# Multimedia over ATM: Progress, Status and Future

**Bing Zheng** and **Mohammed Atiquzzaman**
Department of Electrical and Computer Engineering
The University of Dayton, Dayton, Ohio 45469-0226.
E-mail: *zhengbin@flyernet.udayton.edu, atiq@engr.udayton.edu*

## Abstract

Multimedia is an emerging service which integrates voice, video and data in the same service. With the progress made in high speed large capacity multimedia servers, high speed networks, cost effective QoS, acceptable service category and cost effective set top boxes, it is currently possible to carry multimedia over high speed networks cost effectively and effeciently. This paper surveys the progress made and the future of efficiently carrying multimedia over ATM networks.

## 1 Introduction

The Asynchronous Transfer Mode (ATM) network which provides the high speed and large capacity trunk interconnection of the various networks is very suitable for transporting multimedia, with the high speed video server technology, it is expected that applications such as video on demand and video conferencing will be the most active services over ATM networks. This is evidenced by the strong interest in standardizing of carrying multimedia over ATM networks. ATM Forum has already released its standard on transporting compressed video over ATM backbone networks [1] and IEEE Communication Magazine has contributed a feature topic on VoD [2]. The following four basic elements are crucial in realizing a VoD service.

- *Video Server:* a high capacity video server which can be randomly accessed at a high speed as video sources;
- *High speed Multiplexing Network:* a high speed multiplexing network which can serve as a transport channel to support a large number of users;
- *Cost Effective Set Top Boxes:* a cost effective set top box which can perform interactive access and trick modes and also serves as the video decoder;
- *Cost Effective Service Type:* a cost effective service which can offer acceptable quality of service.

The concept of VoD was presented more ten years ago [3] in which transporting video over fiber was discussed. In [4], the author discussed the system architecture for a large scale VoD based on the broadband telecommunication network.

The objective of this paper is to review the progress made and the future of running VoD over ATM networks. Here we discussed the basics of ATM, the interactive video, and the suitability of the various ATM service categories for carrying interactive multimedia. Strategies and results of operating interactive multimedia over ATM networks have been presented also.

In Section 2 we discuss ATM networks and the service categories followed by VoD and video compression techniques in Sections 3 and 4. The suitability of ATM for video transmission is presented in Section 5 followed by recent studies on carrying highly interactive VoD over the ABR service of the ATM network. Our concluding remarks are given in Section 7.

## 2 ATM Technology

ATM network is based on statistical multiplexing which allows carrying real-time and non real-time services in the same network. There are four reasons which make implementing ATM advantageous over other networks.

- Because of the difference in the nature of the traffic, currently we use separate networks to carry voice, video and data. With the standardization of the B-ISDN, we can expect to form a universal network with a collection of ATM technologies to integrate the three type of traffic into one.
- Because of the difference in the network technologies, interconnecting the LANs and WANs to form the large scale nation wide and even international networks is complex and difficult. The standardized ATM technology is a practical solution which can be used as the basis of LAN, WAN and even as the backbone in LAN/MAN interconnection.
- For the ease of developing applications, a world wide standard is needed to allow interoperability of networks, regardless of the type of information



0-8186-9014-3/98 $10.00 © 1998 IEEE

Table 1: Characteristics of the three service types.BW for *Bandwidth*, Param.Negot. for *Parameters negotiated when connection setup*, Re for *Renegotiation*, Gu for *Guarantee*, Uti for *Utilization*, and v- for *Very*

| Serv Type | Param. Negot. | BW. Re. | BW. Gu. | QoS Gu. | BW. Uti | Cost |
|---|---|---|---|---|---|---|
| CBR | PCR | No | Yes | Yes | v-low | v-high |
| VBR | PCR,SCR | No | Yes | Yes | low | high |
| ABR | MCR,PCR ICR,etc | Yes | MCR | acpt | v-high | v-low |

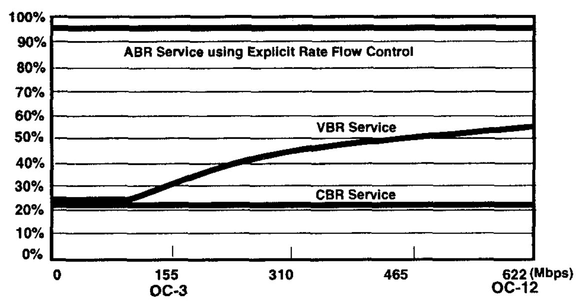


Figure 1: Bandwidth utilization for transmission of MPEG video over the three service types of ATM.

and the type of network. ATM offers a single standard for all types of applications.

- With ATM having several service types to satisfy different application and requirement of service quality, ATM is an ideal solution for multimedia over high speed networks.

### 2.1 ATM Service Categories

Currently, ATM Forum has defined and standardized four type of service: Constant Bit Rate (CBR), Variable Bit Rate (VBR) including real time VBR and non-real time VBR, Available Bit Rate (ABR) and Unspecified Bit Rate (UBR). Their specifications and characteristics are shown in Table 1.

Figure 1 shows typical trunk utilization for MPEG video transported over the three ATM service categories [5]. It is shown that ABR service has much higher network bandwidth utilization.

### 2.2 Quality of Service

One of the strong features of ATM is its quality of service guarantees like delay, cell loss and bandwidth. Real-time voice and video (and applications such as VoD) require stringent delay and bandwidth guarantees for acceptable picture quality which makes ATM very suitable for carrying multimedia traffic.

## 3 Video on Demand

Video on demand (VoD) enables users (clients) to request a video through a network from a video server with a large scale video data base. Generally, the VoD system consists of: the video server, the network, and the client. A VoD system running over an ATM network is shown as Figure 2. Depending on the usage pattern, VoD systems can be classified into three categories:

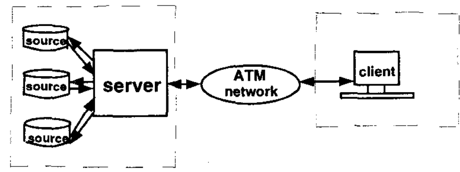


Figure 2: A video on demand system.

- Near Video on Demand (NVoD): In NVoD, the client/user sends a request to the video server and waits for the server to broadcast the requested video at a later time. During the playback, the user is a passive viewer. This is somewhat like a cable TV, the difference is that the client/user can request what he/she wants.

- Interactive Video on Demand (IVoD): The IVoD has VCR-like functions between the server and client. The client/user sends request to the video server, and the client receives the video from the server and starts the playback. The client/user can performance the fastforward (FFW), fastbackward (FBW) or pause operation. There are two main issues to be solved in IVoD systems. The first is the client/user buffer requirement. Since the normal playback and FFW/FBW have different data consuming rates, the buffer needs to be properly dimensioned to prevent underflow or overflow at the client/user buffer. The second issue is the determination of the required network bandwidth and network parameters.

- Highly Interactive Video on Demand (HIVoD): HIVoD has a much higher level of interactivity between the server and client/user. The client/user frequently performs FFW/FBW operation which results in two consequence: the first is that the client/user buffer will be depleted/refreshed frequently. The second consequence is that the server will frequently ask for different bandwidths in order to minimize the startup delay.

## 4 Video Compression

Video is a set of related pictures whose quality is decided by the number of pixels, in uncompressed case, video signals contain a significant amount of redundancy. The first type is the statistical redundancy

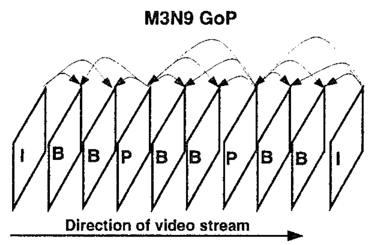


Figure 3: Group of pictures (GoP) in MPEG-2.

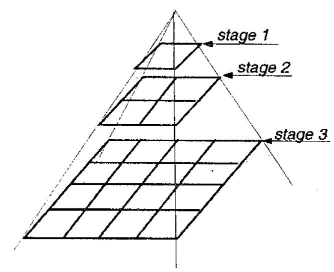


Figure 4: Hierarchical image representative in JPEG.

where each frame has information which are related to its adjacent frames and can be borrowed from others. The second type is the perceptual redundancy which is coming from the signal which can not be perceived by human eyes. Therefore, it is possible for us to reduce the bit rate of video by compressing the video with little or no effect on video quality. Currently, there are two widely used compression techniques: MPEG and JPEG.

## 4.1 MPEG Digital Video Compression Standard

MPEG (Moving Picture Experts Group) compression is accomplished in four basic steps: Pre-processing to filter out perceptual superfluous information; Temporal Prediction and Motion Compensation takes advantages of the fact that each frame in a video may be similar to the preceding and following frames; Finally, the Quantization Coding converts the coefficients of the Discrete Cosine Transfer (DCT) of the residual difference between frames into more compact digital representation. Therefore, MPEG frames are divided into three type: an I frame (or intra-frame) acts as a reference for restoring subsequent frames. I frames have the highest data rates. A P frame (or predicting frame) is predicted from the nearest I or P frames; a B frame (or bidirectional frame) uses the predicting information from its adjacent I and/or P frame. MPEG frames are arranged in groups called Group of Pictures (GoP) which is represented by MmNn meaning that there are $n$ frames in a GoP, of which there is one I frame at the beginning of the GoP, following by m-1 B frames between I and P frames (see Figure 3). For example, a M3N9 GoP has the structure IBBPBBPBB. Typically, MPEG technology has a very high compression ratio ranging from 20 to 100 for video.

## 4.2 JPEG Digital Video Compression Standards

Another widely accepted video compression technology is JPEG (Joint Photographic Expert Group) which has been developed for color still-image compression. Similar to MPEG, JPEG also employs lossy compression scheme to filter out superfluous information in images, encode the residual information based on DCT. Unlike MPEG, the JPEG standard has been developed for still-photographic image compression; so, there is no frame group structure. Instead, JPEG employs hierarchical image representation in which a single JPEG still-image is composed by superposition of several stages of images with different resolution, from top to bottom. The resolution is doubled at every stage, and each stage encodes only its difference from the upper one.

# 5 ATM for Video Transmission

The following sections describe techniques to transport video over the different services of an ATM network.

## 5.1 Video over ATM with CBR

A CBR service requires specifying the Peak Cell Rate (PCR) of the connection and doesn't allow bandwidth renegotiation during the connection. If we open a CBR circuit for video, the PCR should be equal to the highest bit rate of the video stream. However, because of the bursty nature of the video stream, the bandwidth utilization for CBR is quite low. The service cost is very high. The relationship between transmission rate, the client buffer size to avoid underflow or overflow for CBR has been reported in [6].

## 5.2 Video over ATM with VBR

To reduce the service cost and improve the network bandwidth utilization, it is possible to transport video over the VBR service category. In the VBR service category, a connection is specified by a PCR and a Sustainable Cell Rate (SCR). All bandwidth negotiations have to be done during the connection set up phase. Previous work on video over VBR have dealt with the following issues:

*Traffic Shaping and Rate Control*: Traffic shaping and rate control are two related aspects whose aim is to smooth the video stream. Rate control is used to reduce the burstiness of the transmission by controlling the transmission rate. Unfortunately, this results in jitter due to the expansion of the transmission time. Buffer is usually used as compensator by accumulating the spread data. In [7, 8], the authors discussed the traffic shaping of VBR traffic with the leaky bucket model and rate control.

*Bandwidth Allocation/Management*: In [9, 10], the authors investigated the feedback control scheme in bandwidth allocation and management. It was shown that feedback control is effective in transmitting video over an ATM network. For stored video, allocation of bandwidth by predicting future requirement based on current parameters is discussed in [11]

*Congestion Control*: Congestion occurs when the sum of the bandwidths of a number of video streams exceed the total capacity of the network. Congestion degrades the performance of the network and affects the QoS of the connections. The feedback based congestion control is to control outgoing rate based on feedbacking signal, the feedback signal can be either the rate parameter or the buffer occupancy, this scheme was discussed in [12, 13].

*Buffer and Memory Requirements*: A buffer with proper size is used to soomth the fluctuation of video stream. If it is too small, there will be overflow/underflow; too large a buffer size increases the cost of the client. Buffer and memory requirements are discussed in [14, 6].

*Source Model and Behaviors*: A proper video source model is essential for the analysis of VoD systems, bandwidth allocation, etc. The video models fall into three categories: the first is to use the Autoregressive Processes [15], the second uses Markov chains [16] and the third uses a combination of the above two methods [17].

*Coding Technology*: In [18], the author studied the relationship between the coding method and the traffic characteristics of video. It was found that the traffic characteristics of a video stream is very sensitive to the coding algorithm used.

*Quality Control*: Quality control mechanism for video over ATM with VBR service has been discussed in [19].

### 5.3 Video over ATM with ABR

Unlike CBR and VBR connections, when an ABR connection is set up, the server negotiates Minimum Cell Rate (MCR), Peak Cell Rate (PCR) and a number of other parameters. During transmission, only the MCR is guaranteed by the network. The sources can ask for bandwidth allocation between the MCR and PCR.

Work on transporting video over ABR is at its infancy. In [12], closed-loop feedback was applied to control congestion, and a non-zero MCR was set to obtain the guaranteed service. In [20], the authors discussed the explicit rate feedback control scheme to transport compressed video, and investigated the effect of network round trip time on the efficiency of the feedback scheme.

## 6 Video on Demand over ABR

Until now, the work on VoD over ABR is very limited. Because, the ABR service does not guarantee a high bandwidth to a video server and the quality of service for video is sensitive to the delay, the *buffering requirements* and *network congestion* are two important aspects which must be investigated. The first step toward these problems is presented in [21, 22]. In [21], a GoP bandwidth negotiation scheme was proposed, and the buffering requirements and QoS under different network congestion status were discussed. In [22], a fast buffer fillup scheme was presented to reduce the network bandwidth requirements and to simplify the bandwidth allocation management. Optimum dimensioning of client buffer and reducing the startup delay corresponding to different server/client interactivity level were also investigated.

### 6.1 Buffer Requirement

Both for IVoD and HIVoD, buffering is an important issue because of the burstiness of video and the cost of hardware. For interactive VoD with VCR-like functionality, we can illustrate the operation mode of a client by the state diagram in Figure 5. By using the Real Time Dynamic Equilation (RTDE) and Markov chain analysis [22], we obtain the minimum client buffer size for fast buffer fill-up scheme as:

$$C^{C}_{\min} = 2(k-1)T_d E[\lambda] + T_f(E[\beta_I] - E[\beta]) \qquad (1)$$

where $E[\lambda]$ is the expected rate of video, $k$ is the speed factor of FFW/FBW, $T_f$ is the time duration for a single frame of MPEG-2 video, $E[\beta_I]$ and $E[\beta]$ are the I-frame rate and the average rate respectively of the MPEG-2 compressed video, and $T_d$ is the fixed round trip time (FRTT) of the network. When the network bandwidth is negotiated at every GoP, only the first term in Equation(1) is needed. The numerical results corresponding to different level of interactivity and FRTT are presented in the rest of the section. We notice that the minimum client buffer size increases linearly with the FRTT because more buffering is required to maintain the display when the round trip time is large. Note that an M3N9 GoP has a higher bit rate than an M3N15 GoP. For large values of FRTT (corresponding to a WAN/MAN ), the buffer requirement for M3N9 GoP is larger than for M3N15 GoP. On the contrary, for small FRTT (corresponding to a LAN), M3N15 GoP requires a larger

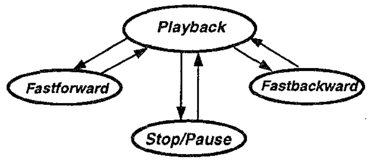


Figure 5: Illustration of the client operation mode.

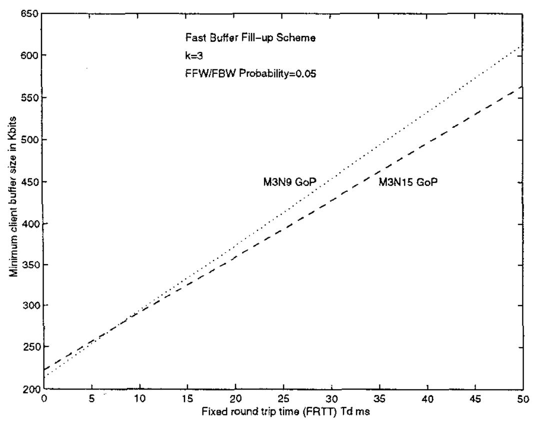


Figure 6: The minimum client buffer size versus the FRTT for FBF scheme.

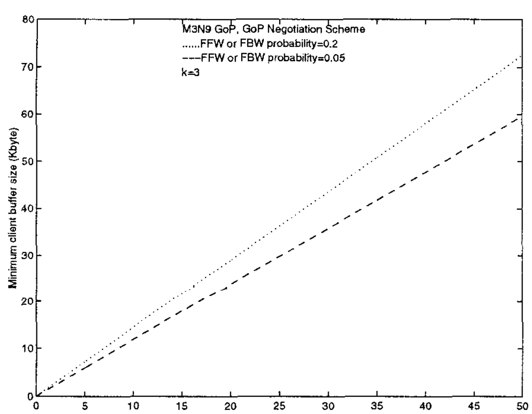


Figure 7: Minimum client buffer size versus the FRTT for M3N9 GoP with GoP negotiation scheme.

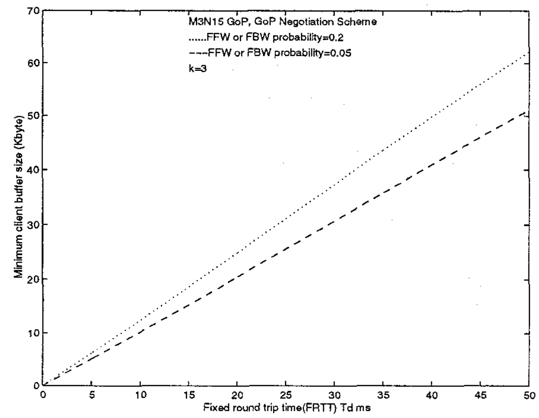


Figure 8: Minimum client buffer size versus the FRTT for M3N15 GoP with GoP negotiation scheme.

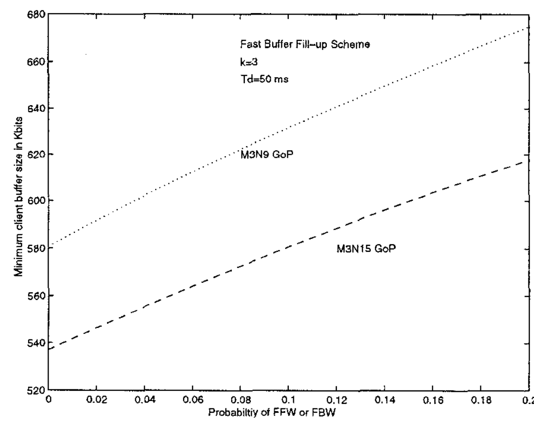


Figure 9: Minimum buffer size of client versus the FFW/FBW for the FBF scheme.

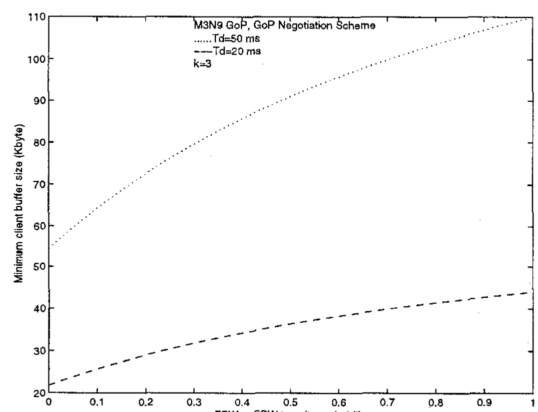


Figure 10: Minimum client buffer size versus FFW/FBW for M3N9 GoP with GoP negotiation scheme.

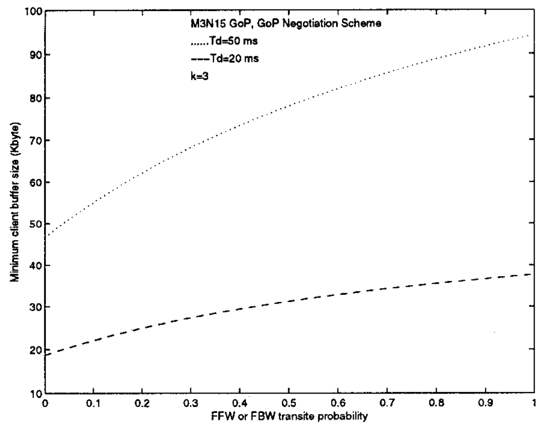


Figure 11: Minimum client buffer size versus FFW/FBW for M3N15 GoP with GoP negotiation scheme.

buffer than M3N9. This is explained by the fact that for a small FRTT, the first term of Equation(1) is small and the buffer size is dominated by the second term of Equation(1) which is the rate difference between the I-frame and the average rate. Figure 9 illustrates the minimum client buffer size corresponding to FFW/FBW probability in the range of 0-0.2 with fixed FRTT. The client buffer size increases quasi-linearly with the FFW/FBW probability. Since the M3N9 GoP has a higher data rate than an M3N15 GoP, it requires a larger buffer. Comparing the two bandwidth renegotiation schemes in Figures 7, 8, 10 and 11, we notice that they have similar pattern. In the first scheme, the bandwidth is renegotiated for every GoP, the buffer size requirement is small at the price of increasing the complexity of network bandwidth management.

## 6.2 Network Parameter and the QoS

Transmission of video over ABR requires the determination of MCR and PCR which are related to the particular video being transmitted and has impact on QoS. The average rate of an MPEG-2 video with MmNn GoP can be expressed as:

$$E[\beta] = \frac{\beta_I + \beta_P(n/m - 1) + \beta_B n/m(m-1)}{n} \quad (2)$$

For no overflow/underflow at the client buffer for the entire playback period, the long term dynamic variation of the buffer accumulation should be zero:

$$\sum_{allGoP} \delta(\text{buffer accumulationofaGoP}) = 0 \quad (3)$$

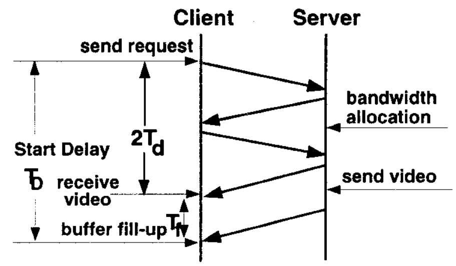


Figure 12: Illustration of the startup delay.

The expected value of ACR (denoted by $E[ACR]$) must satisfy the following condition:

$$\sum_{allGoP} E[ACR]nT_f = \sum_{allGoP} \beta_I T_f + \beta_P(n/m - 1)T_f + \beta_B(m-1)n/mT_f \quad (4)$$

Therefore, to obtain acceptable QoS, we choose MCR=$E[ACR]$. In the fast buffer fill-up scheme, to fill the client buffer as soon as possible, we set a very high PCR which is several times of MCR. For video on demand, *startup delay* is an important parameter which affects the QoS. The fast buffer fill-up scheme is designed to reduce the startup delay by exploiting the bandwidth renegotiation feature of ABR when the user uses one of the trick modes. We define the startup delay as the time between the user pressing the Playback (from the stop state) to the start of the video play, or the time between the user pressing FFW/FBW to the start of the video. The startup delay ($T_D$) consists of two parts: a fixed part arising due to the FRTT of the network and the dynamic part $T_f$ which is required to fill the client buffer to the minimum level to start playback as shown in Figure 12. To evaluate the startup delay, the expected fill-up rate during FFW/FBW is employed which can be calculated by the following formula under the assumption that the network has an exponential distribution for bandwidth request $\mu_r$ between the MCR and PCR [22].

$$E[\mu_r] = \frac{\int_{\mu_m}^{\mu_p} \mu_r P(\mu_r) d\mu_r}{\int_{\mu_m}^{\mu_p} P(\mu_r) d\mu_r} \quad (5)$$

By introducing the parameter $q = \mu_p/\mu_m$ as the ratio of the $PCR$ to $MCR$, we obtain the expected allocated rate:

$$E[\mu_r] = (1/\alpha + \frac{1 - qe^{-\alpha}}{(q-1)(1-e^{-\alpha})})(q-1)\mu_m \quad (6)$$

The expected dynamical start delay $E[T_f]$ can be obtained by:

$$E[T_f] = \int_{\mu_m}^{\mu_p} \frac{C_{\min}^C}{\mu_r} P(\mu_r) d\mu_r \quad (7)$$

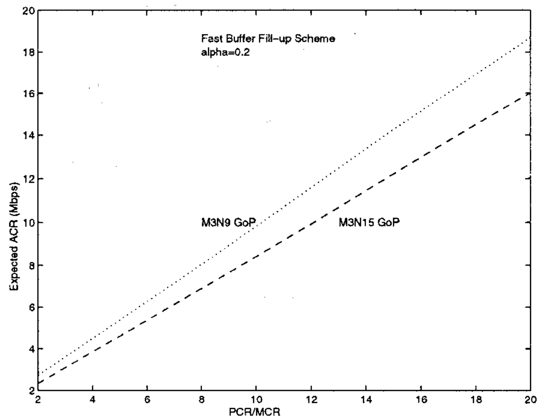


Figure 13: Expected bandwidth allocation versus $PCR/MCR$ for a constant network congestion status.

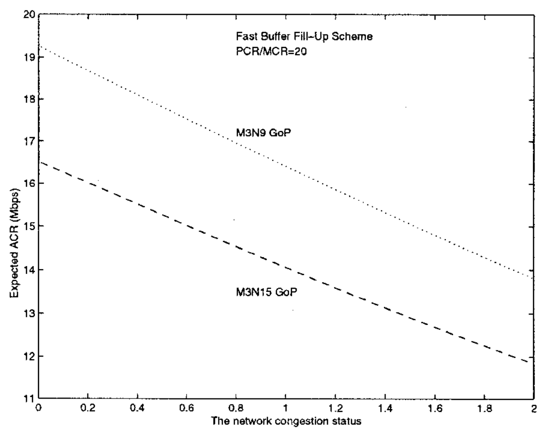


Figure 14: Expected bandwidth allocation versus the network congestion status for fixed $PCR/MCR$.

Since there is no close form results for the above integration, it can only be calculated by numerical integration. Numerical results for expected fill-up rate and expected dynamic start delay are shown in Figures 13, 14 and 15.

Figure 13 shows that the expected ACR during the FFW/FBW linearly increases with increasing PCR for a constant network congestion status. This implies that we can reduce the startup delay by setting a high PCR at connection set up. Figure 14 shows that the expected ACR during the FFW/FBW decreases rapidly when the network becomes congested resulting in a degradation in the QoS. The expected value of the dynamic startup delay versus the difference, PCR-MCR, is calculated from Equation(7) and is shown in Figure 15. We notice that M3N9 and M3N15 GoPs

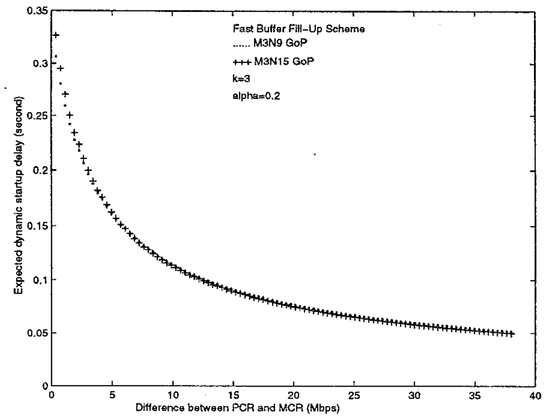


Figure 15: Expected dynamic startup delay versus the $PCR$ rate.

have the same dynamic startup delay for the same level of network congestion. Moreover, the dynamic startup delay decreases exponentially with higher values of PCR.

## 7 Conclusion

In this paper we have discussed the work which have been done so far and those in progress in the area of multimedia over ATM, taking video on demand over ATM as an example. We have provided an overview of the possible services and their suitability in running video on demand (video in general) over an ATM network. Because of the cost effectiveness of the ABR service of the ATM network, different strategies to run video on demand over ATM has been discussed in detail. Issues like client buffer dimensioning as a function of user interactivity level and effect of round trip time on running video over ABR, traffic and connection parameters required during connection setup have been discussed in the paper. We have shown results relating the buffer size required at the client versus the round trip delay in an ATM network. We discussed and compared two operating schemes for running interactive VoD over ATM. In one of the schemes, called the fast buffer fill-up (FBF) scheme, complexity of bandwidth allocation and network management is greatly reduced at the price of increasing the buffer size in the client.

The results presented in this paper prove that interactive video can be cost effectively run over the ABR service of ATM.

## References


[1] ATM Forum Technical Committee, "Video on demand specification 1.0," tech. rep., ATM Forum, Dec 1995.

[2] Daniel Delodderе, Willen Verbiest, and Henri Verhille, "Integrated video on demand," *IEEE Communication magazine*, vol. 32, no. 5, pp. 82–88, May 1994.

[3] S. Lederman, "Video on demand: A traffic model and GOS technique," *Proceedings of GLOBECOM'86*, pp. 676–683, 1986.

[4] W. D. Sincoskie, "System architecture for a large scale video on demand," *Computer Networks and ISDN System*, vol. 22, no. 2, pp. 155–162, September 1991.

[5] Lawrence G. Roberts, "Can ABR service replace VBR service in ATM network," *Proceedings of the COMPCON'95 Conference*, Piscatway, New Jersy, pp. 346–348, 1995.

[6] Jean M. Macmanus and Keith W. Ross, "Video on demand over ATM: constant-rate transmission and transport," *Proceedings of INFOCOM'96*, San Francisco, pp. 1357–1362, March 1996.

[7] Marcel Graf, "VBR video over ATM: Reducing network resource requirement through endsystem traffic shaping," *IEEE INFOCOM'97*, Kobe, Japan, pp. 48–57, April 7-11 1997.

[8] Maher Hamdi, James W. Roberts, and Pierre Rolin, "Rate control for VBR video coders in broad-band networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1040–1051, August 1997.

[9] Marwan Krunz and Satish K. Tripath, "Exploiting the temporal structure of MPEG-2 video for the reduction of bandwidth requirement," *IEEE INFOCOM'97*, Kobe, Japan, pp. 143–150, April 1997.

[10] C. J. Beckman, "Dynamic bandwidth allocation for interactive video application over corporate network," *Proceedings of IEEE COMPCON'96*, pp. 219–225, 1996.

[11] Partho Pratim Mishra, "Fair bandwidth sharing for feedback controlled VBR video traffic," *Proceedings of IEEE GLOBECOM'95*, Singapore, pp. 1102–1108, Nov 1995.

[12] Byungchan Ahn, Ki-Ho Cho, Hyojeong Song, and Jaehyung Park, "Design of rate-based congestion control scheme for MPEG video transmission in ATM network," *IEEE GLOBECOM'97*, Phoenix, pp. 1690–1694, Nov 1997.

[13] Hemant Kanakia, Partho P. Mishra, and Amy R. Reibman, "An adaptive congestion control scheme for real time packet video transport," *IEEE/ACM Transaction on Networking*, vol. 3, no. 6, pp. 671–682, December, 1995.

[14] A. D. Gelman, S. Halfin, and W. Willinger, "On buffer requirement for store-and-forward video on demand," *Proceedings of IEEE GLOBECOM'91*, Arizona, pp. 976–980, 1991.

[15] R. Grunenfelder and J. P. Cosmas, "Characterisation of video codes as autogressive moving average processes and related queueing system performance," *IEEE J.Selected Areas in Communication*, vol. 9, pp. 284–293, April 1991.

[16] D. P. Heyman, A. Tabatabai, and T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM network," *IEEE Trans. Circuits and Systems for Video Technolgoy*, vol. 2, no. 1, pp. 49–59, Jan. 1992.

[17] G. Ramamuthy and B. Sengupta, "Modeling and analysis of a variable bit video multiplexer," *Proceeding of IEEE INFOCOM'92*, pp. 817–827, 1992.

[18] Pramod Pancha and Magda El Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Communication magazine*, vol. 32, no. 5, pp. 54–66, May 1994.

[19] Wenjun Luo and Magda El Zarki, "Quality control for VBR video over ATM networks," *IEEE Journal on Selected areas in Communications*, vol. 15, no. 6, pp. 1029–1039, August 1997.

[20] T. V. Lakshman, Partho P. Mishra, and K. K. Ramakrishram, "Transporting compressedvideo over ATM networks with explicit rate feedback control," *IEEE INFOCOM'97*, Kobe, Japan, pp. 38–47, April 1997.

[21] Bing Zheng and Mohammed Atiquzzaman, "Video on demand over ATM: system design and networking requirements," *ENCOM'98-The Enterprise Networking and Computing'98*, Atlanta, June 7-11 1998.

[22] Bing Zheng and Mohammed Atiquzzaman, "Multimedia over high speed networks: reducing network requirement with fast buffer fillup," *IEEE GLOBECOM'98*, Sydney, Nov.8-12 1998.