THE COMPLETE TECHNICAL PAPER PROCEEDINGS

FROM:

NCTA Technical Papers

A New Paradigm for a Multi-services Cable Communications System

C. M. Breevoort, Ph.D.
ANTEC/ESP

Abstract

This paper proposes a flexible architecture for a multi-services cable communications system that not only supports telephony, data, and digital video services, but also potential future services. A unique network interface unit (NIU) at the home entry point functions as a hub to the in-home cable network, isolating the outside cable plant from any possible ingress from the home. It passes high-speed video data uninterrupted in the downstream direction, while remodulating low-speed data to an area in the 5-40 MHz band. The NIU always remodulates data flowing in the upstream direction between end-terminal devices and the headend. Adoption of an in-home network protocol based on the Universal Serial Bus (USB) modified for coax cable is an inexpensive way to form a link between the NIU and a plethora of USB devices that are appearing on the market. The USB protocol on the in-home network uses an efficient modulation scheme.

INTRODUCTION

The last few years the cable industry has actively pursued the introduction of new digital services over the hybrid fiber-coax (HFC) plant to supplement their more traditional revenue stream of the familiar analog broadcast and pay-per-view (PPV) services. Stand-alone offerings to subscribers include cable telephony, data, and digital video services. All of these services require different amounts of data with various transfer data rates and protocols. Some can tolerate delays, while others would suffer a loss in quality. They share the spectrum above the analog broadcast channels for the downstream direction (to the home), and below it for the upstream direction (to the headend).

Quadrature Amplitude Modulation (QAM) is one of the modulation schemes used in discrete bands in the downstream direction. Each 6 MHz analog channel supports approximately a 27 Mbps data stream for QAM-64, in which one symbol represents 6 bits, or a 35 Mbps data stream for QAM-256, in which one symbol represents 8 bits. The downstream data may represent a broadcast signal intended for all homes as in the case of digital video, or an encrypted message intended for one particular receiver.

The headend controls the upstream data flow using a time division multiple access (TDMA) approach to allow multiple subscribers to share the same frequency spectrum. The upstream signal path uses a Quadrature Phase Shift Keying (QPSK) or 16-QAM modulation scheme. The upstream data rate for each 2 MHz analog channel amounts to approximately 2.56 Mbps for QPSK. The cable plant reserves the 5-40 MHz band for the upstream return channels. Unfortunately, the cable return path presents some engineering challenges [1]. MSOs often use the lower 10 MHz of this band for network maintenance or find it unusable due to ingress. It effectively leaves only a 25 MHz band per node, still exposed to burst noise and ingress, for all services

The increasing demand for interactive services over a two-way cable plant creates problems in network management, channel allocation, maintenance, and operational support. As newer services, such as energy management, digital audio, home security and automation, electronic games, and others not yet imagined, supplement digital video, cable telephony, and data services, it becomes paramount to utilize the available upstream spectrum effectively. Sharing of frequency bands becomes mandatory in order to achieve acceptable upstream data rates.

Different economic models apply to the services offered over the HFC network. The MSO understands the models for digital video, telephony, and data very well. Often they are looked at as three individual services sharing one common transport medium. Although there are many components shared among the end-terminal devices, i.e., the components that convert the digital data streams into a usable format for each service, few companies are considering integrated solutions. Some companies are contemplating the combination of two services, such as a set-top box with a data port, or a digital phone with a data port. Others focus on a powerful box at the side of the house that combines video, data, and telephony services.

The lack of an integrated approach greatly hampers the deployment of additional services. Especially low bit rate services, such as energy management, home security, or home automation are prohibitively expensive to introduce as a cable service if they require a costly QAM receiver and QPSK transmitter. Cable companies will not get access to these services, and their corresponding revenue streams, as long as a simple and cost effective way to interface them to the cable plant is missing.

Headend equipment becomes more and more complex with the addition of each new service. Does the new service obtain its own upstream channel? What are the consequences in case of high bit error rates? How is spectrum management arranged? Which service controls the upstream channel allocation? What about power calibration and ranging between the headend and the many devices in the home? As the MSO adds services to the two-way cable plant, it becomes more difficult to manage a reliable communications link.

The first section of this paper focuses on the requirements and existing architectures for expanded cable services. The following section proposes an integrated architecture implementation that offers advantages over existing individually deployed cable services. The final section addresses some implementation issues for the proposed architecture.

MULTI-SERVICE REQUIREMENTS.

Expanded cable service requirements.

The characteristics of the various digital data services differ in significant ways. The wide range of applications requires different data rates, data packets, data flows, and error handling. Table 1 summarizes the important features of present and some future services. The table lists the service type with the associated data rate and data format. Other categories listed include the packet transfer delay, i.e., the time difference between the packet generation and the delivery; the effect of the variation in this delay; the effect of loss of packets; and means of error handling.

We may safely assume that the amount of data flowing toward the home far exceeds the amount of data originating from the home. This is especially true if digital video services are provided. As Table 1 reveals, the listed services can be divided into two broad categories. The first includes low to medium bit rate services that can be processed by moderately advanced general purpose processors. Even their combined data throughput rates easily fit into one downstream QAM channel. Simple to moderately complex software programs are suitable to process the associated data streams.

The second group includes high bit rate services, such as digital video, which require large amounts of data to arrive at the home. They presently require dedicated processors to convert the digital data to the proper output format. Their data pipes occupy significant amounts of a downstream QAM channel.

Presently, cable companies install equipment for each individual service. This may not be a cost effective approach for the long term. An integrated solution results in improved spectrum management, easier network design, simpler maintenance and troubleshooting, fewer network management and operational support issues, and a much improved architecture for expansion [2].

Another important point is that a significant amount of end-terminal equipment requires cable company support to install and service. Deployment of set-top boxes and cable modem equipment requires a significant amount of support. The desire is to relegate the cable company to its original role of a service provider of analog or digital information over the cable.

Multi-service architectures at the home

There are several approaches to design multi-service architectures at the home. Reference [3] describes an integrated box at the side of the residence. A recent paper [2] proposed one with approaches based upon:

1) a digital selection and multiplexing of all downstream services in the headend,
2) remodulation of the QAM signals into Bipolar Phase Shift Keying (BPSK) signals in the home operating in the 930 MHz range,
3) BPSK for upstream signals in the home, also at 930 MHz, and
4) isolation between home wiring and HFC network to ensure network spectral integrity.

The Network Interface Unit (NIU) in this configuration passes the analog video signals and terminates the HFC plant with a tuner for the QAM receiver. It demultiplexes the incoming signal to feed the data, telephony, and digital video interfaces. The NIU becomes a "super box" with support for the most popular services. It may also require additional wiring between the outside box and the end-terminal devices inside the home, such as phones, set-top boxes, or PCs.

Service Type	Average Data Rate and Data Format	Packet Transfer Delay (Δt source - destination)	Packet Delay Variation	Packet Loss Effect	Error Handling
Digital Video	High (2-8 Mbps), MPEG-2 transport stream. (sometimes embedded in ATM.)	Moderately sensitive. Remote control commands must result in quick channel change.	Constant bit rate minimizes receiver buffer size.	Error concealment and correction possible to a certain extent.	Masking of errors by video and audio decoders.
Voice Telephony	Low (64 kbps), isochronous PCM data.	Very sensitive. Cannot tolerate large delays, otherwise requires echo cancellers.	Very sensitive.	Somewhat sensitive. Dropped data can be calculated from neighboring samples.	Simple interpolation routines.
Data	Variable data rates (100 kbps - Mbps), ATM packets.	Insensitive, as long as data streams not used for videoconferencing or voice calls.	Insensitive. Buffers are normally available.	Critically sensitive. Cannot tolerate missed data. Requires retransmission.	Application program runs error checking protocol.
Energy Management	Low (< 1 kbps.)	Insensitive.	Insensitive.	Sensitive	Error checking built-in in protocol.
Digital Audio	Medium (< 500 kbps), MPEG-2 transport stream, Dolby AC-3, etc.	Sensitive.	Sensitive.	Moderately sensitive.	Audio decoding process performs error masking.
Home Security and Automation	Low (< 1 kbps.)	Insensitive.	Insensitive.	Moderately sensitive.	Retransmission.
Video Conferencing	Medium (< 384 kbps), H320, H324, T.120 formats.	Very sensitive.	Somewhat sensitive. Normally sufficient amount of memory available.	Moderately sensitive. No retransmission for video and audio.	Video decoder can mask lost data. Quality of service is reduced.
Video-on-demand	(Similar to digital video.)	Moderately sensitive. Remote control commands must result in response.	Constant bit rate minimizes receiver buffer size.	Error concealment possible to a certain extend.	Masking of errors by video and audio decoders.
Internet Access	Medium (500 kbps - Mbps.)	Insensitive, as long as data path not used for video or audio playback.	Somewhat sensitive.	Sensitive. Requires retransmission.	Application program performs error checking.
Video Game Delivery	Low.	Insensitive.	Insensitive.	Sensitive. Cannot tolerate erroneous data.	Retransmission.

Table 1. Different Service Requirements over Cable.

There are some inherent limitations associated with this architecture. It deploys a single tuner in the NIU for all downstream signals. It forces the multiplexing of all requested services. This approach is probably not a problem for the low data rate services, but it causes problems for high data rate services such as digital video. Suppose that 100 homes are watching the same digital video stream on their set-top box. Their NIU tuner tunes to the same frequency. Suppose now that out of this group ten homes have a second set-top box required to be tuned to ten distinct digital video streams. Suddenly, the bandwidth is not there to multiplex eleven digital streams into one QAM channel. This shortcoming multiplies with an increasing number of services and homes on-line.

Other limitations of the NIU supporting multiple services are (1) the difficulty of powering a complex NIU through the cable network; (2) limited expansion capability for new or updated services; (3) sometimes proprietary modulation schemes incompatible with end-terminal devices; and (4) sooner or later the NIU will not have enough resources to support the plethora of requested services, especially high-speed video streams.

A NEW MULTI-SERVICE ARCHITECTURE AT THE HOME

The earlier attempts of integrated architectures made significant strides towards sharing of cable plant modem components. The combined tuner and QAM demodulator, and one upstream QPSK modulator, result in a significant cost saving for the deployment of multiple services.

The proposed architecture follows the same basic principle of sharing cable plant modem components, but distinguishes itself in several important ways. It still deploys an NIU at the side of the home, but it contains an inexpensive (bi-directional) modem operating in a narrow band in the 5-40 MHz range as an additional feature. This modem supports the Universal Serial Bus (USB) [4] protocol. The USB is a bus originally designed for the Personal Computer (PC) environment.

A Cable Termination Node (CTN), which consists of a similar modem with several standard Universal Serial Bus ports connected to end-terminal functions (telephone, speakers, PC, etc.), is placed at the end of the in-home cable. The CTN is basically a Universal Serial Bus (USB) hub with an upstream port modified for the coax cable.

Figure 1 shows a block diagram of the NIU. The main hardware characteristics are:

1) Simple NIU with minimum amount of hardware, fit to be powered by the cable plant, and independent of any particular cable service.
2) It terminates the HFC plant with a tuner for the QAM receiver in the downstream direction, but also passes a complete (buffered, amplified) downstream signal carrying the analog AM-VSB channels as well as all digital data channels into the home.
3) The NIU is responsible for remodulation of all services, except the high-speed ones, onto the in-home cabling in the downstream direction.
4) The NIU is responsible for remodulation of all services between the in-home cabling and headend using QPSK or QAM-16.
5) The NIU functions as a master device to in-home devices through a bi-directional link in the 5-40 MHz band deploying a standard Universal Serial Bus protocol.

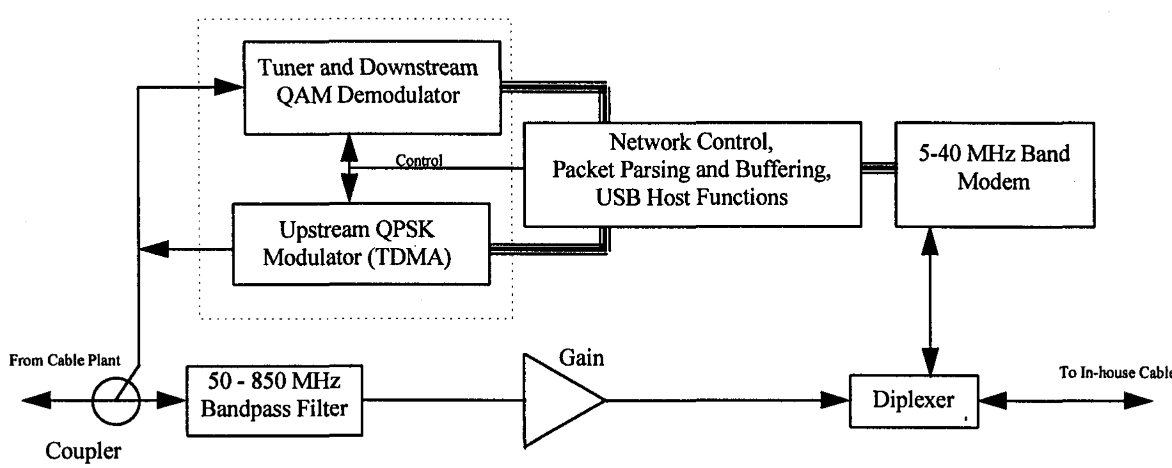


Figure 1. Network Interface Unit (NIU).

The following sections explain in more detail the reasons and justifications for the proposed architecture, and the NIU and CTN implementation.

Simple hardware implementation.

One of the objectives of this architecture is to make the NIU hardware independent of any of the possible offered services. A direct benefit of this choice is its simplicity and the associated low power requirements. The intention is to keep the NIU independent of any service, but use downloadable software programs to "customize" the NIU. The NIU QAM demodulator receives the low speed data services. The host microprocessor in the NIU forms the interface between the modem connection to the cable plant and the modem connection to the in-home cable. It functions as a USB host controller to the in-home network. A small integrated RISC processor handles the required processing load.

The simplicity of the NIU and the limited number of standard connections make it possible to install the NIU under ground level to protect it from weather extremes. Regular access to the NIU after installation is not required.

Buffered by-pass path.

The advantage of the high-speed by-pass path is that multiple set-top boxes (or digital TVs in the near future), each with their own tuner and QAM receiver, can be used in a residence. The headend does not require to perform any multiplexing of high-speed video streams for a particular residence. Another advantage is that future (high-speed) services can enter the home. The NIU does not prevent the downstream data flow from entering the home in any way.

Remodulation of all low-speed downstream services.

The motivation behind the selection of the USB for the cable environment is similar to the one for the PC environment. The intention is that the standard interface encourages consumers to use standard USB devices and connect them to the in-home cable without fear that the standard will become obsolete or will lose compatibility. It releases the cable companies from specifying end-terminal devices or maintaining them.

Some of the criteria for the definition of the Universal Serial Bus, which focuses on computer telephony integration (CTI), consumer, and productivity applications, certainly apply also to the in-home cable network:

1) Ease of use for peripheral expansion.
2) Low-cost solution that supports transfer rates up to 12 Mbps.
3) Full support for real-time voice, audio, and compressed video.
4) Protocol flexibility for mixed-mode isochronous data and asynchronous messaging.

Remodulation of all upstream services.

Communication between the NIU and the headend uses a QPSK or QAM-16 modulation scheme. The headend only needs to control ranging and power level management with respect to this transmitter and not with in-home devices. The NIU isolates the cable plant from the in-home network.

The NIU functions as a host USB Controller.

The Universal Serial Bus is a cable bus that supports data exchange between a host computer and a wide range of simultaneously accessible peripheral devices. The host uses a token based protocol to share the available bandwidth with up to 127 devices. The PC normally functions as the USB host. In the cable environment, the NIU takes over this function. There is only one host on any USB system, which may be implemented in any combination of hardware, firmware, or software. The NIU is very well suited for this function.

The function of the host is executed through the host controller. It detects the attachment and removal of USB devices to the cable, manages control flow between the host and USB devices, and collects status information. Unlike in the PC environment whereby the host provides a limited amount of power to attached USB devices, the NIU does not provided any power through the coax cable to any attached CTNs.

The Cable Termination Node (CTN).

The Cable Termination Node (CTN) is required in order to modulate and demodulate the narrow band signal in the 5-40 MHz band to a standard USB signal. The CTN functions as a generic USB hub with several downstream USB ports, but with the distinction that the upstream port is remodulated for coax cable transmission.

Figure 2 shows a block diagram of the CTN. The CTN consists of a standard USB controller chip, a USB driver chip, and a power management section in addition to the interface to the coax cable, formed by a modem identical to the one found in the NIU.

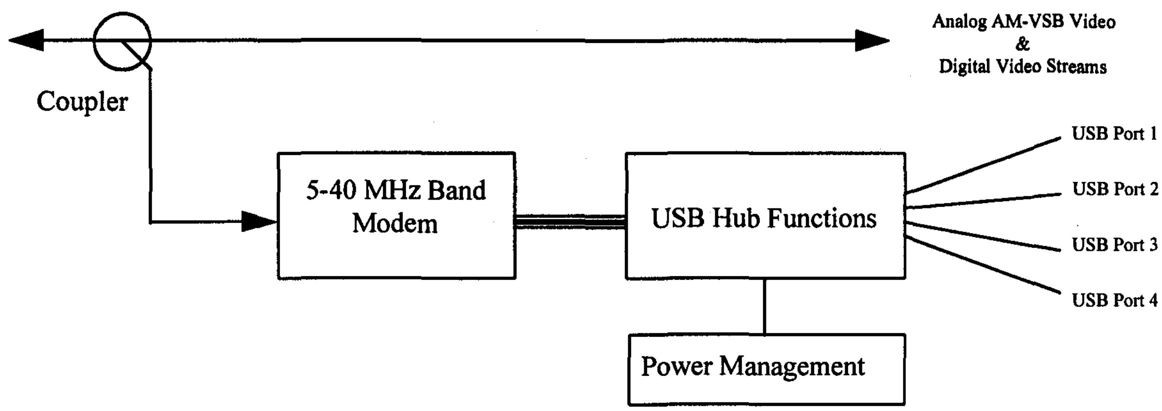


Figure 2. Cable Termination Node with Four USB Port Interfaces.

The CTN is a self-powered hub, i.e., it has a local power supply that furnishes power to any embedded functions and to all downstream ports. The disadvantage of this arrangement is that the NIU has no visibility of the hub when the CTN's power supply is off. It is therefore not possible to differentiate between a disconnected and an unpowered CTN. The NIU assumes that an unpowered device is a disconnected device. An analog path in the downstream direction guarantees that analog AM-VSB signals and QAM digital signals can bypass the CTN. The CTN is small enough to make it fit inside a standard cable outlet box.

USB OVER IN-HOUSE CABLE ISSUES.

The adoption of the Universal Serial Bus for the in-home network applies an existing bus and protocol to a new transport medium. Other in-home buses, such as the Consumer Electronics (CE) bus, cannot meet the requirements for the desired data rates over the in-home coax network. As these buses develop an interface to the USB serial bus, they can be interfaced to the proposed network.

The USB system describes three areas: (1) USB host; (2) USB devices; and (3) USB interconnect. The NIU performs the function of a USB host. The USB devices are divided in hubs and functions. A hub provides additional attachment points for the USB. Installation of USB devices for the cable environment should be as straightforward as in the PC environment: they attach to a CTN inside the home and will be recognized first by the attached CTN and then by the NIU.

The USB physical interface is based upon a four wire cable with two wires for power and ground, and the remaining two wires for signaling over point-to-point segments. The cable carries +5V on each segment to deliver power to devices. The USB specification limits the maximum cable length to several meters in order to meet IR drop specifications and guaranteed input voltage levels. The signals on each segment are differentially driven.

The proposed architecture requires a modification of the USB physical interface between the NIU and the CTNs. The standard physical output of the NIU is a coax cable entering the home. This cable may split several times before terminating in several rooms.

The Universal Serial Bus is a standard bus defined for the Personal Computer environment. The PC, as the host, uses several ports to communicate to hubs or functions. There is only one physical cable entering the home with passive taps going to multiple endpoints. The USB specification describes the bus attributes, protocol, types of transactions, bus management, and the programming interface to design and build systems and peripherals that are compliant with this standard. An extension to a different physical medium, such as the coax cable, requires some modifications.

The protocol relies on differential signals to indicate bus states and device conditions. We do not recommend duplication of these (DC) signals on the in-home coax cable due to the likelihood that a DC path may not exist. The NIU can circumvent these issues by interrogating from time to time all attached CTNs.

The USB specification defines two different data rates: a high-speed one at 12 Mbps and a low-speed one at 1.5 Mbps. Support of the low-speed data rate is not required in the cable environment since the NIU only communicates directly with the CTNs functioning as generic hubs attached to the end of the in-home cable.

Modulation scheme for the in-home network.

The in-home network must implement the Universal Serial Bus protocol over the coax cable. The required maximum data rate is 12 Mbps. The most suitable band available is between 5 and 40 MHz. The use of other bands, mainly above the analog video bands and above the digital QAM channels, would require expensive modulation and demodulation stages in the receivers and transmitters. What we are looking for is an implementation that is suitable for low-power CMOS ASICs.

Traditional modulation schemes come to mind. An implementation is possible with FSK, or BPSK, but these are expensive in bandwidth. For the requested data rate, the BPSK modulation scheme requires at least 15 MHz of bandwidth, thereby almost absorbing all the available bandwidth. QPSK is less demanding in bandwidth. Using direct-digital-synthesis, it is possible to implement transmitters in single chip CMOS circuits. QPSK receivers are more complex, requiring an analog-to-digital converter and some digital signal processing.

A recently proposed modem modulation method called INTRA (INformation TRAnsformation) is based on wavelet theory and the principles employed in multi-rate quadrature mirror filter banks (QMFs) [5,6]. Wavelet functions are functions that are localized in frequency and time. This method achieves data rates comparable to QAM data rates, but can easily be implemented in standard CMOS circuitry without the need for analog-to-digital converters or complex digital signal processing. The required bandwidth is also comparable to QAM.

The INTRA transmitter consists of a shift register and a ROM look-up table. The ROM stores the values for the wavelets, i.e., the short bursts of RF energy. A high-speed digital-to-analog converter converts the ROM digital output to the analog domain. The semiconductor industry has shown with QPSK modulator chips that it can implement high-speed digital circuitry and a digital-to-analog converter on a single die. The INTRA transmitter requires even less circuitry than QPSK transmitters.

Matched transversal digital filters receive the transmitted orthogonal wavelets. Since the wavelets are orthogonal, there is no inter-symbol interference. The receiver correlates the received waveform with the known waveform shapes. The simplicity of the INTRA receiver makes it also attractive for the in-home modem.

The NIU may also output the INTRA signaling over twisted pair wires as an alternate to the coax cable. It allows home owners to hook-up USB devices through the CTNs using less expensive wiring or to add additional wiring.

Software and cable applets (capplets).

The NIU and CTN require a certain amount of software in order to operate a service independent architecture as much as possible. The CTNs inside the house perform a function similar to a standard USB hub. Presently, there are small USB hub controllers available from multiple vendors that implement this function. The expectation is that these controllers will find a place in the CTN without many modifications to the existing software.

The NIU performs the function of a USB host controller and requires software to support this function. The attachment of USB devices in the home triggers the NIU with a request for acknowledgment of the function and possible additional software support in the form of a device driver or application program. A request to the headend by the NIU for support of the new USB function results in the download of a so called cable applet, or capplet for short. This capplet orchestrates, in cooperation with the embedded USB host software, the proper operation of the attached device. Its function is to transforms the downloaded data stream into a usable format for the device.

SUMMARY

This paper has suggested a new architecture for an integrated multi-service cable communications system. The Network Interface Unit (NIU) at the side of the customer premises functions as a Universal Serial Bus (USB) host device to support present and future end-terminal devices inside the home. The architecture allows for integration of future services through the support of the popular USB bus. All the USB devices available to the PC environment, become available to use as attachments to the cable.

The NIU communicates to the headend similar to a traditional cable terminal device using QAM demodulation in the downstream direction and QPSK modulation in the upstream direction. It communicates with the in-home Cable Termination Nodes (CTNs) using the Universal Serial Bus protocol modified for the in-home cable environment. This paper proposes to use a wavelet based modulation scheme, called INTRA, between the NIU and each CTN as an inexpensive and robust in-home modem.

The architecture isolates the in-home network from the outside cable plant. An added benefit of the achitecture is that it greatly simplifies cable plant design, and reduces maintenance.

REFERENCES

1) N. Himayat, *et al.*, "Characteristics of Hybrid Fiber-Coax Return Systems," Society of Cable Telecommunications Engineers Cable-Tec '96, pp. 191-199, Nashville, June 1996
2) A. C. Young, M. F. Davis, and J. B. Terry, "Multi-services Cable Telecommunications Systems," Society of Cable Telecommunications Engineers Cable-Tec '96, Nashville, June 1996.
3) J. Barthold, "Cox To Test Voice-Data-Video Integrator Box in Okla. City," Cable World, p. 10, November 18, 1996.
4) Universal Serial Bus Specification, Revision 1.0, January 15, 1996. See "http://www.usb.org/".
5) B. Miller, Document #IEEE802.14-96/127, INTRA Characteristics, IEEE P802.14 WG, April 24, 1996.
6) B. Miller, Document #IEEE802.14-96/097, Analog Interface for Upstream INTRA, IEEE P802.14 WG, March 8, 1996.

A PROPOSED METHOD FOR QUANTIFYING UPSTREAM INGRESSING CARRIERS

David Large, Media Connections Group
Rex Bullinger, Hewlett Packard

Abstract

The authors propose a methodology for analyzing ingressing carriers and intermodulation products independently from electrical transient interference. The evaluation results are presented in a simple form which can be used to rate system performance, judge the effects of mitigation measures, and predict changes resulting from changes in operating parameters.

INTRODUCTION

Upstream (subscriber-to-headend) transmission in broadband distribution systems differs materially from that in the downstream direction due, in part, to the frequency ranges chosen for reverse transmission and the branching nature of the network.

The reverse spectrum includes frequencies used by many over-air transmitters, including citizens band, amateur radio operators and short wave broadcasters. Also, second-order intermodulation (IM) products from mixing among downstream signals are strongest at the upstream frequencies. Finally, electrical transients (from a multitude of sources) have a spectral content which is stronger at frequencies below 30 MHz.

Network branching and the lack of switching cause signals introduced at all points in the system to combine at the optical node or headend, increasing the chances of interference.

PREVIOUS STUDIES

Several authors have reported the results of studies of ingress, including extensive work done by Rogers Cable Television and CableLabs.[1] The results of such studies have frequently used detailed tabular and/or graphical information on ingressing signal levels and frequencies over various time periods.

This information has been invaluable in gaining a general understanding of the challenges associated with upstream data transmission. Unfortunately, the typical complex presentation of the data is hard to relate to network performance and many of the studies involved complex analyses of the raw data and/or special instrumentation.

REQUIRED SEPARATION OF TRANSIENTS AND CARRIERS

One difficulty has been that much data has been taken with spectrum analyzers so that "events" included the effects of both short electrical transients and longer term ingressing radio carriers and distortion products.

Since the duration of electrical transients was typically short compared with the sweep rate of the analyzer, they appeared as narrow-band interfering signals at whatever frequency the analyzer happened to be receiving at the time of occurrence, whereas the spectrum of most transients includes the entire return band.

In addition to the problems of accurately quantifying the impact of transients in such mixed presentations, they need to be measured and treated separately because the tools for dealing with these impairments are very different. Discrete frequency problems can be avoided by moving upstream transmitter frequencies, while short broadband transients can best be avoided through error correction and/or interleaving techniques, combined with re-transmission when required.

SOLUTION REQUIREMENTS

A useful tool for evaluating discrete product interference should have the following characteristics:

1) It should have good statistical processes which will average out the quasi-random nature of ingressing signals, so that general ingress conditions in a plant can be determined consistently enough that the effect of, for instance, various mitigation measures can be measured.

2) It should provide sufficient data to enable users to choose channel frequencies and/or frequency ranges and to predict the ingress-related performance that will result.

3) It should allow the user to evaluate the service availability effects of changing such parameters as data channel level or frequencies.

4) Given that many transmitters are frequency agile, it should be able to predict the availability gain resulting from various degrees of mobility.

5) Perhaps most importantly, it should provide a limited number of "figures of merit" which can be used to judge the overall ingress situation without further analysis.

The proposed methodology meets all these criteria.

OVERALL PROJECT

This tool was developed as part of an extended and detailed study of data transmission through HFC cable systems. Previous reports have included a general status report on the study,[2] as well as a detailed study of downstream laser clipping.[3]

Overall goals of the study have included the providing of guidance to product designers on the quality of the transmission path and the development of field engineering tools and methods for evaluating cable systems' ability to carry bi-directional digitally-modulated signals. The development of simple methodologies for quantifying all types of ingress has been a major challenge.

AVOIDING ELECTRICAL TRANSIENTS

As with other researchers, we used a spectrum analyzer as our basic instrument for gathering ingress data. The analyzer was programmed to make measurements over a period of time which were then downloaded to a standard personal computer for analysis.

In order to understand how to collect information on ingressing carriers that was not "polluted" by electrical transients, we first evaluated transient events as they appeared at the headend. Although methods for evaluating transient ingress are not discussed in this paper, an understanding of their general nature is important.

The waveforms of representative transients were recorded using a fast digital storage oscilloscope directly attached to the upstream optical receiver in the headend. Using the instrument's internal fast Fourier transform (FFT) capabilities, both the time and frequency domain data were saved for analysis. Figures 1-4 show the general range of observed events. The upper trace in each picture is the time signature of the waveform, while the lower trace is the same event plotted in the frequency domain.

As can be seen, most transients take the form of "bursts" lasting no more than a few microseconds. By contrast, ingressing radio signals are present for much longer time periods. We therefore "filtered" out transients by reducing the spectrum analyzer's video bandwidth to only 300 Hz. Thus the rise time of the video circuitry was about 1.2

milliseconds - many times the typical transient duration. By slowing down the sweep rate so that the dwell time per frequency was sufficiently long, accurate recording of discrete carrier levels was still preserved.

DATA GATHERING

The spectrum analyzer was connected to the headend optical receiver output. The node analyzed passes approximately 2,000 homes, with a forward cascade of four amplifiers, and contains between 40 and 50 active devices. The reverse passband extends from 5 to 30 MHz. The only intentional upstream signal present was a bursted QPSK test signal centered at 24 MHz.

Approximately 1300 basic customers are served by the system, which is fully two-way to every tap. No filters or drop equalization are used, so the node is "wide open" to ingress.

The analyzer was set up to record peak and "average" (see below) signal levels at frequency intervals of 100 kHz and over time windows of one hour. While this limited the number of data points to be analyzed, shorter time periods would result in a more useful analysis and we expect to do a more detailed time-axis analysis in the future.

The resolution bandwidth was set to 300 kHz and the video bandwidth, as discussed above, to 300 Hz. With these settings, the sweep rate was less than one per second, but still fast enough to get a statistically valid number of samples in each hour.

At the end of each measurement time period the data was downloaded to the connected microcomputer, so that two matrices were created, each with time on the horizontal axis and frequency on the vertical axis and with the peak levels in the cells of one matrix and the "average" levels in the other.

DATA ANALYSIS

Figure 5 shows the "raw" data from the peak value chart. This is not unlike those published by several other researchers. The modem signal at 24 MHz is easily seen as are the highly variable peaks associated with CB radios in the 27 MHz region. This system is remarkably free from common path intermodulation distortion products which would have shown up at 6 MHz and its harmonics. Finally, it shows the typical rise in overall ingress levels at lower frequencies. The chart is relatively free of the random spikes caused by transients, while data taken with wider video bandwidths was dominated by large amplitude spikes.

Determining Threshold Interference Levels

Since the aim of the tool was not just to collect level information, but to relate that information to service disruption potential, the next step was to carefully measure the maximum level of interfering carriers that would not disrupt the QPSK-modulated signals under consideration. This was measured under both noise-free conditions and with a carrier-to-4 MHz noise ratio of 29 dB. The results were consistent within about 1 dB. As shown in Figure 6, the receiver was essentially immune to signals offset by more than 900 kHz from the channel center, but a carrier-to-interference (C/I) ratio of approximately 14 dB was required for signals closer to the channel center to maintain a BER below 10^{-5} (our arbitrarily defined failure level). Tests also showed that the BER changed by a little over one order of magnitude for each decibel change of interfering signal level.

Given the required C/I ratio and the desired signal level, we next computed the threshold interference level. Clearly, for any frequency, if the peak level did not exceed the threshold during a given measurement period, the data could be ignored.

In cases were the peak value of a signal exceeded the threshold, we attempted to approximate the percentage of time the carrier was present. In our early analyses, we have used a peculiarity of the particular analyzer model which is that its "average" power reading is actually an average of the screen positions on successive sweeps, rather than true average power. While this is a problem for some measurements, we used it to advantage. In particular, if the carrier level, when present, is relatively constant (within 10 dB or so) the fractional time present can be approximated using:

$$\frac{On\ Time}{Total\ Time} \simeq \frac{(Peak-Noise)}{(''Average''-Noise)}$$

where: Peak is the maximum recorded power level, using the analyzer's peak detector function, at a particular frequency and over all the sweeps during a time interval;
Noise is the observed system noise floor at the resolution bandwidth used; and
"Average" is the average of the screen positions on successive sweeps.

This will be a much better approximation when time intervals are relatively short. Alternately, at the expense of more post-processing, we could save the data from each sweep separately.

Frequency Unavailability

Using the above method, we can create a chart showing what percentage of time each frequency is impaired. Figure 7 shows this transformation of the raw data.

Note that, although the peak levels in the 27 MHz region were very high, since any individual carrier is generally present a relatively small fraction of the time, the percentage unavailability is low compared with the weaker, but longer duration, signals at the lower end of the spectrum. Also, note that the 24 MHz modem signal is present, but at a level which reflects the light duty cycle of the test sequence running at that time.

Channel Unavailability

Although the frequency unavailability chart is certainly more readable than the raw signal levels, it is still not useful in predicting system performance. To do that, we need to examine each possible communications channel over its susceptibility bandwidth. As discussed above, we found 2 MHz QPSK channels to have a "susceptibility bandwidth" of about 1.8 MHz. Thus, an ingressing signal will affect any channel whose center frequency is within ±900 kHz.

Figure 8 reflects this by determining the unavailability of communications channels as a function of channel center frequency and time. As can be seen, rather broad sections of spectrum can be affected by combinations of closely-spaced ingressing signals. For instance, if the 27 MHz CB interference is to be avoided, the only channel frequencies available between the 24 MHz test signal and the top of the spectrum are at 26 and 29 MHz.

The spreadsheet, in addition to the chart, includes a table of average channel availability as a function of center frequency, so that users can match channel quality to service requirements.

Service Unavailability

Although the above chart shows the data for fixed channels, many proposed upstream services employ frequency agile transmitters and include logic to direct communications to clear channels within a specified service bandwidth.

In order to estimate the possible "availability gain" from such a system, we added a Service Bandwidth definition to the model. It then calculates the service unavailability in each time slot by choosing the least impaired channel frequency. The gain relative to any fixed channel can be significant. Figure 9 shows the minimum possible unavailability of a service extending from 9-14 MHz, as an example. Also plotted are the unavailability of the worst, best and average of all channel frequencies in the service range. As can be seen, even the best fixed channel has an unavailability exceeding 10% at some times, while the overall service availability is 100% except for one time slot for which unavailability rose to just over 10%.

Overall, the defined service unavailability was just 0.25%, compared with 5.3% average for all possible channel frequencies in the service group - a reduction factor of greater than 20.

It should be stressed that this improvement factor is the theoretical maximum and assumes that:

A) the system responds instantaneously to interference conditions and,

B) that it correctly chooses the best new frequency for communications.

Nevertheless, it is a useful tool for judging the relative performance possible as a function of spectrum utilization.

Summary Data

While the graphical data presented in this paper is useful for understanding the analysis process, the system generates several overall quality values that may be more useful for system maintenance purposes:

A) The percentage unavailability of all frequencies, averaged over time and frequency.

B) The percentage unavailability of all possible channel center frequencies, averaged over time and frequency.

C) Given a defined spectrum for a particular service, the average unavailability of all possible channels within the service group.

D) The minimum theoretically possible unavailability for a service using agile transmitters.

These numbers can be used as a reference to compare systems, to judge overall quality trends within a system and to evaluate the results of ingress mitigation measures.

Parametric Analysis

Another use for the tool is to evaluate the effects of technical options. For instance, the level of the data carrier can be changed and the effect on unavailability seen immediately. Similarly, if an operator wished to evaluate whether the system were suitable for carrying more aggressive modulation schemes, such as 16 QAM, all that is required is to change the required C/I ratio. Finally, the quality of both fixed-frequency channels and variable service bandwidths for agile systems can be examined. With further work, the model may accommodate different channel bandwidths.

RELATIONSHIP OF INGRESS TO OVERALL PERFORMANCE

The overall BER of an upstream data channel is affected by broadband noise, transient electrical interference, the ingress effects and IM products discussed in this paper, and a host of other effects which fall primarily in the realm of failures or mis-adjustments. Our test results to date have indicated that, absent discrete carrier and/or IM in-channel interference, electrical transients are the primary cause of data errors. Often transients are of sufficient amplitude to cause the upstream laser to go into hard clipping. Depending on their duration, various forward error correction techniques may be useful as defensive measures.

By contrast, discrete carriers have no effect unless they fall in a channel and are of sufficient amplitude. In that case, however, the slope of error rate vs level is very steep, so that a discrete carrier is likely to either cause no problem or a total loss of communications. Given that these signals are present for relatively long time periods, their presence is more easily described in terms of availability than BER.

A somewhat surprising result of field tests is that even if an out-of-band ingressing carrier is of sufficient amplitude to cause fairly severe clipping, it has no effect on data integrity unless a harmonic of the interfering carrier should happen to fall in-channel with sufficient amplitude.

When it does occur, the only defense against a long-term interfering carrier is frequency agility. However, the dynamics of ingressing signals may be most useful to product designers in determining optimal algorithms for control of upstream transmitter frequencies.

SUMMARY

The methodology and model described offer a method of quantifying the performance of cable systems with respect to upstream discrete carrier ingress and common path IM products, independently of the effects of electrical transients. The authors feel that is essential because of the different nature of those forms of signal degradation and available countermeasures.

The procedure uses a standard spectrum analyzer, similar to those in common use in cable systems. The post-analysis of the recorded data is performed using a simple (if somewhat large) Excel® spreadsheet. The model allows users to judge the effects of various operating conditions, and can aid operators in choosing operating frequencies and allocating spectrum among services.

The authors expect to further refine the data gathering process in the near future, in order to strike an optimum balance between resolution and spreadsheet size.

ACKNOWLEDGMENTS

The authors wish to thank Lars Stock for his initial work in developing data gathering and analysis tools, Sean Large for further tool refinement and in taking and analyzing the data, and Hewlett-Packard Company for their continued sponsorship of this study.

END NOTES

1.CableLabs, "Two-Way Cable Television System Characterization," internal CableLabs document, 1995.

2.Rex Bullinger and Dave Large, "Status Report: Hewlett Packard Study of Bi-Directional Data Transmission Conditions in Cable Television Systems," *HFC '96 Conference*, IEEE/SCTE, September, 1996.

3.Rex Bullinger and Dave Large, "Downstream Laser Clipping: Field Measurements and Operational Recommendations," 1997 Conference on Emerging Technologies, SCTE, January 1997.

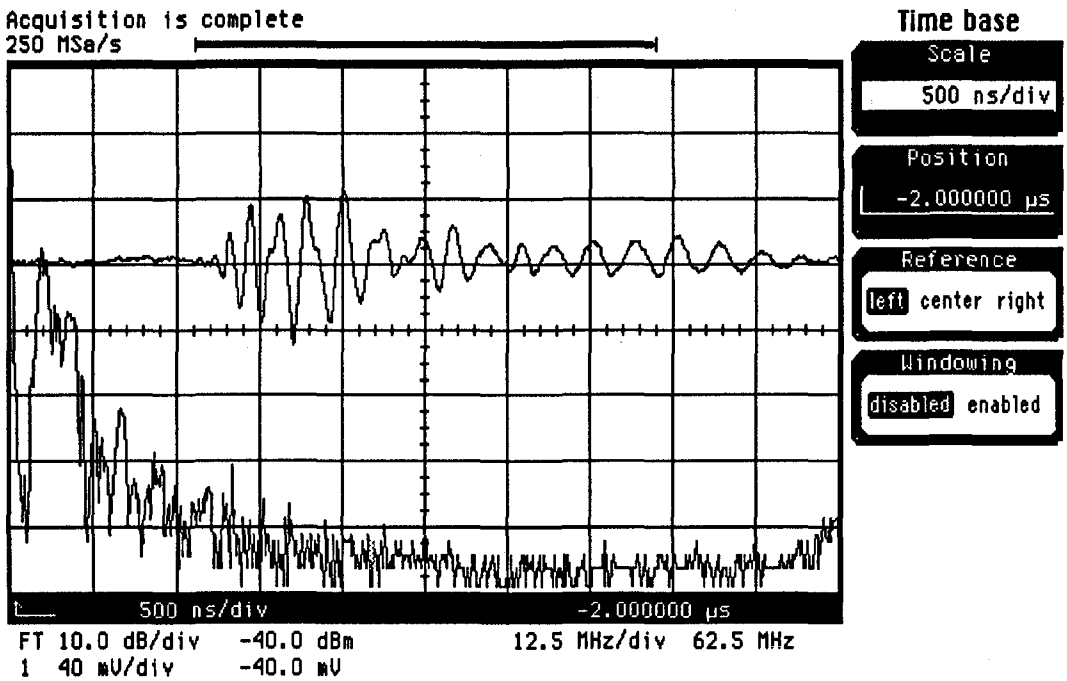


Figure 1: Typical electrical transient and its Fourier transform

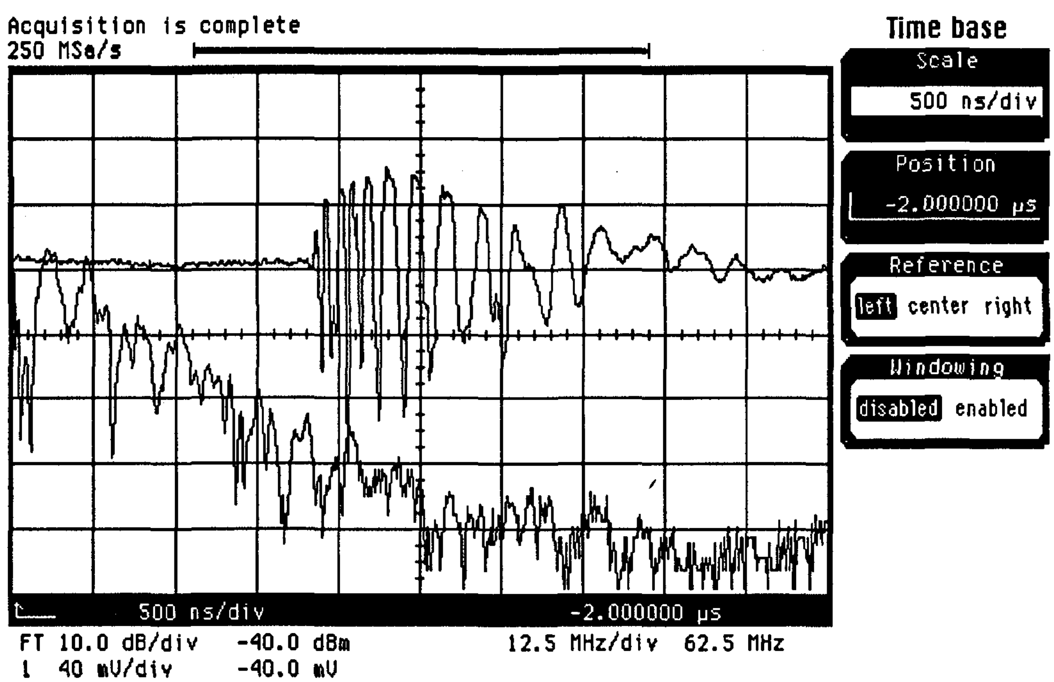


Figure 2: Large transient with clipping

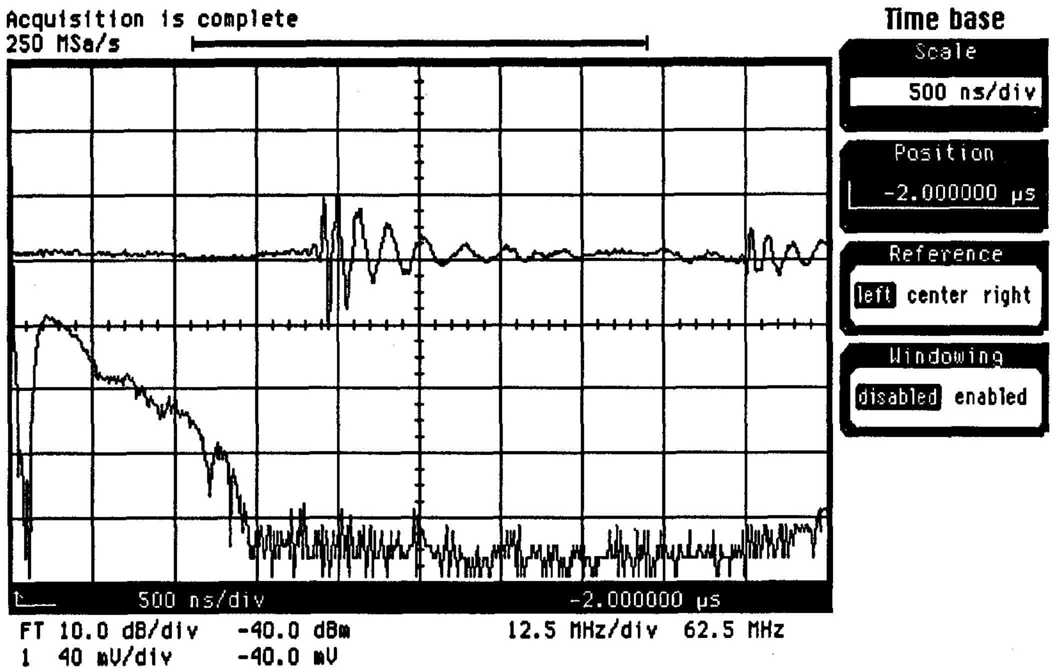


Figure 3: Short transient with smooth frequency signature

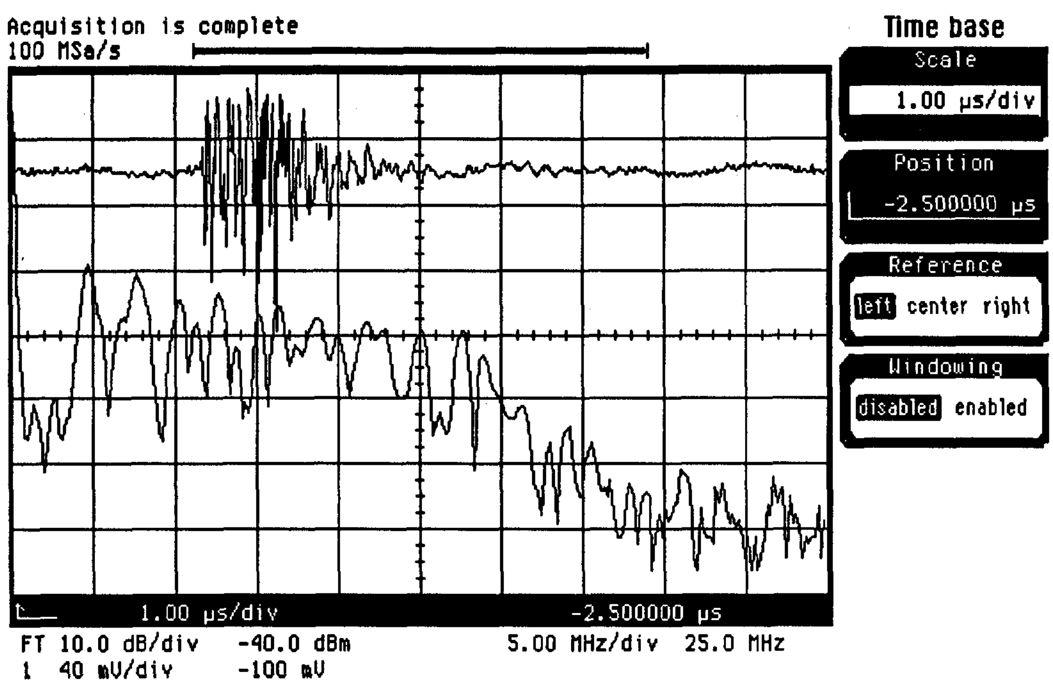


Figure 4: Transient with flat frequency signature

Figure 5: Peak Levels vs Frequency and Time

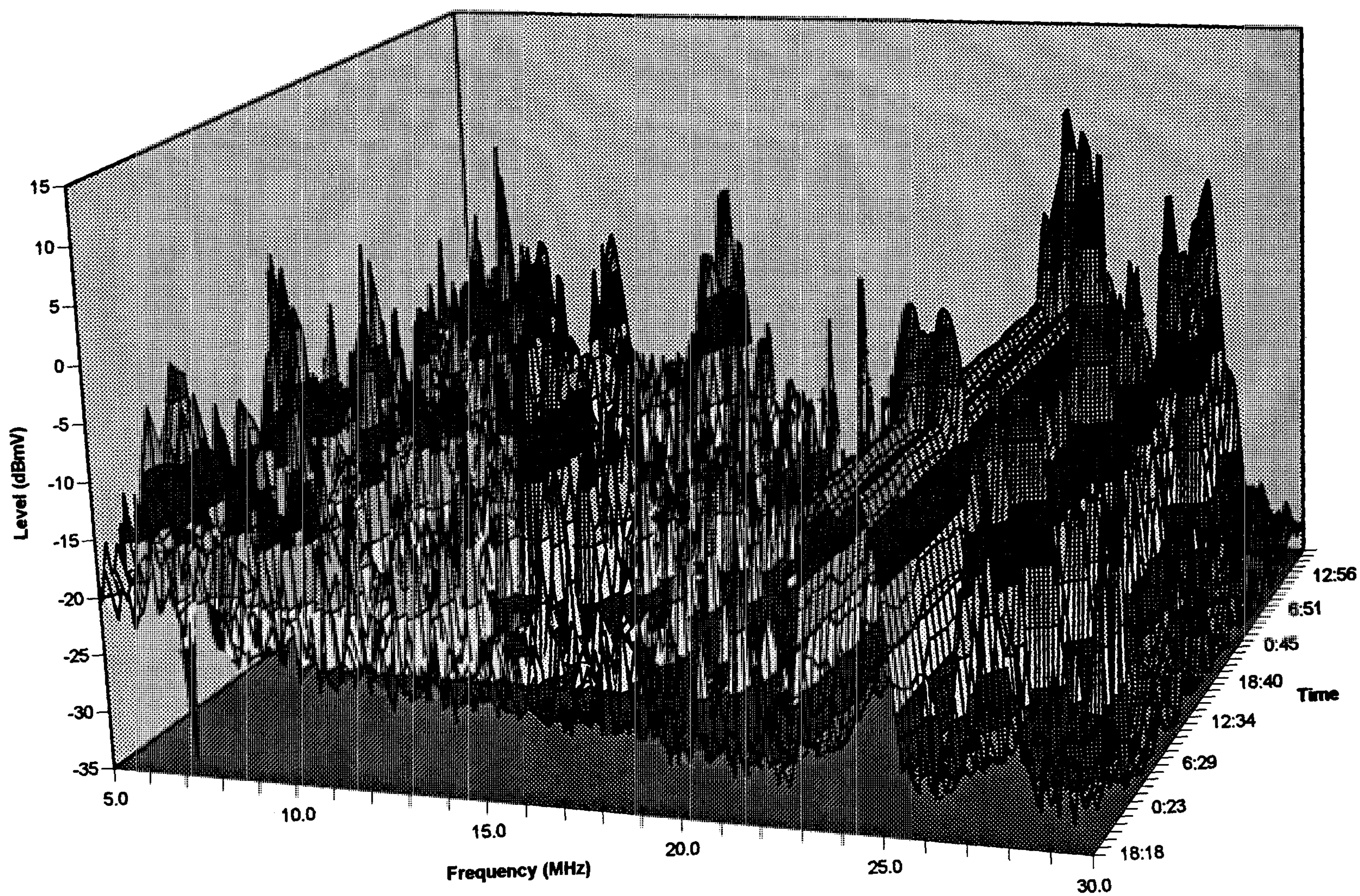


Figure 6: C/I Threshold for Discrete Interfering Carriers

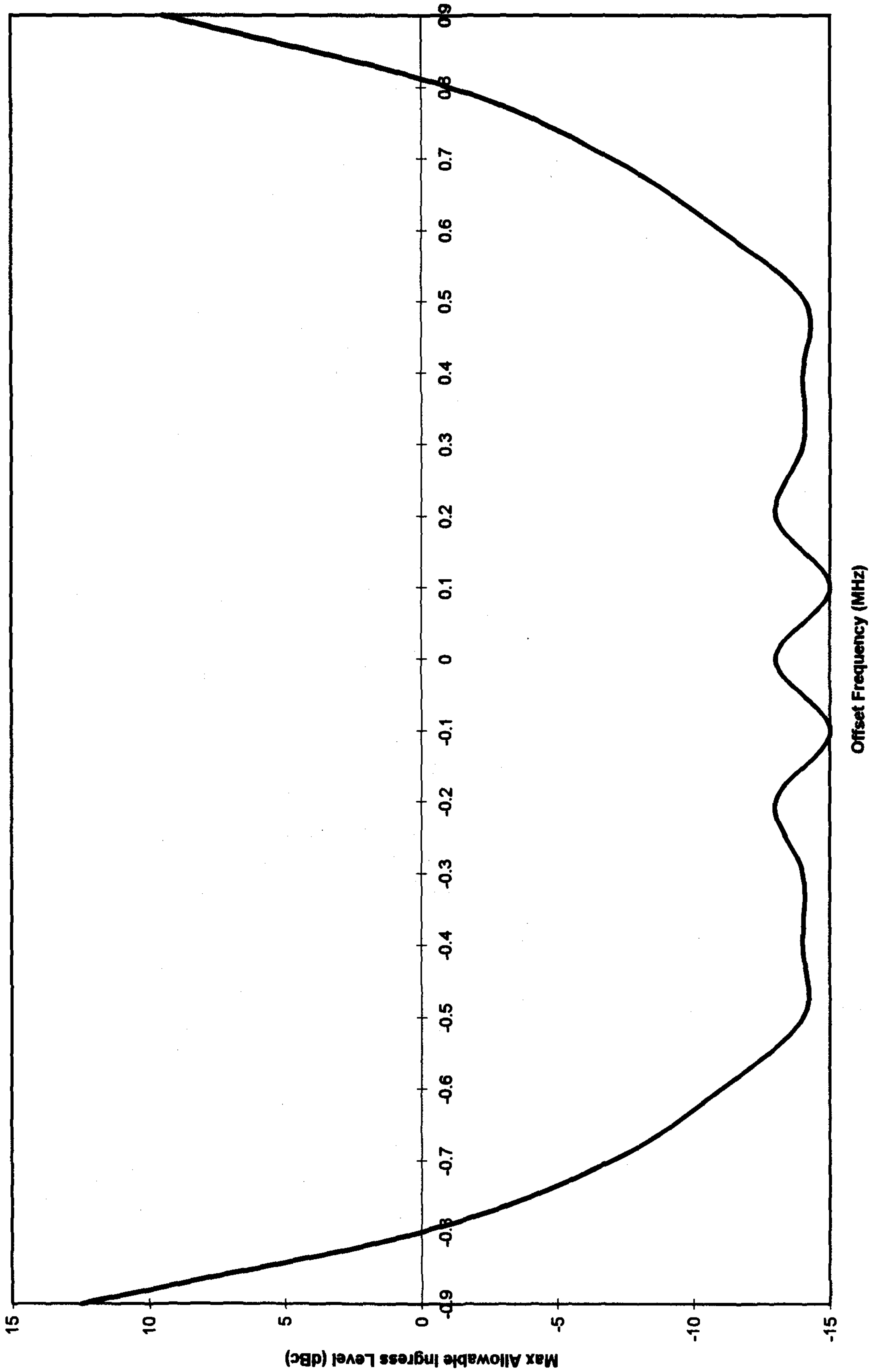


Figure 7: Frequency Unavailability vs Time

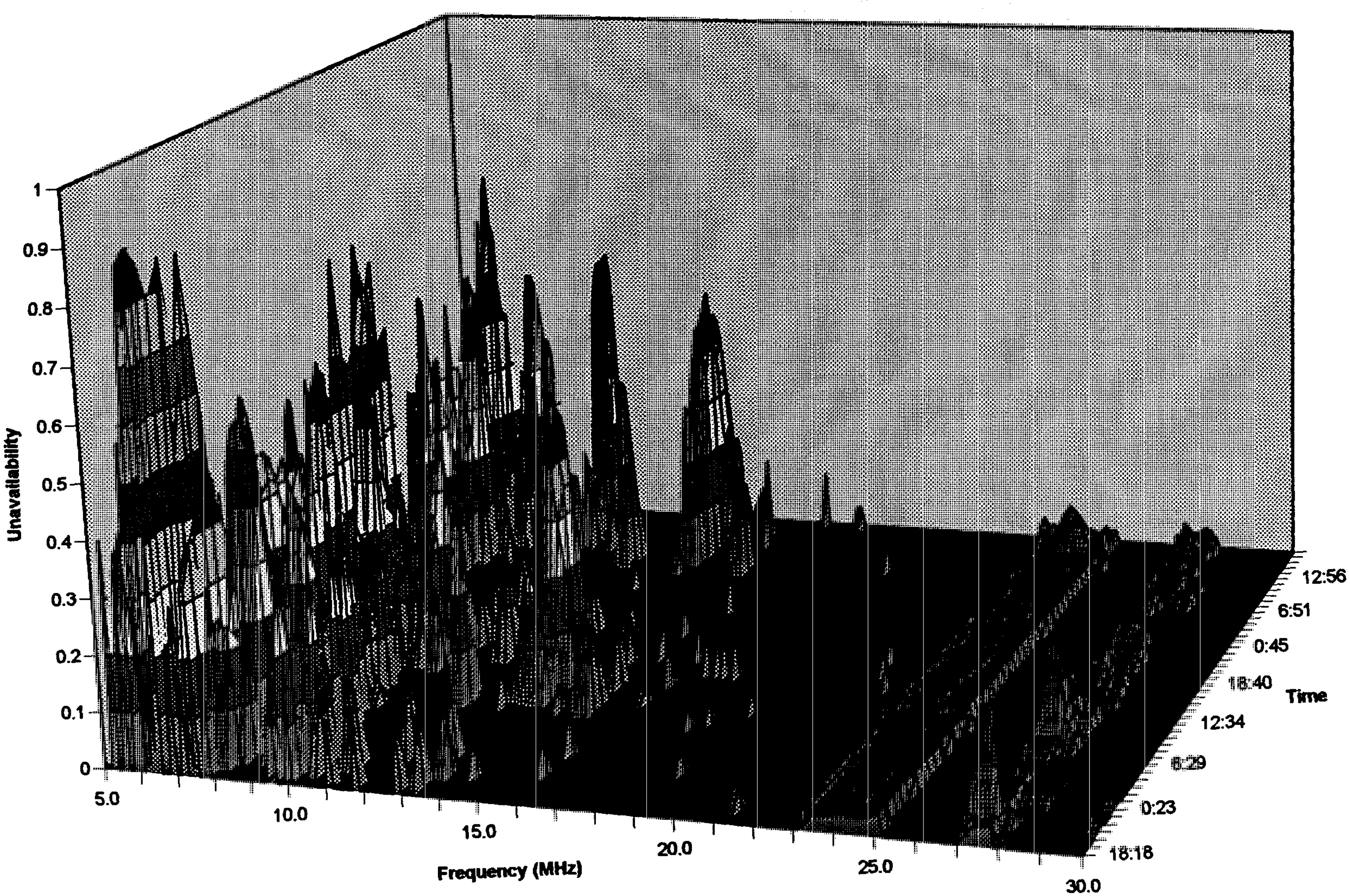
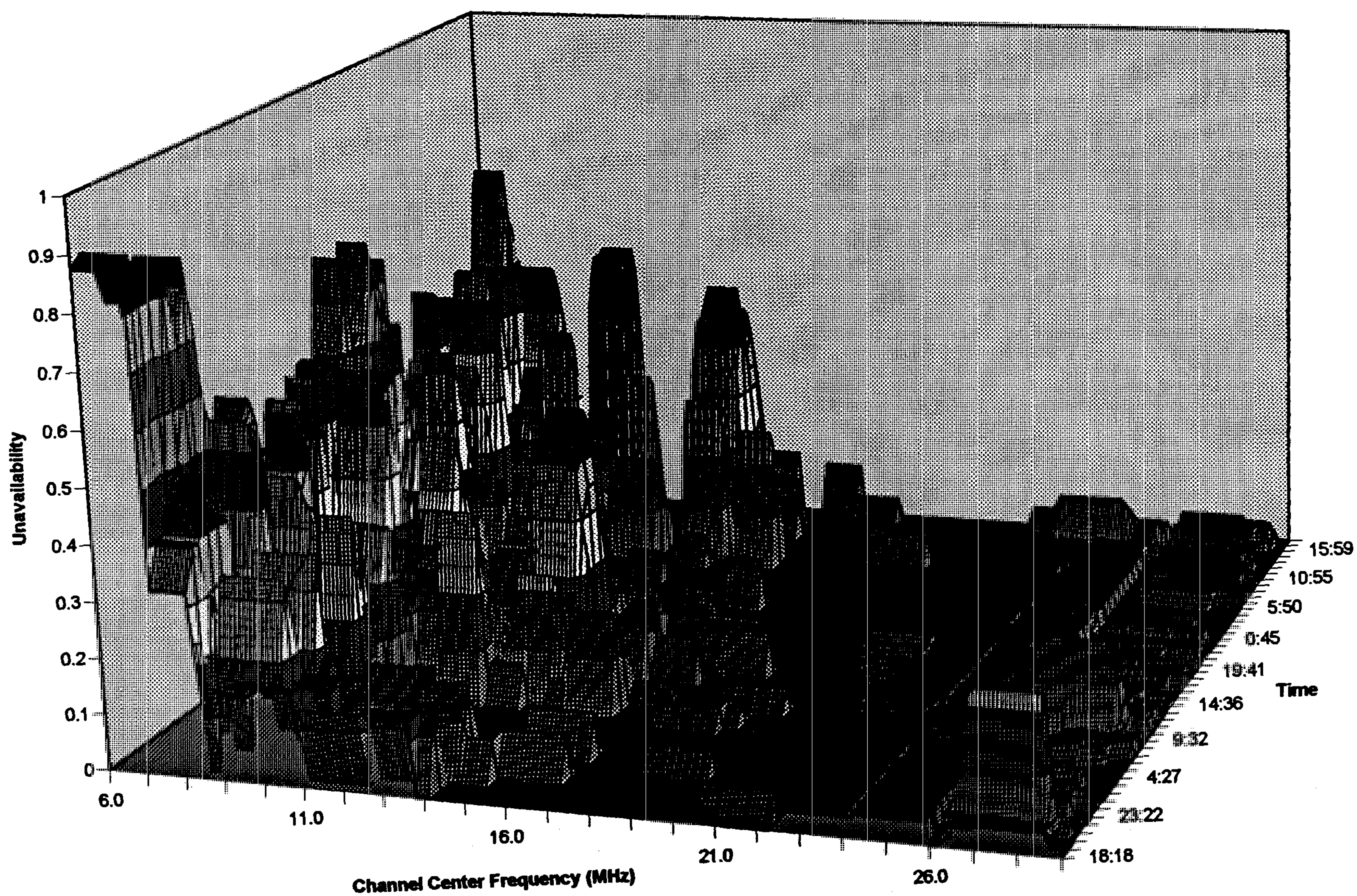


Figure 8: Channel Unavailability vs Frequency and Time



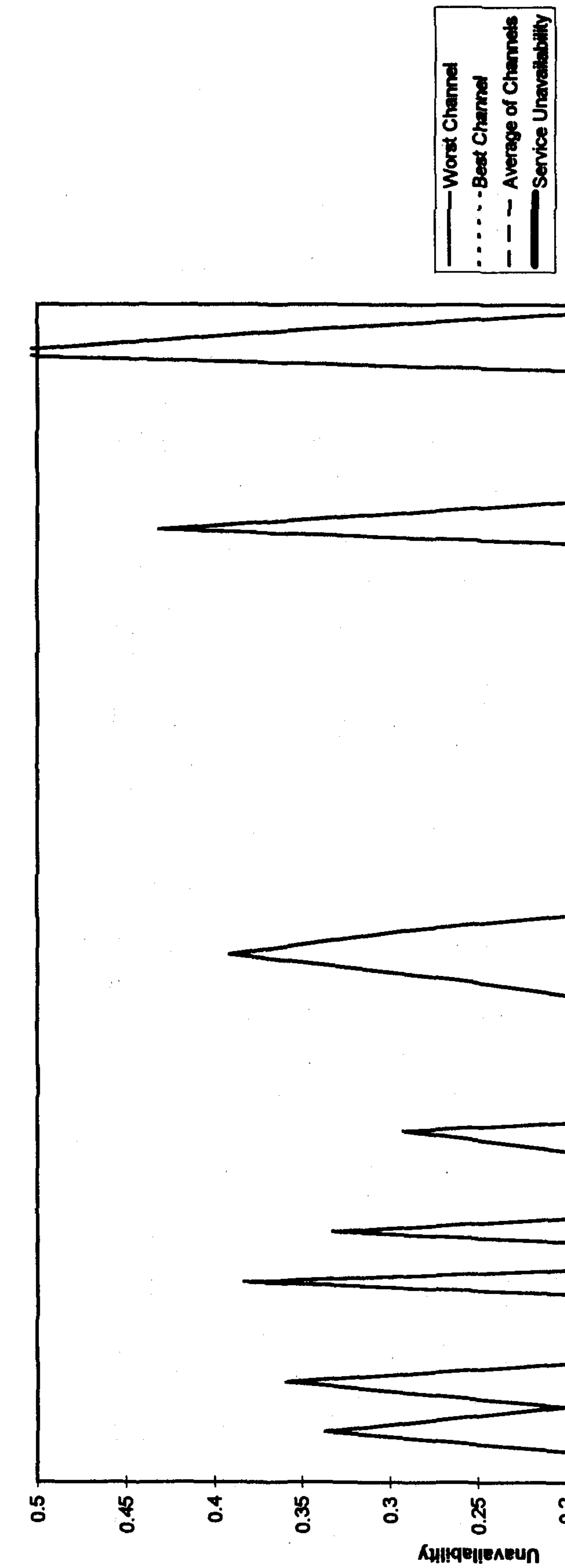


Figure 9: Unavailability of Service Group and its Channels vs Time

Figure 10: INGRESS EVALUATION CALCULATOR: USER INPUT AND SUMMARY RESULTS

Spectrum Analyzer Settings	
Start Frequency	2 MHz
Stop Frequency	42 MHz
Resolution Bandwidth	300 kHz
Video Bandwidth	300 Hz
Reference Level	30 dBmV
Scale Factor	10 dB/div
Measurement Period	60 Minutes
Measurements/sweep	401

Data Collection Parameters	
Start date	1/1/97
Start time	6:18:24pm
Collection period	48 hours
Location	Node 3
Noise Level at RBW	-35 dBmV
Notes	24 MHz modem signal (intermittantly)

Data Service Parameters	
Normal data carrier level	-2.3 dBmV
Required C/I ratio	14.0 dB
Data channel bandwidth	2.0 MHz
Service start frequency	9.0 MHz
Service stop frequency	14.0 MHz
Modem Tuning Resolution	125.0 kHz
Susceptibility Bandwidth	1.8 MHz
Return Spectrum Start Freq	5.0 MHz
Return Spectrum Stop Freq	30.0 MHz

Intermediate Calculated Parameters	
Interfering carrier threshold level	-16.3 dBmV
Minimum channel center frequency	10 MHz
Maximum channel center frequency	13 MHz
Number of usable channel frequencies	31
Equivalent screen position noise value	1500
Possible min channel center frequency	6 MHz
Possible max channel center frequency	29 MHz

Summary Results					
Average service unavailability	0.25%	Average service availability	0.997517	Avail. gain vs average svc chan	21.38265
Average unusable frequencies	2.33%	Average chan unavail (ttl)	6.83%	Average service chan unavail	5.31%

An Evolutionary View of ATM Based Cable Television Residential Architectures

Gaylord A. Hart
Vice President Technology
XEL Communications, Inc.
17101 East Ohio Drive
Aurora, CO 80017-3878
gahart@ix.netcom.com
(303) 369-7000

ABSTRACT

As more services are transported digitally in the HFC network, an increasing number of them will be delivered via ATM (Asynchronous Transfer Mode) technology. At some point, after voice, video, and data are all carried digitally on the HFC network, all network transport will likely be via ATM. ATM, due to its efficient statistical multiplexing, guaranteed Quality of Service parameters, and unique adaptation layers for each type of service, is widely recognized as the best digital transport technology for the integrated delivery of voice, video, and data over a single network. In the evolution of the HFC network, ATM will likely migrate outward from the headend toward the subscriber, and ultimately into the subscriber's residence.

This paper looks at the integrated delivery of ATM based voice, video, and data services directly to the subscriber's home via an HFC network. An evolutionary examination of various residential CATV architectures which may be used to deliver ATM based services within the home and the technical and consumer issues and tradeoffs relating to each of these architectures is provided. The unique requirements imposed by ATM delivery and the unique opportunities afforded by ATM delivery are discussed. The evolution of service offerings made possible by such residential architectures is also examined.

INTRODUCTION

Competition to provide new services, ongoing deregulation of the telecommunications industry, competitive threats to existing services, migration from analog to digital signal delivery, and rapid growth of the internet and data services have dramatically changed the telecommunications landscape. CATV systems must in the near future provide a combination of analog television, compressed digital television, HDTV, telephony, data services, internet access, and numerous other interactive and multimedia services.

At the same time, there is an increasing economic incentive to integrate delivery of these services within the network itself, and ultimately to the subscriber. An integrated delivery system reduces maintenance and installation costs as well as capital equipment expenditures by transporting all services through common channels and equipment. At some point, it will no longer make sense to offer disparate services which share the HFC transport path but which essentially operate independently and use different headend and customer premise equipment, transmission formats, and signaling mechanisms.

For many reasons, true integration of service delivery throughout the network cannot be accomplished until all services are transported digitally. Television signals are still primarily analog on HFC networks, but over the next few years these signals will be displaced increasingly by compressed digital NTSC signals, which are already being carried in some systems. Many other services are already provided digitally or will be in the near future: CD quality audio, internet access, and telephony. Currently, these digital services are all operated independently on the network, sharing only a common RF distribution path in the HFC network. And in most cases these services are delivered without the benefit of modern packet switching, relying instead on more traditional frequency and time division multiplexing methods. However, growing competition, the ongoing migration of analog signals to digital transport, and the increasing need for integrated delivery of voice, video, and data services will ultimately drive HFC network transport toward a packet switched architecture.

Packet switched networks are more suitable for integrated service delivery than circuit switched networks. Packet networks use statistical multiplexing in the transport path and thus can greatly increase transmission bandwidth efficiency. These networks also provide bandwidth on demand to the subscriber without any network re-engineering, which gives the overall network much greater flexibility for providing new services and evolving gracefully. Using modern fast packet technologies, it will be possible to deliver voice, video, and data services directly to the home using a single, common delivery mechanism, possibly even over a single digital data stream.

Carrying voice, video, and data on the same network through the same digital pipes is not an easy task, however. Each of these services has unique transmission requirements. Data transmission tends to be bursty, but transmission delays can usually be tolerated. Television requires a lot of bandwidth to the home, but little or none back to the headend. Telephony requires symmetrical transmission, but will not tolerate long transmission delays. Data is somewhat tolerant of transmission errors because errored packets can be retransmitted, but voice signals do not allow retransmission of errored packets because the corrected audio packet would arrive too late to be placed in the proper speech sequence. A purely digital network delivering all these services must provide the mechanisms and flexibility to meet each of these unique transport requirements.

Of the fast packet technologies available today, only ATM (asynchronous transfer mode) was designed from the beginning to simultaneously support the unique transmission requirements of voice, video, and data. Unlike other packet technologies, ATM supports reserving network resources and defining a Quality of Service (QOS) for a transmission path when the connection is initially established. This is critical for delivering a diverse set of applications with a wide range of transmission requirements. Reserving network resources guarantees adequate bandwidth will be provided by the network to transport the service intended for the link. Defining the QOS for a link further tailors the network transport requirements specifically for the application to be transported. Typical QOS parameters include Cell Error Rate (CER), Cell Loss Ratio (CLR), Cell Transfer Delay (CTD), and Cell Delay Variation (CDV).

As ATM is introduced into the HFC network, it will most likely be used in the network backbone to connect headends or to route signals within headends. As the network evolves, ATM will migrate outward from the headend toward the subscriber. Early in the evolution, services will be offered to end-users via ATM on a service specific basis independently of other services. As more services are delivered to the subscriber via ATM, these will ultimately be integrated throughout the network through common ATM switching, signaling, and transport platforms. Only ATM can provide the flexibility, efficiency, and low cost necessary to integrate the delivery of voice, video, and data services throughout the network. However, for both technical and economic reasons such a network is still a future prospect. ATM will be deployed incrementally in the network, and services will be integrated on an evolutionary basis as hardware becomes available and economic factors allow. Of course, network management becomes much more critical in this system for operations, administration, maintenance, and provisioning, and extensive software support systems must be deployed in parallel with the ATM transport hardware.

ATM TRANSPORT

ATM transport is accomplished with fixed length cells, each cell being 53 bytes long. Of these 53 bytes, 5 are used for the header, and 48 are used for the user information field. The header is strictly used for network transport functions such as destination routing, error detection, and cell delineation. The user information field is used for carrying service specific information and the application payload. Once created, ATM cells are time division multiplexed onto the appropriate transmission channel for routing to their final destination. As cells travel through the network, they may encounter additional switching nodes, at which point they will be switched to the appropriate channel and on toward their final destination.

ATM is a connection-oriented packet transmission protocol. As the name implies, connection-oriented networks require that a logical connection be established between two endpoints before data may be transferred between them. These connections are made via virtual circuits and require setup operations to establish each connection, its routing path, service class, QOS, and bandwidth. The term 'virtual circuit' simply means such a channel appears to the end-user just as a real circuit connecting the two end points. Several virtual circuits (and their ATM cells) can share the same physical channel between two points in the network.

On an HFC network, the physical channels are actually RF channels which are frequency division multiplexed across the upstream and downstream spectrums of the network. Transport may be over fiber or coax. The ATM cells themselves are time division multiplexed onto these RF channels, but with no inherent knowledge that RF transport or a particular modulation method is being used. RF transport simply serves as the physical transport layer in the distribution portion of the network.

The fully integrated ATM network must be capable of delivering all services via virtual connections and with unique content to each home. Such a network will truly allow the subscriber to select services and content on demand. Frequency division multiplexing of RF channels in the network and time division multiplexing of virtual circuits within each RF channel allow flexible coupling of services with RF channels and virtual circuits. For example, it becomes possible to put all services for a single home onto a single RF transport channel, with each service using a different virtual circuit within that RF channel. This allows only a single RF receiver to be used in the subscriber home to recover all services that subscriber takes. When a subscriber is not using any of these services, he need not be allocated any RF spectrum, and this unused spectrum may then be allocated to another subscriber. It may even be possible at some point to use a scaleable bandwidth RF transmission system that occupies only enough spectrum to transmit the services currently being used by an individual subscriber.

Taking the opposite approach, individual RF channels may be dedicated to single services. In this case, for example, a single RF channel could be used to deliver digital audio services, with each active subscriber assigned a separate virtual circuit on that channel to receive the specific music he has selected. This would typically require a separate receiver for each service the subscriber takes. Of course, when a subscriber is not using a service, no virtual circuit or bandwidth need be allocated to that subscriber, thus allowing the unused bandwidth to be allocated where it is needed. For broadcast type music services, several subscribers would simply "listen" to the same virtual circuit, much as they might tune to the same radio station.

In a similar fashion, the return path may be handled by a single residential upstream transmitter supporting several upstream services, each service being assigned its own upstream virtual circuit. Or the return path may be handled with several service specific upstream transmitters, each assigned a unique frequency or time slot on the same frequency. In either case, upstream transport will become increasingly necessary in an integrated network, not only for bi-directional services such as telephony, but for provisioning, status monitoring, maintenance, and customer interaction.

Given the technical complexities of operating an integrated on-demand network, a robust mechanism for spectrum and virtual circuit management will be essential. The way services and virtual circuits are assigned to RF channels in the HFC network, both upstream and down, will also place economic and technical limitations on the ways residential architectures may be realized.

HFC ATM RESIDENTIAL CATV ARCHITECTURES

The following discussion assumes that the HFC network has evolved to the point that one or more services are delivered to (and possibly from) the home via ATM over RF channels. In the downstream direction, the RF signals must at some point be converted to baseband digital signals for recovery of the virtual circuits and the ATM cells they contain, and possibly for further routing within the network or subscriber's home. The opposite is true in the case of upstream signals. For both the up and downstream directions, the residential CATV architecture will determine how and where these RF conversions will take place and how signals will be distributed within the home. This architecture may take several forms, but will likely follow an evolutionary sequence determined by economics, home wiring issues, service offerings, and consumer electronic equipment. Of the myriad number of possible residential architectures, three basic approaches are outlined below to explore potential residential issues and solutions.

Even without ATM, today's residential delivery system already poses some of the most daunting challenges to providing new, reliable, and cost-effective services: powering residential equipment from the network vs. the home, interfacing with existing consumer electronics, and addressing residential wiring issues such as return path ingress. These challenges will only increase as the HFC network migrates toward ATM and as service offerings and consumer electronics become technically more complex. Open interface and transmission standards for both network equipment and consumer electronics will be essential to the success of any integrated service delivery system.

Service Specific ATM Interface Architecture

Initially, ATM will make its way to the residence over a single service. The likely first candidate is an ATM based cable modem for internet access. Since internet access is not a lifeline service (at least not today), powering from the network is not an issue, and this device will be powered by 120 VAC as other residential electronic devices. Similarly, the cable modem will provide a standard RF interface to the cable system (via an F connector) and another standard data interface to the computer (most likely ethernet or ATM). Because the modem bridges together a network and network device that traditionally in the past

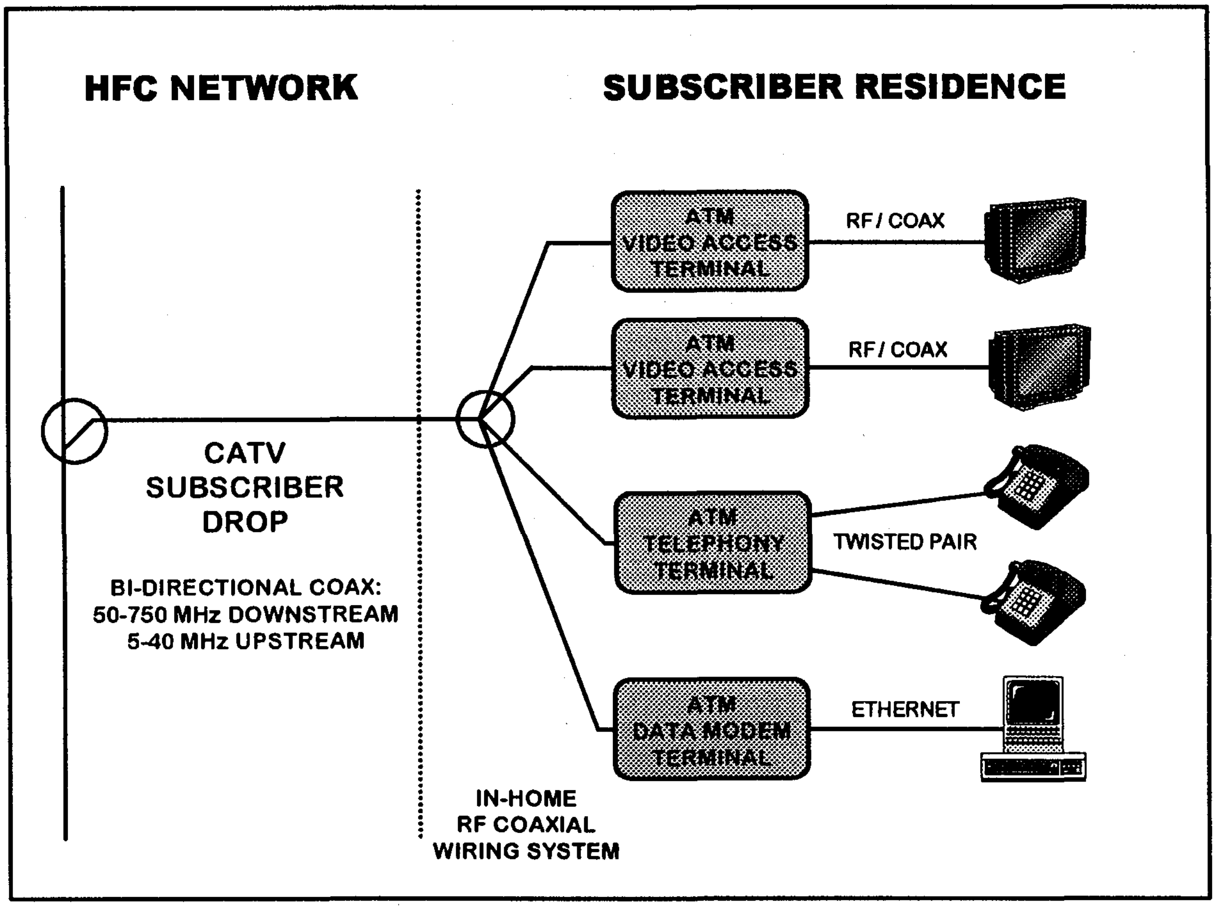


Figure 1 -- Service Specific ATM Architecture

had no interaction, the modem will likely be a standalone device, though versions can easily be made which plug into the computer.

There is also a desire to keep installation costs to a minimum, so it is very undesirable to make significant changes to the existing home wiring. Since many homes are already wired for CATV distribution, but very few for ethernet, signal distribution in the residence for the modem service will be via RF over the existing home CATV wiring or extensions of it. This means the cable modem will typically be located next to the computer it serves, with only a short data interface cable running between the modem and computer.

As the evolution of the HFC network proceeds, more services will begin to be delivered to the subscriber via ATM. A second likely service for ATM delivery to the home is compressed digital NTSC, which can take advantage of the full switching capabilities of ATM to deliver true video on demand. Other services will follow, including basic telephony, and this will probably occur on a service by service basis. Of course, each of these new services will impose new requirements on the residential architecture and wiring. For example, telephony will require that powering issues be resolved to ensure that phone service remains available even during power failures at the residence. Figure 1 depicts a residential architecture based upon specific interface devices for each service.

Because ATM deployment will be incremental and because not all subscribers take all services, it is likely that a specific RF / ATM interface device (sort of an ATM set-top converter) will initially be required for each new service delivered via ATM. Of course, the logical evolution would be for these interface functions to migrate into the consumer electronic devices themselves, and this may very well happen if issues such as open standards, signal security, and feature enhancement can be worked out between the CATV and consumer electronics industries. But there will be a long period of time in which it will be necessary to support legacy consumer electronic products with outboard interface devices, just as we have for years with RF set-top converters. If open standards can be developed for such ATM interface devices, it should be possible for consumers to buy them at the local electronics store.

Regardless of the service being supplied, each ATM interface device must maintain those virtual circuits (if more than one is required) associated with its particular service at this particular residence. Since each service requires a unique ATM access device, this architecture can be considered to be distributed. For some services, such as television, it may be necessary to have an interface device for each consumer device connected to the network (much as is the case today with set top converters). For other services, such as telephony, a single ATM interface device may serve several consumer devices in the home. Because this residential architecture is distributed, an RF receiver will be required in each interface device in the home, and in the case where return transmission is also required, each device must have an upstream transmitter.

ATM Residential Gateway Architecture

Another approach to residential access is the residential gateway, which is shown pictorially in Figure 2. The residential gateway is an integrated access device providing a single interface to the HFC network and multiple service interfaces to the home. In all likelihood, the gateway will consist of a small chassis which accepts plug-in modules. Such an approach allows modules to be installed only when actually needed to provide service and supports ongoing upgrades and evolution of the platform as a whole. The residential gateway architecture can easily accommodate existing consumer electronic devices through standard interfaces but will also support future consumer devices and their interfaces through new service interface modules.

The gateway may be owned by the service provider or by the subscriber, or by both. In this last case, the service provider may own the chassis and common equipment which is critical for service delivery, signal security, and other functions. The subscriber might then provide plug-ins for specific services he subscribes to. As such, the residential gateway could function as a standardized network interface for the home while providing a demarcation point between the network and customer premise equipment. One great advantage of the residential gateway is that it can provide operator controlled access to the HFC network, which eliminates system ingress caused by residential wiring systems or consumer electronic equipment.

Using a non-distributed architecture, the residential gateway also offers the advantage of allowing common equipment to be used to support several services. For example, since all services are fed from this single chassis, a single power supply can be used for all service interface devices in the box. Similarly, a single RF transmitter can be used for all services requiring upstream transport. The residential gateway also supports migration of services to ATM on a service by service basis simply by plugging new cards into the residential gateway. As pointed out earlier, ATM deployment will be incremental. The residential gateway should provide a lower cost solution to delivering ATM services when compared to using multiple

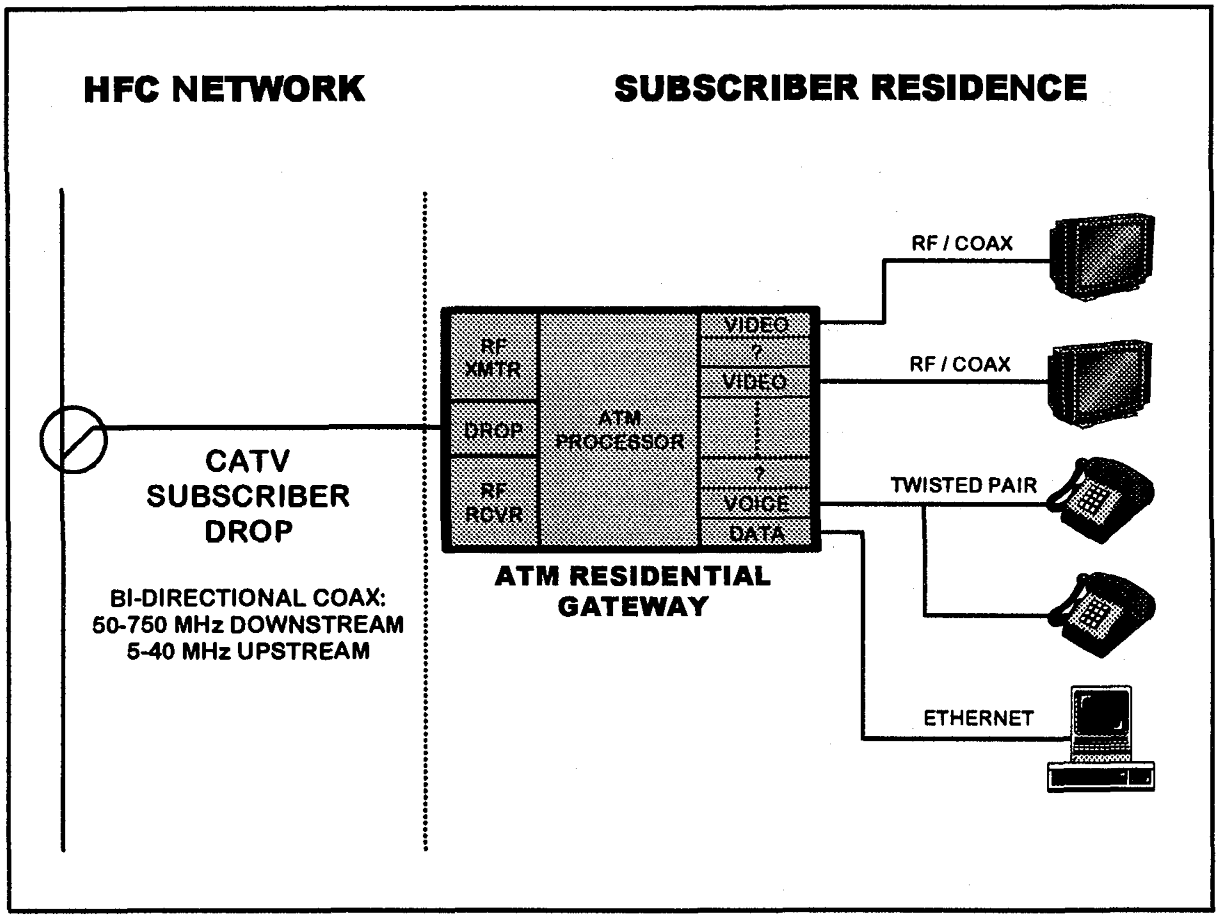


Figure 2 -- ATM Residential Gateway Architecture

service specific access devices, but more careful planning must go into the design of the gateway to ensure flexibility as services and the residential architecture evolve.

From an evolutionary perspective, the residential gateway addresses several of the concerns already discussed above under service specific interface devices, but without using a distributed architecture. Several architectures exist for the residential gateway itself. In the example shown here, a single RF receiver interfaces with the subscriber drop, then demodulates the high speed RF data stream. This stream is then fed to an ATM processor which maintains each of the virtual circuits associated with this particular residence. Several virtual circuits are present in this one data stream. Some of these virtual circuits may be intended for other residences, and these ATM cells are discarded by the local ATM processor. For those virtual circuits intended for this home, the ATM processor delineates which service is associated with each virtual circuit and forwards the ATM cells for each virtual circuit to the correct service interface module.

In all likelihood, these virtual circuits will be a combination of permanent virtual circuits (used for continuously connected services) and switched virtual circuits (used for sporadic or demand based services). Some of these virtual circuits will not be associated with services, but with performing other local functions such as status monitoring and provisioning of the gateway as a whole. Each service interface module is responsible for converting its virtual circuits to electrical signals which can be distributed independently through the residence and then used directly by the subscriber's TV, telephone, computer, etc. In many cases, the existing home wiring can be used for this final distribution.

The reverse path operates in much the same way. Each service interface module formats its return data as ATM cells then forwards the virtual circuit containing these cells to the ATM processor. The ATM processor then time division multiplexes all these virtual circuits onto a single data stream which is fed to the upstream RF transmitter. Again, the upstream path will likely contain both permanent and switched virtual circuits, and one or more of these will be dedicated for maintenance and provisioning of the gateway itself. Not all services will require an upstream connection.

ATM Residential LAN Architecture

One common component in the residential gateway, the ATM processor, may at some point also begin to assume significant additional functions. Since all virtual circuits for all services in the home flow through the ATM processor, the processor may assume switching functions local to the residence itself. Local virtual circuits between consumer devices could be supported, and this would allow consumer electronic devices within the home to communicate with each other and exchange signals, services, and information. For instance, a VCR in one room might provide video via the gateway to a television set in another room. Or computers within the home could exchange files or other information through the gateway.

An early version of such a residential gateway LAN could evolve from the architecture discussed above and could support existing consumer electronic devices over their standard interfaces. However, because most of today's consumer products were not meant to intelligently interact with one another, only limited capabilities can be realized with this approach. Any intelligence in such a LAN would reside primarily in the gateway itself and would be transparent to the devices connected to it (with the exception of home computers, of course).

As service offerings, consumer electronics, and the network further evolve, increased functionality and interoperability will spread throughout the residential architecture and its consumer devices. In the long view, the residential ATM interfaces will likely migrate from the residential gateway into the application device being served, whether it be a computer, telephone, TV set, thermostat, or garage door opener. Once this occurs, the residential gateway will serve as both a local ATM switch for the home LAN and as the HFC ATM network gateway.

Figure 3 shows how such an ATM LAN may serve the residence. In this example, all home devices are connected to the gateway LAN via home runs on shielded twisted pair running 100 Mb/s ethernet. All devices maintain bi-directional communications with the LAN gateway both for service content and for configuration information. Devices may either communicate with each other or the HFC network, but all communications must flow through and be switched by the gateway LAN. All communications in this example are still via ATM, but through cells transported over ethernet. Other residential architectures could be derived which provide similar functionality. For instance, the home wiring could be structured around a bus architecture, where the home devices could communicate directly with each other.

The residential LAN allows great simplification of home wiring. Since all devices in the home would be served off an identical digital interface, only one home wiring system may be required for all residential devices. One could just as

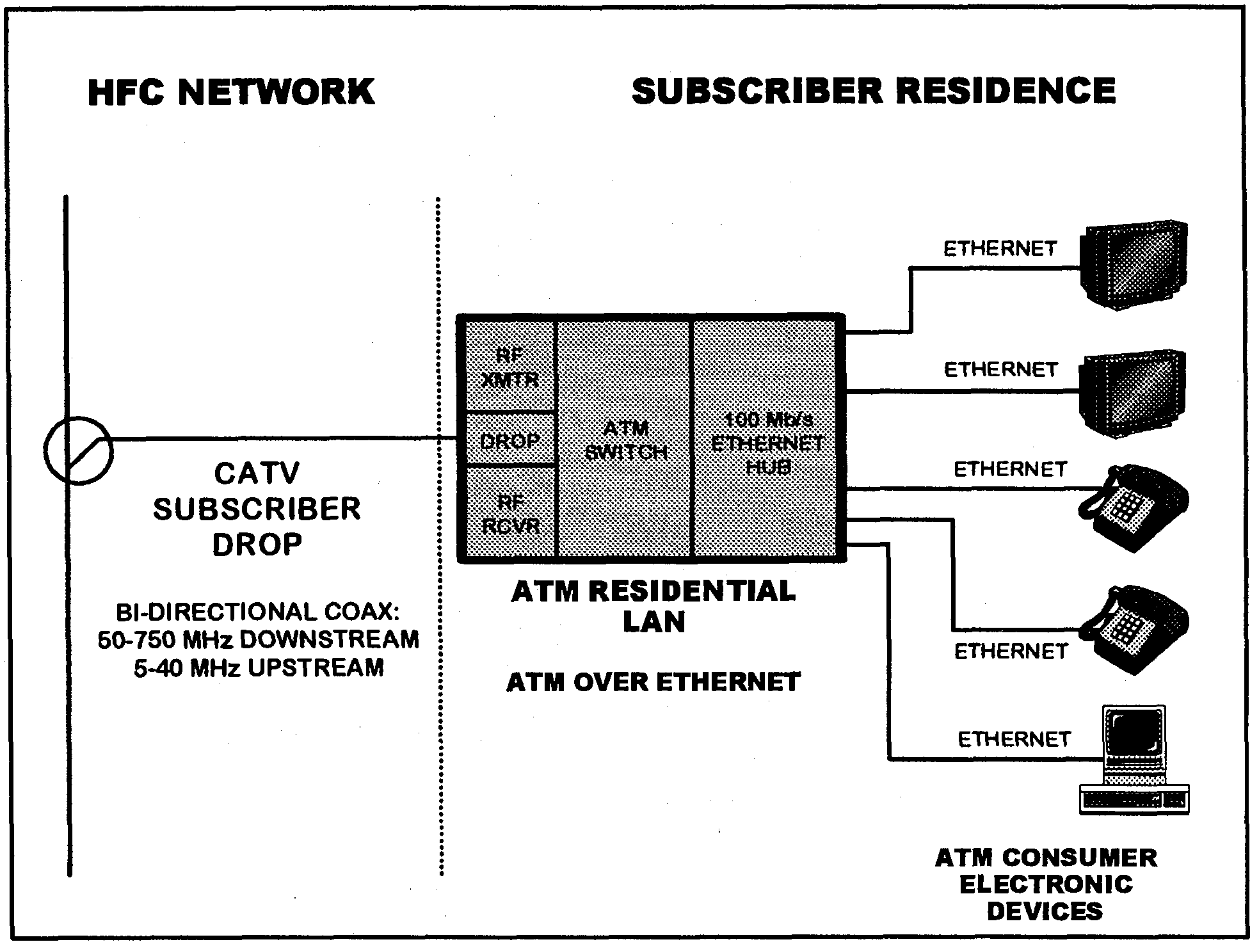


Figure 3 -- ATM Residential LAN Architecture

easily plug a computer or a television into the same outlet in any room in the house. In fact, a single connector could be used for both the LAN digital interface and for powering the local device.

Many transport technologies exist for creating LAN's: twisted pair, fiber, coax, etc. And many protocols exist for creating LAN's: FDDI, ethernet, token ring, etc. The ATM LAN discussed here is too far into the future to predict which residential transport technology or protocol will prevail or even exist when the time comes to build such a LAN. But it is obvious that such an ATM LAN cannot exist without rigorously defined open standards for hardware vendors to build their products around.

SUMMARY AND CONCLUSIONS

Analog signals will continue to disappear from HFC networks as digital delivery of traditional services becomes more common and as new digital services are offered. Digital signals not only offer better service quality, but can offer other benefits such as better spectrum utilization. As the migration toward an all digital network continues, it will eventually make sense to transport these digital signals over an integrated delivery platform. ATM is the best technology for integrating and delivering voice, video, and data services over a single digital network. Operational, economic, and competitive considerations will drive the deployment of ATM in HFC networks. Initially, ATM will be used in HFC systems for backbone transport. But as services, technology, and the competitive environment evolve, ATM transport will migrate over the HFC network toward the subscribers' home, where ATM's full capability will be realized as new service offerings and applications are made possible. It is in the subscriber's home, however, that some of the most daunting challenges exist for fully digital, integrated delivery of services directly to the subscriber. Issues such as powering, legacy consumer products, home wiring, and new applications must be addressed if the full capabilities of an integrated network are to be realized.

New services will also be made possible or economical by ATM technology, and these too will be integrated into the network as it evolves. ATM's switching capability opens up many possibilities for economically delivering local ad-inserts or video-on-demand services from a single, centralized video file server center. As the ATM switching capability extends to individual subscribers, TV ads may literally be targeted to individual homes--in fact, you may be able to choose your own ads! Some day all television viewing may be video-on-demand operating over virtual channels, but this will be transparent to the subscriber, who only knows that he can watch whatever he wants whenever he wants.

REFERENCES

Black, Uyless, *Data Link Protocols*, PTR Prentice Hall, Englewood Cliffs, NJ, 1993.

Black, Uyless, *Emerging Communications Technologies*, PTR Prentice Hall, Englewood Cliffs, NJ, 1994.

Hart, Gaylord, "Architectural Considerations for Overlaying Fast-Packet Networks on Hybrid Fiber/Coax Systems," *NCTA 1996 Technical Papers: Proceedings of the 45th Annual Convention and International Exposition of the National Cable Television Association*, pp. 360-374, NCTA, Washington, D.C., 1996.

Hart, Gaylord, "Deploying ATM in Advanced HFC Networks," Communications Technology, December 1996, Vol. 13, No. 10, pp. 34-51.

Kessler, Gary C., *ISDN: Concepts, Facilities, and Services*, 2nd Edition, McGraw-Hill, Inc., New York, NY, 1993.

Onvural, Raif O., *Asynchronous Transfer Mode Networks: Performance Issues*, 2nd edition, Artech House, Norwood, MA, 1995.

Spohn, Darren L., *Data Network Design*, McGraw-Hill, Inc., New York, NY, 1993.

Stallings, William, *ISDN and Broadband ISDN with Frame Relay and ATM*, 3rd Edition, Prentice Hall, Englewood Cliffs, NJ, 1995.

AN OPEN SPECIFICATION FOR HYBRID FIBER/COAX OUTSIDE PLANT STATUS MONITORING EQUIPMENT

Gordon L. Bechtel, Jeffrey A. Masucci and L. Kathy Bechtel
Stout Technologies, Inc.

and

Pamela Anderson
CableLabs

Abstract

A new, open specification for Hybrid Fiber/Coax (HFC) based outside plant (OSP) status monitoring equipment is becoming available. It details the physical and data link layer protocols, the core message set and the transponder electromechanical interface to be used by HFC status monitoring equipment. It has been developed by CableLabs and its member companies through a cooperative process with the vendor community.

The status monitoring specification provides the structure necessary for vendor interoperability, while allowing vendors 'room' for innovation and product differentiation. Communication is based on two-way burst packet FSK (Frequency Shift Keying) transmission and employs a combination of poll-response and contention based multiple access schemes. The message set provides the means and the structure for communicating information to and from outside plant status monitoring equipment. It is simple, optimized for speed, and extensible. The electromechanical specification provides versatile interconnection between the transponder and the monitored equipment.

INTRODUCTION

A new, open standard for Hybrid Fiber/Coax (HFC) based outside plant status monitoring equipment is becoming available. The specification is a key step toward implementing competitive and reliable HFC networks needed by Cable Multiple Systems Operators (MSOs) for new interactive services, such as high-speed data and telephony. It details the physical and data link layer protocols, the core message set and the transponder electromechanical interface to be used by HFC status monitoring equipment. The standard has been developed by CableLabs and its member companies with strong cooperation from the vendor community.

The CableLabs specification for outside plant status monitoring equipment provides a mechanism for low-cost, interoperable network monitoring that satisfies both the network operator and the vendor communities. It is an essential step towards the delivery of the highly reliable, competitive services essential for next generation MSO networks. Because the status monitoring effort is so important to the operator community, several important goals have been set for it.

STATUS MONITORING GOALS

The goal of the status monitoring effort is to define a standard for HFC network status monitoring equipment that achieves interoperability between different vendors' equipment, satisfies the functional requirements of the network operator and is inexpensive.

Figure 1 shows a generic layout of status monitoring equipment for a HFC network. There are two important parts to the status monitoring equipment: the headend OSP EMS (Element Management System) and the OSP transponders. The transponders monitor various network parameters such as voltage and current, create alarms, when appropriate, and communicate this information to the headend OSP EMS equipment. The OSP EMS gathers this information and relays it onto higher level network management entities.

Technical Functionality

The technical functionality required by the status monitoring system strikes straight at the issue of cost-effectively operating a highly reliable HFC network. Several high-level requirements are needed to achieve the desired levels of reliability and operational savings.

One-to-Many Forward Channel: The status monitoring system needs to operate on the broadcast based HFC forward (headend to user) plant.

Noisy, Many-to-One Return Channel: One of the most difficult requirements to achieve is operation on the many-to-one return channel characteristic of HFC plants. This portion of the HFC plant is very noisy, forcing the modulation technique to be robust and the data rates to be lower. Also, a mechanism for allowing multiple units to communicate with a single headend is re-

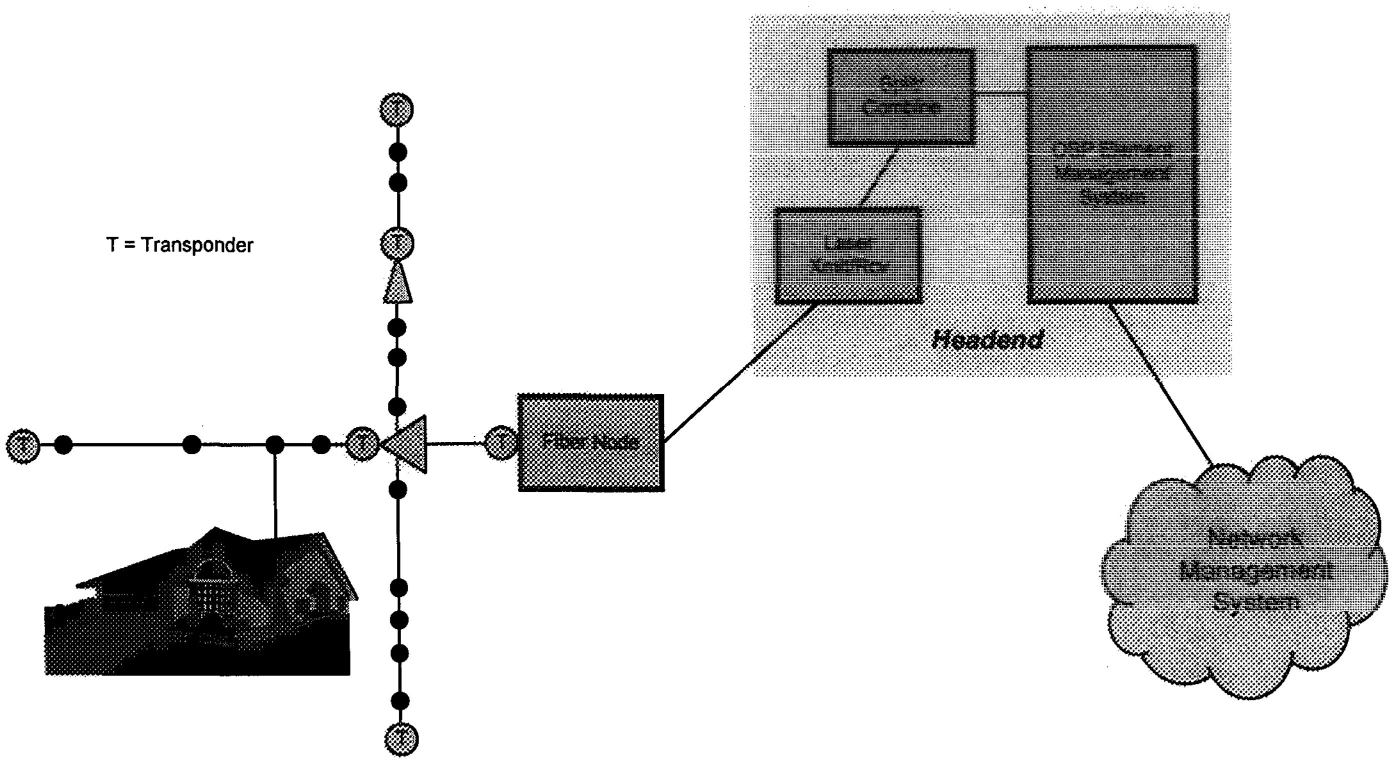


Figure 1: A Generic Layout of a Status Monitoring Equipment on a HFC Network

quired.

Quick Alarm Reporting: When an fault occurs in the network, the status monitoring network needs to present an alarm to the headend within seconds of the fault's occurrence.

Unresponsive Units: When an individual transponder fails, the status monitoring network needs to detect this within a few minutes of the failure.

Catastrophic Failures: A catastrophic event, such as a power failure or a lightning hit, that effects a large portion of the network, may cause multiple alarms to occur virtually simultaneously. The status monitoring system must recover quickly and robustly from such events.

Power Level Calibration: In some cases, the network operator would like to have a transponder provide a signal with an extremely accurate power level for use in calibrating amplifiers and fiber nodes on the HFC plant.

Room for Innovation: From both the vendor and the network operator points-of-view, the standard must not be so rigid as to eliminate future innovation in the status monitoring system. It is important to vendors to be able to differentiate themselves from their competition. It is important to operators in order to continually improve their operations and maintain their competitive edge.

Some Cake and a Fork, Too

Network operators obviously want to have their cake and eat it, too. The ideal is to write a standard that helps achieve an inexpensive transponder with timely availability and interoperability with other vendors' transponders.

A standard is an important step in achieving these goals. A properly written specification can obviously achieve the goal of interoperability between different vendors' equipment. This is important to network operators because it allows them to purchase equipment from the vendor of their choice, as long as that vendor adheres to the standard. Also a carefully crafted and well reasoned standard can help reduce expense by providing clear incentives for mass production of key parts, subassemblies and even entire transponders. The standard must be written carefully, though, with focused attention paid to vendor commentary on cost. Finally, in several ways, a standard gets in the way of achieving timely introduction of status monitoring systems. A standard takes time to complete and this time could be spent in bringing proprietary status monitoring solutions to market. However a standard is essential in achieving the first two goals. Therefore the standard writing process must progress quickly and efficiently.

THE PROCESS

The goal of the Outside Plant Management specification effort is to quickly achieve a workable standard which would facilitate cost-effective implementation by vendors. CableLabs and Stout adhered to four basic philosophies to achieve this goal.

Don't Reinvent the Wheel

There was no need to begin the development of a standard from scratch. The vendor and operator community have too much experience to begin at ground zero. Therefore, the process began by requesting from the vendor community information on their existing and proposed Physical, Data Link Layer and Message Set specifications.

Table 1: OSP Management Standardization Milestones			
Document	**RFI**	**Draft**	**Interim**
Electromechanical Specification	December 9, 1996	—	—
Physical Layer Specification	August 12, 1996	October 28, 1996	December 9, 1996
Data Link Layer Specification	August 12, 1996	September 25, 1996	November 14, 1996
Message Set Specification	August 12, 1996	December 9, 1996	—

A similar approach was used to get started on the Electromechanical specification.

A working group composed of CableLabs staff, CableLabs member companies' staff and Stout Technologies' staff was organized to compare and analyze vendor responses to assess their workability relative to the CableLabs member companies' operating plans and experience. Individual specifications were chosen to meet the needs of network operators while at the same time maximizing the match with products already offered by the vendor community. This approach should satisfy the network operator, while minimizing the impact on the vendor community.

Move Quickly

A long standardization process was one of the primary concerns of the network operating companies. They needed a status monitoring standard right away in order to remain competitive. Status monitoring is a key part of deploying new services in a reliable manner. It has been estimated by TCI that 5 to 10 minutes can be erased from the Mean Time to Repair (MTTR) by installing a status monitoring system.[1] It is also important to increase the overall reliability of the HFC network and reduce operating expense. Coupling the strategic importance of status monitoring with the short term pressure of the competitive environment results in a resolute requirement for quickly turning around a status monitoring specification.

The strategy for achieving the goal of quick turn around was multifaceted. Not only does it require dedication from the people and companies involved, but it also requires a different process than taken in a traditional standards setting arena. Specifically a proactive posture is required from the group with the most interest in having a specification completed: CableLabs and its member companies. The key milestones achieved thus far are shown in Table 1.

Proactive, Interactive Process

CableLabs has maintained an aggressive, proactive role throughout the status monitoring standardization effort. First, an energetic schedule was formulated. Milestones for delivering draft and interim versions of individual documents were set and have, to a large part, been met. Of course, when the aggressiveness of the schedule conflicted with achieving a specification which was acceptable to both the vendor and network operator community, the schedule was relaxed in order to achieve a more acceptable specification.

The second and most important part of the process is the proactive role taken to produce draft, interim and final specifications. As Figure 2 diagrams, rather than wait for standard contributions from interested parties, CableLabs has taken the information available from the vendor community, assessed it relative to the needs and

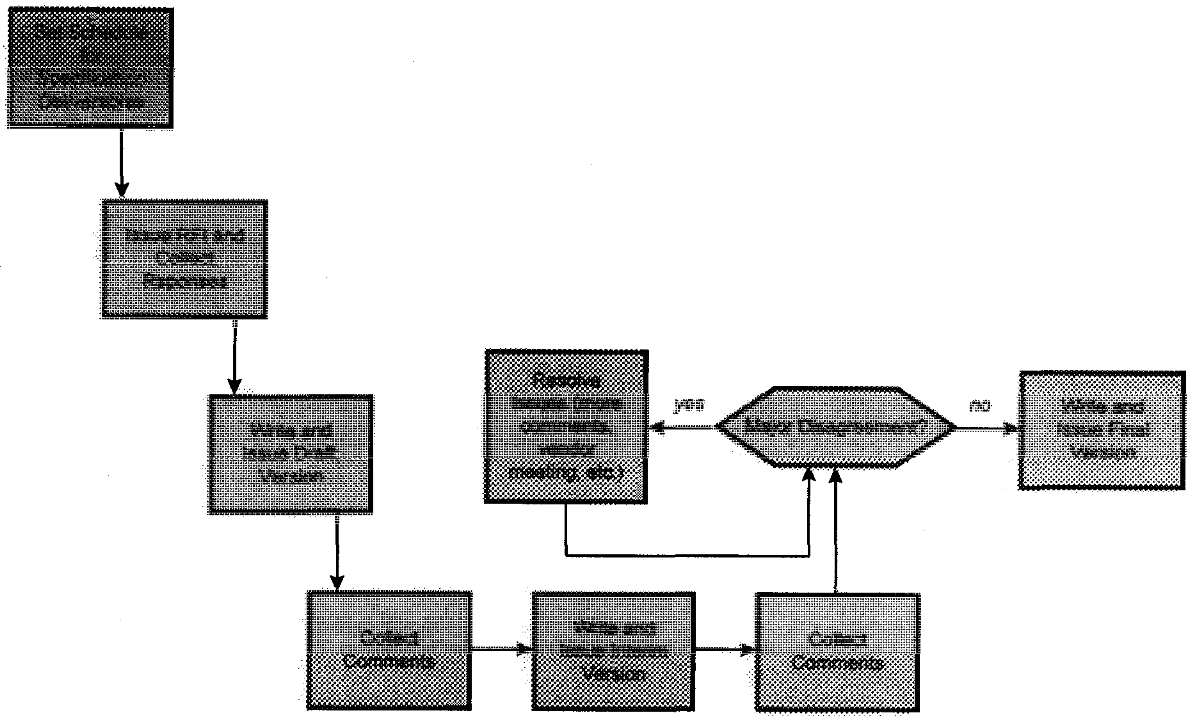


Figure 2: The Specification Preparation Process

experience of the network operators and written appropriate draft specifications. These drafts were then circulated for comment, the collected comments assessed and, if suitable, built in to the interim version of the specification. Differences regarding the interim version were then resolved through a combination of further commentary and general vendor community meetings. A proactive approach has helped to achieve the quality and the speed desired for the status monitoring standard.

Facilitate, Not Dictate

In order to serve the interests of all involved (MSOs and vendors) the best approach to the status monitoring specification development by CableLabs and its chosen contractor, Stout Technologies, was facilitation, not dictation. A dictated specification would not be an acceptable solution for all parties, would never gain widespread industry support and would not be useful. The only process leading to a widely supported specification is an inclusive one. All interested parties must be involved and proper facilitation is required to make sure that that involvement is complete and equitable.

At each step of the development of the status monitoring specification, all vendors' comments were carefully considered and weighed as to their contribution toward delivering a quality product to the network operators. Of course, there were several conflicting requirements, including low cost, time to market, high functionality and differing approaches to the same problem. The facilitation process attempted to make rational decisions to achieve the ultimate goal of a functional and a timely status monitoring solution.

ELECTROMECHANICAL SPECIFICATION

CableLabs is striving for interoperability on the communication level as well as the mechanical level. Interoperability means that a network operator could replace one vendor's status monitoring device with another vendor's without having to modify the communication mechanism it uses or without having to add or remove any special hardware. As a first step toward achieving the mechanical aspects of interoperability, CableLabs has issued a mechanical and electrical RFI.[2] The goal of this RFI is to obtain proposals, comment, and insight from the vendor community to aid the development of a cost effective and common form factor, connector, and electrical interface for status monitoring devices in Hybrid Fiber/Coax outside plant.

Each respondent has been requested to provide the reasoning, including technical and financial concerns, behind their responses so that an accurate assessment can be made. Respondents have also been requested to identify cost tradeoffs and issues on each of the following items, as well as for the overall proposal.

- Form Factor Description and Connector Location
- Temperature Range
- Relative Humidity
- Power Supply Voltages and Currents
- Weight
- Analog Monitor Inputs (Types and Number)
- Digital Monitor Inputs/Outputs/Bi-directional (Types and Number)
- Expansion I/O (SPI, I^2C, other)
- Local Craft Access
- RF Interface
- Surge Protection
- ESD Grounding

An initial proposal for using a PCMCIA form factor and connector looks promising, but more vendor comment and analysis are required before a decision can be made. The complete electromechanical RFI can be downloaded from the CableLabs home page, "http://www.cablelabs.com/."

PHYSICAL LAYER SPECIFICATION

All of the network operator requirements affect the Physical Layer, but there are two key requirements which have the strongest impact. First, since the Physical Layer details the requirements for the modem, it has a direct and substantial impact on the transponder's Bill of Materials and, therefore, direct and substantial impact on the cost of the transponder. Second, since the Physical Layer comes in direct contact with the cable plant, including the noisy return channel, it must be designed to overcome the detrimental effects of impulsive, ingress noise.

The following is an overview of the Physical Layer specification.[3] The complete specification can be downloaded from the CableLabs home page, "http://www.cablelabs.com/."

Robust, Inexpensive

To achieve robust performance while maintaining an inexpensive implementation, low-rate FSK (Frequency Shift Keying) was chosen for both the forward and return channels. FSK is very simple to implement and quite robust in the presence of noise. In fact, the specification requires that the system provide a 10^{-7} bit error rate (BER) when the received signal's CNR is as low as 14dB when operating only in thermal noise. Levels like this are easily achievable on today's HFC plants. The frequency shift chosen is +/-67 kHz relative to the center frequency. +67 kHz corresponds to a binary '1' and -67 kHz corresponds to a binary '0.'

This specification requires a minimum rate of 19.2 kbps with 9.6 and 38.4 kbps as rates that vendors can optionally provide. The overall bandwidth of a single channel is 400kHz. Such a wide bandwidth relative to the maximum data rate of 38.4kbps allows the modem's front-end filters to be designed simply and inexpensively.

Catlike Agility

MSOs require new systems such as a status monitoring system to be frequency agile. Their plants are in flux as new services are added and dropped in an effort to respond to market pressures. In order to operate efficiently in spite of this spectral churn, systems must be dynamically agile so that they can be reassigned to different frequencies at a moment's notice. Ideally, the agility should span the bulk of the available spectrum to give the operator the greatest range of flexibility. Unfortunately, delivering such a wide range of agility is expensive.

The Physical Layer specification currently compromises by setting the return channel spectral range at 5 to 25 MHz. However, this may be too wide for inexpensive implementation. A proposal is being considered to split this range into two smaller pieces — 5 to 10 MHz and 10 to 20 MHz. The transponders would be provisionable to operate over either range, but would operate only on one range at a time. Hopefully this proposal will provide the required agility at the necessary cost target.

Hermits Not Allowed

The status monitoring system will not be working by itself on the HFC plant. It must coexist with analog video, digital video, cable modems, analog set top return channels, telephony and other future services. Therefore, it must not interfere with the performance of another service, and its own performance must not degrade when another service is placed close to it on the HFC spectrum.

Interference with spectral neighbors is minimized by the spectral emission and spectral tone generation requirements. Any tones generated by the status monitoring system must be 50dB below the in-band transmitted power. Since status monitoring systems will usually be operated 10dB below analog video, this gives a 60dB carrier margin for analog video—plenty to maintain high quality analog video. The wide band spectral emission of the status monitoring system must 41dB below the in-band power when normalized to a 6MHz bandwidth. Again, since these systems will typically be operated 10dB below the analog, this gives a 51dB CNR on the analog video system—plenty to give high-quality analog video.

Susceptibility from spectral neighbors is minimized by compliance with the Selectivity specification. The Selectivity requirement states that the status monitoring system performance must not degrade as long as carriers between 200 and 400 kHz away from center are 30dB or less relative to the received in-band power; 400 to 800 kHz away are 40dB or less; and over 800 kHz away are 50dB or less.

Packets, Packets, Packets

The status monitoring system communications is packet based. Packets are used in both the forward and return directions. In addition, transmitters may be turned off during idle periods. Turning off is optional on the forward channel, while it is required on the return channel. As described in the Data Link Layer Specification section (below) many-to-one multiple access is achieved on the return channel through a rudimentary form of TDMA (Time-Division Multiple Access). The Physical Layer must therefore be able to turn on, transmit a packet and turn off in a specified manner.

The ramp times for the turn-on and -off are specified as 50μs. This is roughly two bit periods at 38.4kbps and is short enough to negligibly affect throughput performance but long enough to eliminate the need for complicated spectral emission control solutions which would be needed if very short 'square' edges were used. An extinction ratio of -50dB guarantees that when the transmitter is off, its output power is 50dB below its output power than when on.

Variety is the Spice of Life

Two flavors of transponders are being considered. Network operators feel that they need two particular features that could dramatically impact the cost of the transponder. First, they occasionally need a transponder that can transmit a particularly powerful signal. An application of such a transponder occurs when a transponder is placed at the end of a long coaxial run to monitor the health of the entire run, including the portion from the last amplifier to trunk termination. Second, they also have a limited need for a very accurately controlled power output when using the transponder to aid power calibration of amps and fiber nodes. Both of these features can increase the cost of a transponder up significantly.

A proposal to solve this problem is to specify two flavors of transponder: standard and premium. The standard transponder would meet all the nominal specifications given in the Physical Layer standard. The premium version would also meet the nominal specifications but would exceed them in maximum transmit power and transmit power accuracy. The proposal would have its maximum transmit power increase by 10dB from the nominal of +40dBmV to +50dBmV. The transmit power accuracy would tighten from the nominal of +/-2dB to +/-0.5dB.

The two version solution allows the purchase of the cheaper standard version for the bulk of applications, while allowing the purchase of the more expensive premium version for those few applications where extended functionality is required.

DATA LINK LAYER SPECIFICATION

The Data Link Layer (DLL) specification describes the tools which must be implemented by all status monitoring devices and which will be used by the headend EMS to allocate bandwidth and establish communication links. The goals of this specification are to specify a simple but functional MAC (Media Access Control) and LLC (Link Layer Control) that can be quickly implemented with inexpensive off-the-shelf components. The specification as described below allows simple polling with an ALOHA based contention scheme. Other variations of ALOHA and polling may be implemented. The specification is the result of several merged vendor responses.

Note that the following is an overview of the Data Link Layer. The complete specification can be downloaded from the CableLabs home page, "http://www.cablelabs.com/."

As part of the process of choosing and supplementing vendor responses to the Data Link portion of the CableLabs Outside Plant Management System RFI,[4] specific system and cost requirements had to be identified since, at the time, there were no hard requirements. One of the cost requirements specific to the DLL was that it must be sim-

Table 2: Forward and Return Path Data Link Layer PDU

Length (bits)	Name	Description
8	Sync	Identifies beginning of message
8	Control	Protocol, Contention enable
48	Address	Unique MAC address or group address
8	Message Length	Length of the payload field in bytes
n	Payload	
16	CRC-16	CRC-16 of entire packet

ple enough to be implemented by an 8 bit processor driven by a slow speed clocks. This forced complex synchronization techniques such as those used by TDMA systems (e.g., ranging) to be abandoned. Forward error correction was also abandoned to reduce cost and because the FSK modulation specified in the Physical Layer is already quite robust in noisy environments.

The Package and Its Contents

The Packet Data Unit (PDU) format for the forward and reverse paths is identical. It is described in Table 2 and Table 3. The only difference between the two forward and reverse formats lies in the control byte; a contention bit located in the control byte in the forward path PDU is not present in the return path PDU.

The sync byte identifies the start of the MAC layer PDU and is set to 0xa5. The sync byte also serves as an idle code in headend implementations that require idle codes. The control byte consists of a 7 bit protocol field and a 1 bit contention on/off indicator. The protocol field indicates the type of protocol used for the payload field of the Data Link PDU. This enables the use of multiple Message Layer protocols if, in the future, other types of devices on the network implement the Physical and Data Link layers. The use of the contention bit is described further below. The address is a 48 bit long IEEE MAC layer address. Status monitoring vendors will solicit the IEEE for a range of MAC addresses. The MAC address can be used for broadcast, multicast, or unique addressing. The message length identifies the size of the payload field from 0 to a maximum of 255 bytes. The payload field contains the information to be communicated between devices. The last field is a CCITT CRC-16 over the entire PDU excluding the CRC-16 itself.

Slicing the Bandwidth Pie

The Data Link layer specification allows access techniques ranging from polling to pure ALOHA to be implemented by the Headend EMS. This is accomplished through the use of several rules that govern

Table 3: PDU Byte Ordering

Start					End
Sync	Control	Address	Message Length	Payload	CRC-16

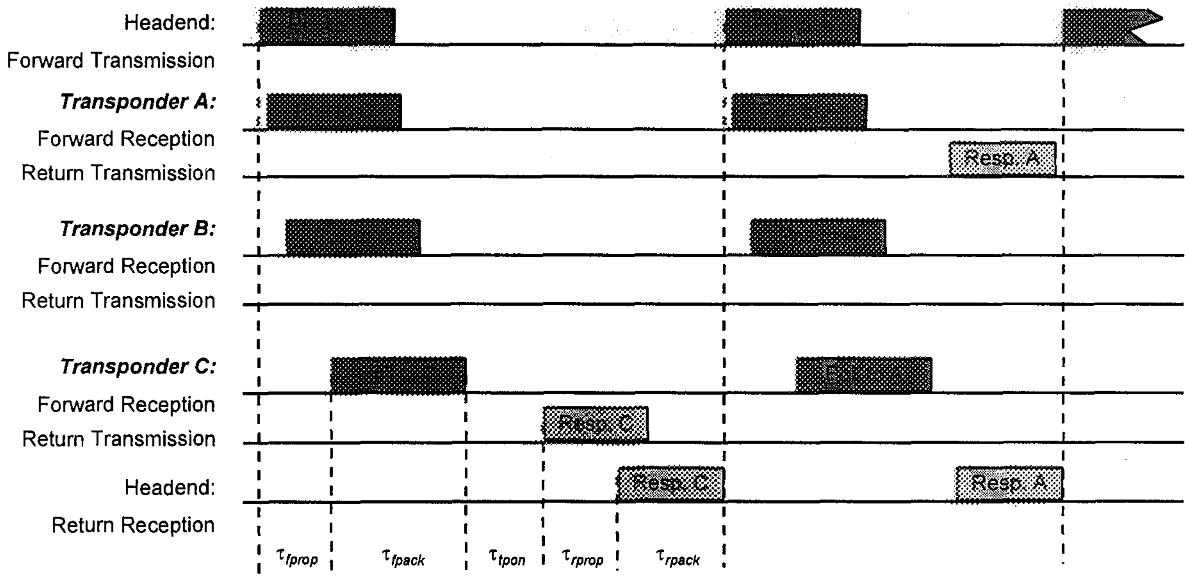


Figure 3: Forward and Return Packet Timing Relationships

message response and transmission times and through the use of a contention access indicator in the forward path PDU.

- Transmissions on the forward and return channels can occur independently of each other.
- The headend can transmit as often as it desires on the forward channel.
- A message received by a status monitoring device on the forward channel must be responded to in 10 ms or less.
- A message sent autonomously by a status monitoring device, such as might occur in the event of an unanticipated amplifier failure, must be acknowledged by the headend EMS within a backoff window, defined at the message layer, or it will be retransmitted. (The reader will recognize this MAC technique as pure ALOHA.)
- Autonomous messages may not be transmitted by status monitoring devices when the contention indicator is off and the message is not addressed to it.
- If the message is addressed to it, the status monitoring device may transmit its message regardless of the contention bit's value.

Upper layer processes handle all the message flow control and reliable transmission. Collisions on the reverse channel are detected by the lack of a response from the headend. No specific collision indication is provided at the data link layer.

This set of rules, as well as those discussed in other CableLabs documents,[5,6] allows the headend EMS to implement a variety of access schemes. Straight polling can be implemented by disallowing contention access, pure ALOHA can be implemented by allowing contention access but not polling, and a combination of polling with a pure ALOHA overlay can be implemented by selectively allowing or disallowing contention access. Figure 3 shows straight polling.

If the contention bit in the forward path PDU is turned on then status monitoring devices that need to autonomously transmit can do so at their convenience. This may

Figure 4: Message Structure		
Message Header		Message Payload
Message Tag	Control/Command	Payload
1 byte	1 byte	n bytes

lead to a collision, for example between transponder C and transponder B, which will be resolved through a binary backoff algorithm. An access scheme such as this is best described as polling with an ALOHA overlay.

The theoretical throughput of a polled access technique approaches 50% for forward and return paths as propagation and processing delays approach zero. The theoretical throughput of a pure ALOHA access technique is approximately 18%. A hybrid implementation will result in a throughput somewhere between these two values and its exact value will depend on several factors including the percentage of time spent in ALOHA mode vs. the time spent polling.

Slotted ALOHA was initially considered instead of pure ALOHA since its theoretical throughput is approximately 37% for the same 'slot' size. But the slot size used for a poll and wait scheme is more that twice that required for pure ALOHA, since a 'slot' in an ALOHA scheme is only the length of a single packet. Therefore, the theoretical throughput for pure ALOHA using the rules presented offers better efficiency than a slotted ALOHA approach. It also scales better as the transmission bit rate increases.

The set of rules presented in the Data Link layer specification provides for a flexible and efficient means of offering bandwidth and establishing communication links. Access techniques such as polling, pure ALOHA, or polling with a pure ALOHA overlay can be implemented by a headend EMS adhering to this specification.

MESSAGE SET SPECIFICATION

Upper layer communication is achieved through use of the message set. The message set provides the means and the structure for communicating information to and from outside plant status monitoring equipment. It is simple, optimized for speed and extensible. The following is an overview of the message set. The complete message set document can be downloaded from the CableLabs home page, "http://www.cablelabs.com/."

The Toolbox

The messages are the tools used by the system to communicate information to and from the outside plant transponders. They are used primarily to read and write individual parameters at the transponder, but also can be used to give direct commands to a transponder. For example, to change the transmit power level, the `SetParameter` message would be used to write the parameter `TransponderTxmtPowerLevel` with the desired value.

Each message has the same basic structure, as shown in Figure 4. The header of each message includes a tag number to help correlate responses with the originating query and a control/command byte which includes bits to indicate continuation, encryption and perform SAR (Segmentation and Reassembly) of packets too large to be transmitted in a single DLL packet.

The payload of each message contains message specific information, such as the message name, the parameters it addresses (if any) and the data being communicated (if any). The payload is variable length as different messages require differing amounts of information to do their job. The most common messages are kept very short so as to optimize communication speed. Specifically, the `GetStatusMajor` message is the most commonly transmitted message from the headend to the transponder population. It simply requests all current major alarms from a specific transponder. Normally, when there are no alarms present (hopefully this is normal!) the transponder would respond with the `ReportStatusNormal` to indicate not only that there were no alarms but also that the transponder was functioning correctly. These two messages, because they are used so often, have payloads of a single byte, giving them a total length of 3 bytes.

The full list of messages and their meanings can be found in the Message Set specification[5].

The Nuts and Bolts

The parameters are the nuts and bolts of the communication system. The messages operate on the parameters to accomplish a desired task. Each parameter looks like a record in a database. It is composed of several fields which can be read and written by the appropriate message. The fields are detailed in Table 4 and include information about the parameter's actual value as well as information on how an alarm is triggered.

The parameters hold all the information which can be communicated to and from a transponder. They are arranged into groups according to the equipment being monitored by the transponder. These groups are:

- *Common:* Parameters common to all transponders. Examples are `UnitAddress`, `TransponderTxmtPowerLevel`, and `Temperature`.
- *Amplifiers and Line Extenders:* Parameters associated only with amps and line extenders. Examples are `ForwardAmplifierCurrent` and `DCPowerSupplyVoltage`.
- *Fiber Nodes:* Parameters associated with fiber nodes. Examples are `ReturnLaserPower` and `DCPowerSupplyVoltage`.

Table 4: Transponder Parameter Record

Field	Purpose
Parameter Name	Byte code of monitored parameter
Data Type	Type of Data (see Table 3.2.3-1)
Data Value	Actual value of parameter
Monitor Point	Number of the physical monitor point in transponder - this Parameter corresponds to I/O port number.
Alarm Direction	Indication whether alarm point is positive going or negative going
Major Alarm Threshold	Data Value at which Major Alarm occurs
Minor Alarm Threshold	Data Value at which Minor Alarm occurs

- *Power Supplies:* Parameters associated with power supplies. Examples are `InputACVoltageLevel` and `ChargingCurrent`.

Of course, the total list of all the parameters which might be found across the population of transponders on a large HFC plant is quite long. (For details, the specification can be downloaded from the CableLabs home page.) Fortunately, each individual transponder only needs to support a small subset of the parameters, including the common parameters and those parameters connected with the equipment it is monitoring. This makes the total amount of information monitored by a single transponder reasonably small and therefore cost-effective.

Simple or Complex: You Make the Choice

The depth of the message set and the width of parameter set to be implemented is up to the vendor. They can balance the desire to monitor everything under the sky with the cost of doing so. The message set and its associated parameter list has been designed with the flexibility to allow this choice.

Extensible

Finally, the message set and the parameter list are extensible to allow vendors and MSOs to adapt the system to changing networks and technology. It is difficult if not impossible to foresee all the possible items to be monitored by the status monitoring system. As new equipment and technology are implemented and find their way into the HFC network, new voltages, currents, switch positions, and many as yet undetermined items will need to be monitored. The parameter set is dynamically extensible to accommodate such new data.

CONCLUSION

The CableLabs HFC outside plant management specification effort is producing a useful set of specifications by which vendors can quickly and cost-effectively get products to market. Accordingly, the cost of deploying such specifications-based systems will drop thereby increasing the demand and increasing the total market. As usage of status monitoring systems increases, Cable MSOs can more effectively manage their networks in the ever increasing competitive telecommunications marketplace.

REFERENCES

1. Roger Brown, *"Are HFC Networks at the Breaking Point?,"* CED Magazine, December 1996.

2. CableLabs, *"Mechanical and Electrical Specification for Hybrid Fiber/Coax Outside Plant Monitoring Systems Request for Information,"* December 9, 1996.

3. CableLabs, *"Physical Layer Specification for Hybrid Fiber/Coax (HFC) Outside Plant Monitoring Systems Interim Report,"* December 9, 1996

4. CableLabs, *"Hybrid Fiber/Coax Outside Plant Management System — Request for Information,"* August 12, 1996.

5. CableLabs, *"Specification of Layer 3 Message Set for Hybrid Fiber/Coax (HFC) Outside Plant Monitoring Systems - Draft Report,"* December 9, 1996..

6. CableLabs, *"Data Link Layer Interim Specification for Hybrid Fiber/Coax Outside Plant Monitoring Systems,"* November 14, 1996.

APPROACHES TO SECURITY AND ACCESS CONTROL FOR DIGITAL CABLE TELEVISION

Claude T. Baggett
Cable Television Laboratories, Inc.
Louisville, CO 80027-1266

Abstract

This paper addresses the several design trade-offs which should be considered by a system operator in selecting a security and access control sub-system for protecting digital television signals on cable television systems. Factors are discussed which have a strong impact on the strength, adaptability, and cost of such systems. The principles discussed in this paper apply whether the primary source of security and access control in the subscriber home is found in a set-top decoder, a decoder interface unit, a home server, or as an insertion in an MPEG-capable computer or television receiver.

SCOPE

This paper will address the options which system operators should consider in choosing a security and conditional access system for their digital television signals. It will not address cryptographic and key handling processes in depth, but only as required to explain the trade-offs in system selection.

BACKGROUND

There are a number of options and design trade-offs which must be considered in order to optimize the digital security and access control functionality to cable system needs. Some of these decisions have only minor impact on the quality of the security sub-system, but others will have a pronounced affect on the robustness, integrity, usefulness, cost, and extendibility of the chosen system and its deployment. A "one size fits all" approach to the functionality and features of the security and access control system is not appropriate because of the great variation in cable system size, services, operating philosophy, and access to capital.

In analog television, there is basically only one decision which must be made for access control; that is, whether to descramble the incoming signal or not. The same descrambler in the set-top converter is used whether the system has one scrambled channel or many. The sophistication and depth of the scrambling mechanism is limited because the analog signal is very difficult to reconstruct without leaving unacceptable artifacts in the visible picture. Therefore, with only one, very limited, process to protect the analog television signal, meaningful security is difficult. Some later systems, which use line shuffling under the control of a modern cryptographic system and hard-encrypted audio, are much superior in this aspect, but are costly and have come too late relative to the advent of digital television.

To protect digital signals, four basic mechanisms are utilized. First, a cryptographic algorithm is defined for the headend which will take the digital signal and scramble the binary characters so thoroughly that no practical amount of analysis will regain the original signal. This same mechanism is used to reconstruct the signal in the subscriber home. If this were all that was possible with digital scrambling it would still be of great benefit because of the depth of the obfuscation of the signal and the fact that it can be restored to its original, pre-scrambled, condition without degradation. However, digital security is much more versatile than that.

The second basic mechanism is the electronic key—literally a long binary word—which controls the scrambling and descrambling processes in the cryptographic unit. The cryptographic algorithm itself is useless without the electronic key. Furthermore, unlike analog scrambling, a different electronic key can be used for differentiated services. Multiple keys can be used to protect a tier of services, a single digital channel, or a specific string of digital data such as a control channel, a pay-per-view event or a multimedia display. Keys can be symmetric, meaning that the same exact binary word is used to scramble and descramble the digital data; or they can be asymmetric, meaning that the key used to descramble the signal is different from the one used to scramble it. Keys must be generated, managed, protected, transmitted, and utilized, thus forming the key management and distribution system which is the most critical part of the security process.

The third basic part of this system is called the entitlement/authorization message or matrix. Originally, this was just a matrix of two columns by the number of rows equal to the number of occupied channels on the cable system. If a given channel had a binary one in the second column it was authorized for descrambling, and if a zero was placed there the customer was not authorized to receive it. This matrix was encrypted for transmission and stored in the secure microprocessor in the access control system. Now, a cryptographic system, somewhat like the master key systems used in buildings, can be devised which convert the simple authorization matrix into a cryptographic process, thus making it much more difficult to defeat.

The fourth part of the system is a secure signature mechanism. Secure signatures make it possible to guarantee the identity of the sender of the digital message and to verify that the message has not been modified en route. This is especially useful in key distribution, certain control messages, and in purchasing.

The specification of these four mechanisms involves the consideration of the intended application, technology issues, threats and countermeasures, governmental policy and regulation, costs, and desirable additional features.

DEFINITION OF TERMS

The following terms are either used in this paper or are considered important to the understanding of digital security systems.

Algorithm: A mathematical process which can be used for the scrambling and descrambling of a data stream.

Authentication: The process by which one party can ascertain with certainty the identification of another party.

Authorization Coding: A digital word which describes the personality or service access capability of the subscriber decoder unit. This code word, which is based on the service access authorized by the billing system, determines which keys are distributed to each customer, and is required at the subscriber decoder to authorize the descrambling of any specific program.

Conditional Access System (CA): The complete system for ensuring that cable services are accessible only to those who are entitled to receive them, and that the ordering of such services is not subject to modification or repudiation.

Cryptanalysis: The science of recovering the plain text of a message without access to the key (or electronic key in electronic cryptographic systems).

Cryptographic Duty Cycle: The maximum secure capacity of a cryptographic process, based on total data throughput on a single key vector.

Descrambling: The process of reversing the scrambling to yield usable pictures, sound, and data services on a cable system.

Electronic Key: The term for data signals which are used to control the descrambling process in subscriber decoders. There are at least three types of electronic keys referenced in this Recommendation, including those used for television signal streams, those used for protecting control system operations, and those used for the distribution of electronic keys on the cable system. While the Authorization

Coding is effectively a key, it is treated separately in this section.

Encryption: The process of scrambling digital signals to avoid unauthorized access.

Host: A device with generalized functionality where modules containing specialized functionality can be connected.

Integrity: The ability of a function to withstand being usurped for unauthorized usage, or modified to yield unauthorized results.

Intrusion Resistance: The ability of a hardware object to deny physical, electrical, or irradiation-based access to internal functionality by unauthorized parties.

Module: A small device, not working by itself, designed to run specialized tasks in association with a host.

National Class Laboratory: The primary source for cryptanalysis in a national government.

Non-Repudiation: A process by which the sender of a message cannot deny having sent the message.

One-Way Hash: A mathematical process or algorithm whereby a variable-length message is changed into a fixed-length digital word, such that it is very difficult to calculate the original message from the word, and also very difficult to find a second message with the same word.

Pay-Per-View: A payment system whereby the subscriber can pay for an individual program or programming period, rather than for a full-period terminated service.

Piracy: The act of acquiring unauthorized access to programming materials, usually considered for the purpose of reselling such access to illegal subscribers.

Public Key Cryptography: A cryptographic technique based upon an asymmetric two-key algorithm, private and public, wherein a message is encrypted with the public key but can only be decrypted with the private key. Knowing the public key does not reveal the private key. Therefore, Party A would devise such a private and public key, and send the public key openly to all who might wish to communicate with Party A, but retain the private key in secret. Then, while any who have the public key can encrypt a message for Party A, only Party A with the private key can decrypt the messages. Also known as a Private-Public Key (PPK) system.

Scrambling: The process of using an encryption function to render television and data signals carried on a cable system unusable to unauthorized subscribers.

Secure Signature: A mathematical process by which the origin and integrity of a transmitted message can be ascertained. This means that the originator cannot deny having sent the message, and the receiver can determine if the message has been modified.

Transport Stream: An MPEG-2 (Moving Pictures Expert Group) or other data digital transport stream.

DESIGN CONSIDERATIONS AND TRADE-OFFS

Since this paper will not address cryptographic algorithms in depth, it appears useful to dispose of this topic first. For many years now, the approved civilian cryptographic process in the United States has been the Digital Encryption Standard, or DES. DES has several operational modes in order to be as widely applicable as possible. The National Institute of Standards and Technology (NIST), the government agency responsible for civilian cryptography, has approved DES for use through approximately 2003, but plans to have a new encryption standard in operation by that time. This does not mean that the DES equipment in the field will suddenly cease to function or become more vulnerable, but that NIST will no longer support it.

DES is only one of many hundreds of electronic encryption algorithms which have been devised, each with its target application, strengths and weaknesses. Many of these are appropriate for the encryption of digital television signals, therefore a choice of these on merit would be difficult indeed. The North American cable industry has decided to begin their digital television transmissions using the GI DigiCipher 2® system, which is DES-based. This does not preclude an operator from deciding on another algorithm if so desired, or starting with DigiCipher® at the outset and changing to another algorithm at a later rebuild.

The important issue for being empowered to choose from multiple algorithms is to be able to select or change systems without undue cost or operational impact. An operator may decide to change cryptology for several reasons, including:

1) The security has been compromised and pirating has begun;

2) Unit reliability has deteriorated to an unacceptable level; and

3) The operator wishes to add new services which the existing suite of security equipment cannot protect.

This issue leads to the first design trade-off question in this paper, which is:

Removable and replaceable versus built-in security

In the past, analog descrambling converters have always had security-related circuitry as an integrated part of the unit. Removable and replaceable descrambling circuitry has never been important in analog television because that circuitry represented an important portion of the overall cost of the box, and a pluggable interface to the descrambler is not cheap or easy. With digital decoders, the security represents a relatively small portion of the cost of the unit, and the interface for digital data is a well known and practiced science.

There are three basic approaches to the architecture of the security system as far as its placement is concerned. In case one, the digital security circuitry can be integrated into the digital circuitry in the MPEG-2 decoder unit. Case two has the security functionality totally placed in a removable module, such as a PCMCIA card, with an open architecture interface to the host device. Case three is a hybrid of the first two wherein the security sub-system is built into the decoder unit, but an open architecture interface is provided so that the internal system can be replaced by a pluggable, replaceable one.

The fully integrated decoder unit with no external socket for replaceable security probably represents the path of least initial cost to the operating system. However, if and when the security requires replacement, the entire decoder unit must be replaced, representing hundreds of dollars rather than tens of dollars for a PCMCIA removable security module. The other problem with this approach is that it is not responsive to the Telecommunications Act of 1996. This law states that set-top decoder units must be available at retail to the subscriber, but not the security element used by the cable system. This law necessitates making the security circuitry removable if the cable operator is not to totally lose control of the security of the cable programming. If the security circuitry is not removable, then the consumer electronics manufacturer will decide which security is adequate to protect the cable business, and there will be only one algorithm available.

The totally removable security element is responsive to the Telecommunications Act, and provides the cable operator a cost-effective way to replace the functionality when the need occurs. The proposed PCMCIA module has adequate capability to provide all of the needed functionality and additional features demanded by the marketplace and has an interface which is more than adequate for the needed control and data transfer. There are three separate efforts underway to standardize the interface between the removable module and the host device. These are the Digital Video Broadcasting (DVB) Common Interface Specification being planned for Europe, which uses a pin-depleted version of the PCMCIA card, the National Renewable Security Standard (NRSS) of the United States which includes specifications for both the PCMCIA module and the ISO-7816 chip card, and the security working groups within the Society of

Cable Television Engineers (SCTE) Digital Video and High-Speed Data Subcommittees of the SCTE Engineering Committee.

To define an interface between the security module and the host system, three separate issues must be determined. First, the physical form factor of the module must be specified along with the number and placement of the connector contacts or pins which provide the physical interface between the two entities. Second, the electrical specifications regarding powering, grounding, logic levels, and clock speeds must be set. Finally, the format of the data, and the command set which controls the interface and integrates operations must be determined.

The NRSS specification stops at this point and makes no attempt to define the exact security functionality contained within the removable module, preferring to leave this determination to the marketplace interaction between the cable operator and the equipment vendor. The DVB specification attempts to go further and define the exact security algorithm and feature set to be used by the operator. If DVB were accepted worldwide it would mean that every cable system, broadcaster, satellite deliverer, and MMDS operator would have the exact same cryptographic function for security, but that they use an electronic key unique to their system. The best advice from cryptoanalysts in the US and elsewhere is that this is a foolish undertaking as it would dramatically increase the worldwide vulnerability to the pirating of communications signals. However, there appears to be no strong reason why the interface specification between the NRSS and DVB standards cannot be harmonized and efforts are underway in the NRSS subcommittee at this time to accomplish that task.

The specifications generated by the SCTE subcommittees will no doubt reflect some of the work which has already occurred in NRSS and DVB, with further tailoring of the requirements to cable industry needs. However, the DVB and NRSS work specifically targets the interface of MPEG-2 signals and may not be optimized for other data transmissions used on cable systems. Therefore, efforts are beginning in the Security working group of the High-Speed Data Subcommittee of the SCTE to define a new interface, perhaps somewhat like NRSS and DVB, which will specifically meet the needs for the data transmission infrastructure on cable. It is probable that the results of the SCTE efforts will yield the proper purchase specifications for the cable television and data industry.

In-band, out-of-band, and hybrid control channels

Traditionally, the control channel for analog descrambling converters has been carried on an out-of-band separate carrier located in the downstream cable passband. Since that portion of the control channel dedicated to security and access control primarily was dedicated to a single issue, allowing the descrambler to turn on or not, overhead on the channel was minimal relative to that required for digital security. Satellite, MMDS (Multichannel Multipoint Distribution Service), and SMATV (Satellite Master Antenna Television System) delivery systems for MPEG television signals use an in-band control channel scheme wherein the packets for security or other control functions are inserted into the transport packet stream. At the receiver, a Program Identification Filter (PID Filter) examines the headers of the incoming packets and routes them according to their content, video, audio, or control. Packets called EMMs (Expanded Memory Manager) and ECMs (Error Correction Mode), which contain key and authorization codes, are routed to the cryptographic processor. It is also possible to do an amalgamation of these two systems where certain control functionality is carried on an out-of-band carrier, and security packets are contained in-band.

Note that the in-band signals carried with each MPEG transport stream apply only to that channel whereon they are carried, and each MPEG television channel has its own peculiar security packets. On an out-of-band control channel all of the keying and authorization data applying to all digital channels are carried on the single carrier. There are pros and cons to each of these approaches. The in-band system requires a packet inserter in the headend for each encrypted digital channel. While this seems like an unsupportable complication, it must be remembered that each encrypted digital channel also has a digital encryption device at the headend. Adding the ability to insert packets at that point represents an extremely small increase in complexity. For programming sources, such as Headend in the Sky™ (HITS) where re-encoding at the headend may not be necessary, the insertion of ECMs and EMMs can be accomplished at the uplink facility and no additional complexity is required at the cable headend. The out-of-band carrier represents a system vulnerability in that if the signal is jammed or otherwise fails, the entire system, including every encrypted channel, also fails at the next key change. Whereas with the in-band system, the failure to rekey only affects the single channel where the problem occurs. There also is some question as to whether cryptographic synchronization can be supported during rapid key change intervals with the single out-of-band carrier.

All of this relates back to a basic security consideration for multichannel digital television on cable. As explained earlier, if you have a proper functioning decoder unit, there are still two key elements required to decrypt a program; the cryptographic key and the authorization code. Suppose a security and access control system uses the same cryptographic algorithm and electronic key to secure every differentiated channel on the system. This means that the correct key is present in the home terminal equipment to decode every program on the system. The only feature preventing that from happening is the authorization code which sets the personality of the decoding unit. This therefore negates much of the advantage of digital security by making it work just like existing analog descrambling converters, meaning, turn them on and they decrypt, turn them off and they don't. The biggest advantage is found in using different cryptographic keys for each differentiated service or tier of channels. This means that each differentiated service or tier is encrypted completely differently from any other, and if a subscriber doesn't take a certain service, the key to decrypt it is not even present in the home terminal.

Now, what does this mean for in-band versus out-of-band control channel architecture? If each pay service, each pay-per-view, and each differentiated tier of channels has its own unique cryptographic key—which is changed on a fairly rapid basis during each day—a huge overhead is placed on the out-of-band channel during key update periods, which is fairly continuous. Additionally, an extra burden is placed on the home decoding unit to ensure crypto-synchronization during key change periods. Since the proper key is sent in conjunction with the secured video and audio in the in-band case, crypto-synchronization is virtually automatic.

It is also unclear at this time how cable systems which have implemented ATM (Asynchronous Transfer Mode) transport structures for voice, data, and video can accomplish out-of-band control, since, by definition, all control is in-band.

There is a hybrid solution which may actually be the best choice of all in these design considerations. This is where the EMM and ECM packets are sent in-band with the MPEG transport stream, but all other control signals are sent over an out-of-band channel. This takes the burden for key distribution off of the out-of-band control channel, but still facilitates other control functions. Besides the home terminal unique control signals sent on the out-of-band channel, other signals such as pay-per-view promotionals, local clock, channel maps, purchasing communications, program guide updates, security and usage audits, messaging, etc., could then be easily transported on this channel in a global fashion. Sending certain control keys over the out-of-band channel may facilitate the compartmentalization of large systems into sub-key regions, to reduce the marketing area of the subscriber/pirate, and to provide a unique point of leverage in disenfranchising cloned home terminals.

Public-private and symmetric key systems

In Public-Private-Key (PPK) systems, an algorithm is used which has two different keys, one called private and the other public. The public key is sufficient to encrypt a message for transmission, but the private key is required to decrypt that message. So, if Party A wishes to receive communications from Parties B, C, and D, Party A will send them the public key part of the key-pair, while keeping the private key strictly to himself. Whenever B, C, or D wishes to communicate a message to A, they encrypt it with the public key and transmit it to A. Note that while B, C, and D, all have the same public key, they cannot read each other's messages to A because they do not have the private key. Party A can read any message from B, C, or D because the private key has been retained. In a cable system, Party A would be the headend, and Parties B, C, D, etc., would represent the subscribers with digital descrambling converters. The headend would send to each subscriber its public key so that each subscriber terminal could communicate toward the headend either on the return plant or via telephone return, if two-way communications is desired. Each subscriber terminal would also send its public key to the headend, or do so at the time of converter issuance and activation, so that the headend could communicate uniquely to each of the subscriber terminals. This would allow the unique control of each unit individually. It would also be possible to have a second PPK system in which the public key from the subscriber terminals are all the same, thus allowing a single global message to be sent containing material that all receive, such as clock, channel map, etc.

Symmetric key means that the exact same single key is used for encryption and decryption. This is much simpler than the PPK approach, but only if you have a method for delivering that symmetric key to each legitimate customer without it being revealed to unauthorized subscribers. With symmetric key, each differentiated channel or tier would have just one key per key period and the problems of key distribution in a PPK environment are eased. The probable best answer is once again an amalgamation of the two systems. Suppose a cable operator used a PPK system to protect the current operating symmetric keys during their distribution to those subscribers who are authorized to receive them. The symmetric key, encrypted by the public keys from authorized subscribers could then be delivered without unauthorized subscribers being able to discern them. Since the changing of symmetric keys is not required on a continuous—but more of an intermittent—basis, the overhead in the control channel from the PPK system can be lessened.

Taking this one step further on any given system there are a limited number of perturbations in the possible service personalities available to the subscribers. Then, each home terminal could be differentiated based on the service personality group into which it falls. All home terminals of the same service class could be keyed similarly, if pay-per-view and purchasing is split out and handled separately. This advantage would fall apart in a completely a'la carte service rendering.

Anytime a rekeying message is directed to the home terminals (if it is sent under a secure signature mechanism) at least the terminal can ascertain that the message is from the headend and that it has not been modified en route.
One other problem with PPK systems as the single system for encryption in cable television is that the operations are more complicated and require more processing and transmission time. This overhead increase could become an important factor in control channel access in a PPK-only system.

There are a number of iterations in design concept based on the use of PPK and symmetric key systems. The important thing to the cable operator is to understand how the system works, and the amount of overhead left in the control channel to allow future subscriber-base expansion without a required rebuild of the CA system. It must be said that the use of any of these systems does not obviate the clone terminal threat. In that case, the clone terminal is configured cryptographically just like the pirate's legitimate terminal and each time the legitimate terminal is given a new key; the clones are likewise updated. Other countermeasures are required in addition to the ones discussed here to resolve this threat.

Form factor trade-offs

This subject was partially discussed above, but a few more words are required for completeness. There are two worldwide accepted module formats which are candidates for use in cable operating systems, the PCMCIA card and the ISO-7816 chip card. Either one could be made to work for the most simple conditional access applications. The ISO-7816 card is basically limited to a single 25 square mil integrated circuit, which is nonetheless adequate for a straightforward decryptor, such as DES. It may not be adequate to house a simultaneous decryptor/encryptor and secure signature unit with adequate storage for several keys, even with limited or no additional features. There are only eight contracts on the card for interface, which means that the data must be streamed onto and off of the card in serial format. This card has been tested at 50 Mbps serial input/output with good reliability. However, since the security card is intended to be inserted into the socket and left there for long periods of time, perhaps years, it is not known how corrosion at the point of contact between the socket and the contact pad would impair this data rate.

The PCMCIA card has sufficient contact pins so that the input and output data can be sent in byte-parallel, or at least nibble-parallel, format, thus reducing the impact of corrosion impairment on a single pin. The module also has considerably more volume than the ISO card, thus facilitating more complex security schemes and multiple features.

Another option for cable would be to define its own unique module form factor and pin configuration. While this has some desirable features, it presents almost insurmountable problems at the cable/consumer electronics/computer interface, and in the end, would not likely represent any real advantage over using the proven and widely deployed PCMCIA card.

DEPLOYMENT OPTIONS

While we are discussing options for moving into the digital security era on cable, a few words regarding deployment are in order. No matter what approach is chosen, everyone in the industry understands that this transition is going to be costly and given to a certain amount of operational disorder.

Here are some options to consider, with probable results:

Table 1: Deployment Options for Digital Scrambling

Option	Probable Result
Ignore digital, stay analog	Be driven out of business.
Co-carry both analog and digital scrambled services along with analog unscrambled basic.	Inventory proliferation; Pirates still steal analog services; Wasted spectrum; Reduced incentive for customer to move to digital tier.
Keep unscrambled analog basic only; Carry all scrambled services on digital system.	No descrambling analog converters required; All pirating stopped or deferred; Improved pictures with digital incites customer to move to digital tier.

For example, suppose a system was capable of 66 channels, being divided into 30 channels in a basic unscrambled analog tier, with the balance scrambled and apportioned into an expanded basic tier, and some number of pay and pay-per-view services. If you chose to approach your digital transition according to the third option above, you might consider the following channel breakout:

1) 30 Unscrambled Analog Basic Tier Services

2) 16 Channels (96 MHz) of Open Spectrum Dedicated to New Digital Data Services such as High-Speed Data or Telephony

3) 120 Digital Scrambled Channels (20 - 6 MHz Channels at 6:1 SDTV Compression)

Only those subscribers taking pay services of any kind are required to have the digital decoder units, which they can rent or purchase from the cable company, or purchase at their consumer electronics retailer. In either digital case, the cable operator is responsible for furnishing the removable security and access control module to each pay customer. In this scenario, the operator has had a net increase of 100 standard definition digital channels, has retained the 30 channels of analog basic programming for the transition period, and has netted an additional 96 MHz for further digital service development. A similar case can be made for systems with fewer channels which result in similar proportional gains. To be economically feasible, this scenario supposes that the system operator acquires the digital programming channels in a format suitable for available home terminal units without having to decode and re-encode each digital channel in a new MPEG format, but does require that trans-encryption is performed for every scrambled digital channel.

CONCLUSIONS

Competitive issues require cable operators to make the move to digital as expeditiously as possible. Enhancements in security and picture quality will directly or indirectly off-set some of the transition costs. Digital transmissions from competitive sources are already operational so cable is behind in deployment. It is absolutely necessary that cable deploy digital services as soon as possible to stop erosion of the customer base. There are a number of options for security and access control which should be considered by a cable operator based upon unique and individual system and company requirements.

When considering the purchase of a security and access control system with which the cable operator has no direct experience, it would be well to consider using the services of a communications security certification company to examine the proposed system for compliance to the purchase specification and to ascertain any unadvertised vulnerabilities which the system may possess.

Cable Networks and Distributed Video Repositories

Shahram Ghandeharizadeh
Matsushita Information Technology Laboratory
Panasonic Technologies Inc.
2 Research Way
Princeton, NJ 08540
shahram@research.panasonic.com

Abstract

Cable networks stretch to many homes in the United States. The available bandwidth of this network exceeds the amount of transmitted content (programming). One way to increase revenue is to employ the idle bandwidth to transmit on-demand programming for a nominal charge either on a pay per view or a monthly flat fee basis. This service would require deployment of a distributed video server across a geographically distributed region, in order to provide programming based on demographic characteristics of each region. This system could provide the necessary infrastructure for the cable companies to expand their home programming with Internet services. This paper provides a global perspective of such a system.

1 Introduction

The cable network with its extensive outreach to homes can conceptually be compared with either the Internet or the telephone network. Currently, the programming potential of the cable network far exceeds the current broadcasted programming it carries to homes. This is partially due to lack of content that is not appropriate for broadcasting. This study proposes extensions to the cable network to provide on-demand service to homes. To realize this objective, it proposes the use of a distributed continuous media server to segment the cable networks of a service provider into partitions. A partition for a specific region strives to service the customers located in that region with the appropriate content.

The envisioned continuous media server consists of a central server and a number of regional servers. The regional servers are a client of the central server. Moreover, they are geographically located such that the load imposed by the customers of each region on each regional server is approximately the same. Each server can support continuous display of audio and video clips. In a typical configuration, the central server would have a higher throughput than each of the regional servers. The central server is a library that contains a copy of all the programs available on the system. Each regional server caches a fraction of the programming available on the central server. This caching is done intelligently based on demographics and the programming expected to be of interest to majority of the residents in a geographical region.

To provide on-demand service, the startup latency observed by a customer should be minimized. *Startup latency* is defined as the amount of time elapsed from when a customer requests a video clip until its display starts. To minimize this latency, the system should connect a user with a regional server that is in close proximity. This would also reduce the number of participating cable links in order to maximize the number of simultaneous displays supported by the cable network. Of course, the regional server should contain the programming desired by a client. Otherwise, it may pursue two alternative possibilities. First, it may forward the client to be serviced by either another regional server that contains the desired programming or the central server. Second, it may elect to cache the requested program from the central server in order to service the client. In both scenarios, the client will observe a higher startup latency. The first approach is appropriate if the requested program is unpopular (not frequently accessed) with that region. With the second scenario, the client would cause the regional server to contain a copy of the referenced program. This caching is appropriate if the requested program is expected to be frequently referenced by the other clients of this regional server. Note that the regional server might have insufficient space and be forced to delete some other program in favor of materializing the requested program.

This paper describes *scalable* servers that can participate as either the central or regional servers. In addition, it hints at *caching* techniques to minimize the amount of storage required at a regional server. A scalable server is desirable for several reasons. First, it separates the software from the physical parameters of its underlying hardware, e.g., the number of mass storage devices. Second, the software does not restrict the number of simultaneous displays that can be supported by the hardware. As the hardware grows, the software enables the system to support a higher number of displays (as long as the hardware platform does not hit a physical limitation). Third, it provides for an incremental investment approach where a service provider starts with a small hardware configuration and expands the system based on customer demand.

Intelligent caching techniques minimize the amount of storage required at a regional server. This reduces the number of storage devices that in turn reduces costs. In addition, it reduces the mean time to failure of mass storage devices in order to improve the availability of data. To illustrate, at the time of this writing, the mean time to failure of a single disk drive is once every 600,000 hours (or 70 years). With a system that consists of ten disks, the mean time to failure of *some* disk is once every 60,000 hours (or 7 years). The mean time to failure of some disk in a system consisting of one thousand disks is once every 600 hours (or 25 days). Of course, these are theoretical expectations with almost always a lower practical expectations.

2 Scalable Servers

Recent advances in computer processing and storage performance and in high speed communications has made it feasible to consider continuous media (e.g., audio and video) servers that scale to support thousands of concurrently active displays. The principle characteristics of continuous media is their sustained bit rate requirement. If a system delivers a clip at a rate lower than its pre-specified display rate without special precautions (e.g., pre-fetching), the user might observe frequent disruptions and delays with video and random noises with audio. These artifacts are collectively termed *hiccups*. For example, CD-quality audio (2 channels with 16 bit samples at 44 kHz) requires 1.4 Megabits per second (Mbps). NTSC quality video (640 $\times$ 480 pixels per frame, 24 bits per pixel) at 29.9 frames a second requires 210 Mbps. At the time of this writing, a variety of HDTV quality video are available with display bandwidth requirements ranging from hundreds of Megabits to Gigabits per second. These bandwidths can be reduced using compression techniques due to redundancy in data. Compression techniques are categorized into *lossy* and *lossless*. Lossy compression techniques encode data into a format that, when decompressed, yields something similar to the original. With lossless techniques, decompression yields the original. Lossy compression techniques are more effective in reducing both the size and bandwidth requirements of a clip. For example, with the MPEG standard, the bandwidth requirement of CD-quality audio can be reduced to 384 Kilobits per second. MPEG-1 reduces the bandwidth requirement of a video clip to 1.5 Mbps. With some compression techniques such as MPEG-2, one can control the compression ratio by specifying the final bandwidth of the encoded stream (ranging from 3 to 15 Mbps).

The size of a compressed video clip is quite large by most current standards. For example, a two hour MPEG-2 encoded video clip, requiring 3 Mbps, is 2.6 Gigabytes in size. (In this paper, we focus on video due to its significant size and bandwidth requirements that are higher than audio.) To reduce cost of storage, a typical architecture for a video server employs a hierarchical storage structure consisting of DRAM, magnetic disks, and one or more tertiary storage devices. As the different levels of the hierarchy are traversed starting with memory, the density of the medium and its latency increases while its cost per megabyte decreases. It is assumed that all video clips reside on the tertiary storage device. The disk space is used as a temporary staging area for the frequently accessed clips in order to minimize the number of references to tertiary. This enhances the performance of the system. Once the delivery of a video clip has been initiated, the system is responsible for delivering the data to the settop box of the client at the required rate so that there is no interruption in service. The settop box is assumed to have little memory so that it is incumbent on the server and network to deliver the data in a "just in time" fashion.

Due to a sequential display of a movie at a home, the data can be retrieved from the available disks one block at a time and delivered to the cable network for display at a customer's television. A server typically consists of a many magnetic disks. **The key to realizing a scalable server is to distribute the load imposed by a display evenly across the available disks.** This is attained by dispersing the blocks of a movie across all the disks using a pre-defined data layout function [BGMJ94]. The pre-deterministic nature of data layout brings about a regular schedule for activation of disks in order to enable the system to maximize the number of serviced customers. When additional disks are introduced into a system, the data layout must be modified to preserve the original pre-defined function. In [GK96], we describe a technique that re-organizes the data while the system is servicing customers.

Mitra [GZS+97] is a research prototype, designed and developed at the University of Southern California, that demonstrates the feasibility of a scalable continuous media server.

3 Pipelining to Minimize Startup Latency

With a geographically distributed server, when a request references an object that does not reside on the regional server, one approach might materialize the object on the regional server drives in its entirety before initiating its display. In this case, the startup latency is determined by the allocated bandwidth (by the network, the central server, and the regional server) to materialize the referenced object and its size. With continuous media (e.g., audio, video) that require a sequential retrieval to support its display, a better alternative is to use pipelining in order to minimize the latency time [GS94, GDS95]. Briefly, the pipelining mechanism groups the blocks of a referenced object X into s logical slices ($S_{X,1}$, $S_{X,2}$, $S_{X,3}$, ..., $S_{X,s}$) such that the display time of $S_{X,1}$ ($T_{Display}(S_{X,1})$) eclipses the time required to materialize $S_{X,2}$ from the central server ($T_{Materialize}(S_{X,2})$), $T_{Display}(S_{X,2})$ eclipses $T_{Materialize}(S_{X,3})$, etc. This ensure a continuous display while reducing the latency time because the regional server initiates the display of an object once $S_{X,1}$ becomes available.

In [GDS95], we studied the pipelining mechanism for two possible scenarios: the bandwidth allocated to materialize object X is either (1) lower or (2) higher than the bandwidth required to display X. We refer the interested reader to [GDS95] for details.

4 Partial Caching

Assume that the number of customers supported by both the central server and the cable network (i.e., their aggregate throughput) is significantly larger than that of each regional server. Based on the discussions of Section 3, a regional server no longer must cache an entire copy of a presentation. Instead, it caches the first few sections of each object (say $S_{X,1}$ and $S_{X,2}$ for object X). With this approach, the regional server can provide the illusion of having a entire copy of X by initiating the display of $S_{X,1}$ and establishing a pipeline between the central server and the customer. The startup latency observed with this technique is equivalent to the scenario where object X is cached on the regional server in its entirety. When compared with a technique that caches a copy of X in its entirety, partial caching has the following advantages. First, given a fixed amount of storage, each regional server can now cache many more objects to provide its region with more content. Second, with this caching technique, the amount of storage required at each regional server is minimized, reducing cost. (As discussed in Section 1, this also enhances the fault tolerance characteristics of the regional server.) Third, it provides for a more centralized control of the entire system. Fourth, for those programs where a customer starts to display and then aborts in the middle of presentation, this paradigm saves storage space at the regional client (because the last slices are neither retrieved from the central server nor staged on the regional server).

However, partial caching suffers from a number of limitations. First, each display requires the participation of the main server. Second, pipelining utilizes the bandwidth of the cable network from the central server to a customer's home. If not managed properly, the bandwidth of either the main server or the cable network might be exhausted, reducing the number of customers that can be serviced simultaneously.

The tradeoff of this caching technique must be analyzed carefully before designing a system around it. Briefly, the results of [GS94] demonstrate that this caching technique is inappropriate when the peak overall system load exceeds the bandwidth of the central server because the central server becomes a bottleneck.

Ideally, the system should employ a hybrid caching solution: A regional server caches the frequently accessed objects in their entirety and employs partial caching for those objects whose popularity cannot be established. Partial caching can be designed to behave in this manner.

5 Conclusion

This paper proposes the deployment of scalable continuous media servers in the existing cable networks to provide on-demand service to homes. The basic skeleton provided here can be extended with a variety of services. One might envision intelligent agents that monitor the television programs viewed by each customer in order to build a profile. These profiles can be used to suggest to a customer on-demand programming that is similar to those they watch regularly.

While the technology has matured to a stage to make the ideas proposed in this short papers feasible, there are social and legal issues that require a careful study. As a simple illustration, the concept of monitoring the programming selected by users' in order to suggest similar programs is trivial. However, the users' right to privacy must also be considered where a customer might refuse the service because the system is collecting has his or her personal information. It is worthwhile to note that *computer* users are starting to become less sensitive to such privacy issue (e.g., see "As the Web Expands, So Do Surveillance Tools," New York Times, Monday, January 6, 1997, Page D5). If the Internet users are employed to model the future cable network warriors, these legal issues might not pose a significant challenge.

References

[BGMJ94] S. Berson, S. Ghandeharizadeh, R. Muntz, and X. Ju. Staggered Striping in Multimedia Information Systems. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 79–89, 1994.

[GDS95] S. Ghandeharizadeh, A. Dashti, and C. Shahabi. A Pipelining Mechanism to Minimize the Latency Time in Hierarchical Multimedia Storage Managers. *Computer Communications*, March 1995.

[GK96] S. Ghandeharizadeh and D. Kim. On-line Reorganization of Data in Scalable Continuous Media Servers. In *Seventh International Conference on Database and Expert Systems Applications*, September 1996.

[GS94] S. Ghandeharizadeh and C. Shahabi. On Multimedia Repositories, Personal Ccomputers, and Hierarchical Storage Systems. In *Proceedings of ACM Multimedia Conference*, 1994.

[GZS+97] S. Ghandeharizadeh, R. Zimmermann, W. Shi, R. Rejaie, D. Ierardi, and T. Li. Mitra: A Scalable Continuous Media Server. *Kluwer Multimedia Tools and Applications*, January 1997.

Characterization of Return Path Optical Transmitters for Enhanced Digitally Modulated Carrier Transmission Performance

John J. Kenny, Ph.D.
Principal Engineer

Antec Technology Center
Atlanta, Georgia

Abstract

Noise loading is becoming an accepted way to characterize the noise and intermodulation performance of return path active components. This paper focuses on understanding how over-driven optical transmitters degrade digitally modulated carriers. Analysis and experimental findings have uncovered several factors which influence the performance reverse path optical transmitters used for transmitting digitally modulated carriers.

BACKGROUND

Noise loading is becoming accepted as way to characterize the noise and intermodulation performance of return path active components[1]. The basis for this stems from the primarily digital transmission use of the return path and the fact that the power spectral density of digital signals resembles the power spectral density of a band of noise. In a situation where the upstream payload consists of many, similar level, digitally modulated carriers, the probability distribution function (pdf) for the amplitude of the composite signal approaches the gaussian distribution associated with thermal noise.

Noise loading, as applied to return path active components, provides guidance to the system operator in the choice of appropriate nominal operating signal levels within the dynamic range of each component. Multiple digitally modulated carriers may then be carried and allow significant tolerance in their levels. Selection of the nominal level may also take into account, and provide headroom for, unaccounted for signals such as high level ingress.

There is a long history of noise loading of frequency division multiplexed telephony coaxial cable and microwave transmission systems. Analysis of the resulting noise and distortion characteristics derived from noise loading those systems produced the thermal noise, second order intermodulation distortion and third order intermodulation behaviors as a function of RF drive level[2].

Attempting to analyze return path amplifier and laser distortion characteristics by those methods immediately shows that when the intermodulation distortion exceeds the thermal noise by just a few dB, the device is severely distorted and the intermodulation noise can not be characterized by second and/or third order intermodulation.

QPSK and QAM signals are characterized by their signal states in the "phase plane", sometimes referred to as a constellation diagram. Noise loading is a scalar measurement. It is related to how much a signaling state of a digital signal is spread, but it will not indicate the direction of that spreading and how close that spreading comes to the decision thresholds in the demodulator. This was the motivation for experimentally investigating the distortion of a signaling state when the laser is over driven.

DISTORTION MECHANISMS

There are several basic impairments in QPSK and QAM transmission that may be viewed with a phase plane or constellation diagram. The usual situation includes some or all of the following properties in the phase plane.

If a laser only experiences tone-like intermodulation products when noise loaded or stressed with high level CW interferences, the signal states would be smeared symmetrically about their undistorted locations.

If there is only compression when the laser is overloaded, the signal states will be moved radially towards the center of the constellation diagram.

The transmitted carrier can undergo a phase shift and cause a rotation of the signal state. For instance, this could occur if an over driven laser experienced excessive chirp as it clipped the RF. Chirp in conjunction with fiber dispersion can result in a carrier phase rotation.

In a laser transmitter the laser diode DC current, I_{DC}, is set by a power control loop and the AC modulating current is superimposed from an RF path. These functions are shown in the block diagram, Fig. 1.

Normal level RF carriers will have peak currents less than I_{DC} - I_{th} (see the L-I characteristic in Fig. 2) so the laser diode current is in a range such that the laser always emits light. The output light is proportional to the laser diode current above threshold, i.e. P is proportional to I - I_{th}. The ratio of the peak of an RF carrier to I_{DC} - I_{th} is called the optical modulation index (omi).

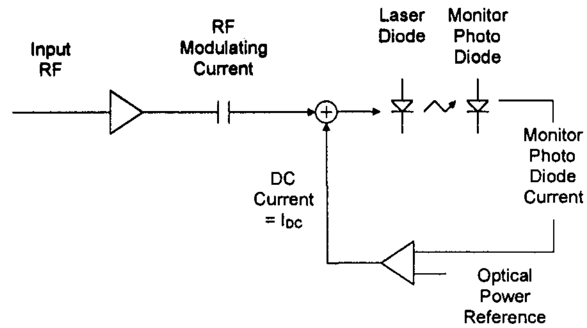


Figure 1 Return Path Optical Transmitter Block Diagram

When there is a modulating current comprised of a wanted digitally modulated carrier and a very high level out of channel signal (omi much greater than 100%), two distinct effects occur.

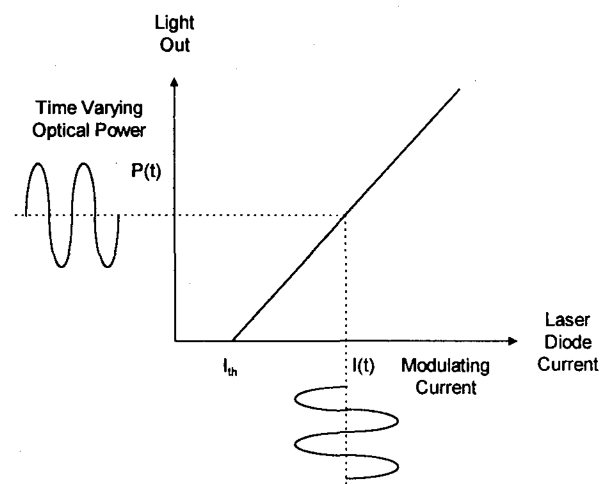


Figure 2 Laser L-I Characteristic

1. Intermodulation products are generated and some may fall into the channel producing interference.
2. The laser is turned off for part of the negative excursion of the high level interference, resulting in an amplitude compression of the wanted signal.

Generally both occur, but it is worthwhile to consider them individually.

At first one might say that in the limit of extremely high omi, the laser would be turned off for half of the time, resulting in a 6 dB RF level drop of the wanted signal. This is usually not the case. The power control loop maintains constant average output optical power. This results in a further reduction of the laser's duty cycle, as shown in Fig. 3.

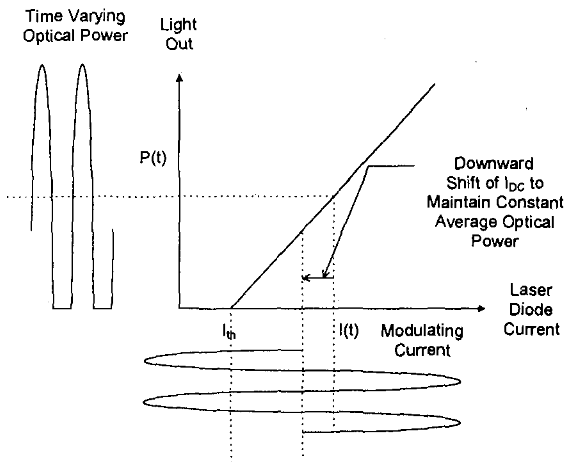


Figure 3 Laser Light Output When the OMI is Greater than 100%

EXPERIMENTAL WORK

An apparatus arrangement as shown in Fig. 4 was assembled. It permitted the flexibility to do both noise load testing and signal constellation distortion measurements.

Noise load testing

Three lasers were characterized for Carrier to Noise plus Intermodulation Noise [C/(N+IMN)] vs. RF drive level in a manner similar to that used in reference 1. Here, however, a 42 MHz band of noise was notched by using a T7 channel deletion filter. A CW carrier around 9 MHz was placed within the notch as shown in Fig 5. The level of that carrier was set to equal the noise power missing from the notch. The motivations for inserting the carrier were to provide a reference level and to estimate the potential for digital signal degradation. The noise generator and spectrum analyzer options of Fig. 4 were used.

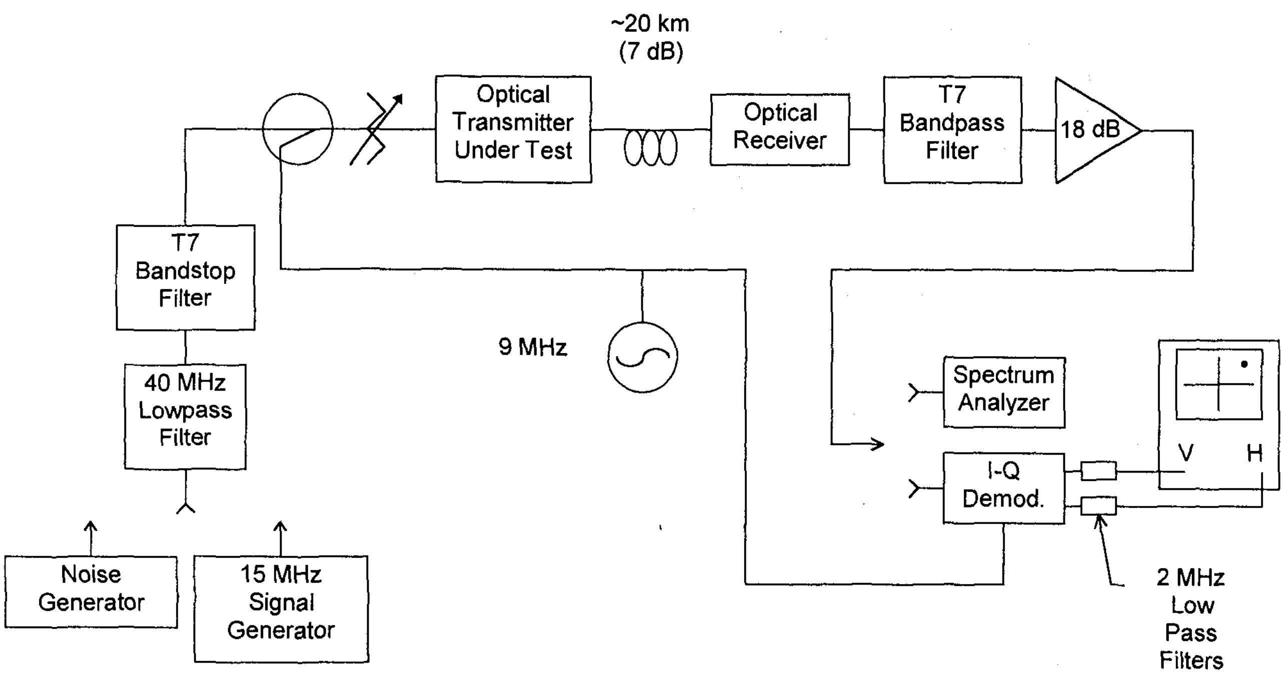


Figure 4 Experimental Test Set Up

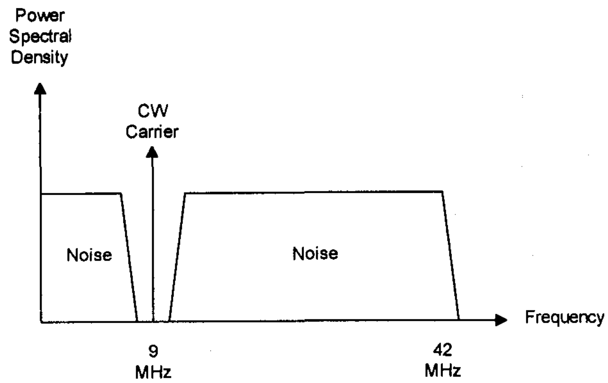


Figure 5 Notched Noise Plus CW Carrier

The three lasers (all isolated and uncooled) characterized were:

Laser	Type	Output Power
1	DFB	+4 dBm
2	DFB	+1 dBm
3	FP	0 dBm

The test results follow in Figures 6, 7 and 8. To the left of the peak C/(N+IMN), C/(N+IMN) is dominated by relative intensity noise, shot noise, optical receiver front end noise and double rayleigh backscatter noise. To the right of the peak, C/(N+IMN) is dominated by noise-like intermodulation distortion. To the left of the peak, lasers 1 and 2, both DFBs, improve C/N at a rate of about 4 dB per 3 dB increase in RF drive. This occurs because, as the RF drive to the laser is increased, not only is the received RF increased relative to various noise sources, but also the laser wavelength chirps more and the double rayleigh backscatter noise is reduced. Laser 3, a Fabry-Perot type, improves at a rate of 1 dB per 1 dB increase in RF drive.

The behavior of lasers 1 and 3 for RF drive levels to the right of the C/(N+IMN) peak show rapid decrease of C/(N+IMN) with increasing RF drive and are generally similar to those reported in reference 1. Laser 2, although it exhibits a smaller dynamic range than laser 1, degrades more gradually as the RF drive increases.

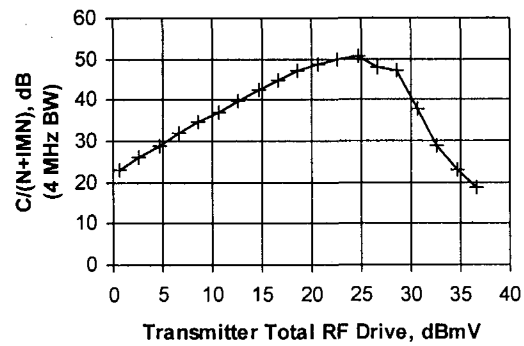


Figure 6 Laser 1
+4 dBm, Isolated, Uncooled DFB

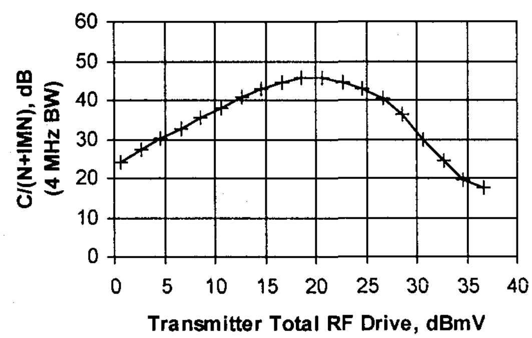


Figure 7 Laser 2
+1 dBm, Isolated, Uncooled DFB

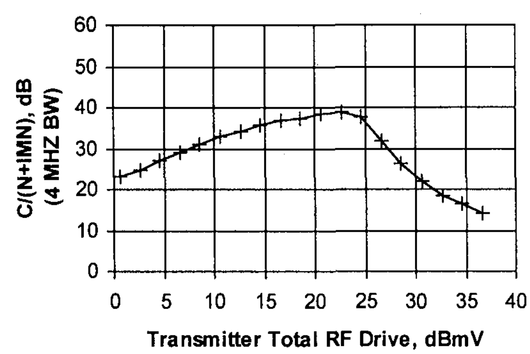


Figure 8 Laser 3
0 dBm, Isolated, Uncooled FP

Signal constellation distortion

The other configuration of Fig. 4 uses an I-Q demodulator to view the signal constellation in a manner similar to the CW Tester™ reported by Prodan[3]. It displays the location of one signaling state on an oscilloscope. It simulates the continuous transmission of the upper right hand quadrant state of QPSK or the upper right hand quadrant corner state of a QAM signal.

The frequency of the wanted carrier to be demodulated is 9 MHz. In addition there is either the notched noise as shown in Fig. 5 or a high level ~15 MHz, 100% modulated carrier, 15 dB higher in level than the wanted carrier. The exact frequency was intentionally chosen so that there were no noticeable beat products within the ±2 MHz bandwidth of the demodulator.

The signal constellation diagram is shown in Fig. 9. The signaling state is shown in the upper right quadrant. If it is being demodulated as a QPSK signal, then the decision thresholds are ideally coincident with the I and Q axes. If it is being demodulated as a 16 QAM signal, the decision thresholds are much closer. The digital signal will be correctly demodulated if the effects of noise, interference and distortion do not displace the signal state beyond the thresholds.

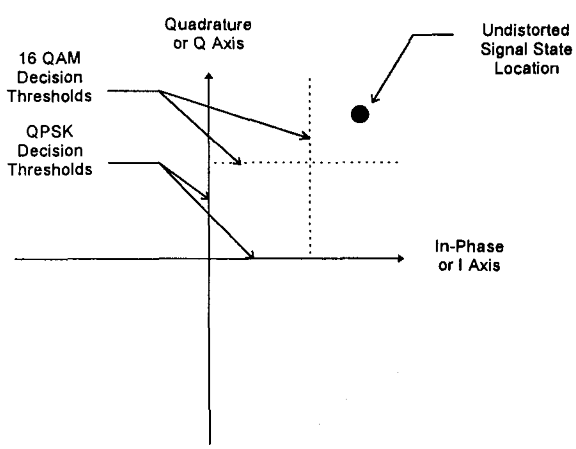


Figure 9 Signal Constellation Representation of a Signaling State Showing Decision Thresholds

When a laser is driven at low RF levels, the signaling state will be somewhat blurred, symmetrically about its intended location. The subject of interest here is what happens to the signal state when the laser is driven at very high levels. Lasers 1, 2 and 3 were driven at the level of the highest data point of Figs. 6, 7 and 8. The C/(N+IMN) was less than 20 dB in each case and dominated by intermodulation noise. An example oscilloscope photograph (Fig. 10) shows significant blurring of the signaling state. All three lasers exhibit an elliptically shaped distortion of the state with the major axis of the ellipse oriented along the diagonal line. This shows that the dominant distortions are noise-like intermodulation products and compression. No phase rotation is noticeable. The elongation of the signal state in the direction of the origin is a very favorable situation for QPSK.

The second test signal was comprised of a 9 MHz wanted carrier and a ~15 MHz, 100% modulated carrier, 15 dB higher in level. The 9 MHz carrier was set to the normal input level for these transmitters (+18 dBmV). For all three lasers, the effect of this out of channel signal is an almost identical, radial, compression of the wanted signal. An example measurement is shown in Fig. 11. The peak compression occurs at the peak of the modulated signal (6 dB above its carrier). When the level of the modulated signal is at its minimum, the signaling state is in its normal location. Further increases in the level of the CW carrier can reduce the wanted signal essentially to zero, as shown in Fig. 12.

LESSONS LEARNED

Several conclusions can be drawn from this work:

1. Signal constellation impairments during noise loading have shown no peculiar effects.

Both product generation and compressive effects are taken into account.

2. Type testing a return path optical transmitter using a constellation analyzer provides valuable insight into its capability for carrying digitally modulated carriers.

3. Smaller wavelength chirping of DFB lasers at lower RF drive levels increases the slope of the C/(N+IMD) characteristic curve.

4. Some laser types have "softer" degradation at high RF drives.

5. Lasers are extremely tolerant of QPSK transmission because laser distortions and QPSK distortion tolerance complement each other.

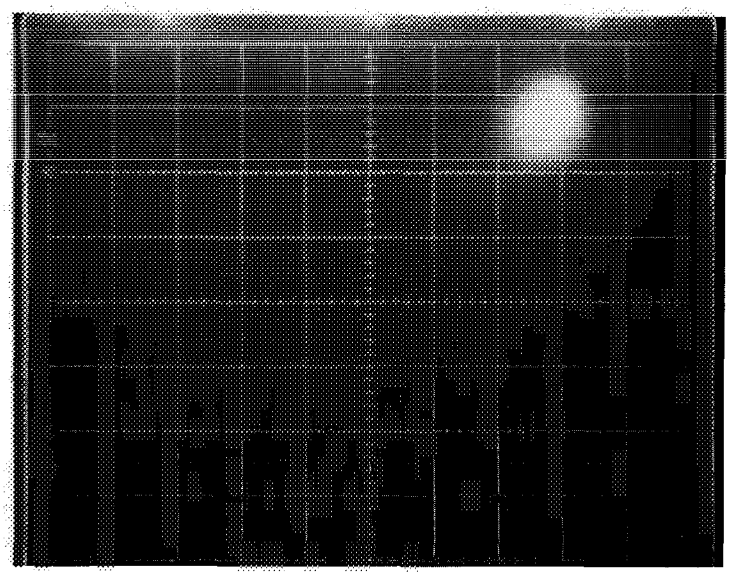


Figure 10 Signal State Spreading During Noise Load Testing at High Levels

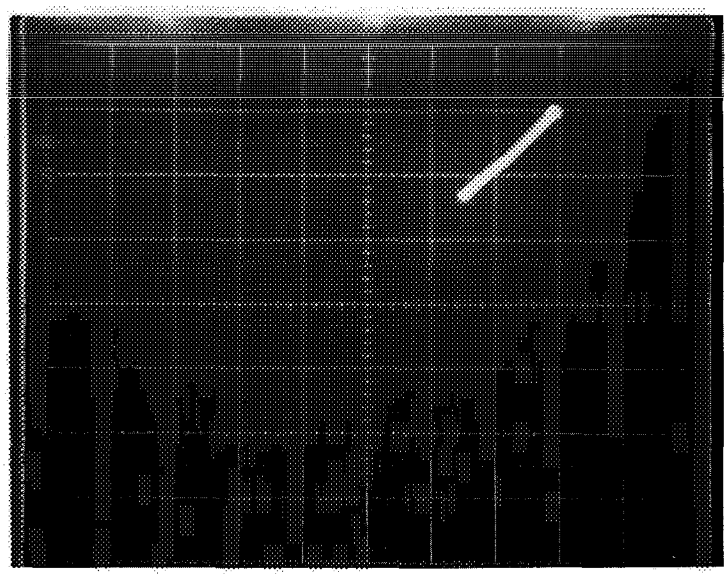


Figure 11 Signal State Compression in the Presence of a High Level 100% AM Modulated Carrier

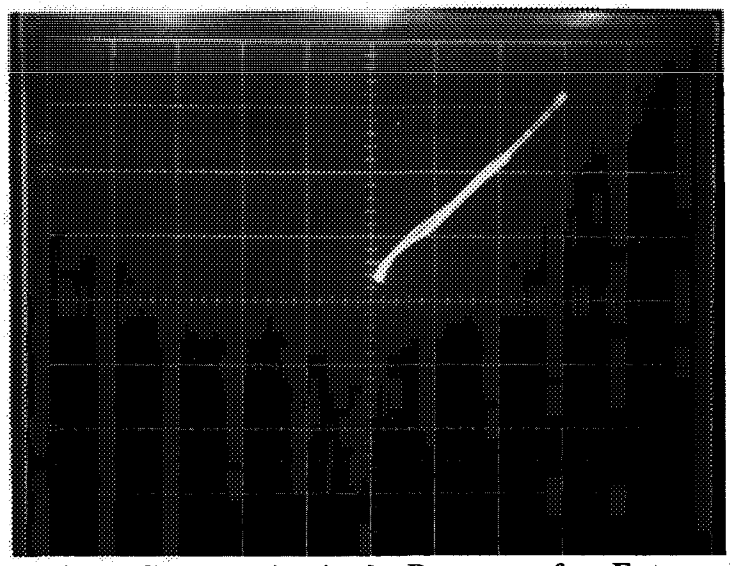


Figure 12 Signal State Compression in the Presence of an Extremely High Level 100% AM Modulated Carrier

ACKNOWLEDGMENTS

The author thanks Jim Farmer for his continuing support and the members of the SCTE Engineering Subcommittee addressing upstream transportation issues for their open and stimulating discussions. An advance copy of reference 1 sent to us by Oleh Sniezko provided valuable background information.

My thanks are also extended to James Street who expeditiously assembled the test facility and collected most of the data.

References

[1] O. Sniezko and T. Werner, "Return Path Active Components Test Methods and Performance Comparison," 1997 Conference on Emerging Technologies Proceedings Manual, January 8-10, 1997, Nashville, TN, pp. 263-294.

[2] Technical Personnel, American Telephone and Telegraph Company, Telecommunications Transmission Engineering, Volume 1 - Principles, 2nd edition, Western Electric Company, Winston-Salem, NC, 1977, pp. 450-453.

[3] R. S. Prodan, M. Chelehmal and T. Williams, "Analysis of Two-Way Cable System Transient Impairments," 1996 National Cable Television Association Technical Papers, pp. 65-78.

DATA/INTERNET SERVICE PLANNING IN THE SHADOW OF AN ANTICIPATED CABLE MODEM STANDARD

Terry L. Wright
Chief Technology Officer
Convergence Systems Incorporated (CSI)
3800 Holcomb Bridge Road, N.W. Suite 204
Norcross, Georgia 30092-2230
Ph: 770-416-9993 Fax: 770-416-9994
Web: http://www.convergence.com Email: tlwright@convergence.com

ABSTRACT

The anticipated completion of a formal international cable modem technical standard (i.e., IEEE 802.14 or a potential SCTE offering), or the potential "de facto" popularity of an MCNS specification, introduces significant strategic and operational implications for the cable operators contemplating entry into the Internet/Intranet/on-line data services market place.

INTRODUCTION

The emerging cable modem standard/ specification imposes substantial challenges to the cable industry in the areas of

- cable modem selection
- distribution channels,
- alignment (strategic fit) with planned service offerings,
- end-user satisfaction,
- subscriber mobility concerns,
- support and deployment issues,
- capital equipment depreciation considerations, and
- total solution integrity.

What changes to the existing infrastructure, and what impact on the aspects of service deployment mentioned above, can be expected once a cable modem standard/de facto specification is available? What implications do these and other related changes have on strategic planning challenges facing cable operators today? How can cable operators navigate the complex terrain between today's proprietary solutions, and the standards-based solutions expected in the future? What market dynamics and other competitive forces are defining and characterizing the timeframe for the cable industry's window of opportunity in on-line services? Just exactly what will it mean to have a cable modem standard, and how should these expectations shape strategic thinking today?

These and other related questions and issues must be thoroughly addressed if the cable industry is to effectively compete in the Internet/on-line data services marketplace. While viewed as a positive development for the cable industry over the long term, the anticipated completion of cable modem standards presents significant challenges to moving forward in the present. With formal standards-based cable modems not expected before early 1998, what can cable operators do to establish a foothold in the cyberspace market today?

This paper addresses the questions and issues mentioned above, as well as other challenges facing cable operators today, as their industry evolves towards the provisioning of high-performance Internet, IntraNet, and other broadband-based networked data services.

STANDARDS, SPECIFICATIONS & CONSORTIUMS

The primary difference between formal standards, and specifications, lie in the areas of accreditation, legal posture, and contributors. (Participants in accredited standards-making bodies are typically immune from antitrust statutes for topical work done in committee.) Of course, an informal specification can end up as a "de facto" standard due to its popularity among vendors and users (e.g., the RSA, Microsoft Windows/DOS, Unix). A formal standard is sanctioned ultimately at the United Nations level. There are a variety of standards-making organizations and study groups. Traditional standards groups exist for most communications technologies with universal application. Examples include the International Telegraph and Telephone Committee (CCITT), Institute of Electrical and Electronics Engineers, Inc. (IEEE), International Standards Organization (ISO), International Electrotechnical Commission (IEC), American National Standards Institute (ANSI), European Computer Manufacturers Association (ECMA), Electronic Industries Association (EIA), and the Internet Engineering Task Force (IETF).

There are also telecommunications and information systems industry consortiums such as the ATM Forum, Frame Relay Forum, SMDS Interest Group, MCNS, North American ISDN Users Forum, X/Open, Open Software Foundation, X-Windows Consortium, and Microsoft MAPI. Together with various testing organizations (e.g., NIST-OIW, COS, EOTC) and coordinating bodies (e.g., T1AG, ECSA), the standards-making process is accomplished. Note that there are varying degrees of cooperation and coordination among all these groups These standards-setting groups operate in a wide variety of ways from traditional formal processes (e.g., ISO-IEC, ITU, ITC, ANSI) to more aggressive, less formal, and more accommodating (and rapid) processes like the IETF. Many are also hierarchical in organization (e.g., IEEE is a study group under ANSI).

It should be noted, however, that all these study groups, accredited committees, consortiums, testing entities, and coordinating bodies interact heavily in the development of global technical standards. For example, the IEEE Working Group responsible for developing a cable modem standard has recently incorporated much of a physical layer contribution from the MCNS LP industry group.

Importantly, it should be noted that standards groups focus on primary *requirements* for a particular technology, and avoid implementation-specific issues.

CABLE DATA STANDARD

For the delivery of data services over cable/HFC networks, the ANSI organization has designated the IEEE 802 Working Group to develop an international standard. (A formal Project Authorization (PA) issued by ANSI provides the authority and direction for the Working Group's effort.) This is the same standards body that developed the Ethernet/802.3, Token Ring/802.5, Logical Link Control/802.2, Spanning Tree/802.1d, and a host of other popular world-wide standards. For development of a standard to support cable modems and related head-end equipment required for delivery of high-speed data, voice, and video over cable/HFC networks, the IEEE formed a dedicated Working Group designated as the IEEE 802.14 WG.

The IEEE 802.14 WG is well attended with more than 90 voting members (as of 1-1-97). A variety of industries and disciplines are represented on the 802.14 committee including telecommunications equipment providers, inter-exchange carriers, utilities, academia, software companies, component providers (e.g., silicon chip developers), systems integrators, and consultants. The 802.14 standard has been under development for approximately the past two (2) years.

The IEEE 802.14 WG is focused on defining the Physical componentry (termed the PHY), media access control function (termed the MAC), layer service protocols and interfaces, message formats, management interfaces, relationships to other existing standards and recommendations, security provisions, and a variety of related topography recommendations and other details required for a formal standard. The Working Group expects to produce a draft document of sufficient quality for external peer review and Letter Ballot in the summer of 1997.

WHY A STANDARD?

A well-defined standard enables a variety of benefits for a variety of entities. The primary benefits of such a standard are

- ubiquitous product application in standards-compliant environments (i.e., product interoperability),
- lower equipment costs due to standards-compliant componentry and volume,
- end-user mobility, and
- expanded product availability through distribution channels.

These benefits can be readily seen in the deployed information infrastructures of most commercial and institutional organizations and end-user environments. For example, there are upwards of a million or so Ethernet LAN implementations deployed globally, millions of Unix-based servers, countless personal computers running some version of Windows and its evolving forms, and millions of Hayes-Command Set compliant modems connecting these PCs to the Internet and other online services and corporate networks. These environments are rich in either formal or de facto standards-compliant equipment and services. It benefits of standardization have been well-demonstrated to date, and it is easy to understand how the IEEE 802.14 standard will benefit the cable/broadband network service provider industry.

Among the primary benefits listed above, however, the last bullet describing *"expanded product availability through distribution channels"*, introduces an interesting complexity into the data/Internet services deployment agenda for the cable/broadband network services industry. It also provides a good segway into the data/Internet services *planning in the shadow of a cable modem standard* thrust of this paper.

IMPLICATIONS OF A CABLE MODEM STANDARD FOR THE CABLE INDUSTRY

As has been the case with other industries deploying standards-based information and telecommunications infrastructure, the cable industry will no doubt enjoy all the benefits of standardization once a cable modem and related equipment standard matures and is supported by product vendors and the cable industry. The extent and timing of when these benefits will be realized is a function of many dynamics including

- manufacturing and volume economics associated with standardized components,
- equipment ramp-up schedules and production capacity of equipment vendors,
- cable operator deployment schedules, and
- market demand for cable's data service offerings.

However, there are at least two aspects of the arrival of a supported cable modem standard that warrant further examination. These are explored below.

THE SERVICE PORTAL PERSPECTIVE

As mentioned above, the expanded product availability benefit (through distribution channels), associated with the emergence of a cable modem standard, warrants further examination. It is not difficult to understand why, until recent years, the telephone industry preferred to "rent" their telephones to customers instead of supporting customer-owned telephones. By controlling the end-user service device, the telephone companies controlled the nature of the service being delivered. That the telephone companies (the old Bell system) held onto this control for so long is understandable. With the FCC's 1982 Carterfone decision and other related legislation (Computer II), this control was obviated in the spirit of competitive benefits and service evolution.

The cable industry's entry into data/Internet services poses a similar decision with respect to control of the *service portal.* Absent any legislation to the contrary, the cable industry is free to determine the nature of the data services it offers on its systems, including the issue of *cable modem ownership.* It is not a trivial decision in that it impacts

- data service business models,
- data service pricing,
- data service portfolios and service classes,
- and business strategies.

In addition, when one considers the underlying historical factors behind the dominance of television as the entertainment and news medium of choice (e.g., user preference for audio *and video* vs *audio-only* for delivery of news and entertainment), it would appear that cable has an extraordinary opportunity to redefine the service portal for information services (away from telephone modems to cable modems/Internet appliances). This is due to the growing popularity and development of multimedia and multmodal applications (e.g., virtual reality and immersive data visualization enviroments) versus traditional text. The decision of cable modem ownership (e.g., provided by cable operators or user-owned through retail distribution) could significantly impact the cable industry's ability to take advantage of this strategic opportunity.

In a practical sense, the issue of cable modem ownership is further obscured by implications explored below.

BELLS AND WHISTLES

For all the benefits of technology standardization, the nature of competition and capitalism imposes a complex wrinkle into an otherwise straightforward environment.

Once a standard has been developed, formally approved, and is supported by equipment producers and consumers, product vendors nearly always add their own proprietary features to a standards-based product in order to create and enjoy some form of competitive advantage. These whistles and bells, as they are generally referred to, involve functionality generally considered "implementation specific" or are beyond what is considered to be core technical requirements, and hence, are not addressed in formal standards. This creates certain challenges to cable operators deploying data services in several areas. These challenges are best illustrated through examination of deployment scenarios very likely to emerge once a cable modem standard takes root in the industry.

Consider the scenario where CableCo A decides to offer high-speed Internet and other data services to its subscribers. For the sake of example only, let's assume that these services target residential-level subscribers only. This scenario occurs at a time after a cable modem standard has been adopted, and a variety of cable modem vendors offer products compliant with this standard. Let's further assume that each cable modem vendor has also incorporated special (non-standard, proprietary) features and functions it believes will positively differentiate its product offering.

Next, let's assume that CableCo A has opted to support subscriber-owned cable modems. Let's now make the assumption that CableCo A's available spectrum for data services is limited such that all data services will occupy the same forward and return channels.

Finally, let's assume that CableCo A has selected a standards-compliant cable modem preferred platform it believes will facilitate deployment of the highest value, most robust, most manageable, and most appealing data/Internet services. CableCo A expects to leverage certain custom features inherent in its platform choice that are unavailable with any other cable modem platforms. (Adoption of a preferred cable modem platform by cable operators could be quite commonplace as they are compelled to maintain a service capability posture that will allow them to maintain service parity with, and hopefully, a competitive advantage over, the local telco down the street.) CableCo A has launched a respectable PR and advertising touting all the neat features and functions its service will offer. With this scenario in place, let's examine some of the issues CableCo A might encounter.

Since the local computer retailers distribute only standards-compliant modems, CableCo A is comfortable with its ability to accommodate all service requests regardless of the cable modem acquired by the subscriber. However, Joe Bithead, who didn't read the fine print on the televised ad for the services, enthusiastically subscribes to the data service, expecting to try out all those neat features as soon as possible. Joe Bithead runs right out to Computer City or other local retailer and buys the least expensive cable modem available. (Remember, they are all standards compliant.) CableCo A's first customer service call for its new data services is likely going to be Joe Bithead. Why? Because the virtual reality-oriented service that really caught Joe's eye to begin with, is only available via the preferred cable modem platform adopted by CableCo A.

The second customer service call is again from Joe Bithead. This occurred when Joe tried to connect his cable modem to his new Ethernet hub in order to connect his kid's older PC to the Internet. Joe didn't realize that multiple IP numbers are only available through the preferred platform (support for a single IP might be all the standard required).

So Joe Bithead, disillusioned and frustrated, returns his low-cost cable modem to the retailer, and exchanges it for the CableCo A preferred platform modem.

The third customer service call for the new data services is, again, from Joe Bithead. Joe is just a little irate since he still can't make the gee-whiz virtual environment application work on his PC. Again, Joe didn't read the fine print of the advertisement, and so didn't realize that this service was only available with the standard Client Kit CableCo A offers its customers. Joe assumed, since he'd been on the Internet for years via his dialup modem, that he'd have access to all the services he'd seen advertised. Joe further assumed, since he already had a Freeware copy of "*Gee Whiz*" that he'd been running for years, he could use it for the service. (Joe hadn't realized that, in order to leverage its superior network access capacity, CableCo A has contracted with a commercial software group to tweak the *Gee Whiz* package accordingly.)

As it turns out, Joe needed special TCP/IP stack software anyway that interacted appropriately with CableCo A's custom server stack in order to utilize the gee-whiz virtual reality application due to the need for specialized RSVP support passed on by the headend server due to some peculiar interaction required with the cable modem's headend equipment. Even worse, Joe Bithead hadn't noticed that he needed to acquire the more expensive service class offering as well in order to utilize the gee-whiz application.

The first customer service call for *CableCo B* is from Joe Bithead's friend Bob Supertechie. Bob had recently moved into the area, but into a different franchise area operated by CableCo B. Bob had listened to Joe's story, and went out to the same retailer to get his cable modem. Bob had carefully listed out all the components and service classes he needed from Joe to get the most from his new high-speed Internet service.

As it turns out, CableCo B had selected a different preferred cable modem platform than CableCo A. The ads were very similar, so Bob just assumed the related service environment was the same. After all, they're both cable companies, and support virtually identical programming lineups. So Bob exchanges his modem for the preferred one, and is once again dismayed because he can't interact with Joe in the *Gee Whiz* virtual reality environment. As it turns out, CableCo A offers this application only as a local service to cable modem users; it did not have a sufficient Internet feed capacity to support the bandwidth requirements of an Internet-based *Gee Whiz* application. Unfortunately, neither Joe nor Bob, nor CableCo A, nor Cable B, realize that even had both subscribers done everything right to begin with (and both CableCo's supported the Internet-based version of "*Gee Whiz* virtual reality") Joe and Bob would still have been ultimately frustrated when Joe discovered that he didn't have enough RAM in his PC to support "*Gee Whiz*", and Bob's Client Kit-provided TCP/IP stack wouldn't support the application until the next release currently scheduled for late summer.

SO WHAT'S THE POINT?

A better question is "what are the points"? One key point is that, although the above scenario is purely hypothetical, it is not beyond the realm of what cable operators can expect to deal with if they intend to be serious data/Internet services players.

A second major point is the coordination among distributors of cable modems, and cable operators providing data services, that will be required to prevent much of the confusion illustrated in our scenario. (Retailers are not accustomed to asking customers what cable franchise they reside in, nor are they used to providing custom Client Kits consistent with customer addresses.

A third crucial point is that the cable modem ownership issue is a key factor in this situation, at least until such time that cable-based data service delivery matures to the point of blending into the background telecommunications infrastructure. (Cableco-provided cable modems and client kits equate to a more controllable service delivery environment that is more easily managed and supported, and translates into fewer customer surprises.) It may make sense, if supported by service provisioning economics, for cable operators to retain control of the cable modem (at least in the near term).

A fourth critical point is the risk of relying on custom (non-standard) product features as a basis of the service portfolio and related competitive advantages. Competitive dynamics in high-speed data services may leave cable operators little choice but to take his risk. Given solutions to the issues shown in our scenario may lessen this risk significantly.

Finally, it is clear from our hypothetical scenario above that customer expectations need to be very well managed. This requirement is inconsistent with traditional entertainment service models as most subscribers expect and receive television programming; the complexity of what is being delivered has not warranted special attention in this area other than in service packaging.

CONCLUSIONS

This brings us back to the cable modem standard, and the data/Internet service planning occurring as the standard-making effort is progressing. Probably the most critical point to be made from our scenario exercise above is that the emergence and support of a cable modem standard does not eliminate the cable operator's need to be diligent in all the related areas of providing a data/Internet service. (Engaging a reputable, and importantly, *cable-experienced* systems integrator or similar partner may be a wise move to facilitate service planning and deployment.) As market forces now suggest, the only thing riding on the cable industry's data/Internet services agenda is the continued and future viability of the cable industry

So what about the modems available *now*, and the anticipated standard most won't be compatible with in the future? The alternatives are few:

1. Delay entry into the data/Internet services space (while managing market perceptions of your intent, and no doubt missing significant near-term opportunity

2. Limit deployment to selected controllable subscriber areas, and leverage the experience to enhance, acquire, and align internal capabilities as necessary to support these advanced services

3. Acquire "soft" modems that can be upgraded in the field to support the standard (let the author know if you find any where its manufacturer is willing to commit to this on paper)

4. Wait until the standard is approved, and standards-compliant products are available, before deploying.

Alternative 2 above, augmented by a well-conceived and well-managed PR campaign (e.g., tell the market what your doing in order to serve their needs best), seems a viable choice for the moment.

SUMMARY

This paper has tried to illuminate some of the underlying and important issues associated with data/Internet service planning and deployment in the shadow of a cable modem standard. It by no means pretends to offer a definitive assessment of these complexities, but does attempt to identify the nature of those complexities, and potential problems associated with them, as place-holder for further examination and research by cable operators.

The primary message this paper hopes to deliver to the cable community is that the emergence of a cable modem standard represents more of a beginning, than a culmination of something long awaited. While the standard will certainly enable the benefits discussed herein for the industry, this paper has attempted to show that the realization of some of these benefits may have ramifications on business issues and strategies that are not readily apparent to the unsuspecting.

Cable modems are but one aspect of successfully provisioning sophisticated high-speed data/Internet services. Fortunately, there is an emerging cable modem standard that should enable many of the anticipated benefits, and many of the other functions and services required for high-speed data services delivery already abide by various formal and de facto standards. In addition there are specialized resources available that can help cable move forward with its data/Internet services agenda.

However, many of the most exciting aspects of high-speed data/Internet services involve freshly-charted territory the standards process has yet to embrace, and perhaps never will. In addition, much of the market's attraction to the Internet has to do with its constantly changing capabilities and the dynamically-evolving information frontier it represents. A cable modem standard is a significant step forward and should be viewed as such. It will not, however, obviate cable's need to embrace all the other business and technical aspects of high-speed data services. The emerging cable modem standard should be viewed initially for what it truly represents: an acknowledgment by the traditional technical standards-making bodies that cable has arrived on the data/Internet services scene, and is going to be a player.

Digital Data In Analog Signals

Walter S. Ciciora, Ph.D.
Technology Consultant
Southport Connecticut
WCICIORA@AOL.COM

Abstract

The NTSC television system was designed primarily to make television receivers affordable in the era of vacuum tube circuits. In the age of digital and analog integrated circuits, much more aggressive signal processing is possible. This allows some of the signal inefficiencies accepted when NTSC was created to be overcome. It is now possible to add significant amounts of digital data into an analog signal in a compatible manner. The FCC has authorized the use of several of these systems and has encouraged the development of others.

Some of these systems simply use the Vertical Blanking Interval, VBI. Others add additional carriers. At least one system takes over portions of the video screen and adds digital data in its place.

This paper will review the methods proposed for compatible data insertion *into NTSC signals and explore possible applications of this approach.*

INTRODUCTION

Ancillary signals are additional signals which may or may not have anything to do with the video program content. Ancillary signals are a "bonus" which allows additional information and services to be provided. These signals were initially impractical due to the expense of early receiving circuits. Ancillary signals are possible because the spectrum utilization of the television signal is relatively inefficient. The goal of the designers of the television system, the National Television Systems Committee (NTSC) was to facilitate affordable consumer television receivers while maximizing the number of channels possible. This, of course, is a difficult trade off. A consequence of this difficult challenge is that there is spectrum capacity to harvest now that more complex electronics can be brought to bear.

The Data Application

Data transmission was the initial application for ancillary signals because data rates which fit into the under utilized spectrum spaces could produce useful results while not placing unrealistic demands on circuit design.

One of the first ancillary signal applications was Closed Captioning for the hearing impaired. The conservative design that characterized consumer electronics was extended to the captioning system. A very low data rate was employed. Fortunately, that is all that is required for this application.

Teletext is a more aggressive data application which puts text, graphics, and even photographs on the television screen. Additionally, Teletext can "download" software to computers either internal or external to the television receiver. While Teletext has failed in the U.S., it has enjoyed considerably greater reception in Europe.

The successors to Teletext are the systems currently becoming available for access to the Internet using television receivers or set top boxes. In other related applications, the signal directly feeds a computer. Because of the massive computing power available, much more interesting images can be delivered. Personal computers are now available with one or more tuners built in.

The Video Application

Recent developments in the television industries have focused upon the transmission of High Definition Television (HDTV) which requires a substantial increase in transmitted information and hence would greatly expand the required analog video signal bandwidth. This is unacceptable. Digital approaches are the solution. Great progress has been made in the area of digital TV bandwidth compression and a national standard has been selected by the Federal Communications Commission in December 1996. These HDTV developments have, by a combination of techniques, substantially reduced the bandwidth required for fully digital transmission of the video information. For instance, a single HDTV channel can be transmitted within the analog Broadcast TV channel assignment of 6 MHz rather than the tens of megahertz once thought necessary. In the case of cable's well-behaved spectrum, double the data transmission rate is possible. Two HDTV signals can be carried in 6 MHz.

This same technology which makes HDTV in 6 MHz possible allows multiple standard definition digital signals to be stuffed into 6 MHz. Movies have several advantages over video in this regard. Movies have twenty five frames per second versus video's thirty. This alone is a twenty five percent reduction in data requirements. Movies have the further significant advantage that they can be processed iteratively. That is, the movie is run through the processor several times with adjustment of the processor made to minimize the creation of artifacts on a scene by scene basis. Very good results have been obtained with movies at data rates of 3.0 Mbps. Quite acceptable results have been seen at 1.5 Mbps. When compared to the video obtained from a commercially recorded VHS cassette, the digital results have some advantages. Since the HDTV transmission rate is around 19 Mbps (in 6 MHz), six 3.0 Mbps movies can be carried in the same spectrum. At 1.5 Mbps, double that number is possible. Since cable has a more controlled spectrum, it can approximately further double these numbers leading to perhaps twenty four movies in 6 MHz. This is even more practical in systems that use statistical multiplexing.

The development of HDTV and its acceptance as a future broadcast standard has led to the requirement for a transition period between broadcasting the present analog TV to that of compressed digital HDTV.

It is quite likely that digital transmission of the current NTSC (standard definition) television signals will also become attractive for some applications, particularly in the transition period. Digitally compressed NTSC is now being delivered in several media with bandwidths much less than the traditional 6 MHz per channel. Since the transmission of standard analog NTSC will remain for many years before complete transition to digital high definition, the availability of a technique allowing simultaneous, non-interfering transmission of digitized NTSC-resolution signal(s) within the same spectrum as an analog NTSC signal could result in a two (or more) times expansion of channel capacity in the existing broadcast frequency assignments. If more efficient means of bandwidth compression emerge, even the transmission of HDTV simultaneously with analog NTSC is an attractive, though distant, possibility.

VBI TECHNIQUES

The original display device used in television receivers was the Cathode Ray Tube, CRT, which uses an electron beam to stimulate a phosphor coating on the inside face of a vacuum tube. This vacuum tube is appropriately called a picture tube. The electron beam is scanned horizontally to form lines and vertically to paint a complete image. The strength of the electron beam is in inverse proportion to the strength of the television transmitter power and regulates the amount of brightness in the picture. The deflection of the electron beam can be accomplished by electrostatic forces or magnetic forces. Most television display devices used magnetic deflection. Magnetic deflection circuits require time to move the electron beam back to the left side of the screen. During this time, the electron beam must be turned off or blanked to prevent unintended stimulation of the phosphor screen and the resulting interfering light. The period of time during which the electron beam is turned off is called the "horizontal blanking interval." When the electron beam reaches the bottom of the screen, it must be returned to the top of the screen to continue the process of making pictures. Just as in the horizontal case, the electron beam must be blanked to prevent disturbing light patterns on the screen. This period is called the Vertical Blanking Interval (VBI). The VBI is much longer than the horizontal blanking interval. The combination of the two blanking intervals constitutes around twenty five percent of the total time. This time cannot be used to convey pictures, but it can be applied for other useful purposes.

Closed Captioning

The first United States attempt to use the VBI for ancillary purposes was in 1970 when the National Bureau of Standards (NBS) proposed to use it for the distribution of precise time information nationwide. The ABC television network was a partner in that effort. While this initiative did not result in a service, ABC recommended a captioning service for the hearing impaired. The First National Conference on Television for the Hearing Impaired met in Nashville, Tennessee in 1971. This was followed with a demonstration by the NBS and ABC at Gallauded College in early 1972. In 1973, the engineering department of the Public Broadcasting System (PBS) initiated development funded by the department of Health, Education and Welfare (HEW). As a result of this work, the FCC reserved line 21 of field one of the television signal for the transmission of closed captions in the United States in 1976. In 1979, the National Captioning Institute (NCI) was founded to caption programming and to further the cause of captioning. In the early 1980s, Sears Roebuck stores carried a captioning decoder in set top box configuration selling for about $250. In 1989, NCI contracted for ITT Semiconductor Corporation to develop a cost effective caption decoder microchip for use in television receivers. In 1990, Congress passed the Television Decoder Circuitry Act mandating new television receivers of thirteen inch diagonal display measure or greater to include caption decoding circuits after July 1, 1993. Approximately twenty million television receivers per year are covered by this requirement. 1992 saw NCI work with the FCC and the Electronic Industries Association (EIA) to develop captioning technical standards. The 1996 Telecommunications Act requires the FCC to institute rules requiring closed captioning on video programming but allowing exemptions for programming that would suffer an "undue burden".

The Closed Captioning (CC) system is called "closed" because it is turned "on" or "off" depending on who is using the television receiver. Those without hearing impairments and who understand the spoken words need not be disturbed by text on their screens. The CC system uses very low speed data in order to minimize the impact of transmission path problems such as reflections and interfering signals. The data rate for the CC systems is 503.5 Kbps of binary (two level) data. This allows only two eight bit characters to be transmitted per VBI line. If only field one is used, there will be thirty of these lines per second. This yields 480 bps or 3,600 characters per minute. If the average word is five characters long with a space, then 600 words can be conveyed per minute. The rest of the line is occupied with both a burst of seven sine wave cycles of 503.5 KHz clock run-in and a unique "start bits" pattern placed at the beginning of the line. These signals synchronize the detector circuitry. Since only Line 21 is protected for captioning by FCC rule, the rate of transmission is slow, but perfectly adequate for the purpose. The on-screen display consists of a maximum of fifteen rows of thirty two characters each. The captions usually appear only on rows one through four and rows twelve through fifteen. The middle rows are usually transparent to show the action. A text mode provides scrolling text. Further details can be found as part of the Electronic Industries Association (EIA) standard number EIA-608.

The closed captioning signal carries four components. There are two captioning "channels" and two text channels. The first captioning channel is synchronized to the video programming so that the words carefully match the video. The second captioning channel is not synchronized.

The EIA filed a petition with the FCC to expand the captioning standard EIA-608, to allow use of line 21 field 2. This adds two more captioning channels and two more text channels. A fifth channel has been added to carry Extended Data Services (EDS). EDS will carry a wide variety of additional information. Finally, the original intent of the NBS will be realized. Precise time information will be transmitted to set VCR clocks (and other clocks as well). The channel's name and call letters are included along with current program information such as title, length, rating, elapsed time, types of audio services and captioning services and intended aspect ratio. Also included is the data for the "V-chip" which is intended to control children's access to programming. Public service announcements such as weather and emergency advisories are also transmitted. Cable system channel layout information is provided so that the channel number indicator can use the more familiar channel identification number rather than the number associated with the frequency utilized. This facility will bring the same "channel mapping" benefits subscribers have enjoyed in their cable set top terminals to consumer electronic products.

Teletext

Teletext is a more aggressive form of data transmission which has been successful in Europe, but has failed to enjoy commercialization in the US. Teletext originated in Great Britain with experimental transmission commencing in 1972. The British Broadcasting Corporation (BBC) branded their service "Ceefax" while the Independent Broadcast Authority (IBA) called their service "Oracle". France developed a packet based Teletext system called Antiope based on a transmission system called Didon. Later Canada developed another system called "Telidon" which featured higher resolution graphics. The Japanese system, called "Captain" featured "photographic coding" to accommodate the Chinese Kanji characters and the Japanese Kana set.

There are a number of reasons for the difficulties in the US. Principle among these was the failure to find a strategy which made money. Without this, the system could not be supported. Additional difficulties included the high cost of memory at the time of implementation. While a Teletext page requires only about a kilobyte of storage, that small amount of memory was expensive then. Further problems centered around the quality of the graphics. The less expensive World System Teletext (WST) had crude "Lego-style" graphics in its basic form. The other contender, the North America Presentation Layer Protocol System (NAPLPS) used a higher resolution graphics system which painfully painted itself on the screen resulting in excessively long delays which tried the patience of the user. Still another complication was the FCC's 1983 decision to allow two standards and a marketplace resolution of the winner. One of the systems is based on the British approach and is called World System Teletext (WST) and the other is the NAPLPS evolution of Antiope, Telidon, and efforts by AT&T. A final problem was reliability of data reception. In a test in the Bay area of San Francisco, only about twenty five percent of installations of the NAPLPS system were trouble free. The remainder suffered from various degrees of multi-path impairment.

Both U.S. Teletext systems have a data rate of 5.727272 Mbps which is 364 times the horizontal rate and 8/5 of the color subcarrier. The data signal has a Non Return to Zero (NRZ) binary format. The WST data line consists of eight cycles of clock run-in (sixteen bits) followed by a unique eight bit "framing code" followed by sixteen bits of control codes and a payload of forty eight-bit display words. Since the page format is forty characters by twenty rows with an additional "header row", twenty one field lines are required per page of WST Teletext. The payload of 320 bits per line allocated means that if one VBI line in each field is allocated, a data rate of 320 x 2 x 30 = 19,200 bps is obtained. Ten lines of VBI are possible (Line 21 is reserved for captions and the first nine lines form the vertical synchronization pulses) yielding a maximum of 192 Kbps for full VBI utilization.

The WST system maps the data location in the VBI line to memory locations and to screen locations and always puts data in the same memory place. This allows for a very simple error protection scheme. Received data is compared with data already in memory. If it agrees, confidence builds that it is correct. It is possible to use a "voting" approach to obtain very robust transmission.

Packetized Teletext

The fundamental difference between the WST and the evolving set of Antiope, Telidon, and NAPLPS systems is that the latter systems all used a packet structure. They have been characterized as asynchronous because there is no mapping between the transmission scheme and memory and screen locations.

PBS has developed a packetized data delivery system based on Teletext called the "PBS National Datacast Network". The standard Teletext data rate of 5.72 Mbps is used yielding 9600 baud per VBI line allocated per field. The Datacast network distributes the same signal nationally. The goal is to generate revenue to help support the PBS network. There are a wide variety of commercial applications for this signal. Currently, the StarSight Electronic Program Guide (EPG) signal is distributed via PBS.

WavePhore

WavePhore utilizes lines 10 through 20 in each field for a data speed of up to 150 Kbps. WavePhore added substantial error detection and protection bits to its structure to protect against multipath and other transmission problems.

SUB-VIDEO TECHNIQUES

There are under-utilized portions of the NTSC spectrum which can be employed to "hide" data. In many cases, the process of hiding the data in incomplete and results in artifacts under certain conditions. In other cases, the preparation of the NTSC signal so that hiding is more effective, by itself reduces video quality. So the challenge is to both hide the data and not impair video quality while retaining signal robustness and the potential for an economic implementation.

The National Data Broadcasting Committee (NDBC) was formed in 1993 to establish a single standard for data transmission in video. It issued a request for proposals (RFP) and narrowed down the selection process to two contenders: WavePhore and Digideck. Laboratory tests were conducted by the Advanced Television Test Center (ATTC) in Alexandria Virginia in December 1994. In April of 1995, the NDBC selected Digideck for field testing. In June, WavePhore convinced the committee to re-test their system after improvements made based on the results of the lab tests.

Meanwhile, the FCC issued a Notice of Proposed Rulemaking (NPRM) in April of 1995. On June 28, 1996, the Federal Communications Commission (FCC or the Commission) approved digital data transmission in the video portion of broadcast television transmission in its Report & Order (R & O), "Digital Data Transmission Within the Video Portion of Television Broadcast Station Transmissions", in MM docket No. 95-42. This R & O amends FCC rules to allow ancillary data *within* the video portion of the NTSC signal in four formats. Two of the formats, by Yes! Entertainment Corporation and A. C. Nielsen Co. place low data rate signals in the overscan region of the picture. The other two systems, Digideck and WavePhore, embed the digital signal into the video signal. Both Digideck and WavePhore participate in the voluntary standards committee titled The National Data Broadcasting Committee (NDBC) sponsored by the National Association of Broadcasters (NAB) and the Consumer Electronics Manufacturers Association (CEMA). NDBC has conducted field tests of these systems in Washington, D. C. on WETA, channel 26 and WJLA, channel 7.

WavePhore

WavePhore encodes about 300 Kbps into the baseband portion of the video.

The WavePhore system begins by reducing video luminance and chrominance bandwidths. The "luminance" is reduced from its theoretical value of 4.2 MHz to 3.9 MHz and the upper sideband of the color signal is reduced by about 300 KHz. It is then possible to insert a data signal in this region at a carrier frequency of approximately 4.197 MHz above the video carrier and a strength approximately 20 dB above the noise floor of the video system. The data is synchronous with the video carrier and thus with the horizontal line frequency. As an odd multiple of one-quarter the horizontal scan frequency, it interleaves between the luminance and chrominance bundles of spectral energy. Data is not sent during the vertical and horizontal blanking intervals. Thirty bits of data are sent per video line. There are 240 available lines per field (not counting the VBI during which the signal is blanked). This yields a raw data rate of 431.6 Kbps. After error correction coding, the raw date rate is reduced to approximately the T1 rate divided by four or 384 kb/s. This one quarter the telephone T1 data rate and so WavePhore calls their system TVT1/4.

WavePhore shuffles the data before applying bi-phase modulation and filtering out the lower sideband. Shuffling the data reduces its visibility in the video. An adaptive equalizer is used in the receiver. A major advantage of the WavePhore approach is that once inserted into the video, it can be conveyed through and video path without giving it further attention. The WavePhore VBI system and the WavePhore sub-video system can be combined to provide over 500 Kbps.

There is some degradation of pictures using this system. However, it appears that the regulating body, the Federal Communications Commission, is willing to let the broadcaster determine what his individual marketplace values and to respond to that decision.

Digideck

The Digideck system adds a Differential Quadrature Phase Shift Key

(DQPSK) signal carrying about 500 Kbps. placed one MHz below the video carrier. In this regard, it is similar to the European NICAM system for adding digital audio to analog television broadcasts. This places the new carrier in the Vestigial Side Band (VSB) region of the signal. To accommodate this, the lower VSB slope is increased. Rather than starting at the traditional 750 KHz below picture carrier, in the Digideck system, it starts 500 KHz and drops more rapidly. The carrier is about 36 dB below peak power and has a raw capacity of 700 Kbps. Forward error correction and other overhead burdens reduce the data capacity to around 500 Kbps. Digideck calls the new carrier the "D-Channel". The data signal is clocked synchronously to the television signal for ease of recovery and for better hiding in the video.

The Digideck receiver also depends on an adaptive equalizer. A consequence of the D-Channel is that it must be inserted at the transmitter site and brought there by an alternate path.

Like the WavePhore system, Digideck introduces some artifacts. A marketplace approach will allow the Broadcaster to determine acceptability.

OVERSCAN TECHNIQUES

The Yes! Entertainment Corporation's system introduces a pulse in the video between 9.1 and 10.36 microseconds after the start of the horizontal synchronization pulse. The data rate is very low, about 14 Kbps. Its application is to deliver audio to a talking toy teddy bear! A.C. Nielsen uses line 22 of one field of the video for transmitting a program source identification. This ID is used to measure the viewing population for statistical purposes. A fifth system, by En Technology was denied permission at the time of the R & O. This system allowed data to extend from the VBI into all areas of the picture with the image being constrained to a variable size box surrounded by the "snow" caused by the data. This system was judged too intrusive.

APPLICATIONS

There are a variety of applications for data in analog signals. Data can be supplied as just data. Alternatively, if sufficient capacity is available, data can be used to deliver digital video or digital audio services. The data can be used with personal computers, special television sets or set top boxes or versions of the "net computer".

The January 1997 Winter Consumer Electronics Show in Las Vegas was dominated by two developments: the Digital Video Disk (DVD) and the World Wide Web on television set top boxes. This latter application has a great deal of equipment manufacturer excitement associated with it. Time will tell whether the marketplace catches the same degree of excitement.

A related application involves Datacasting. This is the inclusion of data in the broadcast television signal for use with a personal computer. The most aggressive such implementation is Intercast whose main partners include Intel and NBC. HyperText Markup Language (HTML) formatted Web pages are delivered in the VBI of the television signal. HTML is a method of linking information. Highlighted words or phrases can be "clicked on" and relevant information appears on the screen. In some cases, this is achieved by going to another location in the same document. In still other cases, data from another document is displayed. In other cases, locations on the World Wide Web are automatically accessed and information retrieved. A personal computer with a television tuner receives the signals and displays the video in a small window. The rest of the screen displays the HTML pages. The computer's hard disk can capture and store pages of interest. Since HTML pages are around 50 Kbytes each and most personal computers now come with at least a 1 GB hard drive, capacity is not a problem. In a major application of this technology, the pages downloaded pertain closely to the video programming. When several hundred pages are downloaded, the access speed is governed by the hard disk, not a modem. In affect, the server is built into the personal computer! The HTML nature of the pages makes accessing different parts of the data downloaded easy and familiar to any Web surfer. The HTML can include embedded links to related Web sites accessed with the computer's regular phone or cable modem. Access to these sites is automatic.

CABLE VS BROADCAST DATA

Since cable's spectrum is much more well behaved than the broadcast spectrum, several significant advantages accrue. A time domain equalizer may not be necessary. If one is included, it may have relaxed specifications leading to lower cost. There is no "airplane flutter", i.e. Doppler effect from approaching or receding aircraft. Because the spectrum is better behaved, less error detection and correction is required for a given level of performance. This was well demonstrated in the Advanced Television Grand Alliance's modulation scheme. While 8-VSB is used for broadcast, 16-VSB was developed for cable allowing two HDTV signals in 6 MHz on cable. 16-VSB does not have twice the data capacity of 8-VSB. The doubling of payload comes because 16-VSB requires significantly less data protection. If this same approach is applied to the techniques proposed for data carriage in analog television signals, more of the raw data capacity can be harvested for payload purposes. This approach has not been well explored and offers a significant opportunity. An additional advantage is cable's availability of multiple channels to carry data. The data carrying capacity of a cable system is just huge!

THE COMPATIBLE DIGITAL CABLE UPGRADE

Most plans to migrate to digital video do not include wholesale replacement of all channels because of the horrific expense of the digital set top boxes. Instead, there is the intention of converting a few of the channels to digital and leaving the remainder as analog. In this strategy, the channels converted to digital will have previously been occupied by low penetration services. Subscribers wishing to continue with those services will need a digital set top box. If these subscribers take no new services, just the ones they had previously taken, their costs go up considerably while there is no increase in revenue. Subscribers who do not wish the new advanced services will not receive a new digital set top box. However, they will lose programming previously carried on the analog channels which are converted to digital. This can be a proportionally serious loss for low capacity cable systems.

An alternative is to use techniques which hide the data in the video for carriage of digital signals. Since television tuners are relatively inexpensive, multiple tuners can be provided so that data can be collected from more than one channel. That data can then be assembled to provide the MPEG streams needed to create new synthetic channels. In this approach, all of the analog channels are preserved for those who are satisfied with the existing service. Only those willing to pay for more will incur the extra cost of the new set top box.

THE FUTURE

There is an effort underway by a new start-up company, EnCamera Sciences Corporation, to raise the capacity of data hidden in television signals to in excess of 3 Mbps per video channel. That capacity could support two MPEG video streams at 1.5 Mbps each or one higher quality stream at 3 Mbps.

Additionally, massive amounts of data can be carried for Web type applications.

While full digital television is on its way, the interim can see a lively and cost effective data service while continuing to serve the analog base of receivers.

THE AUTHOR

Dr. Ciciora is a technology consultant specializing in Cable Television, Consumer Electronics, and Telecommunications.

Most recently he was Vice President of Technology at Time Warner Cable. Walt joined American Television and Communications, the predecessor to Time Warner Cable, in December of 1982 as Vice President of Research and Development. Prior to that he was with Zenith Electronics Corporation, starting in 1965. He was Director of Sales and Marketing, Cable Products, from 1981 to 1982. Earlier at Zenith he was Manager, Electronic System Research and Development specializing in Teletext, Videotext and Video Signal Processing with emphasis on digital television technology and ghost canceling for television systems.

He has nine patents issued and several more pending. He has presented over one hundred papers and published about fifty, two of which have received awards from the Institute of Electrical and Electronic Engineers (IEEE). His papers have been translated into Japanese, Chinese, German and Spanish. Walt wrote a monthly columns for Communications Engineering and Design (CED) magazine and for Communications Technology (CT) magazine for three years each. He continues in alternate months for CED.

He currently serves on the Executive Committee of the Montreux Television Symposium. He was a member of the board of directors of the Society of Cable Television Engineers (SCTE) for six years. He was Chairman of the Technical Advisory Committee of CableLabs for four years and Chairman of the National Cable Television Association (NCTA) Engineering Committee also for four years. He was president of the IEEE Consumer Electronics Society for two years and is a past chairman of the IEEE International Conference on Consumer Electronics. He chaired the Joint Engineering Committee of the NCTA and the Electronic Industry Association (EIA) for eight years. He has served on several industry standard-setting committees. He currently co-chairs the Cable Consumer electronics Compatibility Advisory Group and its Decoder Interface subcommittee.

Walt is a Fellow of the IEEE, a Fellow of the Society of Motion Picture and Television Engineers (SMPTE), and a Senior Member of the SCTE. Other memberships include Tau Beta Pi, Etta Kappa Nu, and Beta Gamma Sigma.

Current interests center on competitive technology, the consumer electronics interface with cable, Digital Video Compression, Interactive Television, Multimedia, and High Definition Television.

Walt received the 1987 NCTA Vanguard Award for Science and Technology and was named "1990 Man of the Year" by CED magazine. CED also named him "1993 Man of the Year". He was the Fall 1994 Levenson Memorial Lecturer at the National Cable Television Center at Penn State.

Walt has a Ph.D. in Electrical Engineering from Illinois Institute of Technology (IIT) dated 1969. The BSEE and MSEE are also from IIT. He received an MBA from the University of Chicago in 1979. He has taught Electrical Engineering in the evening division of IIT for seven years.

Hobbies include helping his wife with her horses, reading, wood working, photography, skiing, and a hope to someday become more active in amateur radio (WB9FPW).

Optical Network Technology: Future Impact on CATV Networks

John Holobinko & Bill Hartman
ADC Telecommunications, Inc.

Abstract

Future CATV networks may be able to transport video, data and voice services over large areas made possible by managing individual optical wavelengths within a single fiber, each wavelength carrying a different service type or going to a different location. The capability of optically routing, switching, provisioning and otherwise controlling various services without intermediate optical/electrical/optical conversion will enable creation of "All Optical Networks". This paper discusses the current state of optical technology required for these networks, where this technology is first appearing within existing CATV infrastructure, and how it may positively impact the capital, operations and maintenance costs of future CATV networks.

Introduction

Wave Division Multiplexing (WDM) is the ability to transmit two or more optical signals independently through the same fiber, utilizing different optical wavelengths. Although transmission of 1310 nm and 1550 nm wavelengths have been used for many years, it has been the advent of commercially available optical amplifiers which have made it is now technically feasible to transmit tens of optical signals simultaneously on the same fiber within a relatively narrow optical window of approximately 30 nm. This is referred to as Dense WDM Transmission, or simply dense WDM for short. Figure 1 illustrates a point to point dense WDM fiber link. Although the ability to transmit a dense WDM stream and amplify its multiple signals with a single optical amplifier is a key element of future All Optical Networks, commercial development of a number of new optical devices will be required in order to take full advantage of the potential benefits of All Optical Networks.

All Optical Networks may provide the following economic benefits to CATV service providers:

- Lower Fiber Plant Cost Through Significantly Reduced Fiber Counts
- Shared Signal Transmission and Switching Through Common Active Optical Components, Reducing Electronics Costs
- Improved System Reliability Through Reducing Overall Network Active Devices
- Faster Fiber Restoration After Cuts, Through Significantly Reduced Fiber Counts
- Reduced Future Costs For Network Capacity Expansion By Further Sharing Common Plant and Equipment

In addition to these benefits, technology improvements may allow each hub to economically serve significantly larger areas in terms of homes passed per hub, thereby allowing further consolidation and reducing operations costs.

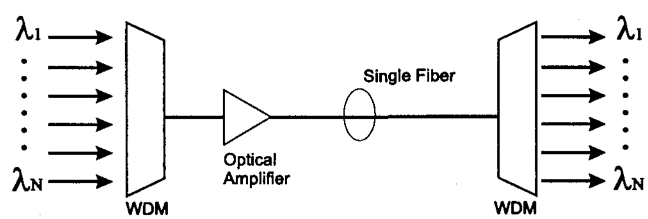


Figure 1. Dense WDM link. "N" can be 4-32 commercial applications today.

Elements of an All Optical Network

Dense WDM technology enables the creation of multiple optical circuit paths within a single fiber path (Figure 2). In relative terms, it is easy to compare an All Optical Network to a fiber network in the following way: An optical cable consisting of multiple fibers within a sheath becomes a "superset". An individual fiber within the cable can be thought of as a virtual fiber cable. An individual optical wavelength within the fiber can be thought of as a virtual fiber. Management, redundancy and routing can all be readily understood by translating requirements in conventional

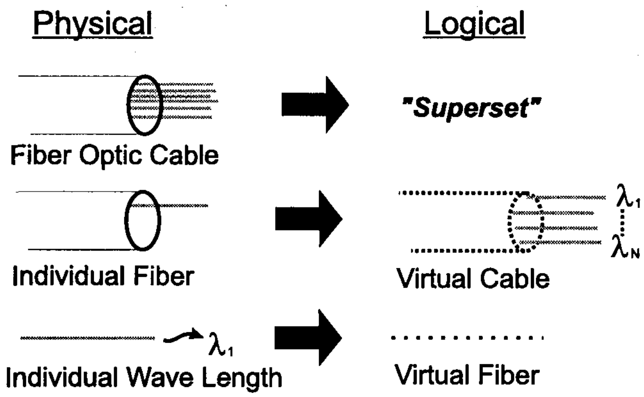


Figure 2. All optical network equivalents

networks between cables, fibers, and their virtual counterparts in an All Optical Network.

To attain the economic and operational benefits derived from implementing the future All Optical Network, a number of optical elements will be required which are currently not available in commercial quantities for wide scale deployment. Figure 3 is a compilation of these devices. From left to right shows progression in time of the anticipated evolution of these devices.

Dense WDM transmission is not without technical challenges. Operating at the 1550 nm window, attention must be paid to issues such as the dispersion performance of the optical fiber, the flatness of amplifier gain in the optical bandpass, and the optical stability of fiber devices. Recognizing these challenges, this paper is primarily focused on the potential application of All Optical Networks in CATV systems.

Large urban/suburban CATV networks consist of a series of hubs/subheadends (or video end offices) connected to one or two master headends (or video serving offices) usually via a redundant fiber optic ring, commonly referred to as the fiber backbone system. While most large backbones are exclusively digital, medium sized rings may be a combination of digital and high powered linear systems. A hybrid fiber/coax distribution architecture is used to distribute signals bidirectionally between the hub and the serving areas via linear optical transmission between the hub and optical nodes, and linear RF transmission within each serving area between the node and subscribers. Figure 4 illustrates a typical backbone system architecture while Figure 5 illustrates the distribution system architecture.

Dense WDM systems offer potential economic advantages in both digital and broadband linear (aka: analog) CATV transmission. The first area of anticipated use

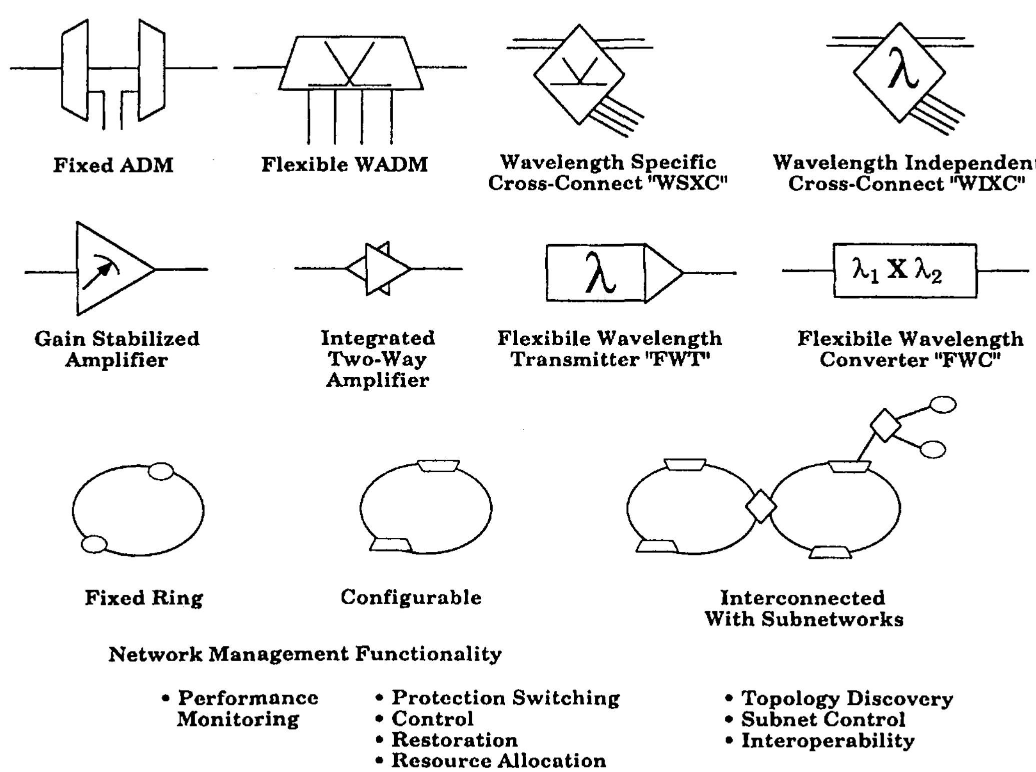


Figure 3. Current and future elements for All Optical Network: Source Bellcore

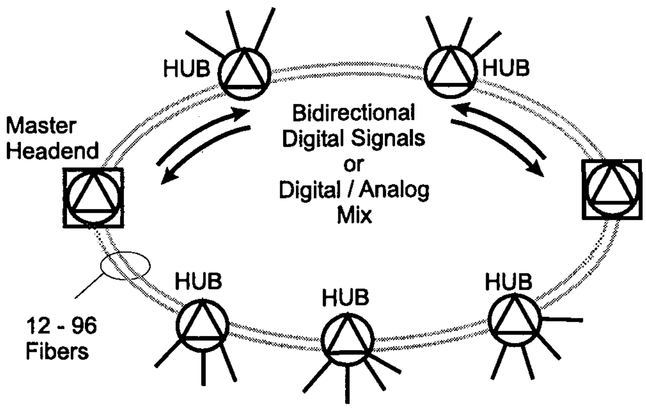


Figure 4. Backbone System. Fiber transmission is typically uncompressed digital for large networks, and a combination of digital and high power 1550 nm AM for smaller systems.

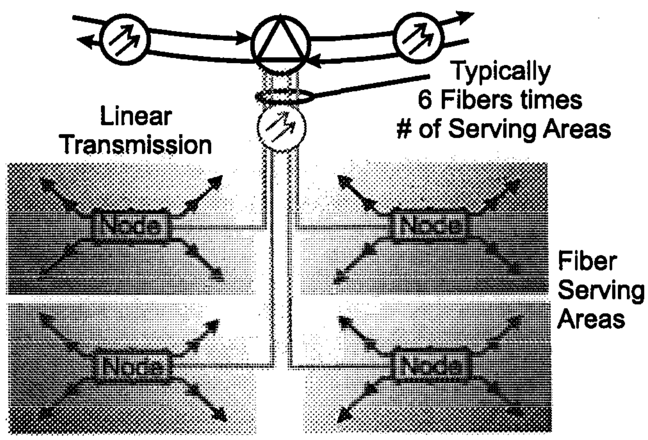


Figure 5. Distribution System. Fiber transmission is broadband linear between the hub and the optical nodes in the serving area.

of dense WDM technology in CATV networks is in fiber backbones, followed by potential implementation in the HFC distribution system.

Dense WDM in Backbone Systems

Fiber backbone systems generally cover long distances. The longest systems in the United States now cover over 500 kilometers. Given the expansion of CATV networks into the delivery of high speed data, telephony and other digital services, these systems usually employ uncompressed digital fiber systems to transport all video, data and voice services, or a combination of uncompressed digital and conventional telecommunications digital systems to transport and remultiplex various combinations of channels to create custom service delivery configurations at each sub-headend.

Shorter systems sometimes employ a combination of digital transmission systems (for voice and data services), and 1550 nm high power linear transmission systems (for broadcast video services) on separate fibers, which are then combined at the hubs.

Each fiber in backbone systems represents a significant capital cost because of the long distances traversed. The initial system may be even more expensive if requirements dictate leasing fiber(s) over a limited right of way such as a long bridge spanning a body of water. The ability of combining multiple digital signals using many optical wavelengths has already been demonstrated and commercial products are already available. For example, ADC Telecommunications demonstrated eight wavelengths of its DV6000 uncompressed digital system at the 1996 SCTE Expo as shown in Figure 6.

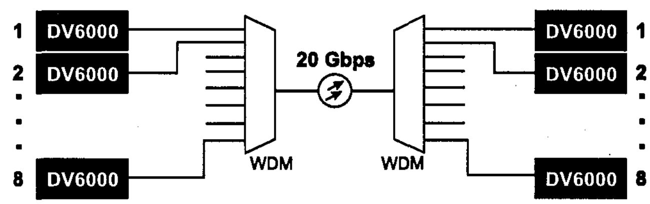


Figure 6. ADC Telecommunications' DV6000 Uncompressed Digital Transmission System employing DWDM for 20 Gbps transport.

This provides a single fiber capacity of 20 Gb/s, which translates to 128 analog CATV channels, 256 DS3 channels, 128 MPEG2 QAM multiplex signal streams (with up to 20 MPEG multiplex channels per stream) or a combination of these signal types. Other vendors currently provide systems which can transmit eight or more simultaneous wavelengths containing digital information, over the same fiber. A counter rotating ring can be implemented with no loss of capacity on 2 fibers, as shown in Figure 7. Alternatively, it is technically possible to implement a bidirectional WDM system with redundancy on a single fiber at 10 Gb/s.

Today, the application of WDM is primarily limited to point to point transmission between hubs on the ring. This is due to the fact that only "hard wired" fixed wavelength optical splitters are available. This provides the benefit of reduced fiber count, but not full optical add/drop capability. In order to dynamically drop or add a wavelength at any hub, a dynamically

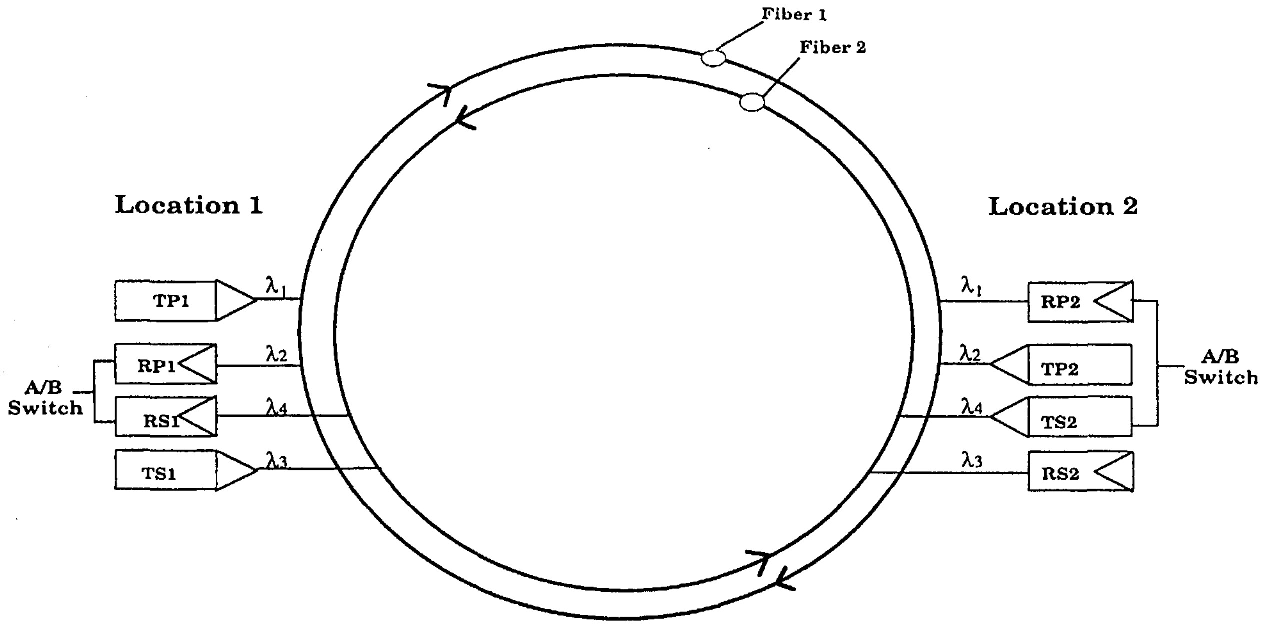


Figure 7. Counter rotating ring with full fiber and electronics protection

wavelength selective optical WDM device will be required. For high speed digital telephony systems, there may be a significant cost advantage in being able to do a drop and add function in optics versus electronically, due to the high cost of the digital add/drop multiplexor. Optical switching can perform virtually the same functions as TDM switching. The only optical limitation in multiplexor functionality s that there is not an easy way to make a drop *and pass* device.

The benefits of dense WDM are not as obvious in a smaller backbone system that employs 1550 nm high power linear transmission of CATV video channels. Although it has been demonstrated that digital and linear optical signals can be simultaneously passed through the same optical amplifier[1], the optical link budgets of high speed digital transmission systems are so dissimilar as to make this probably impractical in common implementation. For example, a 2.4 Gb/s operating at 1550 nm has a link budget of 29- 31 dB, translating to over 100 km in length, which is within the range of distance between 90% of all hubs. A typical high power linear system has a link budget of 12 - 14 dB. Therefore, the linear system will require amplifiers after a distance of 12- 14 dB which is less than half of the link budget for the digital system. Operating the digital system through these amplifiers will provide no advantage to the digital system, and therefore if digital and high power linear signals are mixed on the same fiber, extra cost must be incurred to provide splitters to the digital signals in order to bypass these optical amplifiers.

Dense WDM in HFC Distribution Systems

While there are clear capital and operational savings to be had today from building a new digital fiber backbone system employing dense WDM technology through saving of fibers and electro/optic repeaters, there are potentially larger future savings in the forward and reverse path of the HFC distribution system.

Consider that the average hub serves between 20 and 80 optical nodes. It is typical to provide 4-6 home run fibers between the hub and each node, since at least two fibers are normally required for bidirectional transmission, and extra fibers are installed to support future migration of nodes closer to the subscriber, or additional services close to the node. Multiplying the nodes times the fibers per node calculation means that as many as 480 fibers are required at the hub. Given that WDM would allow multiple linear signals to be transmitted on the same fiber, the number of fibers at the hub could be reduced by 67% while maintaining the ability for future expansion of nodes closer to the subscriber. Even greater savings are possible if fiber branching is allowed. For example, if 16 nodes could be served using 16 wavelengths originating on one fiber from the hub, then it is possible to serve 80 nodes two-way using only 10 fibers total. Figure 8 illustrates this concept. Note also, that in building a metropolitan system that there may easily be ten or more hubs, so that the savings realized is factored by the number of hubs (i.e. distribution networks) in the overall system. To accomplish this savings technically requires future development of both the lasers which can support 110 channel linear transmission at various wavelengths around 1550nm, as well as WDM optical devices which demonstrate excellent long term stability and performance while installed in an outdoor, unprotected environment.

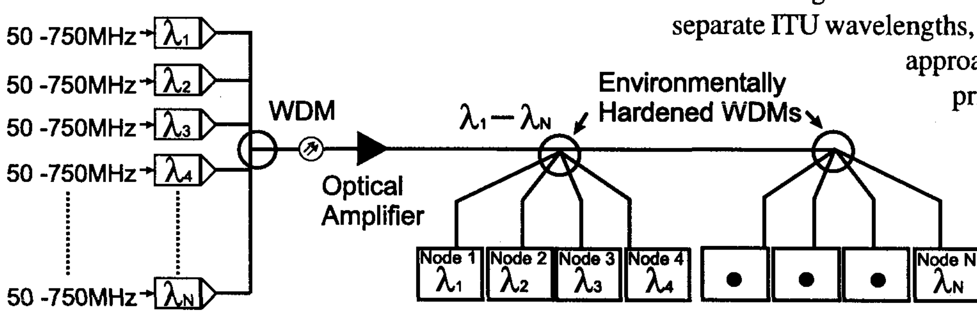


Figure 8. Potential Future Dense WDM in HFC Forward Path

Another alternative that has been postulated is to transmit in the forward path from the hub to nodes the common broadcast channels (typically from 50 - 550 MHz) at one common wavelength, and then to send the narrowcast signals to each node on a different wavelength where they are combined in the optical receiver. Although this is possible from an optical perspective, the combining of signals at the node receiver may be a more difficult approach in actual implementation. This is because it is difficult to combine the signals at each node with the correct RF level. Without going into a detailed technical explanation in the space of this paper, suffice to say that the RF output level of a signal out of an optical receiver is proportional to the input optical power of the received signal and the square of the depth of modulation. Trying to match two different optical transmitters' outputs coming into the node receiver with two power levels and two depths of modulation would probably be difficult.

Dense WDM may also hold future promise for return path expansion. Today, block upconversion is the most often proposed means of taking up to four 5-40 MHz return paths and transporting them across a single fiber back to the hub. In the future it may be possible to optically multiplex the return paths instead of frequency multiplexing these signals. Of course this will require both the multiple optical wavelength transmitters which are cost effective, and the WDMs which meet the environmental rigors required to be installed in a strand mounted unprotected optical node.

All Optical Network Challenges

The current barriers to All Optical Networks center around devices and availability. The ITU has suggested standardized channel spacing based on specified optical wavelengths in the 1550nm bandpass region of optical amplifiers. In the digital realm, high speed 1550 nm wavelength lasers are becoming available for 40 separate ITU wavelengths, at prices which are rapidly approaching standard 1550nm pricing. Work is being done by some vendors on creating multi-wavelength laser arrays to further drive down pricing. However, using 40 different channels brings up the real world problem of transmitter sparing. Currently, this limits the flexibility. The ideal solution is a tunable wavelength laser, if not for all transmitters, then at least as a universal spare. This is a comparable problem to that which the CATV industry had in the 1980's, when VSB/AM modulators were fixed channel. The advent of tunable lasers will significantly accelerate the implementation of WDM systems. Correspondingly, similar breakthroughs are required in broadband linear optical devices.

Another key requirement is the availability of low cost/ highly efficient WDM devices, both fixed wavelength and tunable, suitable for indoor and outdoor installation. In this area, technology is moving forward very quickly, and the emergence of these devices appears on the horizon.

Conclusions

Dense WDM technology and the ability to create All Optical Networks will allow further improvements in the cost, flexibility and reliability of CATV networks.

Availability of the optical components necessary to create these networks will occur within 2-3 years, giving network providers additional means of providing better service and additional capacity to their customers.

References

[1] Chinlon Lin, Keang-Po Ho, Hongxing Dai, Janyi Pani, Hermann Gysel, Mani Ramachandran, "Hybrid WDM Systems for High Capacity Video Trunking Applications", National Fiber Optics Engineers Conference Proceedings, September 1996, p. 261.

Downloadable Firmware in Advanced Settop Terminals

Samuel Reichgott

General Instrument, Advanced Network Systems

Abstract

There are many reasons why both cable systems and equipment suppliers benefit from downloadable settop terminals, but there are many decisions to be made in developing a downloadable settop system. Downloadability can take place at several levels of complexity. A downloadable firmware system should be designed for open development. While settops should accept firmware incrementally, this poses a challenge to the system controller and its operator. There are new considerations for reliability, separate from those of classical communications. The design should also cover the use of inband video channels for downloaded firmware, and be mindful of various consumer acceptance issues.

BENEFITS OF A DOWNLOADABLE SETTOP SYSTEM

Both cable system operators and equipment suppliers have strong motivations for deploying downloadable settops. Some of these motivations are identical, or are at least two sides of the same coin. Other reasons driving downloadability reflect the differences in their business priorities. Still, both cable systems and equipment suppliers stand to benefit from the deployment of these new settops.

For the Cable System Operator

Brand Identity Whether at the level of a logo or trademark, menu text, or entire applications, systems can use downloadability to distinguish themselves from their competition. This is especially important when neighboring or overlapping systems use similar equipment.

Revenue Generation Downloadability provides a new opportunities for tiered or individual subscription services. Data for downloaded applications, or entire applications, could be downloaded on a pay-per-use basis.

Low Cost Upgrades When equipment suppliers and third-party developers, or Independent Software Vendors (ISVs), improve or develop new applications, a downloadable settop allows a system to distribute these upgrades to their customers without sending an installer or requiring customers to return settops for upgrade.

For the Equipment Supplier

Speed to Market Previously, a settop's firmware had to be complete, in its entirety, before a single settop could be shipped. Now, a reliable platform and basic settop functions are all that are required to field a downloadable settop. Additional features are added as they are developed, while the basic features are already being sold.

Reduction of Risk Regardless of the number of "focus groups" used to help define a product, there are always customers, marketing, and engineering personnel thinking of improvements. With a downloadable system, equipment suppliers can produce a product embodying their best attempts to satisfy customer requirements, knowing that they can still download the "final version." Needless to say, this is also a valuable tool when it comes to fixing "bugs."

Inventory Consolidation Instead of producing different ROMs for different markets, the equipment supplier can really keep its software soft. No matter what the mix of applications, all similar hardware platforms can have the same ROM and defer the differences to download time.

Flexibility It's this very concept that differentiated the early PCs from word-processors. While both had about the same computing power, the PC won the market, even among single-application users, because of its ability to change. Customers are naturally attracted to the product that promises a hedge against obsolescence.

LEVELS OF DOWNLOADABILITY

Settop downloadability is not an all-or-nothing proposition. Downloadability can take place at several levels, each with its own concerns and complexities. Equipment suppliers need to examine their market requirements and determine which levels of downloadability to include in their product. These levels may be described as follows:

Text Level This is the simplest level of downloadability. No executable firmware changes, and downloaded text (or graphics) replaces default text (or graphics) in existing applications. At this level of downloadability, one can change fonts and logos, customize menus by renaming standard features, display station call letters, or do basic foreign language text substitution. Text Level downloading is a simple, yet powerful, way to provide a customizable user interface.

Data Level As in Text Level downloading, the executable firmware in the settop does not change. Data Level downloading is distinguished from the simpler Text Level by its ability to present a variety of data to the consumer. This level of downloadability has been available for several years in on-screen messaging and downloadable barker applications. More sophisticated uses of Data Level downloading are in today's on-screen program guides, which include database search and sort capabilities, timers for future tuning and (where supported by the cable system), future purchasing and recording.

Interactive Data Level This is similar to the Data Level, except that at the Interactive Data Level the consumer can exercise control over the data that is downloaded. This level of downloading is typified by the various browsing applications that are beginning to appear. It will continue to gain popularity as the internet becomes accessible from more settop terminals. Another current example of Interactive Data downloading is the Infra-Red Blaster application, in users select their VCR manufacturer and model, and the settop loads the appropriate data to control its built-in IR Blaster. Note that two-way interactivity is not strictly required for a system to operate at the Interactive Data Level; as long as users can control which one-way data is loaded, the requirements for this level are met. Without two-way interactivity, the user must select from a limited number of pre-programmed data streams. As true two-way interactivity is added to the system, consumers may select more precisely the data that enters their households.

Application Firmware Level This is the first level at which executable firmware is changed. Application Firmware Level downloading is the most powerful tool available to achieve the cable system operator's goal of brand identification. The entire look-and-feel of the settop can be changed, but that is the extent of the change allowed. Downloaded application firmware satisfied with the settop platform's fixed capabilities.

Platform Firmware Level Settops that are downloadable at the Platform Firmware Level can actually change the services available to application firmware. New device drivers or other platform services can be loaded to enhance communication capabilities on existing hardware. Depending on how much of the platform is downloadable, part or all of the Application Programmer's Interface (API) can be modified. Downloadability at the Platform Firmware Level is essential to meeting the equipment supplier's speed-to-market goal because it allows a generalized hardware platform to be fielded before the details of its use have been fully realized.

The remainder of this paper discusses the issues associated with Application and Platform Firmware Level downloading. While Text, Data, and Interactive Data Level downloading are powerful tools with their own interesting concerns, these are becoming commonplace in the industry. The ability to download and change executable firmware in a settop requires unique solutions to business and technical questions that warrant their own discussion.

OPEN OR CLOSED SYSTEM?

The decision to offer an open platform for downloading firmware is equally a business and an engineering issue. Entirely new business problems arise, such as with whom to team, how to market ISV applications, how to organize for ISV technical support, and how much proprietary information to share. These issues have long been understood in the personal computer arena, but are novel in the cable settop market. In addition to the development required by cable operators, the computer-savvy ISVs require still more development to match their visions of advanced user interfaces and interactivity. Many would argue that these are good problems to have but they are, nevertheless, serious challenges to a business organization built on the traditional cable industry model. Still, it is difficult to argue with the decision to provide an open platform, given the advantage of hindsight in the example of the PC software industry in the 1980s.

Given the business decision to be open, it is the engineering department's job to respond. Here are a few of the areas that must be addressed when making the leap from proprietary to open development.

The API ISV applications need an interface to the settop platform. The APIs should be provided and documented by the settop manufacturer's firmware development team. Minimally, the API hides the settop's hardware details, making it easier for ISVs to write their applications; but this isn't its only benefit. Even if the API is only a firmware-to-hardware interface, it reduces code size and improves reliability because this interface appears only once, and may be thoroughly tested.

While absolutely necessary, a hardware interface API is probably insufficient. Higher-level features should also be provided by the API, with the same benefits of reuse and reliability. Higher level API functions for cable settops might include password maintenance, favorite channel lists, parental controls, downstream data access and upstream data transmission facilities, purchasing, program guide database access, and a host of other features. The biggest challenge, aside from the implementation of these features, is to create a rich enough set of APIs to satisfy applications that will be written in the future.

The Kernel Old-fashioned proprietary code development was often based on the "super-loop" program architecture. A single thread of control would handle each settop operation, one at a time, then start over again at the top of the loop and repeat forever. This was fine for the closed system with at most a simple menu interface; it was easy to get everything done in time. But the super-loop is clearly inappropriate in an open system, where the different operations are variable and unknown when the platform is created.

What is called for is a small operating system; and a multitasking kernel is a good choice. A multitasking kernel allows different jobs (called tasks) to be dynamically created as required. It orchestrates the concurrent operation of all tasks in the settop. Multitasking lets the processing of secure subscription information (program guide data collection, etc.) take place at the same time as downloaded applications monitor the remote control and run the on-screen display. Each of these tasks can be developed independently, with little or no concern for the details of the other tasks in the settop(provided a well documented API is available).

Many off-the-shelf kernels are commercially available, or a custom kernel may be developed. When considering the buy-or-build option, the firmware manager has to ask several important questions: What kinds of debugging tools need to be developed? How long will it take to develop and test? When these questions are seriously examined, the answer should be obvious to all but the die-hard do-it-yourselfer: Buy, don't build. There are plenty of good kernels across the price spectrum. They come with kernel-aware debugging tools, which are already developed and ready to use. They have customer support departments ready to handle your startup problems. Most importantly, they're written by people in the kernel business, not the cable business. Save your development resources for what you do best!

The Development System The development team which creates the settop platform and "native" applications needs tools to do the job. Some of these tools will be purchased, some will be developed by the firmware team itself. Even the purchased tools will have to be adapted to use the settop as a target system. It stands to reason that ISVs will need the same kinds of tools. If it is not in their culture to begin with, the team will have to learn how to package the work they do, creating their toolbench for outside use. This might seem like extra work at first, but it's actually worth the effort because the resulting documentation becomes a lasting reference, which might otherwise not exist.

MONOLITHIC OR INCREMENTAL DOWNLOADING?

When developing a system for downloading firmware, a fundamental question arises. Should the firmware be loaded in a single, monolithic, chunk or in small, incremental, pieces? The answer to this question affects the rest of the design of both the settop loader firmware and the system controller's downloading software.

Fortunately, it is possible to make the monolithic-versus-incremental decision separately for the settop and the controller. If the settop is designed for the more general, incremental loading, the controller can still download all the firmware monolithically. As explained later, this mixed system can have both speed-to-market and usability advantages.

In the settop, incremental downloading promises the ability to add features without disturbing other executing firmware. New kernel tasks can enter the settop, and be started by the loader, even while the consumer is busy changing channels or browsing the guide. Monolithic downloading, on the other hand, assumes that the entire set of replaceable firmware is loaded at once, leaving the settop temporarily unusable or with limited capabilities during load time.

Once the settop is capable of incremental loading, this ability can be generalized from loading firmware, which is stored in non-volatile memory (typically EEPROM), to loading software, which is held in volatile memory (RAM) for a relatively short time. Transient applications can be loaded in RAM, executed, and then erased. This capability can be used for interactive commercials or infrequently used diagnostics. It can also be a means of cost reduction. Since code is only held in the settop when it is needed, total memory requirements may be minimized.

FIRMWARE REQUIREMENTS FOR INCREMENTAL DOWNLOADING

When firmware is to be incrementally loaded, it is best to assume it can be loaded in any order. While this is easily said, it represents a significant challenge when designing the firmware loader and the downloadable firmware itself.

Relocatable versus Position-Independent

Loading firmware in any order implies that the location in memory where the firmware will reside is unknown when the code is developed. This is true of both the code space, probably in EEPROM, and the data space in RAM. The goal is then to get the code into the settop and allow it to use the platform services without knowing where it will actually land.

There are two ways to achieve this goal. The code can be "relocatable," or "position-independent." Relocatable code is, essentially, incomplete. It carries with it references to other code and data that it requires to do its job, but these references are called "unresolved" and are maintained in a symbolic form. In traditional firmware development, relocatable code modules are linked together at development time and given absolute memory addresses in order to resolve all unresolved references. In a downloadable system, the linking must occur in the target (the settop) because that's where the other required pieces already exist. The settop's loader must maintain all symbols, and their resolving addresses, in memory in case another module is loaded and requires some existing information or function.

Position-independent code, unlike relocatable code, has no unresolved references. Instead, it accesses its own code and data through references relative to values held in the microprocessor's address registers. For example, the code may find a particular variable "at the address that is 10 bytes away from the address stored in the A5 register." Position-independent code, therefore, need not carry symbolic references to its own code or data because the address register values are established after the code and its data have been located in memory. The resulting size of the downloadable file is significantly smaller than the same file in relocatable format. This is important in conserving both downstream bandwidth and settop memory. However, it does not solve all the problems. There still must be a way for the downloaded module to access the functions of other code already existing in the settop, including the kernel. This is the job of the trap libraries, which are explained later.

The decision to go relocatable or position-independent is a trade-off between flexibility and cost. With relocatable code, all relevant symbols are available to all code on an as-needed basis. There is much less need for foresight, because a developer can count on the availability of any function or data to already be in the settop. However, the cost (in terms of downstream bandwidth and memory) to transport and store all the symbols can be enormous. Conversely, position-independent code is very efficient; only the executable code, in its binary form, needs to be downloaded and stored. Great foresight must be used to provide the trap libraries (or other mechanism) required to access all the functions that are already present. If essential functions aren't given the proper visibility, some efficiency and reliability may be lost in having to download the same function more than once. This is the same kind of foresight required when developing the settop's API. Luckily, if the API requirements analysis is truly complete, the trap library requirements should be well understood.

Trap Libraries

A "trap" is a mechanism by which a code module ceases executing in its own code space, forces the processor to execute code in another module's code space, and then returns to resume execution at the point of the trap call. Most commercially available kernels provide a trap library to access kernel functions. Settop platform developers need to create their own trap libraries in order to provide access to their API. There are other alternatives, but a trap library is a very code-efficient means of providing an API interface when using position-independent downloading.

MIGRATION TOWARD AN INCREMENTAL DOWNLOAD SYSTEM

If the settop platform supports incremental downloading, the cable system controller can still download monolithic code releases. From the controller's point of view, it transmits a single file to its terminals. From the settop's point of view, this single file actually consists of separate pieces of the settop firmware. This approach significantly simplifies the controller's development while leaving flexibility for the future.

One of the concerns for the controller is to track which firmware is loaded in each subscriber's settop. As an example of why firmware tracking is necessary, consider the fixed capacity of each settop's downloadable memory. It would be an error to try to download more code than the settop could hold and expect all the features to function. Unless the controller keeps track of all separately loaded modules in each settop, such errors cannot be avoided. It's relatively easy to group settops with identical hardware capabilities together and download the same firmware to all of them. It's a much more complex tracking problem if each settop can be sent different firmware. It also stands to reason that a monolithic download environment presents a simpler user interface to the controller's operator.

Still, a need for greater variety among tiered feature sets may call for incremental downloading. Given the limited memory capacity of each terminal, it is impossible to load all features and enable them in tiers. Different subscribers may someday require different firmware mixes, as the number of available applications increases. Eventually, the industry will have to respond by solving the associated database and user interface issues.

PROVIDING A RELIABLE DOWNLOAD ENVIRONMENT

For professionals in the communications industry, it is unnecessary to restate the need for a "clean" distribution plant, appropriate data packetizing, and data error detection/correction. These classical problems are well understood and don't require further treatment. However, there are other reliability issues that must be addressed for a firmware download system to work in the cable environment.

No matter how good your plant, no matter how good your error correction, there will still be data errors. Especially in a one-way cable system (an open-loop system) these errors could have dire consequences for downloaded firmware. The firmware loader in the settop has to be intelligent enough to prevent bad code from getting loaded. The controller must periodically repeat the downstream transmission of the firmware so it is guaranteed to be eventually received, error-free.

With the possibility of data errors, there is the possibility of parts of the download being absent from the settop. There are two ways to handle this problem, and both may be necessary to guarantee that the settop doesn't "crash."

First, the loader may have to know the minimum requirements for system startup. After firmware is incrementally loaded, the loader should determine that a minimum functional set of firmware is present before it starts executing any of the downloaded code. Second, the firmware itself has to be fault-tolerant. This means that if any firmware that is executing cannot find an essential library function, a platform service, or some important data, it must come to a graceful halt. Ideally, a partially loaded system functions to the fullest extent possible. Minimally, it must remain intact and eventually complete the download and proceed with full capabilities.

As two-way systems mature, settop firmware loaders must likewise mature. The loader should take advantage of an upstream channel to request missing pieces of the download. It should also report loader error conditions for diagnostic purposes.

OUT-OF-BAND VERSUS INBAND DOWNLOADING

Inband data transmission for Data Level and Interactive Data Level downloading is already the norm. This promises to set a trend for Application and Platform Firmware Level downloading. In the analog video environment, there is enormous bandwidth available on video channels (compared with the limited bandwidth on a typical out-of-band service channel). This is a great motivation for designing a system for inband firmware downloading. With so many video channels, it is possible to put different applications, or tiers of firmware, on different channels. However, there are system-level and consumer acceptance considerations needed in making this design decision.

A major issue is where to find the data. The settop's loader must be provided with some way to tune the correct channel for the firmware it should be loading. A reasonable approach is to provide just the channel tuning information on the service channel and allow the loader to find the firmware on an inband channel using this information.

Once the developers solve the problem of where to find the firmware, the problem of when to load it arises. Consumers simply will not understand, or tolerate, a settop that suddenly changes the channel just to do a firmware upgrade. There's the classical argument stating: "Just do it at 2 in the morning." Unfortunately, this really doesn't cut it with fans of the Late Late Show. A creative solution is called for -- one which the TV-centric, computer-phobic consumer can handle.

CONSUMER ACCEPTANCE ISSUES

As mentioned before, consumer acceptance must be considered in designing a downloadable settop. The biggest problems stem from the limited memory in a low-cost settop. Unless memory is infinite, you'll eventually have to erase some of it to make room for newer, better firmware. And therein lies the problem: Once you erase firmware, the settop just won't be fully functional for a while.

The settop designers need to determine a minimal set of functions that will consumers happy while their settop firmware is being upgraded. Is channel-surfing enough, or does the program guide need to stay intact? It comes down to a decision of what can be spared until the upgrade is complete. These decisions will determine how much ROM is required to hold the minimum feature set, versus how much EEPROM (or other downloadable memory) is provided for upgrades.

Another related problem that must be solved is consumer notification. Naturally, people accept change better if they know it's coming. Depending on how consumers are notified, this could be costly for a cable operator. Will the consumer read the note at the bottom of the monthly statement? Probably not. Is a direct mailing the answer? Not for many households that simply trash their junk mail. However, if the settop is smart enough to download firmware, why can't it do the notification? An on-screen message application can be used by cable operators to inform their customers of a pending upgrade. A standard message can be addressed to all affected subscribers. It can provide plenty of advance notice, stating the date and time of the proposed upgrade. In addition, an "emergency" message can pop up on the screen an hour before the upgrade is scheduled to begin. Still, operators will have to expect a few phone calls, no matter how hard they try.

Another area of concern is the retention of user preferences and other settings through a download. Users may be given menu-selectable options, for example, a time display or channel number display on the front-panel LEDs. If such preferences are erased with the firmware, even a successful download in the dead of night will result in phone calls the next morning. The problem gets worse if programmed timers for future purchases are lost. This means lost revenue to the cable operator, as well as upset consumers. Settop designers must be mindful to create a separate area of memory to hold these preferences and timers so they are not lost when the firmware is erased. The time spent to properly engineer this issue is well worth the effort, because it avoids problems for consumers and cable operators alike.

CONCLUSIONS

Downloadable settops offer advantages to cable system operators and equipment suppliers alike. Cable systems gain the advantage of improved brand identity, revenue generation, and low-cost upgrades. Equipment suppliers benefit from speed-to-market, reduced risk in deploying new features, inventory consolidation, and the ability to market more flexible products.

Downloadability can be realized at several levels. Text or data can be downloaded without changing the executable code in the settop, or the settop can change its feature set by downloading new code. Downloading can be interactive -- achieving maximum selectability in full-featured, two-way cable systems.

An open system has distinct advantages. Many third-party software developers can add features to the settop, making the equipment more marketable without the expense of additional development resources. The development team must provide facilities to make open development easier, including an API and well-documented development tools. An open system may also call for a multitasking kernel in order to make new code easier to plug in.

Settops can download firmware incrementally while the controller migrates from monolithic to incremental downloading. This will eventually be required as more third-party applications become available.

The vast bandwidth available on video channels tempts settop designers to use them for downloading firmware. However, system and consumer acceptance problems must be solved before inband downloading can be implemented satisfactorily.

Finally, the consumer's impressions of the system must be addressed when designing a downloadable firmware system. Resolving these questions is an important key to providing a successful system.

Fusing in Modern HFC Network for Improved Network Availability

Oleh J. Sniezko
Director
Transmission Engineering
TCI Communications, Inc.
Terrace Tower II
5619 DTC Parkway
Englewood, CO 80111-3000
(303) 267-6959
FAX (303) 488-3210

Abstract

Fusing or protection against over-currents in the cable TV coaxial distribution network is an area in which many misconceptions found a way into coaxial plant powering design. The origins of these misconceptions are numerous, but most of them can be traced to the interpretation of the rules of fusing in the power plant and power installations. A direct implementation of these rules results in lower reliability of the network without meeting the objectives of equipment and personnel protection, and fault isolation.

This paper presents the fusing analysis from the point of view of the two primary objectives of fusing:
1. to protect components and equipment from costly damage caused by over-currents; and
2. to isolate a faulty part of the network from the remainder of the network once the fault has occurred.

These two direct objectives are subordinate to the paramount objective of increasing system availability through:

1. *improved or at least maintained current reliability of the network, and*
2. *reduced repair time.*

The first decision during the fusing design is whether a fuse is required in a particular location and whether it can meet the objectives of fusing. If the answer to these questions is positive, the following criteria will decide about type, value, and other parameters of the fuse selected:

- *Maximum operating load current,*
- *Ambient temperature,*
- *Aging,*
- *Operating voltage,*
- *Frequency of the powering voltage,*
- *In-rush characteristics of the network, surge currents, etc.,*
- *Available maximum fault current,*
- *I^2t characteristics of the devices to be protected,*
- *Allowed voltage drop across fuse,*
- *Applicable standard agency,*
- *Ergonomics,*
- *Humidity, and*
- *Standardization.*

The paper analyzes all these parameters and presents a detailed fusing design for HFC networks, supported by the up-to-date principles of fuseology. Finally, it presents an example of fusing design in modern HFC networks.

INTRODUCTION

A hybrid fiber-coaxial (HFC) network consists of active and passive elements. To distribute RF signals, the network elements must also distribute low frequency energy to the active components. Moreover, the network operates in an environment that generates unwanted low frequency energy. This unwanted energy causes electrical stresses such as voltage transients and current surges. These electrical stresses and additional mechanical factors can lead to overcurrents. Therefore, most of the HFC networks include a low-frequency system comprising of:

- energy supply system (powering)[1];
- transient and surge protection elements (surge arrestors, amp clamps, grounding, etc.);
- safety elements (grounding, isolators); and
- overcurrent protection.

All these elements have specific characteristics to meet the following objectives:

1. provide for high level of service availability[2], and
2. ensure public safety.

The overcurrent protection system serves to meet the first one by providing

- improved reliability of the network, and
- reduced repair time.

Despite the importance of this system, its design was usually neglected by engineers designing the powering system. This void was filled up by front-line crew with a trail-and-error approach. The paper tries to remedy this situation by presenting a more rigorous approach to the design of the overcurrent protection system.

To protect the network against overcurrents, the elements of the protection system must react first to excessive currents. Hence, by definition, they are the weakest links of the network. This characteristic alone can lead to a dramatic degradation of network reliability if the overcurrent protection devices are not appropriate or the entire system is not designed properly. The best solution would be to avoid these devices altogether. Unfortunately, we somehow need to protect equipment against overcurrents and to isolate failures to minimize failure groups and speed up troubleshooting efforts.

Objectives of Fusing in HFC Network

The two primary objectives of fusing are:

1. to protect components and equipment from costly damage caused by overcurrents; and
2. to lower failure groups.

These objectives are subordinate to the main goal of improving network availability. Both elements of the availability — reliability and mean time to repair — must be addressed during overcurrent protection design.

To improve network reliability, the number of fuses must be limited to the minimum and they must be selected to minimize nuisance blowing caused by voltage transients, current surges, and temporary cable short condition while still protecting the network elements and isolating power faults. To minimize the amount of time required by a technician to locate a fault, fuse locations and their values must be selected according to rules of selective coordination in order to explicitly isolate faults.

Many system operators are trying to meet the objectives of the overcurrent protection system by using circuit breakers. However, circuit breakers have two characteristics that make their use difficult:

- tripping under overcurrent conditions, and
- high thermal dependency (significantly higher than time-delay fuses).

The first characteristic appears to help avoid nuisance outages but actually leads to increased annoyance of customers with intermittent service. The second characteristic makes it difficult to select elements that meet the inequalities for value selection in most locations that would require overcurrent protection to meet the objectives of the overcurrent protection system. Therefore, at the present status of technology, we considered fusing elements the best choice to meet the overcurrent protection system objectives. This decision was based on theoretical consideration and field experience. When used properly, fusing elements can provide better network performance than circuit breakers.

FUSEOLOGY[34]

There exist general misconceptions about fusing in cable TV coaxial distribution network. The origins of these misconceptions are numerous but most of them can be traced to the interpretation of the rules of fusing in power plant and power installations. A direct implementation of these rules results in lower reliability of the network without meeting the objectives of equipment protection and fault isolation.

Selection Criteria for Fuses

The first decision is related to the location of the fuse. This decision will depend on the answer to a question whether a fuse in the particular location can meet the objectives of fusing. If the answer to this question is positive, the following criteria will decide type, value, and other parameters of the fuse selected (the most important criteria are bolded):

***1. Maximum operating load current*:**

The fuse must have an ability to carry the maximum operating current of the network. This ability can be affected by several factors. One of them is the way the fuse is installed. For example, the material of the fuse holders (clips) must have superior retention (such as beryllium copper) to avoid increase in contact resistance. Moreover, the printed circuit board (PCB) traces or conductor size connecting the fuse to other circuit components must be designed to dissipate heat.

***2. Ambient temperature*:**

Ambient temperature is one of the major factors affecting the fuse rated current. The impact of the temperature will be different for different fuse types. For time-delay fuses, the rated current at

60°C is only 80% of the rated current at 25°C, whereas for fast-acting fuses, the rated current at 60°C is between 95 and 97.5% of the rated current at 25°C.

3. Aging:

Aging of a fuse causes the nominal value of the fuse to decrease due to the effects of the continuous current and overcurrents through it. Overcurrents are currents above the continuous rated current but are either not quite high enough to blow the fuse or are higher than nominal but do not last long enough to blow the fuse. These overcurrents can exist in a fuse for several seconds before a lower-value fuse, closer to the short, finally blows. So aging is a combination of time period for which a fuse operates and overcurrents that blow other fuses first but can affect the fuse in the cascade by reducing its current passing capability.

4. Operating voltage:

Generally, the voltage rating of the fuse must be greater than or equal to the operating voltage. This is critical only when the fuse is trying to open. A fuse must be able to quickly extinguish the arc after the fuse element has melted and prevent the system open-circuit voltage from restriking across the open element and passing energy to the protected element.

5. Frequency of the powering voltage:

For current passing fuses and power pack input fuses, the frequency will be 60 Hz or lower (note trials with 1 Hz or DC powering). Hence, most fuses will be suitable for the application. Fuses on the output of switch mode power packs can be subject to additional heat caused by harmonics of the nominal frequency.

6. In-rush characteristics of the network, surge currents, and maintenance procedures:

These factors will cause nuisance blowing if the fuse type and rated current are not adequate. Switch mode power packs used in modern RF equipment for higher power efficiency cause in-rush current at re-starts (up to 10 times of the nominal current). Current surges on coaxial network and incidental temporary shorts caused by improper maintenance practices are common. Time-delay fuses (or Slo-Blo) are ideal for these conditions. Fast-acting fuses may have to be rated at 150% to 300% of maximum operating currents to avoid nuisance blowing. The best parameter describing the ability of the fuse to withstand these conditions is I^2t or "Ampere Squared Seconds". This parameter describes how much total energy the fuse can dissipate before melting. This ties with the aging of the fuse.

7. Available maximum fault current and long term overload levels:

A fuse must be able to open under fault current or long-term overload conditions without damaging the fuse case. The interrupting rating of a fuse is the maximum current at the rated voltage that the fuse can safely open without rupturing or cracking the fuse case.

8. I^2t characteristics of the devices to be protected:

A fuse must be able to limit the energy passed before it completely opens to the level lower than the energy withstand rating of the device being protected.

9. Allowed voltage drop across the fuse:

Fuse resistance (cold and hot) should be negligible to avoid voltage drop that would affect network operation.

10. Applicable standards/agency:

UL listing in the USA is required.

11. Ergonomics (ease of removal, axial leads, visual indication, physical size, etc.):

These factors, often neglected by manufacturers and equipment evaluation/approval personnel, often affect the other factors (mean-time-to-repair — MTTR, incidental shorts, etc.). Physical size of the fuse can be limited by available space and packaging density, making handling during replacement more difficult.

12. Humidity:

In extremely humid conditions, sealed fuses should be used.

13. Standardization:

Standardized fuses are readily available to maintenance crews.

An additional issue of selective coordination must be considered if more than one fuse (or other current limiting device) are cascaded. Coordination is the act of isolating a faulted part of the network from the remainder of the network. Selectivity means positive coordination over the entire range of fault currents, assuring that faulty part is cleared by the first fuse between the fault location and the power source, counting from the fault, and that other parts of the network are not affected.

All these parameters and design rules should be taken into account in setting the fusing design rules for HFC networks.

CONDITIONS IN HFC NETWORK

Maximum Operating Load Current

A nominal maximum load current in most networks is limited to 15 A but may be as high as 20 A. Selecting fuses based on this assumption would lead, however, to a significant degradation in the network reliability. The cable TV network is subject to such phenomena as sheath currents and other disturbances that randomly fluctuate (increase or decrease) current load but are quite normal and common. The equipment is designed to withstand such operating conditions. Moreover, although the main line equipment current passing capability is specified for the nominal operating currents of 15 or 20 A, it can withstand overload currents of 25 A for a prolonged period of time (sufficient to remedy the fault conditions) albeit at derated performance. Hence, only limited protection is required for the main line equipment.

Increased protection may be required for distribution lines that use less robust equipment (lower current passing capacity). Operating currents for these runs typically do not exceed 8 A (may approach this number in urban areas) and are usually significantly lower.

Temperature & Aging

The fuses will be subject to temperatures as high as 80°C for summer months inside actives (the most likely location of fuses) or as low as –30°C for winter months if placed in line passives. The fuse selection process must take into account the highest operating temperatures. Aging would also be a significant factor for fuses in cable TV network.

Operating Voltage & Frequency

A modern HFC network will operate at 60 or 90 VAC nominal RMS voltage, 60 Hz (or possibly 1 Hz).

In-rush characteristics of the network, surge currents, and maintenance procedures

Time-delay fuses must be deployed. Alternatively, the rating of the fast-acting fuses must be increased to provide the same delay time for short-duration current surges to avoid nuisance blowing. For detail analysis of this issue refer to Figure1.

Available maximum fault current and long term overload levels

The fault current in coaxial section of the HFC network is limited by current limiting characteristics of power supplies or by cable resistance and is lower than 25 to 35 A.

I^2t characteristics of the devices to be protected

Newer main line equipment will withstand 25 A current for 2 hours. Older equipment and distribution equipment (line extenders and multitaps) will withstand 135% of their nominal current passing capacity (13-18 A for trunk amplifiers and LEs; 10-12 A for multitaps) almost indefinitely.

FUSING DESIGN

Fuse Locations

In a network with newer equipment designed for high fault currents, a fuse-link protection against overcurrents is required only for distribution (multitap) runs. All other network elements are protected by a current limiting at the power supply locations and by coaxial cable resistance and are designed to withstand these currents.

The other objective — isolating a fault — may require additional fusing on these main line legs that serve less dense areas or branching runs with lower number of homes passed.

The following locations are the only locations where fuses would serve their purpose:

1. Branching (sub-ordinate) main lines;
2. The last fusible location before multitap run and other equipment with limited current passing capability (usually a minibridger output port or line passive power directing port);
3. Output of a second active (second line extender) in tap runs in locations remote from the power supply.

Fuses should be installed in these locations only if the protection improves the plant reliability and improves MTTR through effective fault isolation. Actual fuse placements in these locations will depend on further considerations listed below.

Based on these objectives, the following practical guidelines for fuse locations are recommended:

1. No fuses to be placed in any location in which open or short would cause an outage to more than 50% of homes passed in the power supply area.

2. No more than one fuse to be placed between the power supply and a location in which open or short would cause an outage to more than 25% of homes passed in the power supply area.
3. No more than two fuses to be placed between the power supply and a location in which open or short would cause an outage to more than 10% of homes passed in the power supply area.
4. No more than three fuses to be placed between the power supply and a location in which open or short would cause an outage to less than 10% of homes passed in the power supply area. The third fuse to be placed only in long cascades.

Fuse Values

The fuse values in the locations listed above will be determined by the operating current, available fault current and the type of the fuse.

$$FuseRating \geq \frac{NormalOperatingCurrent}{0.75 \bullet K_{HT}}$$

where

K_{HT} is max. operating temperature correction factor (fuse type dependent).

$$FuseRating \leq \frac{Max.AvailableFaultCurrent}{K_{2h} \bullet K_{LT}}$$

where

K_{2h} is percentage of fuse rating for 2-hour opening time (fuse type dependent)

K_{LT} is min operating temperature correction factor (fuse type dependent).

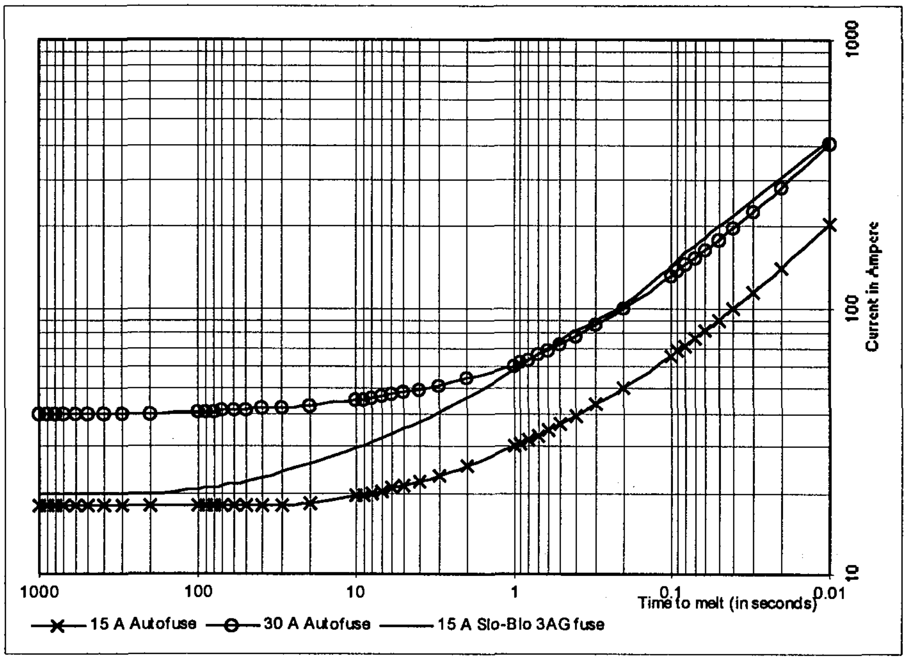
These two limitations decide about a rating of the fuse for a particular location. If these inequalities cannot be met for the location, either the fuse should be relocated or not placed at all.

Fuse Types

Time-delay fuses are preferable for the application in the HFC network and should be the only ones used in that environment. However, many OEMs shifted from 3AG and other type fuses available in the time-delay design to automotive type fuses of different fuse holder design. The automotive fuses are not available in the time-delay design. An alternative is to use fuse adapters that would accommodate time-delay fuses in equipment with automotive-type fuse holders.

A comparison between Fast-Acting and Time-Delay fuses is presented in Figure 1. Two characteristics: time to open at a given current and immunity to transients and surges (I^2t) are compared.

Figure 1: **Comparison of Fast-Acting and Time-Delay Fuse Characteristics: Average Current-Time Curves and I^2t (ampere-squared-seconds)**



	15 A Slo-Blo Fuse (Littelfuse #313 015)		15 A Autofuse Fuse (Littelfuse #257 015)		30 A Autofuse Fuse (Littelfuse #257 030)	
Current Overload	Opening Time	Nominal Melting I^2t	Opening Time	Nominal Melting I^2t	Opening Time	Nominal Melting I^2t
16.5 A	4 hours min.	1870 A^2sec	100 hours	340 A^2sec	infinity	1510 A^2sec
20.25 A	1 hour max.		½ hour max.		infinity	
30 A	5 sec. min.		0.15 sec. min.		infinity	
33 A	10 sec. av.		0.6 sec. av.		100 hours	
40.5 A	5 sec. av.		0.35 sec. av.		½ hour max.	
60 A	1.4 sec. av.		0.1 sec. av.		0.15 sec. min.	

Comments:

1) The 15 A Slo-Blo 3AG-type fuse can withstand 100,000 current pulses with 80 A peak value (corresponding to in-rush current with switch mode power supplies) of exponential shape with a decay time of 0.3 s each (20 cycles) or almost unlimited number of pulses of 10 kA peak value with a decay time of 10 µs (corresponding to lightning strikes). It can be used effectively for fault isolation.

2) The 15 A Autofuse can withstand significantly lower number of pulse overcurrents (1 pulse at 80 A peak current of 0.3 s decay time or approximately 20 pulses at 10 kA peak current of 10 µs). It can be used to isolate faults but at limited pulse immunity in comparison to 15 A Slo-Blo fuse that is also used for this purpose.

3) The 30 A Autofuse provides similar immunity to pulse overcurrents as 15 A Slo-Blo 3AG-type fuse but will not react to any current overloads existing in the HFC network (power supplies will limit the available short current to 25 A).

Selective Coordination[5]

To meet the objective of selective coordination, the following minimal fuse value ratios must be maintained for fuse cascades:

- for time-delay fuses of the same type - 2:1
- for fast-acting fuses of the same type - 3:1 in most cases

- fast-acting followed by time-delay - 1.5:1 in most cases
- time-delay followed by fast-acting - 4:1 and higher in most cases.

These ratios assume that the fault currents at the fuse locations are not significantly different.

The data show that fuse mixing or using fast acting fuses will significantly limit the number of fuse values that can be used in cascade if the rules of selective coordination are applied. Alternatively, the values of fuses farther in the cascade would be low and would result in their nuisance blowing. For example, if fast acting fuses are used and the protection and fault isolation rules ask for three fuses, their values would be 15 A, 5 A, and 1 A (next lower value available in Autofuse style). Note that the locations in which the 1 A fuse could be used would be very limited (low nominal operating current) and its immunity to transients would be dangerously low (I^2t=0.4 A^2Sec).

Voltage Rating

The voltage rating of the fuses should be 125 V or higher. Unfortunately, Autofuses are available only with 32 V rating, and 3AG Slo-Blo fuses are available with 250 V rating for values of up to 8 A. For higher current rating, the voltage rating is limited to 32 V. Fortunately, fuse selection guides allow for exceptions to the voltage rating rules if the maximum power available at the fuse under "dead short" is limited (currents lower than ten times the nominal operating fuse current).

Fusing Design Steps

The fusing design for cable TV plant shall follow the order described below:

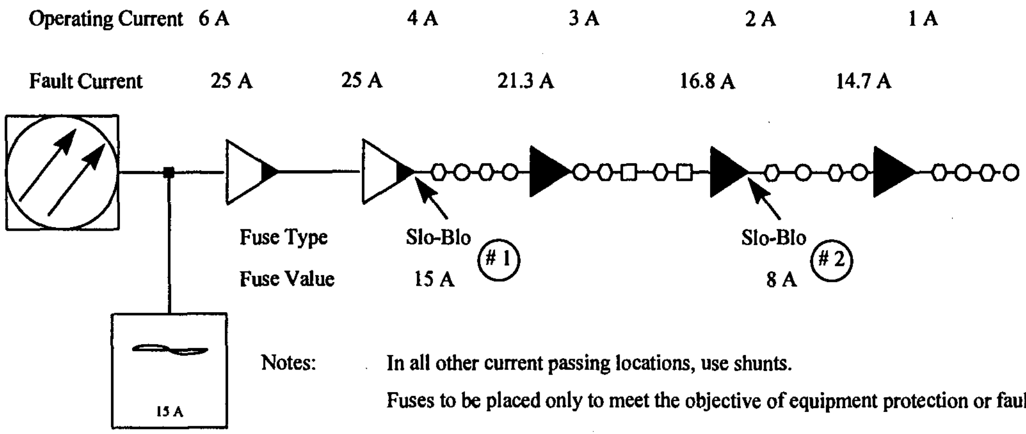
1. Select fuse locations to meet the overcurrent protection objectives.
2. Calculate the allowed fuse rating range for all the location.
3. Starting from the location farthest from the power supply, select cascaded values for the fuses within the ranges calculated in step 2 and according to the rules of selective coordination.
4. Repeat this procedure for all fuse locations.

EXAMPLES OF FUSING DESIGN

The following examples of fusing design are applied to the HFC network with node design with distributed and centralized powering.

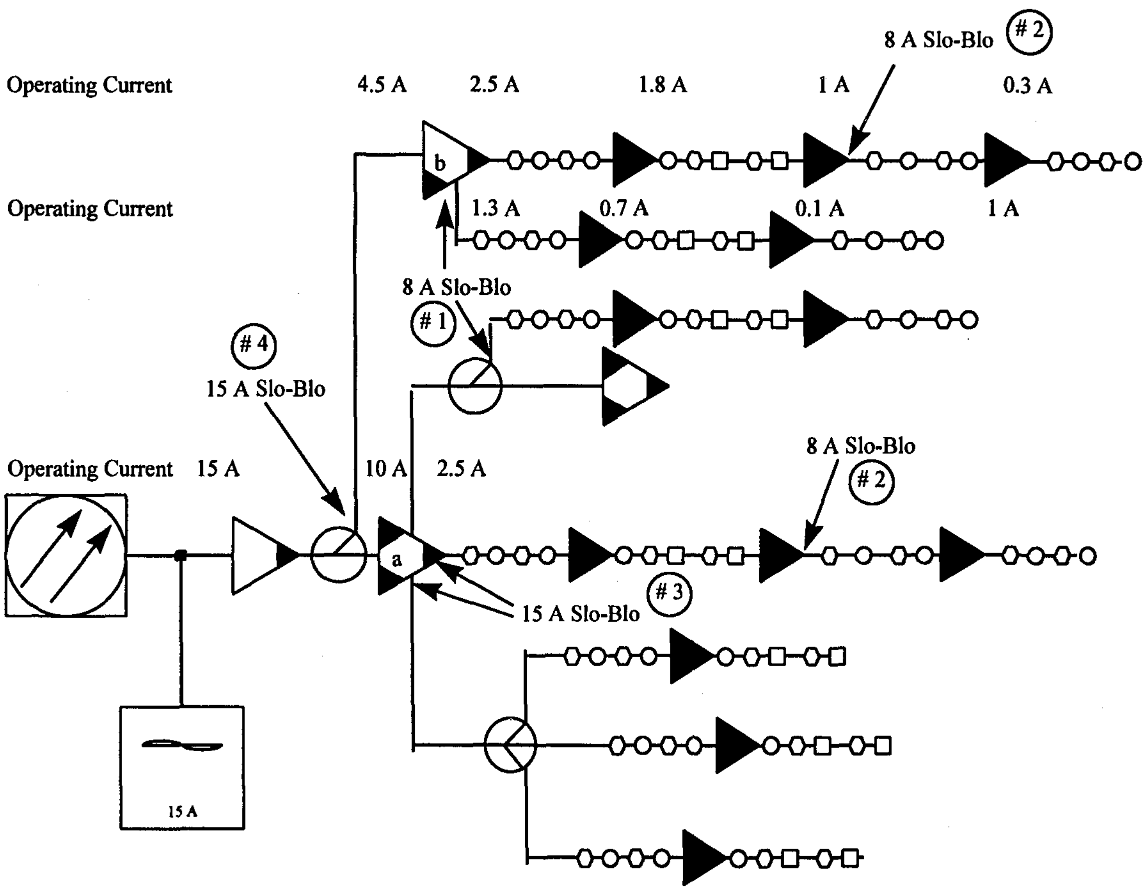
All the rules described above were applied. These rules were also applied to all TCI systems including traditional tree-and-branch designs. Although complete comparative data have not been collected yet, first results indicated significant decrease in fuse related outages. The implementation effort was huge due to the fact that each system developed its own procedures of deploying fuses, circuit breakers, and other means, often with the advice of the manufacturers. These piece-meal solutions, directed mostly towards protecting equipment of lower than required robustness, never accounted for the powering design specifics and never applied the system-type approach to the overcurrent protection design.

Figure 2: Simple Example of Fusing with Centralized Powering



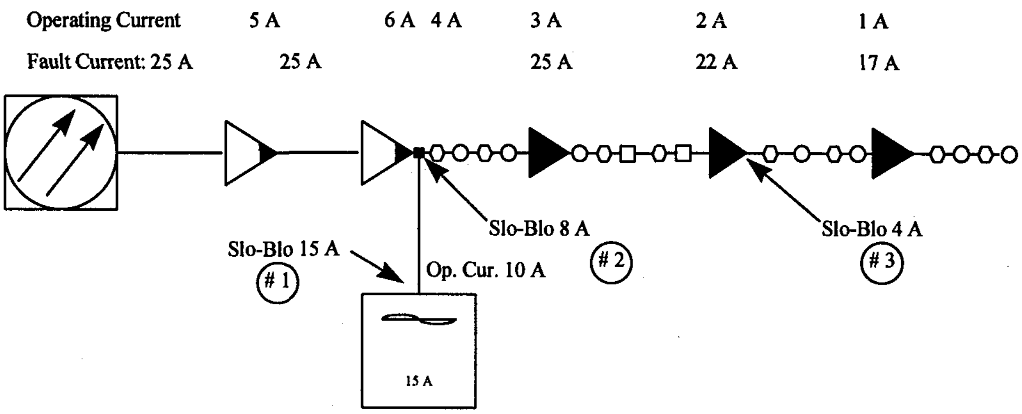
Notes: In all other current passing locations, use shunts.
Fuses to be placed only to meet the objective of equipment protection or fault isolation

Figure 3: HFC Node Fusing Design for Centralized Powering



Notes: In all other current passing locations, use shunts.
Fuses to be placed only to meet the objective of equipment protection or fault isolation

Figure 4: Simple Example of Fusing with Distributed Powering

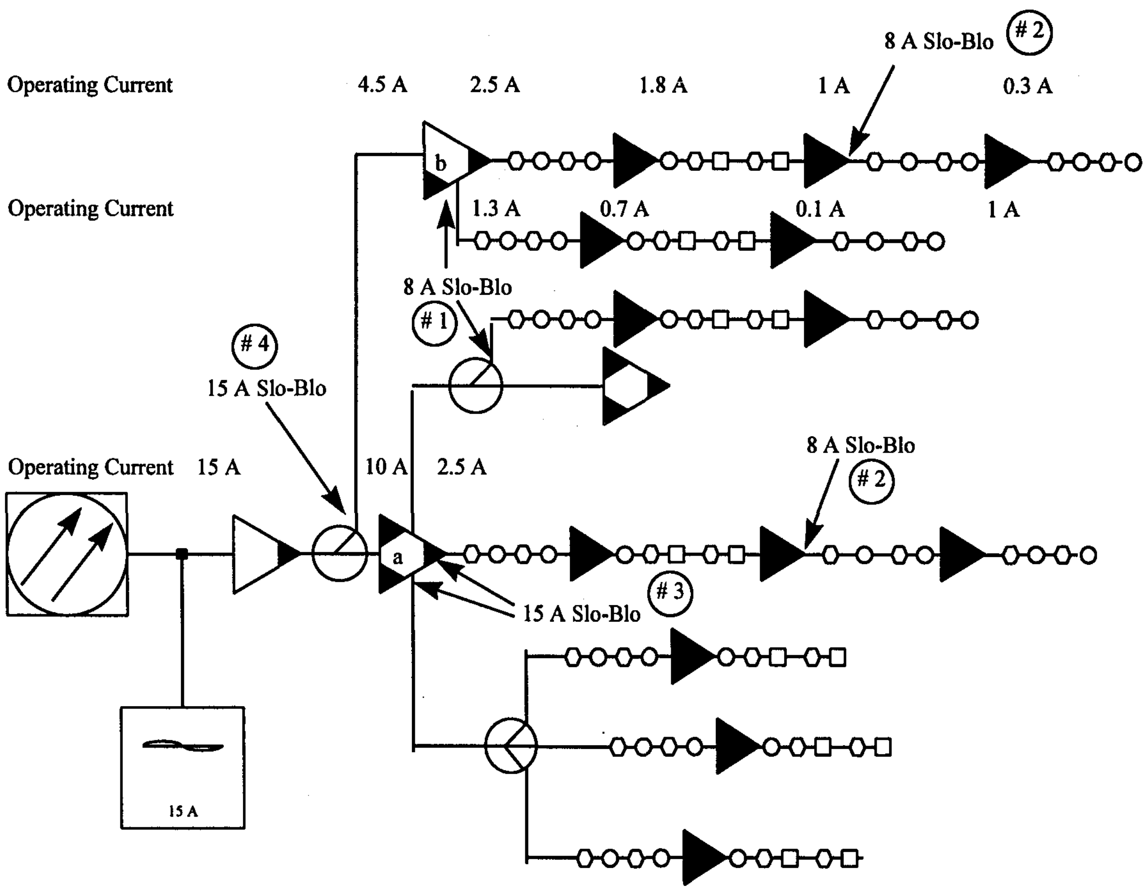


Notes: In all other current passing locations, use shunts.

Fuses to be placed only to meet the objective of equipment protection or fault isolation.

Use only shunts between the power supply and the node if the node is powered from this power supply.

Figure 5: HFC Node Fusing Design for Distributed Powering



Notes: In all other current passing locations, use shunts.

Fuses to be placed only to meet the objective of equipment protection or fault isolation

SUMMARY OF FUSING DESIGN

Time-Delay Fuses vs. Other Overcurrent Protection Elements

Time-delay fuses have characteristics that make them the most desirable for overcurrent protection systems in coaxial sections of the HFC network. 3AG type fuses have the highest values of I^2t for any fuse rating

Fuse Values and Fuse Cascades

The two most suitable values to meet the value selection equations and requirements of selective coordination are 15 A and 8 A. One more value, 4 A, for locations remote from the power supply can be added to the two selected. Lower values would result in increased nuisance blowing due to too low I^2t values.

The 15 A fuses can be used for operating currents of up to 8 A, 8 A fuses for operating currents lower than 4.5 A, and 4 A fuses for operating currents of up to 2.25 A.

ACKNOWLEDGMENT

The author wishes to thank Tom Shirk, a TCI filed engineer, for his work on creating practical guides on selection of fuse locations based on the calculations and fusing examples.

[1] Doug Combs, Oleh J. Sniezko, Tony E. Werner, Powering: Architecture and Voltages, 1996 NCTA Technical Papers.

[2] Tony E. Werner, Oleh Sniezko, "Network Availability Requirements and Capabilities of HFC Networks," SCTE 1996 Conference on Emerging Technologies, San Francisco.

[3] Electronic Designer's Guide, Littelfuse, Form No. EC101-A

[4] Electronic Designer's Protection Handbook, Bussman, Bulletin EDPH

[5] Selective Coordination of Overcurrent Protective Devices for Low Voltage Systems, McGraw-Edison/Bussman, Bulletin EDP-2

Hardware and Operating Considerations of The Decoder Interface

by
Joseph B. Glaab
General Instrument

Abstract:

The hardware specification of the Decoder Interface is finished. The control software is almost finished. Products incorporating the Decoder Interface will soon appear on the market. This paper describes the details of what will be in the TV, VCR, and decoder, and how these units will interact to create a consumer friendly system which permits watching one program while recording another, recording successive programs on different channels, and doing this whether the channels are in the clear or scrambled. In addition, none of the features of the TV, such as picture in picture (PIP), are kept from working as originally intended.

Introduction:

The Decoder Interface is being standardized with two degrees of complexity. The Level I, "Cable Ready" standard is intended to satisfy the requirements for basic "Cable Ready" as outlined in the "Cable Television Consumer Protection Act of 1992" which are:

1 - Ability to watch a program on one channel while simultaneously using a VCR to tape another program on a second channel.

2 - Ability to tape two consecutive programs that appear on different channels.

3 - Use of advanced picture generation and display features.

4 - All of the above while realizing "the need for cable operators to protect the integrity of the signals transmitted by the cable operator against theft or to protect such signals against unauthorized reception".

The Level II "Advanced Cable/Media Ready" standard has a less formal set of requirements, but adds remote control pass through from the TV or VCR to a subsequent user as well as adding a variable number of the features from the tool box of AVBus functions. An example of this would be a VCR or TV with video and/or audio <u>outputs</u> via the Decoder Interface connector and cables. The "Cable Ready" version Decoder Interface hardware requirements, as specified in EIA-IS-105.1, sets minimum requirements, while permitting greater functionality. There are some pieces specified to accommodate a full system, while other elements are optional or reserved for the Level II Decoder Interface.

The Bus Cable:

The thirteen twisted pairs of conductors, of which the cable is composed, are not shielded. Connectors on the ends of the cable are a small "D" shaped shell which latches onto the mating connector, and uses "ribbon" contacts. The maximum length of the cable, or the sum of the pieces which can interconnect up to ten devices, is ten meters. The cable cannot be "star" connected and the pairs specified for video must be terminated at both cable ends. The cable is the same for Level I and Level II systems. Level II is a part of the AVBus system which allocates the thirteen pairs as; 7 video, 4 audio, 1 control, and 1 reference.

Tuner IF Cable:

The key to consumer friendliness is the use of as much of the TV or VCR hardware as possible, while maintaining all of the TV or VCRs' operating features. With this goal, the channel selection is done by the TV or VCR tuner. The tuner output is split electrically so that the TV or VCR can process clear channel video and audio signals, and the clear audio part of a scrambled signal when the video portion is scrambled. The portion of the tuner IF sent to the descrambler is level controlled by the decoder, via the tuner AGC when the decoder is being used. The maximum length of the IF cable is two meters. If two to four decoders share use of a TV or VCR tuner, an "IF Expander", which includes optional decoder powering, is defined.

Cable Ready TV:

As described above, the cable ready TV has a minimum of Tuner IF output, with tuner AGC returned on the cable, and the following decoder interface bus receivers; V1 - Video Input, A1 - Audio Input, Control bus and the Reference pair. The video and audio is delivered to the TV on twisted pairs and is converted from differential to single ended. From this conversion on, the processing is as if the signal arrived to the TV jack panel. Control is the key to this system and the detailed operation will be discussed later.

Cable Ready VCR:

This model follows the Cable Ready TV description above with the exception that the bus receivers are V3 - Video Input, and A3 - Audio Input.

Decoder:

The decoder can be used to supply signals to the TV or VCR by using the Decoder Interface Bus. While the CATV operator supplied descrambler is the decoder most often envisioned, the decoder could just as well supply video and/or audio from another source such as ADSL telephone lines, video with overlays from program guide decoders, etc. The minimum configuration requires outputs to be available on either or both V1, A1-A2, and V3, A3-A4. The decoder requires only bus transmitters. Notice that while TVs and VCRs only have to have a single audio input, Decoders must supply signals on both of the audio pairs, even if the signal is monaural. If audio privacy, or stereo audio privacy is in use in the system, the decoder must have an audio demodulator, and possibly a stereo demodulator.

Shared Decoder:

The Decoder Interface specification defines a shared decoder. This decoder has the ability to accept input selectively from two IF Tuners, and to supply an output to either of the tuners' host TV or VCR. This will permit either watching or recording one scrambled signal while recording or watching either the same signal, or a clear signal. This decoder will possibly be a popular version since this is very close to the single converter useage today. The control of this unit has a special control interest because of the necessity of establishing which decoder IF Tuner port is connected to which Tuner.

Two Decoders - Two tuners:

When two identical signal type decoders are connected to two tuners, each operates as if the system is not interconnected. When the decoders are different, analog vs. digital for instance, a cross connection methodology must be developed. The situation could be that the TV viewer wants the digital program, while it is desired to record the descrambled analog signal. The reverse situation could be desired for the next program session. This cross connection requires an IF switch which cascades decoders, or terminates (Power blocked) the unused IF tuner port if only one decoder is connected. In a similar manner, the decoders used in this configuration must also have audio demodulators, including stereo demultiplexing, if audio and/or stereo audio privacy is in use. Of course, if the decoder is digital, audio reconstruction is always necessary.

IF Switch:

The IF switch will probably be built into the decoder, although that is not mandatory. The switch must maintain DC continuity to the IF paths since the tuner AGC passes that way. Seventy five Ohm impedance and greater that 60 dB of isolation between ports must be maintained which should not be too difficult at the 41-47 MHz IF operating frequency.

Control:

The control follows the CEBus model of pulse width modulation. A "one" is 100 microsecond (μsec) long, a "zero" is 200 μsec, "end of field" 300 μsec, and "end of packet" 400 μsec. The control system is "not asserted" or esentially open circuit when idle. If a unit desires to send, it can do so if the bus is idle. Bits are alternately asserted and not-asserted. Since the system is short and the bits are long there is no possibility of "flying bits", (bits sent and in transit which are not yet received by a unit desiring to send). If the second sender, sends its' first bit , a collision occurs. Both senders detect the collision, stop sending for random lengths of time, and statistically, on a subsequent transmission the second unit detects that transmission and refrains from sending. In a short system such as the Decoder Interface, two units could possibly start sending simultaneously. Since both monitor the line continuously, the first unit to send a "non-assertive" bit while a second unit is still asserted will recognize the collision, and stop sending. In this way, the first unit does not even see the contention and continues its' sending. The control drivers and receivers are relatively simple and are specified as "fail safe" non-assertive even when the unit is unpowered.

Protocol:

The protocol follows the CAL-60 format, which is very extensive. Since the Level II Decoder Interface is AVBus compatible, the packets include addressing to units and network extensions to zones, called house codes. This permits operation of multiple systems in an area when some of the control system are bridged onto a common transport medium such as the power lines. The network interface has been implemented in relatively simple microprocessors. They receive the packets from the control data receiver, establish network and address compatibility and pass the message content on to the device controller. In a similar manner, a unit can send data to a target address, packaged up and transmitted along with an error detection checksum by the data link microprocessor.

Network Architecture:

Establishing what devices are available, device compatibility, and device interconnections are the subject of this section. Assuming that the components are TV, VCR, and decoder(s), the consumer connects the units together with appropriate coaxial cables for IF connections, and AVBus Cable for the Video, Audio, and Control. Upon completion of the connections, the units are plugged into the AC Power. When this occurs, probably at a somewhat staggered rate because all the plugs will not be inserted simultaneously, all of these otherwise intelligent units must evaluate their environment. The individual units know their own capability, and are able to query other devices as to their capability. The first step in such an operation is to acquire addresses, equivalent to names, so that targeted questions and answers can follow. Generally low number addresses are desired so as to minimize the number of bits necessary to be sent to complete an addressing sequence. The zone or house address must be established for all devices. On a Level I network, and probably most other networks intially, the home will be #1. Devices then "hail" for addresses. Certain preference addresses are recommended in the standard, however the general rule is to select an address and "hail" to see if anyone already has that address. Since only one unit can access the control bus at a time, if no response contesting that address is forthcoming, that is the unit address. The analogy could be likened to a group of people in a dark room, with some of those people roped together in pairs. Everyone can listen, however only one at a time can talk. The challenge is to sort out the configuration. After everyone knows the name of everyone else, it becomes necessary to establish device types such as TV, VCR, and Decoder. In the case of decoders, the IF interconnection must be established. A shared decoder must have two addresses, one for each IF tuner.

IF Interconnections:

After addresses are established, the decoder(s) must decide which tuner they are connected to. The algorithm to resolve interconnections is left to the decoder designer. One method would assume an interconnect and for instance re-tune the TV or VCR and see if the video changes. An alternate would be to control the AGC to reduce the signal level, and see if the signal level changed. With an IF Expander box present, there could be multiple decoders connected to a tuner IF. Not all decoders need to control the tuner AGC, so interconnect analysis may vary with decoder.

Hail for A/V lines:

When the units are initializing, it is necessary to evaluate video and audio line availabliity. Some minimum configuration availability is mandated, as previously discussed. In a simple situation, V1 would be allocated for decoder to TV interconnect, while V3 would be allocated for decoder to VCR. If the decoder was digital, or analog with audio privacy, then A1-A2 for the TV and A3-A4 for the VCR would be allocated. If the TV was monaural, the decoder would plan to feed both A1-A2 with the composite audio, while if the TV was Stereo capable, the decoder would stand ready to feed stereo if available. While the exact algorithms are not in place, a decoder would not actually acquire and enable the video and/or audio lines until the first time that they are required. The recommended practice for either holding onto the lines, or releasing them back to TV set operation, for instance when channel surfing, has not been established. It might be expected that the methodology of converter operation might be emulated. For instance tuning to an authorized scrambled channel would cause the descrambler to acquire the necessary A/V lines, while tuning to a "not-authorized" scrambled channel might cause the descrambler to retune the TV tuner, unlocked from display so the TV still shows the selected channel, to a barker channel. Optionally the descrambler might acquire A/V lines and put up a "Barker" graphics or just a blue screen. The system has great flexibility!

Problems to be Solved:

One problem occurs when a shared decoder is already switched to a scrambled IF and the second tuner lands on a scrambled channel. The decoder is not available. The general prioritization would be that a VCR has a higher priority, however the TV viewer of a scrambled channel should probably be given an option when the pre-programmed VCR requests the descrambler switch over to the VCR IF. Then too, how does the VCR know that the channel is scrambled and authorized if no IF is available to analyze?

Another problem for the designer will be how to utilize the functions and features available in the hardware at hand. For instance, does a descrambler have to assume that an IF expander box is present, or can an assumption be made that the IF port is its' alone? By listening to other decoders who think that they are connected to the same host IF tuner, a deduction of multiple decoders can be made. Does the decoder firmware have to accommodate this from the start?

Some of the advanced thinking is that with a full AVBus system, it will be possible for a TV receiver to output a composite video signal, which a decoder separates into "Y,U,V,". These components then go to a processing decoder which overlays video information and passes the processed "Y,U,V," back to the TV for display.

Summary:

While some of the potential hardware and operations of the Decoder Interface have been discussed here, the number of actual combinations of this powerful network structure have hardly been scratched. The first Decoder Interface equipped TVs and VCRs should appear on the market shortly, and decoders of the descrambler variety will work together to make the TV appear to have descrambling capability. The purchasers of these new systems should be pleased with the progress that the Cable and Consumer Industries are making in working together.

References:

EIA-IS-105.1 Draft Decoder Interface Standard
Revision: 6.1 September 8, 1996

EIA-IS-105.2 Draft Decoder Interface Control Standard
Revision: 2.1 September 5, 1996

Cable World special reprint of "Cable Television Consumer Protection and Competition Act of 1992", Page 12A, October 12, 1992
Reference "The Communications Act of 1934, amended "Sec.624A Consumer Electronics Equipment Compatibility"

Joseph B. Glaab, "History and Present Status of The Decoder Interface: 1995 NCTA Technical Papers, Pages 346-349

High Speed Internet Access Using Cable Modems with Telephone Return

Jonathan Fellows Ph.D., Victor T. Hou
General Instrument Corporation
Communications Division

Abstract

For many current cable service areas, telephone return cable modems may be a viable long term technology for delivery of high speed data services. The first generation of commercially deployed telephone return cable modems and associated headend equipment have now been installed in multiple cable plants. This system, incorporating 64-QAM transmission in the downstream and telephone return in the upstream, provides high speed access to the Internet and other data applications.

Many of the network design issues of telephone return cable modems, such as network addressing schemes, asymmetric bandwidth, and system management, will also apply to two-way systems. Other issues, such as integration with dial-up networks, and asymmetric routing, apply uniquely to one-way systems. Experience with performance tuning of TCP in asymmetric systems is described, as well as field experiences with installation and configuration of the modem in customer PCs.

TELEPHONE RETURN RATIONALE

This paper asserts that cable modems with telephone return paths are a practical option for delivery of data services in broadband cable plants, both in the near term and in the future. The remainder of the section motivates this assertion.

One of the dominant motivations for telephone return cable modems is that, in an era of capital scarcity, the cost of cable plant upgrades favors telephone return schemes. Due to the robust error correction capabilities of modern digital television modulation schemes, telephone return cable modems are capable of operating at high speed in practically any current cable plant.

Another motivation is that cable return path technology is not yet fully mature - no de facto standard has emerged in the marketplace. The engineering challenges of the cable uplink are not reduced to common practice. Finally, while there are proof of concept systems in use, there are currently no working prototypes of any of the proposed cable return standards.

Existing web-based Internet services have highly asymmetric bandwidth usage profiles, with much more data traffic directed toward the client than returning from the client. This traffic pattern is well suited to a cable forward, telephone return architecture. Telephone return cable modems support existing Internet services, such as the World Wide Web, at speeds that are at least two orders of magnitude higher than telephone-only Internet access.

The speed achievable using telephone return broadband distribution techniques is a major market discriminator - people will pay for existing services, only faster. The market does not need to wait for new "killer applications".

TECHNOLOGY OVERVIEW

Telephone return cable modem technology brings high bandwidth data connectivity to existing broadband cable networks. Host connections are established in the downstream direction through either internal or external cable modems. Host communications software can use either a telephone modem built into the cable modem or an existing telephone modem within the user PC to establish a dial-up session for return path data traffic. A set of telephone return cable modems that share access to a common broadband channel can be interfaced to a public or private internet through a headend cable router. Current generation downstream modulation techniques are capable of providing a 27 Mbps data subnet in a single 6 MHz TV channel, using 64 QAM modulation. Figure 1 shows the relationship between the above described components.

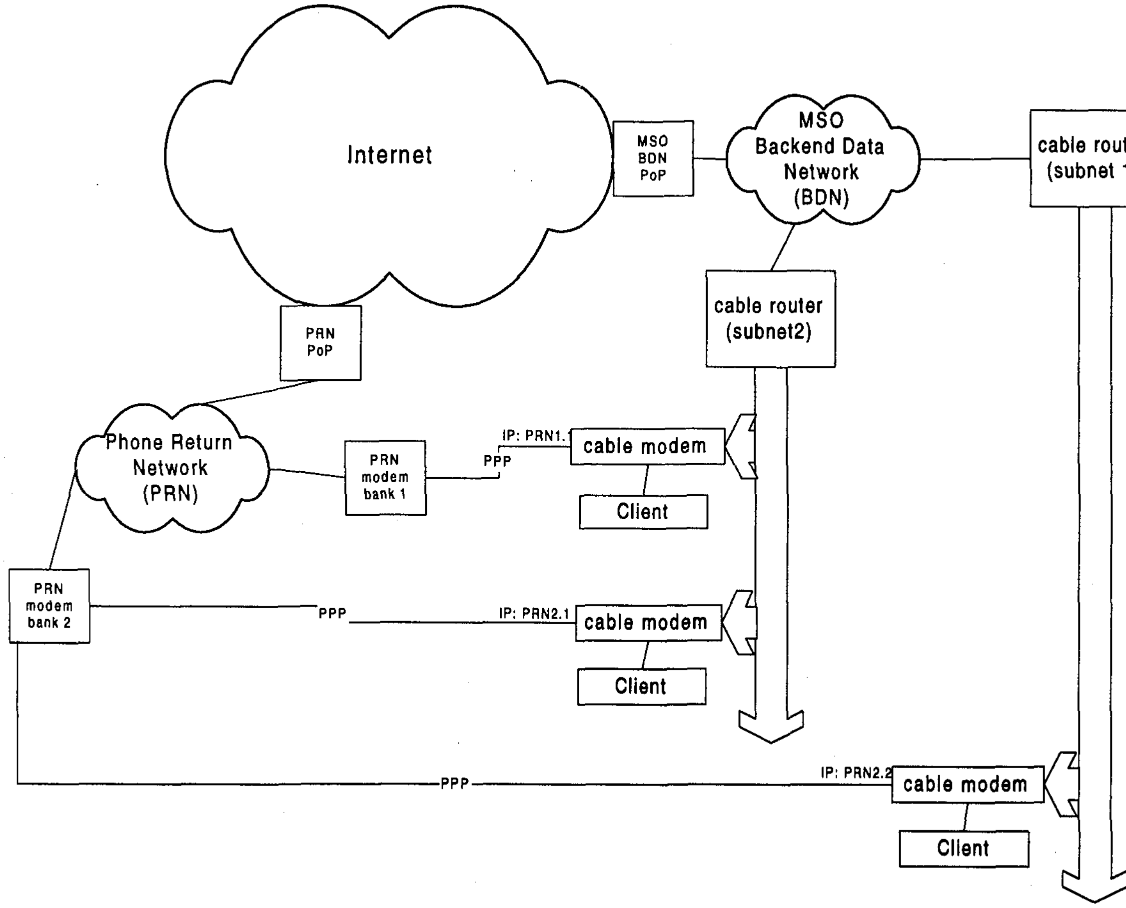


Figure 1 - Telephony Return Cable Modem Network Architecture

DESIGN ISSUES & APPROACHES

This section explores the basic design issues that have emerged for telephone return cable modems, and describes the choices made by General Instrument in its current commercial product line.

Internal vs. External modem

This choice, as it has evolved in current product offerings, involves more than a simple form factor. Current telephone modem products come in both external and internal form factors without changing the way in which they integrate with communications systems. The reason is that in both cases the telephone modem serves as a network interface device for a host that acts as an end system in a network. General Instrument has chosen this approach for its internal cable modem product. This choice offers the advantages of straightforward integration with existing Internet protocols, but requires the product to integrate with a wide range of PCs. This issue is revisited in a later section.

External telephone modems communicate with their attached host using serial I/O interfaces, which does not change the modem's status as a local network interface device. Many external cable modems have chosen to interface with their attached host using an IEEE 802.3 interface [1]. This choice forces the view that the host network interface is the 802.3 Network Interface Card. The cable modem becomes an intermediate system - either a bridge or a router. In either case, the cable modem is required to interoperate with a more complex set of protocols than is the case with an end system interface.

Broadband PHY and Data Link Framing

In late 1996, the North American cable industry agreed on major elements of interoperable digital cable systems. One of these agreements is the use of a downstream transmission standard that conforms to the International Telecommunications Union (ITU) standard ITU-T J.83 Annex B [2]. The specifications in Annex B include 64 QAM modulation, concatenation of a Reed-Solomon block code with an inner Trellis code, interleaving, and a contiguous serial bit-stream.

There are several sensible options for HFC data link framing and Media Access Control (MAC). The options with the best basis in existing commercial practices are Ethernet (IEEE 802.3), ATM, and MPEG-2. Of these, MPEG-2 is currently the best integrated with the ITU J.83B Physical Layer. Neither ATM nor Ethernet currently have widely accepted mappings to HFC Physical Layer protocols. The choice of MPEG-2 also leaves open the possibility of multiplexing data with video services. MPEG-2 is a particularly good choice for telephone return cable modems, since it was designed for one-way operation over HFC cable plants. As a practical matter, choosing MPEG-2 allowed development efforts to be focused on data networking issues.

Addressing and Data Path Selection

Mapping IP addresses onto data link addresses is a key issue in implementing IP over a new data link protocol. This mapping is needed to support IP forwarding in routers and hosts. For IEEE 802 MAC protocols, this mapping is provided by a broadcast-based Address Resolution Protocol (ARP). Use of an MPEG-2 based data link protocol requires a different approach to data link address resolution. This can be based on the association of an IP subnet with each MPEG Transport Stream used as a broadband data channel. The IP subnet is defined using variable length subnet masking, with a maximum subnet size of 4096 host addresses. (The remainder of the 13-bit PID address space is reserved for multicast and system management uses.) The hostid part of an IP address is then directly mapped to a MPEG Packet IDentifier (PID).

Routing vs. Bridging

This is not an issue for internal modems - an internal cable modem is a Network Interface Card (NIC) for an HFC subnet. An HFC NIC filters the HFC MPEG-2 Transport Stream for the PID that corresponds to its host IP address. It then forwards selected MPEG packets to its host for reassembly. A host with an internal cable modem always forwards outbound datagrams to its PPP connection.

It is possible to design an external cable modem as either a bridge (Layer 2 forwarding) or a router (Layer 3 forwarding). After extensive analysis and debate, the authors have come to the conclusion that either configuration can be made to work in two way cable plants, and that the two configurations are comparable in complexity. However, for one way cable plants and telephone return, IP routing provides significantly more flexibility for asymmetric data paths. We have therefore chosen a router based interface for our external cable modem.

The following forwarding conventions are proposed for an external telephone return cable modem acting as a router:

a. Outbound datagrams are forwarded to the cable modem PPP connection.
b. A cable modem filters its MPEG Transport Stream for the PIDs that correspond to any of its configured IP address. It reassembles IP datagrams and forwards them to the attached host.
c. A headend cable router always forwards inbound datagrams from its assigned subnet, using the PID derived from the IP address.
d. Routers in the Backend Data Network have a static route for the broadband subnet that points to the headend cable router.

IP Encapsulation

The choice of MPEG, with its 188 octet fixed size frames requires segmentation and reassembly of payload data from higher protocol levels, in order to achieve Protocol Data Unit (PDU) sizes that are efficient for data communication applications. Fortunately this has already been addressed in the MPEG community with the use of DSM-CC private sections. This entails the use of RFC 1483 encapsulation of payload datagrams[4], which includes the use of LLC/SNAP to denote the protocol type of the PDU being encapsulated These conventions provide for direct encapsulation of IP datagrams of up to 4K octets.

Quality of Service

There has been much discussion of multimedia applications and the Quality of Service support that such applications will require. Multimedia applications are rapidly being adapted to existing Internet delivery methods. Methods for real time delivery of digital sound, which can be compressed to fit today's telephone modems, have improved impressively in the last year. When cable modems provide the bandwidth needed for video applications, Internet delivery methods for video can be expected to improve in a similar fashion. An open question is whether these methods will build on experience with MPEG-2, which requires link level Quality of Service guarantees, or whether they will use methods that rely on more extensive use of buffering.

The currently proposed Internet approach for real time data services is based on the Real Time Protocol (RTP) and the Resource Reservation Protocol (RSVP). The HFC data link is not necessarily involved in either RTP or RSVP. Instead it can provide raw bandwidth which can be managed by the RTP/RSVP protocol implementations in hosts and routers.

Customer LANs

Local Area Networks (LANs) on the customer premises can be accommodated in two ways. When an internal cable modem is in use, the host can be configured as a dual homed host with interface to the cable network and to a local 802.3 network. All that is required is to obtain IP addresses for the LAN (this is easier if addresses are assigned statically) and enable IP forwarding in the dual-homed host. This works today on hosts running the Microsoft NT operating system, and can be made to work with Microsoft Windows 95 with third party software. It is also possible to configure the home LAN so that local hosts have private IP addresses.

When the cable modem is external, support for a home LAN is straightforward. The cable modem acts as a bridge and/or router and supports multiple locally attached hosts. The same addressing choices are available as were described above.

Integration with Dial-up Networks

One of the advantages of an IP-centric network design is that IP networks "just work" under an amazingly wide set of circumstances. Separate routes for incoming and outgoing traffic poses no problems to a standard IP network. This means that integrating a cable modem with a telephone return is simply a matter of setting default IP forwarding addresses appropriately.

System Management

The major system management issues for telephone return cable modems are IP address management, SNMP monitoring, and diagnostic support.

IP Address Management A host with an internal cable modem always has two media interfaces, one telephone, the other broadband. It is possible to configure the host with either one or two IP addresses. A host attached to an external cable modem needs only a single 802.3 media interface and a single IP address.

In the case where the host with an internal modem has one IP address, it is necessary to coordinate IP address assignment in the Telephone Return Network with the broadband data channel being used. This can be accomplished either through static address assignment conventions, or by adding IP address management functions to the dial-in network authentication servers.

In the case where the host with an internal modem has two IP addresses, there is no coordination required between telephone address assignment and broadband address assignment. This is very convenient when the dial-up telephone network is outsourced to an existing Internet Service Provider.

An external modem can be configured with an unnumbered interface on its telephone interface, or it can be configured with a routable IP address. The broadband interface of an external modem can be configured as an unnumbered interface or it can be configured with a routable IP address. So the external modem has two media interfaces and zero to two IP addresses depending on how it is configured

SNMP A telephone return headend cable router should support an SNMP MIB that includes MIB - II. Proprietary extensions are capable of supporting data collection on a destination address basis for usage based billing. MIBs and agents for both cable modems are under development. These MIBs will support HFC transmission parameters and cable modem diagnostics.

EXPERIENCE

Performance

In data communications, two of the most important performance metrics are delay and throughput. For the primary applications of Web surfing and file downloading, the important aspect of throughput is downstream throughput.

One of the characteristics of both two-way cable modem systems and cable modems with telephone return is bandwidth asymmetry. The data rate in one direction is substantially greater than the data rate in the opposite direction. The 64 QAM downstream is capable of an effective data rate of 27 Mbps. With a telephone return modem of, for example, 28.8 kbps, there is a high degree of asymmetry.

Testing shows that telephony return cable modem throughputs for ftp downloads can achieve 3.7 Mbps using a 133 MHz Pentium PC running Windows 95. With simultaneous TCP connections to multiple PCs, an aggregate downstream throughput of approximately 24 Mbps has been achieved. This is nearly 100% link utilization, after accounting for MPEG framing overhead and segmentation quantization effects. It should be emphasized that these are data rates as seen at the application level and not at a lower layer where additional overhead has been added.

Now, we examine downstream latency. It is noted that transmission from the headend is approximately 445 μs for a 1500 byte packet based on a data rate of 26.97 Mbps. The headend cable router buffers downstream traffic, so there is a random waiting time. The propagation delay is proportional to distance and is roughly 375 μs for a reasonable maximum distance of 50 miles from the headend. The delay due to the downstream FEC (Forward Error Correction) interleaving is approximately 4 ms. This is a constant delay. So the downstream latency for a 1500-byte packet to be transmitted by the headend and received at the receive end has a minimum of about 5 ms, dominated by the interleaving delay.

Upstream latencies depend on the distance, the telephone modem rate, the fixed latencies in the modems, and the latencies of the terminal server and routers in the headend network. Of course, there are additional latencies for negotiating through the Internet to the communicating server or much smaller additional latencies for communicating with a local content server in the headend network.

Network Integration

Our design philosophy has been heavily Internet-centric, and this has paid off with a nearly flawless record of smooth integration of the headend cable router into existing IP networks. All that is required is the installation of a static route from an existing router that points to the headend cable router for the HFC subnet.

Client Installation & Configuration:

Installation of an internal cable modem into PCs imposes no particular requirements other than those that would normally be associated with installation of any Network card into a PC. With a "Plug & Play" compliant device, systems running Windows95 will automatically detect the presence of the cable modem on boot up. When installing on a Windows 3.1.X system installation requires the use of an external Plug and Play utility, such as Intel's ICU. Installation in most cases is straightforward for either platform provided that precautions are taken in advance to insure the presence of free IRQs and I/O address space for use with the cable modem. Some "off the shelf" PCs have a default configuration where all IRQs are already assigned leaving none available for the cable modem. For these systems it is necessary to remove a device to free up an IRQ. The most common device selected for removal is MPU 401 (MIDI emulation).

CONCLUSION

The basic point of this paper has been that cable modems with telephony return constitute a technology that meets the needs of today's Internet applications, in a way that is fieldable on today's broadband cable plants. This represents an opportunity for immediate deployment of Internet services that offer end users throughput that is at least two orders of magnitude higher than current generation technologies.

REFERENCES

[1] ISO/IEC 8802 3: 1996 (ANSI/IEEE Std 802.3, 1996 Edition), Information technology - Local and metropolitan area networks - Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications.

[2] ITU-T Recommendation J.83, 10/95, Annex B, Digital Multi-Programme Systems for Television, Sound, and Data Services for Cable Distribution. Note that [SCTE DVS-031] Mode 1 for 64 QAM is essentially identical.

[3] ISO/IEC International Standard 13818-1, Generic Coding of Moving Pictures and Associated Audio: Systems, 13 November, 1994.

[4] Heinanen, J., Multiprotocol Encapsulation over ATM Adaptation Layer 5, IETF RFC 1483, July 1993.

HIGH-SPEED DATA SERVICES AND HFC NETWORK AVAILABILITY

Esteban Sandino, P. Eng.
Corinna Murphy, Engineering Technologist
Rogers Cablesystems Ltd., 853 York Mills Road, Don Mills, Ontario M3B 1Y2, CANADA

Abstract

The provisioning of high-speed data services over HFC cable networks demands a high degree of reliability from the physical transport infrastructure making maintenance of the network infrastructure essential. This maintenance requires defining and implementing procedures which ensure the tracking and subsequent prompt servicing of all physical network failures; and adherence to strict control procedures for scheduling these maintenance and repair activities during the times of lowest customer activity.

This paper describes the network management activities performed by Rogers™ Cablesystems to support the delivery of Internet access services through its Rogers™ WAVE™ product offering. More specifically, it provides a close examination of the service availability statistics collected during the first year of Rogers™ WAVE™ service delivery with special emphasis on the primary causes of HFC network failures and their effect on service availability targets. Also, a look at the changes in the Rogers'™ network management process now being implemented and their impact on service availability will be included.

This paper is an update to an article submitted for publication in the January, 1997 issue of Communications Technology magazine. The article presents service statistics from January to October, 1996. The updated paper will include January to December, 1996 Rogers™ WAVE™ service availability statistics.

INTRODUCTION

The provisioning of high-speed data communications services over HFC networks presents an unique set of challenges to today's cable operators. Physical network failures, including equipment and plant failures, which prevent the delivery of data services are detected by cable data customers immediately. This requires the definition and implementation of procedures to ensure the tracking and subsequent prompt servicing of all physical network failures; and adherence to strict control procedures for scheduling these maintenance and repair activities during the times of lowest customer activity to minimize service interruptions.

In December of 1995, Rogers Cablesystems in Canada launched an Internet-access-over-cable service in its 16,000-subscriber Newmarket, Ontario system under the commercial name of Rogers™ WAVE™. We will focus on the service availability statistics collected during the first year of delivery of this service with particular emphasis on the primary causes of HFC network failures and their effect on service availability targets. A brief overview is provided of the state of the current network management process at Rogers™, planned improvements, and their expected impact on current WAVE™ service availability.

ROGERS™ WAVE™ SERVICE AVAILABILITY TARGETS

The goal for service availability of Rogers™ WAVE™ is 99.90% after accounting for all service interruptions, or a maximum of 525.60 minutes of downtime per user per year as stated in the Canadian Cable Television Association's (CCTA) service quality guidelines. Eventually, the goal is to reach a service availability level of 99.99%. The key to achieving this target is having a short Mean-Time-To-Restore (MTTR) after a fault is detected.

Service availability is determined by:

1. Physical network availability. This is determined by the reliability of the cable plant components, their failure rates, and MTTR of the physical network; and

2. Service data network availability. This is determined by the reliability of non-transport related network components, such as Internet servers and data routers and their corresponding MTTR.

The relative contribution to the overall service availability is apportioned to each of the above components as follows:

- 99.94% for HFC physical network availability, or a maximum of 315.36 minutes of downtime per user per year; and

- 99.96% for data service network availability, or a maximum of 210.24 minutes of downtime per user per year.

THE ROGERS™ WAVE TECHNICAL ACTION CENTER (WTAC)

Since the inception of Rogers™ WAVE™, a WTAC was established with the mandate to monitor continuously the status of all physical and data network components. Such a mandate included the following tasks:

- Problem alerting and downtime notification;
- Problem isolation and escalation to the appropriate technical support group (e.g. having cable technicians address physical cable network problems and data analysts address data network and server problems);

- Tracking and scheduling of all network maintenance activities such as cable upgrades and system reconfigurations to minimize service disruptions (i.e. change control); and

- Tracking of all network failures and issuance of network problem reports and recommendations to help meet established service availability targets.

WAVE™ TECHNICAL ACTION CENTER'S MONITORING AND TROUBLESHOOTING TOOLS

The main tools which WTAC employs to detect, isolate, and track physical and data service network problems are:

1. Rogers™ Integrated Network Management System (INMS) terminals which perform two functions. They monitor the data generated by each of the status monitoring transponders (SMT) located in the coaxial trunk amplifiers; and they control the reverse bridger switches at the coaxial trunk level. Alarms are automatically generated whenever the operational levels of the trunk amplifiers deviate from pre-set thresholds. Figure 1 illustrates how trunk station status is displayed on an INMS terminal.

2. A set of independently controlled bridger switches in the 5-18 MHz and 21-42 MHz return bands allow for control over which trunk or feeder areas should feed reverse signals back to the headend. This

capability allows for quick isolation of problem feeder areas, and, in turn, expedites troubleshooting and service restoration activities. In addition, INMS terminals allow switching of in-line 6-dB attenuators in the reverse path at the trunk level to aid in the isolation of noise and other interference originating from the cable plant.

3. Reverse noise monitoring stations which allow for remote control of a spectrum analyzer directly connected to the reverse feed areas at the various WAVE™ service headends. The spectrum analyzer display at the headend is available on a video display at the WTAC for the continual monitoring of reverse noise levels as illustrated in Figure 2.

4. Simple Network Management Protocol (SNMP) stations which provide the monitoring and control of SNMP-devices, such as data routers, servers, and the current generation of Zenith's cable modems used to deliver Rogers™ WAVE™ services.

5. Zenith's cable modem management utility which provides remote control of functions such as modem frequency and power output level settings.

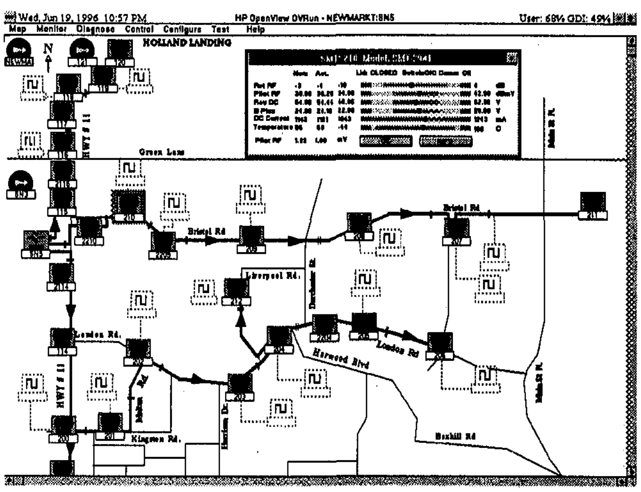


Figure 1. Rogers™ INMS Terminal C-Cor Trunk Stations

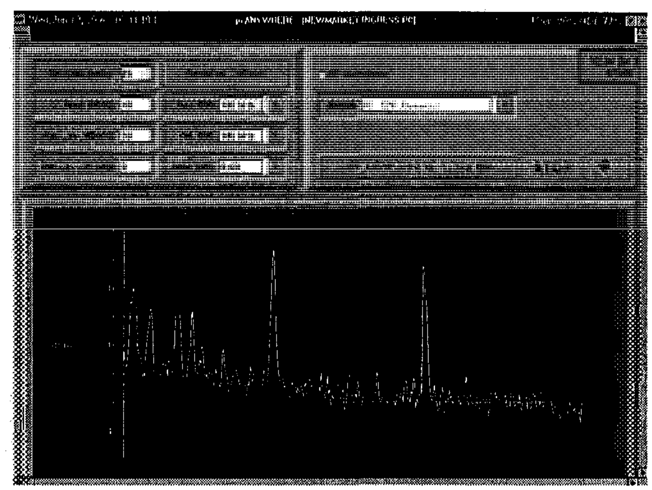


Figure 2. Rogers™ Spectrum Analyzer Utility

HFC NETWORK AVAILABILITY

The Rogers™ WAVE™ TAC has been collecting service availability statistics for Newmarket daily since January of 1996. Figure 3 illustrates the Rogers™ Newmarket HFC network availability for the period of January to December, 1996. These statistics include all instances of service interruptions arising from the following sources:

- Network maintenance activities both scheduled and unscheduled. For scheduled maintenance, the maintenance window for Rogers™ WAVE™ service is restricted to Sundays between 2:00 a.m. and 6:00 a.m. However, the related downtime is still included in the calculation of network availability;

- New plant construction activity;

• Headend and fiber related equipment failures;

• Trunk and distribution failures. No distinction is made at this time between trunk-specific problems and those arising from line extenders and other distribution equipment. However, drop-related problems have been excluded from this analysis;

• Power failures affecting both trunk and distribution which result in service downtime; and

• Reverse noise. This source of service interruption includes all impulse and ingress-related events that result in the degradation of the service.

For the purposes of performing network availability calculations, each episode of network downtime has been normalized to the total cable subscriber base in Newmarket of 16,000. Individual downtimes are multiplied by the number of customers affected and divided by the total cable subscriber base. Total downtime for any period is the sum of normalized downtimes for the measurement period. HFC network availability is then given by the total time in the measurement period less the sum of normalized downtimes for the same period. It is expressed as a percentage of the total time.

Figure 3 indicates HFC network availability ranging from a low of 99.13% in March (i.e. 390 minutes of downtime) to a high of 100% in May. The major contributor to the March figure is maintenance activities which resulted in a network downtime of 250 minutes. From Figure 3, the year overall HFC network availability for 1996 was 99.78%, or a total normalized downtime of 1 160 minutes. This exceeds the maximum allowable downtime of 315.36 minutes required to ensure 99.94% availability.

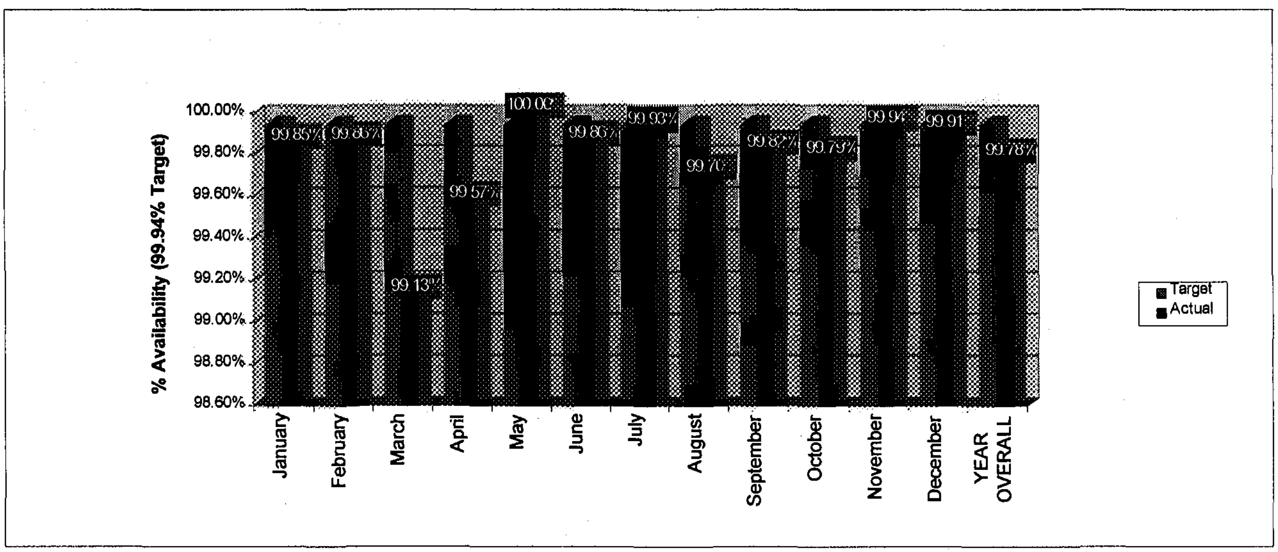


Figure 3. Rogers™Newmarket 1996 HFC Network Availability

SOURCES OF HFC NETWORK DOWNTIME AND STEPS TO PREVENTION

What are the sources of network downtime and how can they be minimized? Figure 4 illustrates the relative contributions to HFC network unavailability from all of the sources previously mentioned.

Overall, maintenance activities account for 27% of all incidents resulting in network downtime. Other major contributors are trunk and distribution problems (24%), power outages (20%), and excessive reverse noise levels caused by ingress and impulse noise events (14%).

Maintenance-related downtime can be minimized by ensuring that all activities affecting HFC networks are restricted to a single maintenance window. At Rogers™, proper change control procedures are being established to ensure that service back-up options are in place prior to performing any network maintenance. Furthermore, appropriate work releases are issued and scheduled work is completed on time.

Proper HFC network operation also involves the tracking of field equipment operation and failure trend analysis. This in turn enables forecasting of equipment failures and avoidance of costly network downtime which can be caused by power supply and trunk station failures. Lastly, directed physical plant maintenance is required to prevent excessive reverse noise levels on the HFC network which can result from defective passive components such as cable and connectors.

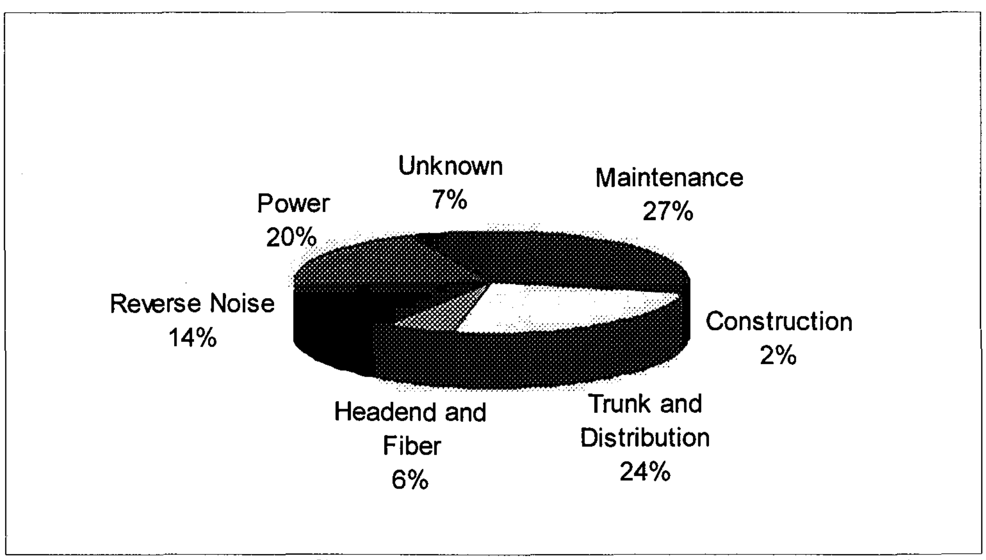


Figure 4. Relative Contributions to HFC Network Unavailability Rogers™Newmarket System - January to December 1996

HFC NETWORK AVAILABILITY AND MTTR

HFC network availability is also determined by the time it takes to restore network services after a fault has been detected and technicians have been dispatched to resolve the problem. Mean times to restore service from various failure types for the Rogers™ Newmarket system are illustrated in Figure 5.

A single construction-related incident resulted in overall slow downs in Internet access speeds for all Rogers™ WAVE™ customers. This particular fault took over three days to isolate and resolve, as the appropriate data monitoring tools were not in place to detect the resulting traffic overloads.

Therefore, even though Figure 4 shows that construction-related incidents of HFC network downtime account for only 2% of all incidents, the excessive mean-time-to-restore for a single incident contributes to an overall high MTTR of 1 473 minutes for construction-related downtime in Figure 5.

A much better measure of the network's capability to recover from a fault is given by the distribution of the actual time taken to repair that particular fault. An illustration of such a distribution is shown in Figure 6. This data reveals that in 122 out of a total of 131 recorded incidents affecting HFC network availability, it was possible to restore the network to full operation in less than four hours with 92 of these being resolved in less than one hour. The collected data indicates only one instance during which full network restoral required over three days to complete.

This high degree of success in meeting Rogers'™ initial average MTTR objective of less than four hours stems from the effective use of network monitoring tools that not only enable early fault detection, but also fast fault isolation resulting in the rapid dispatching of a technician.

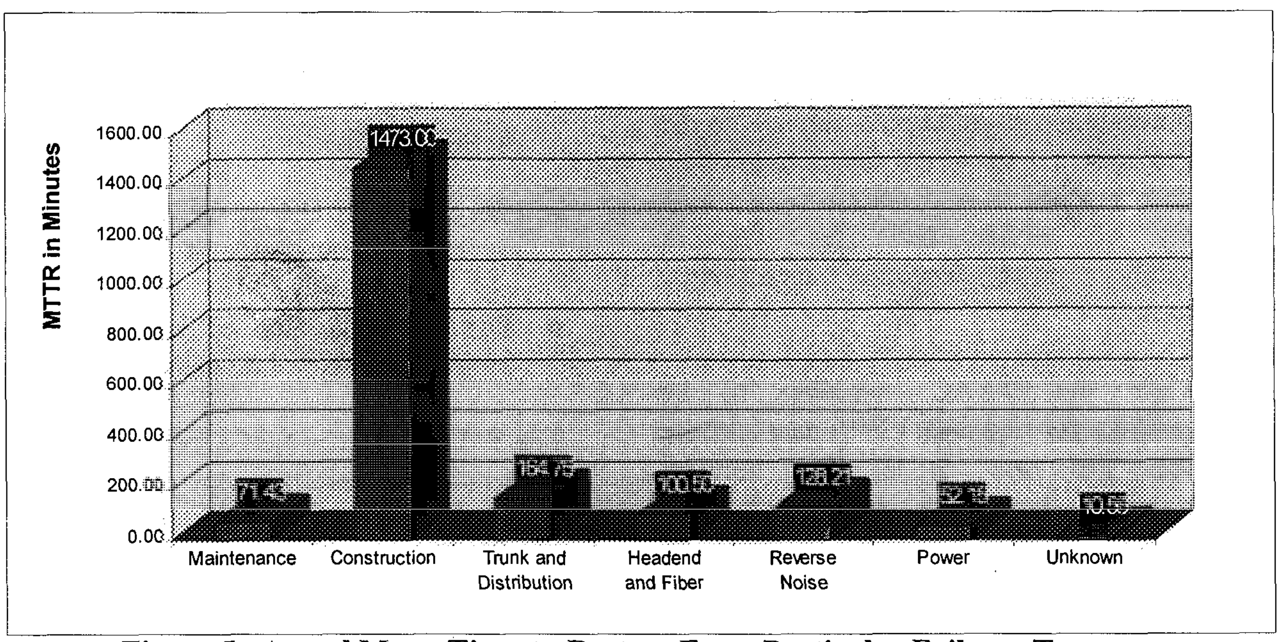


Figure 5. Actual Mean-Time-to-Restore From Particular Failures Types Rogers™Newmarket System - January to December 1996

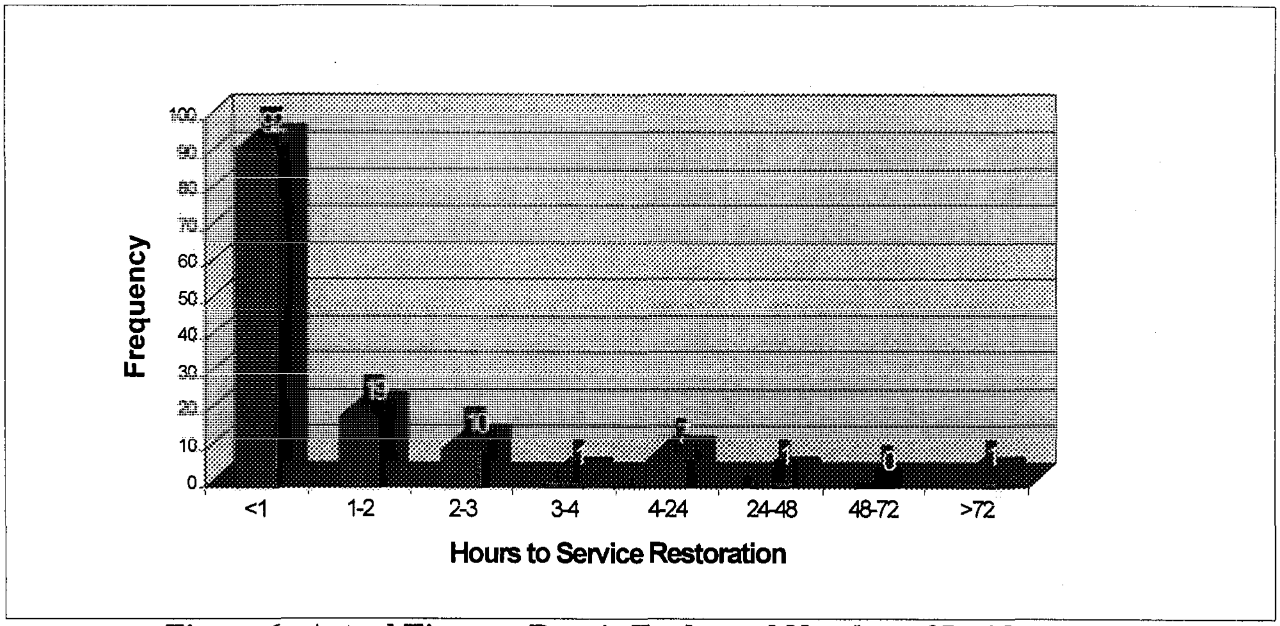


Figure 6. Actual Times to Repair Faults and Number of Incidents Rogers™Newmarket System - January to December 1996

IMPROVING ROGERS™ WAVE™ SERVICE AND HFC NETWORK AVAILABILITY

Until November 1996, the WTAC was responsible for both the HFC physical network and the WAVE™ data service network. Now, in anticipation of increases in service penetration and the preparation for the introduction of additional services to the Rogers' systems, the needs for more efficient problem resolution, and maintaining a high quality of service and HFC network availability have become increasingly apparent.

While the WTAC continues to be responsible for the WAVE™ data service network, a new model for network management has been implemented. It involves the establishment of regional Network Operations Centers (NOCs) to look after all aspects of the Rogers HFC physical network. Both the WTAC and the NOCs work together to achieve the same network management objectives:

- A minimum of 99.90% service availability which translates into a maximum of 8.75 hours of outage time per user per year;
- Resolving all outages within four hours of detection;
- Ensuring that 0% of reported problems develop into outages; and
- Ensuring that 0% of service quality problems develop into customer calls.

The WTAC and the NOCs have well-defined roles and responsibilities for achieving these objectives. The WTAC manages the WAVE™ data service network by:

- Continuous monitoring of the Rogers WAVE™ network components (i.e. routers and servers) using the reverse noise monitoring stations, SNMP stations, and Zenith's cable modem management utility. Each has been described in the section entitled WAVE™ Technical Action Center's Monitoring And Troubleshooting Tools;

- Problem alerting and troubleshooting activities;

- Notification of WAVE™ network and service changes;

- Problem escalation;

- Maintaining problem logs; and

- Enforcing change control procedures for WAVE™ services.

The NOCs manage the HFC physical network in their respective regions by:

- Continuous monitoring (i.e. tracking and auditing the alarms) of the HFC plant from the headend to the tap using the Rogers INMS terminal, the set of bridger switches in the 5 - 18 MHz and 21 - 42 MHz return bands, and the reverse noise monitoring stations. Each has been described in the section entitled WAVE™ Technical Action Center's Monitoring And Troubleshooting Tools;

- Problem alerting through the issuance of trouble tickets and the activation and coordination of service restoration activities during the failure;

- Initiating the notification/escalation procedures as required to clear the network trouble tickets within established timeframes; and

- Maintaing the problem logs through the generating of failure reviews and daily network reports.

As the WTAC continues to perform its monitoring and management functions of the Rogers™ WAVE™ data service network, the recent establishment of the NOCs ensures that availability targets for the HFC physical network are met. This means that HFC network maintenance activities become more focused on taking preventative action to avoid network failures.

ACKNOWLEDGEMENTS

The authors wish to thank everyone at Rogers who is part of the Rogers™ WAVE™ team, and whose efforts and tremendous enthusiasm have taken this new service to its current level of success.

REFERENCES

S. Rochette and T. Werner, "HFC Network Management", *1997 Conference On Emerging Technologies Proceedings Manual*, 301 - 312.

INTERNET ACCESS PROVIDER CABLE TELEVISION'S BUSINESS PLAN

Joseph P. Duggan
Senior Vice President
Waller Capital Corporation

ABSTRACT

The cable television industry has a history of conceptualizing, creating and offering innovative programming to entertain and inform its subscribers. From it's earliest days as a reception medium through satellite delivered programming, pay-per-view and interactive gaming, the industry has developed new technologies and risked significant capital. The latest and most innovative form of "programming" is available on the Internet. Cable television's challenge is to deliver this new "programming" through a competitively priced, superior Internet access product by leveraging off it's superior bandwidth.

The deployment of Internet access over cable television systems represents a substantial opportunity for creating a new revenue stream using existing infrastructure. This will also create a strategic advantage for cable companies as they are threatened by new competitors for the basic video product.

The Internet represents access to the largest library of information in existence. The merging of the cable television industry and the Internet (which heretofore had been associated with the computer industry), is based on the premise that the Internet is rapidly becoming a multimedia environment. Historically, information on the Internet consisted primarily of simple text. Today, the Internet is shifting to an increased dependence on very graphically rich content which takes an interminable amount of time to download using standard telephone dial-up modems. Internet users are demanding higher access speeds to reduce the amount of time that is consumed simply waiting for information. This shift in the composition of Internet related content has created an opportunity for broadband providers (cable) with superior speed over narrowband (dial-up) providers.

This paper offers a layman's introduction to the Internet, defines the business opportunity available to an Internet Service Provider ("ISP") and constructs a basic business plan for a cable television company in the Internet access business.

INTERNET SERVICE PROVIDERS

To those who are unfamiliar, the Internet is a connected set of computer networks running a standard protocol with a globally administered address space. Simply put, the computer networks can all talk to each other and information can be directed from one computer to another by defining its location using commonly accepted directions. The Internet has no beginning or end and as networks are added or deleted or if a failure occurs in a network it has no effect on the rest of the Internet. Most importantly, no one owns the Internet, it is a shared resource.

An Internet Service Provider ("ISP") provides residential and/or business access to the Internet through a transmission system. An ISP connects a computer through telephone or other data transmission lines (i.e., coaxial cable) to a central facility which then routes the connection to an Internet gateway. There are several well known, publicly traded ISPs including American On-Line ("AOL"), CompuServe and NetCom. These companies provide the same product; however, they seek differentiation through subscriber access availability, technical support and proprietary content that is not generally available on the Internet.

There are currently over 70 million personal computers ("PCs") and home PC penetration is estimated at 60% of households by the year 2000. With a computer plus peripherals costing more than $2,500, it is expected that on-line subscriptions (i.e., internet access) at a relatively cheap $20-$30 per month will reach 100% of PC equipped homes. In addition, the roll-out of cheap Internet "appliances" (Internet access only computers) under development by Oracle Corp. and Sun Microsystems, Inc. and Sony's WebTV will increase the demand for Internet access.

A cable television company in the internet service business has three distinct advantages:

- ☐ The company has a relationship with its subscriber base
- ☐ The company has an in-place infrastructure (offices, back-office staff, technical and marketing expertise)
- ☐ The cable operator can offer multiple Internet access options not available to non-wired competitors by using cable plant for either one-way or two-way transmission.

Access Options

An ISP provides a subscriber with access to the Internet through three primary methods; telephone dial-up modems ("Dial-up"), hybrid cable modems ("Hybrid") or high speed cable modems ("HSCM").

Telephone: Dial-up

The vast majority of ISPs are limited to using the local exchange carrier's ("LEC") network to offer access to the Internet. The transmission lines are either wholly-owned (as in the case of an ISP operated by a LEC) or lines (numbers) are leased by the independent ISP. Current transmission speeds are between 28.8kbps and 56.6kbps. LECs also offer ISDN lines which operate at up to 128kbps but at a significantly greater lease line expense. This expense will decrease over time as LECs upgrade their systems. LECs have also made significant progress on asymmetrical digital subscriber line ("ADSL") technology which promises even higher speeds. However, ADSL modems are still being field tested and require significant improvement in transmission facilities and expensive specialized modems.

Cable: Hybrid and High Speed Modems

As noted, in terms of access options, a cable television operator has a distinct advantage over Dial-up ISPs. The cable operator can offer multiple services by using the LECs lines for Dial-up service and cable plant for either one-way or two-way transmission.

Hybrid/Asymmetrical

In a Hybrid environment a modem is connected to both the cable television system and the telephone system. The cable television system provides the downlink path from the Internet and the telephone system connects the PC back to the Internet. This method is fairly efficient, resulting in a significant increase in access speed because most Internet surfers are downloading graphically rich content comprised of a tremendous amount of data while only returning relatively small amounts of information. This does, however, require a hybrid modem (which is not a retail product and is currently only available from a cable television company) and either a shared or second telephone line which increases the customer's access expense.

High Speed Cable

High speed cable access to the Internet requires HSCMs, cable television plant with a clear return channel and sufficient bandwidth. This offers the maximum speed to and from the Internet. Exact comparisons of speed are difficult because access to the bandwidth available for the Internet over cable is dynamically allocated based on available unused capacity. Therefore, system design and demand will determine access speed. The speed differential over Dial-up access is so great, however, that full cable connectivity could take up to 60% of the on-line market share.

Other: MMDS and Satellite

There are other technologies available that are struggling to create access to the Internet via satellite and MMDS delivered downlinks connected to telephone return paths. These are considered niche competitors due to satellite transponder shortages, terrestrial interference and propagation footprints.

The Market

Dial-up modems now come as standard equipment on most PCs or they can be purchased separately from any computer retailer. ISPs offering Dial-up service come in many flavors including local, regional and national providers. Local providers are typically small businesses that were started over the last few years. On a regional basis most LECs offer Dial-up Internet access as well as growing availability of speedier ISDN service.

There are also multiple national ISPs including American On-Line ("AOL"), CompuServe and NetCom. These companies have created nationwide backbones connected to local points of presence ("POPs"). POPs must be located near every market where the national provider intends to offer competitive service, as subscribers will not pay long distance toll charges to connect to the Internet. The tremendous investment required to create sufficient POPs has limited many national providers footprints to more densely populated communities.

Hybrid and HSCMs are currently available in limited quantities and are not subject to a common operating standard (i.e., Dial-up modems are interoperable on any telephone system. However, Hybrid and HSCMs only work on the cable system that supplied the modem). CableLabs, the cable television research and development group, is working on a set of standards that will be released in 1997. The interoperability standards will shift the Hybrid and HSCM modem to a retail environment and away from the ISP's balance sheet.

There are several large cable companies rolling out HSCM service using brand names like, Time Warner's, RoadRunner, US West Media's, Highway1 and TCI affiliate, @Home. All of these services have moved beyond trial stage and are marketing service in those cable systems where they have completed two-way rebuilds.

A CABLE TELEVISION ISP BUSINESS PLAN

For a cable television system operator there are numerous reasons to consider becoming an Internet Service Provider including: (i) if a cable company is not selling Internet access, someone else will be, and it could be a competitor for basic video service; (ii) it provides an additional revenue stream over which to spread the cost of rebuilds and upgrades; (iii) it provides a strategic product alternative to market against competitors for video service; (iv) it will increase cash flow and subsequent exit valuation in the event of a sale; (v) as the Internet changes to more closely resemble "programming", it will eventually steal some of cable's high-end video customers; and (vi) many cable television channels are offering interactive web sites which are becoming an important part of the programming mix.

Once the decision has been reached to consider the business, a cable company should consider the following three step plan:

- ☐ Market Evaluation,
- ☐ Financial Projection, and
- ☐ Service Deployment.

Market Evaluation

A market evaluation must be completed to ascertain if the environment is right for the service. A prospective ISP should focus on the following:

ISP Competition

Major urban markets have a plethora of small ISPs as well as nationally based providers struggling to attract subscribers. Smaller markets will necessarily have fewer competitors as well as few national providers which can provide a local phone number.

Market Demographics

The Wall Street Journal reported that the average household income of on-line users was over $48,000 or 9% over the national average and online users also tend to outpace the general education levels as compared with U.S. adults. However, as the cost of access declines and PC penetration climbs, education and income will fade in importance as the market bellwethers.

Local Telephone Service

The largest variable cost for the ISP is the local telephone line expense. A Dial-up ISP must maintain a sufficient number of local numbers to ensure a minimal number of busy signals when subscribers attempt to log on. Several ISPs reported an average of 8:1 to 10:1 (subscribers to lines) as the appropriate level. Monthly rates for telephone lines from LECs vary widely from market to market. Cable operators offering Hybrid service will have to maintain the same number of lines as Dial-up; however, HSCMs which bypass telephone lines, will not incur this expense.

Joint Venture Opportunities

The opportunity to joint venture with a local business such as a newspaper or radio station can be strategically important. This kind of partner can provide local content (school schedules, local sports, local issue bulletin boards) as well as auxiliary advertising. This is increasingly valuable as ISPs proliferate and attempt to differentiate themselves.

Educational and Business Institutional Presence

Educational institutions such a colleges and universities attract a great deal of Internet usage among students and researchers. In addition, many businesses are large users of the Internet for communications, E-mail and research. This built-in demand can jump start an ISP.

Financial Projections

A financial plan for the ISP business includes: type of service offering (Dial-up, Hybrid and/or HSCM), subscriber penetration projections, operating expenses including LEC line charges for local, long distance and high speed connectivity to the backbone, license fees, software expenses, subscriber billing, promotion, marketing, customer service, staffing and general office expense. A simplistic outline of a five year plan is attached as Exhibit 1.

Capital expenditures both at the Internet control center and for modems, both Hybrid and HSCM, are specifically not addressed herein due to the vast differences in existing cable system architectures and resulting rebuild or upgrade costs and in the differences in capital required by different vendors (some but not all of which would have a direct relationship to system performance).

Type of Service

As noted there are three types of Internet connectivity available to a cable based ISP, Dial-up, Hybrid or HSCM, however, while a cable television based ISP is limited by the condition of its existing plant, it should offer as many services as possible. Even the largest cable operators which are launching HSCM service (RoadRunner, @Home and Highway1) are also quietly offering Dial-up service for those subscribers who don't require the speed of cable modems or simply want the lowest cost option. These subscribers are also the best source of migratory growth as they become more adept on the Internet and can migrate easily onto either Hybrid or HSCM service.

Deployment

After assessing the market and the viability of offering Internet access, an operator must choose whether to purchase an existing ISP or build the business internally.

The location of the target acquisition or the site of a new build must be carefully considered to take into account availability of scaled-up telephone service, access to the Internet backbone provider, and location relevant to any other cable systems that may be locations to launch ISP service. For example, a cable operator that has clustered several small systems can tie back to the ISP control center via T-1 (high speed, high capacity telephone lines) lines and avoid duplication of Internet headend equipment while expanding the universe of possible customers.(See Figure 1)

Independent of which method or if all three methods of connectivity are offered, all of the services require a central processing point, the Internet cable headend equivalent. It is here that the connection from both the subscriber and the Internet are interfaced on the ISP's computers (referred to as servers), which provide storage for e-mail, news, and other content which is stored at the local system level. In addition, a server stores account information and access control (password) information.

The headend is connected to local data (telephone) circuits to subscribers in a hybrid and Dial-up environment and to the cable plant when using HSCM. The headend is also connected by high capacity links, typically T-1 lines to an Internet backbone provider. Internet backbone providers are analogous to wholesalers of Internet access while the local ISP is the retailer.

In terms of valuing potential acquisitions there are no solid "benchmarks of value" as are used in the cable television industry. The variables involved in the decision include: size of the ISP, financial strength, operating history, quality of existing equipment, employees and market niche. What little information available suggests that prices

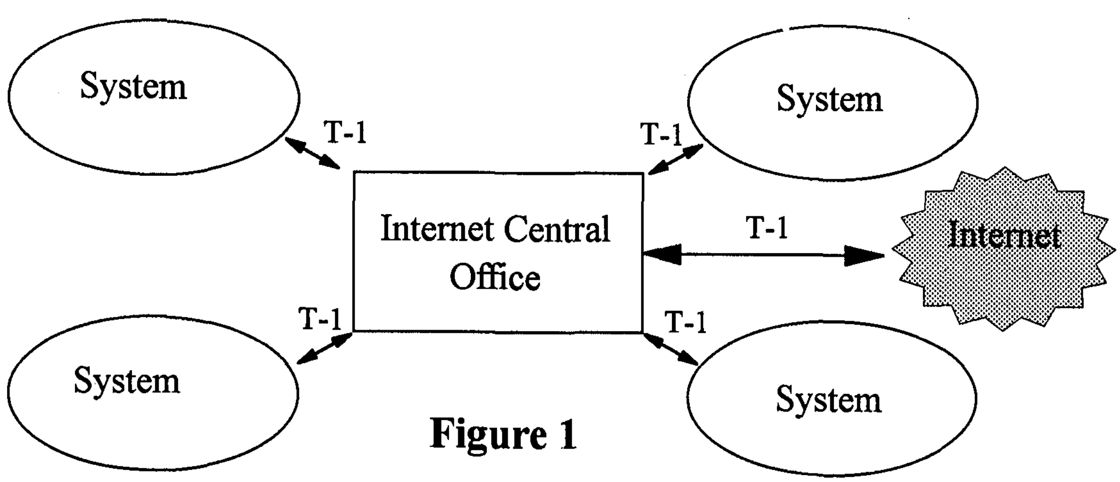


Figure 1

have ranged from one to one and one-half times annual revenue or in some cases just a small premium over the cost of the equipment.

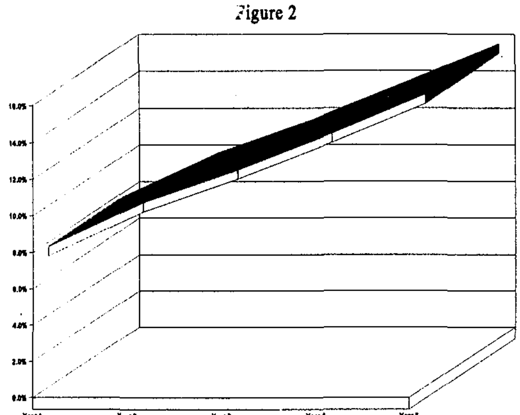
Subscriber Growth

On-line penetration is expected to reach 100% of PC equipped homes and, within the next three years, PC penetration is projected to rise to 60% of all households. For a typical Dial-up ISP, penetration is a function of the percent of PC equipped homes in its market times the percentage of modem equipped (Dial-up) homes times the percentage of expected market share.

Many cable operators start by calculating ISP penetration as a percent of cable subscribers (assuming a natural advantage with current customers). Figure 2 highlights projected penetration over 5 years:

Five Year Penetration Growth

Figure 2



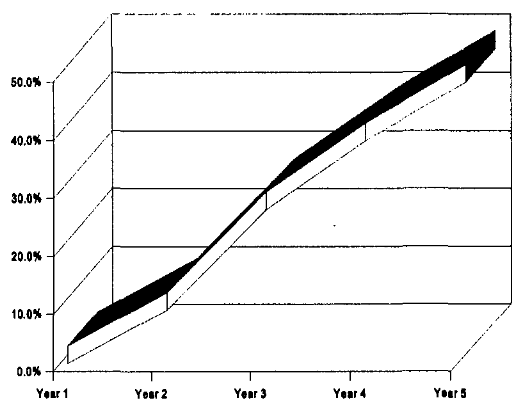
The projection above assumed 50% PC penetration, 50% modem penetration and 30% market share. PC penetration climbs to 70% by the fifth year and modem penetration climbs to 75% while market share remains static at 30% over the period. This is a conservative assumption. An operator offering Hybrid or HSCM service would not be limited by Dial-up modem penetration. In addition, market share, due to multiple products and cable's compelling product differentiation, would be considerably higher.

Operating Margin

The projection assumptions include, penetration as described above, market based pricing of $19.95 for Dial-up, $29.95 for Hybrid and $35.00 for HSCM (not launched until the fourth year of the projection to permit the operator to rebuild plant to two-way specifications). Expenses are composed of telephone charges which include LEC charges (assuming an 1:8 ratio for inbound lines to subscribers) as well as one T-1 expense. SG&A was calculated assuming fixed staffing and a per subscriber charge. This combination of subscriber growth, service mix and expense assumptions results in an operating margin that climbs from negative in the first year to 49% in the fifth year. See Figure 3 for the five year operating margin trend.

Operating Margin

Figure 3



Staffing

A key to success as an ISP is to find employees experienced with computers and the Internet. The most common complaints regarding ISPs relate either to busy signals when attempting to go on-line, or to poor service from the Internet provider's "help desk". A help desk requires a 24-hour a day staff available to advise subscribers. If the operator chooses to outsource this component there are several companies which can provide this service for a flat rate or on a per call charge.

Customer service costs are variable with the number of subscribers. As penetration increases, the ratio of customer service representatives needed will decline due the larger installed base. Several ISPs shared projections which scaled help desk staff from 500:1 (subscribers per staffer) to 1,500:1 by the fifth year of operation.

In order to improve customer service one cable ISP instituted a bulk PC purchasing program and subsidized employee purchases of PCs for home use. This resulted in a more computer/Internet friendly staff capable of providing better service and information on the Internet.

Outsourcing

There are "turnkey" companies that will design and install the components necessary to operate an ISP as well as provide management, help desks and maintenance.

Service Launch

Deployment of the service is the final step of the business plan and consists of three components:

- Marketing,
- Pricing, and,
- Installation.

Marketing

The most effective form of marketing for ISPs is word of mouth. The news of easy access, a customer-friendly help desk and popular local content will quickly spread through a community.

Pricing

Market knowledge gained in the initial feasibility analysis will guide pricing. However, when AT&T launched its $19.95 per month flat charge for unlimited access it set the standard for Dial-up pricing.

Hybrid and HSCM service is less dependent on price as it is currently the exclusive provenance of the cable operator. Most cable operators are adding discounts tied to cable service in order to induce migration and for use as a marketing tool for both services. In addition, ISPs have traditionally accepted credit cards subject to pre-approval.

Installation

There is usually no physical installation expense tied to Dial-up services as it is a simple software adaptation for a modem equipped PC. The subscriber needs to download the software from a diskette provided by the ISP to begin service.

Hybrid and HSCM service will require physical installation to configure the computer's hardware for the modem. This will be a growing issue and liability for operators.

Valuation

The value enhancement to a cable television operator for owning and operating an ISP is not readily calculable. Wall Street analysts are adding 20% to MSOs offering Internet access service with additional value for those MSOs interfacing with a national backbone provider, particularly @Home.

In the attached model, at the end of the fifth year the ISP is generating $146,410 per year in cash flow. The cable operator at the end of the fifth year is projected to have 6,495 cable subscribers generating approximately $38.25 in monthly revenue or $2,984,201 in annual revenue. Assuming a 45% cable operating margin the operator has $1,342,890 in cable cash flow. Cable/ISP cash flow totals $1,489,300, of which 10% is derived from the ISP.

Over time the assumption is that ISP service will be perceived as another form of programming similar to HBO or pay-per-view, and operators will receive an exit valuation at the same multiple as is applied to any other cash flow. A cable operator therefore can project a 10% increase in exit valuation on the combined entity a the time of sale even with these very conservative projections. The service will also protect value over time from subscriber loss to competitive providers

CONCLUSION

The Internet is a vast tool which has become the standard for information and entertainment access. Along with the radio, television and telephone, the Internet has made significant changes in how we communicate, entertain and learn.

ISP service is not a silver bullet for a cable operator. However, it is a way to create a new revenue source. Cable television operators would be remiss not to consider the ISP business and take advantage of their competitive superiority in terms of customer contact, infrastructure and superior bandwidth.

ACKNOWLEDGEMENTS

The author wishes to thank Greg Kriser of Helicon Corp., Mark Stephan at Mediacom L.L.C and Jack Lawrence of Chambers Communications Corp. for their help in preparing this paper.

References

The Wall Street Journal, December 9, 1996

Morgan Stanley, U.S. Investment Research: Deploying High Speed Cable Modems, June 21, 1996.

A CABLE TELEVISION ISP SIMPLIFIED STATEMENT OF OPERATING CASH FLOW	Year 1	Year 2	Year 3	Year 4	Year 5
Cable:					
Homes Passed	10,000	10,200	10,404	10,612	10,824
Basic Subscribers	6,000	6,120	6,242	6,367	6,495
Penetration	*60.0%*	*60.0%*	*60.0%*	*60.0%*	*60.0%*
Internet Service:					
Dial-up	315	424	475	480	410
Hybrid	135	182	256	349	410
HSCM	0	0	0	44	205
Total Subscribers	450	606	730	872	1,026
Penetration	*7.5%*	*9.9%*	*11.7%*	*13.7%*	*15.8%*
Monthly Rates:					
Dial-Up	$19.95	$19.95	$19.95	$19.95	$19.95
Hybrid	$29.95	$29.95	$29.95	$29.95	$29.95
HSCM	$35.00	$35.00	$35.00	$35.00	$35.00
Revenue:					
Dial-Up	$37,706	$88,472	$107,592	$114,254	$106,561
Hybrid	$24,260	$56,922	$78,599	$108,638	$136,461
HSCM	$0	$0	$0	$9,159	$52,257
Installation	$13,500	$4,676	$7,386	$9,330	$6,153
Total Revenue	$75,465	$150,071	$193,578	$241,382	$301,433
Expenses:					
Telephoney	$43,500	$48,176	$51,911	$56,169	$60,784
S,G & A*	$86,750	$87,919	$90,022	$92,020	$94,238
Total Expenses	$130,250	$136,096	$141,933	$148,189	$155,023
Operating Cash Flow	($54,785)	$13,976	$51,645	$93,192	$146,410
Operating Margin	-72.6%	9.3%	26.7%	38.6%	48.6%

* A portion of SG&A that could be capitalized has been treated as an expense for cash flow purposes.

EXHIBIT 1

IP ADDRESS PROVISIONING IN A CABLE TV DATA NETWORK

Christopher Grobicki
Director of Product Management
LANcity Cable Modem Division.
Bay Networks Inc.
200 Bulfinch Drive
Andover, MA, 01810 USA
Chris@lancity.com

Abstract. As subscriber penetrations grow in cable TV based networks, the availability of IP addresses for the connected workstations becomes an issue. This paper discusses different methods for assigning IP addresses in this environment as well as upcoming technologies in the TCP/IP networking community that address this issue.

Data communication and information access using the cable TV infrastructure provide cable operators with a new source of revenue at a time when their subscriber base has leveled or started eroding. This new service offering is not available with telephone modems or satellite broadcast services. It features high data rates and application and topology independence. Not only can it prevent existing subscribers from going elsewhere for service, but it can also attract new subscribers. The operator is in a unique position to provide data applications just as varied as the homes and businesses through which the cable passes. These applications can be used to help a community run its businesses, increase productivity, teach its students, help its citizens in occupying their leisure time, and generally, improve the quality of life.

Internet access, work at home, web page cruising and provisioning, and local area network connectivity provide the wide range of applications necessary to interconnect a community to itself and the rest of the world (see Figure 1.)

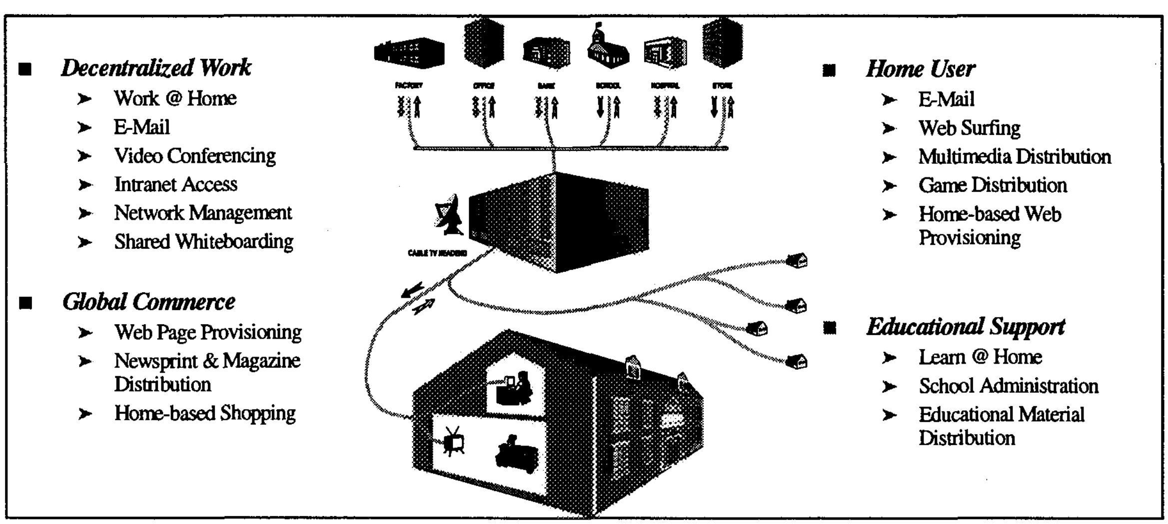


Figure 1. What Data Over Cable TV Enables

TCP/IP has become the standard communication protocol of the Internet, and for seamless connectivity, the cable TV based data network must follow suit. TCP/IP enables worldwide connectivity across the many different physical networks that comprise the Internet. Cable TV is a new addition.

Netscape Navigator and Windows 95 have TCP/IP stacks built into them. Users are asking for services like Web cruising, email, and multimedia distribution. TCP/IP enables these applications.

TCP/IP is more than simply a communication protocol. It provides functionality that allows services to be provided across a neighborhood, town, state, country, and even the world (see Figure 2). A person can communicate with someone in his neighborhood or access a Web site in China, all because of TCP/IP.

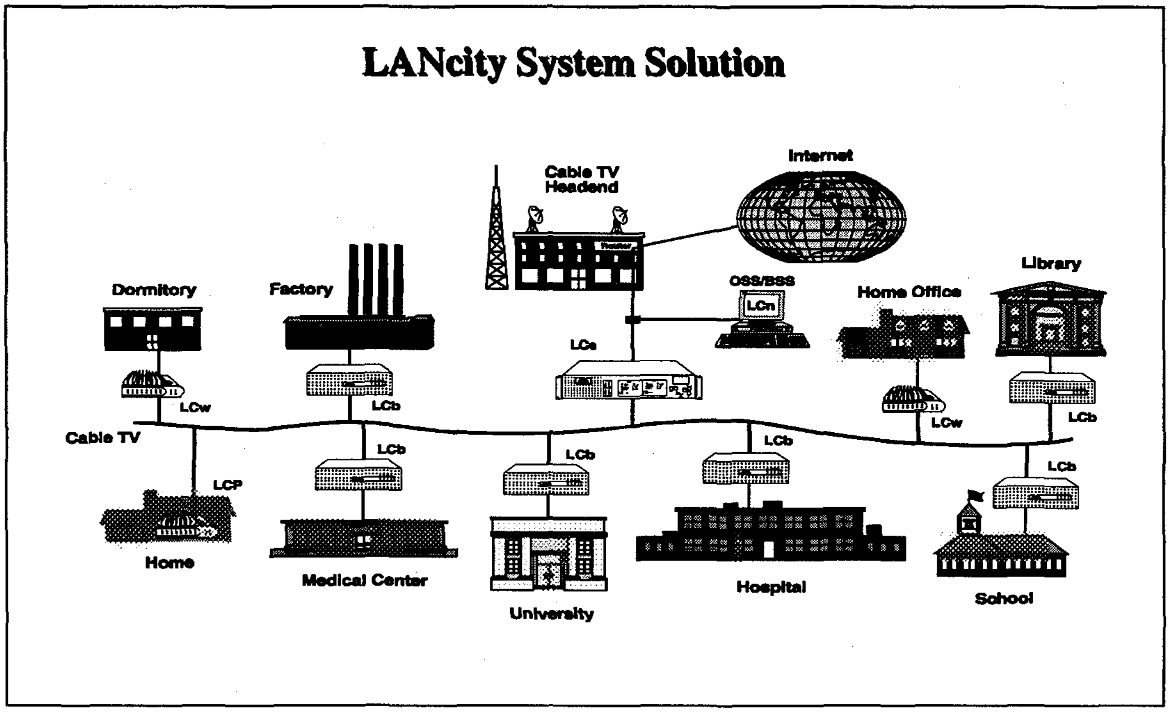


Figure 2. LANcity Solving the Challenges of Data Connectivity

Other communication protocols can provide connectivity for the Internet, but only TCP/IP scales from small networks, isolated to a single building, to very large networks of millions of users. One key reason for the scalability is the elegant addressing structure used in IP. Communication protocols like IPX and Appletalk rely heavily on broadcast addresses to form the basis for intercommunication. While the broadcast function is adequate for Local Area Networks with a small number of PCs, it is much less desirable for metropolitan and wide area networks with potentially millions of users.

While TCP/IP provides an efficient and scaleable choice for large internetworks like the Internet, it also requires more configurations than are required for small LAN based networks. The network address, host address, the domain name, the subnet mask, the IP address of the routers to get on and off the network, the address of the Domain Name Servers, and additional parameters are required for efficient operation.

These parameters must be programmed into the subscribers' personal computers and workstations before they can access the Internet.

Similar to a person's phone number, which identifies that person in the telephone system, the IP address uniquely identifies a host computing device on the Internet. This address enables information to flow robustly and efficiently from the source to the destination, regardless of whether the path is across town or across the world.

NETWORK ADDRESS HIERARCHY

Data addressing in TCP/IP networks has a hierarchy which provides unambiguous, yet efficient transport of information from one point to another, while isolating the complex layers from the subscriber:

Data Link address: The lowest level address used by a communication protocol to gain access to the local network. Since the Internet is made up of many different types of networks, such as Ethernet at the subscribers home, UniLINK over the cable TV plant, FDDI or ATM for headend interconnectivity, and T1 or T3 for connectivity to the Internet, TCP/IP traffic can traverse many different types of Data Link protocol types, each with their unique addressing scheme. These variables render the Data Link address unsuitable for end to end global communication.

A subscriber should not have to know the Data Link type being used by his communication partner, but because of the different requirements in providing data connectivity in the local, metropolitan, and wide area networks it is impossible to build a universal network from a single hardware technology. Data Link addresses are generally permanently assigned to a communication device and will stay with that device for life. Standardized conventions are used to assure that each communicating device has a unique Data Link address, regardless of where it was manufactured.

IP Address: Internetworking connects many different types of physical networks into a global Internet, providing end to end connectivity without requiring a detailed knowledge of the path traversed. TCP/IP uses the IP address convention to uniquely identify the participants in the global Internet regardless of whether they are connected to an Ethernet, Token Ring, Frame Relay, FDDI or whatever type of network with a global addressing scheme. IP addresses are generally assigned by manual entry or by using an automated and central distribution mechanism which will be described later. IP addresses are not bound to the device nor the user.

Domain Name: Although IP addressing provides unambiguous and compact representations for the source and destination of information flow across the Internet, for a human being used to dealing with names as opposed to numbers, it can be daunting. In addition IP addresses are not fixed to a subscriber; they can change. Therefore, it is important to incorporate an addressing convention that is easy to administer and understand and yet can remain with the subscriber for life.

The addressing service that satisfies this need is called the Domain Name System (DNS). For example, john.doe@company.com is a syntax that most people Internet users are familiar with. A domain name server administered by the Internet service provider is an efficient, general purpose mechanism for mapping these easy to understand names to the hosts-assigned IP address.

NEED FOR IP ADDRESSES

Unlike AOL and Compuserve, which provide indirect Telco access to the Internet through their server banks, a subscriber to cable TV based data services has direct access to the Internet. This provides the benefit of faster and unlimited access to information, including highly graphical web pages and multimedia applications. Accessing the Internet directly requires the subscriber's PC to become a TCP/IP host on the Internet, thereby requiring an IP address.

The proper administration of IP addresses to the subscriber base is one of the key criteria in providing a reliable and revenue generating service. IP addresses are used not only by applications like Netscape to access the Internet but also in the management of cable modems that use SNMP. This administration can be relatively painless or full of frustration and headaches.

Cable modems that provide a high degree of manageability incorporate an imbedded SNMP agent. This agent standardizes the manner of accessing operational characteristics and provisioning configuration parameters using industry standard management platforms such as HP Openview or Bay Networks Optivity. The SNMP agent requires an IP address to uniquely identify it to the SNMP manager.

CENTRAL ADMINISTRATION OF IP ADDRESSES

There are multiple ways to provision IP address to the subscriber's PC and cable modems. The LANcity Cable Modem Division of Bay Networks encourages the central administration of IP addresses from a server that is under the control of the cable operator or system administrator. This enables better control of the assignments, improves security, and reduces operational maintenance. It is not desirable to "roll a truck" every time a subscriber changes the PC's configuration. Remote reconfiguration is less labor intensive and, therefore, lower cost mechanism.

Three industry accepted mechanisms for the assignment of IP addresses to communication devices are described. The exact method used is determined by the equipment manufacturer's implementation of their products and the cable operator's operational requirements.

Local Assignment: This is the least desirable methodology unless the number of hosts is small because it requires going to the location the host resides in and manually entering the IP address. If the address changes then another truck roll is required.

BOOTP: A TCP/IP based protocol which provides a static mapping of IP address to the Media Access Control address. It provides a remote mechanism for entering the IP address into a host. This method is simple and robust, ideal for assigning a host with a private IP address that will not be used for accessing the Internet. It is less desirable, however, for public IP addresses which are used to access the Internet because it statically binds them to the Data Link address whether it is actually being used or not. BOOTP is a good allocation protocol for assigning IP addresses to communication devices with SNMP agents because static binding might be desirable to display network configuration maps.

DHCP: BOOTP is slowly being replaced by DHCP which provides the best of all possible worlds. It allows a static mapping of Data Link-to-IP address for management purposes, or a dynamic mapping where the assignment is allocated only during operation or for a specified lease period. This prevents allocating precious IP addresses to workstations not being used and, therefore, allows the address to be reused.

With DHCP a range of IP addresses must be reserved specifically to the DHCP server. If more than one DHCP server exists, each one must be assigned a unique address space.

A knowledgeable Internet Service Provider (ISP) can ease the headaches of IP address and domain name maintenance. Normally ISPs can handle IP address and domain name registration services by working through the proper authorities and agencies, taking over this responsibility from the cable operator.

OVERVIEW OF IP ADDRESSES

The TCP/IP based network, or Internet, is a virtual network built by interconnecting multiple physical networks of different types with routers. Each device on this network must be identifiable with a unique address so that information can traverse from the source to the destination efficiently and robustly. It would be impossible for every router that interconnects the Internet to know where a specific destination address is and know when to forward and when to filter the information. Therefore, something more efficient than a flat address space must be used.

The TCP/IP folks were clever in designing an address that consists of a network address and a host address (see Figure 3).

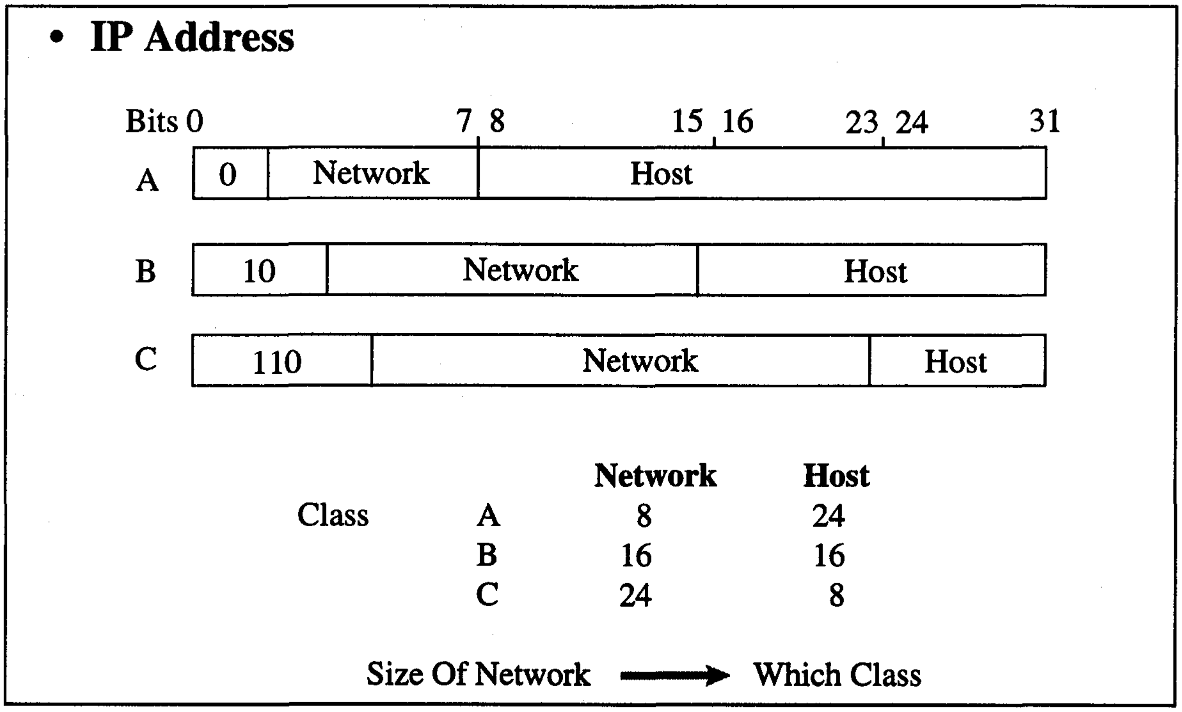


Figure 3. TC/IP - IP Addressing Structure

With this architecture, routers only need to know that they are part of a path to get data they receive to a specific network address and then forward it. Once the packet gets onto the local network, each host looks at the destination address to decide if the packet is destined for it, with all hosts on the network using the same network number.

There are three primary classes of IP network addresses: class A, class B, and class C. Class A addresses allow more than 65,536 hosts, a very large network indeed. These network addresses are assigned to providers like ATT and MCI. Class B addresses are for networks up to 65,000 hosts and are also subscribed to very large users. Class C addresses allow up to 254 hosts and are normally allocated to operators who want to provide Internet services to their subscriber base. Multiple class C addresses can be allocated to a single operator with large numbers of subscribers.

Because the IP address is 32 bits long, it is easy to see that with the explosion of the Internet, network addresses, though still plentiful, must be allocated carefully. Two mechanisms are used to help manage the allocation of class C address space. (Remember, class A , class B, or hard to obtain new service providers are not discussed.) One mechanism is called subnet masking and the other is called supernetting or supernet addressing.

A class C address allows up to 255 hosts on a physical network. What happens when there are fewer than 255 hosts and yet we want to conserve those precious IP addresses? Subnet masking is a standardized, and now required, mechanism that uses the first few bits of the host address field to specify a local physical network. Because the subnet addresses are derived from the main networks address, subnet masking does not require official assignment since the host addresses belong to the organization with that assigned class C network address.

The administrator, as part of the BOOTP or DHCP address assignment, specifies the subnet mask that indicates what bits are to be used for local network addressing. Therefore, as an example, a single class C address can be divided into two local physical networks that contain up to 126 hosts each.

Supernetting is the opposite of subnet masking but equally important in providing address conservation in a data over cable environment. For example, if fiber legs that are part of an HFC plant extend into multiple neighborhoods, the number of subscribers might be greater than 255, beyond the range of a single class C address. In this case, multiple contiguous class C networks can be joined together to form a single address space, thereby, the allocation of the very precious class B addresses is not required. Classless Interdomain Routing solves the problem of routers needing to understand this concept and collapses a block of contiguous class C addresses into a single router entry specified by the lowest address and a 32 bit mask.

WHERE DO IP ADDRESSES COME FROM?

The network portion of a public IP address is assigned by a central authority to assure its uniqueness. Duplicate IP addresses on the Internet can cause addressing havoc. Therefore, those who access the Internet should follow good procedure to ensure this does not happen.

To get a block or blocks of class C addresses, a network administrator must apply to the Internet Network Information Center in Herndon, Virginia. Class A and B addresses are very hard to obtain and are reserved only for the largest of networks. Class C addresses are assigned in normal situations with contiguous class C addresses assigned for larger networks.

Private IP addresses are used to address IP hosts that do not communicate directly with the Internet. They conserve public IP addresses and yet provide full network connectivity between hosts in an enterprise. In a data over cable architecture, private addresses are used for SNMP management of the cable modems. The address allocation for private internets is specified by Request for Comments (RFC) 1597, issued by the IETF. The addresses fall into the following ranges:

10.0.0.0	to	10.255.255.255
172.16.0.0	to	172.31.255.255
192.168.0.0	to	192.168.255.255

These addresses will never be assigned as public network addresses. Therefore, the use of these addresses does not have to be authorized. The hosts that have these addresses assigned can connect to other hosts inside of the domain but are expected to be filtered out by the routers connecting to the Internet.

The mapping of domain names to IP addresses is handled by the Domain

Name System. Network names at the root level (e.g., lancity.com) must be registered with the NIC to assure uniqueness in naming. The name to the left of lancity.com is administered locally from a DNS server that maps the name to the IP address.

ISSUES WITH IP ADDRESS ASSIGNMENT

Address allocation has come a long way in terms of ease of use. However, the operator must be wary of several pitfalls. Some of these issues exist because of holes in the mechanisms, others because of the distributed, yet public nature of the protocols. The IETF as well as the MCNS organizations are resolving these issues to improve and refine the methodology.

DHCP does not currently support a backup server architecture unless each one has its own IP address space. A backup server is valuable in situations when the primary server crashes and the connected hosts are unable to get their IP address and other parameters for coming online.

Having a backup server prevents this problem but also wastes precious address space. A backup server protocol is being developed to allow backup servers with the same IP address space as the primary server.

Currently, there is no mechanism available that updates IP address changes in a domain name server, thereby requiring a cumbersome manual update to the database. This problem is especially important if one is using the dynamic address provisioning feature of DHCP in a very large network.

Because of the large area of coverage and hard to control subscriber access, security in data over cable networks is a concern and must be addressed.

IP address server spoofing, where an illegal server provisions parameters to the data over cable TV network is being addressed by the cable industry MCNS group.

CONCLUSION

IP address allocation in a data over cable infrastructure is not extremely difficult to understand. However, care must be taken to do it correctly and follow procedures. Otherwise, the service will be hard to maintain and unreliable.

This paper has given a brief overview on the subject. Many excellent TCP/IP texts are available with more details on the subject. Though issues still exist, they are being resolved by multiple organizations that understand the value of robustness and security.

Lower Cost Alternatives to On-Demand Network Architecture

Louis D. Williamson
Senior Project Engineer
Time Warner Cable

ABSTRACT

1997 will be the year that broadcast digital services finally becomes affordable to consumers via cable. But to give the cable industry an edge over it competitors, the industry needs to be able to offer on-demand services, such as Video On Demand, to its customers. Unfortunately, the cost of the elements of an on-demand network are still too expensive to offer on-demand services. To make on-demand a reality different network designs and cost reduction efforts are still necessary. This paper will describe how the Asynchronous Serial Interface (ASI) standard can help to reduce on-demand network costs.

OVERVIEW OF ASYNCHRONOUS SERIAL INTERFACE

Asynchronous Serial Interface (ASI) is one of the physical layer specifications developed by the Digital Video Broadcasting (DVB) committee for the interconnection of headend equipment. The interface was developed to carry MPEG-2 data signals, although it can carry any type of data signal. An ASI port is specified to be an unidirectional link which operates at a data rate of 270 Mbits/second. There is approximately 216 Mbits/second of payload available on an ASI link after error detection and framing is removed.

An example of a Quadrature Amplitude Modulation (QAM) modulator implemented using ASI ports is shown in Figure 1. Each QAM modulator that is connected to an ASI link receives the entire 270 Mbits/second data stream. The QAM modulator processes the data stream to filter out the MPEG-2 programs that are addressed to it. The QAM modulator must also repeat or regenerate the data stream and pass in on to the next device in the chain.

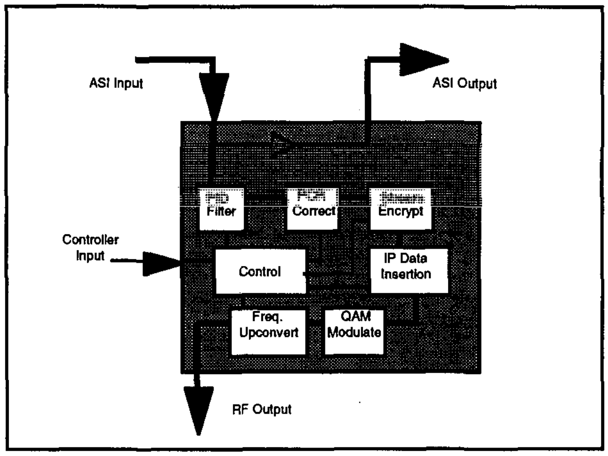


Figure 1 QAM Modulator with an ASI port.

ASI ports can be implemented with either a coaxial cable or a multi-mode fiber optic cable. Since the standard was developed for interconnection of headend equipment, these interfaces should operate over sufficient distances. If longer distances are required, for example to interconnect remote hubs, ASI ports could be implemented using single mode electronics since the specification is based on fiber channel electronics which also supports single mode electronics.

ON-DEMAND NETWORK ARCHITECTURE

The use of elements with ASI ports can help reduce on-demand network costs. To illustrate this, examine the typical on-demand network architecture which is shown in Figure 2. Two components of this network that can be replaced without loss of functionality are the ATM switch and the interactive cable gateway (ICG). The purpose of the ATM switch is to route signals from the servers making them available to any customer in any node. The main purpose of the ICG is to perform rate conversion. The ICG converts from the ATM switch's data rate of 155 Mbits/second into 38.8 Mbits/second which is the rate required by a 256 QAM modulator. The ICG also performs routing, stream encryption, security key insertion, and IP data insertion.

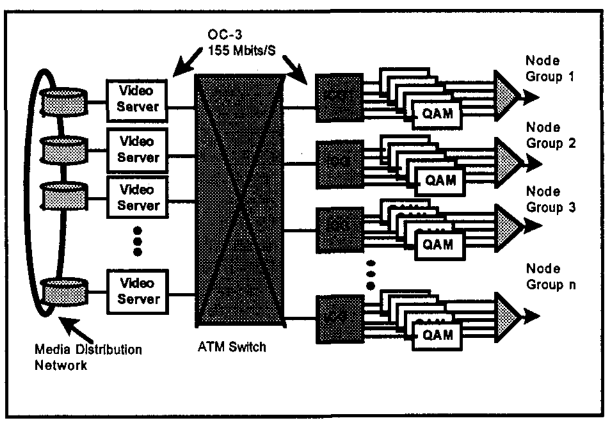


Figure 2 Typical on-demand network architecture.

By using ASI ports on the servers and the QAM modulators, both the ICG and the ATM switch can be eliminated from the on-demand network. Let's examine how a device with ASI ports can maintain the functionality of the network and help to reduce the overall cost.

ATM Switch Elimination

Media needs to be routed from any server to any customer. An ATM switch appears to be the ideal element to handle this task. ATM switches were designed to handle the task of routing many different types of data, such as video and IP data, from any input to any output. But the ATM switch may not be the ideal solution for routing information to customers in an on-demand network. The ATM switch has one big disadvantage, it was designed to be bi-directional. Most the information in an on-demand network flows from servers to customers, only a small amount of information flows in the reverse direction. Therefore, the ATM switch's reverse direction is used very little, which wastes money. The small amount of traffic that flows in the reverse direction is best handled by IP routers, which is usually the case even in networks that have ATM switches.

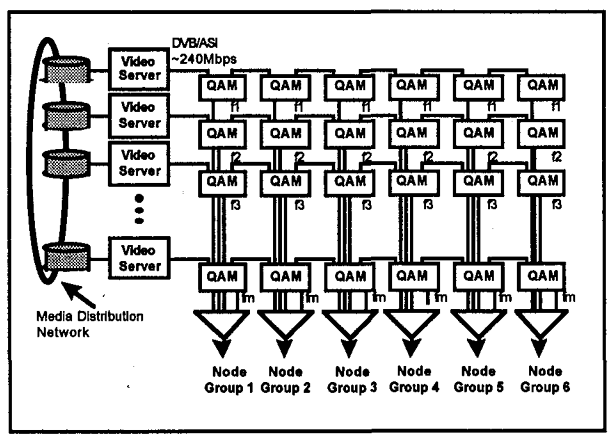


Figure 3 On-demand network architecture with ASI devices.

Clever QAM modulator output combining and the use of ASI ports on the servers and QAM modulators will eliminate the cost of the ATM switch. Figure 3 is a network based upon servers and QAM modulators implemented using ASI ports. In this network, media from any server can get to any customer in the same way as in the network with the ATM switch, but this network was designed without an ATM switch. It may first appear that the ATM switch network has advantages over the network with device using ASI ports, but it does not. With a proper media distribution, which is required in both networks, the network featuring ASI devices can perform any routing combination that can be performed in the network with an ATM switch.

In addition to eliminating the cost of the ATM switch, using ASI ports on the servers should also be cheaper to implement than OC-3 ports on the servers. OC-3 ports are bi-directional. Since most of the information is flowing out from the servers, the return side of the OC-3 port is wasted. Elimination of this portion of the interface port will reduce the cost of implementation.

Interactive Cable Gateway Elimination

As stated earlier, the main purpose of the ICG is for rate conversion. The ICG converts from the ATM switch's rates into rates that are required by the QAM modulators. But this is a very expensive way to perform rate conversion. The input to an ICG is typically OC-3 or 155 Mbits/second. The output of the ICG is usually 26.97 or 38.8 Mbits/second, which are the rates required by 64 and 256 QAM modulators operating in 6 MHz wide channels. Given OC-3 inputs, an ICG can supply data to five 64 QAM modulator or four 256 QAM modulators. Assuming 3.5 Mbits/second for a MPEG-2 video and audio stream, an ICG is required for every 40 MPEG-2 video and audio streams. The average cost of an ICG is approximately $25,000. This adds $625 to the cost of every MPEG-2 video stream that is needed in an on-demand network.

Rate conversion is easily implemented in the network built using element with ASI ports. Each QAM modulator on an ASI link is commanded to accept only the MPEG-2 programs that are addressed to it. It doesn't matter whether it is a 64 or 256 QAM modulator, the modulator will accept only the proper amount of information to modulate. The network controller will insure that a QAM modulator never receives more data than it can modulate, which is the case in both networks.

The other functions that the ICG performed; encryption, security key insertion, and IP data insertion, now need to be performed by the QAM modulator. Adding these functions to the QAM will likely increase the cost of the QAM modulator. But it is unlikely that the addition of these functions to the QAM modulator will increase the cost per stream more than the $625 per stream of the ICG.

In addition to the cost savings, elimination of the ICG has other benefits. In a network built using ASI port devices, the data signals in the headend are in the clear until they are modulated. This aides in troubleshooting the network. QAM modulator redundancy can be implemented by commanding spare QAM modulators to process the MPEG-2 stream of a failed QAM modulator. In addition, ASI is a standard which will make it easier for many different manufacturers to build accessory devices such as; Ad insertion equipment, test equipment, and emergency alert equipment, which are required for successful implementation of on-demand networks.

SUMMARY

Asynchronous serial interfaces on headend equipment will help reduce the cost of on-demand networks. ASI also has several advantages over existing headend network designs. The use of ASI on the servers and QAM modulators will eliminate the need for 2 expensive components, the ATM switch and the interactive cable gateway, that were being considered for use in the headend. ASI is a standards based. This will make it easier for many manufacturers to develop devices such as, ad insertion equipment, test equipment, and add/drop multiplexers. The unidirectional nature of ASI also reduces cost and is better optimized for the data traffic carried on an on-demand networks. This makes using ASI a better choice for headend interconnections. The cable industry must continue to rethink and optimize the way it will support on-demand services. On-demand network architectures still need further cost reductions to make them affordable for deployment, but using new approaches to signal distribution and continued cost reduction of elements will eventually make these networks a reality.

REFERENCES

Interfaces for CATV/SMATV Headends and Similar Professional Equipment, Digital Video Broadcasting Document A010, October 1995.

MPEG-2 Digital Program Stream Compatibility: Programming to DCT/Pegasus/Telcos

John Beyler
Home Box Office

Abstract

In the fast-paced world of digital television today, the "universality" of MPEG is not a slam-dunk to get programming from "point A" (the uplink) to "Point-B Telco"; another Point-B (Pegasus headend); another Point-B (DBS redistributor); and , finally, a Point-B "classic cable operator".

This paper describes the issues which must be considered by a programmer desiring to provide digital programming to all users. Issues such as conditional access, statistical multiplexing, system information tables, and video profiles will be discussed.

INTRODUCTION

Until now, universal interoperability has been very easy: NTSC is universally understood and may be scrambled, transmitted, and displayed on everything from Grandma's tube-type TV to hand-held LCD sets. NTSC has provided an easy interface for forty years, with two upgrades along the way for color in the early 1960's, and stereo sound in the mid 1980's.

Easy interoperability has changed with the introduction of digital distribution in the cable industry. Digital delivery platforms start from a common MPEG video compression standard, but quickly turn to proprietary extensions which have challenged cross-platform interoperablity.

HBO began digital distribution of four feeds with such a proprietary system in 1992 using GI DigiCipher-I™ technology. Though this system used many of the components of MPEG-2 video compression, such as block-based processing using the Discrete Cosine Transform (DCT), it predated the MPEG-2 standard, and so used largely proprietary technology.

In December, 1996 HBO completed the replacement of this DC-I distribution system, upgrading its digital network to MPEG-2 video compression with AC-3 audio compression. In the process, HBO's digital distribution focus has changed from a single digital satellite transponder delivering four feeds (for analog redistribution to the home– the DC-I system) to two digital satellite transponders delivering sixteen feeds capable of analog or digital redistribution to the home (the MPEG-2 system).

Analog television has survived forty years (with the two "tweaks" given above) and continues strong today, while the first generation of digital television lasted only four years, and was replaced by MPEG-2 technology last year. And now ATSC Digital TV (DTV - standard and high definition digital television) is rapidly approaching.

DIGITAL CABLE BANDWIDTH

To set the stage for the discussion of interoperablity hurdles for digital systems, here is an overview of digital cable bandwidth, digital satellite technology, and digital delivery architecture.

Cable systems inherited the six MHz channel spacing of broadcast NTSC transmissions, and that same six MHz channel spacing is used for digital cable and broadcast DTV. One six MHz channel can deliver a single analog service, scrambled or in the clear.

One six MHz channel can deliver numerous digital programs, from four to ten, or more, given current compression technology. The motivations for adding digital channels to an analog cable plant are twofold: increase the number of programs available, and increase the picture and sound quality of those programs.

From the analog perspective, delivery capacity or bandwidth is measured in megahertz (frequency domain). In analog, more bandwidth means more quality. For example, broadcast FM radio uses more bandwidth than AM radio to achieve a higher quality.

From the digital perspective, bandwidth is measured in megabits per second (Mb/s), and, in general, more Mb/s means more quality. The building block for digital signals on a six MHz cable system using 64 Quadrature Amplitude Modulation (64 QAM) is 26.97 Mb/s. This payload of 26.97 Mb/s can be allocated among the different program services (audio, video, data) being delivered on that channel.

Digital cable systems (unlike green field DBS platforms which are entirely digital) are using a mixture of analog and new digital capacity. Basic channels with wide distribution will continue on analog channels which can be recovered with a $100 analog converter. Digital services, typically adding premium and PPV programming, will require a more expensive (by three to four times) hybrid analog and digital set-top box.

A distinguishing factor for digital systems is whether they provide broadcast or point-to-point delivery. In the digital broadcast model, all homes within a node receive an identical digital signal package. DBS uses the broadcast approach, and digital MMDS platforms seem to be developing 100% digital broadcast plants for multichannel video. An example of point-to-point systems (which use packet switching architectures) is the Time Warner Full Service Network in Orlando, Florida.

The channel increase by converting analog channels to digital can be calculated as follows. Rebuilt 750 MHz plant can support 116 analog channels. If the top 200 MHz is converted to digital service, the delivery capacity becomes 78 analog channels, and approximately 890 Mb/s for digital services:

33 channels x 26.97 Mb/s = 890 Mb/s

If you devote, for example, 4.5 Mb/s per program service, 890 Mb/s accommodates about 200 programs. Note that the entire 750 MHz bandwidth does not convert to 6 MHz channels because the entire 750 MHz spectrum is not available for downstream video services (there are gaps for return path, FM radio, and data services).

DIGITAL SATELLITE FORMAT

HBO, along with Showtime, Headend In The Sky (HITS), and TVN, has selected a GI DigiCipher core (conditional access and System Information) for digital distribution. HBO uses two-channel AC-3 audio compression and generic MPEG-2 video compression set for Main Profile at Main Level (MP@ML). The MPEG-2 Main Profile compression is configured to use B-frames and P-frames in a Group Of Pictures of fifteen, where N=15 and M=3. HBO's video compression frame sequence is shown below.

I B B P B B P B B P B B P B B I

HBO's decision to use Main Profile with B-frames was dependent on our affiliates selection of set-top boxes with the additional memory required to decode Main Profile streams.

HBO has chosen a digital multiplex format which delivers two "bundles" of program services on each of two digital transponders. Total digital bandwidth through a satellite transponder is dependent on the choices for modulation format, symbol rate, and error correction rate for the digital multiplex. HBO is using a symbol rate of 29.27 mega symbols per second (Ms/s), where each symbol carries two bits using Quadrature Phase Shift Keying (QPSK). This symbol rate and modulation establish a total digital bandwidth of 58.54 Mb/s through the transponder, which gives an occupied bandwidth that fits within the 36 MHz C-band transponders we use on Galaxy-IR.

The total bandwidth of 58.54 Mb/s is not available for digital television data since some bits must be used for forward error correction (FEC) to compensate for channel errors. Our digital signals use error correction in two main steps: a packet-based Reed-Solomon code which adds sixteen bits of error correction to each 188 bit MPEG packet, and a convolution code rate of 7/8 which adds one error correction bit to each seven data bits. Backing these error correction ratios out of the gross bit rate gives a net information rate of 47.2 Mb/s:

$$29.27 \text{ Ms/s} \times 2 \text{ b/s} = 58.54 \text{ Mb/s}$$
$$58.54 \text{ Mb/s} \times 188/204 \times 7/8 = 47.20 \text{ Mb/s}$$

This information rate is further split into two digital multiplexes using an I/Q split modulation (In-phase and Quadrature-phase) to yield 23.6 Mb/s per bundle of HBO digital programs. For reference, DC-I and MPEG-2 transmissions for Ku-band transponders typically deliver an information rate of 26.97 Mb/s.

HBO's digital program distribution uses two transponders on Galaxy-IR with program feeds allocated as shown below in Table 1.

Galaxy-IR Transponder 23
L-band Frequency 990 MHz
Virtual Channel Table 205

I Multiplex	Ch.	Q Multiplex	Ch.
HBO E	101	HBO W	111
HBO 2E	102	HBO 2W	112
HBO 3E	103	HBO 3W	113
MAX E	121	MAX W	131
MAX 2E	122	MAX 2W	132

Galaxy-IR Transponder 18
L-band Frequency 1090 MHz
Virtual Channel Table 206

I Multiplex	Ch.	Q Multiplex	Ch.
HBO M	106	HBOF E	104
HBO 2M	107	HBOF W	114
MAX M	126	future	---
MAX 2M	127	future	---
future	---	future	---

M = Mountain Time Zone HBOF = HBO Family

Table 1. HBO Digital Transponder Loading

More on the (virtual) channel number and the Virtual Channel Table below in the section on System Information.

Other premium programmers and pay-per-view services have converted to the MPEG-2 video/AC-3 audio format or have plans to convert to this format from analog or DC-I transmissions. Including some basic channels, there will be approximately 140 digital feeds in this format available directly from the programmers (including the sixteen given in Table 1 available from HBO).

HITS (a third party packager) is also transmitting digital programming in the GI format, offering packages of program services, along with other value-added services such as set-top management, authorization, and electronic program guide (EPG).

DIGITAL DELIVERY ARCHITECTURE

Figure 1 below gives some details about digital headend architecture, and interface points which may cause problems for interoperablity.

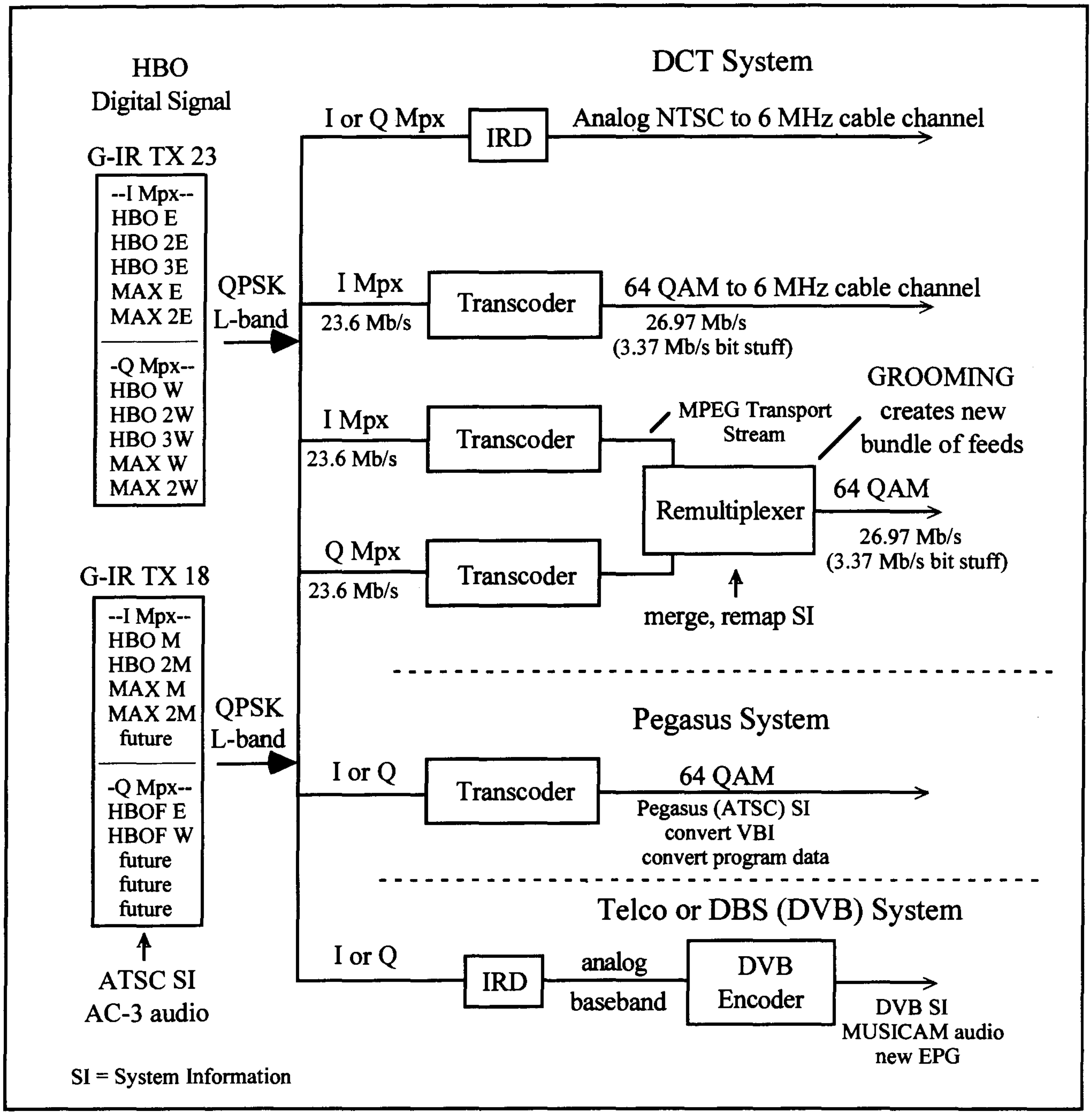


Figure 1. Simplified Digital Headend Architecture

INTEROPERABILITY HURDLES

A primary factor motivating interoperability is avoiding the equipment expense to compress the television signal a second time (up to $80,000 per channel), but there are also quality issues. Programming producers and suppliers would prefer to compress their programming a single time, and retain complete control over the compression algorithms and parameters. One pass through audio/video compression is good. Two passes is not as good, but still viable with cautious selection of compression parameters. Three passes through compression, especially involving encoders with different motion compensation algorithms, may produce excessive compression artifacts.

Digital program multiplexes originated from GI MPEG-2 encoders (whether direct from programmers, PPV operators, or third party packagers) will easily interoperate in a digital headend to send 64 QAM signals downstream to a GI DigiCable™ Consumer Terminal (DCT) set-top box. There are pros and cons involved in acquiring digital programming directly from a programmer or through a third-party packager such as HITS, but the technical interface is straightforward.

It is much more difficult for the digital multiplex to easily translate from the GI MPEG-2/AC-3 format to the Digital Video Broadcasting (DVB) format, used by some domestic platforms. The European DVB Project has produced specifications for digital video systems incorporating MPEG-2 video compression. Two US DBS operators and some telco video systems use DVB technology. The main areas of friction preventing interoperability between these two digital formats are these:

- Audio Compression
- System Information
- Conditional Access

Audio Compression

DVB systems use MPEG MUSICAM (Masking-pattern Universal Subband Integrated Coding And Multiplexing) audio compression. The United States Advanced Television Systems Committee (ATSC) has set a standard for audio compression using AC-3. ATSC digital audio compression is used in DTV, for North American DVD disks, and by the GI MPEG-2 system. AC-3 and MUSICAM use different algorithms and there is no direct translation available for the digital bit streams.

To help eliminate the need to completely reencode video and MUSICAM, HBO has developed a specification and identified a supplier for an AC-3 to MUSICAM converter which translates the audio compression (by way of baseband) while preserving the MPEG-2 compression video packets. This interface requires modification of the System Information to identify MPEG audio, and to correct the Presentation Time Stamps (PTS) and Program Clock References (PCR) to adjust for processing delay. To date, this approach has not been used to interface HBO digital signals to DVB platforms; instead the multiplex is decoded to baseband, and audio and video are then recompressed to the DVB standards. DBS operators who operate a single encoder for millions of subscribers have not indicated a desire to save encoder costs.

System Information

System Information (SI) is the reference data which describes and identifies all the pieces of a digital television signal. The SI consists of a number of tables which are used by MPEG-2 decoders to identify and recover data streams. Some of the main tables used are:

- Carrier Definition Table–gives digital carrier frequency.

- Modulation Mode Table–gives type of modulation used.
- Satellite Definition Table–identifies satellite, if applicable. There are other tables with additional satellite parameters.
- Source Name Table–gives text name of program source.
- Virtual Channel Table–cross reference which coordinates program selection.

The Virtual Channel Table must be downloaded to an MPEG-2 decoder before the decoder can select a virtual channel. HBO uses Virtual Channel Tables (VCT) numbered 205 and 206 for digital transponders on Galaxy-IR, transponders 23 and 18 respectively. This satellite System Information is modified when a digital multiplex is transcoded for cable distribution, for example, the satellite information is not meaningful and the modulation mode would change from QPSK to 64 QAM.

On a digital cable system, a simplified representation of SI tables would appear as shown in Figure 2 below.

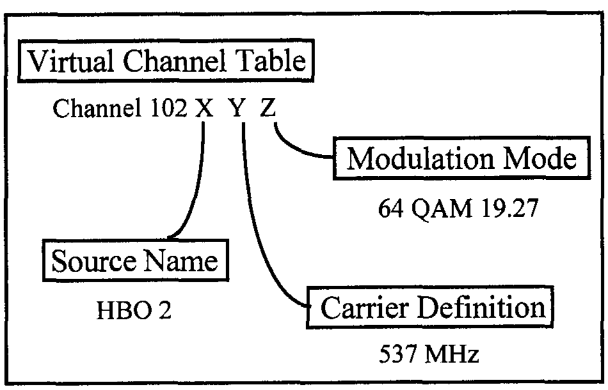


Figure 2. Simplified SI Tables

The System Information must be adjusted at each point the digital multiplex crosses a physical medium boundary; for example: satellite to cable or server to cable.

Harmonization of System Information between ATSC, DVB, and proprietary systems has had limited progress in standards committees. What is essentially a conversion mapping function is frustrated by redundant, reserved, or simply irrelevant data space between the various tables.

Conditional Access

Conditional Access is perhaps the most difficult area of interoperablity for digital systems for a simple reason–network operators do not want their security systems to cooperate. Programmers expect that access and encryption must be segmented in the different layers of a delivery system; different for the satellite and cable plant. Also note that even though the Europeans have developed a common encryption algorithm for DVB systems called Super Scrambling, conditional access remains autonomous.

Other Problem Areas

While audio compression, System Information, and conditional access are perhaps the major areas of interoperability problems, there are other areas of friction preventing a seamless interface.

Carriage of line 21 closed captioning is not just a good idea in the US, "it's the law". The MPEG-2 standard does not dictate the method for handling VBI data, and proprietary techniques have been used to transport captioning information as MPEG private data. The MPEG-2 decoder reconstructs the closed caption data on it's video output.

Different delivery platforms use different rating systems. HBO currently uses a combination of MPAA ratings and content advisories for its programming. These must be mapped to several different scales on different delivery platforms, with the real potential for consumer confusion.

Variable Data Rate encoding for video compression allows digital feeds sharing a multiplex to receive variable allocations of the total available bit rate, depending on their content. This approach can be described as "robbing Peter to pay Paul". An encoder channel with difficult material (high detail, fast motion) can temporarily receive more of the digital bandwidth within a multiplex. This can pose a serious problem for grooming architectures, where independently varying rates from different multiplexes must be combined. HBO currently uses constant data rate video compression to make remultiplexing easier, but is closely monitoring affiliate use of remultiplexing.

Program guide information can be supplied within the digital multiplex of the signals arriving at a headend, in different formats for each delivery platform. This information must be conformed at the headend to a unified format acceptable the set-top, or an alternative approach is to use a third-party guide provider to feed the electronic program guide.

CONCLUSION

In the first days of digital compression the possibility of multiplexing multiple video feeds on a single channel was a real break-through in the crowded bandwidth on satellite and wireline systems. There were technical trials and products launched using the MPEG-1 SIF (Source Input File) format, and at that time some felt that "VHS quality" was good enough.

As more digital systems become available, the consumer appetite and expectations for multichannel video are shifting from sheer quantity back towards the quality of viewing experience. Consumers think anything with the "digital" nomenclature must provide a superior quality experience.

A factor which impedes harmonization of digital video is that the market pressure to deploy often overruns the pace of technical committee standardization work. While committees debate the details not covered by MPEG-2, proprietary systems have been deployed. As of December 31, 1996, there were 4.3 million DBS subscribers using digital compression but less than one tenth were receiving "standard" MPEG. Standards committees and vendor cross licensing still must deal with the fact that a deployable cross-platform delivery platform does not exist.

Vendors must cooperate and compromise to achieve interoperability. The hard work is still ahead.

ACKNOWLEDGMENT

The author would like to thank Craig Cuttner, chair of the SCTE Digital Video Standards Working Group #1 on Audio/Video, for his assistance and insight.

REFERENCES

1. System Information for Digital Television, ATSC Standard Doc. A/56, January 3, 1996.

2. Digital Audio Compression Standard (AC-3), ATSC Standard Doc. A/52, December 20, 1995.

3. ETS 300 468 DVB System Information Table.

4. ISO/IEC 13818-3 (MPEG-2 Audio Layer II)

5. EBU Technical Review No. 266, Winter 1995.

MULTI-LAYER HEADEND COMBINING NETWORK DESIGN FOR BROADCAST, LOCAL, AND TARGETED SERVICES

Oleh J. Sniezko
Director
Transmission Engineering
TCI Communications, Inc.

Terrace Tower II
5619 DTC Parkway
Englewood, CO 80111-3000
(303) 267-6959
FAX (303) 488-3210

Abstract

In modern hybrid fiber-coaxial (HFC) networks, the target subscriber groups for various signals originating from headends and hubs are different. Some of the signals are of broadcast type destined to all customers, the others target a limited number of customers. These differences must be reflected in the design of the headend/hub combining/splitting networks to take advantage of spectrum re-usability.

This paper describes headend and hub combining/splitting networks designed to meet the new requirements of narrowcast and targeted services. It also presents a combining/splitting network for reverse signals that require a distinctly different approach from the forward signals.

GENERAL

Headends and hubs serve as major signal sources and processing centers in hybrid fiber-coaxial networks. The quality of signals originating from these centers is a reference line for the quality of signals delivered to our customers. These signals will never be better than at these origination points.

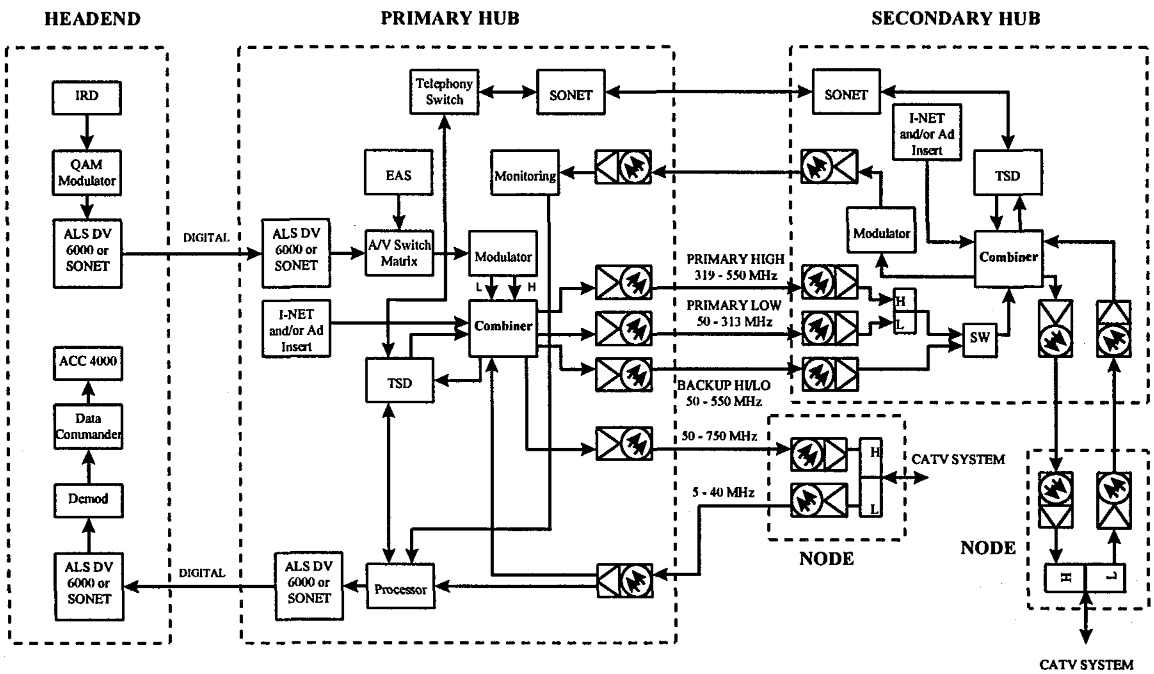
Due to the complex character of the modern headends and hubs and their critical impact on network availability and signal quality, an adequate design for all headend and primary hub forward combining/splitting networks is extremely important. Headend reverse combining network design should be flexible enough to accommodate new services. This should be achieved without future service disruptions.

To enhance the flexibility of the design, the combining/splitting network is divided into sublevels. Each sublevel is designed independently of the other sublevels, with precisely defined interface parameters for compatibility. Headend and primary hubs include combining/splitting networks with a maximum of four functional sublevels. The number of sublevels at a particular location depends on the output level from optical receivers and on the number of optical nodes served by the hub.

SIGNAL ACQUISITION AND PROCESSING CENTERS

The HFC network is supported by processing and routing equipment grouped in headends and hub facilities. These facilities are categorized as headends, primary hubs and secondary hubs depending on their function and location.

Figure 1: Headend Configuration — Example



The headend serves as an entry point in the primary hub ring and may include proprietary or SONET transport system interfaces. Satellite signals and signals from other sources are received here, processed and distributed to the other sections of the system.

The primary hub functions as an acquisition, modulation, combining, and distribution center in the primary hub ring. The primary hub may also deliver signals to the secondary hub ring using fiber transport systems, and to the optical nodes served directly from that primary hub. Once signals are received from the primary hub ring, they are routed to modulators and then into the combining system. Signals originating from other sources also enter the forward combining system. Reverse data, telephony, status monitoring and PPV signals are processed, combined, and routed to their destinations.

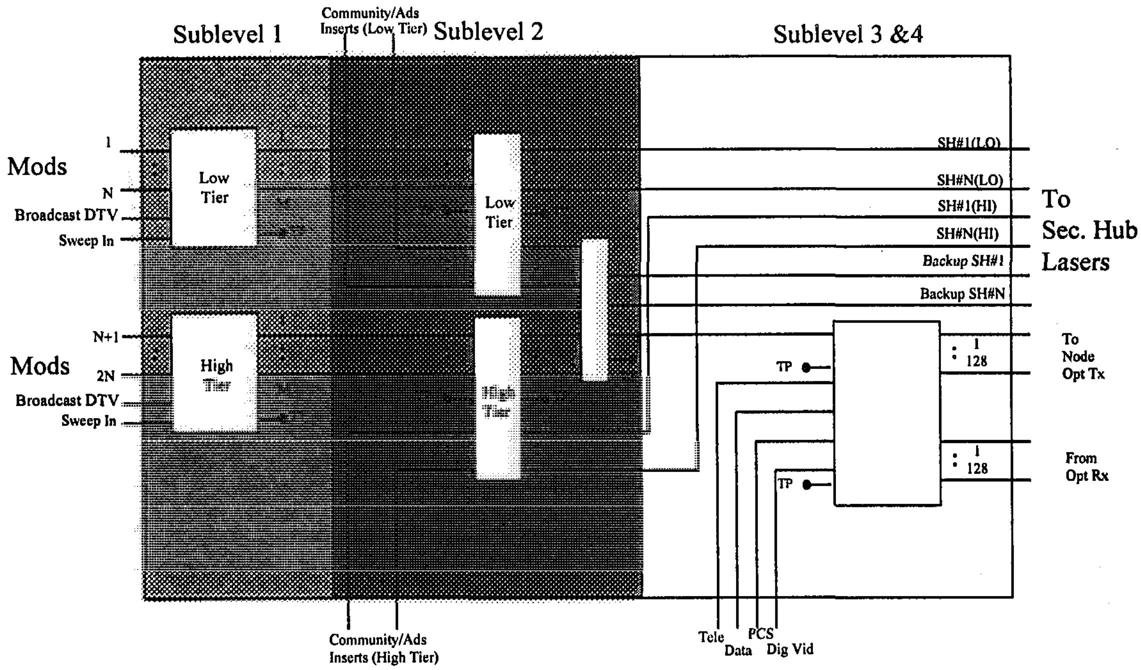
The secondary hub interfaces to the secondary hub ring and serves local nodes. Redundancy switching precedes combining. Telephony and data signals interface through the targeted service delivery access points. Monitoring signals also flow through the secondary hub on their way to the primary hub.

A particular HFC network may not include all these facilities depending on size of the market and alternative architecture design but most will contain at least two elements in large and medium size markets.

FUNCTIONAL DESCRIPTION

The headend and primary hubs include a combining/splitting network with up to four functional sublevels (see Figure 2). The secondary hubs include a combining/splitting network with two or three functional sublevels. The number of sublevels depends on the function of the secondary hub, signal processing complexity and on the number of optical nodes served.

Figure 2: Combining/Splitting Network — Overview



Sublevel 1

Sublevel 1 will function as an input combining/splitting unit. It uses a dual-tier approach with low and high tiers. Two optional combining configurations and two optional splitting configurations exist. The choice of the optimal combination of the configuration depends on the type of modulator bank used and on the number of secondary hubs served from the location. The following recommendations will help in the selection of the appropriate input combining option:

1. Input combining Option 1 combines up to 6 pre-combined modulator inputs for a total of 48 modulators or other signal sources per tier. This option is recommended for all applications with manufacturer-pre-combined modulators. Modulators are usually pre-combined to manage spurious and out-of-band noise that can be dominant in agile units. In many cases where local ad insertion is required, empty slots will be left in the pre-combined channel groups to accommodate the local channel insertion in Sublevel 2.
2. Input combining Option 2 allows for combining of up to 48 single modulators or other signal sources per tier.

In both cases, two additional inputs in each tier can be used for combining sweep signals and broadcast digital TV signals.

The following recommendations will help in the selection of the appropriate output splitting option:

A. Output splitting Option A provides up to 7 outputs for each tier and serves up to three secondary hubs (both principal and back-up feeds) and local optical nodes. If the local nodes are not served from this primary hub, an additional secondary hub can be served (total of 4).
B. Output splitting Option B of Sublevel 1 provides up to 15 outputs and serves up to seven secondary hubs (both principal and back-up feeds) and local optical nodes. If the local nodes are not served from this primary hub, an additional secondary hub can be served (total of 8).

Sublevel 2

Sublevel 2 functions as an ad insertion combining unit for both the low and high tiers. Sublevel 2 additionally provides signal amplification. It has a backup combining unit in case the principal Sublevel 2 combining unit fails. Moreover, Sublevel 2 provides combining network for local distribution to optical nodes.

Sublevel 3

Sublevel 3 provides splitting and amplification network for feeding up to 128 optical nodes directly from the primary hub. It also combines TSD service signals with broadcast and local signals from Sublevels 1 and 2. Sublevel 3 network is also implemented in secondary hubs after the redundancy switch (see Figure 1).

Sublevel 4

Sublevel 4 provides combining network for targeted service delivery signals in the forward direction and splitting network for target service delivery signals from the reverse path optical receivers.

SIGNAL COMBINING/SPLITTING NETWORK — DETAILS

Sublevel 1 (Figure 3)

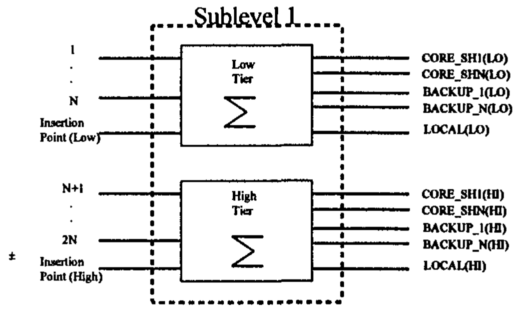
Almost all elements in this level of the combining network are non-redundant. All components, except for the first combining gear following the modulators and the last attenuator before pre-amplifiers in Sublevel 2, are shared among all customers served by the headend/primary hub. The purpose of this sublevel is to combine signals from all broadcast sources, including broadcast digital TV, and split the combined signals to feed the secondary hubs and local optical nodes.

Design Objectives

The following objectives should be achieved:

1. Sublevel 1 section of the network must be designed for all future expansion in the number of secondary hubs served. Future reconfiguration should not cause service disruption.
2. All elements must be reliable and be assembled and cabled to achieve the highest possible level of reliability. Only approved elements should be used. The number of connectors must be limited and cables kept as short as possible.
3. Output levels of Sublevel 1 must be optimized to minimize distortion introduced by Sublevel 2.

Figure 3: Sublevel 1 of Combining /Splitting Network



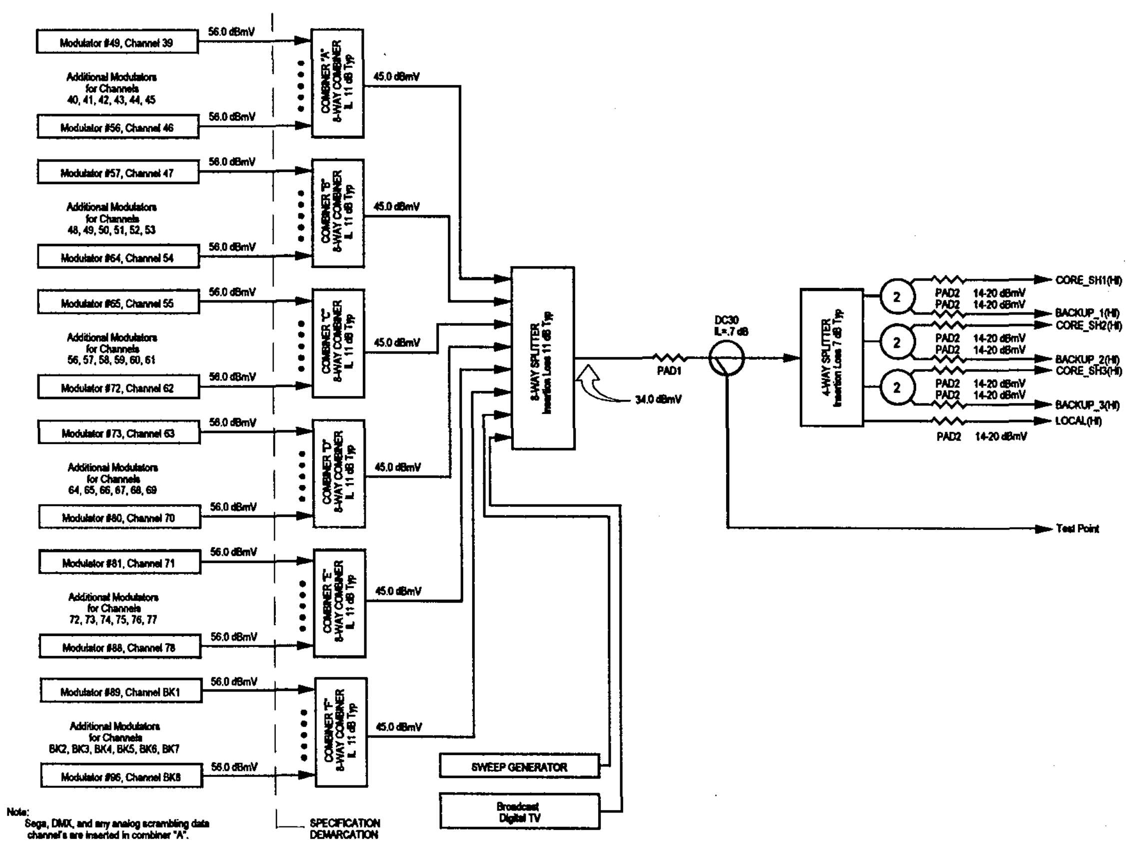
Detail Description — Input Combining Options (see Figure 4)

Low tier combining serves channels 2 through 38, and high tier combining serves channels 39 through 78. A single tier of Option 1 input combining in Sublevel 1 combines input from up to 6 pre-combined sources. A single tier of Option 2 input combining combines outputs from up to 48 single modulators or other sources. In Option 2, the combining is achieved by two layers of 8-way combiners.

NOTE 1: In Option 2, channels into the first layer of combining should be arranged in a non-consecutive order to minimize spurious generated in the output stages of the modulators.

NOTE 2: Output signal levels from modulators should be pre-emphasized to account for diplex filter crossovers.

Figure 4: Sublevel 1 — Example: High Tier Combining Using Single Channel Modulators Feeding 3 SH Lasers (Option H2A)



Broadcast digital TV signals are pre-combined for an equivalent of up to 16, 24, or 32 analog channels (96, 144, or 192 MHz). The pre-combining is achieved with 8-way combiners followed by a 2-way, 3-way (balanced) or 4-way combiner. These signals are combined with the analog broadcast signals in Sublevel 1. The combining is achieved through a spare input of the 8-way combiner. The digital signal levels (average power per channel) at the output of Sublevel 1 combining network are set to the required level in relation to the analog signal levels (peak-value). These levels can be set by inserting a pad of an adequate value before combining with the analog channels.

NOTE 3: If any of the digital TV signals occupies a channel between 2 and 78, it can be injected in the combining network in place of that analog channel at an adequate level as per the design for the particular system.

A sweep generator is directly coupled to one of the 8-way combiner spare inputs.

Detail Description — Output Splitting Options (see Figure 4)

The signals are next directed to a splitting section of Sublevel 1 via an optional pad (usually omitted) and 30 dB forward test point coupler. Two different options exist in the splitting section: Option A serves 3 secondary hubs and a number of local nodes, and Option B serves 4-7 secondary hubs and a number of local nodes.

NOTE 4: Future removal or replacement of the attenuators due to level requirements would result in service disruption. Hence, it is recommended that PAD1 not be placed.

In Option A, the combined signal from each tier is fed to a dedicated 4-way splitter. Three outputs of this splitter are split again and the outputs of the 2-way splitters feed principal and backup sections of Sublevel 2 via dedicated attenuators (PAD 2). The last output feeds pre-amplifiers of Sublevel 2 local node links via level adjusting attenuators (PAD 2). If no local optical nodes are served from the primary hub, the number of secondary hubs served can be increased to four.

Option B is similar to Option A with the 4-way splitters replaced with 8-way splitters.

NOTE 5: If the current or anticipated number of secondary hubs served is higher than 3, use only 8-way splitter (splitting Option B) in the splitting section of Sublevel 1.

Detail Description — Output Level Alignment (see Figure 4)

Sublevel 1 output levels should be within 17 dBmV ± 3 dB. For low gain pre-amplifiers (17 dB), they should be as close as possible to 20 dBmV. For high gain pre-amplifiers, output levels should be as close as possible to 14 dBmV. Guidelines for using low or high gain pre-amplifiers are presented in the description of Sublevel 2. The levels are adjusted with PAD 2 attenuators.

NOTE 6: Future removal or replacement of the attenuators in PAD 2 location of the local optical node (ON) path would result in service disruption. Hence, it is recommended that this attenuator value be selected based on the current and anticipated number of optical nodes served. The attenuators on SH paths are redundant.

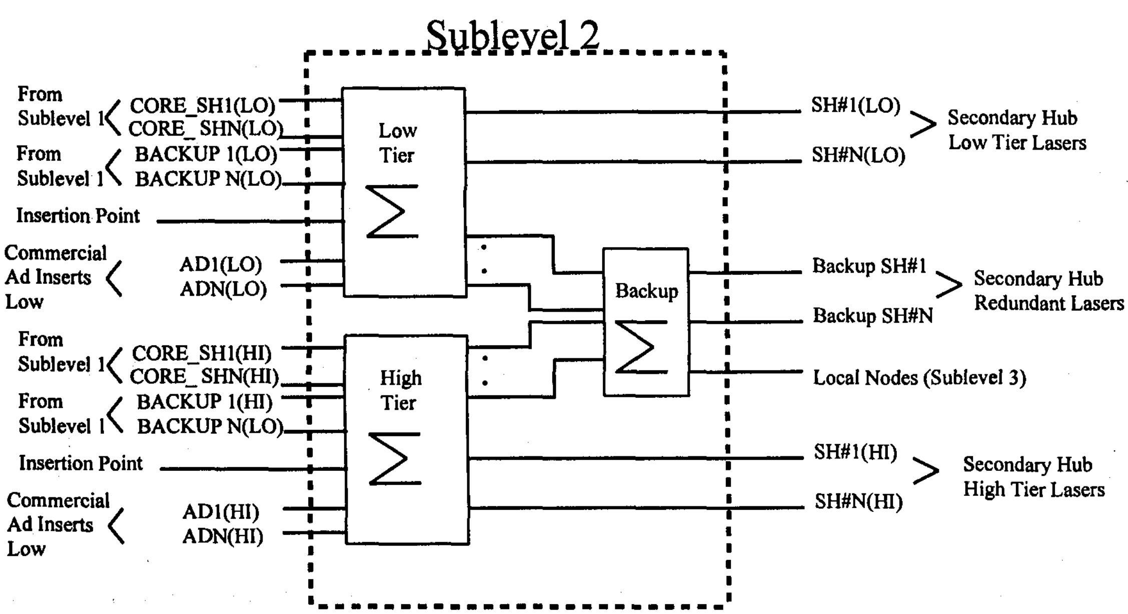
Table 1: Design Choices — Sublevel 1

Network Elements	Choices	Selection Criteria	Note #
Input combining network for analog signals	Option 1	Pre-combined modulators	
	Option 2	Single modulators in primary hubs	
Input Combining network for digital TV signals	Combined through an 8-way combiner	Frequency allocation above analog channels	
	Injected directly as analog modulators	Frequency allocation between analog channels	Note 3
Pad value in PAD 1	0	Level requirements	Note 4
Output splitting network	Option A	For 3 secondary hubs and up to 128 local nodes served or for up to 4 secondary hubs served	
	Option B	For up to 7 secondary hubs and up to 128 local nodes served or for up to 8 secondary hubs served	Note 5
Pad value in PAD 2 in local ON path	Lowest possible	Low input lasers, Sublevel 1 output ≤20 dBmV	Note 6
	Highest possible	High input lasers, Sublevel 1 output ≥14 dBmV	Note 6

Sublevel 2 (Figure 5)

The elements of this level of the combining network are redundant for secondary hub links but are not redundant for local optical node links except for the pre-amplifiers for optical nodes. The purpose of this sublevel is to amplify the combined broadcast signal to levels adequate for secondary hub lasers or for Sublevel 3 in optical node section. Sublevel 2 in secondary hub section allows for inserting signals targeted to a particular community (ad inserts, local programming, and other community targeted services). Sublevel 2 also combines tiers in back-up secondary hub and optical node sections. The tier combining can also be performed in principal secondary hub sections if the design calls for single-fiber principal secondary hub links. However, the tier configuration is maintained even in this case to lower the distortion contribution of the combining network amplifiers.

Figure 5: Sublevel 2 of Combining /Splitting Network



Design Objectives

The following objectives should be achieved:

1. Sublevel 2 section of the network must be designed for adequate levels to the secondary hub lasers and to Sublevel 3 for local ON.
2. Optical node section of Sublevel 2 must be designed for all future expansion in the number of optical nodes served. Future reconfiguration should not cause service disruption.
3. Output levels of Sublevel 2 in the optical node section must be optimized to minimize distortion introduced by Sublevel 3.
4. All elements must be reliable and be assembled and cabled to achieve the highest possible level of reliability. Only approved elements should be used. The number of connectors must be limited and cables kept as short as possible.
5. Active elements of Sublevel 2 must be selected to minimize distortion.
6. The design of the secondary hub section of Sublevel 2 must provide adequate isolation between local programming signal sources.

Detail Description — SH Principal and Back-Up Amplification (see Figure 6)

The input signals from Sublevel 1 are amplified to the level adequate for the secondary hub link laser inputs. The gain of the amplifier should be higher for lasers that require a higher input level in order to achieve adequate isolation between local distribution signals. The amplifiers are followed by attenuators in PAD 3 locations. The values of the attenuators are maximized to achieve the required isolation between local ad insertions and local programming for different optical links. The attenuators are followed by directional couplers for local ad insertions and local programming, and by DC-30 test points. These test points serve as reference points for signal flatness alignment for the entire headend/hub. For single-fiber principal secondary hub links, diplex filters for tier combining are installed before the test point couplers. Since back-up links are a single-fiber type by design, diplex filters for tier combining are always installed.

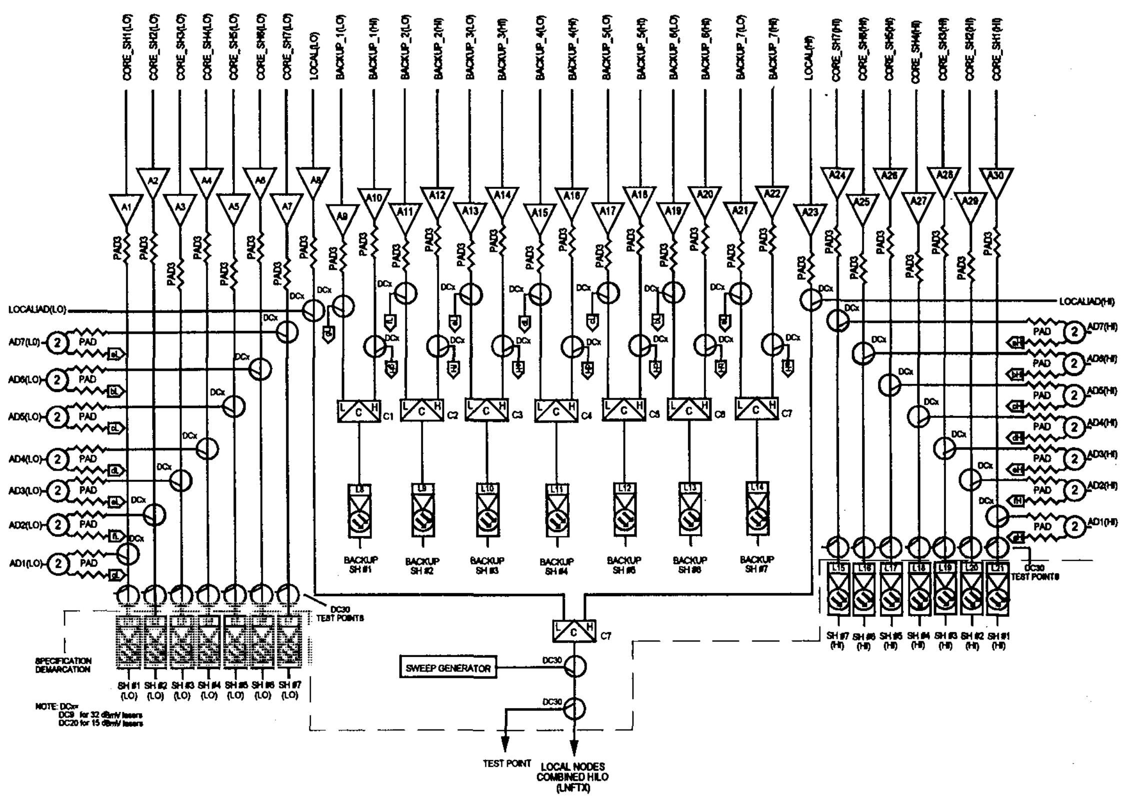
NOTE 7: The tier arrangement is maintained even in the case of single-fiber principal secondary hub links to minimize the loading of active components of the combining network and noise accumulation from modulators and active components.

Detail Description — Optical Node Link Amplification (see Figure 6)

The input signals from Sublevel 1 are pre-amplified to the level adequate for the Sublevel 3. The gain of the amplifiers is higher (not higher than 21 dB) in hubs that feed more than 32 optical nodes with ON lasers that require higher input level. The amplifiers are followed by attenuators in PAD 3 locations. The values of the attenuators are maximized to achieve the required isolation between local ad insertions and local programming for different optical links. The attenuators are followed by directional couplers for local ad insertions and local programming. After the couplers, the signals from the two tiers are combined with diplex filters. The diplexers are followed by DC-20 sweep insertion and DC-30 test points.

NOTE 8: Amplifiers in local optical node links should be configured in a parallel setup or be redundant (with A/B output switching). They should be also monitored for failure. These amplifiers may serve tens of thousands of customers.

Figure 6: Sublevel 2 — Example: High/Low Tier Combining Feeding 7 Secondary Hubs



Detail Description — Local Community Programming & Ad Combining Network

Local service signal combining network consists of 8-way combiners followed by 2-way combiners. The 2-way combiner cannot be used for ad inserts if the lasers in the secondary hub links require 32 dBmV or higher input level. The combined signals are fed into two-way splitters (splitting for principal and back-up links). No splitter is required for optical node links. The outputs of the splitters are followed by attenuators for independent level adjustment to the optical lasers. After attenuation, the signals are combined with the broadcast signal using directional couplers (DC-9). Broadcast or local digital TV signals can be inserted at this point to a community.

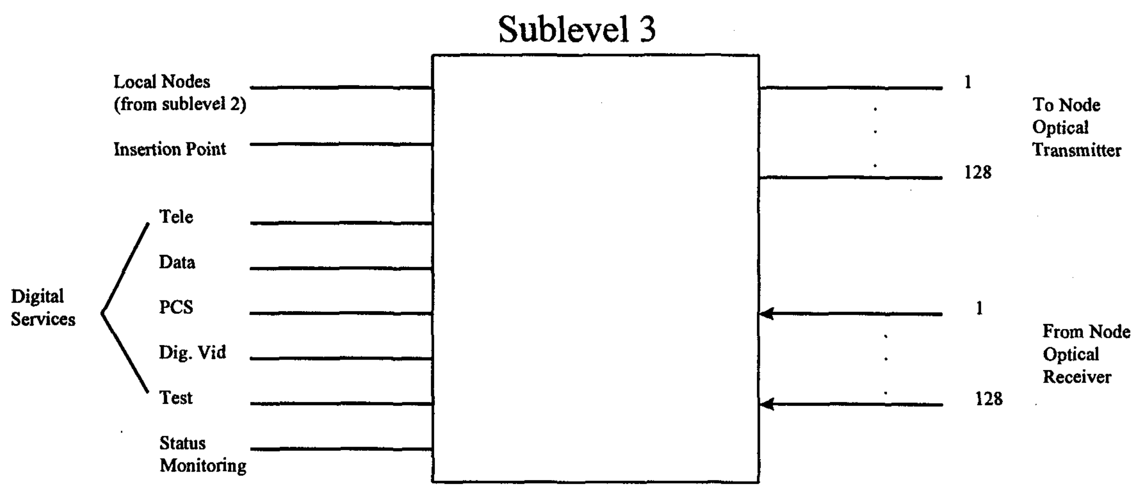
Table 2: Design Choices — Sublevel 2

Network Elements	Choices	Selection Criteria
Pad value in PAD 3	Highest possible	Level required at the laser input or at the input to Sublevel 3
Combining network for local ad inserts and programming	16 or 8 channels per tier	Level required at the laser input
Diplexer in principal SH links	In primary or secondary hub	Performance of the secondary hub lasers either allow or not allow for single-fiber principal SH links
Amplifier gain	17 dB or 21 dB	The number of secondary hubs and optical nodes served, and the input level requirements of the optical node lasers (17 dB gain amplifiers are sufficient in all situations where all lasers require no more than 18 dBmV input)

Sublevel 3 (Figure 7)

The elements of this level of the combining network are not redundant. The purpose of this sublevel is to split and amplify the combined broadcast and local signals to provide an adequate number of outputs and sufficient input levels to optical node lasers. Sublevel 3 allows inserting target delivery service signals and digital TV signals to a particular node.

Figure 7: Sublevel 3 of Combining /Splitting Network



Design Objectives

The following objectives should be achieved:

1. Sublevel 3 section of the network must be designed for adequate levels to the optical node lasers.
2. First splitting section of Sublevel 3 must be designed for all future expansion in the number of optical nodes served. Future reconfiguration should not cause service disruption.
3. All elements must be reliable and be assembled to achieve the highest possible level of reliability. Only approved elements should be used. The number of connectors must be limited and cables kept as short as possible.
4. Active elements of Sublevel 3 must have sufficient bandwidth capacity and be selected and aligned to minimize distortion.

Detail Description — Input Splitting (see Figure 8)

Signals from Sublevel 2 are fed into an 8-way splitter and then directly into amplifiers (for 32 optical nodes) or into 2-, 3- , or 4-way splitters for up to 64, 96, or 128 optical nodes served.

Detail Description — Amplification for ON Lasers (see Figure 9)

The outputs are fed into single- or dual-output pre-amplifiers followed by four- or two-way splitters if required. The gain of the amplifiers will be defined by the input level requirements of the optical node lasers. Next, the signals are routed to the optical node lasers via attenuators and TSD-insertion directional couplers followed by 30 dB test point couplers.

Detail Description — TSD Combining (see Figure 9)

The TSD signals from different service terminals are combined using 8-way combiners. The outputs of the combiners are fed to the insertion couplers via 30 dB test point couplers.

Figure 8: Sublevel 3 — Example: Local Node Configurations for 32, 64, 96, and 128 Laser Nodes

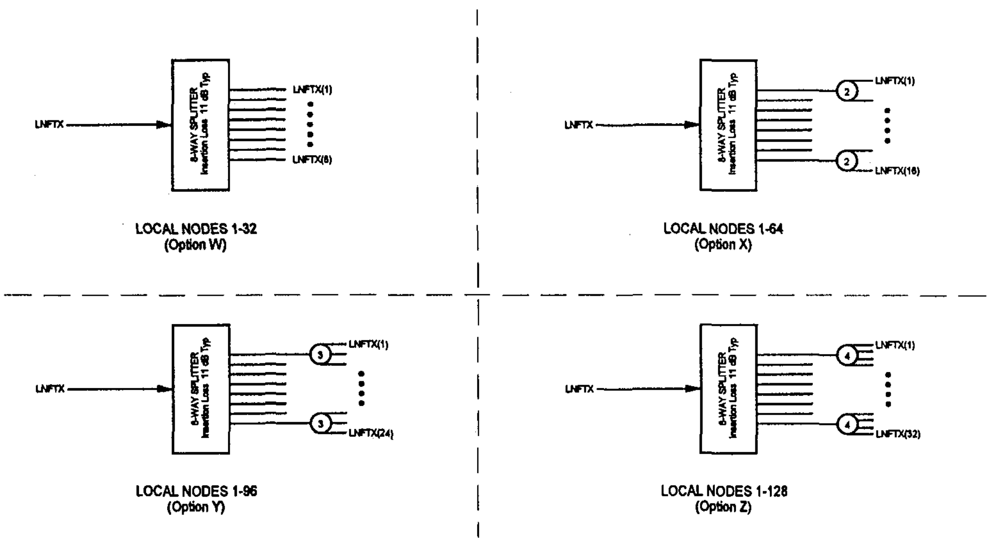
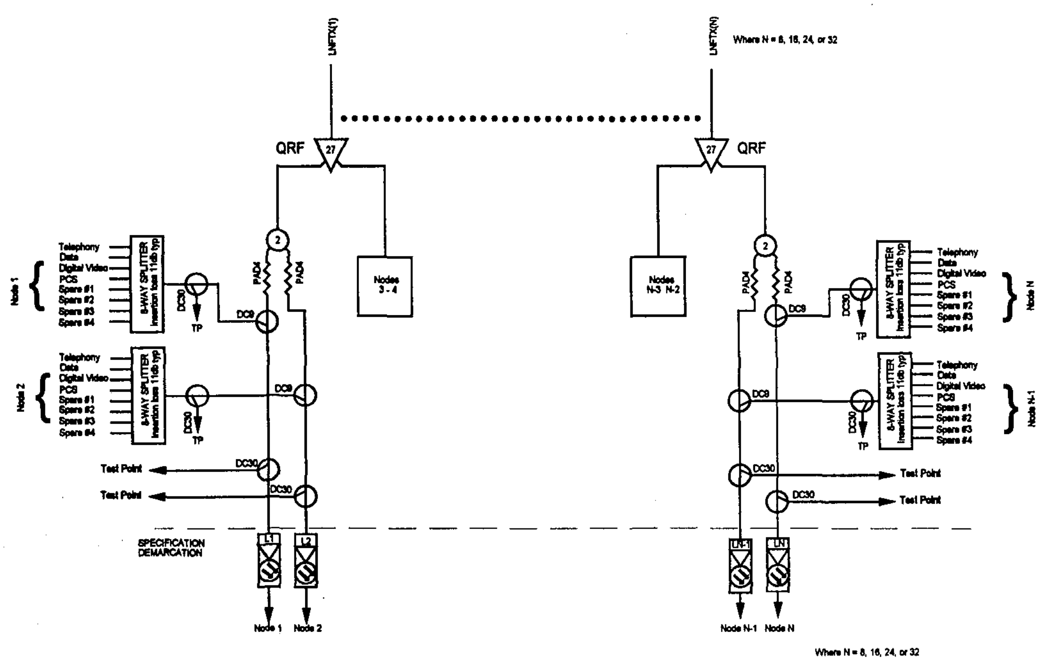


Figure 9: Sublevel 3 — Example: 96 Node Fiber Distribution



Option	Inputs	Outputs
W	8	32
X	16	64
Y	24	96
Z	32	128

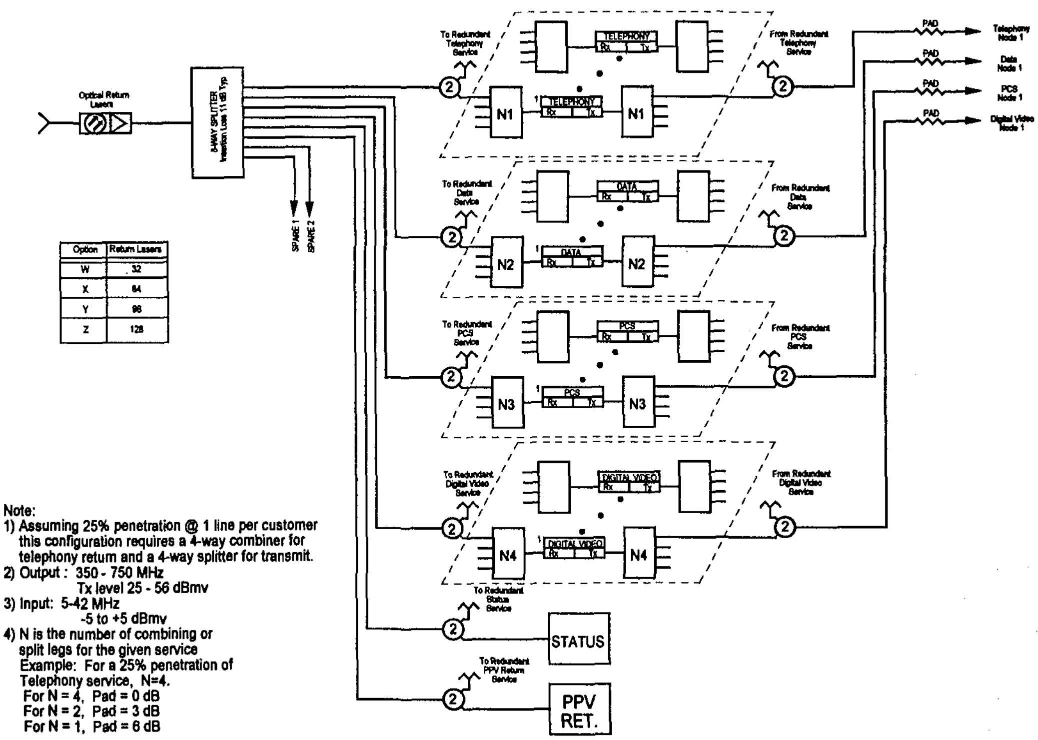
Table 3: Design Choices — Sublevel 3

Network Elements	Choices	Selection Criteria
Input splitting network	8, 16, 24, or 32 outputs	The number of optical node links served from the primary/secondary hub
Amplifier gain for local optical node links	17 dB or 25 (27) dB	The input level requirements of the optical node lasers (17 dB gain amplifiers are sufficient in all situations where all optical node lasers require no more than 18 dBmV input)
Pad value in PAD 4	Highest possible	Level required at the laser input

Sublevel 4 (Figure 10)

The elements of this level of the combining network are not redundant. However, each individual service combining/splitting network contains two splitters (one for forward and one for reverse) for non-disruptive network rearrangement. The purpose of this sublevel is to combine or split the targeted service delivery signals in the reverse and forward directions dependent on traffic engineering.

Figure 10: Sublevel 4 — Example: Return Path Combining



Design Objectives

The following objectives should be achieved:

1. Sublevel 4 section of the network must be designed for adequate levels to terminal equipment (reverse direction) and to optical node lasers (forward direction).
2. The network reconfiguration causing service disruption to all customers served by local optical nodes are limited to a replacement of the faulty elements.
3. The elements of the section must be reliable, and assembled to achieve the highest possible level of reliability. Only approved elements

should be used. The number of connectors must be limited and cables kept as short as possible.

4. Adequate isolation between target service delivery signals received from and directed to different target customer groups must be achieved.

Detail Description — Input Splitting in Reverse Path

Signals from reverse optical node receivers are fed into an 8-way splitter followed by 2-way redundancy splitters. After the two-way splitters, the signals are fed into individual service's terminal equipment.

Detail Description — Output Combiner

The outputs of individual service's terminal equipment are fed into a 2-way combiner via a splitting network. The outputs of the two-way combiners interface with Sublevel 3.

INPUT/OUTPUT SPECIFICATIONS

The following paragraphs present a short summary of the extensive analysis performed to optimize levels and allow seamless interface between the sublevels.

Sublevel 1 I/O Specifications

Input Combining Option 1

Pre-combined modulator output levels should be set at 43.0 dBmV. The minimum input CNR to the Sublevel 1 combining should be 65.0 dB. Spurs should be no greater than -65.0 dBc and the isolation minimum of 20.0 dB port-to-port.

Input Combining Option 2

Analog modulator output levels are nominally set to 56 dBmV. This setting allows for a frequency response alignment margin. The minimum input CNR from the modulators should be 67.0 dB. Spurs should be no greater than -65.0 dBc and the isolation minimum of 20.0 dB port to port.

Broadcast Digital TV Signals

Digital up-converter output levels are nominally set to 56 dBmV.

Output Signal Performance

A minimum of 63 dB CNR shall be achieved at the input to Sublevel 2 with spurs no greater than -65 dBc and a minimum isolation of 20.0 dB port to port. The levels shall be within 17 dBmV ±3 dB.

Table 4: Input/Output Specification for Sublevel 1

	Inputs		Outputs	
Options	**Option 1**	**Option 2**	**Option A**	**Option B**
Parameter				
Number of total analog inputs or outputs	12 with pre-combined modulators	96	16 or 18	32 or 34
Number of ports per tier	6	48	8 or 9 (3 or 4 principal, 3 or 4 back-up, 1 or 0 local ON, TP)	16 or 18 (7 or 8 principal, 7 or 8 back-up, 1 or 0 local ON, TP)
Number of additional ports (sweep, DTV) per tier	2	2		
Analog TV levels	43 dBmV	56 dBmV	17±3 dBmV	17±3 dBmV
Broadcast DTV levels*	33-37 dBmV	35-39 dBmV	6-10 dB lower	6-10 dB lower
CNR	65 dB	67 dB	63 dB	63 dB
Spurious	-65 dBc	-65 dBc	-65 dBc	-65 dBc
Port-to-port isolation	20 dB	20 dB	20 dB	20 dB

* At the input to the 8-way combiner.

Sublevel 2 I/O Specifications

Pre-Amplification

Input levels are predetermined by Sublevel 1 of the combining/splitting network. These levels are optimized for the CNR and distortion performance. The signals are then pre-amplified by high quality power doubling or feedforward amplifiers working at light loading (half of the 77 or lower channel loading) for better distortion performance. The pre-amplifiers' main purpose is to provide for an adequate isolation between local programming signals injected into different optical links. For example, the isolation (more accurately, the ratio of wanted to unwanted signals) between the local signal injected into the laser that requires 32 dBmV input and the local signal injected into another laser may be lower than 50 dB if the attenuator values are low (for low gain amplifiers) and their locations are not selected in an optimal way. The isolation between the local signal injected into the laser that requires only 15 dBmV input and the local signal injected into another laser may, on the other hand, be as high as 85 dB. Amplifiers of higher gain will be required for the high input lasers. Otherwise, isolation objective would not be achieved. On the other hand, the use of higher gain amplifiers compromises the distortion performance. The amplifiers also amplify signals to the level required by SH lasers, especially lasers with high input level requirements.

Table 5: Input/Output Specification for Sublevel 2

	Principal SH		Back-up SH		Local ON		Local Signals	
Parameter	**Input**	**Output**	**Input**	**Output**	**Input**	**Output**	**Input**	**Output**
Broadcast analog, level (broadcast digital 6-10 dB lower)	17±3 dBmV	15, 18, or 32 dBmV ±1.5 dB	17±3 dBmV	15, 18, or 32 dBmV ±1.5 dB	17±3 dBmV	20 to 31 dBmV	NA	NA
Levels from local programming sources	29 to 46 dBmV	16, 19, or 33 dBmV	31 to 48 dBmV	18, 21, or 35 dBmV	33 to 45 dBmV	23 to 35 dBmV	Max 59 dBmV	29 to 46 dBmV
CNR (min)	63 dB	60 dB	63 dB	60 dB	63 dB	60 dB	67 dB	63 dB
CTB (max.) for Options 1B and 2B with 32 dBmV lasers	NA	-86 dBc	NA	-75 dBc	NA	-86 dBc	NA	NA
CTB (max.) for the remaining options	NA	-86 dBc	NA	-86 dBc	NA	-86 dBc	NA	NA
CSO (max.) for Options 1B and 2B with 32 dBmV lasers	NA	-75 dBc	NA	-67 dBc	NA	-75 dBc	NA	NA
CSO (max.) for the remaining options	NA	-75 dBc	NA	-75 dBc	NA	-75 dBc	NA	NA
Flatness (at the SH laser TP)	NA	±0.5 dB, including slope	NA	±0.5 dB, including slope	NA	NA		
Ratio of wanted to unwanted signals between local signals in different links		min 55 dB, preferably >65 dB		min 55 dB, preferably >65 dB		min 55 dB, preferably >65 dB		
Spurious							-65 dBc	-65 dBc
Port-to-port isolation							20 dB	20 dB

Local Programming Combining Network

Single analog modulator output levels are nominally set to 56 dBmV. This setting allows for a frequency response alignment margin. The minimum input CNR from the modulators should be 67.0 dB. Spurs should be no greater than -65.0 dBc and the isolation minimum of 20.0 dB port to port. Other signal sources (for example digital TV signals for local distribution or I-Net) are set at lower levels as per the system design.

Sublevel 3 I/O Specifications

Input Splitting Network

Sublevel 2 output levels to Sublevel 3 inputs are predetermined by Sublevel 2 optical node pre-amplification network. The signals are then split to feed 32, 64, 96, or 128 optical nodes.

Optical Node Pre-Amplification Network

The signals from the input splitting network are pre-amplified by high quality power doubling amplifiers. The pre-amplifiers' main purpose is to provide adequate levels to the ON lasers. A single amplifier should feed a limited number of homes passed to limit failure groups. The analog outputs to the optical node lasers should be set at either 15, 18, or 32 dBmV with digital outputs set at the design relative levels.

TSD Signal Combining Network

Digital TSD levels to the 8-way combiner should be adjustable between 27 and 48 dBmV. This requires a 59 dBmV/6 MHz level capability from TSD transmitters. If this level is not available, lower input lasers must be deployed, TSD level must be lowered or transmitter splitting must be limited.

Table 6: Input/Output Specification for Sublevel 3

	Broadcast and Local Input Splitting Network		**TSD Input Combining Network**		**Amplification Network**	
Parameter	**Input**	**Output**	**Input**	**Output**	**Input**	**Output**
Level (broadcast and local analog, broadcast and local digital 6-10 dB lower)	20 to 31 dBmV	9 to 13 dBmV	NA	NA	9 to 13 dBmV	15, 18, or 32 dBmV
Levels from TSD sources			27 to 48 dBmV	6 to 27 dBmV		
CNR min	60 dB	60 dB	NA	NA	60 dB	56 dB
CTB max.	-86 dBc	-86 dBc	NA	NA	-86 dBc	-70 dBc (-86 dBc for 18 dBmV input lasers)
CSO max.	-75 dBc	-75 dBc	NA	NA	-75 dBc	-66 dBc (-75 dBc for 18 dBmV input lasers)
Flatness (at the SH laser TP)	NA	NA	NA	NA	±0.5 dB, including slope	±1.0 dB, including slope
Spurious			-55 dBc	-55 dBc		

Sublevel 4 I/O Specifications

Table 7: Input & Output Specification for Sublevel 4

	Forward		**Reverse**	
Parameter	**Input**	**Output**	**Input**	**Output**
Maximum level (average power/6 MHz)	59 dBmV	27 to 48 dBmV	20 to 32 dBmV	-5 to 10 dBmV

WIRING AND CABLING PRACTICES

All cables within the combining /splitting network should be routed between their attachment points the shortest possible way. All cable bends must be performed in sweeps with a radius >2.5" for individual cable jumpers and with a radius >5" for cable bundles for series 6 and 59 cables and for video cables. The cables in bundles must be neatly dressed. All cables must be secured to provide reliable support. In places exposed to damage (chafing, abrasion, compression, impact), the cables must be adequately protected by placing them in concealed channels or by covering them with adequately rated raceways. All RF and IF cables should be of quad-shield design and approved as headend cables. All baseband video cables must be of video type.

All unused ports of all combiners and splitters must be terminated.

SUMMARY

The described above combining /splitting network for headends and hubs is an example of the network design for a particular HFC network architecture. However, the concepts presented above are universal to many alternative HFC architectures addressing different market sizes and network configurations. The main concepts are:

- a modular design with separate modules for broadcsat signals, local programming signals, targeted service delivery signals, and reverse signals;
- setting clear objectives for each module and optimizing the module design to meet these objectives;
- clear specifications for module interface parameters for design compatibility.

The complexity of the modern HFC networks and the variety of service types (broadcsat, local, targeted) require a rigorous approach the headend and hub facility design, including their RF combining/splitting networks. The paper presents TCI's experience and choices with an attempt to make them as universal as possible.

ACKNOWLEDGMENT

The author wishes to thank Ed Cooper from Integration Technologies, Jim Farmer and Blair Schodowski from ESP, Patrick Butler from TCI Engineering, and TCI headend staff in Hartford and Fremont for contribution in preparing internal TCI documents that became a basis for this paper.

Multiple Conditional Access Systems

Michael Adams
Sr. Project Engineer, Advanced Engineering
Time Warner Cable

Tony Wasilewski
Chief Scientist, Software Systems
Scientific Atlanta

Abstract

There are a number of well-developed standards currently being used to deliver digital video over cable systems in the area of modulation, forward error correction, transport and compression. However, encryption and conditional access systems are typically, by their very nature, highly proprietary systems. This presents a road-block to the support of multi-vendor set-tops in cable systems. One possible solution is the adoption of a common, standard, service encryptor at the lowest level combined with multiple conditional access systems that provide key distribution and control functions.
This paper will describe the technical challenges in building multiple-conditional access systems and argue that, in practice, other standards are required to support cost effective deployment of multi-vendor set-tops.

Introduction

Conditional Access (CA) systems, by their very nature, tend to be proprietary. However, Cable Operators would like to have the choice of set-top terminals from many different suppliers. One approach would be to settle on a single system design and have all manufacturers license that one system. This would have negative implications for all parties concerned, however, because there is little incentive for feature innovations from the alternate suppliers, since they are locked into a sub-licensed design.

The digital age offers the promise of supporting a highly standard, multi-vendor environment. This includes the possibility of having more than one conditional access system at work within the same network simultaneously. Because the industry has focused on and agreed upon the use of standards such as MPEG-2 and DAVIC[1], it is now feasible to finalize agreements that permit complete interworking of products from different suppliers while still reaping the benefits of digital compression.

The MPEG-2 systems layer[2] provides various hooks to support the co-existence of multiple CA systems within the same digital channel. This allows decoders, using different CA systems, to gain access to the same services with no need to simulcast the MPEG-2 payload.

So that the MPEG-2 payload need only be sent once, it must be encrypted with a standard 'service' encryptor. The multiple CA systems then effectively provide different key and entitlement delivery systems. Because only the CA key and entitlement delivery information needs to be simulcast, this adds relatively little overhead. We will show that, in practice, this is less than 1% per CA system.

Definition Of Terms

The following terms will be used in this paper:

- **Conditional Access (CA)** System - the software and other components necessary to provide for selective access or denial of specific services in a network. The CA system is used to establish the means by which subscription or other payments may be collected from users of a network for use of a service. A conditional access system includes mechanisms for payload encryption, secure key delivery, addressed messaging, secure entitlement delivery and appropriate links to administrative gateways or billing systems.
- **Key Delivery** - the mechanism by which various keys are delivered to the set-top terminal in a secure manner (so that the service cannot be pirated).
- **Key Hierarchy** - a key hierarchy is usually defined in a broadcast security system. At the

lowest level is the Control Word which is the key used with the Service Encryptor.

- **Service Encryptor** - the encryption algorithm performed on the MPEG-2 payload bytes. Note that the MPEG-2 transport system and adaptation headers are always sent in the clear.
- **Control Word** (CW) - the key used with the service encryptor to provide confidentiality of the delivered services. It is changed at a rapid rate to increase the security of the content.

First we will describe an example of a single CA system before turning to multi-CA systems

An Example System

Traditionally, CA systems for broadband networks have been intended to protect primarily against signal theft for the benefit of the network operator. With the advent and migration to digital compression and two-way services, security issues have greatly expanded and so has the list of beneficiary parties. For example, content owners, service providers, billing providers, and end users now all have security concerns in addition to network operators. In addition to signal security, examples of these emerging concerns include:

- sensitive or private data accessed and transmitted in cable modem applications
- authenticating service providers in a multi-provider network
- multiple entitlement agents ("gatekeepers") in one decoder
- authenticating messages in forward and reverse directions
- protecting software and application downloading to Home Communication Terminals (HCT's), including virus protection
- two-way services
- shopping services and E-commerce, E-cash
- subscriber identification and digital signature
- subscriber privacy, for example, credit card numbers.

Scientific-Atlanta's PowerKEY System is the broadband industry's first CA system to support both public key and secret key cryptography. PowerKEY's use of public key (RSA) cryptography allows it to address the issues discussed above in a unique way that traditional secret key-only CA systems cannot match.

The requirements of a robust CA system are met by the following PowerKEY system components:

- Stream Encryption & ECM Streamer Module
- Control Suite
- Transaction Encryption Device (TED)
- Service Decryptor Module
- Security Manager
- Home Communication Terminal (HCT) Secure Element

Figure 1 depicts the configuration of these components within a typical system and the interfaces that must be established with other components and subsystems for delivering secure digital services. In this figure, the HCT receives signals from a Broadcast Center - the dotted lines indicate that the transmission could be either over-the-air or on a wired network. An "out-of-band" data path can be activated, which could be, for example, a QPSK or telephone transmission. This path can be used for impulse pay-per-view returns or with a suitable form of modulation, highly interactive services. The LAN Interconnect Device would typically be an IP router.

EMMs may be sent on this "out-of-band" path. In fact, since PowerKEY EMMs may be encapsulated within IP packets, they can be selectively routed to specific Broadcast Centers. This not only conserves EMM broadcast bandwidth, but considerably complicates the business of the "clone pirate", since there may be many broadcast centers. The "pirate" must maintain legitimate HCTs in each of these broadcast centers to enable clone reception.

PowerKEY System Interfaces for Digital Services

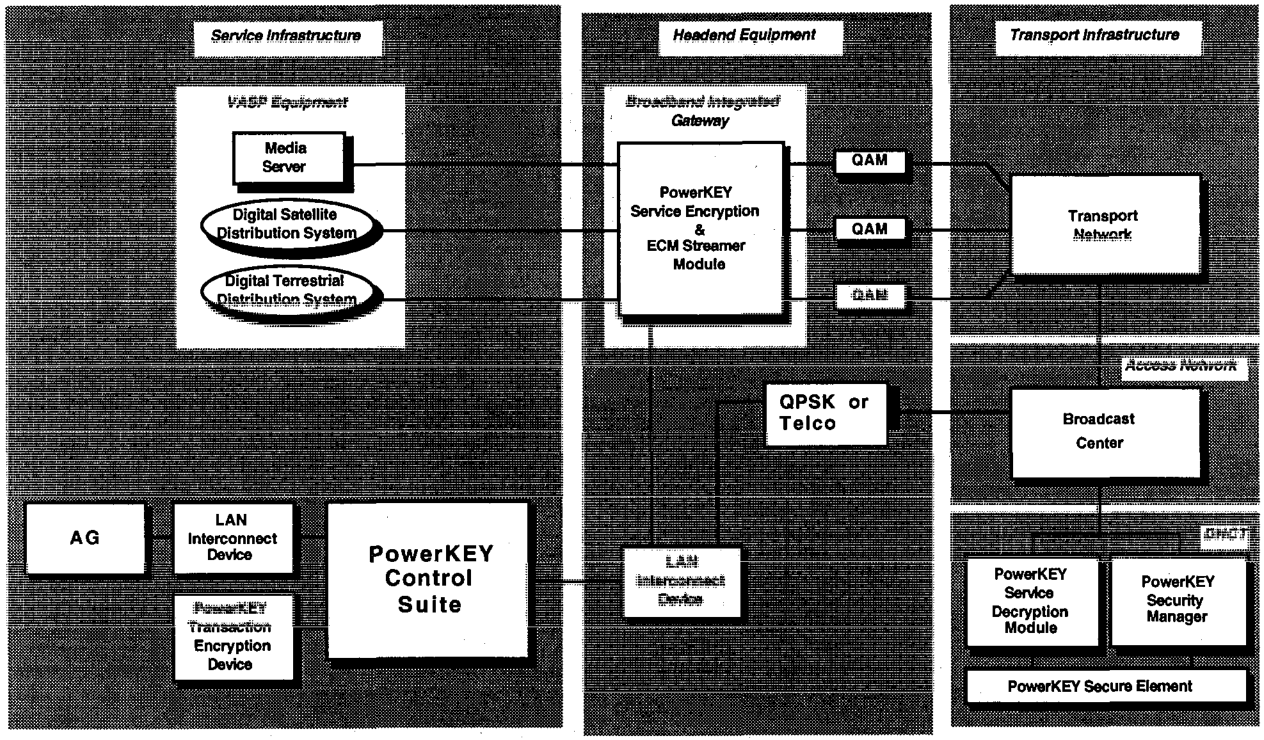


Figure 1. PowerKEY System Interfaces

The PowerKEY CA system employs a multi-level key hierarchy. Control words are fast-changing keys used to encrypt the services (video, audio, data). Mid-level keys called multi-session keys are used to protect the control words so that they can not be discovered in transmission, except by authorized units. The multi-session keys are sent to individual decoders using messages (EMMs) that are encrypted with the RSA public key algorithm. These EMMs are also digitally signed by an Entitlement Authority.

The CableLabs Agreement

In October 1996[3] some major elements of an interoperable digital cable systems specifications were agreed by CableLabs and its members:

- The agreement was based on existing standards (DES encryption, MPEG-2 systems layer).
- The agreement was deliberately defined to be the minimum intersection of multiple CA systems:
  1. The adoption of a standard service encryption algorithm based on DES standards[4,5].
  2. A common control word generation method.
  3. Use of existing features in the MPEG-2 systems layer to allow multiple CA systems to co-exist within a single digital channel.

This agreement represents the final and the most difficult step in long history of standardization. Because the CA system is typically the most feature-rich it significantly differentiates one vendor's product from another. By separating the CA system into two parts (the service encryptor and other components), each vendor is still able to innovate and add features to its CA system without introducing incompatibilities at the service encryptor level.

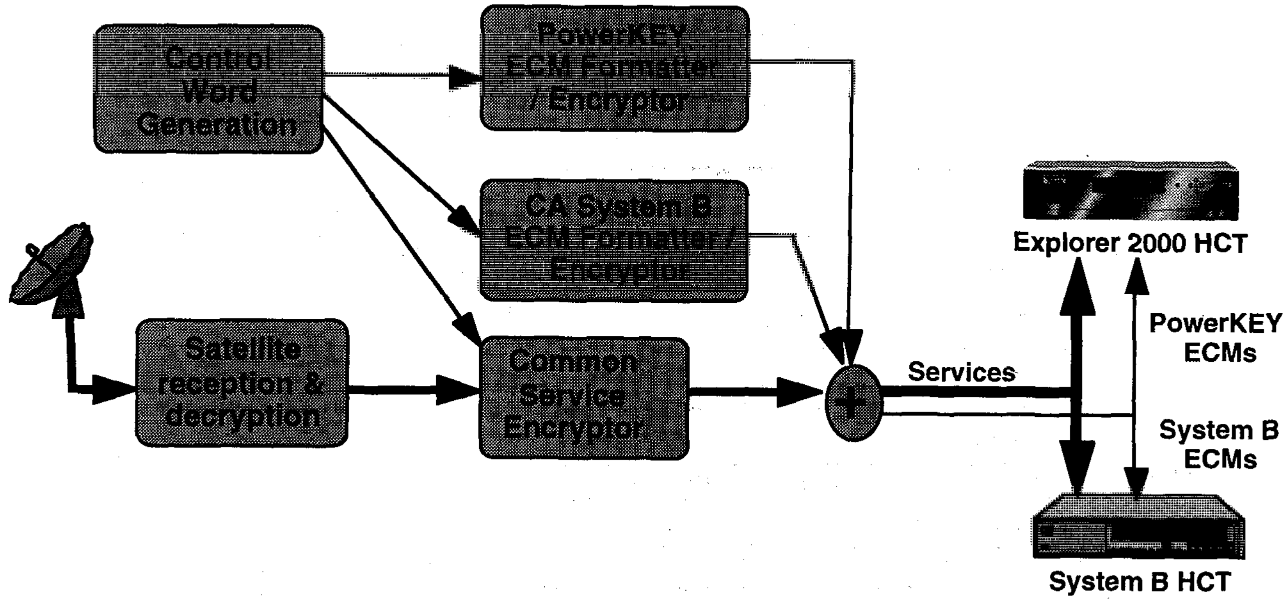


Figure 2. A Dual-CA System

Figure 2 illustrates an example multiple conditional access system. A common control word is used with a common service encryptor to encrypt the MPEG-2 payload. Each of the two conditional access systems independently deliver the control word to the two vendor's Home Communications Terminal (HCT). Each HCT receives and operates only on the ECM stream that it 'understands'.

MPEG-2 (ISO/IEC 13818-1)Transport Stream

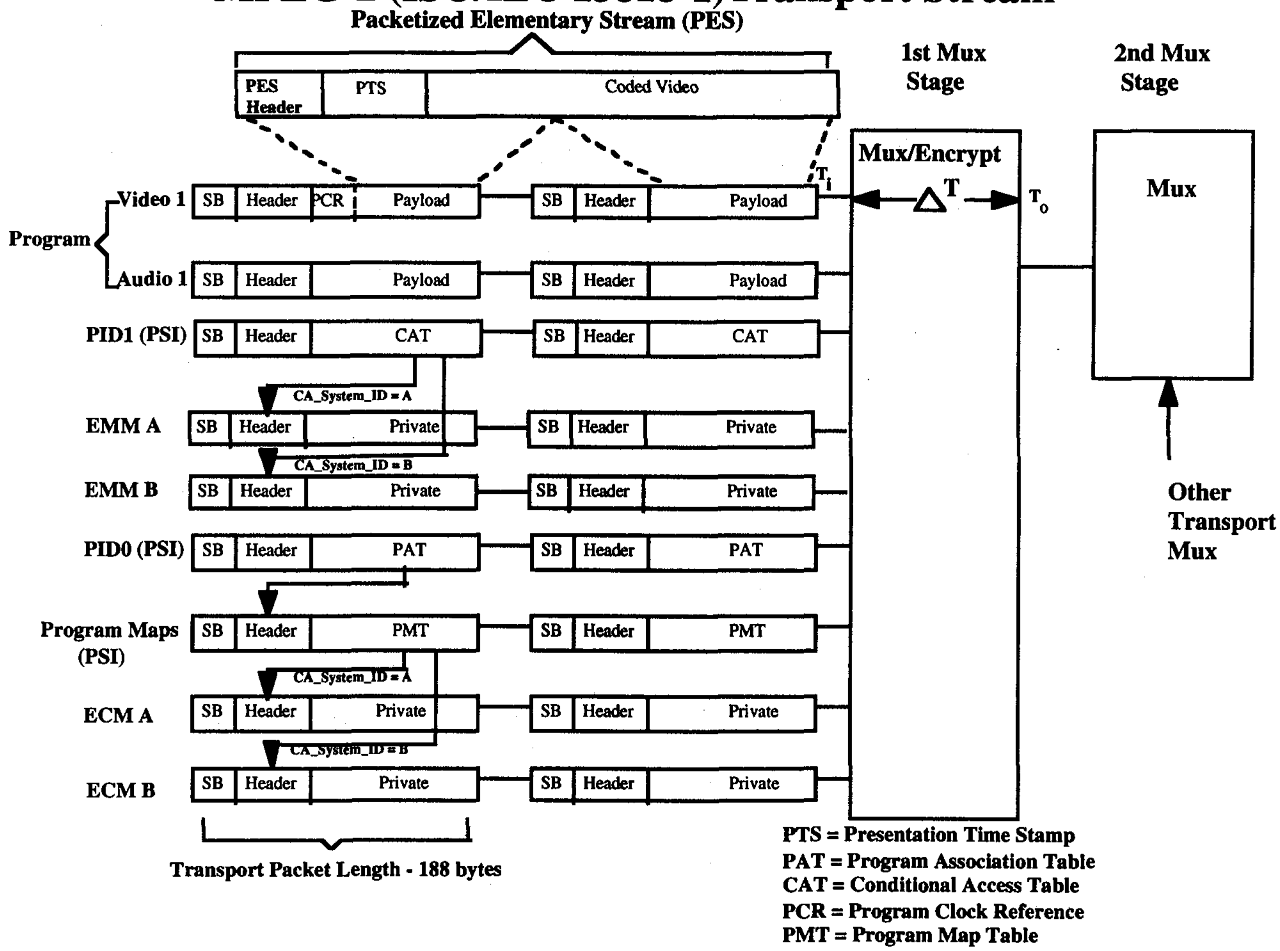


Figure 3. MPEG-2 Systems Layer

Figure 3 illustrates how the MPEG systems layer supports multiple conditional access systems.

- The Conditional Access Table (CAT) provides pointers to multiple Entitlement Management Message (EMM) streams.
- The Program Map Table (PMT) provides descriptors to multiple Entitlement Control Message (ECM) streams.

Increased Overhead of Dual-CA System

What is the overhead of operating a dual-CA system? If we assume 100 Kbps for the additional ECM stream this amounts to less than 0.4% of the digital channel. Taking an estimate of 100 Kbps for the additional EMMs < 0.4% this also amounts to less than 0.4% of the digital channel. (Note that in a cable system EMMs are typically delivered in an out-of-band, QPSK channel.) Therefore the total overhead is less than 0.8%.

In any case, it is the exception rather than the rule, that both CA systems would be active in a single system at the same time. The benefits of a multiple CA strategy of second sourcing, CA system evolution and CA replacement are more important than placing two set-tops, which require different CA systems, side-by-side in the a cable system.

Future Work

Much work still remains to be done to develop multiple conditional access systems:

1. CA system interworking - there are many problems to solve:

- Program schedules and program guide information need to be synchronized. Program guide information must be delivered in a form that all HCTs can access.
- Billing interfaces must become more standard so that the two conditional access systems can be supported by a single billing system.

2. Security Extensions - a standard API is needed to support secure applications, for example, secure WEB transactions, electronic commerce, games, etc.

Summary

The framework to implement multi-CA systems within was initially established by the MPEG-2 systems layer and has been further defined within the CableLabs agreement. However, there is still much work that remains to be done.

There is only a minimal and reasonable overhead to operate a dual-CA system. This represents less than 1% in a cable system.

The conditional access system significantly differentiates one vendor's product from another. By separating the CA system into two parts, each vendor is still able to innovate and add features to its CA system without introducing incompatibilities at the service encryptor level. Therefore, multiple conditional access allows interworking without reducing CA to the lowest-common denominator.

REFERENCES

[1] Digital Audio Visual Council (1996), DAVIC 1.1 Specification Part 8: Lower Layer Protocols and Physical Interfaces (draft as of September, 1996).
[2] ISO/IEC 13818-1 (1994), Information Technology — Generic Coding of Moving Pictures and Associated Audio: Systems.
[3] "Cable Industry Agrees on Key Elements of Digital Systems Specification", CableLabs Press Release, October 3, 1996.
[4] Data Encryption Standard (DES), NIST FIPS PUB 46-2, January 1988.
[5] DES Modes of Operation, NIST FIPS PUB 81, December 1980.

NVOD
THE PREMISE BEHIND THE PROMISE

Dom Stasi
Request Television

ABSTRACT

This paper will detail one programmer's approach to digital, multichannel origination and transmission, and the technical implications of such highly processed signals to CATV reception and distribution.

As a preparation to that discussion, the text will include an overview of digital compression and coding philosophies. 4:2:2, 4:2:0, and 2:1:0 video sampling will be compared. Coding, storage, compression, filtering, decoding and distribution of compressed digital video signals will be described. Serial, and parallel studio interface standards will be explained. And finally, MPEG-1, MPEG-2, distribution protocols will be discussed with an emphasis on operational effects.

INTRODUCTION

Overview

With 1997, the much anticipated implementation of digital video from source to sink will be a CATV reality. Digital compression will bring the capability to distribute hundreds of video channels to those cable homes equipped to decode them.

The hype surrounding this transition emphasizes that the new digital signals will pass relentlessly through your existing physical plant with few, if any, modifications to traditional cable architecture and equipment. As such, programmers are readying their origination facilities, while cable operators await the windfall of impervious signals.

But despite that the transition to digital modulation is being greeted as an industry-wide opportunity to increase revenue and compete more effectively, to those of us responsible for distributing these signals, the reality is, of course, more complex.

For example, in origination, programming will not simply be digitally encoded, but often stored and distributed via non-linear devices such as virtual recorders or file servers to the exclusion of traditional videotape. In distribution, NTSC analog video signals will be supplanted by MPEG-2 compressed digital packets. Ku band satellites will be utilized alongside traditional C-band for distribution to headends. In reception, macroblocks will replace sparklies in a compromised satellite signal. Phase linearity will supplant noise temperature as the most important parameter of your LNB. Rain will cause received signals to fade as much as -13dB, routinely. Even the ubiquitous channel modulator will be nudged aside to make room for the digital upconvertor. While at your subscribers' homes, concatenation error will appear in poorly processed video signals, and entropy-loss will cause as many subscriber complaints as triple beat.

If it is to compete effectively with recent entrants into the video to the home market such as telco and the myriad DBS entities, the CATV industry understands that it must re-evaluate its position as the multichannel video provider of choice. It has done this in the past with great success. And, if it is to maintain its dominant position, CATV will simply continue to do what it has traditionally done: deliver more channels of high-quality video programming than would be available through other reasonable means.

Of course "... more channels of high quality video programming" is a subjective phrase at best. Filling this wealth of new channels with appropriate fare posses a challenge to service providers that extends across artistic, commercial and technical lines. But one thing seems clear, high-quality-video-programming will not mean - as we might be led to believe - video programs whose audio consists primarily of little beeps and booms, and whose pictures get by on 50 pixel resolution. If we're going to rise to the competitive challenge, we'll have to do a lot better than that.

For hard won experience has shown us that high quality video programming means what it has always meant: traditional entertainment fare. So, for purposes of this paper and to better understand the network it describes, entertainment fare means programming born of major effort. The stuff requiring scripts, talent, crews, production and distribution: comedy, drama, documentary and athletic endeavor - things that are hard to create on demand, are costly, must be delivered in real-time without substantive impairments, and that eat bandwidth. Additionally, virtually all of this high-end video product is accessible by our subscribers elsewhere. When a subscriber decides to purchase a pay-per-view movie or event, he is saying in effect that he will forego seeing this product live or larger than life on a 50 foot screen in surround sound. Therefore, what we deliver better be as good as we can make it.

This is not reason to lament. Entertainment programming and the bandwidth it devours have served CATV well. Further, given that CATV bandwidth is limited only by our collective imaginations and not by skin effect or government regulation, we in the CATV industry enjoy a substantial advantage over aspirants to our market. So, however fondly we might wish to emulate other industries, programming-and-bandwidth remain CATV's stock-in-trade. System operators provide the latter; programmers the former. And, if this year is remembered for anything at all, it will be remembered as that in which no effort was spared to increase the effective bandwidth of CATV systems through rebuilds, upgrades, and of course, digital compression. What remains to be determined is how well we *programmers* will respond to the need to fill this new space.

This paper will outline one programmer's efforts and their effect on CATV physical plant performance and subscriber/viewer perception. Given that it is the most extensive and technically prescient effort launched in 1997, it covers most of the origination and distribution permutations that might be found to vary from the conventional.

Given that this session is aimed at operating system personnel, and in order that its content be of relevant interest to a broader range of readers, I have included a review of those aspects of coding and compression that to date have been the domain of manufacturers and program networks.

Multichannel PPV

On January 1, 1997, Request Television, the ten year old pay per view network, began distribution of some 40 channels of high-quality feature film and original production sports and variety video programming to cable systems nationwide. In a conventional fm analog modulation scheme, this level of output would require about 300 videotape players feeding signals adequate to occupy 1440MHz of C-band satellite bandwidth.

Of course, dedication of such resources would be economically unfeasible at best. So, in order to accomplish its objective, Request would find it necessary to radically alter and enhance its networks' origination and distribution architecture.

An early adapter of video encoding and compression, Request has been illuminating two C-band transponders with five video signals since early 1993. More recently, with the availability of MPEG-2 compression, it became viable to encode, process, store and distribute an additional 35 video channels in a wholly non-linear, tapeless environment, and to transmit those incremental channels to five Ku-band transponders representing 120mhz of satellite bandwidth.

An evaluation of the operational specifics follows. All of the concepts considered in this paper are currently in operation in the Request TV multichannel distribution architecture.

DIGITAL VIDEO ENCODING

Fundamentals

Television signals nearly always begin and end as analog entities. Digital encoding of these signals at some point or points along their journey from the image target of a telecine or studio camera to the CRT of your subscribers' home TV receivers does nothing to alter that. Further, digital signals require far more bandwidth than do analog signals in order to convey comparable information rates. These are substantive liabilities. Their perception is compounded by a subscriber base that has been conditioned to expect *improvements* to picture quality with the introduction of digital signals. Yet, in reality even the most conservatively coded digital television frame will be no better than the sum of its native analog impairments. Digital encoding can do nothing to *improve* a signal.

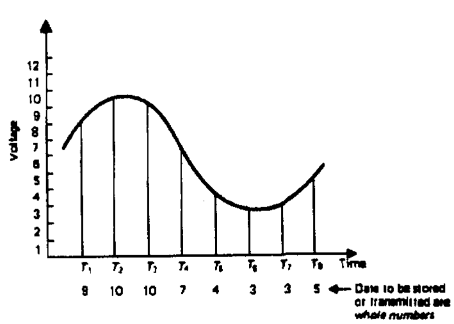
What digital encoding *can* do is limit a signal's susceptibility to further degradation. Efficient digital processing will allow signals to be produced and stored on non-linear media, be copied and post-produced with impunity to generational degradation, and be randomly accessed for distribution. Digital encoding can also virtually eliminate contributions from many transmission channel impairments, impairments that would visibly and perhaps objectionably affect many analog signals of comparable strength. Further, a correctly coded non-linear signal will improve receiver transfer function dramatically over f.m. These are substantive *assets*. They will enable us, if we implement them intelligently, to effectively halt our signal's degradation. If we maintain a disciplined approach to the design of our encoding and distribution systems, we will, for the first time, have at our disposal the wherewithal to limit a signal's impairments to those peculiar to its creation, those inherent to its analog form when presented to the digital encoder and those contributed by the analog medium on which it is exhibited. Simply stated, digital processing will allow programmers to deliver video signals to the CATV headend in a condition similar to origination quality. Further, it will allow a cable operator to pass that quality through to the terminals of his digital subscribers' television receivers at cascades well beyond his physical plant's analog design limits. But these extraordinary performance advantages come with a price: bandwidth. And it is how we employ that most valuable of all transmission and distribution system resources that will determine whether the implementation of digital signals improves our product, or simply imposes another, more objectionable, set of signal impairments.

Sampling & Bandwidth

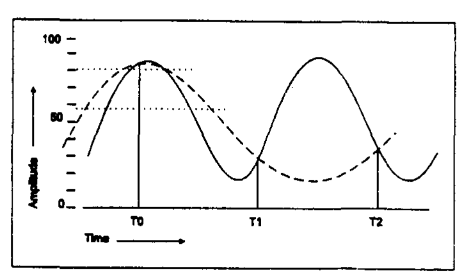
A review of sampling theory will remind us that if an analog signal is to be digitally encoded and reproduced accurately, it must be sampled at a rate not less than twice its highest analog-baseband frequency (f_b). This is the well known Nyquist limit. Therefor, as was stated earlier, digital encoding imposes a bandwidth penalty equal to NLT 2X that of its analog counterpart. However, in practice, to counteract effects such as aliasing, sampling is usually carried out at several times the baseband frequency (fig-1).

Sampling is merely a mechanism to convert signals from the continuous time and voltage domain characteristic of analog representations, to the periodic. As such, sampling signals are clocked at regular intervals, and their amplitude assigned an integer value. If we know the sampling frequency (f_s) and are able to read the sampled values for each point in time, we can transmit and receive periodic representations of any signal as a series of bits, and upon reception restore that sampled signal to its original form. This requires that sampling signals be pulses of limited duration, occurring at discrete frequencies, thus yielding the channel's fundamental data rate.

Figure-1 shows an analog wave sampled at $8f_b$. For interval T-1, the analog-to-digital converter would record an integer value of 8. For period T-3, the instantaneous recorded value would be approximately 10, and so on. Upon recovery, this coarse integration would yield a stepped-function waveform. In practice, however, received signals are restored to their original form through filtering. The higher the sampling rate, the more accurately the reproduced waveform will resemble its original counterpart.



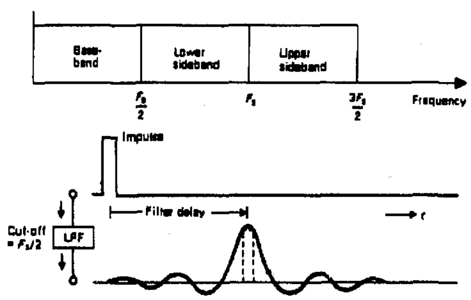
{fig-1}



{fig-2}

Conversely, when a signal is sampled at too low a rate, there are simply too few samples to accurately reconstruct the original analog waveform (dashed line, fig-2).

Therefor, when encoding video whose highest baseband frequency is (for example) 4.2MHz, the sampling frequency, and thus the initial data rate, must be at least 8.4Mbps. A simple interpretation of Carson's bandwidth rule reveals that a channel must be at least twice that frequency, yielding a minimum channel bandwidth in excess of 16MHz (fig-3).

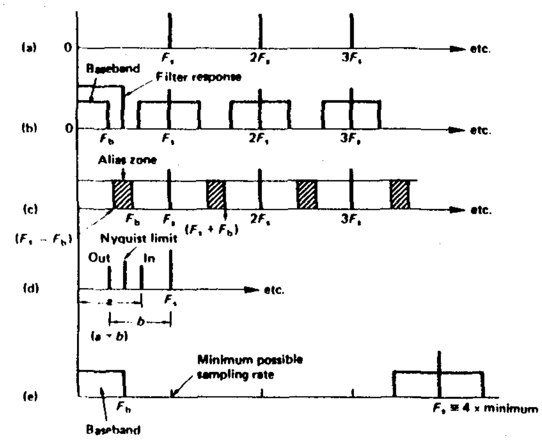


{fig-3}

Again, in reception, the periodic data signal can be reconstructed and returned to the continuous time domain by passing it through a low-pass filter. Practical filters, of course, have imperfect forms and roll off according to a slope rather than exhibiting brick-wall cut-off response. In the sampling process, sum and difference products are generated across a range equal to the sampling frequency plus the highest baseband frequency sampled and the sampling frequency minus the highest baseband frequency:

$$[(f_s + f_b)+(f_s - f_b)]$$

Consider, then, how when an f_s equal to $2f_b$ is heterodyned with that baseband, its highest frequency excursions will produce difference products that fall within the reconstruction filter's passband (fig-4). These difference signals will be reproduced as aliasing artifacts, visible in the reproduced analog video.

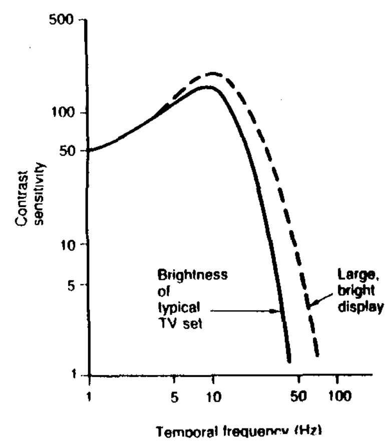


{fig-4}

A commonly employed solution to the problem of aliasing effect is achieved through over sampling. By raising the sampling rate to a frequency well above twice the highest baseband we not only achieve finer integration of the sampled waveform, but place (f_s - f_b) correspondingly outside the reconstruction filter's -3db response. This, however, widens the required channel bandwidth farther still. Add to this the necessity of including overhead bits such as forward error correction, control, and ancillary data to the transmitted digital signal, and both data rate and bandwidth requirements quickly obviate the transmission advantages of digital signals. These disadvantages are overcome through recently developed techniques of digital video compression. However, before embarking on a discussion of compression as applied to Request TV's multichannel origination and distribution model, a discussion of some of the variations of digital video encoding and the specific sampling structures employed is perhaps appropriate.

Component Video Sampling

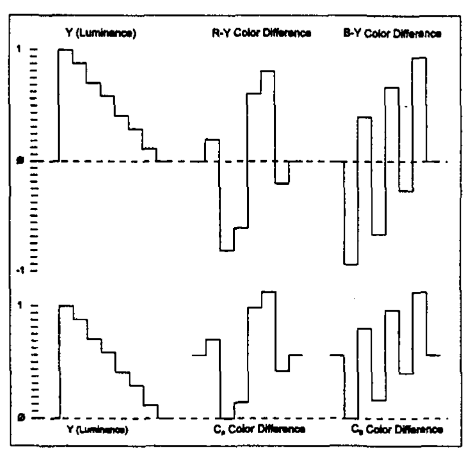
In addition to signaling components, CCIR-601 sets forth the protocol for separating an analog video waveform into its respective luminance (Y) component, which varies directly with a recorded scene's instantaneous brightness, and its two chroma difference signals (R-Y, B-Y) which convey the color information.



{fig 5}

Since the human eye is sensitive to light level changes of about 1% (see fig-5), a television system must accommodate an end-to-end gray-scale of approximately 100:1. If a digital system is to reproduce a visually linear luminance scale that accords with the commonly used IRE (Institute of Radio Engineers) brightness scale, while remaining at least comparable to the analog system's minimum limits, it would seem that it must be capable of quantizing a sampled video signal to 100 discreet values. This is accommodated, in fact exceeded, in professional systems through eight-bit or ten-bit sampling. Eight bit (2^8) sampling yields 256 discreet binary levels, while ten-bit (2^{10}) yields 1024. Both are found in common practice since linear quantization is used in most applications. In Linear quantization each sampling step is of identical size. Results, however, depart from the linear when we consider again that the human eye can detect light changes of 1%. A 1% change at a luminance level of 2 equals 0.02 IRE, while a 1% change at a luminance level of 90 corresponds to 0.9 IRE units.

Each of the component signals (Y, R-Y, B-Y) is sampled at either eight or ten bit levels. Since the color difference signals are bipolar (fig-6), they are either D.C. offset or twos-compliment coded and assigned values which are always positive. To differentiate between encoded color difference components and their bipolar counterparts, color difference signals are called C_r and C_b as opposed to R-Y and B-Y, respectively.

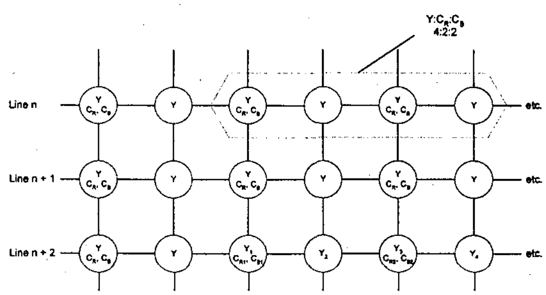


{fig-6}

Luminance is quantized such that values between 0 and 100 IRE units correspond to eight-bit values ranging from 16 to 235, and ten-bit values ranging from 64 to 1023.

Chroma values extend from 16 to 240, and from 64 to 960, respectively.

Given that human visual response to color or hue changes is coarser (less sensitive) than it is to brightness variations of comparable magnitude (see fig-5), the chroma difference signals are sampled at some fraction of the luminance rate. This is commonly 1/2 the luminance sampling rate, but several variations have been introduced and are growing more common as video compression systems proliferate.



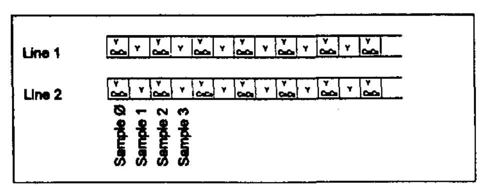
{fig-7}

Conventional CCIR-601 component video encoding yields the sampling profile of fig-7. In this example of a sampled horizontal scan line sequence, we see that Line Segment-1 has a luminance sample (Y) and a pair of chroma samples (C_r , C_b). Segment-2 also has a luminance sample, but no chroma is sampled during its period. The cycle repeats for segments three, four, etc. This cadence is known as 4:2:2 sampling and is the most common sampling pattern encountered in digital video processing: for each 4 luminance samples, 2 each of chroma are taken for a total of eight samples.

Signal processing in the digital domain represents an opportunity to eliminate the vexing interpolation errors common to analog scan conversions. Yet in digital component post-production, a practical difficulty arises from the line (word-length) differences encountered between 525/60 (in actuality the NTSC field rate is 59.94) and 625/50 systems.

Both circumstances have been accommodated in CCIR-601 by choosing a sampling frequency that is an even integer multiple of each format's line period. The H line period of 625/50 systems is 64^{-6} seconds, while that of 525\59.94 = 63.555^{-6} sec. Both of these periods yield whole integer numbers of samples when clocked at multiples of 3.375MHz. Applied in the 4:2:2 sampling protocol , this rate is well above Nyquist, yielding the sampling pattern and frequencies illustrated in fig 7.

Exactly 858 luminance samples are taken in per each 63.555 microsecond period and 864 per 64 microsecond line in the 625/50 system. Chroma is sampled at 1/2 the Y rate. {fig-8}



{fig-8}

These simple sub-multiple patterns contain samples of timing and reference signals such as sync pulses, which are discarded, reducing the number of relevant samples (those representing active picture) to 720 each (Y, C) per active line, hence full CCIR-601 horizontal resolution is equal to 720 pixels.

Composite Video Sampling

In its distribution architecture Request TV makes use of not only component digital video, but composite as well. Described in Society of Motion Picture and Television Engineers (SMPTE) standard 244M, this serial digital interface standard allows for direct digitization of the signal in its composite (intercarrier) form. The distribution standard is produced by sampling a composite baseband at 4fsc, or:

$$4X3.58^{6}\,Hz = 14.3^{6}\,Hz.$$

Samples are taken by D.C. offsetting the bipolar video signal, setting sync-tip (-40IRE) equal to digital 16, and sampling the entire horizontal line including sync pulses, color burst, and vertical serrations. Nine-hundred-ten (910) samples are taken per line.

The technique is profoundly simple, allowing for the use of compact video tape formats such as D-3. As with component coding, the D-3 digital video tape format enjoys virtual impunity to multi-generation degradations, making it attractive for modest-cost post production use. But more important, in a PPV environment, D-3 permits fitting feature length (up to 180 minutes) program material easily onto a single video cassette, thus minimizing VTR machine-count by at least half.

Composite digital video conveys few advantages, and is not well suited to digital compression, thus limiting its practical use as a transport protocol.

Serial Digital Interface Protocol (SDI)

For a digital signal to be more than a post-production convenience, it must lend itself to distribution. Intra-plant distribution between the various Request TV post-production, control room, file server and compression facilities is carried out utilizing both conventional analog, or wherever feasible SMPTE 259M serial digital transport.

The 259M interface protocol was developed by Sony Corp. and allows for full bandwidth CCIR-601 digital signals to be time multiplexed at manageable data rates. The Y, $C_{r,}$, C_b samples are integrated serially at 27MHz. For ten bit serial transmission this results in a 270Mbit data stream, easily transported over coaxial cable for intra-facility distribution.

It is this 270 megabit data stream that represents but a single digital video signal at the input to an encoder.

However, given that this paper presumes to discuss the practice of distributing several digital video signals within the bandwidth normally allocated to but a single 27MHz transponder, or 6 MHz analog channel, some discussion of digital compression basics is in order.

DIGITAL VIDEO COMPRESSION

Fundamental Precepts

Digital compression does not supplant sampling theory. In order to accurately encode an analog signal that signal must initially be sampled at the Nyquist rate or higher. This remains independent of the level of compression employed. In fact, accurate sampling at a high data rate will ultimately aid in optimizing compression.

What digital compression mechanisms exploit in order to reduce data rates is a combination of signal redundancy and the anomalies of human visual perception. In most encoded intelligence the sampling rate, thus the coding bit rate, is constant, while the information rate of the signal it is sampling, varies. The difference between sampling bit rate and information rate is a signal's redundancy. To an extent, redundancy is predictable. The difference between a signal's information content and that predictability is entropy. Compression signals isolate these properties of entropy and redundancy, eliminating the latter from the coding process to the extent possible.

Redundancy in a television signal on a frame by frame basis is substantial. It consists at the very least of highly predictable repetitive signaling components such as horizontal and vertical synchronizing signals, equalization pulses, and vertical serrations. Redundancy also exists in interframe scene similarities characteristic of the 60Hz field scanning rate of NTSC television. Statistically there exists a great deal of scene content correlation between successive video frames. If signaling and scene redundancies can be isolated, they needn't be coded or transmitted every frame, instead, redundancies need theoretically to be transmitted but once and memorized at the decoder for inclusion to subsequent received frames.

Further, the 60Hz field rate itself is a deliberate exploitation of the human eye's light retention characteristics. The eye does not respond instantly to light stimulus, but requires between 150 and 300 milliseconds to transfer an image along the visual cortex. The brain then retains that image for about 100 milliseconds. This combination of latency and persistence is exploited in all television signals, whether analog or digital, full bandwidth or compressed. As fig-5 illustrates, the human visual system's response to rapid changes in contrast, i.e. light level, falls off sharply above 10 Hz, with virtually zero response to flicker above 50Hz. Thus, in the frequency domain, a 60 field per second signal, interlaced at 2:1, is sampled at a rate well above the critical flicker frequency of the human eye, is displayed at a repetition rate well within its 100ms retention period, and is reproduced within a bandwidth merely half that required for a comparable *information* rate in a progressively scanned system.

This level of sampling results in acute redundancy between the majority of subsequent frames in the analog NTSC signal. When subjected to compression, such redundancy yields an advantage that NTSC developers never intended, but which remains an advantage nonetheless.

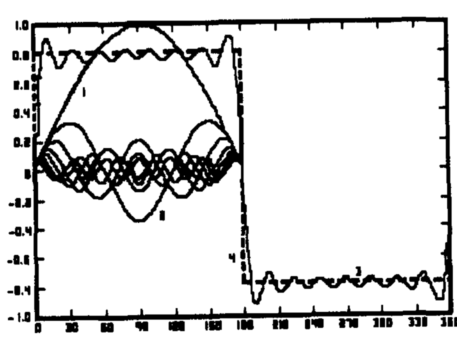
Transform Coding

One way of isolating redundancy, while performing bandwidth reduction occurs through the use of transform coding. In the context discussed here, a transform is a mathematical process for examining data in one domain (in the case of television signals that is either the continuous or periodic time domain), and expressing it in another. To distribute signals in the continuous time domain of analog, requires the bandwidth discussed previously. As such it must be assumed that a signal will traverse all frequencies within its range over time. Digital encoding merely considers such a signal in a *periodic* time domain. Therefor a video channel assumes, at the very least, the entire spectrum from d.c. to 4.2MHz.

Since under no practical circumstances do all frequencies within a given band exist simultaneously, it follows, that if intelligence can be transformed to the frequency domain, and examined periodically, only a code representing the instantaneous frequency of the sampled wave need be transmitted. A substantive coding gain is achieved through such transformation.

In sampled systems, the frequency domain is entered through the Fourier transform. Fourier analysis makes practical use of our understanding that any periodic waveform can be reproduced by adding together some arbitrary number of sinusoids, provided those sine waves are each of a specific instantaneous frequency and phase and are combined in the exact order required to produce our sampled waveform.

While this is of limited practical value, the compliment can be accomplished routinely (fig-9).



{fig-9}

Through Fourier analysis we can examine a specific *existing* waveform, and derive its instantaneous frequency and phase components. If we do this periodically and frequently we can derive enough frequency domain information to reproduce a likeness of that waveform by recombining those myriad samples in an inverse transform. Unlike the continuous time domain of analog, in digital systems the waveform is represented periodically. Through Fourier analysis this produces a finite, or discrete number of samples, resulting in a discrete number of sampled freqüencies. In sampled systems, this type of transform application is known as DFT, or Discrete Fourier Transform, and it is the DFT which forms the basis for data reduction through transform coding as implemented throughout Request Television's compression ensemble.

In order to reconstruct a periodic wave from the linear recombination of sampled frequencies, its instantaneous amplitude, frequency, and *phase* must be known. The DFT performs a spectrum analysis of the sampled input waveform by searching separately for each fundamental frequency each sample contains. This is accomplished by comparing each sample to a series of sine waves of known frequency and phase called basis functions. The instantaneous phase is determined by a wave series, shifted 90^0 to the first and of identical frequency. Sine values are canceled leaving additive cosine components relating to instantaneous amplitude, frequency and phase of the sampled wave. Known as the DCT or Discrete Cosine Transform, this form of sampling is the data-reduction principle upon which most current compression systems are built. Again, transforming from the time to the frequency domain, allows the generation of numerical coefficients representing only the instantaneous frequency and phase characteristics of a sampled wave. Frequencies not present are not transmitted.

It follows further, that if such a periodic signal can be stored and compared with its predecessors, redundancy can be reduced, yielding greater coding gains.

MPEG Coding

Of the MPEG (Moving Picture Experts Group) compression standards that exist today, MPEG-2 is that which will prevail and which has found its way into what is fast becoming widespread commercial network distribution application. However, in order to discuss MPEG-2, those aspects of its genealogy specific to MPEG-1 should be mentioned.

The isolation of redundancy between consecutive frames is the greatest contributing factor in the reduction of data rates when endeavoring to compress full motion video material. The MPEG compression techniques make optimal use of this powerful process. Additionally, MPEG signals are packetized, allowing switched distribution through asynchronous transport protocols and the manipulation of encoded video as digital files.

MPEG-1 is a compression standard developed to produce full motion video at 1.5Mbps. It does so by subsampling which results in noticeable loss of entropy (that information remaining in a video scene or digital file after all redundancy is removed) on fast motion or highly detailed scenes. Additionally, MPEG-1 operates at half the vertical and horizontal resolution of full CCIR-601.

Fig-7 recalls that in standard 4:2:2 sampling the chroma is sampled at 1/2 the H resolution of luminance, while vertical resolution is unchanged. MPEG-1 subsamples all components, ignores alternate fields, and produces a 2:1:0 sampling lattice.

In the context of a discussion of PPV or premium CATV network distribution, what is important to understand about MPEG-1 is its contribution to the evolution of its successor standard, MPEG-2. An examination of the sampling lattice reveals, however, that vertical resolution is the same for both Y and C components. This is exploited in MPEG coding.

MPEG coding reduces redundancy through the use of motion estimation, bi-directional, and both intraframe and interframe coding. MPEG-1 also drastically limits resolution in order to achieve its 1.5Mbps data rate. The protocol is best characterized by its use of I, B, and P frames. These are additional sorts of video frames that allow motion estimation through the comparison of pixel locations between previous and subsequent frames. I frames are complete pictures sent every 10 frames or so, which allow for switching.

B frames are bi-directional predictors which are transmitted and stored at the decoder.

P frames make use of motion vectors and allow the received image to be shifted spatially in reception. All of these features reduce data rate, but add to the memory requirements in the decoder.

Sampled frames are downloaded and buffered in groups that allow for non-sequential processing. Interframe attributes such as motion or redundancy are identified in this manner. The first picture in the group will be intra coded and applied to the DCT. Once transformed and quantized, it is inverse quantized, inverse transformed and buffered for reference. This forms an *I* frame. A *P* frame is then processed by reference to the stored *I* frames and motion information is detected and subtracted from the frame being processed to generate a predicted picture. This frame is transform coded to form a motion compensated, compressed *P* frame which is stored in encoder memory. The encoder now accesses intermediate frames skipped when the *P* frame was created. The encoder calculates forward motion vectors from the locally buffered *I* frame to the *B* picture, and reverse from the *P* frame created immediately prior. Intermediate *B* frames are processed and encoded next, and the process repeats.

In the decoder, the process is synchronized by the recognition of *I* frames, and the inverse transform applied.

MPEG-2 evolved from the precepts of MPEG-1. It is in fact more a set of rules within whose context compression designers may work than it is a rigid standard. But despite that MPEG-2 is a set of rules, observing those rules affords designers and users of the standard an extraordinary range of freedom, not previously available in analog systems.

MPEG-2 proffers a layered coding structure. Such an approach allows for scaleable levels of quality. These are determined by both the programmer and the cable operator. Both encoder and decoder are capable of latitude within the layers of a protocol. Programmers are free to send full resolution, high bit rate data, while in the selection of decoders, the operator is free to choose less complex receivers, his decision governed by his equipment cost and/or the quality level he wishes to purvey to subscribers.

Error concealment, signal to noise ratio, resolution, and entropy propagation are all within the control of MPEG-2 system designers and operators. As such, the Request TV signal assumes many forms and quality levels between production and exhibition.

Contribution quality video is 4:2:2 encoded to 720 pixel resolution, utilizing an 18Mbit data rate. That same signal might reach a digital set top decoder in 4:2:0 format, decoded at 2.5Mbits with a horizontal video resolution of 360 pixels. All fall within the parameters of MPEG-2.

For these reasons, however, when a signal is described as "MPEG-2 compatible," it too often means a great deal less than its purveyors might have us think. MPEG-2 is neither a performance specification, nor an indication of quality. It is more a set of syntactical guidelines, much like those used to make language convey information properly and understandably. And, as with language, there is ample room within its syntax to allow the creation of Hamlet or Disco Duck.

THE NVOD NETWORK

Origination Architecture

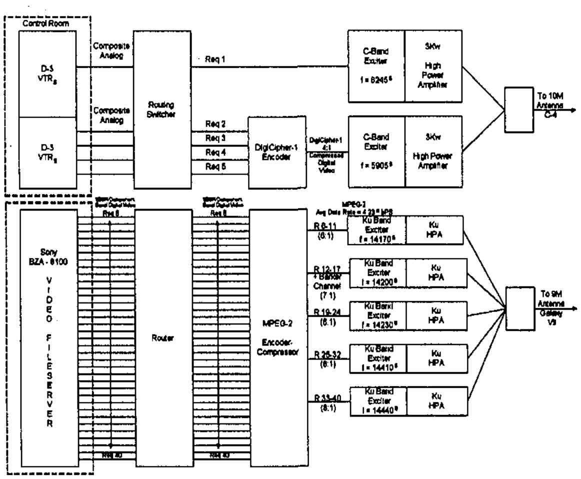
Operating from TCI's National Digital Television Center in Colorado, the Request Television multi-channel NVOD networks comprise some 40 video channels.

Occupying seven satellite transponders, operating in both the C and Ku bands, the network ensemble is designed to produce and distribute a cross-section of signals, accommodating most CATV earth station configurations.

The C-band network group comprises five video channels occupying two 36MHz transponders. Transponder 16, Satcom C-4 remains analog f.m, VideoCipher-II-Plus encoded. Transponder 2, Satcom C-4 is DigiCipher encoded, split-multiplex QPSK modulation, and compressed 4:1.

The Ku band group occupies five 27MHz transponders aboard satellite Galaxy-VII. These are MPEG-2 encoded, DigiCipher encrypted, and QPSK modulated.

Examining the origination architecture (fig-10), it is apparent that expansion to this level of origination followed an evolutionary rather than revolutionary path to full digital distribution. Request-1, like the C-band, analog earth station segment it is designed



{fig-10}

to serve, remains little changed in its 10 year history. Tape based, and functionally analog, R1 employs composite digital encoding only where that scheme might limit post-production generational degradation.

Most incoming source material originates as 24 frame per second (fps) feature length film. Averaging 118 minutes in length, programs are directly digitized and recorded in the D-3 composite digital format. This accommodates either playback to air, or encoding to the file server. If this material is destined for non-linear storage it will be compressed at 2:1 on D-Beta component digital videotape. Component signals are routed via 259M, 270Mbit serial digital interface protocol.

Following the routing of fig-10 it is seen that a common control room is utilized for channels R1 through 5. Though all are D-3 composite encoded, the channels are converted back to analog NTSC prior to input to the router, distributed NTSC throughout the plant, and are split only prior to upconversion at the exciters.

It was determined early in the development of the expanded PPV channels represented by R6 thought 40, that for simplicity of operation, and to accommodate the high degree of title duplication that characterizes NVOD programming, a tape based origination architecture would become unnecessarily complex.

NVOD would differ from conventional pay-per-view in that perhaps 40 incremental movie and event programs (*titles*) would be offered on a monthly basis. Of these, some 30 would be on-air simultaneously. Mostly feature length or longer, programs would be stagger-started. Thus, many titles would require duplication for simultaneous play to air. The initial decision was to make major, popular, and/or premiering titles available for purchase and play at 30 minute intervals. This meant that at any point in time, a feature title would be running on perhaps three networks simultaneously. Only through non-linear storage could a single movie title to be encoded once and randomly accessed for playout. The alternative, 250 video tape players, was an unattractive prospect that would undoubtedly prove both untenable and unreliable in operation.

This of course mandated the use of a file server. However, unlike the sort of server that might occupy a commercial spot playback function in a CATV headend, the type of device necessary for high-end network origination must be of a sort designed to operate within the context of a conventional broadcast automation environment, and more importantly, must be capable of the massive memory requirements needed to store perhaps 160 hours of programming, and to store them at *contribution quality* coding levels.

Contribution Quality Coding

As stated earlier, the scalability of MPEG-2 allows for variable quality levels at appropriate points in a distribution network. This becomes an indispensable attribute given the realities of network processing. In most distribution schemes, digital signals are decoded or transcoded several times along their way to the user. If quality is to be maintained, then, source material must be coded at much higher levels than video at points in the distribution chain closer to the receive sink. Video coded at low bit rates will introduce random-like square response noise to video signals. This noise will be coded, consuming precious bits to the detriment of entropy in subsequent stages if an expected coding gain is to be maintained. Lost entropy cannot be reproduced in later stages. In the case of an underdesigned network, where unrealistic coding gains are assumed, this cycle is repeated causing each subsequent stage to introduce its own artifacts. The aggregate effect is known as concatenation error. Digital artifacts, once coded, will be reproduced in the final picture when decoded at your subscribers' set tops.

Thus, video program material stored as MPEG-2 files at a headend server may be coded at a much lower level than that intended for distribution at the programmer's server. In fact, video material coded at 3.0 Mbps at a headend file server might produce perfectly adequate video quality at the set top, since such coded files need only be robust enough to overcome the normal degradations of a CATV cascade. However, the earlier in the network's hierarchy the signal is coded, the more error correction it must contain, the less entropy loss it can support, and the lower its noise floor must be if it is to produce signals with adequate margin to be coded at 3.0Mbps at the headend. Such are "contribution quality" digital video signals.

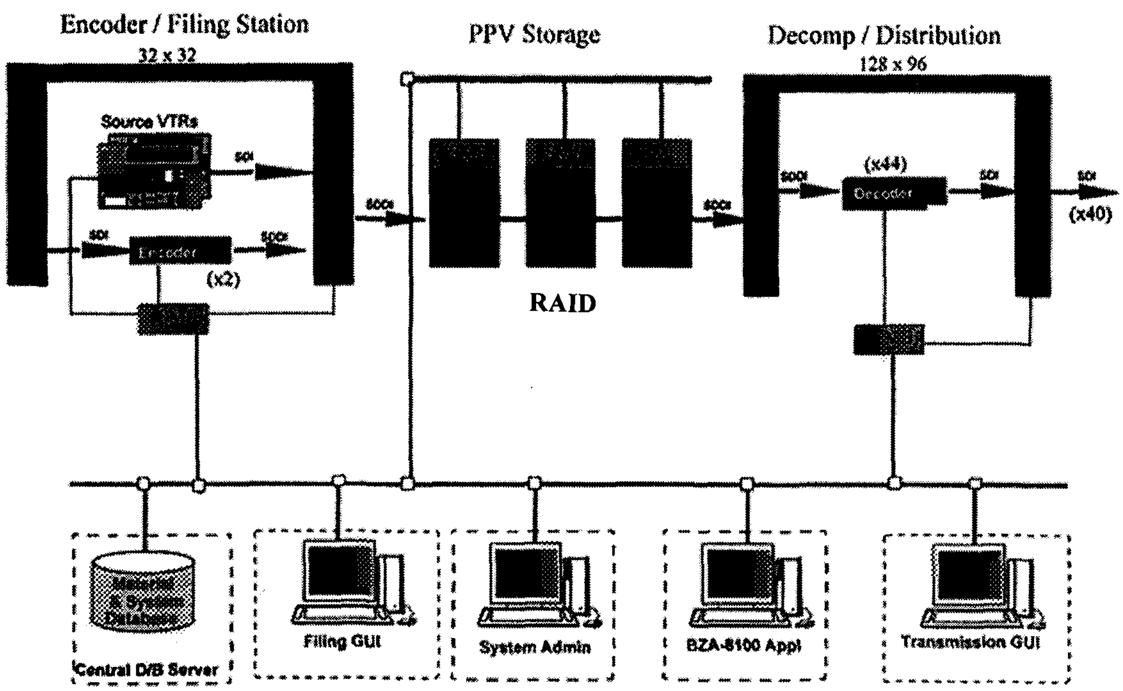
Once a decision is reached regarding the level of entropy loss that can be sustained, the various coding levels (hence memory that must be defined in hardware) that will be accommodated at each point in the distribution chain, can be determined.

Coding some 160 hours of programming at contribution quality bit rates requires massive storage capabilities in an origination server. TCI and Request TV decided on the Sony BZA-8100 video server to fill their storage and quality specifications.

Signals are encoded and stored at main level, main profile, 4:2:2 MPEG-2 coded and distributed in memory for 1:1 redundancy. Coding levels are scaleable at between 5 and 18 megabits depending on application and equipment configuration. Initially, primary video files will be coded and stored at 10 megabits while backup files will be encoded at 5Mbps. The requirement for the nearly 2 terrabits ($1x10^{12}$) of memory needed to support these coding levels was met by the Sony BZA-8100 (fig-11).

and router frames, eliminating single point source(s) of failure.

Encoding into the server represents the first data reduction operation of the chain. Each encoding station will accept analog NTSC, analog component, S-VHS or SDI. For purposes of this discussion, it is here, where SMPTE 259M serial digital video, incoming at 270Mbit is real time encoded



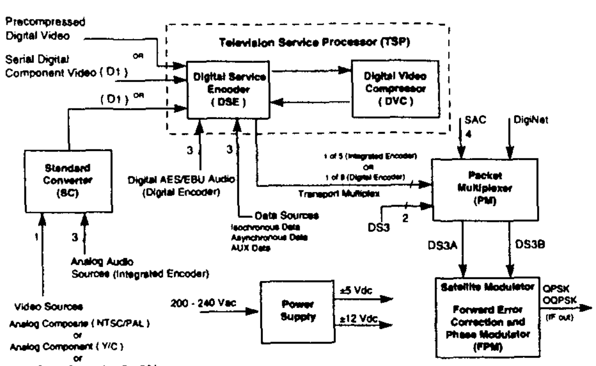
{fig 11}

and compressed to 18 megabit, 4:2:2, "studio" MPEG-2 files. Studio MPEG is the highest "contribution" quality compressed signal we will encounter. In early 1997, Request signals will be encoded at 10 megabits, with redundant (backup) files stored at half that level (5Mbit/4:2:0).

As configured, the distributed, scaleable architecture of the BZA-8100 server will store program material as MPEG files across some twenty RAIDs (redundant array of independent drives) each configured at 80 gigabits (80^9bps). While initially intended as a spot storage system, the scaleable nature of the device allows both short and long form material to be stored into RAID devices with the long form files distributed automatically across multiple disk arrays

It is anticipated that in the backup mode, entropy loss and the inevitable encoding of quantizing square-noise will propagate degraded signals to subsequent network stages. It is believed that video encoded at 5Mbits so early in the chain will not provide adequate margin to cantatenation errors and result in subscriber signals exhibiting objectionable artifacts. Later in 1997, Request signals will be encoded at full studio MPEG-2 (18^6 bps, 4:2:2 sampled). Tests are being conducted to determine the precise contribution level coding that best suits PPV origination. It is projected that a coding level of 9Mbits, 4:2:2 will provide the best compromise, allowing both adequate storage efficiencies to accommodate contribution quality signals, and storage efficiencies appropriate to both primary and backup video files.

The server is capable of 60 discreet outputs. Recalled video files are routed to MPEG-2 decoders and decompressed. Each video output from the decompressors will again be 270Mbit SDI (SMPTE 259M compliant) component digital video. Audio is AES digital encoded at 48Khz and embedded in the serial digital bit stream. This group of signals is routed to the standards converter of the DigiCipher-II encoder (fig-12).



Encoder Functional Elements

{fig 12}

Transport Quality Coding

At the DigiCipher-II encoder multiple channels of serial digital video, each representing one Request video service, are input to the standards converter. MPEG-2 encoded and compressed at the television service processor, these signals now exist at data rates that vary in proportion to the operating bandwidth of the satellite transponder they'll occupy, divided by the number of services multiplexed. For a galaxy-VII Ku band, vertically polarized transponder, signaling data rates and coding budgets would break down as follows:

CODING BUDGET

8:1 Video Compression:	
Satellite Data Rate..................	39.02^6 bps
FEC Data Rate	12.05^6
Information Rate	26.97^6
Messaging & Timing..............	1.62^6
Number of Video Svcs.	8
Data Rate per Svc...................	3.16^6
Audio Data Rate per Svc.	0.40^6
Average Video Data Rate......	**2.77^6**
Required E_b/N_0 Performance.	**3.5dB**
7:1 Video Compression:	
Data Rate per Service	3.62^6 bps
Audio Data Rate......................	0.40^6
Average Video Data Rate......	**3.22^6**
6:1 Video Compression:	
Data Rate per Service	4.225^6 bps
Audio data Rate	0.40^6
Average Video Data Rate	**3.83^6**
5:1 Video Compression:	
Data Rate per Service..............	5.07^6 bps
Audio data Rate	0.4^6
Average Video Data Rate.......	**4.67^6**

Satellite Distribution:

A digital receive signal's required E_b/N_0 of 3.5 compares very favorably to the 9.5dB C/N ratios we're accustomed to dealing with in the analog domain. But what we gain in earth station performance through digital coding and modulation, we lose in required margin above receiver threshold. Shorter wavelengths make Ku frequencies susceptible to atmospheric attenuations (rain fade). Our 50 watt Ku downlinks will vary in receive EIRP as much as -13dB in the presence of heavy rains. This is far greater than anything we've experienced on the CATV C-band earth station network. While it's not common practice, a C-band receive station can be designed with margins above rf threshold of perhaps 2 or 3dB and never experience a discernible atmospheric fade.

Economics and good operating practice dictate that robust digital coding, shorter, Ku, wavelengths and higher EIRP levels should accrue to the advantage of receive station antenna aperture.

Antenna gain varies as the log of frequency. A parabolic receive antenna of given aperture will exhibit substantially higher gain at Ku than a comparable device at C-band. The following equation shows the relationship:

$$G_a = (20_{\log 10} f)+(20_{\log 10} d)+7.5\text{dB}$$

where: Ga = Antenna Gain

f = Frequency in Ghz

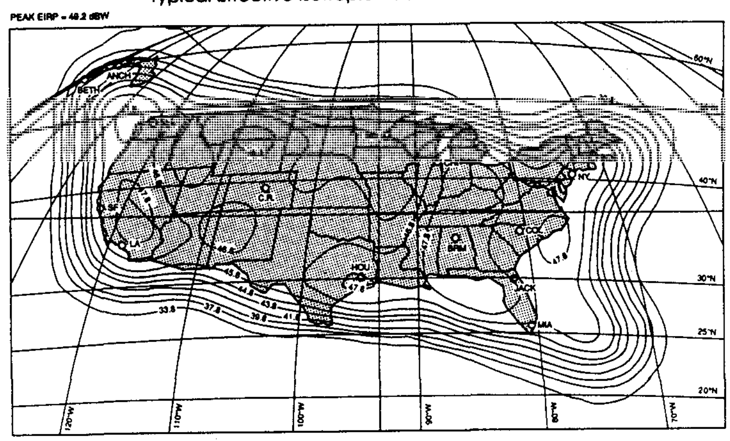
d = Antenna Diameter in Feet

Efficiencies of 55% are assumed

Medium power Ku signals such as those typical of Galaxy-VII will allow the use of antennas as small as 3.6m in most parts of the United States yielding 'four-nines" availability (52 minutes per year).

Galaxy VII (91° West Longitude)

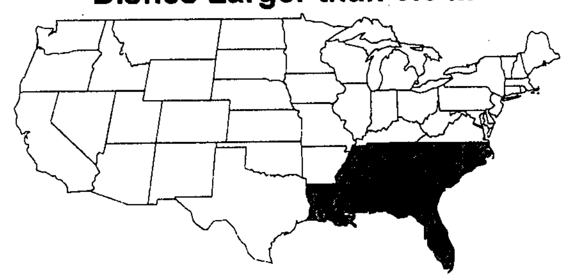
Typical Effective Isotropic Radiated Power (EIRP)



{fig 13}

Other areas, such as those depicted in fig 14 will require larger R.O.s.

Approximate Area Requiring Dishes Larger than 3.6 M



{fig 14}

CONCLUSION

Examining these system architectures and the various data rates employed, it becomes apparent that their determination should be as much a product of economics as of engineering. As was stated at this paper's beginning: bandwidth is precious. But let me add now, that it's precious only to the extent that it can be marketed. As subscribers become sensitized to higher and higher levels of video quality, largely through their exposure to high resolution, progressively-scanned, noise-free computer displays, they make sub-cognitive comparisons to their television picture.

Further, as subscribers migrate toward ever-larger-screen television displays, they grow accustomed to the levels of resolution delivered through well maintained, broadband analog systems. Reducing resolution in a digital system to increase effective bandwidth will yield predictable results, results that are understood intuitively, and are not unlike what can be expected in an analog system. Digital pictures, however, are more complex than that. With MPEG coding we've been handed more latitude to change our product than has ever before been ceded to the technical sector. And our choices are rife with pitfalls that the system designer must understand cognitively as well as intuitively. Nowhere is a lack of such understanding more apparent than in the misguided pursuit of coding gain.

A review of the data rates and levels of coding gain we have discussed in this paper, should leave us all with a sense of control. The feeling, however, is too often one of discomfort as well. For the first time, we, in the CATV community, have at our disposal the wherewithal to exchange bandwidth for content, and do so without apparent loss of quality. We can stuff more and more channels into a given space by forcing higher coding gains, greater entropy losses, and more frequent artifacts to become part of our picture. The too-frequent comparison is made to VHS tape and its consumer popularity. The comparison is grossly invalid.

VHS simply exchanges bandwidth for entropy, rolling off gently and remaining that way.

In the digital domain we not only sacrifice entropy when we sub-sample, we introduce artifacts: arbitrary images not part of the original picture. How many and how severe they appear, is critical.

As stated earlier, the human eye is easily fooled. But only to a point. Response to temporal frequencies, the amount of detail passing the eye per unit time, falls off at roughly the same rate as it does to rapid light changes. Thus, motion artifacts existing at temporal frequencies above 50Hz are easily ignored. However, the eye compensates for this - effectively raising its temporal frequency response - by scanning. The eye scans by simply moving along with a passing image. To illustrate this point, try a simple exercise. Place your open hand eight inches from your eyes. Without moving your eyes, move your hand in a moderately slow waving motion. This creates a high temporal frequency, and the fingers are a blur. Now do the same thing, but follow the movement with your eyes. This lowers the temporal frequency. You see the detail of your hand.

This relates directly to how we perceive motion artifacts in a reproduced digital video picture. When we view a picture on a large screen, and from a moderate viewing distance, the spatial expanse of the image is great enough for our eyes to scan it. Thus, artifacts that might have slipped by on a small studio or control room monitor, will be visible to our large-screen subscribers. Artifacts just below the noticeable level are even more troubling. Artifacts we don't cognitively notice, cause our eyes to fatigue as they try and try to scan, and discern them, especially when experienced over the course of a feature length movie. Eye fatigue causes viewer discomfort. The subscriber recalls the viewing experience as an unpleasant one. Without knowing why, he hesitates before buying again. No marketing survey ever created can identify this potentially devastating circumstance.

We have attempted, therefor, to anticipate what would be perceived as excellent quality signals, while maintaining an aggressive approach to channel loading. The final assessment will rest with our subscribers.

REFERENCES

It is said that copying another's work is plagiarism, while copying the work of several others is called *research.*

The following reference works have been indispensable to the *research* I've committed in developing this paper. I recommend them.

Issues in Advanced Television Technology
by: S. Merrill Weiss

Compression In Video and Audio
by: John Watkinson

Broadband Telecommunications
by: William Grant

Interoperability On The Information Superhighway
by H. Allen Ecker
1994 Technical Papers
NCTA

Television Image Quality
Society of Motion Picture and Television Engineers

Reference Manual for Telecommunications Engineering
by: Roger L. Freeman

NAB Engineering Handbook,
Eighth Edition, 1996
National Association of Broadcasters

OPERATIONAL MANAGEMENT OF DIGITAL CONTENT

Yvette M. Gordon
Time Warner Cable

Abstract

In the development of digital services such as high speed data access, near video on demand, and video on demand, the management of digital content becomes critical. Most content, consisting of compressed images, MPEG video/audio, text, and data will require frequent turnover. This paper addresses hardware and software solutions to enable successful content manipulation in a digital environment.

THE DIGITAL ENVIRONMENT

Content Management Today

Management of content in an analog environment has been centered around the hardware for receiving, encoding, scrambling, amplifying, etc. video signals as depicted in Figure 1. With the addition of digital services, our focus is changing. No longer is content just transmitted, but now it must be stored and transported on demand. We now have to ask questions such as how much disk storage do we need, what are the bandwidth expectations at peak usage hours, and what are the staffing impacts on operational management of the digital hardware and software.

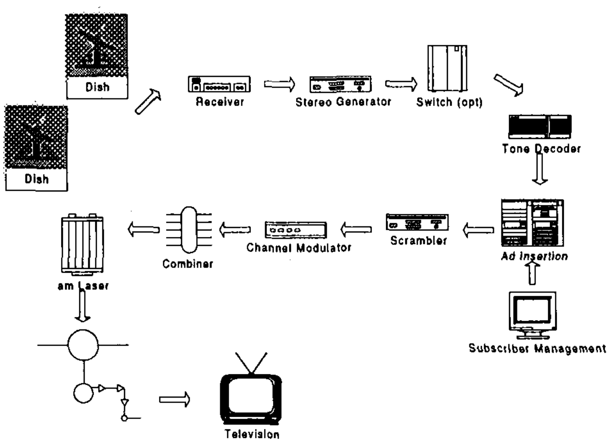


Figure 1. General Headend Flow Diagram

Digital Content Management

As digital services such as high speed data access through cable modems and video on demand are implemented, headend hardware may tend to reflect Figure 2 and Figure 3 respectively. In addition to traditional headend hardware, we are now faced with the management and administration of multimedia servers, routers, and large quantities of disk and/or tape drives.

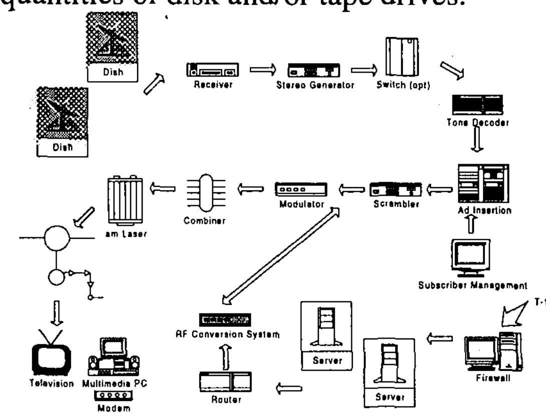


Figure 2. Potential Cable Modem Headend Flow

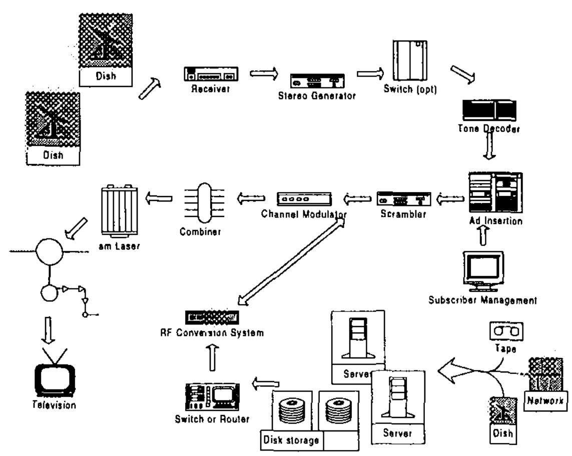


Figure 3. Potential VOD Headend Flow

OPERATIONAL MANAGEMENT

Operational management of digital content is currently very manual; monitoring of content usage, deletion and addition of content onto digital servers must be maintained at that server and, depending on the robustness of the operating system, may require experienced system administrative personnel. The main areas affected: disk space management, failover recovery, bandwidth management, and supporting viewing percentages can be automated in software. We will look at each of these areas and the operational impacts. As software develops to address digital content turnover, these are the key areas that will need to be focused on.

Disk Space Management

With the addition of digital content we are now faced with storage of movies, sporting events, ads, HyperText Markup Language (HTML) files, graphics, and more. Since Video on Demand (VOD) and cable modem services are both on-demand based, we have to store all available content on digital storage media. Considering a full library of movies at approximately 4 gigabytes (1 billion bytes) each, it is easy to imagine a headend consisting mostly of large storage vaults containing terabytes (1,000 billion bytes) of digital storage.

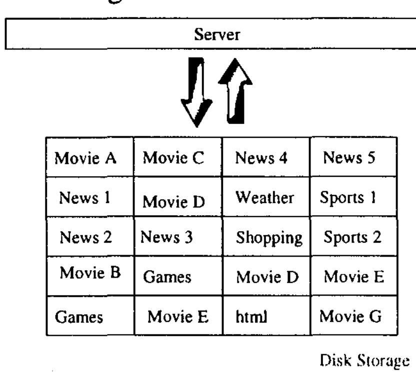


Figure 4. Sample VOD Disk Layout

Figure 4 represents a sample shelf of disks and their stored content whereas Figure 5 represents a similar setup for providing cable modem services. For demonstration purposes, we will analyze the VOD disk layout. What would a typical VOD content turnover look like in one day? A potential schedule may be:

- Remove Movie
- Capture sporting event C at 10am
- Add Movie G from tape
- Update 50 shopping TIFF files and data to display new catalog items
- Add new movie trailers (ads)
- Update NEWS every 1/2 hour

We must also note that digital content turnover is still a very manual process today and each of these items would involve a headend staff member working at a server to complete the tasks. Software solutions are, however, possible and being developed for current trials such as Orlando's Full Service Network.

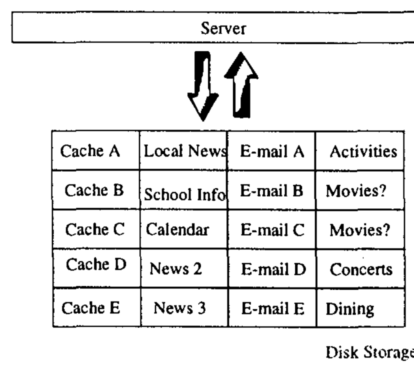


Figure 5. Sample Cable Modem Disk Layout

All networks have many areas for potential throughput bottlenecks. One of these areas is disk and SCSI (Small Computer System Interface) bandwidth. Consider that a single view of a MPEG compressed movie may take up 3.5 Mbps (megabits per second), but the SCSI controller connecting the server to the disks may only support 20 Mbps. Even though a common configuration involves sharing, or striping across many disks, there are still limitations to the number of simultaneous views of a single copy of a movie or other digital content. Depending on the settop or PC application, the result of having no disk bandwidth may be screen blocking or denial of content to the subscriber. What this means to the operational staff is the need to monitor content requests and to make copies of digital content as required by viewing percentages. A recommended threshold may be to copy a digital asset when the bandwidth is 70% used. Note that copying of content also uses bandwidth, as is addressed in the bandwidth section of this paper.

Considering the addition of monitoring and copying, the daily operational schedule may now look like:

- Remove Movie
- Capture sporting event C at 10am
- Add Movie G from tape
- Update 50 shopping TIFF files and data to display new catalog items
- Add new movie trailers (ads)
- Monitor all viewing of content 24 hours a day
- Copy movies whenever viewing percentage exceeds 70% of digital bandwidth

We can now see that the operational impact of digital content management could be significant; however, if the digital system is designed with failover and smooth error recovery in mind, all of the above tasks can be automated in software. More about failover will be addressed in the failover section of this paper.

Hierarchical Storage Management

Although the cost of digital memory and disk drives has been exhibiting typical "Moore's Law" behavior, the cost of storing terabytes of digital content on this media is not yet cost effective. However, access times to other media devices, such as tape drives, do not lend themselves to true video on demand. One potential solution: Hierarchical Storage Management (HSM). HSM has traditionally been designed to allow NVOD or access to digital data on a "wait" basis as the data is moved from tape drives to on-line media.

Assuming that a 90 minute MPEG compressed movie requires 4 gigabytes of storage space and that a suggested file size maximum is 500 megabytes, a digital movie would consist of approximately 8 files representing 11 minutes each. As opposed to storing all 8 files on expensive, fast-access storage, Hierarchical Storage Management would also allow us to store only the first 11 minute file on disk and stream the remaining files as the first file is playing. The only disadvantage of this would be the lack of fast-forward capability for some period of time, until the complete title can be moved to disk. To maximize space with this storage method, the streaming software could delete viewed files from disk; therefore using a maximum of 1.5 gigabytes of disk storage per movie as displayed in Figure 6.

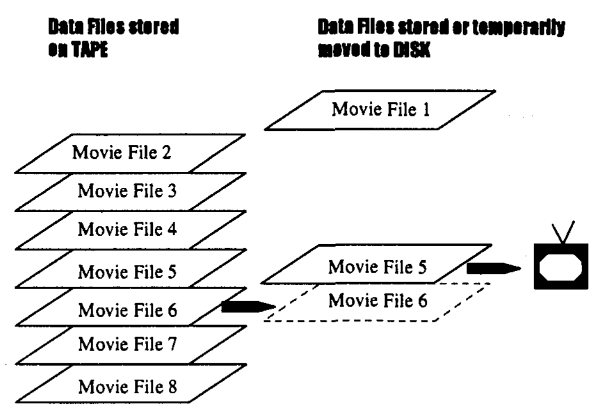


Figure 6. Hierarchical Storage Management Implementation

Network Management Systems

Considering the quantity and complexity of hardware and software to support digital services, monitoring can quickly become overwhelming. Network Management Systems are designed to:

a) reduce downtime by ensuring all components are monitored constantly and consistently thus allowing problems to be addressed immediately and,

b) reduce cost by automating the monitoring of network components, and
c) reduce mean time to repair.

Simple Network Management Protocol (SNMP) was designed as the standard network management protocol for TCP/IP (transmission control protocol/internet protocol) based networks. Management information bases (MIBs) are created by vendors that describe the network objects and SNMP allows the *monitoring* and *control* of each object through a management station containing a user interface for operational staff.

Network management systems have long been used in elaborate monitoring of telephony networks; however, the unique combination of hardware and new, customized software for digital cable services do not allow for easy, complete management with existing management software and MIBs. The investment in customized MIBs and agents to support software and hardware monitoring are worth it for day to day operations considering the monitoring of :

- digital servers
- headend hardware (routers, switches, hubs, bridges, etc.)
- server network traffic
- SCSI and RF bandwidth
- viewing percentages
- custom operating software
- server operating system
- client and field devices

The automation of this monitoring will allow headend staff to isolate problems quickly and focus on maintenance and repair as opposed to elaborate monitoring.

Traditional network management systems will monitor components, determine where a fault occurs and notify the user through a graphical user interface. As expert software develops, we have the additional capability of having the system automatically fix the problem and reconfigure hardware and software to isolate the troubled area.

Failover Recovery

Another area critical to the operational management of digital content is how to handle failover. Assuming we have implemented a network management system that finds network problems early, what do we do about it? Proper management of failures begins in the design of the digital system. Few cost models for digital failover exist; however, we must keep in mind that subscriber downtime must be minimized and recovery of disk or server failures can be very time consuming.

Server Failures, or commonly called server crashes, should be accounted for in the design of the operational software. All digital services should maintain a shared memory area with a backup server so when one server fails, another can continue serving it's subscribers and functions.

Software Failures can be difficult to quantify. Applications for a PC or settop should be developed so software failures have user friendly recovery programs for the subscriber and allow continued navigation. When possible, software should utilize common libraries to ensure standards with failover are properly met.

Network Failures should be minimized through duplicate network paths between servers and configurable RF paths. If, for example, a modulator were to fail or be taken out of service, the system routing/connection scheme should be configurable to easily send that data through another modulator for the same node or area. Accounting for periodic maintenance of network hardware must be built into the operating software design.

Disk Drive Failures can be worked around by implementing existing technologies such as redundant array of independent disks (RAID) which uses a predetermined amount of disk space as overhead for redundancy; therefore allowing disk failures with minimal impact to streaming of data.

BANDWIDTH MANAGEMENT

Figure 3 shows a network topology which includes two media servers with disks attached that contain digital content. Data is switched and converted to MPEG2 for transport to the home. Digital bottlenecks in this environment shift continuously. We can have bottlenecks in moving data through the SCSI controllers from disks to the servers, in moving from the servers to the switch, and in sending burst data through the network, upstream and downstream. It is critical that all available customer bandwidth must be allocated on an as needed basis, creating a dynamic utilization environment.

However, even by having a dynamic bandwidth topology, what happens when no bandwidth is available? For example, as depicted in Figure 7, the same SCSI bandwidth may be used to populate unused drives as well as to stream video to subscribers. When the device capacity is maximized, a bandwidth manager has to be able to prioritize based on business issues. Each piece of digital content should be associated with meta data such as the application this content is associated with, time of day weighting factors, type of media, usage expiration date, and priority of use. When bandwidth contention occurs, these business factors can be used to prioritize bandwidth access. In the evening, for example, news content may be prioritized higher than children's programming and vice versa in the morning hours. Newly released movie titles may take precedence over the placement of older titles onto disks. This type of prioritization can only be accomplished with a software-based, automated content management system. Development of such systems is currently in its infancy.

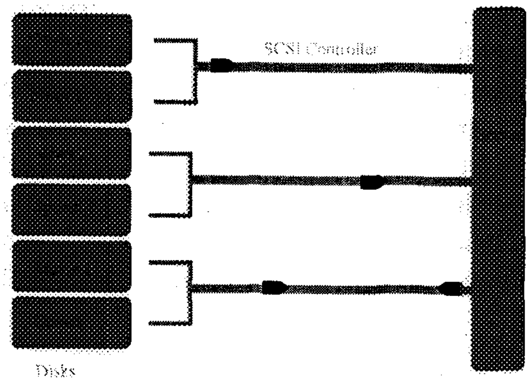


Figure 7. Sample SCSI Controller Assignment

Supporting Viewer Percentages

In addition to having an automated system managing bandwidth based on business and contractual factors, we must consider the use of bandwidth to support viewer percentages. If all bandwidth is being used to stream data to customers, we have no server bandwidth left to make additional copies of the content or to populate more content to disks using the same SCSI controllers.

There are several things we can implement to plan ahead for peak usage times. Figure 8 shows us sample VOD bandwidth usage. We can see that the average bandwidth usage is low over each SCSI controller, however, the peak use is quite high. When bandwidth usage (read and write) exceeds 70% or some pre-determined mark, bandwidth should be allocated for copying the popular content to another disk. By having an automated system for content placement which also monitors and allocates bandwidth for disk writes, we can be guaranteed not to create poor video quality (colored blocks on the television set) for subscribers when populating new content.

Another effective method of maximizing bandwidth is to offset the bandwidth and storage needs of popular content with less viewed content. In Figure 7 we see that SCSI bandwidth is shared amongst many disks. If news is popular in the evenings and children's movies are popular in the mornings, these two types of content would be well suited to utilize the same controller, therefore maximizing the available bandwidth for each at various times of day.

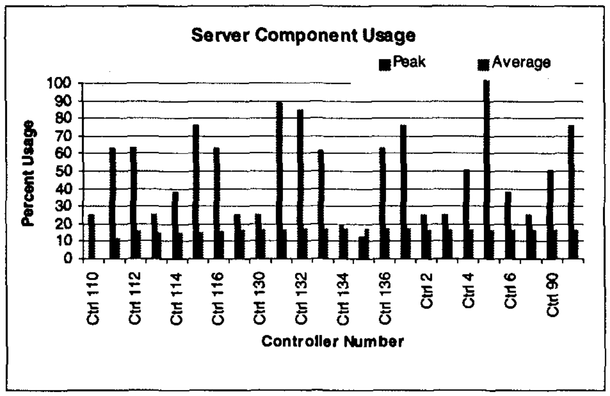


Figure 8. SCSI Bandwidth Usage

SUMMARY

Supporting the operational management of digital content poses new challenges. There is currently no digital content management software that tracks content access, predicts content access based on past and current usage, tracks and manages content distribution and storage, tracks and manages storage availability, manages scheduling of asset distribution based on predetermined criteria such as licensing constraints, manages optimal bandwidth utilization, and supports hierarchical storage management. By implementing the provided solutions for disk, network, and bandwidth management we can manage digital content today, maintain control of the system operations, and ensure quick failover recovery.

Subjective Effects of Bit Error on MPEG-2 Video

Mukta Kar, Majid Chelehmal, Rhonda Hilton, David Eng, and Richard Prodan
Cable Television Laboratories, Inc.

Abstract

The technology of digital communication systems allows more robust transmission than analog systems, but it is not completely error-free. During transmission, bit errors are introduced into the signal stream by the terminal equipment or the transmission media or both. Bit errors are unwanted but often unavoidable. Burst noise distributes bit errors due to the process of data interleaving before radio frequency (RF) modulation and transmission.

An attempt has been made to study the subjective effects of bit errors on the MPEG-2 (Moving Picture Experts Group) digital video stream. After studying syntax elements at the system layer and the video layer of the MPEG-2 standard, subjective effects due to bit errors have been estimated. The experiment at the CableLabs laboratory introduced errors into the bitstream at random locations using a software bitstream editing tool. The resulting bitstream was demultiplexed and decoded. Visual impairments were subsequently observed on a TV monitor and bit errors were analyzed using CableLabs' conformance tools. In most cases, the experimental results matched estimated ones.

INTRODUCTION

The MPEG-2 standard has been established to facilitate the large scale delivery and exchange of compressed audio/video information using communication networks and digital storage media. Another implicit goal of this standardization is to achieve interoperability among the equipment that will be used in this process of delivery and exchange. In the MPEG-2 standardization, syntax and semantics have been developed to represent compressed audio/video signals so that after transmission they can be recovered uniquely. Some errors are unavoidable and are present as channel noise in transmission system or as bit/byte dropouts in storage media. These errors are often termed "burst errors."

BACKGROUND

If the samples of a signal are sent sequentially in a transmission system or stored sequentially in a storage system, the effects of burst errors are overwhelming. The majority of the samples are lost due to burst errors and are not recoverable. To minimize the effect of burst errors, a scheme known as data interleaving is employed. The process of interleaving is shown in Figure 1.

In interleaving, a number of sequential symbols are assembled into code words. A number of sequential code words are ordered along rows in memory. When the memory is full, the symbol sequence is reordered over the medium by reading down columns. If burst errors occur on such interleaved data, full samples are not necessarily lost. The net effect of interleaving is to spread out the burst errors as bit errors over a large number of symbols instead of affecting one or more contiguous symbols. The original symbols are recovered by the inverse process of interleaving known as de-interleaving.

For the above reason, the authors investigated the effect of bit errors in the MPEG-2 bitstream on viewing or subjective quality.

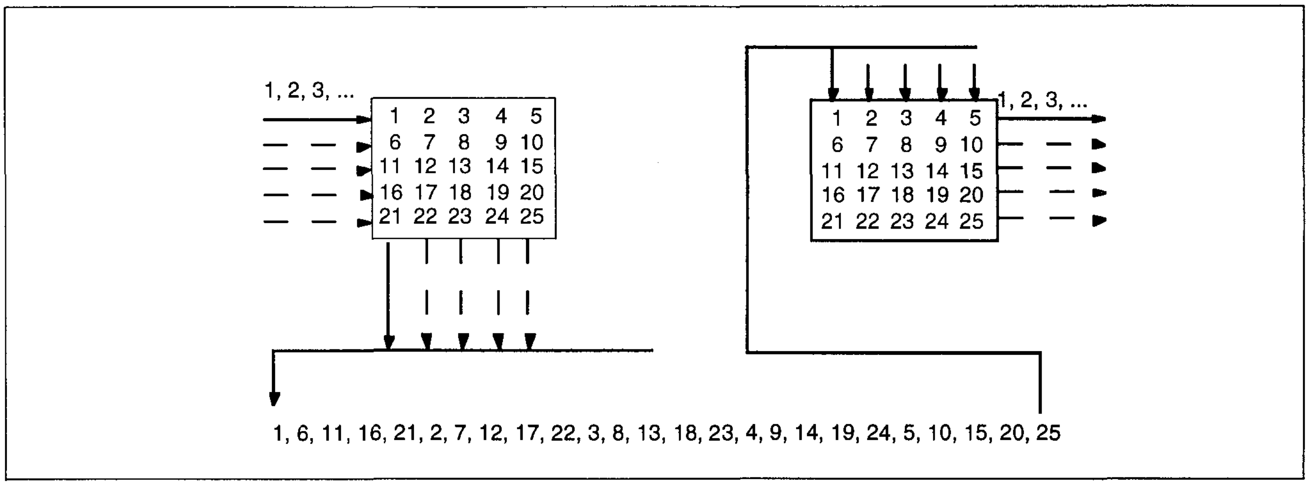


Figure 1: Interleaving and De-interleaving

MPEG-2 STRUCTURE

In order to understand better the effect of bit errors on picture quality, it is worthwhile to review the fundamental structure of MPEG-2 bitstream syntax. The MPEG-2 standard is divided into two distinct layers - the transport layer and the compression layer as shown in Figures 2 and 3.

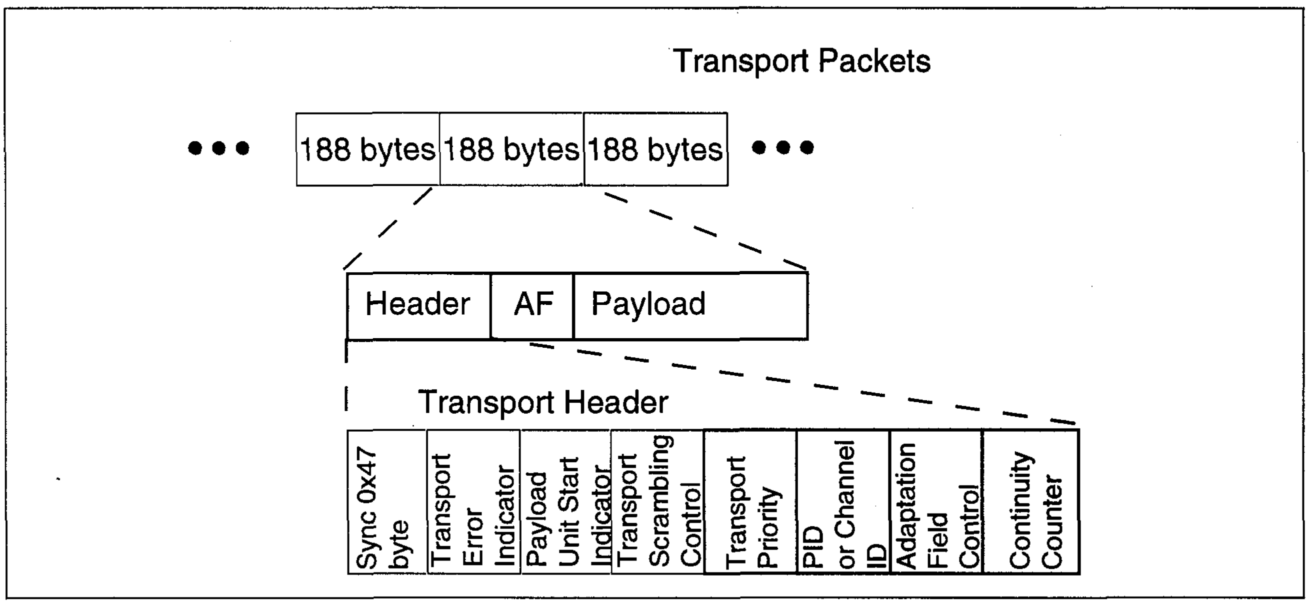


Figure 2: The Basic Content of TS Packet

Optical Network Technology: Future Impact on CATV Networks

John Holobinko & Bill Hartman
ADC Telecommunications, Inc.

Abstract

Future CATV networks may be able to transport video, data and voice services over large areas made possible by managing individual optical wavelengths within a single fiber, each wavelength carrying a different service type or going to a different location. The capability of optically routing, switching, provisioning and otherwise controlling various services without intermediate optical/electrical/optical conversion will enable creation of "All Optical Networks". This paper discusses the current state of optical technology required for these networks, where this technology is first appearing within existing CATV infrastructure, and how it may positively impact the capital, operations and maintenance costs of future CATV networks.

Introduction

Wave Division Multiplexing (WDM) is the ability to transmit two or more optical signals independently through the same fiber, utilizing different optical wavelengths. Although transmission of 1310 nm and 1550 nm wavelengths have been used for many years, it has been the advent of commercially available optical amplifiers which have made it is now technically feasible to transmit tens of optical signals simultaneously on the same fiber within a relatively narrow optical window of approximately 30 nm. This is referred to as Dense WDM Transmission, or simply dense WDM for short. Figure 1 illustrates a point to point dense WDM fiber link. Although the ability to transmit a dense WDM stream and amplify its multiple signals with a single optical amplifier is a key element of future All Optical Networks, commercial development of a number of new optical devices will be required in order to take full advantage of the potential benefits of All Optical Networks.

All Optical Networks may provide the following economic benefits to CATV service providers:

- Lower Fiber Plant Cost Through Significantly Reduced Fiber Counts
- Shared Signal Transmission and Switching Through Common Active Optical Components, Reducing Electronics Costs
- Improved System Reliability Through Reducing Overall Network Active Devices
- Faster Fiber Restoration After Cuts, Through Significantly Reduced Fiber Counts
- Reduced Future Costs For Network Capacity Expansion By Further Sharing Common Plant and Equipment

In addition to these benefits, technology improvements may allow each hub to economically serve significantly larger areas in terms of homes passed per hub, thereby allowing further consolidation and reducing operations costs.

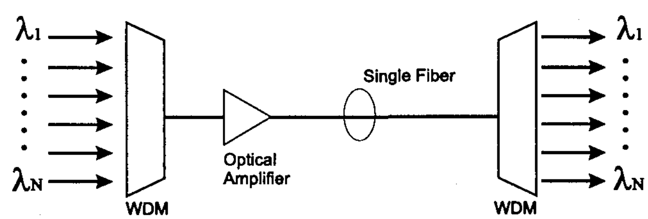


Figure 1. Dense WDM link. "N" can be 4-32 commercial applications today.

Elements of an All Optical Network

Dense WDM technology enables the creation of multiple optical circuit paths within a single fiber path (Figure 2). In relative terms, it is easy to compare an All Optical Network to a fiber network in the following way: An optical cable consisting of multiple fibers within a sheath becomes a "superset". An individual fiber within the cable can be thought of as a virtual fiber cable. An individual optical wavelength within the fiber can be thought of as a virtual fiber. Management, redundancy and routing can all be readily understood by translating requirements in conventional

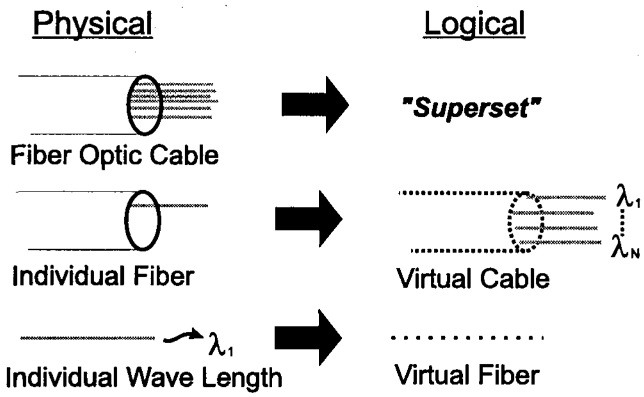


Figure 2. All optical network equivalents

networks between cables, fibers, and their virtual counterparts in an All Optical Network.

To attain the economic and operational benefits derived from implementing the future All Optical Network, a number of optical elements will be required which are currently not available in commercial quantities for wide scale deployment. Figure 3 is a compilation of these devices. From left to right shows progression in time of the anticipated evolution of these devices.

Dense WDM transmission is not without technical challenges. Operating at the 1550 nm window, attention must be paid to issues such as the dispersion performance of the optical fiber, the flatness of amplifier gain in the optical bandpass, and the optical stability of fiber devices. Recognizing these challenges, this paper is primarily focused on the potential application of All Optical Networks in CATV systems.

Large urban/suburban CATV networks consist of a series of hubs/subheadends (or video end offices) connected to one or two master headends (or video serving offices) usually via a redundant fiber optic ring, commonly referred to as the fiber backbone system. While most large backbones are exclusively digital, medium sized rings may be a combination of digital and high powered linear systems. A hybrid fiber/coax distribution architecture is used to distribute signals bidirectionally between the hub and the serving areas via linear optical transmission between the hub and optical nodes, and linear RF transmission within each serving area between the node and subscribers. Figure 4 illustrates a typical backbone system architecture while Figure 5 illustrates the distribution system architecture.

Dense WDM systems offer potential economic advantages in both digital and broadband linear (aka: analog) CATV transmission. The first area of anticipated use

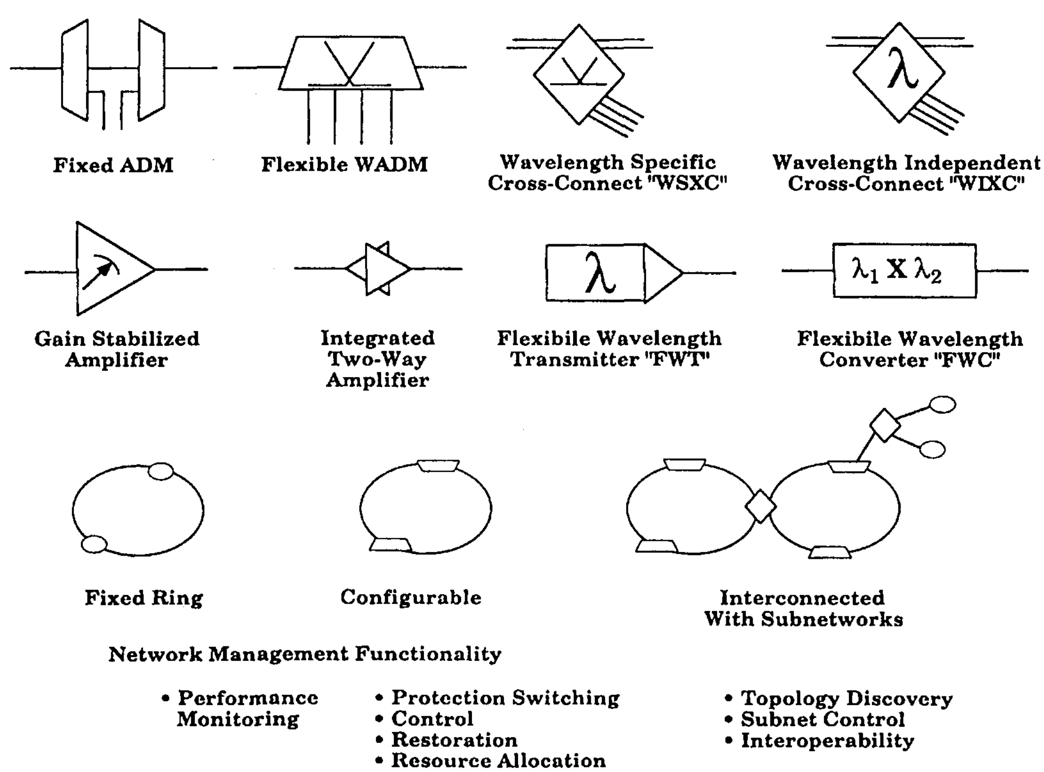


Figure 3. Current and future elements for All Optical Network: Source Bellcore

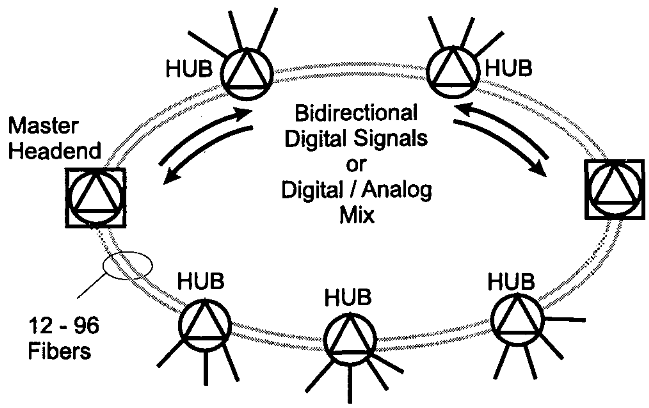


Figure 4. Backbone System. Fiber transmission is typically uncompressed digital for large networks, and a combination of digital and high power 1550 nm AM for smaller systems.

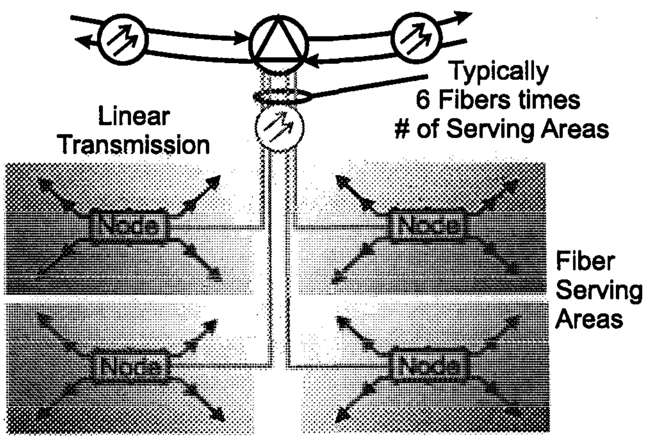


Figure 5. Distribution System. Fiber transmission is broadband linear between the hub and the optical nodes in the serving area.

of dense WDM technology in CATV networks is in fiber backbones, followed by potential implementation in the HFC distribution system.

Dense WDM in Backbone Systems

Fiber backbone systems generally cover long distances. The longest systems in the United States now cover over 500 kilometers. Given the expansion of CATV networks into the delivery of high speed data, telephony and other digital services, these systems usually employ uncompressed digital fiber systems to transport all video, data and voice services, or a combination of uncompressed digital and conventional telecommunications digital systems to transport and remultiplex various combinations of channels to create custom service delivery configurations at each sub-headend.

Shorter systems sometimes employ a combination of digital transmission systems (for voice and data services), and 1550 nm high power linear transmission systems (for broadcast video services) on separate fibers, which are then combined at the hubs.

Each fiber in backbone systems represents a significant capital cost because of the long distances traversed. The initial system may be even more expensive if requirements dictate leasing fiber(s) over a limited right of way such as a long bridge spanning a body of water. The ability of combining multiple digital signals using many optical wavelengths has already been demonstrated and commercial products are already available. For example, ADC Telecommunications demonstrated eight wavelengths of its DV6000 uncompressed digital system at the 1996 SCTE Expo as shown in Figure 6.

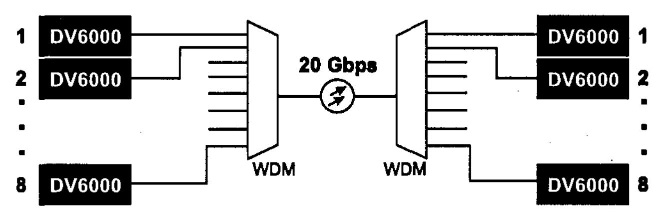


Figure 6. ADC Telecommunications' DV6000 Uncompressed Digital Transmission System employing DWDM for 20 Gbps transport.

This provides a single fiber capacity of 20 Gb/s, which translates to 128 analog CATV channels, 256 DS3 channels, 128 MPEG2 QAM multiplex signal streams (with up to 20 MPEG multiplex channels per stream) or a combination of these signal types. Other vendors currently provide systems which can transmit eight or more simultaneous wavelengths containing digital information, over the same fiber. A counter rotating ring can be implemented with no loss of capacity on 2 fibers, as shown in Figure 7. Alternatively, it is technically possible to implement a bidirectional WDM system with redundancy on a single fiber at 10 Gb/s.

Today, the application of WDM is primarily limited to point to point transmission between hubs on the ring. This is due to the fact that only "hard wired" fixed wavelength optical splitters are available. This provides the benefit of reduced fiber count, but not full optical add/drop capability. In order to dynamically drop or add a wavelength at any hub, a dynamically

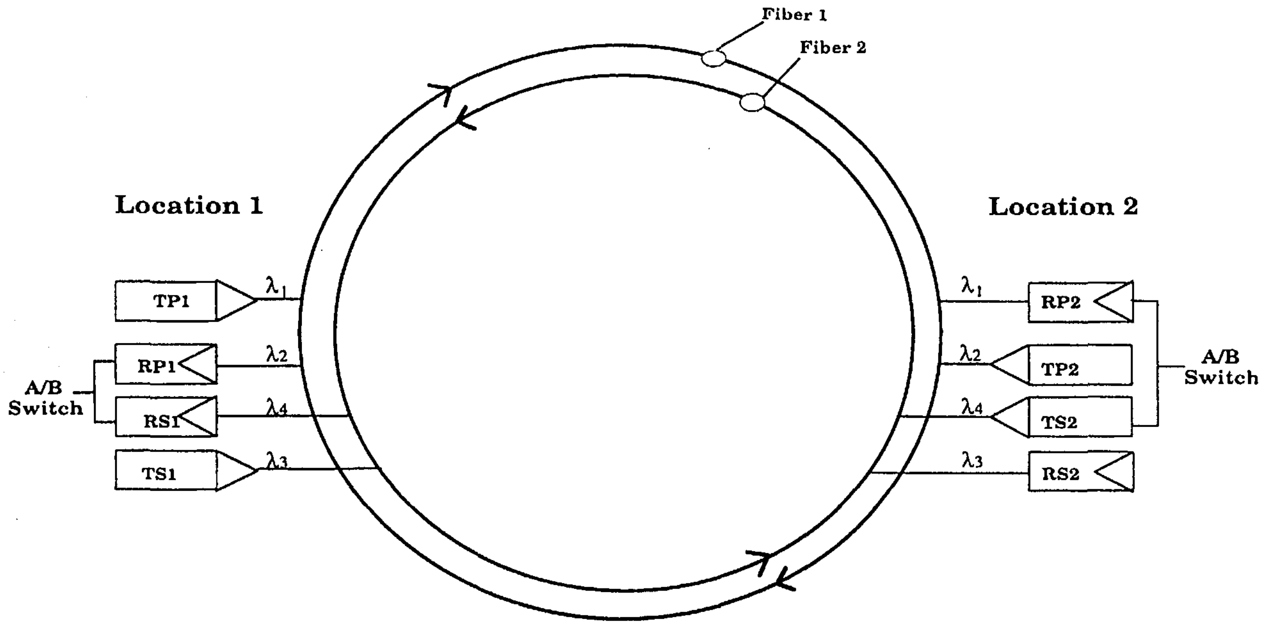


Figure 7. Counter rotating ring with full fiber and electronics protection

wavelength selective optical WDM device will be required. For high speed digital telephony systems, there may be a significant cost advantage in being able to do a drop and add function in optics versus electronically, due to the high cost of the digital add/drop multiplexor. Optical switching can perform virtually the same functions as TDM switching. The only optical limitation in multiplexor functionality s that there is not an easy way to make a drop *and pass* device.

The benefits of dense WDM are not as obvious in a smaller backbone system that employs 1550 nm high power linear transmission of CATV video channels. Although it has been demonstrated that digital and linear optical signals can be simultaneously passed through the same optical amplifier[1], the optical link budgets of high speed digital transmission systems are so dissimilar as to make this probably impractical in common implementation. For example, a 2.4 Gb/s operating at 1550 nm has a link budget of 29- 31 dB, translating to over 100 km in length, which is within the range of distance between 90% of all hubs. A typical high power linear system has a link budget of 12 - 14 dB. Therefore, the linear system will require amplifiers after a distance of 12- 14 dB which is less than half of the link budget for the digital system. Operating the digital system through these amplifiers will provide no advantage to the digital system, and therefore if digital and high power linear signals are mixed on the same fiber, extra cost must be incurred to provide splitters to the digital signals in order to bypass these optical amplifiers.

Dense WDM in HFC Distribution Systems

While there are clear capital and operational savings to be had today from building a new digital fiber backbone system employing dense WDM technology through saving of fibers and electro/optic repeaters, there are potentially larger future savings in the forward and reverse path of the HFC distribution system.

Consider that the average hub serves between 20 and 80 optical nodes. It is typical to provide 4-6 home run fibers between the hub and each node, since at least two fibers are normally required for bidirectional transmission, and extra fibers are installed to support future migration of nodes closer to the subscriber, or additional services close to the node. Multiplying the nodes times the fibers per node calculation means that as many as 480 fibers are required at the hub. Given that WDM would allow multiple linear signals to be transmitted on the same fiber, the number of fibers at the hub could be reduced by 67% while maintaining the ability for future expansion of nodes closer to the subscriber. Even greater savings are possible if fiber branching is allowed. For example, if 16 nodes could be served using 16 wavelengths originating on one fiber from the hub, then it is possible to serve 80 nodes two-way using only 10 fibers total. Figure 8 illustrates this concept. Note also, that in building a metropolitan system that there may easily be ten or more hubs, so that the savings realized is factored by the number of hubs (i.e. distribution networks) in the overall system. To accomplish this savings technically requires future development of both the lasers which can support 110 channel linear transmission at various wavelengths around 1550nm, as well as WDM optical devices which demonstrate excellent long term stability and performance while installed in an outdoor, unprotected environment.

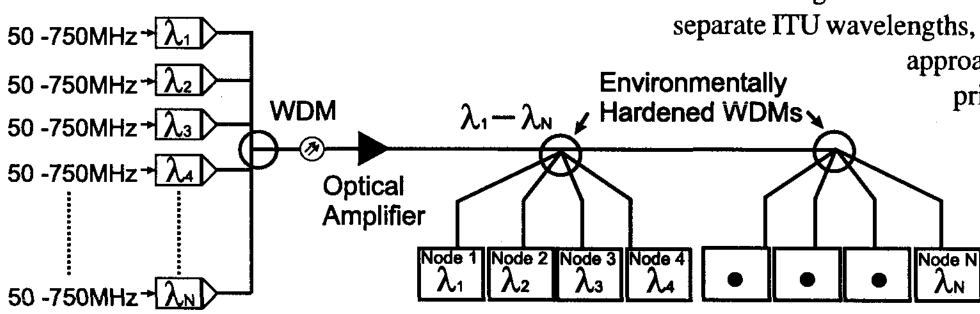


Figure 8. Potential Future Dense WDM in HFC Forward Path

Another alternative that has been postulated is to transmit in the forward path from the hub to nodes the common broadcast channels (typically from 50 - 550 MHz) at one common wavelength, and then to send the narrowcast signals to each node on a different wavelength where they are combined in the optical receiver. Although this is possible from an optical perspective, the combining of signals at the node receiver may be a more difficult approach in actual implementation. This is because it is difficult to combine the signals at each node with the correct RF level. Without going into a detailed technical explanation in the space of this paper, suffice to say that the RF output level of a signal out of an optical receiver is proportional to the input optical power of the received signal and the square of the depth of modulation. Trying to match two different optical transmitters' outputs coming into the node receiver with two power levels and two depths of modulation would probably be difficult.

Dense WDM may also hold future promise for return path expansion. Today, block upconversion is the most often proposed means of taking up to four 5-40 MHz return paths and transporting them across a single fiber back to the hub. In the future it may be possible to optically multiplex the return paths instead of frequency multiplexing these signals. Of course this will require both the multiple optical wavelength transmitters which are cost effective, and the WDMs which meet the environmental rigors required to be installed in a strand mounted unprotected optical node.

All Optical Network Challenges

The current barriers to All Optical Networks center around devices and availability. The ITU has suggested standardized channel spacing based on specified optical wavelengths in the 1550nm bandpass region of optical amplifiers. In the digital realm, high speed 1550 nm wavelength lasers are becoming available for 40 separate ITU wavelengths, at prices which are rapidly approaching standard 1550nm pricing. Work is being done by some vendors on creating multi-wavelength laser arrays to further drive down pricing. However, using 40 different channels brings up the real world problem of transmitter sparing. Currently, this limits the flexibility. The ideal solution is a tunable wavelength laser, if not for all transmitters, then at least as a universal spare. This is a comparable problem to that which the CATV industry had in the 1980's, when VSB/AM modulators were fixed channel. The advent of tunable lasers will significantly accelerate the implementation of WDM systems. Correspondingly, similar breakthroughs are required in broadband linear optical devices.

Another key requirement is the availability of low cost/ highly efficient WDM devices, both fixed wavelength and tunable, suitable for indoor and outdoor installation. In this area, technology is moving forward very quickly, and the emergence of these devices appears on the horizon.

Conclusions

Dense WDM technology and the ability to create All Optical Networks will allow further improvements in the cost, flexibility and reliability of CATV networks.

Availability of the optical components necessary to create these networks will occur within 2-3 years, giving network providers additional means of providing better service and additional capacity to their customers.

References

[1] Chinlon Lin, Keang-Po Ho, Hongxing Dai, Janyi Pani, Hermann Gysel, Mani Ramachandran, "Hybrid WDM Systems for High Capacity Video Trunking Applications", National Fiber Optics Engineers Conference Proceedings, September 1996, p. 261.

Wireless Telephone Industry Opens Doors for Cable

Matthew Waight
General Instrument Corporation

Abstract

The wireless industry is a leader in the development of new technology for low-power communications equipment. The integration of microwave frequency circuitry and new SMD filters for wireless has created new opportunities for other industries. These new components offer tremendous opportunities to improve and simplify cable products. A 860 MHz Dual-Conversion tuner is described which uses wireless technology to meet the requirements of all cable transmission formats, including NTSC, PAL, and 64 QAM, while eliminating all adjustments.

Cable Tuner Technology

Dual-Conversion tuners are used in cable systems in order to achieve the required composite distortion performance. These tuners have traditionally been designed using discrete oscillators, balanced diode mixers using ferrite baluns, and aperture HI-IF filters. General Instrument's 550 MHz tuner, using traditional technology, required tuning of the HI-IF aperture filter, Up-Converter mixer and oscillator, Down-Converter oscillator, and IF filter section. A total of seven separate adjustments were required for each tuner.

Predicting the performance of mixers is difficult when mixing multiple signals (1), (2) regardless of the technology used. Eliminating the variable of discrete diode based mixers and replacing them with a MESFET based Gilbert-Cell mixer makes the performance more predictable. Using GaAs technology (3), we were able to integrate an oscillator with the Up-Converter mixer and a differential RF amplifier in a single RF ASIC, eliminating the need for a separate pre-amplifier. Figure 1 shows a differentail amplifier driving a Gilbert-Cell MESFET mixer. This device takes advantage of a FET's superior third-order distortion performance while the Gilbert Cell's structure improves second order distortion. This change was implemented in 1992 and reduced the number of components in the Up-Converter section from 124 to 64 while eliminating two adjustments. Over 20 Million tuners have been manufactured using the integrated GaAs Up-Converter IC, making General Instrument a leading user of GaAs ICs.

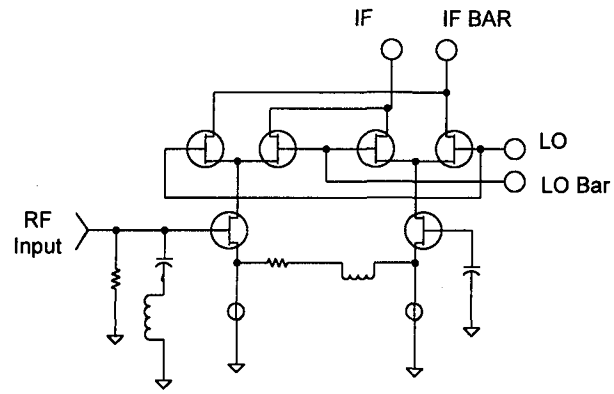


Fig. 1 Gilbert-Cell Mixer with Diff-Amp Input

The trend towards integration continued in our 1 GHz tuner which used an integrated Up-Converter, Down-Converter, and two single-chip synthesizers. As 1 GHz cable systems proved to be more marketing than reality this product was modified to a 860 MHz tuner. While we had increased the bandwidth and achieved a significant amount of integration we still retained a single-sided PCB and aperture HI-IF filter design.

General Instrument had two basic tuners (550 MHz and 860 MHz), with different versions for NTSC, PAL B, or PAL I output. Each version had different tuning procedures and Bill of Materials (BOM). These products used single-sided PCBs and a wave-soldering process which required a significant amount of inspection and touch-up. The total cost of supporting these products was becoming non-competitive due to labor costs.

A goal was set to design a single tuner for all converters, regardless of format, including digital terminals such as the DCT-1000 (64 QAM). The tuner needed to significantly reduce the direct labor requirement and eliminate all adjustments. In order to achieve this goal all cost items were considered, including manufacturing costs, test requirements, alignment cost, and indirect labor costs. Switching from a single-sided PCB to a double-sided PCB, which was traditionally rejected automatically due to the increased cost, was left open to consideration if the total cost was reduced. The basic block diagram can be seen in Fig. 2, showing the key sections of the tuner.

Pegasus Network Architecture

Michael Adams
Sr. Project Engineer, Advanced Engineering
Time Warner Cable

Abstract

This paper will describe the architecture of a digital video delivery system that is designed to complement Hybrid-Fiber Cable (HFC) networks. The paper will discuss the evolution of the network from a pure digital broadcast system into an interactive system and ultimately into a true video-on-demand capable system. This paper will include significant learning's about interactive network system design from the Orlando Full Service Network as well as integration of these designs with analog and digital broadcast systems.

INTRODUCTION

NETWORK ARCHITECTURE GOALS

The Pegasus Network Architecture is designed to meet the following set of key goals:

Integrated Service Delivery

This is achieved by integrating broadcast and interactive services into a single network architecture. The same mechanisms used to deliver digital broadcast services[1] will be used to provide advanced services like video-on-demand and home shopping. In this way, initial deployment for broadcast services sets the foundation for on-demand services. Incremental investment in the headend shall be the only requirement to provide on-demand services.

Cost Effective at Low Peak-Usage

It is critical that early deployment of interactive services can be cost effective even when there is a low utilization of those services. It must be possible to spread the cost of shared resources at the headend across a sufficiently large customer base (and revenue stream) to justify the investment in those resources.

The Pegasus network architecture is designed to be very flexible with respect to peak usage. The broadcast nature of the HFC plant is used to provide aggregation of demand across a large area (the size of a distribution hub or approximately 20,000 homes passed). As demand increases, more resources can be added as required.

Standard Service Encryption

Historically, analog cable systems have been tied to a single vendor solution by the access control mechanism; each vendor implementing their own proprietary scrambling and headend control systems. To promote multi-vendor digital set-tops, a standard service encryptor must be agreed. This allows multiple, independent providers to control access to their services while employing a single service encryption algorithm.

All services need access to a common, secure transport layer. This is provided by service encryption in hardware. A single encryption algorithm reduces hardware complexity. The access control mechanism must support a wide range of tiered services from broadcast to interactive.

In October 1996[2] some major elements of an interoperable digital cable systems specifications were agreed by CableLabs and its members:

- The agreement was based on existing standards (MPEG-2[3], ATSC Systems Information[4], ITU-T J.83 Annex B[5], DES encryption[6],[7]).
- The agreement was deliberately defined to be the minimum intersection of multiple CA systems:
  1. The adoption of a standard service encryption algorithm based on DES standards.
  2. A common control word generation method.

3. Use of existing features in the MPEG-2 systems layer to allow multiple CA systems to co-exist within a single digital channel.

This agreement represents the final and the most difficult step in long history of standardization. Because the CA system provides many features, it significantly differentiates one vendor's product from another. By separating the CA system into two parts (the service encryptor and all other components), each vendor is still able to innovate and add features to their CA system without introducing incompatibilities at the service encryptor level.

Support for a wide range of services

There are a wide range of possible services and service providers. There is no one platform that is optimal for all service providers - some services are broadcast, some are on-demand, others narrow-cast. Therefore it must be possible to connected many different service providers and servers to the network.

Growth of Server Capacity.

Application and media servers represent a significant investment and this will be true for some time. As more cost-effective servers are developed, it must be possible to deploy them while still gaining return on initial server investment. This means that all servers must support common interfaces and operating environment. Equally, as demand increases it must be possible to expand server capacity in economic increments.

Distributed Client-Server Communications

The client/server architecture is a powerful software paradigm that is well suited to large-scale interactive networks; the server provides shared resources and the client provides a rich user-interface.

The Pegasus Network Architecture supports a client/server architecture by providing real-time, two-way data communication between the client and the server.

Efficient Use of Network Resources

The HFC upgrade was initially designed to provide reliable, high-quality broadcast services. To enable interactive networks, additional bandwidth must be allocated to two-way communications. The efficient use of this bandwidth is crucial because the headend infrastructure costs (switching, modulation, and server capacity) scale directly in proportion to required bandwidth.

SERVICES

The Pegasus Network Architecture is designed to allows the integration broadcast, interactive and on-demand services. First we will define what we mean by these terms:

Broadcast

Digital broadcast services are a direct evolution of analog broadcast. The transmission medium is digital and content is compressed to make more efficient use of spectrum, but otherwise it is the same medium.

Examples of digital broadcast services are higher-quality versions of their analog counterparts; for example, HBO provided by a direct-broadcast, digital satellite service.

Interactive

Interactive services require a two-way, real-time connection from the set-top to the headend. The service can become much more responsive, allowing the subscriber to interact with an application program.

Examples of interactive services are the World-Wide-Web, multi-player games, and home-shopping services.

On-Demand Interactive

On-demand services extend the level of service again; the subscriber is given interactive random-access to the entertainment medium itself. The Orlando FSN is testing the marketplace for on-demand services.

Examples of on-demand services are movies-on-demand, news-on-demand, and sports-on-demand. We expect on-demand services to follow rapidly on the heels of digital broadcast deployment and, because the incremental cost of an interactive set-top is relatively small, it makes good business-sense to deploy interactive-capable set-tops.

Digital Communication Channels

The Pegasus network architecture defines three digital communication channels in addition to the conventional analog broadcast channel. These are:

- the Forward Application Transport (FAT) Channel
- the Forward Data Channel (FDC)
- the Reverse Data Channel (RDC)

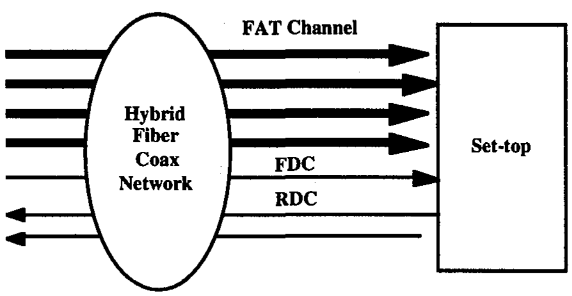


Figure 1 Digital Communication Channels. The arrow thickness is in proportion to the channel bandwidth.

Figure 1 illustrates the three types of channel.

- Forward Application Transport (FAT) channel. The set-top terminal can select any FAT channel by tuning to it.
- Forward Data Channel (FDC). The set-top terminal can always receive the FDC, even while tuned to analog services.
- Reverse Data Channels (RDC). The set-top terminal can only transmit in one RDC. However, more than one RDC may be defined per node for capacity reasons.

All channels are shared by a number of set-top terminals. A FAT channel can carry broadcast digital services in which case it is shared by all set-top terminals, or a FAT channel can carry on-demand services in which case it is shared by relatively few set-top terminals.

DIGITAL BROADCAST

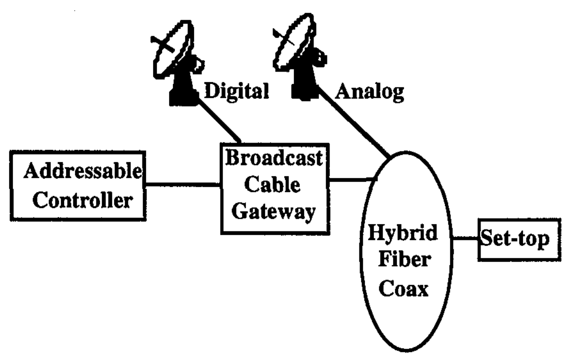


Figure 2 Digital Broadcast Architecture

A possible digital broadcast architecture is illustrated in Figure 2. The digital services are received from satellite in MPEG-2 transport stream format. The satellite feed is sent to a Broadcast Cable Gateway (BCG) which transforms the signal for distribution over the Hybrid Fiber Coax (HFC) network. Although this architecture provides channel expansion, it has some serious limitations:

- Interactive Services are not supported.
- High-speed Data Services are not supported.

- The out-of-band signaling supports impulse pay-per-view only.
- The network remains closed and proprietary - effectively locked-in to a single vendor.
- Proprietary networks are actually ***more*** expensive than standards-based networks.

PEGASUS PHASE 1

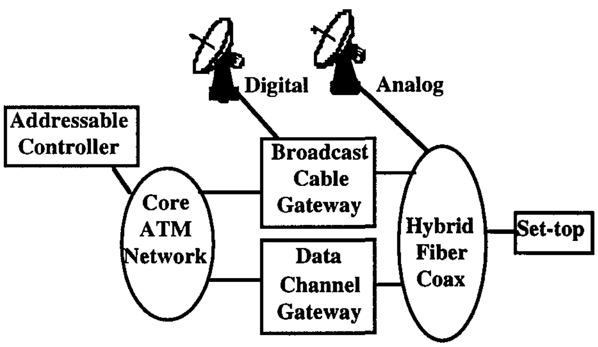


Figure 3 Interactive Digital Architecture

Pegasus Phase 1 adds a two-way, real-time data communications infrastructure. Figure 3 shows the addition of a Data Channel Gateway (DCG) to support two-way data communications. The DCG supports the Forward Data Channel (FDC) and the Reverse Data Channel (RDC). These channels provide a two-way, Internet Protocol (IP) datagram service between the headend components and the set-top terminal. IP is chosen because it is an open, industry-standard, protocol suite that can support interactive services, as well as management, signaling and application download.

Implementation

A typical Time Warner Cable division is shown in

Figure 4 after the HFC upgrade is complete. There are one or more headends and a large number of distribution hubs. Each distribution hub serves, on average, 20,000 homes-passed.

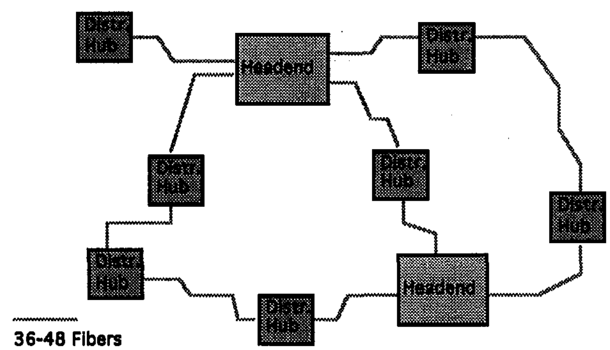


Figure 4. A Typical Upgraded Division

Figure 5 shows the actual components that will be deployed in a Phase 1 headend.

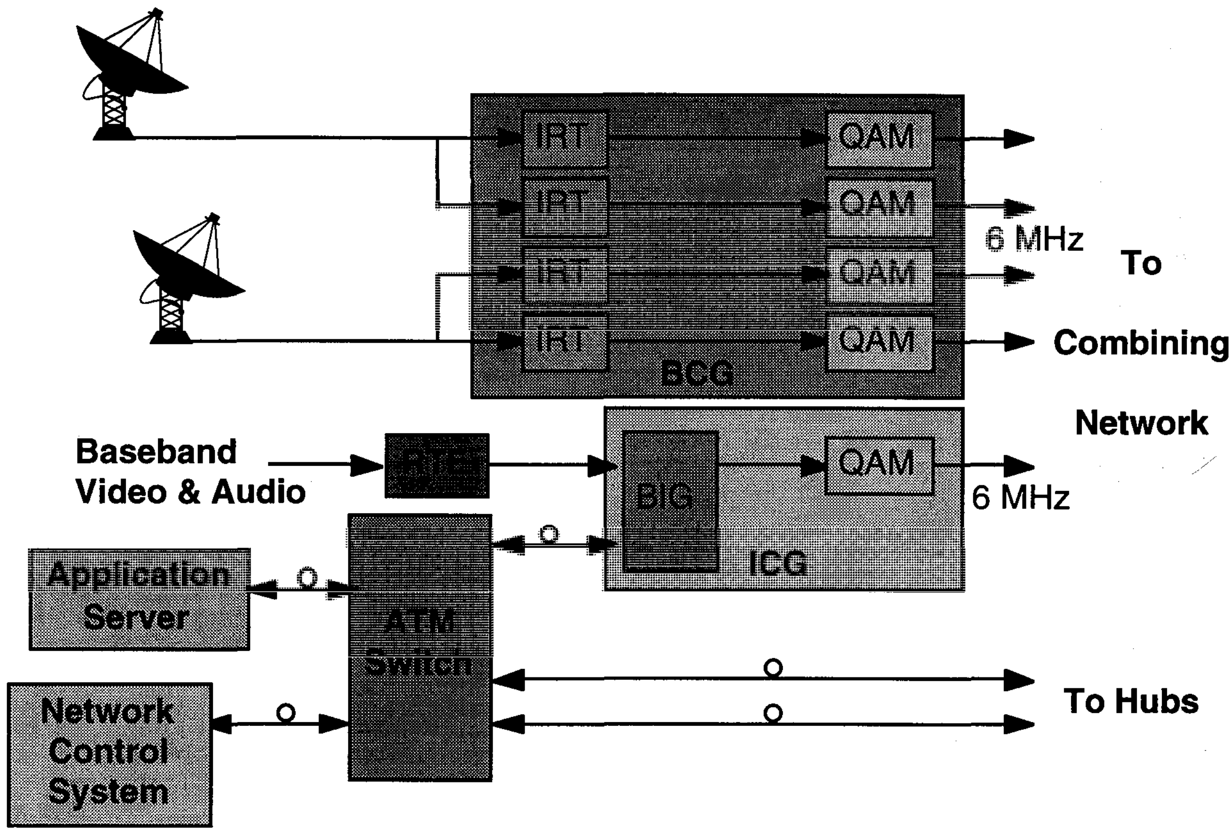


Figure 5. The Pegasus Phase 1 Headend

The components in Figure 5 will now be described in more detail:

Broadcast Cable Gateway (**BCG**)

The Integrated Receiver Transcoder (IRT) receives an QPSK signal from satellite and decodes it to provide a clear MPEG-2 Transport Stream. This is fed into the QAM Modulator which performs service encryption, inserts Entitlement Control Messages (ECMs) and inserts ATSC System Information.

Each IRT/QAM pair adds 6-8 digital channels in 6 MHz of cable spectrum.

Network Control System

The Network Control System supports the Pegasus Set-top Terminals and also provides Element Management for Pegasus Headend and Hub components.

The Network Control System runs on a standard UNIX platform and is located in a secure location within the Business Office.

ATM Switch

The Asynchronous Transfer Mode (**ATM**) switch concentrates out-of-band traffic from the Distribution Hubs. All connections are provided by OC3 single-mode fiber interfaces.

The ATM switch also switches out-of-band traffic to Distribution Hubs from the Network Control System and the Applications servers. A 10 Gbps ATM switch can support up to 60 Hubs.

Interactive Cable Gateway (**ICG**)

A Broadband Interactive Gateway (BIG) receives data via an OC3 single-mode fiber interface from the Network Control System and Application Servers.

The QAM Modulator adds service encryption and conditional access. In Phase 1, the ICG supports the DAVIC Data Carousel. This is used to provide system data to the Pegasus set-top terminal, for example, software updates and program guide information.

Real-Time Encoder (RTE)

The Real-Time Encoder receives a baseband NTSC video and baseband stereo audio input. A Real-Time Encoder is required per *locally-encoded* digital channel.

Real-Time Encoders cost approximately $50,000 per channel and so they are not recommended unless it is essential to encode programming locally because it is not available in MPEG-2 format from satellite.

Application Servers

Application Servers will be introduced in Pegasus Phase 1.1. (Examples are interactive program guide servers, WEB servers, and game servers.)

Application Servers require a UNIX platform and an ATM OC3c interface.

Distribution Hubs

Figure 6 is a diagram of the equipment located at the Distribution Hub.

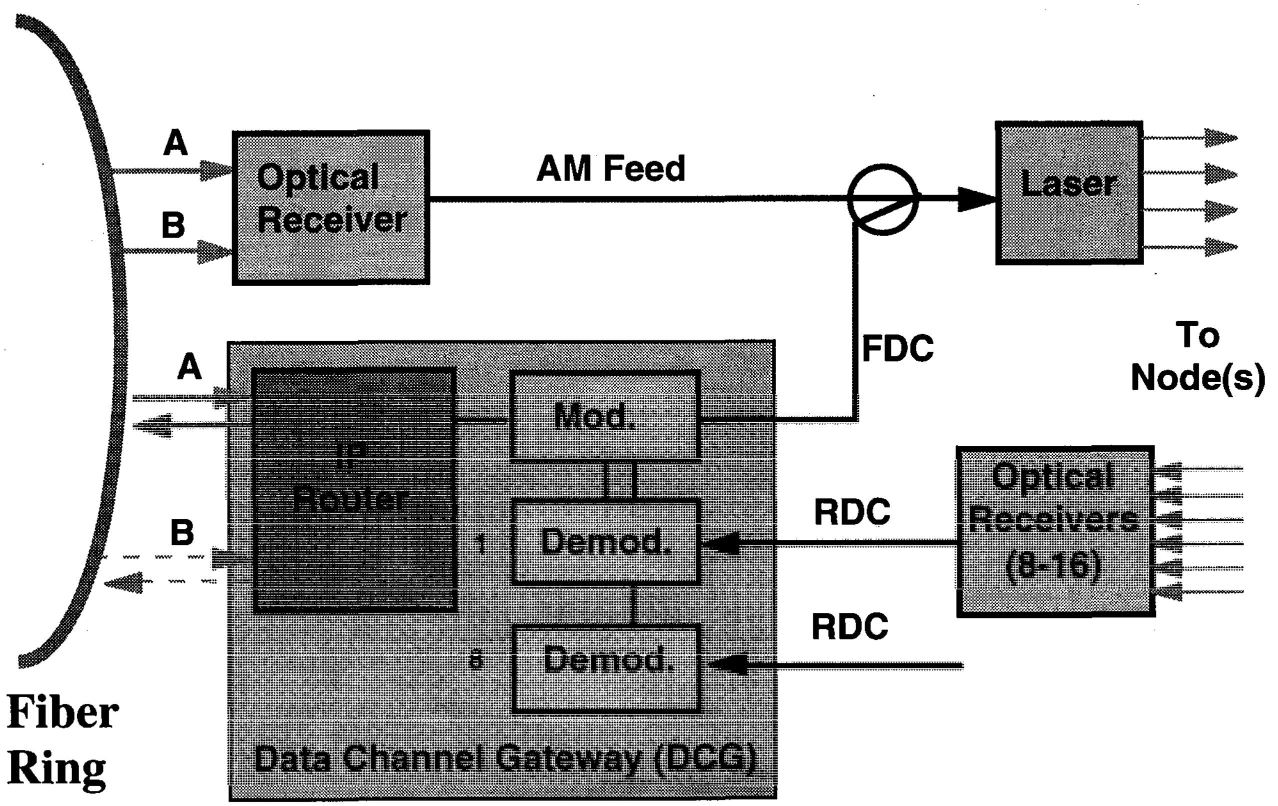


Figure 6. The Pegasus Distribution Hub

The components in Figure 6 will now be described in more detail:

Data Channel Gateway

The Data Channel Gateway consists of a standard IP Router and a number of DAVIC-standard[8] QPSK Modems. The IP Router transfers out-of-band signaling, control and data between the Headend components and the QPSK Modem. The IP Router is connected to the ATM switch at the headend by an OC3 single-mode fiber interface. A reach of up to 60 Km is possible (without repeaters) at 1310 nm.

The QPSK Modem is actually a combination of a single Modulator and up to 8 Demodulators working together to provide a high-speed, real-time, two-way path to the Pegasus set-top.

The QPSK Modem is located in the Distribution Hub for the following reasons:

• To control noise aggregation in the reverse path.

• To allow the FDC and RDC bandwidth to be progressively increased by sub-dividing the fiber nodes into smaller groups. More FDC and RDC bandwidth will be required as Pegasus set-top penetration increases and as interactive services are widely deployed.

Pegasus Phase 2

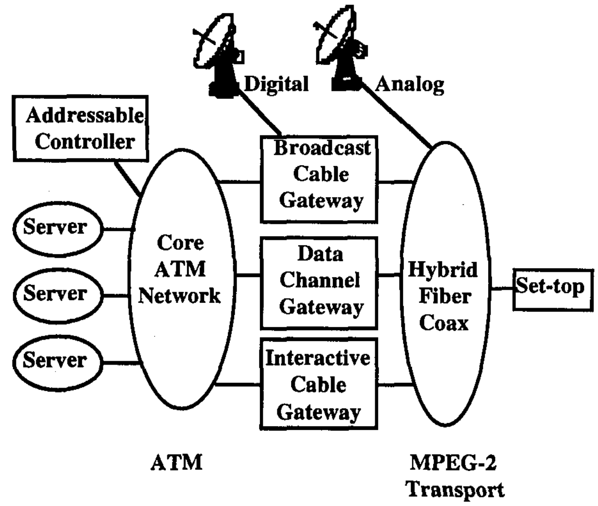


Figure 7. Pegasus Phase 2

Figure 7 shows the addition of on-demand services. Servers may be connected to a core Asynchronous Transfer Mode (ATM) network. An Interactive Cable Gateway (ICG) translates from ATM into MPEG-2 transport[9].

This provides the following advantages:

- Smooth migration to Video-On-Demand.
- Video-On-Demand can be supported even at very low penetration (e.g. <1%) because server investment can be shared across an entire division.
- Standard server interface enables multiple server vendors.
- Seamless integration of broadcast, interactive and on-demand services.

Implementation

As on-demand services are added, the only location affected is the Headend. Figure 8 shows the **additional** components in a Pegasus Phase 2 Headend.

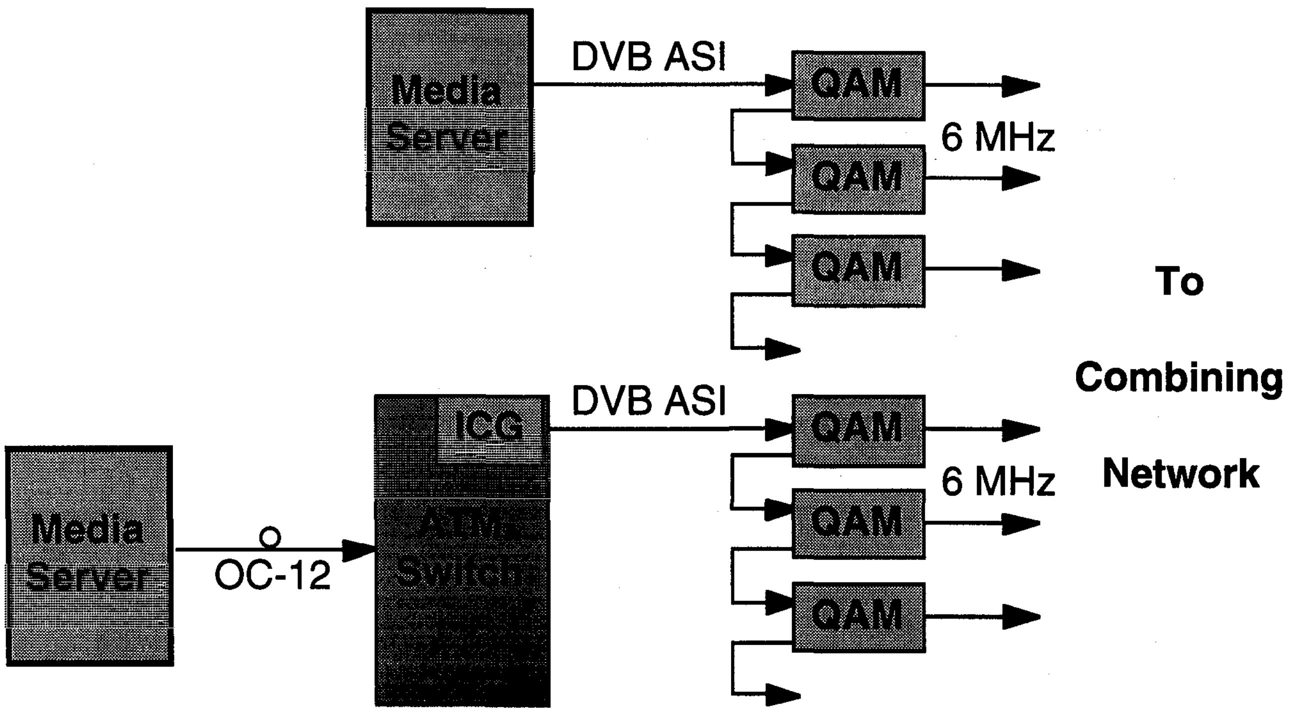


Figure 8. Phase 2 Pegasus Headend

The components in Figure 8 will now be described in more detail:

Media Servers

Media servers are required to deliver on-demand programming to the subscriber. Two types of media servers are envisioned:

1. Small-scale Video-On-Demand media servers which provide a DVB ASI[10] (Asynchronous Serial Interface). No ATM switch or ICG is required for a limited number of titles and where the server is located within the headend.

2. Large-scale media servers which provide an OC12c interface. This interface allows a large-scale server to be located remotely - for example, at a regional library center.

The Interactive Cable Gateway (ICG) will be packaged into an line-card in the ATM switch to reduce cost.

SUMMARY

Pegasus Network Architecture realizes a genuine communications network, allowing new services to be added as they become available.

Standard protocols (IP routing and ATM switching) are used to build an open network at lower cost than existing proprietary approaches.

MPEG-2 Transport is used to supports ***integrated*** digital broadcast and interactive services.

ATM is to ***switch*** interactive multimedia only where justified by the application.

REFERENCES

[1] Digital Video Transmission Standard for Cable Television, SCTE DVS-031, Society of Cable Television Engineers - Digital Video Standards, October 15, 1996.
[2] "Cable Industry Agrees on Key Elements of Digital Systems Specification", CableLabs Press Release, October 3, 1996.
[3] ISO/IEC 13818-1 (1994), Information Technology — Generic Coding of Moving Pictures and Associated Audio: Systems.
[4] System Information for Digital Television, Advanced Television System Committee Standard Jan 3, 1996.
[5] ITU-T Recommendation J.83 Annex B (1995), Digital Multi-Programme Systems for Television Sound and Data Services for Cable Distribution.
[6] Data Encryption Standard (DES), NIST FIPS PUB 46-2, January 1988.
[7] DES Modes of Operation, NIST FIPS PUB 81, December 1980.
[8] Digital Audio Visual Council (1996), DAVIC 1.1 Specification Part 8: Lower Layer Protocols and Physical Interfaces (draft as of September, 1996).
[9] Communications Technology, "ATM and MPEG-2", M.B.Adams, Dec 1995, Feb 1996.
[10] "Interfaces for CATV/SMATV Headends and similar professional equipment", Digital Video Broadcasting document A010 Annex B, October 1995.

Pegasus Set-top Terminal

Ralph W. Brown
Time Warner Cable

Abstract

The Pegasus set-top terminal has many new features and functions. It is the first cable set-top converter that is built on open standards. These standards help to create a platform that can be used to enable many new services. This paper will discuss these new features and how standards will be employed to enable future digital services.

THE PEGASUS STRATEGY

On March 6, 1996 Time Warner Cable released an RFP for the Pegasus Program [1]. The Pegasus set-top terminal is the foundation of Time Warner Cable's strategy for providing advanced analog and digital services to our subscribers. The Pegasus Program itself is the culmination of several years of experience creating and operating the Orlando Full Service Network™ [2,3,4,5].

Phased Implementation

It is clear that we cannot envision today all of the services that could potentially be provided over a Hybrid Fiber/Coax (HFC) network. Some services, such as Video-on-Demand (VOD) may not be cost effective today, due to server and switching costs. Consequently, the Pegasus set-top terminal must provide a platform that allows new services to be added to the system as they are created and become cost effective.

The Pegasus Program is a phased approach to deployment of services. The initial Phase 1.0 implementation provides analog and digital broadcast services, including Impulse Pay-Per-View, Interactive Program Guide, and Digital Music Service. Phase 2.0 represents the deployment of streaming video services, such as Video-on-Demand. Between Phases 1.0 and 2.0 is a continuum of applications that rely on connectionless IP based communications, such as Internet TV, Interactive Shopping, Interactive Games, etc. This phased approach allows for incremental investment in infrastructure and subsequent recovery of this investment through new revenue streams.

Address Competitive Digital Offerings

In its initial phase, the Pegasus set-top terminal is intended to address the competitive offerings of the DBS and MMDS providers. Currently, DBS service providers are promoting more channels and higher audio/video quality provided by digital broadcast. In order to address this competitive threat it is necessary that cable operators provide superior services. The greatest advantage cable operators have at their disposal is a dedicated, high-bandwidth, two-way connection to the home. The Pegasus set-top terminal is capable of supporting all of the services currently available on DBS platforms and more.

Open Standards

The Pegasus set-top terminal must be competitive on the cost side as well. A integral part of the Pegasus Program is the requirement for multiple vendors providing compatible products. In order to achieve this goal it is necessary to adopt open, standard interfaces at logical points throughout the system. The specific standards that were selected are described in subsequent sections.

Superior User Interface

The acceptance of the Pegasus technology by our subscribers will be directly affected by the quality of the User Interface (UI). If the UI is unappealing or difficult to use, the acceptance of this product will be poor as well. The graphics capabilities of the Pegasus set-top terminal, as described later in this paper, provide for a rich user interface. It is also possible to download a new UI into Flash ROM, as improvements are made in the UI over time.

Real-time Two-way Communications

One of the most critical features of the Pegasus set-top terminal is its real-time two-way communications capability. Real-time two-way communications are required in order to support interactive applications such as Internet Television and Video-on-Demand. The Pegasus set-top terminal makes use of an out-of-band channel supporting TCP/IP and UDP/IP communications protocols.

Application Platform

Finally, the Pegasus Program is ultimately a software strategy. The Pegasus System, and the Pegasus set-top terminal in particular, represent an application platform. The creation and deployment of new applications is facilitated by the Pegasus system. One aspect of this strategy is an attempt to eliminate applications that are dependent on specific hardware components, such as Sega Channel and Digital Music Service. The use of standards facilitates this effort.

PEGASUS HARDWARE ARCHITECTURE

Figure 1 is a block diagram of the Pegasus set-top terminal.

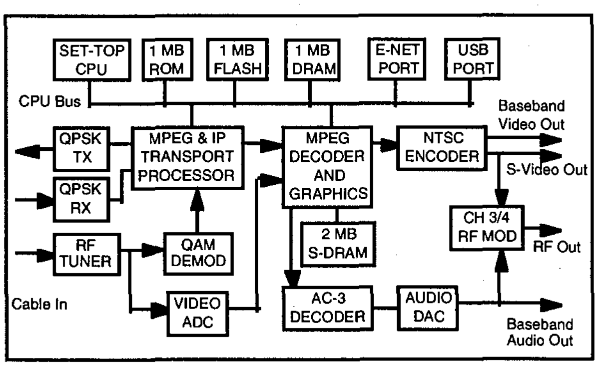


Figure 1 - Pegasus Set-top Terminal

The Pegasus set-top terminal uses QPSK modulation for out-of-band communications and a combination analog and QAM channel for video delivery. An MPEG and IP transport processor off loads much of the network protocol processing from the set-top CPU. The transport processor performs the task of MPEG transport demultiplexing and Packet Identifier (PID) filtering. The selected video and audio PIDs are passed on to the MPEG decoder and graphics processor. IP traffic from the out-of-band channel and potentially carried in MPEG private data are passed over the set-top CPU bus using Direct Memory Access (DMA) into DRAM memory.

Analog video is digitized through a video analog to digital converter (ADC) and passed to the graphics processor. Consequently, both the digital and analog video are processed in the same way. Graphics can be overlaid and blended with both the analog and digital video. The graphics processor also has the ability to scale the resulting digitized video.

The output of the MPEG decoder is passed to an NTSC encoder for baseband, S-Video, and RF modulated outputs. The output of the AC-3 decoder is passed to baseband stereo and RF modulated outputs.

The set-top CPU is interconnected with the set-top memory, MPEG decode/graphics, and two peripheral ports, a Universal Serial Bus (USB) port and 10BASE-T Ethernet port. These ports can be used to connect the set-top to other peripheral devices, such as keyboards, joy sticks, game players, and PCs.

Network Interfaces

The Pegasus set-top terminal is intended for use in an HFC network. Figure 2 shows the spectrum allocation for the Pegasus set-top terminal over the HFC network.

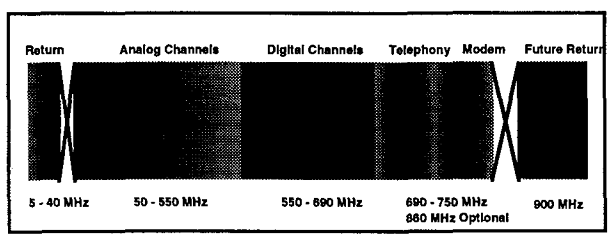


Figure 2 - Spectrum Allocation

The return path uses the range from 5 - 40 MHz. Analog broadcast is generally allocated to the 50 - 550 MHz range. A 140 MHz range is allocated for digital entertainment services and 30 MHz each is allocated to cable modem and telephony services. Future return capacity is available in the spectrum above 900 MHz. This spectrum allocation is not absolute and may vary based on programming choices and traffic requirements on individual headends.

The Pegasus set-top terminal has a single tuner for tuning both analog and digital services. Multiple tuners are possible, but it must also be possible to implement the set-top with only a single turner. Analog services use traditional broadcast NTSC signals. Digital services are implemented over a number of data channels:

- Forward Application Transport Channel (FAT)
- Forward Data Channel (FDC)
- Reverse Data Channel (RDC)
- Vertical Blanking Interval (VBI)

The Forward Application Transport Channel delivers MPEG-2 Transport Streams [6] to the Pegasus set-top terminal. The MPEG System Information Tables are defined by the ATSC standard[7]. The FAT Channel supports both 64 and 256 Quadrature Amplitude Modulation (QAM), using the ITU-T, Annex B QAM standard [8] as updated by SCTE-DVS [9]. This channel provides 26.97 Mbits/Sec with 64 QAM and 38.81 Mbits/Sec with 256 QAM in 6 MHz.

The Forward Data Channel delivers IP messages to the Pegasus set-top terminal. The FDC uses QPSK modulation providing 1.5 Mbits/Sec in 1 MHz bandwidth. The FDC is located in 70 to 130 MHz region and is frequency agile. The FDC is used to transmit commands and IP Messages. The advantage of the FDC is that it has 100 % availability, unlike the FAT which may be tuned to a different digital or analog channel. A second tuner could potentially make available a dedicated QAM channel for data transmission.

The Reverse Data Channel transmits IP messages from the Pegasus set-top terminal. The RDC uses QPSK modulation providing 1.5 Mbits/sec in 1 MHz bandwidth. The RDC is located in the 5 to 50 MHz region and is frequency agile.

The FDC and RDC communication channels conform to the DAVIC 1.1 standard [10] for the lower layer protocols. Higher level signaling, messaging, and download functions are performed using the Digital Storage Media Command and Control (DSM-CC) protocols [11].

A limited bandwidth channel is available through the use of the Vertical Blanking Interval (VBI) in analog channels. The use of the NABTS [12] standard for carrying data permits applications to receive data while tuned to an analog channel. This is particularly useful for augmenting analog broadcast video with program synchronous information, such as enhanced advertisement, sports, or weather information.

Unified Memory Architecture

One of the critical aspects of the Pegasus set-top terminal is its flexible memory architecture based on a unified memory address space. The MPEG decode memory is accessible to both the set-top terminal CPU as well as the MPEG decompression engine. This permits the MPEG decode memory to be used for application data or code when it is not being used for MPEG decompression. Larger, more graphically rich applications can run on the set-top by trading off video resolution with application size. This flexibility would not be possible in a split memory architecture. The net result is a lower cost set-top.

The memory space is made up of the following components:

- 1 Mbyte Mask ROM
- 1 Mbyte FLASH ROM
- 1 MB DRAM dedicated to set-top CPU
- 2 Mbytes DRAM shared between CPU and MPEG decoder

This memory architecture has the advantages of being downloadable, less DRAM required, and more application memory when the MPEG video resolution is low.

The set-top CPU includes a memory management unit (MMU) to prevent applications from corrupting the memory of other applications or the set-top operating system. This MMU is not intended to support a demand page swapping mechanism typical of general purpose computer systems.

Figure 3 shows the memory layout of the Pegasus set-top terminal.

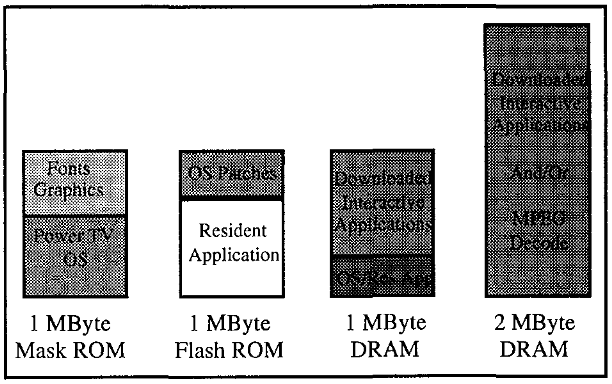


Figure 3 - Pegasus Memory Layout

The Mask ROM is used to hold the set-top operating system as well as many of the fonts and graphics used by the Resident Application. The Flash ROM contains the Resident Application itself and any OS patches that may be required. The Flash ROM is downloadable over the network and permits bug fixing and updates the User Interface over time. The DRAM dedicated to the set-top CPU is used for OS and Resident Application runtime variables and any transient applications that are downloaded to the set-top over the network. The DRAM associated with the MPEG Decoder and Graphics processor provides MPEG decode space and/or additional space for downloaded applications.

Video Processing

The Pegasus set-top terminal supports both analog and digital video. In the digital form it supports MPEG II Main Level / Main Profile [13]. This includes display resolutions of:

- 704 X 480
- 352 X 480
- 544 X 480

Supported aspect ratios are:

- 4:3
- 16:9

The supported MPEG data rates range from 1.5 to 15 Mbits/Sec.

Analog video is digitized and can be combined with graphics in the same way as digital video. This provides a consistent user interface across analog and digital video services. This consistency is important particularly when interactive applications could be performed in conjunction with both analog and digital video. The video can also be scaled by the graphics processor.

The Pegasus set-top terminal has the following video output ports:

- Baseband Video Output
- RF Modulated Channel 3/4 Output
- S-Video Output

An optional RF bypass module is available to enable direct bypass of the set-top to the RF Modulated output. With this option, a VCR can record the output of the set-top terminal while a TV can either display the output of the set-top terminal or tune to any clear (unscrambled) analog channels. This is the most basic form that allows recording on one channel while watching another. The Pegasus set-top is also capable of Master/Slave operation permitting two set-top terminals to be linked permitting VCR recording from one set-top and TV viewing from another.

The Pegasus set-top terminal can receive Closed Caption material, when present in the MPEG data stream on a selected digital channel, and insert it onto line 21 of the vertical blanking interval prior to routing the signal to the baseband, RF, and S-video outputs. On analog channels VBI line 21 is passed through to the video outputs.

Audio Processing

The Pegasus set-top terminal supports a variety of audio formats for both analog and digital services. The audio formats supported for analog services are,

- Monaural
- BTSC Stereo
- Secondary Audio Program (SAP)

The audio formats supported for digital services are:

- Dolby™ AC-3
- Set-top synthesized PCM audio

The synthesized audio can be mixed with either the analog or digital audio. For both the BTSC and AC-3 audio 5.1 channel AC-3 encoded audio can be multiplexed into 2-channel Dolby Pro Logic™ encoded audio. Audio encoded in Dolby Pro Logic™ surround sound is not processed by the set-top and is passed transparent through the set-top terminal. An external Dolby Pro Logic™ surround sound processor is required. Stereo output is available through the baseband left and right outputs. The RF channel 3/4 output modulator always outputs monaural audio.

The Pegasus set-top terminal is also capable of synthesis of audio within the set-top. This locally generated audio can be used for sound clips, alarms, and alerts. The synthesized audio is mixed with the audio associated with the analog or digital video content.

Secondary Audio Program (SAP) can be selected for analog and digital services when available from the program source. Analog services support only a single secondary audio track, while digital could potentially support multiple audio tracks.

Graphics

The Pegasus set-top terminal has a powerful graphics processor that is integrated with the MPEG Video Decoder. This graphic processor is capable of graphic overlay onto both analog and digital video. The graphics can be combined onto the analog or digital video using opaque, transparent, and translucent graphics (translucency and alpha blending). The graphics processor provides a hardware blitter with alpha blending and chroma key applied on blit operations. The graphics processor supports 8, 16, and 24-bit color depths selected from a full color palette of 16 million colors. These color depths permit rendering of shaded objects as well as realistic images. As a result the UI presented by applications can be very rich.

The graphics processor is also capable of video scaling by decimation and can use arbitrary vertical and horizontal scale factors. Frames of the video can be captured to memory from either digital or analog video. The captured video can then be manipulated by the set-top CPU.

To overcome limitations of interlaced NTSC display the MPEG decoder/graphics processor incorporates an anti-flutter and horizontal smoothing filter. To improve the appearance of text on the screen anti-aliased font support is provided. The hardware blitter can also be synchronized with the vertical sync pulse to enable smooth scrolling of display and other animation effects.

Additional Peripheral Ports

In addition to the above mentioned video and audio output ports the Pegasus set-top terminal has two additional output ports:

- Universal Serial Bus (USB) [14] port
- 10BASE-T Ethernet [15] port

These ports are used for connecting the set-top terminal to a variety of peripheral devices. The type of devices that can be connected to the set-top via the USB port include:

- IR transmitter for VCR control
- Keyboard for textual input
- Joysticks and other pointing devices for game play or Web browsing
- Printer for hardcopy output

By using the USB standard interface additional devices can be created and integrated into the Pegasus system. There is no need for specific hardware implementations to support new devices.

The type of devices that can be connected to the set-top via the Ethernet port include:

- Video Game Players

- Personal Computers

The Ethernet interface makes an ideal interface to video game players. A single standard interface can support a variety of game players without custom modification to the devices or the set-top. Games could be interactively downloaded to a particular game player and the Ethernet port enables interactive network games.

Since the Pegasus set-top uses IP as its data transport for the out-of-band communications channel, the Ethernet interface to a computer is a logical extension to the set-top terminal. In this configuration, the set-top terminal can function as a symmetric cable modem and a normal set-top terminal at the same time. Because the bandwidth is limited on the FDC this is not a replacement for cable modem, but can give people a taste of the capabilities offered by a cable model.

Security

The Pegasus set-top has a slot for a smart card and a built in secure-micro for support of security services. This includes key management for decryption of digital services and local authorization of IPPV purchases.

PEGASUS SOFTWARE ARCHITECTURE

Figure 4 is a diagram of the set-top software architecture.

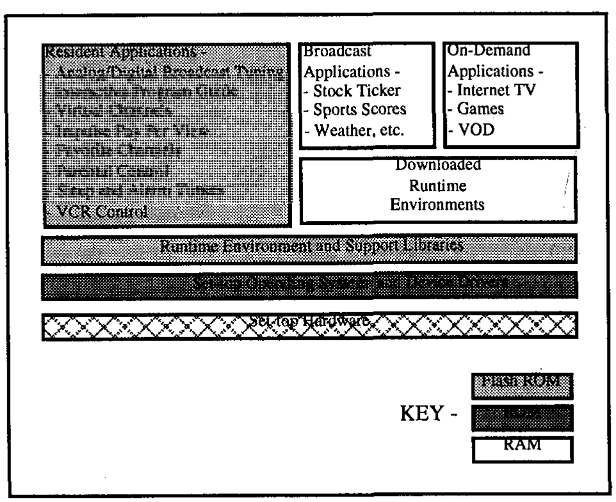


Figure 4 - Software Architecture

At the lowest level in this diagram is the set-top terminal hardware described in the previous sections. The next level is the set-top operating system and device drivers. Above the set-top operating system are the applications. The diagram also shows how these software components are mapped into the set-top memory architecture.

Set-top Operating System

The set-top operating system provides the interface layer between the applications and the set-top hardware. The Pegasus set-top terminal uses the PowerTV™ [16] operating system. This operating system kernel provides the following features:

- Priority-based, preemptive task scheduling, supporting multi-threaded applications
- Unified events system
- Dynamic linking and dispatching of functions and modules
- Interprocess communication facilities
- Interrupt and exception handling services
- Efficient dynamic memory management

In addition to these features, the operating system provides a set of support modules appropriate for set-top applications.

Resident Application

The Pegasus set-top terminal supports the resident application in Flash ROM. The resident applicant provides basic navigation and those additional features that must be immediately available when the set-top is turned on. In Phase 1.0 the Resident Application provides:

- Analog and Digital Broadcast Tuning, Channel Up/Down, Volume Up/Down, and Mute
- Reservation and Impulse Pay-Per-View (IPPV)
- Interactive Program Guide (IPG)
- Digital Music Service
- Parental Guidance Control
- Favorite Channels

- Sleep and Reminder Timers
- Messaging (including Emergency Alert System)
- Set-top settings (including AC outlet control,
- VCR Control through an IR link

After Phase 1.0 the Resident Application must also permit the download of interactive applications that provide other interesting services. These interactive applications fall into one of two broad categories, Broadcast Applications and On-Demand Applications.

Broadcast Applications

Broadcast Applications typically are program synchronous, i.e. they relate to the video content along with which they are broadcast. Examples of Broadcast Applications include:

- Enhanced advertisement
- Stock tickers
- Sports information
- Weather information

These applications are also typically broadcast in-band, either in the VBI channel for analog services or in MPEG private data for digital services.

On Demand Applications

On Demand Applications are typically downloaded to the set-top terminal at the time the subscriber requests the service. They are also typically session based or use client/server interaction. Since the Pegasus set-top terminal integrates real-time, two-way communications, interactive applications will be supported without hardware modification in subsequent Phases. The types of on demand interactive applications that are envisioned for future phases include:

- Internet TV
- Interactive Shopping
- Community Based Information Services
- Interactive and Networked Games
- Video-On-Demand (VOD)

The Pegasus set-top terminal will support true Video-On-Demand (VOD) as the capital investment becomes cost effective. The cost of video servers, headend switching and modulation equipment

SUMMARY

The Pegasus set-top represents more than just a digital set-top, it represents a strategy for moving forward into the competitive age of digital entertainment services. The key points of the Pegasus strategy are:

- Phased Implementation - deploy services as it makes sense
- Address Competitive Digital Offerings - meet and exceed the digital services provided by DBS and MMDS providers
- Open Standards - standards permit interoperability and multi-vendor suppliers
- Superior User Interface - a powerful graphics processor provides a rich user interface and Flash ROM provides the ability to update the UI over time
- Real-time Two-way Communications - essential for on-demand interactive applications, such as Internet TV and VOD
- Application Platform - to succeed in bringing new services to market, the Pegasus set-top provides an open application platform

REFERENCES

1. *Time Warner Cable - Pegasus Program Request For Proposal and Functional Requirements*, Time Warner Cable - Engineering & Technology, March 6, 1996.

2. Michael Adams, *Real Time MPEG Asset Delivery over ATM*, Proceedings National Cable Television Association - Cable '95 Conference, Dallas, TX, May 1995, pp. 315-326.

3. Louis D. Williamson, *Conditional Access and Security Considerations of Transactional Broadcast Digital Systems*, Proceedings National Cable Television Association - Cable

'95 Conference, Dallas, TX, May 1995, pp. 207-212.

4. Michael Adams, *Fault Tolerance in the Orlando Full Service Network*, Proceedings National Cable Television Association - Cable '95 Conference, Dallas, TX, May 1995, pp. 279-286.

5. Ralph W. Brown and J. Callahan, *Software Architecture for Broadband CATV Interactive Systems*, Proceedings National Cable Television Association - Cable '95 Conference, Dallas, TX, May 1995, pp. 270-278.

6. ISO/IEC 13818-1 (1994), Information Technology — Generic Coding of Moving Pictures and Associated Audio: Systems.

7. System Information for Digital Television, ATSC Standard, Document A/56, January 3,1996.

8. ITU-T Recommendation J.83 Annex B (1995), Digital Multi-Programme Systems for Television Sound and Data Services for Cable Distribution.

9. *Digital Video Transmission Standard for Cable Television*, SCTE DVS-031, Society of Cable Television Engineers - Digital Video Standards, October 15, 1996.

10. Digital Audio Visual Council (1996), DAVIC 1.1 Specification Part 8: Lower Layer Protocols and Physical Interfaces (draft as of September, 1996).

11. ISO/IEC 13818-6 (1995), Information Technology — Generic Coding of Moving Pictures and Associated Audio: Extension for DSM-CC.

12. Joint EIA/CVCC Recommended Practice for Teletext: North American Basic Teletext Specification (NABTS) (ANSI/EIA-516-88) (May, 1988).

13. ISO/IEC 13818-2 (1994), Information Technology — Generic Coding of Moving Pictures and Associated Audio: Video.

14. Universal Serial Bus Specification, Version 1.0, Compaq, DEC, IBM, Intel, Microsoft, NEC, Northern Telecom, January 19, 1996.

15. IEEE Std 802.31-1992, Type 10BASE-T Medium Attachment Unit (MAU) Protocol Implementation Conformance Statement (PICS).

16. Information on the PowerTV™ operating system can be found on the World-Wide-Web at: `http://www.powertv.com`.

PERFORMANCE OF CABLE MODEM SYSTEMS

Thomas E. Moore, Kenneth R. Martin, Robert S. Cruickshank III
Cable Television Laboratories, Inc.

Abstract

For the most part, the public telephone network is used to access the Internet from remote locations and data are transmitted at speeds ranging from 2,400 to 28,800 bits per second (2.4 – 28.8 kbps). With cable, however, data may be transferred at speeds up to 10 million bits per second (10 Mbps) — about 1,000 times faster than by using telephone wire. Cable offers a faster, more powerful alternative to telephony systems, and the result is an impressive increase in computer communications power.

Cable modem technology is in a unique position to harness the demands of users seeking fast access to information services, the Internet, business applications and cablecommuting. With the growing interest in data-over-cable service, there has been an increased focus on the issue of performance.

Cable data services are deployed on a shared access network. Thus a 30-Mbps service is shared by multiple users simultaneously. Due to the shared nature of these networks, there is interest in how the performance of the network as seen by a single user changes, if at all, as the number of users increases. This document will show that such degradation of service is minimal with proper design, and that cable modems will reliably deliver data to all users hundreds of times faster than existing phone modems.

ASSESSING THE CABLE MODEM BUSINESS

Over the final months of 1996, several cable operators began introducing high-speed data services on a limited commercial basis. An extended period of technology and market trials across the industry had yielded the same consistent results: People love the enhanced performance of cable modems compared to their much slower telephone dial-up modems. In fact, cable modems were found to vastly exceed the expectations of customers testing the devices. Once they experienced the tremendous increase in transmission speed and the benefits of being always connected, customers confirmed operators' hope that the public would embrace the service and be willing to pay for it.

The cable industry fully realizes it has a tremendous market opportunity to gain an early foothold in the burgeoning Internet access business. It is also known this window may be limited given the pace of developments in alternative access technologies by cable's competitors. Pent-up customer demand for more communications power is ripe for the taking, but operators also realize they must proceed with extreme caution. Cable's nay-sayers continually point to the failure of the industry to deliver on promises of new advanced services, as well as a past reputation of poor customer service. In order for cable to capitalize on this vast opportunity for new and lucrative revenue streams, we have to get it right the first time. It is absolutely critical to the industry that operators be able to offer reliable service and have the skills necessary to manage proactively this new data environment.

Cable Television Laboratories, Inc. (CableLabs) and many cable operators have taken a leadership role in helping the industry recognize the many challenges in running a data business. The fundamental problem is that cable modem service is new and not well understood. Many cable modem products are emerging from a variety of vendors, but few established products exist upon which operators may base their business. Virtually no empirical, quantitative performance data exist on these products. The end result is that the limitations, strengths and weaknesses of each vendor's product are not generally known. Further, there is no interoperability among the proprietary modem products from each vendor.

Scalability is another challenge which must be resolved. The effects on network performance of adding new users, increasing traffic loads and variances in traffic patterns are largely unknown. How these issues are addressed will have enormous network design and management implications. Added to all of this, fundamental aspects of the Internet and data services in general are in flux. The economic model of all-you-can-eat for $19.95 a month cannot continue. A metamorphosis of the business is underway, evidenced by the vastly different approaches of providers such as America Online, the Microsoft Network, NetCom and other Internet service providers.

Following is a synopsis of findings on cable data modem performance, as collected from member field trials and tested and simulated in the laboratory at CableLabs. Performance issues are examined through a series of steps. First, we look at network configuration (i.e., dedicated circuit vs. shared access) with a discussion of the advantages of a shared access network. Second, we study the medium access control (MAC) protocol and why it is the key element for determining performance. Third, we describe the high-speed. data-over-cable environment and how it was simulated and tested in the lab, with results provided. Fourth, there is a discussion on the importance of designing flexibility into the network when considering growth of users and deployment of enhanced services. Finally, support for enhanced services is discussed for ensuring that the network evolves gracefully as more demanding applications are added.

NETWORK CONFIGURATIONS: DEDICATED CIRCUIT VS. SHARED ACCESS

Two of the most prevalent network configurations in the communications world include a dedicated circuit connection and a shared access connection (shown in Figure 1).

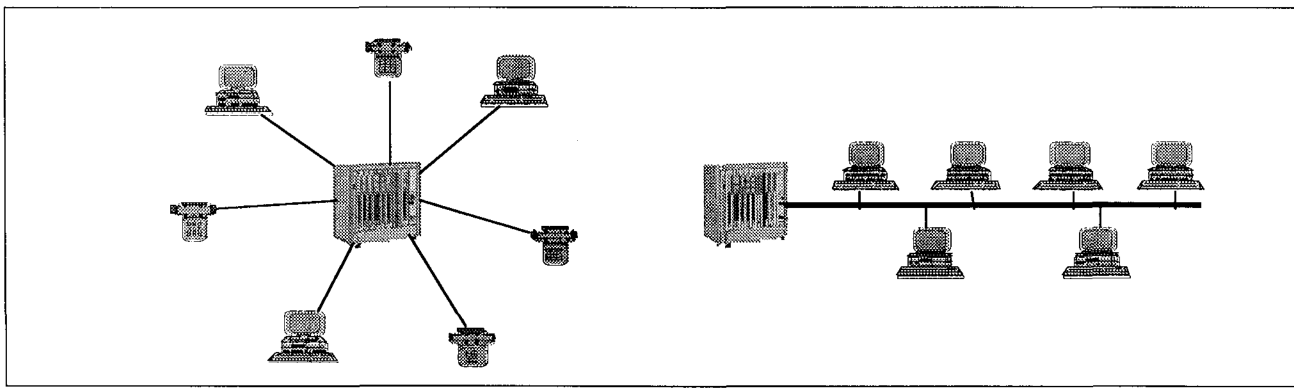


Figure 1: Dedicated Circuit Connection (left) vs. Shared Access Connection

The dedicated circuit architecture is used by telephone companies for services such as voice, telco data modems, and ISDN. In this configuration, a user establishes a dedicated, non-shared connection with a host computer on the other end. A single connection is set aside for each user's sole benefit until it is terminated at the user's discretion (e.g., user hangs-up). Typically, these connections offer maximum throughput of around 64 kbps to 128 kbps using ISDN, or 2.4 kbps to 28.8 kbps with a standard telephone modem.

A shared access connection similar to that employed in many local area networks (LANs) as well as in a data-over-cable environment is quite different. In the shared access connection, there is a single high bandwidth "pipe" shared among many users. This pipe might provide throughput capacity of anywhere from 10 Mbps up to 30 Mbps in a single 6-MHz channel that currently carries one video program. Since multiple users all share a common pipe, there is a need for a common set of rules followed by all users to share that resource in a fair and efficient manner. This set of rules is the MAC protocol and will be described later in some detail.

MYTH VS. REALITY IN A SHARED ACCESS WORLD

People grow up with the knowledge that a single apple pie shared between two people yields one-half a pie to each. Given this knowledge, it seems reasonable to assume that with respect to data communications, a single 30-Mbps pipe shared among, for example, 10 users yields effective throughput to each user of around 3 Mbps.

Fortunately, for the users of corporate LANs and data-over-cable services, such simple mathematics do not apply to shared access networks. People who have access to a 10-Mbps Ethernet LAN in the office know that such simple division does not work. It is commonplace to have 50 or 100 users on such a network and yet the individual performance seen by a single computer is much greater than 1/100th of the 10-Mbps connection. File transfers or Web pages representing several million bits of information pop up on the screen in a second or two even though the simple math described above would predict such transfers would take half a minute or more.

Why is this so? The answer lies in the inherent "bursty" nature of the data communications traffic that traverses such shared access networks.

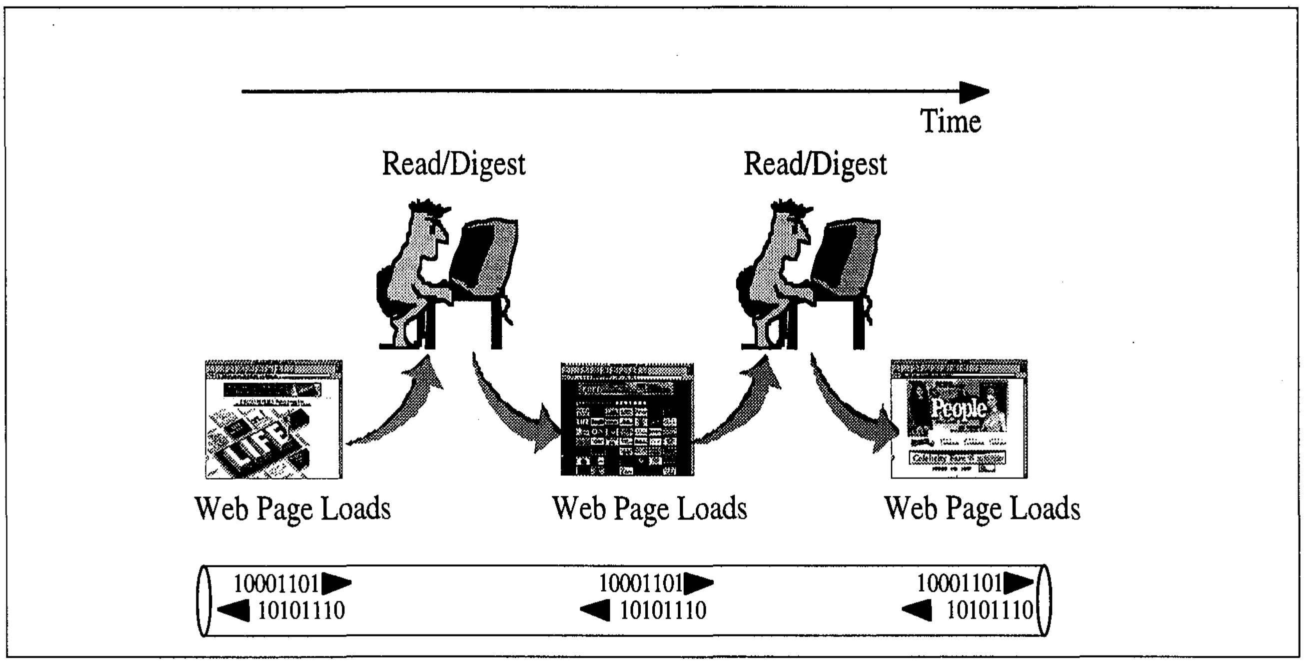


Figure 2: Web Surfing Process

Figure 2 shows a typical Web surfing sequence and the associated information flow in the network. The user requests a Web page. During the download of this page, bits of information flow on the wire. Once the transfer is complete (i.e., the page arrives), the user reads the page and during this time no data are flowing on the wire. Soon the user selects another link and information starts flowing again. This concept of short bursts of information flow followed by periods of quiet time is typical of many data services that run over the Internet.

To understand better the nature of this information flow, engineers at CableLabs designed a WWW (World Wide Web) surfing experiment. In a controlled laboratory setting containing both the PC performing the surf and the server holding the pages being requested, the engineers performed a typical 60-second surf of a popular Web site. During this surf, the engineers analyzed the traffic flowing on the broadband network connecting the two machines. Figure 3 shows the traffic flowing during this experiment. There is an obvious burst of information flow each time a Web page is requested and then a quiet time for several seconds while the page is read before the next page is selected. During this typical 60-second surf, 1.26 Megabytes of information flowed "downstream" from the server to the PC and the PC returned approximately 69 kilobytes of information in the opposite direction. This "upstream" traffic represents user mouse clicks and acknowledgments from the PC that it received the downstream traffic correctly.

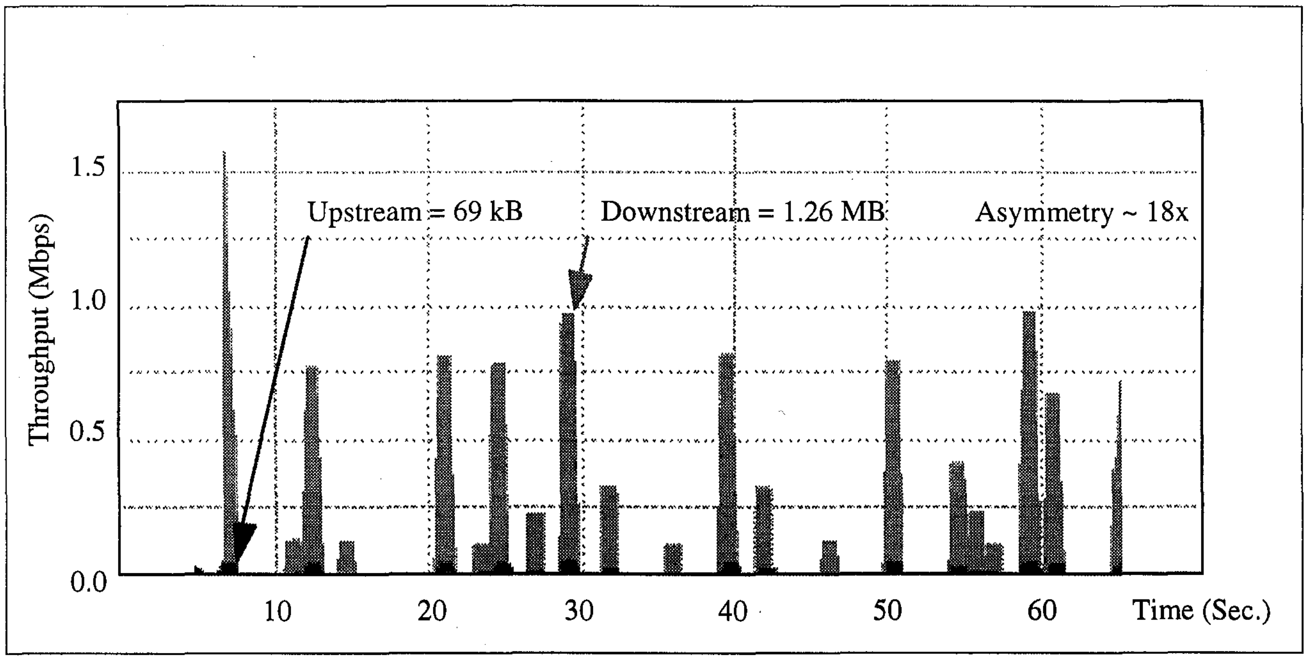


Figure 3: Typical Web Surf Throughput

This bursty nature of traffic flow is a big part of the reason shared access networks work so well in supporting multiple users. When a user wants to download a Web page, the user accesses the broadband network and transmits for a short burst at a high rate and then goes quiet. The next user then transmits during the first user's quiet period, and so on. This phenomenon is called statistical multiplexing of network traffic. During a time window of one or two seconds, it is not a requirement that only a single user be transmitting. In fact, with a 30-Mbps data-over-cable service, tens of users may simultaneously receive information at any instant in time. With a dedicated connection, the user has a low bandwidth connection tied up whether or not it is being used effectively: With a shared access network, the user is only using the network resource at the precise moment it is needed and then relinquishing control of that resource for others to use.

MEDIUM ACCESS CONTROL

The key element for supporting a large number of users on a single shared access network is split-second timing and coordination. Such coordination is supplied by an important piece of the modem technology called the MAC protocol. It is a set of rules followed by all network users to ensure that they share the bandwidth fairly and efficiently without performance degradation for others. The protocol allocates bandwidth to the users, arbitrates among users, and keeps track of all users' activities so that each user receives the desired throughput and assures the network is performing optimally. In order to understand how performance will change, if at all, as large numbers of users are added to the network, an understanding of the MAC protocol is necessary.

PREDICTED HIGH-SPEED DATA-OVER-CABLE ENVIRONMENT

In advance of results from full-scale deployment, many techno-savvy cable consumers, as well as cable's high-speed data competitors, are asking just how many customers can cable data service support? To answer this question, CableLabs has been using a combination of modem performance simulation and laboratory tests of actual modems.

First, a sophisticated computer model of various modem manufacturers' MAC technology was constructed. Connected to this MAC model were a number of simulated Web surfers all performing a 60-second surf very similar to the example described above. The model was verified with actual laboratory test results on 30 real modems to ensure that the simulation was producing valid results. The number of simulated users was then increased better to reflect full-scale deployments. The results are very encouraging.

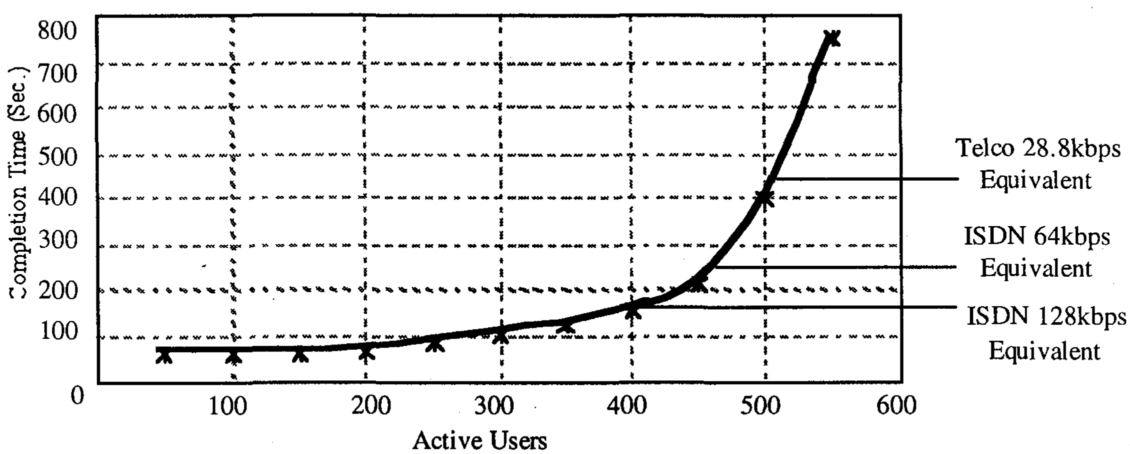


Figure 4: Predicted Data-Over-Cable Network Performance

Figure 4 shows the time it took to complete the typical Web surf as the number of simultaneous active users was increased from 50 to 550 users. One can see that with up to 200 simultaneous users, the typical surf took only the allotted 60 seconds to complete (i.e., the user saw no degradation in performance even when 199 other active users were on the network). As the number of users continues to increase above 200, the time to complete the surf begins to increase slightly, and when the number of simultaneous users reaches 400, the typical surf takes twice as long to complete. It is important to note, however, that even with 400 cable modem users simultaneously requesting service, the network performance still exceeds that of a 128 kbps ISDN connection, and the cable modem network performance with 500 simultaneous users exceeds the performance of the fastest telephone connection typically available today.

DESIGNING FLEXIBILITY INTO THE NETWORK TO SUPPORT FUTURE GROWTH

So what do the simulation and testing results mean? A network that only supports several hundred simultaneous users seems ridiculously ineffective in a city of several million potential subscribers, right? Not necessarily. The answer to this question exists in the design of to-day's modern cable plant. Most cable operators deploying advanced services such as high-speed data are doing so on upgraded 750-MHz hybrid fiber/coax (HFC) architecture. This approach (as shown in Figure 5) combines the use of optical fiber from a central headend to a fiber node, and then coaxial cable from the fiber node to the house. Such design allows the operator to subdivide a large city into many different sub-networks each with between 500 to 2,000 potential subscribers.

Each group of homes served by a single fiber node functions as a stand-alone entity and can be thought of as an individual sub-network.

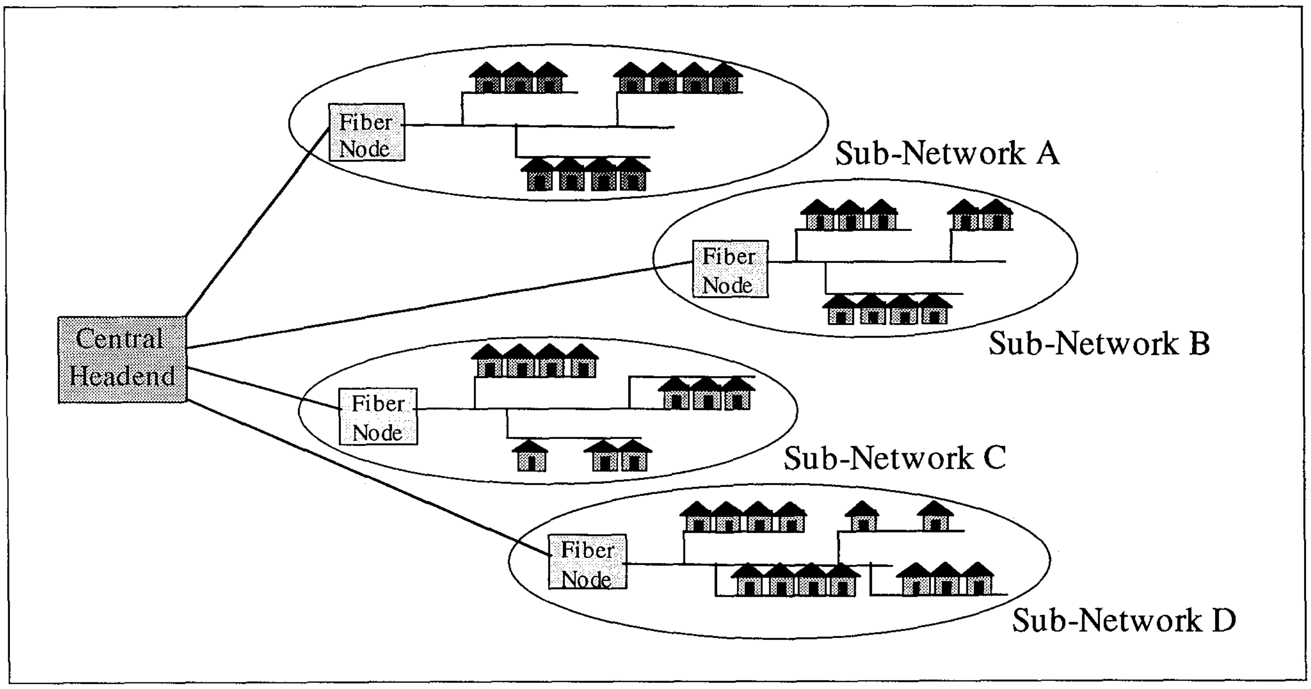


Figure 5: Sub-networks on an HFC Architecture

Thus, the cable operator need only design a single fiber node sub-network to provide effective service and then simply repeat that to reach millions of subscribers. The benefit of such a design means that if high-speed data over cable works for a single fiber node, then it can be repeated multiple times to support a large city. The following breakdown shows that even some of the largest nodes supporting 2,000 homes will likely see data service requested by only 129 simultaneous users during the peak period (busy hours) in the foreseeable future.

Homes on a Single Fiber Node =		2,000
% of Homes Taking Cable =	x	65%
# of Cable Subscribers =		1,300
% of Cable Subscribers Taking Data Service =	x	33%
# of Cable Data Subscribers =		429
Data Subscribers Online During Peak Usage =	x	30%
# of Simultaneous Active Users =		**129**

The cable operator's ability to sub-divide or segment the larger network into many small sub-networks provides a great amount of flexibility to evolve as demand for services such as data over cable increases. In addition to the ability to sub-divide the network, the use of the 750-MHz platform offers some 115 downstream channels on each fiber node that can be used for video, data, telephony, or other services. The operator has the ability easily to add additional data channels (i.e., replace an existing video channel with data service) or further sub-divide the network by running another fiber line from the headend if demand grows faster than expected. With digital video compression beginning to be deployed, additional channels will become available, relieving currently crowded channel line-ups.

THE BENEFITS OF A SHARED ACCESS NETWORK

The combination of existing HFC plant design and the flexibility to reallocate resources as demand grows means that the concern about performance degradation with an increased number of users is a non-issue. With this concern eliminated, one can focus on the numerous benefits of a shared access network.

First, a data-over-cable shared access network provides superior information burst capacity. When a user clicks on a link, he/she wants that page downloaded immediately. The ability to deliver that page in a timely manner is known as the network's burst capacity. On a dedicated circuit, such as a 64-kbps ISDN connection, there are only 64 kbps of throughput burst capacity available to a single user. It is impossible for that user to borrow additional unused capacity from other idle users. With a shared access network, however, a user has the ability to grab a big piece (i.e., several Mbps) of the shared 30-Mbps pipe to download the requested Web page and then release that resource for allocation to the other users. This ability to use resource only when needed provides great performance benefit, as well as inherent economic benefit to both the service provider and the customer. So cable provides not only greater average speed, but much greater maximum speed.

Second, a shared access network gives the data-over-cable subscriber the ability to be online all the time without the hassle of setting up a network connection every time he/she wants to send an e-mail or surf for information (i.e., waiting for computer to dial, server to respond and set up connections, etc., which can take 30-60 second every time). With a dedicated circuit connection, the resource is tied up regardless of whether the user is actually delivering or requesting information, thus the service provider must charge for the time the user is connected to the network. This provides the user the incentive to disconnect when he/she is done with a session and go through the time consuming process of reconnecting later. With a shared access network the modem is only using resource at the split second that information is requested. Thus users can be online 24 hours a day and only consume bandwidth when actively transferring information across the network.

Third, all users on a shared access network are connected to the same information pipe. This gives the content provider the unique ability to broadcast data streams (i.e., send one data stream down the pipe and have hundreds of users see it simultaneously). This can be an extremely efficient and effective way of delivering services such as streaming stock tickers, news feeds, multi-player games, software downloads, etc. In the dedicated circuit world, such services can only be provided by making a copy of such information for each user and repeating that stream on every dedicated circuit, which is inefficient.

SUPPORT FOR ENHANCED SERVICES

The exploding popularity of the Internet, combined with users' desire to do more online, has fostered a rapidly evolving application environment. Several years ago, the Web didn't exist and the predominant Internet application was text-based information retrieval and e-mail. Today, text and still-image based Web surfing is rapidly being augmented with Internet phone, broadcast audio and video, video conferencing, and shared collaborative applications. These evolving applications put greater demands on all networks. Fortunately, the MAC protocols being deployed today and being envisioned for the future have the flexibility necessary to effectively and efficiently support a mix of more demanding services on cable data networks. Design engineers are constantly redesigning and upgrading these protocols (which can be done easily with software changes in the headend) to handle new and different advanced services. This flexibility will be the key to success for cable operators deploying these services.

Glossary

Term	Definition
Dedicated Circuit Connection	a designated circuit reserved for a specific user
Downstream	the signal from the headend to the user terminal
Ethernet	a LAN used for connecting computing devices within the same building
Headend	in cable TV systems, the point at which the signal originates
HFC	hybrid fiber/coax
kbps	thousand bits of information per second
LANs	local area network
MAC	medium access control
Mbps	million bits of information per second
RF	radio frequency
Shared Access Connection	a single network shared by multiple users simultaneously
Upstream	the signal from the user terminal to the headend

Wireless Telephone Industry Opens Doors for Cable

Matthew Waight
General Instrument Corporation

Abstract

The wireless industry is a leader in the development of new technology for low-power communications equipment. The integration of microwave frequency circuitry and new SMD filters for wireless has created new opportunities for other industries. These new components offer tremendous opportunities to improve and simplify cable products. A 860 MHz Dual-Conversion tuner is described which uses wireless technology to meet the requirements of all cable transmission formats, including NTSC, PAL, and 64 QAM, while eliminating all adjustments.

Cable Tuner Technology

Dual-Conversion tuners are used in cable systems in order to achieve the required composite distortion performance. These tuners have traditionally been designed using discrete oscillators, balanced diode mixers using ferrite baluns, and aperture HI-IF filters. General Instrument's 550 MHz tuner, using traditional technology, required tuning of the HI-IF aperture filter, Up-Converter mixer and oscillator, Down-Converter oscillator, and IF filter section. A total of seven separate adjustments were required for each tuner.

Predicting the performance of mixers is difficult when mixing multiple signals (1), (2) regardless of the technology used. Eliminating the variable of discrete diode based mixers and replacing them with a MESFET based Gilbert-Cell mixer makes the performance more predictable. Using GaAs technology (3), we were able to integrate an oscillator with the Up-Converter mixer and a differential RF amplifier in a single RF ASIC, eliminating the need for a separate pre-amplifier. Figure 1 shows a differentail amplifier driving a Gilbert-Cell MESFET mixer. This device takes advantage of a FET's superior third-order distortion performance while the Gilbert Cell's structure improves second order distortion. This change was implemented in 1992 and reduced the number of components in the Up-Converter section from 124 to 64 while eliminating two adjustments. Over 20 Million tuners have been manufactured using the integrated GaAs Up-Converter IC, making General Instrument a leading user of GaAs ICs.

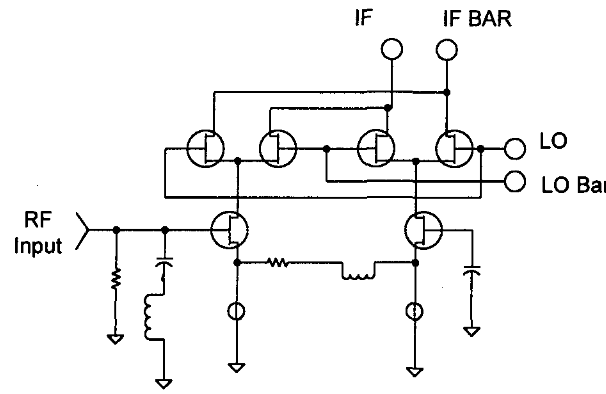


Fig. 1 Gilbert-Cell Mixer with Diff-Amp Input

The trend towards integration continued in our 1 GHz tuner which used an integrated Up-Converter, Down-Converter, and two single-chip synthesizers. As 1 GHz cable systems proved to be more marketing than reality this product was modified to a 860 MHz tuner. While we had increased the bandwidth and achieved a significant amount of integration we still retained a single-sided PCB and aperture HI-IF filter design.

General Instrument had two basic tuners (550 MHz and 860 MHz), with different versions for NTSC, PAL B, or PAL I output. Each version had different tuning procedures and Bill of Materials (BOM). These products used single-sided PCBs and a wave-soldering process which required a significant amount of inspection and touch-up. The total cost of supporting these products was becoming non-competitive due to labor costs.

A goal was set to design a single tuner for all converters, regardless of format, including digital terminals such as the DCT-1000 (64 QAM). The tuner needed to significantly reduce the direct labor requirement and eliminate all adjustments. In order to achieve this goal all cost items were considered, including manufacturing costs, test requirements, alignment cost, and indirect labor costs. Switching from a single-sided PCB to a double-sided PCB, which was traditionally rejected automatically due to the increased cost, was left open to consideration if the total cost was reduced. The basic block diagram can be seen in Fig. 2, showing the key sections of the tuner.

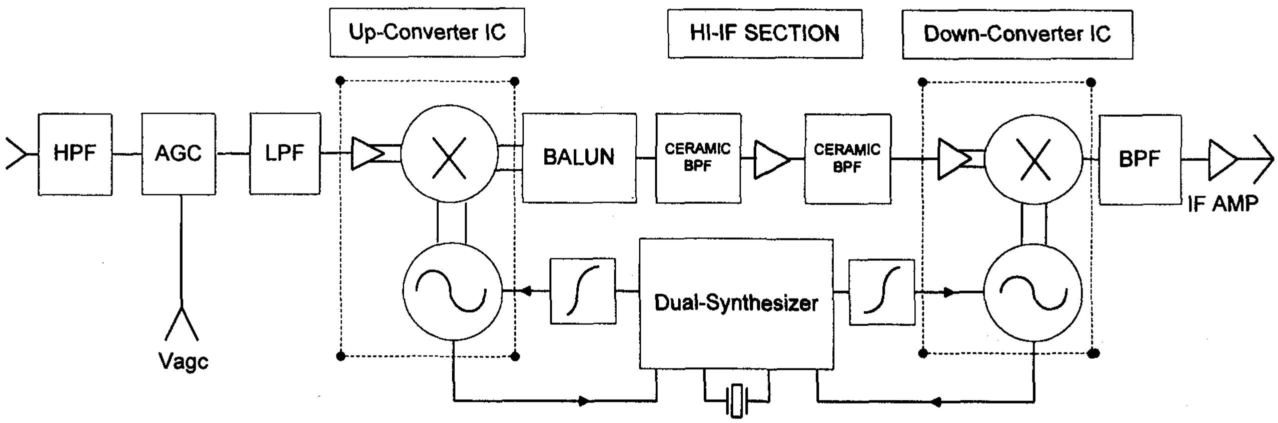


Fig. 2 V-860 Block Diagram

RF Transistors

The tuner being described has two key amplifier requirements. The HI-IF amplifier is used in-between the ceramic filters, (to be discussed), and the IF amplifier which follows the Down-Converter section. The traditional narrow-band high-frequency amplifier uses an air-coil between the supply and the collector and this part is often hand inserted. Using design notes for wireless amplifiers (4), (5), a quarter-wavelength transmission-line design was used to eliminate the air-coil. Using both voltage feedback and a constant base current source, emitter resistors were eliminated which helped achieve a noise figure performance of less than 2.5 dB with a low-cost bipolar transistor (Siemens BFP 183).

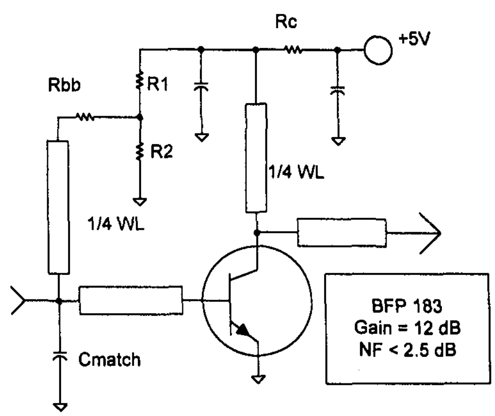


Fig. 3 HI-IF Amplifier Circuit

The IF amplifier typically uses a 3 dB pad on the output to improve the output return-loss but the wide-band filter design and 5V supply make the loss of 3 dB of gain impractical. The new design used an old design note, (6) which describes both series and parallel feedback techniques. By using non-linear models of the transistor (Siemens BPF 193) with HP's MDS RF design system, the circuit met all requirements with a 5V supply. The design achieved an output return-loss of 15 dB, gain of 15 dB, and met the X-MOD distortion requirement while eliminating the pad and offering a 6 dB improvement in third-order distortion performance.

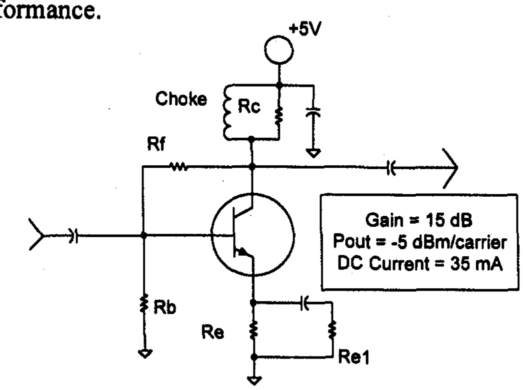


Fig. 4 IF Amplifier Circuit

RF ASICS

Our GaAs IC supplier, ANADIGICS, is also supplying the wireless industry which typically requires 5V technology. In 1995 they developed a Down-Converter ASIC which required only 5V instead of both 5V and 9V supplies, simplifying the design and lowering the power dissipation. As the first active device, the Up-Converter IC essentially determines the noise figure and distortion performance of the tuner as well as the phase noise degradation of the signal. The 64 QAM system used with our DCT-1000 digital terminal limits the phase noise degradation due to the tuner (7). Since the Down-Converter is narrowband, the phase noise performance is typically 9 dB better than the wideband Up-Converter oscillator circuit. The distortion requirements have previously required the differential amplifier/mixer section to use 9V.

By using new bias techniques, a 5V Up-Converter ASIC has been developed, allowing the tuner to use only 5V for all devices except for the varactor tuning voltage. This greatly simplifies the tuner design by minimizing the number of external inputs while further reducing the power dissipation in the tuner. By using advanced test equipment each RF ASIC is tested for critical RF and DC parameters including phase noise, 2nd and 3rd order distortion, gain, and noise figure. By using tested ASICs the number of parameters tested at the tuner level is reduced. As the wireless industry grows, vendors such as ANADIGICS should be able to further lower the price of GaAs devices which are still widely regarded as too expensive for consumer electronics.

Frequency Synthesizers

The use of phase locked loops to synthesize the oscillators needed in tuners is not new (8). The early TV and cable tuners placed a pre-scalar IC in the tuner while more recent designs have the entire PLL circuit, including the synthesizer IC, in the tuner itself. Bipolar technology was traditionally used for lowest cost. The wireless industry has generated a large market for very low power synthesizers which has been filled by BiCMOS devices having power consumption of less than 100 mW (9), by vendors such as National Semiconductor's LMX233X line of devices. Replacing two bipolar devices with one dual-synthesizer device can lower the power consumption by over 700 mW while lowering the parts count. As the wireless industry expands, these BiCMOS devices will continue to be lower in cost as manufacturers move to ever smaller device geometries and larger wafers. By adopting this technology the V-860 will enjoy the cost benefits of the wireless market while offering vendors a more stable demand for existing technology.

Ceramic Filters

While aperture filters can have excellent performance they require complicated mechanical designs and have to be tuned by hand. This limits the throughput of the line and adds a human variable to the product quality. While the cable tuner does not have stringent size requirements, any reduction in size is helpful. The wireless industry has severe size limits which has created a large market for ceramic filters which use high dielectric materials to realize very small filters with the required response (10). While these filters can not easily achieve the desired image attenuation in a single device they can be used in pairs with an amplifier in-between to more than achieve the required response. By using existing technology which is used in large volumes by the wireless industry, the cable industry can greatly simplify their products in a cost effective manner.

The V-860 was able to use an existing filter product and by slightly modifying the center frequency and bandpass limits, within the vendor's limits, achieve an acceptable response for use of two filters with an amplifier in-between. By buying a filter with a specified response, rather than manually tuning an aperture filter, all tuning was eliminated while improving the variance in the response. With the tuner's flatness over 6 MHz typically better than 0.5 dB, the ceramic filters have been a critical element in meeting this requirement. Each filter has a typical image rejection of 40 dBc, a 1 dB bandwidth of 35 MHz, and an insertion loss of 1.5 dB.

Another critical element is being able to work with PAL B (7 MHz bandwidth) and PAL I (8 MHz bandwidth) systems. Using a wide filter response is necessary to work with all of these systems. The same wider response is also needed for digital QAM systems to avoid amplitude and group delay distortion. A wider passband does increase the second and third-order distortion requirement of the Down-Converter but the result is a very repeatable product with no adjustments at a competitive price.

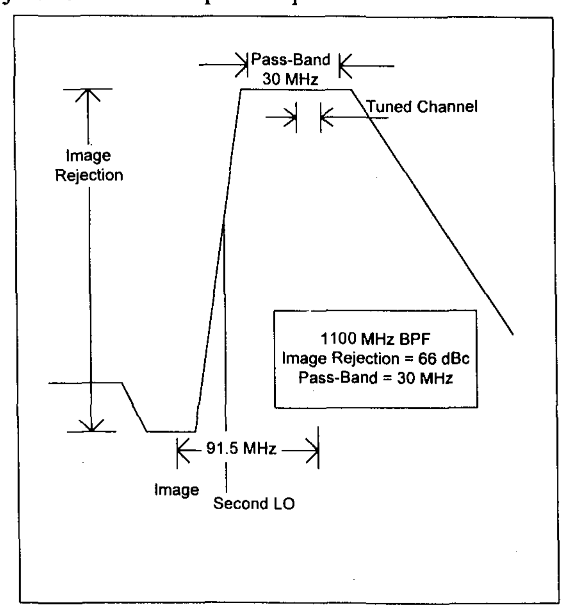


Fig. 5 HI-IF Response for Two Ceramic Filters

Design for Manufacture

Any attempt to lower the number of hand-inserted components and the number of adjustments will likely increase the bill of material (BOM). While the total cost of a product is what should be measured it is the BOM which is most easily measured and as a result is the most commonly monitored item. This can result in significant friction to any attempt to dramatically change the product design to reduce the labor cost. By gathering all costing information prior to proposing any change, quick rejections of ideas were prevented. Items such as direct labor rate increases, material overhead, failure rates, and indirect labor costs were be investigated. A key argument was comparing an increasing labor cost to a decreasing BOM over time and volume. This allowed us to argue for increasing the SMD percentage and investing in auto-insert equipment to reduce the amount of labor. By working with the manufacturing engineers from the start, better decisions were made early in the product development, avoiding cost increases later-on.

General Instrument's new tuner achieved 100% auto-insert of electrical components by primarily using SMD components with some auto-insertable chokes and crystals. Compared to the 860 MHz tuner it replaced, the V-860 resulted in a significant reduction in labor cost while increasing the build rate on a shorter production line. This was achieved by using a double-sided PCB and reflowing all SMD's while wave-soldering only the leaded components. All inductors less than 50 nH are printed to reduce the number of chokes. The use of 0603 chip resistors was introduced to decrease the size of tuner. The number of components was reduced with a significant reduction in leaded components.

The process of building a tuner starts with screen printing solder paste on a panel of PCBs, followed by placement of all SMDs. The panel is then reflowed. The leaded components are then auto-inserted. The PCB's can then be separated and inserted into a chassis and the assembly is then wave-soldered. Covers are added and the finished tuner is auto-tested. The new process eliminates all glue and all hand-insertion of components. There are no adjustments to slow the line rate and by eliminating the wave-soldering of the SMD components the amount of inspection and touch-up is minimal. As important as improving the cost structure of the tuner is, the increase in build capacity is also critical. By having one tuner for all products, forecasting is easier and delivery times are improved.

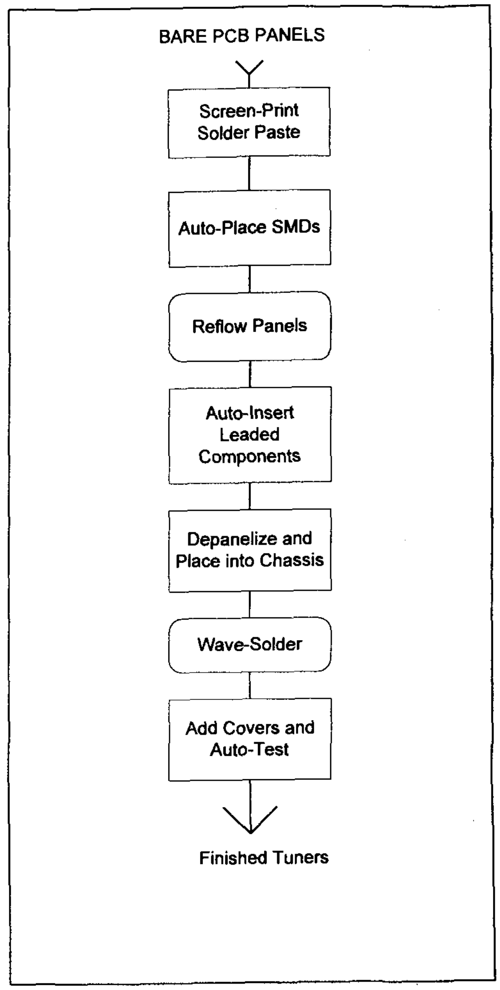


Fig. 6 Process Flow for Tuner

Universal Tuner Design

While wireless cable may sound illogical, wireless for cable makes a lot of sense. By using devices developed for the wireless telephone industry a 860 MHz Dual-Conversion Cable tuner has been developed which eliminates all adjustments and all hand-inserted components. The filters and RF ASICs have allowed for significant reductions in direct labor while reducing the parts count. The wireless technology has allowed the tuner to be used with all broadcast formats, including digital QAM systems. If a mature technology, such as dual-conversion tuners, can be dramatically improved by using technology from the wireless industry, one can only wonder what other advances are possible for the cable industry if we bother to look.

References

1. Barrie Gilbert, Rupert Baines, "Fundamentals of Active Mixers" Applied Microwave & Wireless, Winter 1995, pp. 10-36.

2. Michael Leffel, "Intermodulation Distortion in a Multi-Signal Environment" RF Design, June 1995, pp. 78-84.

3. N. Scheinberg, "A GaAs Up Converter Integrated Circuit for a Double Conversion Cable TV "Set-Top" Tuner", IEEE Journal of Solid-State Circuits, Vol. 29, No. 6, June 1994, pp. 688-692.

4. Terry Cummings, "Designing a Low Cost GPS LNA Using the NE68519", RF Design, June 1994, pp. 36-45.

5. Hewlett Packard Application Note 944-1.

6. Richard S. Hughes, "Unique Properties of the Series-Parallel Feedback Video Amplifier", EDN.

7. Kamilo Feher, Advanced Digital Communications, Prentice-Hall,New Jersey, 1981.

8. J.G.N. Henderson and E.N. Monacchio, "Frequency Synthesis for Television Receivers" RCA Laboratories, ER-166, January 1975

9. Jack Browne, "BiCMOS PLLs Conserve Current in Wireless Designs", Microwaves & RF, October 1996, pp. 136-140.

10. Darioush Agahi, "Monolithic Ceramic Block Combline Bandpass Filters", RF Design, March 1989, pp. 35-41.

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