ARCHITECTURAL CONSIDERATIONS FOR OVERLAYING FAST-PACKET DIGITAL NETWORKS ON HYBRID FIBER/COAX SYSTEMS

Gaylord A. Hart
XEL Communications, Inc.
17101 East Ohio Drive
Aurora, CO 80017-3878

ABSTRACT

This paper describes basic fast-packet switching network attributes, providing background about the operation and advantages of fast-packet technologies. Fast-packet networks achieve their high throughput by lowering the processing overhead required for transport and shifting this responsibility to the terminal equipment at either end of the network. At the same time, these networks achieve great efficiency by statistically multiplexing data from several users onto a single transport path. Frame and cell relay technologies are discussed, with a brief overview of the relative merits of each.

This paper primarily focuses on architectural issues unique to providing fast-packet transport over hybrid fiber/coax (HFC) networks. Various transport topographies and multiplexing methods over the HFC network are described and discussed. Particular attention is given to the return path, which presents unique challenges for achieving high throughput efficiency just where spectrum is currently most limited. Congestion considerations and the effects of BER on system throughput are also discussed.

INTRODUCTION

Traditional telephony network architectures are circuit switched. In this case, an actual circuit or channel of fixed bandwidth is assigned to connect two endpoints in the network. This connection may be set up on a permanent basis (as with leased lines) or on a real-time basis (as with a telephone call). The circuit is dedicated to these endpoints for the duration of the connection, even if no traffic is flowing between them. This requires that additional channels be provided in the network for all other active users, one for each connection.

If the demand for connections exceeds the network's channel capacity, no further connections can be made until a channel is released. This condition is referred to as blocking because access to transport is blocked. To eliminate blocking, enough channels have to be provided for every possible connection in the network under peak load conditions, but this is not economical since average traffic loads are typically far below peak loads. Most switched circuit networks are designed to minimize blocking, but not to eliminate it. Queuing theory is used to statistically derive how many channels are required to maintain a certain probability that blocking will not occur or will occur less than a certain percentage of the time.

The fixed channel bandwidth restriction of circuit switched networks is a disadvantage if data or other services with mixed or varying bandwidth requirements are being transported. The end user may wind up spending more for a large, yet mostly under-utilized transport channel just to meet his peak demand transmission requirements. On the other hand, another user may end up spending less for a

channel which is utilized very efficiently most of the time, but which fails to meet peak demand requirements because the transport channel bandwidth is too small. Since switched circuit networks cannot provide bandwidth on demand they are not very flexible in this regard.

For voice applications, switched circuit networks are efficient. Voice traffic tends to utilize a channel 100% of the connection time, and the bandwidth requirements do not change over time. However, this will not be the case where different services may be offered over the same pipe at different times (each with its own bandwidth requirements) or with data transmission (which tends to be bursty). Channel utilization will likely be less than 100%, and different channel bandwidths will be required depending on the application and the peak to average transmission requirements of that application.

For modern multimedia networks, which must carry voice, video, and data, circuit switched topologies are highly inflexible and inefficient. In a purely digital network, voice, video, and data applications each have unique transport requirements which demand flexibility from the network. Ideally, a single digital pipe to the home should provide all these services. An advanced digital network must somehow provide enough built-in flexibility to provide the unique transport requirements of each of these applications while at the same time using a common transport mechanism for all. Also, an advanced digital network must provide for efficient and graceful evolution of the network and network applications. Fast-packet switched networks will provide the flexibility, efficiency, and low cost necessary to accomplish these objectives.

FAST-PACKET NETWORKS

Packet switched networks have been around for quite some time, typified by protocols such as X.25. Early packet transmission protocols were developed when transport speeds were relatively low and transport errors relatively high. These protocols were developed with an emphasis on error-free transmission, not speed, and typically require an elaborate handshaking whereby each network link has to verify or repeat transmission until error free reception is achieved before a packet can be sent on to the next link, where the process is repeated once again. This handshaking demands considerable link overhead and thus makes transport more difficult at the speeds required today.

New packet protocols (referred to as fast-packet protocols) have been developed which eliminate most of the transport overhead and error processing in the transmission path itself, thus allowing higher transmission throughput. These fast-packet protocols have been developed to take advantage of the virtually error-free transmission of today's networks and the use of low-cost, intelligent terminals at the network endpoints.

When transmission errors are rare, it makes no sense to provide elaborate mechanisms for detecting and correcting errors at each link in the network. Consequently, most fast-packet networks perform little or no error processing at the link layer (OSI transport model layer 2). In many cases, if this processing exists at all within the network, it consists only of detecting errored packets and discarding them. After all, it makes no sense to send an errored packet any further and risk congesting the network.

Nevertheless, providing reliable service requires that error checking and correction be done somewhere. In fast-packet networks this function is typically performed at the transport layer (OSI model layer 4) by intelligent endpoint terminals. Processing errors at the end points eliminates transport overhead and allows the network to transport data quickly and efficiently. The higher throughput of fast-packet networks is accomplished to a high

degree by eliminating lower level layers from the OSI transport model or by transferring the functional responsibility for these lower level layers to higher level layers in the model. Errored packets are simply thrown out by the network without any attempt to correct the error or notify anyone of the error. The end terminals must themselves determine if packets have been received with errors or have been lost and then provide for retransmission.

Packet switched networks further increase throughput by statistically multiplexing several users' digital information onto a single transport channel. This information may consist of voice, video, or data applications, or a combination of these. In fact, the packet payload may itself be packets using a different transport protocol. Unused channel capacity from one user is then allocated to another user on a real-time basis, thus taking advantage of the dynamic nature of channel capacity requirements. Fewer transport channels are now required in the network since each individual channel is more fully utilized. This results in more efficient use of network resources and lower transport costs.

Multiplexing is carried out by partitioning each user's data into small packages, or packets, for transport. Each packet has a strictly defined structure, determined by the particular packet technology in use, containing the user data and additional overhead for performing other critical transport functions: 1) packet boundary delineation, 2) packet routing to the intended destination, 3) congestion control, and 4) error detection. The packet overhead is kept to a minimum since it consumes transport capacity.

Individual packets from each user are buffered at nodes in the network, then time-division multiplexed into the network when transport capacity becomes available. Packets are transported and routed within the network as indivisible units. At the far end, the packet overhead is stripped off, and the payload data bits are reassembled in the correct order before handing off to the end user or terminating application.

Unlike circuit-oriented connections which have fixed data rates, packet-switched transmission has the unique advantage that packets do not inherently have any data rate associated with them. Providing that the transport channel has sufficient bandwidth, the actual channel speed is transparent to the end users. Since packet transport is not locked to a fixed data rate, bandwidth may be allocated to users on demand and without hardware changes. Higher bandwidth is simply allocated by allowing a given user to transport more packets in a given period of time.

Unlike switched-circuit networks which may have connections blocked while circuits are unavailable during heavy traffic loads, packet networks can still accept packets under heavy load conditions. When congestion occurs, a packet network will simply experience greater transmission delay. The nodes in a packet network are connected by fixed size transmission pipes. Under normal conditions, the number of packets entering one of these pipes is not enough to fill the pipe, and packets are allowed to enter the pipe as soon as they are available.

Congestion occurs when more packets are trying to enter the pipe than the pipe has capacity for. When this occurs, packets must be buffered at each node while awaiting a slot in the pipe for transmission. The buffers thus serve to mitigate peak demand by spreading it out over a longer period of time. As the buffers begin to fill up with packets, the delay increases for each packet before transmission to the next node. However, the buffers are also finite in size. If the network experiences extreme congestion (i.e., the packet buffers overflow), significant delays will result because those packets lost in the buffer overflows must now be transmitted again, thus adding more traffic when it is least desired. Once congestion begins to

occur in a packet network, performance tends to degrade rapidly.

For data services, delays may be unimportant. But for video and voice services any such delay will likely be intolerable. However, techniques exist for mitigating congestion and system delay. Dynamic routing can help reduce congestion by balancing the transmission loads of the various links in the network. Packet networks usually also allow packets to be prioritized, which means that more important services or services requiring real-time transport (e.g., television signals) can be given greater access to the network when desired. An alternative approach allows the network to discard less important packets when congestion occurs. Careful network planning is required to control and minimize congestion.

Packet transport may be further characterized as being connection-oriented or connectionless. Regardless of whether a network is connection or connectionless based, switching functions must be provided at each node in the network for routing packets on to the next node.

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 and its routing path. Virtual circuits use a pre-defined routing path for all packets traveling between two network endpoints. These routing paths are defined by logical and physical paths through the network from one endpoint to the other. Since all packets follow the same path through the network, they also arrive at the end node in the same sequence as originally transmitted (providing no packets have been discarded due to errors).

The term 'virtual circuit' has been coined because such a channel appears to the end user very much like that provided by switched circuit networks. Unlike switched circuit networks, however, several virtual circuits (and their packets) can share the same physical channel between any two nodes internal to the network. Virtual circuits may be further characterized as switched virtual circuits (SVC's) or permanent virtual circuits (PVC's). SVC's are analogous to switched circuit network dial-up connections in that the setup and teardown of the connection is done on a demand basis. PVC's are analogous to leased line connections and must similarly be provisioned to establish the connection..

For connectionless networks, no previously established connection between endpoints is required, and thus no pre-defined data path exists through the network. Packets are simply put into the network with a final destination address inserted into the packet at the originating node. Datagram transmission is used within the network, whereby each individual packet (or datagram) is routed through the network independently of any preceding or following packets.

Each successive receiving node examines the packet address to determine if this node is the final destination. If not the final destination, the node takes into account its position in the network and the possible paths to the destination and then routes the packet on to the next node. Each packet may take a different path through the network and may actually arrive at the far end out of sequence. In this case, the receiving terminal is responsible for sequencing the packets in the correct order. Datagram networks allow dynamic fault recovery by routing packets around damaged links and on-the-fly congestion control by sending packets over lower utilized routes in the network.

Fast-packet transport is typified by two technologies: frame relay and cell relay. The primary difference between frame relay and cell relay is that frame relay uses variable length

frames whereas cell relay uses small, fixed length cells. Both have fixed transport overhead associated with each packet. For frame relay, five octets of overhead are required per frame, but the payload may consist of one to 4096 octets. Frame relay typically incurs a lower transport penalty for its overhead since more user data may be transported per packet than with cell relay. On the other hand, variable length frames are more difficult to process because of their variable length. Longer frames, simply because of their size, are also more likely to take errors, which means that a larger frame must then be retransmitted for error recovery.

Cell relay is best typified by ATM, which uses a 53 octet transmission cell comprised of five octets of overhead and 48 octets of payload. Fixed size cells are inherently easier to process because the location of each component in the cell is always the same, which readily allows direct implementation of cell processing in silicon. Because fixed size cells allow greater control and predictability of transmission timing and delays, cell relay is more easily optimized for low delay, high bandwidth applications and some versions are suitable for voice, video, and data.

Other differences between frame relay and cell relay exist as well, with each technology providing unique advantages and disadvantages. Both frame relay and ATM are connection oriented fast-packet technologies supported and defined by industry-wide standards.

HFC PACKET NETWORK ARCHITECTURES

Traditionally, CATV systems have existed as isolated islands receiving signals via satellite. It is common for a large metropolitan area to have several CATV systems operated by different MSO's, each system providing service to a particular geographic section. Because of the broadcast nature of traditional CATV television services, little need existed for building large interconnected CATV networks between these systems. Even within systems, individual headends typically only needed to be connected to transport video signals between them. This was accomplished with AML links or more recently with fiber.

All of this has changed with the introduction of new services, deregulation, and competition to provide existing and new services. Today's HFC network is envisioned to provide in the near future analog television, compressed digital television, HDTV, telephony, data services, internet access, and numerous other interactive and multimedia services. Many of these services cannot be provided without being connected to a larger universe of other networks, indeed, both national and international networks. It will no longer be possible to operate systems as islands.

Nor will it be possible in the long run to operate HFC networks as providing disparate services, each sharing the HFC transport path but essentially using different transmission formats and signaling schemes. The first stage of this evolution, replacing analog signals with digital signals, is already well along. Many operators already simultaneously carry FM radio signals and CD quality digital audio music services. In a short time, compressed digital NTSC services will be common. The benefits of digital technology, including better signal quality and bandwidth efficiency, will accelerate this change.

As more signals on the HFC network go digital, there will be greater economic incentive to process and transport these signals through common channels and equipment. After all, bits are bits, whether they encode voice, video, or data. Such an integrated network will be capable of delivering all these services over a single data stream with unique content and connections for each home. This network can best be built using modern fast-packet technologies.

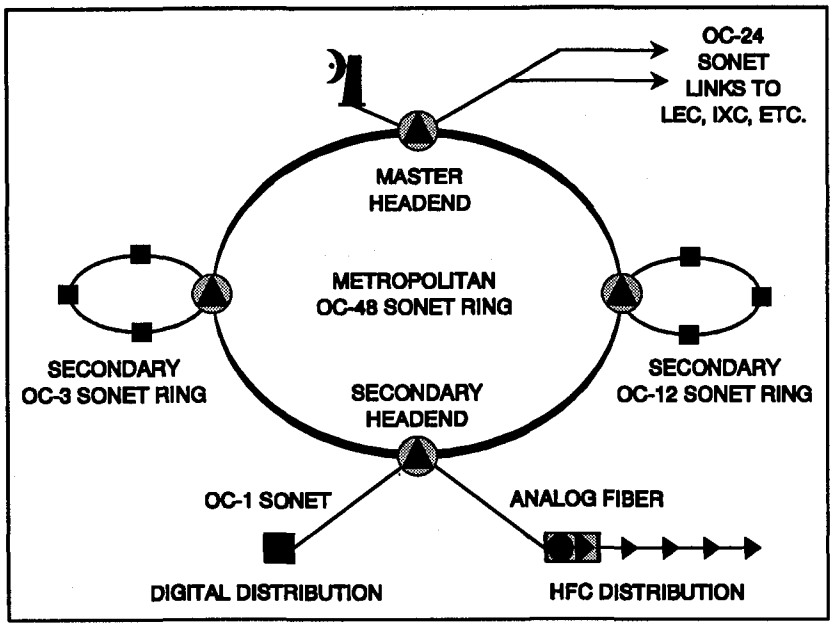

Figure 1--Advanced HFC Metropolitan Network

Advanced Metropolitan HFC Networks

Figure 1 shows a large metropolitan HFC network which consists of a master headend for gathering and processing signals and several secondary headends. All of these headends are connected by a SONET ring which is used to transport digital signals between the headends. Secondary SONET rings may also be present at each headend for local transport of digital signals to businesses or secondary hubs attached to each headend. Each headend will in turn have several HFC nodes to support local distribution of services to residential and small business areas. Point-to-point SONET links may also be used to distribute digital signals to individual businesses or other sites. The overall metropolitan network has additional SONET links which connect to the larger network universe: LEC's, IEC's, and other CATV systems.

Though not ubiquitous, networks of this type have already been built. From a functional standpoint, the overall SONET ring architecture has been used to transport digital television and audio signals from one headend to another, much as with earlier AML systems, and for local ad insertion and distribution. The ring has also been used to support parallel alternate access telephony services and data transport services, though these have typically been operated separately from the CATV network itself. More recently, some operators have been installing telephone switches in headends to accommodate basic telephony services as well, and these switches use the SONET ring for connecting to customers and other carriers.

In most cases, however, these networks are primarily characterized by multiplexing equipment and have been deployed using traditional circuit based technologies. The various services are multiplexed onto and off of the SONET ring using dedicated channels. As these networks migrate toward packet based transmission systems, the overall metropolitan architecture will not change, but more of the SONET transport capacity will be used as the physical transport layer to carry packets between switching nodes. The fundamental changes will take place at the switching nodes, which will be comprised of the regional headends and other hubs where signal distribution takes place. Here packet switches will be added after the SONET multiplexers to

route packets onto and off of the SONET ring and into and out of the local distribution system.

The functions described above already being performed by the circuit based version of this architecture will continue to be performed, but through virtual circuits in the packet network. Television and audio signals will still be distributed between headends, but as packets. Imagine how easy it will be with a packet switched network to distribute local adds to particular headends or even to a particular node on a headend. In the long run, ads may literally be targeted to individual homes.

New services will also be made possible or economical by packet networks, and these too will be integrated into the network as it evolves. For example, video on demand can readily be implemented since video packets may be switched and routed between any two points on the network. Ultimately, this switching capability will extend to the individual subscriber, allowing true virtual channels. Additional services such as basic telephony and internet access will at some time also be carried by the packet network. Since both of these services depend heavily upon switching, the inherent switching capability of the packet network provides an efficient, integrated approach to providing these services.

HFC Headend with Packet Network Overlay

Figure 2 is a block diagram for a headend in the packet switched metropolitan HFC network, showing the additional digital and packet transport components. The traditional analog television and audio distribution components have been left out for simplification, but these would also feed each of the HFC nodes. The SONET multiplexers used to interconnect this headend with other headends and secondary sites in the packet network are shown on the left. These multiplexers allow signals on the SONET ring to be dropped off at this headend and signals originating from this headend to be added onto the ring.

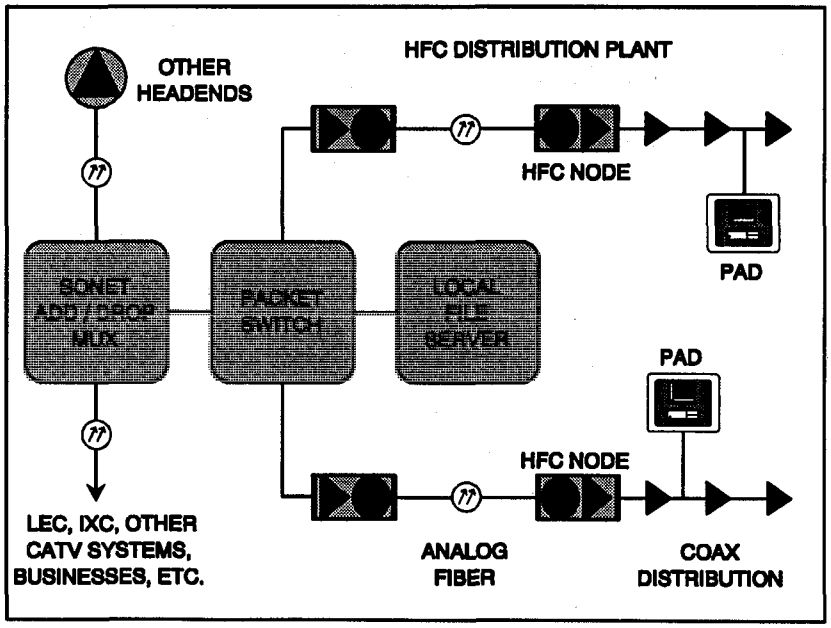

Figure 2 -- HFC Headend with Packet Switch

Since the headend is a major routing node in the network, a packet switch is central to the headend and is shown here with ports feeding the SONET multiplexer, a local file server, and all the HFC nodes on the network. This switch will examine every packet entering one of its ports and route each packet out the appropriate port towards its final destination. For packets being sent out to HFC nodes, this will typically be the last switch the packet sees in the network. For packets coming in from an HFC node, this may be one of many switches the packet will pass through before reaching its final destination.

The file server shown in this headend might support any one of a number of applications. This could be an internet file server or a video file server for video on demand services. In all likelihood, a headend will contain many file servers, each designed to support a specific application. But each of these servers will be connected to the network through the packet switch.

PAD's, otherwise known as Packet Access Devices or Packet Assembler/Disassemblers, are shown on the right hand side of this diagram. These would be located in the subscriber premises, whether residential or business, and would be served by a drop from the coaxial distribution plant just as any other service. For the purpose of this paper, a PAD may be considered any device providing a packet network service to an end user. This could be an internet access port for a computer, an ethernet port connected at the other end to an office for telecommuting, a set-top box providing compressed digital NTSC to a TV, or even a telephone.

HFC Node with Packet Network Overlay

Figure 3 shows the HFC node with packet network overlay in greater detail, though still greatly simplified. A standard HFC distribution architecture is presented. Here you can see the actual forward and return RF paths of the node.

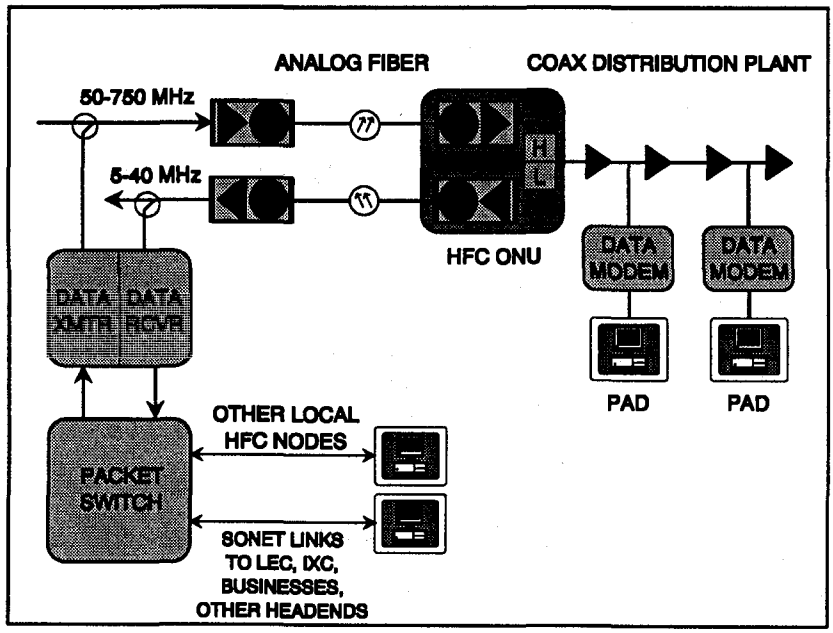

Figure 3 -- HFC Node with Packet Network Overlay

Typical RF frequencies for these paths are used (50-750 MHz downstream; 5-40 MHz upstream). An RF packet transmitter is shown at the headend, where packets to be sent downstream are coupled with other RF signals prior to being fed to the downstream analog laser. An RF packet receiver is also shown at the headend. Upstream RF signals are coupled into this receiver from the analog RF optical receiver in the headend.

Since the packet transmitter and receiver represent different physical paths in the network, each is connected to a separate port on the packet switch. In reality, several headend RF packet transceivers will probably be connected to each node, one for each RF packet channel on the node. Each of these will also be connected to the packet switch. The packet switch shown here is the same as in Figure 2.

Two subscribers are shown connected to this node. At each subscriber's end, an RF data modem serves as the interface between the PAD and the coaxial distribution plant. In all likelihood, this modem will be integral to the PAD itself, and the two together will be considered a subscriber network unit (SNU). A subscriber may have several service specific SNU's in his home, one for each service being taken. One SNU could support telephony, while another might support internet access or compressed digital NTSC television service. Over time, the SNU may actually be a packet gateway to the home, providing several services from a single device. As with other fast-packet networks, the SNU has the responsibility for making sure that errored or lost packets get retransmitted.

If the two subscribers shown in Figure 3 have SNU's that support data connections and they wish to communicate directly via computer, first they must establish a connection. Establishing this connection will consist of defining type of service, gaining access to transport capacity, negotiating bandwidth and quality of service, and setting up the routing path. Once this is accomplished, data packets will flow upstream from one subscriber, enter the packet switch, and then be routed downstream on the same node to the other subscriber. Of course, connections could similarly be made with other terminals anywhere on the network, either on other nodes or nodes served by other headends. Similarly, other services can be supported via the same network.

HFC NODE CONSIDERATIONS

Given the flexibility and transparent transport capability of the HFC architecture, overlaying a basic fast-packet network on the HFC network is largely a matter of installing the appropriate terminal equipment at the headend (switches, file servers, and RF data modems) and customer premises (RF modems, routers, and PAD's). However, deploying advanced, fully integrated packet networks capable of delivering voice, video, and data over the HFC architecture may not be so easy.

The downstream path of the HFC network is well understood and should provide adequate bandwidth and RF transport performance for advanced packet services. The upstream path is less understood, and many technical questions and challenges remain in this direction. The two primary concerns are error performance and transport capacity. Packet networks rely on excellent transmission error performance to help avoid congestion, and advanced integrated packet networks will require significant return spectrum to support the many services envisioned. In the long run, changes to the basic HFC system may be required to ensure adequate error performance and available spectrum in the return path.

Since businesses and homes will use this network, and advanced data services are envisioned which will involve financial transactions or other exchange of confidential

information, security is a critical issue. By virtue of the HFC architecture, all downstream packets are transmitted in a broadcast mode over a given node, and all downstream drops have equal access to this RF data channel. This means confidential information from other users, albeit difficult to access in a meaningful form due to the complexity of the RF and packet transport media, will be present in every business or home on the node.

The return path is less troublesome from a security standpoint. Return path signals are considerably isolated from individual drops by the directional couplers used in taps, and due to the reverse tree and branch structure in the return path not all return signals pass every drop. However, steps must be taken to ensure each subscriber's privacy both in the upstream and downstream direction, whether that subscriber is at a business or a home. Encryption and decryption will be required for some services, if not for all. Encryption can take place at the transport level or at the applications level.

Downstream Path

To realize the full benefits packet switched networks have to offer, efficient statistical multiplexing of packets is necessary. In the downstream direction, multiplexing data packets from several users or services onto a single, high-speed RF data channel at the headend is relatively simple and inexpensive to accomplish. All downstream SNU's then simply monitor this data stream for their intended packets and then extract them.

In most cases, the RF data channel and the packets sent down it will be unique to each HFC node, and each RF data channel will be associated with a specific port, and hence routing path, on the packet switch in the headend. As traffic requirements increase, more RF data channels are simply connected to other ports on the packet switch and added to the HFC node via frequency division multiplexing.

On the other hand, increased traffic demand may be dealt with via fiber division multiplexing by subdividing the HFC node into two smaller optical nodes. In this case, a given RF data channel would then support roughly half the number of subscribers as before, cutting by half the traffic demand on the channel. Of course, a new RF data modem and packet switch port would still be required to support the new node generated by the subdivision. The flexibility of the HFC architecture and the packet network itself allows economical and efficient evolution of the network.

The second requirement to realize the benefits of fast-packet networks is that a low bit error rate (BER) must be maintained to keep the network from filling up with retransmitted packets. The downstream carrier-to-noise ratios and transmission performance of the HFC network are such that very high order modulation schemes may be used to achieve excellent spectral efficiency with very little sacrifice in BER performance. If necessary, tradeoffs may be made between BER performance and modulation spectral efficiency. Of course, forward error correction (FEC) may be used in the RF transport path to improve BER performance for any modulation scheme, but this requires additional overhead bits in the data, which in turn reduces spectral efficiency. In many cases, good BER performance may be achieved without requiring FEC at all in the downstream path.

Upstream Path

The upstream path in an HFC system is subject to noise funneling brought about by the noise summation of all the return legs on a node. This noise is a potential problem for maintaining acceptable bit error rates in the return path. Fortunately, this effect is minimized by modern

fiber architectures, which allow the use of smaller node sizes, and hence less noise to be summed. The return path can also be susceptible to ingress (short-wave signals, CB and ham transmitters, spurs and LO products from consumer devices connected in the home) and impulse noise (appliances, lightning strikes, etc.). Studies have shown that most of these secondary problems occur not in the hard-line portion of the plant, but in the drop or customer premises wiring. These parts of the plant must be brought under control, either by hardening the drop or using filters to limit unwanted signals from entering the plant.

A fast-packet network achieves considerable efficiency by unloading most of the lower level transport error processing and its overhead. Virtually error-free transmission in optical networks has made this possible. When errors are rare, discarding an occasional errored packet and retransmitting it does not consume much bandwidth. But if errors are frequent, retransmitting packets can consume considerable transport capacity. A single bit error requires an entire packet to be retransmitted. HFC networks must provide excellent BER performance if fast-packet networks are to be operated over them.

If necessary, forward error correction (FEC) may be used at the physical layer to reduce the loss of transport capacity due to errored transmissions. However, FEC also incurs a bandwidth penalty as extra bits must be transmitted to accomplish FEC at the receiving terminal. A trade-off exists between bandwidth consumed by retransmitted packets vs. bandwidth consumed by additional FEC bits. The choice of modulation schemes will also make a difference in BER performance. Lower order modulation methods are more robust when faced with poor transmission channels, but these methods are also less spectrally efficient.

One approach to overcoming noise and interference problems is simply to increase transmitter power to a level where good BER performance is achieved. Within reasonable limits, this is perhaps the most economical and spectrally efficient approach, but requires return amplifiers and lasers with very good linearity to withstand the higher signal levels. In any case, the return plant will have to be designed and balanced with as much care and attention paid to performance as with the downstream path. Careful analysis must be done in the system design to optimize BER performance with the least impact on spectral efficiency. Higher overall throughput is the primary goal.

As also pointed out, packet networks derive increased throughput by statistically multiplexing several users onto a single transmission channel. Multiplexing packets efficiently from separate SNU's on the same channel in the upstream direction on an HFC fiber node is not easy to accomplish. Normal multiplexers operate by taking in several data streams and combining them into a single high speed stream. In the upstream direction on an HFC network, each SNU functions independently, but typically transmits on a common return channel with other SNU's. No single device performs the multiplexing function, and in essence all the SNU's on the RF return channel comprise a distributed multiplexer of sorts.

Statistical multiplexing can only be achieved in the HFC return path by providing a mechanism for controlling each SNU's access to the upstream RF channel to prevent collisions which will occur if two or more transmitters become active at the same time. This mechanism must also minimize any overhead penalty imposed by this coordination and minimize any periods of inaccessibility to the channel. Otherwise, channel throughput will be limited.

Unfortunately, due to the HFC architecture, an SNU cannot monitor or coordinate the return path transmission activity of other SNU's on its RF channel. One simple solution is to provide

an independent return RF channel for each SNU (frequency division multiplexing). Unfortunately, this approach is the same as the dedicated channel approach used by switched circuit networks, and none of the benefits of statistical multiplexing are achieved. Dedicated channels are inefficient from a spectral perspective and costly from a hardware perspective. Another mechanism must be used for statistically multiplexing several SNU's onto individual RF channels to realize transmission efficiency and bandwidth on demand.

Another approach is to use CDMA transmission (Code Division Multiple Access, a form of spread spectrum), which allows several users' transmission spectrums to overlap simultaneously. CDMA offers the benefit of noise immunity when high processing gains are used, a desirable characteristic for return path transmission. But CDMA has limitations for this application, especially if any services or users require large data throughput, which is likely to be the case for advanced services.

The return path's available spectrum is too small to allow effective spreading, and hence processing gain, for anything but the narrowest of return data channels. Even so, the entire return path would likely have top be used for CDMA, which would rule out sharing this spectrum with other types of services. Any attempt to allocate a smaller subset of the return spectrum for CDMA signals alone will only make good processing gains more difficult to achieve. Finally, CDMA is still a channel oriented transmission method, and another means would have to be used along with CDMA for statistically multiplexing several SNU's onto individual CDMA channels.

There are four fundamental ways to coordinate multiple SNU transmission access on a single upstream RF channel: time division multiplexing, polling, token passing, and collision detection. Combinations of these methods may be used as well.

With time division multiplexing, each SNU upstream transmitter has ownership of a fixed time-slot in which to transmit. Reference timing and slot assignment are provided by the downstream data path for the service link. The subscriber buffers data until his time-slot becomes available, then bursts this data back to the headend within the allocated time. Non-transmission guardbands must be provided before and after each time slot to ensure no two transmissions overlap, and these guardbands consume some channel transport capacity. The primary disadvantage to this approach, however, is that the subscriber gets a time-slot even if he has no data to transmit, and thus bandwidth may go unused.

The benefits of statistical multiplexing are not possible under this scheme unless dynamic access to all time slots is provided to all SNU's. This would allow slots to be assigned only to SNU's that need transport capacity, and multiple slots could be assigned to SNU's needing extra capacity. This implies that a central, intelligent channel manager is coordinating slot assignment and that this manager has a means of determining each SNU's requirements. This can only be accomplished through some form of polling, which entails additional transmission overhead as the time slot manager queries SNU's for their needs and assigns resources accordingly.

The CATV industry has used device polling for years to coordinate upstream set top converter communications. An upstream channel manager, or host terminal, located in the headend informs each SNU in turn that transmission access has been granted to the upstream RF channel. If the remote SNU has no data to transmit, it informs the host terminal of this fact with a very short message or by failing to transmit within a pre-defined time span. If the SNU has data to send, it then seizes the transmission channel until all its data has been sent or until a defined limit is reached, thus preventing any single SNU from hogging

bandwidth in the return path. Polling itself may require a significant amount of transmission path overhead, depending on the complexity of the polling mechanisms. Due to the nature of polling, transport capacity is consumed by overhead in both the downstream and upstream paths.

Token passing is another means of coordinating individual SNU access to the return transmission channel. In this case, only the current holder of a token (a software authorization to transmit) is allowed to transmit in the return path. Only one token is allowed to be shared among all the SNU's using a given return path channel, and care must be taken to ensure channel recovery if the token is somehow lost. Typically this token is passed from terminal to terminal on a bus, usually in a ring architecture. However, direct communication between SNU's is not possible in an HFC network, so any token passing has to take place through the packet switch in the headend. This implies that the token must either be sent to a channel manager for forwarding to the next SNU or that each SNU has knowledge of the other SNU's on the channel and knows which SNU is to receive the token next. This type of system is complex to implement on an HFC network and will create considerable delays as the return path remains idle while the token is passed. Since the token must be passed up to and down from the headend packet switch, polling may be the more attractive alternative.

Collision detection is yet another mechanism of regulating multiple access to a channel (ethernet is a good example of collision detection). In this case, each terminal on a bus monitors the channel for inactivity. Once silence is detected, any terminal that wishes to transmit may do so. The channel is then monitored for transmission errors to make sure that only one terminal seized the channel. If more than one terminal did transmit, all terminals cease transmission. Each terminal then generates a short, but random time-out period in which it cannot transmit. Usually, one terminal will emerge from its time-out period before the others, and this terminal will then attempt to seize the channel again, and the process repeats until a single terminal gains transmission access. Once a terminal gains this access, it transmits until it has no more data to send or until a defined limit for channel access is reached.

Collision detection is not very efficient for HFC networks. Due to the HFC architecture, collisions cannot be directly detected on the return path by an SNU. The return path can only be monitored for collisions at the headend. This means that collision detection must be mediated through a headend channel manager and communicated back downstream, thus adding mediation overhead and delays to transport access. Because collisions are unpredictable, access delays may be highly variable, and this is not desirable for isochronous applications such as voice or video. The complexity of a collision detection system for an HFC network is also not desirable. Again, some form of polling appears to be a more attractive alternative.

No simple solution exists for multiplexing several remote terminals onto the same return channel with minimal system delay or overhead. But with careful design, channel overhead may be reduced, and delays may be minimized and made acceptable for service transport.

Bandwidth capacity is the final and perhaps greatest concern with the HFC return path. Typical HFC networks have been built with 5-30 MHz return paths. More recently, networks have been built using a 5-40 MHz return. In either case, this provides little spectrum for advanced services, and this may be even more limited since recent studies have shown that the region below 10-15 MHz may be difficult to use due to noise and ingress. In all likelihood, this lack of bandwidth is the greatest barrier to deploying full service fast-packet networks on HFC systems. Five solutions to this bottleneck

are readily apparent, but other solutions exist, as well.

First, mid-split systems may be deployed rather than the current sub-split design. As an example, the upstream band might occupy 5-150 MHz, and the downstream band 200-750 MHz. A 50 MHz guardband would be provided between these. Of course, this takes away from downstream capacity, but with the use of compressed digital NTSC television signals instead of analog NTSC, considerable downstream bandwidth can be freed up.

Second, the coaxial legs coming into the optical network unit at each node may have their individual sub-split return paths kept separate. These would then individually be block converted to a unique frequency and then combined for transport over a single fiber back to the headend. Of course, one of these coaxial legs need not be block converted since its return signals may be transmitted on their original frequencies. For example, an optical node with four coaxial arms radiating from it may block convert one 5-40 MHz return leg to 65-100 MHz, a second leg to 125-160 MHz, and a third leg to 185-220 MHz. One leg would continue to occupy 5-40 MHz. When combined for transmission back to the headend over a single laser, a 25 MHz guardband is provided between each frequency grouping for filtering purposes.

Third, the return path spectrum can be placed above the downstream path spectrum, perhaps occupying 850-1000 MHz. At these frequencies, the necessary diplex filtering between the forward and return paths is more difficult to accomplish, so additional guardband between the two will be required. On the other hand, the noise and ingress performance at these frequencies is much better than in the 5-40 MHz range.

Fourth, a completely separate return path may be added to the coaxial portion of the HFC network (the return path is already separate on the fiber side of the node). In this case, a dual cable approach is taken, one cable for the forward spectrum, one for the reverse. This eliminates the need for diplex filters and guardbands between the forward and reverse paths and allows symmetrical or asymmetrical spectral capacity to be provided. Adding bandwidth to either direction is also easier to accomplish if network demands require it. While attractive from a technical perspective, this approach may not be cost effective.

Fifth, an optical node may be subdivided into two or more optical nodes when either the forward or return path runs out of spectrum. Of course, each new node will still operate over the same frequencies as the original node, but now fewer subscribers will be sharing this spectrum, thus reducing the demand for capacity.

SUMMARY AND CONCLUSIONS

Fast-packet technologies are capable of efficiently delivering advanced digital services, including voice, video, and data, over an integrated transport network. This is made possible by reducing the transport overhead normally associated with error processing and by statistically multiplexing several services and users' data over single transport paths within the network. Such a network can most cost-effectively be implemented using HFC architectures. To support a fast-packet network capable of delivering voice, video, and data, an HFC system must provide significant bandwidth and excellent bit error rate performance, both of which are possible.

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.

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

Pecar, Joseph A., Roger J. O'Connor, and David A. Garbin, *The MGraw-Hill Telecommunications Factbook*, McGraw-Hill, Inc., New York, NY, 1993.

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.