MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

A Unified Traffic Control Scheme for ATM Networks

Qin Zheng, Hugh Lauer, John Howard, Chia Shen

TR94-14 July 1994

Abstract

The paper proposes a generic traffic control specification at the UNI unifying the characteristics of credit-based and rate-based schemes, and allowing implementation of both, along with other traffic control schemes, simultaneously, in an ATM network.

1 Introduction

A common requirement for multi-media network services is that they be provided on and integrated with local area and campus networks that also have to provide traditional services such as file transfer, network file systems, remote terminals, electronic mail, etc. However, multi-media and traditional services are so very different that they pose conflicting requirements on the ability of the network to handle traffic efficiently and effectively.

For example, video streams compressed with MPEG can vary in their bandwidth requirements by a factor of ten or more in a short period of time. Let several of these streams share the same network link, and their peak requirements can combine to effectively shut out all traditional traffic from a link for milliseconds at a time. In older, slower networks, this was not much of a problem because (1) there was plenty of buffering in the nodes at each hop, relative to the bandwidth of the network, and (2) packet times were measured in milliseconds anyway.

High speed networks such at ATM networks, however, operate at an altogether different point in the bandwidth-delay space. The unit of transmission is much smaller – 53 byte cells instead of 1500 byte packets – but so is the number of buffers typically found in each switch node. At 155 megabits per second, for example, it takes over 10,000 cells of buffer to absorb the effect of several I-frames coinciding and consuming all of the link bandwidth for 1/30 second. This is an order of magnitude more buffer space that is typically provided in an ATM switch.

As a result, the burstiness of multi-media traffic in an ATM network can naturally cause severe data losses of non-multi-media and traditional traffic in the same network. Of course, traditional traffic has used end-to-end error detection and flow control to avoid such data losses. However, the operating points of algorithms such as TCP/IP are far outside of the envelope imposed by bursty multi-media traffic in high-speed ATM nets. Thus, to be able to combine multi-media and traditional traffic on the same ATM local or campus network, better flow control is needed, at least for traditional traffic and possible for multi-media traffic as well. It is in the interest of both applications to find such a better method.

2 Rate-based and Credit-based Flow Control

In the ATM Forum, an industry body which both promotes ATM technology and recommends ATM network standards, there are two main philosophical approaches to flow control of non-real-time, non-multimedia traffic [1] in the presence of real-time multimedia traffic which is usually transfered at a higher priority. One approach is called "rate-based" control and is an end-to-end scheme in which sources adjust the rates of transmission as a result of feedback from destinations [1, 2]. When constructing the feedback, destinations take into account such factors as the congestion of the network (as reported by upstream nodes), data losses, etc. In the currently favored method, a source gradually decreases its rate of transmission unless it gets positive feedback from a destination that it can increase it again. The destination sends such feedback so long as it sees an unbroken data stream arriving with no congestion indication. Together, they adaptively try to find the best data rate for their particular connection and the current network conditions.

The other approach is called the "credit-based" control, a link-by-link scheme in which no node of the network forwards cells to the next node without a positive indication that there are sufficient buffers at that next node to receive those cells [3, 4]. Each node must actively

[1] Such traffic is called "ABR", meaning traffic demanding the Available Bit Rate after all other commitments are met.

manage its buffers and regularly transmit credits upstream to the next node for each connection. Thus, the effective data rate of a connection is exactly equal to the rate at which credits are transmitted in the opposite direction.

Each scheme has its advantages and disadvantages. The RB scheme relies on end systems to do traffic control, thus very little intelligence and participation are required on the part of ATM switches. Also, the rate-based paradigm is in accordance with the current standards for ATM traffic management to be implemented on many network interface cards (NICs) [3]. However, the issues of performance and its ability to react to congestion have been raised about the RB scheme. There is no analytical proof or real world experience showing that the RB scheme will be able to provide satisfactory performance and minimize cell losses of ABR traffic, especially in the presence of wildly varying VBR traffic. In particular, we speculate that it connot respond fast enough to the case of of several I-frames of MPEG video completely taking over a link for 1/30 second.

The CB scheme, on the other hand, has a solid proof that lossless transmission can be achieved. With its static buffer allocation algorithm, it can also achieve the maximum link bandwidth utilization. However, since the CB scheme requires coordination and active participation of all switches and NICs in a path, significant changes to the architecture of the existing ATM switches are needed. The switch implementation cost (in terms of processing power and buffer requirement) is also in dispute. Thus a main concern about the CB scheme is whether or not the scheme is mature enough to force it into a standard, possibly obsoleting existing investments and requiring new levels of engineering and development in the switch industry.

As of this writing, neither scheme seems strong enough to prevail over the other. Although there are plenty of simulation results with artificial traffic models, there is not enough convincing real-world experience to evaluate either scheme in depth, especially in the presence of sophisticated, interactive multi-media applications. Thus, there is room for another approach.

In this paper, we propose a unified traffic control specification for the ATM UNI. The proposed specification is generic and flexible enough to implement RB, CB, or even other, yet to be invented, traffic control schemes. Moreover, it does not add significant implementation complexity at a network interface card over that of either a RB or CB scheme alone. This would allow users to simultaneously use different traffic control schemes in the same network. Moreover, it would give vendors the maximum freedom to experiment with innovative approaches to the traffic control problem for which no complete solutions have been found so far. On the other hand, by specifying a set of generic and powerful traffic control functions, vendors will be able to move ahead and at the same time be guaranteed that products complying with the specification will interoperate with each other and be relatively future safe.

3 Description of Unified Scheme

Figure 1 is a sketch illustrating a single connection from a source illustrating a source end system to a destination end system, and cell flows. In this unified system, the source end system

- maintains a credit count C for each VC and update C upon transmission of cells and reception of credit update (CU) cells, and
- controls the cell transmission rate of a VC not to exceed an allowed cell rate (ACR) for that VC and implement a mechanism which dynamically updates ACR.

A destination end system

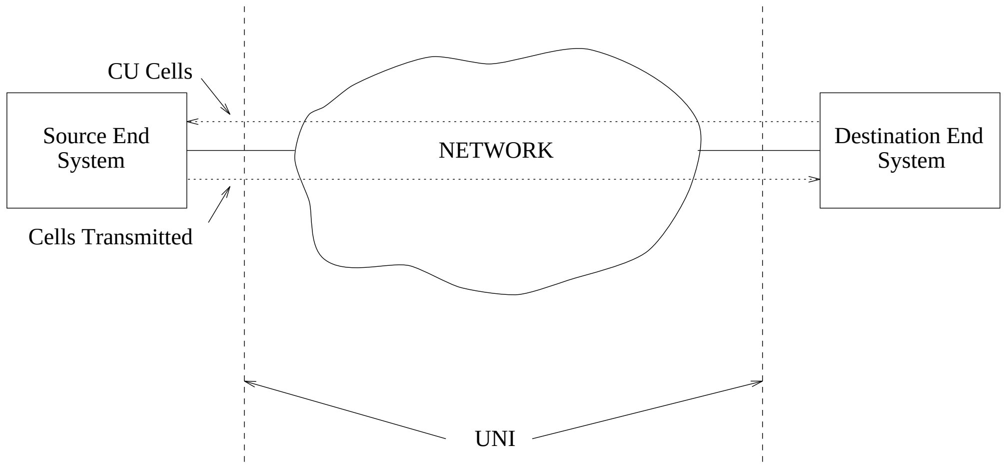

Figure 1. Traffic control at the UNI

- keeps a record for each VC of the numbers of cells received and their congestion status (i.e., whether or not the EFCI bit is set),
- assembles CU cells for each VC according to the number and/or status of cells received and the number of cells forward, and
- sends CU cells back to a node specified by that VC.

In addition,

- a congested switch may set the EFCI bits of cells passing through it to convey congestion information, or remove CU cells, and
- a switch may implement functions specified for source and destination end systems such that it is capable of behaving as a virtual source and destination.

The traffic control functions of a source end system are to be conceptually implemented with three modules as shown in Figure 2: a credit manager (CM) maintaining a credit count C for each VC, an ACR generator (`ACR_G`) calculating the current allowed cell rate for each VC, and a rate controller (RC) controlling the transmission rate of a VC not to exceed its current ACR.

```
(1) Pseudo-Code for CM:

    if (receive CU cell)
        derive the credit update value dC
        C:=min (C_max, C+dC)                          ! credit update
        send C to ACR_G    ! initiate an ACR update

    if (receive a cell_out signal from RC)          ! one cell is transmitted
       C:=max(0, C-1)                                ! credit update
```

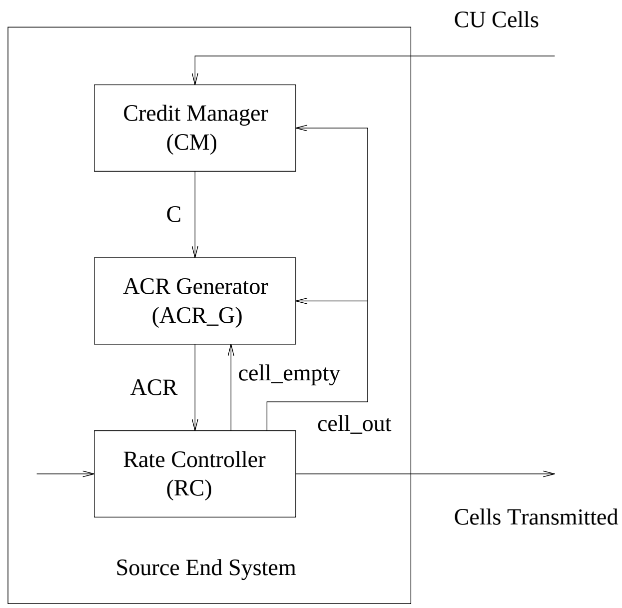

Figure 2. Traffic control at a source end system

```
(2) Pseudo-Code for ACR_G:

    if (C updated from CU cell)                     ! ACR calculation for a new
       ACR:=f1(C)                                   ! credit value

    if (C updated from cell_out signal)             ! ACR calculation upon
       ACR:=f2(C)                                   ! transmission of a cell

    if (receive an cell_empty signal)               ! ACR calculation at the
       ACR:=f3(C)                                   ! next cell transmission
                                                    ! time when a VC is idle

    where f1(), f2(), f3() are specified as discussed below in section 4.

(3) Pseudo-Code for RC:

    if (now >= next_cell_time)
        next_cell_time := now + 1/ACR
        if a cell is available
                transmit a cell
                send a cell_out signal to CM and ACR_G
        else
                send an cell_empty signal to ACR_G
```

As shown in Figure 3, a destination end system is conceptually composed of two modules for traffic control: an Incoming Cell Accountant (ICA) and a CU Cell Generator (CU_G). The ICA maintains three counters for each VC: the total number of cells received N_in, the total

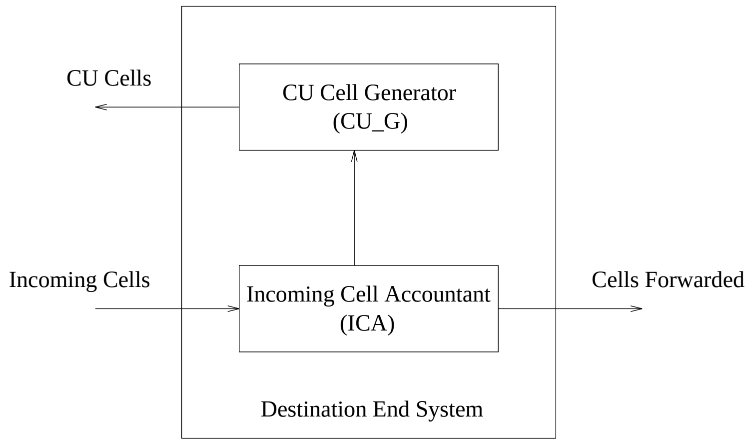

Figure 3. Traffic control at a destination end system

number of cells forwarded `N_out`, and the number of continuous congestion-free (EFCI=0) cells received `N_cf`. According to these variables, the `CU_G` generates CU cells using a set of rules selected by the VC.

```
(1) Pseudo code for the ICA

    if (receive cell)
N_in := N_in + 1
        if (EFCI = 0)
N_cf := N_cf + 1
else
N_cf := 0

    if (forward cell)
N_out := N_out + 1

(2) Pseudo code for the CU_G

    if (RULE(N_in, N_out, N_cf, N_in_old, N_out_old, N_cf_old) = TRUE)
     assemble a CU cell
        send the CU cell to a node specified by the VC
        N_in_old := N_in
        N_out_old := N_out
        N_cf_old := N_cf

    where RULE() is specified as discussed in section 4.
```

Since the proposed specification has the potential of supporting a large variety of traffic control schemes, it is important to have a protocol to select a scheme agreed upon by a source, destination, and intermediate systems at a VC setup time. This protocol will allow the parties to decide whether the VC will be RB, CB, or some other scheme for each individual connection.

Accordingly, the source node will adopt specific functions for f1(), f2(), f3(), and the destination node will adopt a specific RULE().

4 Discussions

The flexibility of the proposed specification comes from the freedom in implementing f1(), f2(), f3(), and RULE() by vendors. We first show that the proposed RB and CB schemes can be implemented by a NIC complying with the proposed specification.

Implementation of the RB scheme:

```
RULE() = (((N_in mod Nrm) == 0) & (N_cf > 0))

f1() = min (ACR + Nrm*AIR + Nrm*ADR, PCR)
  where ADR = ACR/MDF

f2() = max (ACR - ADR, MCR)

f3() = max (ACR - ADR, MCR) if ACR > ICR and ACR otherwise

cell sequence number = N_in
```

where, as described [1], ACR is the Allowed Cell Rate, ADR is the Additive Decrease Rate, AIR is the Additive Increase Rate, MCR is the Minimum Cell Rate, the minimum for ACR, ICR is the Initial/reset Cell Rate for ACR, PCR is the Peak Cell Rate, the maximum rate for ACR, MDF is the Multiplicative Decrease Factor, and Nrm is a constant number which determines the frequency that feedback cells are generated.

Note that a pure RB scheme can be implemented with f1, f2, and f3 independent of the credit count C.

Implementation of the CB scheme:

```
RULE() = (N_out - N_out_old >= N2)

f1() = f2() = PCR if C > 0 and 0 otherwise

f3() = ACR

cell sequence number = N_out
```

where PCR is the Peak Cell Rate, the maximum rate for ACR, and N2 is a constant number which determines the frequency that feedback cells are generated.

The proposed specification does not add significant implementation complexity at a NIC which implements a CB or RB scheme alone. Referring to Figure 2, at a source end system, the specification requires the replacement of a RB NIC's ACR update circuit with CM and `ACR\_G`, and an enhancement of a CB NIC's round-robin scheduling mechanism with a RC, which can, for example, be implemented with a timing chain approach. At a destination end system, per-VC accounting is required to implement a credit update protocol.

The proposed specification can be viewed as a RB scheme enhanced with a credit handling capacity to provide more flexibility in choosing an ACR update policy. Considering the frequency and magnitude of changes that the current RB baseline is experiencing, this flexibility is absolutely necessary to cope with the inevitable problems to be discovered when a RB scheme is implemented and tested in real systems. Adding a credit handling capacity also makes the specification capable of supporting a CB flow control scheme.

The proposed specification can also be viewed as a UNI implementation of a CB scheme enhanced with a rate control to allow an end-to-end implementation of the CB flow control scheme. It has the advantage of not requiring replacement of all existing ATM switches with CB ones at once. An evolutionary approach seems to be the best way that people could possibly accept a new flow control scheme such as the one proposed in the CB baseline.

Because feedback includes credit value, other traffic control schemes are possible. In particular, a source could derive a rate of transmission from time derivative of credit value feedbacked.

5 Conclusion

This paper presents a traffic control system unifying the characteristics of credit-based and rate-based schemes, and allowing implementation of both, along with other traffic control schemes, simulaneously, in an ATM network. We have shown that by adopting specific functions for f1(), f2(), f3(), and RULE(), we can precisely implement either RB or CB scheme. Moreover, we have shown that both schemes can be used in the same network at the same time, on a connection-by-connection basis.

References

[1] ATM Forum/94-0438R1, "Closed-Loop Rate-Based Traffic Management,", July 1994.

[2] N. Yin and M. G. Hluchyj, "On closed-loop rate control for ATM cell relay netorks," in *Proceedings of IEEE INFOCOM'94*, June 1994.

[3] ATM Forum/94-0632, "Credit-Based FCVC Proposal for ATM Traffic Management (Revision R2),", July 1994.

[4] H. T. Kung, R. Morris, T. Charuhas, and D. Lin, "Use of link-by-link flow control in maximizing atm network performance: Simulation results," in *Proceedings of IEEE Hot Interconnects Symposium, '93*, August 1993.