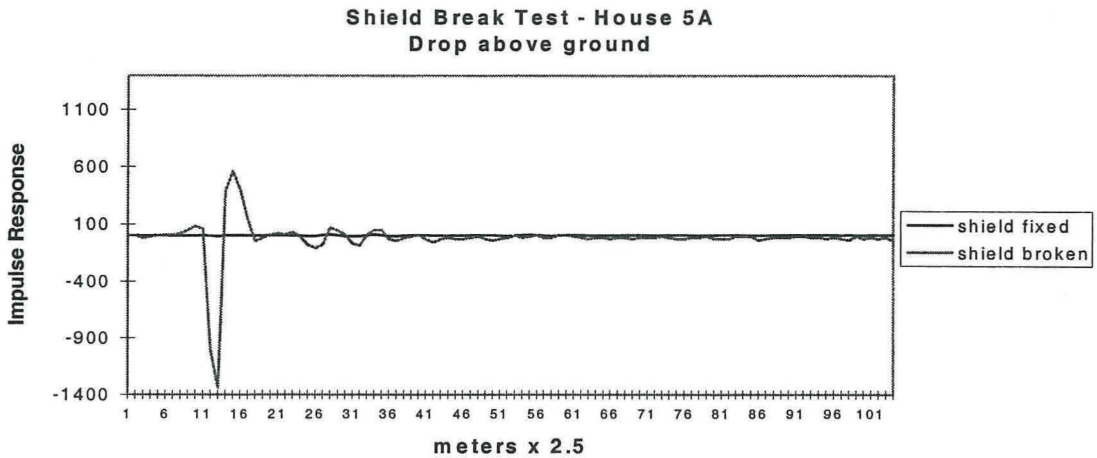
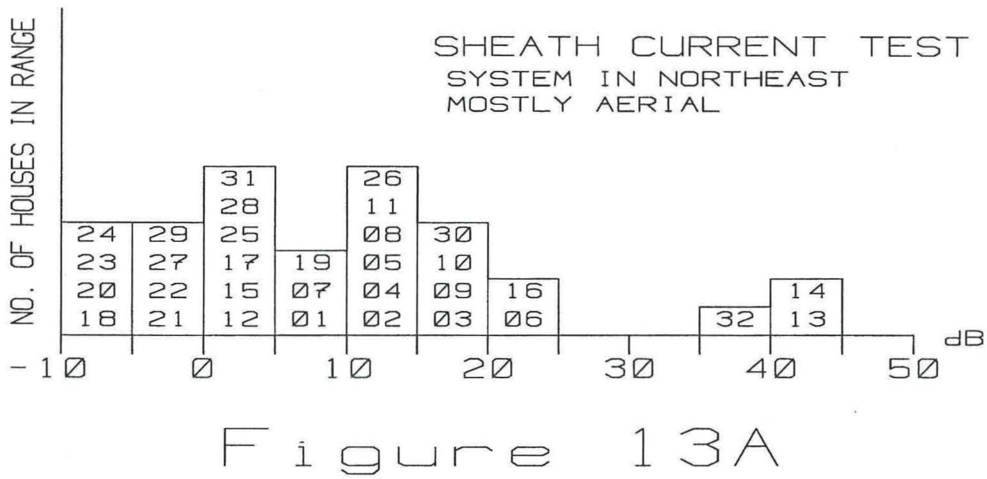


Fig. 12 Re-test of Fig. 11 with an aerial drop. Gross power =60.1 dB broken, =23.0 dB unbroken



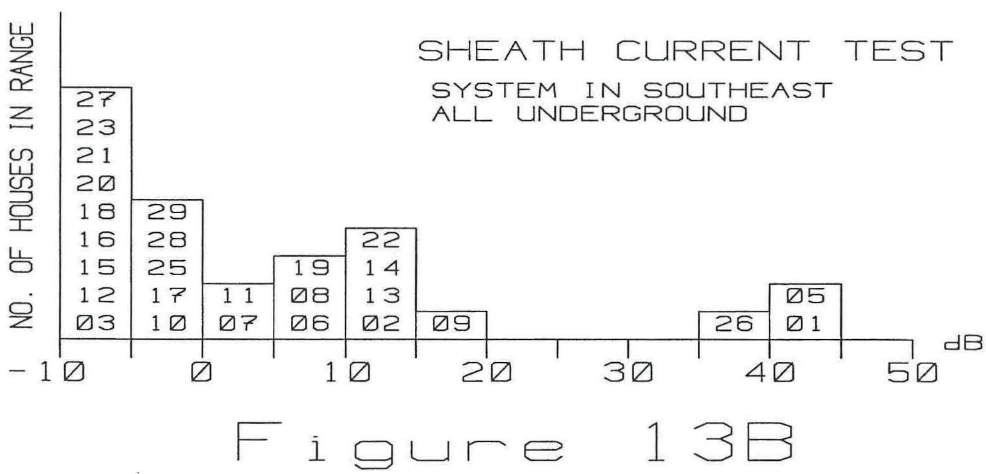


Fig. 13 Histograms of sheath current induction tests done in other cities. Shield breaks were not induced.

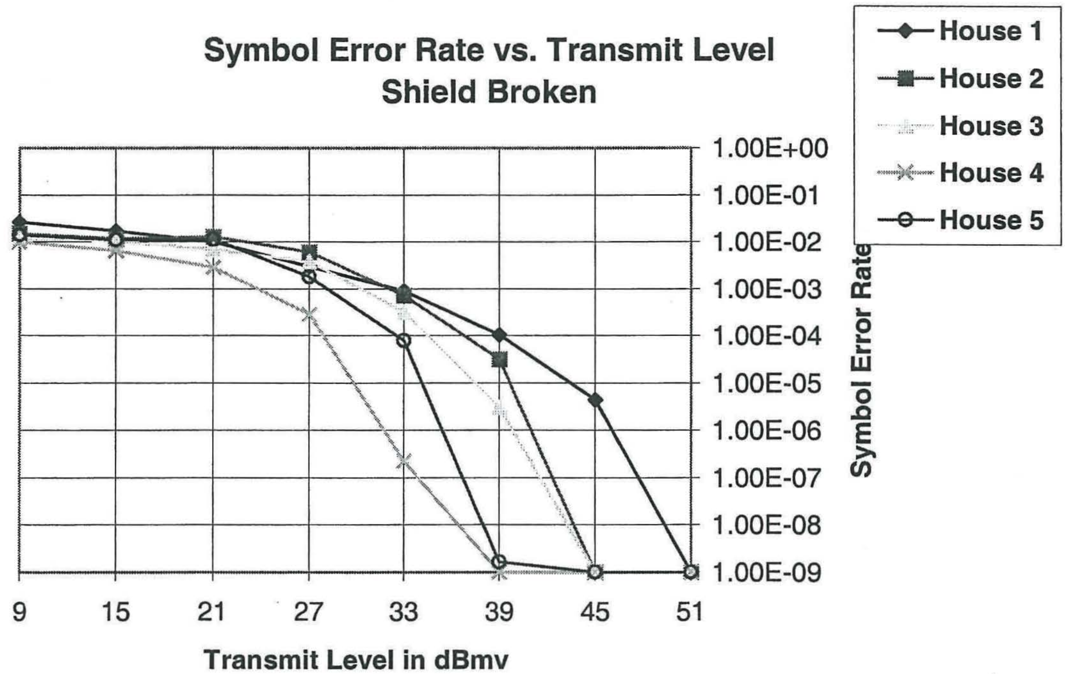


Fig. 14 Required QPSK transmit level at a tap port for a given symbol error rate in the presence of burst noise generated on house's AC power wiring. Shield intentionally broken at the ground block.

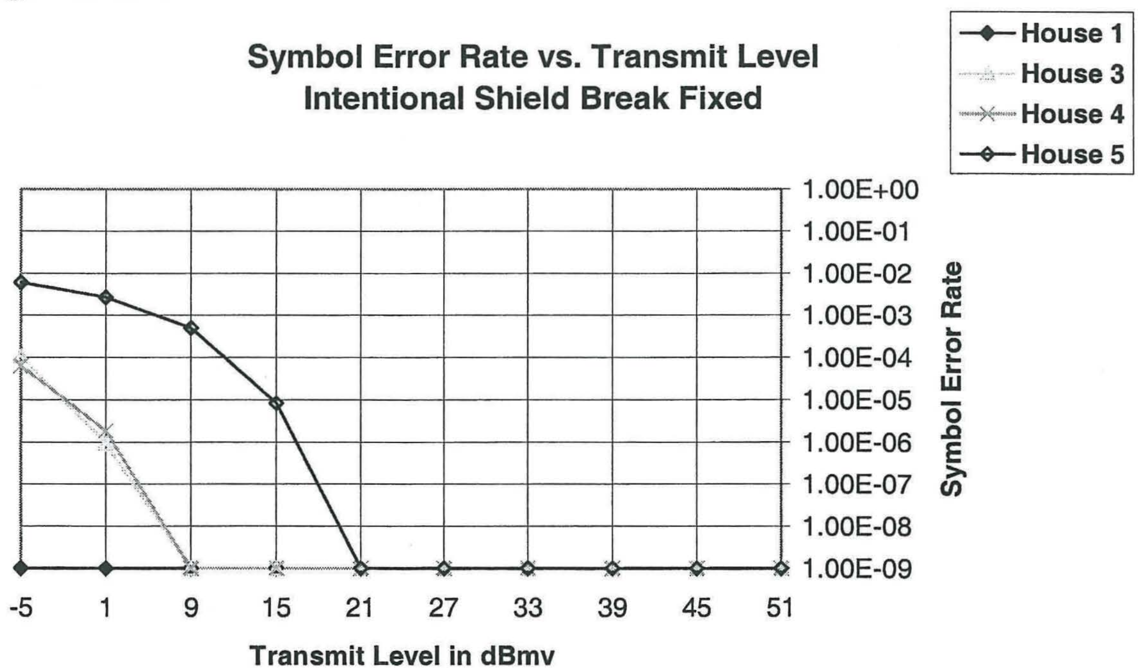


Fig. 15 Required QPSK transmit level at a tap port for a given symbol error rate in the presence of burst noise generated on house's AC power wiring. The intentional shield break at the ground block is repaired.

CUSTOMER MANAGEMENT SOLUTIONS FOR DOCSIS

Daniel Reilly
Solect Technology Group

Abstract

Operators, hardware vendors and billing companies are actively developing provisioning solutions to compliment the value of DOCSIS interoperable platforms. The value of this development will be realized in the form of reduced overhead support costs and higher penetration rates. Cable operators that include the necessary provisioning applications in preemption of DOCSIS will have the means to extract the value from this standardization. Those that do not address this issue will be faced with delivery and service problems that will diminish penetration rates and subsequent revenues.

INTRODUCTION

As cable companies continue to evaluate business and implementation strategies for the upcoming Data Over Cable Service Interface Specification (DOCSIS), it is clear that the companies which will realize the greatest market opportunity and highest penetration levels are those that can provide the most diverse services to the subscriber. Typically, this value proposition is correlated with local and national content, pricing strategies and quality of service. Given the advent of DOCSIS and the eventual retail availability of cable modems, this value proposition must include a viable Customer Management solution that will streamline customer registration and empower the subscriber to utilize a wide array of products and services. As such, a Customer Management solution can not be viewed simply as billing application, but as a means to reduce costs and increase profitability.

Even though a paradigm shift has not occurred in the market, it is a foregone conclusion that all service providers will be faced with new market competition, whether from ISPs, DSL or wireless providers that will offer high-speed access services which include end to end provisioning solutions. When faced with this scenario, cable companies will be competing on an uneven playing field that will result in lower penetration rates and revenues. Clearly, a cable modem will offer a higher level of service than most competing services, but for many consumers, convenience and simplicity of service activation will be the deciding factor in selecting an IP service provider. If given the choice between a service that takes two weeks or more for installation and one that offers self-registration, the shortest path to service activation will generally be chosen.

The interoperability standard set forth by DOCSIS does provide an economic model that will benefit both the cable operator and subscriber through the retail availability of cable modems. This will effectively reduce the consumer price levels of the cable modems, but it does not solve the problem of remote activation and provisioning of cable modems. Eliminating the price barriers associated with cable modems will be a significant step in driving penetration, but the next critical step in this process is developing the back-off systems and interfaces that will simplify customer activation, provide up-sell potential to new services and maximize market penetration. As the competitive market continues to migrate to creative services, such as VoIP, Customer Management solutions will play a more significant role in customer acquisition, retention and revenue.

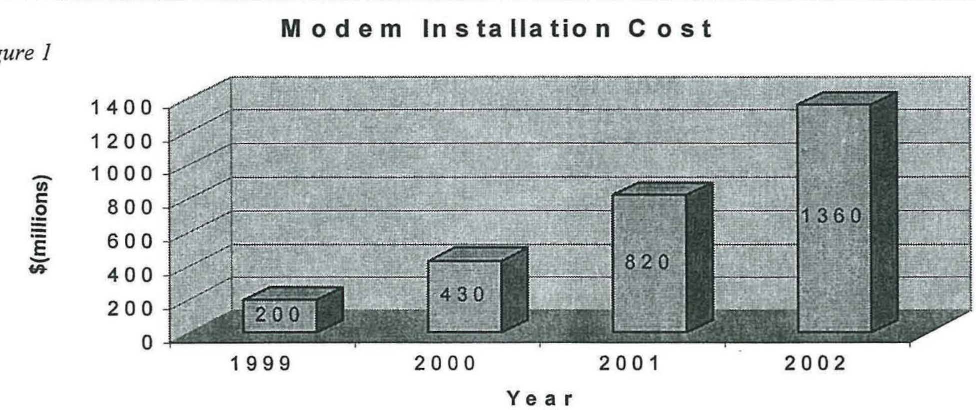
COST TO SERVE

Market Statistics and Risk

Depending on which study you read, it is generally accepted that cable modems will be highly successful in capturing IP market share. By some estimates, such as those found in the Forrester Report, it is speculated that over 10 million cable modems will be installed by the year 2002.[1] DOCSIS and the retail deployment strategy will facilitate this astounding growth, but this is based on the assumption that downward pressures on price, including access rates and equipment costs, will dictate penetration rates. Nevertheless, it must be stated that the greatest risk to the future success of cable modem service will be the inability of a cable operator to satisfy market demand. If cable modems do saturate the market, the single most important variable cost that is not factored in this equation is the expense of service activation and 'truck rolls' that are necessary elements in the current process. Even though this cost is in many cases passed on to the consumer, the future competitive environment may eliminate this option.

For any given cable system that is offering a high-speed data service, frequent 'truck rolls' to either upgrade a subscribers coaxial plant, install an Ethernet card or simply provide cable modem activation is, and will continue to be commonplace. For purposes of this paper, it is assumed that the average cost per truck roll is approximately $100. Given this conservative cost estimate, the total aggregate overhead costs associated with cable modem activation by the year 2002 could exceed $1.3 billion. If the end to end work flow were to be added to this estimate, including calls to a Customer Service Telephone Center to initiate a work order for new service, a more realistic cost estimate would be at least $1.5 billion. *Figure 1* provides an overview of this analysis.

Figure 1



Aside from the overhead cost estimates associated with 'truck rolls' and Customer Service intervention, a secondary issue that cannot be ignored is the availability of in-house and third party field technicians to meet market demand. It has been estimated that without a means to provide remote modem provisioning capabilities, that many customers will have to wait weeks for modem installation. If this is the case, many customers will simply search for solutions and services that can be delivered now, and not in weeks or months. Without a viable solution to these issues, many perspective high-speed modem customers will opt for the readily available IP solution, which in many cases could be a competing IP service provider.

CUSTOMER MANAGEMENT SOLUTIONS

Overview

The ideal Customer Management solution is one that will not only minimize human intervention by the cable company, but also one that will provide a fully automated system to properly manage an IP subscriber. This process is dependent on empowering the subscriber to self-manage an account, from the selection of particular modem to the choice of service levels and finally to the payment method.

The current process is one that is driven by manual activities that are neither cost effective or efficient. The challenge that the cable industry is faced with is implementing the necessary systems and provisioning applications that will create a self-managed end to end solution. The deployment of a self-provisioning solution will not eliminate all of the manual processes. For example, some subscribers could be data-only customers that will require a cable drop and plant upgrades. Nevertheless, the basic goal of a provisioning solution will be to minimize overhead costs, lessen stress on human processes and create additional value to the end-user.

In a DOCSIS environment, the strategy and benefit of this standardization will be to have a variety of interoperable cable modems in retail outlets. As such, the shared vision is that a prospective customer can purchase a cable modem at the retail store, take it home, plug it in the computer, the cable outlet, create a new account and initiate service. Even though this process is not much different than what is currently offered by dial-up IP service providers, it is made more difficult by the wide range of system configurations, interfaces and services that can be offered through a high-speed cable service and mutually exclusive cable systems. Clearly, the intention of the interoperability model is to minimize these differences, but the reality is that the challenges and final solutions will be more complicated than those found in the dial-up world.

An example of these differences can be seen in Figure 2.

Figure 2

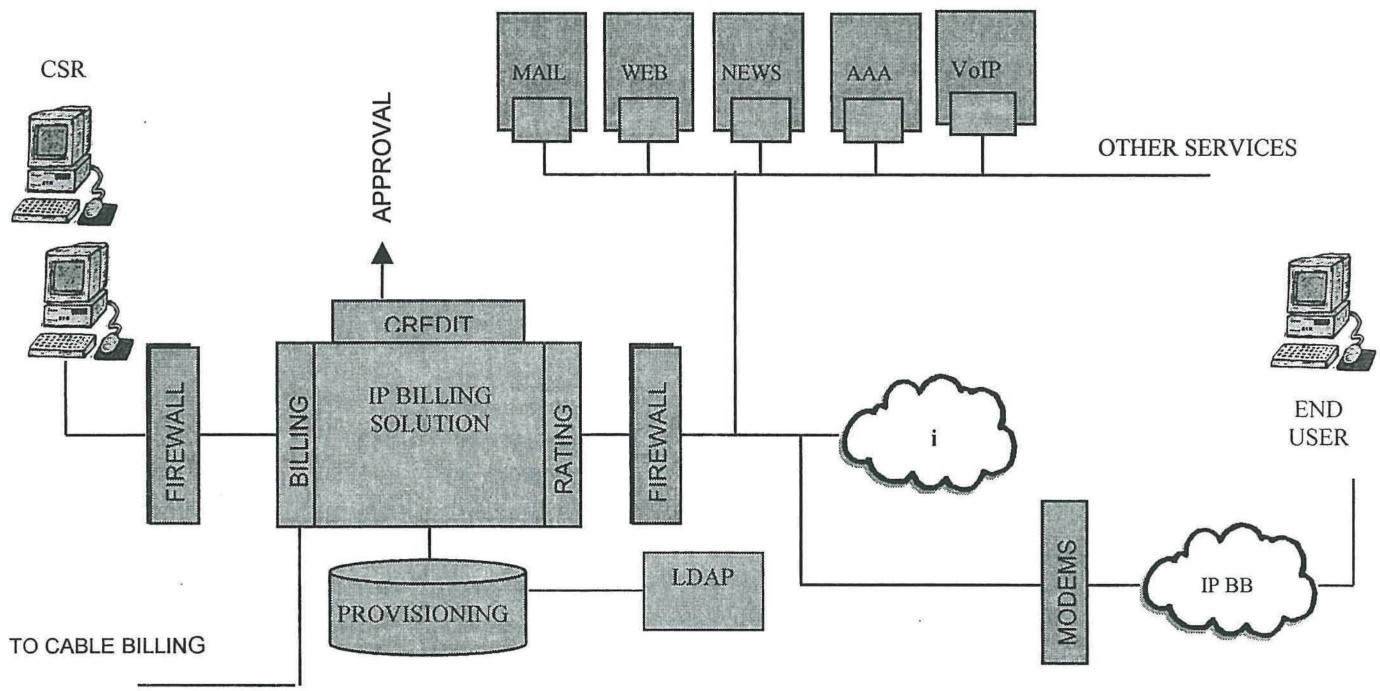
Dial-Up Registration	**Cable Modem Registration**
User-Interface for Auto-Registration	User-Interface for Auto-Registration
Account Creation to include:	Account Creation to Include:
Domain Name	Domain Name
Demographics	Demographics
E-mail Account	E-mail Account
Web-Hosting	Web-Hosting
Ancillary Services(News)	Ancillary Services(News, VoIP, VOD, et..)
Credit Card Verification	Credit-Card Verification or Single bill for video, data and telephony.
Service Level	Service Level (Usage, Flat-Rate)
Enable/Disable Account	Cable Modem Registration
	Remote Configuration for QoS
	Work Flow
	MAC address acquisition and collection
	Enable/Disable Account

Provisioning Attributes

Given the expected proliferation of cable modems in the market, it is clear that the various customer-care and billing companies will be asked to develop and deploy new interfaces, platforms and applications that will facilitate modem provisioning and customer self-management within a cable environment. In the future, a core functional requirement of both the IP service provider and end-user will be a solution, or service, that will deliver the following items:

- Remote online-modem registration. This will minimize the overhead costs associated with additional technicians and CSRs. Only those subscribers with plant upgrades will require a truck roll.
- Online bill presentment and Customer self-care modules. This will minimize Customer Service intervention.
- Interfaces to a DOCSIS certified headend control unit. This will ensure interoperability for new cable modem subscribers, and for those that migrate to other cable systems.
- Rating capabilities for usage and a wide variety of services. The expectation is that flat-rate billing is not a viable future solution.
- Interfaces to existing Legacy systems. This will provide congruence and consistency for both the operator and subscriber for billing and new service offerings.
- Rapid time to market for new services, such as VoIP.

The inclusion of these elements in a service offering will give the cable operator a competitive advantage in the IP market. It provides the flexibility that will be required to capture and retain customers, react to market pressures and provide a cable modem subscriber the means to self-manage an IP account. The following illustration provides an example of the diverse services, systems and interfaces that would be inclusive in this solution.



CONCLUSION

By the end of 1999 it can be assumed that multiple provisioning solutions and methodologies will be available for cable operators. The reality is that the value of DOCSIS will not be leveraged until prospective cable modem customers have the ability to 'plug and play'. This fact is well known by the operators, hardware vendors and billing suppliers. As such, it is imperative that these applications be developed, tested and deployed even before DOCSIS begins to drive cable modem penetration rates.

In recent months the IP industry has seen emergence of several new DSL and wireless IP service offerings targeted at prospective high-speed customers, both residential and commercial. In a sense, the land-grab has begun and those IP providers that can provide:

- A rapid time to market for the consumer
- Autonomy and self-account management.
- Creative new services as they become available.
- Greatest value for the money.

The companies that do satisfy these requirements will end up with the largest portion of the property at end of the day. Those IP companies that to do not enter the market with these provisioning capabilities, will spend millions of dollars in customer acquisition and retention.

[1] Forrester Research, Inc., "Broadband Hits Home", August, 1998

Deep Fiber Networks: New, Ready to Deploy Architectures Yield Technical And Economic Benefits

Donald Sipes and Bob Loveless
Scientific-Atlanta, Inc.

Abstract

The cost of running fiber deeper—to the extent of achieving passive and near passive cable architectures has been dramatically reduced due to recent advances in HFC access network components. This paper explores various technical and economic aspects of "deep fiber" architectures, and how the rapid evolution of 1550 nm EDFA's, optoelectronic receivers, passive optics and an optimization of HFC network elements such as transmitters, nodes, and the return path can now yield substantial network improvements for a deep fiber environment.

ADVANCED SERVICES: THE NEED FOR BANDWIDTH

The advent of digital technology in HFC networks has opened up a myriad of opportunities for MSO's. MPEG compression of video streams has allowed for a near order of magnitude increase in the video capacity of a single channel such that more and varied basic broadcast programming can be offered in the same channel space. Video Servers have made possible the distribution of prior pay per view movies in the near-video-on-demand (NVOD) format and the soon to be available video-on-demand (VOD) service. Video in the HDTV format is now beginning to be available from broadcasters, and its availability will only increase with time.

In addition, digital technology has allowed for non-video services to flourish. The DOCIS standard has created an avenue for receiving high speed internet access, and both circuit switched and IP based telephony services are either now available or in advanced trials. Synergistic services, which make use of voice, video, and data simultaneously are in the concept and advanced development stages and should be available within 5 years.

The introduction of these advanced services comes at a cost: namely, the need for increased capacity, and especially increased reusable bandwidth. In HFC networks all services are ostensibly broadcast: the prime difference between services being the footprint over which these services are broadcast. Channel Lineups for broadcast video services typically cover the largest area. Advertising zones are typically second, usually on the order of a typical 20 k home hub. For initial penetrations for high speed data services such as cable modems, a typical hubsite will be divided into several sectors using a single 6 MHz channel. Telephony services are broadcast over the smallest area, typically a 6 MHz Channel for each node. Naturally as penetration of these services increase, the broadcast area for each of these services will also decrease.

Video: Multiple Customer Types

As interactive digital set top boxes are deployed, and as digital ready TV's become available, the types of customers served by MSO's will increase. An MSO today really only has to be concerned with two types of subscriber. Those with set top boxes and those without. With digital service rollouts the types of customers will increase to cover the permutations of those subscribers with analog or digital TV's and those with analog, digital or no set top box at all. Since Cable TV is still a fairly high penetration service, MSO's will still need to offer video services to all types.

This point makes its impact in several ways. First, even after the rollout of digital services, large numbers of purely analog subscribers will remain, accustomed to the 80 to 96 channels of analog programming available to them, so it will be difficult to remove those channels for use as digital programming. Second, pending must carry rules for new DTV and HDTV programming could require MSO's to provide these channels also. Finally, off air digital programming will be in a multilevel VSB format while HFC channels from the headend are modulated via QAM formats. Some simulcasting of digital broadcast channels may be required to satisfy the needs of these two types of subscribers. The net of these changes is that a large amount of spectral channel allocation will be taken up by broadcast services leaving less room for advanced interactive services.

Data and Voice: The Need for 2 Way Bandwidth

As data and voice services are deployed over the next few years and the penetrations of these services grow, the need to allocate capacity and hence bandwidth to these services will also increase. For network simplicity reasons most MSO's are allocating one or at most two 6 MHz channels each for cable modems and telephony. Therefore, as capacity needs increase, the area over which each of these services are broadcast will have to decrease.

For cable modems this need for additional capacity is further exacerbated by the developing use and maturity of the Internet. Current cable modem deployments are allowing for a minimum available data rate of about 128 kbs allowed per user. While this rate is a large improvement over current telephony based modems, the availability of downloadable video clips will put pressure on MSO's to increase dramatically this minimum available data rate.

For cable telephony, several providers are discussing plans for offering as many as 4 telephone lines per subscriber and offering at least one of them as an "always on" data channel. For current systems allocating one telephony channel per typical 500 home node, the "logical node size", i.e., the area over which this service will be broadcast, will have to be reduced even further.

These constraints are illustrated in figure 1. As can be seen, the inclusion of large numbers of analog channels for non set top subscribers, HDTV channels in possible must carry scenarios, advanced services such as NVOD, and VOD, data, voice and other services can push the required bandwidth in excess of 1 GHz! Since most systems are being upgraded to 750 MHz and possible 870 MHz,

Service	Fwd B/W Range (MHz)	Reverse B/W Range(MHz)
Broadcast Analog	500 - 700	2
Broadcast Digital	18 - 24	2
NVOD	36 - 96	2
VOD	24 - 36	2
HDTV	12 - 66+	—
WEBTV	6	2
Television Based Data Services	6	2
PC Based Data servicess	6	4
Worldgate	—	2
Targeted Advertising	12 - 36	TBD
Cable Telephony	12	12
IP Telephony	3 - 6	1 - 3
IP Videoconferencing	3 - 6	6 - 12
Multimedia	TBD	TBD
	638 - 1000MHz	35 - 44MHz

Figure 1. Service Spectrum Allocation

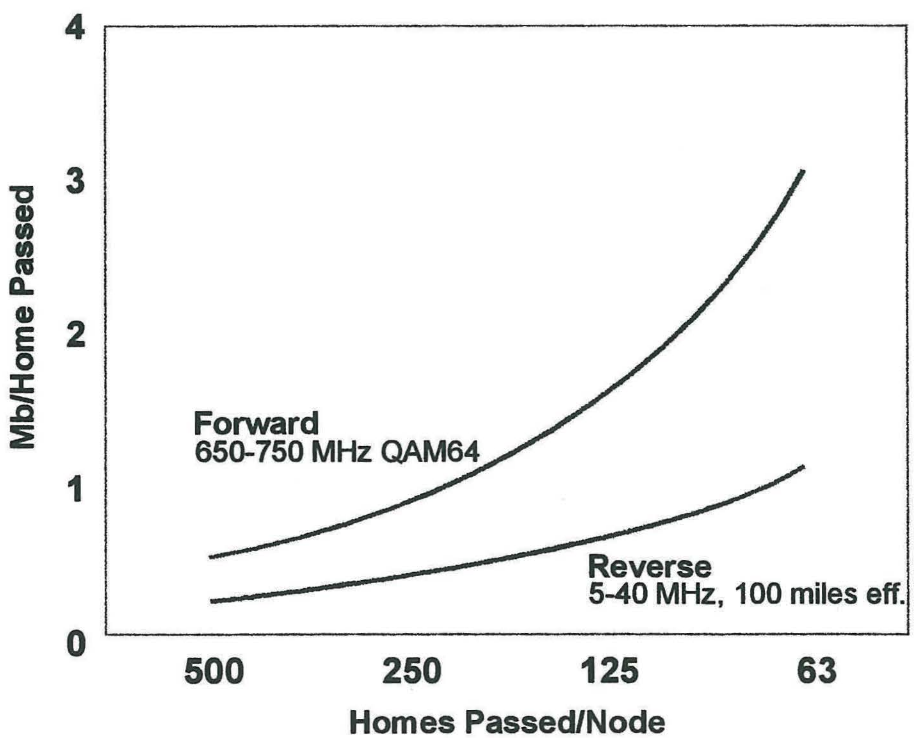


Figure 2. Available Bandwidth/Home Passed

clearly something will have to be done to accommodate these demands. One logical solution is to allocate less bandwidth to the interactive channels and reuse this spectrum over and over again by reducing this logical node size for the interactive channels. From the standpoint of the reverse channels, similar problems are seen, causing the need for smaller logical node sizes to be even more pronounced. In addition, for ingress and channel redundancy issues, not all of the 5 MHz to 40 MHz bandwidth is available. Figure 2 shows the available reusable bandwidth as the node size is reduced from 500 homes to 50 homes passed in a typical 80 – 100 home per mile plant. It is envisioned that nearly 2 Mbs per home passed will be required within 10 years. If we consider optimistic projections for service rollouts and penetrations, this number could be considerably higher.

ADVANCED SERVICES: THE NEED FOR RELIABILITY

In addition to increased capacity in both the forward and reverse path, increased network reliability is also a primary consideration. Two key means for increasing the reliability of a network is to reduce the amount of coaxial cable in the plant and to reduce the number of active devices in the field. Figure 3 illustrates the number of active devices such as RF amplifiers in a plant as the physical node size is reduced. A minimum in the number of actives (and hence the maximization of the reliability of the plant) is achieved at the passive cable point: the point at which no additional actives are needed. As the node size is reduced further, additional fiber optic receivers are needed and the number of actives increases. Also in figure 3 the amount of power needed per mile of plant is shown.

DEEP FIBER ARCHITECTURES

A typical Deep Fiber Architecture is shown in figure 4. While from the diagram it is very similar to standard HFC networks in several respects, it contains important differences as well. First, typical HFC networks are based on logical node sizes in the 1000 to 2000 home range. As logical nodes sizes approach the passive HFC point, the amount of independent (non broadcast) information delivered per mile of plant goes up dramatically. For this to occur the processing equipment at the hubs, or the DWDM networks required to remote the hub equipment to the headend must become more extensive. Second, the increase in the number of fibers employed in these fiber deep architectures must be handled by in field splitting or in creating smaller hub sizes. Finally, driving fiber deeper in the network results in the reduction in the numbers of RF amplifiers in cascade so the performance required by each of the elements in the network can be re-optimized to this new configuration in both the forward and reverse.

DEEP FIBER TECHNOLOGY

Since Deep Fiber architectures employ substantially more opto-electronic components than standard HFC architectures, the steep improvements in both the cost and performance of optoelectronic technology and optoelectronic devices have had a major impact towards pushing fiber deeper. HFC transmitters and

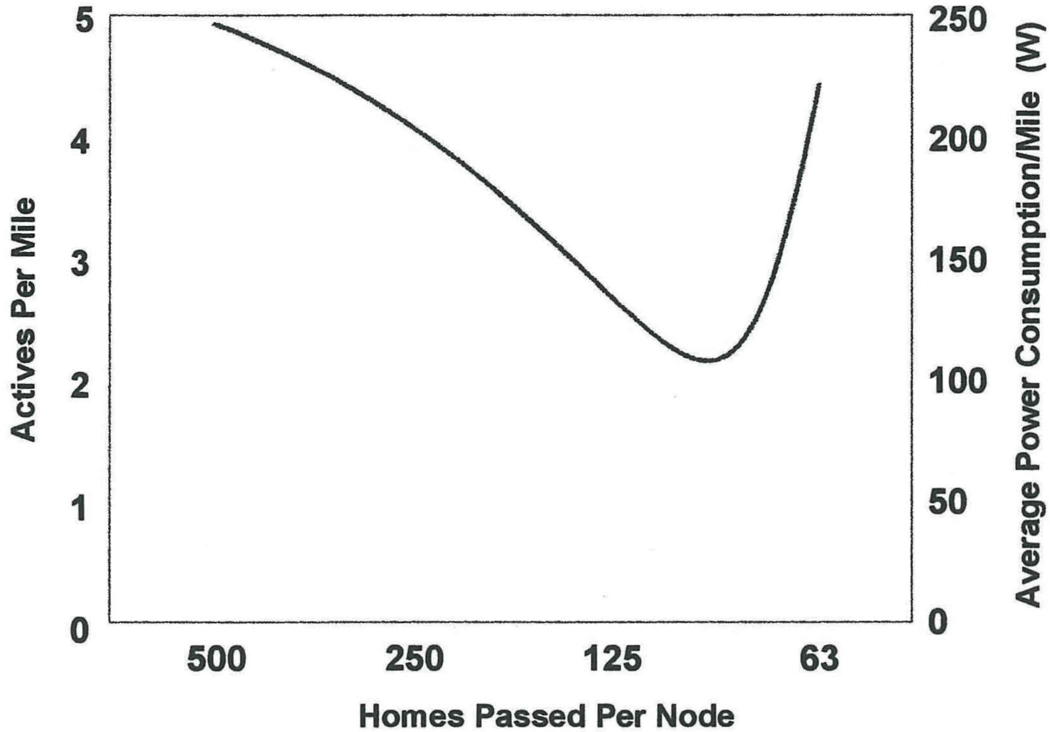


Figure 3. Actives and Power Consumption Per Mile as Fiber Goes Deeper

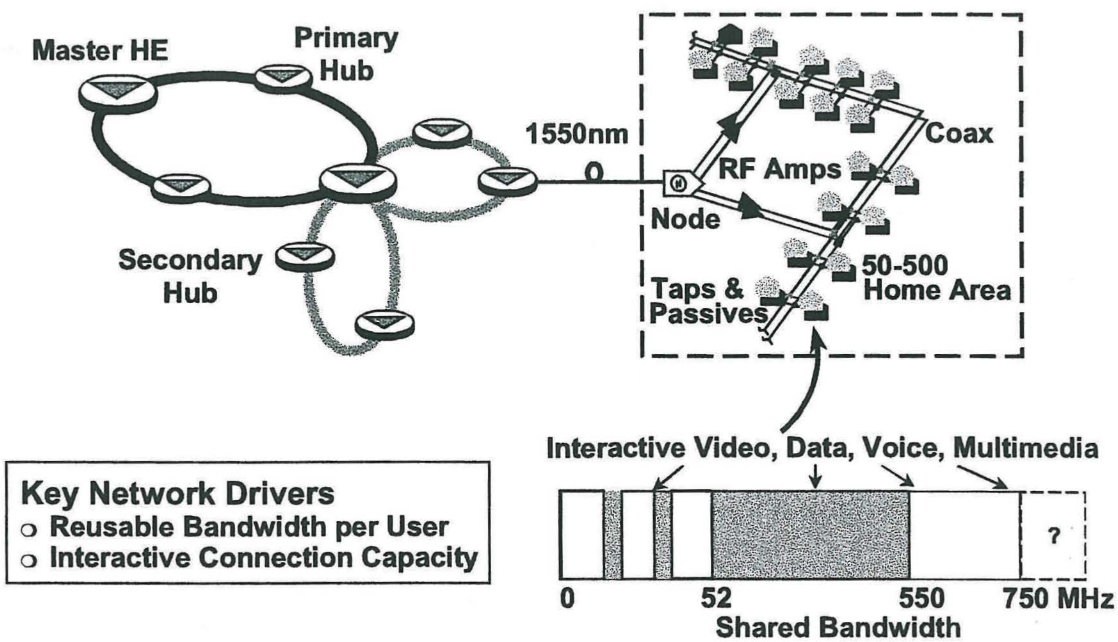


Figure 4. HFC Deep Fiber Architecture

receivers, as well as fiber optic passives such as splitters and multiplexers have seen major improvements in terms of performance, reliability, and cost.

DWDM Technology and High Power Optical Amplifiers

Another area where opto technology development is having a large impact in the deployment of deep fiber systems is in the introduction of 1550 nm transmission technology with high power optical amplifiers, DWDM transmission, and WDM overlay technology. Figure 5 shows the progression of the cost of transmitting light at 1310 nm and at 1550 nm. As will all laser based systems, the cost of generating light per mW decreases as more of it is created. Secondly, the reductions in cost per mW year over year are the result of technology learning curve improvements. It can be seen that 1550 nm transmission technology is on not only a much steeper learning curve, but also moving to higher powers and hence lower $/mW transmission costs.

Since in nearly all HFC networks the majority of the traffic is broadcast, systems can be designed which take advantage of the low costs afforded by 1550 nm high power optical technology. For narrowcast traffic, DWDM transmission and overlay technology allows the narrowcast portion of the HFC spectrum to be transmitted separately and hence each portion of the transmission can be optimized independently. At the headend, the QAM modulated interactive narrowcast channels are mapped to an optical wavelength and transmitted to the hub via DWDM. At the hub, the narrowcast DWDM channels are optically routed to their respective node or groups of nodes, where the narrowcast and broadcast portions of the spectrum are combined optically on a single photodetector. The different RF spectrum of the broadcast and narrowcast traffic allows for this to be accomplished without interference.

Performance Improvements

In Deep Fiber HFC networks, there are up to 10 times as many nodes as in a traditional 500 home node network. At first glance, one would assume that the optical transmission system would have to supply 10 times the optical power for the broadcast traffic to maintain adequate CNR levels. Figure 6 shows that this is not the case. As fiber goes deeper into the network, and the RF amplifier cascade is eliminated. The specifications usually reserved for the fiber optic node become the same as the end of line spec. In this configuration, the optical modulation index of the 1550 nm transmitter can be increased such that the fiber optic link performance matches the end of line requirement. With this, the optical power required at the node to maintain adequate CNR levels may be reduced substantially. In fact, for an order of magnitude increase in the number of nodes in a Fiber Deep Network, only 3 dB extra fiber optic broadcast optical power is required.

DEEP FIBER ECONOMICS

Despite the advantages gained in terms of network capacity, performance and reliability when fiber is pushed closer to the home, the high cost of such systems has been the major stumbling block to the adoption of Deep Fiber Networks. When most operators began their plant upgrade plans a couple of

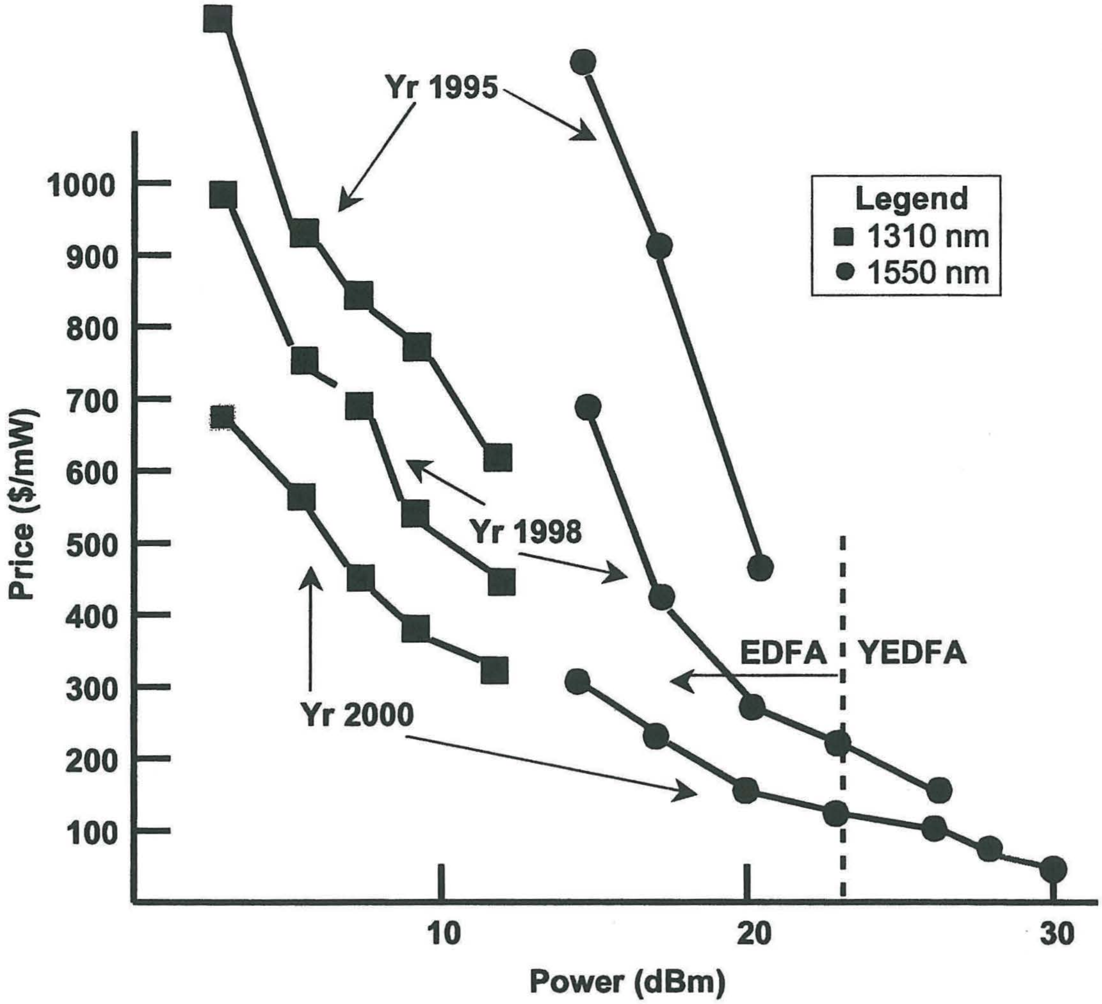


Figure 5 Transmission Power/Cost Trends

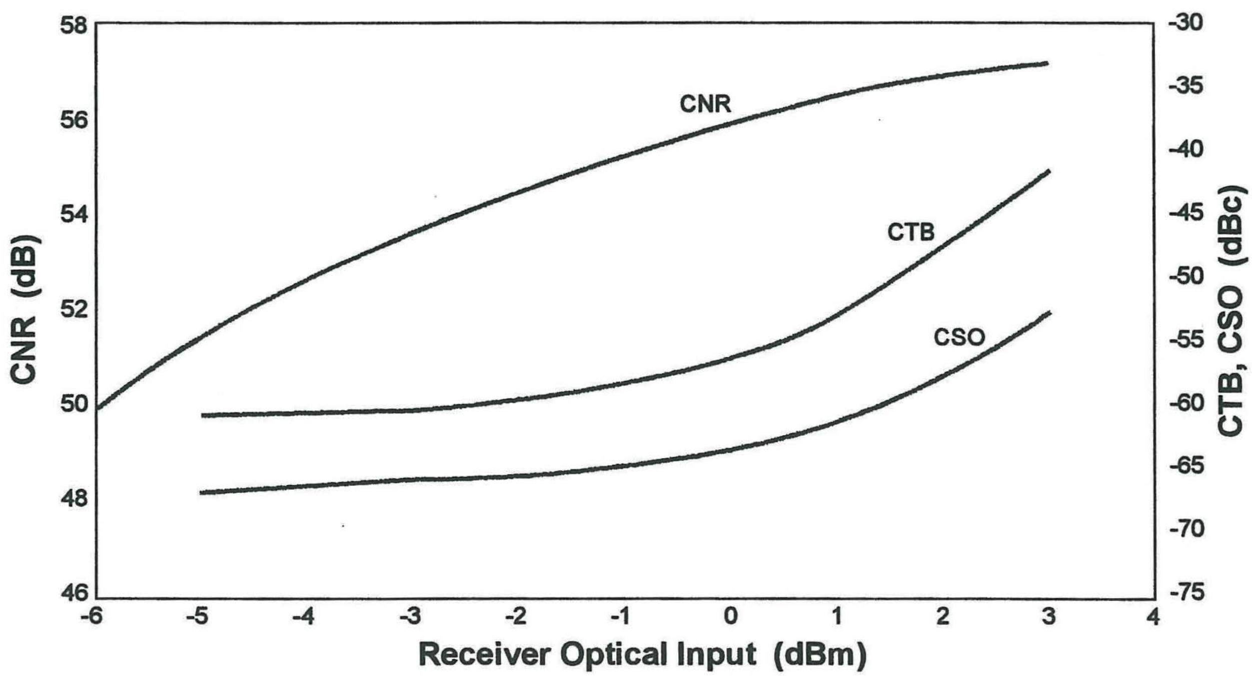


Figure 6. Fiber Deep Node Performance

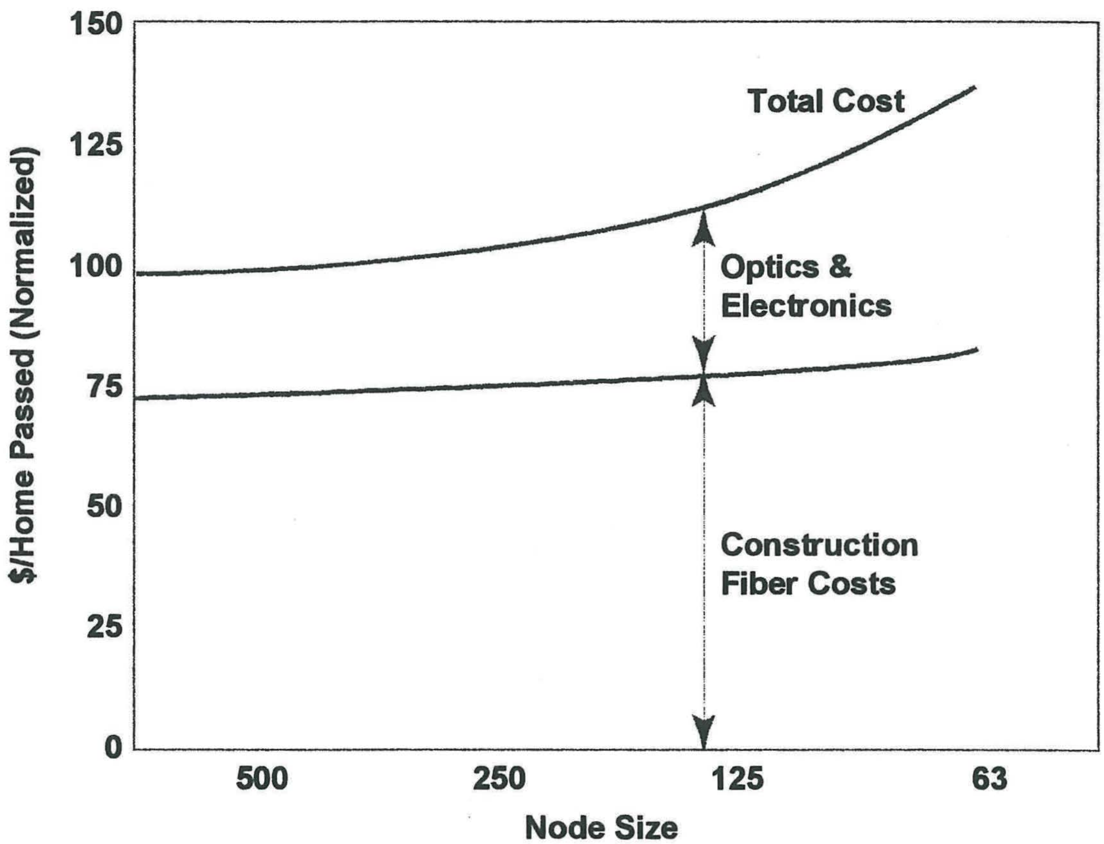


Figure 7. Plant Costs as Fiber Goes Deeper (100 = $/Home Passed for 500 Homes)

years ago, 500 home nodes were the limit for how deep fiber could go in an HFC network. Improvements in optoelectronic components, new transmission techniques, and reduced cost of optical fiber have caused this fiber limit to be pushed closer to the home. Figure 7 illustrates these improvements. From this graph, we see that the premium for driving fiber to the 125 home near passive HFC architecture is less than 10% the cost of rebuilding a 500 home HFC network. For passive networks, the premium is about 40%. Also from the graph, it appears that about three quarters of the cost is in fiber and construction, while the contribution from electronics cost grows as fiber goes below the near passive point. As the optics and electronics costs continue to fall, it is expected the costs of passive HFC networks will drop to levels close to current HFC plant costs.

CONCLUSIONS

The arrival of advanced interactive services will place great demands on improving network capacity, reliability and cost. Continuing improvements in optoelectronic components, network architectures, and transmission techniques such as DWDM, are allowing fiber to be driven closer to the home, achieving the network improvements such as increased reusable bandwidth per subscriber, higher reverse path performance, reduced network power consumption, and higher network reliability all at a cost inline with current HFC plant costs.

DEPLOYMENT OF GROOMING WITHIN THE COMPRESSED DIGITAL DOMAIN

Adam S. Tom, Ph.D.
Imedia Corporation

Abstract

"Be careful what you wish for" could be the watchword for the cable industry with digital television. After years of advocacy for the technical superiority of digital compression and modulation, digital broadcasting, from standard definition television (SDTV) to high definition television (HDTV), should reign supreme by the middle of the next decade.

The greatest challenge for cable operators here will be to respond to the market-driven, multiple-sourced expansion of demands to deliver customized programming with ad insertion, video on demand (VOD) and even datacasting, Internet access, telephony and interactivity capabilities for the individual cable operator.

MPEG-2 multiplexes from remote and local sources must be processed through a demultiplexer, transcoded and then remultiplexed to form the outgoing multiplex. To transmit a coherent outgoing program transport stream with good picture quality, this process must include parsing, synchronized scheduling, stream transcoding and splicing, and analysis and service information editing.

To gain the increased revenue opportunities of customized programming, this process must also maximize bandwidth. Imedia's statistical re-multiplexing optimizes compression ratios without feedback, separates the encoder from the multiplexer and seamlessly switches and splices MPEG-2 programming in real time. This supports the capability to selectively groom programs to create a new statistical multiplex customized for a cable system and transmit it over the space of a single analog channel.

The reasons for using digital representations of video derive from the flexibility and power of digital compression and processing with integrated circuit technology via the reduction of video to the common denominator of the bit.

Digitally compressed video is increasingly replacing analog for capturing, representing, storing, and transmitting visual sources for selective display. Digital compression technology also permits multiple TV programs per each transmission channel.

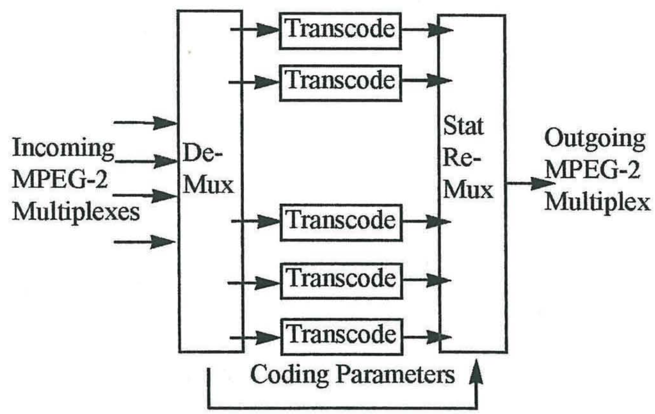


Figure 1: *Fundamental statistical re-multiplexing architecture*

Figure 1 represents the fundamental architecture to statistically re-multiplex digital television for cable. A multi-program transport stream can originate from a variety of incoming sources such as local feeds like video servers, data servers, tape and live-encoded video, and remote feeds like satellite, off-air broadcasts, and fiber networks. Transport streams are fed into a Demultiplexer that splits apart all the constituent video streams and audio streams and other services into the individual services.

Each service that is selected and passed to the output is re-multiplexed and transcoded, if necessary, to fit in the output fixed-rate band width. Coding parameters generated during the original encoding process are extracted from the incoming streams and passed over to the Statistical Re-Multiplexer when controlling the transcoding process. The input sources are then combined into a newly-formed statistical multiplex that can be transmitted over a single analog channel.

algorithm-driven package which can efficiently choreograph all of their actions to most directly address customized grooming for quality output.

The first step is to input the multi-program transport stream, represented in Figure 2 by the thick arrow. Its source could come

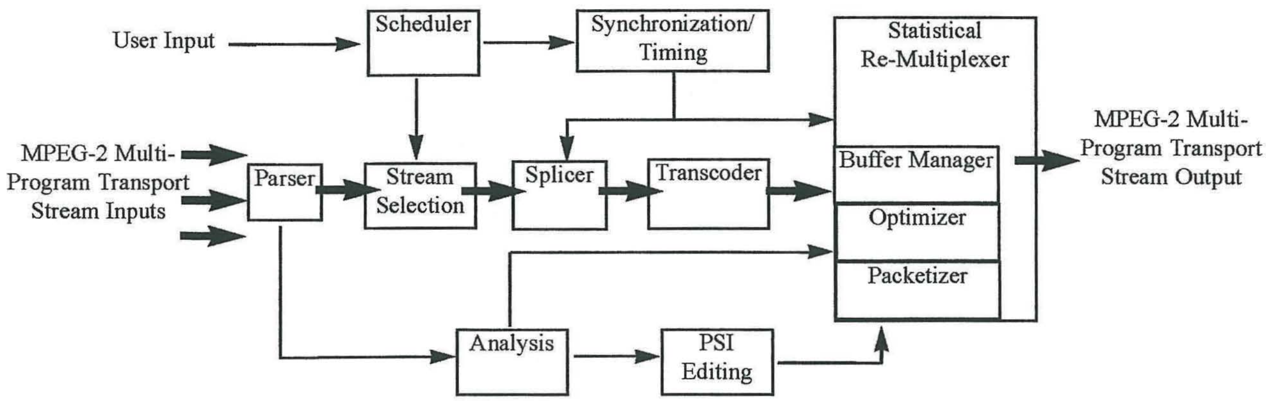


Figure 2: *The building blocks of statistical re-multiplexing for cable*

Cable operators, large or small, must aim for the highest picture quality, most flexible, bandwidth efficient, and cost-effective digital video implementations to reach their various goals, like:

- Add desirable new programs (channels) of interest to viewers,
- Use bandwidth efficiently to permit introduction of other services,
- Perform digital insertion of local advertisements and programs,
- Combine enhancement and transaction-oriented data with video streams, and
- Filter satellite-contributed program statistical-multiplexes efficiently to select only desired programs (grooming).

The key to each goal is stand-alone statistical re-multiplexing. The Imedia CherryPicker stand-alone statistical re-multiplexer technology reflects a comprehensive view of these combinant processes and integrates them into an

from one encoder, a video/ad or VOD server, or a satellite transponder such as one of the pods on HITS carrying 12 programs. A program - or service - is made up of streams which in turn can contain video, audio, and/or data.

The transport streams go directly into a Parser. A parser is a demux that looks at the program identifiers and service information in the MPEG-2 systems layer as it analyzes the composition of the streams. The next step is user input from the Scheduler through a control GUI, where the stream selection and configuration processes are performed.

Here is where decisions are made for what's going to be on the output and the point at where bit rate can also be controlled. If, say, you want one program from input #1 and three programs from input #2, this is where that selection will be confirmed.

Here also is where the user interfaces to decide which stream at this time, and what stream at that time, and also where scheduled ad insertions are triggered. If its VOD, this is where information will be entered as to when and how often a movie should be shown.

After stream selection and configuration, the Splicer is where one stream will be spliced into another stream. Splicing is where two different streams are combined together into one stream, like advertising or a traffic alert into a network feed. This Splicer will do seamless, frame-accurate MPEG-2 splicing between video streams and between audio streams, enabling the insertion of digital ads without black frames or freeze frames in between.

The next block is the Transcoder, where the bitrate adjustments occur. The Transcoder is required because bandwidth output is always fixed, so you are required to adjust the bitrates of each individual stream to be able to interlock all of the chosen streams together into the given output bandwidth. This, in combination with the intelligence of the Statistical Re-Multiplexer, optimizes bandwidth and preserves picture quality.

The Statistical Re-Multiplexer is really the heart of the CherryPicker system. Its job is to perform rate control and to monitor how the Transcoder and the Splicer operate. These three elements collaborate to ensure that relative timing is maintained, that the target decoder buffers are managed correctly, and that the Packetizer addresses all the streams.

Imedia has pioneered advanced statistical re-multiplexing techniques that manage the challenges of statistical re-multiplexing - (1) not exceeding the bit-rate capacity of the transmission channel, (2) accepting remotely encoded inputs, (3) optimizing channel bandwidth utilization, (4) maintaining high picture quality, and (5) packetizing all data forms into the output multiplex.

In a statistical multiplex, each program dominates the average bit rate capacity at one point or another, when it takes much more than its "share" of the average bit-rate available. To achieve a constant picture quality for all streams, the bitrate must vary for each stream. This creates a multiplex in which it is difficult to separate out a program and combine it with other variable-bit-rate programs from other statistical multiplexes; that's why it must be statistically re-multiplexed.

The Synchronization and Timing block serves the critical tasks of monitoring the timing and switching between the streams and assuring that all the streams are synchronized together. To guarantee proper decoding and proper presentation in time at the set-top box, timing elements within the MPEG-2 syntax such as Program Clock References (PCR), Presentation Time Stamps (PTS), and Decoding Time Stamps (DTS), need to be monitored and updated as all of the transcoding and multiplexing operations occur.

In order to make decisions about which streams are to be transcoded at what times, the statistical re-multiplexing system here needs to know information about each individual stream and it needs to know this on an instantaneous basis. The Analysis block is always monitoring the stream and deciding what to do. The information from the Analysis block goes to the Statistical Re-Multiplexer and also to the PSI Editing (Program Specific Information) block, which is a function that describes and reconciles the contents of a transport stream.

The PSI Editor also makes sure that all the information in the output stream is described correctly and that the information is constantly updated. When you go from one stream to another, program-specific information will change, and that information may be arrayed across several programs in a transport stream as you're switching between programs.

All of that needs to be reconciled and compiled together so that all the different pieces of information are brought together to describe the resultant single transport stream. That information is then fed back into the Packetizer and the Re-Multiplexer and included as part of the outgoing transport stream. These building blocks of advanced digital deployment are already enabling a wide variety of cable and satellite providers to build value and revenue opportunities, each choosing unique combinations of multiple services into one unified stream to reach and serve their customers.

Figure 3 represents a local control headend of a small MSO located in a rural community serving a subscriber base of over 100,000. Taking full advantage of the versatile flexibility of their Imedia installation, this customer wants to cherry pick programs from HITS transponders to create a new program lineup for their cable system. Here they are choosing programs from HITS1 and HITS7 and putting this new multiplex onto a single 6Mhz channel output and doing the same operation with HITS3 and HITS9. Both inputs to the CherryPicker are DHEI.

They are also experimenting with revenue-enhancing ad insertion into the HITS1 and HITS7 channels using an ad server connected to the CherryPicker via the DVB-ASI input. The CherryPicker does the insertion. The output of that CherryPicker then goes to an IRT for encryption, and through a C6U for upconversion. Encryption in this system is locally controlled by the GI DAC6000 Network Control System. Both inputs and outputs to the CherryPicker in this system are 27 Mbps.

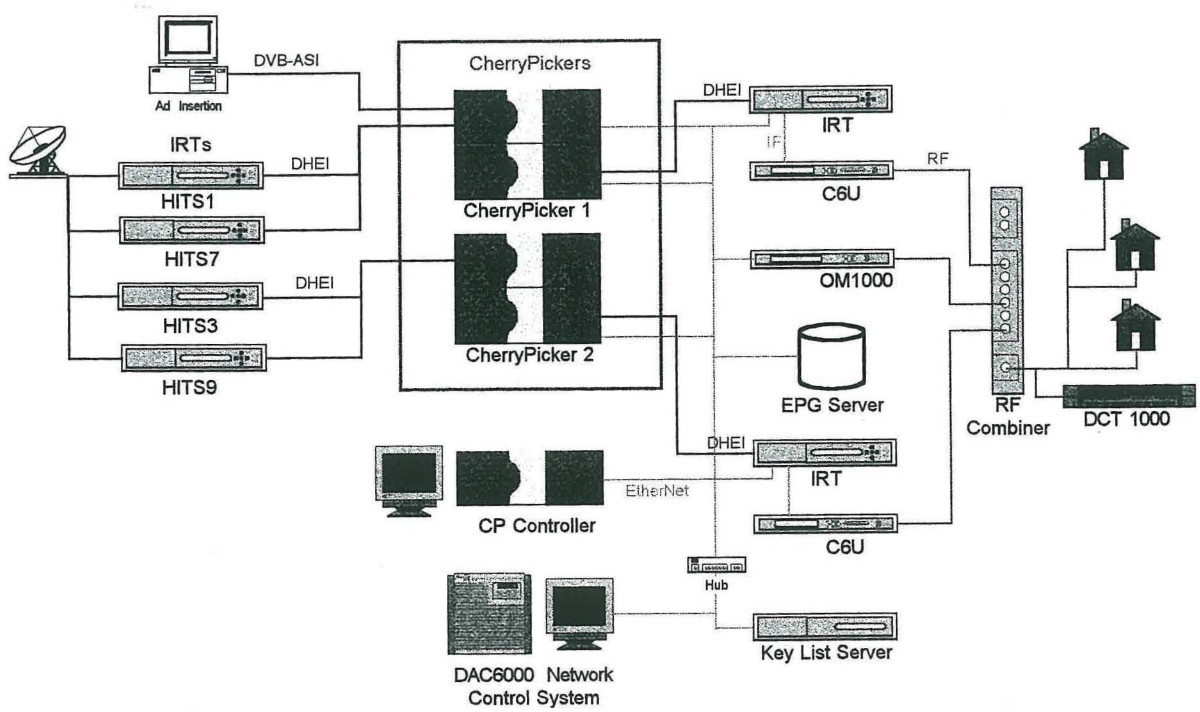


Figure 3: *Customer A: GI Local Control Headend*

After upconversion, the multiplex goes to an RF Combining Network which then transmits out onto the cable system to the subscribers' homes which are using DCT 1000 MPEG-2 Digital Consumer Terminals.

Both of these two CherryPickers are controlled by the CP Controller and all of the elements are hooked together by Ethernet including the Electronic Programming Guide Server and the Out-of-Band Modulator (OM1000.) With these aggregate gains in broadband efficiency, this provider is able to enjoy the benefits of more ad buying power.

Figure 4 represents a large MSO operating in the southeast with a subscriber base of 350,000, and is employing the Imedia technology for grooming and local program insertion. Here they are doing a CherryPicking operation, but they are also able to combine that process into an output multiplex containing video streams from a video server of local content at their local headend. This video server contains video programs and interstitials. They are combining it with pay per view (PPV) programming from TVN.

The remaining structure is similar, a locally controlled headend which is going through the same process as the previous one. What this shows is that there's broad flexibility in being able to bring in feeds from video servers and that also, in the future, VOD and real-time encoded feeds could just as easily be included in the CherryPicker. The CherryPicker can also be used in nationally controlled headends where the DAC6000 would not be needed.

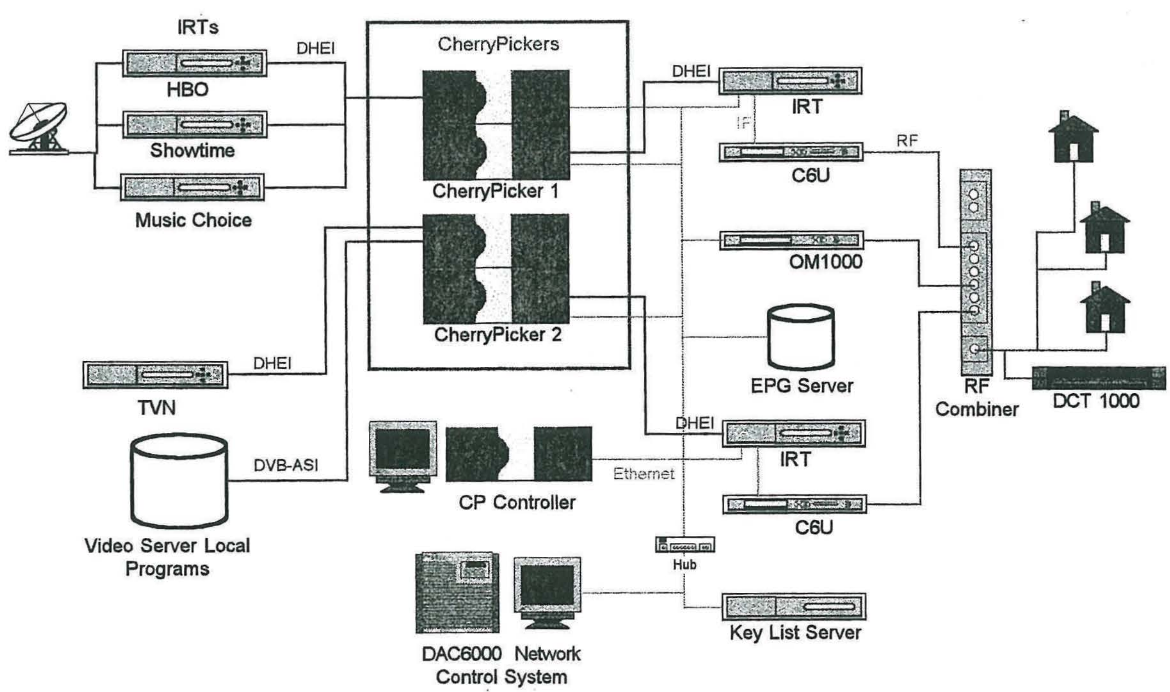


Figure 4: *Customer B: GI Headend with Video Server*

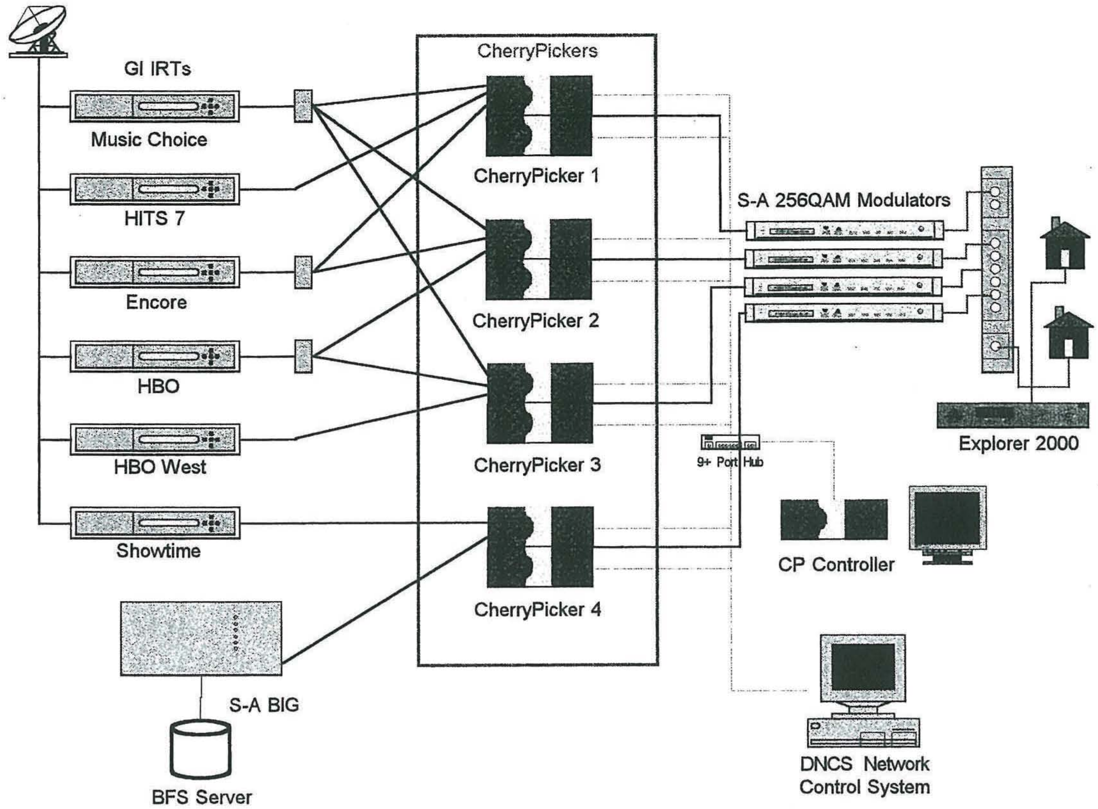


Figure 5: *Customer C: GI Inputs with S-A 256-QAM Outputs*

Located in upstate New York, this small MSO with 85,000 subscribers has built a S-A (Scientific Atlanta) headend environment that supports 256-QAM. They have six satellite feeds coming into their headend. When they need to cherry pick certain programs from those six feeds, they can output that onto four 6Mhz channels.

There's a measurable bandwidth difference between the incoming 27 Mbps and the outgoing 38.8 Mbps. In Figure 5 they are dividing the cable bandwidth to the maximum because of the bandwidth difference, so they need to take these 27 Mbps streams and add other programs on top of it to get it to fill up and complete the 38.8 Mbps stream. They are gaining a high rate of efficiency here in the mismatch between the data rates and it is the CherryPicker that makes that feasible.

Here you can see the types of interconnects they have established linking the four CherryPickers. This reveals what sort of breakup in the programming they're going to do, and how it then goes out to four S-A 256-QAM Modulators. These units not only enact the encryption, but they also perform the modulation. This customer is doing a translation from a GI front end to a S-A back end. The IRTs are General Instrument but the cable modulators are Scientific Atlanta. This system is controlled by the Scientific Atlanta DNCS Network Control System and is also taking in some information from their BIG (Broadband Integrated Gateway) and BFS Server, ultimately going into a combining network and out to the subscribers' homes via an Explorer 2000.

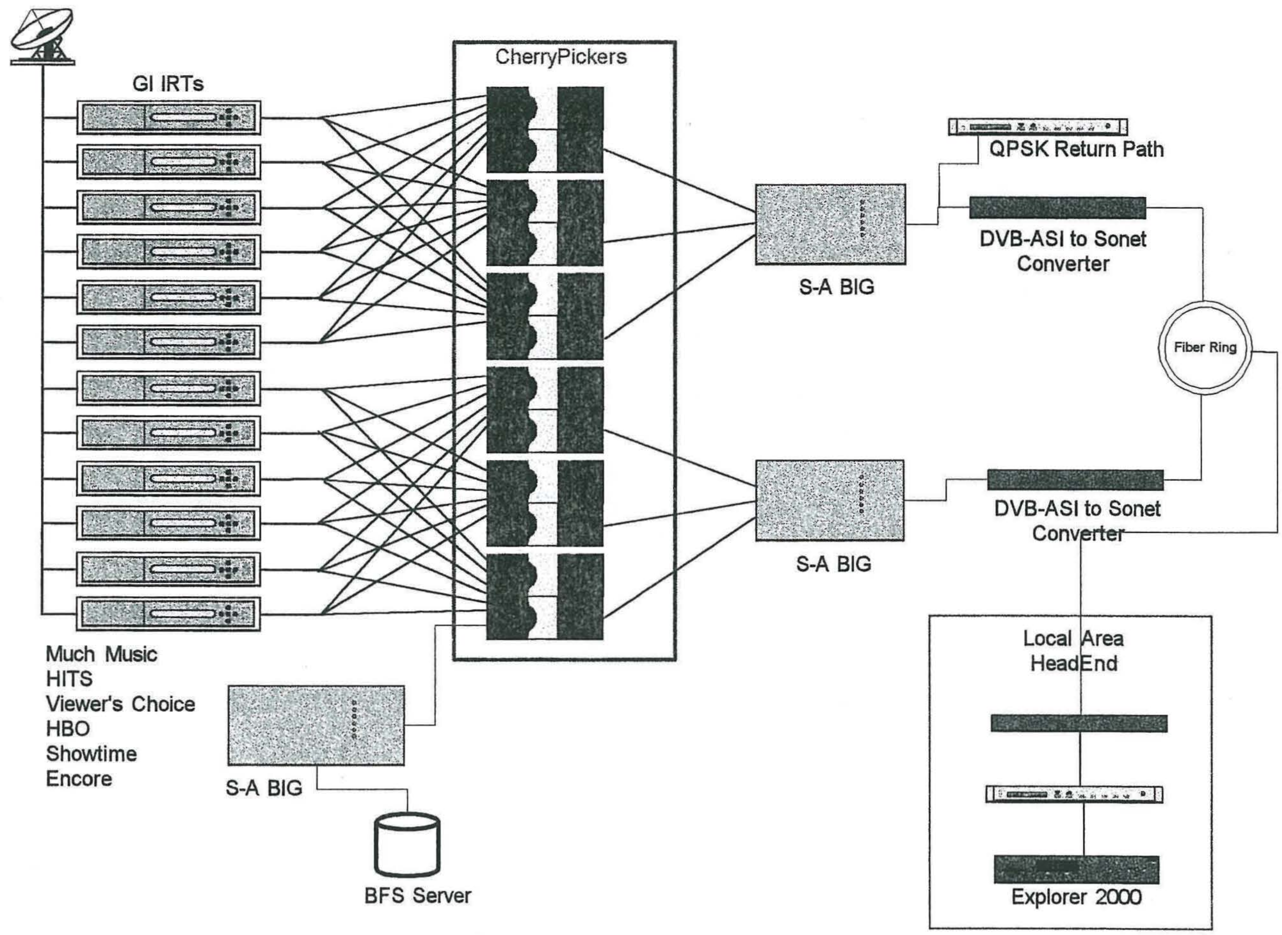


Figure 6: *Customer D: Fiber Ring Distribution*

This New England-based customer has exploited the scalable expandability of the Imedia system to direct a dozen IRTs with six CherryPickers to transmit and distribute through a Fiber Ring. In Figure 6, the CherryPickers take feeds going from GI IRTs at one data rate to an S-A BIG at a different data rate, but the remultiplexing operation for redistribution is output over a Fiber Ring rather than going out over a standard cable system.

The S-A BIG is needed because it can take in many multi-program transport streams from different sources and output it into the type of format for distribution over a Fiber Ring. A local area headend can take off the programs from the Fiber Ring and distribute it over the headend. This provides an optimal network for distribution to a lot of headends, and serving an even wider constituency of subscribers. Their unique SONET Fiber Ring configuration efficiently encircles Vermont through six nodal cities, effectively grooming the entire state.

Every operator that deploys digital television will face a different set of challenges, each delivering new limitations and opportunities. What these customers share in common is their proven abilities for problem solving through building installations that support flexible and effective methods for switching, splicing and grooming MPEG-2 digital video streams and the ability to optimize bandwidth efficiency while preserving picture quality using statistical re-multiplexing.

DIGITAL TELEVISION GLOSSARY

CONSTANT-BIT-RATE (CBR)
a value where one TV program is allocated a certain number of bits-per-second at all times (no statistical multiplexing)

CHANNEL CODING
tries to minimize channel transmission errors by using error-correction and error-detection algorithms

DATACASTING
transmission of digital data

DCT
Digital Consumer Terminal

DHEI
Digital Headend Expansion Interface

GROOM
assemble programs from a variety of multiplexes and feeds and create a customized group of programming delivered as single or multi-program transport stream

GUI
graphical user interface

HITS
Headend In The Sky

IRT
Integrated Receiver Transcoder

Mbps
Megabits-per-second

MPEG-2
a decoding and compression standard, also known as ISO 13818 or H.222.0, developed by the Moving Picture Experts Group, using mathematical algorithms which represent TV frames composed of macroblocks and blocks, and taking advantage of predictability in the picture and in picture motion

MODULATION
the method of putting the bits on the channel

MULTIPLEXING
involves combining multiple TV programs together before modulation and transmission over the channel

MUX
a Multiplexer

PARSER
A function that identifies the individual parts of an MPEG-2 compressed transport stream including systems, video, and audio layer information

QAM
Quadrature Amplitude Modulation

QPSK
Quadrature-Phase-Shift-Keying modulation

SOURCE ENCODING
assigns bits to represent video (and audio), taking into account redundant or predictable characteristics

SPLICING
where two different streams are combined together into one stream

STATISTICAL MULTIPLEXING
permits each TV program in a multiplex to share the multiplex according to its moment-by-moment bit rate requirements

STATMUX
a Statistical Multiplexer

TRANSCODING
the ability to modify the incoming data rate of a video stream to a different (lower) outgoing data rate

VARIABLE-BIT-RATE (VBR)
Compression bit rate allocation that is much more efficient because it allocates only the number of bits required at a certain time to a TV program to achieve a target picture quality

VOD
video on demand

Digitized Return Path:

Extending the Applicability of Hybrid-Fiber Coax Architectures for Full-Service Networks

Farr Farhan
Scientific-Atlanta, Inc.

Abstract:

In this paper, the concept of digitizing the Hybrid Fiber Coax Reverse (or Upstream) path is introduced. The Pros and Cons of such a concept are compared against the existing and traditional methods. The cost, performance and evolution potential of this concept are presented, as well as preliminary lab trial data. Some of the concepts are explored further noting some implementation challenges. In closing, some other evolutionary possibilities for the CATV Hybrid-Fiber Coax architectures are discussed.

INTRODUCTION: THE PROBLEM

As cable operators continue their aggressive efforts to upgrade their networks, they are constantly looking for ways to make full use of existing equipment to control costs, while still attempting to increase network capacity. This is especially evident in the implementation of upstream transmission of bandwidth-consuming, interactive multimedia traffic.

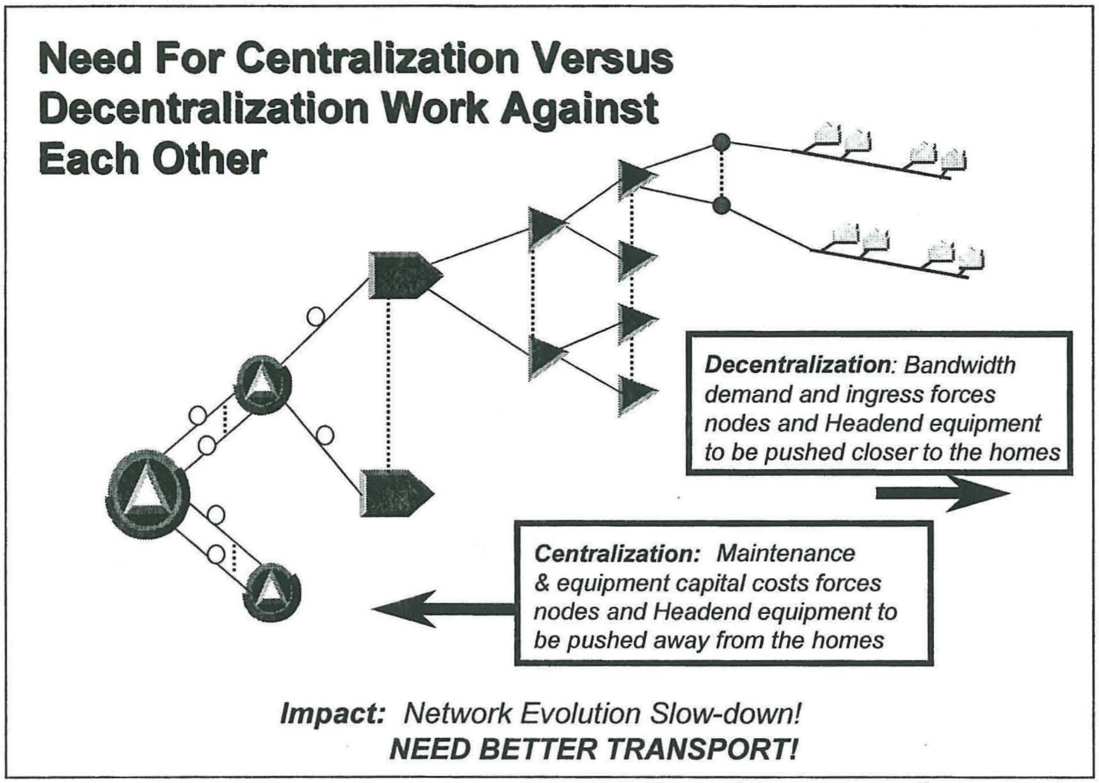
The deployment of the reverse path has brought about several major challenges for cable operators. To deliver new revenue-generating services -- such as multimedia, Internet access and telephony -- and to increase subscriber penetration, operators must dramatically increase their network capacity. They need to allow more traffic over a single fiber while retaining their existing architectures when possible in order to control costs. Operators also need to transmit signals in both directions over longer distances, and allow extra bandwidth for future upgrades -- all while improving network performance and reliability to keep their subscribers satisfied.

In this situation, two main forces are competing against each other (Figure 1). The need for additional bandwidth forces nodes and headend equipment to be pushed closer to the homes. In the meantime, the need to control maintenance costs and capital investments force the equipment away from homes, so a larger number of subscribers can share the same capital investment. The answer to this dilemma is: Better Transport. This means transmission equipment that is lower in cost, better in performance and more importantly, more efficient.

Reverse Optics Issue 1:

Within traditional analog optical network architectures, the performance of upstream signals are usually unpredictable, suffering from the effects of optical noise as well as temperature fluctuations at the node and at the hub (See Figure 2). The fiber-optic portion of return path HFC is still considered the weakest link. This transmission is often handled by Fabry-Perot lasers, which are relatively inexpensive but unable to

Figure 1: Two Forces Working Against Each Other



effectively handle large volumes of multimedia traffic. Uncooled distributed feedback (DFB) lasers, offer better dynamic range performance but at a higher cost. These lasers also exhibit a variation in slope efficiency and noise floor as a function of temperature.

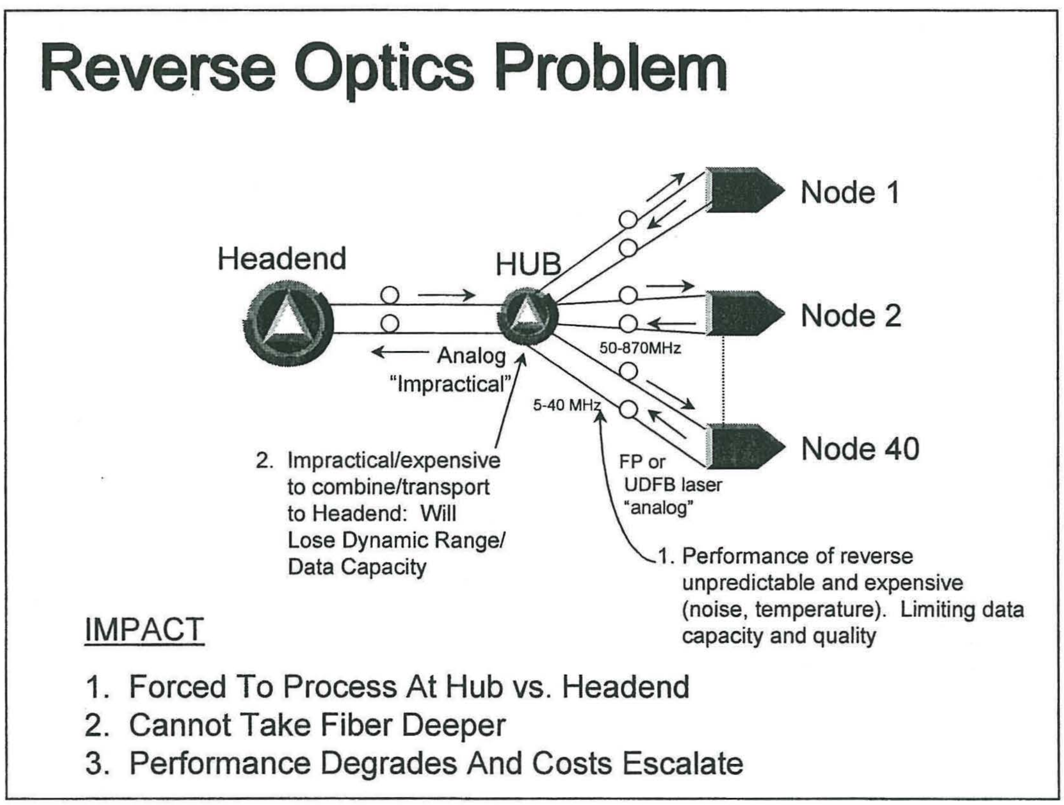


Figure 2: Reverse Optics Problem

Although many vendors have done a great job adding circuitry to reduce the dependency of Carrier to Noise Ratio (CNR) on temperature variation, there is still a measurable dependency and, of course, unpredictability.

Reverse Optics Issue 2:

With so many signals to handle, and because the analog signal cannot be sent over long distances without signal degradation (CNR degrades as a function of distance), many operators have been forced to configure their networks to process the upstream traffic at the hub (Figure 2). These hubs have by necessity become larger than desired buildings that require additional right-of-way, staff, power, etc., making them expensive facilities to operate.

To transmit the signal from the hub to the headend, some operators have implemented some form of multiplexing or combining scheme such as, RF combining, block conversion or reverse dense wave division multiplexing (DWDM). In all these schemes noise is added to the signal in addition to other specific artifacts. Reverse DWDM has provided the best choice for preserving signal quality, however still at a significant price increase.

THE PROPOSED SOLUTION: APPLYING DIGITAL TECHNOLOGY

The proposed solution to the dilemma facing operators is the application of digital technology in the reverse path. Specifically, by digitzing the link between the node and the hub, and the link between the hub and the headend (Figure 3). While the use of digital has been technically possible for some time, it only recently has developed to the level of feasibility needed by cable operators in performance and to some extent in cost.

The solution involves digitizing the upstream spectrum at the node, utilizing an Analog to Digital Converter (A/D) followed by a serializer (Figure 4) and lastly a digital laser to transmit the digital stream. At the Hub (or even at the node), a number of these digital streams can be combined using Time Division Multiplexing (TDM) techniques (Figure 3 and 6) to create a higher speed digital signal. This higher speed digital signal is then transported to the final destination for O/E conversion, followed by demultiplexing and de-serialization and lastly by a Digital to Analog Converter (D/A) for reconstruction of the original upstream spectrum. Figure 6 shows that the signal can remain in the digital format, from the node all the way to the headend, while providing the added flexibility of TDM for trunk reduction.

By use of digital techniques including TDM, the cable operator can now greatly improve reverse bandwidth and consequently improve subscriber penetration. In effect this proposal solves both of the issues raised previously with today's reverse optics as follows:

1) By keeping the signal in the digital domain, there is no longer a penalty in signal quality due to distance or combining. As a matter of fact, as long as the converted signal is maintained in its digital format, the degradation remains zero. Because of the lower power requirements and in general the resilience of digital transmission, the return path payload can be transmitted from the node all

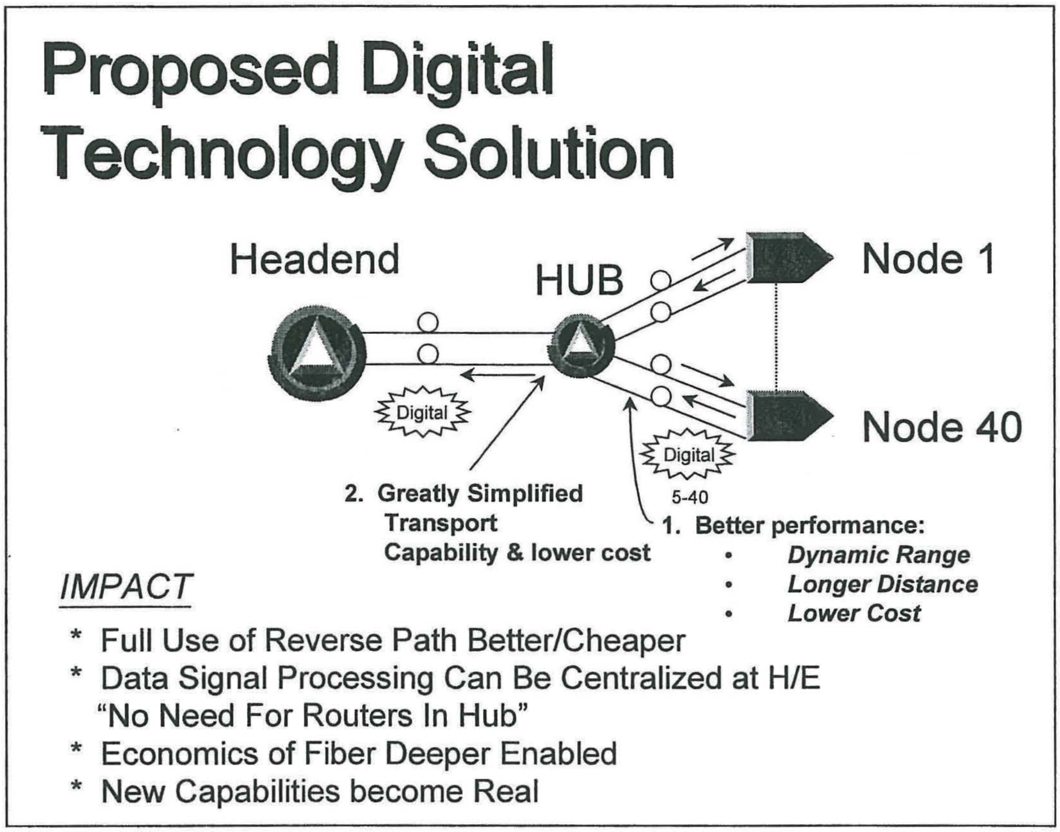


Figure 3: The Proposed Solution

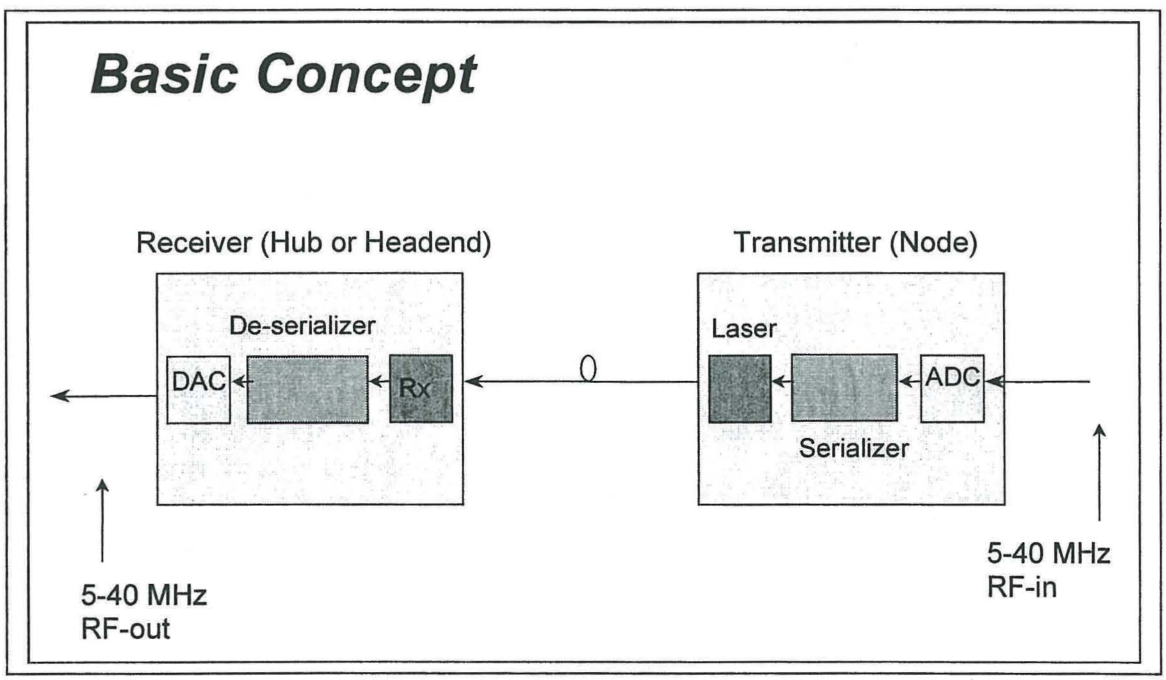


Figure 4: The Basic Concept

the way to the headend, creating the option of a passive hub (See Figure 7)

2) The use of TDM creates a new dimension in multiplexing capability. TDM augments Wave Division Multiplexing (WDM). Both TDM and WDM can coexist in perfect harmony. As a matter of fact, this is exactly what is happening in the world of telecommunication outside cable.

Some operators prefer to use their existing hub architectures, while others choose to build networks without hubs. Digital technology has the flexibility to meet both their needs allowing them to select where they want to process the information. Operators can often retain much of their current architecture by simply plugging in digital instead of analog transmitters or receivers.

COMPARISON WITH ANALOG SOLUTIONS

1) Performance

Because digital transmission is more robust than analog transmission, the susceptibility of the upstream signal to optical noise and temperature variation of the network is greatly reduced. As mentioned previously, once the return band is converted to digital format and retained, its quality is preserved. Recent lab trials have shown that with existing conversion technology, Fabry-Perot performance can be exceeded (See Figure 5). Using digital methods, it can be proven that the performance of Uncooled DFBs can be met or even exceeded. Furthermore, we have demonstrated that the susceptibility of system gain to temperature variation in the plant was literally removed. We also found that CNR was also significantly less susceptible to temperature variations in the plant. It is interesting to note that all these improvements in signal quality are achieved without any dependency on the quality of the optical components used. This latter benefit has a very significant implication: The cost of an optical link is no longer determined by the quality of the optics used. When this benefit is combined with the independence of signal quality over distance (as long as the payload is kept in digital format), the future promise becomes even greater.

2) Capacity

The added capacity, enabled through TDM, allows a much greater number of subscribers to be served by one fiber. It also makes the network much easier to manage.

If propagation delay is not a concern, the use of digital transmission also enables many components, such as demodulators and routers, to be moved upstream to the headend, creating economies of scale at the headend as well as simplifying maintenance. It also lowers costs by eliminating the need for large and expensive hub facilities.

3) Cost

While the performance capability of digital technology has ever been increasing, its cost has been continually decreasing. The prices are expected to continue to fall in the coming future. While the performance of the key components for building a digital optical link capable of carrying digitized cable return spectrum has been increasing, overall, their price has been decreasing.

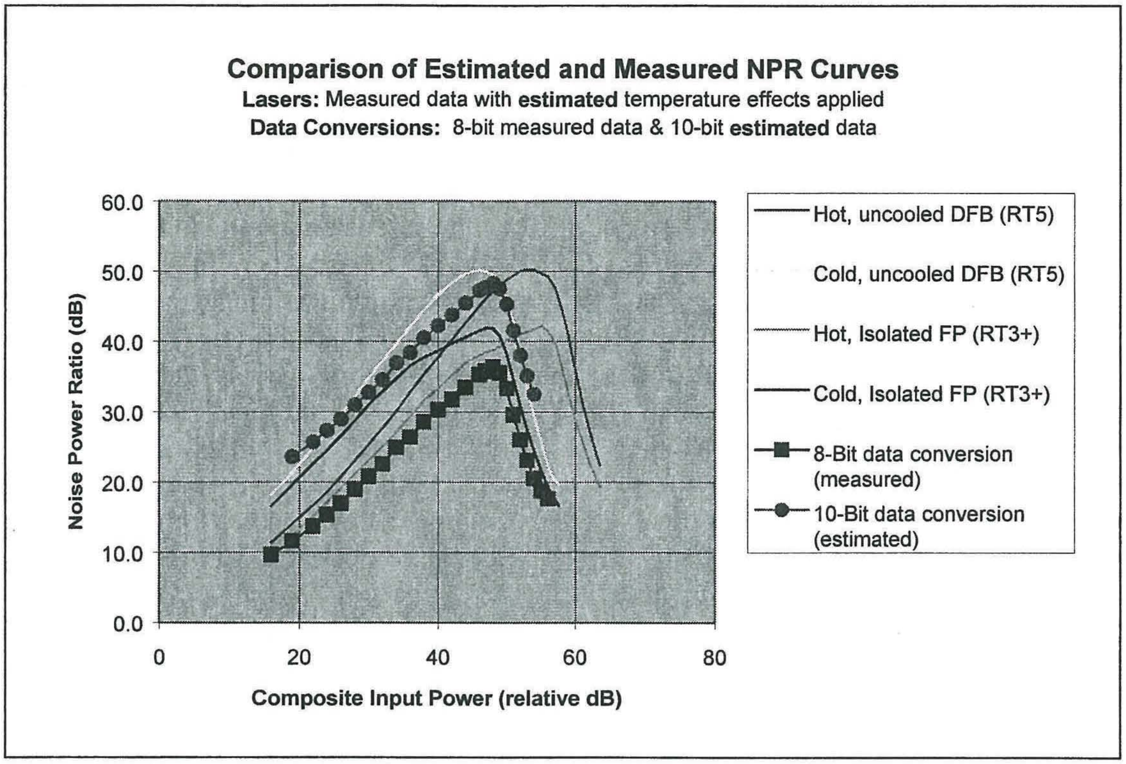


Figure 5: Preliminary Lab Results

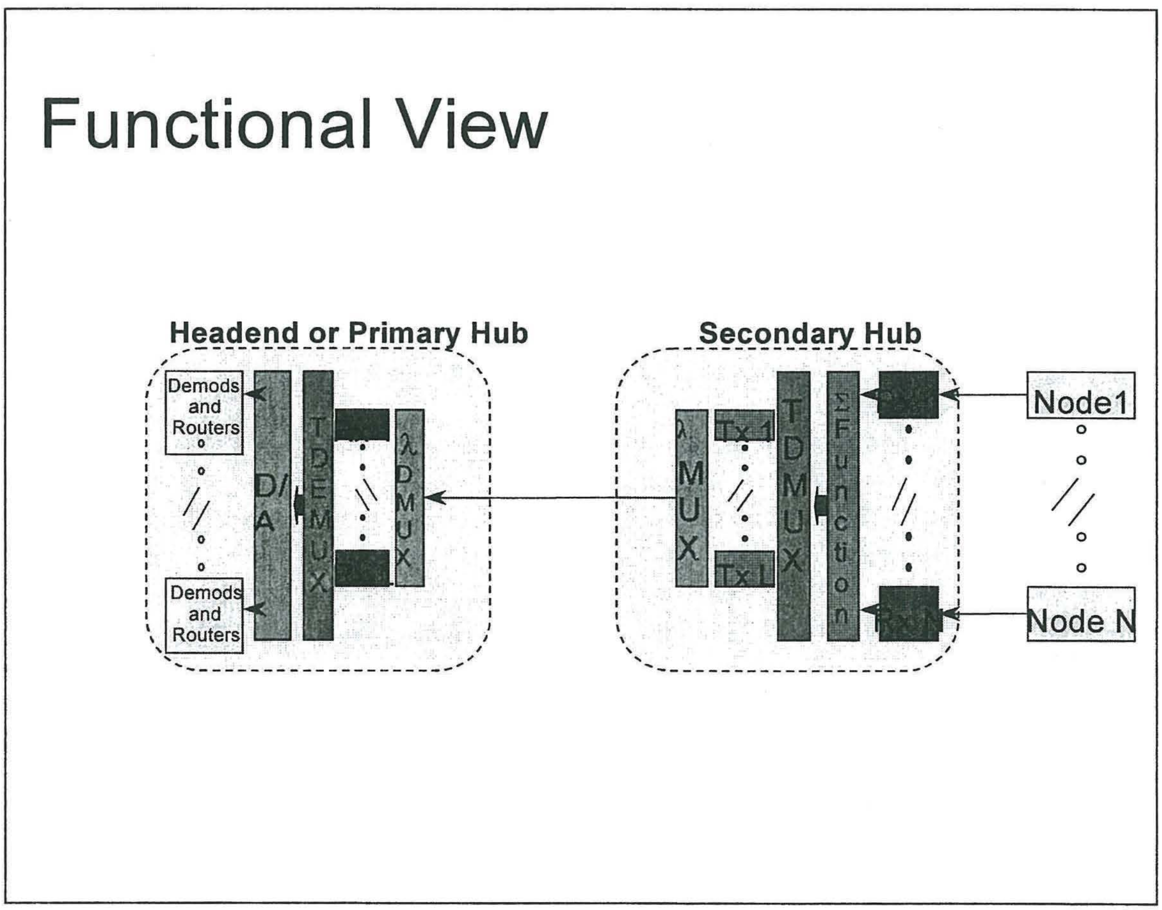


Figure 6: Functional View

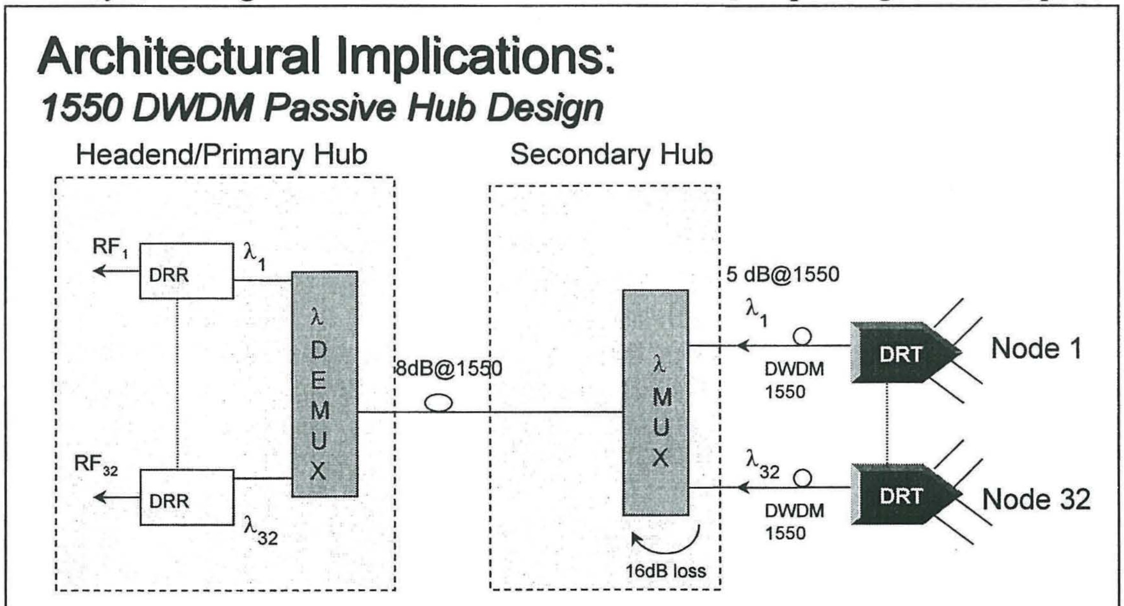
The key cost advantage behind a digital reverse link is the fact that it will rely on the technological advancements of digital components, while enjoying their declining cost curves. Everyone would agree that the global deployment of digital components will continue to outweigh that of analog components. Hence it is logical to conclude that digital implementations will always have a cost advantage over analog implementations.

Implementation Challenges

While there are always challenges with the introduction of new concepts, we expect that the challenges with the introduction of this technology to be far less than those experienced recently with other key technologies. We have learned that the quality of clock is extremely important. Any noise introduced in the time domain or in the voltage domain can lead to degradation of signal to noise ratio. The reason is that uncertainty introduced in the clock sampling instant, translates to uncertainty in the converted signal, hence to noise. Though this noise is bounded, the signal to noise ratio will be impacted. Of course there is still a lot more to be learned in this area, however so far, experimental results have closely matched the underlying theory.

Architectural Implications

The flexibility and power of digital technology make it suitable for a variety of situations. In existing networks, receivers can be installed at the hub or headend, depending on the operator's



Fiber Node to PH	
CNR (NPR)	40 dB
Dynamic Range	15 dB

Figure 7: Architectural Implication

preference for where the information is processed.

During the construction of a new network, the operator can choose to dramatically lower hub costs by designing the network for transmission directly from the node to the headend, ultimately eliminating the need for processing equipment at the hub (Figure 7).

Evolution Potential

This new technology allows much more information to be placed on a fiber, increase signal quality, increase network reliability and at the same time reduce network cost. The high bandwidth capacity provides the ability to transmit a range of services in the reverse path, including cable modem traffic, telephone service and other interactive content. It also offers the extra bandwidth necessary for future network expansions including digital payload transport. As a matter of fact, the latter application will convert HFC to the only network capable of simultaneously transporting analog and digital payloads.

By applying digital techniques to reverse signals, it is expected that the effects of undesirable plant noise and ingress will be significantly mitigated. The latter schemes have the potential to erase the negative criticism of cable's "noisy" return plant.

CONCLUSION

Virtually every industry is realizing the advantages of digital solutions, and in the coming years almost any solution that can be converted to digital, will. With the introduction of an array of new, interactive services and ever-expanding channel line-ups, the cable industry is even in more need of an advanced and long lasting transport solution for the upstream path. Digital technology can solve this issue. It is the long awaited solution for the legacy issues of the cable return path.

While preliminary results show that the performance of traditional Fabry-Perot lasers has been exceeded, we claim that the technology is capable of providing the performance of uncooled DFB at the cost of Fabry Perot. Furthermore digital reverse can serve up to four times as many subscribers on a single fiber, compared to traditional cable networks. All this combined with a simplified network architecture with fewer components, means more reliability, less operational and capital cost with better signal quality. This technology has the potential of lifting the cable industry, as fiber did to trunk and feeder plants. We are about to witness the beginning of a significant turning point for the cable industry!

Distortion Beat Characterization and the Impact on QAM BER Performance

D. Stoneback, R. Howald and T. Brophy

General Instrument

O. Sniezko

AT&T Broadband & Internet Services

Abstract

This paper discusses measurement and analytical results of distortion beat characterization. This analysis received significant level of urgency recently when field tests showed that the traditional methods of characterizing nonlinear distortions (see NCTA Practices[i]) generated in multichannel HFC networks do not allow for an easy prediction of the behavior of digital signals in the same environment. Especially the introduction of higher, more sensitive modulation schemes (256-QAM) made characterization of the impairments of the HFC channel very valuable.

Beat distortion requirements for analog video are based on subjective video perceptibility correlated to the levels of impairments measured according to a standardized test method. For QAM, BER is a more quantifiable measure. To properly characterize the BER degradation analytically, a statistical description of the intermodulation distortion is necessary. The description required is the probability density function (PDF), an appropriate mathematical description for the evaluation of steady state BER performance. Beyond the PDF, the nonstationarity of the distortion process must be considered to capture possible variations of BER as carrier frequency drift varies the statistics.

Using these techniques, BER performance can be analyzed as a function of the variables that contribute to the distortion's parameters - channel map, analog-to-digital signal power ratios, relative carrier phasing, time, and the nonlinearity performance of the channel. This analysis and supporting measurements can be used to optimally align the forward path when the spectrum consists of a composite of analog and digital signals.

INTRODUCTION

The forward path in HFC networks has been proven able to successfully support digital video signals, transporting the information on RF carriers using 64-QAM modulation. This digital modulation format is also called out for carriage of forward path cable modem signals in DOCSIS compatible modems. The power allotted to the QAM signal is implemented by careful consideration of several issues. Among them are the tolerable degradation of analog performance, the minimum levels for proper performance of the QAM signals, and the necessary S/(N+IMD) for 64-QAM link budget. The payload data in 64-QAM applications is augmented with a powerful forward error correction (FEC), mitigating much of the concern with respect to the last item. However, the 256-QAM modulation scheme, to be implemented for higher bandwidth efficiency of the forward path, is more sensitive to impairments. Moreover, as the field tests indicated, higher layer protocols (for example, protocols for acquisition time) can affect this sensitivity if not implemented properly.

Field Testing

The field testing was conducted in several systems at AT&T Broadband & Internet Services (B&IS) before introducing digital TV 64-QAM signals to ensure adequate quality of the new service. The test results (see examples in Table 1) showed that in the presence of the normal level of other impairments (noise, reflections), the digital signal was not affected if the digital signal level was 50 dB higher than nonlinear distortion levels measured per NCTA practices. For the ratios between 40 and 46 dB, the subjective performance of the digital signals ranged from very good to tiling. For ratios below 40 dB, the digital pictures were either locked or showed heavy tiling.

Digital TV Margin Testing (450 MHz Cascade)						
	Levels			Impairments	Digital Quality	Comments
Test Point	Analog NTSC	Digital TV (50-61)	Beat Level: CTB, CSO	Digital/CTB; /CSO		
	dBmV	dBmV	dBmV	dB		
#1	52	42.5	-36	78.5	Very good	
#2	const	37.7	-25.6, -17.6	63.3; 55.3	Very good	
#2	const	37.7	Both @-8.4	46.1	Tiling	CSO as high as CTB, two interferers resulted in higher total interference power
#2	const	37.7	CSO=-18	52.7	Very good	CSO as high as CTB
#2	const	43.7	CSO=-12.3	56	OK	CSO higher than CTB (affected by digital levels)
#2	const	45.7	-9.4	55.1	OK	CSO higher than CTB (affected by digital levels)
#3	ch 36=38	30	-10.3	40.3	Close to tiling	Misaligned, before realignment
#3	ch 36=38	25	-10.3	35.3	Locked	Misaligned, before realignment
#3		25.5	-19.5	45	Very good	After realignment
#3		25.5	-6	31.5	Locked	After realignment
#3		25.5	-13	38.5	Tiling	After realignment
#3		29	-19.4	48.4	Very good	After realignment
#4		40.5	-13.5	54	Very good	Channel 53 has periodicity (5 dB dip)
#4		40.5	-0.9	41.4	Tiling	Channel 53 has periodicity (5 dB dip)
#4		40.5	-7	47.5	Good	Channel 53 has periodicity (5 dB dip)
#4		40.5	-4	44.5	Good	Channel 53 has periodicity (5 dB dip)
#4		46.5	CSO=-10	56.5	Good	Channel 53 has periodicity (5 dB dip)
#4		49.5	CSO=-6.7	56.2	Good	Channel 53 has periodicity (5 dB dip)
#4		40.5	-20.1; -15.6	56.1	Very good	Channel 53 has periodicity (5 dB dip)
#5	ch 36=40.5	ch 50=31.8; ch 61=24.5	-16.1	47.9	OK	10 dB slope (2/47) and roll-off
#5	ch 36=40.5	ch 50=31.8; ch 61=24.5	-3.5	34.7	Tiling	10 dB slope (2/47) and roll-off
#6	ch 36=34.0	ch 50=25.5; ch 61=16.3	-25.5; -28.1	51	Very good	Flat from 2-47, rolling off for digital
#6	ch 36=34.0	ch 50=25.5; ch 61=16.3	-32; -28.1	57.5	Very good	Flat from 2-47, rolling off for digital
#6	ch 36=34.0	ch 50=28.5; ch 61=19.3	-24.7; -25.3	53.2	Very good	Flat from 2-47, rolling off for digital
#7		ch 50=35; ch 61=31.5	-13	48	Very good	
#7		ch 50=35; ch 61=31.5	-4	39	Tiling on 50 and 61	
#7		ch 50=38; ch 61=34.5	-13	51	Very good	

Table 1: CTB & CSO Impairment Sensitivity of 64-QAM Signal (Field Results)

Lab Tests

The field test could not separate other impairments from the nonlinear distortions. To remedy this shortcoming, the tests were repeated in a controlled environment at AT&T B&IS lab in the test setup presented in Figure 1.

The following tests were performed:

1. Sensitivity threshold (minimum input level);
2. Immunity threshold to:
   - Noise (minimum CNR),
   - Single CW carrier (minimum C/I),
   - CTB and CSO generated by unmodulated (CW) carriers, and
   - CTB and CSO generated by NTSC modulated carriers.

The setup allows for separation of the channel on which the tests are performed from the path in which the impairments are generated. This avoids any influence of the impairment generator on the quality of the reference signal. Moreover, the composition of the interference remains the same since the relative level to the reference QAM carrier is adjusted after generation. The only limitation of the test setup was the fact that the impairments were generated in relatively short cascade but they were generated in a well-behaved region of the amplifiers (far from clipping and saturation points).

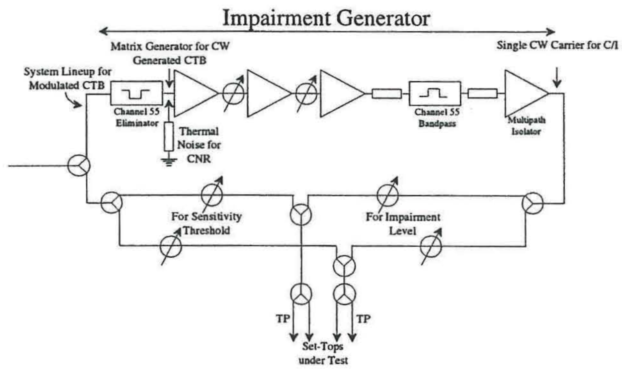


Figure 1: Impairment Test Setup

Impairment Type			Box Type						
			Type 1		Type 2	Type 3		Type 4	
			#1	#2		#1	#2	#1	#2
	Conditions	Units	Impairment Level for Just Perceptible Visibility						
CNR	at -6 dBmV signal to box	dB	22.3	23.0	23.2	22.1	22.8	23.3	24.5
C/I (CW)	at -6 dBmV signal to box	dB	22.8	20.9	23.0	24.0	24.6	22.6	22.1
C/CW CTB	at -6 dBmV signal to box	dB	30.7	NA	36.6	NA	NA	NA	NA
C/Modulated CTB	at -6 dBmV signal to box	dB	34.6	38.8	40.4	35.0	33.8	40.2	40.9
Plant Conditions: CNR=49.9 dB; C/CTB=59.7 dB; C/CSO=65.7 dB (all referenced to an analog carrie 10 dB higher than QAM carrier)									

Table 2: Lab Test Results for 64-QAM Impairment Sensitivity

The tests were repeated for several types of settops. The differences were related to the signal re-acquisition algorithms. The results are presented in Table 2. They clearly depict the problem: sensitivity to single CW interferer is significantly lower than the sensitivity to the CTB and CSO impairments, which are narrowband in nature. Moreover, the sensitivity to CTB distortions generated by modulated carriers is consistently higher than the sensitivity to CTB distortions generated by CW carriers. Both these results show that the NCTA based method of characterizing nonlinear composite distortion is insufficient for digital signals. Additionally, the tests showed that the sensitivity to the composite nonlinear distortions could be affected negatively by careless selection of the signal acquisition algorithms.

Field Tests with 256-QAM Signals

The field test were repeated jointly by AT&T B&IS and GI in a different plant (high quality) with 256-QAM signals. The test setup is presented in Figure 2. The 256-QAM signal was set at the headend to -6 dBc relative to adjacent analog video carriers. Upper and lower adjacent channels contained analog video. A Wavetek sweep was on the system during the first night of testing but was removed for a repeat test the second night.

The results of this test are presented in Table 3. The only problems were experienced on the node that showed CTB and CSO higher than –60 dBc (node 76) as referenced to the level of analog carrier. This corresponds to 54 dB ratio between the level of digital carrier and the level of impairments (note that both CTB and CSO are high and their effect is cumulative). The repeated test showed the same results. This unexpected outcome (relatively good CTB and CSO) supports the need for additional characterization of the bursty behavior of the composite nonlinear distortions.

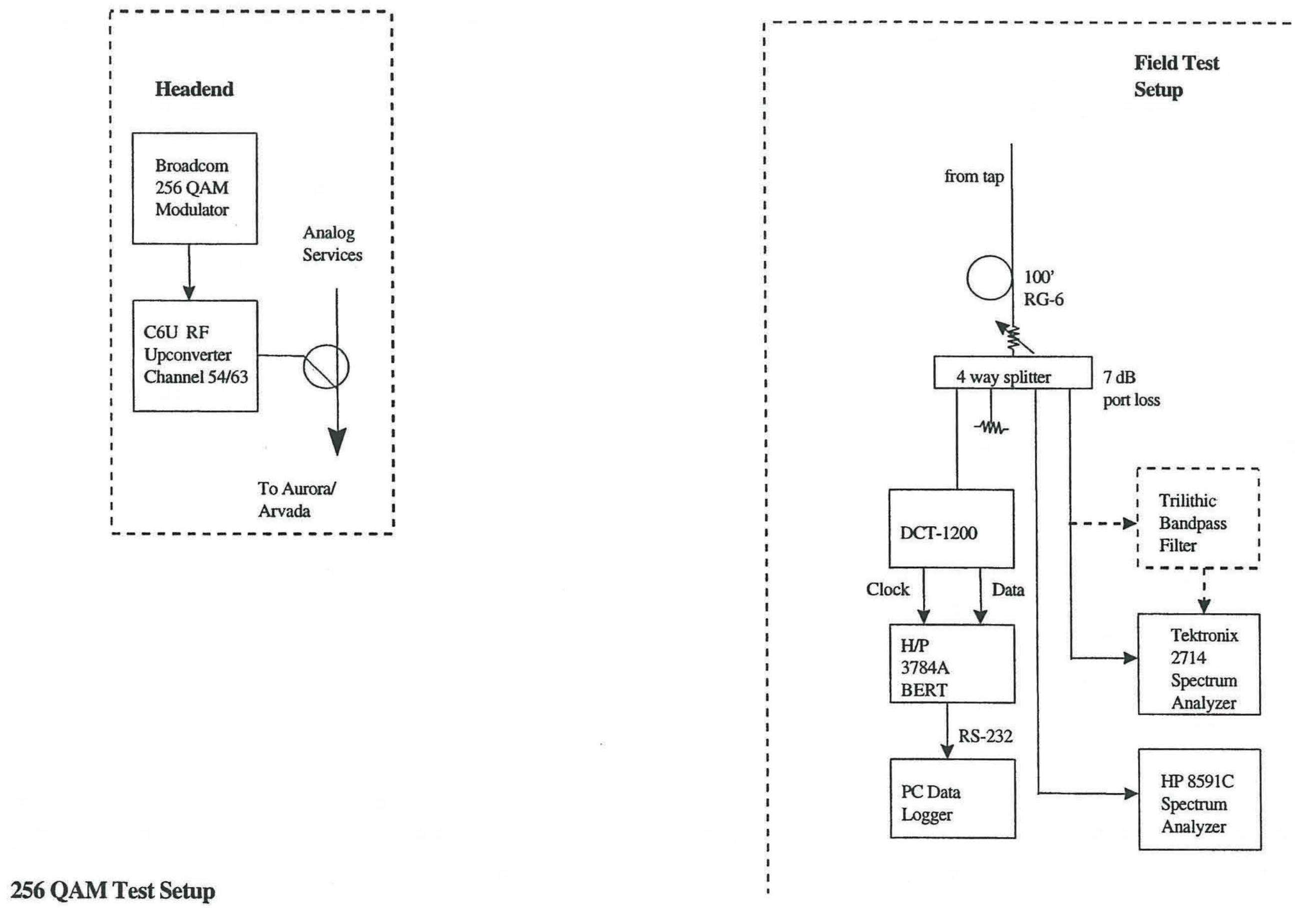


Figure 2: Field Test Setup for 256-QAM Signals

256-QAM test results summary									
	Node 24	Node 32[1]	Node 39	Node 42	Node 43	Node 76	Node 95	Node 99	Node 76
Digital C/N in 6 MHz b/w using channel 63	37.2 dB	>32.0 dB[1]	39.1 dB	35.9 dB	40.3 dB	36.4 dB	36.4 dB	39.6 dB	retest
Digital level into DCT for BERT test, in 5.35 MHz b/w[2]	-3.6 dBmV	-12.9 dBmV	-2.6 dBmV	-1.9 dBmV	-2.5 dBmV	-4.1 dBmV	-7.9 dBmV	-2.6 dBmV	
BERT test duration in sec	1800	1800	1800	1800	1800	43133	1800	1800	55229
Errored seconds	0	0	0	0	1	22	1	0	29
Total bit errors	0	0	0	0	26198	166023	23751	0	230070
% error free seconds	100	100	100	100	99.94	99.95	99.94	100	99.95
Average digital level relative to channel 61	-7.3 dB	-6.5 dB	-7.4 dB	-6.3 dB	-6.5 dB	-7.5 dB	-6.2dB	-7.2 dB	
Channel 3 carrier level[2]	1.8 dBmV	-4.3 dBmV	5.1 dBmV	3.5 dBmV	9.6 dBmV	7.0 dBmV	1.0 dBmV	5.8 dBmV	
Channel 31 carrier level[2]	3.9 dBmV	-3.2 dBmV	6.8 dBmV	6.4 dBmV	6.5 dBmV	5.2 dBmV	1.2 dBmV	7.6 dBmV	
Channel 61 carrier level[2]	3.7 dBmV	-6.4 dBmV	4.8 dBmV	4.4 dBmV	4.0 dBmV	3.4 dBmV	-1.7 dBmV	4.6 dBmV	
Analog C/N: channel 3, measured in service	45.3 dB	45.1 dB	47.8 dB	45.6 dB	46.2 dB	45.2 dB	46.3 dB	45.2 dBmV	
Analog C/N: channel 31, measured in service	46.9 dB	45.2 dB	45.8 dB	47.4 dB	48.8 dB	47.8 dB	47.0 dB	46.2 dBmV	
Analog C/N: channel 61, measured in service	46.1 dB	44.0 dB	46.8 dB	45.0 dB	46.1 dB	46.1 dB	44.9 dB	45.8 dBmV	
Analog CTB: channel 63	64.1 dB	*[1]	67.9 dB	65.0 dB	63.9 dB	59.8 dB	63.3 dB	66.7 dB	
Analog CSO: channel 63	65.0 dB	*[1]	65.8 dB	61.0 dB	64.3 dB	58.4 dB	63.6 dB	66.0 dB	
number of actives	node + 2	node + 4	node + 2	node + 2	node+ 5	node + 4	node + 1	node + 4	

Table 3: 256-QAM test results summary

[1] Noise and distortion readings are not accurate because of low signal level

[2] All measurements are through 100'of RG-6 drop cable and a four-way splitter

---

MEASUREMENT SETUP

To measure the time domain characteristics of a distortion beat, we decided to use a spectrum analyzer as a tuned receiver with the video output from the spectrum analyzer feeding an oscilloscope. In order to fully characterize the time domain information of the waveform, a search was conducted for the fastest oscilloscope with an amplitude histogram feature. A new line of Tektronix oscilloscopes, called Digital Phosphor Oscilloscopes (DPO) is perfectly suited for this task. The DPO feature allows the scope engine to process incoming data at the full rate of the digitizer, without waiting to update the display. The DPO can record about 500 Million points in 45 seconds. Additionally, this scope has the unique ability to output the entire statistics array to a data file. This data file contains information about how often every voltage was "hit" during the test.

A block diagram of the basic measurement setup is shown in Figure 3. The test signals are fed into the Unit Under Test (UUT), filtered, amplified, and displayed on a spectrum analyzer -- just as in a conventional CTB or CSO measurement. The video output connector of the spectrum analyzer is connected to the input of the oscilloscope. After measuring the CTB or CSO on the spectrum analyzer via the normal "NCTA" method, the analyzer is then set to zero span, which makes it function as a tuned receiver, with the output being at the video connector on the back panel. This detected "video" signal is then measured by the oscilloscope.

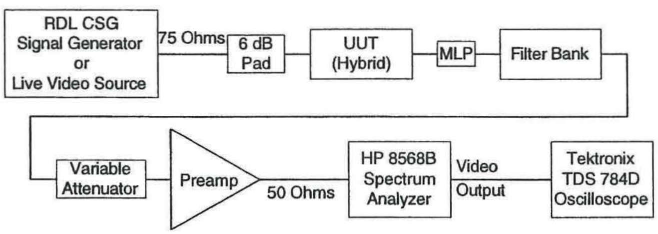


Figure 3: Basic Measurement Setup

Sample oscilloscope displays are shown in Figures 4 and 5. Figure 4 is triggered on a 10 MHz sinewave, which can be seen across the middle of the display. Notice that the tops and bottoms of the sinewave are fuzzy. The DPO records all voltages during the measurement period, so any noise on the waveform will be recorded. In this case 414.97 Million points were measured in about 15 seconds. The left side of the display contains the amplitude histogram. The histogram indicates that the majority of the energy is at the maximum and minimum voltages, which is as expected for a sinewave. Figure 5 shows the same sinewave, but with a free running trigger. To accomplish this, the trigger is set to an unused channel (channel 4) and is set to a high voltage (1.2 V) so that noise will not cause a trigger. Triggering is then set to Auto. Whenever a non-periodic waveform is being measured, it is desirable to not trigger on a particular voltage, since such triggering would cause the histogram data to favor voltages that occur near the trigger voltage. All the CTB and CSO beat measurements were done with a free running trigger. Figure 5 shows that the histogram does not change, but no exact waveform is visible. The color (or shade of gray) of the waveform indicates how often a particular point on the display is hit, with blue being very seldom and red being often.

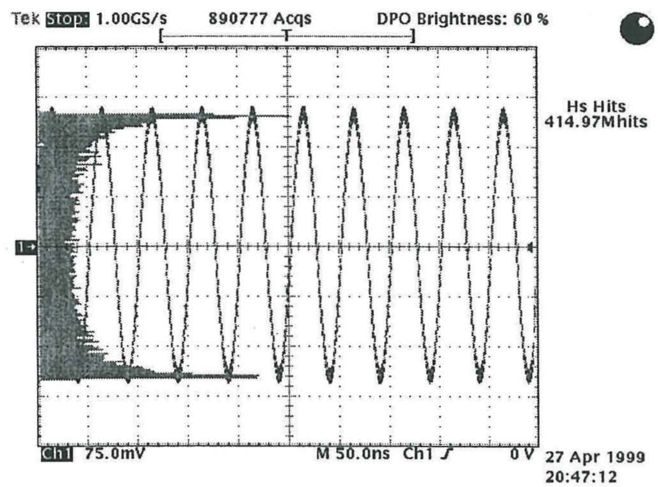


Figure 4: Oscilloscope - Triggered

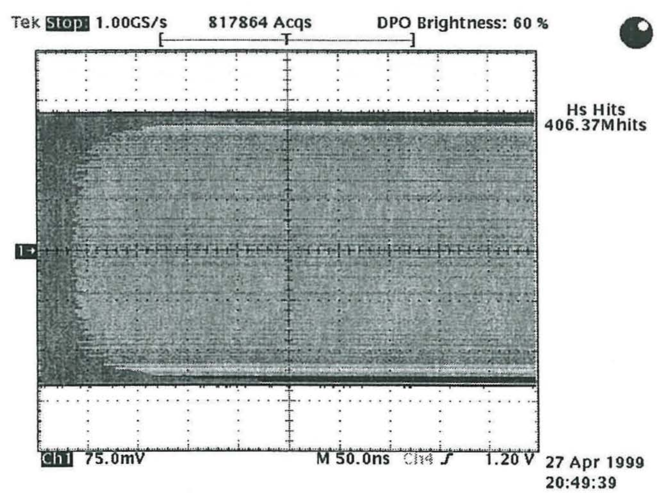


Figure 5: Oscilloscope - Not Triggered

The jaggedness of the histogram displays seem to be caused by inaccuracies of the scope's A/D converter. We did several measurements of various waveforms, including several amplitudes of noise and discovered that the peaks and valleys of the histogram display are always at the same pixels, regardless of the input signal or the volts/division setting of the channel. We concluded that least significant bit (LSB) accuracy of the A/D was to blame for this phenomena. Although LSB inaccuracy has a minor effect on the amplitude of a signal, having wider and narrower pixels (due to LSB inaccuracies) will be immediately evident on an amplitude histogram.

A second measurement recorded in addition to a histogram is a long time domain record of a distortion event. We choose to save a 130,000 point waveform. For each type of distortion being characterized, several 130,000 point waveforms were saved at various trigger levels. This data is useful for an FFT analysis of the spectral components of the distortion and will be discussed later.

Special Equipment Features

There are several important equipment features and setup tricks in addition to the already mentioned special features of the Tektronix DPO oscilloscope. Many of these irritating gotchas were not noticed until many weeks of data had been collected. The most serious issues are mentioned below.

Getting a reliable, calibrated waveform out of the "video" port of a spectrum analyzer is not a trivial exercise. Very few analyzers provide an accurate, high bandwidth, analog representation of the displayed waveform. We first tried to use an HP 8591C portable spectrum analyzer's video output port. (The video port we are referring to here is the one with an analog representation of the detected waveform, not the composite video port used to display the signal on an external TV monitor.) Everything seemed to be working fine, until we started changing reference levels and dB/division. We quickly realized that the voltage at the video port was not tracking the amplitude displayed on the CRT. This led to several conversations with the experts at HP, where we learned that the video output port on the 8591 is before scale factors and calibration data are applied. As a result, it is essentially useless for the measurements we wanted to make. Upon further questioning we discovered that the 8568 (and 8566) are some of the few units that actually do all the scaling, log amplification, and calibration corrections in the analog circuitry. As a result, the video output port accurately matches the displayed values over all scale factors and reference levels. All the newer equipment has "cheaper" analog circuitry, which is corrected by digital processing. So, there is a reason to keep that old test equipment around!

Another caveat is that the video output connection on the back of the 8568 spectrum analyzer is not designed to drive a low impedance load. For many weeks we collected what we thought was valid data, but were unable to solve some correlation problems. We then realized that the oscilloscope had been set to terminate the video connection in 50 ohms. Normally, this would be good engineering practice. However, the output impedance of the 8568 is not low enough to support a 50 ohm termination, and as a result, the values displayed on the scope did not match the values displayed on the spectrum analyzer.

The voltage gain settings on the scope are also very important. The video signal from the analyzer can go a couple millivolts below zero volts. Since our goal was to develop accurate probability functions and, ultimately, to take FFT's of time domain waveforms, we did not want to clip the waveforms. Thus, we set the ground reference of the scope to be 0.5 graticules from the bottom, instead of being at the bottom. Since there are 8 vertical graticules on the scope (with 7.5 being used) and 10 on the analyzer, there are 0.75 scope graticules per analyzer graticule. Thus, if the analyzer is reading 10 dB/div, the data will appear on the scope as 13.33 dB/div.

We had a couple other issues with the analyzer. The first 8568 we used had an abnormality in the video output connection. The output signal favored certain low voltages. We discovered this by measuring noise and noting a spike in the response. We never found the cause -- we just got a different 8568 and started over. Also, the 8568 applies correction factors to the analog circuitry. Thus, more accurate results can be displayed on the screen and exported via the video connection if the user calibration is performed and is turned on.

Measurement procedure

The first step to making measurements is to calibrate the equipment. This includes making sure all factory calibration is within specifications, performing any necessary user calibration routines, and setting the proper gains and levels. All user calibration routines should be run. The HP8568 routine (Recall 8, Recall 9, Shift W) must be specifically turned on to be active. Every time the "instrument preset" is used, the calibration data is turned off. The user must turn it on again.

Next, the scope must be set up. The gain of the scope must be adjusted so that the values displayed on the scope match the values displayed on the analyzer. The vertical offset must be set to -3.5 div, and the vertical gain adjusted such that a signal displayed on the reference level of the analyzer is displayed on the top graticule of the scope. Triggering must be set to an unused channel at a high voltage and be in "auto" mode.

Data can now be collected. The spectrum analyzer should be tuned to the distortion and the level of the distortion measured according to the NCTA Recommended Practices (we generally use Resolution Bandwidth (RBW) = 30 kHz, Video Bandwidth (VBW) = 30 Hz, and Span = 500 kHz). The difference between the distortion level and the noise floor should be noted. Then, the RBW and VBW are set as desired for the histogram measurement and confirmation is made that the noise floor is still sufficiently below the signal being measured. Next, the reference level and dB/div (if in log mode) are adjusted as desired and the histogram is measured on the scope. It will take several attempts to get a reference level that allows signals to fill the scope display without allowing some peaks to go past the top of the display. Never forget to record the reference level of the spectrum analyzer. It will be required for data processing and does not show up in the scope plot.

Processing the data

In DPO mode, the oscilloscope stores a two-dimensional array, which contains the number of times each display pixel was "hit" by the waveform. This histogram file (called an "imh" file) can be converted to a tab delimited file with a conversion program. When the data is imported to a spreadsheet, each voltage is in a column, with the lowest voltage in the left column and the highest voltage in the right column. There are 200 columns which map to the 8 divisions on the screen. Thus, there are 25 pixels (columns) per vertical graticule. Each row represents one pixel across the horizontal axis (time axis) of the oscilloscope screen. There are 500 rows which map to the 10 divisions on the screen . Thus, there are 50 pixels (rows) per horizontal graticule.

To process the data, the sum of each column is calculated. This gives a one-dimensional array containing the number of hits at each voltage. The voltages in the array are the scope's voltages, not the voltage of the waveform being measured. Each scope voltage must be mapped to its equivalent signal voltage at the spectrum analyzer, by accounting for the fact that there are 7.5 divisions on the scope and 10 divisions on the analyzer. Additionally, if the analyzer is in log mode, the dB values must be converted to voltages. Be careful to calculate using the correct impedance (50 or 75 ohms).

To normalize the data, the number of hits at each voltage is divided by the total number of hits. This yields the probability for each voltage, $P(x_i)$, where x_i = the voltage at each data point. The total probability equals 100% ($\sum_i P(x_i) = 1$). The cumulative distribution can be obtained by summing the probability $P(x_i)$ from the first column to the current column.

The probability density function (PDF) requires several calculations to assure unity area under the curve. When log data is converted to linear data, the linear points will no longer be evenly spaced (x_i will not be constant). Without correction, closely spaced x_i will be under-represented on the PDF curve, while widely spaced x_i will be over-represented. To correct this, each data point, $P(x_i)$, must be divided by the difference between the current voltage and the previous voltage, Δx_i. The PDF is obtained by plotting the result ($\frac{P(x_i)}{\Delta x_i}$) for each x_i. By plotting the PDF per millivolt as described above, linear data with a constant Δx_i and logarithmic data with a variable Δx_i will both appear equal. In addition, unity area is guaranteed since $\sum_i \frac{P(x_i)}{\Delta x_i} * \Delta x_i = \sum_i P(x_i) = 1$.

Finally, it is necessary to scale the measured voltages so that CTB and CSO beats with different average powers can be normalized to each other. A CTB with a power of -62 dBm will have a different PDF than a CTB with a power of -65 dBm. This can be corrected by always hitting the analyzer with the same beat power or by mathematically correcting the data.

We chose the latter because making all measurements hit the analyzer at the same level seemed to be problematical. To scale PDF data taken in log mode, we add or subtract the dB correction factor to all the powers. To scale PDF data taken in linear mode, we calculate the multiplication factor associated with the desired correction and multiply all the voltages by that factor.

$$\text{Multiplication Factor} = 10^{\left(\frac{\text{Correction Factor (dB)}}{20}\right)}$$

CTB AND CSO BEAT FINDINGS

One of the first steps to measuring CTB beat characteristics was to choose a proper resolution bandwidth for the spectrum analyzer and sweep time for the scope. Figure 5 shows PDF's of the CTB produced by a hybrid at 547.25 MHz when loaded by 82 channels from 55.25 to 541.25 MHz. Unless otherwise noted, all of our tests used the same 750 MHz push-pull hybrid. After rescaling the amplitude data to account for the slightly higher total average power measured by wider RBWs, all the PDF's fall on top of each other. This is shown in Figure 5.

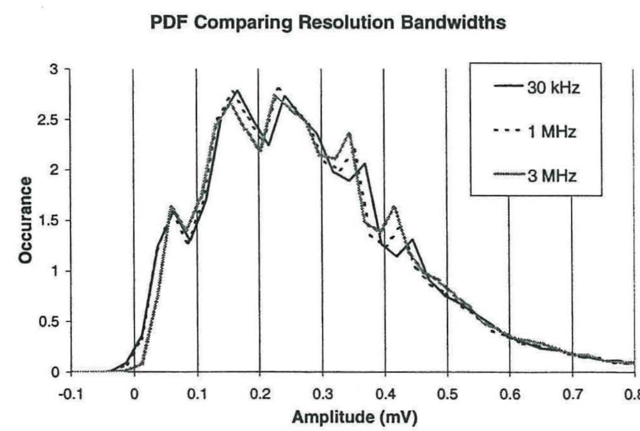


Figure 5: Resolution Bandwidth Comparison

We decided that in order to capture high frequency transient events, it is desirable to use a large RBW. However, larger RBWs allow more noise into the measurement and also capture energy from adjacent beats. As a compromise, we chose a 1 MHz resolution bandwidth. This bandwidth is wide enough to capture the character of the distortion (including high frequency components), but narrow enough to reject, at least partially, adjacent distortion beats. In addition, we settled on a sweep speed of 5 uSec/div (100 nSec/point) for the scope. This allows frequencies up to 5 MHz to be processed.

A major goal of this paper was to compare the PDF's of CTB and CSO through various devices when stimulated by different loadings of CW or live video signals. This proved to be a huge task. The procedures mentioned previously require very accurate and meticulous measurements. Although the process has been continually refined since its inception over 5 months ago, valid data has only been measured in the last several weeks. In particular, properly scaling the measured voltages so that CTB and CSO beats with different average powers can be normalized to each other is critically important. A 1 dB error here can make one draw the wrong conclusion from a PDF comparison. Most measurements of average beat power are done with the "NCTA" method, using RBW=30 kHz and VBW = 30 Hz. However, the average power of a beat changes when the RBW is increased from 30 kHz to 1 MHz. The amount of this increase is variable depending on the signal being measured, but is generally 1 to 4 dB. PDF's should be scaled according to the power in the 1 MHz measurement, since the data and the scale factor should come from the same conditions. When these findings are compared to the NCTA method, this difference in power will need to be incorporated.

Figure 6 compares two drive levels of live video, one being 3 dB higher than the other All the data is scaled to an average power of -67 dBm (approximately 0.1 mV in 50 ohms). The PDF's clearly show that 0.1 mV is the most likely value. The data at the low drive level was taken several times to assure repeatability. No difference can be found between CTB or CSO at high or low output levels. Figure 7 Compares CW to Live Video. Since our concern is with the high amplitude peaks, which are not readily visible on a linear scale, Figures 8 and 9 show the same PDF's on a log scale. Now the difference between Video and CW loading is evident. Later in this paper, we will discuss how the frequency content of these two signals is also different. We must immediately add that our data collection is continuing and we hope to verify these recent findings during the paper presentation at "Cable 99." A sample of the Oscilloscope display during a histogram measurement is shown in Figure 10.

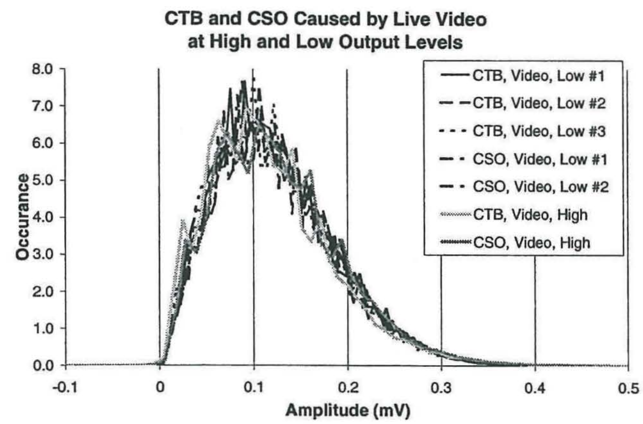


Figure 6: PDF Caused by Live Video

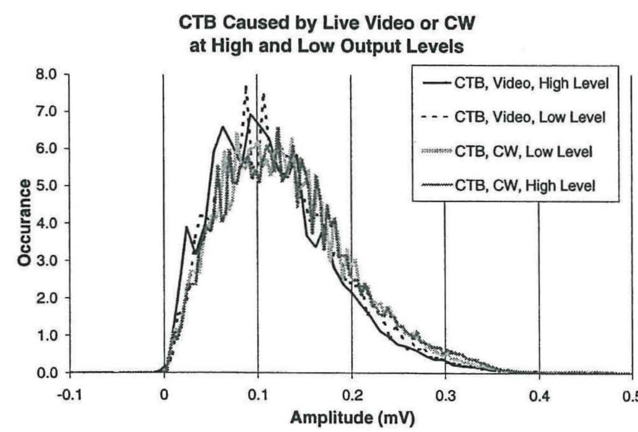


Figure 7: PDF Caused by Live Video or CW

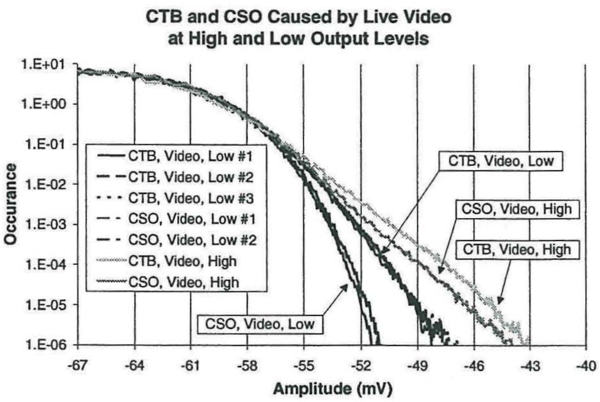


Figure 8: PDF Caused by Live Video (on Logarithmic Scale)

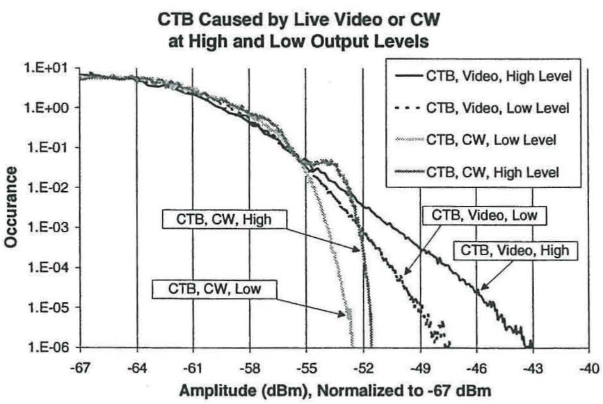


Figure 9: PDF Caused by Live Video or CW (on Logarithmic Scale)

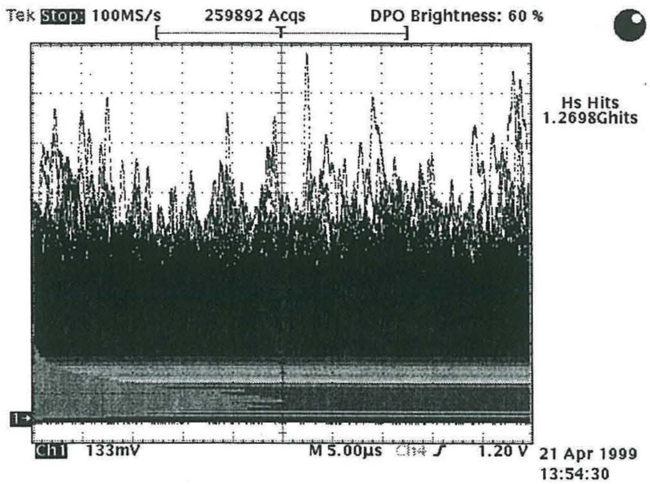


Figure 10: Oscilloscope Histogram Display (CTB of video at high output level.)

BER INTERFERENCE TESTS

In addition to testing the amplitude characteristics of distortion, we did some measurements of the susceptibility of 64QAM and 256-QAM to CTB and CSO interference. The block diagram of the test set-up is shown in Figure 11

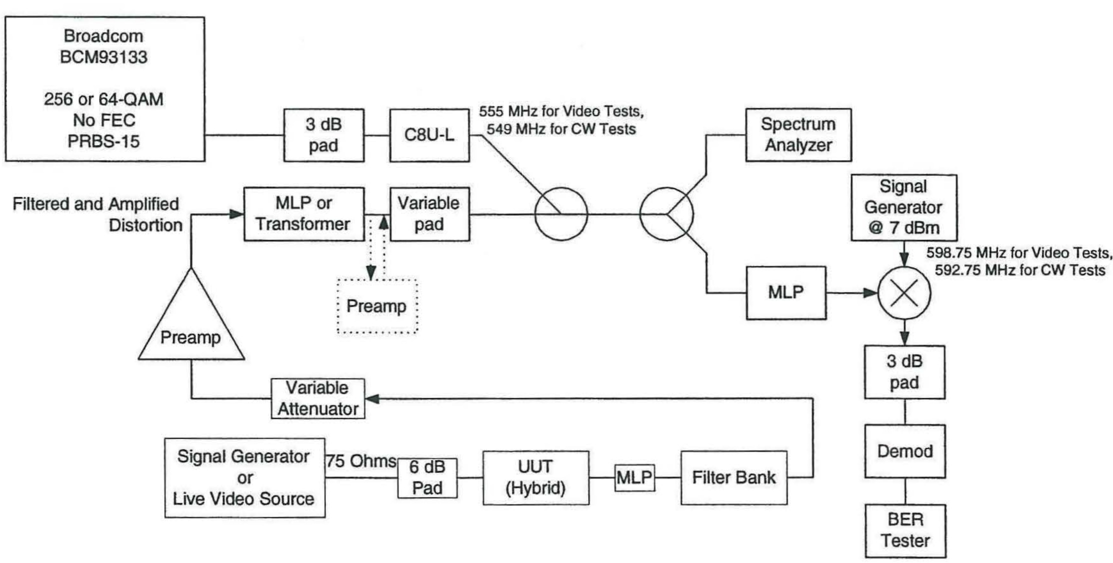


Figure 11: Block Diagram of BER Interference Test

The same UUT (hybrid) was used to create distortion from either a CW or a live video feed. This distortion was filtered, amplified, and combined with the QAM signal at varying levels. At each level, the BER and interference were measured. To get the best possible measurement of the total channel power and the total power of the distortion within the channel, the digital channel power measurement feature of the HP 85721A Measurement Personality in conjunction with an HP 8591C spectrum analyzer was used. This measurement feature provides an accurate measure of the total integrated power between two measurement markers (the 6 MHz channel in this case). The results are presented in Figure 12.

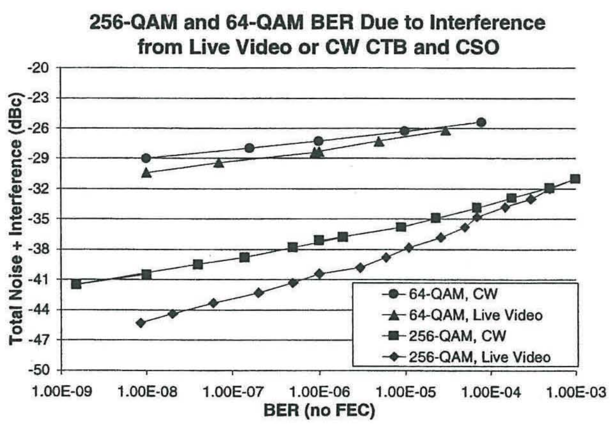


Figure 12: 256-QAM BER

Several important conclusions can be drawn from the results in Figure 12. Most importantly, although 256-QAM is only 6 dB more sensitive to additive white Gaussian noise (AWGN), at BER=10^{-8} it is 14 dB more sensitive to CTB and CSO from live video and 11 dB more sensitive to CTB and CSO from CW. In addition, 256-QAM is increasingly more sensitive to live video (compared to CW) as the interference level drops. Plots of the Live Video or CW distortion compared to the QAM channel at BER=10^{-6} are shown in Figures 13 through 16.

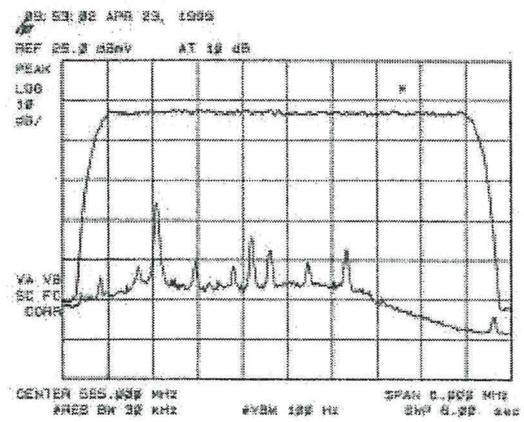


Figure 13: 256-QAM with Live Video CTB and CSO Interference

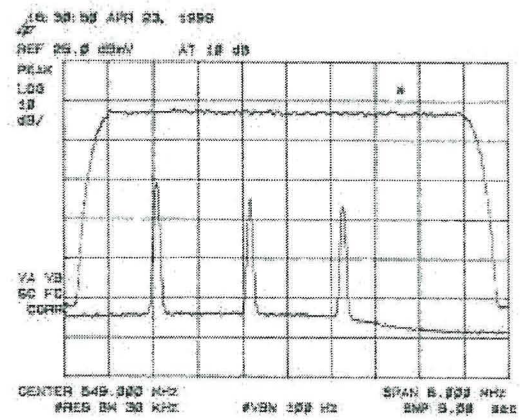


Figure 14: 256-QAM with CW CTB and CSO Interference

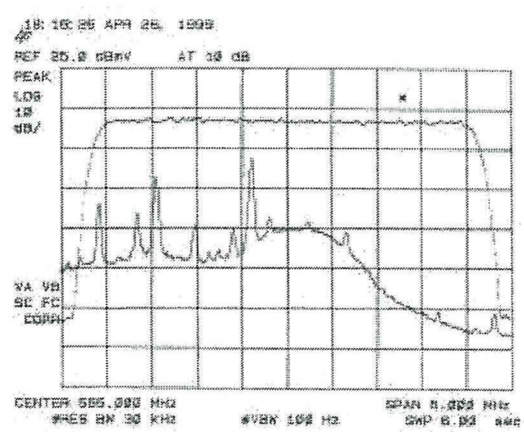


Figure 15: 64-QAM with Live Video CTB and CSO Interference

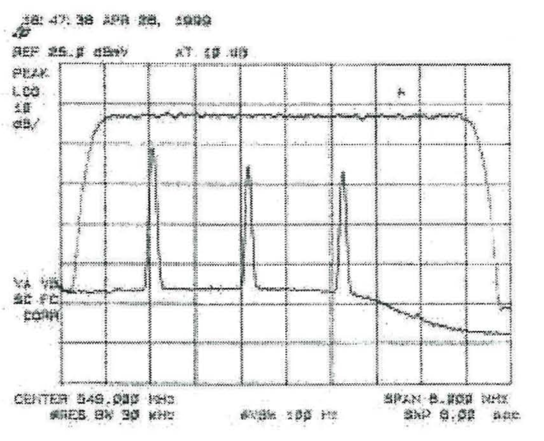


Figure 16: 64-QAM with CW CTB and CSO Interference

PROBABILITY DISTRIBUTIONS

BER

To analyze BER for any degradation type, including simply additive white Gaussian noise (AWGN), a statistical description of the impairment is needed. The "Gaussian" in the AWGN case is that description for common thermal noise. Although seemingly closed form solutions exist for the BER of digital signals in AWGN, these solutions are, in fact, numerical integral solutions. The Gaussian probability density function (PDF) (also called a Normal density), and its cumulative distribution function (CDF) are so commonly used that the expression for writing solutions using the latter has its own function name. By various definitions, it is referred to as the error function. The Gaussian PDF is the best known in the communications industry and many others, and is expressed

$$f(x) = [1/\sqrt{2\pi\sigma^2}]\, \exp[-(x-\mu)^2/2\sigma^2],$$

where

μ = mean (average)
σ = standard deviation.

The two parameters, μ and σ, completely specify any PDF described as Gaussian, a very important property when used in conjunction with other Gaussian properties in mathematically processing random signals. As such, it is often the case that a Gaussian random variable is abbreviated as $N(\mu,\sigma)$, with "N" standing for Normal. In communications, the noise is AWGN, and typically $\mu = 0$. This is another way of saying that thermal noise has a zero average voltage. The noise power in this case is then σ^2.

For digital communications, the related function often used for evaluating BER is denoted Q(x), where

$$Q(x) = {}_x\int^{\infty} [1/\sqrt{2\pi}]\, \exp(-z^2/2)\, dz.$$

Q(x) is basically the complement of the CDF for Gaussian statistics. For QAM signals, the variable x becomes a function of the SNR, or of the SNR per bit, E_b/N_o. For example, in the 256-QAM case being studied here, Q(x) becomes $Q[(8E_b/85N_o)^{½}]$ and the complete theoretical BER expression for performance in AWGN only becomes

$$Pb(256\text{-QAM}) = (15/32)\, Q[(8E_b/85N_o)^{½}].$$

Composite Triple Beat

The case of Gaussian noise is well-known and understood, as AWGN is unavoidable in any system, although the source of the noise fluctuations may differ. What separates the BER analysis of different impairments types in digital communications systems is the modulation employed and the statistical nature of the impairment. It is not always easy to characterize an impairment stochastically. Working with PDF's mathematically is straightforward, although at times cumbersome. However, determining what PDF is to be used is not so straightforward. The degradation that occurs may be associated with nonlinear phenomenon, and a function of many variables themselves of which are not easily described mathematically. If there are enough random variables, none of which are particularly dominant, and they are independent, it is often the case that the Gaussian assumption is valid. This assumption is grounded in the Central Limit Theorem.

For the analysis at hand, the PDF characterization is a two-fold process. First, the distortion beat was captured on a high-speed digital oscilloscope following detection in a spectrum analyzer, and processed as previously described (direct RF detection can also be performed, but would require removal of the deterministic bandpass component of the signal). The scope itself has the capability to perform the histogram function, which has essentially the same information as a PDF, achieved via numerical means. Second, the data captured is exported to a PC and curve fit to determine an appropriate mathematical description. The raw data from the oscilloscope, accessed via software interface, can be loaded into a mathematical analysis program, such as MathCad or Matlab. The data also dumps nicely into Excel, which is also sufficient for much of the processing directly. However, the intricate statistical functions for PDF determination are not as easily implemented, aside from the inconvenience of working with large equations in such a spreadsheet program. Thus, the data captured was then moved into MathCad for further processing.

There are several ways to proceed towards a PDF description. Those familiar with statistics and population sampling may be familiar with a pair of tests known as the Chi-Square test and the Kolmogorov test. The general approach of these mathematical routines is look at a set of data, compare it to an estimated PDF, and determine whether the hypothesized PDF is a good match. There is much more too it than that, as the "confidence" level of the test and "power" of the test are important variables involved in making a determination. These are quantitative parameters. But, the general idea is to utilize known data and compare it to how it should be distributed if it had a certain PDF. The histogram plays an important role in "goodness of fit" testing, as it provides the clues to what PDF's are worthwhile to test. Actually envisioning the PDF fit in a figure is more enlightening. Figure 17 shows an example of captured data (the rough curve), captured and filtered (for smoothness) data, and some overlaid PDF candidates (lognormal, Rayleigh, and F distributions).

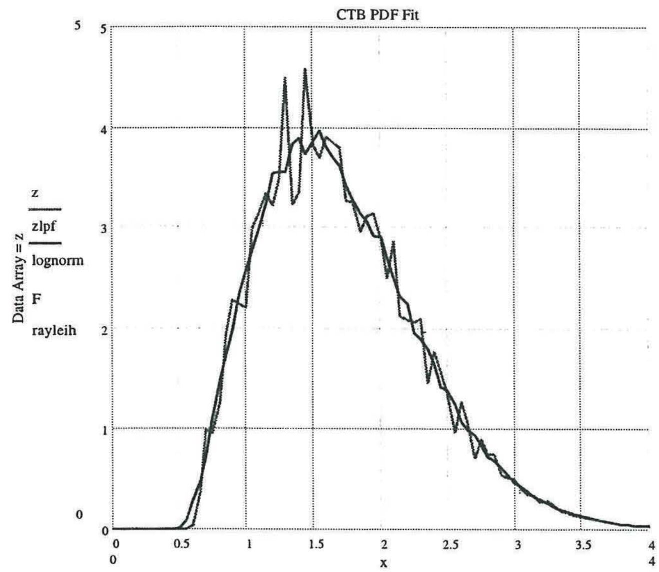


Figure 17: CTB Beat Statistical Fit

As can be seen in the figure, a PDF called the lognormal matches the sampled data well. The lognormal PDF is given by

$$f(x) = [1/\sqrt{2\pi x\sigma^2}] \exp[-(\ln(x)-\mu)^2/2\sigma^2].$$

Another PDF with matching characteristics is the Rayleigh distribution, which is given by

$$f(x) = [x/\sigma^2] \exp[-x^2/2\sigma^2].$$

In this case, the distribution is also shifted on the x-axis away from the origin. This is likely an artifact of the use of linear detection of the spectrum analyzer for capture. This causes poor resolution for very small (near zero) amplitudes, but better response at high amplitudes. The higher amplitudes are the ones of most interest for this analysis, as this is what generates errors.

The shapes of these two PDF's vary as a function of their moments (mean and variance are related to the moments of the Gaussian PDF), so it is not unusual that both can be selected to provide a good fit.

It is also valuable to note that, as has been seen previously, the PDF characteristics vary for live video. As was seen, a longer "tail" exists at the higher amplitudes, which is a characteristic more associated with the slower-decaying lognormal function, which rolls off as $[\ln x]^2$ instead of just x^2.

INTERFERENCE MODEL

CW Interference

The impact on BER due to an additive impairment requires determining how detection is impaired when the additional disturbance is included. In the analysis for CTB beat distortion it is easily recognized that the distortion component is an additive effect, inserting an undesired in-band transmission to the 256-QAM channel. There are two other simplifications that will be made to make the analysis tractable:

1. Assume the interference is narrowband
2. Assume the interference is slow

The first item allows us to proceed by beginning with a signal-to-interference (S/I) problem, where the interference is CW, and predicting and measuring BER performance vs. S/I. It is assumed that the CTB beat is sufficiently narrowband enough to behave in terms of bandwidth as a CW signal. Indeed, the 256-QAM symbol rate is about 5.2 Msps, whereas the CTB beat bandwidth is in the tens of kilohertz. The signal BW is thus over 50 times wider than the CTB beat.

The second assumption is not as strong. The amplitude variations of the CTB beats themselves are slow relative to the symbol rate, which is simply repeating that the noisy modulation on the distortion waveform is narrowband. However, consider just a CW interferer. When converted to baseband for symbol detection, the interference represents an additive component with an offset frequency from the center of the channel, creating a sinusoidal component of the noise disturbance. The offset for a CTB beat would be 1.75 MHz, about a third of the symbol rate. This is obviously slower than the symbol rate, but not necessarily to the neglect of the effects of variation during a symbol. However, fortunately, the slow assumption is a conservative one for a noncoherent interferer, since sinusoidal interference would average towards zero by its fluctuations during a symbol period. Coupled with this conservatism is the simplification of the analysis to a tractable problem.

It is worthwhile to repeat again the narrowband property. Because of this assumption, even a noise-like PDF could not be treated in a customary AWGN fashion, in which the analysis is based on an underlying noise density across the bandwidth. In this case, the noise does not have this property.

Geometry of S/I

A CW interferer generates the familiar "doughnut" shape in a constellation. Figure 18 shows this effect for a real 64-QAM signal. The S/I in this case was -23 dBc, and 64-QAM was selected because of the clarity with which the impairment can be seen for 64-QAM, compared to the plot when 256-QAM is used on the same equipment. The doughnut shape represents the phasor rotation of the additive interference component attached to the end of each vector signal component, which terminates at a constellation point. It is important to note that for this additive impairment, all symbols are

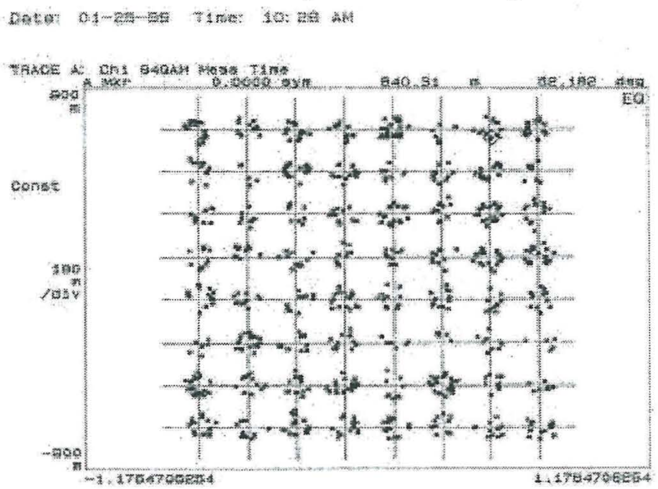


Figure 18: 64-QAM with S/I = 23 dBc and SNR = 40 dB

Modeled performance of 256-QAM is shown in Figure 19 and Figure 20. In Figure 19, virtually noise-free 256QAM is impaired by a CW interferer of -30 dBc. Despite this very low interference power, the disturbance is visually apparent. This is not surprising when it is recognized that, for 256-QAM, the ratio of boundary distance (the distance between a point on the constellation and the "wall" separating it from an adjacent symbol bin) to the average power is about 22 dB. From this figure, we indeed would expect to see some BER impairment when noise is added to the equation. However, it is important to point out that, without adding noise, this scenario would ideally be error-free. That is, for this CW case, the amplitude of the interferer does not change (nor, equivalently, does the S/I). This error free situation emphasized above considers *only* the detection aspects of the receiver design. In other words, despite the presence of large interference, it assumes otherwise ideal behavior of the receiver right up to the point of making a decision on the symbol. In fact, there are other receiver functions for complex modulation like

effected equally, which simplifies analysis. Some symbols have error boundaries on two sides or three sides, but the effect of the interferer on the location of the constellation point is identical.
256-QAM that can be sensitive, including synchronization functions.

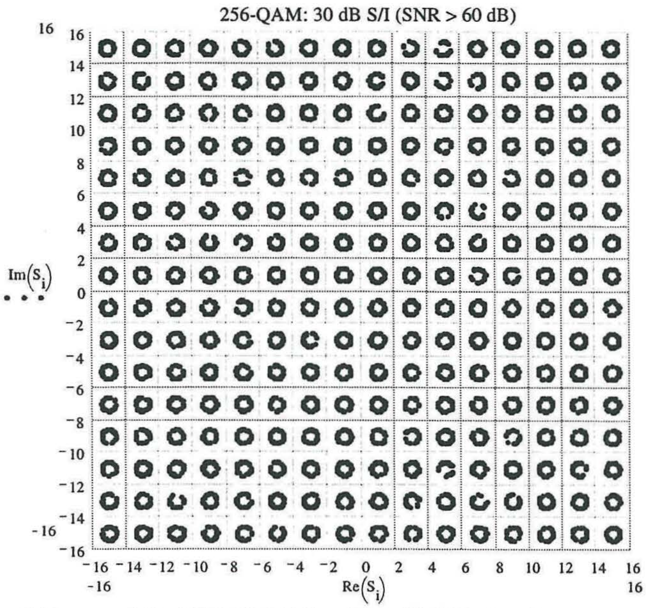


Figure 19: 256-QAM with CW Interference of S/I = -30 dBc and SNR > 60 dB

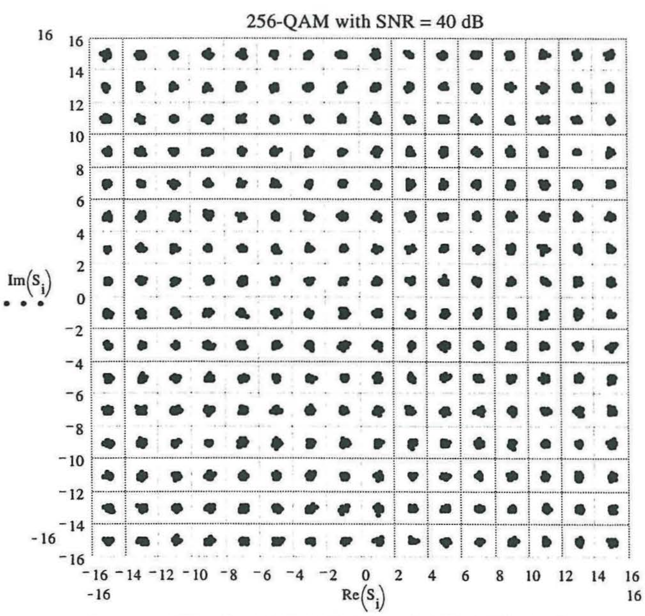


Figure 20: 256-QAM in AWGN with SNR = 40 dB

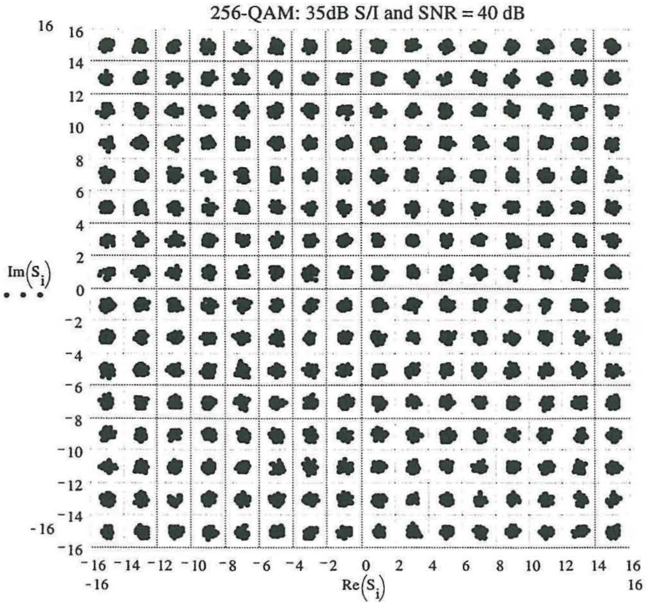


Figure 21: 256-QAM with S/I = 35 dBc and SNR = 40 dB

Figure 20 shows a 256-QAM constellation with AWGN only, with an SNR of 40 dB. In Figure 21, the same SNR is shown, but this time with an additive interferer creating an S/I = 35 dBc. It is apparent that the constellation has been degraded significantly by this amount of interference.

Analytical BER Approach

With a CW interferer, BER analysis can be approached through constellation geometry. Under the slow assumption of the interference effect, the equivalent representation in the constellation is that one point on the "doughnut" represents the impact of the tone on the performance during detection of one symbol. It is further assumed that any point on the circle is equally likely – and thus the phase of the rotating interference phasor is Uniformly distributed.

The result of the interference is that, for a point at some phase ϕ on the circle, the constellation has been shifted towards one decision boundary by [Ki cos ϕ], while the shift has moved it further from the opposite decision boundary by the same amount. Of course, the effect of the moving closer to a boundary dominates the BER effect. This shift can be accounted for by modifying the theoretical BER expression in AWGN by this change in decision boundary distance, which is embedded in the Q(x) argument of the 256-QAM expression.

Since most points in the constellation have four boundaries, we can talk about the above two movements as the in-phase, or I-channel degradation. The quadrature channel, or Q-phase degradation is similar, only based on sin ϕ. Any one of the boundaries being exceeded causes an error, so they are added for composite BER for this symbol. To evaluate the effect of all possible points on the phasor circle means averaging the BER expression over the Uniform phase PDF. That is, BER is calculated for each possible phase and averaging the results over the likelihood of that phase. If it is assumed that all symbols behave the same - not quite the case because of the number of boundaries for the outer symbols in the constellation - then this is all that is necessary to evaluate the symbol error rate. The BER is approximately one-eighth of this, assuming Gray encoded symbols (each adjacent constellation point is different by one bit). Assuming all symbols have four-sided decision boundaries creates a tight upper bound. This is the approach taken, which generated the BER vs. S/I performance plot in Figure 22. We can compare this model to the S/I performance measured on the 256-QAM modem system using CW tones, which is discussed in the laboratory measurement section below.

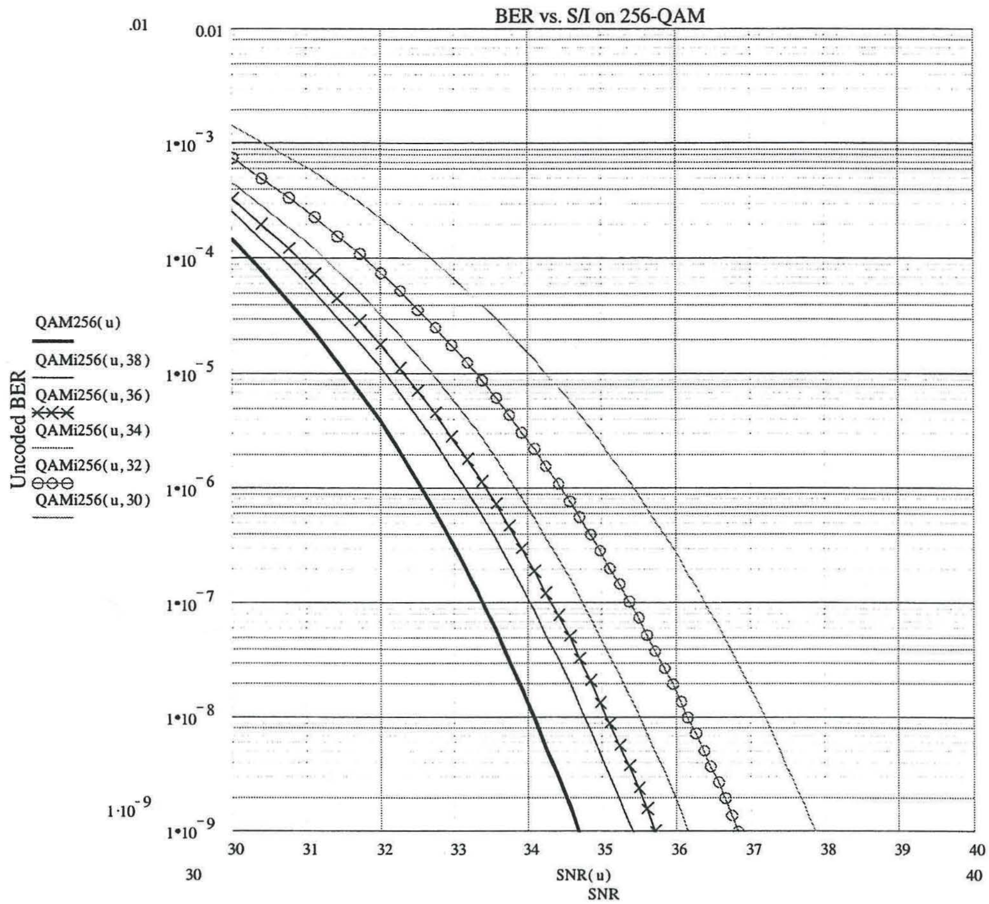


Figure 22: Theoretical 256-QAM vs. S/I

Migrating to CTB

The analysis approach described above is easily adjusted to handle CTB under the narrowband assumption. The adjustment necessary is one that accounts for the amplitude variations of the distortion beat. Since these are, indeed, slow relative to the symbol rate, they can be incorporated to the analysis in the same fashion. Now, however, rather than integrating the modified BER expression with just a change in argument for the interference phase, we include a variable to accommodate another averaging of the expression over a variable amplitude. Thus, the solution becomes a double integral over these two, assumed independent, random variables. The result is an expression in this form:

$$Pb = \int \left[\int Pe(x,z)\, f(x)dx \right] g(z)\, dz,$$

where f(x) and g(z) represent the phase and amplitude PDF, respectively, of the interference. This latter approach is used to handle the narrowband interferer when CTB is that disturbance.

LAB BER VS. S/I PERFORMANCE FOR CW INTERFERENCE

Measured Performance

Prior to the full-blown RF cascade tests, 256-QAM characterization was performed in terms of signal-to-interference (S/I) testing of 64-QAM and 256-QAM. The point of this testing was to begin to quantify the effects of distortions falling beneath the digital channels, an essential part of any composite signal optimization program. Table 4 shows a subset of these measurements. Two cases are shown. The first set indicates the effect of a carrier placed right on the center of the channel, where it is most likely also to disturb synchronization of the receiver. This is, therefore, the expected and measured worst case. The second set shows measurements taken at the CTB offset frequency.

64-QAM fo		256-QAM fo		64-QAM fo-1.75Mhz		256-QAM fo-1.75Mhz	
S/I(dbc)	BER	S/I(dbc)	BER	S/I(dbc)	BER	S/I(dbc)	BER
-50	0	-50	0	-50	0	-50	0
-45	0	-45	0	-45	0	-45	0
-40	0	-40	0	-40	0	-40	0
-35	0	-39	0	-35	0	-39	0
-30	0	-38	1.10E-08	-30	0	-38	0
-25	0	-37	1.10E-08	-25	0	-37	0
-24	2.2E-09	-36	9.90E-08	-24	0.00E+00	-36	2.40E-10
-23	8.0E-07	-35	4.70E-07	-23	0.00E+00	-35	2.50E-09
-22	6.7E-06	-34	2.20E-06	-22	3.20E-09	-34	2.20E-08
-21	5.0E-05	-33	7.40E-06	-21	3.50E-07	-33	9.00E-08
-20	4.4E-04	-32	2.20E-05	-20	7.80E-06	-32	4.80E-07
-19	No Sync	-31	6.60E-05	-19	4.30E-05	-31	2.20E-06
		-30	2.00E-04	-18	5.40E-04	-30	9.00E-06
		-29	No Sync	-17	6.50E-03	-29	3.60E-05
						-28	1.30E-04
						-27	4.00E-04
						-26	2.00E-03

Table 4: 64-QAM vs. 256-QAM Signal-to-Interference (CW) Performance

The results of the S/I tests were quite informative, showing 64-QAM, in general, being about 12-14 dB more robust to an interfering CW signal. Note that this is significantly larger than the theoretical 6 dB AWGN difference. This is not necessarily unusual, in that the CW interferer does not represent a Gaussian disturbance. Typical 64-QAM sensitivity to interference shows BER thresholds of error free operation occurring at S/I of between -21 dBc and -24 dBc, depending on location in the band. For 256-QAM, the variation was significantly degraded, to between -34 dBc to -38 dBc. A very small part of this may be modem system implementation loss between 64-QAM and 256-QAM. Measurements were made with the interference placed across the digital band. BER improved as the interference moved away from the center. The results of the cascaded amplifier distortion performance will be compared to these CW interference tests performed at the CTB frequency elsewhere in this discussion.

BER Analysis vs. Measured (CW)

It is of interest to compare the measurements in Table 4 with the results of the BER analytical model developed based on constellation geometry, as described previously. These measurements were done on a clean channel – in other words, the only losses associated with performance are those related to the non-idealities generated by the modem test system itself. There were no distortion beats or optical impairments. The 256-QAM test equipment is as has been described previously. Implementation losses measured in prior characterizations using Modulation Error Rate (MER) concepts (fidelity of the constellation relative to ideal) and estimated detection SNR at the receiver place this number at about 35 dB. Thus, to compare the measured data with the model requires sliding up and down the SNR = 35 dB line on Figure 22.

The figure plots BER performance for ideal 256-QAM, and for S/I = 30, 32, 34, 36, and 38 dB. Obviously, the closest to the ideal curve is the 38 dB case, and so on down to 30 dB. Table 5 shows the comparison in terms of BER value, as well as in the delta in SNR (x-axis difference) between measured and modeled. Note that for zero BER as measured in the first row, the BER value 1E-12 was used to generate an SNR comparison. At error rates this low, very little additional SNR buys another order of magnitude improvement, so it is not particularly significant what very low value is used. For example, just an additional .4 dB of SNR provides a theoretical BER of 1E-13 in AWGN.

The only unusual result seems to be that at worse S/I, the measured data actually is worse than the model, which should be a conservative value. It is suspected that the reason for this is simply that the model overlooks all but the symbol detection aspects, and at lower S/I, other receiver functions are being effected by the interference also. Additionally, care must be taken in accurately assessing the implementation loss of the receiver, such as in controlled tests looking at the vector analyzer MER as done here, or evaluating carefully in a known pure AWGN link.

BER vs. S/I Comparison			
S/I	Analytical BER	Measured BER	SNR Delta
38 dB	4.00E-09	0	1.8 dB*
36 dB	1.00E-08	2.40E-10	1.0 dB
34 dB	4.00E-08	2.20E-08	0.3 dB
32 dB	3.00E-07	4.80E-07	0.2 dB
30 dB	3.00E-06	9.00E-06	0.7 dB

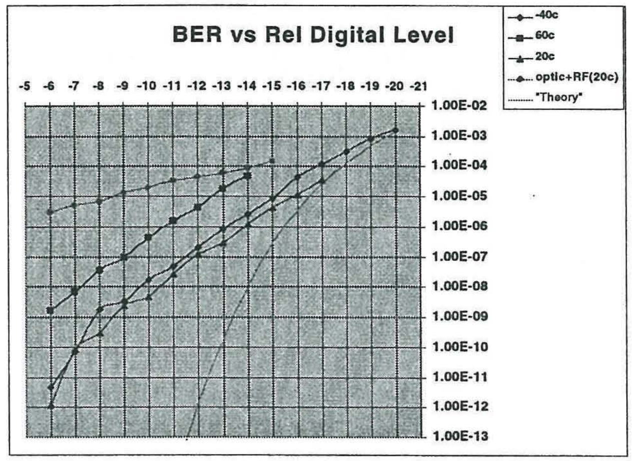
*considering error free as 1E-12
Assumed Imp "Loss" of 35 dB SNR

Table 5: Modeled vs. Measured 256-QAM S/I Performance

Next, a real hardware chain generating actual CTB is considered.

CASCADED RF BER PERFORMANCE TESTS

During the past year, 256-QAM BER measurements have been performed during hardware qualification runs to take advantage of complex chamber setups. This section describes some of these results. All BER results noted are *uncorrected* 256-QAM BER measurements, and we assume a correctable threshold boundary of about 1E-4. This boundary is used as the acceptable BER before error correction that will still provide 1E-8 after error correction. Figure 23 summarizes the results on one particular example cascade. Note that for the RF-only case, SNR is very high, such that BER effects due to AWGN are insignificant for most of the range of digital back-off.



Note: Theory Curve is offset by 3 dB to allow ease of shape comparison, representative of an inherent modem implementation loss. Also, it is based on CNR of a 6 dB fiber link, with CIN treated as an additive Gaussian impairment, and ignoring IMD contributions.

Figure 23: 256-QAM BER, RF and (RF+Optics)

RF Cascade BER Measurements

Low Temperature – Measured digital signal to CTB (S/CTB) = 53 dB for digital @-6 dBc. Distortion performance is best at cold temperatures, and this is reflected in BER. At -40° C, while BER degraded significantly, the QAM level was able to be lowered to the minimum obtainable in the test setup, and the demodulator was still able to lock and hold synchronization. This level was a QAM level relative to analog of about -20 dBc (S/CTB = 39 dB). This is far below what minimum level can be implemented in real plant conditions because of other link degradations, and because of input levels required in the settop box.

The BER degradation curve followed a shallow sloped curve relative to a theoretical plot, which is an unsurprising result considering the overall end-to-end link implementation loss of the system (the dB offset of the curve from ideal). This number, relative to an ideal theoretical curve (as opposed to the one adjusted here for modem and system implementation losses), is about 7 dB @ 1E-8.

The threshold to which the link maintained a 1E-4 level of uncorrected performance is down to a QAM relative level of -17 dBc (S/CTB = 42 dBc). A 256-QAM level of -6 dBc (S/CTB = 53 dB) resulted in an uncorrected BER of about 4E-12. The complete dynamic range over which 256-QAM can run with FEC applied is then about 12 dB with respect to the RF plant.

Room Temperature – The BER performance at ambient closely follows the cold temperature data, with some slight improvement. There is an increase in thermal noise as temperature increases and there was a slight CTB degradation (.5 dB). Generally, however, the difference in performance between cold and ambient is not significant, varying less than .5 dB on the BER curve to a maximum of perhaps 1 dB.

Over the complete operational range of the link, BER performance was between 1.2E-12 (-6 dBc) and 3.5E-5 (-17 dBc). Thus, the dynamic range of the ambient BER is not degraded compared to cold (correctable dynamic range = 12 dB).

High Temperature – There is an obvious CTB degradation at hot temperature at some frequencies (about 3 dB at the frequency of interest compared to cold), as well as the thermal noise increase.

The additional CTB degradation correspondingly results in less dynamic range of the 256-QAM link. Additionally, the best case BER is now about 1.7E-9 (-6 dBc, S/CTB = 50 dB). The theoretical SNR difference between two extremely low error rates, such as between 1E-9 and 1E-12, is quite small - less than 1.5 dB - so this is not alarming. What it does indicate is that, since CTB is quite low on average in absolute terms, it must have significant peak-to-average effects to cause errors, much like noise, but in a narrowband.

At the minimum QAM level, the BER degrades only to 5E-5, a correctable value, before losing synchronization capability at -14 dBc (S/CTB = 42 dB). This represents a dynamic range window of 9 dB on the RF plant, and still is greater range than within the recommended operational digital "window" of -6 dBc to -10 dBc.

Measured C/I Results and Correlation – As previously mentioned, as part of the effort to completely understand and quantify S/CTB performance in forward 256-QAM links, S/I performance was taken using a narrowband interferer summed with the 256-QAM signal. Measurements show, in the worst case (at center frequency) that when the interferer is in the -38 dBc range, error-free transmission begins to get impaired, and it degrades significantly as S/I moves slowly higher. At what S/I errors begin to occur is within a dB or two across the bandwidth. These results turn out to be consistent with the measured cascade data.

As an example, at hot temperature, measured cascade CTB of about 56 dB occurs. For a digital channel @-6 dBc, this is S/CTB = 50 dB. Although SNR was very high, errors were still counted under all cases of digital level, including this maximum digital level. This is because, unlike CW interference, CTB peak-to-average values are noise-like, measuring in the 12-15 dB range. Of course, the measurement technique used in the cascade runs is average CTB per NCTA. Thus, an S/CTB of 50 dB average would correspond to peak CTB excursions as high as the mid-to-high 30 dB's – precisely where CW tests showed error counts to occur. A peak to average of 15 dB would place the CTB peak at S/CTB of 35 dB. In the hot example above, the measured BER for this -6 dBc setting was 1.7E-9. In the S/I data taken using a CW interferer placed at the CTB offset frequency, a BER of 2.5E-9 corresponded to the S/I setting of exactly 35 dB.

On the minimum digital level side, which decreased 8 dB in this test, S/CTB degrades to 42 dB average, and 27 dB peak, with a measured BER of 5E-5. The S/I = 27 dB measurement for the CW test was 4E-4. These are about 1.4 dB different on an ideal curve versus SNR.

As one more data point, consider ambient at -10 dBc. Measured BER is about 5E-9. Ambient CTB is 49 dB for digital @ -10 dBc. Peak is then as high as 34 dB. In CW tests where S/I = 34 dB and the tone placed at the CTB beat frequency, measured BER was 2.2E-8. In terms of SNR difference, this is only about .4 dB. The fact that the CW case performs as the slightly worse of the two in each of these examples is explained by the fact that the CTB impairment is an amplitude varying one. Its average level is, obviously, considerably lower, so most of the time the CTB beat spends below its possible peak. But, as expected, it is the peak disturbances that heavily weight the error rate performance. This is the usual situation for transient BER effects, where, for example, the arithmetic mean of 1E-6 and 1E-9 is very much closer to 1E-6.

BER Analysis vs. Measured (CTB)

As in the S/I section for a CW interferer, a BER analytical model was developed. The modification relative to the CW interference case is the use of a random distribution of the amplitude as previously discussed, which was incorporated via a double integral solution over a modified Q(x). This model and measurements made during the RF cascade testing described above is shown in Table 6. The correlation is quite good, and the results indicate the same pattern as the CW case in terms of modem performance. That is, while the modeled performance is based on a conservative bound, the measured performance was actually the worse of the two as the S/CTB decreased. Again, it is expected that this is the result receiver performance in aspects other than detection was also being impacted by the beat disturbance. The modeled performance is based on the same modem system used in the CW case, and thus the same implementation loss.

It is well-known that a Gaussian noise envelope has a Rayleigh PDF, and a convenient property of the beat's noise-like amplitude envelope is the known relationship between the parameters of such noise, and the Rayleigh characteristics. That is, for a known Gaussian PDF, the Rayleigh PDF is completely known also. This is helpful in the analysis, because it is easier to measure the noise statistics, compared to capturing, processing, and scaling the detected envelope. Then, the known envelope amplitude statistics are used to modify the interference "doughnut" radius' relationship to the detection boundary. This relationship is easily calculated as it is a function of the CTB level, the signal power, and the relationship between boundary distance and signal power.

A complete BER vs. S/CTB plot can also be generated as in the CW case. However, convergence problems of the numerical double integral restrict the range of SNR and CTB over which this can be calculated before the calculation must be modified. Thus, this data is observed best available in tabular form at this point. Further numerical methods are continually being considered to enhance calculations efficiency and accuracy for these complex calculations. The case of CTB effects on 256-QAM BER using randomly distributed CW frequencies looks to be a predictable quantity, and important step as the analysis moves into different frequency patterns, optical links, and load variations.

BER vs. S/CTB Comparison			
S/CTB	Analytical BER	Measured BER	SNR Delta
48 dB	6.90E-10	1.20E-12	1.2 dB
46 dB	1.40E-09	3.10E-10	.4 dB
44 dB	3.80E-09	4.80E-09	.6 dB
42 dB	2.60E-08	1.20E-07	.5 dB
40 dB	4.10E-07	1.20E-06	.5 dB
38 dB	6.70E-06	1.20E-05	.4 dB

Assumed Imp "Loss" of 35 dB SNR

Table 6: Modeled vs. Measured 256-QAM S/CTB Performance

One-Hop Optical Link and RF Cascade

For completeness, and to inspire further study, the final 256-QAM BER measurement in the test suite included a single 1310 nm optical link, with a 6 dB link loss, in front of the cascade. The optics added an expected SNR loss of between about .5 dB and 1.5 dB. The other noticeable change was that the BER curve slope became shallower, indicative of another impairment generating its own set of dominant error-generating statistics. The result of this is that the best BER at the -6 dBc value, was only as low as 3E-6. Clearly, another primary impairment - laser clipping - is at work, considering that the ambient RF-only case was 1.2E-12. These points are about 3.5 dB apart on a theoretical curve. The worst case BER that could maintain modem synchronization was a digital signal back-off from analog of -15 dBc, which measured 1.5E-4.

Results for correctable BER *dynamic range* actually do not vary from the ambient case. In fact, the range measured is about a dB wider.

The BER curve behavior can be explained by clipping. While uncorrected BER is impaired significantly, this is, in fact, somewhat encouraging. The characteristics of the clipping events tend to be handled well by sophisticated FEC employed in the forward link. Beat distortion degradation, by contrast, is not as well handled because of the duration of the peak events of the intermodulation, which challenges the capability of the interleaver and Reed-Solomon coding. Again, this is yet another way of pointing out that the CTB envelope is slowly-varying, thus effecting multiple symbols in series.

FFT INFORMATION

Yet another way to crunch the CTB data in a useful way is through the Fast-Fourier Transform (FFT). The FFT tool is valuable for establishing the frequency content in a detected beat. This content immediately points out the "slowly varying" nature relative to the symbol rate, which is also recognizable directly from the spectrum analyzer itself. The FFT tool is also valuable as the frequency distributions are varied. The ability to store, evaluate, and manipulate the frequency content information of the distortion at baseband is useful for several reasons. First, analyzing the characteristics in post processing is easier, flexible, and can be done with varying parameters. Second, regions of interest in the spectrum can individually be focused upon in more detail. For example, any particular region of the spectrum can be integrated to determine the power associated with some portion of the frequency content of the impairment, such as is done with phase noise impairment in carrier tracking loops. Third, it becomes easy to determine the impact on the spectrum through further processing steps, such as in a digital receiver, or through further RF processing steps in a plant. Finally, discrete components extract themselves in an FFT when swept and captured with very low resolution bandwidths not conveniently observed in analog mode.

The FFT of an input analog comb, processed directly by the high speed oscilloscope at RF and through MathCad is shown in Figure

24. A 649 MHz distortion beat is shown in Figure 25. Along with this FFT is a time domain plot of the same from the oscilloscope. On the screen with the time domain trace is the scope's version of an FFT.

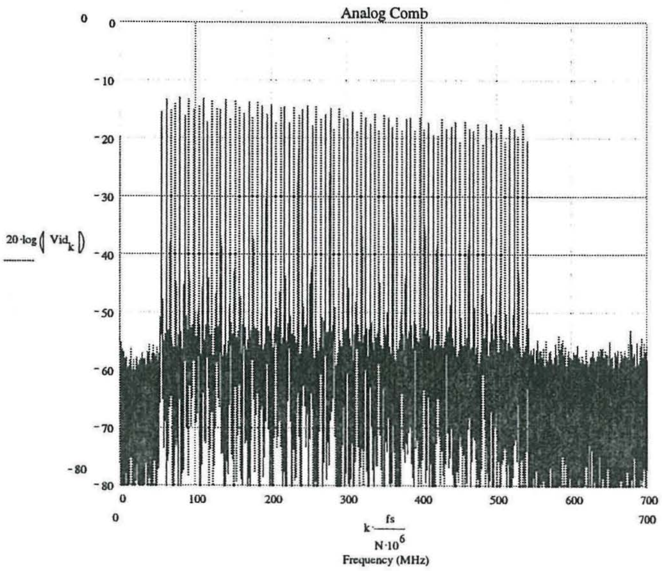
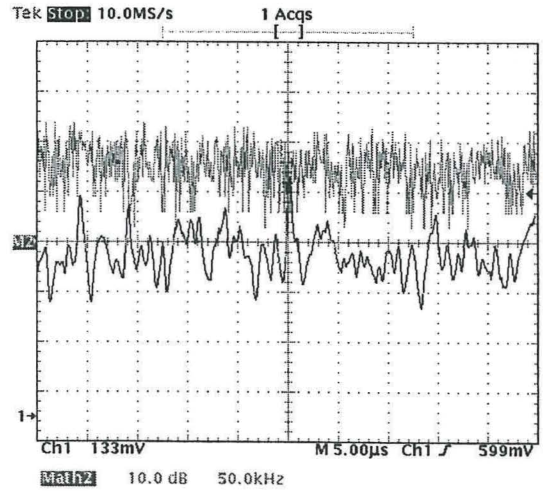
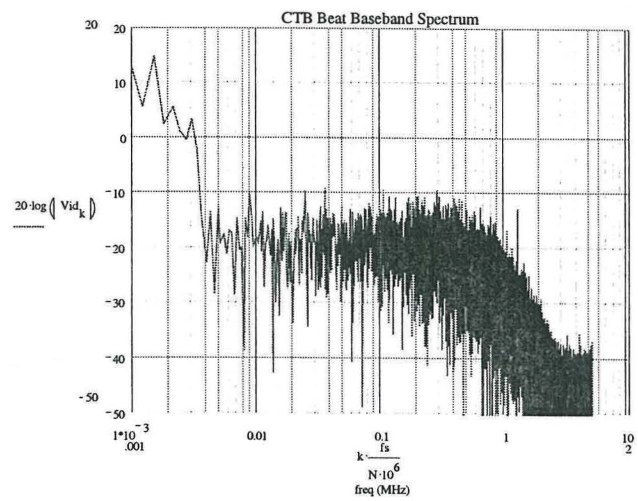


Figure 24: FFT of Forward Analog Comb From Matrix Processed Directly

Of particular note in the processed FFT is the CSO beat at 1.25 MHz. Also, note that the noise floor roll-off occurs at 1 MHz, corresponding to the RBW of the spectrum analyzer detecting the beat. Finally, note that the energy of the CTB is concentrated in the low frequency range, consistent with the fact that the test used a new matrix generator (with non-aged crystals), which thus had frequencies closely aligned (in ppm offset from initial) from one another.



(a)



(b)

Figure 25: Time Domain Plot (a) and FFT (b) of 649.25 MHz Distortion Beat with CW Carriers

Figure 26 shows a CTB beat at 553.25 MHz when live video is used. Again, the CSO beat and roll-off are apparent. Also appearing in the spectrum is the 15 kHz synchronization information that accompanies real video signals.

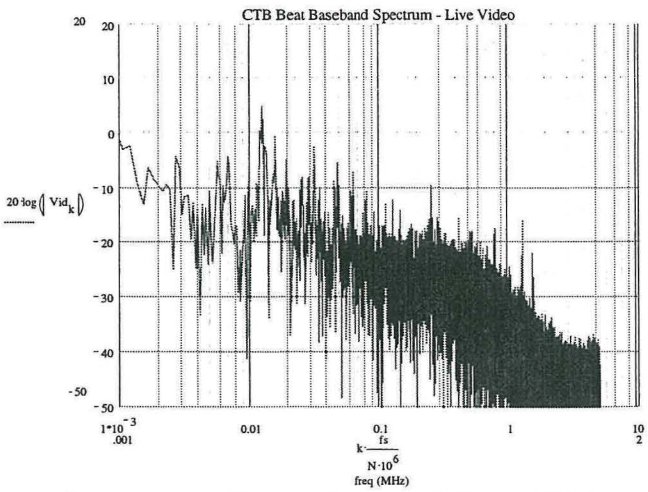


Figure 26: 553.25 MHz CTB Beat FFT using Live Video

As frequency distributions and load levels vary, the effect on the beat amplitude and spectrum - which relates to the duration of beat amplitudes and subsequently the BER - can be evaluated.

CONTINUING EFFORTS

This modeling, analysis, and measurement effort is a work in progress, with many items scheduled for quantification in the months ahead. A subset of the avenues of study are described below.

Frequency Distribution

The majority of data taken for this paper was performed using randomly distributed frequency offsets on the analog channels. Equipment is available to effect the frequency and phase relationship of the channels in any manner desirable. It is known that, as frequencies drift in time, the composite waveform peaks can vary as the drifts align themselves and misalign themselves. This is another way of saying that, statistically speaking, there is a finite, but very small probability that a randomly selected HE will exhibit high peak to average values for some periods of time during its lifetime. The characterization of the distortion when purposely aligning the frequencies and phases is continuing, and the expected higher peak values at more periodic increments has been shown.

These time domain aspects are important in two ways. First, the periodicity implies a deterministic aspect of the peaking that can be used for optimization of the link. Second, they provide information about the dynamics of the symbol error mechanism, which leads to the proper design of the error correction capabilities of the physical layer transmission. That is, for transmission standards of today, the depth of the interleaver and the error-correcting requirements of the Reed-Solomon FEC can be determined if the dynamics of the burst-error generation is known for all cases. As the frequency distribution varies, the power required of the FEC required to mitigate the beat effects also changes.

Further quantification of the phenomenon associated with the distribution of frequencies is in progress.

Optics

Most of the measurements taken so far have been done with distortion generated via RF saturation, in some cases combined into the QAM channel, and in some cases existing because of a loaded cascade. The RF signature of the beat distortion represents an important piece of the puzzle. However, the addition of the optical link enables two important impairments: clipping and second order (CSO) distortion. Second order distortion has been looked at, but has not been the main focus of the study, as this is less significant of a problem in RF-generated distortion. However, because of the importance of second order effects in optics, more focus on CSO is expected.

In terms of CTB, a shift in PDF characteristics is expected as the soft clipping distortion mechanism of RF amplifiers is augmented with the hard clipping distortion. Both soft clipping and hard clipping are functions well studied regarding their effects on signal statistics. Furthermore, nonlinear circuits (pre-distorters) designed specifically for CTB mitigation can impact characteristics.

NCTA Definitions

A primary inspiration for these studies is to develop, crudely, a test procedure modification, at it simplest a "fudge factor" that can be used to correlate measured distortion performance via the NCTA definition to successful 256-QAM performance. Indications are that, because there are multiple impairments that 256-QAM exhibits high sensitivity to, this may be difficult except under very controlled circumstances (like can be achieved in a lab but not in the field).

Other Variables

The nature of either of the distortion beats can be characterized using the methods described here. There are compelling reasons to evaluate the same system implications of the CSO, particularly as measurements move away from RF only. Taking BER measurements with pure CSO and pure CTB would be informative.

Other obvious variables in the composite signal load include ratio of analog to digital levels, frequency, and composite load variation.

SUMMARY AND CONCLUSIONS

For systems that carry both traditional analog video channels in combination with digital QAM signals, the usual NCTA method of measuring distortions are inadequate to ensure robust error-free reception of the digital channels. The statistical properties of the distortion beats as described by the probability density function and its associated cumulative distribution function play a role in determining the acceptable level of the distortion beat falling within the digital channel.

We began with some field data to show the level of visible impairments in the pictures, and followed it with early lab evidence to separate the influence of intermodulation distortion. These early experiments gave evidence of one broad conclusion, that the sensitivity of QAM channel to CTB created by modulated carriers was higher than that of the CTB generated by CW tones. They further showed the need to characterize the bursty nature of the composite distortions.

We began the more thorough characterization with the development of a measurement technique to capture the data from distortion beats in a manner which facilitated analysis and experimental comparison. The body of the paper contains the details of these setup issues and caveats for their use, as well as recommended spectrum analyzer bandwidth and mode settings. The essence of the technique is to use the spectrum analyzer as a tuned receiver and capture the video data on a fast oscilloscope with histogram capability for further evaluation. The instantaneous amplitude of the distortion beat outputs were compared to several "typical" distribution functions, the best matched of which was the Rayleigh distribution. This was true regardless of average value measured via NCTA methods. This same output was analyzed with mathematical analysis software to determine its frequency content via FFT. The FFT data clearly show the band limitations selected by the measurement equipment and the additional CSO beat just beyond the band edge.

Quantitative results from the experimental data comparing 64 and 256-QAM are found most readily in Figure 12. To reiterate, 256-QAM is 6 dB more sensitive to AWGN compared to 64-QAM. However, under the conditions tested here it showed 14 dB more sensitivity to CTB and CSO from live video and 11 dB more sensitivity to CTB and CSO from CW tones. This sensitivity is heightened as the interference level drops.

Analytical interpretations and the qualitative match with experimental data were considered next. It is here that the Rayleigh distribution is fit to the data, and background and significance are explained. Models of various types of interference were described, along with the assumptions and their limitations. Geometric interpretations of CW single interferers and AWGN for 64-QAM and 256-QAM were also given. These were followed by analytical BER evaluations and an approach to understand CTB in terms of the analysis, and hence BER as a function of the CTB.

Lab measurements of BER for various levels of S/I were presented, and show that that 64-QAM is approximately 12 dB more robust to a single interfering tone regardless of whether the tone was at the center frequency of a channel or at a location consistent with a CTB beat. These data, the earlier field data, the analytical indications, and the comparison of live video to CW tones show fairly consistently that for low levels of live video interferers, 256QAM will be more sensitive by this 12-14 dB. Measurements of RF amplifier cascades at various temperatures show the functional dependence of 256-QAM BER on level of interference, and the addition of a fiber optic link demonstrate the importance of considering all the sources of impairment . In particular, laser clipping limits the attainable level of BER for a given level of interference compared to that of an RF cascade alone.

In conclusion, we are beginning to make headway toward answering the question of how to define completely the relative level of a distortion beat that impairs a digital (QAM) channel. Instantaneous distributions of distortion beats are certainly a factor in answering this question, but not totally independent of the type and source of interfering signals. In addition, the methods of forward error correction will have more or less influence again depending on the various types, sources, and levels of the interference.

Looking ahead, it is clear that further work will be required in fiber optic links, especially to quantify the impairment induced by clipping. Expected variations in the PDF of distortions need to be quantified, as does the influence of the source frequency variation. Plans for further work also include the influence of various channel plans, number of analog versus digital channels, relative phasing of the frequency sources, and quantification and interpretation of time domain data including number, distribution, and duration of peaks leading up to and following a "trigger" from a particular amplitude interfering event. Finally, more complete data on the effects of CW to QAM power and the relationship to average power of a distortion beat as measured by NCTA methods will be pursued, as will the influence of cascade depth and optical links. It is anticipated that a comprehensive interpretation of influences on QAM BER will lead to greater understanding of today's components, systems, and their interaction and will help to suggest elegant solutions to overcome limitations.

ACKNOWLEDGEMENTS

We would like to acknowledge the following individuals for their valuable contributions to this paper, and their continued efforts on this program: Javier Borlando, Vipul Rathod, Mike Short, and Joe Waltrich

---

[i] NCTA Recommended Practices for Measurements on Cable Television Systems, Second edition, revised October 1993

DWDM: MATCHING TECHNOLOGY ADVANCEMENTS WITH BUSINESS REQUIREMENTS

Venkatesh Mutalik
Staff Engineer, Advanced Fiber-optics
Philips Broadband Networks

Abstract

Dense Wavelength Division Multiplexing (DWDM) technology has been used in digital communications systems for several years, but its use in analog CATV networks is only recent. Although the components for DWDM technology exist, their use in CATV systems is quite different from that of digital systems, and the requirements are quite often complicated and conflicting.

In general, factors such as number of optical hubs, required bandwidth capacity, transport distances, required RF level differential between the broadcast and narrowcast inputs, choice of return transmission scheme determine the shape and transparency of the envisioned network

Proper analysis of the above parameters determines the number of optical wavelengths needed, level of transparency of the optical hub, the choice of a proper multiplexing scheme for the return transmission system, and the number of nodes that may be combined on each return DWDM transmitter.

These choices are, in turn, ultimately limited by fiber effects such as Stimulated Raman Scattering (SRS) cross-talk, fiber dispersion, EDFA noise and gain-tilt, and available passive and active component specifications, like DWDM laser chirp.

This paper will examine the state of the art in DWDM technology from a system architectural point of view. The goal of the paper is to explain several competing architectures within the framework of DWDM technology in an effort to insure a cost effective and optimally functional network design.

INTRODUCTION

We start with the question "How much fiber is enough?" A Network Manager recently explained to me that his company started a particular network about 3 years back with 12 fibers connecting a ring of hubs with the main headend. The next year, the demand was for 24 fibers, and this year, the demand is for 48! "Nobody can tell me why we need 48 this time around, but it looks like 48 is the magic number this year" he said. Although this may be an extreme case, the truth is most MSOs today realize that headend fiber counts are limited as compared to potential demand. By the time they allocate fiber for broadcast, telephony and internet access using standard equipment, most MSOs struggle trying to determine how much additional fiber is needed for broadcast and interactive services.

DWDM offers a way out. In this technology, several wavelengths of light in the 1550 nm low-loss wavelength window are multiplexed into one optical fiber, thereby increasing its capacity many times over. Although this technology has been used by digital systems for many years; it has only recently been introduced in the analog CATV realm. Yet in less than a year, DWDM has found many passionate proponents. However, as with any other technology, operators and designers face the challenge of determining when DWDM makes sense and when it does not.

DWDM SYSTEM ISSUES

This paper will provide a quick summary to give substance to the discussions that follow. The term "*RF channel*" is used to designate regular sub-carrier multiplexed systems (such as 77 NTSC channels etc). This is to prevent confusion with "*wavelength channel*" that describes multiple optical wavelengths multiplexed using the DWDM technology over one fiber.

Fiber Dispersion

Fiber dispersion along with laser chirp generates CSO and CTB in multi RF channel optical transmission systems. The dispersive effect is dependent on the number of *RF channels* over single wavelength, RF frequency of operation, fiber link length and the laser chirp. Higher chirp lasers over longer distances of fiber have worse CSO and CTB and so on.

Figure 1 shows a DWDM transmitter's performance with a 50 km fiber link and an equivalent passive link without fiber. In each case, the receiver input power is maintained constant. It is seen that CSO (located between 50 to 200 MHz) is enhanced due to fiber dispersion, and although not clearly visible in this graph, there is CTB degradation as well.

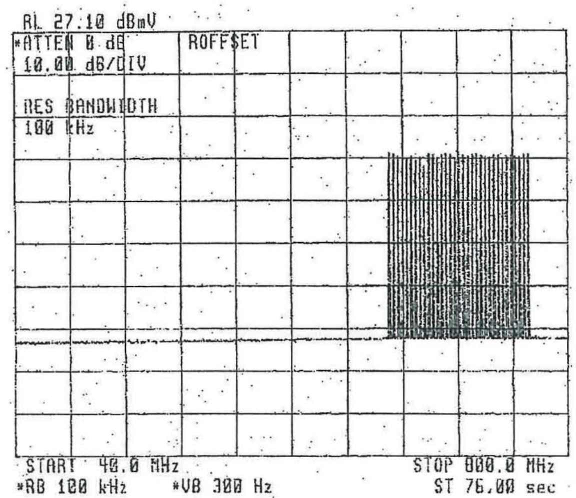


Figure 1a

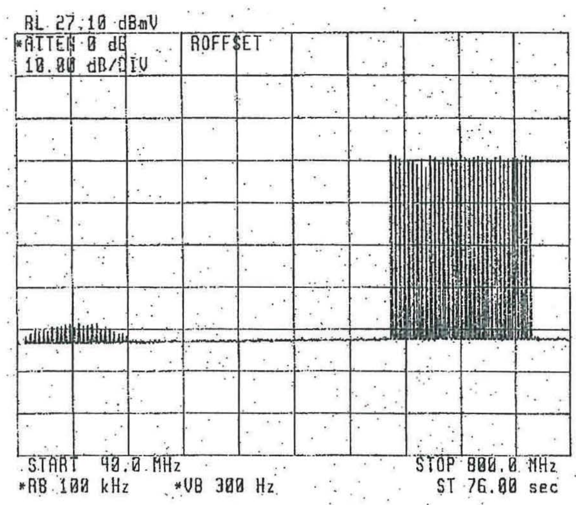


Figure 1b

Figure 1. RF spectrum on DWDM transmitter with passive attenuation (Fig. 1a) and over 50 km of fiber (Fig. 1b)

In long fiber link systems, fiber dispersion is one of the most severe limiting factors. Several compensation techniques exist that reduce this effect. The most obvious is the externally modulated transmitter which because of its almost non-existent chirp generates very little fiber induced CSO and CTB. Other ways include dispersion compensating fiber (DCF) and fiber Bragg gratings. All of these technologies are lossy and expensive. Proper *RF channel* allocation to eliminate in band

CSO (which happens in overlay systems since the narrowcast signals are from 550-750 MHz), combined with patent pending techniques to compensate for fiber dispersion using electronic means offer the least expensive solution today.

Fiber Non-Linearities

In addition to fiber dispersion, several fiber non-linearities exist due to the small core size. The most prominent that concerns us is Stimulated Raman Scattering (SRS) cross talk. SRS cross talk in fiber links appears in multi wavelengths systems where the longer *wavelength channels* get amplified at the expense of *shorter wavelength channels*. It depends upon RF frequency, fiber link, EDFA spacing and output power and polarization.

Figure 2a shows equipment set-up to investigate the effect of polarization effect on SRS. The dark well defined line maximizing at –13.70 dBmV is the SRS crosstalk when the system polarization is set-up to maximize SRS. Next, with the whole system entirely intact, the polarization paddles are moved 50 dB! Polarization effect (which occurs with fiber movement) on SRS is so strong that it could completely mask SRS effect in a potentially SRS limited system. Any experimental set-up will generally severely underestimate SRS cross-talk if proper care is not taken to account for the worst-case polarization states. This holds particularly true for more than 2 wavelengths DWDM systems.

Accordingly, the only technically acceptable way to characterize SRS in DWDM systems is by design and calculations confirmed by testing worst case SRS scenarios. SRS is measured in a two-wavelength system and the SRS effect is prorated for multiple wavelengths and for different EDFA fiber links. Since polarization state of the optical link cannot be predicted beforehand – it may be influenced by the climate inside and outside the plant that is dynamically changing - the design must be effected for the worst case SRS effect. The SRS effect changes rapidly with RF frequency. The severe RF frequency dependence of SRS may be understood as an

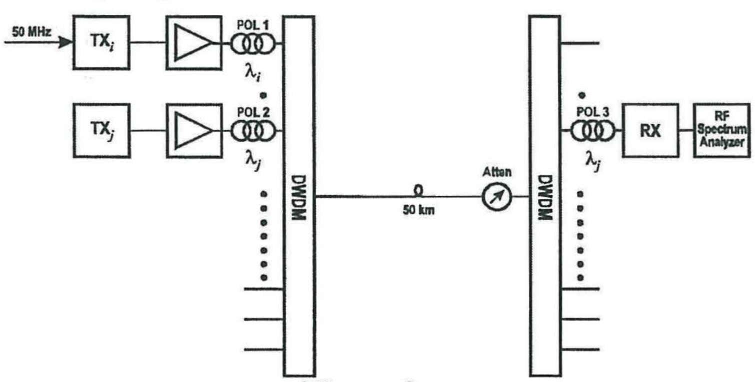


Figure 2a

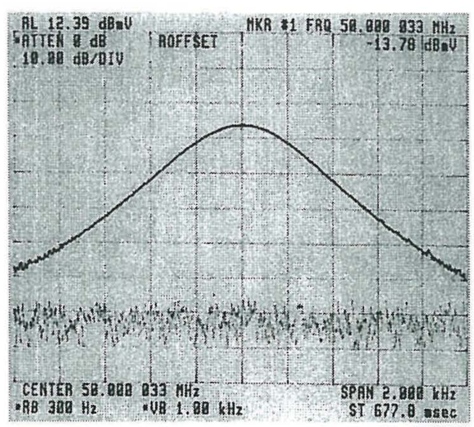


Figure 2b

Figure 2. Setup to measure fiber induced SRS effect. Notice three polarization controllers (Fig. 2a). Polarization effect on SRS. It is possible to completely mask the effect of SRS by changing polarization

wavelength 1550 nm transmission still apply here, and proper design must be incorporated to limit this. Four Wave Mixing (4WM), so called because two optical wavelengths interact in the fiber and create two additional wavelengths around themselves is much below the SRS effect in most multi-wavelength systems, due to the high fiber dispersion.

It is clear that fiber dispersion is a necessary evil. Were it not for fiber dispersion, SRS and 4WM effects would dominate and severely limit the system performance. However, dispersion induces distortions in the RF domain, which could limit system performance anyway. The techniques to limit dispersive effects while still gaining the needed benefits is a balancing act.

Because of the above discussion decisions regarding the use of different fiber types such as DSF (dispersion shifted fiber) that has zero dispersion at 1550 nm, must be evaluated carefully. Since the core size of DSF is smaller than the regular SMF 28 fiber, the resulting fiber non-linearity is larger to begin with. When combined with the lack of adequate dispersion however, SRS and 4WM are significantly enhanced. Other fibers such as the NZDF (non-zero dispersion shifted fiber) that has small amount of dispersion around 1550 nm primarily to reduce 4WM may not adequately reduce the SRS effect. Fibers such as the LEAF (larger core size) do reduce non-linearities, and the AllWave fiber eliminates the water peak at 1400 nm. These may be employed for future purposes. One of the elegant options however is to accept the dispersion for reducing the non-linearity and to compensate for the resulting distortions.

EDFA Gain-flatness

Traditional EDFAs do not have a "flat" gain bandwidth with respect to wavelength. This means that a "flat" multi wavelength input would not result in a flat output. Large wavelength tilts are undesirable not only due to SRS considerations, but also because system design in complex DWDM networks is considerably complicated as receiver input powers cannot be calculated and guaranteed accurately.

Presented in Figure 3 is a measured graph of a 20-wavelength channel DWDM system through two EDFAs and 105 km of fiber. It is seen that the system has a peak to peak flatness of about 7 dB EDFA gain equalization in such cases must be attempted

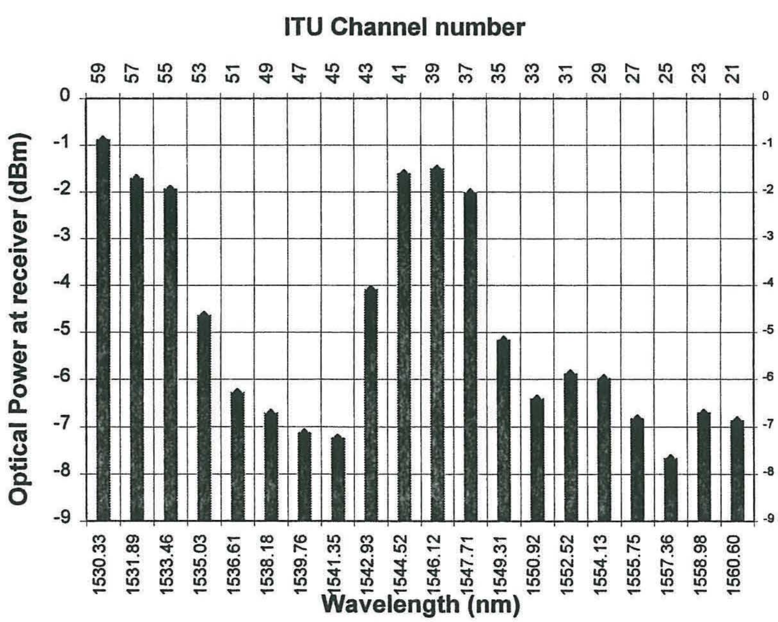


Figure 3. Measured EDFA gain tilt through two EDFA and 105 km of fiber for a 20 channel DWDM system

where the input wavelengths are pre-emphasized to achieve the required output flatness. Sophisticated computer programs are used to estimate the pre-emphasis as the gain-flatness for EDFAs dynamically changes with the optical input power. In a typical system employing several wavelengths, it is almost always necessary to pre-emphasize at least a few wavelengths.

Internally Gain-flattened EDFAs for particular input levels are very popular for long haul digital interconnect (2.5 GB/s) operations. These EDFAs could be restrictive to CATV system designers as in most cases, amplifier spacing in CATV networks is dictated by varying hub locations where maintaining precise optical input levels may not be possible.

Other Effects

Since the fiber core guides the shorter wavelengths better than the longer wavelengths, minimum optimum fiber bend radius for 1550 nm fiber management is much larger than it is for 1310 nm systems. Accordingly, proper fiber routing inside the node where space is limited is extremely important. This is particularly important for DWDM system where a precise RF level differential must be maintained and trouble shooting is complicated.

DWDM transmitter Loading, SNR and Clip Margin

Loading of DWDM transmitter is of the form of QAM64 or QAM256 signals. It is generally assumed that each RF QAM channel is spaced 6 MHz wide (with 5 MHz of noise bandwidth) and 8 such QAM Channels fit in a 50 MHz bandwidth or 32 such channels fit in a 200 MHz bandwidth. Generally broadcast transmission is from 50 to 550 MHz, and the RF plant in the US is usually 750 MHz. Therefore, 550 to 750 MHz loading represents the maximum possible loading in most typical cases.

Digital signal performance is represented by the signal to noise ration (SNR) as opposed to the CNR. In this paper, all data for narrowcast transmission is presented assuming a 5 MHz noise bandwidth.

A Word about Clip Margins

The CNR that we commonly understand for analog networks is the ratio of the "average carrier power" to the "average cumulative noise" within the video bandwidth. This concept works very well for analog video signal transmission over the optical links. However, QAM64 and QAM256 signals are usually described as a constellation where there is considerable variation between their peak powers and their average powers. This is particularly severe when a "large number" of *RF channels* (such as 32 QAM channels in the 200 MHz loading) modulate the laser.

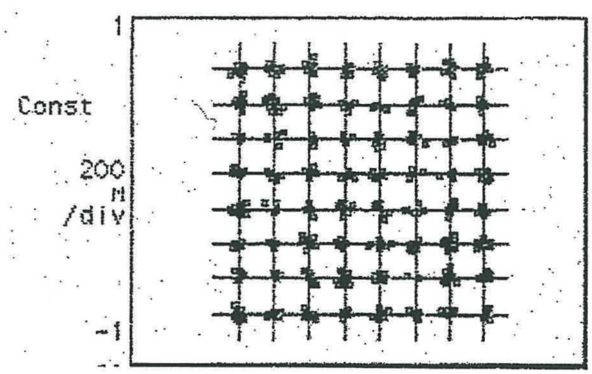


Figure 3a. QAM64 Constellation diagram illustrating peak and average power

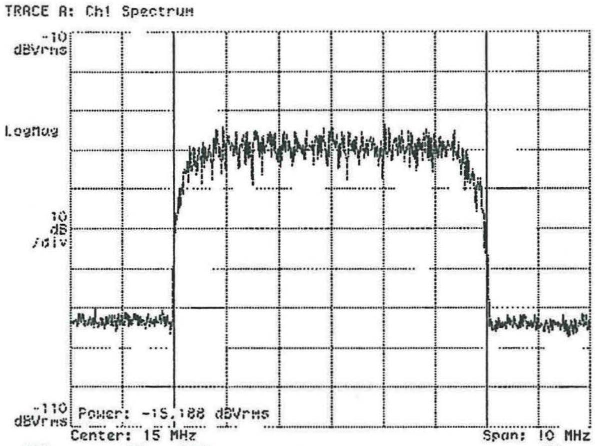


Figure 3b. The average power normally seen on an RF power meter or the vector signal analyzer

While the combined average video carrier power does not clip the laser, their peak powers clip lasers several times. In analog video transmission, these clip time periods are small and do not affect the picture quality significantly because the eye averages out minor glitches in the picture. However for digital systems, when the laser is in even very modest clipping, (the laser light is essentially off), bit errors result. Empirically it is found that for a "large number" of QAM channels modulating the laser, the total RF input to the laser must be run 2 to 3 dB below the normally understood value of clipping. This is designated as the *clip margin*. All discussion to follow includes a 3 dB clip margin for the narrowcast transmission, and the graphs to be presented must be understood in this context.

DEVICE PERFORMANCE

DWDM transmitters must be maintained at very closely held optical frequencies set by the ITU channel numbers. DWDM performance is dependent on three factors

Temperature performance

Temperature performance over specified temperature range usually 0C to 40C and long term stability and drift of the DWDM laser and its temperature control circuitry. Such control circuitry is crucial and must be tested and verified.

Transmitter Performance

The yield for higher chirp devices (>300 MHz/mA Chirp parameter for 10 mW) is larger, and such devices are less expensive. However as noted above, higher chirp devices in collaboration with fiber dispersion generate large amounts of fiber CSO and CTB thus limiting their application, unless something can be done to compensate for dispersion.

Passives Selection

Passive DWDM components are usually thin film based or grating based or AWG design based. Their insertion loss, temperature performance and passband flatness is of interest for DWDM performance.

In the current state of the art, we have transported 20 optical wavelengths, over more than 100 km, with thirty-two QAM256 channels per wavelength, after taking into account all fiber non-linearities and dispersive effects over standard single mode fiber. This represents a 200 GHz spacing on the ITU raster, and is essentially limited by the gain bandwidth of the EDFAs. With some effort, this spacing may be reduced to 100 GHz, thereby increasing total capacity to a 40 wavelength DWDM system.

DWDM IMPLEMENTATION

Single Receiver architecture

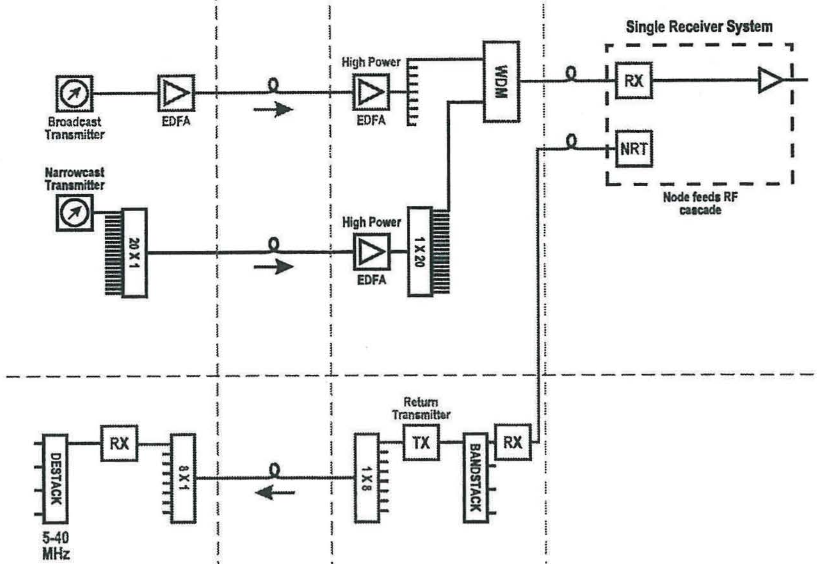


Figure 4. A single receiver system with narrowcast and broadcast optical signals combined on one single receiver at the node

As is seen in the above diagram, the each receiver receives the cumulative input of the broadcast and narrowcast input. The receiver now has an aggregate noise from both the transmission media including the shot noise, laser RIN, EDFA noise and fiber

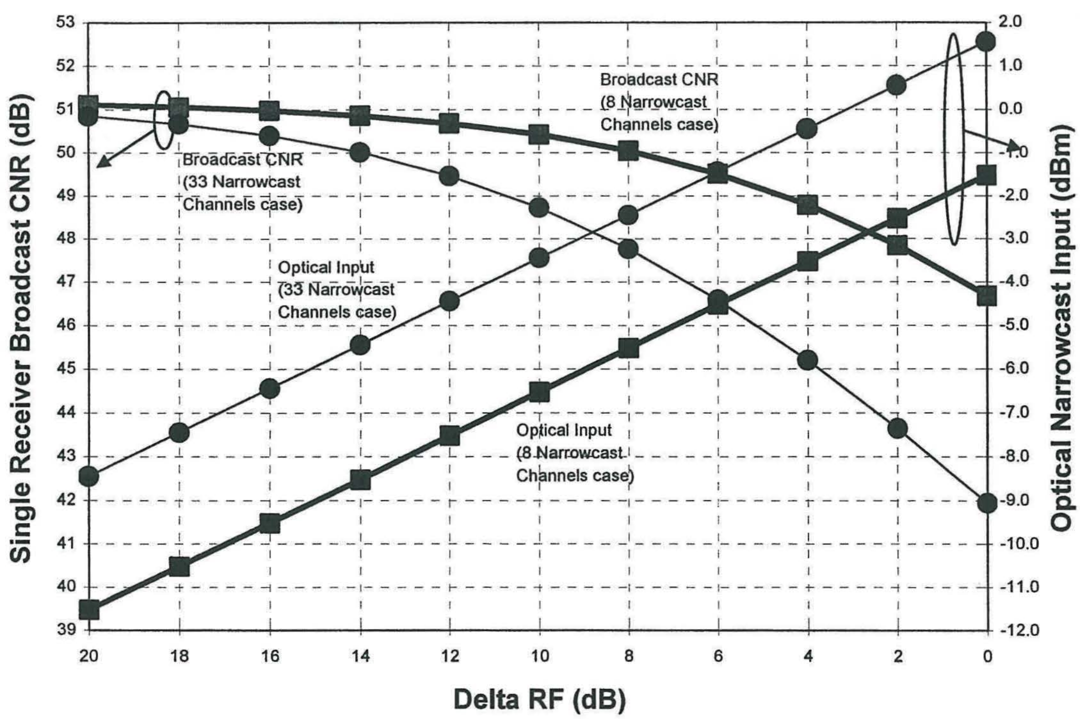


Figure 5. CNR penalty in the single receiver system. Notice that the CNR penalty increases as the narrowcast optical input increases, Analog input = 3 dBm (optical), total fiber link =65 km (45 km long haul+15 km distribution)

noise for the broadcast along with similar terms for narrowcast. These are then combined with thermal noise of the receiver

Such noise sources can be quantified quite accurately and the resulting CNR can be evaluated very accurately. Usually, the operator specifies a certain RF Level differential between the broadcast and the narrowcast. Since the number of channels for transmission and the RF drive level for a particular system is fixed, one ends up having to maintain a constant ratio of optical power differential to obtain the required RF delta

Figure 5 is a graph of performance in a single receiver system. It is seen quite clearly that in the above architecture higher optical input from the narrowcast input results in a large CNR penalty for the broadcast. From the graph, one can see that precise optical levels for both broadcast and narrowcast must be maintained not only to generate the specified CNR and SNR, but also to maintain the specified RF Delta. This is a non-trivial problem especially given the varying gain tilts of the EDFAs. In general, optical attenuators must be used for each node to satisfy the various power requirements. For higher number of channels (for example 32 channels from 550 to 750 MHz), the CNR penalty is quite severe. In the graph presented in figure 5, it is seen that even with a +3 dBm input to the receiver from the broadcast receiver, the final broadcast CNR is only 49.5 dB. To maintain a 6 dB RF Delta, the narrowcast optical input must be –1.5 dBm. Such high optical levels for narrowcast and broadcast require receivers that can handle high optical power. The end result is also an inefficient use of system optical power budget.

From Figure 1, it is seen that fiber induced CSO and CTB distortions are prevalent in systems having even modest fiber link (50 km). These distortions which are generated by narrowcast QAM signals are spread over the lower RF band from 50 to 200 MHz and could really limit the CNR of channels located at those places even further.

For all of the above reasons, single receiver systems can generally be used effectively when the number of narrowcast channels are quite limited (for example 8) and when the required RF level differential is quite high (for example 10 dB).

Return System: A typical return system is shown in Figure 4. Since the loading on the return transmitter is limited (in the US from 5 to 42 MHz), many return links from the node to the hub may be "stacked" together over one DWDM transmitter for purposes of transportation back to the headend.

Figure 4 shows a typical case where 4 such return streams are multiplexed. DWDM transmitters used for return systems could potentially be identical to the forward transmitter if a proper block conversion scheme were chosen.

If the return transmission scheme was QPSK (as it is the most prevalent), both the transmitter capacity and fiber capacity could be increased significantly. For higher modulation formats such as QAM16 and OFDM signals over long links, return system design is complicated and challenging and sometimes could be the limiting factor.

Alternatively an A/D scheme may be effected within the node to digitally sample the return band and transmit the bits over available node return transmitter. Multiple links of this type could be aggregated at the hub for transportation to the headend.

Dual Receiver Architecture

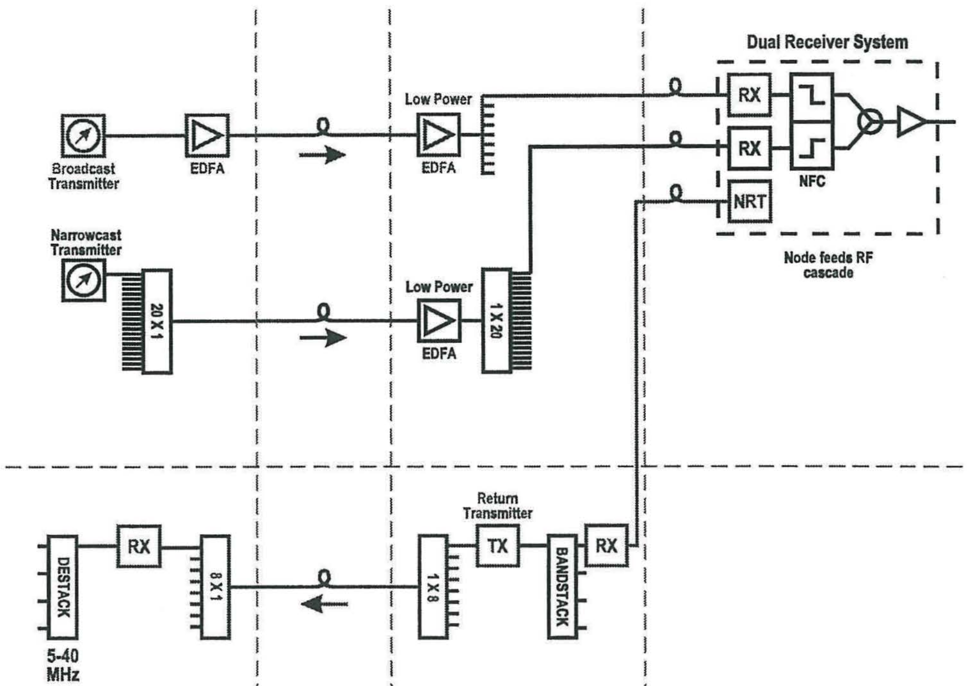


Figure 6. Dual receiver system, and NFC (network forward combiner) within the node combines broadcast and narrowcast signals

In the dual receiver system, the transmission is achieved by using two receivers in the node (one for broadcast and one for narrowcast) and a network forward combiner (NFC). Here, the RF level differential can be de-coupled from optical levels, which is a significant simplification.

Since appropriate filtering and amplification/attenuation can be done *after* the receivers, broadcast CNR and narrowcast SNR are reasonably independent of each other. Lower optical input power may be applied to the individual receivers and there is no conglomeration of noise normally seen in single receiver systems.

The component NFC is similar to a forward diplex filter and must be designed to have a crossover at the frequency band of interest such that minimum impact occurs over the entire broadcast and narrowcast band. Also, because of the NFC, CSO distortions shown in Figure 1 do not affect CNR of the broadcast link in this architecture thus enabling longer fiber links.

Figure 7 is a graph, showing performance for a 32-channel dual receiver system the same link as the single receiver case. It is seen that a 6 dB RF level differential is maintained with 0 dBm into the broadcast receiver and –10 dBm into a narrowcast receiver, where the CNR and SNR are similar to those attained by the single receiver solution. A quick comparison illustrates that there is a 3-dB reduction of optical budget for the broadcast an 8.5-dB reduction in the narrowcast power budget. This is a very persuasive argument for the dual receiver concept since it results in EDFA cost reductions.

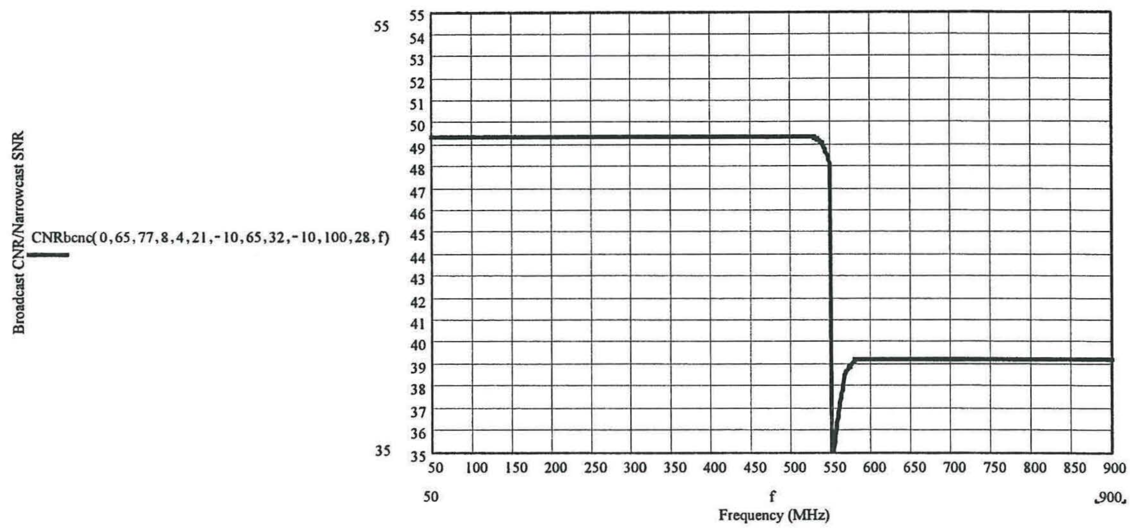


Figure 7. Graph illustrating the CNR in a dual receiver system. The shape at the crossover frequency is due to the shape of the NFC

The main limitation of this dual receiver concept is that the fiber counts and node configuration must change from its present structure. Appropriate space in the node platform must also be accommodated. In many cases, the cost advantage of lower system power budget is offset by the added cost of the additional node receiver and the NFC.

Digital convergence over the next several years is another issue that must be carefully evaluated. In this situation, the amount of broadcast analog frequency content may steadily decrease until it becomes necessary to change the NFC within the node. If the serving area were very large, it would be inconvenient to change the NFC within the node several years down the line.

The dual receiver concept conserves system budget and enables easier implementation of the DWDM architecture, particularly when the number of narrowcast channels are large (for example 32 QAM Channels) and the required RF level differential is small (6 dB). In these cases, in addition to ease of implementation, significant cost reductions are possible.

Ultimately decisions like digital convergence, size of the serving area, fiber counts, temperature performance of the node and space constraints along with performance considerations would determine when the dual receiver concept is applied.

Dual Hop Architecture

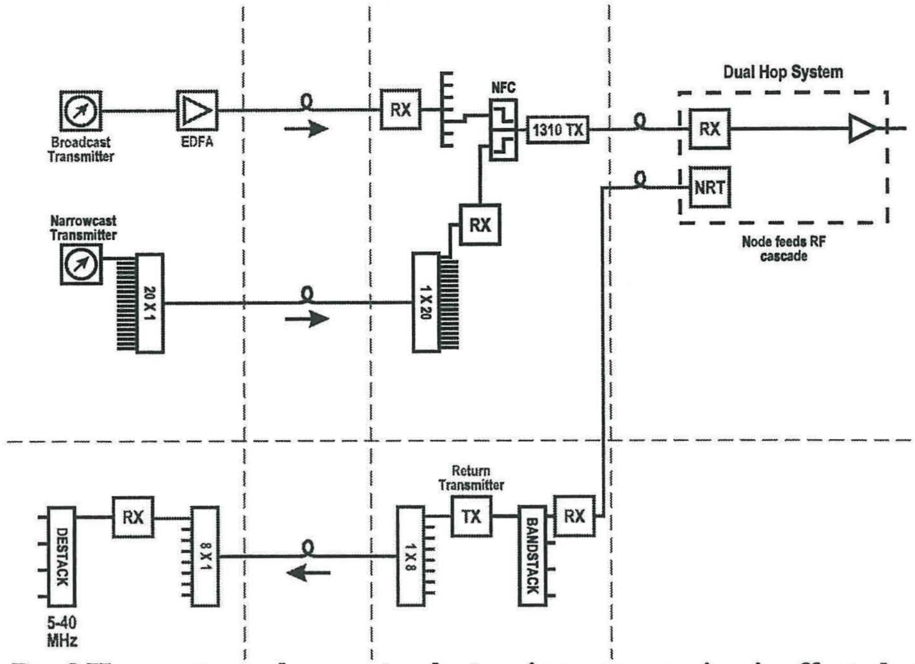


Figure 8. Dual Hop system where opto-electronic re-conversion is effected at the hub.

Here, a 1550 nm broadcast and 1550 nm narrowcast are propagated over fiber as normally envisioned. However, at the Hub, the receivers convert both the broadcast and narrowcast to RF and appropriate broadcast and narrowcast levels are connected to one single 1310 nm transmitter. These are then sent by conventional means over optical fiber to the node.

Analyzing a dual hop system for broadcast transmission is similar to any other dual hop operation with the exception that a *clip margin* must now be applied to the 1310 nm distribution transmitter. This is to protect digital QAM transmission over the same laser. Narrowcast transmission may be analyzed in a similar fashion.

Figure 9 shows the end of line results for a dual hop system. The thick lines show the EOL as a function of the distribution broadcast link CNR. However, a 1 to 3-dB *clip margin* may need to be applied to the 1310 nm transmitter resulting in lower EOL specification than that given in the graph. Note however that the distribution CNR may be maintained even with the clip margin for example by increasing the optical level to the receiver.

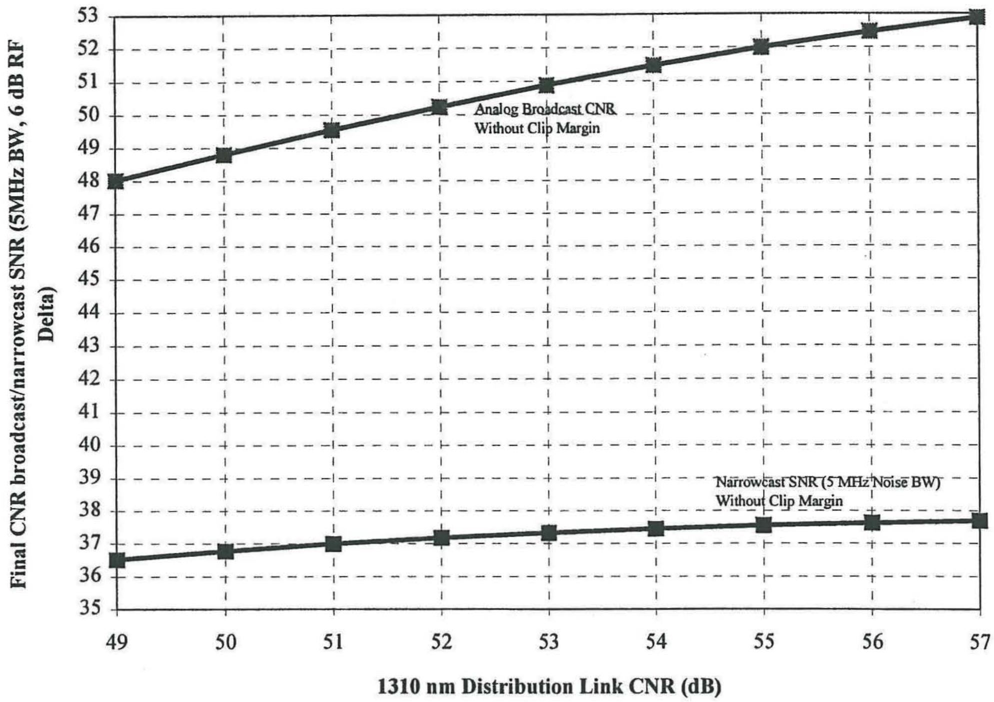


Figure 9. Final dual hop broadcast transmission depends on the CNR of the long haul link and the distribution link CNR and the Clip margin The broadcast long haul CNR is assumed to be 55 dB, and the narrowcast SNR in a 5 MHz bandwidth is 38 dB. The graph shows a calculated final CNR and SNR value as a function of the CNR of the distribution link CNR, without *Clip Margin.*

The dual hop operation is very resilient against various digital convergence paradigms. If it were necessary to change the digital and analog balance, it would all be done at the hub location as opposed to the node in the dual receiver case. True local ad insertion not normally possible in the single and dual receiver models are possible in this scheme. The dual hop operation is equally applicable for cases with small and large number of channels and for small and large RF level differentials.

Cost wise, the system has even fewer EDFAs than both the single and dual receiver systems. However, one extra receiver and one extra 1310-nm transmitter, would potentially offset the cost advantages. Fiber counts from the hub to the node are conserved while additional space and equipment is needed at the hub.

Among operators with an established 1310 nm base, this concept has found a lot of favor. The advantage of the dual hop system is that the node is unaffected, thus making laying of the cable plant a completely transparent operation. Since 1310 nm is a mature technology currently experiencing cost reduction exercises the distribution plant for DWDM requirements may be built with little extra investment.

CONCLUSIONS

Operators are concerned about capacity constraints and of over building. Over built systems represent an opportunity cost that presumably could have been spent on other lucrative options. Even so, DWDM represents the best way to invest in a system, since it enables the subscribers to use advanced services while deferring the costs of expensive hub builds to the future when such services actually take-off and are able to produce money on their own. In other words, DWDM potentially represents a "faster to revenue" way while still maintaining all options for the future.

All three architectures presented here have their own unique applications in the DWDM universe. Another significant advantage of DWDM technologies is the "transparent" hub, and the associated cost savings that it brings with respect to the real estate and manpower. In all cases, the capacity of the long haul fiber is considerably enhanced.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the long cooperation of Dr. M. Schemmann in the physical definition and test of concepts presented in this paper. Special thanks to Christina Lumbreras who provided a lot of help in the testing of devices and architecture definition.

REFERENCES

a) Grafting WDM on to existing cable systems-Enhancing cable TV using WDM technology, Venk Mutalik, CED February 1997

b) Non-linear Fiber Optics, G.P Agrawal, Second edition, Academic Press

c) Fiber-optic Communication Systems, G.P Agrawal, Second edition, John Wiley and Sons Inc.

d) Broadband Return Systems for Hybrid Fiber/Coax Cable TV Networks, Donald Raskin and Dean Stoneback, Prentice Hall PTR, 1998

CONTACT INFORMATION

Venkatesh G. Mutalik
Staff Engineer, Advanced Fiber-optics
Philips Broadband Networks
100 Fairgrounds Drive
Manlius, NY 13104
P (315) 682-9105
F (315) 682-2279
venkatesh.mutalik@pbn-us.be.philips.com

E-COMMERCE OVER CABLE: PROVIDING SECURITY FOR INTERACTIVE APPLICATIONS

Tony Wasilewski
Scientific-Atlanta, Inc.

Abstract

Digital CATV networks now being deployed offer the promise of a rich application environment that goes beyond broadcast and I/PPV.

E-commerce applications that extend the revenue-generating possibilities of the network also bring new issues and challenges. Many of these challenges are security-related and create requirements for authentication, encryption, simplified key management and message integrity.

Public key cryptography can help meet these new requirements as well as provide the basis for the "many-to-many" security relationship that is necessary to support scalable, spontaneous E-commerce applications.

A contemporary approach to CATV security will include support of a Public Key Infrastructure (PKI).

BACKGROUND AND NETWORK OVERVIEW

Security in broadband networks, including CATV HFC networks, can be enhanced by means of a technology called public key cryptography. Cryptographic methods have been used for many years to secure networks of various types, but numerous advantages are afforded by using a public key method in conjunction with more traditional secret key approaches. In particular, contemporary security needs include support of electronic commerce. Public key cryptography is particularly useful in supporting and improving this particular application. Furthermore, the body of standards for public key technologies is growing, and the development of the Worldwide Web has brought about some commonality in implementations.

There is already a broad base of commerce activity occurring electronically. Home shopping networks, such as QVC and the Home Shopping Network, where goods are purchased via television advertising and telephone ordering, is a multi-billion dollar per year industry. On-line Internet shopping in 1998, totaled $8 Billion in consumer transactions according to Forrester Research. Business-to-business transactions on-line are considerably higher. This indicates a strong interest by consumers and businesses in shopping/procuring through networks.

While the convenience of "mouse and network" shopping continues to lure an increasing number of users, security is still a major concern. According to *Internet World*, in a survey of 1000 Americans carried out by ISP NetZero, 53% cited "privacy and security" as the top concern about on-line shopping.

As cable modems and digital set-tops become increasingly deployed, routed IP addressing schemes are becoming the norm on broadband CATV networks. In the past, to communicate with home terminals, connectionless "single-wire" addressing was typically used. In this approach, equipment in the headend merely appends the terminal's address to a message and puts the message on the (usually single) downstream carrier. Then all terminals in the system must filter through all the messages in order to find the one(s) addressed to it. In modern broadband networks, the communications model has become more complex.

New network models have placed new security demands on digital cable networks This is being driven by new connection-oriented and IP subnetted configurations as the schematic in Figure 1 illustrates.

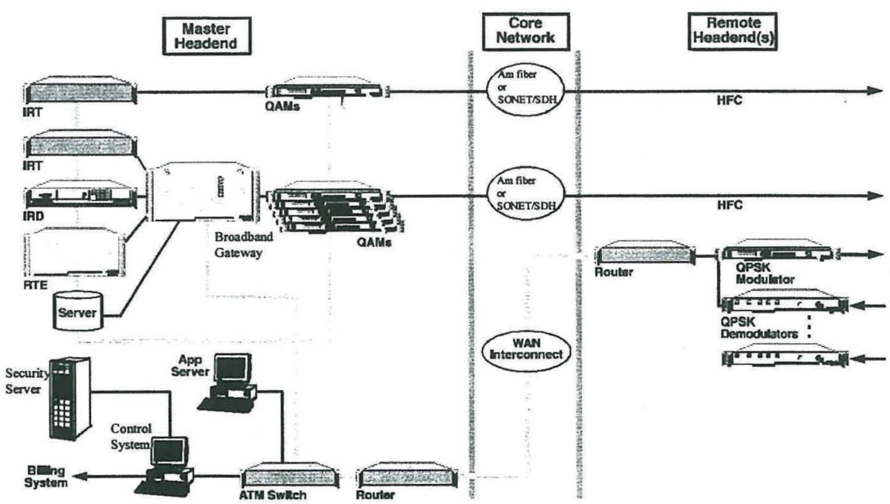


Figure 1 - Digital Broadband Network

New services such as E-mail, interactive shopping and video-on-demand (VOD) are utilizing architectures in which an ATM (asynchronous transfer mode) or SONET (Synchronous Optical Network) may be employed as the wide-area transport supporting servers of all types that transmit digital files as well as entertainment content. Such applications can be hosted in regional centers which send information through the network using various types of broadband gateways and local and wide-area routers. That information is further routed through smaller nodes which handle both downstream and upstream information to/from televisions and/or personal computers at the consumer side of the network.

These trends and the services they support contribute to changes in basic network security relationships. In broadcast-only services (whether digital or analog), such as satellite or traditional cable, there exists a one-to-many security relationship in which the operator of the network establishes an account with each subscriber in the network and provides authorization for each. In most cases, this is accomplished using secret-key-based security systems.

With the advent of two-way interactivity, this relationship is changing. Subscribers want to access a scalable and dynamic electronic-commerce environment. To support this type of interaction, the security relationship must become many-to-many since each subscriber may want to make transactions with a different set of merchants or service providers. The secret key-only approaches that have been traditionally deployed in many conditional access systems do not support this new relationship mode as well as public key methods.

Further inspection of Figure 1 illustrates how the network offers new possibilities. In the center, a large-scale transport system such as AM fiber, SONET and/or ATM is linked with various gateways that can bring in digital satellite, off-air, other locally-encoded signals and server-based digital content. Internet services can, of course, be delivered in such a network as well.

These services can be multiplexed for transmission by broadband gateways and then processed for further transmission through modulators in the access network. In the figure, quadrature amplitude modulation (QAM) is shown, and the signal runs over the HFC network to the subscriber premise.

Finally, at the subscriber premise, a receiving decoder (set-top) using MPEG-2 and other standards is employed. These set-tops can support reverse path communications with many servers or service providers. This multitude of new connection possibilities makes support of a many-to-many security relationship desirable.

EXAMPLES OF EMERGING E-COMMERCE APPLICATIONS

Examples of possible e-commerce services that can be offered over CATV networks is shown in Figure 2.

- On-line ticket sales
- Home catalog shopping
- E-auctions
- E-gambling
- Local fulfillment (pizza, flowers, stamps)
- Travel services
- Affinity (loyalty programs, cross-selling)

Figure 2 - Possible CATV E-commerce Services

The popularity of E-commerce over the Internet has been a proving ground for the introduction of similar applications into the home. However, with the advent of sophisticated HFC networks that support IP traffic and home terminals that can run the operating system, middleware, applications and security services to support them, the cable industry is perhaps better positioned than ISPs which use the PSTN to deliver these services to the consumer.

WHY LEGACY SECURITY APPROACHES ARE NOT SUFFICIENT

It is important, when preparing to launch new E-commerce applications, to examine legacy security methods to understand to what extent they do or do not provide adequate support. Three examples are used here to illuminate some of the issues. The first case involves communication over the Internet; the second looks at key management in CATV networks; and the third reviews the transmission of sensitive consumer information such as credit card numbers over a network.

Case 1): The basic Internet protocols exhibit many security weaknesses. Security was not foremost in the minds of the original designers of transmission control protocol/Internet protocol (TCP/IP). First, there is an inherent lack of privacy both in IP itself and in other layers upon which it typically depends. For instance, lower-layer protocols such as Ethernet are broadcast and session-oriented protocols such as File Transfer Protocol (FTP) provide no protection of content while in transit. Second, authentication is lacking. In general with IP, the user sends packets labeled with a source and a destination ID. The recipient has to trust that the packets really come from the labeled sender because there is no means to authenticate the sender of the message. In addition, the authentication and integrity of the data itself is in question because other than simple checksums for basic error detection, there are no means available to safeguard against malicious tampering.

Case 2): Key management in cable TV networks is typically achieved in an unauthenticated manner. Many cable TV systems have security systems that employ secret-key or "symmetric" encryption. In these systems, when messaging home terminals, the same key is required at both at the encryption and decryption sites. Thus, a network operator must maintain a database of all secret keys for each set-top. These databases can be vulnerable to attacks that can compromise system security through unauthorized access to the keys. Even if the operational databases are reasonably well-protected, keylists can be stolen during shipping. Using this key information, clones of legal terminals can be made that have the same keys in them. These clone terminals will then respond to messages intended for the legal device, such as the authorization of services, etc. Thus, protection (secrecy) of these home terminal key databases is of utmost importance.

Employing such a secret terminal key database also inhibits multi-site control of the security function, because distributing the secret keys to many locations makes them potentially more vulnerable. A distributed control scenario simply has more places to physically protect since copies of the key database are replicated. This may also have negative implications for "retail" scenarios.

Case 3): As noted above, consumers' concerns about the protection of vital personal or financial information is the major concern of current on-line shoppers. In order to complete an on-line purchase with a credit card, the buyer's credit card number must somehow be transmitted to the on-line merchant. This could be done by telephoning the merchant in advance or during the transaction to send the number. However, by requiring such actions, some of the main advantages of on-line shopping are lost. In particular, the spontaneity of impulse buying is disrupted and scalability is limited since it would be impossible for the buyer to be "introduced to" all possible merchants, in advance.

HOW ADVANCED CRYPTOGRAPHY CAN ENABLE E-COMMERCE

Public key cryptography is an approach that can provide many advantages in support of emerging applications on HFC networks. Figure 3 shows the fundamentals of public key cryptography and contrasts these with secret key cryptography.

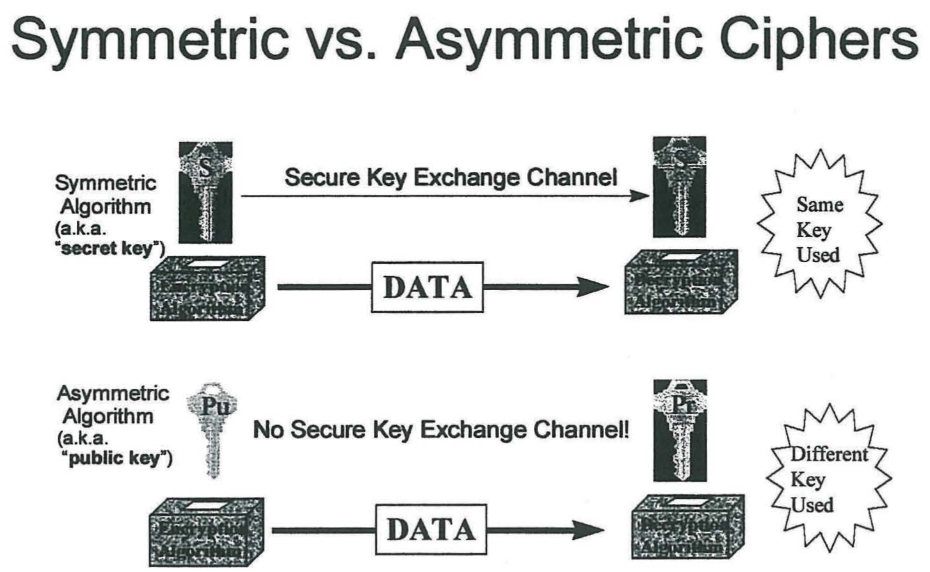


Figure 3 - Public Key vs. Secret Key Cryptography

Figure 3 shows that, in a secret key cipher such as the data encryption standard (DES) or DVB Superscrambling, the same key must be used by the sender and the receiver to encrypt and decrypt the message. The process begins with plain text (i.e., an unencrypted message). The key is then applied and the plain text is run through an encryption algorithm. This produces cipher text which can be sent with confidentiality over a network. At the receiving end, the same key must be applied by the receiver to recover the plain text again. The implication is that a secure channel is needed to get this key from the sender to the receiver. If the key were transmitted openly, it could be recovered by unauthorized entities through simple network "snooping" and then used to read messages encrypted with it.

Public/private keys -- which are called asymmetric because the encryption and decryption keys are different -- are mathematically related to each other. What one key encrypts, the other matched key can decrypt. Anyone with the private key can decipher any message encrypted with the corresponding public key and vice versa. The RSA algorithm is an example of this method. Usually, the private key is securely stored so that it may not be easily discovered or altered. This allows messages to remain private to the holder of the private key and also supports another function called digital signature, which will be discussed later.

A major advantage of public key methods is that the public component of the key pair can be published. It can be known by all parties and does not have to be kept secret. It can be put in a global directory where is can be easily accessed on an as-needed basis. Knowledge of a public key can not be used to derive the value of its matched private key.

The public key method provides much better support of a multiple service provider or merchant scenario in the sense that no pre-established relationship between each service provider/merchant and each user is necessary. Security issues associated with secret key databases are eliminated because the public keys can be shared and published openly. Thus, they cannot be stolen because they are already well known.

Use of Public keys is also the best way to provide digital signature services and authentication of users. Digital signatures and authentication allow spontaneous connections between users and service providers or between users in a peer-to-peer mode even though the parties may not know of each other in advance. If they are registered in a public-key database, they can exchange secure messages and be sure that the message author is truly the entity it purports to be.

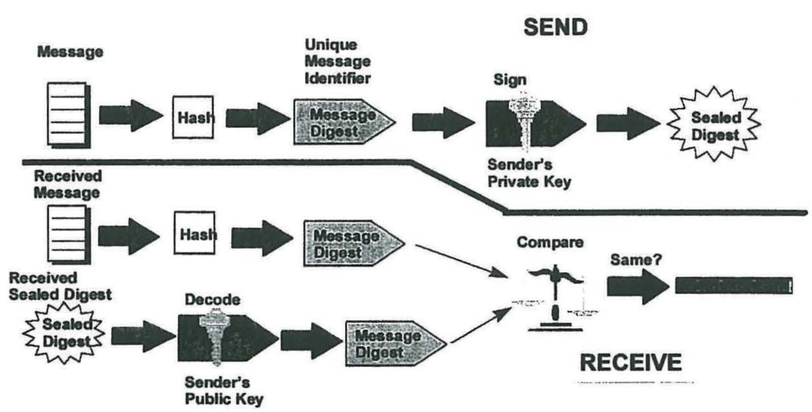


Figure 4 - Fundamentals of Digital Signature

Figure 4 illustrates the authentication process. A message -- it could be a long message, a file, or maybe just an e-mail -- is run through a one-way hash function, This function produces as output a much smaller token (typically 128 or 160 bits) which is called a message digest. This token is a unique identifier of the original message. However, merely knowing the digest does not allow discovery of the message. In addition, because of the design of the hash function, an adversary cannot even formulate an alternative message which produces the same message digest. This makes an attempt to provide a false message extremely difficult.

The message digest is validated and bound to the sender by digitally "signing" it. This is accomplished by encrypting it with the sender's private key. A message encrypted with the private key can be unraveled with the corresponding public half of the key. Because the sender's private key is used -- the one that is not published and that only the sender knows – this provides a way for the sender to put his or her unique digital signature on the message. Each user's private key transforms the message digest in a unique way to produce a "sealed digest". This sealed digest is appended to and sent along with the message.

To verify a digital signature, the receiving terminal simply goes through the inverse motions to process it: receive the message, calculate the message digest in the same way as the sender, and use the public key of the sender to decode the sealed digest. The recipient then compares the transmitted digest with the locally calculated one based on the received message content. If these two quantities are the same, then the recipient knows two things: 1) the identity of the sender is that of the owner of the public key used to decode the "sealed digest" and 2) the original message has not been altered. Note that the message itself need not be encrypted in order for the digital signature to work.

Secure hash functions require very special design. Some examples are MD5 (RFC 1321) or SHA-1 (FIPS PUB 181-1).

While, public key methods offer distinct advantages, it should be noted that there are licensing and intellectual property issues that must also be considered. In contrast, many secret key ciphers are in the public domain and may be used royalty-free. There is also a requirement to provide digital certificates, a certification authority and a public key infrastructure to effectively us public key methods.

CERTIFICATES, CERTIFICATE AUTHORITIES AND THE PUBLIC KEY INFRASTRUCTURE

Data integrity and authentication of correspondents are great advantages offered by public-key cryptography. However, to trust digital signatures, users must have reliable means of obtaining public keys to use in signature verifications. This is done using digital certificates.

Digital certificates are tamper-proof bindings of a public key and the owner of that key. They usually include a "distinguished name" related to the owner, an expiration date, and other data. The "distinguished name" can be a set-top address rather than a subscriber name. Billing information can then be used to link a subscriber with the terminal.

Trust is established by a Certificate Authority. The Certificate Authority is the entity that applies its digital signature to each certificate. Since there are relatively few Certificate Authorities, their public keys can be trusted by publication in open venues, via software distribution (such as in Web browsers) or over the Internet. Then, users can determine the validity of any given certificate by checking that it bears the signature of a trusted Certificate Authority. Clearly, the private key belonging to a Certificate Authority must be guarded carefully.

All of the functions needed to effectively use a public key-based system are collectively known as a "Public Key Infrastructure" or PKI. The principle functions of a PKI are:

- Key Generation
- Storage
- Key transfer (shipping)
- On-line public key repository
- Key renewal (changing)
- Data Recovery (lost keys)
- Retiring keys

Figure 5 shows a structure of a PKI suitable for use in contemporary CATV networks.

A CATV PKI includes not only set-top keys but also keys that the network operator can use to exert control over the set-tops and to enforce its digital signature. In fact, an MSO can have both a corporate-level signature and a site-specific signature. The corporate-level signature can ensure complete control over set-tops in all its systems and to prevent "migration" of set-tops to systems of other operators. The site-specific signature can be used to differentiate control in one headend or system from another.

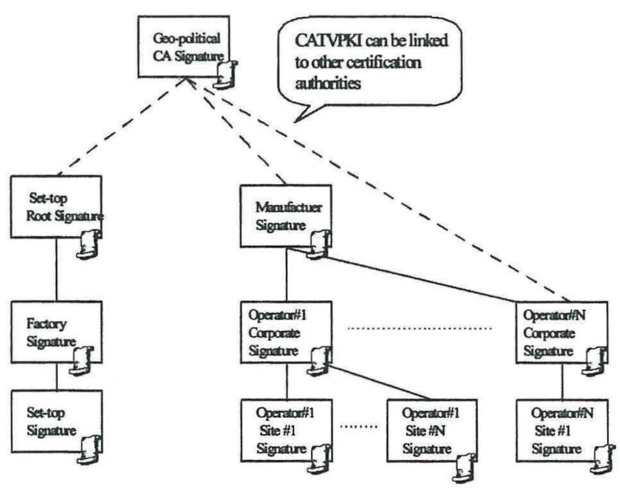


Figure 5 - CATV Public Key Infrastructure (PKI)

The set-top is initialized with corporate-level and site-specific "key certificates". It then can check signatures of messages and accept those from the network operator, but reject those from other sources. This is a significant tool in preventing authorization spoofing by pirates but has the further advantage that the operator can also allow or deny access to the set-top by other service providers or merchants. This can be accomplished when the operator validates the digital certificate of a merchant. The set-top can then access and trust the merchant, but otherwise will not attempt transactions with that merchant.The PKI can also be used to link the network operator's signatures and those of the set-tops with a larger universe of E-commerce. Indeed, through the signature mechanism "chains of trust" can be established. For example, a "geopolitical" certificate authority such as Verisign, GTE Cybertrust or governmental agencies certify the keys of MSOs, set-tops and merchants so that by checking a short sequence of signatures, spontaneous E-commerce relationships can be formed quickly. These relationships can also be dynamic.

Flexible PKIs allow changing the "chains of trust". Thus, even though the manufacturer may participate in the signature PKI (as shown in Figure 5), it is perfectly feasible to eliminate this connection. In this case, however, the manufacturer can no longer provide services such as re-keying set-tops when needed, such as in the sale of a headend or system to another operator.

USEFULNESS OF A HYBRID APPROACH

As noted, there are many advantages derived from public-key cryptographic methods, but traditional secret key approaches still have important application in CATV networks. High-speed data, such as Internet data or compressed digital video, still benefit from secret key algorithms such as Harmony DES or DVB Superscrambling because they are much faster. DES, which is a symmetric key algorithm, will run perhaps 100 to 1,000 times faster in the same implementation (hardware or software) than a public key algorithm will. The optimum approach, then, is to use both public-key and secret-key technologies in the appropriate combination. Such a hybrid approach, benefits from the authentication and digital signature capabilities of public key methods in areas such as key exchange and verification of software downloads. At the same time, the speed of secret key algorithms can be exploited to provide confidentiality of large quantities of high-speed data.

LICENSING AND INTELLECTUAL PROPERTY ISSUES

Public key cryptography has several patents associated with its use. Having been invented in 1976, the patents typically have expiration dates ranging from 1997-2000. Licensing of these patents is required for commercial use of public-key technologies. Check with your conditional access vendor, since these licenses may already be included in the product.

OTHER ISSUES TO CONSIDER

As with other technologies, public key cryptography must be implemented properly to deliver its full benefits and selecting standards-based implementations promotes interoperability and economies of scale. Thus, two areas should receive particular attention when the use of public key methods is contemplated:

1) use/existence of relevant standards
2) secure packaging

STANDARDS

Relevant standards in this technology area include a few that are stable, but also many that are still under development. The Internet Engineering Task Force (IETF) is actively engaged in this quest and has defined a security architecture for the IP layer (RFC 2401). Called IPSEC, more information can be found at: http://www.ietf.org/html.charters/ipsec-charter.html. Some of the important methods they are considering at the network layer include: authentication headers (AH), encapsulated security payload (ESP) and Internet Key Exchange (IKE). At the session layer, the well-known secure sockets layer (SSL/TLS) has become widely used in Web applications. At the application level secure multipurpose Internet mail extensions (S/MIME) has gained some acceptance.

The ITU-T X.509 series recommendation already includes a standard for public key certificates. DAVIC, the Digital Audio Visual Council, has published Part 10 of DAVIC 1.2 which includes general security interfaces and tools for multimedia applications. MasterCard and VISA have been leaders in specifying e-commerce secure electronic transactions (SET - http://www.setco.org/).

An "informal" but widely referenced standard is the public key cryptography standard (PKCS) published by RSA Laboratories http://www.rsa.com/rsalabs/pubs/PKCS/index.html. Developed in conjunction with representatives of many computer and communications firms, it gives excellent recommendations on how to use a wide array of public key techniques and includes such important topics as message padding. A more formal effort to establish procedures governing the use of public key cryptography is the Institute of Electrical and Electronic Engineers (IEEE) P1363 which passed its first ballot on April 2, 1999. (http://grouper.ieee.org/groups/1363/)

SECURE PACKAGING

To promote interoperability and retail availability of good security for networks, physical packaging becomes important. For physical security devices a popular standard is ISO 7816. This is the most universal reference for smart-card technology. It now comprises a series of six parts, covering mechanical, electrical, and protocol interfaces of these hardware tokens.

The Personal Computer Memory Card International Association (PCMCIA) package is also a choice in this area. Indeed, the DVB Common Interface, NRSS-B and OpenCable POD all use this basic form factor. Also known as PC-Card, this standard provides a uniform and convenient physical form factor and a flexible 68-pin interface with considerable provisions for software-based configuration of the package.

Certificate Authority equipment must also have highly tamper-resistant packaging.

CONCLUSION

With good protocols and an appropriate combination of public key and secret key approaches, CATV security systems can effectively and safely enable many classes of broadband networks to securely deploy digital services and robustly support E-commerce applications.

Tony Wasilewski is Chief Scientist, Subscriber Networks Sector at Scientific-Atlanta, Inc. His contact information is:

5030 Sugarloaf Parkway
Lawrenceville, GA 30042
Phone: 770-236-5004
Fax: 770-236-3080
E-mail: tony.wasilewski@sciatl.com

EMERGING TRENDS IN HOME NETWORKING

Dan Sweeney
Intel Home Networking Organization

THE HOME NETWORK

There is a growing phenomenon in the home computing environment that Intel believes will dramatically alter the home of the future. It is the multiple-PC home. Based on 1998 Dataquest* research, Intel estimates that the number of multiple-PC homes by the end of 1998 will be approximately 18 million. This number represents about 37% of the 49 million U.S. homes with PCs (GRAPH 1). Dataquest estimates that in the year 2003, 28 million households will have at least two PCs. And, Forrester Research recently predicts that home networking will generate annual sales in excess of $1 billion by 2002 (GRAPH 2).

Intel believes that home networking will dramatically impact the way technology is used in the home. This paper explores the multiple-PC home trend, Intel's vision of the optimal features of a home networking solution, and the implications of home networking for today and the not-so-distant future.

A PRECIDENT OF MORE THAN ONE...

There was a time when the idea of one television set per home was considered a luxury. Today, according to a December, 1997 Odyssey* study, approximately 76% of all U.S. households have two or more televisions. It can be argued that the PC is following the same pattern. Consider three factors that contributed significantly to multiple TV set ownership within the home:

- Purchase of a newer, bigger, better television
- Purchase of an additional television to reduce conflicts in the house over TV use
- Purchase of an additional television for use in a second or third room (kitchen, bedroom, etc.)

Replace the word "television" with "PC" in each of the above statements and they fit the multiple-PC phenomenon. After all, rapid advancements in PC technology result in "newer, bigger and better" every holiday season. PCs undoubtedly take the potential for conflict one step further than television.

Unlike TV viewing, which can be a multi-person experience, personal computer use is distinctively personal. In most cases it is impossible for users to share.

During Intel-sponsored focus group research, one respondent lamented that, because his teenage daughter has replaced the telephone with the Internet for long distance communication with her boyfriend, he had to purchase another computer for his use. Finally, it's not a leap to hypothesize that as PC use increases, more home users are likely to demand the convenience of accessing the computers' capabilities from the home office, the kitchen, the bedroom, or wherever they happen to be.

TODAY'S MULTIPLE PC HOME

What does today's multiple-PC home look like? Perhaps less "techy" than might be imagined. Notably, among the hundreds of two-PC household decision-makers that Intel interviewed, few consider themselves early adopters of technology. In fact, in focus group discussions, most indicated a modest level of PC knowledge. What they communicated was that they find the PC a useful, integral part of life - *for more than one member of the household.* Intel also learned the following about these consumers:

- 85% have two or more adult users in the home and 50% have at least one child user
- The *primary* users are typically adults between the ages of 25 and 54
- Multiple-PC homes don't appear to be a trend of only the highly affluent. Annual household income ranged widely from $20,000 to over $100,000, with the most prevalent income cluster being in the $20,000 - $70,000 range

The only *significant* difference between these households and one-PC households is the tendency to be connected to the Internet. According to research Intel conducted in May of 1998, 86% of multiple-PC homes have Internet access compared to 47% reported for single-PC homes in a January, 1998, Odyssey study. Notably, this access is used by nearly everyone in the home. Ninety-eight percent of adults and sixty-nine percent of children who use a PC in these households report being on the Internet at least once per month (GRAPH 3). More than half of this group reported that *someone* is on the Internet in their home at least 10 hours per week (GRAPH 4).

Another significant piece of information is that multiple-PC household members spend a lot of time on their computers.

The majority of primary users (typically adults) spend about 19 hours per week on the PC – mostly engaged in work applications and on-line access.

Secondary users (more likely to be adults than children) spend about nine hours per week split between work applications and entertainment. The tertiary user (typically a child) uses the PC about six hours per week for educational or entertainment purposes. In many households there are four, five, and even more users competing for PC time. Critically, most of this PC use occurs in the evening when everyone in the family is home. That can result in conflict for valuable resources such as printers and Internet access - which leads to the need for home networking.

GETTING THE MOST FROM YOUR PCs

If you buy a PC for everyone in the house, all PC-related conflicts are over, right? Well, maybe. In a number of Intel-sponsored focus group discussions across the U.S., one issue was apparent. Home consumers don't consider the PC a stand-alone device any more than businesses do. To get the most value from their computers, home users need to be able to access PC resources (like printers and Internet access) from any PC they are working on. For most home consumers Intel spoke with, multiple PCs in their homes only partially addresses their needs. To get full use of their PCs, these consumers have turned to a variety of methods.

A very small percentage of U.S. multiple-PC homes have actually installed a traditional office network (GRAPH 5). Other consumers have purchased additional printers and/or Internet access accounts and phone lines for their additional PCs. Those who haven't made such purchases (the majority) are grudgingly living with makeshift solutions, like running floppy disks from the non-printer-connected PC to the printer-connected PC, and simply waiting for the Internet-connected PC to be available before accessing their favorite website. While most felt they could live with such compromises, when presented with an easy home networking option, the majority indicated they would jump at the chance to take it.

The Idea of Home Networking

Intel asked consumers how appealing a product providing the following benefits would be:

- ***Printer sharing*** - enabling all PCs in the home to access the best printer in the house

- ***Simultaneous Internet access from a single phone line*** – enabling all PCs in the home to access the Internet at the same time through one Internet account
- ***File sharing*** - enabling all PCs in the home to access and share files
- ***Multi-player gaming*** – enabling all PCs in the home to participate in multi-player games

The response clearly indicates a strong consumer need. More than 50% of consumers surveyed indicated that they would find such a product highly appealing (rating appeal 8-10 on a scale of 1-10) (GRAPH 6). Nearly 70% indicated that such a product would be at least somewhat appealing (6-10 rating). Critically, among those indicating a lower appeal level, one of the biggest concerns was whether they would be able to install and operate it. (See next section.)

Those who found the concept of home networking appealing were then asked which benefits (of those presented) would be most important to them. Although printer sharing and Internet/Modem sharing were perceived as the greatest potential benefits of home networking, file sharing and multi-player gaming were also appealing. In-depth discussions with consumers have indicated that the total package of networking, not one single benefit, is what they find most interesting (GRAPH 7).

BUT FEW ARE NETWORKED TODAY

Nearly 18 million consumers have more than one PC in their home. At least half of these have indicated they would be very interested in a solution that would give them the major benefits of a home network. Why aren't multiple-PC consumers flocking to retailers for the currently existing network-in-a-box solutions? Two reasons:

First, consumers perceive a network as difficult to install and maintain. GRAPH 8 shows that 50% of multiple-PC household decision-makers use a networked PC at work. Among those who don't use a LAN (local area network), focus groups indicate that they are very familiar with someone who does. These people are not strangers to the idea of sharing data and PC resources on an office network. But, familiarity has bred contempt. Throughout Intel's investigation of the multiple-PC home, few discussion topics were as lively or emotional as those surrounding consumers' impressions of office networking:

> *"The network is always going down."*
> *"It takes a whole department to run the network."*
> *"Networking is a hassle."*

All of these are similar to the types of responses heard across the U.S. from consumers who have (or plan to soon have) at least two PCs in their homes - a fairly PC-literate group. Few argued the overall benefits of networking, but all perceived the cost of those benefits (in hassle and frustration) to be very high. Almost none felt compelled to pay that high a price to install a *traditional* network in their home.

The second reason multiple-PC owners aren't rushing to install home networks is very practical. Most don't want to drill holes in their walls to install network wiring. The option of stringing loose wire from room to room across the carpets or hardwood floors leaves something to be desired from an interior design standpoint.

ENTER HOMELINE-BASED HOME NETWORKING

Existing telephone wiring is an excellent medium for networking PCs within the home without adding new wires. The average multiple-PC household has 4-5 telephone jacks, and most are near existing PCs. Phonelines also provide a secure environment for data transmission (GRAPH 9, 10).

THE OPTIMAL HOME NETWORK

Following are key criteria based on input that Intel has gathered from thousands of consumers :

Home Networking Criteria

- ***Easy to use*** - Given the current perception of networking as "difficult," Intel believes that consumers will adopt a home networking solution when it is extraordinarily simple for the typical multiple-PC owner to install and run. Intel's consumer line of products will be extraordinarily simple from both an installation and operation standpoint.
- ***No new wires*** - A successful home networking solution won't require the installation of any new wiring. It will work within today's typical home with no structural, cabling or other modifications. Intel's home networking products will offer fast, reliable connectivity between home PCs through ordinary phone wiring already in homes.
- ***Accessible from anywhere in the home*** - As noted earlier in this paper, computing is increasingly likely to take place in multiple rooms of the home. An optimal networking solution will allow PCs to be easily added or moved wherever people most want them to be.

As noted earlier in this paper, most homes have multiple phone jacks located in rooms that most frequently contain PCs (kitchen, master bedroom, den, other bedroom).

- ***Low price*** - Priced comparably with popular peripherals such as printers, digital cameras and scanners, research indicates that a home networking solution would thrive. Intel's consumer home networking line will be affordable – offering consumers the opportunity to get more out of their PCs with less investment in extra peripherals or Internet access.
- ***Fast*** - A key necessity for a home network will be high bandwidth. The emerging solutions to bring data to the home (e.g. cable modems, UADSL, satellite) are promising multi-megabit rates. As consumers adopt these speedy Internet connections *outside* the home, fast (at least one megabit per second) solutions will be required *inside* the home to avoid bottlenecks. Intel's home networks will offer 1 mbps data speeds. This is fast enough for today's printing, file transfer and Internet sharing applications. After all, 1 mbps is over 18X faster than a 56 kbps (kilobit-per-second) modem.
- ***Works with popular networking protocols and high bandwidth Internet access (DSL and cable modems)*** - The optimal home network would be fully interoperable with existing applications and protocols such as TCP/IP, the Internet's standard protocol. It would also allow consumers to connect to DSL or a cable modem. Intel's consumer line of products will meet both of these requirements.
- ***Becomes an industry standard*** – Strong industry standards, like those being pursued by the Home Phoneline Networking Alliance, will foster third party development, consumer acceptance, and growth of the overall home net-working category. Standards will also assure consumers of interoperability between products. Intel's consumer product line will be full compliant with the standard being proposed by the Home Phoneline Networking Alliance.

Tomorrow's Connected Home

How does Intel project that the multiple-PC, networked home is going to change PC usage in households of the future? Short-term, it will make PC use more convenient for the multiple-PC household. Users could access files, printers, modems, and more from the network without regard to whether the accessed device was physically connected to the PC in front of them or a PC located somewhere else in the house. A home network would make it possible for all non-printer-connected-PCs in the home to share one high-quality color printer without the hassle of copying files to disk (an increasing challenge given that many of today's files require more than one floppy) and interrupting work on the printer-connected PC. It would also make it possible for one PC user to send e-mail over the Internet while another user accessed the Internet for stock quotes or homework research (with only one phone line, one modem, and one Internet connection).

Additionally, connected home PCs would provide the benefit of file sharing. A file from a laptop in the living room could easily be sent to the home office PC as a means of backup - or vice versa. For those "fun and games" PC users in the family, a home network would allow multiple-player gaming from all the connected PCs in the home. Imagine a whole house full of teenage Quake* players! But, the way a home network will change households in the very near-future pales in comparison to the way it will change homes in the not-so-distant future. The space-age home of sci-fi movies is on the brink of becoming reality. Consider the following scenario in which phoneline and radio frequency home networking devices work together:

It's a typical Monday morning. You wake up a half-hour before the rest of your household to make coffee and absorb the morning news. "Local Weather. Audio and Visual", you say as you toss an English Muffin into the toaster.

The eight by six inch communication pad affixed to your refrigerator immediately displays a local weather map with high and low temperatures for the day. At the same time the pad plays a 60-second audio clip from an Internet broadcast. Both actions are nearly instantaneous because the device is connected via radio frequency to every PC on the network in your home. Your den PC is constantly accessing the Internet and caching the four key areas you're interested in (weather, traffic, top news stories and stock prices), so the data is available anywhere, anytime in your home - on command. When your muffin and coffee are ready, you detach the communication pad from the fridge and use it to read stories from all of the major news organizations as you eat.

At 6:30AM, it's time to wake your daughter. "Wake-up and music. Katy's room", you say as you pour yourself another cup of coffee. Upstairs, the PC in your daughter's room receives the command over the home network. Immediately it's monitor displays a video image of you "Rise and Shine, Katy. Breakfast in twenty minutes".

From downstairs you hear the music that your daughter directed her PC to play for wake-up -- courtesy of the Spice Girls.*

A few seconds later you hear your daughter on the home network inter-com system (each room has an unobtrusive microphone/ speaker device connected to the network for easy communication between family members, or PCs and electronic devices using voice recognition soft-ware). "Don't forget that you need to drop me off at school today", she reminds.

Oops. You had forgotten. "Den PC, update work and family calendar. Print to kitchen." The commands are transmitted to the Den PC via the home network. Then the den PC accesses the calendar from your work-office PC twelve miles away, combines it with the calendar your family keeps at home and prints out a copy for you on the kitchen printer. Unfortunately, you scheduled an 8:00AM meeting for this morning, which will leave no time for messing with traffic.

"Monitor best route to work from Katy's school", you say. Then you begin the mad rush to get everyone in the house out the door on time.

At 7:30AM, after dropping off your daughter, you connect the communication pad that you brought with you into your cell phone and dial home. The call is answered by the answering machine connected to the home network. Voice-printing immediately identifies you to the system and your request to have the latest "best route to work" transmitted to you in your car is fulfilled. Your den PC has been monitoring traffic conditions via the Internet continuously since your request earlier that morning. The data is sent pad and you quickly agree with the PC recommendation that the best route avoids a major accident on Interstate 5.

At 7:55AM you arrive at the office for the meeting with your manager. Before you even have time to take your coat off, your boss' assistant calls. Meeting will be delayed a half-hour. Your manager is stuck in traffic. Oh well, you think as you pull the communication pad out of your bag. At least now you can finish catching up on the news.

This example illustrates the ease and convenience of home computing to operate security systems, access the Internet, communicate with family members, printers and more, all through a home network, connecting several PCs in different locations in your home. But the additional benefits that home networking will make possible are limitless. Given the aggressive drive to bring bigger and better technology benefits to the home from outside, there is little doubt that innovators will search for ways to bring technology benefits *throughout* the home once a widespread infrastructure exists. The optimal home network described in this paper would provide that infrastructure.

INTEL'S ROLE IN HOME NETWORKING

Intel Corporation is committed to increasing the value of PCs in the home. The research outlined in this paper indicates that a telephone-based home network is a logical next step to achieving this goal for multiple-PC consumers. Research also indicates that the number of these consumers is likely to grow dramatically over the next few years. Our role is to meet the home networking needs of consumers through high quality, easy-to-use products, and to meet the needs of the PC and consumer electronics industries by providing the home networking silicon enabling them to provide networking functionality in a wide range of consumer products.

For more information regarding Intel's consumer product line visit http://www.intel.com/home/network

For information regarding Intel's 21145 Phoneline/Ethernet LAN controller see document HNO102, "Intel 21145 Phoneline/Ethernet LAN Controller Product Brief: Single-chip solution for home and office networking" on the web at http://developer.intel.com/design/network/new21/21145.htm

**ABOUT INTEL-SPONSORED RESEARCH IN THIS PAPER

Following is an overview of the key Intel-sponsored research cited in this paper:

- May, 1998 survey of 300 multiple-PC (or single-PC with plans to buy a second within twelve months) homes. Survey conducted by Market Strategies, Inc.
- February, 1997 survey of 400 multiple- PC (or single-PC with plans to buy a second within twelve months) homes. Survey conducted by Market Strategies, Inc.
- 1997 Focus group research of multiple-PC owners.
- August, 1996 survey of 210 multiple-PC (or single-PC with plans to buy a second within eighteen months) homes. Survey conducted by Market Strategies, Inc.
- Numerous informal interviews of multiple home PC owners. This paper can be viewed online at http://www.intel.com/anypoint line and at http://www.intel.com/business/anypoint

Gain in Statistical Remultiplexing of MPEG-2 Transport Streams

Ji Zhang, Sangeeta Ramakrishnan & Fang Wu
V-BITS, Inc.

1.0 Abstract

In this paper, we study the effect of statistical remultiplexing on improving bandwidth utilization of multiplexed MPEG signals. The study is done through simulation results of 2 and 12 channels of compressed bit stream feeds into a real-time rate remultiplexing systems with optimal combination of transcoding and buffering strategies. Our results indicates that larger number of channels, when rate remultiplexed together, results in reduced need for transcoding and improved overall video quality preservation.

2.0 Overview

Statistical multiplexing of MPEG-2 transport streams has been widely recognized as the effective way to optimize the quality of compressed video signals under the constraint of a fixed total transmission bandwidth. The resulting improvement can be described in terms of the so called 'statistical gain', which can be measured in several ways. It can be measured either in terms of the improvement of subjective visual quality of the decoded video, or in terms of the objective average reduction in quantization value, or in terms of the reduction of the quantization value variability (statistical variance), etc.

Statistical remultiplexing, also called statistical rate remultiplexing, is a process which performs statistical multiplexing of signals already in compressed format. Statistical remultiplexing allows the operators to re-allocate the bit budgets among different video channels by performing transcoding on the pre-compressed signals. Statistical remultiplexing systems are used as a way to remove the dependency of the transmission channel on real-time encoding. When used in video-on-demand systems, digital cable headend systems and digital advertisement insertion systems, statistical remultiplexing can improve the overall system efficiency, resulting in better bandwidth utilization and reduced transmission cost.

In this paper, we study the overall bandwidth reduction due to statistical remultiplexing. We will study the bitstreams and their associated bit rate distributions to understand the effect of statistical remultiplexing on the quality of the video signals. We will also consider issues on the amount of transcoding as it related to the bit rate and the number of bit streams at the input of the statistical remultiplexer. The simulation results are presented to show that statistical remultiplexing gain can be greatly affected by several important factors including input bit rates, number of inputs, and the amount of transcoding allowed.

2.1 Quality Issues in MPEG-2 Signals

MPEG-2 signal is the output of an MPEG-2 encoder. It is a representation of the real-time video sequence, and audio signal. However, in general the MPEG encoding process is a lossy process, namely, although the MPEG decoding process can revert the encoding process to produce the original digital video and audio sequence, the decoded result differs from the original video image samples. Therefore, MPEG encoding and decoding processes are not regenerative processes, i.e, applications of such process results in irreversible difference between the original video and the decoded video. This can be mathematically abstracted as the introduction of noise into a information communication process, as shown in Figure 1. The equivalent noise component in the output decoded video signal can not be determined without the complete knowledge of the original signal. Therefore, the equivalent noise model says that it is impossible to remove the noise component from the decoded output video once it is introduced into the system. The noise component is introduced in the encoder by the quantization process. The decoder introduces at most minimal amount of error and is generally considered a noise-free operation. In addition to the motion estimation process, the quantization process largely determines the number of bits used to represent a coded picture, and as a consequence, determines the overall bit rate of the encoded video signal. In general, larger quantization steps produce larger quantization distortion, i.e., lower decoded video quality. Larger quantization steps also reduce the encoded bit rate. This relationship is often described by rate-distortion curves.

Many different metric can be used to measure the amount of distortion between the original (pre-compression) video images and the decoded video images. These include signal to noise ratio, peak signal to noise ratio, mean squared errors, absolute errors, etc. Less distortion generally

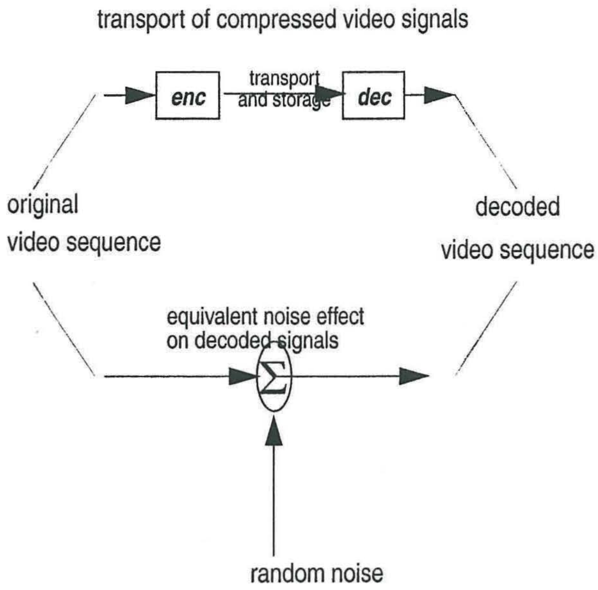


FIGURE 1. Quality loss due to encoding and decoding can be viewed as a noise addition process

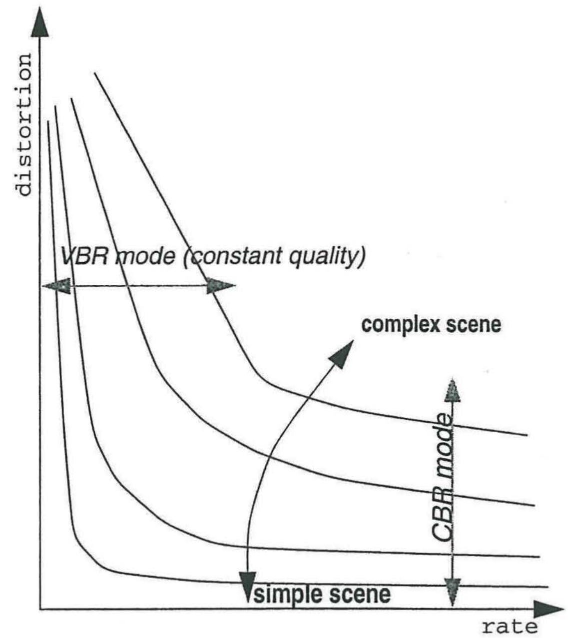


FIGURE 2. Rate-distortion curve of compressed video signals

results in more faithful reproduction of the original video source and results in better decoded video quality and vice versa.

MPEG video compression uses variable length coding and motion compensation. These techniques significantly reduces the number of bits required to represent video sequences. However, it also introduces the variable bit rate nature of the resulting bit stream. Specifically, there is no longer a fixed boundary between coded picture units. Variable bit rate (VBR) may result: higher bit rate for high motion or complex scenes and lower bit rate otherwise. This can also be reflected through the rate-distortion curves, as shown in Figure 2. As seen from the figure, VBR mode compression intends to create a bitstream that has constant quality, i.e., constant distortion, by varying the bit rate. VBR signals create significant difficulties for real-time transport because communication channels are typically constant bit rate (CBR) transport. MPEG video can be encoded at constant bit rate, which generally results in variable quality, as is shown in Figure 2. As mentioned earlier, there is a direct relationship between the bit rate and the quantization steps, but other factors, such as pre-filtering, frame/field mode decision, picture coding type structure, also significantly affect the quality of the video signals.

2.2 Transcoding of MPEG Video Signals

Transcoding of MPEG video signal (or signals based on similar techniques, such as H.26x video conferencing signals) is a process applied to already compressed video signals. To put it simply, transcoding first decodes a compressed video signal and then encode the signal back to compressed domain. This can be illustrated via Figure

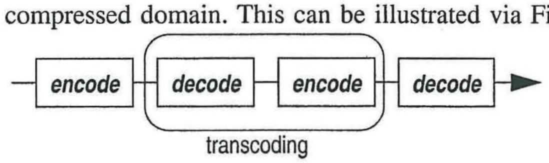


FIGURE 3. transcoding can be viewed as concatenated encoding and decoding process

3. As we have discussed, the additional steps of decoding and encoding may introduce additional distortion that can not be reversed. This effect is also called generational quality loss.

Because there is no need to display the decoded video images during the transcoding, several techniques exist to significantly eliminate the computations required to completely reconstruct the video signal before encoding it back (for example, see Ref [3]). The transcoding process generally refers to the process in which the bit rate is transcoded *downward* to a smaller value, resulting in some degradation due to additional distortion introduced. For converting the bit rate to a higher value, the problem is much more straightforward. For example, MPEG-2 video standard allows for insertion *filler* data at transport stream layer (NULL transport packets) or at elementary stream layer (stuffing bits). Inserting and/or removing filler data generally is a trivial and reversible process. In addition, it does not introduce generational quality loss to the resulting video signals.

Transcoding allows the operator to re-adjust the bit rate usage of the compressed video signal that is different from that of the original signal. This capability, together with the statistical multiplexing algorithms, allow system operator to optimally allocate bandwidth among different applications.

2.3 Statistical Remultiplexing Gains

Statistical remultiplexing combines the traditional statistical multiplexing with the transcoding operations. The key objective here is to re-organize the bit rate of encoded video signal across multiple channels of independently encoded materials. This eliminates the dependence of digital cable service operators on the bit rate of pre-encoded materials, which often were generated remotely and done at a different time. The stat remuxing allows the operators to utilize the available bandwidth most efficiently, both in terms of bit rate allocation, channel lineups, and in terms of demographic based programming. The capability to perform the stat remuxing is key to effectively deploy the digital cable broadcast service and other broadband services.

It is well known that statmuxing results in overall efficiency gain in bandwidth utilization. Similar phenomenon can be seen in many different scenarios - from circuit utilization in voice networks, or bandwidth utilization in data networks, to multiplexed video transport over digital cable networks. One way to measure the statistical multiplex gain of digital video signals is by using the probability distribution of the instantaneous bit rate. The bit rate can be measured, for example, in terms of the number of bits divided by the time intervals between two time-stamps, called program clock reference (PCRs). This is a rather simplified view of the bit rate behavior of compressed video streams, as it ignores temporally correlated bit rate information. Nonetheless it provides a good measure of the statistical behavior of bit rate usages. A bit rate probability distribution reflects how frequent the bit rate of a stream has a particular range of values. In what follows, we use the real-time digital feeds, courtesy of TCI Headend-In-The-Sky (HITS), as sample bitstreams to obtain simulation results in statistical remultiplexing gains. For the sake of simplicity, we study the bit rate of video transport stream only.

The bit rate distribution of a typical video channel can be shown in Figure 4. Depending on the type of program content, the shape of the distribution varies somewhat from channel to channel. The peak of the distribution indicates the most common bit rate values, although sometimes, there may be more than one peak in the distribution, indicating that the bit rates may have two or more common values. If multiple video channels are multiplexed without transcoding the resulting multiplexed bit rate can also be represented by another probability distribution. This can be expressed as follows:

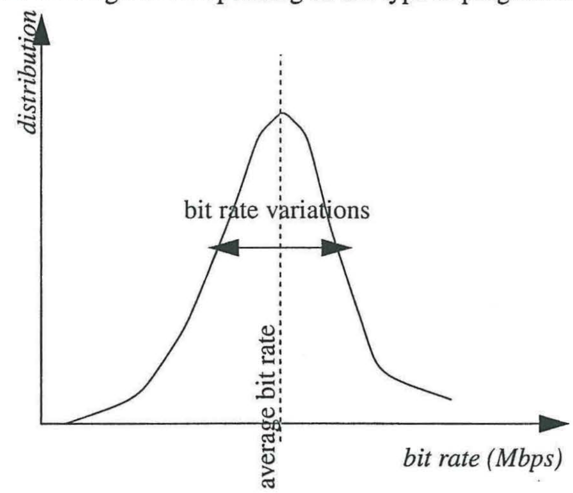


FIGURE 4. Bit rate distribution of a typical compressed video signal

X_i : bit rate for the i-th input stream **(EQ 1)**

M : number of input streams **(EQ 2)**

$$X = \sum_{1 \le i \le M} X_i : \text{bit rate for the multiplexed stream} \quad \textbf{(EQ 3)}$$

σ^2 : the bit rate variance for a typical channel (assuming all channels have similar variation) **(EQ 4)**

The statistical multiplexing gain is derived from the fact that the bit rates for different signals tends to be uncorrelated. As a result, the total bit rate variance is given by $M\sigma^2$, in stead of $M^2\sigma^2$ in the case of completely correlated signals. The effective bit rate variation allocated from the total multiplexed channel to each of the individual channels is given by $\frac{\sigma^2}{M}$. In other words, the peaks and the valleys of the bit rate tend to complement each other and reduces the overall variance of the resulting bitstream. However, the value $\frac{\sigma^2}{M}$ is generally greater than zero for independent feeds from different transponders, the objective of the statistical rate remultiplexing is to further reduce the variation of the bit rate distribution of the resulting multiplex down to zero. For non-real-time data streams, this is usually done by buffering. For real-time MPEG signals, however, precise delivery constraints by the multiplexer must be enforced to ensure correct decoder operation. This often requires a combination of buffering and transcoding of the input bitstreams, which effectively alters the shape of the bit rate distribution. Given that transcoding reduce the number of bits used to represent the coded pictures and thus generally degrades video quality, it is preferred to avoid transcoding by applying buffering to the statistical multiplexing as much as possible.

In the first example, we show the effect of rate remultiplexing of 12 video channels on the bit rate distribution (see Figure 5) is to reduce the variance of the bit rate distribution of the output.[1] Note that the input total bit rate ranges from about 20Mbps to over 28Mbps. The output bit rate, as a result of the statistical remultiplexing, is narrowly centered around 25Mbps. In real-time systems, the inclusion of audio, data packets and the occasion inclusion of NULL transport packets will make up a near constant bit rate transport stream, although this is not shown in the figures.

In the second example, similar plot is generated for only 2 independent video channels (see Figure 6). Next we consider the amount of transcoding that has to be performed

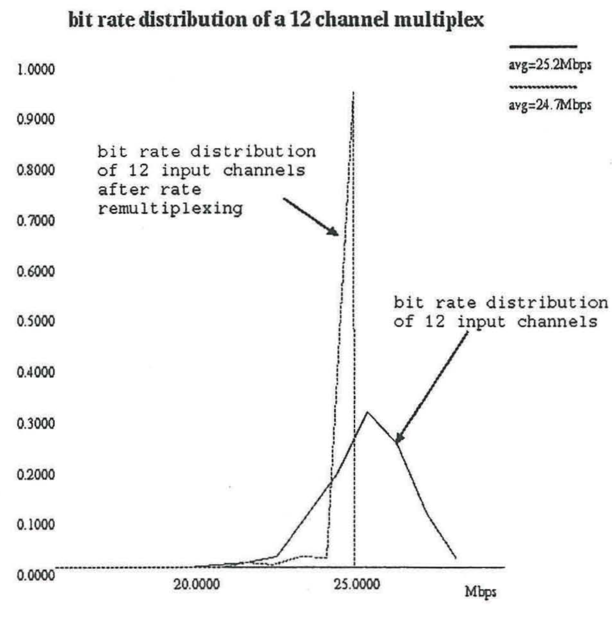


FIGURE 5. bit rate distribution for 12 independent (HITS) channels from two independent transponders (6 channels in each transponder)

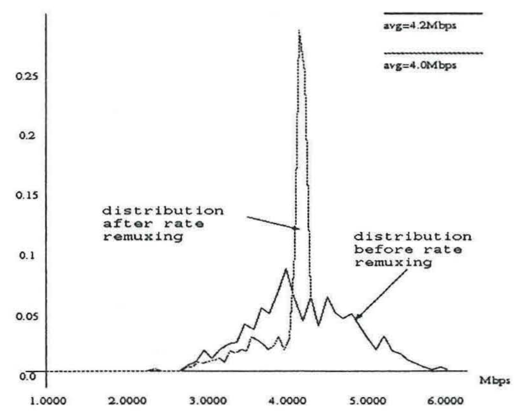


FIGURE 6. bit rate distribution for 2 independent (HITS) channels from two independent transponders (1 channel in each transponder)

on individual channels in order to achieve the above results. Note that the simulation requires certain selective real-time transcoding capability specific to the products used to generate these results.

1.Note that in order to accurately measure the bit rate, we do not include any audio, data or NULL packets as part of the bit rate calculation.

For the case of 12 channels, this is shown in Figure 7.

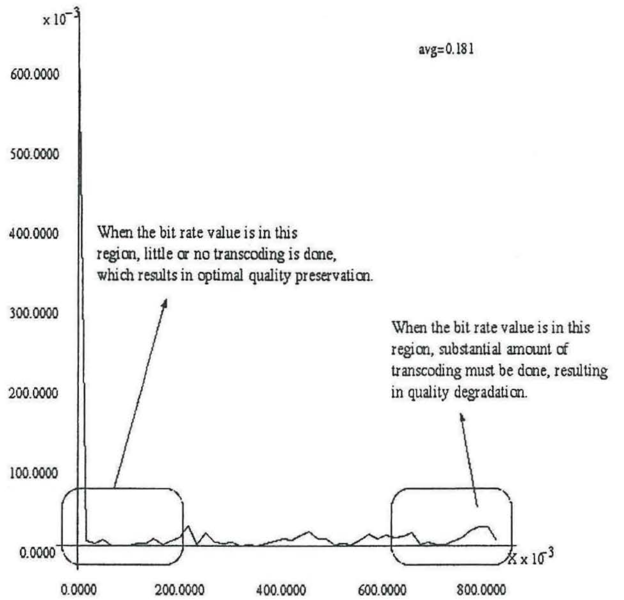


FIGURE 7. distribution of the percentage of transcoding on the input channels for a 12 channel multiplex

From this we see that the average percentage of bitstreams to be transcoded is about 2.7%.

However, if we generate the similar plot for a 2 independent channel multiplexing, the result is quite different (see Figure 8). In this case, an average of 18% of the bitstreams must be transcoded. Although the overall reduction ratio in output bit rate is lower for the 2 channel situation at $\frac{4}{4.2} = 95\%$, as compared with $\frac{24.7}{25.2} = 98\%$ for 12 channel case. This shows that in general less transcoding is required for larger number of channels to be rate remultiplexed. Subjective observation of the video quality indicates that consistent moderate amount of transcoding is more visually transparent than a temporarily aggressive[1] transcoding, even for a short duration. Therefore, effective rate remultiplexing must strive to minimize the aggressive transcoding by intelligently spreading out the processing to more channels.

Also note from both Figure 7 and Figure 8 that even in the case of 2 channel remultiplexing, well over 60% of the bit-

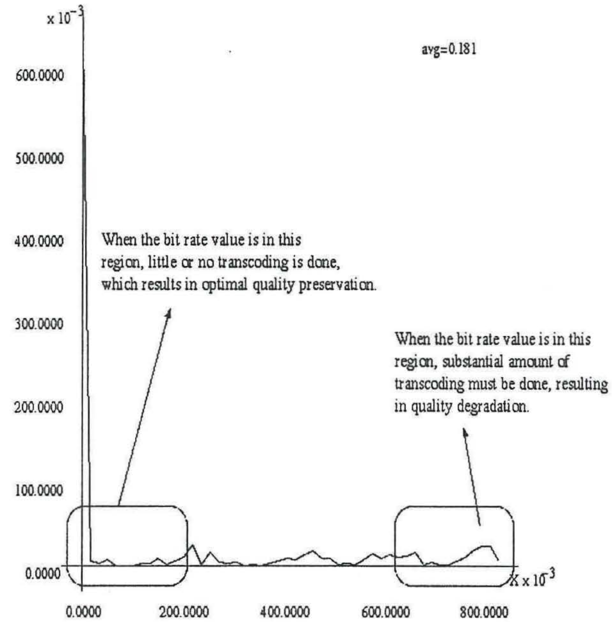


FIGURE 8. distribution of the percentage of transcoding on the input channels for a 2 channel multiplex

streams do not need any transcoding, i.e., the signal shall be multiplexing only through proper buffering. This implies that signal quality can be faithfully preserved.

Next we take a look at the bit rate distribution of one typical input channel and see how it is affected by the rate remultiplexing. Figure 9 shows the bit rate distribution of a single input channel as part of a 12 channel rate remultiplexing. Figure 11 shows another 12 channel rate remultiplexing with total output bit rate lowered by 2Mbps. Figure 11 shows the case for a 2 channel remultiplexing. In this example, we try to keep the overall bit rate reduction ratio to be as close to that of the 12 channels as possible. Yet we see from Figure 11 that there is a substantial bit rate reduction in the higher bit rate (2-3Mbps) regions. This implies a significant degradation to the input signal when the number of input channels is low.

3.0 Conclusion

In this paper, we studied the effect of statistical rate remultiplexing in terms of the distributions of bit rates before and after the system processing and the distributions of the percentage of transcoding. The general conclusions are as follows:

1.Here by *aggressive* we mean the relative bit rate reduction result from transcoding.

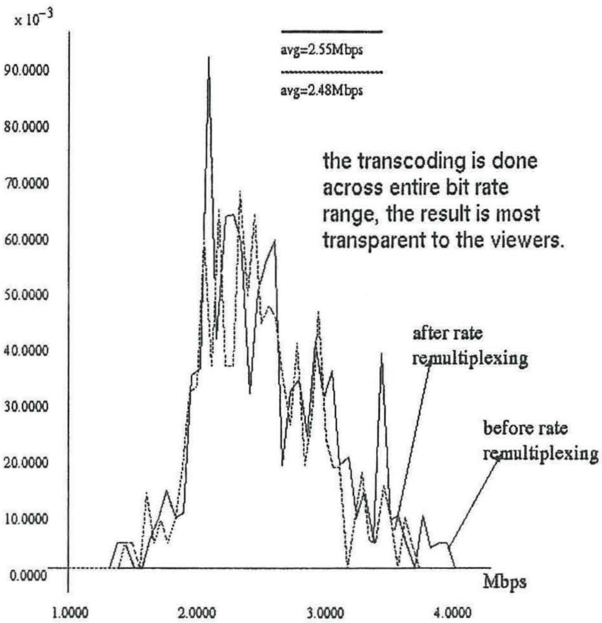


FIGURE 9. distribution of the bit rate for a single input channel as part of a 12 channel rate remultiplexing

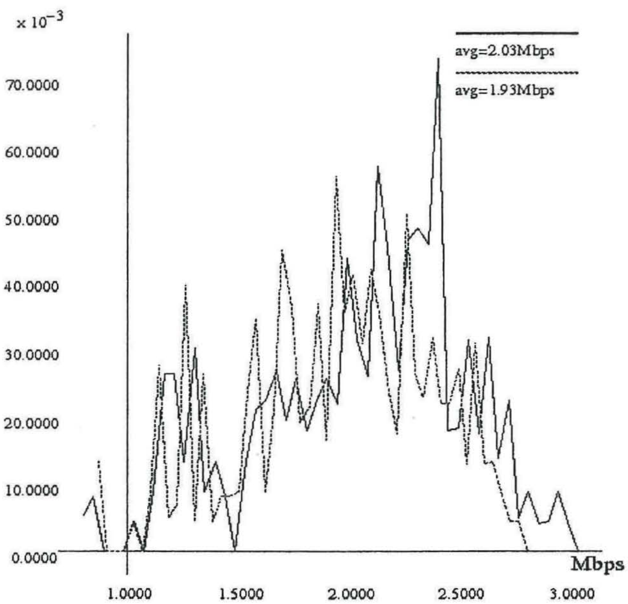


FIGURE 10. distribution of the bit rate for a single input channel as part of a 12 channel rate remultiplexing

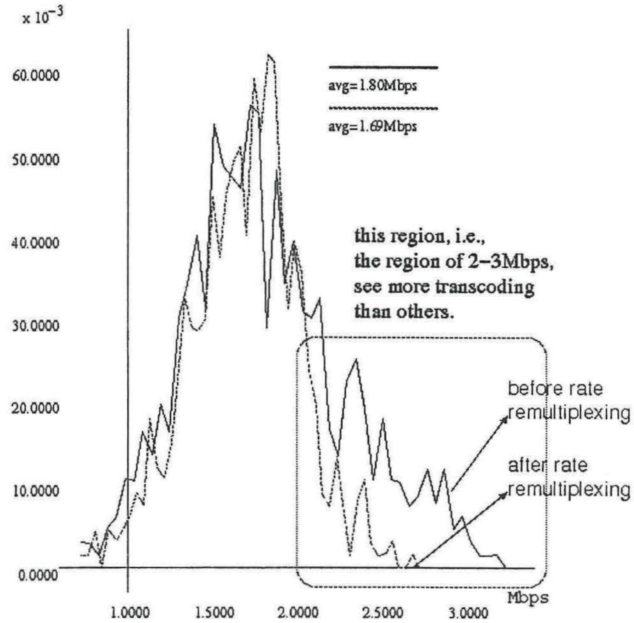


FIGURE 11. distribution of the bit rate for a single input channel as part of a 2 channel rate remultiplexing

- less transcoding is needed if larger number of input channels are remultiplexed together for form a multiplex with correspondingly higher output rate
- more uniform and less aggressive transcoding at wider bit rate range is possible for rate remultiplexing of larger number of channels
- transcoding is generally not required across all channels at all times, as it will unnecessarily degrade the video quality without improving the bandwidth utilization

All of the above indicates that the rate remultiplexing of larger number of channels results in improved video quality and effective transcoding strategies can optimally preserve the signal integrity yet maintain the best bandwidth utilization. For example, 12 independent video programs rate remultiplexed into a QAM64 channel at 27Mbps results in less transcoding, i.e., better quality. In addition, it is conceivable that 16-18 video programs rate remultiplexed into a QAM256 channel at 38Mbps shall results in even less need for transcoding than the case of QAM64 channel, resulting in more bandwidth efficient and better video quality preservation.

4.0 References

1. *ISO/IEC 13818, "Generic Coding of Moving Pictures and Associated Audio Information".*
2. *IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 4, August 1998, L.J. Lin & A. Ortega, "Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics".*
3. *IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, No. 2, April 1996, H. Sun, W. Kwok, & J.W. Zdepski, "Architectures for MPEG Compressed Bitstream Scaling".*

Authors:

Ji Zhang, Sangeeta Ramakrishnan, Fang Wu, V-Bits, Inc., 2199 Zanker Road, San Jose, CA 95131, web: *www.v-bits.com*

HFC ARCHITECTURE IN THE MAKING

Oleh Sniezko, Tony Werner, Doug Combs and Esteban Sandino
AT&T Broadband & Internet Services

Xiaolin Lu, Ted Darcie, Alan Gnauck, Sheryl Woodward, Bhavesh Desai and Xiaoxin Qiu
AT&T Labs
Rob Mclaughlin
AT&T Broadband Services Engineering

Abstract

Many architectural realizations of the HFC network successfully support the services ranging from broadcast video (analog and digital), through pay per view and video on demand, to telephony and high-speed data. However, the HFC network and its architectural implementations continuously evolve to fulfill the increasing demand for reliability, high quality, and sufficient capacity to provide a broad array of services to an increasing number of customers.

This paper presents an analysis of various architectural implementations of the HFC network. These implementations range from a traditional Fiber-to-the-Serving-Area approach, through fiber overlay for digital services, to fully-passive coaxial networks fed by dedicated fibers.

INTRODUCTION

In the last several years, cable network operators have embarked on extensive upgrades and on the transition to hybrid fiber coax (HFC) architectures. The main goal has been to evolve the infrastructure from a broadcast-type trunk-and-branch plant to a high capacity, superior quality and reliability, and two-way network ready to deliver advanced telecommunications services.

The need for this evolution is apparent from the list of advanced services, which demand increased network capacity and quality. Even for video services, the competition from alternative broadcast entertainment providers and transition from analog NTSC based video to digital (standard and high-definition) video require additional attention from cable operators.

Services	*Architectures*
Traditional Services: ▪ Broadcast TV ▪ Broadcast Radio ▪ Addressable Services: - Addressable tiering - PPV - Games - Digital radio ▪ Home Shopping **Advanced Services**: ▪ Targeted advertising ▪ Targeted entertainment ▪ Telephony ▪ High speed data ▪ Full multimedia **Competition**: ▪ DBS ▪ LEC ▪ Multimedia mergers	**Requirements**: ▪ Superior reliability ▪ Superior quality ▪ Competitive price ▪ Interactivity (two-way) **Architectural changes**: ▪ Fiber supertrunking and backbone ▪ Regional hub ring ▪ Redundancy (secondary hub rings) ▪ Deep fiber deployment & segmentation

Table 1: Services, Requirements & Solutions

Although cost competitive in comparison to the upgrades of other access networks to the same level of capacity and performance, upgrading from trunk-and-branch coax plant to HFC network requires significant financial effort from cable operators. Therefore, while satisfying the needs listed in Table 1, it is critical that the upgrades accomplish the following objectives:

1. Establish a future-proof network
2. Lower operating cost

3. Significantly improve network reliability (MTTR: mean time to repair; and MTBF: mean time before failure)

All these considerations lead to our 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.

ARCHITECTURAL STUDY OBJECTIVES

This section outlines the scope and objectives established and followed by our team for this advanced architecture study.

Services Supported By the Network

The network must be capable of supporting a wide variety of services:

1. Analog video (basic, premium, PPV and IPPV
2. Digital video (broadcast and IPPV)
3. Interactive video (e.g., video on demand (VOD))
4. High speed data access
5. Telephony
6. Telemetry

Architectural Choices

For practical reasons, the analysis concentrated on several architecture alternatives:

1. Fiber-to-the-Serving-Area (FSA) with fiber node segmentation supporting forward and reverse communication services[12]
2. Trunk-and-branch coaxial architecture for traditional services with fiber overlay for advanced services[3]
3. Deep fiber penetration to eliminate RF actives (Multiplexed Fiber Passive Coax — MFPC)

Impact on Powering

The analysis considered the impact of all architectures on powering options for the plant and for terminal equipment. The following issues were analyzed:

1. Impact on overall network power consumption
2. Terminal equipment powering alternatives
3. Optimal network powering architecture (level of centralization)
4. Power distribution in the existing plant and in plant extension or new-builds

Acceptable Performance

All the architecture alternatives were analyzed for the same levels of performance:

1. Forward path end-of-line performance
2. Reverse path performance
3. Network availability

Equipment

The following issues were considered during the analysis:

1. Technology feasibility
2. Equipment availability
3. Equipment price

Fiber Count and Redundancy

To provide the level of reliability required for the advanced services, the network must allow for cost-effective redundancy and low mean time to repair. To achieve these objectives, the following issues were analyzed:

1. Number of fiber in single cable sheet
2. Redundancy in different network sections
3. Number of homes served per optical cable route without redundancy
4. Restoration time

Network Characteristics

While optimizing network topology to support different services with different characteristics, it is desirable to maintain transparency and flexibility in introducing new services. The team therefore defined and evaluated options and balances of:

1. Ultimate topology for backward compatibility and migration feasibility
2. Network transparency between terminal unit locations
3. Functionality of network terminals
4. Centralized vs. distributed network terminals

Cost Comparison

After taking into account the considerations listed above, the analysis concentrated on cost comparison for the architecture alternatives. The initial capital costs were compared against benefits and life-cycle savings. Some additional cost elements were also analyzed qualitatively (for example, terminal equipment cost).

Additional Considerations

The following additional items were considered:

1. Terminal equipment standardization
2. Time to market for the architectures (construction time and equipment development time)
3. Suitability for MDUs and plant expansions

This paper will provide a snap shot of this study with emphasis on the architecture alternatives, related features, and cost comparison. It will be seen throughout the paper that the differentiation among the alternatives are related to:

1. Bandwidth per home passed
2. Network power consumption
3. Network availability (reliability and MTTR)
4. Efforts of network alignment and certification
5. Maintenance cost

ARCHITECTURAL ALTERNATIVES

Generic HFC Architecture

Most of the existing HFC networks in markets exceeding 100,000 homes are based on ring configurations (single or dual). Traditional headends are consolidated, and are interconnected by the market rings (primary hub rings) with additional processing centers dubbed primary hubs (PH). The secondary hub rings are introduced to enable the headend consolidation. In this configuration, secondary hubs (SH) serve as signal concentration and distribution points to limit the number of fibers between primary hub ring and secondary hubs to achieve cost reduction, improve MTTR, and allow for cost-effective backup switching (redundancy). These two elements are common to most of the HFC networks used by cable operators. The differences are limited to technological choices in these rings.[456]

This generic architecture is depicted in Figure 1. The network sections marked as 'C' and 'D' (access network) provide the last links to the customers. The architectural solutions for this network section differ from operator to operator. However, most of them deploy fiber-node-based configuration followed by active RF coaxial networks. The differences are reflected in the node sizes and in the level of redundancy. This access network is a subject of continuous architectural analysis by almost all major HFC network operators.

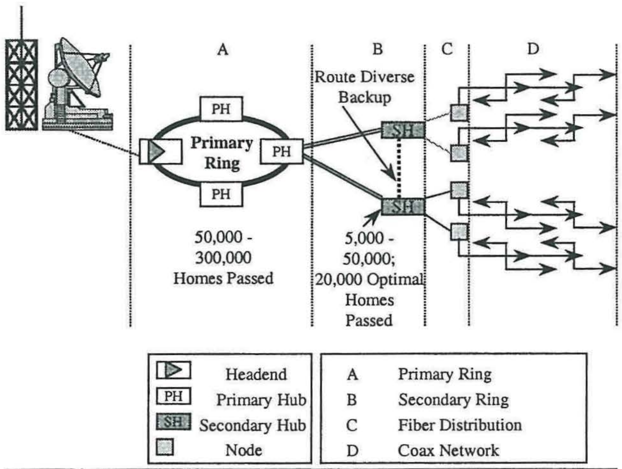


Figure 1: Modern HFC Network: Generic Configuration

Access Network Alternatives

Physically and logically, the last-mile access network can be categorized into two categories: Point-to-point (PTP) and point-to multi-point (PTMP). Examples of the PTP architecture include copper-pair based double star networks (traditional local telephone networks), certain FTTC networks, and WDM-based PON. Among the PTMP networks are such networks as HFC-based networks, certain TDMA-based PON, and wireless networks. Each network was historically defined to support certain type of services with their intrinsic characteristics. The challenge we are facing today is how to evolve the network architecture to support a wide variety of services with characteristics that the embedded system was not originally designed for.

The HFC network was originally designed for broadcast services with PTMP architecture. Utilizing emerging lightwave technology with different levels of fiber deployment, our study concentrated on defining upgrade alternatives to support new service needs that may be optimized with certain degree of virtual PTP configuration. The alternatives were:

1. Fiber to the Serving Area with fiber node segmentation supporting forward and reverse communication services
2. Trunk-and-branch coaxial architecture for traditional services with fiber overlay for advanced services
3. Deep fiber penetration to eliminate RF actives (Multiplexed Fiber Passive Coax — MFPC)

FSA with FN Segmentation

This network architecture (Fig. 2) has been used by many HFC network operators. The differences are mostly related to the node sizes, with particular emphasis on the design effort (optimization for power consumption, time-to-market, or end-of-line performance and bandwidth) and the level of redundancy (refer to Cox's ring-in-ring topology). In the analysis performed by the team, this architecture was used as a baseline[7] and was characterized by the following parameters:

1. Node size is between 600 and 1,200 household passed (HHP)
2. Each FN can be segmented with up to four 300 HHP buses
3. Number of fibers from secondary hub to the node are between 4 and 6,
4. Number of amplifiers in cascade are between 5 and 8,
5. Upstream is in 5-40 MHz, and downstream is in 50-750 MHz with 50-550 MHz being allocated for analog video and the rest of the bandwidth for digital services

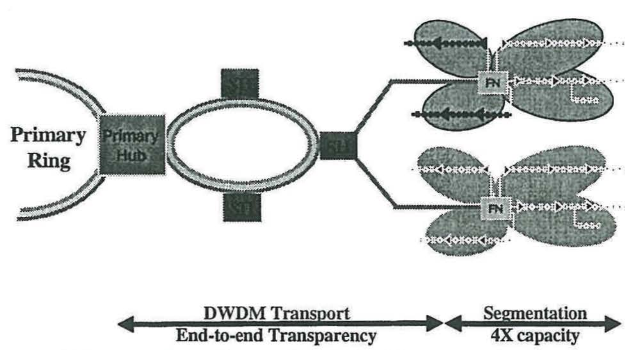


Figure 2: FSA with FN Segmentation

Mini Fiber Node for Fiber Overlay

To resolve upstream limitations (ingress noise and bandwidth) and to simplify terminal operation, we proposed and evaluated the mini fiber node technology. Using emerging lightwave technology, the existing network is overlaid with a fiber-to-the-bridger architecture to exploit the large ingress-free bandwidth at high frequency for two-way digital services. As shown in Fig. 3, independent of existing systems, the mFNs couple directly into the passive coax legs (with drop taps) after each distribution coax amplifier (i.e. line extender). Each mFN contains a low-cost laser diode and a low-cost PIN diode, and is connected to the headend with separate fiber.

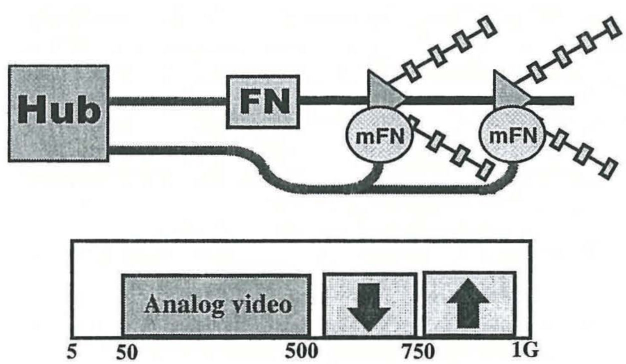


Figure 3: Mini Fiber Node for digital overlay

Based on this strategy, the mFNs subdivide the FN serving areas into small cells (typically 50 HHP/mFN) and exploit the clean and large bandwidth at high frequency for both upstream and downstream transmission. The mFN therefore creates a new path for digital services without affecting analog TV services carried by conventional FN/amplified-coax paths. All services are then merged over passive coax distribution legs.

Features of the mFN architecture:

By exploiting the clean and large bandwidth, this strategy increases overall system bandwidth beyond current coax amplifier limitations, for new digital services, without replacing coax amplifiers and changing amplifier spacing. It also avoids the complexities of noise reduction (e.g., frequency agility) and related signal processing and RF techniques. This therefore simplifies system operation and reduces terminal cost.

Also, because mFNs only carry digital subcarrier signals over a clean high-frequency band, low-cost, low power consumption and space-saving optical and RF components can be used in the mFNs and also at the headend.

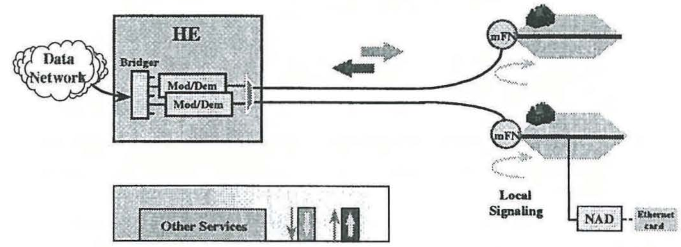


Figure 4: mFN Local Access Control Protocol

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 (Fig.4). This can be accomplished by incorporating a simple out-of-band signaling loopback scheme such that users know the upstream channel status prior to transmission. This enables the use of standard, but full-duplex, Ethernet protocol (CSMA/CD), and therefore the use of standard and low-cost terminals (modified Ethernet transceiver, Ethernet bridger and Ethernet card). No ranging is needed, and the headend becomes virtually operation-free. The relatively small round-trip delay between each user and the mFN (~2000ft) also substantially increases bandwidth efficiency and reduces contention delay. This is appealing for VBR (variable bit rate) type of services. For CBR (constant bit rate) services, certain scheduling or priority provisioning may be necessary, and can be easily added to the above protocol (Fig. 5).

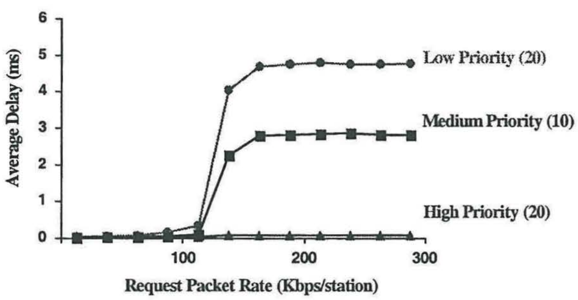


Figure 5: mFN-based priority provisioning protocol

The large bandwidth supported by the mFN infrastructure also enables the use of efficient but much simpler modulation schemes such as multi-level FSK or even ASK. This therefore provides a low-cost, low-power-consumption alternative to the current cable modem technique.

Open Issues:

The mFN technology explores a radical path to resolve the network limitations. Unfortunately, the existing system for broadcast video and DOCSIS-modem-based services is left behind with no benefits from the mFN strategy.

Bringing fiber deeper into the network incurs certain incremental cost. Reducing the cost of fibering (material and labor) then becomes critical. The mFN strategy will simplify system operation for new services and reduce terminal cost. However, it will be more compelling if the front-end cost can be justified by operational savings over the *entire* network, both embedded and mFN overlay.

Convergence: Multiplexed Fiber Passive Coax Networks

To resolve those issues and to establish a platform that can improve the performance of the embedded system while also evolving to meet future needs and simplifying operations across the entire network, we proposed a new architecture called Multiplexed Fiber Passive Coax (MFPC). As shown in Fig. 6, instead of overlaying over existing coax amplifiers, mFNs eliminate all the coax amplifiers. (our design indicated that 2-3 coax amplifiers will be eliminated by one mFN). Between each mFN and customers, passive coax plant is used to carry both current and new services.

Fibers connecting multiple mFNs will be terminated at the MuxNode that resides either at the original fiber node location or at location that "consolidate" 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.

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. To future-proof the network with more capacity and simple terminals, this architecture can also support a distributed-processing strategy for new, purely IP-based services enabled by the mFN and MuxNode, and maintain all the benefits of the initial mFN strategy. The development can be partitioned into two phases.

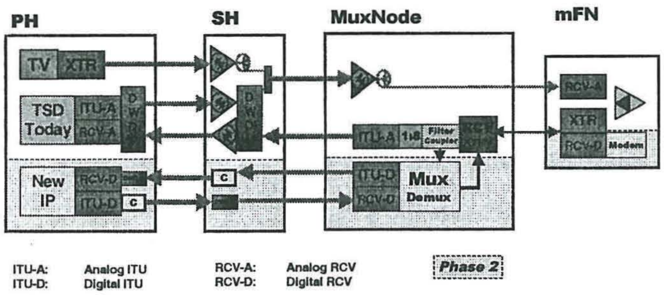


Figure 6: Multiplexed Fiber Passive Coax (MFPC) architecture

Phase1: Current Service Delivery

One preferred approach is to transmit downstream broadcast signals, including analog and digital video, over dedicated fiber from the primary hub to the secondary hub. The existing narrowcast and switched signals, such as telephony and high-speed data using DOCSIS-based cable modems and set top boxes will be DWDM multiplexed at the PH and transmitted over another fiber to the secondary hub. After DWDM demultiplexing, the narrowcast and switched signals will be optically combined with the broadcast signals (in a different RF band) and transmitted to the MuxNode over optical fiber.

At the MuxNode, those combined downstream signals will be optically amplified and distributed to multiple mFNs over optical fiber. Given the short distance between the MuxNode and the mFN, it is also feasible to move the EDFA from the MuxNode to the secondary hub and only keep the optical splitter at the MuxNode, therefore simplifying the MuxNode. Other options, such as re-lasing at the MuxNode, are also promising.

The mFN performs the same function as that of a typical FN. It receives downstream signals and distributes them to customers over passive coax cable. In the upstream direction, the mFN combines upstream signals from all the coax branches it serves and transmits them to the MuxNode over optical fiber.

The MuxNode further performs O/E conversion to those upstream signals from the mFNs, RF combines them, and transmits to the secondary hub using a wavelength specified laser. Upstream signals from multiple MuxNodes will be then DWDM multiplexed at the SH and transmitted to the PH. Besides DWDM multiplexing, another option is to re-lase upstream signals at the SH.

To expand capacity of the system, one can frequency shift the upstream signals at each mFN such that when they are combined at the MuxNode they will be at separate RF bands. If the re-lasing option is used at the SH, the frequency stacking could also be deployed at the MuxNode.

Phase 2: Add-on Distributed Processing Platform

As shown in Fig. 6, a distributed processing platform can be transparently added when it is needed. The new IP-based services could be delivered to the MuxNode in baseband format over a DWDM link separate from the one carrying current services. As an alternative, the baseband signals and RF passband signals could be transmitted over the same DWDM link. At the MuxNode, the baseband signals are demultiplexed and further distributed to multiple mFNs.

At each mFN, the downstream digital baseband signals are received and modulated onto RF carriers. They will then be combined with the received downstream RF signals and transmitted over the coax buses to customers. In the upstream direction, customers transmit the new IP-based upstream signals in a high frequency band (900 MHz – 1 GHz) to the mFN using simple FSK or QPSK modulation scheme. At the mFN, the MAC function is performed for those high frequency signals as discussed before. Those signals are also demodulated to baseband. They will then ASK modulate RF carriers at frequencies above the current 5-40 MHz return band. These signals will be combined with the 5-40 MHz band, and transmitted to the MuxNode over a single fiber. Another option is to frequency shift the 5-40 MHz band and keep the baseband signals untouched.

Between the MuxNode and the mFN, coarse WDM (1.3μm/1.5μm) is used for single fiber bi-directional transmission. At the MuxNode, the combined upstream signals are received and separated. The ASK signals are demodulated and multiplexed, and the 5-40 MHz band signals are RF combined (or frequency stacked). Both signals will be transmitted to the PH over the same or separate DWDM trunks.

COST COMPARISON

Front-end Cost

To evaluate implementation feasibility, we completed more than 600 miles network design of the above three architectures, and compared the front-end cost. An example is shown in Fig. 7. The design was over a medium-density system with 80 HHP/mile. The analysis was based on the existing and emerging lightwave and RF technology, and the parameters used in the cost model, such as labor cost, were based on commonly used industry averages.

It is interesting to see that, with the help of the MuxNode and by eliminating all coax amplifiers, the cost of the passive coax design is comparable to that of initial mFN design. Of course, the value of this architecture is far greater than any other alternative.

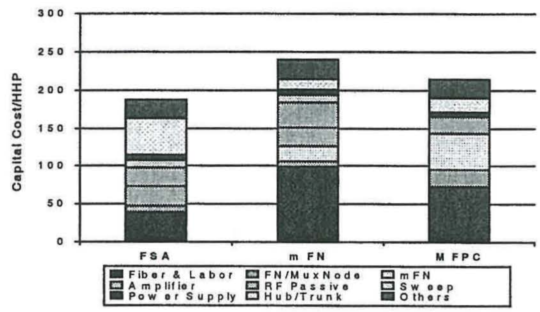


Figure 7: Front-end cost comparison

Life Cycle Cost

One of the biggest advantages of this architecture is the elimination of all the coax amplifiers. This results in substantial reduction in active components and therefore reduction in overall network power consumption (Fig. 8 and 9). In addition, sweeping and maintenance of active components in the field will be reduced, which leads to further operation savings. It is estimated that annual operation savings, including reduction in customer calls. etc, could reach $11/HHP.

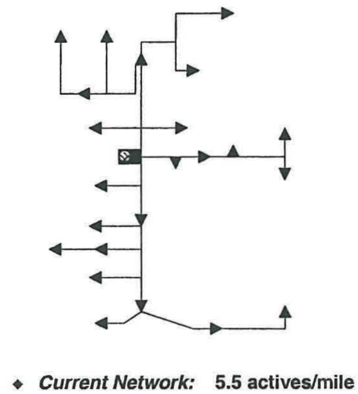


Fig. 8 Current FSA HFC

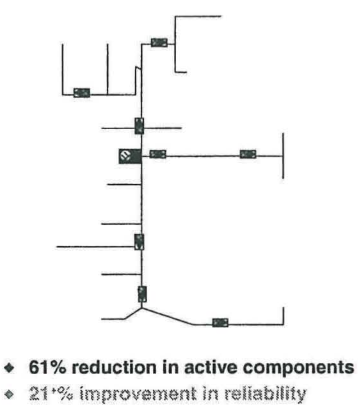


Fig. 9 MFPC

CHARACTERISTICS OF MFPC NETWORKS

Powering

Due to a significant reduction in the number of active devices and the progress in RF active component technology, the power consumption of the network elements will be decreased by at least 50%. This will allow for powering architecture optimization and for less challenging powering of the customer terminal devices.

Network Availability

The network availability increases dramatically, especially after the implementation of phase 2. In phase 1, elimination of the active cascade and the lower number of actives will allow for reduction in number of failures per MuxNode area. Moreover, due to the star configuration, the failure group size will be significantly lower (especially in the case of optoelectronics failure). Additional improvement in network availability will be realized thanks to lower MTTR (easier network troubleshooting). This feature will be significantly enhanced after implementation of phase 2 (dedicated bandwidth to smaller group of homes). Optimized powering will further improve the network availability.

Flexibility

The MFPC architecture maintains the transparency of the HFC network. In phase 1 of the implementation, the architecture does not differ from the traditional node-based architecture except for the fiber depth. When phase 2 is deployed, the network combines all the benefits of the HFC network and the cell-based multistage multiplexing architecture. Any new service can be implemented by addition of the terminal equipment at the customer premises and in few network centers.

This MFPC supports today's terminals, while paving an easy migration path to enable high-performance terminals and simple operation scheme. The DOCSIS-based cable modems can operate normally with better performance due to reduced bandwidth sharing and potential elimination of the noisy 5-15 MHz band. The simple modems, based on simple modulation and local access control ("Ethernet-like") protocol, can be added on when they are needed.

Capacity Growth and Future-Proofing

In phase 1, the network can deliver several (up to 10) TSD channels downstream with either 64 or 256 QAM modulation (presented in reference [1]). The reverse capacity per home passed can be increased by 50% if 64QAM modulation can be used (the MFPC architecture will provide significantly better reverse performance from the point of view ingress and interference management).

In addition, this architecture provides several unique paths for bandwidth expansion that other networks cannot match. In phase 1, this can be accomplished by:

1. Expansion of the traditional reverse bandwidth to 45/48 MHz (no filter cascading)
2. Wider implementation of concentration at the MuxNode (instead of conventional RF combining)

More important, the MFPC ensures that the major part of the network stretching between the primary hub ring and the mini fiber node is future-proof and provides large bandwidth capacity (a bandwidth upgrade for this part will require upgrading some equipment in the primary hub, secondary hub, MuxNodes and mFN). In the passive coax plant, up to 100 Mbps capacity can be added and shared among 100-200 homes passed. This capacity can be provided with QPSK modulation and will be symmetrical. An increase in this capacity can be further achieved by deploying more efficient modulation schemes. Positioning the mFN closer to customers further paves a simple path to even bring fiber to the homes when it is needed.

CONCLUSION

HFC is no doubt the first economically viable means for broadband services. The need to further upgrade networks for future-proofing motivates industry to explore new architectural solutions. We studied three architecture alternatives, analyzed their capabilities, and compared their costs. The results demonstrated that, by utilizing multi-stage multiplexing topology and emerging lightwave and RF technology, deep fiber penetration offers a future-proof network. The new MFPC architecture supports all current system operation with better performance, and enables abundant noise-free bandwidth for future growth. The advantageous architecture reduces the number of active components in the field, therefore simplifying network powering and operation. The incremental front-end cost can therefore be justified by life-cycle savings. More operation savings can be further realized with simplified terminal equipment. Yet the capacity, reliability, and performance of this new network are far better than that of other alternatives.

ACKNOWLEDGMENT

The authors would like to thank their colleagues: Adel Saleh, former division manager of AT&T Labs, and Mark Dzuban, Cameron Gough, Marty Davidson from AT&T Broadband Services Engineering for their support and assistance to the advanced architecture study.

The authors also would like to express their gratitude to Bogdan Lomnicki and his team at Quasar Inc for designing the advanced architecture over our cable networks.

[1] Oleh J. Sniezko & Tony E. Werner, Invisible Hub or End-to-End Transparency, 1998 NCTA Technical Papers.
[2] Tony E. Werner & Oleh J. Sniezko, HFC Optical Technology: Ten Years of Progress and Today's Status, Choices and Challenges, CED, September 1998.
[3] Xiaolin Lu, Ted Darcie, Alan Gnauck, Sheryl Woodward, Bhavesh Desai, Xiaoxin Qiu, Low-cost Cable Network Upgrade for Two-Way Broadband, SCTE 1998 Conference on Emerging Technologies.
[4] Oleh J. Sniezko, Video and Data Transmission in Evolving HFC Network, 1998 OFC Conference.
[5] Thomas G. Elliot and Oleh J. Sniezko, Transmission Technologies in Secondary Hub Rings — SONET versus FDM Once More, 1996 NCTA Technical Papers.
[6] Oleh J. Sniezko, Reverse Path for Advanced Services — Architecture and Technology, 1999 NCTA Technical Papers.
[7] Network Architecture, Report TS-NETW-981001 (internal AT&T report)

Abstract

HOMEPNA - ENABLING BROADBAND DISTRIBUTION THROUGHOUT THE HOME

Thomas L. Funk
Manager of Strategic Marketing, Compaq Computer Corp.

The Home Phoneline Networking Alliance (HomePNA) was founded in June 1998 by AMD, Intel, IBM, Compaq, 3Com, Tut, Epigram, AT&T Wireless, Lucent, Conexant and Hewlett-Packard to speed the adoption of a unified home phoneline networking standard for data rates up to 1Mb/s (1.0 specification).

Since its founding, the Alliance's membership has grown to include more than 70 leading manufacturers and vendors from across the technology and consumer electronics industries. Many of these companies have successfully launched product lines based on the 1.0 specification and the Alliance has begun work on a second specification (2.0) that will raise data rates up to 10Mb/s. Member companies expect to ship products based on 2.0 within the calendar year.

In the broadband access world, cable companies, RBOCs, CLECs, and satellite providers have focused their energy on the consumer "last mile" solution. These efforts include initiatives for cable modems, xDSL, and wireless technology. While a clear high-speed access winner has yet to emerge (and in fact a single technology may never capture the entire market,) the members of HomePNA recognize the need to build out what Gary Arlen has dubbed the "last five meters."

Growing consumer interest in home networking and in broadband access demonstrate that healthy stand-alone markets exist for each. However, in both cases, the influence and demand for one product set - either home networking or broadband access - will greatly enhance the demand and sales for products from the other.

For example, broadband access on its own has compelling applications for a single consumer accessing the Internet through their stand-alone PC. But, when this access can be easily distributed to multiple information appliances within the home (i.e. other PCs, a set-top box, Internet telephone, and almost any other device with an IC,) the Jetsons-like vision of the networked home suddenly becomes very real and achievable.

Similarly, a stand-alone home network offers consumers compelling applications such as print and file sharing, multi-player gaming, as well as dial-up Internet access. But when these applications are combined with high-speed, always-on connections to the Internet (as promised by broadband access), the phrase "killer app" springs to mind.

While broadband access and home networking can both succeed in the absence of the other, rapid adoption of one is supportive of the growing adoption of the other. This is the mutual opportunity we should pursue. Always-on high-speed access distributed throughout the home enables a multitude of product and service opportunities, while the next generation products that will provide these new services will increase the consumer desire for broadband access.

In-Service Measurement of Composite Triple Beat and Other Proof-of-Performance Enhancements

Robert V.C. Dickinson
Edwin L. Dickinson
Dovetail Surveys Inc.

Abstract

This paper covers some innovations developed to streamline the execution of proof-of-performance testing and the generation of the required reports. These include automated cyclic test signal application and in-service testing techniques for measuring CTB, CSO, C/N and in-band spurs without producing visible disruption of subscriber viewing.

Background

Proof-of-performance testing has become another of those necessary but annoying parts of the cable operator's life and business. Having done many proofs since the inception of this requirement we have struggled to improve the procedure for better measurements, quicker execution and less disruption of the cable system especially in terms of customer dissatisfaction. The memory of the first series of proofs still lingers. Data was taken laboriously with available equipment and recorded with pencil and paper. This just couldn't last very long. Then computer programs were developed to capture the data and ultimately to direct the spectrum analyzer and other equipment to do their functions semi-automatically. These improvements were welcomed and helped to systematize the procedure and reduce the time required, but still proofs are a burden sopping up man hours and communications channels while interfering with subscriber viewing.

The Traditional Way

The proof-of-performance requires the following measurements to prove compliance with Part 76.

1) To be measured on all analog channels
   - Visual carrier levels both absolute and relative
   - Visual carrier level stability
   - Aural carrier levels relative to the associated visual carrier
   - Aural carrier frequencies relative to the associated visual carrier

2) To be measured on channels designated as test channels;
   - In-band flatness
   - Visual carrier to noise ratio
   - Visual carrier to composite triple beat, composite second order and spurs
   - Low frequency disturbances
   - Terminal isolation (usually derived from manufacturer's specs.)

3) To be measured at the headend every three years on the designated test channels;
   - Differential gain
   - Differential phase
   - Chrominance-luminance delay

The measurements under 1) above are accomplished manually or automatically employing standard test equipment and procedures. Those listed under 2) are measured at the headend and at each test point and are normally those which require special test signals and procedures which can be the most labor intensive and disruptive to CATV viewing. The items under 3), although required only triennially and only at the headend, are easily performed along with the tests in 2).

The traditional way to make the measurements listed under 2 & 3, is to insert and remove test signals at the headend. To measure C/N, CTB, CSO and spurs without interference from the visual carrier and the program material on the test channel requires that the test channel be turned OFF while the field measurement is executed. Measurement of in-band flatness may be done with a non-intrusive vertical interval test signal. Some remove the video modulation to measure low frequency disturbances.

The tests requiring synchronized operations between the field and the headend represent an area where the most time and aggravation is experienced. When a test signal must be inserted or removed or a channel turned ON or OFF, a person must be present at the headend and communication must be carried on between the field and the headend to coordinate the action with the commensurate measurement. In addition to the personnel needs, a radio or telephone channel is required. The availability of a communication channel is taken for granted until the regular system operational traffic loads the channel to where the testing procedure is slowed down waiting for openings to make the required transmissions. To make matters worse, there are spots where radio communications are poor and extra time is required to make up for this deficiency.

A More Efficient Way

One approach at a "more efficient way" is to automate the testing cycle at the headend. This amounts to permanently installing programmable equipment which does the test signal insertion, switching, etc. in a preset sequence with adjustable timing to match the field data recording needs. This "permanent" feature eliminates setup and knock down time after the initial installation. With such a system in place no personnel are required at the headend and the cyclic operation of the headend tasks continues without reference to the operations in the field. This arrangement immediately eliminates the requirement for at least one person (at the headend) and for any type of field/headend communications. The field task then resolves to moving to the next test point, plugging in and waiting for the next testing sequence to begin and then taking the required data. There is some time lost (about one half of the period of one sequence on average) but this is often less than that lost due to communication problems, etc.

The above is a great improvement on the traditional way but still requires the repeated removal of test channels to accomplish the C/N, CTB, CSO, and spur measurements. This is perhaps the most onerous problem since it causes subscriber dissatisfaction due to viewing interruptions. Not only are the subscribers unhappy but the CSRs must be alerted and given the proper responses for the call-ins and, of course, the volume of traffic increases substantially. This problem has been ameliorated by announcements and other cable operator innovations but certainly is still a problem that we could all do without.

In-Service Testing

The first approach to in-service testing is the normal use of vertical interval test signals, which is often employed. In-band flatness has long been tested by use of the Multiburst signal where the amplitudes of the bursts are compared to indicate the flatness of the channel. This is a good and acceptable test but suffers from the specific frequency nature of the packets and sometimes from the addition of noise in remote locations making the results less accurate. A variation on this approach is offered by use of the "Line Sweep" signal. The Line Sweep signal is not available from all video test signal generators. It is a video signal on a single line where a full amplitude sine wave signal is swept from 500KHz to 5 MHz linearly over the active video portion of the single video line. This signal can be applied in the vertical interval in one or both fields. Since this signal is full amplitude it is always as great or greater than any normal video component and therefore produces a maximum response. This signal appears 60 times per second if inserted in both even and odd TV fields. To detect this the spectrum analyzer is set to sweep slowly across the 6 MHz channel (usually about 10 seconds) in a peak detection mode. This displays a solid line indicating the response across the channel. A typical in-band-flatness display is shown in Figure 1. The preferred way of quantifying the result is to capture the spectrum analyzer screen picture on a computer and later submit it to computer analysis. A faster sweep may be employed but this can cause narrow dropouts between the peaks although the peaks will still indicate the in-band flatness. These dropouts may be detrimental to the computer analysis.

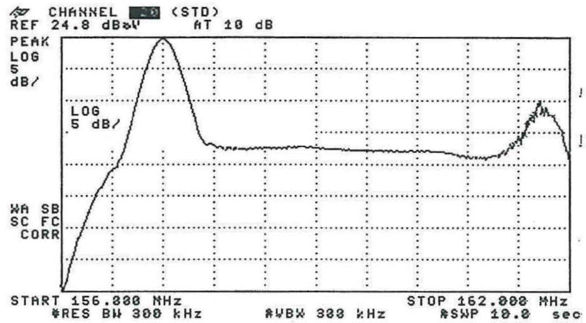


Figure 1 - In-band Flatness

Insertion of vertical interval test signals is also useful in performing the triennial video measurements non-intrusively. In this case the images are captured through a calibrated demodulator and mathematically processed to determine Differential Gain, Differential Phase and Group Delay. This measurement is required to be made only at the headend. Figure 2 illustrates the capture of FCC Composite and Multiburst signals from which differential gain, differential phase and group delay parameters may be calculated.

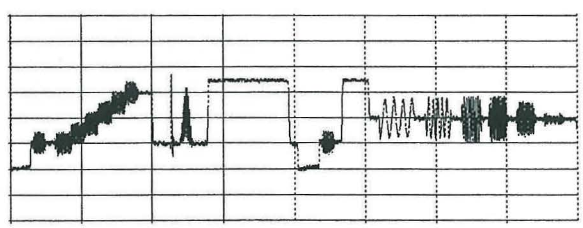


Figure 2 – Detected FCC Composite & Multiburst Signals

The real killer test is to see the composite third order products that lie beneath the visual carrier. Some methods have been developed to extrapolate the desired CTB results from out of band measurements but direct observation is to be desired. Tests for C/N, CSO and spurs may be made on a quiet line in the vertical interval although these measurements are vulnerable to small disturbances on the visual carrier which sometimes degrade or obscure the desired information.

Tests indicate that modern TV receivers are tolerant to loss of visual carrier for up to several vertical interval lines without picture disturbances or loss of sync. This fact suggests a technique of turning the visual (or visual and aural) carrier(s) OFF for short periods of time and observing what is underneath. Based upon this method, "looking underneath" the visual carrier has been pursued coupled with a sampling type of measurement of CTB, C/N, CSO and spurs during this OFF period. This technique makes use of work done by John Huff [1] and Francis Edgington [2] with the addition of the transmitted signal gating.

There are several precautions to be observed in applying this approach.

First, the rise time of the gating signal for the visual carrier must be properly controlled or else spurious products will be generated outside of the channel of interest. What happens inside of the test channel is of less importance since it is in an unviewed time (vertical interval) of the picture format. If the gate has a rise time similar to the TV sync pulse its products will remain within the channel. Figure 3 shows a typical gating signal as seen in a received video waveform.

Second, the detection equipment must provide a similar gating function except that it must be turned ON only during the time that the modulator is turned OFF. The timing for the received sample may be derived from counting TV lines or by sensing the instant of full carrier turn OFF of the test channel.

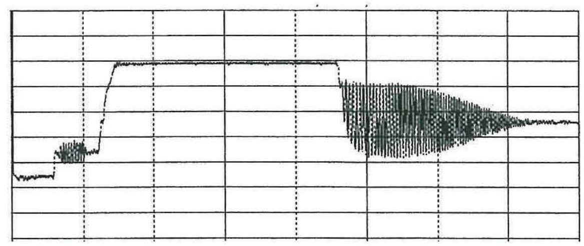


Figure 3 – Detected Video Waveform with Gating Pulse

In order to view the spurious signals, a spectrum analyzer following (or including) the receiving sampling gate, is employed. The spectrum analyzer must be operated in the peak detection mode since data is received for a period of microseconds and therefore have a very low average value. Since the gating of the visual carrier occurs only 60 times per second (when gating both fields of the TV picture) the analyzer sweep must be slow enough to receive at least one sample per horizontal pixel of the display. Typically there are about 500 pixels horizontally in the spectrum analyzer display so the sweep must be no faster than $500/60 \cong 9$ seconds to assure that there is data in each pixel interval. Sweeping for 20 seconds provides at least 2 samples per pixel of which the maximum sample amplitude will be recorded.

The analyzer resolution bandwidth setting will determine the shape of the indicated response. In the case of a CW signal the amplitude will be less than without sampling by an amount determined by the impulse response of the analyzer due to its bandwidth and filter settings. The rise time of the indication on the spectrum analyzer is governed largely by the resolution bandwidth. For instance, the rise time (RT) of a pulse is in the order of the inverse of the bandwidth, i.e., the RT in a 30 KHz bandwidth is about 33 microseconds and reaches 63% (-4 dB) of its full value in this time while a 60 KHz bandwidth would reach 86% (-1.3 dB) in the same time. A more complete discussion is given in Reference 2. Due to this effect a video filter narrower than the resolution bandwidth should not be used since it will increase the rise time of the instrument and thereby attenuate the signal peaks. Figure 4 shows a CW signal lying beneath the channel video in the gated mode.

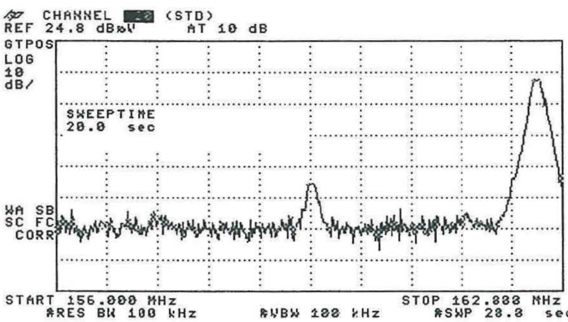


Figure 4 - CW Carrier under Video Signal in Gated Mode

For a "noise-like" signal this technique records the amplitude peaks. The peaks bear a relation to the average noise (which is what we are trying to measure) which has been determined [3] and is a result of the instrument settings. For example using a sampling pulse width of 40 microseconds and a 30 KHz resolution bandwidth, the average of a noise-like signal (C/N, CTB, CSO) is about 4.3 dB less than the indicated peak value. A typical spectrum analyzer display with CTB and CSO products within the test channel is shown in Figure 5. To obtain the numerical result the peak value is recorded and then the correction factor is applied. Since there are many complex variables in the real life situation the simplified model may need some correction. Therefore it is suggested that the system be calibrated by measurement of known amounts of injected CW and noise signals.

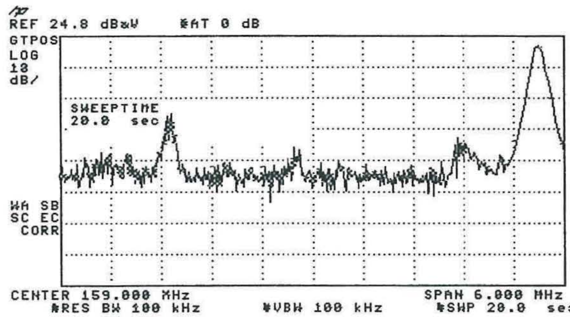


Figure 5 – CTB, CSO, Spurious Signals and Noise under Test Channel in Gated Mode

Putting It All Together

With all of the above in mind, the system would be something like the following including a rack mounted controller, permanently installed in the headend, plus a small encoder module for the modulator of each test channel and special considerations and possibly equipment for the field measurements.

The controller needs to intercept the video signals going to each respective modulator and insert the required vertical interval test signals at the proper times. In addition, it must control the gating time and duration to the encoding gates that interrupt the rf TV signal. The controller must also provide means for configuration of the test parameters such as start and stop dates and times, vertical interval test signals to be used, test signal insertion fields and lines and indicators to show the present status in the testing cycle. Test modes are valuable to check setup conditions and to perform the headend tests that may be done manually while the tester is in the headend.

The encoder module is required to interrupt the rf TV signal. This is preferably the visual IF signal in the modulator. Using the IF signal requires high isolation gating of rf only at relatively low frequencies (IF) which do not necessitate UHF design of the circuits as would be required with VHF and UHF modulator outputs. In addition, not interrupting the aural carrier makes interference with the TV audio less likely. Configuration of both the controller and the encoders must stress precise timing and amplitude control to create "glitchless" switching for high quality in-service testing.

At the receiving end a sampling gate, which has high OFF isolation at all CATV system frequencies, is needed to avoid corruption of the data due to leakage from the ever-present full amplitude system signals. Some analyzers include a time gating function that may be used in this connection. The sampling gate must be triggered in a way to sample only when the modulator gate is OFF. This also is provided in some analyzers which include TV demods and TV line finders. It is also possible to trigger on the no signal condition that occurs when the carrier is gated. This generally requires some kind of receiving device to separate the test channel from the other channels. Whatever analyzer and supporting equipment is used must be fully understood in order to be properly configured and thereby correctly capture and ultimately analyze the data.

A side benefit of the above permanently installed system, beyond the ability effortlessly execute proof-of-performance testing, is that it may be employed for incidental testing without a great deal of setup and knock down a normally experienced.

This system has been employed in the performance of numerous proofs with dramatic improvements in measurement efficiency and without disruption of program delivery. The equipment and techniques described in this paper are the subject of a current patent application.

References:

1) Huff, John L. – "Time Selective Spectrum Analysis – Extending Spectrum Analyzer's Usefulness", NCTA 1983

2) Edgington, Francis M. - "Time-Selective Spectrum Analysis: Non-Intrusive Analysis of a Cable System", NCTA 1992

3) Hewlett Packard Application Note 1303 – "Spectrum Analyzer Measurements and Noise"

Contact info: Bob or Ed Dickinson
Dovetail Surveys Inc. - 961 Marcon Blvd., Suite 450, Allentown, PA 18103
Tel: 610-264-0100, Fax: 610-264-8901
web site http://www.dovetailsci.com
e-mail info@dovetailsci.com

Interactive TV Applications:
Standard APIs for Digital TV Receivers

Petr Peterka
General Instrument Corporation

Abstract

With advanced analog cable settop boxes, early digital satellite boxes and recently with digital set-top boxes, television viewers are getting used to more than just audio-visual (A/V) content. Enhanced broadcast includes graphical and data enhancements to the specific A/V program, such as additional text, graphics, user choices, teleshopping, etc. Standalone applications such as electronic program guides are becoming a norm. All current deployment of such systems is based on proprietary solutions.

The Internet and the Web in particular was enabled by platform independent content formats such as HTML and Java. The same must happen in order to deliver enhanced content and downloadable applications to digital TV receivers of all kinds including terrestrial receivers, cable set-tops, satellite receivers and computers. A platform independent content format is not enough to provide a rich, well-integrated audio/video/data content to all possible receivers. These devices must have a common set of application programming interfaces (API) in order to make downloadable content and applications truly interoperable. The goal of these APIs is to provide access to the receiver functions such as tuning and channel changing, receiver resources such as a return channel and the TV screen, as well as system information necessary for channel navigation and program guides. User-specific data such as user preferences and personal data may also be made available to applications via these APIs.

This paper addresses the current work-in-progress in the ATSC T3/S17 specialist group also known as the DTV Application Software Environment (DASE), specifically the definition of Java APIs. Classification of downloadable applications is presented together with a set of requirements that must be met in order to enable such interoperable applications. Also a detailed description of all DTV receiver system services that are being abstracted by the Java APIs is presented. The focus of this paper is on Java-based downloadable applications that are enabled by the presence of a Java interpreter in the form of a Java Virtual Machine (JVM) and a set of Java APIs providing access to the DTV receiver functionality.

MOTIVATION

As the wide spread use of the WWW was enabled by platform independent content formats such as HTML and Java, we will see more and more enhanced content and downloadable applications available on digital receivers of all kinds including terrestrial receivers, cable set-tops, satellite receivers and computers. A slow processor and a very small amount of memory no longer limit these devices, which was the case until recently. The point of platform independence is essential to the broadcast environment. As opposed to the computer industry where there is a very small number of dominant operating systems and hardware platforms, the embedded device world uses a wide variety of real-time operating systems as well as hardware platforms and CPUs. Content delivered over the air must be consumable by a large number (majority) of receivers since it is too costly to waste the available bandwidth by sending separate formats of the same content to a wide variety of receivers.

The Internet provides a very interesting model but the DTV network can provide much richer user experience. Delivering the Internet experience to a television audience is a worthwhile goal but the potentials are not limited to those of the Internet. The broadcast world has its own characteristics, advantages and limitations. It is primarily a unidirectional multicast network, although a return channel is becoming a common practice in both digital cable networks with a two-way cable as well as satellite networks with traditional telephone return paths. Another differentiating factor is that the primary content is high-quality and high-resolution audiovisual content with a large amount of bandwidth available on each channel. The capability to synchronize applications with the main video program is often necessary.

DOWNLOADABLE APPLICATIONS

Applications described here are only those that are downloaded from the network and use nonproprietary formats and APIs. Although native (i.e. platform dependent) and resident applications are important, they are outside the scope of this discussion.

The wide variety of downloadable applications suitable for the DTV domain may be classified based on different criteria including (1) level of interactivity, (2) level of synchronization with the audio-visual content, (3) style of authoring or programming, and (4) level of persistency.

Level of Interactivity

The simplest applications provide additional textual or graphical enhancement to the traditional audio-visual programming. Examples of such enhancements may be as simple as subtitles or closed captions, or more sophisticated such as graphical illustrations during weather forecasts, statistics sent along with a sports program, or news headlines provided during a news hour. This class of applications allows very minimal or no user interactivity. The next step in the level of user involvement is local interactivity. This means that all receivers are running the same application but the behavior is different based on individual user actions, choices or settings. Examples of such applications may include additional features of applications mentioned in the previous category, such as user customized news search criteria, team or individual player statistics selected by the user, as well as user-selectable camera views during a football or golf game. This class also includes applications such as electronic program guides, local interactive games and commercials. Finally, the highest class of applications based on interactive capabilities is server-based interactivity or sometimes called transactional applications. A return path and a connection to a network server, such as a cable headend or a service provider, are required. This type of applications includes traditional Internet service, interactive commercials with order forms, electronic commerce, video on demand (NVOD), etc.

Level of Dependency on the Audio-Visual Program

Many applications may be related to the current audio-video program. These applications fall into the enhanced audio-video program category. Such applications will most likely be delivered in the same MPEG program stream. Within this category, applications may be either loosely associated with the audio-video streams, such as a news ticker during a political news program, or highly synchronized with the video, such as real-time telemetry data during an automobile race or basketball statistics updated in real time.

Other applications will be completely independent of the current A/V program. Examples may include an electronic program guide, stock ticker, e-mail or home banking. These applications may be superimposed on top of the video program if they don't require the entire screen or may replace the video program. Such applications may be delivered as an independent MPEG program.

Authoring Style

This classification for a DTV application describes the authoring style of the application. There are three possible types of authoring: (1) procedural applications, (2) declarative content applications, and (3) mixed-format applications.

Procedural applications are applications that are generally written in a procedural programming language such as Java while declarative content applications are written using a declarative language such as HTML. Mixed-format applications are those that contain both procedural and declarative content components.

Level of Persistency

Broadcast applications are generally played once and not permanently stored on the DTV receiver. Such applications are usually small or are delivered in the data carousel stream in advance so that they can be played when triggered either by synchronization with the audio/video channel, by the signaling protocol or by user interaction.

An application that is used often can be optionally stored locally on the DTV to save time next time it is used. Some applications are too big to be acquiring and downloading every time they are needed. An example of such an application is an electronic program guide or an e-mail client. Another reason for storing an application is that it may be a paid service; once the user pays for the application, it is decrypted and stored for future use.

CURRENT PRACTICE

Current advanced analog cable settop boxes, digital satellite receivers and digital cable set-top boxes provide television viewers with more than just audio-visual content. Enhanced broadcast includes graphical and data enhancements to specific A/V programs, such as additional text, graphics, user choices and customization, teleshopping, etc. Standalone applications such as electronic program guides are becoming a norm. All current deployment of such systems is based on proprietary solutions. Therefore, such services are limited to a specific network provider and CE manufacturer.

Another common practice is the delivery of the Internet to the television. It is enabled by supporting a two-way communication path using protocols such as HTTP or a mechanism where a large number of web pages, possibly related to the current video programming, are broadcast and the user may browse this limited view of the Internet.

TECHNICAL SOLUTION

Several conditions must be met in order to provide an economically feasible environment for downloadable DTV applications. First, applications must be written in a format understandable by all DTV receivers; therefore, platform independent. Such content format may be HTML [12], Dynamic HTML (HTML, CSS1 [13] and DOM1 [14]) or XML-based XHTML [15], or a procedural programming language format such as Java byte codes [7]. Second, an interpreter capable of decoding the application format is required on the receiver. In the case of HTML or any of its related formats, an appropriate markup language parser is needed. In the other case a Java byte code interpreter such as a Java Virtual Machine (JVM) [8] is required. Third condition is a standardized set of interfaces, which provide access to the DTV receiver functionality that is essential for any downloadable application.

Such an environment enables the deployment of platform independent downloadable applications and still allowing for a relatively high implementation freedom for DTV receivers, including the choice of a platform or middleware implementation language, operating system, CPU and other relevant hardware.

STANDARDS WORK

A platform independent content format is not enough to provide a rich, well-integrated audio/video/data content to all possible receivers. These devices must have a common set of application programming interfaces (API) in order to make downloadable content and applications truly interoperable. There are several standards organizations and company consortia which are trying to do exactly that. An open standard definition of such APIs opens up the market, which is currently dominated by proprietary solutions, for competition. DAVIC [18], DVB [19], ATSC [1], OpenCable [20], Sun Microsystems [9] and other organizations are currently in the process of specifying a set of Java APIs that would enable such applications.

The following sections will discuss the current work in progress of the ATSC T3/S17 [2] specialist group also known as Digital TV Application Software Environment (DASE). This group has been working over a year on a selection of an Application Execution Engine (AEE) and a Presentation Engine (PE). The current draft specification includes the JVM as the Application Execution Engine, XHTML-based parser as the Presentation Engine, and a set of Java APIs.

API DESIGN GOALS

The goal of these APIs is to provide access to selected receiver functions such as tuning and channel changing, receiver resources such as a modem or a return channel if available, a conditional access module and the TV screen, system information (e.g. ATSC PSIP, DVB SI, SCTE SI, etc.) necessary for channel navigation and program guides, as well as user preferences and user-related information.

Another important aspect is manageability of these receivers. The current TV viewer is used to a very reliable appliance which does not put up computer-like error messages, does not require a frequent reboot nor an advanced degree to operate. The TV is becoming a broadcast network computer combining the traditional passive TV experience with an interactive aspect currently expected on computers. It is essential to provide means to remotely diagnose and troubleshoot any problems with no or minimal involvement of the viewer. A number of APIs provides application and resource instrumentation that will enable such remote management.

In order to maximize the implementation freedom of each DTV receiver manufacturer, these APIs must maintain a certain level of abstraction. Consumer electronic manufacturers should be also free to set their own policies for resource and application management.

Security becomes an essential component of a DTV receiver as well. The APIs must be defined so that appropriate security policies may be implemented based on the trust level of each downloaded application, user preferences as well as the network operator's business model.

DASE APIS

This section describes the current work-in-progress in the ATSC T3/S17 specialist group (DASE) [2] and other directly related efforts. First, the DTV reference architecture and a detailed description of all DTV receiver system services that are being abstracted by the DASE APIs is given. Finally, the API definition and a mapping to other related standards such as MPEG-2, DSM-CC, ATSC T3/S13 [4] and T3/S16 [5] conclude the discussion.

Reference Architecture

The purpose of this paper is not to describe the reference architecture of a DTV receiver. The following paragraphs show the basic concepts in order to demonstrate how the DASE APIs relate to other DTV receiver platform components.

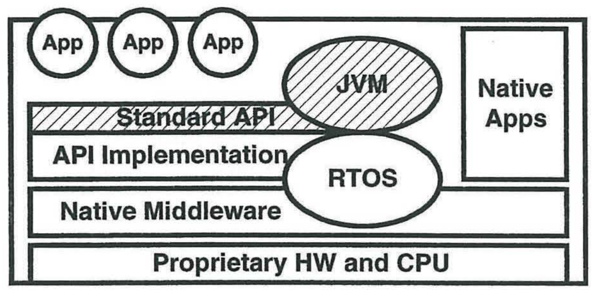


Figure 1 - Reference Architecture

Figure 1 above shows a very simple diagram of a DTV receiver reference architecture just to demonstrate the basic elements of a DTV receiver with respect to downloadable applications and the API they may use. The main point here is that only the shaded components are standardized: the Java Virtual Machine and the Java APIs. Everything else, primarily the API implementation, the real-time operating system, the native middleware, as well as the hardware and CPU may be chosen by the CE manufacturer based on its own architecture.

System Services

The DASE APIs act as an abstraction layer above the operating system's native libraries as well as the receiver middleware. The Java Virtual Machine in the role of an AEE provides a set of APIs which are very generic and typical for operating systems. In order to provide a rich set of DTV receiver specific functions, such as tuning, and a close integration between a DASE application and the DTV receiver, an additional, well-defined set of APIs must be developed.

ATSC T3/S17 (DASE) has identified a number of system services by analyzing the API needs of many example interactive and data broadcast applications submitted by ATSC T3/S17 members. The following system service groups were identified:

1. Network Communication Group
2. Content Management Group
3. PSIP Service Group
4. Presentation and User Interaction Group
5. Application Management Group
6. Environment and User Management Group
7. Resource Management Group
8. Security Service Group
9. Utility Service Group

Network Communication Group

This group represents services related to the network communication between the DTV receiver and external devices. This is primarily the MPEG-2 transport stream in the broadcast environment. It also includes return channel, both batch and real-time. If the DTV receiver communicates with other devices on the home network, this group provides appropriate services to support it. The DTV receiver is expected to be a part of a larger home network. Downloadable applications may use or control other devices on the network. These services should abstract and unify the home network device control functions.

The main services in this group include service selection (i.e. channel changing), access to broadcast data, explicit tuning, access to an optional return channel and home networking.

Content Management Group

This group of services represents content decoding, synchronization of content and media control. It is primarily concerned with audio/video content but includes data of various formats as well as downloadable application code. The content management services may also provide the system with content life cycle, content integrity and stream synchronization functions.

For applications that may need to store content, this service group will also provide content storage and playback functions.

Program and System Information (PSIP) Service Group

The PSIP Group provides the receiver with MPEG-2 Program and System Information (PSI) and the ATSC Program and System Information Protocol (PSIP) [3] data. It is closely related to the Content Management Group because it partially describes the content as received via the network. It supports basic navigation services (i.e. transport streams, virtual channels), program guide services via scheduled program events (i.e. event information tables - EIT) as well as data services via extensions supporting data broadcast as defined by ATSC T3/S13 and object carousels defined by T3/S16 and DSM-CC.

Presentation and User Interaction Group

This group is primarily responsible for presenting content to the end user. Content consists of audio, video, text and graphics. These services provide functions for displaying and controlling the presentation of information to the user.

This group of services also allows the DTV receiver user to interact with the applications executing on the receiver. These include IR remote control events, front panel, keyboard, pointing device, voice control, etc.

Application Management Group

The application management services provide the support for downloadable applications: code verification and authentication, application registration and life cycle, version management, and application management for the purpose of local or remote monitoring and control.

Environment and User Management Group

Environment services collect and provide information about the overall DTV receiver environment and profile, configuration and versioning, system monitoring and usage statistics.

The DTV receiver may be a multi-user environment. This group also includes system services that support user profiles, permissions, preferences, usage and user identification functions.

Resource Management Group

Resource management services provide a uniform mechanism to reserve, release and request status of scarce resources such as tuner, modem, conditional access module and other devices, as well as local storage resources such as memory, disk space, etc.

Security Group

This group of services supports security-related functions which provide the system with capabilities to control and manage access to applications, content, and services: conditional access for controlling access to audio, video and data services; encryption and decryption of content; authentication of applications and users; digital signatures and certificates for controlling application downloads; privacy for protecting user-specific information; content control permissions based on content rating or end user geographic location,

Utility Service Group

The DTV receiver is expected to be a multitasking/multithreading environment where different components need to communicate with each other. These include event routing and dispatching, exception handling, inter-process communication and other forms of signaling.

This group also provides miscellaneous functions such as time and scheduling, mathematical functions, string manipulation, internationalization and localization functions.

Some applications may need to store their state and other data between executions. These services will provide access to the local storage functions. They will work closely with the resource management services to provide coordinated access to the storage device.

API Definition

Access to the above described system services is provided via the standardized set of DASE APIs. There are several high-level requirements related to the API definition: (1) The API must be at such an abstraction level so that it does not dictate any particular implementation or DTV receiver architecture; it is important to allow DTV receiver manufacturers to differentiate their products. (2) The API definition must be consistent across all groups of system service APIs. Event mechanisms, error and exception handling are examples of such areas. (3) The API definition should be minimal but complete. This means that any overlap between APIs should be avoided while all application related aspects of the DTV functionality should be accessible through the API.

The current definition of the complete set of DASE APIs consists of several groups, also called Java packages. Some were designed by other industry consortia and some were developed directly for DASE. Specifically, DASE adopted the Sun Microsystem JavaTV APIs [9] and Personal Java APIs [10], Java Media Framework APIs [11] and HAVi [16] User Interface API.

JavaTV APIs

A substantial portion of the DASE API set comes from the Sun Microsystems JavaTV APIs [9]. It provides a high-level, protocol independent set of APIs that are not adopted only by ATSC but by other standards organizations such as DVB. This fact is very important because JavaTV provides a common subset of APIs for applications intended to run on different systems (e.g. ATSC or DVB). JavaTV includes the following APIs:

1. Abstract SI API provides protocol independent access to the basic information delivered to the receiver via a system information (SI) protocol. This API supports DVB SI, ATSC PSIP (A65) as well as SCTE [6] SI protocol.

This set of APIs is divided into several subpackages which represent different views into the SI database depending on the application's needs. Specifically, the navigation package representing a customizable collection of channels, the guide package representing current and future program events, the pipeline package providing information about groups of related channels and the descriptor package which allows an application to retrieve raw MPEG descriptor data. This is necessary to support extensions to SI protocols such as PSIP.

2. Service Selection API allows an application to select services or change channels. When a channel is selected, this API provides the application with a set of presentation controls depending on the nature of the service provided by the particular channel, e.g. video, audio, subtitles, graphics, applications, etc. An event mechanism is included to provide the calling application with information about changes in the selected service.

3. Data Broadcast API provides an abstraction over different types of data delivered to the receiver. This includes streaming as well as file-based data (carousels), asynchronous, synchronous and synchronized data, as well as different data encapsulation mechanisms including IP.

4. Applications Life Cycle API is represented by an Xlet which is the broadcast equivalent of a Web-based Applet. It can be used to do basic application control functions, e.g. initialization, start, pause, stop. This API also includes a mechanism to report application's immediate state back to the controlling entity, e.g. an application manager.

5. Media API: JavaTV also includes portions of the Java Media Framework [11] API and several extensions defined by DAVIC [18]. Through these APIs applications can control the media presentation, audio and video attributes, and get access to media stream events for synchronization.

DASE APIs

APIs designed specifically for DASE consist of extensions that provide access to ATSC specific features such as PSIP, Data Broadcast protocol, etc. or functionality that was not included in other adopted APIs. DASE-specific APIs include:

1. **SI API** provides extensions to the JavaTV SI API in order to include ATSC specific features as defined in PSIP (T3/S8).

2. **Data Broadcast API** extends the JavaTV data broadcast API in order to provide access to ATSC specific data signaling features as represented by the data broadcast specification defined by T3/S13 [4] including an abstraction of the Data Service Table (DST).

3. **Application Management API** extends the JavaTV Xlet API in several ways. The Xlet is initialized with some ATSC specific objects including the data service information the application is part of. This API also adds a concept of an Application Manager where applications can be registered for a persistent download and which provides access to other running applications, in the form of an application proxy, subject to security policy limitations. A significant functional extension is provided by the manageability instrumentation API, which is based on the ITU-T X.731 state management standard [17]. This API enables content providers or network operators to monitor the activities on each DTV receiver, get notified of problems, collect usage statistics and control the behavior of downloadable applications. This capability will be essential in order to provide a very reliable and fail-safe operation of each receiver.

4. **User Management and Preferences API**. The Preferences API allows applications to get access to the user-defined settings and preferences, and potentially modify such preferences. These may include preferred language, user favorite channels, rating ceiling, etc. If the DTV receiver supports a multi-user environment, the User Management API provides a basic mechanism to create, delete and activate users, associate a list of preferences and permissions with each user and invoke user authentication functions.

5. **System API** is a simple set of properties used to communicate the specific receiver API specification and implementation version numbers, system profile and similar attributes.

6. **Security API** enables the internal security policy to be enforced through any DASE API. This group of APIs defines a set of security-related exceptions and security-related permissions, which may be used in defining the security policy as well as aid in implementing such a policy.

HAVi APIs

Home Audio/Visual Interoperability (HAVi) is a home audio-video networking specification [16]. It includes a set of Java APIs whose subset is currently included in the DASE specification. HAVi User Interface APIs provide additional features not found in Java Abstract Windowing Toolkit (AWT) such as transparency, support for video, remote control events, etc. There are two packages included from the HAVi specification: the User Interface package and the Remote Control Event package.

Personal Java

The entire DASE API set is expected to be implemented on top of the Personal Java platform [10]. Such a platform provides the basic APIs necessary for any device to run downloadable applications and provide basic OS-like functions. Note that some packages or classes may be included only optionally, such as AWT or the Applet package.

CONCLUSION

The purpose of this article was to provide a background on the current standardization work in the area of enabling enhanced interactive content for digital TV receivers, specifically the ATSC effort to define a set of Java APIs providing access to the receiver functionality and user data. A standardized set of APIs is one of several requirements for a platform supporting downloadable interoperable applications.

Details about other parts of the DASE architecture, the Presentation Engine, as well as the draft specification including a detailed JavaDoc documentation may be obtained from the DASE Web site [2]

REFERENCES

[1] Advanced Television Systems Committee (ATSC): http://www.atsc.org

[2] ATSC T3/S17 DTV Application Software Environment (DASE): http://toocan.philabs.research.philips.com/misc/atsc/dase/

[3] Program and System Information (PSIP): http://www.atsc.org/Standards/A65/

[4] T3/S13 Data Broadcasting: http://toocan.philabs.research.philips.com/misc/atsc/t3s13/

[5] T3/S16 Interactive Services: http://toocan.philabs.research.philips.com/misc/atsc/t3s16/

[6] Society for Cable Telecommunications Engineers (SCTE): http://www.scte.org/

[7] Java language specification: http://java.sun.com/docs/books/jls/index.html

[8] Java Virtual Machine (JVM): http://java.sun.com/docs/books/vmspec/index.html

[9] Sun Microsystems JavaTV http://java.sun.com/products/javatv/

[10] Personal Java: http://java.sun.com/products/personaljava/spec-1-1/pJavaSpec.html

[11] Java Media Framework (JMF): http://java.sun.com/products/java-media/jmf/forDevelopers/playerguide/index.html

[12] HyperText Markup Language (HTML): http://www.w3.org/TR/REC-html40/

[13] Cascading Style Sheets (CSS): http://www.w3.org/Style/CSS/

[14] Document Object Model (DOM): http://www.w3.org/DOM/

[15] Extensible HyperText Markup Language (XHTML): http://www.w3.org/TR/WD-html-in-xml/

[16] Home Audio/Visual Interoperability (HAVi): http://www.havi.org

[17] ITU-T X.731 State Management: http://www.itu.int/itudoc/itu-t/rec/x/x500up/x731.html

[18] Digital Audio Visual Council (DAVIC): http://www.davic.org/

[19] Digital Video Broadcasting (DVB): http://www.dvb.org/

[20] OpenCable: http://www.opencable.com/

About the Author

Petr Peterka is a Senior Staff Engineer in the Advanced Development department of General Instrument Corporation. He serves as a vice-chairman of the ATSC T3/S17 specialist group and as a Java API software architect for this group. He has also actively participated in the development of a number of APIs discussed in this paper. Petr can be reached via e-mail at ppeterka@gi.com.

Managing The HFC Environment Using Cable Modem Derived Performance Data

Rex Bullinger
@Home Network

Abstract

Analysts estimate that there are well over 1/2 million cable modems in use today in HFC networks. Due to a lack of adequate installation and maintenance tools a portion of these modems are operating with suboptimal receive and transmit levels.

The results of a survey of a number of modems show that upstream and downstream levels are often not ideal. The result of this is that a part of the network bandwidth in the HFC link layer is potentially being consumed in error correction.

Some tools currently available to the cable modem installer, network troubleshooter, and network maintenance technicians are described.

INTRODUCTION

Transporting data over cable TV networks is not new. For many years I-nets (Institutional Broadband mid-split RF Networks) have carried data using early RF modems from a half dozen or so pioneer vendors. Also not new is the transport of data over impaired environments; ones in which a certain level of data loss is assumed and designed for.

What is new is the developing ubiquity and raw data density of digital data signals 64 and 256 QAM on cable. Also on the very near horizon is the ubiquity of telephony over cable. Though telephony's raw data throughput may not be as demanding as cable modems or digital TV, at least in the downstream, the need to eliminate perceived gaps in speech means it does require greater signal isochronal integrity.

So the suitability of the cable TV environment for the carriage of data has been a subject of discussion for many years and continues to be. The purpose of this paper is to discuss current actual field practices in the evaluation and monitoring of the cable TV physical RF environment with respect to cable modem two-way data carriage.

IMPLICATIONS OF CURRENT PENETRATION RATES

Over the air radio frequency (RF) transmissions as well as RF transmissions over copper paths that are longer than a few tens of feet can and will suffer impairments. To counter these impairments transmission protocols have been developed which contain error correction capabilities. Similar error correction capabilities exist and are utilized in the data signals carried on cable as well[1]. Some of these methods such as FEC (Forward Error Correction) are unidirectional and are used in digital TV transmissions where there is no return path, as well as in cable modem signals. But where a return path exists, other levels of data error correction are possible such as TCP retransmissions. In today's carriage of data over cable, error correction is much in use. But how much? This is not well known.

Error correction is a wonderful capability but it is not free. The price to be paid is network capacity or bandwidth. For FEC this bandwidth hit is up front in that it is built into the signal as overhead whether it is used or not. When a two way path exists TCP retransmissions are possible as another layer of error correction capability. When subscriber penetration is low there can exist heavy retransmissions for error correction and/or packet loss without it being very noticeable to users. But as subscriber penetration grows and the extra bandwidth needed to support them is already being used for error correction, network performance may seem to degrade more rapidly than the network design would suggest.

This means that since we currently don't know the amount of error correction that is actually being employed to support current subscriber levels we don't know how much HFC capacity headroom we have. Clearly, we need means to assess when HFC performance is marginal so attention can be focussed on improving the HFC plant with the goal of driving error correction to zero. This would make both the data path more reliable as well as free up bandwidth being wasted on error correction.

ANALOG TESTING AND CERTIFICATION

A current industry practice is to require certification of each node before releasing it for cable modem service. This certification is based solely on traditional RF measurements assuming that if the plant meets these requirements then data carriage will be successful. This leap of faith has been successful so far. Meeting FCC Proof of Performance specifications and diligence in CLI (Cumulative Leakage Index) seems to be adequate. So, for the downstream, clean pictures do generally mean clean data.

For the upstream path, however, the situation is a bit less straightforward. In spite of this, though, pursuing the same philosophy of certifying with a suite of analog RF tests has been generally working with upstream data signals as well.

However, certification at the time of data service launch in a particular node says nothing about the *ongoing* capability of the node to support data service.

DIGITAL TESTING

The term "digital testing" as used here describes the concept of both retrieving analog parameters through the network (through digital means) as well as the retrieval of purely digital parameters such as packet error rates.

Strip Charts

During some of the first cable modem deployment activities the need for a simple graphical view of performance became evident. In response to this need a single packet error rate parameter was chosen and charted as a time series over a 1-week period. Packet error rates are charted on a per CMTS upstream receiver port basis with data points taken every few minutes[2]. Figure 1 is an example.

Strip charts provide an intuitive graphical display of performance but they still require a human to do the examination and interpretation. The data is taken and the resulting graphs are created automatically, but no limit checking is done nor are out of

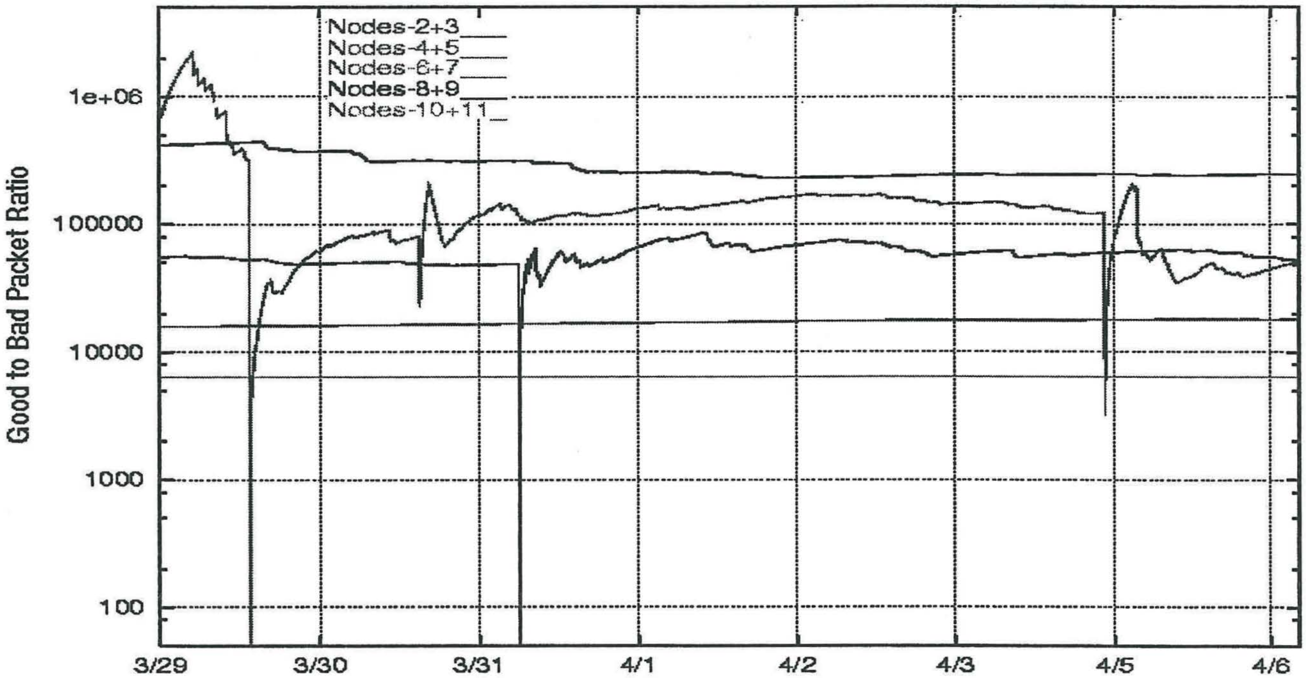


Figure 1 - Strip Chart Example

limit condition alerts generated. Human interpretation and action is still needed.

Perftest

@Home Network engineers have developed a cable modem installation performance acceptance test specifically to evaluate the HFC physical layer. It is called Perftest and measures data transfer speed between the memory of the subscriber's PC and the memory of the local proxy server. This eliminates any disk transfer speed effects as well as limiting the scope of the test to the only part of the network that really matters during the installation.

Because every CMTS in the @Home Network has high speed, local access to a proxy server, testing upstream and downstream data transfer rates between the subscriber's PC memory and the memory of the local proxy server gives a reliable and repeatable result. Figure 2 illustrates the concept.

Perftest reports to the installer the average result of 3 twenty-second test runs in each direction, the standard deviation of these results as well as the number of retransmissions experienced. The results are also logged to a test history file on the proxy server. Perftest can be executed remotely by network operations center technicians. But as with strip charts, no automatic evaluation or condition alerts are generated without human interaction.

Modem Levels

Much can be learned from examining two parameters: the downstream signal level as reported by the modem at its input port, and the level the modem is reporting that it is transmitting back upstream.

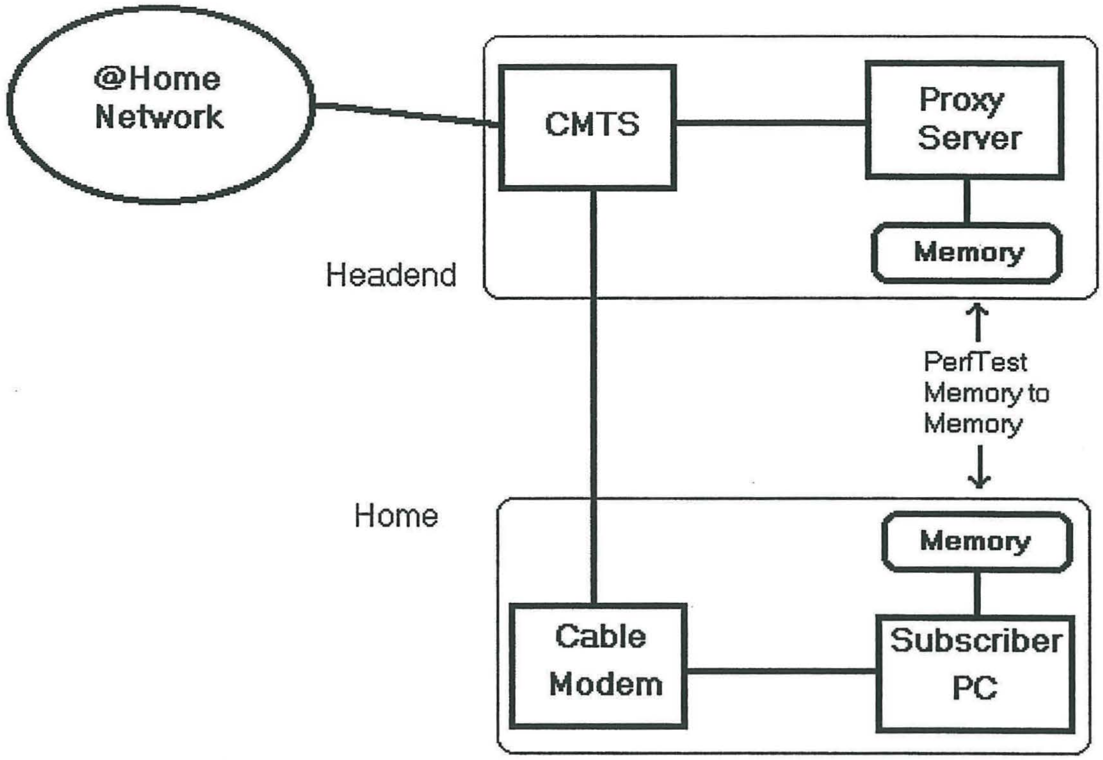


Figure 2 – Perftest Concept

If the modem is reporting that the downstream signal it is receiving is not optimum, it can be inferred that errors are more likely than if the signal was at an optimum level. If errors are more likely then error correction mechanisms such as forward error correction and TCP retransmissions are likely to be in use. Requests for retransmission will reduce available bandwidth between the subscriber PC and the CMTS. Thus, it is desirable from a HFC link point of view to have the downstream received level be solidly in the optimum range.

All modern cable modem systems utilize the "long loop ALC (automatic level control)" technique to manage cable modem upstream transmit levels. This is where the CMTS commands each individual modem to adjust its transmit level, up or down, such that the level received at the CMTS upstream receiver port becomes optimum[3].

If the modem is instructed to reduce its transmit power to a level below about +35 dBmV, experience has shown that the signal to noise becomes less than desired. This can also indicate a plant setup and/or balance problem.

A modem being instructed to raise its transmit power to a level above that which it is capable is also an undesirable condition. This means that the signal to noise ratio at the CMTS upstream receiver port is very likely to be poor as well as in other parts of the plant.

Recently an unscientific survey was done of several hundred modems at various locations around the United States looking for modems that were either transmitting or receiving at levels outside of the manufacturer's specified optimum range. A small number, 3 - 5% were found to have their downstream receive levels outside of the specified range, and a larger number, about 20%, were transmitting above or below the specified range or were even "railed out", transmitting at their maximum power. The vast majority of modem transmit level variations were on the high side. This sampling included 3 MSO's and 3 proprietary technologies. Results across these variables were consistent. No correlations were found to either MSO or technology. If this sampling is truly representative of the general case, then this is a concern and implies some HFC link bandwidth is being used for retransmissions and is thus not available for normal service or new subscribers.

Other Digital Parameters

There are approximately 33 DOCSIS MIB parameters describing various error rates of bytes, packets, frames and synchronization as well as others for proprietary products[4]. How these can be exploited individually or in combinations has yet to be fully explored.

IMPLICATIONS OF DOCSIS

For the DOCSIS market to achieve its full potential, cable modems must be self-installed with automatic provisioning. Ideally then, the optimum RF and MAC MIB parameters would have been identified and would then be polled and checked automatically as part of the DOCSIS installation.

The drive toward self-provisioning clouds the picture for determining where installer tools are going. Ideally, there would be no truck rolls, but when one is necessary, it is desirable for the technician on site to have tools to quickly locate and fix the fault and move on. This implies a box into which the modem drop cable is connected that would quickly evaluate and predict exact modem performance.

CONCLUSION

The current state of the art of evaluation and monitoring of the HFC physical layer can be made better. Just as with the higher layers of the network, automatic monitoring tools with limit checking of modem levels with one or more data transmission quality parameters are needed for the HFC layer.

The tools now available such as strip charts, Perftest, and manual monitoring of modem levels show promise. However, much progress is needed in making these functions as convenient, easy to use and as automated as possible.

The following stages are suggested going forward: 1)achieve modem level convergence to optimum for both downstream as well as upstream, 2)correlate modem level trap thresholds with actual data throughput limits for both downstream and upstream, 3)determine which DOCSIS MIB parameters, individually or in combinations, give the most useful information and create automatic systems to monitor them such as described by Sandino and Kim[5].

These steps would also lay the groundwork for "digital certification" of nodes. Not only could nodes be more automatically tested for certification but the same MIB variables could then be used for ongoing monitoring to assure permanent fitness of the node for data carriage.

Initially, though, it is believed that the biggest "bang for the buck" can be achieved through just step 1 above, the convergence of modem levels to a target optimum level, especially modem transmit levels. This suspicion stems from a year of experience assisting technicians in troubleshooting cable modem installations.

REFERENCES

[1] Thomas K. Fong, Bill Kostka, and John Chen, An Experimental Study of the Return-Path Performance of a DOCSIS System, 1999 NCTA Technical Papers. NCTA, Washington, D.C.

[2] Dan Mckernan, private conversation, Redwood City, CA, April 6, 1999.

[3] Walter Ciciora, James Farmer, and David Large, Modern Cable Television Technology, San Francisco: Morgan Kaufmann Publishers, Inc., pp. 584-86.

[4] Data-Over-Cable Service Interface Specifications: Radio Frequency Interface Specification, SP-RFI-104-980724, Interim Specification, Cable Television Laboratories, Louisville, CO, July 24, 1998, <http:\\www.cablemodem.com>

[5] Esteban E. Sandino and Albert J. Kim, A Structured Approach for the Collection and Analysis of HFC Service Network Statistics Using Advanced Network Terminals, Proceedings Manual, 1999 Conference on Emerging Technologies. SCTE, Exton, PA, January 1999.

MEANINGFUL METRICS FOR RETURN PATH MONITORING

Bill Morgan and Craig Chamberlain
Hewlett-Packard

Abstract

The deployment of advanced data services and the decrease in homes passed per node in order to accommodate more data traffic is putting more pressure on developing a cost effective solution to return path monitoring. This paper will discuss some of the efforts made to date, identify metrics that are currently available for quantifying the performance of the return path, and discuss several new measurements which support this objective. The goal of this paper is provide quality information to the system operator and enable the delivery of higher performance return path services.

What Have We Learned?

Over the last several years, we have participated in several attempts to monitor and ultimately improve the performance of the return path. Large quantities of data have been gathered during this period providing insight about the types of data that directly contribute to improved return path performance. In collaboration with a software partner and a major customer, a monitoring system was developed which used a standard laboratory spectrum analyzer optimized to collect data on a single node as quickly as possible. The software used multi-levels of alarm detection and provided feedback to the user both graphically and statistically. The goal was to plan maintenance for the network based on this feedback and reduce the number of alarms. This was intended to improve the overall quality of the network.

There were several drawbacks with this approach; one was the cost of the system, per node monitored. There is a general belief that if the system had indeed provided a higher quality network, this cost could have been justified. The second and more severe problem with this approach was the low probability of a scanning spectrum analyzer capturing a burst noise event. This probability has been compared to the probability of capturing meaningful data about a rainstorm by placing a shot glass in a parking lot for an instant. A short period of time provides a very poor indication of the amount of rain falling, but after a long period of time the shot glass will provide a good indication of the average rainfall.

Shot Glass in a Rainstorm

To better visualize this, it is important to understand the process a spectrum analyzer uses to scan a frequency band. Because a spectrum analyzer is a narrowband receiver, at any one instant in time it is unable to look at a frequency span wider than its resolution bandwidth. To monitor a wide span of frequencies, the analyzer tunes to a sequential series of frequencies making a measurement at each step. This series of frequencies can be anywhere from 200 to 600 points depending on the analyzer being used.

If this were the only issue, sweeping 400 points the analyzer will be at each specific frequency point for 0.25% of the time. In reality, there are other overheads that reduce this time even further. When an analyzer finishes a sweep it has to retrace to the start frequency. This time is variable depending on the analyzer, but can be from 10% to 20% of the total sweep time. There is also overhead at each data point while the analyzer deals with phase locks and display issues. These overheads can chew up as much as 50% of the time at each data point. So in the final analysis, an analyzer sweeping 400 points at a rate of 20 mSec (an excellent sweep rate for a modern analyzer) may be spending less than 25 μSec , or about 0.1% of its time at any one frequency. This gets reduced even further as the same analyzer is combined with a switch to monitor multiple nodes. It starts to look like a shot glass in a parking lot, doesn't it?

The end result of this effort, which lasted for over two years, was a hard drive full of spectrum response data and no measurable improvement in the quality of the network. We followed this effort very closely and as a result have attempted to draw conclusions and improve the monitoring approach. As a result of this, collecting large amounts of data as quickly as possible is not the correct way to predict future network performance. In addition, spectrum scan data is especially difficult to interpret / correlate and very time consuming to use as a performance monitoring tool. The meaningful information collected from these initial efforts was long-term trend data from measurements made over days, months and seasons.

What is the Proper Perspective?

Another flaw with current approaches to return path monitoring is the assumption that the network is faulty and the monitoring is responsible for documenting these faults. It has been our experience that a properly aligned return path is typically well behaved and the monitoring effort needs to focus on meaningful metrics and performance trends, not fault documentation. It is absolutely critical to any network monitoring effort that the operator begin with a properly aligned network. It is surprising to us how often this is not the case. We need to be in the mode of proactively planning maintenance on the return path rather than reactively responding to faults.

An Alternate Approach to Monitoring

An alternative approach to monitoring is to establish a user-defined set of baseline performance metrics and monitor against these metrics, identifying when measurable performance degradations occur and enabling scheduled maintenance to resolve problems before they impact customers. Ideally, this needs to be coupled with metrics that are available from the terminal and headend equipment, enabling quick response to catastrophic problems requiring immediate attention.

For the purpose of this discussion, it is convenient to group artifacts which impact the performance of the return path into two broad categories.

Short-term events ($< \approx$ 200 mSec) are typically characterized as burst noise and are good indicators of problems in the network. They are usually effectively handled by the error correction algorithms in the data communications channel and do not directly impact the user, but will degrade data throughput.

Long-term events ($> \approx$ 200 mSec) are typically characterized as broadband noise or interfering carriers and in sufficient level can be catastrophic to the network and completely disrupt communications. Narrowband interfering carriers can be addressed by frequency hoping modems if these are available, but there is no place to hop to when the interference is broadband.

A good monitoring system should provide data on both categories of interference. The goal is to develop measurements that augment, but don't duplicate, data that is already available from the terminal devices and headend equipment. The balance of this paper will discuss several new measurements developed to meet these requirements, including carrier-to-noise on TDMA digital channels (Time Domain Multiple Access - signals in the return path emanating from multiple sources and sharing the same frequency spectrum), burst event counting and a dual dwell frequency table sweep.

Carrier-to-noise (C/N) is not a new measurement, but developing a good carrier-to-noise measurement in the return path requires the ability to accurately measure the channel power of a TDMA digitally modulated carrier. This capability has evolved from the ability to measure digital channel power on continuous carriers. The evolution of these measurements began with the development of the digital channel power measurement that was first available in spectrum analyzers in the mid 1990's.

Measuring Digital Channel Power

The NCTA Recommended Practices define digital channel power as "the power level as measured by a power meter which uses a thermal couple as a transducer. That is ... the average power in the signal, integrated over the actual occupied bandwidth of the signal." The NCTA has provided several methods for making a channel power measurement from a single sample in the center of the channel with a correction for channel bandwidth, to full sampling across the band with integration of results. The integrated measurement is the preferred approach since the shape of the digital channel will not affect the accuracy of the result. It is important to understand how this measurement is made in order to understand its evolution.

Because the spectrum analyzer measurement bandwidth (resolution bandwidth or RBW) can be set narrow relative to the channel bandwidth, it enables the measurement of a single digital channel in the presence of adjacent channels. But because of this narrow bandwidth (see Figure 1), multiple measurements must be made across the bandwidth of the carrier and the results integrated and adjusted to arrive at the accurate channel power.

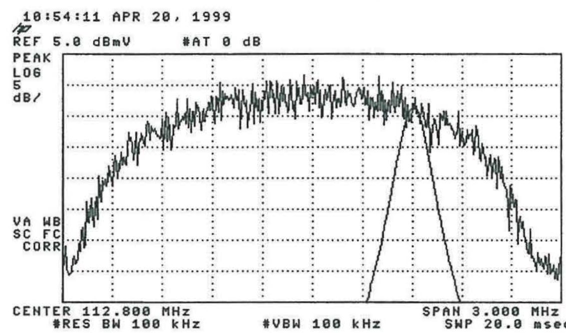


Figure 1 - Digital Channel Spectrum and RBW Filter Response

Using the noise equivalent bandwidth of the spectrum analyzer to define the measurement bandwidth of a single sample, the contribution of each sample to the total power of the channel may be calculated. This allows a simple integration of all the samples to calculate the total power of the channel.

Measuring TDMA Channel Power

The next requirement is to apply the same measurement capability to TDMA digital carriers. Figure 2 is an example of a TDMA digital carrier with a good demonstration of the alternate carrier and noise floor time slots. In order to measure the power of the channel while it is present, a threshold sampling approach is used to capture the TDMA carrier during its ON time. A properly set threshold for sample control will make sure samples are captured only while the carrier is present.

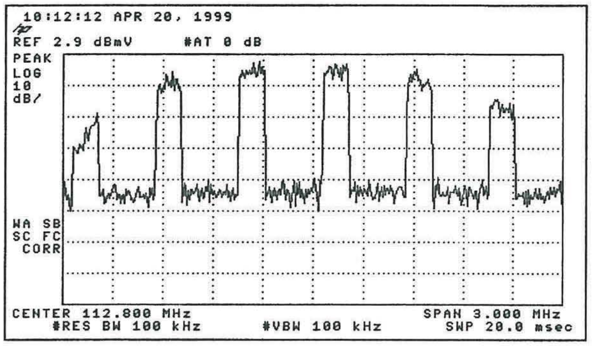


Figure 2 - TDMA Digital Channel Spectrum Response

Once the capability of measuring TDMA carriers using a threshold was developed, it became a natural migration to reverse the sampling algorithm and keep only samples below the threshold. This enabled the measurement of the average noise power in the TDMA channel bandwidth. The combination of these measurements provides the tools required for monitoring C/N in an active return path, providing an extremely flexible measurement for monitoring the performance of a return channel.

Limitations / Variations

There are several limitations to this measurement. The first is that in a TDMA channel slot, multiple transmitter sources are talking and a sampled approach for this measurement captures samples from many of the sources. Therefore, the result is an average of the power from multiple sources. If we assume a properly aligned network these sources which are controlled by a long loop AGC should be operating close to each other in level. The individual transmitter source level is typically available from the headend terminal equipment and this data may be used to monitor individual sources that are operating out of range.

This sampled measurement approach, although extremely accurate, requires a variable amount of time to capture samples, since it is dependent upon how often and for what duty cycle the channel is present. During high traffic time, it may take quite a bit of time to capture enough noise samples, and during late night hours when there is no traffic, it may require a similar length of time to capture channel power samples. But in a system that is designed to deliver long-term trend analysis, this is not a limitation.

In addition to this method, there are several others used for measuring carrier-to-noise in a TDMA environment. A reasonable approximation can be made of TDMA channel power by capturing a single sample and adjusting the result for the channel's bandwidth with a known correction factor. As discussed earlier, the accuracy of this result is affected by the difference in the shape of the channel measured and the shape of the channel used to calculate the correction factor. But it does provide a much faster measurement result since the instrument does not have to wait for the carrier to turn on multiple times. It is also possible to measure the average noise power using a single sample with bandwidth compensation, or measure the noise at a clear area of the spectrum offset from the channel. Both of these are also limited since they do not provide the true integrated noise power under the carrier.

Monitoring carrier-to-noise is an effective way to track longer-term events that affect the return path performance. Small changes in system gain or broadband noise performance show up quickly and can be repaired before network performance is affected. In addition, by measuring each individual channel, the user has the capability of setting different performance parameters on each service dependent on the quality of service required. This tool becomes very powerful as operators are required to provide a higher level of service to different customers (small business, etc.). This methodology has the flexibility to grow as new advanced measurements are defined and implemented.

Addressing Short-term Events

The averaging measurement used to monitor channel power and noise does not do a good job of capturing burst events. To address short-term events, two new measurements have been developed. An additional sampled measurement takes samples at a single frequency and measures the # of occurrences and duration of events that occur above a threshold. Knowing the number and duration of the events over a sample time indicates the "error free" time we could expect from the network. This result may be converted to an estimate of percent availability, a metric familiar to the telecom environment with value for any data communications environment. It is an estimated result since the sampling approach is not capable of capturing every burst transition. Percent availability is an excellent metric to use for trend analysis and also works well for qualifying a new frequency slot prior to the deployment of new services.

Other uses have been conceived for burst event counting. Intentional monitoring of specific operational frequencies can provide information about the use of these frequencies and enable the user to build a picture of usage. For instance, the user will know the approximate timeframe modems "talk" and also the amount of time converters require to request PPV events or to respond to a billing request. This can also be used to give an indication of mode usage and impending bandwidth limitations. In addition, having the duration of the events available provides a valuable troubleshooting tool to the cause of the events.

Another new measurement developed is a multiple dwell table scan measurement which scans a specific frequency table and measures both a short (100 μSec) peak detector dwell result and a long (9 mSec or variable length) peak detector dwell result. The short measurement has a tendency to provide the value of the ambient noise floor and is a reference point. The longer measurement captures the peak level of burst events, since it dwells at the frequency with an armed peak detector for a longer period.

When the two measurements are observed relative to each other, an indication of the increased "noise" from burst events is provided and this can represent an approximation of noise floor to short-term interference ratio. This measurement also gives the user more control over the traditional scanning approach of the spectrum analyzer. By sweeping a frequency table instead of an entire span, the speed of the measurement may be optimized for the needs of the user.

Another use of this measurement is the intentional monitoring of occupied frequencies in the spectrum. This provides rough data indicating if the path has amplitude stability, if the transmitting devices are operating within their designated ranges and if the system has failed during the test period. A lack of signal or even a substantial drop in noise level could indicate an actual failure. Using this information in an alarm window can alert the user to changes in noise and intermittent failure events.

Summary

The combination of these new measurement capabilities may be used in several different modes. Using feedback from the CMTS (Cable Modem Termination System) in the headend, these measurements can be focused on the nodes or channels that have been exhibiting packet or byte errors. This is more of a reactive mode, and is a mode that we prefer to stay out of but improves the efficiency of the monitoring. By making periodic measurements on a schedule, the baseline performance of the nodes may be used as a reference for watching trends. This is a preferred proactive mode that will allow the operator to schedule maintenance as nodes degrade, but before quality of service is impacted. In addition, the frequency of measurements on a node can be adjusted by the relative performance of one node to other nodes.

The data services being deployed in today's CATV network provide several methods of actively monitoring the performance of the return path. It is our goal not to verify or duplicate the results from the terminal equipment, but instead find a way to augment and begin to correlate these results. This can be accomplished by using measurements, which allow the tracking of trends in the performance of the network and predict network problems. This feedback allows maintenance to be scheduled so problems may be fixed before system performance is affected.

Bibliography

Gorin, Joe, Spectrum Analyzer Measurements and Noise - Measuring Noise and Noise-like Digital Communications Signals with a Spectrum Analyzer, Hewlett-Packard Application Note #1303

NCTA Recommended Practices for Measurements on Cable Television Systems, Second Edition, National Cable Television Association, Science and Technology Department

NEW GENERATION NON-COMPRESSED DIGITAL TRANSPORT TECHNOLOGIES ARE CHANGING THE SCOPE OF TODAY'S BROADBAND NETWORKS

John Decker
Product Line Manager – Fiber Optics
Philips Broadband Networks

Abstract

This paper discusses the following trends in Non-Compressed Digital Video Transport:

- *Comparison of current technologies*
- *Interconnects driven by new revenue sources*
- *Economical digital interconnect improves business case*
- *Feature list of new digital technology*
- *Description of new revenue services*
- *Business case structure for interconnects*

DIGITAL TRANSPORT TECHNOLOGY

Modern non-compressed digital transport technology performs the most efficient long-range video transport of any methodology today. For a number of years, non-compressed digital transport has been used to interconnect major markets with many channels of video. The traditional benefits of interconnection include fewer, smaller, simpler headend facilities: fewer costly and unsightly outdoor dish farms; and better use of skilled technical people.

New Revenue Sources for MSOs

With the surging demand for additional revenue-generating services, these interconnect systems are becoming increasingly significant. Operators who rapidly and efficiently introduce these revenue-generating services are able to fend off competition, and command early high rates of return. The consolidation of services in a master headend significantly reduces the amount of equipment and support overhead in each of the original remote headends, as summarized in Table 1.

Thus headend consolidation eliminates all or most of the support overhead expenses associated with each remote headend, savings that fall directly to the bottom line.

MSO Consolidation

The continued trend toward MSO consolidation has enabled the aggregation of adjacent properties that will be most efficiently run through a single master headend.

When the application demands long distance carriage and transparent signal quality, only non-compressed digital technology can deliver the necessary performance.

But this level of quality and performance always carried a price tag, which limited its use to interconnecting major markets, where the cost could be spread over a large subscriber population. Now a new class of non-compressed digital interconnect technology is becoming available which, due to its lower cost, accommodates new applications in an affordable manner. These new systems will provide basic transportation of high quality video by taking full advantage of the newest multiplexing techniques, simplifying the packaging mechanics, and limiting bells and whistles to those features needed to assure reliability, such as status monitoring and path redundancy.

Service	Support Overhead
• Video-on-demand, NVOD	• Servers and support center
• Digital distribution (e.g. HITS)	• Satellite antennae, receivers, headend equipment
• High Speed Data	• Servers, support personnel, user help bank
• Telephony	• Equipment and local support infrastructure
• Commercial insertion	• Equipment and staff
• Video studio services	• Staff, equipment, floor space
• Local venue contribution networks	• Staff, equipment

Table 1.Master Headend Service Consolidation Reduces Equipment and Support Overhead Needed in Original Remote Headends

Trend toward Optical Multiplexing

The new breed of specialized video transport technology can be integrated in a compact package, which occupies limited rack space in already crowded headend facilities. It combines passive multiplexing using Dense Wavelength Division Multiplexing (DWDM) with the traditional Time Domain Multiplexing to carry up to 128 channels on a single fiber.

IF and Baseband

The new breed of digital transport technology carries both baseband video and IF, allowing long distance transport of IF carriers modulated with scrambled video, QAM, or QPSK information.

Channel-by-channel switchable IF or baseband transport provides flexibility to accommodate changing needs of the operator.

Optical Bit Rate and Range Implications

Since most of the multiplexing function can now be accomplished passively in the optical domain, relatively low bit rates can be used while still providing the necessary capacity on a single fiber. All DWDM transmitters operate in the 1550 nm optical window, where fiber dispersion limits distance between active repeaters. Higher bit rates yield shorter range. The effects of chromatic dispersion diminish the effective range of digital systems inversely proportional to the optical bit rate. As a result of this effect, SONET systems operating at 2.4 GB/s for the transportation of video typically have a range under 90km. With the lower modulation rates (typically 1.5 Gbps), range can be extended to hundreds of km with optical amplifiers before a repeater is needed.

Less Repeaters - Longer Distances

Long distance transport, and the expansion of distances between repeaters mean there will be less time and expense in establishing sites for repeaters.
This becomes especially critical when long distances are needed to close the return loop for the path redundant, self-healing ring.

Unlimited Expansion

Since traditional clock jitter problems have been overcome, it is possible to repeat and regenerate the digital signal for thousands of kilometers. This provides an open architecture for the future that would support transmitting some or all of the channels in an interstate network.

Data Transport

The advent of economical optical multiplexing allows integration of dissimilar technologies and data formats to coexist on the same fiber. Data transport technologies and formats are evolving rapidly, fueled by the tremendous decentralized demand created by the Internet. Demand for cable modem data, telephony, and other forms of digitized information are growing exponentially. Few people predicted the pace of Internet demand, as content providers rapidly emanated from decentralized sources.

As a result, data transport formats and equipment will continue to ride the crest of a well-funded development wave for several more years to come. Optical integration of the data carriage assures that the CATV operator will be able to avail himself of the most advanced and cost-effective data transport technologies, using Dense Wave Division Multiplexing, rather than Time Division multiplexing. Data services from various third party sources are integrated via selecting a compatible ITU wavelength laser. A multi-data format product is under development using the same modular optical and element management components as the digital video transport products.

APPLICATIONS FOR NON-COMPRESSED DIGITAL

It is easy to identify many otherwise difficult problems that non-compressed digital technology can solve. Here are some examples of applications for which this transparent transport technology can be useful.

Wide Area Interconnect

Headend facilities can be consolidated in regions that span hundreds of kilometers. As new revenue services appear, the economic critical mass is often inadequate to sustain the necessary equipment, facility, and support staff investment. An economical interconnect system can aggregate a sufficient number of subscribers to make the venture profitable. As a result, only fiber availability, not the cost of electronics, will drive the economics. Operators of classic systems can now consider the benefits of interconnection and headend consolidation.

Supertrunk or Interconnect over Telecommunications Fiber

It is generally not advisable to carry analog (CATV) and binary telecommunications (e.g., SONET) signals on different DWDM wavelengths on a single fiber. Yet there may be opportunities which are best served by transporting video on wavelengths leased from a telecommunications provider. Since non-compressed digital video is fully compatible with existing DWDM telecommunications networks, the technology can be used directly to transport video on unused wavelengths in these networks. In contrast, analog transmission over a leased wavelength is not at all straightforward, and will in many cases be impossible for reasons such as intolerably large (nonlinear) crosstalk effects. Apart from its ease of integration, non-compressed digital video is also the most cost-effective solution for such applications.

Lease Wavelengths to Others for Video Transport

Perhaps your company or an affiliate has telecommunications fiber serving business parks and large enterprises. Consider selling high quality, dedicated video circuits using DWDM wavelengths on the same fiber that carries telecom services. Simply use non-compressed digital equipment to encode and decode the video signals. This approach can generate revenue from service to sport venues, university and medical campuses, and entertainment and advertising video production centers. It allows telecommunications services and many channels of high quality video to be carried over the same fiber strand.

Localized Analog Programming to Complement Analog Program Spectrum

Operators strive to provide a programming lineup tailored to each community. This is difficult to do with conventional supertrunking techniques, which carry the same bulk analog spectrum to each hub. While it is feasible to transport targeted analog services in separate fibers and inject them into the channel lineup at appropriate frequencies in each hub, this is not a flexible approach. Non-compressed digital equipment delivers neighborhood-targeted programming to hubs in baseband or IF form, where a conventional modulator creates a clean RF channel at the proper frequency. Consider this as a method for "supertrunk bypass" to customize the channel lineup for neighborhoods.

Zoned Local Advertising

Local ad insertion is an important revenue stream. In major markets, this revenue can be enhanced by selling available advertising timeslots on a zoned basis -- a special case of the targeted programming application. The top-20 ad revenue channels will be in the analog domain for many years, of course, since advertisers buy into the most widely viewed programming. What is the best way to deliver customized versions of popular programming like ESPN to each neighborhood?

One way is to locate the ad insertion equipment in each hub, but this requires space in the hub for the insertion equipment and a baseband signal which supertrunk transport techniques do not normally provide. A better alternative is to perform the ad insertion at the headend or primary hub to create multiple versions of the top local ad channels and deliver these to hubs using the non-compressed digital supertrunk bypass system described above.

Or, if you prefer to keep the insertion equipment in each hub, use the non-compressed digital feed to provide the network signal. This way it arrives in baseband form as required by the insertion equipment.

PEG and Broadcaster Program Backhaul

Any interconnected region by definition serves a number of communities. The Public/Education/Government (PEG) programming is different for each community, and often originates at locations that are distant from a cable headend or hub (e.g., city hall). This content is created in baseband video form, so non-compressed digital transmission technology is the natural choice to carry PEG programming to cable facilities. Similarly, local over-the-air broadcasters use this technology to supply a pristine version of their broadcast signal for cable distribution.

This is just a sampling of applications for non-compressed video transport systems. As new models come to market, there are many choices which complement traditional video transport techniques and enhance the economic opportunities for cable operators.

ECONOMIC ANALYSIS

The business case for headend interconnects is critical to the decisions of how soon and how widely to establish interconnect systems. Recent advances in digital and DWDM technologies make a compelling case toward establishing super-headends with economical distribution of a full set of revenue services in a primarily optical domain all the way from the super-headend to the node.

Speed-to-Market for New Revenue Services

Time-to-market studies show that more rapid deployment of such revenue services provides the largest market capture rate and the earliest return on investment. Time-to-market is facilitated by rapid scaleable use of existing facilities, and avoidance of new brick and mortar.

Coupled with a DWDM hub distribution architecture, a high bandwidth forward and return system for each of these new services could rapidly be deployed using only existing space and facilities.

Some markets, typically outside of dense metropolitan areas, are candidates for interconnects to add additional services. Areas of lower population density, with smaller headends often do not have enough subscribers to support the business cast for the investment in additional services. Rural cable system owners are anxious to enter into the market for added revenue services such as high speed data, VOD, and cable telephony. Until recently the cost barriers to provide the interconnect were prohibitive. The introduction of more cost effective interconnects, has lowered the barrier to entry. A phased deployment of a multiple headend interconnect might proceed as shown in Table 2. Services following this could be telephony and true video on demand.

A profit-loss analysis (P&L) will demonstrate the economic viability of an interconnect. The P&L should be estimated over periods of one, two, and three years for each of these new service offerings:

- Internet Services
- NVOD
- VOD
- Telephony
- Commercial Insertion
- Venue contribution
- Studio Services
- Leased Transport

The below profit and loss analysis should be conducted for each of the service offerings.

- Added Revenue
- Cost Savings Interconnect
  *Reduced Staff
  *Reduced Buildings
  *Reduced Taxes
  *Reduced utilities
  *Reduced Operating Expenses
  *Elimination of remote headend equipment such as satellite dishes and receivers
- Cost Avoidance Interconnect
  *Obsolete Equipment Replacement
  *Duplication of Equipment in Remove Headends
- Total Revenue and Cost Savings
- Cost of Interconnect Equip & Installation
- Net Gain (Loss)

Phase 1	Installation of interconnect equipment at headends, (fiber in place).
Phase 2	Installation of NVOD servers
Phase 3	Installation of Cable Modem server
Phase 4	Initial NVOD revenue
Phase 5	Initial high speed data revenue

Table 2. Example of a Multiple Headend Interconnect Phased Deployment

Options for High-Efficiency HFC Return Architectures

Eric Schweitzer, PhD
Harmonic Inc.

Abstract

The increasing deployment of Internet access, video-on-demand and other two-way services requires operators to use the most efficient means possible to maximize the limited bandwidth available in most HFC systems. A network is only as strong as its weakest link, which makes the return-path portion of a system critical when preparing for the delivery of two-way services.

The issue of return bandwidth becomes even more crucial in networks with minimal fiber counts. Because the return bandwidth is limited, it is inefficient to dedicate a single fiber to each return. Architectures which address these issues can make use of technologies such as Wave Division Multiplexing (WDM), block up conversion (frequency stacking), and digital transmission.

INTRODUCTION

Providing internet access, video-on-demand, and other two-way services presents a great opportunity as well as significant challenges to system operators. Perhaps the greatest challenge is in the return path.

Historically, the return path of a cable system has not been heavily used. Therefore, the bandwidth available is limited. Also, ingress noise is a significant problem in the frequency band of the return path. Finally, because until recently the return was not heavily used, there is limited experience with the return path in the industry.

The bandwidth limitations and ingress problems can be addressed by segmentation. Limiting the number of homes on each segment of the return system increases the available bandwidth per home and limits the amount of ingress the return system must handle.

The fiber portion of a typical Hybrid Fiber/Coaxial (HFC) cable network is in a star topology — there is a direct connection with multiple fiber from either the head end or a hub to each optical node. The coaxial plant following the optical node is a tree and branch topology. It is difficult, if not impossible, to segment a tree and branch network, while it is relatively easy to segment a star network.

Therefore, the optical node is the deepest point in the network at which segmentation can be achieved relatively easily. Furthermore, nodes with multiple ports can be segmented by isolating the ports. This allows placing unique information on each port, segmenting the network even further.

This segmentation creates a network with many dedicated return links. Finding an efficient and cost-effective architecture to transmit the information on the multitude of return links presents a challenge. There are several technologies which can be applied to create many possible network architectures. This paper will explore four technologies as they are used to create ten network architectures.

TECHNOLOGIES

Multiple Fibers

The simplest technology for transmitting multiple returns is to dedicate a separate return transmitter, return receiver, and fiber for each. As the demands on the performance of the link are modest, low-cost transmitters using Fabry-Perot or uncooled DFB lasers can be used.

However, dedicating a fiber to each return path does not use fibers efficiently. Each fiber is being used to transmit only a small fraction of its total bandwidth capacity. This is not a problem if extra fibers are available.

Block Up Conversion

Each return link uses only a modest amount of bandwidth. One way to use fibers more efficiently is to block up convert the return path to use bandwidth in the higher frequency range. Multiple blocks can be stacked together and transmitted on a single fiber using a single transmitter and receiver.

This block up conversion must be done carefully. It is very easy to introduce unacceptably high levels of phase noise and group delay. Allowing sufficiently large guard bands between blocks makes the design of the required filters much less complex. Therefore, these transmitters must have a bandwidth extending to at least 870 MHz to allow a sufficient number of blocks on a single fiber.

A reasonable scenario is ten return blocks on a fiber link using the band from 45 to 870 MHz. This heavy loading of the optical link requires that higher-performance and extended bandwidth transmitters and receivers be used. These have significantly higher cost than the transmitters and receivers required to support only a single return block.

It is also difficult to design a block up converter with acceptable dynamic range and frequency stability. This analysis assumes that block up converters with acceptable performance will be available for a cost similar to block up converters used for cellular phone applications.

Wavelength Division Multiplexing

By using transmitters with defined wavelengths and the appropriate filters, a single fiber can transmit several wavelengths each carrying different information. The maximum number of independent wavelengths depends upon how well the wavelength of each source can be controlled and on the filters used to combine and separate the light.

The simplest form of Wavelength Division Multiplexing (WDM) is to use a single fiber to transmit two wavelengths — one in the window near 1310 nm and one in the window near 1550 nm. This form of WDM can be called Coarse Wavelength Division Multiplexing or CWDM.

CWDM does not require accurate control of the wavelength of the laser within the transmitter. Therefore, transmitters used for CWDM can use low-cost Fabry-Perot or uncooled DFB lasers. Also, couplers to combine and separate these two wavelengths are readily available at modest cost.

The next level of WDM would be to use several wavelengths near 1550 nm together with 1310 nm. Uncooled DFB lasers with three different wavelengths near 1550 nm are commercially available. I will call this technology Medium Wavelength Division Multiplexing (MWDM). These uncooled

DFB lasers for MDWM are slightly more costly than similar lasers for CWDM.

The uncooled DFB lasers used for MWDM will vary by up to 15 nm over the operating temperature range of a node. Therefore, the wavelengths used must be separated by at least 15 nm. Couplers to combine and separate these wavelengths are more difficult to build, but readily available at modest cost.

The finest level of WDM involves using multiple sources near 1550 nm on a single fiber, each with accurately controlled wavelengths. This is frequently designated Dense Wavelength Division Multiplexing or DWDM. The International Telecommunications Union (ITU) has defined a standard set of DWDM wavelengths for communications applications.

Laser sources on the ITU wavelength grid are readily available. Couplers to combine and separate these wavelengths are also readily available. However, to accurately control the wavelength of the source and response of the filters in the couplers is complex. These devices are fairly expensive today, although costs are decreasing rapidly.

Current technology can achieve acceptable analog[1] performance with eight wavelengths per fiber. This is the assumption used in this analysis. It is clear that this will be increased to 16 and 32 wavelengths in the future, making this option more attractive.

Digital

The most basic digital return is to digitize the analog waveform. Achieving acceptable dynamic range requires an eight or 10 bit analog to digital converter and a sampling rate near 100 MHz. Therefore, a 1.8 G bit per second digital link can transmit two independent returns.

Digital transmission does not require high performance. Therefore, low-cost transmitters can be used over long distances. In addition, digital DWDM lasers are readily available, allowing these two technologies to be combined.

SINGLE-LINK NETWORK ARCHITECTURES

Networks in which there is a single optical link do not require cascaded optical links to transmit information from the node back to the location at which it is processed. There are two common network architectures for which this is true.

In one architecture there is a direct fiber connection from each node to the head end. In the other architecture the hub contains most, not all, of the equipment used to support the dedicated services. An example of the latter architecture in a network supporting cable modem communications would have the Cable Modem Termination System (CMTS) and connection to the internet in the hub.

[1] In this context, analog includes schemes for transmitting digital information on analog carriers. Examples of this are Frequency Shift Keying (FSK), Quadurature Phase Shift Keying (QPSK), and Quadurature Amplitude Modulation (QAM).

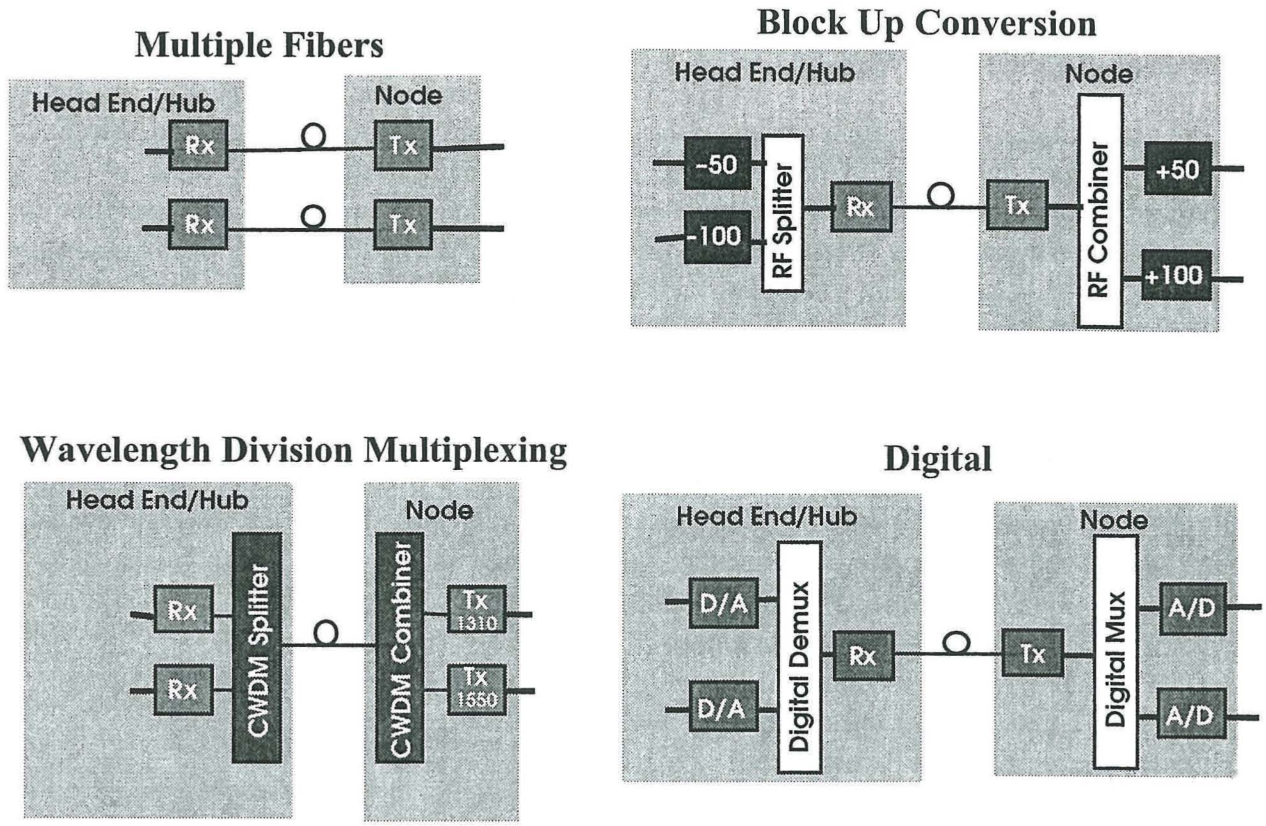


Figure 1 — Block diagrams of four architectures in which there is a direct connection from the node to the site at which the dedicated information is processed.

	Relative Cost		
Architecture/Technology	One Return per Node	Two Returns per Node	Four Returns per Node
Multiple Fibers	1.00	2.00	4.00
Block Up Conversion		8.82	13.77
Wavelength Division Multiplexing		3.07	5.14
Digital		9.56	19.12

Table 1 — Comparison of the cost of the four architectures shown in **Figure 1**. Costs for one, two and four separate returns per node are shown. A single return is a trivial case using only a single fiber. Using any of the more complex technologies has no benefit and is, therefore, not shown. The costs are estimates and include only the equipment. The cost of fiber, patch panels, etc. is not included.

	Fibers Needed		
Architecture/Technology	One Return per Node	Two Returns per Node	Four Returns per Node
Multiple Fibers	1	2	4
Block Up Conversion		1	1
Wavelength Division Multiplexing		1	1
Digital		1	2

Table 2 — Comparison of the fiber requirements of the four architectures shown in **Figure 1**.

There are many feasible network architectures which use the technologies mentioned previously. Four of these will be addressed as being typical. Block diagrams of these four architectures are shown in **Figure 1**. To save space, the figure shows only the scenario with two returns per node. It is straight forward to extrapolate these block diagrams to a single return and four returns.

A table comparing the approximate costs of the four architectures is shown in **Table 1** and the fiber requirements of the four are shown in **Table 2**. The costs are estimates and include only the equipment — transmitters, receivers, bock converters, couplers, platforms, etc. The cost of fiber, patch panels, etc. is not included.

The costs and fiber requirements are tabulated for a single return from each node, two returns from each node, and four returns from each node. This allows comparisons for future scaling.

A single return from each node is an elementary scenario: Only a single fiber is required and applying any of the technologies discussed adds only cost and no functionality. It only makes sense to consider a single link with a return uncooled DFB laser transmitter and a return receiver. This is used as the baseline for normalizing the costs.

Multiple Fibers

Using multiple fibers for each return from each node can be done easily today by using a node designed for this. The costs in **Table 1** are based on using uncooled DFB lasers. If the poorer performance of Fabry-Perot lasers can be tolerated, the costs would be lower.

Block Up Conversion

Block up conversion requires placing the up converters in the node and corresponding down converters at the receiving site. The costs associated with the increased performance demands on the optical equipment are included in the analysis.

Wave Division Multiplexing

CWDM can be used for two returns from each node while MWDM can be used for four returns from each node. Again, the costs are based on using uncooled DFB lasers which will have acceptable performance for this application. If the lower performance of Fabry-Perot lasers can be tolerated, the costs would be lower.

Digital

Using a digital return from the node enables lower-cost optical components optimized for digital transmission to be used. This is offset by the extra cost of the Analog to Digital converters (A/D) digital multiplexers required. Also, due to the high data rate, digital is limited to two return blocks per wavelength.

DUAL-LINK NETWORK ARCHITECTURES

The most commonly-used network architecture is one where there are several hubs served from the head end with multiple nodes served from each hub. This architecture requires two cascaded optical links — one from the node to the hub and another from the hub to the head end.

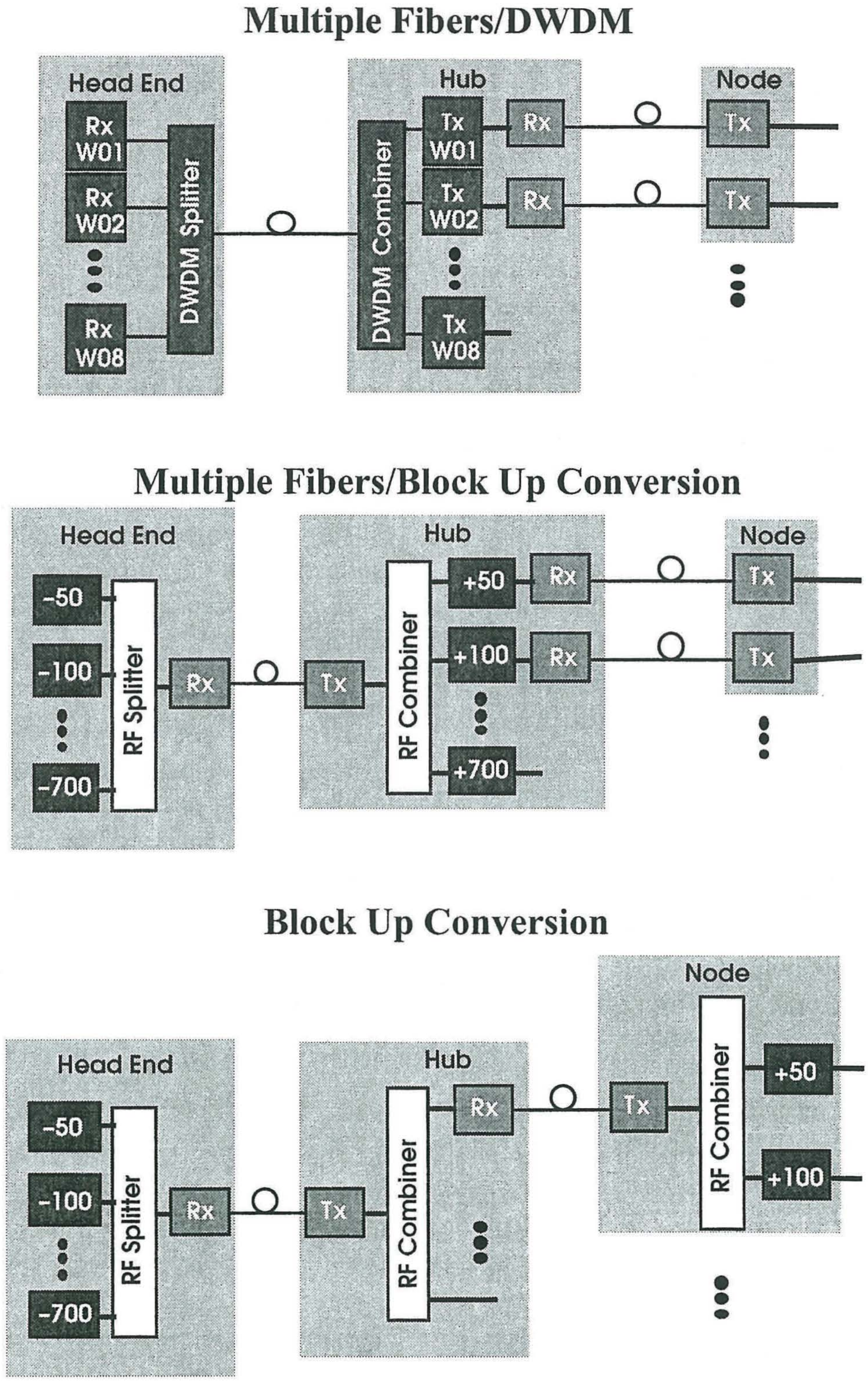
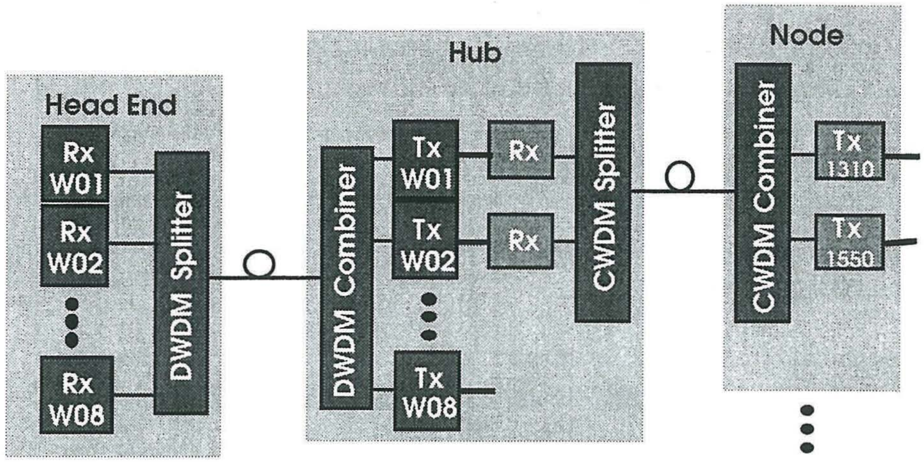
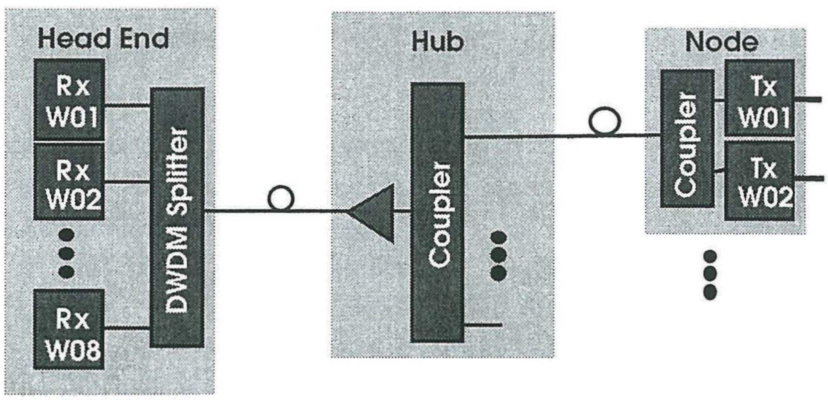


Figure 2 — Block diagrams of three architectures which require cascaded optical links. Block diagrams of the other three which are discussed are shown in **Figure 3**.

Wavelength Division Multiplexing



Dense Wavelength Division Multiplexing



Digital

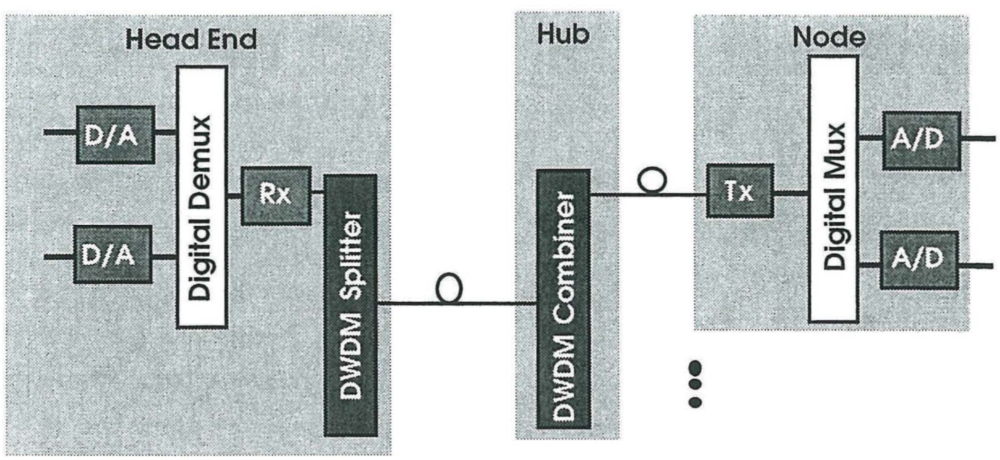


Figure 3 — Block diagrams of three architectures which require cascaded optical links. Block diagrams of the other three which are discussed are shown in **Figure 2**.

The fact that there are two links opens up may options for applying the technologies mentioned previously. Six options will be addressed. Block diagrams of these six architectures are shown in **Figure 2** and **Figure 3**. Again, to save space the figure shows only the scenario with two returns per node. Extrapolating to one and four returns per node should be straight-forward.

The relative costs of the six architectures are shown in **Table 3**. Again, estimated costs for a single return, two returns, and four returns from each node are listed. The costs include only the equipment

As it is possible and desirable to consolidate returns in the hub; there is no elementary scenario to which the others can be normalized. The architecture with a single return from each node and which uses DWDM from the hub to the head end was arbitrarily chosen as the reference.

The number of fibers required from the node to the hub for each architecture are shown in **Table 4** while the number of nodes which can share each fiber from the hub to the head end for each architecture are shown in **Table 5**. Both of these tables include a single return, two returns, and four returns from each node.

Multiple Fibers/DWDM

In this architecture, a dedicated fiber is used for each separate return from the node. DWDM transmitters at the hub are used to consolidate eight returns onto a single fiber. The costs for this architecture are based on using transmitters with uncooled DFB lasers in the node.

Multiple Fibers/Block Up Conversion

This architecture is very similar to the previous. The difference is that block up conversion between the hub and head end is used instead of DWDM.

Block Up Conversion

This architecture places the block up converters in the node. Converters in different nodes can be assigned different frequency blocks allowing direct combining of 10 return blocks in the hub.

These blocks can be combined onto a single transmitter/fiber in the node for the link to he hub. However, this does require return equipment with an extended bandwidth in the node. The cost of this is included in the analysis.

Wavelength Division Multiplexing

This architecture uses some form of WDM for both links. DWDM is used from the hub to the head end in all three scenarios.

The technology used for the links from the node to the hub changes with each scenario. In the scenario when there is a single return from each node, this architecture is identical to using Multiple Fibers/DWDM. When there are two returns from each node CDWM is used is used for these links. MWDM is used when there are four returns from each node.

Architecture/Technology		Relative Cost		
		One Return	Two Returns	Four Returns
Node to Hub	Hub to Head End	per Node	per Node	per Node
Multiple Fibers	DWDM	1.00	2.00	4.00
Multiple Fibers	Block Up Conversion	0.70	1.40	2.80
Block Up Conversion		1.17	1.67	2.68
Wavelength Division Multiplexing		1.00	2.10	4.11
Dense Wavelength Division Multiplexing		0.69	1.40	2.79
Digital		1.18	1.87	3.74

Table 3 — Relative cost of six architectures which use two cascaded optical links. These are diagrammed in **Figures 2** and **3**. Estimates for one, two and three separate returns from each node to the head end are shown.

Architecture/Technology		Fibers Needed from Node to Hub		
		One Return	Two Returns	Four Returns
Node to Hub	Hub to Head End	per Node	per Node	per Node
Multiple Fibers	DWDM	1	2	4
Multiple Fibers	Block Up Conversion	1	2	4
Block Up Conversion		1	1	1
Wavelength Division Multiplexing		1	1	1
Dense Wavelength Division Multiplexing		1	1	1
Digital		1	1	2

Table 4 — Number of fibers needed between the node and the hub for the six architectures which use two cascaded optical links. These are diagrammed in **Figures 2** and **3**.

Architecture/Technology		Nodes per Fiber from Hub to Head End		
		One Return	Two Returns	Four Returns
Node to Hub	Hub to Head End	per Node	per Node	per Node
Multiple Fibers	DWDM	8	4	2
Multiple Fibers	Block Up Conversion	10	5	2.5
Block Up Conversion		10	5	2.5
Wavelength Division Multiplexing		8	4	2
Dense Wavelength Division Multiplexing		8	4	2
Digital		8	8	4

Table 5 — Number nodes per fiber for the link between the node and the hub for the six architectures which use two cascaded optical links. These are diagrammed in **Figures 2** and **3**.

Dense Wavelength Division Multiplexing

This architecture uses DWDM transmitters in the node. The wavelengths are chosen such that eight returns from multiple nodes can be combined. If there is more than one transmitter in a node, the output from the transmitters is combined onto a single fiber within the node. Fibers from multiple nodes can be combined in the hub.

There are both cost and performance benefits to using couplers, rather than a DWDM multiplexer, to combine the light from multiple transmitters. However, a DWDM demultiplexer must be used to separate the light at the receiving site.

Link losses in this architecture can be high. This requires that there be an Erbium Doped Fiber Amplifier (EDFA) in the hub. This

EDFA can have modest power and performance compared to those used for forward path applications. Therefore, the cost of this EDFA is also modest.

Digital

This architecture uses digital DWDM transmitters in the node. The wavelengths are chosen so that eight returns can be combined onto a single fiber in the hub.

As in DWDM, a fiber for each DWDM transmitter is used between the node an the hub. There is sufficient data capacity on each fiber for up to two separate returns. A single return from each node would use only half the data capacity.

CONCLUSION

Providing internet access, video-on-demand, and other two-way services will require one or more dedicated return links from each node. There are numerous architectures using any of several technologies which can be used to support this.

In networks without cascaded optical links, the cost of using extra fibers for multiple returns is always lowest provided the fibers are available. If the fiber are not available, using some form of wavelength division multiplexing is probably less expensive than installing new fibers and seems to be the least expensive option.

In networks with cascaded optical links, several architectures have lower cost than using multiple fibers. Using DWDM from the node has nearly the same cost as using block up conversion — either in the node or in the hub. The least expensive option varies with the number of returns from each node, but the differences are small in all three scenarios.

However, block up conversion places a significant quantity of equipment in the hub. At the very least, a set of receivers and transmitters is required. This requires larger hub sites and makes maintenance and repair more difficult. In contrast, DWDM from the node requires only a low-cost EDFA in the hub. This makes the hub completely optical — no RF equipment is required in the hub.

While today the performance and the cost of DWDM from the node and block up conversion in the hub are very similar, there are significant operational advantages of DWDM. Also, it is clear that the cost of DWDM equipment will be decreasing in the future. These factors make DWDM from the node the best option for a high-efficiency return architecture.

Performance Measures for Cable Data Transmission

Richard S. Prodan, Dan Rice, and Majid Chelehmal
Cable Television Laboratories

ABSTRACT

Different services require different error performance and different data rates. For example, a service such as status monitoring with polling may be able to tolerate more errors and lower transmission speeds than a time-critical application such as video telephony or "twitch" video games where re-transmission of errored frames cannot be tolerated. Each service will require different operating characteristics for satisfactory performance.

Three important indicators used to evaluate the performance of modems in a cable television network, in the presence of continuous or impulsive noise, are carrier-to-noise ratio (CNR), frame-loss-ratio (FLR), and bandwidth efficiency (BWE). These three parameters are related to each other and are bounded by the channel capacity limit of Shannon. Actual cable modem performance in different operating modes can be described parametrically using measurements in these three indicated dimensions. When the value of CNR in the transmission channel is known, the value of BWE at an arbitrarily small FLR can be deduced. The resulting operating point can be compared to minimum service requirements.

This paper describes how these useful performance indicators can be used to evaluate and to compare the performance of different cable modem transmission system modes in the presence of continuous or impulsive noise. Determination of satisfactory regions of operation under various transmission channel conditions is discussed. An example of the use of the described performance indicators for determining satisfactory service delivery is provided.

INTRODUCTION

Accessing information in an efficient and timely manner has risen to the forefront of visibility not only in high tech circles, but also in the everyday lives of people around the world. The cable modem has become one of highest profile communication appliances in the race to provide high-speed data communications. This is a result of Internet growth and the development of the interoperable CableLabs® Certified™ cable modem, the cable modem compliant with the data over cable system interface specification (DOCSIS) by the Multimedia Cable Network System consortium (MCNS), and Cable Television Laboratories, Inc. (CableLabs®). To heighten the understanding of current cable modem technology and its impact on the cable industry, CableLabs has evaluated current modem technologies available for the bi-directional transmission of data through the cable plant. This technique for characterization of modem technology using the three performance indicators, CNR, BWE, FLR has been used at CableLabs to evaluate the performance of prior proprietary modems, the current DOCSIS technology, and possible future extensions to the DOCSIS physical layer.

One of the cable industry's competitive advantages is the wide bandwidth available in the hybrid fiber coax (HFC) network. There are many trade-offs in the best utilization of the bandwidth and power resources in delivering different digital services in a reliable, cost-effective manner. To assist cable service providers in making these trade-offs, evaluations of proprietary commercially available cable modems from several different vendors have been completed. The comparisons presented here between modems, (without identifying the specific manufacturer), are based on laboratory testing performed by CableLabs and theoretical calculations for DOCSIS modems. Some typical measurement results are presented here to demonstrate a technique which can be used to select the modem technology and operating parameters which will meet the requirements of a service based on the quality (lack of transmission impairments) of the HFC network.

PERFORMANCE METRICS

Three important metrics have been used successfully and are recommended to evaluate the performance of modems in a modern HFC cable television network in the presence of continuous or impulsive noise. The metrics are listed below.

Carrier-to-Noise Ratio (CNR): This is the widely used indicator of the noise characteristics of a transmission channel. A metric descriptive of the characteristics of the channel noise can be expressed in terms of additive white Gaussian noise, or of impulsive noise bursts of specified amplitude, duration, and repetition rate. This is an indicator of the performance of the transmission network. Since the signal power is limited in the HFC network by the dynamic range of the active components, the CNR can be improved only by reallocating the power budget, changing the network structure, or improving conditions by proper maintenance. A metric descriptive of the channel noise type and severity in relation to the signal can be expressed in many forms such as:

- Carrier-to-Noise Ratio (CNR);
- Signal-to-Noise Ratio (SNR);
- Energy-per-bit-per-noise power spectral density (Eb/No), i.e., the SNR per bit;
- Carrier to interference (C/I);
- Impulse noise amplitude, duration, and repetition rate statistics.

All of these metrics are useful in their context. This paper will refer to the channel condition with a general metric of CNR because the cable industry is most familiar with this term (and it has been used historically). CNR is typically defined as the ratio of signal power-to-noise power in the channel before demodulation, while SNR is the metric that typically refers to the ratio after demodulation. Eb/No is usually the appropriate metric when comparing modems of different modulation formats, bandwidths, and symbol rates. CNR is an easily measurable metric in the laboratory and the field.

Frame Loss Ratio (FLR): FLR is the ratio of errored data frames with respect to the total number of frames transmitted, when the data frames are transmitted over an impaired channel. This indicator can be applied to a variety of data packet types, including Ethernet-type frames consisting of 64 bytes to 1518 bytes each, but it also can refer to 53-byte frames as used in asynchronous transfer mode (ATM) transmission protocols. The desired accuracy and quality of service that the network should provide for a given service dictates the value of this metric.

Bandwidth Efficiency (BWE): BWE indicates the data capacity that can be transmitted through the channel. It is expressed in terms of the amount of data transmitted per unit of time through the unit of bandwidth (bits/sec/Hz). The value of this indicator is governed by the design of the modem. It should be noted that the transmittable data only includes the useful message data; provision must be made for any overhead needed for forward error correction (FEC) and media access control (MAC) overhead. As overhead is increased, the efficiency is reduced and a lower data rate is transmitted per allocated channel bandwidth. A modem without provision for FEC overhead may have a high bandwidth efficiency, but it will fail rapidly on a noisy transmission channel. A compromise of FLR and BWE must be found case by case for an existing channel condition.

The three indicators defined above are related to each other and are bounded by the law of Shannon. The Shanon-Hartley theorem is expressed as:

$$C = BW \log_2(1+S/N)$$

Where C is the maximum capacity in bits/second for an arbitrarily small error ratio, BW is the equivalent noise bandwidth, S is the signal power, and N is the power of additive white Gaussian noise (AWGN). When the value of CNR is known, the upper limit to the value of BWE = C/BW, for effectively error-free transmission can be computed by applying that theorem.

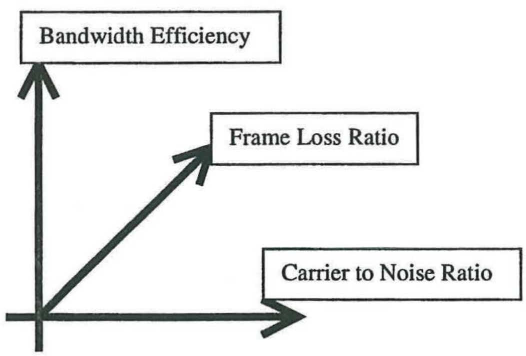


Figure 1: Three-dimensional Modem Performance Space

The values of the three indicators can be plotted on a three-dimensional graph as shown in Figure 1. In practice, cross sections of the three-dimensional graph, perpendicular to the FLR axis, are often used instead for the sake of convenience. The line that represents Shannon's limit also can be plotted on those bi-dimensional cross-section graphs. That line describes the maximum theoretical performance possible for any combination of BWE and CNR at an arbitrarily low FLR approaching zero (with appropriate forward error correction).

The most common type of comparison between modem technologies, which indicates the robustness of the physical (PHY) layer format, is the error rate vs. CNR in an additive AWGN channel. The type of error rate used here is the 64-byte Ethernet FLR. Figure 2 shows the theoretical FLR vs. CNR for a couple of different FEC modes of the DOCSIS 16-QAM-modulation format. In Figure 2, the parameter t indicates the number of bytes the FEC can correct in each codeword. Similar data has been obtained for several commercially available modems and for the other DOCSIS formats.

The transmission robustness in the upstream direction of the HFC network is often determined by the impulsive burst noise characteristics of the channel as opposed to AWGN limitations. Tests also have been performed to characterize a modem's error correction performance in a burst noise channel. A broadband CNR of 0 dB was established during the noise burst. The noise pulse width was set for one of three determined lengths, 1 µs, 10 µs, or 100 µs. The frequency (number of noise pulses per second) was increased and the error rate was recorded along with the noise pulse repetition rate (PRR). Figure 3 shows an example of the interfering signal. Figure 4 shows the result of the impulsive noise test for a 0 dB CNR noise pulse duration of 10 µs.

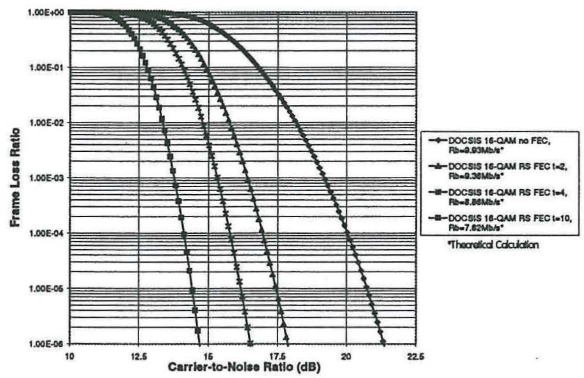


Figure 2: FLR vs. CNR for DOCSIS 16-QAM FEC Modes

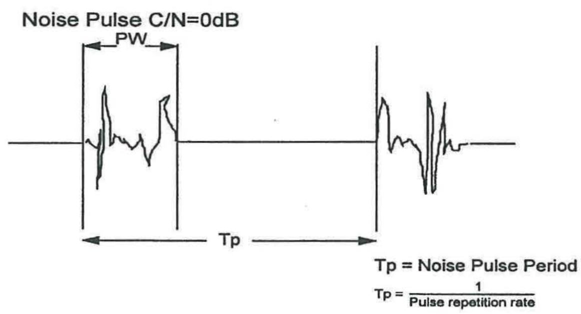


Figure 3: Interfering Burst Noise Signal

BANDWIDTH EFFICIENCY

As stated earlier, the bi-directional bandwidth resource is one of the big infrastructure advantages that cable has over its competitors. Bandwidth efficiency indicates the amount of data, which can be transmitted through a unit of bandwidth per unit of time. In other words, a high bandwidth efficiency makes the best use of the HFC network bandwidth resources. Figure 5 and Figure 6 present data demonstrating the trade-offs made between power robustness and bandwidth efficiency. The data rates used in calculating bandwidth efficiency are corrected for overhead. Graphs of this type can be used to determine what technology will need to be utilized, based on the type of service and the quality of the HFC network. These bandwidth efficiency graphs can be created from the performance data given in a format similar to Figure 2 and Figure 4.

Continuous Noise

Figure 5 is based on an error performance of 1% FLR in an AWGN channel. This type of graph can also be created for other FLR values like FLR=10%. Each point on the graph indicates the bandwidth efficiency and CNR at which the specific modem can achieve a 1% FLR. The bandwidth efficiency for vendor A and DOCSIS 16-QAM is very good, but they require the highest CNR. The CNR for the DOCSIS 16-QAM modem is less than 20 dB, which is reasonable for a return path node. The CNR required for 1% FLR for vendor D is very low, but the sacrifice in bandwidth efficiency is very significant. Increasing the coding depth and robustness typically results in the loss of bandwidth efficiency. Note that if a line is drawn through the points for no FEC, t=2, t=4, and t=10, for the DOCSIS QPSK or 16-QAM modes, it is not linear and starts to fall off quickly. For the AWGN channel half of the possible t=10 coding gain can be achieved by using only t=2. One is always able to get the most out of compromises. Looking at Figure 5, the highest bandwidth efficiency for a given CNR will provide the highest data throughput under impaired transmission conditions. Conversely, the lowest CNR for a given bandwidth efficiency yields the largest noise margin. What this graph does not show is the advantage of extensive coding during high powered burst noise.

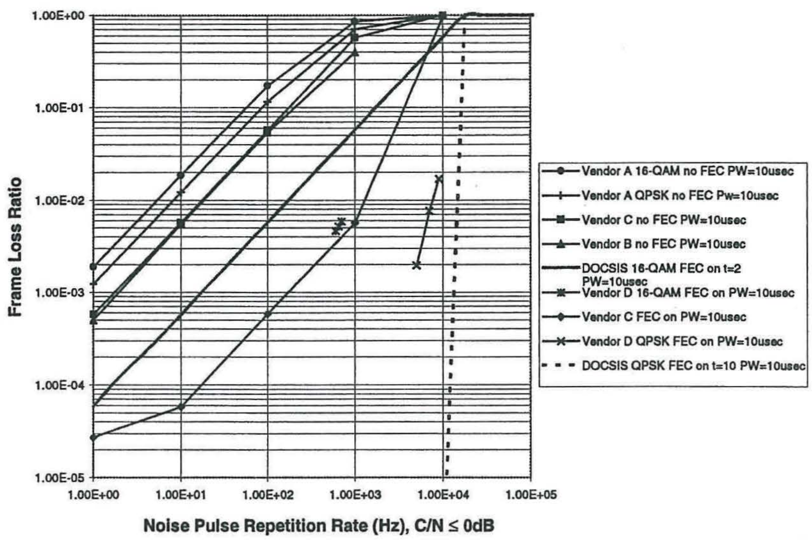


Figure 4: FLR vs. Noise Pulse Repetition Rate for PW=10 μs Burst Noise

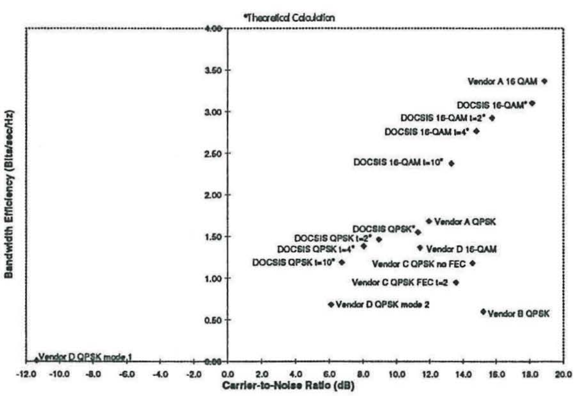


Figure 5: Bandwidth Efficiency vs. CNR for 1% FLR

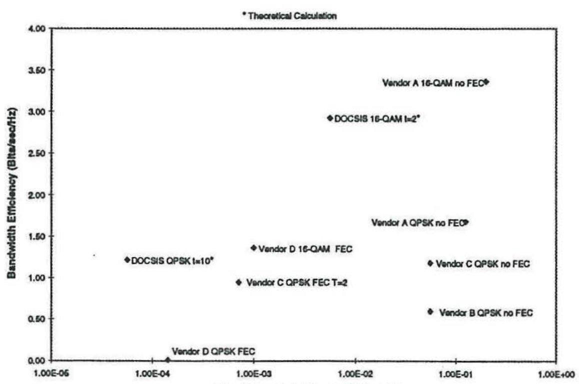


Figure 6: Bandwidth Efficiency vs. Noise Pulse Period for 1% FLR and 10 μs Noise Burst

Burst Noise

Figure 6 is based on an error performance of 1% FLR. This graph presents the bandwidth efficiency vs. the pulse repetition period for burst noise injected into the channel at a 0 dB CNR level (or high enough to cause errors during its duration). The best performance is obtained by maintaining a high bandwidth efficiency and operating at the smallest noise pulse period possible while maintaining a FLR=1%. The data is presented for a noise burst duration of 10 μs. Similar graphs can be created for 1 μs and 100 μs noise pulse widths and for other FLR values such as 10%. This data is obtained from the measurements similar to those presented in Figure 4. Figure 6 verifies the fact that a trade-off for robustness is made with bandwidth efficiency. For this scenario, the DOCSIS QPSK t=10 modem is very robust by reducing bandwidth efficiency. Operation in this mode provides burst-error correction, which gives an advantage in burst-noise mitigation. Because of their FEC, the Vendor C modem with FEC, the DOCSIS 16-QAM, and Vendor D modem in 16-QAM mode also performed well. The Vendor A 16-QAM modem has the highest bandwidth efficiency but, with no error correction, it is not as robust in a high-power burst-noise channel.

USE OF BANDWIDTH EFFICIENCY DATA

Detailed in this section is a technique of how bandwidth efficiency data can be used to choose the appropriate modem technology. Several of these modems have a set bandwidth and FEC capability. Therefore, the system designer is limited to what is available. Some of these modems have varying parameters that can be used to move to different points on the bandwidth efficiency graphs. For the modems with several modes shown on the graphs, such as DOCSIS, the system engineer could add points to create more of a continuous trace and could choose the operating point to be somewhere along the trace. All of these modems operate with a discrete set of parameters. In an operational system, FEC options, data rates, or bandwidths, can be adjusted dynamically to compensate for the changing channel and traffic conditions.

Different services have different error performance requirements and data rates. For example, a service such as status monitoring or polling may be able to tolerate more errors than some time-critical data application, such as video telephony or "twitch" video games, where re-transmission of errored frames can not be tolerated. One can use the information presented here to make decisions on optimal technologies for cable services as shown in the steps below.

Create tables or graphs similar to Figure 5 and Figure 6 corresponding to the error performance needed for the service provided.

Characterize the quality of the HFC plant from a CNR and burst noise distribution point-of-view. If the plant can only support a 16 dB CNR, then draw a vertical line on the graph at 16 dB, as shown in Figure 7.

Determine the minimum acceptable data rate for the service and the amount of bandwidth, which can be dedicated to that particular service. There are several issues associated with this decision.

The bandwidth decision may be limited to a range between strong ingress sources.

It may be a low revenue-generating service and the service provider may not want to allocate a wide channel or multiple smaller channels for the service. It may be a high revenue-generating service for which the service provider is willing to use a large part of the bandwidth in the best part of the return spectrum.

The amount of users supported per channel and the number of homes passed needs to be considered.

Consider how to fit all of the bi-directional services into a limited return band between 5 MHz and 42 MHz.

After deciding on the minimum data rate needed and the maximum bandwidth, which can be allocated, the minimum bandwidth efficiency can be calculated by dividing data rate by bandwidth. A horizontal line can be drawn through this point and any modems falling below it will not meet the required efficiency.

For example, if that line is drawn at 1.3 bits/sec/Hz, as in Figure 7, all the modems falling in the upper left quadrant will meet the minimum needs of the specific service. The system engineer could use a table in the same way by crossing out all the modems, which do not meet the CNR requirement, and then the modems that do not meet the bandwidth efficiency requirement. The ones that are left will meet the requirements of the service.

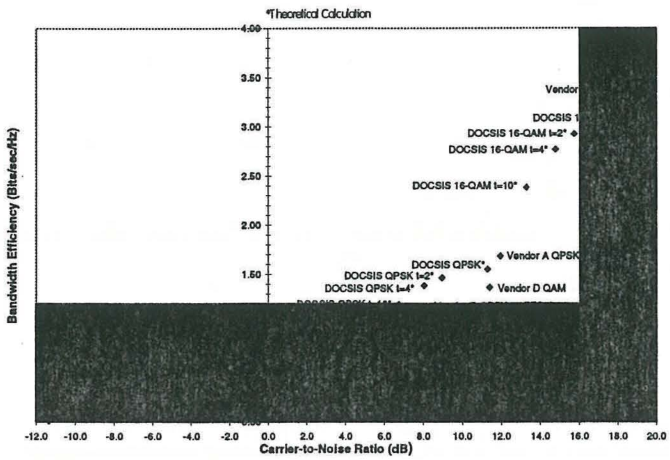


Figure 7: Bandwidth efficiency vs. CNR for 1% FLR

One also needs to consider the quality of the HFC plant from an impulsive noise point-of-view. If knowledge of the average length of time impulsive noise is present, from test equipment such as CableLabs CWTester™, this can be used to determine the amount of error correction needed. A similar procedure, shown in Figure 7, can be used with Figure 4. After choosing FEC depth, it should be verified that the error correction mode still falls into the quadrant of the graph chosen for CNR and BWE. If the constraints of the service will be difficult to implement in the current HFC network, this may be the justification necessary to upgrade the network by dividing an optical node or increasing the maintenance to improve the ingress performance.

CONCLUSION

Cable modem technologies have been compared to each other by the three performance metrics described herein. This data has been presented in a format detailing which technologies make the best use of the HFC networks' valuable bandwidth and power resources. The comparison also demonstrates how trade-offs in bandwidth efficiency must be made for robustness.

The type of analysis presented here can enable service providers to make system design decisions based on knowledge of the services they want to provide. This type of data can be used to help make decisions on which modem technology best suits their needs for the particular service. It can also be used to help make decisions on the amount of time and capital that needs to be spent on upgrading and maintaining the quality of the HFC plant. This data can also provide service providers with the information to determine if they are optimally and reliably utilizing their valuable bandwidth resource. The optimal use of HFC infrastructure will help ensure that the cable modem is the premium communication appliance in delivering high-speed data.

Personal Networking in the Home: Opportunities and Challenges

David Benham
Cisco Systems

ABSTRACT

The New World network will combine Internet data, phone and video services over a single cable line, fundamentally changing the way communication, entertainment/news, education and commerce services, as well as many other services we are only beginning to imagine, are delivered to consumers. At the same time, many new types of Internet-enabled consumer appliances will be used in and out of the home. These will range from web -phones to handheld and counter-top devices based on highly customizable thin and very-thin client architectures where the client's intelligence is derived from the network.

Thus, this New World network build-out represents the first major deployment of phone services over an Internet protocol (IP)-based infrastructure, using cable lines as the transmission vehicle.

At same time, this build-out will also need to be able to host a new type of 'personalized network' for the home, a sort of a plug-and-play local area network, connecting PCs, web-phones, TVs and other consumer Internet appliances. These personal networks begin to lay the groundwork for turning the Internet into the next mass medium, eventually connecting anyone to anything.

This paper will describe the potential revenue opportunities, survey the home area network connectivity options, and discuss what the technical and operational challenges will be for the service provider.

WHERE IS THE REVENUE?

The funding, estimated in the billions, for this New World network build out will primarily come from Technically Advanced Families (TAF) and Near TAF households as profiled in Figure 1.

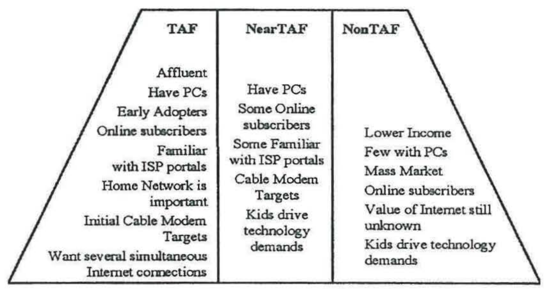


Figure 1: Profile of Families

At a meta-level, the revenue achieved from these households can be broken down into four categories; listed in no particular order below.

Work-at-Home Subsidies

First, work-at-home, or telecommuting, applications that are paid for or subsidized by a household member's employer or government will account for much of the early revenue. Businesses paid for some high-tech consumer products in the past, such as personal computers, fax machines, pagers, and ISDN, used by their employees at home.

This funding benefits product development by reducing the risk of consumer acceptance as it transfers cost away from the consumer. Stimulated by work-at-home subsidies, a market might become large enough to drive manufacturing volume to the point where prices are reduced for purely consumer applications. This way the work-at-home employee will use the New World services for business purposes where the company defrays the expense. After hours, the employee will use the same services for personal use. The opportunity for the service provider is to offer extra services that the employee will want to subscribe to purely for consumer purposes.

Subscription Revenues

Second, we will see subscription fees for value -added communications or entertainment features. Subscription revenue refers to the periodic income received from consumers in exchange for an ongoing service. Subscriptions for most consumer services--such as newspapers, magazines, cable TV service, and basic telephone--are in the relatively tight range of $10-20.

Transaction Fees

Third, add transaction fees for commerce or commerce helpers. In most media today, subscription revenues alone do not generate a profit. Whether it is newspapers, magazines, Internet access, or monthly cable charges, subscription revenues must be augmented with transaction or advertising revenue to make the business profitable. The same is likely to be true for the New World residential network.

Transaction fees are charges paid per specific event or transaction, such as payments for pay-per-view (PPV) screening. But revenue from PPV systems has not come close to living up to expectations. Thus, operators will need to look for ways to charge fees paid to receive information from a consumer's home network while elsewhere on the web (such as viewing baby/nanny cams). Perhaps operators will be able to extract transaction surcharges paid by home shopping where the transaction would be more much difficult without the having the home network outfitted with commerce-helpers connected to the Internet. One example of a commerce helper might be a web pad integrated into a kitchen appliance that reads SIC bar codes on grocery products. In short, transaction fees represent a potentially enormous source of revenue to content and service providers, especially considering that it could cost the consumer nothing to initialize.

Advertising Revenue

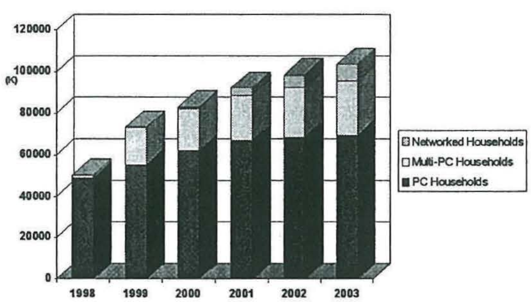
Television advertising is a $47 billion a year business in the United States alone. Total U.S. advertising (newspapers, magazines, and broadcast) is increasing at five percent per year. Even newspapers, thought by many to be obsolete in the digital age, still collect upward of $37 billion in the nation for print advertising and classified ads. Home networks and residential Internet access should eventually command a large enough audience to generate advertising revenue on the level of television advertising.

WHY A HOME NETWORK?

Consumers worldwide have high utilization of telephone and television services over existing telephone, over-the-air, and cable networks. For the most part, the services are of acceptable quality with a reasonable price (except perhaps long-distance phone charges). To justify the multibillion-dollar rollout of a New World network to the home, new services will be required to stimulate consumers or advertisers to underwrite the costs.

Aside from work-at-home connectivity and raw high-speed access, the ability to purchase all the communications services from one vendor, such as local and long distance phone services, unified messaging, data, and video will drive consumers to be interested in paying for broadband networking to their home. Voice-enabled cable modems will allow consumers to get phone service from their cable company, in addition to high-speed Internet access and television service.

The real boost in revenues will come from new services and advertising opportunities that will become available when the consumer can network multiple devices within the house, with each other and with the Internet. This is much more than just sharing the Internet access pipe. Even without the introduction of many new consumer Internet appliances, there is already strong growth in multi-PC homes. See figure 2.



Source: Dataquest, Jan 99

However, these PC will eventually get out numbered by many different types of consumer Internet / web appliances. See figure 3.

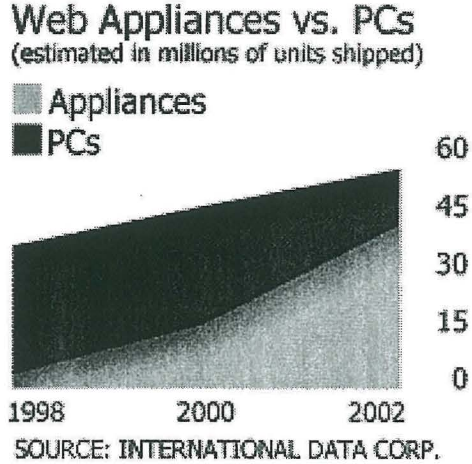


Figure 3

New Application Specific Appliances

As new services become available from the New World network, new devices will be created to make applications and interacting wit the web easier and more widely available (i.e., not requiring a full personal computer). Several of these new devices can be generically described as thin, and very-thin clients, where the client's intelligence is derived from the network. The intelligence can come from the cable operators network, the residential gateway, or home area network hub, in the home.

Some of the new devices we anticipate:

Web Terminals / Web Pads: Small terminals or pads, which may use an integrated LCD screen, or utilize the TV as a monitor. These may be connected to the operators network directly via an integrated cable modem, or be connected via an in-home residential gateway.

Video Telephony Terminals: Small terminals with a camera and microphone that allow telephony or video telephony applications. Again, these may use an integrated LCD screen, or utilize the TV as a monitor.

Gaming consoles: Nintendo, Sony, Sega, and similar devices that could allow opponents to interact over the network.

Home sensors, burglar alarms, baby/nanny cams: These devices will benefit from an "always -on" packet network which allows them to be continuously monitored (even from work), rather than waiting for an alarm signal.

Energy management and telemetry devices: With the advent of an always connected network to the home, new opportunities for more intelligence heavy-duty (high-energy consuming) devices is created. For example, electrical companies today are generally willing to offer preferred rates to businesses and homeowners who promise to reduce their electrical consumption during peak hours. This demand -shaping saves electrical companies money because it prevents them from having to build in extra capacity to handles the peak. As a home area network becomes pervasive, this could enable heavy-duty appliances to be monitored and re-scheduled to operate at off-peak hours whenever possible.

Connectivity within the Home

The primary requirement from the market is "no new wires!" There are primarily three high-speed technologies (generally around 10 Mbps today), or means, developing that can meet this requirement. Solutions are quickly coming to market that connect devices within the home via short-range, multimedia wireless, home -phone -wiring and power/electrical wiring. The standards bodies that defined the USB and IEEE 1394 ("Firewire") technologies are talking about making updates to their specifications for both higher-speed and longer distances. For now, their distance limitations restrict them primarily to peripheral area networking (USB) and entertainment area networking (IEEE 1394).

OPERATIONAL SERVICES AND CHALLENGES

First and foremost, simplicity or "Plug & Play" services are needed, especially for low cost, mass deployment of these new thin-client appliances and home PCs. Some of the functional means to that end will need to include directory services and user authentication.

A universal directory service is a critical middleware function that should reside in the New World broadband network or in the residential gateway, or some combination of both. The most common application of directory are similar to white-pages found in the phone book. Directories can be used to store information about anything, including infrastructure components, services, user preferences, etc. One example of a directory service specification is the Lightweight Directory Access Protocol (LDAP). User information could be stored on in an LDAP compliant server, so that it can be accessed by multiple applications or devices. LDAP provides rich security facilities which should be able to support sophisticated access rules to personalized data.

Universal authentication can be provided using a combination of directory service and password/digital certificate technologies. Collections of certificates technologies giving access to services could be encrypted and stored on a directory server for ready access from anywhere on a network. Applications requiring a higher degree of security could use digital certificates that are issued and stored on a "smart card" that the user can carry and use a with a wide variety of devices. For the work-at-home employee, security features for their employer's Virtual Private Network (VPN) will be extremely important.

Finally, another pair of very important technical challenge exists in the integration of the home network to the service providers network. Each of these interactions need to be transparent to the consumer in the home and will likely be one of the primary functions of the residential gateway. One challenge is properly handling the dynamic allocation of IP addresses in the home, all while using one-to-many private address plan. The other is integrating the quality of service (QoS) attributes of the home networking technologies to that of the service provider's network and the Internet.

CONCLUSION

Home Nets are an extension of Internet! IP is the unifying protocol in the Internet, for business, and now its coming into the home, enabling the convergence of data, voice and video, as well as anywhere and anytime personal communication devices

The challenge for service providers is to adapt their backbone and distribution technologies as well as enable new revenue generating services to be tightly coupled to those home networks and their new thin client appliances, service needs and E-commerce demands!

Author information:
David Benham is Manager of Product Marketing, Home Area Networking-Consumer Internet Devices at Cisco Systems. He can reached at the e-mail address: dbenham@cisco.com

Precise Location of System Flaws Using Egress and Ingress Signal Data Streams Separately and in Combination

Raymond J. Schneider
ComSonics Inc.

ABSTRACT

Precise location of leakage signals is a challenge. Models of leakage detection are presented and used to guide the development of signal processing strategies for improving the accuracy of source localization.

A range of strategies are examined together with both simulated and real data to identify the potential performance of fully developed systems.

PROBLEM SUMMARY

System Flaws

Flaws in a CATV system result in both the leakage of rf energy into the environment and a weakness at which external signals can more easily ingress into the CATV system, posing the potential for upstream and downstream interference.

Leakage monitoring from vehicles is often done either manually or with logging equipment collecting data at relatively low sampling rates and at a single receiver frequency. In an earlier paper a system capable of ingressing position and leakage data on a narrow band carrier into an operating cable system was demonstrated. [1]

Improvements

Preliminary assessments of data drawn from the initial experiments suggested several improvements:

1) Increase the sampling rate to ensure that constructive and destructive interference patterns can be resolved.
2) Explore signal processing methods to estimate the position of system flaws based on analysis of the received egress and ingress data streams.
3) Provide for differential GPS post-processing to anchor data to precise position.

In the discussion that follows we will describe the theoretical framework in which we intend to examine the issue of flaw localization, we will discuss certain fundamental issues which restrict the performance of any flaw localization system, and we will illustrate a set of flaw localization strategies using both modeling and experimental data.

THEORETICAL FRAMEWORK

In the context of CATV system integrity we will define a *flaw* as a non-uniformity in the system where rf energy can both leave (egress/leakage) and enter (ingress/interference) the system. A flaw has a location and a set of properties which define it as an unwanted antenna. Due to these properties it will admit or emit energy in different degrees and with different characteristics as a function of frequency.

Detection of flaws is accomplished by observing phenomena which arise from the presence of the flaw. The primary phenomenon is just the egress or ingress of energy. The use of egress/leakage is the traditional method of finding leaks. Ingress has also been used through the expedient of keying CB transmitters while someone monitors the return path. [2] at the headend.

The flaw may be considered an unconventional antenna, and as such the path between a conventional antenna and the flaw may simply be considered an example of a two way transmission path. In the case of egress, the transmission path is the CATV downstream being broadcast through the flaw, in the case of ingress an external transmission is received through the flaw.

The Friis transmission formula describes the power transfer between a transmitter and receiver in free space.

Eq'n (1): $$\frac{P_R}{P_T} = \frac{\lambda^2 G_T G_R}{(4\pi r)^2}$$

where
λ= wavelength of transmission,
r = distance between transmitter and receiver,
G_T = transmitter antenna gain in direction of receiver,
G_R = receiver antenna gain in the direction of transmitter,
P_T and P_R are respectively the transmitted power and the received power.

In practice this equation is too simple. The transmitted power does not arrive at the transmitter only by the direct path. Instead reflected power adds or subtracts from the signal from all directions depending upon the presence of reflecting, absorbing, and re-radiating elements in the surrounding environment. In addition to these various multi-path signals the transmitter and receiver gain terms vary with direction of arrival. Thus the actual power received is the coherent summation of all possible transmission paths which terminate at the receiver. This general situation is too complex to model. However there are various simpler situations which can be modeled and provide insight into the general process.

Ground Reflection

The simplest augmentation of our first equation is that involving a single reflection with the ground. It is normally the case with CATV egress signals that both the transmitter and the receiver are close to the ground. Thus ground reflections at moderate grazing angles are unavoidable.

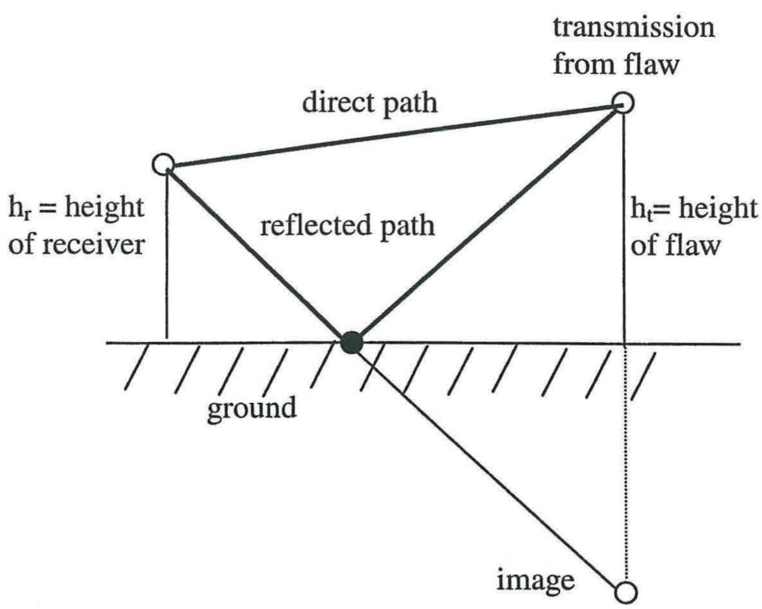


Figure 1. Single Ground Reflection from flaw to receiver

The analysis of a ground reflection can often be facilitated by the idea of a mirror. The conducting ground "images" the flaw and the resulting signal can be represented as the combination of the direct path signal from the flaw and a signal from the image.

As the a vehicle drives by the flaw the egress energy grows rapidly because the power goes as the inverse square of the distance. This variation in power is the dominant or primary effect. Smaller, secondary effects give character to the signal due to constructive and destructive interference between the direct and reflected path signals, and due to details of the radiation patterns of the equivalent flaw and receiving antennas respectively.

If we neglect the antenna patterns by assuming isotropic (omnidirectional) antennas, the magnitude of the electric field from the flaw can be written as:

Eq'n 2

$$|E_r|^2 = \left(\frac{|E_t|\lambda}{4\pi}\right)^2 \left(\frac{1}{r_d^{\,2}} + \frac{k^2}{r_r^{\,2}} + \frac{2k}{r_d r_r} \cos\left(\frac{2\pi(r_r - r_d)}{\lambda} + \phi_o\right)\right)$$

where

$|E_r|, |E_t|$ are the magnitudes of the transmitted and received electric fields. The square of the electric field is proportional to power.

$$r_d = \sqrt{d^2 + (h_t - h_r)^2}$$

direct path distance from flaw to receiver

d = horizontal distance

h_t = height of flaw (transmitter)

h_r = height of receiver

$$r_r = \sqrt{d^2 + (h_t + h_r)^2}$$

reflected path distance from flaw to receiver

k = reflection coefficient

λ = wavelength of the radiation

$\frac{2\pi(r_r + r_d)}{\lambda}$ is the phase difference between the two paths, direct and reflected

ϕ_o = the phase change at the point of reflection

The terms in $\frac{1}{r_d^{\,2}}$ and $\frac{k^2}{r_r^{\,2}}$ as well as the term multiplying the cosine term of equation 2 generate a waveform with a single maximum. The phase variation caused by changing distances as the vehicle drives by the flaw can cause added structure in the signal, broadening the peak and putting ripple into it. Figure 2 illustrates a family of wave-forms generated by varying offsets (20' to 50' in 10' increments) generating the wave-forms one would receive as one drove by the flaw.

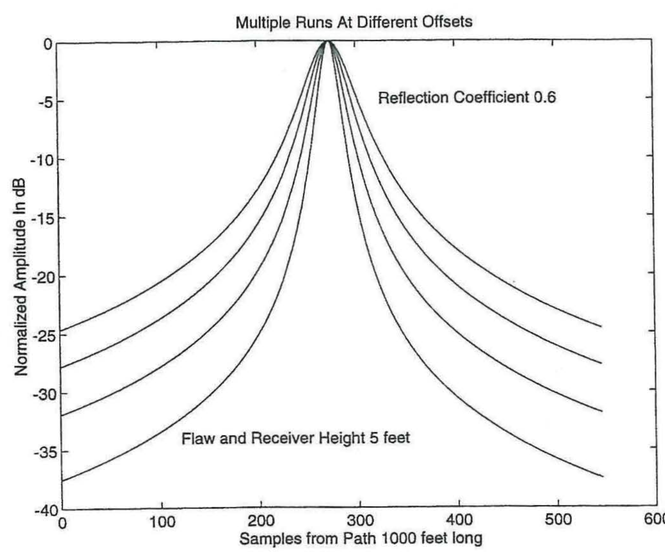


Figure 2. Runs Past Flaw at Different Offsets

This single peak aspect illustrated in Figure 2 is due to the fact that the flaw and the antenna are only five feet off the ground at a frequency of 113 MHz (a wavelength of 8.71'). Thus the phase difference between direct and reflected paths does not change very much. Note however that the paths with greater offsets have peaks that are broader than those with smaller offsets.

If the flaw and the antenna are both raised appreciably (or conversely if the frequency is raised proportionally) one sees many interference lobes. Figure 3 is an illustration of this effect using a height of 100'.

As the height is reduced, the interference lobes are fewer and only in the neighborhood of the peak.

Peak to Trough Ratio

If k is the reflection coefficient (≤ 1), the largest amplitude that the signal can acquire for a single ground reflection is (1+k), constructive interference, and the weakest amplitude would be (1-k) destructive interference.

Eq'n 3

$$dB_{ptt} = 20 \log_{10}\left(\frac{1+k}{1-k}\right)$$

Equation 3 calculates the peak to trough ratio in dB (dB_{ptt}). This is the worst case variation to be expected for a given reflection coefficient. For example, a 0.6 value of k, fairly typical for ground reflection computes to a value of 12 dB for dB_{ptt}.

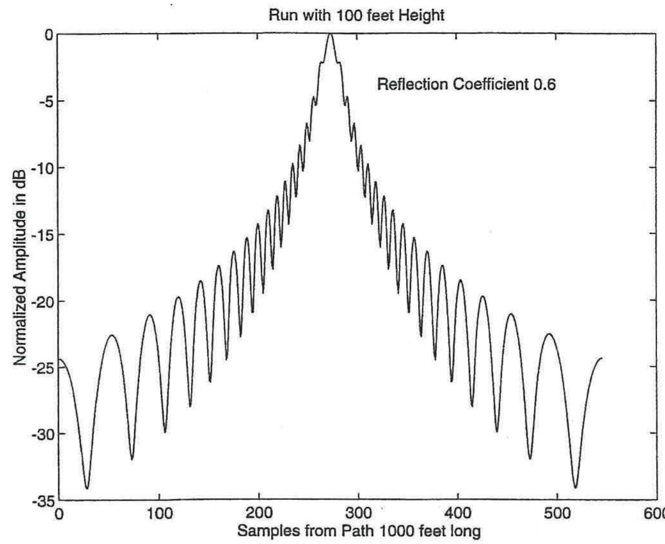


Figure 3. Path with Height of 100'

Faceted Reflections

In addition to ground reflection, large expanses of metal, as found on the sides of many buildings (ex. aluminum siding) provide highly effective reflectors. These reflectors may have reflection coefficients of 0.9 or more. At these reflection levels, multi-bounce reflections can sometimes play a role in the received leakage signal.

Figure 4 is a diagram which illustrates an example of multiple reflection modes based on the ComSonics facility. In the case of faceted reflection, one can use image ideas to calculate the expected signal paths. However the amplitude of the image will be limited by the size of the facet (reflecting area) and the visibility of the image from the line-of-sight being considered.

In Figure 4 the front of the south side of the ComSonics facility is the first facet and it has the effect of displacing the image of the flaw down the road to the north. The second mode illustrated is a corner reflector mode. In corner reflectors, energy is reflected off two walls that are 90° to each other and the energy emerges in the opposite sense displaced but parallel to the incoming sense. Because of the unusual shape of

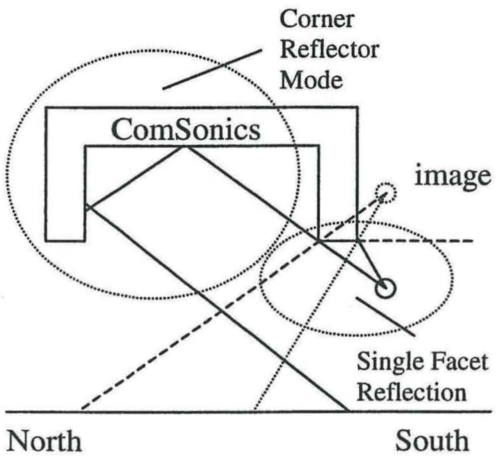


Figure 4. Multiple Facet Reflection Modes

the building a third mode involving parallel plate reflections may exist.

Faceted reflection modes produce more complex interference effects and also produce multiple *images* of the flaw distributed along the track of the surveillance vehicle. These effects are often masked by the fact that the vehicle travels at different speeds as it goes past a flaw, so that even with a graph of the signal it may not be evident that it arises from multiple reflections. Most leakage surveillance sampling is done at relatively slow sampling speeds. To guarantee that all interference effects are observed it is important to sample fast enough to capture the effects of small phase changes.

Figure 5 shows a data sequence including four runs past a large flaw intentionally inserted into ComSonics' model system. The flaw is located approximately 28 meters from the road running past the front of the building.

The vehicle ran past the flaw at varying speeds making two passes to the south and two passes to the north. These passes are annotated as p1s, p1n, p2s, and p2n respectively. Figure 5 shows the raw data prior to any processing. The strong peaks vary in width from pass to pass because the vehicle drove past the flaw at different speeds.

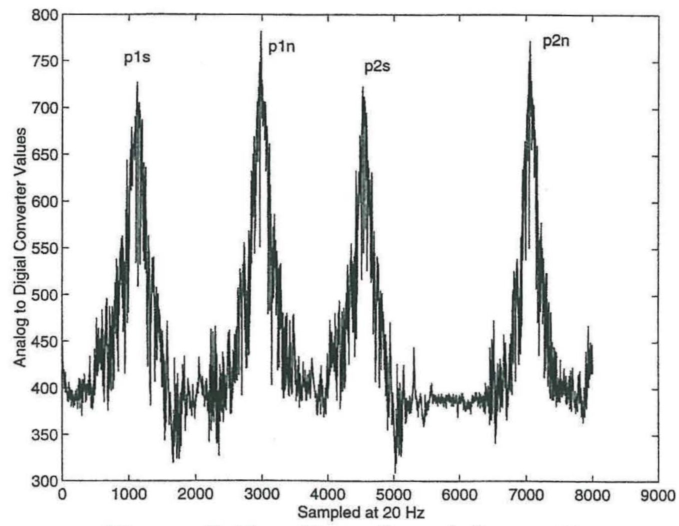


Figure 5. Four Runs Past A Large Flaw

Normally one would think that the spiky fine structure in figure 5 was noise. We compared data from different runs by first normalizing the data to position. Selecting the largest value in each run as a reference point, we calculated the distance of the vehicle from the reference using the GPS data. We interpolated intermediate positions by assuming the vehicle's speed did not change between GPS updates once a second. We could then compare runs directly.

Figure 6 is a direct comparison of p1n and p2n, the two northerly runs past the flaw. To make the visual comparison easier to perform, the data from the two runs has been offset by 100 A/D values by simply adding 100 to the data points of p2n. The data is plotted versus distance from the reference point in meters north and south of the peak. (note this is northerly and southerly since the road is not exactly N-S).

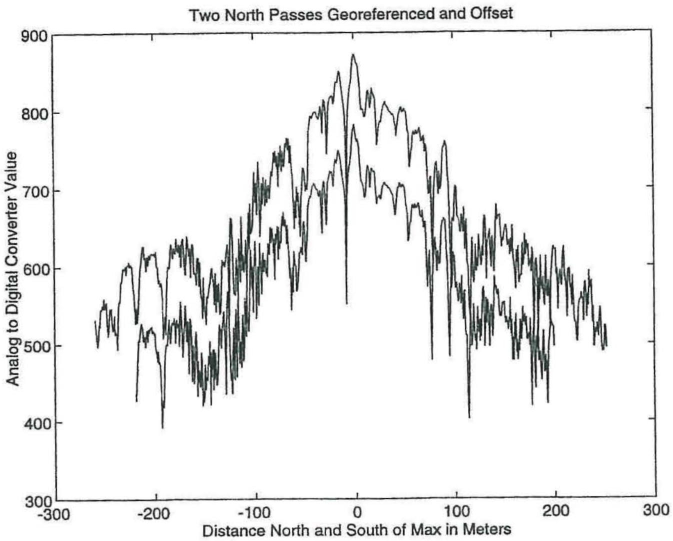


Figure 6. p1n and p2n Plotted Together for Comparison

What is significant in figure 6 is that each feature in the data is repeated in the two runs down to a very minute level of detail. The conclusion is that these details are due to physical and spatial properties of the flaw and its surroundings and not to temporal variations. Thus it should be possible in principle to analyze signal features and draw conclusions about the flaw and the environment in which it is active.

The primary objective is to be able to locate the flaws as precisely as possible. To accomplish that, one must be able to distinguish the apparent leaks caused by reflections and other signal distorting effects from the signal due to the direct path. This might be accomplished by simply picking the biggest signal except that one would need a means of telling that the next biggest signal wasn't just a reflection. Signal analysis methods seem likely to offer a means of making such distinctions.

Finding Reflections Using Convolution

There are a variety of signal processing techniques that can be employed to analyze data collected from flaws. Most involve convolution. One way to think about convolution is that it is like running a template along the data looking for a match. The better the match, the higher the value.

Figure 7 results from convolving a slice of the biggest peak of the data in figure 6 across the full data run. Notice the resulting peaks in the data. The central peak shows the result of the convolution when it exactly matches the peak of the data. Then there are three *image* peaks which are reflections, probably off the ComSonics building.

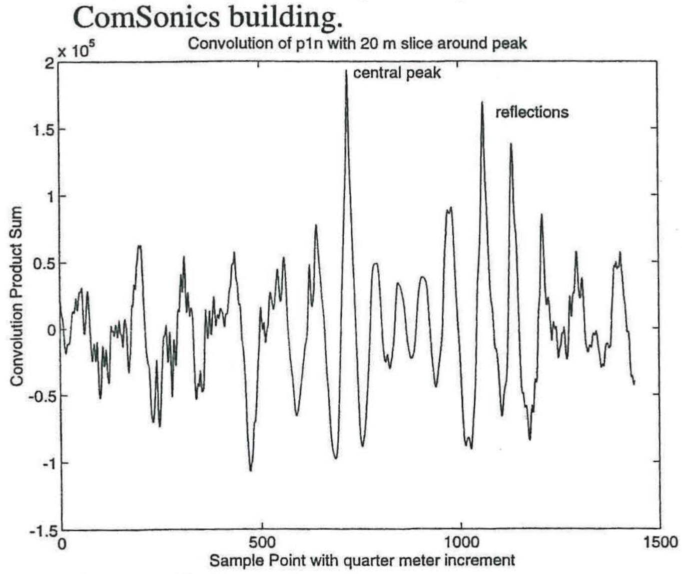


Figure 7. Convolution of p1n Showing Reflections

Ingression

If the surveillance vehicle is transmitting in the return path frequency interval, then ingress will take place in the

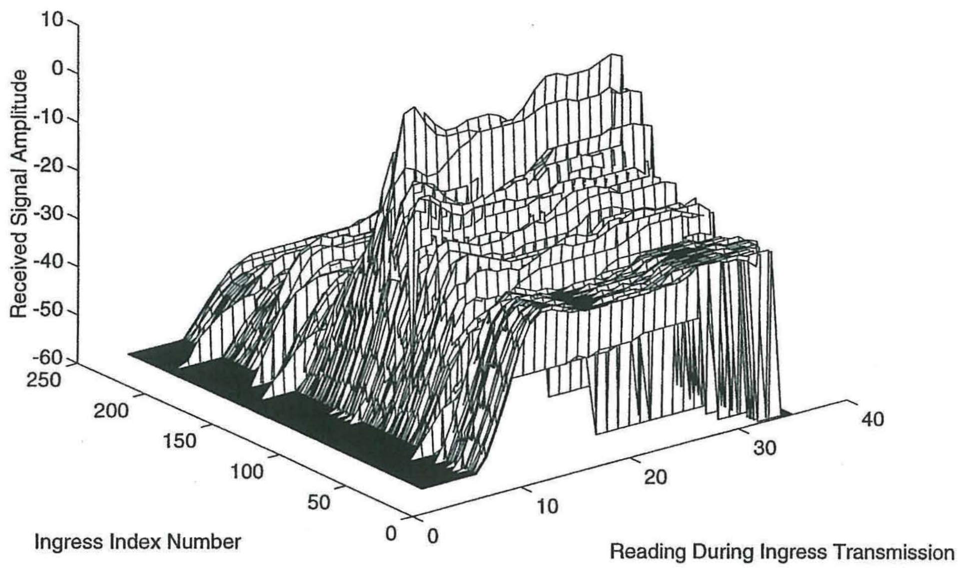


Figure 8. Headend Receiver Readings During Ingression

neighborhood of a flaw. The signal paths of the egress and ingress will be highly correlated due to the reciprocity theorem. Since the wavelength of the ingress is generally much larger than the wavelength of the egress, there will be less structure in the ingress signal. If the transmitted signal is modulated, information can be transmitted into the return path through the flaw. This was demonstrated in an earlier paper. [1]

During the transmission interval, an ingress signal received at the hub or headend location can be measured in amplitude. Fine structure due to path variation during the transmission interval can be observed. This fine structure can provide information about amplitude and phase variation used to infer relative distance to the flaw.

Figure 8 depicts an ingression run with 221 transmission intervals. Each transmission interval is triggered by a GPS reading, i.e once a second. GPS position data and leakage detector readings are transmitted and received and demodulated by a receiver at the headend. During the receipt of the data the carrier level is measured and recorded with the received position and leakage data. In figure 8 the receiver carrier levels are plotted slice by slice. Each interval is a few tenths of a second long and there are up to 35 readings per transmission. The readings taken in the vicinity of the ingression peak provide high density sampling.

FUNDAMENTAL ISSUES

The fundamental issue to be resolved is:

How does one detect and localize a flaw using ingress and egress information?

From the discussion above of the theoretical framework and some practical issues surrounding it, we find that flaws can be characterized by a strong peak due to the inverse square power relationship. This peak can be variously *ornamented* with attending peaks and troughs as a function of multipath constructive and destructive interference. It can also be surrounded by various *reflections* and *images* that confound its location.

Thus to detect and localize a flaw one must:

1) find a way to isolate the flaw from the surrounding *clutter* of unwanted reflections and images,
2) possibly use the detailed ornamentation of the flaw to classify it as to type and location, and
3) use the peaking characteristic to locate it precisely along the surveillance path.

Ingress coupled with egress provides not only an independent look at the flaw, but

4) positively confirms that the flaw is in the system under surveillance,
5) instantly transfers the location and vehicle developed flaw information to the headend, and
6) provides frequency diversity, hence spatially distinct information about the geometry of the flaw.

LOCALIZATION STRATEGIES

We have illustrated the difficulty of localization, even in the case of a large leak by showing what an isolated large leak in front of the ComSonics building looks like. Even this rather complex event is simple by contrast with the signals arising from multiple leaks in urban settings. The ongoing Ingressor development is focusing on the issues listed above. To deal with these issues we require a group of strategies which collectively can deal with the problem of detection, classification, and localization of flaws.

Flaw Isolation

To isolate flaws we are developing a compound strategy:

1) first isolate events which may be flaws using flaw-fitting methods (a group of specialized basis functions in a linear vector space)
2) perform event analysis to assign attributes to the events
3) cluster events that are spatially near one another and create cluster zones that may be related to a single flaw
4) perform the analysis in a time sequential, adaptive manner to minimize *time-late* flaw classification time

Analysis of Ornamentation

Ornamentation characteristics include such entities as semi-periodic fine structure on the inverse square characteristic caused by constructive and destructive interference associated with ground reflection and faceted reflectors.

The current strategy is to characterize these structures by their symmetry with the flaw event cluster peak, their peak-to-trough ratios, their periodicity, and period variations. This information increases the confidence that a flaw is correctly classified and in at least some cases provides estimates of off-path distance.

Peak Normalization

Because the flaw signals exhibit an inverse square peaking characteristic, peaking is a dominant feature of flaw event clusters. However, multiple flaws will add to the complexity of the situations creating cluster overlaps and *ghost* peaks.

The biggest peak in a group can be thought of as a spreading center. Image peaks are smaller than the biggest peak. Thus clusters will normally consist of a strong central peak and a family of smaller peaks. When this symmetry is broken it usually means that there is an overlap from another cluster of peaks centered on another flaw. As the peaks approach the noise level of the system they become more difficult to resolve.

Ingression Alert

Using the ingression path to send not only a unique signal which can be monitored at the headend, but a signal with egress information, and localization information provides a real-time path that absolutely demonstrates the presence of a system flaw.

The lower frequency of ingress produces different interference signals which can be compared with the higher frequency egress signals. Comparing the two signals provides a cross check on the information provided by either one alone.

FINDINGS AND CONCLUSIONS

Since this report is on a work in progress, the findings and conclusions are preliminary. So far this work has accomplished the following:

1) Intentional broadcast of modulated narrow band signals in the CATV return path interval allows egress measurements and localization information to be transmitted readily into system flaws.
2) The information can be decoded at the hub or headend and provides direct confirmation that the system has a flaw.
3) Measurement of the received carrier amplitudes allows comparison with egress levels reported across the link.
4) High speed sampling shows that egress and ingress signals are extremely repeatable, even in fine structure.
5) Signal processing methods can be employed to improve localization by both rejecting artifacts such as reflections and more precisely locating the point of closest approach to the flaw. Further work is expected to provide significant performance in these areas.

ACKNOWLEDGEMENTS

Thanks go to Dennis Zimmerman, the CEO of ComSonics for the idea of ingressing data and measurements back into the cable plant, and to Randy Smith whose untiring efforts provided the data shown in this paper.

REFERENCES

1. R. J. Schneider, "Utilizing Ingress as a Plant Maintenance Strategy," NCTA Technical Papers, 81-87 (1998)

2. R. V. C. Dickinson, "Centralized Cable Leakage Detection, Localization and Measurement," NCTA Cumulative Leakage Index / Flyover Seminar, Feb. 14-15, 1989.

Rate-remultiplexing: An Optimum Bandwidth Utilization Technology

Richard S. Prodan, Mukta Kar, and Majid Chelehmal
Cable Television Laboratories

ABSTRACT

Compression technology, coupled with advances in digital modulation, has many advantages over its analog counterpart, most notably efficiency and flexibility in bandwidth utilization. MPEG standardization, MPEG-2 in particular, has been the key ingredient for interoperability and has been the catalyst in the widespread acceptance of digital audio/video (AV) technology. These factors have contributed to the transmission/broadcast of multiple standard definition television (SDTV) signals, or one high definition television (HDTV) signal, over a single 6-MHz channel. The advantages of digital technology are merging telecommunication services, computing, and digital AV industries into multimedia, where digital video will play a dominant role. New multimedia applications are evolving every day. Even though digital technology has expanded the transmission efficiency, the addition of new applications and services for delivery to homes and businesses will bring new challenges to the cable and broadcast industries. Most of these challenges can be facilitated by an efficient use of the transmission spectrum. We believe that rate-remultiplexing technology can address the issue of optimal utilization of available bandwidth using compressed digital video delivery.

This technology has been researched at CableLabs over the past two years. In this paper, rate-remultiplexing techniques will be analyzed, and their corresponding advantages and disadvantages will be discussed. The implementation of this technology in hybrid fiber coaxial (HFC) cable networks for efficient spectrum management will be discussed in detail. Finally, development activity and the current availability of equipment utilizing rate-remultiplexing will be discussed.

INTRODUCTION

Multiplexing is the sharing of a single resource by more than one application. Multiplexing, such as frequency division multiplexing (FDM), has application in the analog world. However, multiplexing has become indispensable in digital and data communications. Such signals are placed in a time division slot sequentially. The multiplexed signals, along with the multiplexing information, form a serial bitstream, which is then transmitted over a link either as a baseband transport or then modulated over a carrier as shown in Figure 1. At the receiving end, the bitstream is demultiplexed (per multiplexing information) with the bitstream and all constituent signals retrieved. There are many advantages to multiplexing, the most important of which are flexibility and better utilization of the available spectrum. Without multiplexing, each signal will either require a separate physical link or a carrier frequency for transmission.

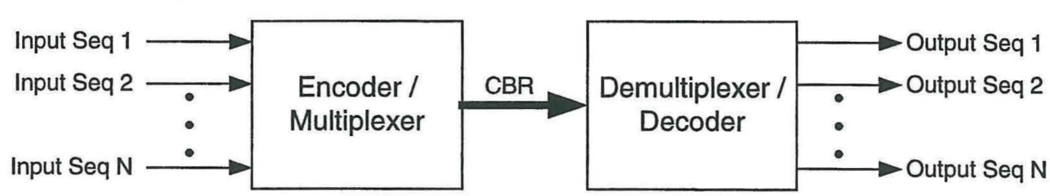


Figure 1: System for Transmitting Multiple Digital Signals over a Single Channel

Until a few years ago, multiplexing was used only for low-speed signals such as voice and data. Advances in compression technology and digital modulation have made digital video communication a reality. Although there are limited applications in the transmission of silent or still images, real-time video with synchronized audio make multiplexing essential. In the analog NTSC system, video and audio can be considered as frequency multiplexed. In a digital communication system, compressed video, audio, and private data are time multiplexed to form a single TV program stream. Simple encoding/multiplexing and demultiplexing/decoding block diagrams are shown in Figure 2 and Figure 3, respectively. In the same manner, multiple video, audio, and other data streams can be multiplexed to form a single serial stream containing multiple TV programs. This multiplexed stream may be transmitted over a short link in the baseband domain or may be modulated over a carrier and transmitted.

Until a few years ago, the use of compression technology and video, audio and data multiplexing was proprietary and was used in limited applications. The standardization effort, by ITU and by ISO/MPEG, was the catalyst in the widespread use of digital compressed AV technology. In particular, the MPEG-2 standardization and advances in digital modulation have contributed to the transmission/broadcast of multiple standard definition TV (SDTV) programs or high definition TV (HDTV) programs over a single 6-MHz channel. It is to be noted that using the analog technology, one 6-MHz channel can be used only to transmit one standard definition TV program. In effect, digital technology has the capability to expand the use of the available physical spectrum significantly.

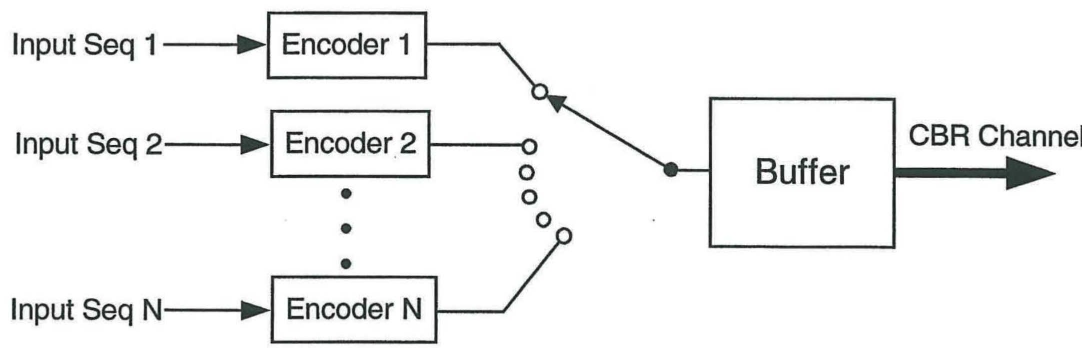


Figure 2: Encoder/Multiplexer Block Diagram

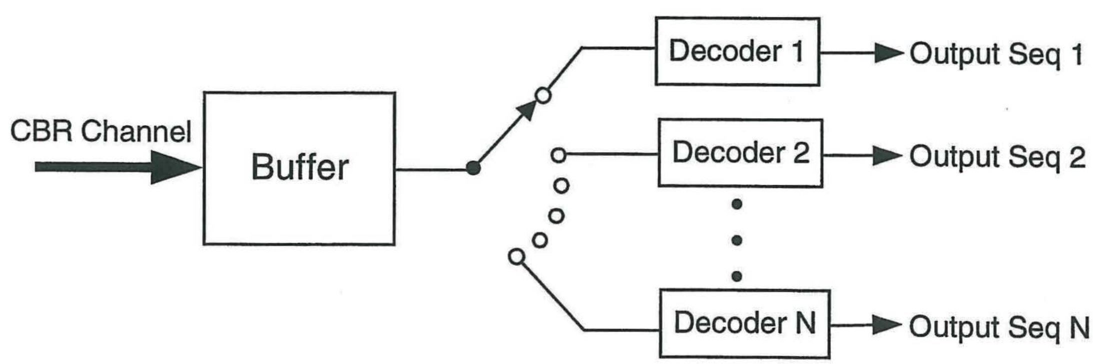


Figure 3: Demultiplexer/Decoder Block Diagram

Advances in compression technology and digital modulation have been merging the telecommunications services, computing and digital AV industries into the multimedia arena where digital video is an important ingredient. New multimedia applications are evolving every day and will compete for the available spectrum for delivery to the user/consumer.

Digital technology is complex and also is more expensive than analog technology. However, digital technology has the potential to add new services in addition to digital TV programming. To offset some of the cost of the investment for digital technology, as well as to increase the overall revenue stream, it will be prudent to add additional multimedia services. All of these services will require new channels, which implies that there will be the need for more spectrum. Using higher order digital modulation (64 QAM to 256 QAM in the case of cable) can expand the spectrum space. The use of a more efficient modulation technique will require upgrades to physical plants, including consumer premises equipment (CPE). An alternative to such upgrades is to use remultiplexing or "grooming" of the available spectrum so that new services may be added more efficiently.

CBR AND VBR STREAMS

With constant bit rate (CBR), a sequence of video frames is compressed such that the output bitrate remains constant while maintaining a desired level of quality. Compressed bits used in each of the encoded frames may vary, but the average number of bits used per second is kept constant. In other words, the bitrate of the compressed stream is constant. Variable bit rate (VBR) allows an encoder to compress every frame to the best extent (with a minimum number of bits), while maintaining the desired quality level. The average output bitrate varies depending on the complexity of various scenes. VBR compression is particularly suitable for storage purposes and, for this reason, it is used in DVD technology. However, in transmitting a single VBR stream over a channel of fixed bandwidth, VBR provides no extra advantage over CBR. But, when VBR streams are created as a result of statistical multiplexing, a better bandwidth utilization is achieved compared to the straightforward multiplexing of CBR streams.

STATISTICAL MULTIPLEXING

Figure 4 shows a block diagram of a statistical multiplexing system. In this system, a number of video sequences are compressed on the basis of what bits are allocated to each encoder. The bit allocation decision depends on the picture statistics of the current frames and on the coding statistics of the previous frames in the input sequences. Picture statistics are derived from picture complexity analysis data. Coding statistics indicate the in-use status of bits allocated for encoding the previous frame, and whether more bits or fewer bits are used than allocated. Encoders 1 through N act as VBR encoders. The video elementary streams (ES) are buffered in buffers 1 through N, and are multiplexed with other elementary streams and program-specific information (PSI). The multiplexer bitrate constraint is that the sum of all ES bitrates, which is less than, or equal to, the constant channel rate (R_c). Statistical multiplexing helps to utilize digital channels better than CBR-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 used in statistical multiplexing.

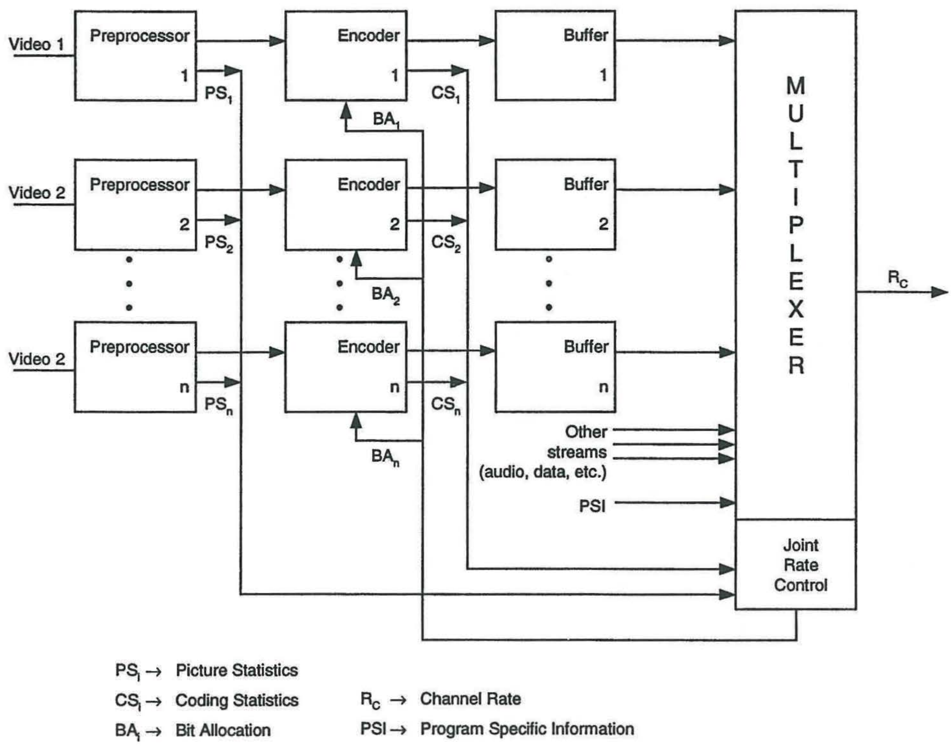


Figure 4: Statistical Multiplexing System Block Diagram

DTV SIGNALS

As is the case with analog TV programming distribution systems, digital television content providers will distribute programming using media such as satellite, off-air, or direct feed. These signals, when appropriately demodulated, will provide MPEG-2 multiplexed baseband transport streams. Each of these multiplexed streams will often contain multiple TV programs. All the programming received in a single multiplex may not be chosen at the headend for downstream delivery on the same 6-MHz channel. A *remultiplexer* receives one or more multiplexed streams as input and creates a new output multiplexed stream from local operator-selected programs. A remultiplexer is needed in the headend to interface multiplexes of previously compressed video, audio and program specific information (PSI) streams. Figure 5 shows a block diagram of a simple remultiplexer.

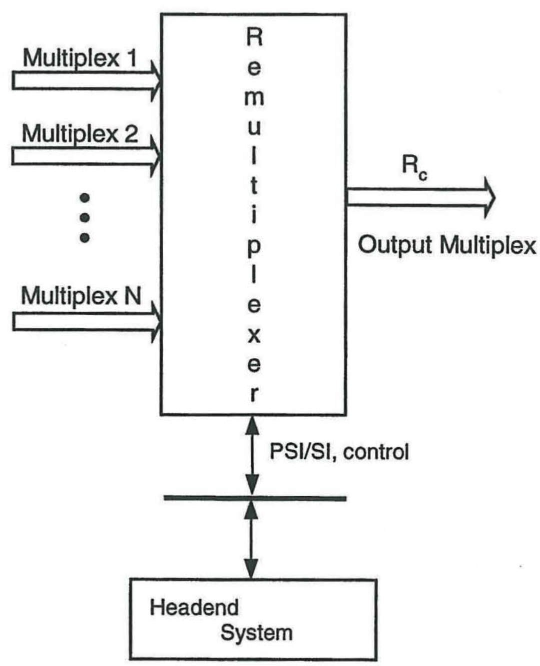


Figure 5: Simple Remultiplexer Block Diagram

Of course, both the input multiplexes and the new multiplex must be compliant to ISO/IEC

13818-1, MPEG-2 transport stream specifications. Additionally, the video elementary streams in the output multiplex must adhere to the Transport System Target Decoder (T-STD) buffer model to avoid overflow or underflow of the decoder buffer. A few of the important functions performed by a remultiplexer are:

1) Demultiplexes the input multiplexed streams (unbundles the individual programs);
2) Creates a new multiplex out of the operator-selected programs and includes PSI for the new multiplex;
3) Maintains the bitrate constraint such that sum of all elementary stream bitrates, including PSI does not exceed the transmission channel rate;
4) Removes jitter in Program Clock Reference (PCR) time stamp values and maintains AV synchronization within the programs;
5) Provides perceptually seamless switching capability from one program to the other without any audible or visual artifacts.

RATE-REMULTIPLEXING

Nominally, a remultiplexer does not alter the ES 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. Rate-remultiplexing may be thought of as an enhanced remultiplexing method. It meets the latter constraint by transcoding individual video ES within the output multiplex. The new functionality added to a standard remultiplexer to achieve bitrate control include: Variable Length (Entropy) Decoding (VLD) of the video elementary streams; examination of DCT coefficient quantization of each input video stream to determine the overall bitrate allocation; coding statistics determination across the multiplex; coded picture bitrate reallocation by altering (re-quantizing) the DCT coefficients; and Variable Length Coding (VLC) of the resultant re-quantized video elementary streams.

Transcoding together with joint rate control and statistical multiplex scheduling are applied to ensure that the output multiplex is an MPEG-2 compliant transport stream, and that each video ES in the multiplex adhere to T-STD buffer model—no overflow or underflow occurs at the decoder buffer. Degradation in picture quality is not noticeable with occasional or slight reductions in the average bitrate of individual video streams. However, significant reduction in average bitrate may cause noticeable degradation in perceived quality.

Currently, content producers create compressed content compliant to the MPEG-2 standard and distribute content to the broadcast affiliates or cable headends. However, headends and hubs may vary in quality (bitrate) based on the allocation of channel capacity for each program. For example, a small hub with limited availability of physical channels may try to pack in as many digital channels as possible with a corresponding compromise in quality. To meet varying needs, content producers can create compressed content at higher quality (i.e., higher bitrate). The affiliates and headends can lower the bitrate as needed before delivery over their systems. Similar situations may occur with digitally compressed commercials and content for video-on-demand (VOD). 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. Without rate-remultiplexing capability, no matter how many different versions are created, a close match between available channel bitrate and the stored compressed commercial can not be guaranteed. Similarly, in the case of video-on-demand based on the quality-of-service (QOS) requirement, the bitrate of the video stream can be reduced to a target different quality level (bitrate). Another case may arise when a compressed video stream is distributed over heterogeneous networks. It may be necessary to reduce the video stream bitrate for a section of the network due to the lack of channel bandwidth availability, or during times of congestion. In such a case, QOS will also degrade. In these cases, rate-remultiplexing compressed video provides a reasonable solution.

TRANSCODING

Transcoding is the technique by which a compressed video stream is translated to a lower bitrate. Figure 6 is a block diagram of a simple transcoder. R_i is the bitrate of the input compressed video, R_o, the rate of the output transcoded video and c(t) may be considered as the user's input constraints.

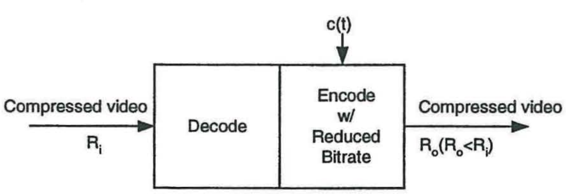


Figure 6: Simple Transcoder

In MPEG-2 compression, the majority of bits are used for representing the DCT coefficients of an 8 x 8 block of pixels (there are many 8 x 8 blocks in a picture). The main saving in bits is obtained by reducing the number of these non-zero DCT coefficients, or by making the value of the coefficients smaller. In the case of a compressed domain transcoder, rate reduction is achieved by representing coefficients up to a certain frequency threshold and setting a zero value for the coefficients above the threshold, or by coarsely quantizing the DCT coefficients. The latter not only causes reduction in the number of non-zero coefficients, but the values of the remaining coefficients also become smaller.

There are two transcoding techniques: open loop and closed loop. A typical open-loop transcoder is shown in Figure 7. It is easy to implement, and rate reduction is achieved in the compressed domain with minimal decoding and computation.

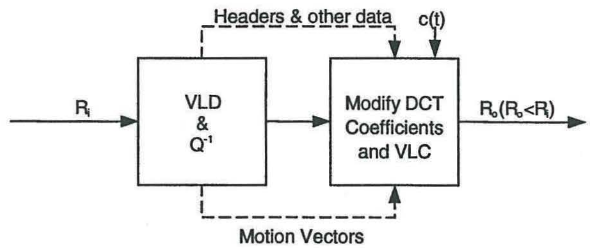


Figure 7: Open-loop Transcoder

Such a scheme provides little control over the output stream's picture quality. Another way to implement a transcoder is to decode the compressed video to uncompressed video and then re-encode it to a reduced bitrate. However, such a method will be highly complex to implement as it requires a decoder and an encoder for each video channel. It also adds more latency in transcoding.

Figure 8 shows a closed-loop transcoder where the re-encoded output is compared with the output of the first decoder. This provides a reasonable estimate of the error and the re-encoding decision may be made based on the estimated error. This is obviously better than an open-loop method where the bitrate is reduced without any computation of error resulting from the reduction in bitrate. Hybrid methods that estimate reconstructed (decompressed) errors in the compressed domain only can be used in a closed-loop system with significantly lower complexity.

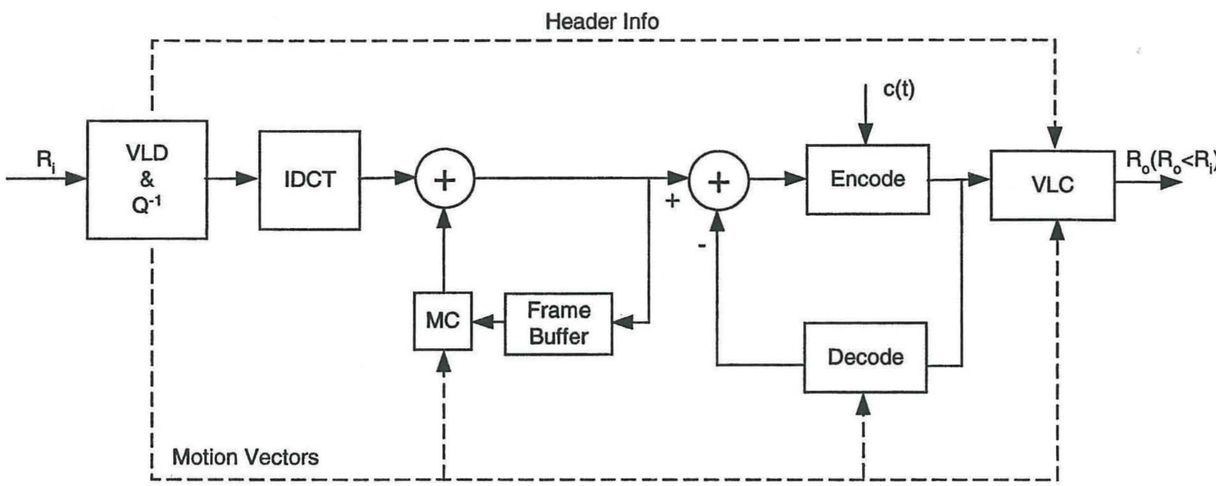


Figure 8: A Typical Closed-loop Transcoder

In MPEG-2 compression, motion vector (MV) computation is the most computationally intensive process. There are many ways to transcode a compressed video stream. In implementing any of these methods, a great deal of computation can be saved if one can reuse the MVs available with the input stream as much as possible. New prediction and MV computations may be needed in some local stream splicing situations such as ad insertion. It may be observed from Figures 7 and 8 that the computation of MVs has been avoided by utilizing the MVs available within the compressed stream. In open-loop transcoding, there could be drift in picture quality, particularly for a long group of pictures (GOP), as the P pictures are encoded and decoded with reference to either the I pictures or the previous P pictures. The error made in transcoding the initial I picture will be added with that of the following P pictures. In this way, the error accumulates in transcoding each consecutive P pictures until the next I picture. At every I picture, error accumulation resets. The problem of drifting in picture quality becomes worse as the number of iterations in transcoding increases. In general, in the open-loop method, the picture quality may degrade more than expected due to cascading transcoding operations on the same stream. In the closed-loop method, the degradation in picture quality due to cascading is deterministic and is reduced.

A rate-remultiplexer, shown in Figure 9, performs all of the functions of a remultiplexer, but has the following added capability. If the sum of the bitrates of all selected streams including PSI exceeds the channel transmission rate limit R_c, each video stream is trimmed or transcoded until the bitrate of the new multiplex is within that limit. This functionality will benefit headend operation in the following ways:

a) If some compromise in quality is acceptable, this later capability will allow the cable headend to pack more channels into a single multiplex.

b) Remultiplexing of multiplexed transport streams containing variable bitrate video elementary streams that have been statistically multiplexed when compressed at the source can be accommodated.

c) Only one bitrate version of a compressed commercial in the server need be stored, even though it will be remultiplexed in many different channels with different bitrates. It is assumed that most of the rate-remultiplexers will have the capability of digital program (local commercial) insertion (DPI). It was mentioned earlier that all remultiplexers should have smooth program switching capability. When the switching capability is already built in the device, it will be straightforward to add the program insertion capability.

d) The same thing will be true for video on demand (VOD) where only one high-quality version of the content will be stored. Then, based on the QOS requirement or available bandwidth, quality may be lowered if necessary.

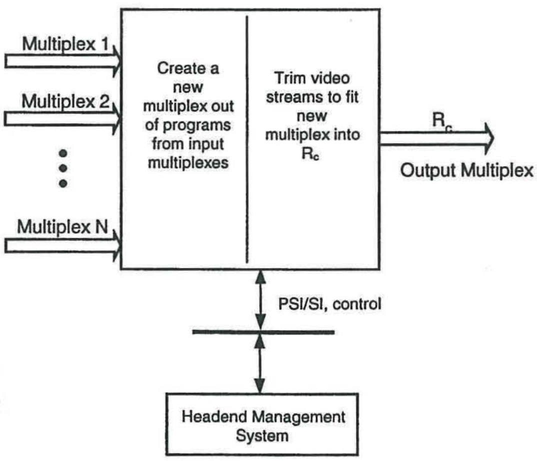


Figure 9: Rate-remultiplexer Block Diagram

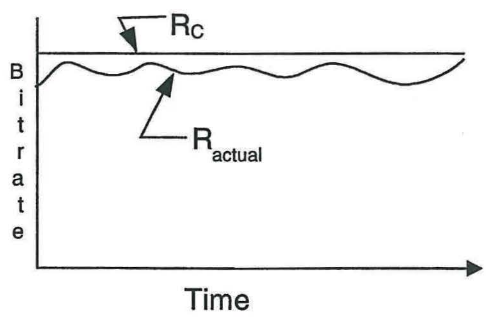


Figure 10: Bitrate of the Output Multiplex from a Rate-remultiplexer

In statistical multiplexing, encoding and multiplexing are performed at the same time and, as such, require all the input video streams to be in uncompressed format. The output bitrate of a statistical multiplexer is kept very close to the channel rate. The rate-remultiplexer can deal with previously encoded video streams, irrespective of constant or variable bit rate encoding. Even if the sum of instantaneous bitrates of all the selected streams for multiplexing is larger than R_c, it trims the video ES such that R_{actual} fits within R_c with minimal degradation in picture quality. The actual output bitrate (R_{actual}) of a rate-remultiplexer may vary (as shown in Figure 10) depending on the information content to the input video stream. Rate-remultiplexing may present an opportunity to deliver non-real-time data using the bits available.

The above advantages of rate-remultiplexing make it a very attractive choice over standard remultiplexing. It provides wider freedom to choose programs from various multiplexes and to pack them in a single output multiplex without having to worry about violating the constraint of fitting them together in a fixed rate channel. This freedom lends itself to greater flexibility by optimizing the use of the available 6 MHz channels in a specific market.

Local ad sales contribute significantly to the total revenue of the cable industry. The capability to insert compressed digital commercials into digital channels, at the headend, is a necessity for the support of local advertising inserted into regional compressed programming. Rate-remultiplexing technology provides that capability to a headend as it 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. As mentioned above, in the case of VOD, it will be much easier to modify content bitrate to meet resource constraints or QOS requirements. Hence, rate-remultiplexing is an excellent tool for the optimal use of the digital channel capacity for digital video services.

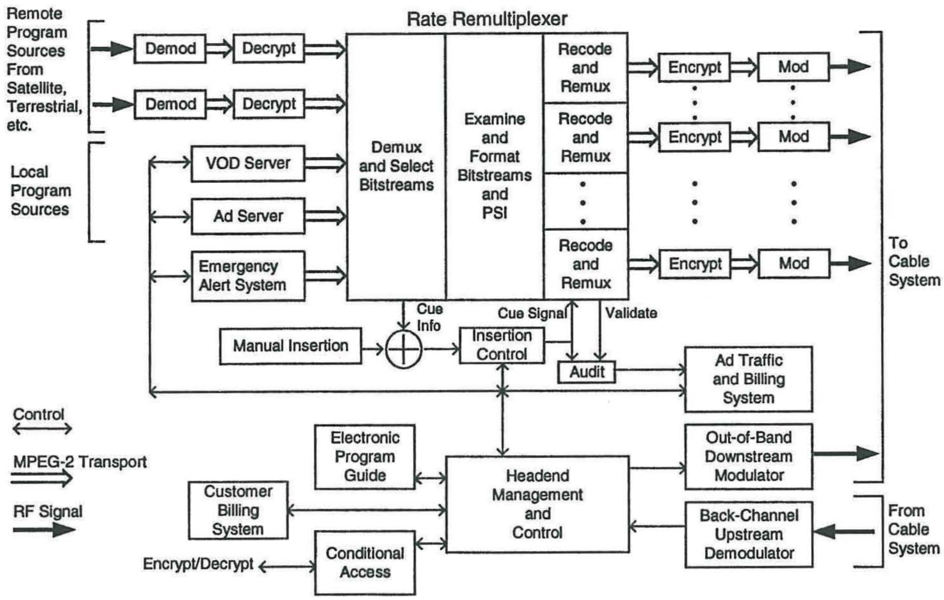


Figure 11: Integrated Architecture of a Headend Using a Rate-remultiplexer

Rate-remultiplexing will be an indispensable technology for most digital headends. A headend architecture where rate-remultiplexing technology has been integrated with the existing analog headends is shown in Figure 12. In this architecture, a number of multiplexes compliant to the MPEG-2 systems layer transport stream specifications (ISO/IEC 13818-1) have been interfaced to the rate remultiplexer. These multiplexes may originate from various sources, e.g., satellites (distributed by HBO, CNN, etc.), terrestrial (ABC, CBS, NBC, FOX, etc.), and direct feed (coaxial or fiber). They are appropriately demodulated and converted back to the baseband transport stream. The sources also include local servers such as VOD, ad, and other content servers. At the rate-remultiplexer, the input transport stream multiplexes are de-multiplexed into elementary streams, and a few of the programs are selected, per user choice, out of a number of programs available with input multiplexes. It also extracts PSI and SI streams associated with each input multiplex, and reformats and updates the information contained in these streams per the headend management system directives. The remultiplexer then remultiplexes the selected programs, including PSI and other data, into a new multiplex and recodes the video elementary streams, if necessary, to match the bitrate of the new multiplex with that of the available channel bitrate. DPI and VOD functions have also been integrated. By using transcoding, a major technical problem of matching stored bitrate with that of the channel bitrate becomes simplified. It is expected that the rate-remultiplexer will have the capability to detect network inserted cue-commands or a connector to interface externally detected cue-commands, including a local manual insertion mode. The output multiplexes from the rate-remultiplexer will be QAM-modulated. Each 6-MHz QAM modulated carrier is up-converted to the desired cable physical channel and is input to a combiner. The combiner sums analog channels with channels carrying digital multiplexes and channels carrying IP-based services. The output of the combiner is then transmitted over the cable system to subscribers. The out-of-band (OOB) channel for information and entitlement services and back channels for interaction on bi-directional systems have been included in this architecture.

CONCLUSION

In summary, the need for rate-remultiplexing to support applications involving compressed video transmission have been explored. Statistical multiplexing at the source encoding level and rate-remultiplexing at the headend will help utilize the available digital bandwidth more efficiently.

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. Transcoding can be divided into two major categories: open loop and closed loop. Open loop is simple to implement, but substantial picture drift may occur. It provides no control over errors made in transcoding a video elementary stream. Closed loop provides better control over errors and provides reasonable control over expected picture quality. Slight reductions in average bitrate may not cause any noticeable reduction in video quality. Substantial reductions in average bitrate may cause noticeable degradation in picture quality, which is expected in MPEG-2 compression.

Several advantages of rate-remultiplexing technology in a cable headend have been noted. A good degree of flexibility in packaging channels is provided. This will allow each headend to utilize their channel bandwidth resources in the system in an optimized way to deliver a selected number of services from multiple sources. Also, solutions to the problem of matching stored content bitrate with transmission channel bitrate in services such as DPI and VOD are accommodated. In other words, no bitrate match is required.

Finally, an integrated headend architecture was presented where rate-remultiplexing will play a major role in providing various digital television services and will coexist easily with existing analog/digital television and other services. Initially, rate-remultiplexing technology may be more complex than standard remultiplexing, but the advantages of rate-remultiplexing will outweigh the cost.

REFERENCES

1. *Statistical multiplexing with look-ahead using MPEG-2 video encoding*, by Lilla Böröczky, Agnes Y. Ngai and Edward F. Westermann, presented at SMPTE 33rd conference on Advanced Motion Imaging, Orlando, Florida, February 25-27, 1999.
2. *A Frequency-Domain Video Transcoder for Dynamic Bit-Rate Reduction of MPEG-2 Bit Streams*, by Pedro A. A. Assuncão and Mohammed Ghanbari, published in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 8, December 1998.
3. *Video Transcoding for Matching Network Bandwidth Constraints*, by Pedro A. A. Assuncão, a Ph.D. thesis, Department of Electrical Engineering, University of Essex, U.K., 1998.
4. *Multiplexing VBR Video Sequences Onto a CBR Channel with Lexicographic Optimization*, by Dzung T. Hoang and Jeffrey Scott Vitter, submitted to ICIP '97.
5. CODING OF MOVING PICTURES AND ASSOCIATED AUDIO, ISO/IEC 13818-1 (Systems), ISO/IEC 13818-2 (Video).

RESIDENTIAL GATEWAYS: FROM THE INSIDE OUT

Paul Pishal
Philips Broadband Networks

Abstract

The Cable TV industry is undertaking one of the most aggressive initiatives in its 51-year history. The digitization of the cable TV network infrastructure is proliferating on a massive scale with equipment deployment, compliance testing, and specification setting all taking place simultaneously. At the same time the consumer electronics industry also has a major digitization initiative occurring. The next generation of home devices will not only be digital, they will also feature "in-home" networking capability. Though monumental, both of these initiatives are occurring somewhat independently. Without a common roadmap of features, functions, and specifications, these two industries will remain out of step in delivering new and innovative services and applications to the digital home.

This paper focuses on the digital network initiatives both the Cable industry and the Consumer Electronics industry are pursuing. Investigating the strategies of these initiatives will demonstrate that without a closer partnership between these two industries, the full capabilities of new digital technologies will not be realized by the end-user. Finally, this paper will address the need for the development of a residential gateway architecture to provide a solution as these two networks continue to evolve.

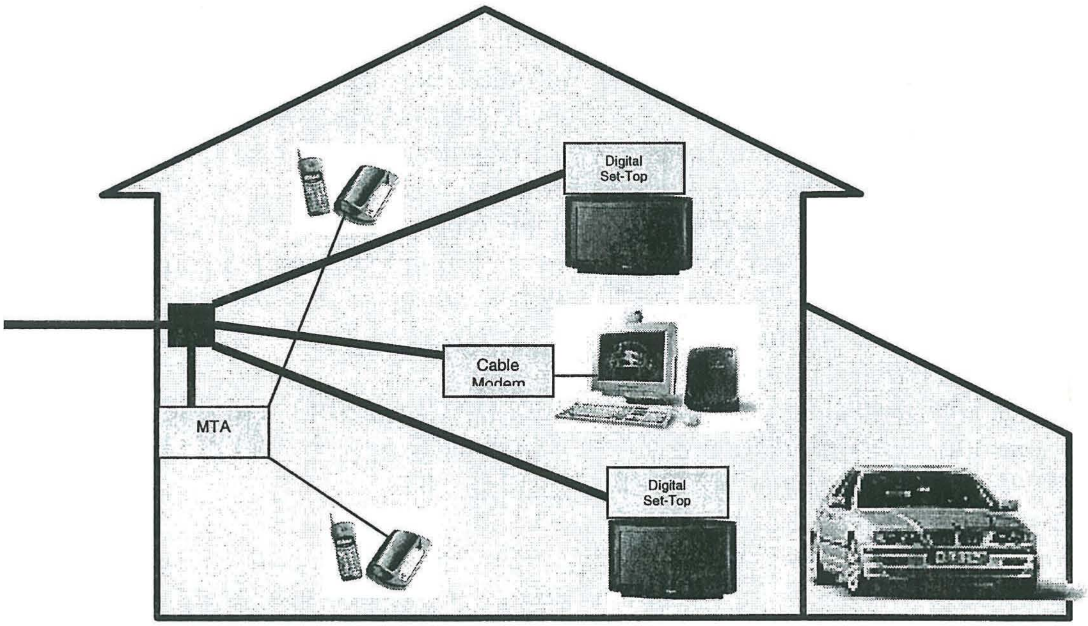
INTRODUCTION

While many cable service providers and consumer electronics vendors will define their own versions of the digital home, it will ultimately be the end consumer who will shape this environment. Most all parties agree that the digital home will be a networked dwelling with connectivity to a digital cable network that delivers advanced services to the home and a consumer electronics network that delivers information and entertainment throughout the home. As these two networks are being defined, designed, and built, each one is evolving with specifications independent of the other. The result will delay successful integration of the digital home and the advanced services being delivered to the home.

THE DIGITAL BROADBAND NETWORK AND NEW DIGITAL TERMINAL

The Cable TV industry is actively progressing in the deployment of digital services. Through the leadership of CableLabs and the support of the major cable operators, the specifications for digital service delivery networks have progressed significantly. Though not completing these specifications at speeds that many would like, to veterans of standards processes, the results achieved to date would be the envy of other industries.

DOCSIS-based cable modems, OpenCable digital set-tops, and PacketCable IP telecommunications devices will all populate the new digital home of the future. Independently specified to offer high-speed data, digital video, and voice services, respectively, these new digital devices will also be developed to deliver multiple services. For example, an OpenCable compliant digital set-top box can be equipped with a DOCSIS modem for two-way communications. This two-way communication would provide the capability to deliver cable modem service, as well as IP telephony and video conferencing. Additionally, cable modems could be used to supply not only high-speed data services but IP telephony services as well. Therefore, connection to the end devices (e.g., TVs, PC, telephones) will then become the next hurdle for the cable operator to overcome. The home wiring and installation challenges play an increasingly important operational role. Added complexity for provisioning and installing the new digital devices can limit the speed of rollout, and the result will be delayed success.



The process for interoperable standards

The time cycle to develop a specification and then utilize that specification for final delivery of interoperable products can take a considerable amount of time. In the case of cable modems, the DOCSIS process, gaining momentum in early 1996, resulted in CableLabs certified modems becoming available in March 1999. Both PacketCable and OpenCable specification processes are now ongoing. Therefore, any additional initiatives that extend beyond current DOCSIS, PacketCable, and OpenCable activities will have to factor in the element of time.

Current interfaces to the home

These digital specifications define many interoperability interfaces in an end-to-end system. Probably the most significant interface point is to and from the home. The interface within the home is usually dictated by existing standards (i.e., defined or de-facto). The current home devices interfaces are NTSC signals to the TV, Ethernet to the PC, and twisted pair to the telephone. The NTSC video interface was defined more that 40 years ago and Ethernet was define more than 20 years ago.

	Analog TV	Digital TV	PC	Telephone
DOCSIS			Ethernet, USB, PCI	
Open Cable	NTSC Coax	IEEE 1394		
Packet Cable			Ethernet, USB, PCI	Twisted Pair

The home interfaces, to date, define connections directly to the TV, PC, or telephone. Connections to intelligent networks within the home are not being defined in the current Cable industry standardization initiatives. Home networks are becoming a growing market segment with the Consumer Electronics and computing industry actively defining specifications and products.

Consumer Electronics Industry

The consumer electronics industry sees market growth for new digital devices in the home. Recent CEMA figures show digital television penetration at 30% of US home in the year 2006. Enhancing the entertainment experience is the primary driver for these devices, though the information component of entertainment will continue to play a larger role in the home.

The sources of content continue to move from broadcast and stored media (e.g., tape and disk) to real-time availability of content and information. Therefore, an increasing reliance of network delivered content and information is impacting the way end consumers expect to receive these services. As a result, the home will grow with new digital terminals, but in order to take advantage of network available content, these new devices will need to interface seamlessly with the adjacent broadband network.

These new devices include

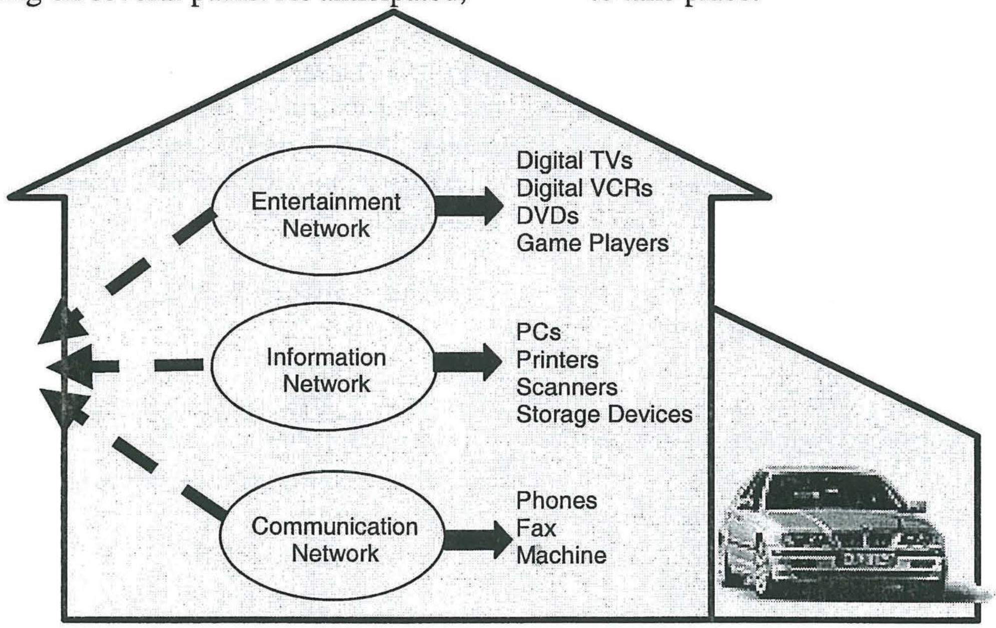
- Digital TVs (including HDTV)
- DVDs (Digital Versatile Discs)
- Digital Cameras (video and still)
- Digital VHS
- Digital Hi-Fi

Containing processor intelligence, these new devices will be capable of communicating with other devices. The goal is to create new levels of user capabilities for entertainment and learning that are not possible today. With increasing deployment throughout the digital home, these capabilities will be expanded to involve home control and security applications.

With an eye toward the future, consumer electronics manufacturers clearly see the need to network these devices in an open and interoperable manner. There are many industry initiatives that will drive the features and functions of the home network. Successful growth for home networks will occur only in an open standards environment. Interconnection with multiple devices from multiple vendors is the ultimate goal.

THE HOME NETWORK

Home networking initiatives are taking on several paths. As anticipated, there are several developments occurring in home networking. Shown below are several home networks that will connect devices in the home. These individual networks are classified as an entertainment network, an information network, and a communication network. Today, the terminal devices drive the architecture of these networks. However, as the differentiation of communication, entertainment, and information begins to subside with advancing technology, the convergence of these networks will start to take place.



The Communications Network consists of the in-home wiring today. Today's in-home twisted-pair wiring is an extension of the local exchange company network, with "dial tone" designed devices to connect with a central office. However, any in-home station-to-station connectivity must be done through the local exchange company switch.

The Information Network market approach resembles that of the office network. This allows PCs and peripheral devices to be networked together as more and more PCs proliferate in homes. Additionally, shared devices such as printers and scanners drive the need for networking. Application-based networking, such as gaming or Internet sharing, are also drivers for the Internet Protocol (IP)-based home network.

The Entertainment Network is in its infancy. Digital consumer devices can be networked together to either process or store entertainment content. Applications are starting to evolve that will allow multimedia to enhance the home entertainment experience. In order to tap the potential of these exciting entertainment services, standards must be set so that interoperability between different brands of digital devices can be interconnected.

Many industry alliances and consortia have been formed to set specifications for home networks. The alliances and consortia, such as Home Audio/Video Interoperability (HAVi), Universal Plug and Play, and Home API, are focused to create interoperable specifications for digital home devices and computing devices. Additionally, these initiatives allow for the development of distributed applications on the home network.

To accomplish this, the requirements for a home network consist of:

- The physical medium to connect the devices (e.g., power line, twisted pair, coax, wireless, optical)
- A set of software API to insure interoperability among terminal devices
- An addressing scheme that allows connectivity
- A defined execution environment that allows for control and 3rd party applications
- Capability to carry both asynchronous an isochronous data
- Easy user installation and management

THE HOME NETWORK ARCHITECTURE

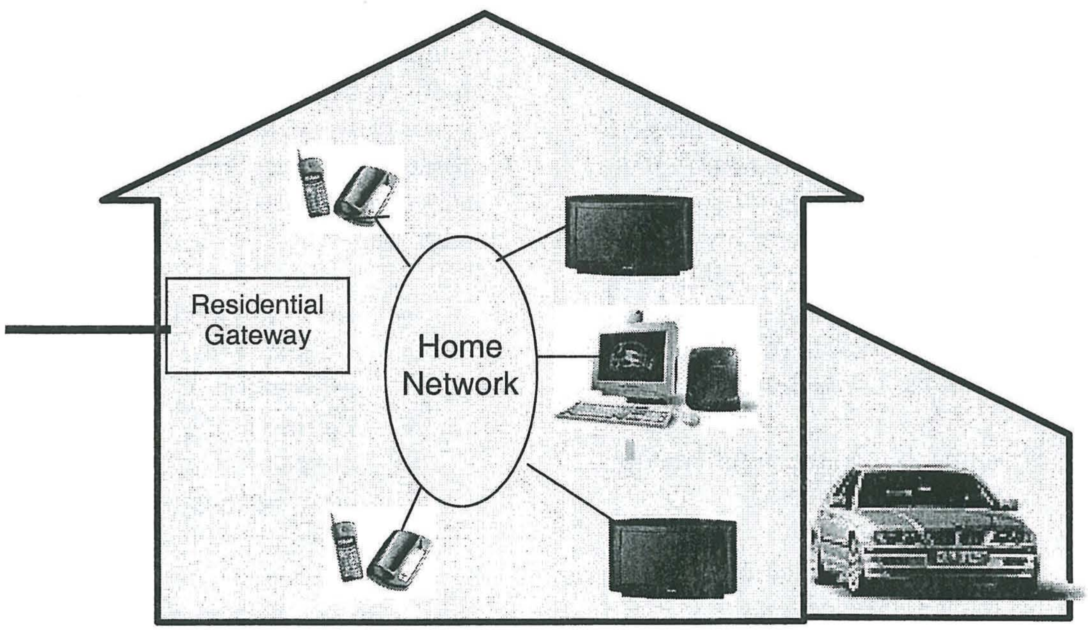
The home network can be a peer-to-peer network or it can have a controller. It becomes both economically and technically important to define the level of interoperability between the home network and the digital cable network. Interoperability must go beyond the physical interface and take advantage of the networking capabilities being defined for the home. Without this interface, network services will only be available to TVs, PCs, and telephones with co-resident network equipment (e.g., set-tops, cable modems, and multimedia adapters). With this interface, the following applications can be realized:

- Receiving a telephone call and displaying caller ID information on a TV window
- Starting a video-on-demand program, pause, and watch the remaining portion in another room
- At your PC, completing research on the web and allowing the results to be viewed by someone watching television in a different room, and saving the results on a common storage device
- On your PC, editing highlights of your daughter's ballet recital recorded on your digital camcorder, and then forwarding them to her grandparents' TV set.

These services become available only when the broadband network becomes interoperable with the home network. These services merely scratch the surface for the type of imaginative applications that can be realized. Before this can occur, several issues must be clarified. Ideally, a gateway architecture needs to be established that will interface the two networks.

THE RESIDENTIAL GATEWAY

The Residential Gateway is positioned as the interface between the HFC broadband network and the home network. A Residential Gateway becomes the point of interoperability between the broadband network and the home network. As changes occur in the networks inside and outside the home, the Residential Gateway can "mediate" the changes without the need to update every home digital appliance.



This becomes the potential collaboration point between the two industries.

The major functional requirements needed for a Residential Gateway are:

- Distribution of different services to multiple digital terminals
- Provide access to home control functions
- Request network bandwidth and Quality of Service from the home network to the broadband network

The technical requirements for a Residential Gateway will help to determine if this gateway is network provided equipment, part of the home network, or a compromise between the home and the broadband network. These technical requirements are:

- Address resolution between home devices and network connections
- Memory and processing requirements to run the home network communications protocols

- Flexibility to interface with yet to be defined/invented home digital terminals

Residential Gateway Architecture

The Residential Gateway will evolve along two different paths. These will be a network centric path and a home centric path.

The standards and specifications of the Cable TV digital plant drive the network centric path. Many operators are looking to digital set-tops with a combined video, voice, and data networking capability as a residential gateway. The connection to the network is already being defined by the efforts of CableLabs and the digital specifications of DOCSIS, OpenCable, and PacketCable. The digital connection to the home is specified today with an IEEE 1394 interface initially defined for HDTV pass through. In order to expand that interface for home network functions will require a clear definition as to how the home network will interconnect the digital set-top.

With the rapid and diverse technical evolution of the home network, the interface from the digital set-top and the home network becomes difficult to define. Beyond the physical IEEE 1394 interface, the communications protocols and API need to be specified. Depending on which home network technology is to be interfaced, even the IEEE 1394 interface may change. Additional processing power and memory need to be allocated within the box to insure that all features and functions can be supported across this interface.

The Residential Gateway may evolve from the dominant home network technology that gets a successful foothold in the residential market. This gateway approach will evolve from the consumer side and will have network requirements that can use the current Cable TV digital specification initiatives. This residential gateway architecture operates as a home controller. The benefit of this architecture is that the gateway can be home network specific, regardless of the type of technology used in the home network.

CONCLUSION

As both the home network and the digital cable network evolve, new capabilities will continue to be introduced resulting in exciting new services, applications, and experiences to the end-user. As the cable industry continues to set forth specifications allowing new digital services to the home, these specifications must be expanded to look into the home, which includes the home network.

In this rapidly changing world of home networks, a residential home gateway will be defined, built, and deployed as an extension of the computing consumer electronics devices that are deployed.

To take advantage of new service opportunities driven by digital devices and home networks, the Cable TV industry must expand the current specification initiatives to take account home networks, which will become pervasive, as well.

Reverse Path for Advanced Services — Architecture and Technology

Oleh Sniezko
AT&T Broadband & Internet Services

Abstract

This paper presents several methods of provisioning for frequency reuse in the reverse path. It analyzes several optical technologies existing today or feasible in near future. These technologies range from a traditional 1310 DFB lasers technology through ITU grid 1550 nm lasers to external modulators with distributed access. These transport technologies are augmented by several methods of multiplexing to achieve frequency reuse:

1) *spatial division multiplexing (physical segmentation of the reverse plant)*
2) *wavelength division multiplexing (coarse or dense)*
3) *frequency division multiplexing (frequency stacking or frequency block conversion), and*
4) *time division multiplexing (either after digitization of the reverse signals or after demodulation of the reverse signals).*

This multiplexing may take place in optical nodes or secondary hubs. Besides allowing for frequency reuse, the multiplexing schemes allow for fewer fibers and make the redundancy switching feasible.

The paper presents architectural implementation of these optical technologies and multiplexing methods. Further, the paper presents the results of the cost analysis for all these methods based on the most up-to-date pricing or estimated cost of the technologies being developed. Finally, the paper provides performance allocation for different segments of the reverse paths for the architectures analyzed, and achievable or projected end-of-line performance of the reverse path.

INTRODUCTION

One of the main drivers for upgrading HFC networks is to provide two-way communication for interactive services. The HFC networks support services that require bidirectional transport system to a dedicated group of potential subscribers. An increasing demand for these services leads to increasing demand for bandwidth in forward and reverse paths. The forward path of the HFC network has a plentiful of bandwidth and can support significantly higher modulation levels, thus increasing bandwidth efficiency. Traditional HFC reverse path, on the other hand, has lower bandwidth resources and can support only lower modulation levels. This shortcoming is partially remedied by the highly asymmetrical demand for capacity in most high-speed data access applications. However, with the advance of the communications services that require symmetrical rates in both directions, the only way to increase the throughput is by frequency reuse in the reverse path.

FORWARD PATH SEGMENTATION TECHNIQUES

Figure 1 presents a generic HFC architecture. It is applicable to large markets but its elements are also present in networks serving smaller markets. This architecture in most major markets is based on dual ring topology. These rings serve HFC plant streaming off the secondary hubs. The HFC network can be based on optical nodes (traditional fiber serving areas) or can deploy fiber significantly deeper.[1] Primary hub ring transport and multiplexing technologies have been described in many papers and are reasonably mature.[2,3] In these rings, technological progress is being implemented as it becomes mature and cost-effective. The other elements of the HFC network are undergoing a much faster evolution.

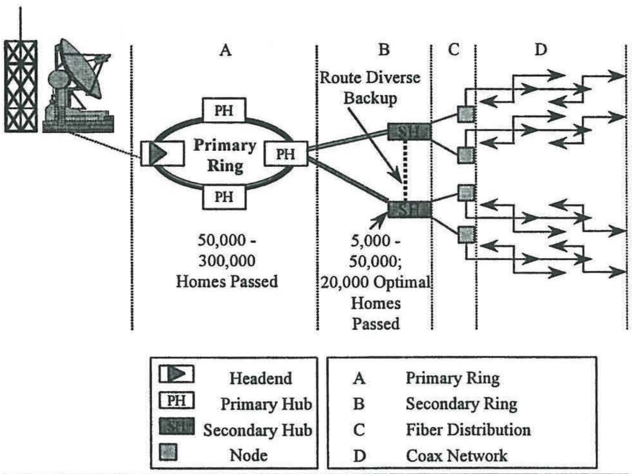


Figure 1: Modern HFC Network

Forward Primary Hub to Secondary Hub Links

The secondary hub rings provide redundancy and eliminate needs for high fiber count optical cables. They serve as signal multiplexing points to limit the number of fibers between primary hub ring and secondary hubs to achieve cost reduction, improve MTTR, and allow for cost-effective backup switching. AT&T Broadcast & Internet Services uses DWDM systems[4,5] in these rings or TDM systems (mostly SONET transport) for locations that could not deploy DWDM for technical or other reasons.

In the most recent architectural study by AT&T[6], a TDM technology is proposed to supplement the DWDM for future-proofing of the HFC network. However, in this application, the TDM technology is transparent to the services provided since it is overlaid throughout the network to the customer premises.

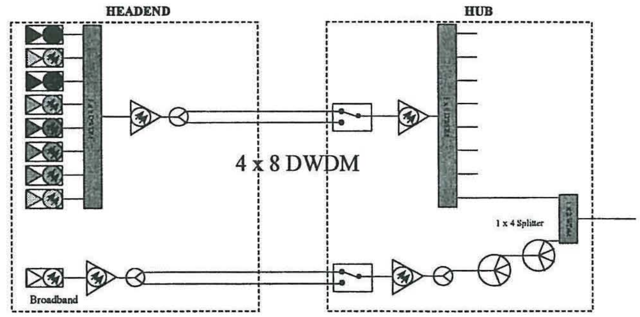


Figure 2: Configuration of DWDM in Forward Direction — Example

Forward Secondary Hub to Optical Node Links

The network sections (access network) marked as 'C' and 'D' in Figure 1 extend passed the secondary hubs and provide the last links to the customers. The architectural solutions for this network section differ from operator to operator. However, most of them deploy optical node-based configuration followed by active RF coaxial network. The differences are reflected in the node sizes and in the level of redundancy. Due to high bandwidth capacity requirements in the forward path, frequency-stacking techniques cannot be used in this section. The multiplexing (frequency reuse) is mostly based on SDM.

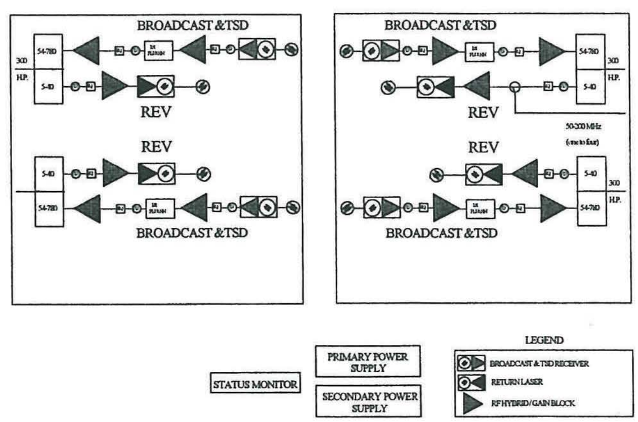


Figure 3: Spatial Division Multiplexing in Optical Node — Example

REVERSE PATH MULTIPLEXING TECHNIQUES

The reverse path in secondary hub rings and in the access network can use several multiplexing technologies to implement frequency reuse. The higher number of choices results from the fact that the reverse bandwidth is narrower. These multiplexing choices are:

1) spatial division multiplexing
2) wavelength division multiplexing (coarse or dense)
3) frequency division multiplexing (frequency stacking), and
4) time division multiplexing.

Reverse Primary Hub to Secondary Hub Links

Pure SDM techniques do not achieve the goals required of these links. They must be used with other multiplexing techniques such as WDM, FDM, or TDM. The advantage of WDM techniques is that they are transparent to any services and signals distributed (they simply mimic the SDM techniques with separate wavelengths) and can be combined with the remaining two techniques. These solutions are presented in Figures 4 through 6. It is apparent that the level of multiplexing can vary depending on availability and cost-effectiveness of the components as well as on the level of required performance. An alternative to the options described above, a baseband TDM technology (for example based on SONET transport used in forward and reverse paths) can be used but then secondary hubs become primary hubs, which in turn results in an increase in operating cost.

Dense Wavelength Division Multiplexing (DWDM)

The number of wavelengths per channel can range from 2 to 16 (or even 32) dependent on the level of multiplexing required and number of available fibers. However, this may lead to a higher cost of spares due to the higher wavelength number unless variable wavelength transmitters are developed and used.

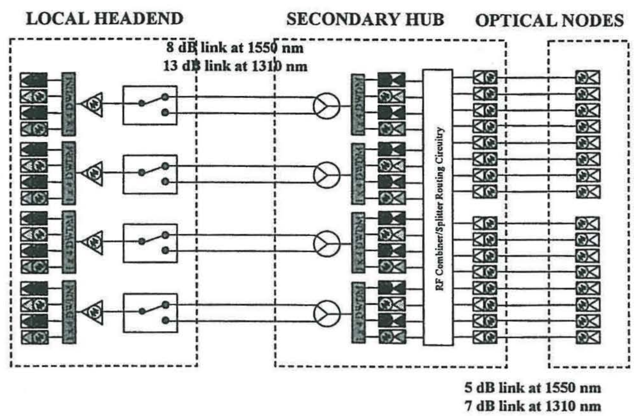


Figure 4: DWDM in Secondary Hub Links

Frequency Division Multiplexing (FDM) a.k.a. Frequency Stacking

The number of channels in frequency stackers/destackers depends on cost, equipment availability and level of performance required. It can range from 2 to 18 channels (up to 1 GHz).

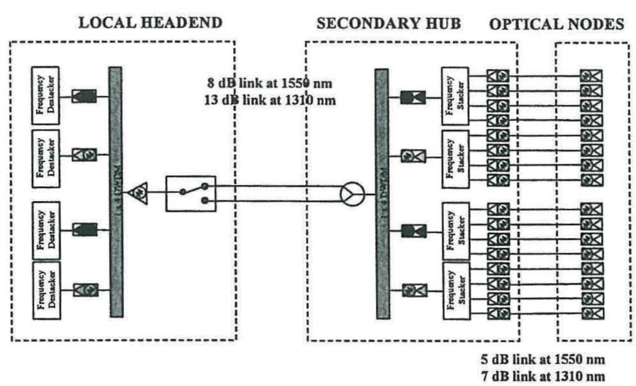


Figure 5: DWDM Combined with FDM in Secondary Hub Links

Time Division Multiplexing (TDM) after Digitization

Similarly, the number of channels multiplexed after digitization depends on the required performance, availability of the technology and its cost. It can range from 2 to 4 (or 16 at OC192 rates of multiplexing). Moreover, with simple arithmetical addition, an equivalent to RF combining topology (without its disadvantages) can be achieved at initial deployment stages before high frequency reuse is required.

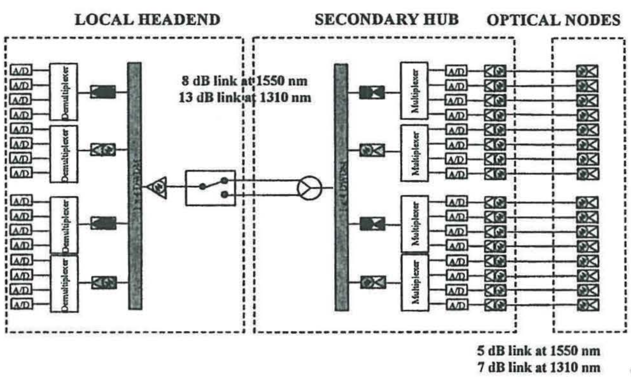


Figure 6: DWDM Combined with TDM in Secondary Hub Links

Secondary Hub to Optical Node Links

The same techniques that are used in primary to secondary hub links can be used in optical nodes. These techniques can be later repeated in secondary hubs for higher level of multiplexing. The choice is purely dependent on the equipment availability and cost as well as on the performance required. Some examples are presented in Figures 7 through 9.

ITU 1550 Lasers in Nodes

The lasers in the node can be 1310 or 1550 nm with wavelengths from ITU grid for transparent multiplexing in the secondary hub. The multiplexing may take place in the node if the number of fibers in the link is limited. The number of ITU lasers can range from 1 to 4.

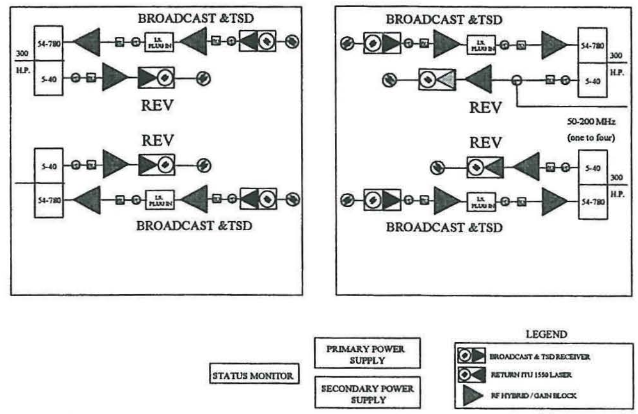


Figure 7: ITU Lasers from Optical Node

Frequency Stacking in Nodes

The number of FDM channels can range from 1 to 4 and they can be carried on 1310 or 1550 ITU laser.

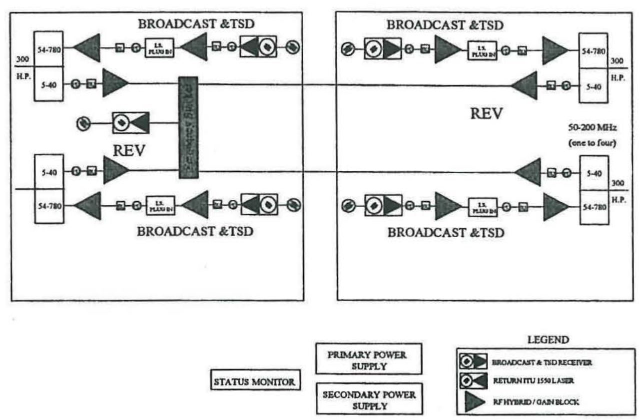


Figure 8: FDM from Optical Node

Digitizing and Time Division Multiplexing in Nodes

The number of channels multiplexed at the node can range from 2 to 4 (or 16 in a MuxNode[7]). They can be transported on 1310 or 1550 ITU lasers of digital quality and can be passed through or further multiplexed in the secondary hub.

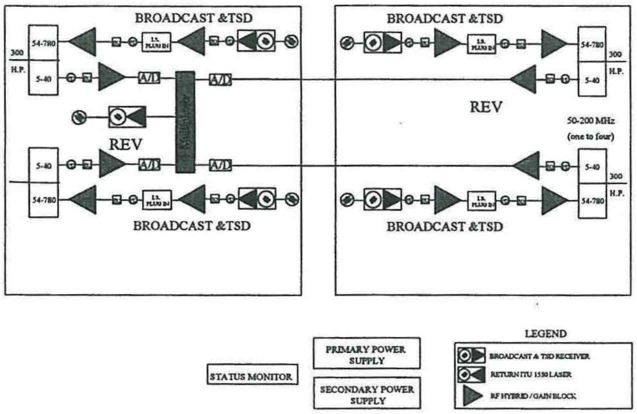


Figure 9: TDM from Optical Node

Multi-Multiplexing between Nodes and Primary Hub

All the multiplexing techniques described above can be used in combination and can be cascaded. Some examples of the combinations of the secondary hub and node multiplexing are shown in Figures 10 through 12.

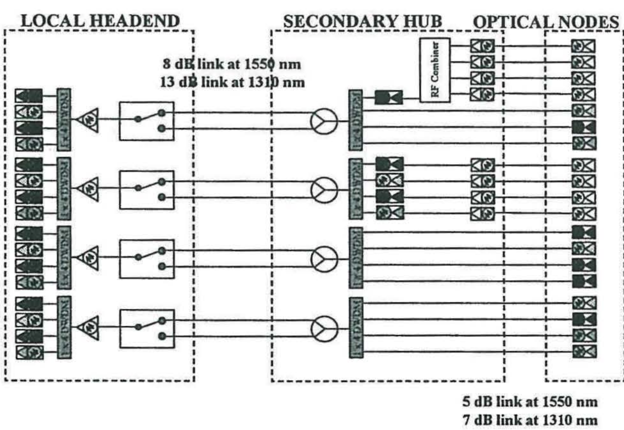


Figure 10: ITU Lasers from Nodes Multiplexed in Secondary Hub

The DWDM techniques can be used with ITU grid lasers installed only in secondary hubs and optical node feeds being repeated (and combined before the ITU grid lasers if needed) or with ITU grid lasers installed in the nodes with secondary hub serving a role of passive multiplexer.

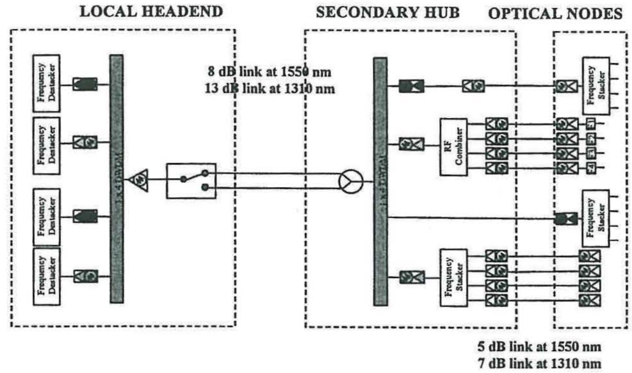


Figure 11: FDM from Nodes with FDM and DWDM at Secondary Hub

Frequency stacking can be performed in nodes or in secondary hubs. Alternatively, individual small nodes can be frequency shifted to different bandwidths and RF combined at the secondary hub to feed ITU grid lasers. The frequency stackers in the nodes can feed ITU grid lasers for passive wavelength multiplexing in the hub or feed 1310 lasers for re-lasing into ITU grid and then multiplexing in the hub. Small nodes can be RF combined in the hub after detection and before frequency stacking.

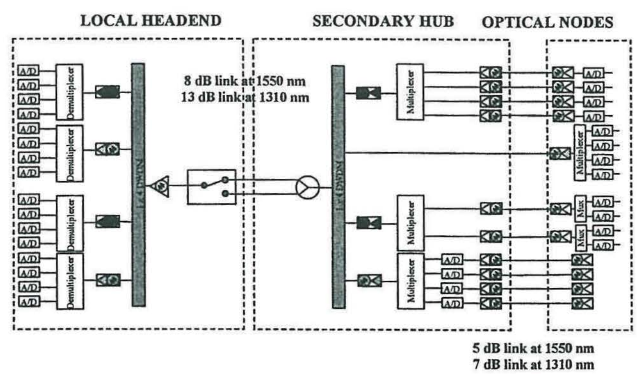


Figure 12: TDM from Nodes with TDM and DWDM at Secondary Hub

Similarly, digitization and TDM can take place in nodes or secondary hubs. Moreover, for small nodes with digitization, TDM can take place in the hub (and initially can be equivalent to RF combining). Additionally, after some multiplexing in the nodes, additional multiplexing into higher rates can be performed in the hub. Digitized and multiplexed signals in the nodes can feed ITU grid lasers for passive wavelength multiplexing in the hub or feed 1310 lasers for re-lasing into ITU grid and then time division and/or wavelength multiplexing in the hub. Small nodes can be RF combined in the hub after detection and before digitization.

Access Network

AT&T analyzed alternative architectures for access network (downstream of secondary hub). The analysis results[8] show that fiber-deep architecture (MFPC) may be the architecture of choice for HFC network upgrades. In this case, mininodes will be deployed and fed off today's nodes (MuxNodes). All the multiplexing techniques are applicable to this new architecture for both phases of deployment and can be implemented in secondary hubs, MuxNodes and mininodes in any combination.

COST COMPARISON

Model Description

The costs of the alternatives were analyzed for a secondary hub with 20,000 homes passed. The detail model description is presented in Table 1. The architecture presented in Figure 4 was used as a baseline for cost comparison. Incremental costs (savings) for all other alternatives were calculated on a per node basis, per serving area basis, and per home passed basis.

Model Parameter	
Area Served	Secondary hub
Homes Passed	24,000
Number of Nodes	32
Number of Nodes at 300 HPs	8
Number of Nodes at 600 HPs	8
Number of Nodes at 900 HPs	8
Number of Nodes at 1200 HPs	8
Maximum Size of Serving Area	600 HPs
Average Node Size	750 HPs

Table 1: Model Description

Alternatives Analyzed

Table 2 lists all of the alternatives considered during the cost analysis. The high number of alternatives for digital technology combined with TDM supports author's assessment about high flexibility of this technology. A combination of this technology with DWDM allows for building a very robust reverse path with limited number of fibers between primary and secondary hubs. Without DWDM, the number of fibers increases dramatically for this technology unless very high multiplexing rates are used (OC192). Two alternatives for FDM systems and three alternatives for DWDM systems were also analyzed. The most commonly deployed alternative in AT&T HFC network was used as a baseline (presented in Figure 4).

Code	Description	PH-SH Link	SH-FN Link (for node sizes)			
			300	600	900	1,200
DWDM B	Baseline	1550 ITU Lasers for 8-Wavelength Mux	1310, combined @ SH for 600 HPs	1310	segmented into 450 HPs, 1310	segmented into 600 HPs, 1310
DWDM P	Pure DWDM	Passive Mux for 8 Wavelengths/Fiber	1550 ITU	1550 ITU	segmented into 450 HPs, 1550 ITU	segmented into 600 HPs, 1550 ITU
DWDM O	Optimized DWDM	1550 ITU Lasers and Passive Mux	1310 combined @ SH for 600 HPs	1550 ITU	1550 ITU segmented into 450 HPs	1550 ITU segmented into 600 HPs
FS SH	Frequency Stacking @ SH	Frequency Stacking into 16 Chs, 1310 nm Lasers	1310 combined @ SH for 600 HPs	1310	segmented into 450 HPs, 1310	segmented into 600 HPs, 1310
FS SH/FN	FS @ SH & FNs	Frequency Stacking into 16 Chs, 1310 nm Lasers	1310 combined @ SH for 600 HPs	freq. shifted, 1310, combined @ SH	segmented into 450 HPs, FS, 1310	segmented into 600 HPs, FS, 1310
DIG SH	Digitized & Muxed @ SH	Digitized & 2 Chs Muxed into OC48, 1310 nm Digital OC48 Lasers	1310 combined @ SH for 600 HPs	1310	segmented into 450 HPs, 1310	segmented into 600 HPs, 1310
HRDIG SH	Digitized & Muxed @ SH	Digitized & 8 Chs Muxed into OC192, 1310 nm Digital OC192 Lasers	1310 combined @ SH for 600 HPs	1310	segmented into 450 HPs, 1310	segmented into 600 HPs, 1310
DIG/DWDM SH	Digitized, Muxed & DWDM @ SH	Digitized & 2 Chs Muxed, 1550 nm Digital OC48 Lasers, 8-Wavelength Mux	1310 combined @ SH for 600 HPs	1310	segmented into 450 HPs, 1310	segmented into 600 HPs, 1310
DIG FN/MUX SH	Digitized @ FN, Muxed & DWDM @ SH	2 Channels Muxed, 1550 nm Digital OC48 Lasers, 8-Wavelength Mux	digitized, 1310 digital OC24 laser	digitized, 1310 digital OC24 laser	segmented to 450 HPs, digitized, 1310 digital OC24 laser	segmented to 600 HPs & digitized, 1310 digital OC24 laser
DIG/MUX Mix	Digitized @ FN, Muxed & DWDM @ FN or SH	2 Channels Muxed, 1550 nm Digital OC48 Lasers, 8-Wavelength Mux	digitized, 1310 digital OC24 laser	digitized, 1310 digital OC24 laser	segmented to 450 HPs, digitized, muxed into OC48, 1550 ITU digital OC48 laser	segmented to 600 HPs, digitized, muxed into OC48, 1550 ITU digital OC48 laser
DIG/MUX O	Digitized @ FN, Summed or Muxed & DWDM @ FN and/or SH	4 Channels Summed, 2 Channels Muxed, 1550 nm Digital OC48 Lasers, 8-Wavelength Mux, Passive Mux	digitized, 1310 digital OC12 laser, summed @ SH	digitized, 1310 digital OC24 laser	segmented to 450 HPs, digitized, muxed into OC48, 1550 ITU digital OC48 laser	segmented to 600 HPs, digitized, muxed into OC48, 1550 ITU digital OC48 laser

Table 2: Reverse Architecture Alternatives Analyzed

Multiplexing Savings versus Baseline

The unit price data for DWDM and FDM (FS) systems are based on recent quotes or price/volume projections for the next several months. The unit price data for digital and TDM systems are based on preliminary expectations. More accurate data will be available during the NCTA technical conference in Chicago in June 1999. They may seem too optimistic at first but are supported by at least one quote from a vendor. Other vendors indicated that they are active in implementing this technology into the HFC reverse path elements. Due to preliminary nature of the data, no conclusions are presented. However, the results lead to believe that the most cost-effective combinations are DWDM with TDM of digitized reverse bandwidth.

The digital technology components experience steep decline in prices. The prices are driven down by high volume and related to it high level of integration. Most of these components are standard and the only additional requirements are higher thermal stability and reliability for the components installed in optical nodes.

Code	Total Cost	Cost/ FN	Cost/ Serving Area	Cost/ HP	Ref. to Baseline	# of Fibers PH-SH
DWDM B	$389,068	$12,158	$8,842	$16.21	100.00%	12
DWDM P	$331,968	$10,374	$7,545	$13.83	85.32%	12
DWDM O	$335,468	$10,483	$7,624	$13.98	86.22%	12
FS SH	$228,674	$7,146	$5,197	$9.53	58.77%	6
FS SH/FN	$250,474	$7,827	$5,693	$10.44	64.38%	6
DIG SH	$192,266	$6,008	$4,370	$8.01	49.42%	44
HRDIG SH	$113,788	$3,556	$2,586	$4.74	29.25%	12
DIG/DWDM SH	$252,814	$7,900	$5,746	$10.53	64.98%	6
DIG FN/MUX SH	$271,154	$8,474	$6,163	$11.30	69.69%	6
DIG/MUX Mix	$234,114	$7,316	$5,321	$9.75	60.17%	6
DIG/MUX O	$220,304	$6,885	$5,007	$9.18	56.62%	6

Table 3: Cost Comparison of Alternatives

PERFORMANCE AND FEATURE COMPARISON

Transparency

The issue of transparency is essential to HFC networks and allows for elimination of terminal equipment from many locations and its concentration in fewer processing centers and on customers' premises. These characteristics also allow for the addition of new services by adding terminal equipment in these few locations and in the customers' homes. The more transparent the network is, the more practical are retail models for customer equipment and the more efficient self-provisioning of the services becomes.

From among the technologies analyzed, the most transparent is pure DWDM. The other two technologies are reasonably transparent and implement little processing. This processing is transparent to the signals transported on HFC network. At network interfaces (processing centers and customer terminals), the signals are present in their native form and hence are compatible with the terminal equipment.

Flexibility

DWDM technology is very flexible to the changes in frequency allocation. Only a few components are frequency limited but these limits are significantly outside of today's frequency operating ranges. The other two technologies may require (dependent on implementation) and upgrade or replacement. The TDM system is easier to upgrade in case the frequency allocation changes (reverse frequency bandwidth increases). FDM technology would require a replacement and the change might affect other network components (filters).

Reliability

Both FDM and TDM technologies introduce additional active components into the reverse path. Therefore, they inherently lower the reliability. To address this concern, they must be carefully designed for long MTBF. The elements of the DWDM technology have been proven in the field and do not differ significantly from the optical link components that have been deployed in the HFC network for the last 10 years.

Other Considerations

Digitization and TDM technology have one significant advantage: practical transparency of this technology from the point of view of signal performance. The link of a quality not sufficient for QPSK or QAM signals is mostly transparent for digital baseband signals. Hence, the end-to-end performance requirements can be allocated to the A/D converter (the first stage of the digital link). Once signals are digitized, the network becomes transparent to them and does not cause a noticeable degradation under normal operating conditions. Moreover, once digitized, the signals can be processed to eliminate impairments contributed before digitization.

Additional advantage comes from the fact that all components of this technology are standard and widely used by the telecommunications and computer industries. Technological progress is also driven by these industries. Due to the volume, the components are on a steep price curve. Moreover, they are being continuously improved upon.

The combination of this technology with DWDM technology is very promising and allows to find a reasonable compromise between the cost of the reverse path at the cost of transparency and flexibility but with some positive characteristics added to the mix.

On the other hand, FDM systems tested by the author showed a range of performance from almost complete transparency to a significant degradation in link quality (main contributor to the reverse path constraints). They are being developed by a handful number of small manufactures and are highly proprietary. In most cases, they require a pilot to maintain synchronization. Despite these facts, they are a valuable tool in lowering the reverse path cost and allowing for space-saving in locations with limited space.

CONCLUSION

The results of the analysis described in the paper show that, although the pure DWDM technology provides the highest level of transparency and flexibility (future-proofing), a reasonable compromise can be achieved by combing this technology with digitizing the reverse path signals and TD multiplexing them. Although some level of flexibility is lost, additional benefits such as increased link robustness and performance transparency offset this lost. Moreover, the cost analysis based on the preliminary estimates indicates a potential of significant saving in the reverse path network cost. The author is participating in the testing of this technology with several manufacturers. Initial test results confirm the theoretical analysis. The technology is in its prototype stages and may be implemented in the field in the third quarter of 1999.

Frequency stacking and FDM technology, if implemented properly, will also allow for some cost and space requirement reduction without noticeable performance degradation. It introduces similar limitations to network flexibility without the additional benefits coming from digitization and TDM technology. The FDM and frequency stacking technology will most likely be ready for implementation at the same timeframe as the TDM technology.

---

[1] 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.

[2] Tony E. Werner, Regional and Metropolitan Hub Architecture Considerations, SCTE 1995 Conference on Emerging Technologies.

[3] Oleh J. Sniezko, Video and Data Transmission in Evolving HFC Network, 1998 OFC Conference.

[4] Same as [3]

[5] Thomas G. Elliot and Oleh J. Sniezko, Transmission Technologies in Secondary Hub Rings — SONET versus FDM Once More, 1996 NCTA Technical Papers.

[6] Same as [1]

[7] Same as [1]

[8] Same as [1]

SECURITY ISSUES FOR REMOTE ACCESS AND VIRTUAL PRIVATE NETWORKS INVOLVING CABLE MODEMS

Daniel Howard
Digital Furnace Corporation
Atlanta, GA

Abstract

The use of cable modems in the small office/home office (SOHO) market may appear to create special considerations when setting up remote access for employees who are telecommuting and virtual private networks (VPNs) for branch offices or small businesses with distributed employees. Shared access of the medium, multiple users and applications of home PC's, inability of telecommuters to properly administer their home networks, and the potential for multiple cable modems in a premise are all issues which impact the threats, policies, architectures, and solutions for secure networking using cable modems. In this paper, the key security issues, sample architectures for VPNs involving cable modems, cable modem security mechanisms (such as Baseline Privacy Plus), and approaches for providing secure remote access and VPNs involving cable modems will be discussed. While the suspected challenges have to do with privacy over the cable network, it will be shown that with the advent of Baseline Privacy in DOCSIS cable modems, the most likely problems have to do with developing and implementing the security policy as it relates to home networks.

Introduction

Cable modems are gaining in numbers and the delays in the deployment of alternate technologies such as xDSL will create a tremendous revenue opportunity for using cable modem networks to address the small office/home office (SOHO) market. Unfortunately, one of cable's perceived drawbacks is data integrity and security over cable modem networks. Further, since cable has the capability to address previously untapped segments of the SOHO market, the types of users, their skills, bandwidth and security needs will vary so widely that much of the challenge will be in marketing to the various customers.

It is the duty of technology developers is to ensure that technology solutions exist and are cost effective for providing whatever capabilities are required and enumerated by marketing personnel. The developers of the DOCSIS standard and cable modem vendors have embraced this philosophy and have provided an arsenal of technologies that can be used to guarantee that cable modem networks have security features that meet or exceed those of alternate technologies. But understanding how each of the available technologies relates to the customer's needs and cost constraints is a new problem for cable modem service providers.

The purpose of this paper is to explore the security technologies and issues associated with using cable modems for the SOHO market. In particular, the available security technologies will be applied to the following two applications: remote access to corporate networks and virtual private networks (VPNs). It will be shown that the capability to securely employ cable modems is readily available, but that the presence of the home in the network creates unique issues that are best addressed via development of proper security policies.

Introduction to Network Security and Security Technologies

The basic security concepts that impact cable modems for the SOHO market include security types, needs, threats, policies, and architectures. Generally, the security threats are used to develop a security policy for the individual or corporation, and the policy dictates the security architecture and the technologies, including both hardware and software, that are required to implement the policy. The policy also includes specification of the types of security to be provided, including privacy, authentication, data integrity (including protection against virus attack), accountability, and robustness. Security robustness defines conditions for non-repudiation of service and the ability to detect and characterize security attacks. Security technologies employed to implement security policy include authorization, authentication, certification, and public and private key infrastructures. Finally, the security architecture includes the functional architecture implementing security, as well as component networks and software to be addressed. For example, if remote access via cable modems is permitted as part of the general security architecture, then it must consider access network security as well as end to end security.

First, consider the security threats. A beauty shop using a cable modem for high speed internet access so that customers can access images of hairstyles may have no real threats to data security if the cable modem client computer is not connected to the computer used for record keeping, billing, and so on. At the other extreme, a large corporation in a highly competitive, international field with government contracts may have threats that include independent hackers, political extremists and groups thereof, industrial espionage and virus attacks, foreign government organizations, and former or current (but disgruntled) employees. It is interesting to note that approximately 70% of corporate security breaches have been from inside the network rather than from outside.

The threats and associated risks and consequences are used to define a security policy which enumerates the security needs of the corporation, the techniques and technologies used to meet those needs, and the process by which the security procedures are monitored, managed, and updated. Further, security policies determine who gets a given level of access, how passwords are distributed and updated, and how new and terminating employees are handled by the security system. The aspects of the security policy of greatest interest for cable modem networks are those which define the forms of remote access permitted, the types of authentication and authorization that are used, and how digital certificates, certification authorities, and public/private key infrastructures are used. Policies for traveling employees will also be relevant to the present discussion if the traveler is accessing his home or branch system via cable modem.

While ideally, cost and convenience of security policies would not enter into the policy specification, it is nonetheless a fact of life that security technologies cost both time and money to implement, and often reduce the ease of use and performance of applications that are running in secured modes. In order to make tradeoffs in cost, convenience, and security achieved, the actual technologies involved must be considered, since computing performance can depend heavily on which layer of the OSI model has the security implementation, as well as how it is implemented. For example, data link and physical layer security measures for filtering the packets attempting to pass into inside the trusted network from outside the firewall will generally produce higher speed and performance in client applications, especially if the security can be implemented in hardware. But hardware implementations can often be limited in flexibility and are not easily upgraded when new technologies are available.

Hence, in order to understand the tradeoffs in cost and performance versus security provided, we must examine the specific technical functions provided by security technologies. We begin with authentication. Authentication in its most generic sense is a means of identifying individuals and verifying their eligibility to receive specific categories of information. In data networks, authentication is the act of identifying or verifying the eligibility of a station (e.g., a cable modem), originator, or individual to access information. Authentication establishes the validity of a transmission, message, station, or originator, and a provides positive identification with a degree of certainty sufficient for permitting certain rights or privileges to the person or station positively identified.

One type of authentication results in access to network resources, and usually starts with user ID's and passwords, but can ultimately result in a chain of challenges and replies involving exchanges of encrypted data and digital signatures. The other type of authentication is source authentication, which usually involves imbedding digital signatures in documents, files, email, and other data to ensure that the items transmitted were from the stated source and have not been altered. Each type of authentication has a process associated with it that varies in complexity depending on the level of protection desired.

For access to network resources, simple user ID and password authentication, such as Telnet and FTP.TCP/IP for TCP/IP networks, have many limitations and vulnerabilities, the majority of which relate to the ability of hackers on the Internet to easily receive unencoded packets with such information. More advanced methods such as Kerberos, cookie implantation during initial registration, public key cryptography systems, the DOCSIS registration procedure for cable modems and Baseline Privacy Plus for cable modems use some form of encryption and are thus more secure. In another technique used in authentication for remote access services, the client computer hangs up after initial logon via dialup connection and is called back at a prearranged number stored in an authentication database. Hardware devices (or authentication equipment) can also be used, where secret algorithms and keys are used in the device to convert a user input into a response that is recognizable only by another machine with the same hardware device. Other implementations of network authentication include cryptographic authentication, time based authentication (access can only be granted at specific times or for specific time durations), peer-entity authentication, self-authentication, and smartcard and/or token authentication.

The heart of all advanced systems used for authentication is encryption, where the methods and algorithms used for encryption vary, as does the way in which cryptographic keys are obtained and how they are used. For example, one of the main differences between techniques such as Kerberos and public key cryptography systems is that the former relies on a trusted third party, the Kerberos server, which if compromised, opens up the entire system to attack. Key based cryptography systems, on the other hand, rely on stored keys which if compromised, only betray the user who's private key is discovered. There are two main varieties of key based encryption: secret key (or symmetric key) systems, and public key systems. In secret-key cryptography, the same key is used for both encryption and decryption. An example of this is the Data Encryption Standard or DES system. Often this system is implemented in hardware to speed up performance.

Key pair systems, usually called public key encryption, require the use of two keys, a public key for encryption and a private key for decryption. A private key can also be used to sign items for the purpose of source authentication, with the public key being used for said authentication. An individual can also encrypt their own items with their public key so that only they can open them.

The so-called Pretty Good Privacy or PGP system uses a public-key system which employs IDEA encryption and RSA encryption. The Digital Signature Algorithm (DSA) is another popular public-key technique used for signatures. There are also cryptosystems based on elliptic curves, and a key agreement protocol called Diffie-Hellman for establishing secret keys over an insecure channel.

Certificates are often used in public key systems to verify that a person and a public key are correctly associated. The most important information in a digital certificate is a public key and a name. Often a certificate also contains an expiration date, the name of the certifying authority, a serial number, and other information. It can also contain the digital signature of the certificate issuer. During authentication, a chain of certificates can be created, each one certifying the previous one until the parties involved are confident in the identity of each other.

Once the user is identified and authenticated, the network authorizes the user and grants to a user, program, or process the right of access to the network resources requested. The authorization provided can depend on the level of access requested by the user, the user type or category, time of day, network loading at the time of authorization, availability of human operators to validate the authorization, and whether or not network intrusion has been recently detected.

All of the preceding is of course specified in the security policy, in which is also specified the security architecture. A proper security architecture should not rely on any particular encryption scheme, but rather be able to insert new encryption schemes as they become available. The reason is that hackers are constantly trying to find ways to break known schemes, and when a scheme is cracked, the security policy should dictate that a new encryption scheme is immediately substituted.

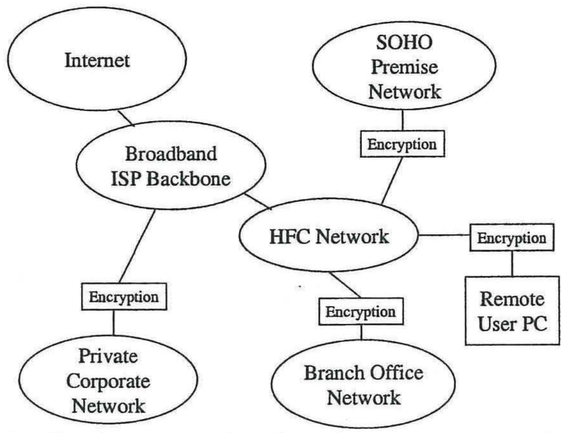
In general, the security architecture specifies the components of the system, the interfaces between components and with the outside world, and desired scenarios and permitted layouts for interconnection of components. The location of trusted and untrusted networks/entities must be defined in the architecture, as well as whether access to the trusted networks/entities is via a single point or via multiple locations. In the case of large corporate networks, it is quite common to have several access points to the network, each of which is protected by a firewall. Small offices, on the other hand, may only have a single access point to the outside world, and the premise of this paper is that this access point will increasingly be provided by a cable modem.

Likewise for homes, although it must be recognized that both single and multiple access architectures are possible. Many homes currently have multiple PC's with analog phone modems. Homes with digital video and cable modems will likely have at least two cable modems: one will be in the home office cable modem and another in the digital set top box. It is nontheless desirable to have a single access point in the customer premise, since this provides the maximum control over the security in the premise. For example, a security architecture developed for homes [Calvert1] uses a single gateway access point and contains the following elements: devices, terminals, authentication mechanisms, identity validator, gateway, enforcement engine, introduction mechanism, policy database, secure channel, participant, and demarcation point or threshold (where the inside, trusted home network meets the untrusted outside network). The architecture specifies how the elements are used and interconnected to provide security for a home network with broadband access. Many of these same elements apply to the case of a SOHO premise in which workers desire secure access to and from the premise. But to characterize how security plays a role in such applications, we must first explore the network and security technologies applicable to SOHO networking involving cable modems.

Network and Security Technologies for Cable Modem Networks and VPNs

First, let us define some network architectures for SOHO applications using cable modems. The two key applications are remote access to a corporate network and a virtual private network or VPN. A VPN is a network that connects remote offices or employees to a private corporate network through a commercial internet service provider, or ISP, instead of through the more traditional private network. For the purpose of this paper, we assume that the commercial ISP is a high speed cable access provider, such as the At Home corporation or ServiceCo. VPNs can involve services such as remote access, data, fax, voice over IP (VoIP), and video conferencing.

The figure on the next page shows how cable modems can be used in remote access and VPN applications. In the most general case, the private network to which the VPN is connected can run IP or non IP packet traffic such as IPX, AppleTalk, SNA or DECnet. The most common method by which the VPN is connected to the corporate or private network through a non secure IP network is IP tunneling, where packets on the virtual network are encapsulated in IP packets for transport over the public network.



Usually the encapsulated packets are encrypted using techniques described previously. Examples of VPN tunneling approaches [Cabletron 1] include:

1. IP tunnels between a remote user and a corporate firewall with tunnel creation and deletion controlled by the user's computer and the firewall
2. IP tunnels between an Internet service provider and a corporate firewall with tunnel creation and deletion controlled by the ISP
3. IP tunnels between sites over the public Internet
4. IP tunnels over a service provider's backbone IP network that is separate from the public Internet

VPNs based on IP tunnels can be deployed and managed by either the corporate user (especially if there are no quality of service requirements), or they can be deployed and managed by the cable ISP. The latter is usually done under a service level agreement that ensures a minimum quality of service to support real time applications such as voice over IP (VoIP) and video conferencing. For option 4 above, the strength of cable ISPs is that they can offer shared bandwidth with complete control over both the IP backbone and the cable access network.

There are several different tunnel protocols available [Cabletron2], including Level 2 protocols such as the Point to Point Tunnel Protocol (PPTP), the Level 2 Tunnel Protocol (L2TP), and a level 3 tunneling protocol, IPSec, which is an effort by the IETF to add standards-based authentication and encryption to TCP/IP. IPSec is an evolving set of specifications for cryptographically-based authentication, integrity, and confidentiality services at the IP datagram layer. IPSec is specified so that many methods of encryption and key management can be supported. Although IPSec is the trend for VPN systems, the standard is still evolving and all vendor equipment employing IPSec is not yet interoperable.

Authentication in VPNs is often via estlablished methods such as Radius (Remote Authentication Dial-In User Server, of the Terminal Access Controller Access System (TACACS), although there are also proprietary methods used for authentication. A firewall based VPN can also be used, where the firewall performs address translation, user authentication and user authorization, in addition to the well known packet filtering function. While there are often significant performance degradations associated with firewalls (lack of support for some services, slow throughput due to using a host based operating system), firewall-based VPNs are probably fine for limited numbers of simultaneous remote access users or relatively small amounts of traffic passing site to site over the network.

There are also software based VPNs such as those designed for use on NT server which are flexible enough for support of both VPN traffic as well as non VPN traffic (e.g., web surfing). The faster encryption performance of hardware systems coupled with the greater protection often afforded by hardware based encryption has also led to hardware-based VPNs. These use hardware encryption modules in between the CPE computers and the access point (cable modem) and which can be built into routers or combined with software systems such as firewalls and authentication servers to provide an integrated solution for the SOHO market.

VPNs with cable modems can also be set up entirely over the cable modem network when all offices, branches, and homes involved have access to cable. In this case, IP tunneling may not be required since the cable network will soon support DOCSIS security measures such as the Baseline Privacy Interface (BPI), a method of encrypting data for transport between the cable modem (CM) and the cable modem termination system (CMTS). In BPI, data is encrypted between the CM and CMTS at the MAC sublayer (which is based on the DOCSIS standard for cable modems). While BPI used only the 6 byte MAC address to authenticate the cable modem to the CMTS (and thus has limited protection against theft of service), BPI Plus addresses that vulnerability by adding digital certificate-based CM authentication to its key exchange protocol.

In BPI, the service ID (SID) is used to identify a security association in the CM. During initialization, the CM requests an authorization key via transmission of the CMs MAC address, the CMs public key, and a list of unicast SIDs corresponding to provisioned class-of-service settings configured for BPI [Judge1] [BPI+]. BPI+ uses a variety of encryption techniques for various stages of the authentication, authorization, and encryption of actual data traffic, including RSA for encryption of authorization keys (using the CMs public key that, along with its RSA private key, is factory installed in the CM). The US Data Encryption Standard (DES) is used for for traffic encryption and for traffic key encryption, although different modes are used for each. Finally, note that while the entire packet data unit (PDU) of the cable modem packet can be encrypted, the MAC header and portions of the extended header are not encrypted.

The result is that DOCSIS cable modems now have the ability to support privacy over the access network that is equal or better than the privacy afforded by the public switched telephone network or leased lines, even though the RF downstream signals of cable modems are seen by other stations. Encryption in both the access network as well as in the IP datagrams provides a level of security that is only limited by the strength of the underlying encryption algorithms.

Security Issues for Remote Access and VPNs Over Cable Modems

A widely publicized "flaw" in the At Home network's cable modem service was the folder sharing capability of Microsoft Windows which permitted other cable modem customers to see files on an individual's internal network if he had folder sharing turned on [Pelline1]. The At Home network administrators had turned this feature off of all systems connected to the cable modem, but users turned it back on in order to set up home networks. A quick solution was found by launching a software upgrade which disabled external file sharing but not internal network file sharing.

This security event points out what is felt to be the most serious challenge to home based cable access for remote access to corporate networks and VPNs: the home environment itself. For example, there is a security issue with home networks that have multiple users on the a single system connected to the cable modem who are not employees of the company. Another issue is the use of home and office networks that employ wireless LAN systems in addition to wired LAN systems. The security of wireless LANs is generally not as extensive as that of the cable modem access network using Baseline Privacy due to the latency and performance degradations which attend more secure techniques. Hence, the use of wireless networks in the customer premise for either data or voice should be seriously considered in the security policy, since it can become the Achilles' heel of the network.

Multiple users on a home based system creates potential problems for employees who typically post their password and user ID near the computer, making it possible for others in the house to gain access to the VPN. The most prudent remote access approach would be for corporate users to use their office laptop computers at home when connecting and state in the security policy that no other home users can use the company laptop. Failing this, the security policy should dictate strict rules for how the home user accesses the corporate network and how he or she should prevent unauthorized access within their home. Over time, there will likely be configuration-oriented methods to prevent other home users from accessing corporate resources, such as logon procedures which rely on smart cards carried by the employee, voice identification, and so on. Many of these tactics are in their infancy now but are expected to be refined over time as more experience with home based high speed access is gained. It should be noted that the 'always-on' characteristic of cable modems creates unique access control problems in that home systems are generally more open to casual visitors than office based systems. Again, it is felt that as the desired capabilities are defined in detail for small office/home office systems used for remote access or VPNs, proper security policies can be developed which address the security needs and provide the required level of protection.

Conclusions

The most important conclusion is that technology exists to make remote access and VPNs using cable modems as secure as necessary for most corporate applications. It is really the behavior of humans in their homes that creates the challenges for cable modems, and indeed any broadband access technology which addresses the SOHO market.

MSOs or cable ISPs interested in addressing the SOHO market will likely have two main types of users: those seeking to set up their own VPN and remote access, and security facilities, and those who wish to rely on the cable ISP to set up, monitor, and maintain the system. For the first type, Cable ISPs should nonetheless work with the customer to assist them in developing a security policy so that when flaws in the policy lead to security breaches, the onus is on the corporate user to update their systems. For the second type, cable ISPs should develop an arsenal of solutions to offer with tradeoffs between security, cost, and performance and services provided. In this manner, the cable ISP can address the many types of remote access and VPN customers for cable modems which are likely to exist.

Further, since cable ISPs rely more heavily on the underlying encryption technologies for security over the access network, they should aggressively seek upgrades of security technologies that address known flaws or enhance the network performance of shared access security systems.

Finally, whatever solution is used, the cable ISP should include monitoring and detection of security events as key to the long-term process of maintaining the integrity of the overall cable modem network. This will assist in developing the perception that cable modem networks are as secure as alternative technologies that are similarly priced. Since VPNs in particular will likely be deployed using a variety of technologies such as xDSL, wireless, and conventional telco solutions, the cable ISP must be prepared to use a thorough understanding of the technology to market it effectively to SOHO customers.

References

[Calvert1] Calvert, K. and Aylesworth, D. "Toward a security architecture for home information infrastructure, Broadband Telecommunications Center technical report, summer 1998.

[Cabletron1] "What is a virtual private network (VPN)?" Cabletron white paper, http://www.cabletron.com/vpn/VPNwhatis.htm.

[Cabletron2] "A ten point plan for building a VPN" Cabletron white paper, http://www.cabletron.com/vpn/VPN10ptplan.htm.

[Judge1] P. Judge, "A review of the baseline privacy interface of the data over cable system interface specification (DOCSIS)" final report CS7100, http://www.cc.gatech.edu/~judge/cs7100/project03/cable.htm

[BPI+] "Baseline Privacy Interface Specification, Draft" SP-BPI-D01-990212, 2/12/99.

[Pelline1] "@Home user files may be at risk", J. Pelline, http://www.news.com/news/Item/0,4,15924,00.html.

SELF SERVICE ACTIVATION
A FIELD STUDY OF HIGH-SPEED DATA (HSD) SELF INSTALL

Bruce F. Bahlmann
MediaOne Internet Services Group

Abstract

In September of 1998, MediaOne began a field study of self install. This study was unique because the customer actually performed the wiring necessary to activate HSD services. This paper will review the details and findings of this study and provide direction for those seeking to further their own self service program.

SELF SERVICE ACTIVATION

Background

An ever-present fact in our business is that Multiple Subscriber Organizations (MSOs) must continue to grow their subscriber base. However, to insure profitability, MSOs must begin to streamline customer activation and technical support obligations by augmenting personnel with technology wherever possible.

Figure 1.0 represents the complex sequence of events that each potential customer must experience. Today, many MSOs must rely on their internal employees to process each phase of the customer care process. If unaltered, this single fact will continue to limit their ability to scale the business.

The long term goal of many MSOs is to provide an efficient self service environment for its customers. Although this self service environment is initially intended for high speed data (the focus of this paper), it could eventually support all products that utilize the broadband pipe (i.e. HSD, core video, telephony, etc). The self service environment is not intended to replace "traditional" MSO installations with self service installations. Instead, the self service environment merely provides additional installation and support options for its customers. The combination of the proposed self-service environment as well as other MSO employee driven efforts to install and support customers will provide a more scalable and cost effective solution for expanding the Internet service customer base. The result of customers choosing some or all of these options will be an increase in overall customer care efficiency. The following list highlights some possible metrics that could be used to develop a self service program:

Figure 1.0 Customer Care Process

- Installation Time
  Time it takes for an MSO field representative to complete a high-speed data installation.

  The addition of a self service (self install) should impact average installation time because the more a customer can do on their own the less time MSO personnel are involved.

- Average Installation Cost
  The average cost to an MSO for each subscriber installed.

Since the average installation cost represents ALL high-speed data installations, this figure should decrease as the number of self installations increase (see note below).

- Percentage of Customers Self-Installed
  The portion of the total number of high-speed data installations that were completed entirely by customers.

  The availability of a fully functional self service environment should enable an increasing number of self installed customers.

- Support Calls per Customer
  The average number of support calls taken per customer.

  As the number of maintenance tools become available through the self service environment the average number of support calls per customer should decrease.

Note that building this technology (i.e. self service environment that supports self installation) does not in itself guarantee that it will be used by either new or existing customers. Therefore a more in-dept analysis should be conducted to determine what actions an MSO should take to encourage the use of self service environment and self service activation (i.e. marketing and other incentives). The field study outlined by this paper merely provides an introduction into the dynamics of self install. It is strongly recommended that further study in this area be combined with developing applications that provide self service functionality (i.e. activation and maintenance tools).

A self service environment must address all aspects of the customer care process and allow customers to independently select and then activate their Internet or other services. Once activated, the self service environment must enable subscribers to make specific maintenance modifications to their Internet service account without a call to a MSO's Technical Support Organization. Subscribers should have the ability to perform all necessary activation and maintenance operations via a convenient platform independent (web browser) application. While this document primarily focuses on self service activation, most of this functionality can also be used to provide self service subscriber maintenance (changing e-mail passwords, swapping network cards, etc). Achieving self service subscriber activation involves automating existing "manual" activation tasks in Table 1.0. It does not include wiring the home and installing the network card in the subscriber's computer. This will be the customer's responsibility (self install) or a MSO's installation technician's responsibility (traditional installation).

Task Type:	Description:
Manual	Wire outlet (if necessary)
Manual	Install NIC
Manual	Create a user ID
Manual	Provision services
Manual	Active service
Manual	Test service
Manual	Train customer
Manual	Complete paperwork/check-in

Table 1.0 Activation Tasks

It is important to understand that simply allowing a customer to activate their Internet service on their own is *not* synonymous with the primary goal of self service activation but rather just an additional way of installing a HSD customer. In the end, the *best* self service activation program is one that supports a variety of self install options for customers with varying amount of support from the MSO. A phased approach could continually add self install options that provide a decreasing degree of MSO support.

The initial phase of this plan was to complete a field study to learn more about the dynamics of self install.

FIELD STUDY

The objective of MediaOne's field study was to provide a means by which technically skilled individuals could wire, install, configure, and activate their HSD service on their own. This option supplies qualified individuals (skilled in wiring and computer configuration) with an install kit and instructions that they could check out from a MediaOne service center. Next they went home and installed the service at their convenience. Once complete, all customers completed a survey in which they were able to provide feedback on their comfort level with the self install. The survey results were used to help determine the proper costs (i.e. motivation/incentives mentioned earlier) for this option as well as how the option may be improved.

Note that a by-product of the field test enabled MediaOne to begin looking at comparative costs of traditional installs vs. self installs to determine what cost MediaOne incurs for each of these options, what are the tradeoffs and any potential savings. This subject will not be addressed in this cover but was explored as part of the study.

Field Study Environment

The field study took place in MediaOne's Minnesota region. The Minnesota region consists of approximately 571,000 homes passed, all of which are HSD capable. Minnesota had recently launched one-way HSD services which were available to all the 571,000 homes passed. Minnesota's one-way HSD service consisted of using a one-way cable modem that received its downstream bandwidth from the cable line with a telephone line to accommodate its upstream bandwidth. The maximum possible throughput of the HSD product was approximately 1,500,000/33,600 bps.

Due to the recent launch of HSD services, new signups were being scheduled up to three weeks in advance. Since a majority of the "earlier-adopters" were technically savvy, many of them asked if they couldn't just do this install themselves. This initial demand by customers prompted MediaOne to begin the field study to help these customers get HSD more quickly, and to answer lingering questions about the viability of self install.

A project plan was developed to conduct a self install field study which included documentation and coordination for the involvement of MediaOne telephone representatives (Tier 1) and field fulfillment.

Goals and Anticipated Results

As MediaOne entered into the study there were two goals in mind:

- Save the company money
- Increase the number of installs the same number of field technicians can do in a day

Self service is a high profile topic within MediaOne and this field test was greeted with much enthusiasm.

Self Install Process

Upon approval of the project plan, more detailed processes were developed to assemble kits, create a survey, and manage the distribution of the self install kits. As a result, several steps were defined for activating a self install. These steps are as follows:

- Qualify customer skills/dwelling

- Initialize billing account and HSD services (select email names, etc)
- Schedule a "traditional" HSD install (see fail-safes below)
- Arrange for pickup
- Customer comes to service site to pick up kit
- Customer signs service contract and equipment release form
- Local coordinator reviews install steps with customer.
- Local coordinator ensures customer's account is established and "ready" (will enable customer to receive Internet service once install is complete)
- Customer returns home to install kit
- Once complete, customer completes on-line survey
- Completed survey informs Tier 1 who closes the install (No Truck), cancels the "traditional" HSD install, and activates the discounted service
- Tier 1 sends notification to customer informing them they received their survey, cancelled their scheduled install, and have applied a discount

Qualifications

Since the installation documents were not geared for casual to average computer users (many troubleshooting steps were NOT included), some level of qualification screening was needed to ensure that every customer was technically skilled to carry out the installation process and that their home was capable of supporting the service (i.e. already had CATV). The following is a sample of some of the questions asked customers to ensure they would be able to perform a self install:

- Does customer currently have MediaOne CATV?
- Is customer comfortable opening a computer?
- Is customer experienced in resolving IRQs?
- Does customer know how to back out of changes made on a computer?
- Does customer have a browser installed in their computer?
- Does customer know how to terminate CATV wire?
- Does customer have access to the necessary tools (listed in manual)?
- Would customer like to participate in the test, complete survey, etc?

If the customer answered yes to all these questions they were accepted into the self install test group and proceeded with the next step in the self install process explained previously.

Fail-Safes

Additionally, several measures or fail-safes were developed to ensure that the self install ended with a satisfied and working customer. The following fail-safes were put in place:

- When the customer selected the self install option they were scheduled for a "traditional" MediaOne HSD install. This date served two purposes. First it preserved this customer's installation priority allowing them to be installed at that date if they elected not to proceed with the self install or were not able to complete it. Second, it served as an end point for the customer's potential opportunity to be part of the test group and thus be eligible for the discount. If the MediaOne install date had arrived and the customer had completed all or part of the install (successfully and correctly) they would only be billed for the portion that MediaOne had to complete, but would not be considered part of the self install test group and would not receive the discount.

- If the customer was unable to complete the install due to some other reason (computer was buggy), they could request a service call from MediaOne to complete the install. In this case, if the service call was minor the customer would just be charged time and materials but would remain a test group participant and thus receive the discount. If the service call was more like a full install, the customer would be charged for an install and be excluded from the test group and its discounts.
- If the customer received the kit but decided that they no longer felt comfortable with the CATV wiring, they could elect to have MediaOne install an additional outlet (AO). Since an AO service order would come from the core video side of MediaOne this request could likely be serviced before the HSD install date. The customer would have to complete the CPE installation and configuration and ensure the CATV was connected per the self install kit instructions. Once complete, the customer would only be charged for an AO service call which is generally cheaper than an HSD service call.
- If, for whatever reason, the customer suspected that any component in the self install kit was defective, the customer could freely replace the component at any MediaOne service site. To replace the components, a customer would present their self install agreement and the suspected defective part to receive a new one.

Providing fail-safes allows more people to become part of the test group while incurring minimum charges in exchange for MediaOne performing some parts of the install. As a result, the self install would still represent a savings to the customer however, it would be less of a savings had the customer performed everything without requiring MediaOne intervention.

Self Installation

An important part of the qualification/scheduling process was that the customer selected an email account over the phone. Once this email account was created, the customer was effectively pre-authorized (advanced authorized) to use MediaOne Internet service. The email username and password selected by the customer allowed them to access the service from their home once cabling and computer configuration was complete. This step was confirmed by the local coordinator before the customer left the service site. Additionally, the Tier 1 person recommended the customer do some general measuring in their home before coming to pick up the kit. The general measuring allowed MediaOne to provide the customer with lengths of wire most suitable to their dwelling. MediaOne offered to provide the customer with any length of wire they felt they needed to proceed with the install.

During the actual installation activity, the customer merely followed instructions contained in the kits and connected the "pre-terminated" wire lengths as instructed. Since the RF portion of the self install contained the most critical steps, highlights of these steps are covered in this document. The goal of the RF installation doc was to have the customer make "home-runs" from where the MediaOne cable entered their premises (MediaOne de-mark was the ground block) to the room where the cable modem would be placed (see Figure 2.0).

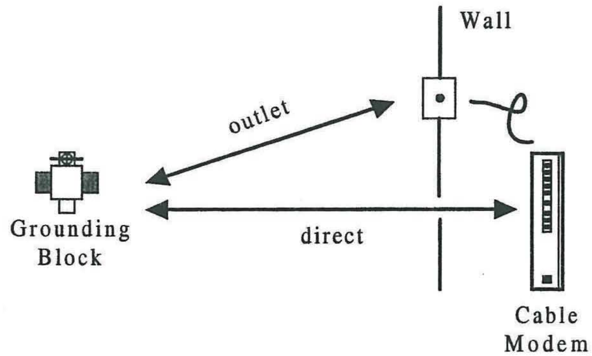


Figure 2.0 Self Install Cabling

By making these cable runs and following additional instructions the dwelling would be capable of both one-way and two-way cable modem service. At the de-mark, the wire leading to dwelling's video distribution system (some kind of splitter) was re-routed through either a direct coupler (DC6) or a standard two way splitter depending on signal strength (see Figure 2.1).

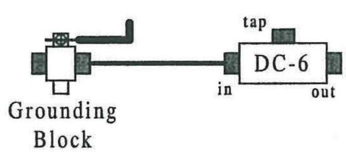


Figure 2.1 HSD Splitter Connections

The splitter/DC6 was then connected to the home run lead from the cable modem and a high-pass filter connected to the remaining receptacle (see Figure 2.2).

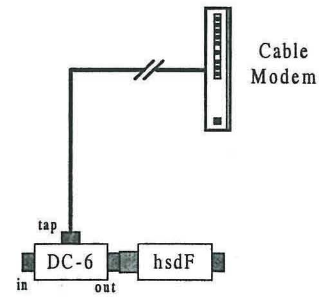


Figure 2.2 Use of High Pass Filter

The use of the high pass filter (an ingress or leakage preventative device) prompted two very strong opinions on the subject. There are those that believe that "good" RF plant operations and maintenance can prevent the need to use filters. There are also those that believe that trying to maintain RF without using filters is unachievable (see additional note below). At present there is no strong evidence that either opinion is correct. However, since MediaOne already is committed to using filters, this practice was rolled into the RF installation instructions.

Note that the benefits of using filters can be overshadowed by the problems they cause for self service. Use of a high pass filter anywhere between the customer cable modem and the broadband plant can terminate the HSD service. It can also prevent self service installs if this filter is placed out-of-reach of the customer (i.e. at the pole). In the proposed next steps section of this document, some ideas are discussed which address this problem and allow high pass filters and self service to exist in harmony.

The other side of the high pass filter is connected to the video distribution system (see Figure 2.3). Note that this area between the high pass filter and the video distribution system is where traps and amps can be placed (use of traps and amps are use by some MSOs to block premium channels or boost signal respectively).

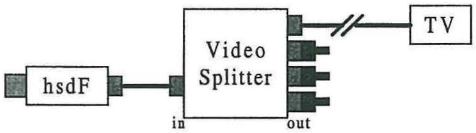


Figure 2.3 Video Distribution System

RESULTS

The self install test program was closed after 83 customers had selected this option, proceeded through the process as it was defined, and became regular paying MediaOne HSD customers. Of the 83 customers only one customer was unable to perform the tasks necessary to activate Internet service. The good news is that this person still proceeded through the fail-safes and later became a regular HSD customer. The results of the remaining customers who actually completed the self install and the survey are explained in detail below.

Survey Summary

Self installation of MediaOne Internet service was a positive experience for its test participants. Surprisingly the ease of the RF installation portion of the self install ranked very close to that of configuring the PC. Cable modem and browser configuration were the most problematic aspect of the self install.

A majority of the test participants completed the installation within two hours. While slightly more than half of the participants experienced technical difficulties during the installation, less than a third of those found it necessary to contact technical support.

The top reasons for electing a self installation of MediaOne Internet service were fast activation (i.e. participants did not have to wait for MediaOne to install the service) and cost savings (i.e. the installation was free AND participants were given $10 off service for the first three months). These reasons were consistent across the voluntary comments and stated importance ratings.

Participants in the test were highly involved in the PC category. A self assessment of skills revealed that participants had a high degree of comfort with opening a PC, installing a computer expansion card, and installing/configuring PC software.

Test participants tended to own multiple PCs, have at least a Pentium processor in their newest PC, and be employed in the PC industry. While all participants were male, their educational background and age varied.

Survey Conclusions

Self installation of MediaOne Internet service was a positive experience for test participants. Not surprisingly, these participants were highly involved in the PC category (e.g. PC industry occupation, multiple HH PCs, Pentium/Pentium II processors, extremely high comfort level with installing/configuring hardware/software on a PC). It is likely that the positive experience with self installation of MediaOne Internet service was directly related to the capability of the person completing the self installation.

General Conclusions

Advanced Authorization was used to set up the customer's account prior to their self install. This pre-activated Internet account enabled the service to be immediately used once cabling and CPE configuration were completed. Field tests proved this was an effective method of leveraging existing practices to expand the number of ways to install HSD customers.

True "No-Truck" HSD installs are possible in a variety of cases depending on the skill of the customer and the type of dwelling. For example, an apartment can be one of the easiest wiring jobs and one that a majority of customers can accomplish. Likewise, a customer with technician skills often prefers to do the install himself.

Self install is viable and has a bright future. Currently about 5% of all installs could be self installs. Some ways to increase this in the future will be explained in proposed next steps section.

Selection of self install candidates is a key to being successful. Asking the right questions to ensure customer is comfortable is worth the time taken.
Motivation to get customers to try this method must be emphasized. Without it, candidates will select the path of least resistance. It also helps to portray like the program as well established with numerous successes.

CHALLENGES THAT REMAIN

For the average, below average, and beginning computer users the HSD installation represents a significant challenge. Although progress is being made towards simplifying this process, the following represent some of the most underestimated and unresolved issues remaining:

Software (Lack-of)

A common mis conception with regard to self activation is that HSD installs are as easy to install as other dial-up Internet service providers. This is true. MSOs provide a network rather than a dial-up connection. Dial-up connections have been around for years, there are several tried and true install applications that enable a dial-up ISP to configure a customer with a variety of telco-modems to connect to their service. These tried and true install applications do not exist for network cards which are the basis for MSO installs. Instead, MSOs must rely on manual configuration of this hardware, reducing the number individuals who can qualify (skill-wise) for self activation. A tried and true network configuration application suitable for use with HSD installs would be at least a one-year effort and would need to keep up with current network technologies and drivers. In other words, network installs will be more difficult than dial-up installs in the short term due to changes in network cards, drivers, and networking technologies. MSOs must demand network installation tools exist.

In-House Wiring

Most rooms used for residential or other Internet computing already have a telephone outlet. However, CATV jacks for access to HSD or having CATV mean HSD Ready is a problem with new houses being constructed. Building contractors and house designers must be aware of the in-house wiring requirements of a "proper" HSD system that will support TV, HSD, and telephony (i.e. quad-shielded RG6 cable). Until this is viewed and treated as a desirable feature in new homes (especially those that are totally finished out) installing HSD will always be more complex that activating dial-up services in the home.

Universal Hardware Support

MSO HSD also suffers from the lack of hardware support, this is common to personal computers already in homes as well as those currently sold. Since a majority (more than 90%) of all home computers do not include a factory installed network card, the install must provide additional hardware and sometimes software to activate this type of Internet connectivity. The addition of this hardware is problematic as the average home owner does not feel comfortable opening up their personal computer and installing additional hardware. Dial-up customers do not face this problem nor do they face the problem of pre-qualifying a personal computer. Qualifying a personal computer is needed because not all computers can handle any additional hardware (slot or IRQ deficiency). The problem of universal hardware support cannot be solved in the foreseeable future. Only after *all* personal computers come *standard* with a built-in network card, USB, or 1394 (firewire bus) will an end point for universal hardware support be attainable. A workable predictor is 3 years after such time as *all* personal computers come standard with one of the items above. Only then will this problem become less of an issue. This is based on a 3 year ownership of a personal computer before it is desirable to upgrade.

Note as some MSOs offer customers support for multiple CPEs limitations in USB and Firewire functionality will limit these technologies from supporting more than a single CPE. As a result, use of USB/Firewire as a subscriber's only link to the cable modem can prevent additional CPEs from accessing the Internet. The ethernet interface (which requires a NIC) will be the only way a cable modem will support multiple CPEs to access the Internet without the use of some additional software.

Billing Issue

Activating the individual components of HSD service (i.e. CM and CPE) has been a relatively easy process because these devices easily provisioned. The hard part of self activation is associating this process with a billing system. The true *success* of a self service installation program should be based how well it performs the necessary checks and balances to qualify a customer financially and interface with the MSO's preferred billing system. A standard API is needed from MSO billing systems that is vendor independent. Without this API, completing any type of self activation without any MSO personnel intervention is *not* possible. This is why a phased approach must be taken to reach some level of self activation.

Summary

The phased approach enables additional installs without consuming valuable field technician time. Technicians can reserved for completing customer installs for people without the spare time or skills to do this themselves and maintain service for those experiencing problems. MSOs must overcome the challenges that they face to streamline existing install options and continue to expand in to new areas that fulfill customer needs. Because the base of customers who *qualify* for self install is limited, the best option for expanding the number of working MSO HSD customers is to provide many different ways for them to get activated (cater installation options to the skills and needs of our customers).

PROPOSED NEXT STEPS

Need a Strategy

As MSOs continually move in the direction of less employee involved installs (more automation) they will become much more reliant on technology to interact with their customers. An often overlooked fact with regard to self service activation is that it is a MSO's first crack at a new customer and thus familiarizes them with an interface that they can later use for other purposes -- say account maintenance. Therefore, MSOs must look at self service as a way provide things such as brand name reinforcement, cross selling, even advertising!

MSOs that don't capitalize on this initial contact with customer (i.e. offer them ALL their services up front and provide customers comprehensive support, co-branded services, as well as special offers [incentives] to continue to return to this interface) will loose out on a gold mine of opportunity. Thus it is critical that each MSO have a self service strategy in place before driving down the road of selecting a vendor or particular application and getting some kind of self service in place and operational. Self service should involve all areas of the business (i.e. video, pay-per-view, HSD, telephony, marketing, customer care, operations, etc) to ensure adequate planning and that the product the MSO will receive will meet *all* the needs of the business.

Develop Models

A recent survey of several MSO regions could not produce a model for how they propose to install and support various broadband services (such as HSD, telephony, Digital TV, On-Demand Video, etc) together via a single customer drop.

The absence of a model allows for variation in the way each MSO as well as each MSO region supports the same broadband service. However, if a model was adopted by the company and all its RF experts in the regions, there would be fewer (if any) variations in the way various services are deployed, installed, and supported. Standardization would be easier and there would be less work involved in deploying new services if the model was followed and there was no need to customize the service to work within various non-standard regions.

This model for all services could be as simple as a diagram showing how each service will connect to a single customer drop. However, a much better model would be one that went into detail on some of the combinations in the field (apartments, duplexes, homes, commercial, etc). In most MSOs, the way each of these dwelling types is wired varies from region to region and even within a region (because regions usually consist of several smaller cable companies that have merged). Once the model is created the next hurdle is to enforce them. This may mean that a 5 minute service call becomes a 45 minute rewire. However, these efforts to achieve some degree of standardization to a model will pay big dividends in the future as more services are bundled and actually dependent on particular wiring practices (e.g. HSD and telephony). The model should also allow for hybrid wiring (e.g. category 5 and quad shielded RG6). This will provide customers the ability to choose how certain services are installed (perhaps the cable modem is installed in the basement and category 5 is run from there to an outlet). In this instance, customers could then connect MSO products to their existing in-home service (i.e. home LAN).

A model should be a public document and shared with building contractors and inspectors to ensure that standards set by the cable company are followed or at least deficiencies pointed out during an inspection so customers are made aware that their home does not meet the cable company's wiring requirements.

One suggestion that was emerged from the self install field study was the idea of an "HSD Ready Model". The idea here is provide support for HSD in everything an MSO does. If every house was wired for HSD then self install would be supported and thus the home would not require additional service calls to activate certain services. To achieve the suggested HSD ready "status" an MSO could do the following:

- On each service call, ask the customer if they have a computer and would consider getting connected to the Internet via their cable line some time in the future (if not already). If so, the technician offers to wire an additional outlet for free. If the customer accepts the offer, the customer's CATV account is flagged as "HSD Ready" and that customer could now receive literature about Internet service in their cable bill.
- New installs could also provide the same service in offering customers an additional outlet that would be HSD Ready.

Many different field organizations within an MSO could be offering customers an HSD ready AO. MSOs could achieve an increasing number of dwellings that are HSD ready. Over several years of this long term initiative, most of the dwellings would be HSD ready and therefore significantly reduce the average time for an MSO to perform a "traditional" HSD install at that location. Cutting down this time as well as allowing more customers to self install opens up many more install options for prospective HSD customers.

<u>Partner with CPE providers</u>

As more homes become HSD ready (similar to homes that already have a telephone outlet in their home office) it is unrealistic that an MSO will be able to provide in-home support for all its customer's CPEs. Between installation and service calls there is too much demand for MSO field organizations to handle the numbers of service requests. Phone troubleshooting, on-line FAQ, etc. will filter much of the calls but as subscribers increase MSOs will not be able to maintain sufficient numbers of technicians to provide this level of support. Therefore, MSOs must partner with service providers that are more capable of providing this service (emphasis on several providers).

The use of CPE service providers would help MSOs focus on maintaining its broadband network without having to keep up with the number of technicians to cover all installs and service calls. Note that as the number of installed customers increase, the number of installs will gradually decline. Since CPE service will be irregular it will be difficult for an MSO to maintain full time skilled technicians to cover this unpredictable work load.

<u>Seek Out Cable Modem Service Providers</u>

MSOs should be seeking out local providers that intend to carry DOCSIS cable modems and ensure that service on the cable modems and installation guides (or installation kits) are made available to people depending on which MSO provides the customer with CATV.

Service partners can offset some trouble calls especially if the customer purchased a cable modem from a particular retailer. In this case, some of these customer problems may go directly to the retailer rather than the MSO. In fact, the MSO should provide some basic level troubleshooting to the customer to make this determination: MSO (provisioning/network problem) or retailer (bad cable modem).

STATISTICALLY MULTIPLEXING/RE-MULTIPLEXING

Limin Wang and Ajay Luthra
General Instrument Corp.

ABSTRACT

This paper presents a statistically multiplexing/re-multiplexing system for both uncompressed and pre-compressed digital video signals. The system consists of encoders and transcoders plus a joint rate control engine. Hence, it is able to handle both the uncompressed digital video signals and the pre-compressed video bit streams. Specifically, encoders encode the uncompressed video signals while transcoders transcode the pre-compressed video bit streams. The rate control engine dynamically distributes the channel capacity among the programs according to the program relative complexities. Computer simulation results are reported.

1. INTRODUCTION

With recent advance in digital video compression, such as MPEG-2 [1], and digital transmission, it is possible to deliver several digitally compressed video programs in the same bandwidth presently occupied by a single analog TV channel.

Fig. 1 shows a multi-program transmission environment where several programs are coded, multiplexed and transmitted over a single channel. Clearly, these programs have to share the channel capacity. In other words, the aggregate bit rate of the programs has to be equal to, or less than, the channel rate. This can be achieved by controlling either each individual program bit rate (independent coding) or the aggregate bit rate (statistical multiplexing, or stat mux [2,3]). In independent coding, rate control can only be performed across the time and spatial dimensions of a program. However, in stat mux, control is extended to an additional dimension, that is, the program dimension, implying more freedom in allocating the channel capacity among programs and therefore more control of picture quality among programs as well as within a program.

Furthermore, it should also be noticed that more and more pre-compressed video materials, such as films, are becoming available these days. In addition, some applications may require to unbundle the statistically multiplexed video signals and re-multiplex (stat remux) some of them with other video streams. Hence, it is desirable that a stat mux system is able to handle not only the uncompressed video signals, but also the pre-compressed video signals. The problem with the pre-compressed video materials is that they could be pre-encoded at any bit rate, either constant (CBR) or variable (VBR). In order to include the pre-compressed video bit streams in a stat mux system, the rates of the pre-compressed video bit streams have to be *changeable* while multiplexing.

This paper presents a stat mux/remux system for both pre-compressed and uncompressed video signals. The system consists of both encoders and transcoders, as shown in Fig. 2. To squeeze the incoming programs into a given channel, the stat mux/remux system either encodes them by using encoders if they are in raw pixel data, or transcode them by using transcoders if they are in compressed bits. Ideally, the input programs should share the channel capacity (bits/s.) according to their relative complexities. That is, more complex programs are assigned more bits and less complex programs less bits. The stat mux/remux system implements a joint rate control that manages the bit allocation among the incoming programs.

The paper is organized as follows. Section 2 introduces the concept of transcoding by presenting two rate-conversion transcoder architectures for MPEG bit streams. Performance comparisons of the two architectures with direct MPEG are also provided in terms of PSNR. Section 3 describes the stat mux/remux system and reports its performance.

2. TRANSCODER ARCHITECURES

A straightforward rate-conversion transcoder for MPEG bit streams can simply be a cascaded MPEG decoder and encoder, as shown in Fig. 3. In the cascaded-based transcoder, the decoder decodes the input compressed MPEG bit stream, reconstructing the video signal, and the encoder re-encodes the reconstructed video signal, generating a new bit stream. The desired rate of the new bit stream can be achieved by adjusting the coding parameters, such as quantization parameter, Q_2, in the encoder. Note that the quatization parameter, Q_1, embedded in the pre-compressed bit stream is decodable, but not changeable. The main concern with the cascaded-based transcoder is its implementation cost: one full MPEG decoder and one full MPEG encoder.

The cascaded-based transcoder could be simplified if the picture types (I, P or B) in the incoming pre-compressed bit stream can remain unchanged during transcoding. Maintaining the picture types also means maintaining the temporal processing of macroblocks (MB) in each picture (intra, inter, forward, backward, or interplative). Because of the similarity between the original and reconstructed video signals, the motion vector (MV) fields embedded in the pre-compressed bit stream should be reasonable good for the reconstructed video signal. Hence, the MV fields can be used for the MC in the encoder, implying that motion estimation (ME) -- the most expensive operation in the encoder can be removed. Fig. 4 shows a cascaded-based transcoder without ME where the MV fields decoded from the decoder are re-used in the encoder.

In evaluating the cascaded-based transcoder architectures, experiments were carried out for a number of video sequences for different rate conversions. Table 1 shows the results, in terms of PSNR, for two sequences of *Market* and *Trapeze* for two different rate conversions (15 → 3 Mbits/s and 6 → 3 Mbits/s.). For comparison purpose, the PSNRs for direct MPEG at the same final rate of 3 Mbits/s. are also provided in Table 1. Direct MPEG can be considered as benchmark for transcoder. Transcoder introduces an additional quality loss, as compared to direct coding, because the signals passing through a transcoder are actually quantized twice. From Table 1, it should be noted that transcoder without ME in the encoder actually performs nearly as well as with ME.

The reconstructed video sequences at the same final bit rates were viewed side by side for subjective assessment. There were virtually no perceptual differences found between two cascaded-based transcoders with and without ME, implying that cascaded full decoder/encoder can be replaced by cascaded transcoder without ME. GI developed a transcoder architecture for MPEG-2 bit streams derived from the cascaded-based transcoder with re-use of motion vector fields. It has been shown theoretically that its performance is identical to the cascaded-based transcoder with re-use motion vector fields (Fig. 4). But, its architecture is much simpler than the cascaded-based transcoder, and it actually saves many function blocks and memories, as compared to the cascaded-based transcoders (Fig. 3 and 4).

2. STAT MUX/REMUX

A stat mux/remux system is developed for both uncompressed and pre-compressed video programs. Fig. 5 shows the main architecture of the system consisting of both encoders and transcoders. Encoders are used to encode the uncompressed digital video signals while transcoders to transcode the pre-compressed video bit streams. The stat mux/remux system implements a joint rate control scheme, which dynamically distributes the channel capacity over programs according to the program relative complexities. This means that given a fixed channel, the rate assigned for a program not only depends on the program own complexity, but also others. Specifically, at each frame, each MPEG encoder/transcoder l receives a target number of bits, T_l, from rate control engine, as shown in Fig. 5. The MPEG encoder/transcoder then encodes/transcodes the frame at that rate by adjusting the coding parameters, such as, quantization parameter. The average quantization parameter, Q_l, used for a frame and the resulting number of compressed bits, R_l, generated for the frame are then sent to the rate control engine. The product of Q_l and R_l can be considered as a frame complexity measure, and used to update the complexity measure for the corresponding picture type. The rate control engine in turn determines a new target number of bits for the next coming frame of each program based upon the type of the frame, types of other frames in the same program and in other programs, as well as the set of the updated picture complexity measures. Joint rate control with dynamic bit allocation can work with the programs of different GOP structures, and it can also address the possible future changes in the program GOP structures. The bits generated from MPEG encoders and transcoders are multiplexed in Mux engine and moved into encoder buffer for transmission.

Experiments were conducted with joint coding of eight video programs at a total bit rate of 24 Mbits/s, or 3 Mbits/s per program on average. The eight programs include film materials, news, sports, and MPEG2 test sequences. All the eight programs have a resolution of 720x480 pixels and a frame rate of 30 frame/s. interlaced with a color-sampling ratio of 4:2:2. Five of them were pre-compressed at 15 Mbits/s. and the rest three were the uncompressed video sequences. Table 2 shows the processing engine used for each program. The stat mux/remux system (Fig. 5) in experiments therefore had five transcoders and three MPEG encoders *running* simultaneously. Two different GOP structures were chosen for the eight programs, that is, GOP(N=15,M=3) and GOP(N=12,M=3) where N is the GOP length and M is the distance between two P pictures. Table 3 shows the GOP structures for each program. For comparison purpose, the eight programs are also independently encoded/transcoded at a rate of 3 Mbits/s. That is, the uncompressed video sequences are encoded by MPEG-2 encoders, and the pre-compressed video bit streams are transcoded by MPEG transcoders, separately.

Table 4 gives the bit rates in Mbits/s. for the eight programs for both independent and joint coding. Note that these bit rates were averaged over a certain number of frames per program. Hence, they were slight higher than the actual rates as the last GOP may contain less number of P/B pictures that usually use less bits than I picture. From Table 4, it should be seen that by independent coding, all the eight programs were, more or less, coded at the same rate of 3 Mbits/s while the bit rates by joint coding could be very different, depending upon the corresponding program complexity. For example, sequence Mobile used more 4 time as many bits as sequence News.

Table 5 shows the average PSNR (dB) for the eight programs for both independent and joint coding. The PSNR for more complex sequences are increased significantly, but at the expense of the PSNR of less complex sequences. However, it should also be realized that the quality loss for the well-coded sequences with very high PSNR might not be as visible as the quality gain for those complex sequences with low PSNR. The quality variation among the eight programs by using joint coding is seen to be much smaller than that by independent coding.

Fig. 6 shows the PSNR with respect to frame number for independent coding, which demonstrates a big difference in quality (up to 17 dB) among the eight programs. For some individual programs, the PSNR may vary considerably from frame to frame (e.g. football1503.snr). On the other hand, joint coding significantly narrows down the difference in quality among the program, as shown in Fig. 7. In fact, the PSNRs for the eight programs are maintained within a range of less than 9 dB, as compared to a quality difference up to 17 dB for independent coding (Fig. 6). Furthermore, it should also be noticed that the PSNR for each individual program by joint coding tends to be more stabilized than by independent coding.

4. CONCLUSIONS

The paper presented a stat mux/remux system with encoders and transcoders. The system is able to handle both pre-compressed and uncompressed video signals. Specifically, the uncompressed video signals are encoded by encoders while the pre-compressed video bit streams are transcoded by transcoders. The rate control engine manages the bit allocation among the input programs. The computer simulation results demonstrated that the stat mux/remux system indeed achieved a relatively uniform quality among programs as well as within a program.

REFERENCE CITED

1. ISO/IEC (MPEG-2), "Generic coding of moving pictures and associated audio", March, 1994.

2. M. Perkins and D. Arnstein, "Statistical multiplying of multiple MPEG-2 video programs in a single channel", SMPTE Journal, Sept. 1995.

3. E. Nakasu, K. Aoki, R. Yajima, Y. Kanatsugu and K. Kubota, "A statistical analysis of MPEG-2 picture quality for television broadcasting", SMPT Journal, Nov. 1996.

Table 1. PSNR in dB

MPEG/Transcoder	Market	Trapeze
Direct MPEG at 3 Mbits/s	34.85	35.42
15 → 3 Mbits with ME	34.43	35.16
15 → 3 Mbits/s without ME	34.37	34.91
6 → 3 Mbits/s with ME	33.71	34.59
6 → 3 Mbits/s without ME	33.72	34.56

Table 2. Processing Engine

	News	Market	Flower	Football	Mobile	Movie	Tennis	Trapeze
Process	MPEG	Trans.	MPEG	Trans.	MPEG	Trans.	Trans.	Trans.

Table 3. GOP Structures

	News	Market	Flower	Football	Mobile	Movie	Tennis	Trapeze
N	15	15	12	15	12	15	15	15
M	3	3	3	3	3	3	3	3

Table 4. Bit Rate in Mbits/s

	News	Market	Flower	Football	Mobile	Movie	Tennis	Trapeze
Joint	1.31	1.70	3.63	3.95	5.65	1.78	3.84	2.40
inde.	3.03	3.04	3.05	3.02	3.04	3.02	3.01	3.01
Δ	-1.72	-1.34	0.58	0.93	2.61	-1.24	0.83	-0.61

Table 5. PSNR in dB

	News	Market	Flower	Football	Mobile	Movie	Tennis	Trapeze
Joint	37.81	32.81	29.22	33.85	27.87	36.55	30.98	33.93
Inde.	41.06	35.94	28.35	31.63	25.16	38.44	30.23	35.15
Δ%	-7.91	-8.70	3.07	7.02	10.77	-4.92	2.48	-3.47

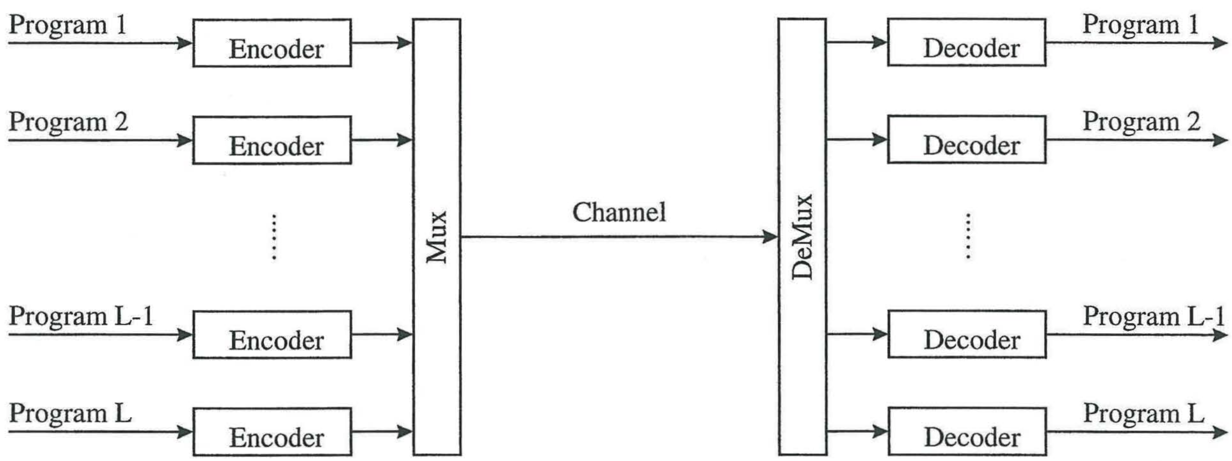


Figure 1. Multi-program transmission: Several programs are coded, multiplexed and transmitted over a single channel.

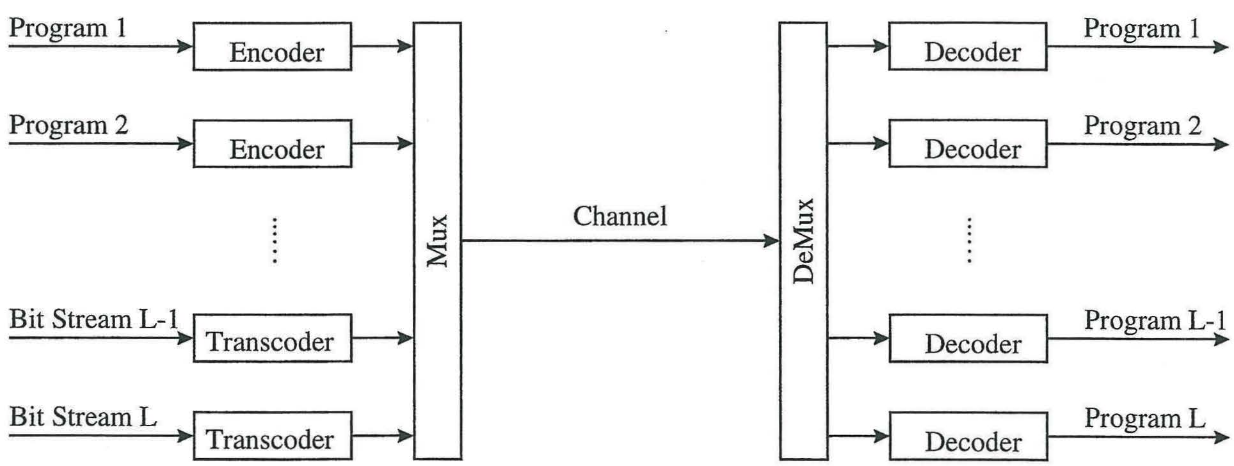


Figure 2. Multi-program transmission: the inputs can be either compressed or uncompressed video programs.

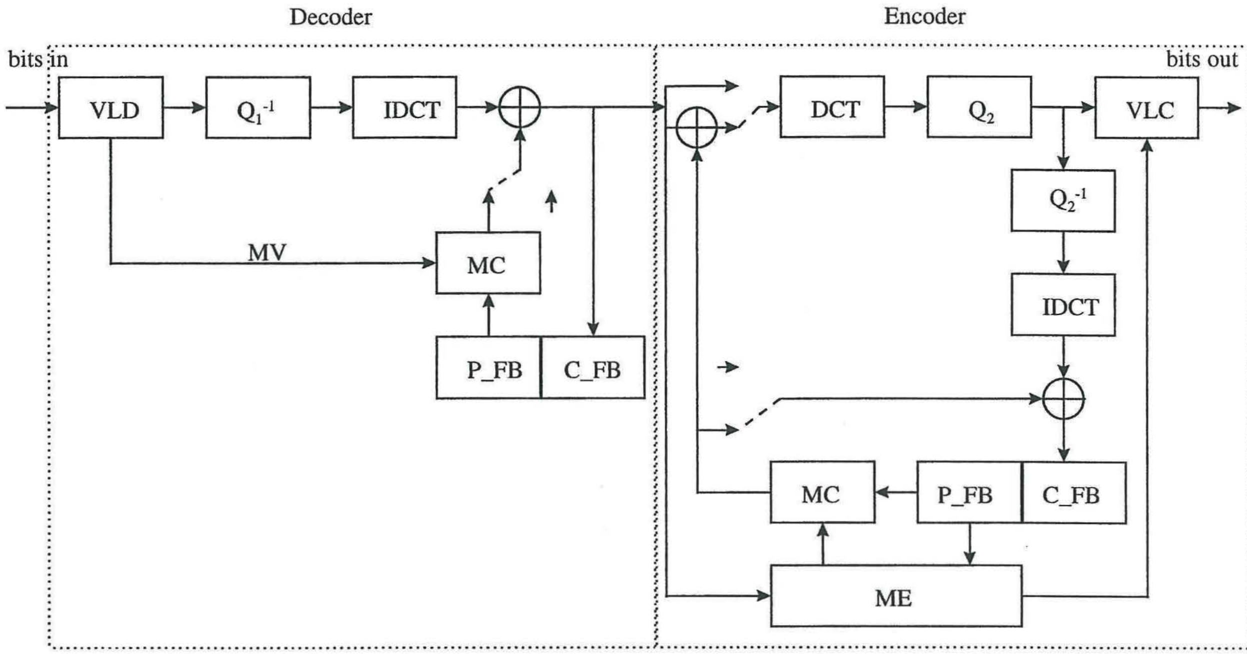


Figure 3. Cascaded MPEG decoder/encoder.

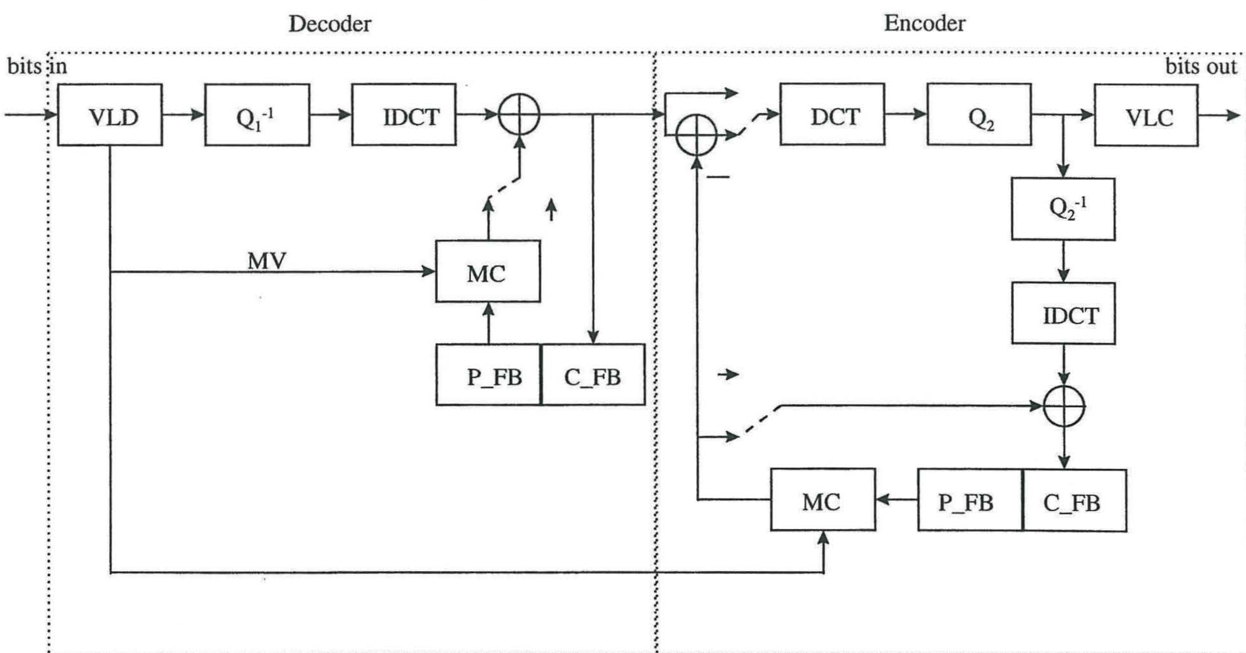


Figure 4. Cascaded MPEG decoder/encoder without ME.

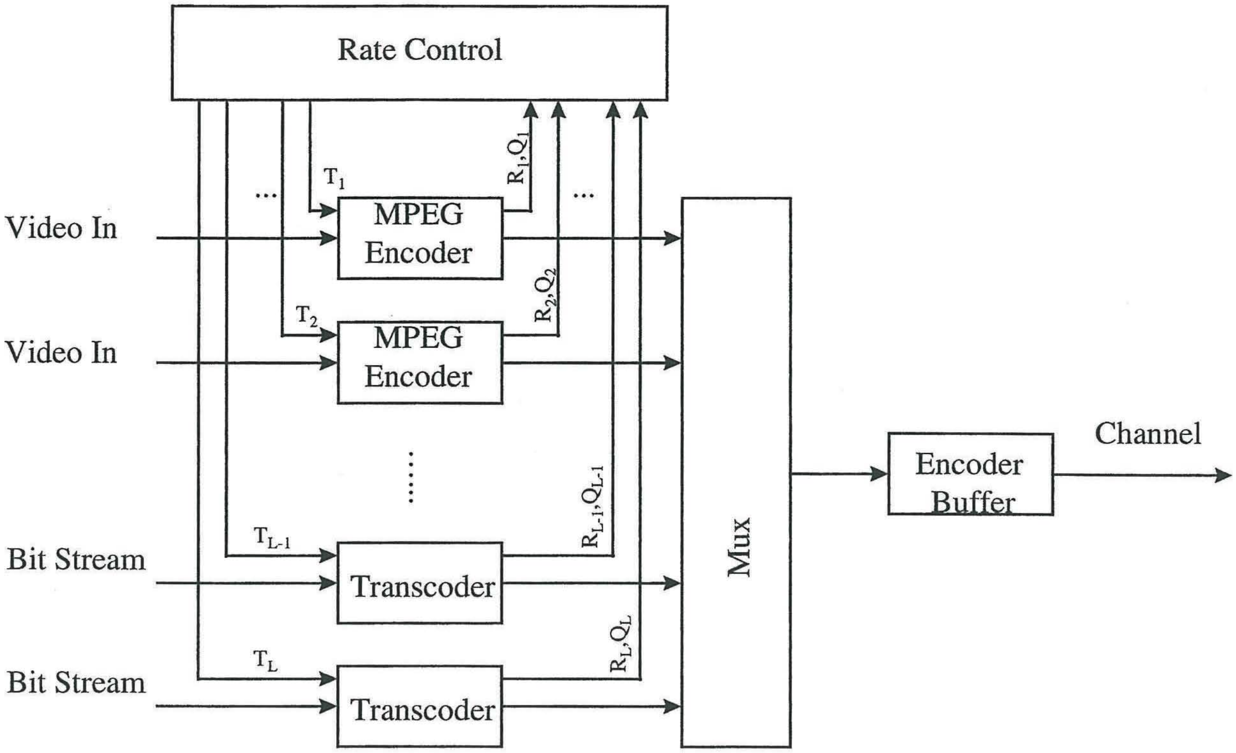


Figure 5. Block diagram of a stat mux/remux system for both pre-compressed and uncompressed video signals.

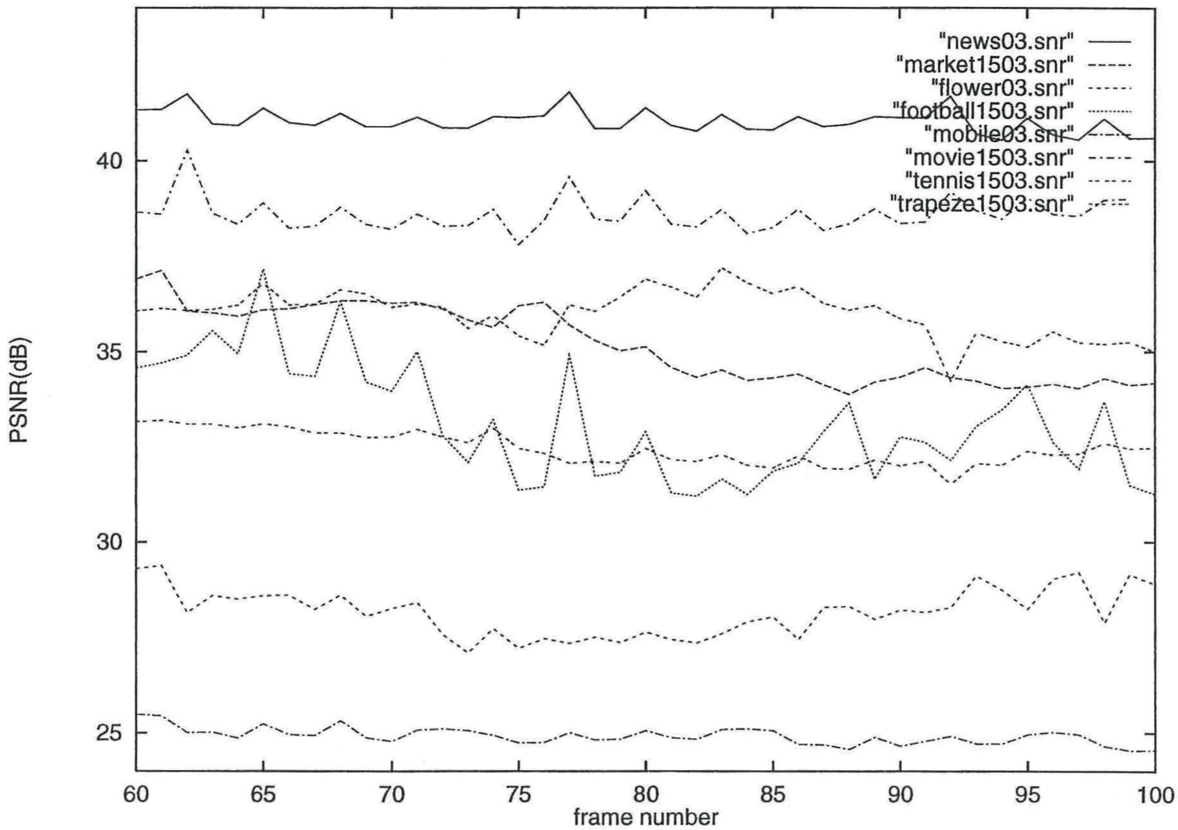


Fig. 6. The PSNR by independent encoding/transcoding for the eight test sequences where "...03.snr" means direct MPEG at 3 Mbits/s and "...1503.snr" means transcoding from 15 Mbits/s to 3 Mbits/s.

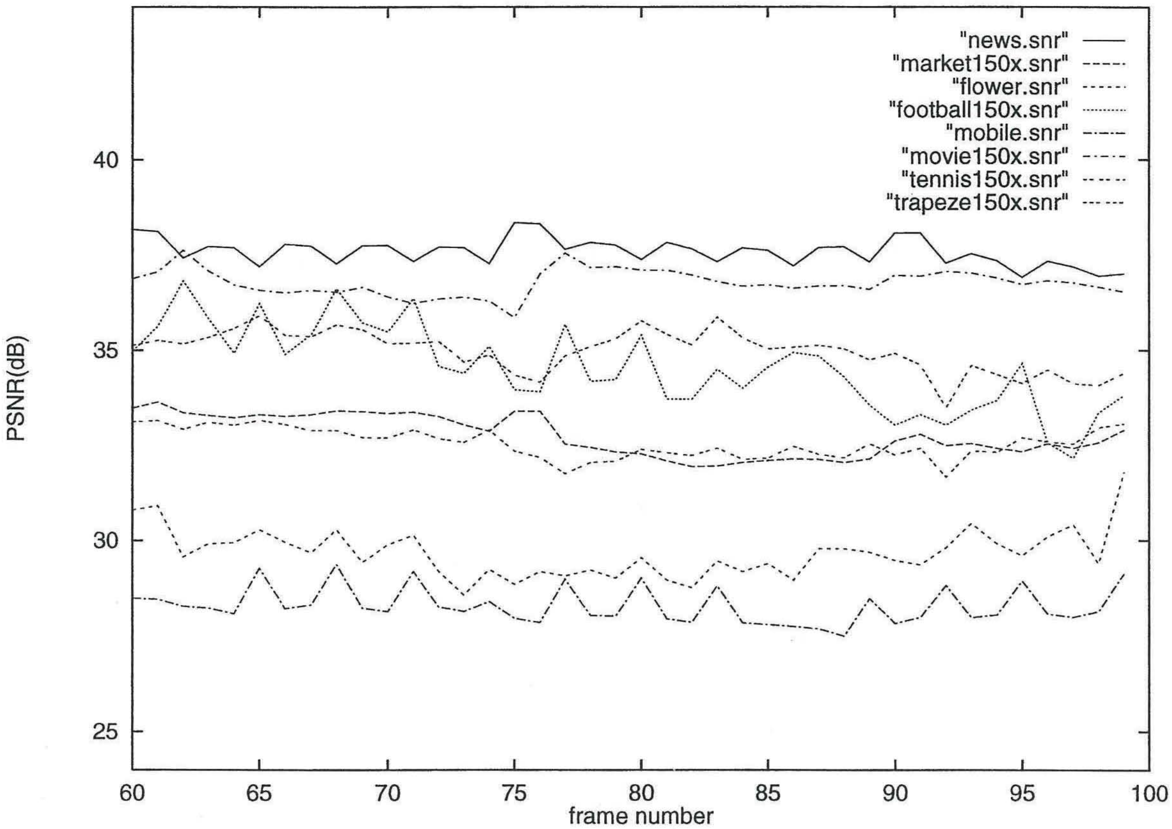


Figure 7. The PSNR by stat mux/remux for the eight test sequences where "... .snr" means uncompressed video signals and "...150x.snr" means pre-compressed video bit streams at 15 Mbits/s.

THE EXPOSED HEADEND: ENABLING A STRATEGICALLY DEFENSIBLE POSTURE FOR BROADBAND TELECOMMUNICATIONS SERVICES

□
Terry L. Wright
Chief Technology Officer
Convergence.com Corporation
3950 Johns Creek Court, Suite 300
Suwanee, Georgia 30024-1265
Ph: 770-416-9993 Fax: 770-416-9994
Web: http://www.convergence.com
Email: tlwright@convergence.com

Abstract

This paper discusses critical service monitoring/management challenges several MSOs are already facing, and many others are certain to encounter within the next several years. If and/or when the FCC imposes service-transport unbundling regulations on the cable industry, these challenges will immediately become industry-wide. Even absent unbundling regulations, as cable operators increase advanced telecommunications service offerings, they will eventually need to overcome these issues.

The basic services monitoring and management challenges are rooted in the shared nature of broadband delivery networks, and the lack of thorough operational information integration among back-office graphical information (GIS) and billing systems. Shared network oriented issues are further aggravated by the finite nature of bandwidth capacity allocated for advanced broadband service delivery.

In a broadband digital services environment where services are concurrently delivered, especially when external service providers are involved, asset control and service integrity (i.e., stability and consistency) complications are introduced. These challenges impose the need for an integrated services monitoring/management system (ISMS). Such a system must, among other critical functions, enable operators to maintain control over their broadband infrastructure in order to ensure the integrity of concurrently delivered services.

If the notion of an ISMS as a basic requirement does not make you uneasy, then it is likely you do not fully grasp the problems an ISMS must solve. As this paper title suggests, implementing an ISMS involves a virtual exposure of key components of the shared broadband service delivery infrastructure, as well as a dynamic proxy representation of those components. In addition, an ISMS must have dynamic access to (often-disparate) back office system databases in order to ensure service reliability and consistency. This paper explores the details of the integrated services management problem. Ideally, readers will gain sufficient insight and understanding as to why an ISMS may need to be the cornerstone of any future advanced broadband services capability.

BACKGROUND

Broadband Internet and other advanced telecommunications services are gaining momentum, and appear to be viable contenders in the race to shape the future of cyberspace. This translates directly into increased revenue per subscriber as these services are developed and deployed. (Through creative service packaging options, a portion of this increased revenue could even come from new entertainment subscribers.) Data/Internet service subscribers are increasingly demonstrating a **preference for performance** with the growing penetration of broadband data services. In fact, in most markets offering advanced broadband data services, attrition rates are approximately the same as or below the rates attributed to death and/or subscriber relocation.

It is especially interesting to note that many broadband-based data service subscribers are getting online **for the first time** (approximately 20% or so[1]), versus outgrowing a dialup connection. This speaks loudly to the convenience as well as the performance of high-speed data via broadband cable. This is the good news.

As is usually the case, success in new and leading edge businesses generally comes with its own set of unique challenges. When those businesses are as complex and dynamic as provisioning broadband Internet and all the other rapidly emerging advanced telecommunications services riding the Internet wave, some of these challenges can be substantial.

Understanding the Challenge

Subscribers appear willing to pay a premium for broadband performance only if other service attributes are at least at parity with competing services in the areas of service reliability and consistency, and customer technical support.

Effective customer support can be achieved through a knowledgeable internal and/or outsourced technical helpdesk capability. In a shared broadband network environment, however, service reliability and consistency present considerable challenges.

Service *reliability* requires broadband operators to attend to the signal quality and stability of the underlying broadband delivery infrastructure in a *proactive* manner. This means an ISMS must enable operators to anticipate and automatically respond, wherever possible, to degradations in the infrastructure that can lead to service failures. (Ideally, service-threatening conditions are remedied **before** failures occur.) However, until service-disabling infrastructure anomalies can be recognized by its correlation engines, and responded to by its automated scripts, an ISMS must enable an operator to quickly identify, isolate, and resolve, service failures.

Problem anticipation and automated response requires an integrated correlation engine that links problem scenarios with many related dynamic performance variables and conditions. Before operators can achieve an anticipatory service management posture, they need system capabilities that enable correlation functions to be developed and refined (as new problems arise) for incorporation into an ISMS. In order to do this operators need an accurate service monitoring/management capability that is granular enough to capture correlation characteristics, and responsive enough to enable timely service restoration prior to the availability of automated system response capabilities.

Service *consistency* requires operators to consider and dynamically manage the effects of concurrent service delivery on their shared infrastructures.

An Integrated Services Management System capability appears to have been largely overlooked in all the excitement surrounding broadband service performance. Absence of this capability could easily become the cable industry's Achilles heel if not addressed in a hurry.

Let us begin our understanding of what an ISMS must do by exploring the details of the problems it must solve. This is probably best done by examining the primary attributes of broadband-based advanced telecommunications services technically, as well as from the typical user's perspective.

DETAILED PROBLEM ATTRIBUTES

The current and future problems an ISMS must solve manifest in the intra-service and inter-service domains, and in the related area of integrated back-office databases (ISMS solution implementation).

Intra-service issues are introduced by the standardized capabilities of DOCSIS 1.0-compliant cable modems. To a certain extent, the introduction of DOCSIS 1.1-compliant cable modems sometime down the road will make the service consistency perception problem somewhat more manageable, in exchange for increased complexity.

Inter-service domain issues are primarily associated with the notion of external entities delivering services on cable networks. However, the eventual introduction of OpenCable-compliant set tops, incorporating any version of DOCSIS-compliant cable modem functionality, could further amplify the both the intra- and inter-service domain issues. Longer term, the arrival of DOCSIS 1.2-compliant cable modems and/or set tops could simply aggravate the problems even more.

The fundamental issues associated with an ISMS solution implementation are its need to access back-office systems' information collateral necessary to perform its functions, and the integrity of this information. This primarily billing and mapping information is essential to service outages notification, identification, isolation, and restoration capabilities an ISMS must provide. Often, the back-office systems that generate this information are not well integrated.

Reliability

In order for subscribers to consider broadband services as reliable, the availability of those services must at parity with competitive offerings. Service availability requires rapid outage awareness, isolation, and resolution. It is therefore imperative that an ISMS have access to operationally accurate information in order to facilitate quick awareness, isolation and resolution of service outages.

In most existing scenarios, however, the reliance of an ISMS on any one back-office database is problematic. (This problem might be all too familiar to many operators.) This is largely because many of these databases, although containing similar information, have evolved in support of different functions (e.g., billing and mapping), and usually under the control of different departments. In addition, a by-product of many cable system mergers and acquisitions has been the introduction of different back-office systems. In the interest of continued operational continuity during the aftermath of these mergers and acquisitions, legacy back-office systems are often left in place.

A significant aspect of building an ISMS solution will therefore be the creation of an accurate, dynamically sustained and normalized abstraction of pertinent back-office system information.

The bottom line here defines requirement one for an ISMS:

- ***Create and use a dynamically normalized abstraction of back-office systems databases containing subscriber and service-delivery infrastructure information considered critical to the rapid isolation and resolution of service-threatening conditions.***

This ISMS requirement also enables continued use of legacy systems, minimizing operational stress (i.e., staff re-training) and related disruption.

Intra-Service Issues

The primary attribute of a broadband service delivery environment is the shared nature of the transport infrastructure. Even a single broadband data service shares the transport network with entertainment programming. Unless dedicated channels are allocated for each distinct data service, multiple data services will share the same spectrum. (Recent studies[2] suggest that 18 MHz of return spectrum will be required to support all the commercial services cable operators are contemplating.)

In a generic Internet access services scenario, multiple subscribers contend for bandwidth within the channel pair(s) allocated to this service class. The access methodology inherent in cable modem solutions is designed to effectively manage the contention for this bandwidth among subscribers. However, during peak traffic periods, even this access methodology can be overwhelmed by demand, resulting in a degradation of service performance *consistency*. In today's competitive telecommunications climate, variations in service consistency are potential problems. (Subscriber perception is reality.) The service consistency issue will eventually apply to all broadband services including Internet access, IP telephony, interactive television, video-on-demand, video-conferencing, interactive games, and so on.

With traditional telephony-based Internet services (i.e., a dedicated circuit for each user), subscribers enjoy a level of service consistency at least to the ISP point of presence (POP). This means that any content (e.g., web pages, newsgroups, files) consumed from servers resident in the POP, assuming the server platform is sufficiently performance-scaled to meet anticipated concurrent load, will be delivered in a fairly consistent manner. (Of course, if the content being consumed is beyond the ISP POP, and especially if it must transit one or more Network Access Points (NAPs) in route, performance inconsistencies are introduced.) However, in apples-to-apples comparisons to circuit-based competitive offerings, a disparity in service performance consistency represents an exploitable weakness. Inconsistent service performance is the first cousin of perceived service reliability. (If the time it takes for your lights to come on varied noticeably by time of day (i.e., subscriber load), how reliable would you perceive your utility service to be?)

Of course the answer to performance inconsistency in a shared network environment is to increase the available bandwidth proportional to the offered load. This can be accomplished via node splitting or simply by devoting more spectrum to data services. Both of these solutions have an associated cost in equipment or loss of revenue previously attributable to re-allocated capacity, or both. (And unless you are using wave division multiplexing (WDM or DWDM) or other fiber-sharing approach, node splitting will consume lots of fiber. Perhaps someone will invent a smart fiber node, with built-in, appropriately scaled CMTS, access router, and neighborhood-sized server, that is capable of taking advantage of fiber-sharing technologies.)

The bottom line here defines requirement two for an ISMS:

- ***Monitor perceived data services (minimizing bandwidth consumption in doing so) well enough to anticipate necessary bandwidth increases in order prevent performance inconsistencies.***

Ensuring the ability to effectively deliver *different* qualities (i.e., classes) of services, among other things, is the essence of DOCSIS 1.1. This emerging revision to the DOCSIS standard should go a long way towards managing perceived service consistency in a broadband environment. However, when all is said and done, the consistency problem will likely only be compounded. In fact, with DOCSIS 1.1-defined bandwidth guarantees for different service classes, service inconsistencies will likely surface sooner as bandwidth is being consumed by the service architecture...whether or not a user is present in the equation. (A QOS guarantee is just another way to spell "circuit", even though it may be a virtual one.) Guaranteed throughputs and latencies associated with various QOS-based services will further stress available bandwidth.

The addition of QOS capabilities defines requirement three for an ISMS:

- ***Monitor perceived data service consistency across and within multiple service classes (again, minimizing bandwidth consumption in doing so) well enough to anticipate bandwidth increases necessary to prevent performance inconsistencies.***

The service consistency issue may seem to be a minor problem to many cable operators. Unfortunately, service consistency *is* a very real concern. The online community is a vocal one. If you believe subscriber comments like "cable modem service is great...when it works" will help facilitate your penetration goals, then you can probably ignore most of the above. I suspect that most operators would find such comments quite troubling, especially after investing all that capital just to enter the hot arena of intensifying telecommunications services competition.

The complexities of maintaining service consistency are amplified in the inter-services domain. Let us move on to explore the interesting implications associated with the inter-services domain of broadband networking.

The Inter-Services Domain

A few large cable operators have already encountered the challenge of delivering advanced telecommunications services from external entities. If and when unbundling regulations (i.e., separation of service from broadband transport infrastructure) are imposed on cable operators, then the entire industry will very likely come face to face with this issue in short order. Probably the easiest way to make the fundamental points here is to speculate on the nature of an unbundled cable industry environment. Whether operators proactively allow external service providers access to their networks, or the FCC imposes access, the problem is essentially the same, as is the solution. The hypothetical post-unbundling scenario presented below should emphasize the critical nature of the inter-services problem domain.

It's the day after the FCC imposes unbundling on the cable industry. Your phone rings persistently for most of the day. You hear from the likes of AOL, Mindspring, Earthlink, Prodigy, and several local garage-shop ISPs. (You may even hear from players like Lycos, MSN, and eXcite.) They all want to know when they can arrange access for access to your broadband in order to reach subscribers over your network. All but the local ISPs probably have exponentially larger budgets to spend on data services (including advertising, applications, and so on) than you do. Even if the FCC was generous to cable with any access pricing guidelines, your worst nightmare is about to begin.

Not only do they want access to your network, all these providers want access to the various gadgets that *control* how data services are delivered on your network! And if you deny them this access, they will all cry foul to the regulators claiming you are denying them the kind of access they need to deliver competitive services. (I told you it was a nightmare.) Precedence for this already exists...the competitive local exchange carriers (CLECs) and long lines players are already using this line against incumbent local exchange carriers (ILECs). Your new job: "Broadband Infrastructure Referee". So how do you cope?

ISMS requirement number four:

- ***Expose manageable service delivery components of your broadband infrastructure to external service providers, through an operator-controlled proxy function that relies on the database defined in requirement number one.***

(This is much easier to say than to actually do, as we will see in the next section.)

The exposed infrastructure will likely include cable modems (CMs), CM termination systems (CMTS), cable set top devices, network interface units (NIUs), bandwidth, data switches, routers, DNS servers, and so on. In essence, all shared manageable components involved in service delivery will need to exposed.

In case you are wondering what rationale these invaders will use to argue their cause, a few immediately come to mind. The simplest argument would be that they need access to the QOS controls of your cable modem solution. They need this access (say, tomorrow at 2:00 PM) in order to facilitate a value-enhanced commercial video conference they've sold to a few of their corporate customers. Equally plausible would be that they need to fool around with your router configuration (yikes!) in order to facilitate some multicasting monstrosity in which they've recently invested the equivalent of your entire current year data services budget. Of course, there will also be all those content-specific players that need to reserve streaming audio and video bandwidth so they can broadcast a live concert of some unknown rock band to all the kids. Speaking of kids...those interactive games can sure chew up the spectrum can't they? Since we are just speculating, you can not leave out the interactive TV stuff the networks are planning. Depending on how attractive those FCC access-pricing rules are, you may still need to be in there yourself with your own suite of cyber thrills. Enough said?

While it is true it may be some time before the imposition of such a scenarios on cable, it could take considerable time to develop and perfect an ISMS with an acceptable "referee whistle" capability. Before we get into the primary ISMS implementation challenge, let's see if what kind of impacts OpenCable and DOCSIS 1.2 devices might have on any of this.

OpenCable-compliant set tops are supposed to include an imbedded DOCSIS cable modem function. It seems reasonable to assume that the primary drivers behind this feature are to leverage existing standards-based technologies, as well as enable in-channel compatibility between DOCSIS cable modems and OpenCable set tops. It seems equally reasonable to assume that OpenCable set tops will deliver applications focused on the interactive television as the viewing medium or service portal. Since no one company seems to have the where-with-all to do everything (not even Microsoft), it seems logical to expect that there will be different service providers attempting to leverage these different portals. And since the portals (DOCSIS CMs and OpenCable STBs) are supposed to have in-channel compatibility, it's almost assured they will be sharing the same data channels in some systems. This model sounds much the same as the external service provider scenario in an unbundled world, with the added consideration of introducing new interactivity dynamics.

DOCSIS 1.2-compliant devices are expected to employ a physical layer technique known as synchronous code division multiple access (S-CDMA), or something similar. This spread-spectrum technique is designed to improve portal device tolerance to a variety of broadband network anomalies (e.g., ingress, microburst noise, poor C/N ratios, and so on) while maximizing the number of concurrent data streams possible. A primary expectation of this future revision to DOCSIS is that it will enable a larger percentage of existing cable plants to carry data services. The expected physical encoding technique will enable greater flexibility in trading off spectral efficiency for increased throughput. DOCSIS 1.2-compliant products will not ease requirements for an ISMS, and in fact, may only increase the number of concurrent data channels requiring management.

GENERAL IMPLEMENTATION CONSIDERATIONS

It is probably useful to approach implementation of an ISMS by considering it within the context of a Network Operations Center (NOC).

NOC vs ISMS

A NOC is for the traditional monitoring, troubleshooting, and managing of network infrastructure, including key functional transport components residing on the network (e.g., routers, switches, hubs, and so on). A network operations center typically contains a suite of technology, both hardware and software, designed to assist network engineers in maintaining a reliable networking environment. A NOC generally contains systems that monitor various components of a network, enable manipulation of those components, sense general service-threatening conditions, and help isolate (and possibly resolve) problems that are discovered. NOCs are necessary for a variety of reasons. The complexities of the underlying technologies, and the number of companies and offerings (e.g., circuits and devices) involved in maintaining a sophisticated network, are good examples.

An ISMS, on the other hand, focuses on ensuring the reliable delivery of multiple services in a shared network environment like broadband cable. It is focused on those contentious aspects of shared-infrastructure service delivery, and individual service consistency, that a traditional NOC does not address. An ISMS should not be viewed as a NOC replacement, but rather as an extension to a NOC.

There will obviously be the temptation to share systems infrastructure between a NOC and an ISMS. This option must be thoroughly investigated, as an ISMS will exhibit much greater probing granularity and frequency than a traditional NOC.

Consumption of revenue-generating service delivery bandwidth should naturally be minimized in any ISMS implementation. This will require a balanced use of out-of-band and in-band capacity to feed correlation functions, probe for performance-related statistics and conditions, and/or to impose configuration changes or other controls necessary to maintain service integrity (reliability and consistency).

An ISMS should be considered as early as possible in the planning of broadband data services deployments. This is because the capabilities of your ISMS will likely govern the number and nature of service policies than can be supported.

SUMMARY

Service management systems for advanced telecommunications services represent capabilities for which the cable/broadband industry has not traditionally had a significant need. In the existing industry environment, most cable operators can still get by without such a capability due to focus on a single data service offering. However, cable/broadband operators allowing multiple concurrent services to be delivered over their networks, especially if from external service entities, probably wish they already had an ISMS.

In a shared network environment where finite capacity is allocated to the delivery of advanced digital broadband services, the potential exists for service consistency to be at the mercy of the collective subscriber community's online behavior. When folks get online, what they do once online, whose content do they consume and where do they consume it from...all these online behavior attributes will affect the consistency of services in a shared network environment.

While it is true that the Internet is a shared network (it's actually well over a million interconnected networks), Internet access is only one of many advanced digital broadband services the industry is contemplating, and in some cases, actively deploying.

As the cable/broadband industry evolves, especially in the digital/data services arena, new service management requirements will accompany service introductions. Sooner or later, broadband operators will need the ability to expose (and manage the exposure of) key service-delivery components to the outside world. When that time comes, they will need an ISMS in order to ensure the basic functionality, quality, integrity, and performance consistency of those services. A good ISMS appears to be an essential enabler of a strategically defensible posture for advanced broadband service providers.

---

[1] Based on feedback from four well-known broadband data service providers.

[2] Courtesy: Digital Furnace Corporation

The Impact of Customer Care and Billing on Broadband Data Services

Craig Hanson, Director, Advanced Technologies
Ragan Wilkinson, Champion of Customer Care, Advanced Technologies
Lanse M. Leach, Ph.D., Senior Vice President, Chief Technology Officer
CableData, Inc.

Abstract

This paper explores the role of Customer Management Systems in the deployment of broadband data services. It highlights emerging requirements and design options for system integration, content, pricing, and customer care. It is intended for content and network providers as well as vendors involved in broadband rollouts.

INTRODUCTION

The trials are over and the rush is on! You have your market strategy, you've chosen your technology, you've secured content. Now, you are faced with the difficult problem of growing and keeping market share in the competitive broadband market. The implications are frightening. Will growth affect the quality of service? Will it lead to increased costs? Will growth related issues slow down your expansion? Can additional services be sold to recover acquisition costs? Can the quality of service be maintained with the inevitable increase in complexity?

To answer these questions, you must consider many factors including the need to manage organizations, establishing processes, and training people. But, should you be concerned with the capabilities of your Customer Management System[1] (CMS)? Does customer care and billing affect your organization's ability to deploy broadband services? Consider the following two scenarios:

An engineer in a communications firm is writing a paper for an industry conference. Knowing that it will be difficult to write the paper while at work, the engineer decides to do it from home. The paper requires access to multimedia data stored on systems at work, so the engineer decides to switch from his current ISP to a broadband service. Knowing that the cable modems are being sold at a local computer store, he travels to the store to purchase a cable modem. At the store, he is greeted by a salesperson who knows all about the service and offers to sign him up and check out his modem prior to sending him home. While taking the order, the salesperson recognizes that the engineer is a premium cable subscriber and can therefore be offered the service at a discount. The cable modem checks out fine and the salesperson is able to log on to the engineer's account using the selected modem before leaving the store. After the engineer gets home, he logs on to the service with his new cable modem. He finds that he already has Email from his cable provider thanking him for signing up and informing him that he's been given a free PPV movie that he may order through their Web site.

A young broadband subscriber has a paper due for school. She attempts to get on the World Wide Web to research wild life in Alaska, but it's not working. She asks her father for assistance, but he can't make it work either. He fiddles with his PC, first checking network settings, then checking to see if all the connections are secure. He can't figure it out either, so he calls the customer support number. After waiting on hold, he gets a Customer Service Representative (CSR) who attempts to trouble-shoot the problem. After another twenty minutes, the CSR is stumped, so the customer is transferred to another number to schedule a service appointment. The next day a cable technician comes to the house and verifies that the cable connection is fine, but still they can not link up to the Web. Another appointment is scheduled for three days later. That night, the father discovers through his own trouble-shooting efforts that the Ethernet cable connecting the cable modem to the PC is defective, so he replaces it and everything works fine. He cancels the second appointment, but on his next bill he is charged for the

[1] The term Customer Management System (CMS) refers to the enterprise software that is responsible for product sales, customer care, billing and commerce, and more.

unnecessary trouble call. A month later he calls to remove service.

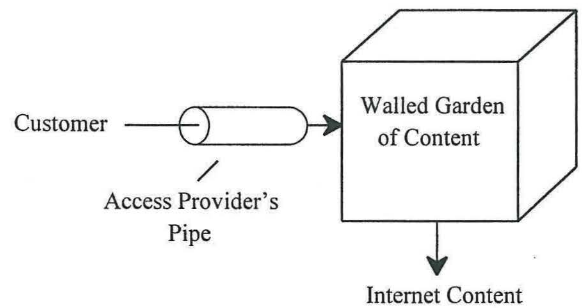
This paper details CMS requirements for providers that want to create a flexible product offering, provide flexible price and discount options, offer compelling content, integrate back office systems, and provide advanced customer care. Additionally, the paper explores the design options that are available for deploying broadband data services.

CONTENT

In the traditional video market, cable TV providers maintain a distributor relationship for the content itself. As these same providers move into the interactive broadband arena, they must think carefully if they are to maintain this relationship. By virtue of the technology itself, the provider in an interactive broadband market is not well positioned to maintain the distributor relationship. The simple act of deploying a greater than 1Mbps connection to a residence now enables the customers to seek and receive content from anywhere. Opening up the broadband network to the millions of Web sites has an enormous up-side: There is instant content that customers want and you can give it to them faster than the competition. There is also a down-side: This situation guarantees that all content flowing on the broadband network is not from the provider. Broadband providers are faced with a choice. Do they want to continue the distributor relationship for the broadband content?

The saying, "Content is king" would appear to be stronger than ever in the interactive broadband market. Never before have the broadband content owners had a wide-open distribution pipe like they get in a narrowband fashion with today's Internet. As the broadband industry blooms, so will broadband content providers.

One content option that seems probable is that the access provider will choose to create a "Walled Garden of Content" similar to Internet providers such as AOL. Then the provider will incent the customer to use its content instead of the competition's. The provider would create customer value through ease-of-use, completeness, discounting, or single-bill strategies.

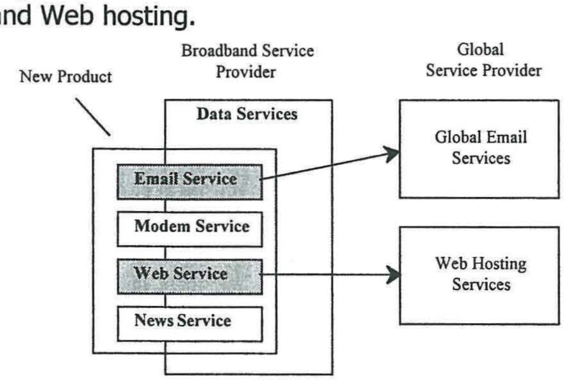


If a provider is marketing both the broadband access and the content, the provider can influence the customer to use its content through price discounts and consolidated billing. This highlights a requirement that the CMS be aware of both access and content pricing. This also highlights the requirement for discounting across products. This is one area where a strong, open CMS can be used as a competitive advantage in the broadband market.

FLEXIBLE PRODUCT OFFERINGS

There are four prevalent models that are used by broadband service providers to enter the market. These include outsourcing to a national service provider (i.e., @Home and Road Runner), partnering for content, partnering with an ISP, and rolling out one's own service. With each of the models, a business must be able to create or acquire marketable electronic services, market the service, sell the service, activate the service, get paid for the service, and provide on-going care for the customers with the service.

Electronic data services currently being deployed are widely varying from localized data content to services such as Email and Web hosting. Regardless of the business model, a provider needs the ability to rapidly deploy new services. In addition, a need exists to efficiently implement value-added services such as Email and Web hosting.



Global Services

The term "global services" is used to describe data services that providers use to add value. These services are often provided for free or heavily discounted to attract and acquire new customers. They differ in that the services require very large economies of scale which service providers can not afford to deploy on their own. For this reason, global services will typically be deployed by a national or global service provider.

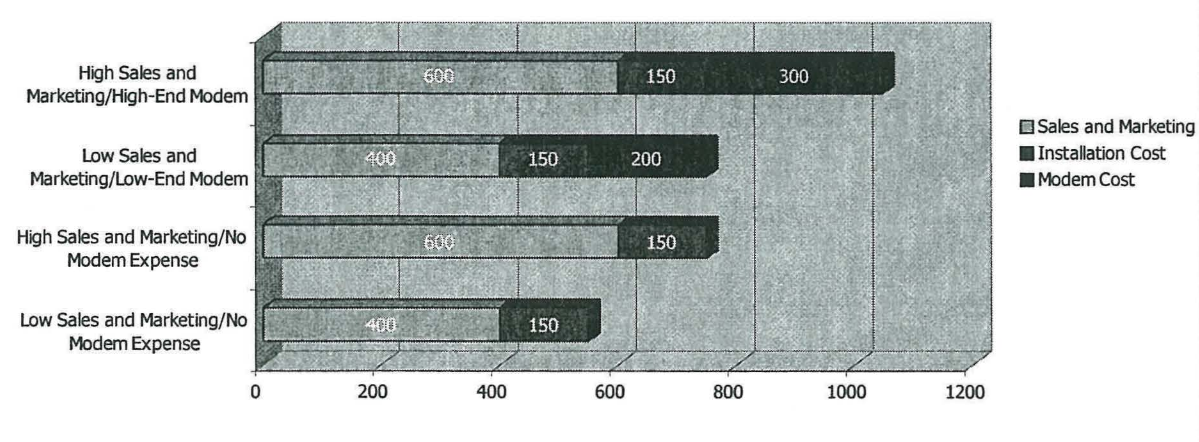
If a service provider's CMS is not aware of the global service as well the global service provider, it will affect the provider in two ways. First, provisioning and billing for the global service will be a two-step or manual process. This will create an unusually high cost-of-acquisition. Next, the process will increase the complexity of resolving customer care calls for both billing queries or trouble tickets. Again this increases operating costs. A provider who can integrate the systems will lower operating costs and increase customer satisfaction.

New Services

New services are services that are not globally offered or are conceptualized and created by the service provider. Often these services have a local flavor or are community oriented. However, the services can be of any nature. For such services, the provider is responsible for content creation, site hosting, and on-going maintenance. The requirement for connecting the CMS to these services is the same as for global services. It is important to note, however, that time-to-market requirements for new services often compromise the integration of all necessary systems. If this could be the case, the provider may decide to choose an off-the-shelf commercial infrastructure product such as Microsoft Commercial Internet Server (MCIS) for deploying the services. These commerical products offer a rapid service deployment model and create a single point of integration between vendors.

FLEXIBLE PRICE OFFERINGS

ISP customers have grown accustomed to the single price service packages. Unlimited access for a flat rate is common with local ISPs and national ISP providers such as AOL. Because broadband infrastructure is expensive, the "one price fits all" model is not cost effective for broadband providers. Service providers must decide which of the startup costs they will cover (PC installation cost, cable modem cost) while determining the level that is required in the areas they service. The Yankee Group estimates the cost to acquire a broadband subscriber ranges between $550 and $1050 depending on if the modem is leased or purchased by the subscriber [Yankee, 98], see figure 1.



Source: The Yankee Group

Figure 1. Cost per Subscriber Acquisition (in Dollars)

A reasonable subscription cost of $50 per month indicates that a provider would need between 10 and 20 months to recover the startup cost. A pricing model based upon the value that customers derive makes the most business sense in the broadband domain, where service is tailored to each customer's needs.

There are several prevalent pricing models that are required of a CMS. These include:

- Service tiers
- Threshold pricing
- Usage based pricing

Within each price model, the CMS must be able to bundle, discount and rebate Internet services in order to provide a competitively priced product.

Service Tiers

A fundamental capability that a CMS must provide is the ability to package and tier services with each package containing one or more broadband services. For example, a Family Internet package may contain 5 Email addresses, 10 Megabytes of Web-oriented disk space, support for two computers, and a cable modem with a transmit rate limited to 128Kbps. Groups of packages, called tiers, can then be created and organized by the CMS so that service tiers are easily identifiable and available when an Internet order is taken.

Service tiers are not unique to data services. Packaging of cable TV service is the primary pricing model used for video services (see figure 2). What makes Internet service packaging a challenge for a CMS is not the packaging, but rather the required provisioning associated with services contained in the package. In the data service domain, the CMS may be required to communicate service to many systems. Using the Family Internet package example, the CMS would be required to provision five Email addresses on a mail server, allocate 10MB of disk space on a Web server, and provision a CMTS[2] or cable modem creating one logical package spanning several systems responsible for delivering individual services. Furthermore, the representation of a data service tends to be more complex than the representation of a video service. Name value pairs, service strings, and DOCSIS MIB variable settings are examples of varying representations of service on different systems. Prior to sending provisioning commands to a CMTS, cable modem or Internet server, the CMS is required to translate the representation of service according to the expectations of the targeted system. For example, the CMS may represent the Family Internet package with more than one billing code. During provisioning, each code must be translated to codes specific to the delivery system. The translated codes must then be routed to the appropriate delivery system.

Since tiers have a one-to-many relationship with services, mapping services to the appropriate provisioning system and then transmitting the provisioning commands are capabilities required of a CMS. These capabilities require complex provisioning software for broadband services, so extending a CMS's existing set-top provisioning software is usually not an option.

By way of contrast, consider the current video services model. The difference between packages simply equates to authorization code settings that are transmitted to an addressable controller. Pick HBO, set a bit in a code. To add a sports option, set ten bits. At the end of the task, the CMS is required to send the authorization code settings to the addressable controller which understands the channel mapping.

Threshold Pricing

Not all changes to Internet service affect the subscription price. The CMS must determine whether a billing threshold has been exceeded before altering the service codes attached to the account. Again using the Family Internet package example, the services included in the package (five Email addresses and cable modem support for two computers) may be changed directly by the customer from a Web site. After the customer adds an additional Email address for a spouse or child, a change of service notification is sent to the CMS. The CMS must bump the subscriber's Email service count, change the subscription price if (and only if) the service threshold has been exceeded, and record information associated with the change.

[2] CMTS is an industry acronym that refers to a Cable Modem Termination system.

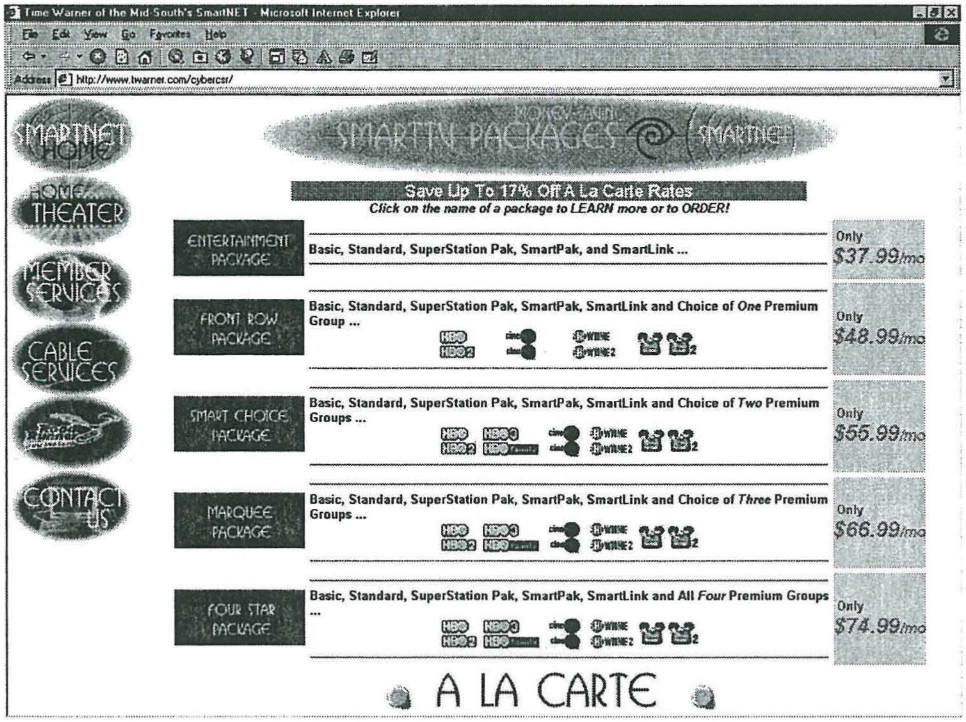


Figure 2. Time Warner's Mid-South Division package of video services as sold over the Internet.

The threshold-based change of service applies to many types of broadband services. Current examples include: disk space quotas, the number of PCs connected to a cable modem, the number of protocols required by the customer (TCP/IP and/or IPX), the number of Email accounts, the number of news groups subscriptions, the number of chat group subscriptions, etc. Since Internet service types and rules vary from provider to provider, a flexible CMS will abstract the service rules, so an operator can configure them.

Figure 3 identifies a software abstraction of an Internet service. The abstraction contains both the data attributes and the threshold change of service algorithm. The change of service function would be provided as an open programmatic interface that could be executed from an external system.

Usage Based Pricing

Three months after the successful introduction of a high speed data offering, a mid-sized cable operator begins receiving calls from customers who are irate because they can not connect to the Internet using their telephony return cable modems. The engineering department determines that the 7-to-1 ratio of subscribers to modem ports is insufficient to meet current demand. To correct the problem, management decides to purchase additional terminal servers to maintain a 4-to-1 subscriber to port ratio, and switch from a flat rate to a usage-based pricing structure that would encourage subscribers to disconnect whenever they are not using the service. After the first month of the new pricing structure, the MIS department is required to manually adjust 10% of the HSDS accounts because the CMS was unable to calculate charges based on usage. The MIS department has to add staff to support this labor intensive, error prone process until rating software could be added to the CMS.

Billing customers based on consumption is nothing new for CMS's that support telephony services, but this capability presents a challenge for CMS's that traditionally support video services and are now being extended to accommodate broadband data services. Usage-based billing takes on two forms: rated and un-rated. Rated transactions involve the consumption of a service such as playing a game or sending an electronic greeting card.

```
                         InternetService Abstraction
Attributes
    ServiceIdentifier       // Email Accounts, Chat Groups, PCs, etc
    Description             // Displayable Description
    CurrentUsage            // Counter
    Threshold               // Change service is threshold is exceeded
    PricePerUnit            // Price for each item in excess of the threshold
    BillingCode             // Billing code to use for change

ChangeOfService(   in AccountNumber, in ServiceIdentifier, in Action,
                   out PrevPrice, out NewPrice)
{
    Find Subscriber's Account
    Find InternetService using AccountNumber and ServiceIdentifier
    PrevPrice = Account.SubscriptionPrice

    Case Action = Add
        Increment InternetService.CurrentUsage by 1
        if (InternetService.CurrentUsage > InternetService.Threshold)
        {
            Add InternetService.BillingCode to Subscriber's Account
            NewPrice = Account.SubscriptionPrice + InternetService.PricePerUnit
        }
        else
            NewPrice = PrevPrice

    Case Action = Remove
        if ((InternetService.CurrentUsage > InternetService.Threshold) AND
            (InternetService.CurrentUsage-1 == InternetService.Threshold))
        {
            Remove InternetService.BillingCode from Subscriber's Account
            NewPrice = Account.SubscriptionPrice - InternetService.PricePerUnit
        }
        else
            NewPrice = PrevPrice

        if (InternetService.CurrentUsage > 0)
            Decrement InternetService.CurrentUsage by 1
}
```

Figure 3. Software Abstraction of an Internet Service

Rated transactions are the simpler of the two for a CMS to handle because the price is already determined. Un-rated transactions involve the collection, aggregation and rating of usage data such as time, packets, or bandwidth. For unrated transactions, the CMS must determine the charge based on the aggregated consumption data. In both cases the CMS must be able to automatically adjust the subscriber's account based on the event charge.

INTEGRATION

If there is one CMS attribute that will significantly improve operating efficiencies and customer service in a broadband rollout, it has to be the ability to integrate with other systems. The systems that are involved in the broadband data services deployment come from a variety of vendors. It is the degree to which these systems are able to act as one system that determines the complexity of the deployment.

To maintain the level of service that customers demand the system must have:

- Automated flow through provisioning associated with order processing
- Automated billing associated with usage data and billable events
- Integrated customer care and trouble shooting tools used to isolate and correct network and PC problems

Integration Requirements

To create a seamless system integration must be accomplished between the CMS and the following types of systems:

1. Order taking systems
    - Internet self provisioning
    - Retail
2. Usage information systems
    - Mediation
    - Content
3. Service delivery systems
    - Cable Modem Termination System
    - Internet server for local services
    - Global service providers

It is important to note that there are often many different vendors for these systems. This increases the complexity of integration.

Order Taking Systems

The ability to add or modify service is also a fundamental capability provided by CMS's for video services. In the video services model, a subscriber typically calls a customer support number and receives assistance from a customer service professional to make desired changes. A small percentage of changes in service may be executed by technology savvy customers using an automated telephone system or a Web site. For broadband changes of service, the frequency changes originating from a phone call is low. Internet customers are accustomed to E-commerce applications and on-line ordering. Whether it's trading stocks on-line or altering their friends and family lists, Internet customers prefer the ability to effect changes without requiring human intervention.

For data services, an order can originate from many different sources. The primary sources that have been identified include retail outlet, provider Internet site, partner Internet site and the traditional customer service representative. To support external change requests, a CMS must provide open programmatic interfaces that allow an external system to:

- Authenticate itself to the CMS
- Authenticate the subscriber initiating the change
- Validate the change request
- Provide pricing information associated with the change
- Verify the account status of the subscriber
- Initiate the change and verify change completion

Usage Information Systems

This section details the requirement for flexible usage based pricing. The implication of this business requirement highlights the need for integrating usage information systems with the CMS. Usage information systems take the form of mediation systems that collect and manage network usage, as well as content systems.

Network usage systems typically are specific to the network elements. The hardware providers provide usage information such as MB used or connect time to the CMS. It is the responsibility of the CMS to determine if the usage effects the state of the account. The CMS could take a variety of actions including altering pricing, discounting or even limiting service.

Content systems provide content usage information to the CMS. Usage statistics such as the number of times used or the duration of usage can be passed to the CMS. The CMS can use this information to determine effects on the account such as discounting or even limiting service. Since there are a variety of content systems, the interface needs to be very flexible.

Since a CMS is typically not involved in the data collection and aggregation functions, CMS integration to a mediation system(s) is a critical issue. At the center of any system integration effort are the following issues:

- What data elements are exchanged?
- What message format will be used?
- What transmission protocol will be used to transport the message?

To answer these questions, we propose a message format (see Table I) that could be used to integrate a CMS with many different types of mediation systems.

Such billable event messages would contain an ASCII string of <name>=<value> pair information. Each name/value pair is separated by a vertical bar ('|') delimiter to accommodate variable data lengths. The message itself is enclosed in a pair of pipes. The tokens in the transactions (names, and values) are case insensitive and order independent to provide

Table I. Billable Event Message Elements

Data Element	Description	Data Type	Required/Optional
AccountNumber, SerialNumber, MacAddress	An account or equipment identifier that can be used to relate the billable event to a subscriber. At least one of three elements must be provided.	Character	R
EventDateTime	Time of event occurrence	Date/Time	R
Provider	Supplier of the service	Character	R
EventType	User defined event classification	Integer	R
Description	Description of the service	Character	O
Rated	Is the event already rated?	Boolean	R
Price	Required if rated=Y	Currency	
UnitsOfMeasure	Required if rated=N Usage units (e.g., seconds, bytes, etc)	Character	
Count	Required if rated=N Number of units consumed	Long	
PricePerUnit	Unit price	Long	O
Username	Login of the subscriber using the service	Character	O

flexibility. A data packet format of this type would provide the flexibility needed for system interoperability, and could be transmitted using a variety of industry standards protocols such as TCP/IP, SNMP or RS232.

Given the above data elements, the following are example billable event messages:

|accountNumber=07807100501 | EventDateTime=April 10, 1999 12:01:05 | Provider=GameFactory | EventType=Game | Description=Internet NASCAR | Rated=Y | Price=4.99 |

|macAddress=080F11223344 | EventDateTime=04/15/1999 17:25:00 | Provider=EDHISP | EventType=ConnectTime | Description=Monthly Usage | Rated=N | UnitOfMeasure=Minutes | Count=1833 | Username=jdoe |

Service Delivery Systems

Service delivery systems have responsibility for turning on and off the services themselves. These take the form of modem termination systems, national service provider systems, partner systems, local service systems, and global service systems. The typical CMS transaction sends a command to start, suspend, or stop service from the CMS to a service delivery system. The service delivery system can act as an order taking system.

Design Alternatives

There are two options for achieving seamless provisioning and billing functionality: 1) Add provisioning capabilities to the billing system or 2) Programmatically integrate the billing and provisioning systems. Both options assume that the billing system is responsible for maintaining customer accounts and the services tied to the accounts

When provisioning software is added to the billing system, following capabilities must be provided by the billing system to be considered a viable business solution:

- The provisioning engine within the billing system must be able to interoperate with a variety of Internet servers and Cable Modem Termination Systems.
- The billing system must be able to accept usage data, adjust customer accounts accordingly and report on the adjustments.
- The billing system must provide an open interface that can be integrated with retail interface and self-provisioning applications.
- The billing system must be able to report on installs and changes in service that originate from the various interfaces.

- The billing system must be able to bundle, rebate, and discount Internet service with other services that are added to a customer's account.

When a billing system is integrated with a provisioning system, the following capabilities must be provided by both systems for this option to be considered viable:

- Both systems must support a standard interface (e.g., TCP/IP, CORBA, Java, etc) to exchange data.
- The provisioning system must provide usage data to the billing system.
- The billing system must provide account and service information to the provisioning system.
- The billing system must provide scheduling information to the provisioning system.
- Both systems must support synchronization capabilities to reconcile services billed versus services provisioned.
- The provisioning system must update the billing system when service is installed or changed from any of its available interfaces.

Option 2, integrating a provisioning system the billing system, initially proves to be the most flexible for rapid deployment of new services. Although this architecture is appealing, over time the systems will need to become tighter in their integration. Since the CMS holds customer's financial status as well as a record of current services, the provisioning system will need to query the CMS before provisioning a service. If this option is chosen, the CMS and the provisioning system should be tightly coupled before moving beyond a trial

CUSTOMER CARE

If price and service are the product attributes that attract customers, customer care is the attribute that will retain them [Whiteley, 91]. However, providing quality customer care for data service products presents a difficult challenge. Unlike traditional video service, customer care for data services requires support from authentication and routing systems, dynamic host control protocol (DHCP) servers, domain name service (DNS) servers, mail servers, Web servers, and more. Table II identifies the added complexities in this area by comparing modems and data services against traditional set-top boxes and video services.

Because of the complexities inherent in broadband data services (i.e., cable modems and the connections to a PC, PC configurations, broadband routers, network management tools, etc), product support becomes a difficult undertaking. A CMS can make the customer care task easier by providing tools that can:

- isolate data service problems
- scheduling the right person for the job
- provide the work order information in the appropriate media format

Fault Isolation and Identification

One way to provide quality customer care is to staff your call center with experienced network engineers and system administrators. The problem with this approach is finding available engineers with customer-oriented skills plus the million-dollar or so requisite increase in staffing budget! The realistic solution to the problem is to require the CMS to provide better broadband support tools capable of isolating problems. Today, the most popular tools used to identify broadband problems are network management tools (e.g., HP Overview, SNMPc, Sun Net Manager, etc). A network engineer can determine if the CMTS and cable modem in question are operational and if network packets are flowing correctly; however, engineers usually don't take customer calls nor do they have access to the CMS. And CSRs typically does not understand network protocols or network management tools.

One effective solution is to have the CMS provide an integrated trouble-shooting application that hides protocol complexities while allowing a CSR to perform network management tasks. For example, when a customer calls in with a problem, a CSR accesses the account and the CMS supplies information about the cable modem and computer to the trouble-shooting application. Using the modem's IP address, CMTS information, and the customer's account information, the trouble-shooting application can attempt to isolate the problem between the PC,

Table II. Data Services vs. Video Services

Data Services	Video Services
Cable modem vs. Set-top box	
Active two-way device	Passive device with limited upstream capability
Consumer determines modem make and model	Operator controls box make and model
Managed via standard network protocols	Managed via proprietary protocols through an addressable controller
Complicated installation	Simple installation
Installation and Account Management	
Service is provisioned on many systems (e.g., CMTS, cable modem, authentication server, DHCP server, mail server, etc)	Service is provisioned on the set-top box through an addressable controller
Many service installation options available (installed directly by the consumer, through a retail outlet, or with the assistance of a CSR)	Service is installed with the assistance of a CSR
Service is managed by the consumer via the Web	Service is managed with the assistance of a CSR.
Problem Isolation and Identification	
Service interruption can be attributed to many possible factors (e.g., problems with the customer's PC, the modem itself, a network outage, Internet servers, etc)	Service interruption can be attributed to a few factors
Requires many skills sets to support (e.g., Network Engineer, RF Engineer, PC administrators, Web masters, etc)	Requires one or two skill sets to support

cable modem or broadband network (see figure 4). Furthermore, the trouble-shooting application can display an interactive image of the customer's cable modem (extremely important in a DOCSIS environment) to allow the CSR to mimic LED behavior that the customer's modem is displaying in order gather additional information.

The goal of the trouble-shooting application is to bridge the gap that currently exists between customer care and engineering thus providing smooth turnovers to the appropriate tier 2 support organization(s).

Scheduling and Work Orders

Another requirement that broadband services place on the CMS is the ability to assign one or more technicians to an installation or trouble call job. Since data services requires both cable and computer skills, multiple technicians may have to be assigned and scheduled accordingly. There is nothing more disturbing for a customer than to have a technician with the wrong skill set show up for a scheduled appointment.

The multiple technician requirement also implies that multiple work orders may have to be generated by the CMS for a single job. Furthermore, today many operators outsource PC support, so work order information must be Emailed to the outsourced business.

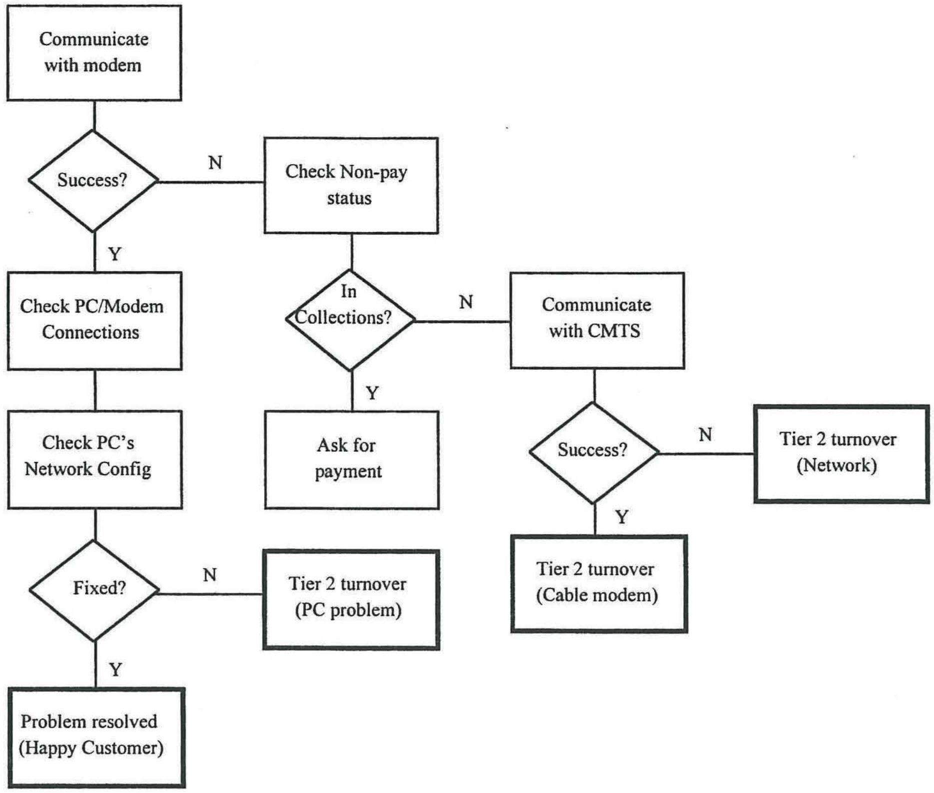


Figure 4. Problem isolation provided by broadband support tool

To adequately support work force management the SMS must:

- Support the assignment and scheduling of multiple technicians to a single job
- Generate multiple work orders that allow the operator to: 1) select the data contained on the report, 2) tailor the report format, and 3) select the destination of the work order (printer, fax, or Email).

SUMMARY

Customer care and billing have a clear impact on a provider's ability to succeed in the broadband marketplace. The Customer Management System used affects every aspect of a rollout from deploying content to provisioning services. A strong flexible CMS enables a provider to create a flexible product offering, provide flexible price and discount options, offer compelling content, integrate back office systems, and provide the advanced customer care that will attract and retain customers in the competitive broadband arena.

REFERENCES

[Whiteley, 91] R. Whiteley, *The Customer Driven Company,* Addison Wesley, 1991.

[Yankee, 98] The Yankee Group, *The Market and Marketing of Residential Broadband Internet Access,* Internet Market Strategies Vol. 4, No. 15—December 1998.

THE LAST FIVE METERS
Five Factors for Home Networking

Gary H. Arlen
Arlen Communications Inc.

Abstract

Home area networks (HANs) – formerly characterized as "home automation" – are finally becoming viable because of increasing functionality and usage of electronics and communications devices. Although households with multiple PCs have become a driving force for HANs (to share peripherals and/or Internet access), many other factors are affecting the development of home networks.

The last five (or maybe ten) meters in a house may become as big a challenge as the "last mile," which has impacted cable providers for decades. Five factors affect the deployment of HANs:

Interoperability
Interconnectivity
Interface
Service/Support
Price.

Cable operators will find them-selves on the front lines in dealing with these factors if/when consumers adopt HANs.

HOME NETWORKING ARRIVES

Establishing a Market

Home Area Networks are on their way to becoming a $1.4 billion industry by 2003.[1] After decades of dreams and promises, ranging from the CEBus (Consumer Electronics Bus) to various "smart appliance" plans, the market has taken on new momentum thanks to the growing number of homes with two or more personal computers. About 34% of the 50 million U.S. homes with PCs have two or more active systems, and the number of multi-PC households is growing faster than the overall PC penetration rate.[2]

At the same time, a proliferation of set-top boxes (cable, terrestrial Digital TV, satellite and other sources) plus home security and digital entertainment devices is fueling consumer interest in the category. Since cable TV is likely to deliver many of the services into the home, cable operators may find themselves in the center of a technical and marketing maelstrom.

Multiple Distribution Formats

The HAN market will be very fragmented, based on consumer appetite and vendor aggressiveness, as indicated in this forecast:

Market Share
Percentage of End-Use By Format

	'99	'00	'01	'02	'03
Phone	56%	70%	65%	60%	50%
Power	7%	6%	8%	8%	8%
RF	7%	10%	14%	22%	32%
Ethernet	18%	7%	5%	4%	3%
Other	12%	8%	8%	6%	6%

Notes: Phone = phoneline; Power = powerline
Other = alternative and emerging wired solutions
Source: Cahners In-Stat Group April 1999

Use of existing home phoneline wiring will become the primary distribution mechanism, but Radio Frequency delivery represents the fastest growing segment. Wireless services can handle transmission at speeds comparable to phoneline connections (1 Megabit per second and faster). However, wireless signals may face interference and congestion, especially in dense housing clusters (townhouses or high-rises).

Plans to boost the speed of power-line distribution (currently operating in the 350 kilobits per second range) may improve the opportunity for providers in that market sector. Moreover, powerline products are generally cheaper and more convenient (e.g. can be placed close to an electrical outlet) than phoneline sources. Wireless sources may prove to be the most attractive because of their "anywhere" capability.

Juggling Multiple Formats

This market fragmentation means cable operators will face growing confusion as consumers try to hook together products that receive cable input (video, datacasts, Internet and voice access).

REACHING THE LAST 5 METERS

Delivery to the last five meters in the home is likely to become as perplexing as outdoor plant services. In particular, cable operators face challenges on five fronts:

Interoperability

Cable systems must be able to function with whatever HAN the customer chooses to install. Industry standards are still being developed in this sector, but some signatories to home networking standards have already indicated they will launch independent ventures. The Shared Wireless Access Protocol (SWAP) for RF systems has just been adopted, although products complying with this standard will not be available until yearend. Cable operators will be challenged to work with a range of customer-selected HAN facilities.

Interconnectivity

Consumers may chose selectively to install products that are very specialized or unique (e.g. streaming media players). Again, cable providers may be asked to support HAN connectivity processes are still works in progress.

Interface

Methods to hook these devices into cable-delivered service must be created in ways that are understandable to end-users. Interface issues have barely been addressed at this time.

Service/Support

Cable operators must be prepared to handle a plethora of questions and problems that will arise as customers try to install, configure and operate HAN components – especially if cable services are funneled through the HAN. Inevitably, consumers will seek to blame someone or figure out the source of problems along the line (in-home or on-network) for installation, operation or functionality. This will make today's cable modem questions pale in comparison.

Price

The initial cost of HAN equipment (ranging from $50 to $300) is a new factor in consumers' growing electronics budgets. Some systems may require additional monthly usage fees. It is not clear how this pricing – or where it will be paid – may affect viewers' adoption of these services.

CONCLUSION

Connecting cable customers to a variety of HANs will pose significant challenges for cable operators. The early customers are likely to be high-end techno-buffs who put sizeable demands on the last few meters of the home communications process. Cable operators will have to work with a variety of systems and services to fulfill customers' expectations for these new technologies.

---

[1] Cahners In-Stat Group, April 1999

[2] International Data Corp. quoted in **Electronic Business** magazine, May 1999.

The New Frontier of Custom Digital Service

Patrick Harshman, PhD
Elisa Camahort
Steve Toteda
Harmonic Inc.

Abstract

The growing subscriber demand for advanced digital services, such as video, internet access and telephony, creates a lucrative opportunity for both cable operators and other service providers. To effectively compete against broadcasters, DTH satellite operators and telcos with xDSL, MSOs must solve the two major issues facing them during a digital service rollout.

The first is efficiently using available system bandwidth to maximize service delivery options. The second is creating customized analog and digital services that can be targeted for specific customer demographics. Achieving bandwidth efficiency and customized service delivery capability is desirable from an operational perspective, providing flexibility and scalability. Equally significant is the improved ability to deliver the quality and quantity of content and service options that customers are coming to expect.

New technologies that enable content customization, digital rate adaption and the opportunistic bit stuffing of data into video streams allow cable operators to address both the bandwidth efficiency and custom service delivery objectives, opening the door to expanded revenue opportunities. The new abilities enabled by these technologies provide cable operators with a differentiable advantage as the broadband service marketplace becomes increasingly competitive with the entry of non-traditional service providers.

CHALLENGES FACING CABLE OPERATORS

It used to be simple. MSOs provided cable television service and telcos offered data and telephone service. Though they had the same customers, MSOs and telcos weren't competing. Now, thanks to deregulation, new network operators and the explosion of internet services, MSOs are competing with telcos, broadcasters and satellite operators to offer subscribers advanced broadband services. Cable operators can offer services, such as telephony and internet access, that consumers traditionally received from other service providers. The challenge is that those other providers have an equal opportunity to move into the video arena, the traditional preserve of cable, and offer consumers the same bundle of services.

Cable operators have several advantages over these other providers. Cable has a significant market penetration edge over satellite and xDSL service, and the subscribership for both technologies has yet to meet early expectations. In addition, cable operators boast a high-speed advantage over traditional telco solutions when it comes to internet access. However, cable operators have several challenges to successfully address in order to maximize their market share and ensure long-term profitability.

Telcos have long offered a variety of custom services to their subscribers that can be bundled as the individual customer desires. This allows targeted marketing. By contrast, cable operators have established a tiering system, in which the customer has two or three levels of service to choose from, plus premium channels. Cable operators may be taking advantage of various available satellite feeds, but find their hands tied when it comes to customizing how they deliver that content. Consequently, there is little difference between the packages offered by various cable operators in different regions.

While operators could address this issue by decoding and re-encoding each feed, if they choose to avoid that expense, they are left with a satellite feed taking a large portion of their bandwidth. Add in local content and advanced services such as internet access or Video-on-Demand, and you quickly have more content than you have room for.

The second challenge for cable operators is significantly expanding their service offerings without a corresponding increase in allocated bandwidth. It is not the most viable option to increase the amount of content by reducing the quality of signal, because cable operators are now sharing a competitive arena with satellite and telco operators that have a reputation for reliability and quality far exceeding that of cable.

Cable operators must take further strides to ensure that they can deliver high-quality, customized service offerings reliably and consistently. System downtime, signal degradation or loss, or taking an inordinate amount of time to implement promised upgrades is not an option.

A major concern is how to provide quality service without exceeding budget plans. Another challenge for cable operators is to expand the volume and variety of service offerings economically, react to subscriber demands quickly and become telecommunications providers, not just cable TV providers, seamlessly.

DRAWBACKS OF CURRENT DIGITAL SERVICE DELIVERY MODELS

The current digital service delivery model is to access remote satellite feeds and transmit the content on to subscribers, however it has inherent drawbacks that prevent it from being the solution cable operators need in the new market environment.

There are currently two solutions for processing remote satellite feeds. The first is to simply transmit the feed as is. While this is an inexpensive solution and certainly provides a comprehensive programming line-up, it puts unwanted limitations on the operator. This unaltered satellite feed is now taking up bandwidth, but only a portion of the program offering may actually be of interest to large groups of subscribers.

The second, traditional solution is to locally encode all programming. In this model, digital satellite feeds are decoded then re-encoded. This model provides the flexibility to customize the feed, dropping individual programs at will. This solution requires an enormous outlay of resources, however, both in the acquisition of expensive encoding equipment, and in the space and powering requirements to operate that equipment. Each hub site now must house complex equipment that will require more maintenance by skilled workers. As a system grows, expenditures will grow at a rate that is not beneficial to the bottom line. The decode/re-encode solution also has a technological disadvantage, as each time the signal is manipulated there is a corresponding loss of signal quality.

The current solutions for handling remote satellite feeds require balancing flexibility versus expense and quality, a choice that operators are more reluctant to make when faced with increasing competition and subscriber expectations.

Another key consideration when trying to achieve the optimum balance of flexibility and efficiency, is that many satellite feeds arrive statistically multiplexed. When re-transmitting this feed as is, it is simply a matter of allocating the bandwidth for the feed as delivered. However, once grooming techniques are applied to the feed, a bandwidth allocation problem is, once again, encountered. Dropping programs from the feed reduces its required bandwidth allocation, and operators must still reserve enough bandwidth to allow for the maximum bit rate that could be needed.

A desire to use every Hz of available bandwidth makes it tempting to compress yet another video program into the multiplex to take advantage of the small amount of bandwidth reserve. Cable service is already vulnerable to charges of lower quality standards, compared to satellite or telco service, so achieving greater bandwidth efficiency at the expense of signal quality is to be avoided.

The current model does not effectively provide the cable operator with both flexibility and bandwidth efficiency, while maintaining quality and good economical value.

Solving these challenges with a patchwork of solutions is possible, but will inevitably cost more and create a more complex, and therefore unstable, architecture. Cable operators can achieve maximum flexibility and bandwidth efficiency with one economical, complete system solution.

Significant new technology developments can be incorporated into existing system architectures to achieve this ideal, balanced system solution. Content customization techniques, digital rate adaption and opportunistic bit stuffing provide a complete solution that allows flexible service customization and efficient bandwidth utilization as part of an integrated and reliable system solution.

ADVANTAGES OF THE NEW DIGITAL DELIVERY MODEL

While satellite sources alone may provide adequate television programming, to become a full-service provider, cable operators must begin to access content from a new range of networks. These include video from servers to provide Video-on-Demand and IP networks to provide IP data, interactive services and IP telephony. And there will still remain the requirement to integrate local content, whether analog or digital at its source.

The first goal is to increase service options, without correspondingly increasing the complexity of headend and hub designs and without investing in overlapping network architectures to handle the transport of the various services.

It is not enough to be able to groom the various transport streams. The ability to multiplex the resulting program streams together, regardless of source, into a single custom multiplex, allows an increase in variables from which to select, and therefore increases freedom when building a business. The ideal system can integrate these functionalities seamlessly.

Figure 1 illustrates the ability to incorporate input from a variety of content sources and create a seamless custom service package for subscribers.

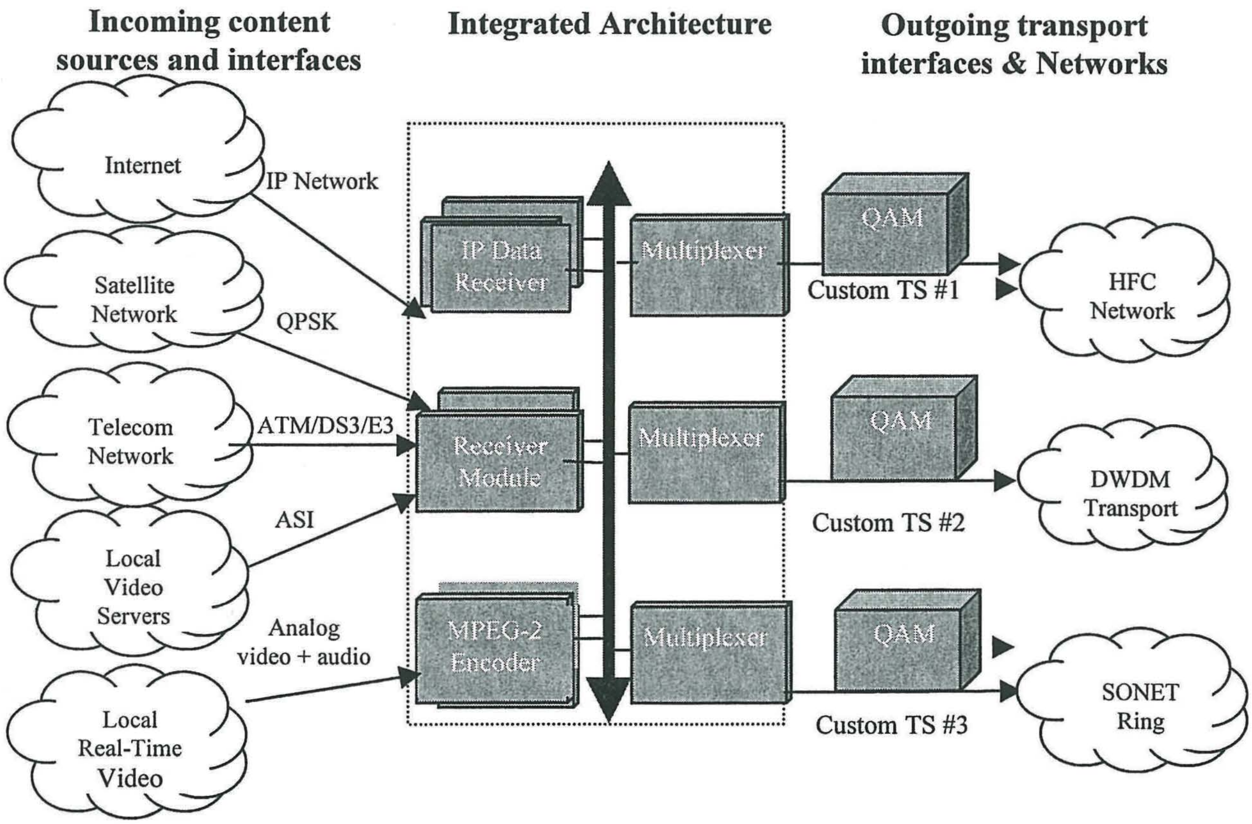


Figure 1: Network Transport Options

While Figure 1 illustrates a wide range of service delivery options available with a single system, Figure 2 illustrates the specific advanced digital services now available for delivery to subscribers with this architecture.

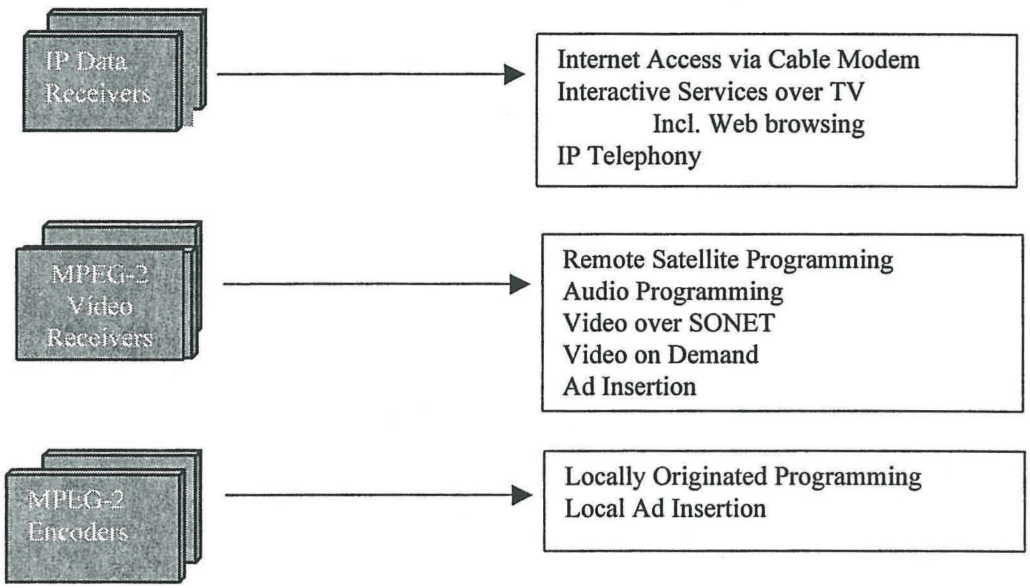


Figure 2: Custom Service Delivery Options

DIGITAL RATE ADAPTION

With all of these service options, the next goal is to create a custom multiplex, utilizing available bandwidth as efficiently as possible, without sacrificing quality.

Two new technologies facilitate this process. The first is digital rate adaption, which takes statistical multiplexing a vital step forward. The second is the opportunistic insertion of data into available bandwidth within video transport streams.

Unlike traditional statistical multiplexing, which is performed on constant bit rate transport streams, digital rate adaption is a new technology that allows the service provider to take a transport from multiple sources and do a final multiplex. The unique capability of this technology is that those transport streams can be either remote feeds or locally generated, and the original sources can be constant or variable bit rate. Cable operators can define bandwidth requirements and quality standards for their programming, and digital rate adaption enables a real-time constant bit-rate statistical multiplex to be created from these disparate sources.

The digital rate adapter analyzes the incoming bit streams, adapts the rates according to user-defined standards and creates a new statistically multiplexed transport stream. This flexibility also allows a higher bit rate per channel.

Figure 3 illustrates a solution for transmitting eight satellite programs and ten locally encoded channels, using traditional statistical multiplexing. This results in an average bit rate per channel of 2.7 Mbps.

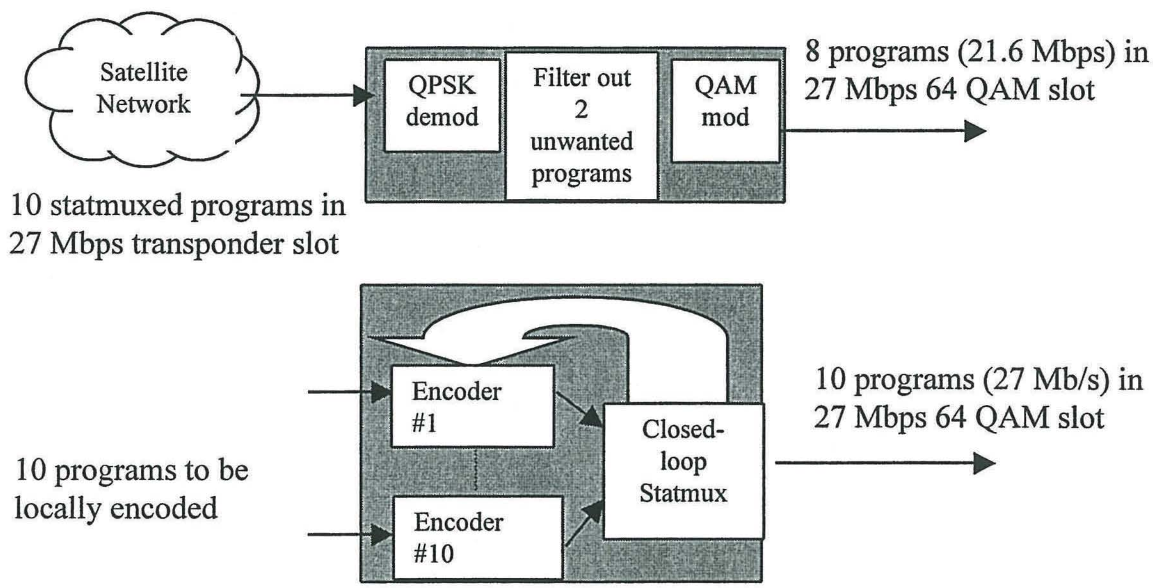


Figure 3: Traditional statistical multiplexing and program customization solution

Figure 4 illustrates how flexible digital rate adaption can provide a higher bit rate per channel and perform the same grooming function. This solution results in an average bit rate per channel of 3.0 Mbps.

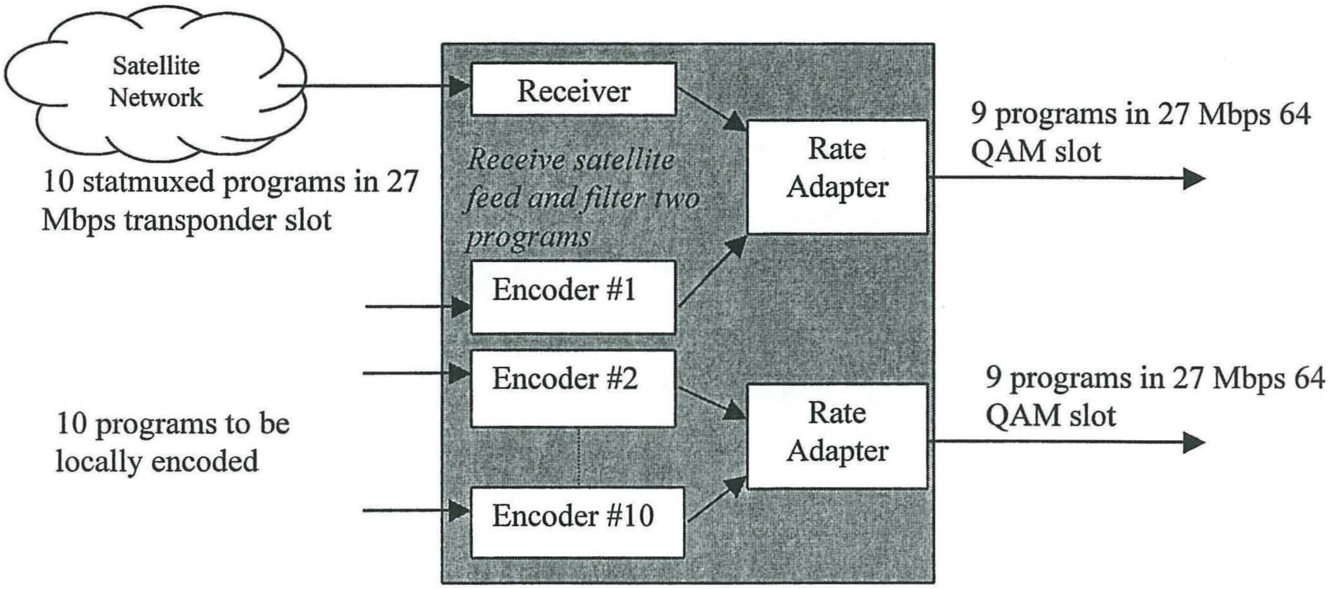


Figure 4: Digital Rate Adaption of both constant and variable bit rate transport streams

Figure 4 depicts the advantage of an integrated system approach to rate adaption and program customization. Being able to flexibly adapt constant and variable bit rate program streams in a one-step process provides, in this example, a 10% increase in program quality, while meeting the same bandwidth requirements.

Figure 5 depicts a typical application, with the cable operator receiving one HITS satellite feed that contains twelve statistically multiplexed programs, of which six will be kept. A Turner Broadcasting feed containing five constant bit rate programs is also being received, of which two CNN channels at 5 Mbps will be kept. Dropping nine channels between the two satellite feeds leaves eight channels to transmit and extra bandwidth.

Three programs are being locally encoded with resulting constant bit rate transport streams. Without adapting the rate of the satellite programs, the amount of bandwidth freed would be unpredictable, and therefore not suited to support these constant bit rate programs.

If the operator defines, however, the bandwidth of its QAM modulator, digital rate adaption will re-multiplex the local programs with the remaining satellite programming to match that defined bandwidth.

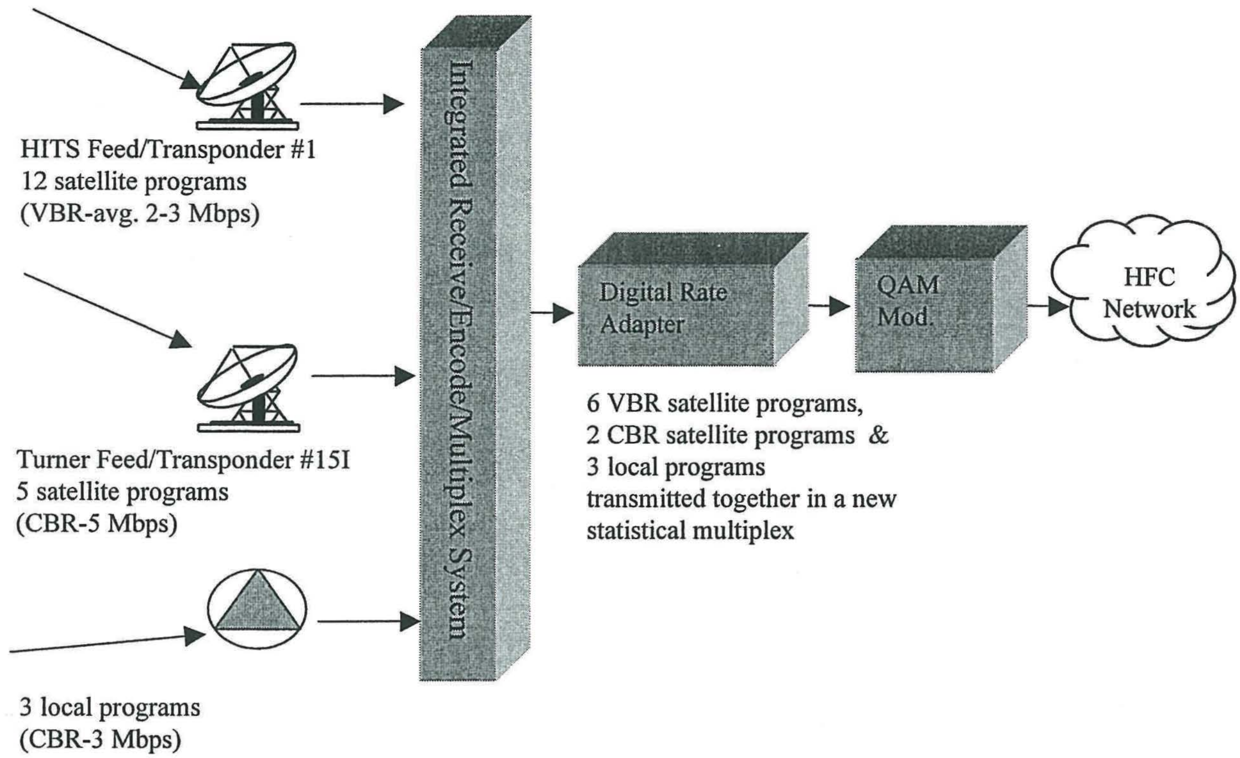


Figure 5: Typical Application

OPPORTUNISTIC BIT STUFFING

Even when optimizing bit rates via digital rate adaption, there is likely to remain a small amount of under-utilized bandwidth. This variable bandwidth reserve may, for example, vary between 1-3 Mbps at any given time.

Faced with this additional variable bandwidth, one option is to add another video program, but this will necessarily reduce the overall quality of every program.

A second option for using this extra bandwidth would be to use the 1-3 Mbps reserve to support IP-based interactive services. For example, the IP data traffic associated with an interactive service might require an average of 2 Mbps. Although the bandwidth reserve may sink below that mark, computers are accustomed to IP data arriving in bursts, not a steady flow, as is required for video.

Combining a variety of custom video and data services is like sending stones of all shapes and sizes down a single pipe. Even when fit together in the most efficient way possible, there will still be a little space left over. Opportunistic bit stuffing is like simultaneously shoveling sand down the pipe. It can fill in all of the little pockets of space available without slowing the delivery of the primary services.

The first step is to encapsulate IP packets into MPEG-compliant packets, assigning a Program ID, so that the data packets can be included in the Program Association Tables accompanying the MPEG-2 Transport Stream. These IP packets can now be inserted into the multiplex as bandwidth allows.

Every operator will want to utilize some combination of these technologies to optimize their own network. In an ideal, open system solution, one can scale in different services and efficiency-improving technology as your system expands.

In addition to optimizing bandwidth in a given headend's channel line-up, this new model can also have a significant impact on the bandwidth efficiency and economy of a larger cable network.

Consider a distributed HFC network with one primary headend and multiple secondary headends. The operator of such a network faces a cost versus quality dilemma. On one hand, the operator would like to centralize content origination, such as satellite reception, Video-on-Demand servers and ad servers, to the primary headend to minimize cost and maintenance requirements. On the other hand, the operator also would like the flexibility to customize the feeds as they are transmitted from each secondary location.

By utilizing rate adaption and bit stuffing technology, both goals are accomplished. For example, the optimum bit rate for a given video program in one secondary headend multiplex may be different from that in another secondary headend. Rather than placing two separate encoders or video servers in two different secondary locations, the video programming source can be located at the primary headend, and digital rate adaption and bit stuffing at the secondary headend location further customize the program.

Figure 6 illustrates a potential architecture, featuring this new model for digital delivery

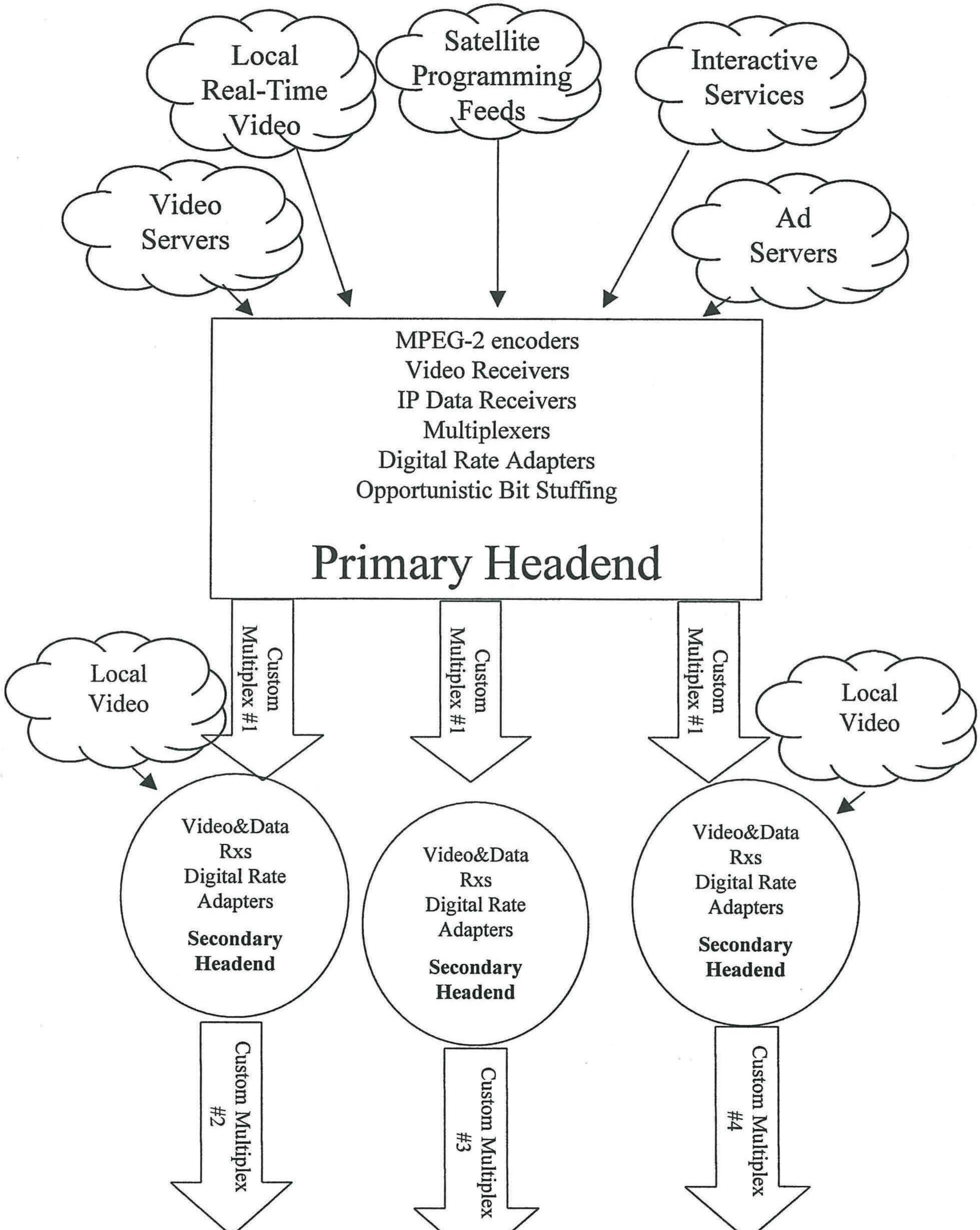


Figure 6: Sample Primary & Secondary Headend Requirements

CONCLUSION

Content customization, digital rate adaption and opportunistic bit stuffing provide the add/drop capability and program line-up grooming that allows targeted marketing to specific groups of subscribers, but without installing expensive decoding and re-encoding equipment at every hub, without a corresponding loss of signal quality and without leaving bandwidth underutilized.

This new model for digital service delivery has technological and business-driven benefits. It allows a new flexibility in choosing content sources and programming line-ups. The ability to combine the different content sources and perform bit rate analysis, remultiplexing and data insertion with a single integrated system ensures the cable operator the freedom to alter the make-up of custom services, as subscriber demands dictate.

Content customization, digital rate adaption and opportunistic bit stuffing allow cable operators to offer custom services, from their choice of content providers, while continuing to utilize their bandwidth as efficiently as possible. In today's increasingly competitive marketplace, taking advantage of these technologies will significantly improve cable operators' ability to respond quickly to market pressures and subscriber demands, taking full advantage of new revenue opportunities.

UTILIZING SRSC™ TECHNOLOGY IN THE TRANSMISSION OF MULTIPLE FORMATS IN THE CABLE INDUSTRY

Thomas W. Myers
SilkRoad, Inc.

ABSTRACT

The availability of interactive media via cable TV delivery systems is just beginning to grow. There are several limitations to this growth, but none more significant than the current limitations of signal transmission.

SilkRoad's SRSC™ technology will help facilitate the leap in interactive television technology by using a single-wavelength to transport multiple varied signal formats. This not only increases bandwidth, but it also expands accessibility and transmission distance.

This paper describes the basic science of this new technology and how it will lead to multiple-format delivery alternatives.

THE TECHNOLOGY

The Basis of SRSC

SilkRoad Refractive Synchronization Communication™ (SRSC) is an application of photonic physics based on the relativity theory of Einstein and the electromagnetic field equations of Maxwell.[1]

This technology allows "packets" of information (SONET, RF, IP, etc.) to be combined on a single wavelength on a single laser.[2] The signal can then be transmitted with all the advantages that a single wavelength offers:

- Simpler product design
- Fewer lasers and lenses[2]
- Lower lifecycle cost
- Fewer potential points of failure
- More amplification without interference
- More distance before signal regeneration

Despite the use of a single laser, SRSC technology is capable of providing more bandwidth than multiple-wavelength DWDM.

Foundations of SRSC Theory

SilkRoad uses the Mach-Zehnder modulator to achieve the optical tweezers effect, causing dielectric particles (photons) to spiral. The principle behind optical tweezers is that a single beam of tightly focused laser light creates an extremely high electric field gradient in the vicinity of its focus.

Under the influence of this gradient, dielectric particles are drawn into the laser beam. The dielectric particles experience a force directed toward the focus of the beam that resembles a whirlpool effect. They develop an *angular momentum* as they are drawn into the laser, resulting in a corkscrew-like movement along the laser transmission path.

Although the light within the laser has a common source and frequency, the corkscrew patterns may be tight or expanded for different photons based on the time dimensions in which they travel. These variations in the tightness of the corkscrew pattern are known as the photon's angular momentum. Different

[1] *This new technology has reformulated Maxwell's equations to solve for wave motion from one medium to another with a non-zero optical index (Maxwell assumes the optical index is zero). In this solution the real part of the optical index (n) and the imaginary part of the optical index (k) in all three orthogonal axes are positionally time dilated in relativistic time.*

[2] *SRSC uses a TEM_{00} laser, with no lenses in the transmission equipment.*

levels of angular momentum[3] are identified as *orders of the Laguerre* within the wavelength.[4]

Photons on different orders of the Laguerre may overtake one another, but collisions never occur because, having no mass, multiple photons can occupy the same space at the same time. Figure 1 shows photons traveling in different orders of the Laguerre.

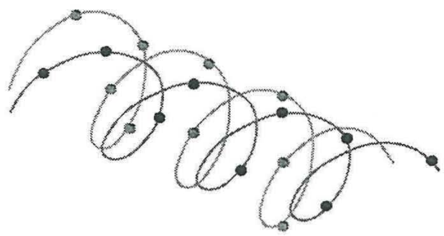


Figure 1 Laguerres in different orders

An orbiting pattern reduces the forward distance photons can travel during a given amount of time, so the degree of stretch in the corkscrew pattern (angular momentum) determines the relative speed at which the photon moves forward.

The highest level of angular momentum is the maximum stretch of the spiral. We refer to this fully stretched pattern as the highest order of the Laguerre. Photons traveling in the highest orders of the Laguerre will travel the greatest distances through fiber-optic cable. At lower levels, the spiral tightens, and the spanning distance declines.

Photons in different Laguerre orders can exist in the same place, but with different relative speeds and different orbits. Each order travels in its own relativistic time.

Since signals of the highest orders of the Laguerre achieve the greatest spanning distance in fiber-optic cable, SRSC uses them as carrier signals for data transmissions (Figure 2). At the receiving end, the Laguerre signals are converted to an electronic signal from which the individual transmissions can be isolated.

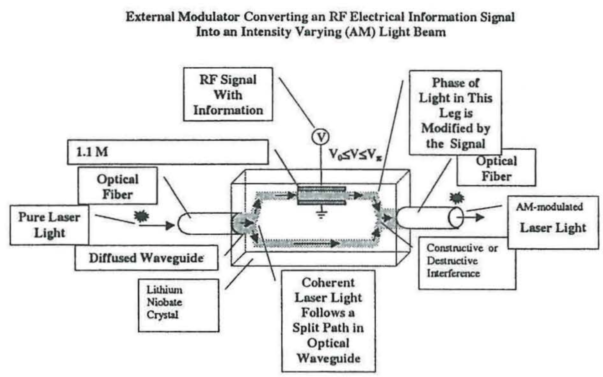


Figure 2 The SRSC Modulation Approach

While SilkRoad uses a broadband RF signal for subcarrier frequency modulation at the electronic level, our laser transceivers use a very narrow spectral bandwidth to transport multiple channels of data. To make an SRSC transmission bi-directional, we shift the ellipsometric phase that is characteristic of the different Laguerre orders.

Optical Refraction

Refraction is the basis of SilkRoad's *Optical Refractive Synchronization*. It is the change of energy, direction, or speed of a light beam that is propagating through a medium when it passes into a second medium.

Snell's Law mathematically represents the angle of both reflected and refracted light traveling from one uniform material through another. It predicts a continuous bending of the light beam and a subsequent change of speed of the light that is proportional to the

[3] In 1936, Beth first demonstrated the angular momentum that is imparted to an elliptically polarized beam of light. Beth's experiments demonstrated that each Laguerre order had increased angular momentum according to ± **L**• • , where **L** is the Laguerre order. Prior conventional wisdom had been that photons had an angular momentum of ±• • .

[4] The Laguerre Gaussian principle predicts that light photons in different Laguerre orders have different characteristics. Planck's theory suggested that a photon's energy is related to its frequency. Einstein used this theory to explain the photoelectric effect, where the speed of ejected electrons is related not to the intensity of light, but to its frequency. Einstein also discovered that light in a vacuum travels at a consistent speed that is independent of the speed of the observer. In other words, the speed of light is the same for someone traveling at several thousand miles an hour as it is for someone standing still. This led to the theory of relativity within four dimensional space-time.

material's index of refraction, and it allows for calculation of the angle of transmitted light after it passes through the boundary between two media.

Passing a light beam through a crystal changes both the speed of the light beam and its index of refraction. The degree of change is based on the physical properties of the crystal and the polarity of its particles. SRSC modulation splits a signal. It then manipulates the polarity of a crystal on one of the resulting light paths to control the optical refraction. Adjustments in refraction alter the relative phase of the split signals so that when they are reunited they either compliment each other or cancel each other out, producing AM modulation.

Lasers

Lasers are optical resonators or oscillators with mode adjustments[5] that can generate light with different frequencies and characteristics.

The mode options fall into two basic categories:

- **Longitudinal modes** control the spectral characteristics of a laser
- **Transverse modes** control beam divergence, diameter, and energy distribution.

SilkRoad uses a standard laser that is set to TEM_{00} propagation mode (Figure 3) with a patented new distributed feedback (DFB) process. To generate high-level Laguerre-order carrier signals, we use external modulation.

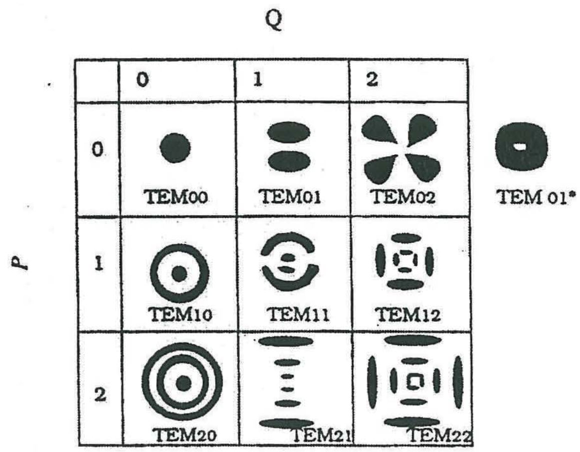


Figure 3 TEM Effects

THE APPLICATION

The simplicity of the SRSC single-laser system fosters high reliability and relatively uncomplicated solutions for add/drop, system redundancy, and other network requirements. SRSC requires less controlling software than competing systems. It also provides more channels than DWDM, faster transmission rates than SONET, and longer distances per hop than other laser-based transmission systems.

SRSC single-wavelength technology helps to control two forms of interference that limit the performance of WDM systems:

- It eliminates four-wave mixing.
- It minimizes the effects of dispersion.

This allows more information to travel further down the fiber, and it minimizes signal degradation, even on fibers that lack the dispersion compensation feature.

[5] Unless efforts are made to limit the number of oscillating modes

The Transmission of Multiple Formats

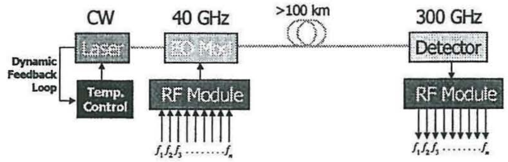


Figure 4 Combining signals on a single wavelength.

Figure 4 shows a single SRSC modulator combining channels with multiple formats on a single laser. The channels can be optical, digital, or analog. Despite these differences in format, the system transmits them as a single signal. Figure 5 shows channel stacking.

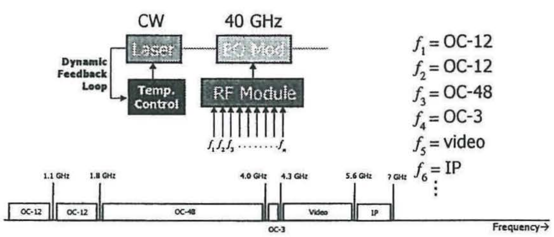


Figure 5 SRSC Simplicity of Stacking Channels

Using this approach, the signal moves in a true optical environment for distribution and management. All signals travel above their baseband in the optical environment and are returned to baseband when utilized at the other end.

This allows the SilkRoad system to use all types of modulation for any and all signal types, including SDH, SONET, RF, HFC, DS and IP.

Bi-Directionality of the Signal

To achieve the bi-directionality required for interactive TV, SRSC switches the optical phase of the signal for each direction. Bi-directional signals of the same wavelength are able to pass through each other without interference due to this difference in phase.

The conversion path on the fiber handles all the information as if it were analog (even if it is shaped in a digital form).

SRSC technology follows this conversion path on the output signal:

$$\text{Optical} \rightarrow \text{Digital} \rightarrow \text{Analog}$$

Upon return, the signal is converted from:

$$\text{Analog} \rightarrow \text{Digital} \rightarrow \text{Optical}$$

This full path is taken only if the initial input is optical; otherwise, the path is shortened depending if the signal is digital or analog. Using SRSC, the received signal is the same as the signal sent.

Span Length and Amplification

The SRSC optical signal has a number of advantages over competing technologies in how the signal is transmitted and managed on the fiber. Being a TEM_{00} optical signal, the system is able to send an HFC signal over 100 kms without amplification. When amplification is needed, a standard EDFA device is used to optically amplify the signal. As Figure 6 shows, the limitation on all EDFAs is in the raising of the optical floor (noise level) along with the signal, so the signal must be regenerated after multiple amplifications.

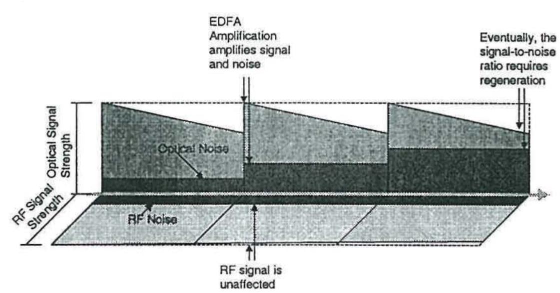


Figure 6 The effect of EDFA Amplification

With no four-wave mixing and fewer dispersion problems, the SRSC signal has less noise to amplify and fewer limitations on the degree to which the signal can be amplified. SRSC also has other means of amplification that can hold down the optical floor and maintain an improved RF signal over a long fiber link. By detecting the signal in RF space and conditioning the signal to be placed on an optical carrier, the optical floor is kept low (Figure 7) as we transport the RF signal needed at the receiving end.

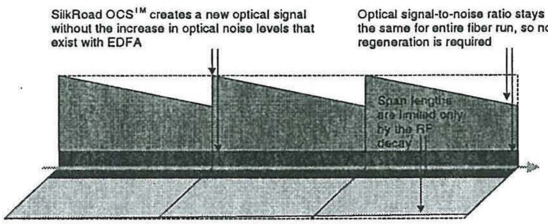


Figure 7 The Effect of SRSC Amplification

This increases the total span lengths and allows the system to maintain and deliver an optical signal matching the highest specifications and standards.

CONCLUSION

SRSC technology is a new approach that offers many benefits for the future of optical transmission signals in all industries.

In cable television, SRSC signals will be able to maintain the specifications for signal integrity over a longer distance and with greater accessibility.

About the author:

THOMAS W. MYERS is the Director of Engineering at SilkRoad, Inc. With a BS in Industrial Engineering and an MS in Electrical Engineering, Mr. Myers has instructed engineering courses at several colleges and universities. He has won several major engineering awards, including an Emmy for technical achievement, has several published books, and has been awarded four patents.

WWW Traffic Modeling for HFC Networks

Dee Denteneer† & Verus Pronk†

Philips Research Laboratories
Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Abstract

Traffic modeling for HFC networks differs from traffic modeling for queuing systems by a greater emphasis on individual connections and on short-range dependencies. These issues are illustrated by means of a model for WWW or web client traffic and a simulation of an HFC network that serves a number of web clients. These simulations offset the number of clients that an HFC network can serve against the quality of service they receive. In addition, these simulations compare the quality of service received by web clients to the quality of service received by comparable Poisson sources.

INTRODUCTION

Currently, teletraffic modeling is one of the hot topics of the telecommunications society and the body of work devoted to this topic is vast and rapidly expanding. It roots firmly in traffic measurements that establish that traffic on modern communication systems (e.g., an Ethernet or the Internet) differs in significant ways from assumptions about traffic that have been traditionally made in performance analysis.

With some simplification, one may say that the focus in this area is on models for *aggregate traffic* that exhibit *long-range dependence.* Interest in traffic due to a single source is secondary and motivated by a search for a physical explanation of this long-range dependence: often this explanation is formulated by means of heavy-tailed distributions (see [10], [14], [15], [18], or, more generally [1]). Here, long-range dependence means that time correlations between traffic loads exist for very long periods of time, i.e., they tail off hyperbolically in stead of exponentially as with the classical short-range dependent models. It is this characteristic of the traffic that is responsible for the excessive waiting times in queuing systems such as the Internet.

The relevance of these new models derives particularly from recent results in queuing analysis. In, e.g., [12], [5], and [17], it is shown that long-range dependence has an enormous impact on both waiting times and cell-loss probabilities in queuing systems. In the same vein, in [2], it is shown that waiting times in queues with heavy-tailed service times are considerably larger than waiting times in queues with light-tailed service times.

So traffic models for queuing systems rightly stress the long-range dependence of the traffic, possibly neglecting short-range dependencies.

Traffic modeling for HFC networks is different from traffic modeling for such queuing systems in two respects. First, in HFC networks, there is a clear relevance of short-range dependencies. Clearly, request mechanisms, such as multiple requests and piggybacking, make it plausible that packets generated 'close to each other' are relevant for throughput and delay. Bursts of traffic can effectively be dealt with, as it is not necessary to go through contention periods for each packet within a burst. Second, traffic modeling for HFC networks is more directly geared to the traffic generated by individual sources than to traffic form an aggregate of users, because it is the aggregation of single-user traffic itself that is the subject of investigation.

Analyzing the performance of HFC networks by means of simulation requires computer routines to generate traffic that mimics actual traffic in such systems. The sensitivity of HFC performance to traffic assumptions has, as yet, not been thoroughly investigated. However, studies that are available by now indicate that correct traffic models are of great concern and that both long-range dependencies and short-range dependencies are relevant.

As to the relevance of long-range dependence: in [8], the authors compare simulations with actually observed Ethernet traffic traces to simulations with artificial traffic, that was obtained by time-permuting these observed traces. Here, the traffic used in these two simulations is identical in one respect: the same data values are used in the simulations. However, the traffic streams differ in their time structure: the observed Ethernet traffic is long-range dependent, whereas the permuted traces are independent. Therefore, differences obtained in these simulations can be attributed to this time dependency. Their simulations show that the performance of an HFC network (measured in terms of average transmission delay) in case of the actual Ethernet traffic is much worse than the performance in case of the artificial traffic. They conclude that the correlation structure plays an important role, also for HFC networks.

In [6], the authors investigate the relevance of short-range dependencies: they compare the performance of HFC networks for traffic streams with exponential interarrival times to the performance for traffic streams with Pareto interarrival times. Again, the traffic assumption (exponential or Pareto) has an enormous impact on the outcomes of the simulations; the delays in simulations with Pareto interarrival times compare favorably with the delays in the simulations with exponential interarrival times.

Thus, there is sufficient evidence that traffic assumptions are relevant to the results (or may even determine these results) and that realistic models are needed. For this reason, we develop a model for web client traffic, which is then utilized, both to assess quality of service versus load and to further add to the current insights regarding sensitivity to traffic in HFC networks. It is the aim of this paper to further contribute to this sensitivity analysis by comparing the delay experienced by Poisson sources with the delay experienced by (very bursty) web client traffic.

The rest of the paper is organized as follows. In the next section, we present a model for web client traffic. The section thereafter describes the simulation configuration. We proceed by describing the results of simulating a typical HFC network that serves web clients, and end with some concluding remarks.

A MODEL FOR WEB CLIENT TRAFFIC

The aim of this section is to informally present a stochastic model that describes the traffic generated by a web client. Realization of this model resembles the actual traffic generated by web clients to such an extent that is not possible to distinguish artificial traces from actual traces.

Basically, the model has the following ingredients:

- The times at which packets are generated by the web client. This can equivalently be described by means of the interarrival times: the time between successive packets.
- The size of these packets.

The model then consists of the probability distribution functions that describe the variability of both interarrival times and packet sizes, and a description of the correlation between these. The familiar Poisson process is an example of a candidate model and substitutes an exponential distribution for the interarrival times, a constant distribution for the packet sizes, and assumes independence of all quantities.

However, the Poisson process falls short of our goals. Actual web client traffic is much more variable than a realization of a Poisson process and artificial traffic can easily be distinguished from actual traffic.

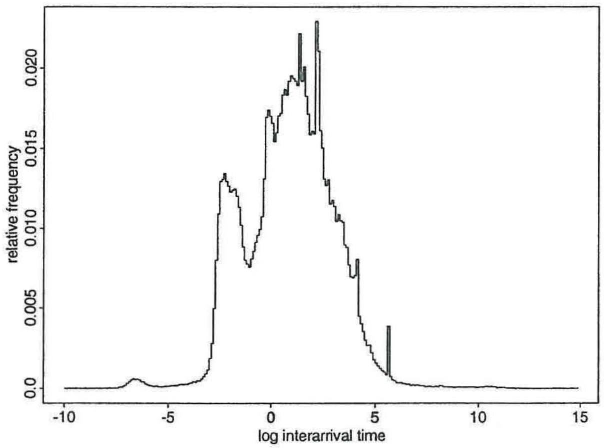


Figure 1: Histogram of the log interarrival times of successive page requests in the user process, based on a set of traces, selected from the UCB home IP usage study. The multi-modality of the histogram reflects the various time scales of the user process.

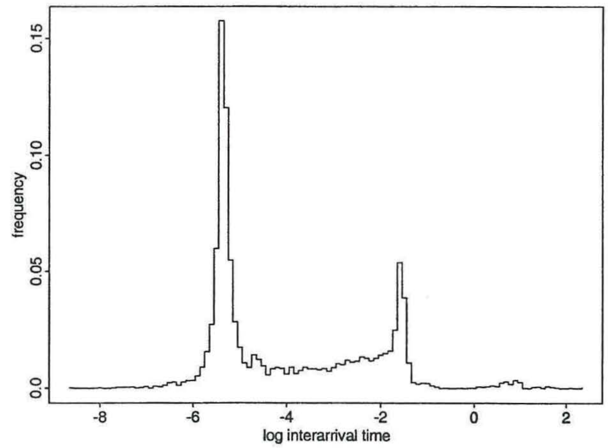


Figure 2: Histogram of the log interarrival times of successive packets in the TCP process, based on one trace from LBL-TCP-3. The multi-modality of the histogram reflects the various time scales of the TCP process.

To understand the deficiencies of the Poisson process, observe that web client traffic is governed by two processes:

The user-process This is the process as perceived by users who are browsing the web. A web browser requests a succession of web-pages, e.g., by clicking with the mouse.

The TCP-process This is the actual packet exchange that goes on between the web client and the web server. After a user requests a page, one or several (as when a requested page contains several images) TCP connections are opened. The web client traffic in each connection consists of an *open connection*, an *information request*, a series of *acknowledgements*, and a *close connection*.

Each of these two processes has its own time scale, as users typically act much slower than computers. Hence, the existence of two time scales makes it untrue that just one, uni-modal probability distribution function will suffice to describe the time between successive events.

Now this argument can be extended. Again, the TCP-level process does not consist of a homogeneous generation of packets with identically distributed interarrival times. Rather, the traffic at this level consists of *flights* of packets. The time between packets in the same flight is largely determined by the speed with which a packet is handled by a computer; the time between successive flights is determined by the round-trip time of the network. Also, a user does not request pages at a constant rate. Rather, he will alternate between various states and these states can be characterized, e.g., as *actively navigating*, *thinking*, and *having a break*. The time between successive page requests will depend on the state the user is in. So, in order to describe the interarrival time at the user level, one probability distribution function is required for each such possible state.

These states can be argued about theoretically, but they can also be observed in measurements: see Figures 1 and 2. They are both based on publicly available data, that can be found at the Internet traffic archives [7]. In Figure 1, the histogram of the interarrival times for requests of the user process are displayed as observed in the UCB home IP usage study: a collection of traffic traces that contain information on home IP usage by UC Berkeley students, faculty and, staff over a period of 18 days in Nov. 1996. Clearly, the histogram has several modes. Each mode reflects one of the time scales of the user process and each mode can be labeled with one of the user states. Figure 2 shows the histogram of the interarrival times of packets of the TCP process from the client to the server, as in LBL-TCP-3: a trace of two hours containing all wide-area TCP traffic between Lawrence Berkeley Laboratory and the rest of the world (see [15]). Again, the multimodality makes it clear that there are distinct time scales that play a role.

Our model can then be summarized as follows. The web client traffic can be characterized by means of a succession of states: either user states or connection states (within flight or between flights). The interarrival time of the packets depends on this state and can be estimated from measurement data by means of algorithms such as EM and Baum-Welch (e.g., see [11]). The duration of these states can also be estimated from these data, and in our model we have arrived at the following rules: user states are equally probable and form an independent succession. The total duration of the connection state and the duration of the between and within flight states can be inferred from the length of the file to be transferred, the round trip time of the network, and the computer speed.

Finally we note that the time between flights of packets corresponds roughly to the round trip time of the connection. This makes the models scalable: the round trip time can be shortened (artificially) so that faster networks or caching-techniques can be investigated.

SIMULATION CONFIGURATION

An HFC network with a single upstream and a single downstream channel is considered with a transmission capacity of approx. 3 and 30 Mbit/s, respectively. The downstream capacity is assumed sufficient not to form a bottleneck in the system. The round-trip delay is set to 2.6 ms., which includes transmission, propagation, interleaving, and processing delays. The transmission medium is assumed error-free.

Before their transmission upstream, application data is segmented into 64-byte *data cells* with a payload of 48 bytes, corresponding to that of an ATM cell. The overhead associated to this segmentation includes the various headers, but also physical-layer overhead. So, at the MAC layer, two immediately successive, 64-byte data cells can be considered as if they are transmitted back-to-back.

Access to the upstream channel is organized with a *request-grant* mechanism, whereby requests are sent in contention and resulting grants guarantee collision-free transmission. For contention resolution, a blocked, ternary tree algorithm is used ([3], [16]). Only after a tree has completed, a new tree is initiated.

The size of a request cell is one third the size of a data cell, so that the 'size' of each node in a tree corresponds to the size of one data cell.

The upstream transmission time is slotted and scheduled on a frame-by-frame basis. The length of a frame is 3 ms., corresponding to the transmission of at most 18 data cells in 18 slots. In each frame, at least a fixed number n of nodes of the tree *can* be scheduled. In case less than n nodes are actually scheduled, the remaining slots can be used for data cells. Conversely, if not all $18-n$ slots are used for data cells, more than n nodes in the tree are scheduled, if available.

Applications, that have requested and are still awaiting grants, are granted collision-free transmission of data cells in a round-robin, cell-by-cell fashion.

When an application runs out of pending grants, i.e., when a *request bound* is crossed, and additional cells have arrived for transmission since its last request, a new request is transmitted.

As a single web client only produces a moderate amount of upstream traffic, i.e., in the order of 1 or 2 ATM cells per second on average, a bulk Poisson source, generating single ATM cells, is used to consume the bulk of the upstream bandwidth. The remaining bandwidth is (partly) consumed by either a number of web clients or an equal number of *equivalent* Poisson clients. An equivalent Poisson client generates the same amount of ATM cells as a web client on average, but with exponentially distributed interarrival times. In this way, pairs of corresponding simulations were carried out.

For the very bursty web clients, however, cell rates observed in simulations differ substantially from the theoretical mean of 1.76/s: in simulations, which typically cover only a few minutes, sample mean rates for web clients were found that differed by more than a factor of 25. So, in order to obtain proper values for the mean rate of the equivalent Poisson clients, the total traffic produced by all web clients during a simulation was used to calculate the mean rate of each of the equivalent Poisson clients for the corresponding simulation.

Table 1 lists the simulation settings. The codes WC and EP stand for web client and equivalent Poisson, respectively. The codes *low* and *high* stand for low and high aggregate load, respectively. Simulations WC and EP *low* give a joint data load of approx. 69%, of which 2.7% is jointly generated by the web clients and equivalent Poisson clients, respectively. In simulations WC and EP *high*, these figures are 84.8% and 11.5%. For the *low* simulations, at least one node of the tree can be (and is) scheduled, corresponding to a load of 5.6%. For the *high* simulation, at least two nodes can be scheduled, corresponding to a load of 11.1%.

simulation	*nodes/ frame*	*sources*	*cell rate*	*request bound*
WC *low*	≥ 1	1 bulk Poisson	4000/s	150
		100 web clients	1.64/s	15
EP *low*	≥ 1	1 bulk Poisson	4000/s	150
		100 eq. Poisson	1.64/s	15
WC *high*	≥ 2	1 bulk Poisson	4400/s	150
		400 web clients	1.72/s	15
EP *high*	≥ 2	1 bulk Poisson	4400/s	150
		400 eq. Poisson	1.72/s	15

Table 1: Simulation settings. Note that the web-client cell rates are time averages observed in simulations and these deviate from the long-term average of 1.76/s.

RESULTS

In this section, we present some results obtained in a series of simulations. Aim of these simulations is twofold.

- To compare the QoS that web clients receive with the QoS received by equivalent Poisson clients.
- To quantitatively offset the number of web clients that can be served by an HFC network against the QoS that they will receive.

Figure 3 illustrates time series of individual cell transmission delays (CTDs) for the bulk Poisson traffic, web client traffic and equivalent Poisson traffic relating to the *high* simulations.

The figure shows a number of notable differences. First, the CTD of the aggregate web-client traffic is significantly lower on average. Second, the bulk Poisson traffic CTD in WC *high* is lower and less variable than in EP *high*.

However, what strikes the most is that the differences are caused by a change in only a relatively small portion of the total traffic.

The large differences in mean CTD are to be contributed to the influence of the more bursty behavior of the web clients on primarily the contention resolution process. This bursty behaviour generally causes successive web-client cells to be generated by a relatively small number of web clients as compared to the uncorrelated generation of successive cells by the equivalent Poisson clients. As a result, fewer web clients with larger requests will contend in a single tree, causing less delay in getting the requests to the scheduler.

For a more detailed analysis, consider Figures 4 and 5, which illustrate cumulative distribution functions (CDFs) of the various CTD series. Figure 4 shows this for the bulk Poisson traffic for all simulations. Figure 5 only considers the WC simulations and shows the CDFs of both web-client and bulk Poisson traffic.

First, Figure 4 shows the unsurprising fact that cell transmission delay increases with increasing load: the CDFs of the *low* simulations lie to the left of the CDFs of the *high* simulations. Second, it can

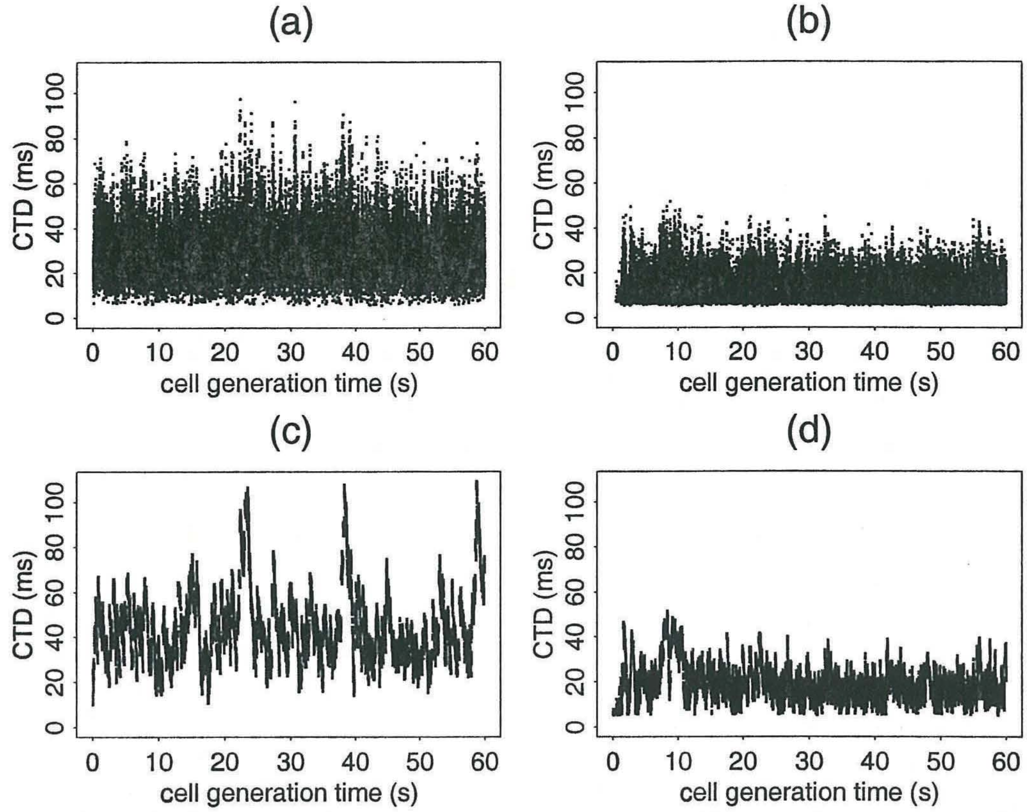


Figure 3: Time series of cell transmission delays (CTDs) for the *high* simulations: (a) equivalent Poisson client traffic, (b) web-client traffic, (c) bulk Poisson traffic that accompanies equivalent Poisson client traffic, and (d) bulk traffic that accompanies web-client traffic.

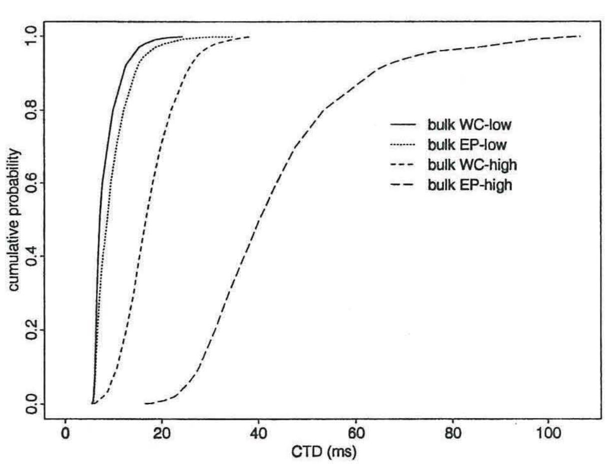


Figure 4: Cumulative distribution functions of the cell transmission delay (CTD) experienced by the bulk Poisson traffic in all simulations.

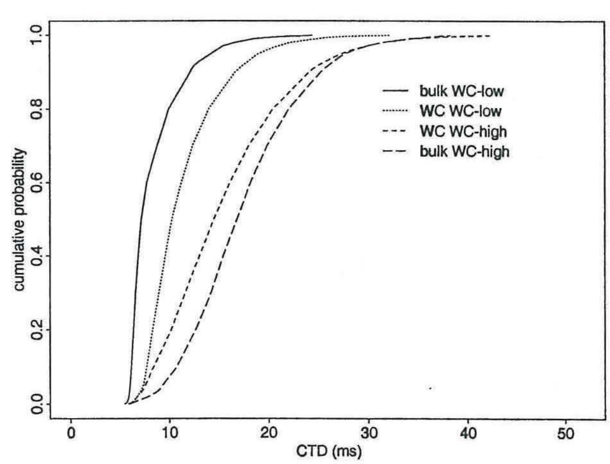


Figure 5: Cumulative distribution functions of the cell transmission delay (CTD) experienced by both the bulk Poisson and web-client traffic in the WC simulations.

be observed that the CDFs of the WC simulations lie to the left of those of the EP simulations. This indicates that transmission delay in case of web-client traffic is lower than in the corresponding case of equivalent Poisson traffic. Equivalent Poisson clients generate requests of size 1 in general, whereas web clients tend to generate larger requests, making the contention process more efficient.

Figure 5 compares the delay experienced by web clients and the bulk Poisson traffic. It shows that bulk Poisson traffic experiences less delay in the *low* simulation, but more delay in the *high* simulation than the accompanying web-client traffic. The difference in the *low* simulation is caused by the contention process: web clients will generally have to contend with the bulk Poisson source during the contention process, whereas the latter will often be the only contender. The difference in the *high* simulation is caused by the build-up of the queue for the bulk Poisson source. Similar behaviour was observed in the EP simulations.

To this comparison of the delays experienced by the various sources, we should add, however, that they also depend on the scheduler in use. In the current simulations, we have used a simple round-robin scheduler. Using fair schedulers, such as weighted fair queuing (see [4] or [13]) or weighted round-robin (see [9]), may significantly alter the results.

CONCLUSIONS

In this paper, we have illustrated the importance of accurate traffic modeling for HFC networks. Using Poisson processes only to describe traffic does not give a clear picture of HFC network performance. There is a clear need for application-specific traffic models for an accurate prediction of QoS versus load in service scenario studies. Most notable in this context is the need to also consider short-range dependencies in traffic, as well as single sources, as they significantly influence the contention-resolution process in HFC networks.

References

[1] Adler, R., R. Feldman, and M.S. Taqqu [1998], *A practical guide to heavy tails. Statistical techniques for analyzing heavy-tailed distributions*, Birkhäuser, Basel.

[2] Boxma, O.J., and J.W. Cohen [1998], The M/G/1 queue with heavy-tailed service time distribution, *IEEE Journal on Selected Areas in Communication*, Vol. 16, No. 5, pp. 749-763.

[3] Capetanakis, J.I. [1979], Tree algorithms for packet broadcast channels, *IEEE Transactions on Information Theory*, Vol. 25, no. 5, pp. 505-515.

[4] Demers, A., S. Keshav, and S. Shenker [1990], Analysis and simulation of a fair queuing algorithm, *Internetworking: Research and Experience*, vol. 1, no. 1, pp. 3-26.

[5] Erramilli, A., O. Narayan, and W. Willinger [1996], Experimental queuing analysis with long-range dependent packet traffic, *IEEE/ACM Transactions on Networking*, Vol. 4, No. 2, pp. 209-223.

[6] Harpantidou, Z., and M. Paterakis [1998], Random multiple access of broadcast channels with Pareto distributed packet interarrival times, *IEEE Personal Communications*, April, pp. 48-55.

[7] `http://www.acm.orig/sigcom/ITA/`

[8] Ivanovich, M.V., and M. Zukerman [1998], Evaluation of priority and scheduling schemes for an IEEE 802.14 MAC protocol loaded by real traffic, *Proc. Infocom '98/97.*

[9] Katevenis, M., S. Sidiropoulos, and C. Courcoubetis [1991], Weighted round-robin cell multiplexing in a general-purpose ATM switch chip, *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 1265-1279.

[10] Leland, W., M. Taqqu, W. Willinger, and D. Wilson [1994], On the self-similar nature of Ethernet traffic (extended version), *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, pp. 1-15.

[11] McLachlan, G., and T. Krishnan [1997], *The EM Algorithm and Extensions*, Wiley.

[12] Norros, I. [1995], On the use of fractional Brownian motion in the theory of connectionless networks, *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 6, pp. 953-962.

[13] Parekh, A.K., and R.G. Gallager [1993], A generalized processor sharing approach to flow control in integrated services networks: the single-node case, *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 344-357.

[14] Paxson, V. [1994], Empirically derived analytical models of wide-area TCP connections, *IEEE/ACM Transactions on Networking*, Vol. 2, No. 2, pp. 316-336.

[15] Paxson, V., and S. Floyd [1995], Wide-area traffic: the failure of Poisson modeling, *IEEE/ACM Transactions on Networking*, Vol. 3, No. 2, pp. 226-244.

[16] Tsybakov, B.S., and V.A. Mikhailov [1978], Free synchronous packet access in a broadcast channel with feedback, *Probl. Peredachi Inf.*, vol. 14, no. 4, pp. 32-59 [*Probl. Inform. Trans.*, vol. 14, pp. 259-280, 1978].

[17] Tsybakov, B., and N.D. Georganas [1997], On self-similar traffic in ATM queues: definitions, overflow probability bound, and cell delay distribution, *IEEE/ACM Transactions on Networking*, Vol. 5, No. 3, pp. 397-409.

[18] Willinger, W., M. Taqqu, R. Sherman, and D. Wilson [1997], Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level (extended version), *IEEE/ACM Transactions on Networking*, vol. 5, No. 1, pp. 71-86.

†The authors can be contacted at
{dentenee,pronk}@natlab.research.philips.com

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