Real-Time Support of Continuous and Variable Bit Rate Traffic on an ATM Network

Aloke Guha and Allalaghatta Pavan

Honeywell Technology Center
3660 Technology Drive, Minneapolis, MN 55418

Abstract

There is growing interest in using switched network technology such as asynchronous transfer mode (ATM) to support both continuous bit rate multimedia data and variable bit rate burst traffic for distributed computing. Most previous work and studies have focused only on the real-time support imposed by the isochronous requirements of the video or audio sources. This paper reports the results of our early study on how ATM can support real-time and prioritized traffic. We describe some basic approaches in handling real-time quality of service parameters within the context of ATM, and observe from simulation experiments that simple priority-based scheduling approaches at the virtual circuit level provides improvement in the deadline guarantee for all real-time traffic, especially the burst traffic. We believe that additional techniques such as deadline-based scheduling will be necessary to provide better real-time support .

1. Introduction

In the recent past, there has been a growing interest in applying ATM packet switching network technology to local networks, as evidenced by the activities of the ATM Forum, the international forum responsible for promoting ATM. The interest in ATM is fueled by its designed support for multimedia, its support high data rates through hardware switching, and the promise of a common standard for local to wide area networking. It is the support for multimedia that motivates us to examine the use of ATM for real-time applications, where the term real-time transcends the needs of the continuous bit rate sources such as audio and video, and includes deadline driven traffic such as from variable rate messaging or bursts such as alarm messages.

Our interests in real-time computing and control is driven by the use of multimedia data in real-time controls applications. These applications imply a broader scope in the definition of multimedia since it encompasses both image data from sources such as video, X-ray, and spectral data from infrared, color, and other optical sensors. In such real-time applications, multimedia information is used for real-time decision making and also for the detection of alarm situations which require immediate attention. While decision making usually implies a human in the loop architecture, many time-critical automated systems might directly use in-the-loop signal or image processing to automate the alarm management and diagnostics process. Thus, the most general notion of a real time system that uses multimedia data encompasses deadlines in delivery of the data, both continuous bit rate (CBR) from multimedia sources as well as variable bit rate (VBR) as expected from bursts of discrete data packets and messages.

The use of multimedia in the distributed real-time systems imposes requirements on the operating system, the storage and data management, the man-machine interface and the communications network. Our specific interest in this paper is on the network support aspects. To determine the specific areas of support required, we have conducted a detailed analyses of the needs of distributed multimedia-based control systems [Guha 93]. A summary of these requirements is given in Section 2. Section 3 covers existing approaches to real-time support in ATM networks, while Section 4 describes our proposed approach. In Seciton 5 we present results from our simulation study. Finally, we conclude in Section 6.

2. Distributed real-time applications

The network requirements were derived by examining the classes of operations that occur in typical distributed real-time applications such as command and control [Guha 93]. The traffic engendered by these applications must provide the following features.

2.1 Priority

For purposes of scheduling tasks, a single attribute (priority) is associated with different communication tasks. In the most general case, priority is composed of such factors as criticality, deadline, laxity, and resource requirements and is usually specified globally by the

1063-6927/94 $03.00 © 1994 IEEE

application. For hard real-time tasks, failure to meet the deadline may be considered a system fault, in which case the priority of such tasks may be very high. In some applications, priority is independent of deadline. For purposes of generality, we treat priority and deadline independently. The key principle in working with prioritized tasks is to avoid priority inversion [Sha 87] wherein a lower priority task is allocated resources before a higher priority one.

2.2 Deadline and synchronization support

To support advanced control features, (such as multiple input and multiple output control), the network must provide I/O synchronization. Here multiple CBR streams and/or VBR data need to delivered within a specified time bound, a deadline, to a common destination.

2.3 Multicasting

Besides directly supporting priority based real-time traffic and I/O synchronization, the real-time network must also support an effective multicast capability. However, within the scope of this paper, we will only examine support for priority based real-time data.

We now briefly examine how these requirements are addressed by the existing body of work on real-time communication networks.

3. Approaches to real-time support in ATM networks

There has been scant work on the support for real-time communications on packet switched networks, much less in the context of ATM networks. The primary emphasis of published work has been the support of CBR stream traffic that require isochronous packet delivery. There has been no notion of supporting deadline based traffic for VBR data communications. Clearly, the nature of ATM plays a role in determining its efficacy in providing these network features.

3.1 ATM Networking

ATM is a switch-based, cell relay technology that can connect individual nodes at dedicated bandwidths [Boudec 92]. By statistical multiplexing of identically sized (53 byte) cells, ATM can handle a mix of data, voice and video traffic. By using small fixed-sized cells, ATM provides the flexibility and efficiency to support both bursty asynchronous traffic as well as a wide range of continuous stream sources such as audio and video. Cells are multiplexed asynchronously and routed to the proper destination by identification of a virtual channel that is assigned to each source requesting a connection. Since ATM provides a connection-oriented service for each source, every connection is assigned a virtual circuit that is defined at each switch by a virtual channel identifier (VCI) and virtual path identifier (VPI) in the cell header. The VCI and VPI combination identifies the path through the ATM switch. All connections in the ATM switch are set up through a process called signaling, before actual information is transferred.

We note two key points that are significant in determining the applicability of ATM for priority-based real-time traffic. First, while the use of statistical multiplexing makes ATM attractive for efficiently handling different data rate traffic, it is by definition designed for fair access to different users. ATM does not inherently support multilevel priorities. Current activities in the standards bodies do not address the specification or handling of multilevel priority traffic. Second, since ATM uses a connection-oriented data transfer mechanism, all data sent is associated with a connection that has been established a priori. Unlike true packet switching, the time of sending a packet is different from the time a connection is set up. Because there is no explicit timing information associated with the the cells, it is difficult to monitor deadlines associated with a given packet cell. How timing information is associated with any cell is an open issue (although as yet unspecified AAL, the ATM Adaptation Layer, 1 and 2 are proposed for maintaining a relationship between sender and receiver for connection-oriented traffic). Coupled with the notion of statistical multiplexing, this lack of timing in a packet makes it difficult to ascertain the determinism and the associated support for real-time data transfer.

3.2 Related work and relevance

The real-time work as present in the ATM-related literature has focused on real-time versus non-real-time in terms of link scheduling but not in the context of deadlines. More recent efforts within the ATM networking community, primarily in the ATM Forum, have begun to address the notion of delay and cell loss (the relative fraction of the cells dropped by the switch in the face of congestion caused by high loads typically due to unprovisoned traffic). Cell loss is related to missed deadlines in the extreme sense, since a loss of a cell implies the loss of information to be transmitted and therefore failure to meet deadlines. However, even if cell loss were not an issue, deadlines can be missed at high loads since the cell of a low-laxity packet can be "stuck" in the switch buffer. (We observe this in our simulation results described in Section 5 where infinite buffers were assumed.) A brief summary of the relevant work in the area of real-time support is worth noting and is provided below. Most of these do not address the real-time problem posed here.

Hyman, et al [Hyman 91] have examined how different quality of services (QOSs) can be guaranteed for two classes of traffic, real-time or non real-time, using link schedulers that schedule arriving cells of varying priorities such that the QOS requirements can be met. However, while the approach addresses mixed real-time and non-real-time traffic, it does not address the general class of multiple levels of priority of multimedia or non-multimedia traffic.

Clark [Clark 92] has described a packet switched network architecture for supporting real-time "playback applications" for stored multimedia. While these applications ensure a degree of control over the source through buffering data or adjusting the playback point, they are not applicable to real-time applications where data sources cannot be controlled.

Ferrari, et al. [Ferrari 90] define a scheme for real-time channel establishment for WANs that they believe is applicable to ATM. However, they assume that delay in each link is bounded, that the peak and average rates are known a priori, and that flow control is rate-based. In many real-time command and control networks, these assumptions on rate based flow control are not always valid.

More relevant approaches in handling priority traffic in the ATM have focused on selective cell discarding based on the priority of the cell [Petr 93, Kröner 91]. We believe these approaches are more applicable to handling priority traffic in switched networks for multimedia, but they do not directly address the real-time needs of the traffic. Other related approaches [Chipalkatti 89, Zhu 92] to real-time scheduling for network traffic has been to use separate queues for real-time and non-real time traffic. The scheduling policy examines how to empty the queues while attempting to minimize the loss ratio (of deadlines). While these approaches work well for a two-class traffic, they may not be directly applicable to the multilevel priority deadline based traffic.

In the context of supporting traffic with low loss requirements, there have been proposals such as the "best effort" service [Cherbonnier 93] that attempts to provide fairness between users by throttling traffic (by backpressure signals) from users that mostly contribute to the overload while protecting the "innocent" connections. However, there is no notion of priority of the users or consideration of deadlines. Other approaches to reduce unwarranted congestion is a rate based approach such as stop-and-go queuing (SGQ) algorithm [Golestani 90] where a source is not permitted to transmit more than prespecified number of bits during a predetermined time interval. As in other rate based flow control approaches, SGQ results in a "best- effort" or "no guarantees" scheduling. Recent simulation studies on the use of SGQ for best effort scheduling [Bagwell 93] show that in the presence of isochronous traffic that is permitted to reserve all bandwidth necessary to ensure QoS (quality of service) guarantees, asynchronous (read VBR) TCP traffic may experience as much as 2% cell loss.

The work we present here addresses the problem of how both CBR and VBR traffic with real-time needs are supported by ATM with and without support for priority in the switch. There are no assumptions on the associated priority of the traffic. Unlike typical ATM switch vendors, we do not assume that the isochronous CBR traffic inherits higher priority.

4. An ATM-based networking approach

To meet the networking requirements of multimedia-integrated time-critical applications, we have been exploring a real-time control network architecture that provides a single integrated network capable of carrying a wide range of services that include network scheduling for real-time traffic, synchronization for control and multimedia data, and end-to-end reliability. Our approach is to develop mechanisms for mapping the communication requirements of the applications onto appropriate QOS parameters, and then implement the appropriate underlying network services. Here we address the support for real-time traffic through network scheduling.

4.1 Priority based scheduling

Scheduling can be employed at different levels: at the cell level, at the virtual circuit (VC) level, or at some higher message level. In the context of ATM, cell-level scheduling entails extremely simple and fast processing. On the other hand, VC-level scheduling can involve more complex decisions, made perhaps during the signaling phase when the VCs are established. In order to provide support for scheduling without compromising the high data transfer rates through the switch, we propose VC-level scheduling while allowing for cell-level decisions for support, for example, in discarding cells which have missed deadlines. There are a number of different aspects to be studied in specifying the needs of these applications and how these needs are met:

Specifying priority: In ATM networks, QOS parameters are specified for a VC during call set-up. We propose extending the QOS specification beyond that of peak and mean flow rates and cell loss to include measures of priority, *defined by the user*, and deadlines. These real-time QOS parameters will be incorporated into the VC lookup tables at each switching point in the network between the end stations (Figure 1). The priority information associated with a VC may be translated into cell loss priority and possibly into the generic flow control field for use in cell level scheduling. The deadline information is used during cell switching to determine the worth of the cell, for instance, to drop the cell if the laxity (the difference between current time and the deadline) associated with its message or packet is zero or less. Note deadlines are more relevant for non-continuous message traffic than for CBR traffic.

VC ID	Source	Destination	Priority	Deadline	...

Figure 1. Example VC lookup table for real-time support

A key issue in specifying real-time parameters in ATM is how to resolve message level priorities into VC level priorities and possibly cell level priorities. Note that cell-level priority does not have to be specified basis since that it be readily obtained from the VC lookup table. The VC level based priority association is preferred since the ATM cell format provides no space to accommodate this added information. The cell loss priority field is only 1-bit wide and cannot represent more than two levels of priority. The 4-bit GFC field can be used, but it is not visible outside the user network interface. Also, we prefer not to specify deadlines at a cell basis for similar reasons. For CBR multimedia traffic, it could be computed on the fly for each cell based on the deadline information for the VC. For non-CBR cells of a switched VC (SVC) that is dynamically set up, we assume that the deadline information is specified with the VC.

Guaranteeing priority: Guaranteeing of priorities takes place at two levels. First, at signaling time when SVCs are set up. That is, a VC request is entertained relative to the priorities of the VCs already granted access. For example, entry of the VC could be based on prespecified queue allocation (as in [Kröner 91] for cells) in a shared physical buffer for different priorities. This will allow preemption of lower priority traffic by higher priority requests. Second, in case of congestion among traffic of existing VCs, we can use scheduling policies based on deadlines. We believe that cells entering the switch could be discarded based on laxity consideration. For example, during VC table lookup, a cell can be tested for discarding if its laxity is less or equal to zero.

However, it is not clear whether the cell discarding should be based on 'hard' deadlines as opposed to 'soft' deadlines. One way of specifying soft deadlines is through a concept similar to time value functions [Jensen 85] that specifies the relative merit of the cell after its deadline is past. However, the evaluation of the time value function may be nontrivial adding to the complexity of switching decisions. Evaluation of our proposed approach requires a comparative evaluation of different cell discarding techniques based on fast computations of "cell deadlines".

PVC and SVC allocation: Network scheduling implies allocating network resources according to priority needs. We have to make a choice between using permanent VCs (PVCs) and SVCs to implement priority. Certain classes of traffic, such as alarm messages are aperiodic (VBR), bursty, high priority, and may have highly variable burst lengths. For such traffic, PVCs will have to be established as there may not be enough time to set up a VC when an alarm stream occurs. Other operations such as requests for video windows for monitoring may occur infrequently and can therefore use SVCs set up on a per call basis.

Flow control, congestion control, and priority guarantees: In ATM, flow rates (average and peak) and the level of guarantee desired may be specified through QOS parameters and implemented using flow control and congestion control policies. As indicated in Section 3.2, current bandwidth reservation schemes as well as traffic policing schemes in ATM are based on providing fair access to all users, with special attention to the isochronous CBR traffic. Our congestion control or flow control policies are based on traffic priority.

5. Empirical results

An extensive simulation model for an ATM-based network was developed using the SES WorkBench simulation toolkit. We have assumed a point to point ATM network with nodes (user stations) attaching to a single ATM switch via a pair of unidirectional fibers. Below is a short description of the key components of the model.

5.1 ATM Switch

A space domain non-blocking ATM switch is generically modeled with both input and output queuing and a speedup factor of S, ($1 \leq S \leq N$), N being the size of the switch. The speedup factor S enables upto S packets to be delivered to the same output port simultaneously. These delivered packets are queued in the output buffers to wait for the service of the output links, while those packets not delivered are queued in the input buffers to wait for the service of the switch. The extreme situations of $S = 1$ and

S = N thus correspond to pure input queuing and pure output queuing switches, respectively.

We model the system as queues in tandem. We logically divide the packets in the head-of-line (HOL) positions into N "virtual queues" where virtual queue Q(V, j), $1 \le j \le N$, is composed of cells in the HOL positions of all input queues Q(I, i), $1 \le i \le N$, destined to port j (Figure 2). Thus, each virtual queue can be thought of as having S servers to represent the S-fold speedup of the switch. The N virtual queues Q(V, j), $1 \le j \le N$, feed into N output queues Q(O, j), with the jth virtual queue feeding into the jth output queue, $1 \le j \le N$.

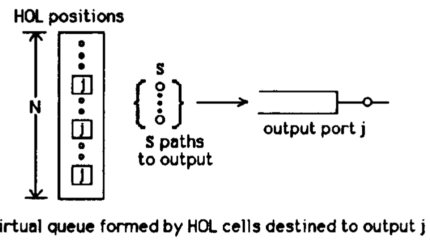

Figure 2. Virtual queue in switch model (HOL: head of line)

We thus have a general model which consists of queues in tandem to describe a family of ATM switches. This has the following advantages : 1) By modeling the switch as a sequence of queues, the analysis is simplified; 2) The model does not tie us into a single switch architecture since the purpose of the model is to study ATM networking approaches and not existing switch designs; and, 3) Moving between a pure input queuing, a pure output queuing, or any hybrid queuing design in between the two is easily accomplished by changing the speedup factor.

5.2 User station

Each user station is assumed to possess a single ATM adapter connected to the switch via a pair of links, running at OC-3 (155 Mbps) speeds in our case. Each station can spawn three kinds of processes representingthree kinds of traffic sources: CBR, VBR and Poisson (generating Poisson arrival of cells). Each source instance generates a cell stream on a separate VC (Virtual Channel).

CBR sources are characterized by the mean bandwidth of their respective VCs. VBR sources are characterized by three parameters: i) the mean bandwidth, ii) the peak bandwidth, and iii) the burst factor of their respective VCs. The burst factor is defined as the total data, in excess of waht would be produced at the mean data rate, during the burst period. A two-state process model is used to represent a VBR source, with states representing "active" and "idle" periods of the sources with the time in each state exponentially distributed with a mean computed from the three parameters listed above. Poisson sources have only one parameter, namely the mean interarrival time.

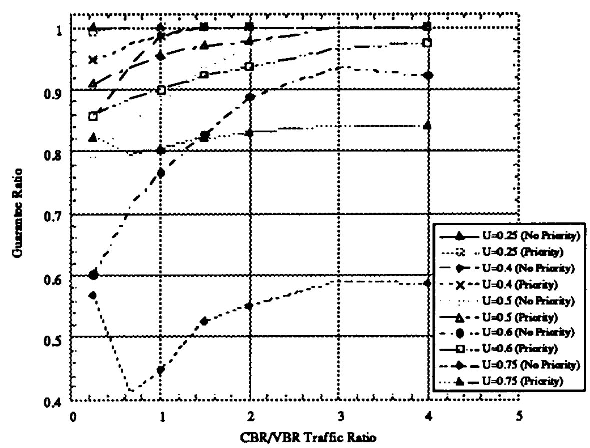

Figure 3. Guarantee ratio for CBR traffic with and without priority support; (U : switch link utilization)

Each user station can spawn any number (upto a maximum) of any of the three types of sources. Thus the traffic sources in the model are really logical sources, each source coming from a separate process in a user station. The adapters thus play the role of multiplexing all VCs emanating from a single user station. Each source is assumed to have an user-defined priority, and also an associated deadline specifying the maximum delay the cell can incur for delivery to the end station. Both criticality and deadline are defined on a per-VC basis. That is, each cell in a given VC has the same criticality and deadline values as defined for the VC. This assignment of deadlines to cells has been done to instrument a measure of the deadline support at the level of cells. In practice, a cell level deadline is not truly meaningful.

5.3 Traffic scenarios

For the current experiments, a number of assumptions were made. We assume the use of PVCs only in the model. Later, we plan to incorporate Switched Virtual Circuits (SVCs) and signaling protocols in our model. In the current scenarios we have also ignored Poisson sources because the interplay of CBR traffic and VBR traffic is of greater interest since they are more common. We have also assumed pure output queuing at this stage, i.e., the speedup factor S is assumed to be N. Presently, priority by the user is assigned randomly, from a set of three possible values, to each VC, regardless of the type of source the VC represents. The deadline value is typically estimated as the tolerable delay for the cells in the VC from the time they are generated to the time they are delivered to the destination. The delay consists of transmission delay, propagation delay and switch processing delay. Presently, we ignore propagation delay since only one switch is modeled in the simulations.

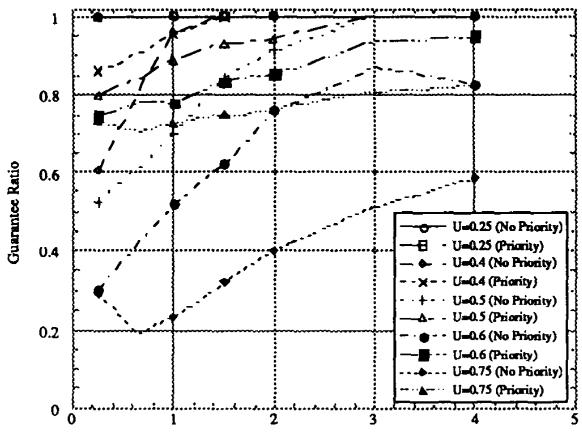

Figure 4. Guarantee ratio for VBR traffic with and without priority support

For the simulation runs, the assignment of ports to the VCs in the switch are presently done randomly, with the bandwidth constraints maintained, i.e., the sum total of mean bandwidths of all VCs on a single input or output port cannot exceed the link speed of 155 Mbps. Since only two types of sources are involved, we vary the relative loads between the two types. For the present set, we consider relative loads of 20%, 40%, 50% 60% and 80% due to CBR traffic (and remaining to VBR traffic) for five scenarios for each loading factor (0.25 to 0.75) for the switch. Further, we run two experiments with each scenario: one where the switch operates in raw ATM mode, i.e., without any priority and deadline handling and one in which the switch operates to handle priority. The current handling of priority is done by a simple non-preemptive priority mechanism to study its effect. We have not yet run any experiments to process deadlines of the cells. We simply measure the number of cells arriving at the destination beyond the specified deadlines and compute a guarantee ratio, GR, for each traffic type that is defined as

$$GR = 1 - \frac{\sum_{i}^{3} D_i C_i}{\sum_{i}^{3} N_i C_i}$$

where D_1, D_2 and D_3 are the number of cells in each priority class that have missed their deadlines; C_1, C_2, and C_3 the three values of user-defined priority; and, N_1, N_2 and N_3 the total number of cells in the respective priority classes delivered. Clearly, the larger the guarantee ratio, the better is the performance of the switch in terms of guaranteeing deadlines.

The data from our simulations averaged over 5 runs corresponding to a near-steady state operation at the given load distribution that occurs over 150,000 cell deliveries. The results are plotted in Figures 3-6. We have examined the guarantee ratio with and without priority for both CBR and VBR, for different link utilizations and different mixes of CBR and VBR bandwidth ratios. The plots show how guarantee ratio is affected by the load on the network in terms of the utilization as well as the interference of the traffic, i.e., VBR perturbing CBR traffic and vice versa. As mentioned earlier, all classes of traffic can assume any level of priority, as we expect in the real-time distributed control and computing environments. Also note that since buffers were assumed infinite, we do not have to consider cell loss issues but can monitor the delays that results in missing deadlines.

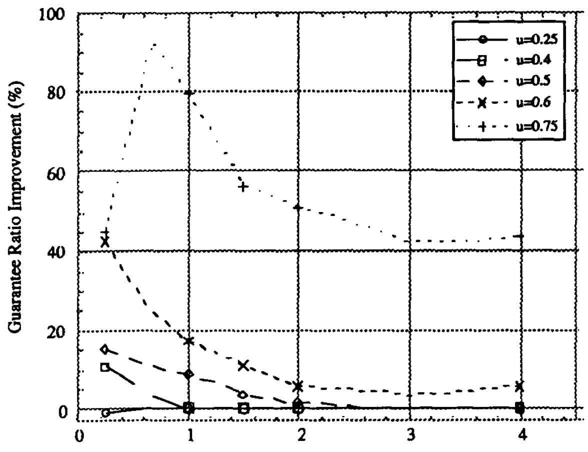

Figure 5. Improvement of guarantee ratio for CBR traffic with priority support

5.4 Interpretation of results and future work

Although our studies are preliminary and more analyses will be undertaken, we note certain trends evident within experimental errors. These are:

- The guarantee ratio for both classes of traffic increases at higher ratios of CBR to VBR traffic, with or without priority support. (Figures 3 and 4). We believe this is attributed to the fact that more CBR traffic implies more deterministic loads and less bursty unscheduled traffic.
- The effect of priority support for both CBR and VBR traffic shows increases in guarantee ratio, especially at the

higher utilizations where scheduling based on priority during more congestion results in better support of higher priority traffic (Figures 5 and 6). The improvement in the guarantee ratio for VBR traffic is much more dramatic at higher ratios of VBR to CBR traffic. As Figure 6 shows, more than 250% improvement occurs at link utilizations (i.e., high loads) when the CBR:VBR bandwidth ratio is less than 1:1 ratio. We would postulate that priority scheduling of the VBR, i.e., the less deterministic messaging traffic, has a larger impact than that of the more deterministic CBR traffic.

The key issues in scheduling traffic in the ATM switch will be that of the deadline based VBR traffic. Our results show that using priority scheduling on VBR and CBR traffic has a much bigger impact on the deadline guarantee ratio for VBR. Cell loss as a metric (as used in [Bagwell 93]) is not an adequate measure for deadline guarantees although increased cell loss implies loss of deadline and reliable delivery.

While clearly further studies are required, the notion of meeting hard deadlines for bursty VBR traffic will require better cell scheduling techniques that use more explicit deadline information, a notion that is absent in the connection-oriented world of ATM. For instance, more deadlines missed in this model were caused by the fact that expired cells stayed in the buffer causing other traffic with shorter laxities of the same priority to suffer unreasonable delays as well..

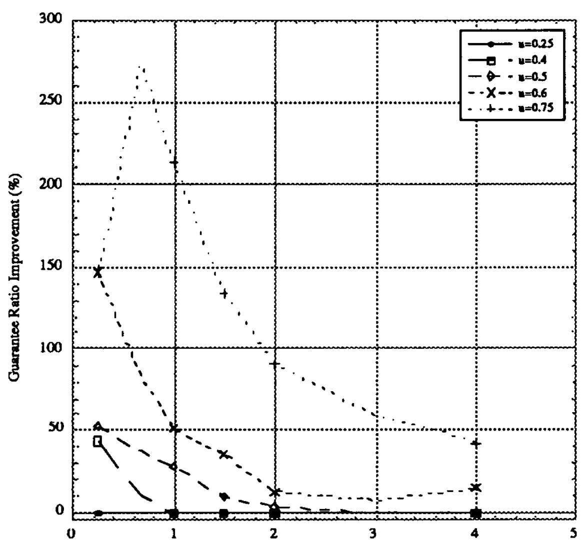

Figure 6. Improvement of guarantee ratio for VBR traffic with priority support

One option towards handling VBR traffic interference is to use cell discarding for expired cells as a way of improving guaranteeing deadlines. This also requires providing network level notion of the deadlines associated with a message. There also associated notions of whether a number of cells (all belonging to a complete message), especially in VBR traffic, or a complete message should be discarded. The latter is media dependent, i.e., typical multimedia data such as video can tolerate some cell loss, with a smaller tolerance in audio. However, cell loss in text messages are not acceptable in most time-critical applications. Also, the sender of the message should be notified in such cases so that alternate remedial steps can be taken.

The second option is to overprovision the VC for a VBR source according to the peak burst rate. While that is not efficient, it may allow for better guarantee ratios. An area we will investigate is what form of provisioning improves guarantee ratio while not sacrificing the full bandwidth to the peak rate.

We are currently implementing techniques for providing information on message deadlines, for both VBR and CBR delivery, to the ATM switch so that the switch can make more informed scheduling. The critical issue there is to use lightweight scheduling decisions so that the high data transmission rates for continuous media sources can be supported.

5.5. Experimental testbed evaluation

The scheduling support between end nodes and within the ATM switch, the sharing of timing information for supporting synchronization, and provision for end-to-end reliable data transfers are currently under development at the Honeywell Technology Center. Both modeling and simulation as well as actual implementation is under way. Our experimental work has begun with collaboration with USWest, AT&T, University of Minnesota, and the Minnesota Supercomputing Center using Honeywell's M*ercuri* ATM testbed (the only ATM-based real-time multimedia testbed that we are aware of in the US presently) and under the auspices of USWest's COMPASS Program..

6. Conclusions

Current ATM networks while suited for supporting multimedia transport, do not address the real-time needs, specifically the need for supporting deadlines and multiple levels of priority. Earlier work and studies have focused only on the real-time support of CBR traffic as imposed by the isochronicity of the video or audio sources. Here we have reported the early results of our study on how ATM does and can support real-time and prioritized traffic. We have proposed both QoS specification schemes and simple virtual circuit and cell-level scheduling approaches to support priority traffic within the context of ATM. We observe that simple switch scheduling approaches based on priorities at the virtual circuit level has a beneficial impact on both CBR and VBR traffic, especially on the VBR traffic when the proportion of VBR traffic in the network is large. We believe additional techniques for scheduling cells with more explicit information on message-level deadlines will be necessary to provide better real-time support. We plan to continue our work in this and related areas in the near future.

References

[Bagwell 93] C. Bagwell and J. Daigle, "ATM Cell Delays and Loss for Best-Effort TCP in the Presence of Isochronous Traffic," Mitre Corp. Proposal to the ATM Forum, September, 1993.

[Boudec 92] J.Y.L. Boudec, "The Asynchronous Transfer Mode: a tutorial," *Computer Networks and ISDN Systems* 24(1992) 279-309.

[Cherbonnier 93] J. Cherbonnier and D. Orsatti, "Base proposal for simple flow control by the network at the UNI, " IBM Corp.ATM Forum Contribution, July 1993.

[Chipalkatti 89] R. Chipalkatti, J.F. Kurose, and D. Towsley, "Approximate Minimum Laxity and Earliest Deadline Scheduling Policies for Real-Time and Non-Real-Time Traffic in a Statistical Multiplexer," *Proceedings of Inforcom'89*, Ottawa, Canada, April 1989.

[Clark 92] D. Clark et al., "Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanism," *Computer Communication Review*, Vol. 22, No. 4, August 1992.

[Ferrari 90] D. Ferrari and D.C. Verma, "A Scheme for Real-Time Channel Establishment in Wide-Area Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 8., No. 3, April 1990.

[Golestani 90] S. Golestani "Congestion-Free Transmission of Real-Time Traffic in Packet Networks." Proceedings of IEEE Infocom '90, pp. 527-536.

[Guha 93] A. Guha, M. Agrawal, and R. Rundquist, "Multimedia-Integrated Real-Time Control Systems," Proc.of the First IEEE Workshop on Real-Time Applications, May 1993, New York.

[Hyman 91] J.M. Hyman, A.A. Lazar, and G. Pacifici, " Real-Time Scheduling with Quality of Service Constraints," IEEE Journal on Selected Areas of Comminication, Vol. 9, No. 7, September 1991, pp. 1052–1063.

[Kröner 91] H. Kröner, G. Hébuterne, and P. Boyer, "Priority Management in ATM Switching Nodes," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 3, April 1991, pp. 418-427.

[Petr 93] D. Petr and J.S. Wineinger, "End-to-End Priority Cell Discarding Analysis for ATM Networks," Proceedings of Infocom Conference, 1993, pp. 1178-1185.

[Sha 87] L. Sha et al., "Priority Inheritance Protocols: An approach to Real-Time Synchronization," Technical Report CMU-CS-87-181, Computer Science Department, Carnegie-Mellon University, 1987.

[Zhu 92] W. Zhu and S.T. Chanson, "Adaptive Threshold-Based Scheduling for Real-Time and Non-Real-Time Traffic," Proc. Real-Time Systems Symposium, Phoenix, AZ, Dec. 2-4, 1992, pp. 125-135.