Traffic Management for Highly Interactive Transactional System

Mario P. Vecchi and Michael Adams
Time Warner Cable

Abstract

Management of heterogeneous real-time traffic is a key subject in the development of a full service network for the delivery of multi-media assets. The requirements of MPEG compressed material (currently constant bit rate) are contrasted with the requirements for other multi-media assets such as graphics, data and programs (variable bit rate). Traffic management in the Orlando Full Service Network^TM is described in some detail, and is used to demonstrate how both types of traffic (i.e., compressed video and audio streams multiplexed together with IP data streams) can be accommodated on a single ATM network. The design and management of such a network is discussed in some detail, using the lessons learned from the Orlando trial to illustrate the allocation of traffic across a switched network, from the server complex, over the hybrid fiber/coax broadband access, to the homes.

Introduction

The Time Warner Full Service Network^TM was designed to provide broadband digital interactive services over the hybrid fiber/coax network. It represents a pioneering step in the development of technology and marketing to introduce a full range of new digital services to home consumers, ranging from interactive TV to telephony, data communications and utility metering. The Orlando Full Service Network trial represents the first instance of the concept, and it is designed to support 4000 subscribers, with an expected simultaneous load estimated at 1000 full-motion audio-video streams. The Orlando Full Service Network is based on a switched, distributed architecture, where individual audio and video media streams - coming from a server complex at the head-end - are switched to the respective Home Computing Terminal (HCT) at the customer's residence. The software to control the operation of the media servers and the implementation of the applications forms a complex distributed computing environment that requires its own extensive communications capabilities to exchange application code and other data.

The Orlando Full Service Network uses a single Asynchronous Transfer Mode (ATM) node to support transmission of application code and data, and compressed MPEG audio and video streams. These traffic types have different characteristics and quality of service requirements [1]:

- Application code and data do not have stringent delay requirements, hence retransmission is possible to achieve a reliable network layer, and some packet losses at the physical/link levels are acceptable.
- MPEG audio and video streams are real-time media, hence they have stringent delay requirement and retransmission is prohibited. Also, they do not tolerate packet losses well. Therefore, MPEG audio and video streams must be accurately rate-controlled to prevent buffer overflow in the network and at the MPEG player, and forward correction methods are necessary to improve reliability of network transport.

It is worth noting some of the assumptions and requirements that are specific to the design of the Orlando Full Service Network trial.

- The network is required to support 1000 simultaneous MPEG audio-video streams at approximately 4 Mbit/s per stream. Hence 4 Gbit/s of forward capacity is required for this purpose. Additional capacity is required for application code and data download.

- The network must provide sufficient bandwidth to allow signaling with low-latency in the reverse direction (from the subscriber to the network). However much less reverse bandwidth is required than forward bandwidth and it was recognized that the network would be highly asymmetric in nature. In practice, approximately 95% of the bandwidth is in the forward direction and 5% in the reverse direction.

- The system would require more than one media server for reasons of capacity and fault tolerance.

- The access hybrid fiber/coax network is composed of multiple access sub-networks (called neighborhoods), based on the allocation of approximately 300-500 homes on a common coaxial plant for each fiber node.

The above assumptions and requirements led to an ATM paradigm. To allow for maximum flexibility, the servers need to be connected to the neighborhoods by means of a switching network that can establish connections from any server to any customer. An ATM switching approach was the only viable approach given the multi-gigabit bandwidth requirements.

ATM is a transport, multiplexing and switching technology based on fixed-length cells of 53 bytes that carry both the payload (i.e., the data to be delivered) and the header. The 5-byte header contains addressing information for each cell consisting of a Virtual Path Identifier and Virtual Circuit Identifier (VPI/VCI) pair for each network link. A given connection is established by setting the proper VPI/VCI pair associated with the source and destination. Hence, a given stream of ATM cells can interleave data from many different connections by carrying cells with the appropriate VPI/VCI pairs.

Network Design

Figure 1 is an overview of the Orlando FSN network. The network design was driven by the following key facts:

- ATM signaling standards and implementations for Switched Virtual Connection (SVC) were immature, and a switching network based on Permanent Virtual Connections (PVCs) was chosen.

- The Constant Bit Rate (CBR) MPEG video streams were mapped into Constant Bit Rate CBR ATM connections. Reserved bandwidth was guaranteed for each video/audio connection.

- Internet Protocol (IP) was selected as the data communications network protocol to allow the re-use of existing networking software for the data communications between network endpoints.

The logical routing and addressing of the various information streams are shown schematically in Figure 2. In the forward direction, the information from the server complex consists of multiple OC3 streams, each containing multiple ATM connections. The ATM switch routes each ATM connection to its appropriate output port operating at DS3 rate. The various DS3 streams are modulated at their corresponding carrier frequencies (see later section for RF spectrum assignment discussion), and delivered over the hybrid fiber/coax to the respective HCTs. The HCT at each individual home receives multiple streams of ATM cells at DS3 rates. In order to select the correct information that has been switched to its specific application, the HCT first tunes its demodulator to the correct RF frequency, and then it selects the individual ATM stream based on its VPI/VCI pair.

In the reverse direction, multiple ATM connections at DS1 rate from the HCTs are merged in each Neighborhood Area Node (NAN) using a TDMA scheme to avoid contention. These ATM connections are multiplexed up to DS3 rates, and then routed by the ATM switch to the appropriate server interface at the headend. A Connection Management Module, part of the distributed control system, is responsible for managing the establishment of the connections between the server complex and all the users HCTs.

As described, the network is asymmetric and the following overview will cover a description of the forward and reverse directions. This contrast is worth noting as most communications protocols assume a bi-directional (i.e., symmetric) physical facility. Considerable work was necessary to adapt these protocols to uni-directional facilities used in the Full Service Network.

Forward Direction

Eight media servers (SGI Challenge XLs) are connected to disk vaults using fast and wide SCSI-2 interfaces. The vaults can be configured to provide a total of 1.7 terabytes of media storage capacity or about 500 movies.

The media servers are connected to a GCNS-2000 ATM switch supplied by AT&T. A total of 48 SONET OC3 links provide 5184 Mbit/s of forward payload bandwidth. Each OC3 provides a net payload of approximately 108 Mbit/s (see later section on OC3 for more explanation).

The media servers are also interconnected with a FDDI ring. This ring is used to transfer media content to the disk vaults and to collect billing records from the servers. A separate FDDI ring was used only to expedite development, and the ATM switched network will be used to support all communications needs in future full service networks.

The ATM switch is connected to a bank of QAM-64 modulators supplied by Scientific Atlanta. There are 152 uni-directional DS3 links to provide a total of 5600 Mbit/s of payload capacity from the ATM switch to the neighborhoods. Each DS3 provides 36.86 Mbit/s of payload capacity after ATM and PLCP overheads are taken into account.

The QAM modulator outputs are defined at carrier frequencies from 500-735 MHz spaced at 12 MHz. This allows the outputs to each neighborhood to be combined into a composite RF signal. Conventional analog television channels in the range 50-500 MHz (spaced at 6 MHz) are also provided. The RF spectrum assignment diagram is shown in Figure 3.

The composite RF signal from 50-735 MHz is then used to amplitude modulate a laser. The laser is coupled to a single-mode fiber which takes the signal out to the neighborhood about 10 miles away. At the neighborhood the optical signal is converted back into the RF domain by the Neighborhood Area Node (NAN) and used to feed a coaxial feeder network which passes about 500 subscribers.

The RF signal enters the subscriber residence and feeds the Home Communications Terminal (HCT) or set-top. The HCT is a powerful RISC-based multi-media computing engine with video and audio decompression and extensive graphics capabilities.

Reverse Direction

The HCT transmits a QPSK-modulated signal in the 900-1000 MHz band. Note that this high-split RF for the reverse direction is a novel feature of the Orlando FSN. Reverse carrier frequencies are defined at a spacing of 2.3 MHz. The QPSK channel operates at DS1 rates, providing a net data rate of 1.152 Mbit/s after accounting for ATM overhead. Each reverse channel is slotted using a Time Division Multiple Access (TDMA) scheme. This allows a single reverse channel to be shared among a number of HCTs. The slot assignments are made at the head-end and sent to the HCT over a forward channel such that no more than one HCT is enabled to transmit in any given slot. By default, each HCT has access to a constant bit rate ATM connection with a bandwidth of 46 Kbit/s. More importantly, the access latency of a typical packet is 25 ms worst-case.

The reverse channels from a neighborhood are transported by the coax plant back to the Neighborhood Area Node (NAN). At this point, the reverse spectrum is used to modulate a laser, which is coupled to a single mode fiber. Separate fibers (in the same sheath) are used for the forward and reverse directions.

At the head-end, the optical signal is first converted back into the RF domain and the fed to a bank of QPSK demodulators. These convert to cell-stream into a ATM-format DS3. The mapping is standard but the DS3 is a uni-directional link.

The outputs of the demodulators are combined by seven ATM multiplexers (supplied by Hitachi). A standard, bi-directional ATM-format DS3 is used to connect each multiplexer to the ATM switch.

ATM Addressing

The ATM switch and multiplexers are configured with a mesh of Permanent Virtual Connections (PVCs).

In the forward direction, Virtual Paths are configured from each OC3 port to each DS3 port. This allows the server to address any Forward Application Channel (FAC) by selecting the appropriate VPI. The HCTs tuned to a FAC ignore the VPI and reassemble connections based on VCI. Thus we have a two-level switching hierarchy, VP switching in the ATM switch and VC switching in the neighborhood.

In the reverse direction, the multiplexers perform a traffic aggregation function from DS3 to DS3 rates. A mesh of virtual paths is provisioned from the multiplexer DS3 ports to the server OC3 ports. This allows any HCT to send cells to any server port by using the appropriate VPI.

Connection Management

The allocation of VP and VC identifiers and of network bandwidth is performed by the Connection Manager. The Connection Manager is a distributed set of processes than run on the media servers. In response to an application request for a connection with a given quality of service, the Connection Manager determines a route, allocates connection identifiers and reserves link bandwidth. The connection identifiers are returned to the server and client applications at the media server and HCT respectively.

ATM Mapping

The mapping of higher-level protocols into ATM virtual channels is outlined in Figure 4. Classical IP mapping over ATM [2] is used with some modifications to support uni-directional virtual channels. MPEG system layer streams are mapped directly into AAL-5, with a separate virtual channels for audio and video. The emission rate of each virtual channel is carefully rate-controlled in the server interface to minimize queue lengths in network buffers [3]. At present, no isochronous streams are supported.

Bandwidth Allocation

From its inception, the Orlando FSN was designed to serve 4000 subscribers with a maximum load of 25% of those subscribers simultaneously accessing interactive, full-motion services, generating a capacity requirement of 1000 video streams. Investigation of available MPEG compression technologies convinced us that a 3.5 Mbit/s compressed video data rate was required for high-quality pictures from a movie source (as good or better than S-VHS). The audio compression rate was chosen to be 384 Kbit/s to provide high quality stereo and matrixed surround sound. This yields to a total of 3.949 Mbit/s per stream after allowing for system layer overhead. (This is rounded to 4 Mbit/s per stream for the calculations in this paper). In addition, any single neighborhood was designed to support a peak load generated by 40% of subscribers simultaneously accessing interactive, full-motion services.

The entire OC3 bandwidth cannot be used for movies on demand because some bandwidth must be reserved for messaging and application loading. (Recall that each audio-video stream requires approximately 4 Mbit/s for audio and video).

In the current design, 4160 Mbit/s, of the 5184 Mbit/s of total OC3 capacity, is allocated for up to 1040 Audio-Video (AV) streams. The remaining 1024 Mbit/s is reserved for messaging and application download. This is almost 20% of total OC3 bandwidth. Why is so much bandwidth reserved for this purpose?

DS3 Bandwidth Allocation

The answer is apparent if we work backwards from the DS3 Forward Application Channels (FAC). Each FAC contains a number of sub-channels which are actually ATM Virtual Channels (VCs). These Virtual Channels are created dynamically by the Connection Manager and can carry audio, video or data traffic. This is illustrated in Figure 5. (Note that for simplicity the audio and video channels are grouped together (AV1 or AV2) in the diagram). The diagram illustrates 6 FACs but in reality a neighborhood will typically carry 9 or more FACs (depending on subscriber count).

Each FAC carries two IP channels, Fast IP and Slow IP. (These are grouped and labeled simply 'IP' in the diagram). A FAC can also carry a number of Audio/Video channel-pairs (labeled AV1 or AV2). There are two video rates:

- AV1 requires approximately 4 Mbit/s to deliver 24 fps movie material and 384 Kbit/s audio.
- AV2 requires approximately 6 Mbit/s to deliver 30 fps video material and 384 Kbit/s audio.

IP Channels

Two IP sub-channels exist on each Forward Application Channel; a 'slow' IP channel of 0.714 Mbit/s and a 'fast' IP channel of 8 Mbit/s. Two IP channels are used so that the HCT can mask traffic on the Fast IP channel during normal operation thus saving CPU bandwidth that would otherwise be used for unnecessary receive processing.

Slow IP is used for general messaging between the servers and HCT. A 0.714 Mbit/s channel has proved to be more than adequate for this purpose.

Fast IP is used to download application code to the HCT when required; for example, when a subscriber enters the movie-on-demand venue. Timely application download is ensured by allocating sufficient bandwidth for the Fast IP channel (8 Mbit/s or 23.6% of the total FAC bandwidth). If a typical application contains 8 Mbits (or 1 MByte) of data, it will take approximately 1 second to download.

Two or more HCTs, tuned to the same FAC, may request application download at the same time. In this case, the Fast IP channel is shared between them and the download time is proportionately longer.

Each FAC has its own IP channels. Early in the design it was thought that a single FAC channel could be shared by all HCTs for download. Unfortunately, this is not possible because it would add significantly to the HCT response time due to the time needed to re-tune. For example, if an HCT were tuned to FAC3 and playing a movie, when the subscriber makes a request for home shopping, the HCT would have to re-tune to the download FAC to request the home shopping application, and then re-tune to FAC 3 to receive the home shopping AV stream. The re-tune time is approximately 200 ms, thus 400 ms of latency would be added. Worse still, during re-tune and application download the HCT can display **only** locally generated graphics and audio. This would place an unacceptable burden on the application developer who has to keep the user interested for typically 1.4 seconds with locally generated cover.

For these reasons, it was decided to allocate bandwidth in every FAC for Fast IP as well as slow IP. At 8 Mbit/s the Fast IP Virtual Channel consumes as much bandwidth as 2 AV1 (movie delivery) streams. When this decision was made, network capacity was increased to carry 1000 AV1 streams after taking IP overhead into account. Each FAC can carry 7 AV1 streams and so DS3 links were added to provide a total of 152 FACs. These provide capacity to carry 1064 (152 x 7) AV1 streams.

OC3 Bandwidth

As previously stated, sufficient OC3 bandwidth must be reserved to support 1000 AV streams, each at 4 Mbit/s. This requires 4000 Mbit/s of bandwidth.

Each OC3 can carry 108 Mbit/s and there are 48 OC3 links, giving a total of 5184 Mbit/s. Thus the bandwidth available for IP is **1184** Mbit/s (5184 - 4000).

However, the total bandwidth required to support all 152 FACs for IP is **1325** Mbit/s given 8.714 Mbit/s per forward application channel. (The reason for this discrepancy is that there is more FAC bandwidth than OC3 bandwidth to allow for per-neighborhood peak loads of 40%).

To solve this discrepancy, we could add more OC3 bandwidth, increasing cost. Instead, an alternative approach was chosen; to use statistical multiplexing at the OC3 links and over-subscribe the available OC3 bandwidth. To do this effectively, the IP traffic must be **flow** controlled to prevent packet loss, and the IP channels must be **rate** controlled to prevent the IP traffic from exceeding the allowed connection bandwidth.

IP Traffic Characteristics

The traffic on the Fast and Slow IP channels is bursty in nature. In particular, the Fast IP channel is only used while a HCT is loading an application. If we take a community of 4000 HCTs, the probability of a significant fraction of them simultaneously loading an application is low.

Rate Control

Each OC3 card has a number of rate-queues which are implemented in the OC3 interface hardware. A rate queue allows ATM cells to be emitted at a constant rate. There are 5 rate queues defined in each OC3 card as follows (in rate order):

- 8 Mbit/s for Fast IP
- 5.56 Mbit/s for 30 fps MPEG compressed material (video)
- 3.56 Mbit/s for 24 fps MPEG compressed material (movies)
- 0.714 Mbit/s for Slow IP
- 0.393 Mbit/s for MPEG audio (MusiCAM)

The rate queues also have high or low priority. All high priority rate queues are serviced before the low priority rate queues. Within a priority class, rate queues are serviced in a round-robin scheme to provide virtual channels which have a constant cell rate.

The IP traffic is tagged as a low priority and the AV traffic as high priority. The AV traffic must be accurately rate-controlled (within about 1%) to prevent buffer overflow or under-flow in the HCT. However, the IP traffic need only be rate-limited; as long as the IP rate does not exceed the bandwidth allocated in the FAC there is no impact to the service except for greater latency in application download.

Flow Control

The MPEG compressed streams are constant bit-rate, which means that their bandwidth allocation is always fully utilized by a constant stream of data. In contrast, the Fast IP and Slow IP channels are variable bit-rate. In fact a particular Fast IP channel may carry no traffic at all for minutes or even hours.

Statistical Multiplexing takes advantage of the averaging effect of many bursty traffic sources when sharing a constant bit-rate link. However, a flow-control mechanism is required to prevent packet loss. This can be seen if the following scenario is considered:

Assume 20 bursty traffic sources each with a maximum bandwidth of 10 Mbit/s are sharing a physical link with a maximum capacity of 100 Mbit/s. On average, less than 10 sources are active at the same time and there is no packet loss. However, when more than 10 sources become active, some packets must be discarded or buffered. Buffering is only a very short term solution at these bit-rates. 1 Mbit of buffering will only survive for 10 ms if all 20 traffic sources become active. If the traffic sources re-transmit discarded packets this effectively increases their offered load to the link. The link can become flooded with re-transmissions and little or no real traffic (i.e. packets carrying useful payload) will successfully traverse the link.

If we plot the above scenario as a graph, of effective throughput versus offered load, it would appear somewhat like the curve in Figure 6 labeled 'packet loss' [4]. As the offered load increases, throughput increases linearly until the link becomes congested. At this point if the load is increased further, the effective throughput falls dramatically as most of the link bandwidth is occupied by re-transmissions.

If sufficient buffering is available in conjunction with a flow control mechanism, the curve in Figure 6 labeled 'no packet loss' would be observed. In this case, the source waits for acknowledgment before sending more traffic and thus regulates its output. The FSN design allocates sufficient buffering at the server to ensure the 'no packet loss' case.

The effect of flow control is to reduce the rate of each source according to the total load on the shared link. There are a number of algorithms that have been developed which attempt to keep the offered load at, or below, the 'knee' shown in the graph. In particular, the TCP 'slow-start' algorithm adjusts the TCP window size according to the success or failure to transfer packets over the link. On startup, the window size is set to an initial value and is increased as acknowledgments are received indicating that the packets were successfully received at the far end. If a negative acknowledgment is received or a time-out occurs waiting for acknowledgment, the window size is reduced.

Using the same numbers as before, 1325 Mbit/s of IP source bandwidth is allocated to 1184 Mbit/s of link capacity, over-subscribing it by 12%. This is very conservative, and as experience is gained with the IP traffic profile, it is expected that much higher statistical multiplexing gain can be used with no perceptible impact to the system response time.

The TCP slow-start mechanism has been employed in the FSN design. Test results show the same download time as TCP implementations without the slow-start algorithm.

It is essential for this scheme to operate successfully that IP traffic cannot increase to beyond its rate allocation and contend with AV traffic. To help ensure this the IP traffic uses a low-priority rate queue. Test results show that the TCP traffic allocated to the low-priority rate queue does not interfere with AV traffic in the high-priority rate-queue.

Conclusions

The Full Service Network uses a single ATM network to support transmission of application code and data and compressed MPEG audio and video streams. These have very different traffic profiles and quality of service requirements:

- Application code and data must delivered in a timely fashion, but some packet loss can be tolerated if a reliable transport layer is employed. Delay requirement is not stringent.
- MPEG audio and video streams do not tolerate packet loss well and must be precisely rate-controlled to prevent buffer overflow in the network and at the MPEG player. Stringent delay requirement and retransmission is prohibited.

A combination of hardware-based rate control and software-based flow control was successfully employed to provide statistical sharing of bandwidth to a large number of subscribers. Rate-control was also used to prevent output buffer overflow in the ATM switch when many AV streams are combined.

In future, larger networks with more general topology the techniques described may not be adequate to regulate the flow of IP traffic, and techniques such as fast buffer reservation [6] may be required. In addition, techniques to bound delay and jitter of multimedia traffic [7] may be required for larger networks.

Acknowledgments

The work described in this paper was made possible with the help of many colleagues in Time Warner, AT&T, C-Cube, Scientific Atlanta and Silicon Graphics. In particular, Louis Williamson and Tom Speeter deserve special credit for their contribution.

References

[1] M. Wernik, O. Aboul-Magd and H. Gilbert, "Traffic management for B-ISDN Services", IEEE Network, September 1992.

[2] P. Boyer, F. Guillemin, M. Servel and J.P. Coudreuse, "Spacing Cells Protects and Enhances Utilization of ATM Network Links", IEEE Network, September 1992.

[3] M. Laubach, "Classical IP and ARP over ATM", RFC1577, January 1994.

[4] V. Jacobsen, "Congestion Avoidance and Control" Proc. ACM SIGCOMM 88, August 1988.

[5] R. Jain, "A Timeout-Based Congestion Control Scheme for Window Flow-Controlled Networks", IEEE Journal On Selected Areas in Communications, Vol. SAC-4, No. 7, October 1986.

[6] J. Turner, "Managing Bandwidth in ATM Networks with Bursty Traffic", IEEE Network, September 1992.

[7] L. Trajkovic and S. Golestani, "Congestion Control for Multimedia Services", IEEE Network, September 1992.

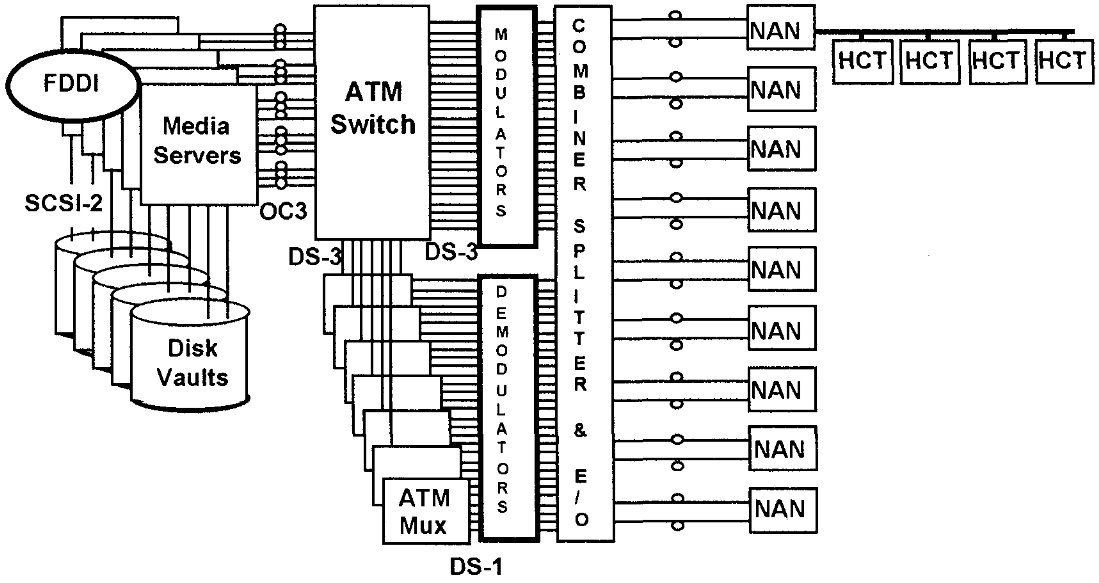


Figure 1. Full Service Network Overview

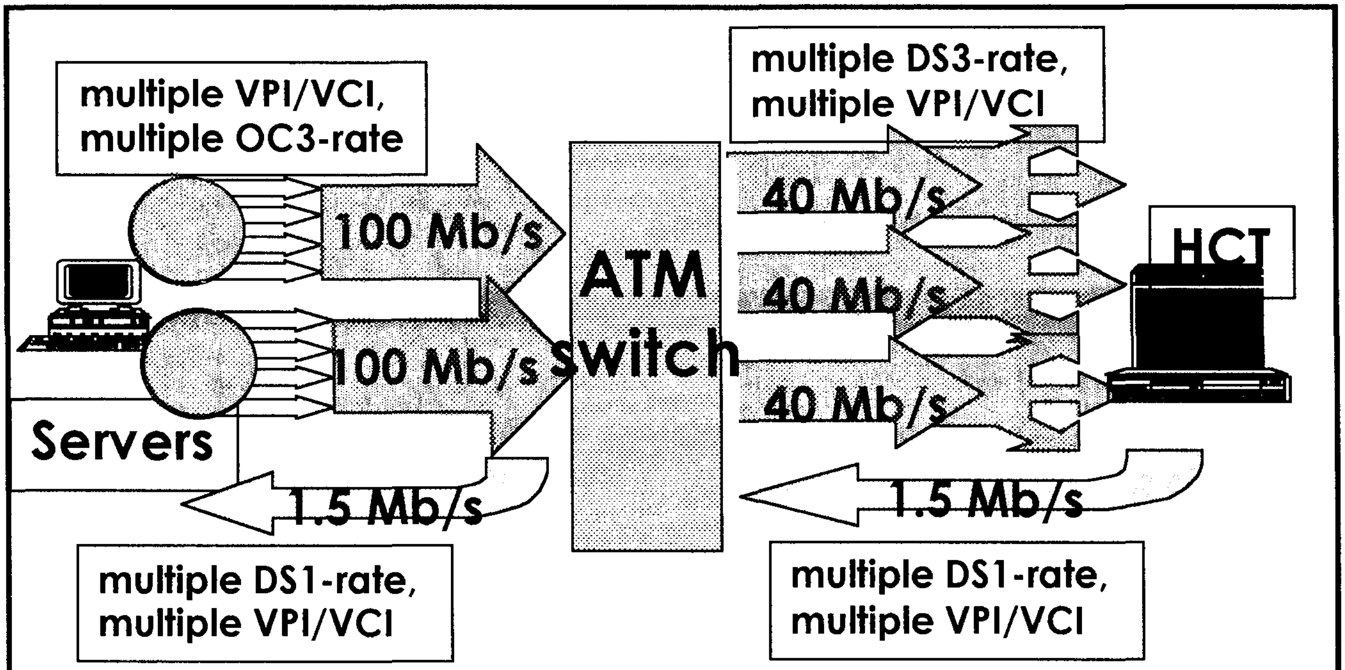

multiple VPI/VCI, multiple OC3-rate
100 Mb/s
100 Mb/s
Servers
1.5 Mb/s
multiple DS1-rate, multiple VPI/VCI
ATM switch
multiple DS3-rate, multiple VPI/VCI
40 Mb/s
40 Mb/s
40 Mb/s
HCT
1.5 Mb/s
multiple DS1-rate, multiple VPI/VCI

Figure 2: Logical ATM Connectivity

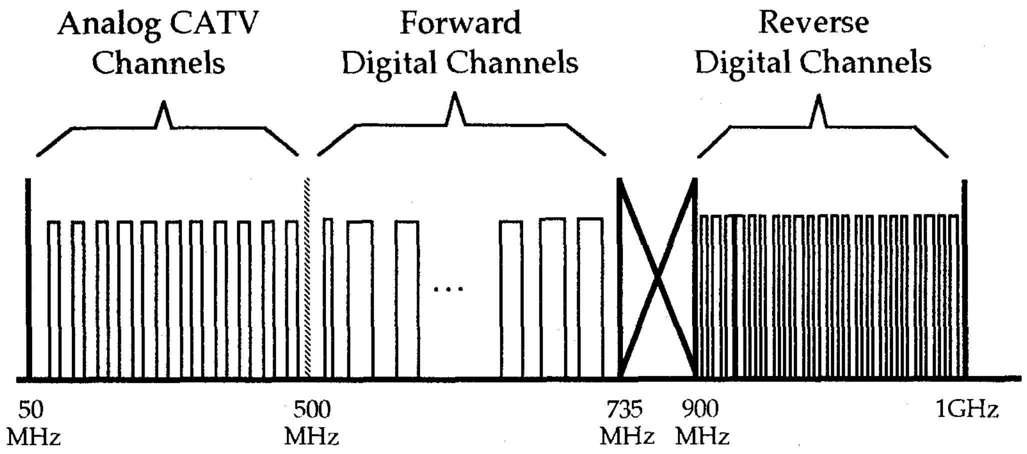
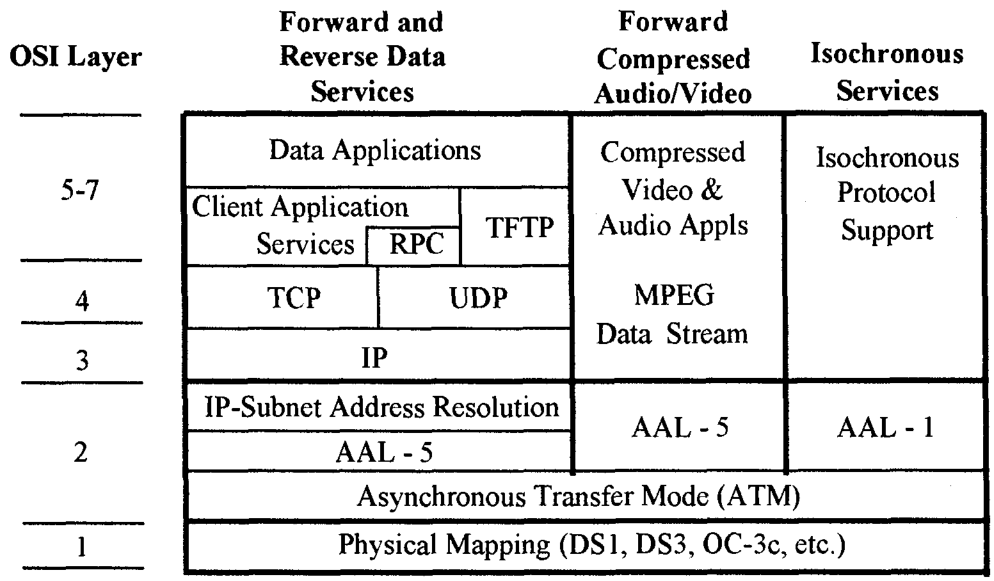


Figure 3. Allocation of Analog and Digital Spectrum



OSI Layer	Forward and Reverse Data Services		Forward Compressed Audio/Video	Isochronous Services
5-7	Data Applications		Compressed Video & Audio Appls	Isochronous Protocol Support
	Client Application Services RPC	TFTP		
4	TCP	UDP	MPEG Data Stream	
3	IP			
2	IP-Subnet Address Resolution		AAL - 5	AAL - 1
	AAL - 5			
	Asynchronous Transfer Mode (ATM)			
1	Physical Mapping (DS1, DS3, OC-3c, etc.)			

Figure 4. Mapping of Higher-Level protocols into ATM

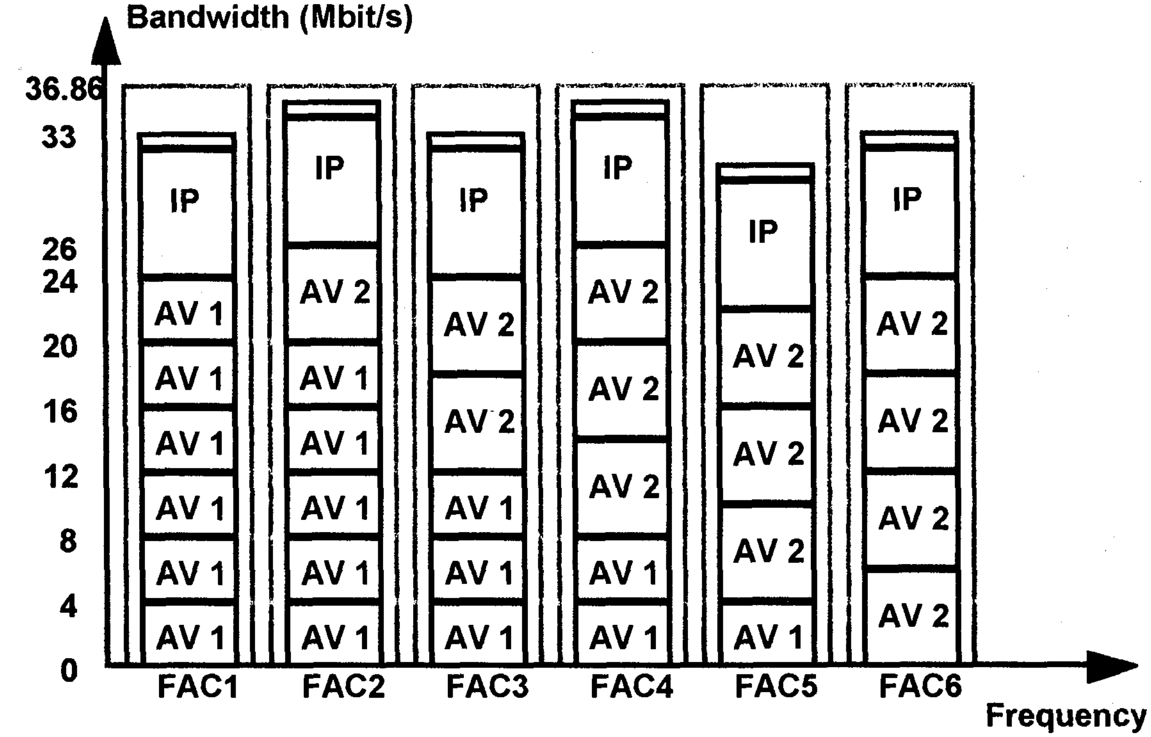


Figure 5. Bandwidth Allocation within Forward Application Channels

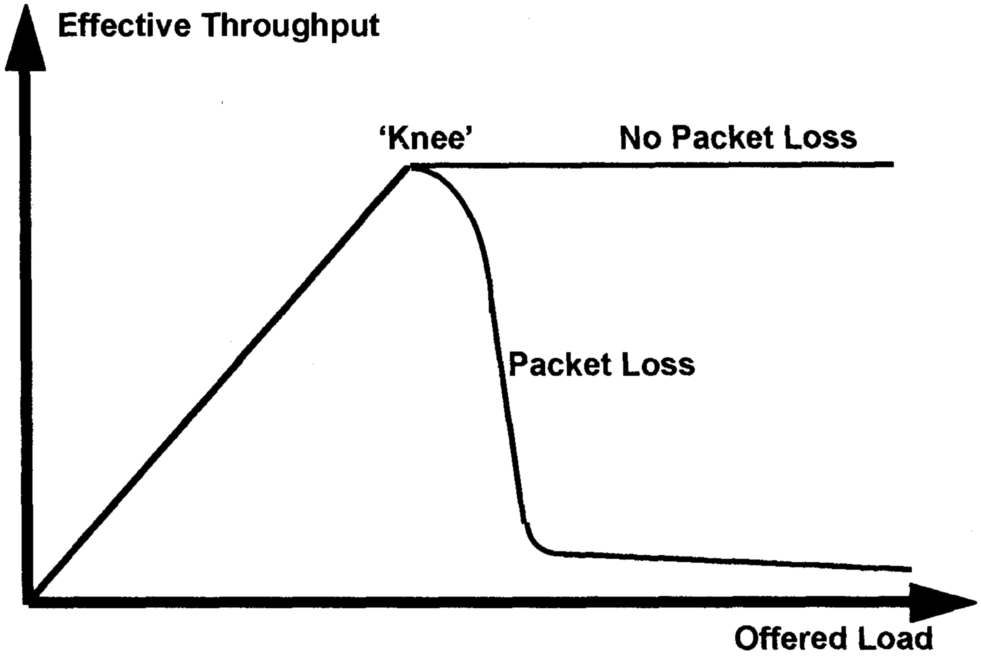


Figure 6. Effect of multiplexing many sources on a shared link