

# Real-Time Multimedia Scheduling Policies for End-to-End Delay Jitter and Loss Guarantees Across ATM Satellite Systems

Hervé Le Pocher, Victor C. M. Leung, *Senior Member, IEEE*, and Donald W. Gillies

**Abstract**—We propose a simple first-in first-out (FIFO)-based service protocol which is appropriate for a multimedia ATM satellite system. The main area of interest is to provide real-time traffic with upper bounds on the end-to-end delay, jitter, and loss experienced at various service queues within a satellite network. Various service protocols, each based on a common underlying strategy, are developed in light of the requirements and limitations imposed at each of the satellite's subsystems. These subsystems include the uplink (UL) earth station (ES) service queue, on-board processing (OBP) queues, and the downlink (DL) ES service queue feeding into a wireline ATM network or directly to an end-user application. Numerous network simulation results demonstrate the tractability, efficiency, and versatility of the underlying service discipline. Key features of our strategy are its algorithmic and architectural simplicity, its non-ad-hoc scheduling approach, and its unified treatment of all real-time streams at all service queues. In addition, the delay and jitter bounds are uncoupled. In this way, end-to-end jitter can be tightly controlled even if medium access requires long indeterminate waiting durations.

**Index Terms**—Asynchronous transfer mode (ATM), jitter, multimedia communication, protocols, resource management, satellite communication, scheduling.

## I. INTRODUCTION

**D**ELIVERY of ATM multimedia services, especially in its initial stages, may be facilitated and enhanced by the deployment of satellite systems. Many of these next-generation systems will provide wide-band multimedia connections, often using ATM-based OBP, on a *Ka*-band satellite platform [1]. In order for ATM satellite systems to be successful, they should conform to, and integrate with, many of the established goals governing the ongoing development of the terrestrial ATM network infrastructure.

Such integration will require the satellite to provide per-virtual-connection (VC) assignments which guarantee upper bounds on end-to-end cell transfer delay (CTD), cell delay variation (CDV), and cell loss ratio (CLR). These three quality-of-service (QoS) parameters will be greatly affected by the medium-access control (MAC) algorithm which allocates UL channel access time and access duration among the

distributed and contending ES's [2], [3]. A multimedia MAC has two conflicting objectives: 1) to coordinate the ES UL transmissions, thus maximizing throughput, and 2) to provide a high QoS commensurate with a multimedia ATM environment.

Relatively little research has been directed toward optimizing satellite performance in terms of queue scheduling objectives. For example, once an ES gains access to the UL channel, which cells (VC's) should it service (transmit) first in order to satisfy its overall multimedia QoS objectives? This problem will continue to gain importance as the number of multimedia VC's, queued within satellite networks, increases. Other compounding difficulties are the expected divergence of traffic characteristics (variable/constant bit rate (VBR/CBR), burst ratio, burst duration, bandwidth) and divergence of QoS requirements among the queued VC's. These service issues are central to the current development of wireline ATM networks [4], and should be incorporated within next-generation multimedia satellite systems.

The delayed frame queueing (DFQ) service protocol was originally developed for wireline ATM networks [5], [6]. It has recently been extended to the wireless mobile ATM environment [7], as well as the ATM satellite environment [8]. The focus of this paper is to explore various QoS features associated with different queue scheduling policies, and to thus optimize the behavior of queues based on their location and function within a multimedia satellite system.

### A. System Overview

In this paper, a connection's QoS is entirely determined by the cumulative delays and cumulative losses experienced at the queues of each satellite subsystem: 1) the UL queue (UL-Q) of the transmitting ES, 2) the switch output queue of the satellite (sat-Q), and 3) the DL queue (DL-Q) of the receiving ES; see Fig. 1. The aim of this paper is to determine the service characteristics best suited to each of the three satellite subsystems. These three queues (UL-Q/sat-Q/DL-Q) need to address different operating conditions as well as different service objectives. We now consider three main types of queues within a satellite system.

The central operating characteristic of a UL-Q is that it only has limited and bursty access to the UL channel. As a result, a UL-Q must buffer its traffic until it is granted access by the MAC. This leads to burst formation which shows up at the sat-Q and the DL-Q. Although undesirable, this burst formation is a necessity in order to achieve good statistical multiplexing

Manuscript received February 28, 1998; revised September 14, 1998. This work was supported by a grant from the Canadian Institute for Telecommunications Research under the NCE Program of the Canadian Government. This paper is based in part on a paper presented at ICC'98, Atlanta, GA, June 1998.

The authors are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, B.C., V6T 1Z4 Canada.

Publisher Item Identifier S 0733-8716(99)01712-6.

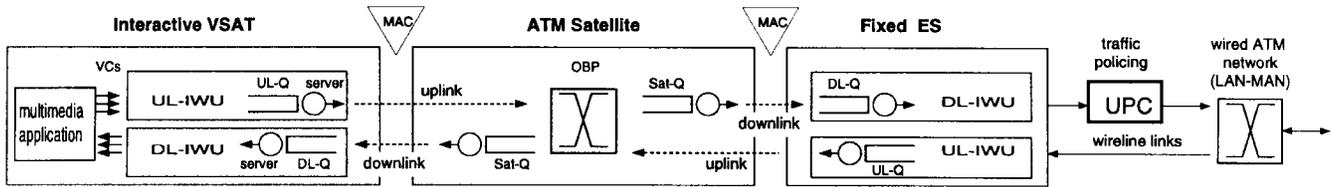


Fig. 1. Multimedia satellite system integrated with terrestrial ATM network.

gain at the UL-Q. Note that a queue service protocol only defines the behavior of the queue server once it has gained access to the channel, but does not define how or when the server gains such access. Thus, our generic protocol can be combined with many types of MAC protocols.

The main determinant of the sat-Q is that it must meet the satellite's numerous physical constraints. In general, this means that the sat-Q must be small, simple, and fast. Its main objective is to optimize the throughput of real-time traffic, regardless of burst formation, i.e., similar to the UL-Q situation. Its major advantages compared to a UL-Q server is that the sat-Q server has continuous access to a dedicated DL channel, and it need not contend with distributed queueing.

Generally, the DL-Q server is not heavily loaded. For example, a direct-to-user ES may only accept and process a single VC, from the much greater aggregate traffic that is broadcast by the satellite. In contrast, an ES feeding into a wireline local area–metropolitan area network (LAN–MAN) may accept a major portion of the satellite's broadcast, and yet its effective load may be low if it is using a high-speed terrestrial ATM link (e.g., 622 Mbits/s). The main objective of the DL-Q is to provide adequate CDV bounds as specified by multimedia applications, or as specified by policing mechanisms such as the usage parameter control (UPC) at the network node interface (NNI) between the satellite network and the terrestrial ATM network.

Previous proposals for integrating satellite systems with terrestrial networks are characterized by a strong coupling between the CTD and the CDV [10], [11]. Thus, the duration over which statistical multiplexing can take place at the UL-Q is severely limited for applications which require tight CDV bounds. Our proposal uncouples these two QoS parameters, and allows maximal UL channel utilization among real-time sessions with diverse delay sensitivities, e.g., teleconferencing versus video retrieval. Finally, all three queues (UL/sat/DL) should be capable of guaranteeing CLR bounds on a per-VC basis, such that some level of QoS isolation is obtained.

## II. DELAYED FRAME QUEUEING (DFQ) PROTOCOL

We begin by presenting the DFQ service discipline which is the underlying strategy used throughout this paper. This service strategy only defines the service priorities, deadlines, and eligibility times of queued cells, but does not define how channel access (e.g., UL channel) is gained. Our presentation is divided into a few stages, each of which further approaches the final desired scheduling system within a satellite environment. In this paper, the propagation, processing, and switching delays of a connection path are assumed to be constants and are set to zero. Thus, queue delays are assumed to be the sole cause of end-to-end CTD and CDV. Similarly, violations of these

two queueing bounds are assumed to be the sole cause of end-to-end CLR. In particular, note that gross queue (buffer) overflow is not a cause of cell loss, only scheduler saturation can cause cell loss.

We start with a simplifying assumption, namely, that each queue server within a network has continuous access to the transmission medium. This reduces the system to a wireline network. Thus, the generic term “node” may refer to a UL-Q, sat-Q, DL-Q, or even a queue within a wireline network. We first present the work-conserving version of DFQ, where all queued cells are always eligible for transmission. Note that “eligibility” refers to a state defined by the DFQ protocol at the server, and is defined independently of a queue's access to the transmission medium. For example, a cell queued at a UL-Q may be eligible for transmission, but cannot be transmitted until the server gains access to the medium. We then present the nonwork-conserving version of DFQ which provides tight jitter control. Finally, we consider the more realistic case, in which a general MAC protocol allows the UL-Q to have only intermittent access to the UL channel.

### A. Work-Conserving DFQ

The work-conserving version of the DFQ protocol is particularly appropriate for the UL-Q and the sat-Q, where high throughput utilization is critical.

*Synchronization:* The time axis of each network link is divided into one continuous *stream* of logical abutting frames of duration  $T$  seconds; see Fig. 4. The significance of the frame period  $T$  is that it defines a standard time unit during which statistical multiplexing can take place. The DFQ strategy assumes a synchronous network, i.e., each node is aware of the time boundaries which define the beginning and end of each link-level frame. We assume that all network links are in *phase* at each node, meaning the arrival and departure times of frames at nodes are synchronized. This is equivalent to setting the phase mismatch constant  $\Theta_l$  to zero in Golestani's *stop and go queueing* discipline [12]. This zero phase assumption is not required, and is only included to simplify the equations, figures, and discussion.

*Service Discipline:* The service queue of each link is organized as a sequential row of  $N + 1$  FIFO buffers,  $f_0, \dots, f_n, \dots, f_N$ , each containing cells which have a service deadline  $(n+1)T$  seconds after the start of transmission of the current outbound link-level frame. For simplicity, we assume that the value  $N$  is common to all of the network queues. Service priority is given to cells buffered in the FIFO's with the smallest index values of  $n$ ; see Fig. 2. If the link speed is  $c$  cells/s, then there are  $v = c \cdot T$  cell slots per frame.

At the start of a frame period, if FIFO  $f_0$  contains more than  $v$  cells, then the excess cells are discarded. In general, if

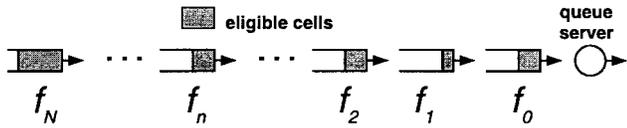


Fig. 2. Basic queue architecture of the DFQ protocol, indicating relative service priorities.

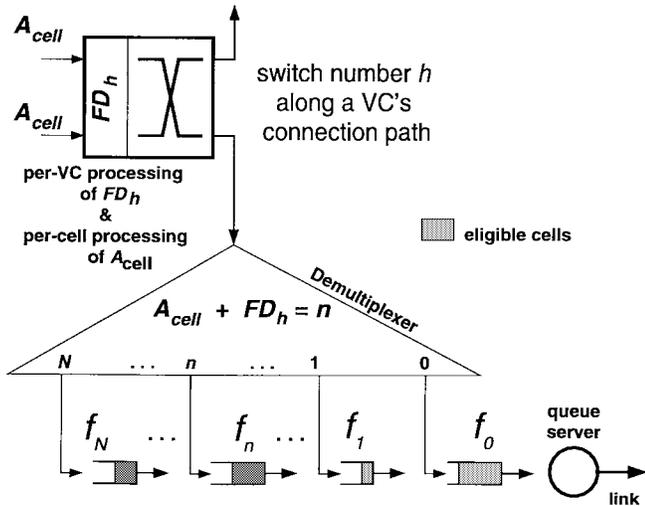


Fig. 3. DFQ buffering protocol based on per-VC  $FD$  assignment and on per-cell  $A_{cell}$  header value.

at the start of a frame period the FIFO's  $f_0, \dots, f_{n-1}$  contain in total more than  $n \cdot v$  cells, then *scheduler saturation* will occur within the next  $n \cdot T$  seconds. The CLR guarantees can be managed by selectively discarding cells from among these  $n$  FIFO's. Conversely, if the first  $n$  FIFO's contain, in total, no more than  $n \cdot v$  cells, then all  $n$  FIFO's will be serviced during the current frame period.

The FIFO identification scheme is periodic, and synchronized with the link-level frames. At the end of each frame period, each FIFO index (subscript) is reduced by one in a modulo  $N$  fashion. As a result, once a new frame begins its transmission, the link server has  $T$  seconds to transmit all cells within logical FIFO  $f_0$ . Once the  $T$  seconds expire, FIFO  $f_0$  is cleared of its remaining cells and renamed  $f_N$ , and it is logically repositioned at the end of the sequence of  $N + 1$  FIFO's, while the index  $n$  of all of the other logical FIFO's is decreased by one, e.g.,  $f_n \rightarrow f_{n-1}$ . Thus, the service priority of a queued cell increases as the index  $n$  decreases for the FIFO in which it is buffered. In graphical terms (Fig. 2), cells which remain queued (unserved) move closer and closer to the link server as time progresses in intervals of  $T$  seconds.

**Buffering Protocol:** We have described how the service priority and deadlines of queued cells are represented by their location among a row of FIFO's. We now describe the protocol that determines into which FIFO a cell is buffered upon its arrival at the queue of the  $h$ th node along its path. This target FIFO is given by  $f_{(A_{cell} + FD_h)}$  where  $FD_h \in \{0, 1, 2, \dots, N\}$  is the local *frame-delay* parameter assigned to the VC at node  $h$  during connection setup, and  $A_{cell} \in \{0, \dots, n, \dots, N\}$  is the index number  $n$  of the FIFO where the cell was buffered at the moment when the cell was transmitted from the upstream

node  $h - 1$ . Thus,  $A_{cell}$  is the number of whole frame periods by which the cell is serviced Ahead of its local deadline at node  $h - 1$ . This buffering protocol is illustrated in Fig. 3.

The first node along a connection assigns the value  $A_{cell} = 0$  to all cells that are entering the network, i.e., by definition, traffic sources (applications) always transmit their cells on time. Also,  $FD_h$  is a constant for each VC, while  $A_{cell}$  will vary from cell to cell. Thus, the earlier the cell gets serviced at node  $h - 1$ , the greater its  $A_{cell}$  value when arriving at node  $h$ , and the further away from the link server it will be buffered at node  $h$ . For a given VC, the minimum subscript for  $f_{(A_{cell} + FD_h)}$  is  $f_{FD_h}$ . This means that all cells of a connection are guaranteed a minimum local eligibility duration of  $FD_h \cdot T$  seconds at node  $h$ . Connections which are assigned long eligibility durations impose a smaller effective load on the queue's deadline requirements. As a result, network management will generally assign large local  $FD_h$  values at congested nodes. Note that at a given node, frame-delay values are assigned on a per-VC basis.

The transmission of the value  $A_{cell}$  entails some amount of transmission overhead. Customized satellite systems generally allow overhead bits to be added to a standardized cell format. We assume that an 8 bit  $A_{cell}$  header ( $N_{max} = 255$ ) is appended to each ATM cell transmitted over the UL and DL satellite channels.

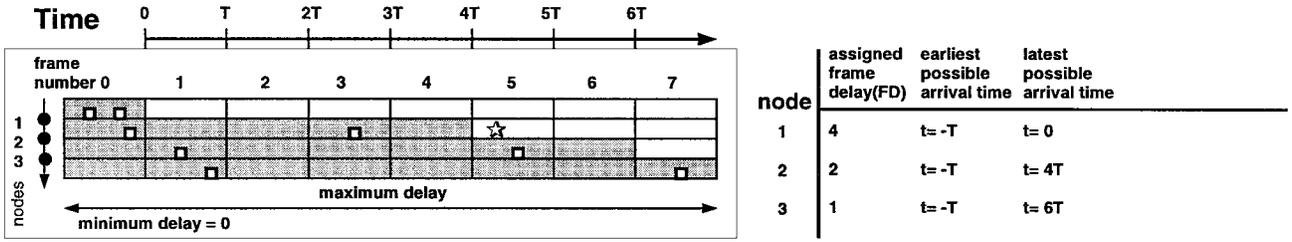
**End-to-End QoS:** The timing behavior of DFQ, across multiple nodes, is depicted in Fig. 4. For illustration purposes, we highlight the transmission of two cells (represented by squares) across a three-node connection. The cells arrive at node 1 within frame number 0. The shaded frames indicate the possible time spans over which the cells may be in transit on each of the four links associated with the three nodes. The end-to-end CTD bound is defined by (1). The minimum delay is achieved if the cell always arrives at empty buffers at each of its  $H$  nodes, and thus experiences zero queue delay. The CDV bound for the work-conserving DFQ is simply given by the upper bound of the CTD. This is called a trivial jitter bound

$$0 \leq \text{CTD} \leq \left(1 + \sum_{h=1}^H FD_h\right)T. \quad (1)$$

In the context of a satellite network, nodes 1, 2, and 3 could be replaced by the UL-Q, sat-Q, and the DL-Q, respectively. Fig. 4 would then suggest that the UL-Q server has continuous access to the UL channel since all of the five frames leaving the UL-Q (node 2) are shaded; see the star  $\star$  symbol. This is not realistic for a multimedia MAC protocol, and will be modified later in Section II-C.

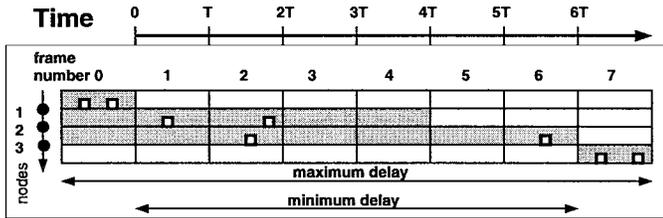
### B. Nonwork-Conserving DFQ

Only a nonwork-conserving protocol is capable of providing a nontrivial jitter bound, i.e.,  $\text{CDV} < \text{CTD}$ . This is achieved by not servicing cells until they become eligible, even if this causes the server to idle while the queue is backlogged. This version of the DFQ protocol is particularly appropriate for the DL-Q server which must directly interact with and conform to the traffic flow specifications requested by end-user applications or policed by UPC contractual agreements.



The VC is assigned FD values of 4, 2 and 1 at nodes 1, 2 and 3 respectively. Two VC cells arrive within frame number 0. Shaded frames indicate the possible time spans over which these two cells may be in transit on each link

Fig. 4. Queue delays across a three-node VC path where all nodes use a work-conserving DFQ protocol.



The VC is assigned FD values of 4, 2 and 1 at nodes 1, 2 and 3 respectively.

Fig. 5. Queue delays across a three-node VC path where the first two nodes use a work-conserving protocol and the last node uses a nonwork-conserving ( $g = 0$ ) DFQ protocol.

Each VC must exit the network, or an administrative zone, at an egress link. By modifying the DFQ protocol at this egress link, a VC can be provided with a tight end-to-end CDV bound. First note that in Fig. 3, for the work-conserving DFQ, any of the  $N + 1$  FIFO's may contain eligible cells. We continue to use the work-conserving DFQ at all but the last node along a connection. At the last node, the nonwork-conserving DFQ protocol does not service a FIFO until  $T$  seconds before its associated deadline. In other words, only cells buffered within FIFO  $f_0$  are made eligible for transmission at the last node. In this way, the new protocol assures that cells which enter the network within a common frame will also exit the network within a common frame; see Fig. 5.

By applying this nonwork-conserving DFQ at the last node  $H$  along a connection's path, the duration over which a cell is eligible for transmission at node  $H$  is  $T$  seconds regardless of the value  $A_{\text{cell}}$  upon its arrival. This indicates that the potential multiplexing gain will be minimal at the last node, this being the cost of nontrivial jitter control. However, the preceding nodes  $(1, 2, \dots, H - 1)$  may maximize their multiplexing gain by maximizing their frame-delay assignments. However, at node  $H$ , the VC should be assigned the minimum possible frame delay, i.e.,  $FD_H = 1$ , as doing otherwise would simply increase the CTD. That is, the CDV and CLR would not be affected if  $FD_H > 1$ . Equation (2) shows that the CDV bound is nontrivial and deterministic, i.e.,  $CDV = 2T$ . The timing behavior of the nonwork-conserving protocol, relative to the

frame-delay assignments in Fig. 4, is illustrated in Fig. 5.

$$CTD = \left( T \sum_{h=1}^H FD_h \right) \pm T. \quad (2)$$

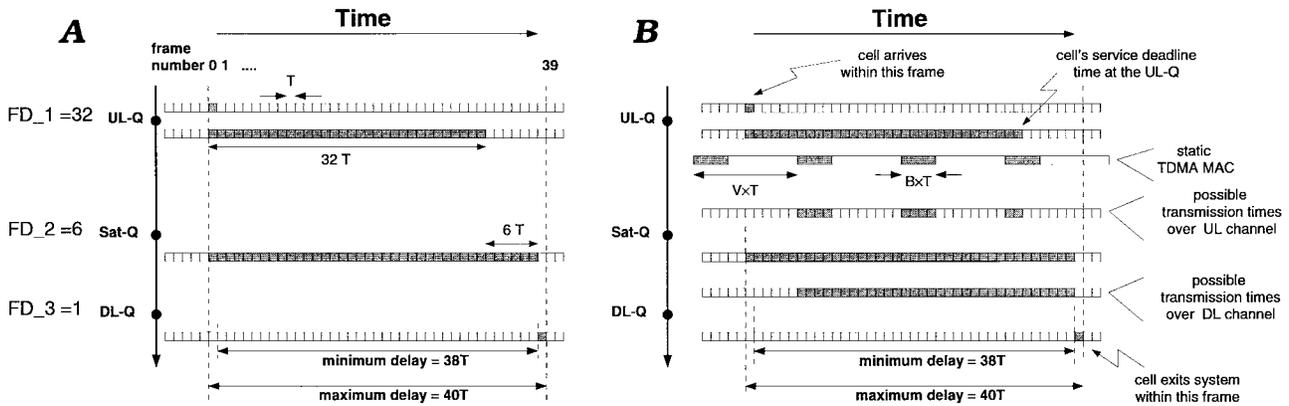
Although the DFQ protocol has been presented here using a simple three-node network, it is equally applicable to arbitrary network topology and transmission speed, much like other frame-based scheduling strategies [13].

### C. Defining a Nominal MAC Protocol

For simulation purposes, we define a nominal MAC protocol which determines the time and duration of UL channel access granted to an UL-Q server. The major purpose of this MAC is to emulate the sporadic nature of the channel access and illustrate how it affects delay and jitter bounds across a satellite network. For simplicity, we assume a static time-division multiple-access (TDMA) MAC protocol where the TDMA frame period equals an integer number of DFQ frame periods  $V \cdot T$  seconds. We also assume that an UL-Q is periodically assigned an access duration equal to an integer number of DFQ frame periods  $B \cdot T$  seconds, where  $B < V$ . Each UL-Q synchronizes its own DFQ  $T$ -frames with the start of a TDMA frame period.

In Fig. 6(a), we present a three-node connection similar to that in Fig. 5, but now the assigned frame delays are 32, 6, and 1 at the UL-Q, sat-Q, and DL-Q queues, respectively. In Fig. 6(b), the same illustration is repeated, but now the static TDMA MAC is superimposed to highlight the times during which a particular UL-Q has access to the UL channel. We consider a cell which arrives at the UL-Q during frame number zero. By performing a logical-AND function between the static TDMA MAC and the 32 shaded frames, we obtain the "possible transmission times over UL channel" during which the cell is both eligible for transmission and has channel access. Once the cell arrives at the sat-Q, it is eligible for transmission until its deadline expires. Thus, the "possible transmission times over DL channel" is a continuous expanse of time which begins at the earliest possible UL access time (as determined by the static TDMA MAC) and ends at the sat-Q's DFQ service deadline.

Fig. 6(b) indicates that a large FD assignment at the UL-Q will allow a VC's traffic to contend for service during several consecutive TDMA frame periods, but that once a



In panel (A), a VC is assigned  $FD$  values of 32, 6 and 1 at the UL-Q, sat-Q and DL-Q, respectively. The DL-Q server uses the nonwork-conserving ( $g=0$ ) DFQ protocol. Panel (B) illustrates the additional transmission constraints imposed by a static TDMA MAC.

Fig. 6. How a nominal MAC affects the allowed transmission times across the UL and DL satellite channels.

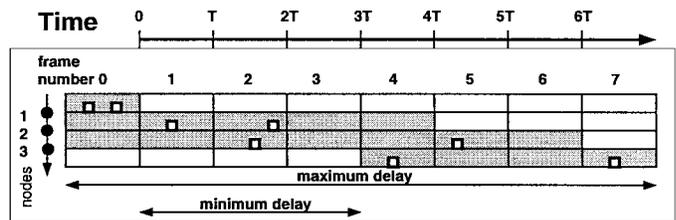
cell's deadline expires, it is discarded, even if the UL-Q has channel access and even if there are no other contending cells. This is what is meant by delay guarantees. We emphasize that this nominal MAC is only chosen for illustration purposes, such that our satellite simulations, presented in Section V-A, incorporate an important characteristic seen among systems using a shared transmission medium.

### III. FLEXIBLE AND MULTIPLE JITTER BOUNDS

The DFQ protocol is primed for several significant network management and service enhancements which can be readily incorporated into the simple, shared FIFO service architecture. We have specified the DFQ protocol such that it provides either trivial jitter control as in (1) or very tight jitter control as in (2). Many other intermediate CDV bounds can be provided by a DFQ service queue by simply modifying the number of eligible FIFO's at the egress link. In addition, multiple CDV bounds can be serviced simultaneously while maintaining a simple queue shared by the VC population.

Consider a set of  $N + 1$  FIFO's where only cells within FIFO's  $f_0, \dots, f_g$  are eligible for transmission,  $g \in \{0, 1, \dots, N\}$ . The upper bound of CTD will remain unchanged  $(1 + \sum_{h=1}^H FD_h)T$ , but the jitter bound will be relaxed to  $CDV = (g + 2)T$ . As an example, the timing behavior of this nonwork-conserving protocol where  $g = 3$  is illustrated in Fig. 7. Here, we assume the same frame-delay assignments as described for Fig. 4, and in particular,  $FD_3 = 1$ . Note that the DFQ FIFO's at the terminating node provide a relaxed jitter bound of  $5T$ . As a result, the network has the option of increasing  $FD_3$  to as much as 4. This would increase the end-to-end delay bound from  $7T$  (Fig. 7) to  $10T$ , and would result in a lower CLR since all cells of that VC would now have a minimum eligibility duration of  $4T$  at node three, rather than  $T$  as in Fig. 7.

If multiple CDV bounds are to be simultaneously serviced at a queue, then each level of CDV QoS requires its own set of  $N + 1$  FIFO's. For example, in Fig. 8, connections can choose from three levels of jitter control: trivial, moderate, or tight. The resulting service queue contains a total of  $3(N + 1)$



VC is assigned  $FD$  values of 4, 2 and 1 at nodes 1, 2 and 3 respectively.

Fig. 7. Queue delays across a three-node VC path where the first two nodes use a work-conserving protocol and last node uses a nonwork-conserving ( $g = 3$ ) DFQ protocol.

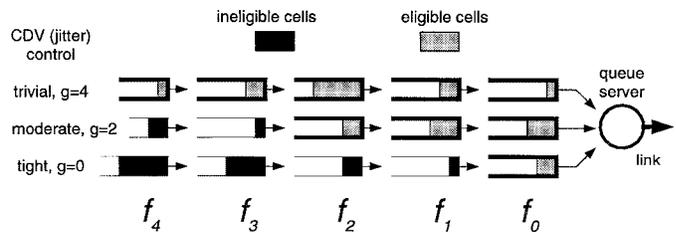


Fig. 8. Expanded DFQ queue architecture such that a server can offer three possible CDV bounds to its VC population ( $N = 4$ ).

FIFO's. The three sets of FIFO's are serviced in *parallel*, meaning, for example, that the three  $f_0$  FIFO's have the same priority and can thus be serviced, in full or in part, in any arbitrary order. As with regular DFQ, the three  $f_0$  FIFO's have service priority over the three  $f_1$  FIFO's, and so on. Note that each VC must be permanently assigned to one and only one set of FIFO's in order to maintain a given CDV QoS level.

As mentioned in the Introduction, this type of CDV control is best suited to the DL-Q which must directly provide jitter bounds to real-time applications [14], [15], and also provide CDV tolerance (CDVT) bounds to the UPC at the network interface. However, providing multiple jitter bounds could also be extended to the sat-Q. This is because a low-power, inexpensive, mobile ES may only be equipped with

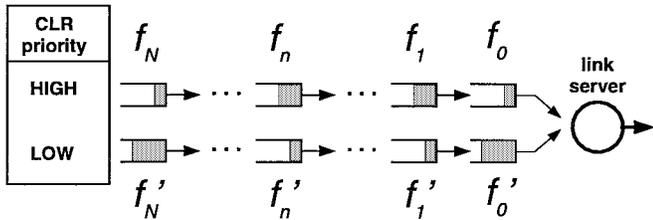


Fig. 9. Expanded DFQ queue architecture such that a UL-Q can offer two CLR bounds and a DL-Q (or sat-Q) can offer multiple CLR bounds to its VC population.

a small DL-Q. Thus, controlling the burstiness at the DL-Q input may be critical for minimizing buffer overflow. Another consideration is that a tight jitter bound limits the input overload at the ES's processor, and thus prevents the occurrence of receiver livelock [16] where bursty processing of newly arriving inbound packets overwhelms the limited CPU bandwidth and interferes with the smooth execution of playback processes.

#### IV. CLR MANAGEMENT

Connection admission control (CAC) algorithms are congestion management tools which rely on the validity and accuracy of user-declared traffic parameters. These algorithms are generally very conservative, and this leads to network underutilization. Finally, they do not provide any practical guarantee of the actual QoS that is being delivered. These issues are the driving force for QoS monitoring and robust CLR management.

In general, ATM traffic will require CLR bounds ranging from  $10^{-2}$  to  $10^{-9}$ . Independently maintaining and isolating these divergent CLR guarantees for individual VC's will require some type of control mechanism. The inherent ability of DFQ to locally detect and diagnose scheduler saturation allows selective cell discard mechanisms to be implemented. Scheduler saturation occurs when one or more queued cells cannot meet their service deadlines due to contention for the link bandwidth. This is the only type of cell loss considered in this paper. Note that if an UL-Q or a sat-Q cannot detect scheduler saturation and selectively discard, then it will transmit cells which are of no value to the user's application, thus wasting a limited satellite resource, bandwidth.

##### A. CLR Strategy for DL Queue and sat Queue

The following CLR management strategy is only applicable in cases where the queue server has continuous access to the channel, i.e., sat-Q and DL-Q. The basic DFQ queue is associated with a single set of FIFO's  $f_0, \dots, f_n, \dots, f_N$  which is shared by all queued VC's. Consider a second identical set of FIFO's which is serviced in parallel with the first set of FIFOs; see Fig. 9. At any given time, each VC is assigned to one of the two sets of FIFO's. The VC's assigned to the high CLR-priority set of FIFO's will be fully protected from moderate scheduler saturation conditions. Conversely, cells will be preferentially discarded from VC's assigned to the low CLR-priority set of FIFO's when scheduler saturation occurs.

In this case, the term *parallel servicing* means that the FIFO index  $n$  is the most significant determinant of service priority, while the CLR priority of a FIFO is only a secondary determinant of its service priority. Thus, the service priorities of the FIFO's in Fig. 9, from highest to lowest, are  $\{f_0, f'_0, f_1, f'_1, \dots, f_n, f'_n, \dots, f_N, f'_N\}$ . In a similar fashion as with the regular DFQ protocol, cell loss only occurs if the queue server cannot service all of the cells within FIFO's  $f_0$  and  $f'_0$  during a given frame period  $T$ .

Queues (nodes) along a VC path will negotiate for a *fair share* of the total  $FD$  and the total CLR allowed by the delay and loss requirements of the real-time application. Thus, at each node  $h$ , a VC is permanently assigned a local frame delay  $FD_h$  and a local  $CLR_h$  such that their sums, over the connection path, correspond to the application's delay and CLR requirements, respectively. Congested nodes will, in general, be allocated the majority of the total allowed  $FD$  and CLR. If a VC's locally measured losses temporarily exceeds its local CLR bound, then that VC is locally assigned to the high CLR-priority set of FIFO's. Otherwise, the VC is locally assigned to the low CLR-priority set of FIFO's. Thus, a VC's end-to-end CLR is controlled in a distributed fashion, and does not require interaction between nodes, except during connection setup.

Proper cell sequencing is not affected when a VC makes a transition between the two sets of FIFO's since both sets of FIFO's are serviced in parallel. The only requirement is that a VC not simultaneously place cells within two parallel FIFO's with the same index, e.g.,  $f_n$  and  $f'_n$ . The upper frequency of these transitions is only bounded by the frame period  $T$ , to maintain cell sequencing. A rudimentary measurement of CLR might consist of recording discarded cells and using the declared sustainable cell rate (SCR) parameter combined with some long-range averaging period. Note that extremely low CLR bounds may require several minutes of monitoring to assess the true mean CLR. However, a VC can be reassigned to the high CLR-priority set of FIFO's at the first sign of cell loss; there are no hard restrictions to the number of transitions a VC may undergo between each set of FIFO's.

In our network simulations, we assign CLR bounds on a per-VP basis. The particular algorithm for determining which of the two CLR-priority queues to assign a VC is as follows. During each frame period  $T$ , the actual CLR is determined on a per-VC and a per-VP basis by using the SCR value. If the actual CLR of a VP exceeds its target bound, then all VC's of that VP are assigned to the high CLR-priority queue, any VC that has a CLR which is greater than its target bound is also assigned to the high CLR-priority queue, and all remaining VC's are assigned to the low CLR-priority queue.

An advantage of this CLR control strategy is that it allows established connections to be isolated from new connections. Assume that the network assigns very large CLR bounds to new connections, i.e., orders of magnitude larger than the actual requested CLR bound. Then, if a node becomes congested, the established VC's will tend to be assigned to the high CLR-priority FIFO's, while the new VC's will remain in the low CLR-priority FIFO's and will experience the majority of cell discards. The new connection(s) can then be discontinued or can renegotiate a lower QoS. If no congestion

is present, this will be confirmed over time by the measured loss rates, and the new VC's can be reassigned their proper CLR bounds.

Note that end-to-end CTD/CDV bounds are intrinsically accounted for since they are guaranteed for all cells which are not discarded. The proposed strategy offers complete service isolation among the VC population, and yet is implemented with a scalable, shared buffer architecture, and allows an arbitrary number of QoS levels for CLR. This management scheme greatly reduces the need for highly accurate or highly conservative CAC algorithms which may result in high connection rejection rates.

### B. CLR Strategy for UL Queue

The UL-Q does not have continuous access to the channel. This situation requires a slight modification to the previously proposed CLR control protocol. Again assume a general queue architecture as depicted in Fig. 9. Once the UL-Q gains access to the channel, the service priorities of the FIFO's from highest to lowest are now defined by three sequential lists. The first list of FIFO's (highest service priorities) are given by  $\{f_0, f_1, f_2, \dots, f_{V-1}\}$ , the second list of FIFO's (intermediate service priorities) is given by  $\{f'_0, f'_1, f'_2, \dots, f'_{V-1}\}$ , and the third list of FIFO's (lowest service priorities) is given by  $\{f_V, f'_V, f_{V+1}, f'_{V+1}, \dots, f_N, f'_N, \dots, f_N, f'_N\}$ .

Recall that  $V \cdot T$  seconds defines the static TDMA frame period that was defined for our simple MAC protocol. Assuming the static TDMA MAC case, the reasoning of the protocol is as follows. Cells buffered within the high CLR-priority FIFO's belong to VC's which have experienced excessive losses relative to their target CLR. Thus, enough FIFO's must be serviced such that no cells will be discarded during the period during which the UL-Q does not have access to the channel. This number of FIFO's is simply the TDMA frame period divided by  $T$ , i.e.,  $V$  FIFO's. This accounts for the first list of FIFO service priorities. Once these have been fully serviced, the server then similarly services the first  $V$  FIFO's in the low CLR-priority queue. This strategy minimizes the losses among the aggregate VC population. Finally, if there remains any time left during the  $B \cdot T$  seconds of the channel access duration, then the remaining FIFO's are serviced in a parallel fashion, as described by the third list of FIFO's.

A second significant modification to the original CLR control protocol must also be made. The FIFO service priority lists  $\{f_0, f_1, f_2, \dots, f_{V-1}\}$  and  $\{f'_0, f'_1, f'_2, \dots, f'_{V-1}\}$  indicate that the FIFO index  $n$  is no longer the main determinant of the service priority, for example,  $f_3$  has service priority over  $f'_1$ . As a result, we can no longer allow a VC to make any transitions between the two CLR-priority sets of FIFO's  $f_n$  and  $f'_n$ ; else, we can no longer guarantee that a VC's cells will maintain their proper cell sequence. Thus, a VC must be permanently assigned to either the low or the high CLR-priority FIFO's. This means that there is no longer an absolute need to measure the actual CLR since these assignments are permanent.

This CLR protocol relies on *a priori* knowledge of the time period between each successive channel access, i.e.,  $V \cdot T$  is a constant. Some MAC's may generate variable channel access

times; in this case, the UL-Q would replace the value  $V \cdot T$  with a worst case (maximum) value, which must exist if real-time service is being provided.

## V. SATELLITE NETWORK SIMULATION RESULTS

### A. CTD and CDV Guarantees Across an (UL-Q, DL-Q) Pair

The purpose of the following simulations is to illustrate how a UL-Q can simultaneously provide multiple CTD bound guarantees to its VC population. In addition, the simulation shows how the DL-Q then provides tight CDV bounds. Only one UL-Q and one DL-Q are simulated. The satellite contains no queues, and does not contribute to queue delay. The DL-Q uses the nonwork-conserving protocol ( $g = 0$ ).

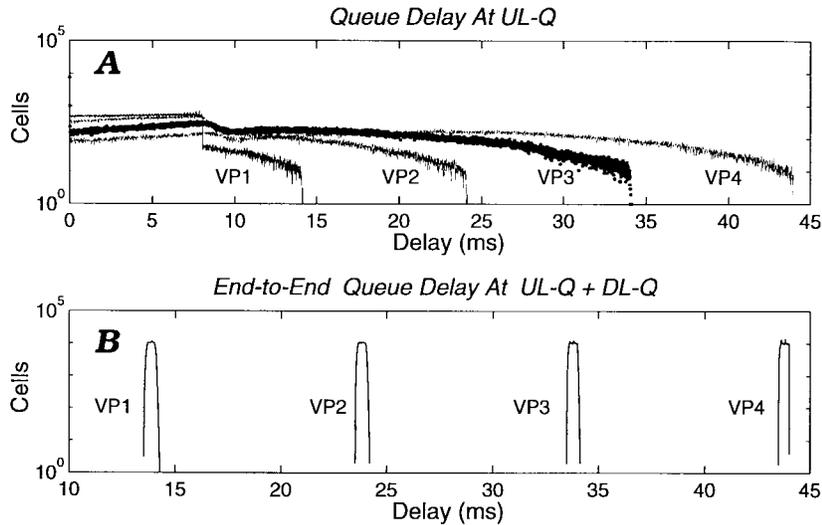
The simulation uses the following parameter values: DFQ frame period  $T = 0.5$  ms (50 cell slots), TDMA frame period  $V \cdot T = 10$  ms (1000 cell slots), and the MAC channel access duration  $B \cdot T = 2$  ms (200 cell slots). The simulation lasts for  $10^7$  cell slot periods. The nominal satellite bandwidth is  $(53 \times 8 \times 50)$  bits/0.5 ms = 42.4 Mbits/s.

Each VC is modeled as a generalized ON/OFF source. The ON duration varies randomly, with uniform distribution, from 0 to 8 ms. The OFF duration varies randomly, with uniform distribution, from 10 to 160 ms. While ON, the source transmits in a CBR fashion, one cell within each 20th cell slot. The UL-Q multiplexes four VP's, and each VP contains 21 VC's. The link utilization of the UL-Q is 0.966. Each VC within a VP is assigned an identical frame delay as described in Fig. 10.

The broad variation in the UL-Q delays for any given VP, shown in Fig. 10(a), indicates how CDV will be greatly affected by the sporadic UL channel access provided by the MAC. The absence of long tail distributions suggests that the UL-Q is not needlessly transmitting cells which have already exceeded their deadlines. The much smaller delay variation shown in Fig. 10(b) confirms that when a cell leaves the DL-Q, its total queue delay (UL-Q + DL-Q) varies by less than 1 ms or  $2T$ . The simulation results show little difference between the upper delay bounds in panel (a) versus those in panel (b). Thus, a tight CDV bound has been obtained at minimal cost in terms of increased CTD.

There are two motivations for providing multiple queue delay bounds on a per-VC basis at the UL-Q. First, some real-time applications (e.g., interactive video) may require smaller delay bounds than others (e.g., video retrieval). Also, some connections may have a much greater wireline propagation delay, and thus benefit from a smaller queue delay across the satellite network. Second, those VC's which can be assigned a greater queue delay will tend to cause less scheduler saturation, and will thus allow greater statistical multiplexing gain. For example, the mean CLR for the aggregate traffic in Fig. 10 is  $2.7 \times 10^{-3}$ . If all four VP's are assigned a frame delay of 28 at the UL-Q, the mean CLR then increases by more than one order of magnitude, to  $3.3 \times 10^{-2}$ . Even with tight jitter bounds, no cells are discarded at the DL-Q since the relative load at the DL-Q is five times smaller.

Queue delays on the order of 10 ms are relatively insignificant compared to the propagation delay (250–280 ms) for



$T = 0.5\text{ms}$ .  $FD$  assignments at UL-Q are 28, 48, 68 and 88 for VPs 1, 2, 3 and 4, respectively. All VPs are assigned  $FD = 0$  at DL-Q. Static TDMA MAC is used at UL-Q.

Fig. 10. Queue delay and jitter bounds across a UL-Q and a nonwork-conserving DL-Q.

geosynchronous earth orbit (GEO) satellites. However, such queue delays are well within the range of propagation delay (20–25 ms) for low earth orbit (LEO) satellites. Thus, the importance of providing short delay bounds will be greatest for LEO satellites. Controlling CDV is critical to both types of satellite orbits, but in particular to GEO satellites, where long propagation delays make dynamic MAC protocols less efficient, thus increasing the jitter at the UL-Q.

A critical feature to discern from this simulation is the fundamental point stated earlier, that the DFQ strategy only defines the service priorities, but does not specify a particular MAC. That is, for any given combination of system parameters (MAC algorithm/traffic load/traffic burstiness/delay bound targets), the DFQ strategy will assure that the UL-Q server prioritizes cell transmissions such that delay violations are minimized for the given MAC protocol. To further optimize the system performance, a dynamic MAC would need to determine which UL-Q (earth station) needs immediate media access as well as the required duration of this access. However, this improved MAC would not change the transmission priorities that the UL-Q DFQ protocol would assign to its queued cells once it gains access to the link.

### B. Flexible and Multiple CDV Guarantees Across a sat-Q + DL-Q Pair

The purpose of the following simulations is to illustrate how a DL-Q can simultaneously provide two different CDV bound guarantees to its VC population. The network simulation consists of a sat-Q and a DL-Q. The DL-Q uses the nonwork-conserving protocol with two sets of FIFO's (Fig. 8) and various values of ( $g$ ). The simulation uses a DFQ frame period  $T = 0.5$  ms (32 cell slots), and lasts for  $10^7$  cell slot periods. The nominal satellite bandwidth is 27.1 Mb/s.

Each VC is modeled as a generalized ON/OFF video source which bursts every 62.5 ms, and has a mean bit rate of 0.54 Mb/s. The source has a random active duration, uniformly

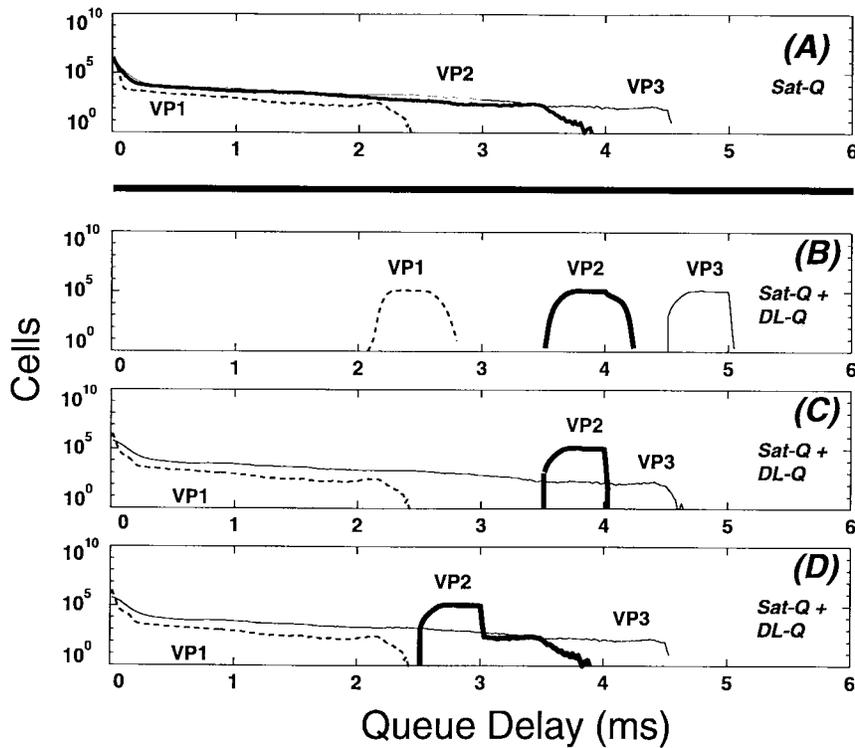
distributed, from 10 to 40 ms in duration. When ON, the source generates a cell within each 20th cell slot. The sat-Q has continuous access to the downlink and multiplexes three VP's; each VP contains 15 independent VC's. The link utilization is 0.9. Each VC within a VP is assigned an identical frame delay as described in Fig. 11.

In these simulations, the DL-Q server has twice the nominal bandwidth (64 slots/0.5 ms). This avoids cell loss when tight jitter control is implemented. As previously mentioned, zero loss at the DL-Q is a reasonable assumption for several reasons. The DL-Q performs various combinations of CDV recovery as described below.

*Queue Delay at Sat-Q:* Fig. 11(a) shows the queue delay distribution for cells at the sat-Q. From (1), we expect absolute delay bounds of 2.5, 4.0, and 5.0 ms for VP's 1, 2, and 3 respectively. The aggregate CLR at the sat-Q, due to scheduler saturation, is  $5.46 \times 10^{-5}$ . No cells are lost at the DL-Q. Fig. 11(b)–(d) shows the total queue delay (sat-Q + DL-Q) as a function of the CDV bounds defined at the DL-Q. Note that the results in Fig. 11(a) remain valid for all of the next three examples.

*End-to-End Queue Delay with Tight Jitter Bound:* In Fig. 11(b), the nonwork-conserving protocol ( $g = 0$ ) is used at the DL-Q; thus, only FIFO  $f_0$  is made eligible for transmission. All three VP's use the same set of FIFO's; thus, they share a common jitter bound of  $2T$  or 1 ms. Fig. 11(b) shows the end-to-end queue delay distributions of the VP's. From (1), we expect upper delay bounds of 3.0, 4.5, and 5.5 ms for VP's 1, 2, and 3, respectively. These delay bounds are also valid for the next two simulations.

*End-to-End Queue Delay with Multiple Jitter Bounds:* In Fig. 11(c), two CDV bounds are simultaneously guaranteed at the DL-Q; thus, there are two sets of FIFO's in the queue. VP's 1 and 3 are serviced by the work-conserving DFQ  $g \geq 9$ , while VP2 is serviced by the nonwork-conserving DFQ ( $g = 0$ ). Thus, all FIFO's within VP's 1 and 3 are eligible for transmission; consequently, they receive a trivial jitter bound



$T = 0.5$  ms.  $FD$  assignments at sat-Q are 4, 7 and 9 for VPs 1, 2 and 3, respectively. All VPs are assigned  $FD = 1$  at DL-Q.

Fig. 11. Flexible and multiple queue delay and jitter bounds across a sat-Q and a DL-Q.

since the minimum queue delay is 0. In contrast, only FIFO  $f_0$  is made eligible for transmission for VP2.

*End-to-End Queue Delay with Moderate Jitter Bound:* In Fig. 11(d), the DL-Q behaves in a similar fashion as in Fig. 11(c). However, now FIFO's  $f_0, f_1, f_2$  are all made eligible for transmission for VP2. This is equivalent to  $g = 2$ ; thus, the jitter bound is now  $4T$  or 2 ms. Fig. 11(d) shows the end-to-end queue delay distributions for all three VP's.

### C. Multiple CLR Guarantees at a DL-Q

The purpose of the following simulation is to illustrate how a DL-Q can simultaneously provide multiple CLR guarantees to its VC population. The network simulation consists of a DL-Q which uses the work-conserving protocol with two sets of FIFO's for the CLR priorities (Fig. 9). The simulation uses a DFQ frame period  $T = 0.5$  ms (50 cell slots), and lasts for  $5 \times 10^7$  cell-slot periods.

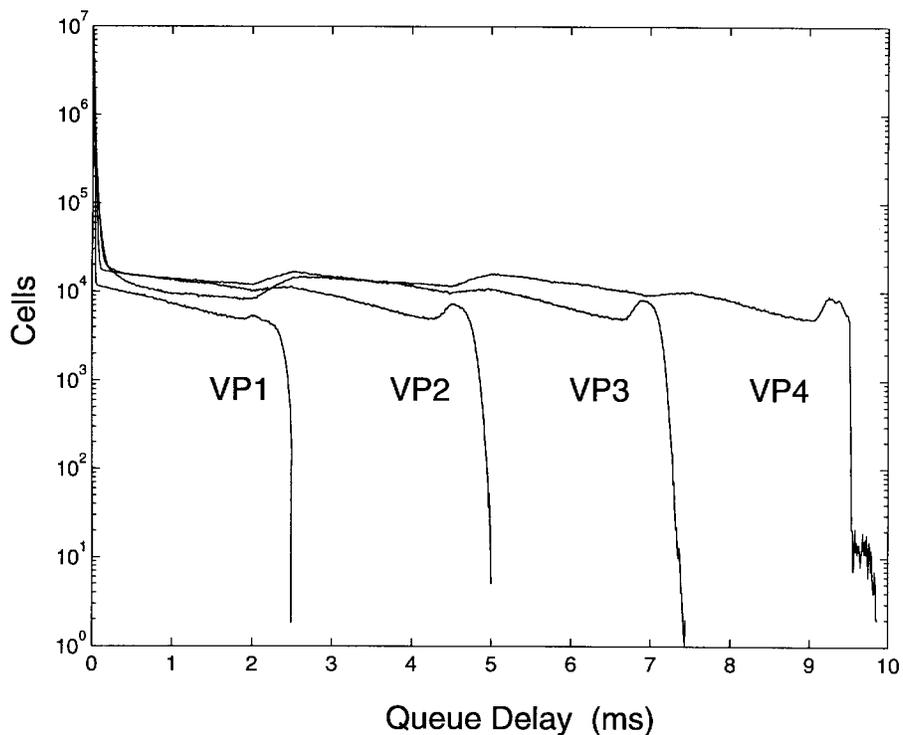
Each VC is modeled as a generalized ON/OFF source as previously defined in Section V-A. The DL-Q multiplexes four VP's; each VP contains 105 VC's. The link utilization is 0.966. Each VC within a VP is assigned an identical frame delay as described in Fig. 12. The DL-Q server performs various examples of CLR control among the four VP's.

Regardless of the individual CLR target bounds that are assigned to each of the four VP's, their queue delay distributions remains relatively constant as depicted in Fig. 12, and the mean CLR of the aggregate traffic remains constant at  $2.458 \times 10^{-3}$ .

*No CLR Control:* In Fig. 13(a), all VP's are assigned a target CLR bound of zero; thus, each VC is permanently assigned to the high CLR-priority queue once it experiences a cell loss. Hence, there is no explicit control over CLR in this simulation. The delay distribution results from this simulation are given in Fig. 12. The resulting CLR's for each of the VP's are shown as four dots in Fig. 13(a). The VP's which are assigned larger frame delays experience CLR bounds which are orders of magnitude lower. This result deserves some explanation. First, we recall that, by definition, a FIFO service discipline requires that the *head-of-line* cell is the first cell to be serviced, while the *end-of-line* cell is the last to be serviced.

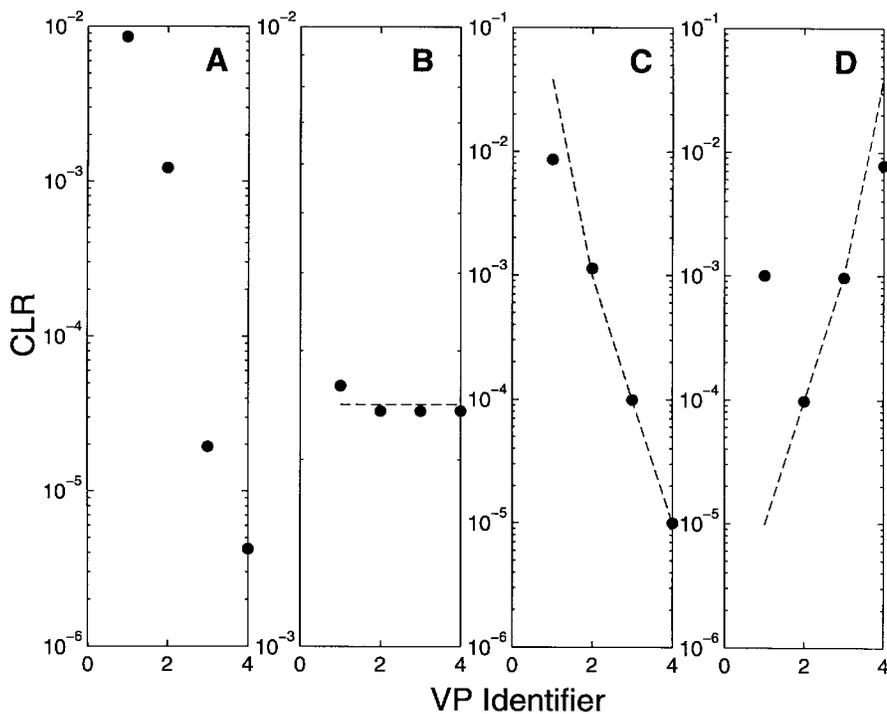
The CLR for VP4 is the smallest because it is assigned the largest frame delay, and thus always places its cells at the head-of-line of its target FIFO  $f_{(A_{cell} + FD_i)}$ . That is, the target FIFO is initially empty before VP4 cells begin entering the FIFO. In contrast, when VP1 cells are loaded into a FIFO, that FIFO may already contain cells from VP's 2, 3, and 4. Thus, VP1 is always the last to place its cells within a FIFO, and its cells are always at the end-of-line. When scheduler saturation occurs at the  $f_0$  FIFO, our particular simulation algorithm always discards the cells located at the end-of-line; hence, VP1 is the most likely to experience cell discards. Note that it would be equally valid for another algorithm to discard head-of-line cells or intermediate cells.

A second reason why VP4 experiences a very low CLR is as follows. The DL-Q server is work conserving, and thus services as many FIFO's as possible during each frame period  $T$ . Thus, in general, VP4 is the VP which has the most



$T=0.5$  ms, *FD* assignment at DL-Q are 4, 9, 14 and 19 for VPs 1, 2, 3 and 4, respectively.

Fig. 12. Queue delay bounds at work-conserving DL-Q.



$T= 0.5$  ms, *FD* assignments at DL-Q are 4, 9, 14 and 19 for VPs 1, 2, 3 and 4, respectively. Simulation results ( $\bullet$ ), and targeted CLR bound (---). Associated DL-Q delay bounds are given in Fig. 12.

Fig. 13. Various CLR target bounds at the DL-Q.

significant fraction of its cells serviced while they are buffered within FIFO's which have indexes  $n$  greater than zero. Since cells are only discarded from FIFO  $f_0$ , VP4 is best isolated from scheduler saturation.

*Uniform CLR Target:* In Fig. 13(b), each of the four VP's is assigned a CLR equal to the mean aggregate CLR ( $2.458 \times 10^{-3}$ ). The resulting CLR's for each of the VP's are shown in Fig. 13(b). The dashed line indicates the targeted CLR bounds. All four VP's experience CLR's which are very close to their targeted bounds.

*CLR Targets that Decrease as the Assigned Frame Delay Increases:* In Fig. 13(c), the four VP's are assigned different CLR target bounds, such that the VP's assigned larger frame delays are given smaller CLR targets. These targets are indicated by the dashed line. The resulting CLR's for each of the VP's are shown with four dots. All four VP's experience CLR's which meet or exceed their CLR targets.

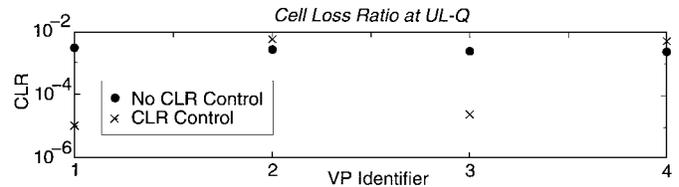
*CLR Targets that Increase as the Assigned Frame Delay Increases:* In Fig. 13(d), the four VP's are assigned different CLR target bounds, such that the VP's assigned larger frame delays are given larger CLR targets; see dashed line. The resulting CLR's for each of the VP's are shown with dots. VP's 2, 3, and 4 experience CLR's which meet or exceed their CLR targets; however, VP1 experiences a CLR which is two orders of magnitude greater than its target.

The simple explanation for the failure of VP1 to meet its target is that this simulation case is the most stringent among the four simulations presented. This is because it is doubly difficult to provide a VC both with a very low delay and a very low CLR. The scheduler is simply overloaded. We point out, however, that our DFQ-based scheduling strategy explicitly makes this overload condition observable to the network management.

#### D. Two-Level CLR Guarantees at an UL-Q

The network simulation described in Section V-A and Fig. 10 is further explored. In particular, the UL-Q CLR results are given in Fig. 14 (dots). Recall that only a single set of FIFO's was used; thus, no explicit CLR control was provided, yet the four VP's obtained similar CLR bounds. This is in contrast to Fig. 13(a), where CLR varies significantly with the assigned  $FD$ . In Fig. 13(a), the DL-Q has continuous media access, and thus cells are only discarded from FIFO  $f_0$ , and end-of-line cells are selectively discarded. In Fig. 14 (dots), the UL-Q has media access only once every  $20T$ , and consequently, cells may be discarded from FIFO's  $\{f_0, f_1, f_2, \dots, f_{19}\}$ . That is, complete FIFO's may be discarded, meaning that a VC is no longer protected from discard even if its cells are generally head-of-line cells.

We repeat the network simulation from Section V-A, now using the CLR protocol proposed in Section IV-B. The UL-Q does not have continuous access to the channel; thus, the CLR protocol can only offer two levels (high, low) of CLR control to the entire VC population, and no explicit CLR targets are defined. VP's 1 and 3 are assigned to the high CLR-priority queue, while VP's 2 and 4 are assigned to the low CLR-priority queue. Fig. 14 (crosses) shows that there is nearly three orders of magnitude difference between the high and low CLR priorities.



Associated UL-Q delay bounds are given in Fig. 10.

Fig. 14. CLR control at the UL-Q.

When we provide two levels of CLR, the mean aggregate CLR is slightly higher ( $2.71 \times 10^{-3}$ ) than when no CLR control is provided ( $2.58 \times 10^{-3}$ ). This small discrepancy is simply due to the fact that the modified FIFO priorities  $\{f_0, f_1, f_2, \dots, f_{V-1}\}$  and  $\{f'_0, f'_1, f'_2, \dots, f'_{V-1}\}$  given in Section V-A, are nonoptimal with respect to servicing low CLR-priority cells before their deadline; however, this modified priority scheme is critical to providing two QoS levels for CLR.

## VI. DISCUSSIONS

Next generation *Ka*-band satellites will need to efficiently service a wide variety of high-bandwidth multimedia applications, requesting a diverse mix of ATM-based QoS guarantees. This situation will induce future satellite systems to take on many of the high-performance characteristics currently associated with wireline ATM schedulers. In particular, queue servers within satellite subsystems will need to address issues such as speed, scalability, complexity, independence between assigned delay bound and allocated bandwidth, QoS isolation, and burst formation [4]. DFQ scheduling is a generic and versatile protocol based on a simple, shared buffer architecture with excellent scaling properties. This protocol and its simple FIFO buffer architecture may be used throughout a satellite's subsystem queues.

At the UL-Q, we propose a work-conserving DFQ which allows high throughput of real-time traffic. This scheduler has been efficiently integrated with a static TDMA MAC protocol, but is also capable of interworking with many dynamic MAC protocols as well. Although the MAC causes large delay variations at the UL-Q, the DFQ scheduler optimizes the successful completion of contractual delay bound guarantees. Finally, the DFQ protocol is modified at the UL-Q to provide two levels of CLR-priorities.

At the DL-Q, we propose a nonwork-conserving DFQ which allows CDV bounds to be guaranteed to end-user applications, and CDVT bounds to be guaranteed at the wireline ATM interface. The DL-Q may offer a wide range of CDV bounds by simply modifying the number of eligible FIFO's in the standard DFQ architecture. In addition, several CDV bounds can be simultaneously offered to the VC population of a DL-Q. Finally, an arbitrary number of CLR bounds can be offered to its VC population.

At the sat-Q, several queue architectures are possible, depending on system needs and satellite limitations. The simplest option, requiring the least buffer capacity, is a work-conserving DFQ. If burstiness is an issue at the input of small DL-Q's, then the sat-Q may require the nonwork-conserving DFQ. With

some additional complexity, an arbitrary number of local CLR bounds can be offered at the sat-Q.

Note that DFQ can be used as an interworking unit (IWU), across an arbitrary satellite system, to provide end-to-end CDV bound guarantees. This means that the sat-Q need not utilize the DFQ protocol. For example, Fig. 1 indicates that the UL-IWU and DL-IWU are integrated with the UL-Q and DL-Q, respectively. The UL-IWU generates the  $A_{cell}$  value which passes transparently through the arbitrary satellite queue. At the DL-IWU, the  $A_{cell}$  value is used to properly buffer the cell within the DFQ architecture. In this situation, the frame delay assigned to the UL-Q must also incorporate the maximum queue delay expected at the sat-Q.

Tight CDV bounds can be freely associated with large or small CTD bounds. Thus, jitter-sensitive real-time sessions which can tolerate longer delays can be more efficiently multiplexed at the UL-Q. Although the DFQ protocol is based on a single frame size  $T$ , it provides a wide range of CTD bounds and CDV bounds. Its statistical multiplexing efficiency compares favorably [8] with that of high-performance ATM schedulers such as self-clocked fair queueing [4].

The numerous network simulation results presented in this paper demonstrate the tractability of the DFQ protocol. The DFQ strategy provides *explicit* CTD/CDV bounds which are not a function of the changing load conditions. Thus, the end-to-end QoS is obtained using very simple expressions, for example, (1). The only QoS parameter which changes with load is the discard rate (CLR). Controlling this last QoS parameter is the task of the distributed CAC algorithm, combined with the proposed CLR management schemes.

## REFERENCES

- [1] J. Gilderson and J. Cherkaoui, "Onboard switching for ATM via satellite," *IEEE Commun. Mag.*, pp. 66–77, July 1997.
- [2] T. Ors, Z. Sun, and B. G. Evans, "A meshed VSAT satellite network architecture using an on-board ATM switch," *Proc. IEEE Int. Performance, Computing, Commun. Conf.*, 1997, pp. 208–214.
- [3] I. F. Akyildiz and S. H. Jeong, "Satellite ATM networks: A survey," *IEEE Commun. Mag.*, pp. 30–43, July 1997.
- [4] H. Zhang, "Service disciplines for guaranteed performance service in packet switching networks," *Proc. IEEE*, vol. 83, pp. 1374–1396, Oct. 1995.
- [5] H. LePocher, V. C. M. Leung, and D. Gillies, "A distributed protocol to schedule and monitor real-time traffic," in *Proc. IEEE PacRim'97*, Victoria, BC, Canada, Aug. 1997, pp. 531–534.
- [6] ———, "An application and management based approach to ATM scheduling," in *Proc. 2nd IFIP Workshop Traffic Management and Synthesis of ATM Networks*, Montreal, P.Q., Canada, Sept. 1997.
- [7] ———, "Providing explicit delay jitter bounds for real-time traffic over wireless ATM networks," *Proc. Wireless Mobile ATM'98*, Hangzhou, China, Apr. 1998, pp. 194–204.
- [8] ———, "Real-time broadband services with jitter control over ATM satellite bridges," in *Proc. ICC'98*, Atlanta, GA, June 1998.
- [9] ATM Forum, Technical Committee, "Traffic management specification," draft, 1996.
- [10] K. Okada *et al.*, "A study on satellite switched TDMA systems for applying ATM," in *Proc. IEEE ICC'92*, 1992, pp. 355–359.
- [11] A. Ohta, K. Okada, and H. Mizuno, "Performance evaluation of ATM/TDMA conversion protocol by using MMPP arrival model," in *Proc. IEEE GLOBECOM'94*, 1994, pp. 1351–1357.
- [12] S. J. Golestani, "A framing strategy for congestion management," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 1064–1077, Sept. 1991.
- [13] L. Trajkovic and S. J. Golestani, "Congestion control for multimedia services," *IEEE Network*, pp. 20–27, Sept. 1992.
- [14] S. Varma, "MPEG-2 over ATM, system design issues," in *Proc. COMPCON'96*, 1996, pp. 26–31.
- [15] D. J. Wright, "Voice over ATM: An evaluation of implementation alternatives," *IEEE Commun. Mag.*, pp. 72–80, May 1996.
- [16] A. Mehra, A. Indiresan, and K. G. Shin, "Structuring communication software for QoS guarantees," *IEEE Trans. Software Eng.*, vol. SE-23, pp. 616–634, Oct. 1997.
- [17] Z. Zhang and A. S. Acampora, "Equivalent bandwidth for heterogeneous sources in ATM networks," in *Proc. IEEE ICC'94*, 1994, pp. 1025–1031.



**Hervé Le Pocher** received the B.Sc. degree in biology and the B.A.Sc. degree in engineering from the University of Ottawa, and the M.A.Sc. degree from the University of British Columbia.

He is currently in the Product Research Department at PMC-Sierra Semiconductor, Burnaby, B.C., and is also currently pursuing a doctorate in electrical engineering at the University of British Columbia. His research interests include real-time networks and service schedulers, magnetic brain stimulation, 3-D electromagnetic modeling, and EM field sensors for transient and pulsed RF fields. He has held research positions at the University of Ottawa, the Department of Communications Canada, and the Department of Health & Welfare Canada.



**Victor C. M. Leung** (S'75–M'79–SM'97) received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (U.B.C.) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at U.B.C. on a Natural Sciences and Engineering Research Council Postgraduate Scholarship, and received the Ph.D. degree in electrical engineering in 1981.

From 1981 to 1987, he was a Senior Member of Technical Staff at Microtel Pacific Research Ltd., specializing in the planning, design, and analysis of satellite communication systems. He also held a part-time position as Visiting Assistant Professor at Simon Fraser University in 1986 and 1987. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He joined the Department of Electrical Engineering at U.B.C. in 1989, where he is an Associate Professor and a member of the Centre for Integrated Computer Systems Research. He is also a Project Leader in the Canadian Institute for Telecommunications Research, a Network of Centres of Excellence funded by the Canadian Government. His research interests are in the areas of architectural and protocol design and performance analysis for computer and telecommunication networks, with applications in satellite, mobile, personal communications, and high-speed networks.

Dr. Leung is a voting member of the ACM.

**Donald W. Gillies** received the S.B. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1984 and the M.S. and Ph.D. degrees in computer science from the University of Illinois, Urbana-Champaign, in 1990 and 1993, respectively.

Prior to his graduate studies he was an Engineer in the Xerox Office Systems Division, Palo Alto, CA. From 1993 to 1995, he was an Assistant Professor of Electrical Engineering at the University of British Columbia, Vancouver, Canada. Since 1996, he has worked in LEO satellite telephony, first at Motorola where he was involved in the IRIDIUM<sup>TM</sup> project, and currently at Qualcomm, where he is involved in the GLOBALSTAR<sup>TM</sup> project. In 1998, he founded Planetwide Software, Incorporated, a consulting firm with research and development expertise in real-time systems and communications software as well as architecture and performance analysis of real-time systems.