Using Rate Based Flow Control to Manage Available Bit Rate Traffic in Asynchronous Transfer Mode Networks

Robert Walthall
Department of Electronic Engineering Technology
Brigham Young University
Provo, UT 84606
Email walthalr@et.byu.edu

Abstract

The ATM Forum voted to implement the Rate Based Flow Control (RBFC) scheme to manage ABR traffic in Asynchronous Transfer Mode (ATM) Networks. The purpose of this document is to introduce the reader to the qualities of traffic in ATM Networks that necessitate a control system, and to describe the evolution of the RBFC scheme.

1 Introduction

This year at Interop hundreds of vendors claim to support ATM solutions for "your high speed networking needs". The ATM Forum now has more than 500 members including most of the manufacturers of networking products. Most all of the Telecommunications Service Providers (Sprint, AT&T...) now provide ATM services. Even the government has jumped on the bandwagon with ATDNet, a very large and advanced ATM network. It would seem that somewhere within ATM's promise of services and scalableness lies the future of networking. One of the biggest complaints about ATM is it's current lack of standards. While the ATM Forum is working hard to speed up the process of developing these standards, the fact is that the task is one of sifting through a lifetime of work in so many different areas of research and development that it is going to take some time. One of these areas of ongoing research is congestion control a critical element of traffic management in ATM Networks, and the focus of this paper.

While there are a multitude of good introductory papers / text on ATM (to which the reader is referred if he/she is in need of an introduction)[1,2,10], most of the papers that have been written to date on congestion control assume a significant amount of familiarity with the topic and it's terms. Difficulty also arises from the fact that research in congestion control has been pursued in seemingly all possible directions. Only one of which (based on the RBFC scheme) will become a

Last Update: October 7, 1995

standard. In this paper, I have endeavored to present the topics required in order to gain an understanding of the need for flow control, and the basic ideas involved in the development of the RBFC scheme in a manner which requires a minimal amount of familiarity with congestion control, queuing theory, mathematical modeling and the like.

2 Types of Traffic

ATM has been designed to support a number of different types of traffic including voice, data, and video. The traffic has been categorized according to its behavior into the following classifications; Constant Bit Rate (CBR), Variable Bit Rate (VBR), and Available Bit Rate (ABR).

- **CBR traffic** is typically used for voice, because while voice is bursty in nature (a person does not constantly speak while he/she is on the phone) it is not appropriate to only transmit while one is speaking. This would cause the background noise to be eliminated removing the lifelike quality we currently enjoy while holding a phone conversation.

- **VBR traffic** is characterized by compressed video such as Motion Picture Experts Group (MPEG). These compression schemes involve sending an initial frame containing all the data, and then sending updates to that frame. The updates contain only the information that has changed since the initial frame was sent. The chart below is representative of a video with a rate of 30 frames per second. The initial frame contains 3.5 Mbits, and its updates are 1, 2.5, and 4 Mbits respectively. [3]

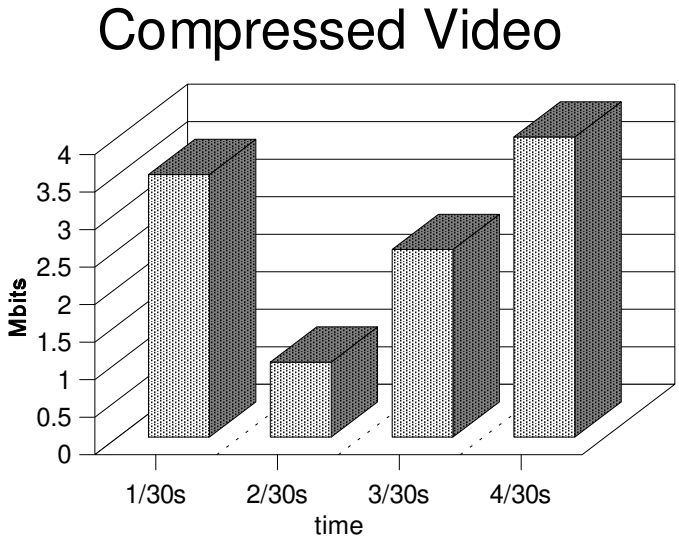

Figure 1

- **ABR traffic** is characterized by applications that are not as sensitive to delay as voice and video. This is not to imply that the user will experience long periods of delay when using ABR service, but simply that to most data applications a few milliseconds of delay will not be noticeable. Most of today's computer applications fall into this category.

The table below summarizes the different types of traffic and their applications.

Best Effort	ABR	data
Guaranteed	VBR	video
	CBR	voice

Table 1

3 Quality of Service

Guaranteed traffic in an ATM network is traffic that can negotiate a Quality of Service (QoS) contract at the time when the connection is being established. This QoS is currently defined by specifying the following parameters. [4]

- **Peak Cell Rate** (PCR) The maximum transmission rate of the end user for the application he/she will be using.
- **Minimum Cell Rate** (MCR) This is a guaranteed rate below which a call will not be required to transmit.
- **Burst Tolerance** (BT) The maximum number of bits that can be transmitted at the PCR.

PCR for figure 1: 4Mb / (1/30s) = 120Mb/s
BT for figure 1: 4Mbits

Other important parameters in the QoS would include Cell Transfer Delay (CTD) the maximum end to end delay experienced by the cell. And Cell Delay Variation (CDV) which is the variation of delay time from one cell to the next. To date most of the switches on the market are not to a point where they are capable of guaranteeing delay parameters other than specifying a maximum possible delay on their specifications sheet. In spite of this, voice, data, and video communications over ATM Networks are prolific in many areas of the country, even the world.

4 Over-Booking the Switches

In order to understand the strategy that is used to over-book the switches it will be necessary to gain a basic understanding of statistical multiplexing. This can be accomplished by first considering the process of time division multiplexing (TDM). TDM is used by the phone companies to carry conversations in T-1 lines. The T-1 line can be divided into 24 64kb/s channels enabling it to carry 24 voice conversations. The 24 64kb/s channels are grouped together into a frame. This frame can be likened to a train [10]. In essence, the frame has 24 freight cars, and it waits a specific amount of time for each of its 24 customers to fill their duly appointed freight cars. If at the end of this time a freight car has not been filled, the frame is sent with that freight car empty.

1 2 3 4 5 ... 24

As was mentioned earlier, people do not usually talk the entire time they are on the phone. So a lot of freight cars are sent empty.

Statistical Multiplexing is simply the ability to use those empty slots. For example, suppose that voice, data, and video are being sent over this connection. Voice requires very little space (bandwidth) when compared with video or data. If the customer on channel one had a lot of data to send, and channels 3 and 4 were being used by voice customers (and were therefore empty most of the time). The data from channel one (or any other busy channel) could be sent in slots 3 and 4 (or any other slots that were empty for that frame).

1 2 1 1 5 ... 24

Because ATM implements statistical multiplexing, and because ATM networks are transmitting VBR as well as CBR and ABR traffic. It is possible to over-book the switches. For example, if the output capacity of the switch is 155Mb/s it is possible to accept 26 customers who specify VBR traffic with a PCR of 10Mb/s.

(26) 10Mb/s = 260Mb/s

260Mb/s will exceed the output capacity of the switch, but since the traffic is VBR (and will therefore not be transmitting at the PCR most of the time) the switches can be over-booked, thus relying on the probability that they will not all burst at once. The cell loss probability can be reduced by buffering the arriving cells so that the switch can tolerate a burst that will exceed its output capacity (155Mb/s). These buffered cells will experience delay while the switch transmits the other traffic, but with small buffers the delay is minimal.

Studies have shown that when using small buffers (4k), 26 clients with a QoS of

PCR 10Mb/s, MCR 1Mb/s, BT 10 can be transmitted in a 150Mb/s system while maintaining a cell loss probability of 10^{-10}. This yields 17% utilization of the switch capacity.[5] Thus, the average transmission rate is 25.5Mb/s

$$(.17)\ 150\text{Mb/s} = 25.5\text{Mb/s}$$

This is obviously undesirable. The switch having an output capacity of 150Mb/s is being grossly under-utilized. By increasing the number of clients the probability of them bursting at once increases which results in an increased cell loss probability. One solution is to increase the buffer size. Then when large numbers of clients burst simultaneously the result is simply increased delay. The study further showed that when large buffers (4Meg) are used the number of clients can be increased to 96 while maintaining a cell loss probability of 10^{-10}, PCR 10Mb/s, MCR 1Mb/s, and BT 10. This yields a switch utilization of 65%. [5]

$$(.65)\ 150\text{Mb/s} = 97.5\text{Mb/s}$$

The objective of ABR service is to utilize this remaining unused capacity of the switch. Because the CBR and the VBR traffic are sensitive to delay, and because they are also capable of specifying a QoS at connection time, the switch is able to guarantee that it will maintain a cell loss probability of less than 10^{-8}. This is the minimum acceptable cell loss probability according to the specification [4].

ABR traffic, on the other hand, cannot predict its QoS requirements, and as was mentioned earlier, it is not adversely affected small amounts of delay.[6] A good example would be LAN Emulation (LANE). Consider the following scenario:

- A server with a number of clients in L.A. is asked to make an ATM connection to the company server in N.Y.
- This server has no way of predicting the requests of its clients.
- But if a client were to choose to download and run windows from NY, a few milliseconds of delay would not be noticeable to the end user.

Therefore this type of traffic is ideally suited for ABR service.

5 Rate Based Flow Control

It is apparent that because the switches will be over-booked with VBR and CBR traffic that will fluctuate between some minimum and over 100% utilization of the switch resources, there is a

need for a means of controlling the amount of traffic that is to be accepted into the network as ABR traffic. The amount of traffic that can be tolerated by the switch will vary as the VBR connections burst their initial frames and sent updates to it. The control mechanism selected by the ATM Forum was Rate Based Flow Control. This control mechanism is illustrated by the following example. (fig 2)

1. Two clients (C1 and C2) are connected through a switch (S1) to a single client (C3). If C1 and C2 are both transmitting at 100Mb/s (100Mb/s + 100Mb/s = 200Mb/s), and the output link to C3 is limited to 155Mb/s, the switch will have to buffer the traffic which exceeds the capacity of the output link.

2. When this buffer reaches a predetermined threshold it will begin to set a bit in the header of each cell. This bit was named the Explicit Forward Congestion Indicator (EFCI).

3. When C3 receives a cell with this bit set it will send a cell back to C1 and C2 instructing them to reduce their transmission rate by some predetermined value (for this example 1/2 will be used).

4. Thus C1 and C2 begin to transmit at 50Mb/s each. This will enable the switch to transmit it's current traffic as well as empty it's buffer with the unused portion of the bandwidth (BW), 55Mb/s.

155Mb/s - 100Mb/s = 55Mb/s

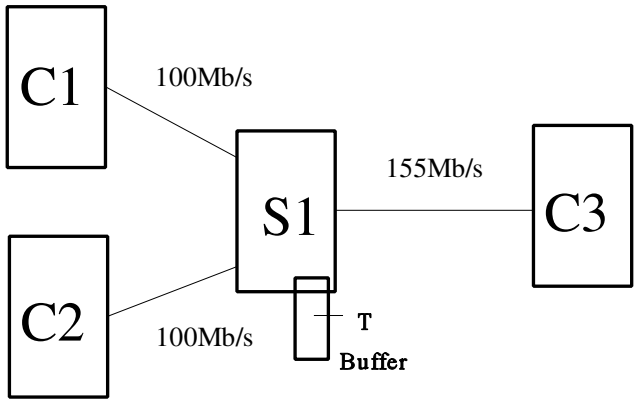

figure 2

In order to maximize utilization, it is preferable to operate continuously in a mildly congested state. Because of this, C1 and C2 will increase their transmission rate until they received a reduce transmission (Tx) rate message from C3. At this time they will reduce their transmission rate by

half (allowing the switch to clear out it's buffer), and then begin to increase the transmission rate again. The flow is demonstrative of an additive increase, and a multiplicative decrease .

In figure 3 the Tx rate is increased by one until the reduce Tx rate cell is received from C3 (time 5) at which time C1 multiplies it's current rate by 0.5 (multiplicative decrease)[2].

$$(0.5)\ 4 = 2$$

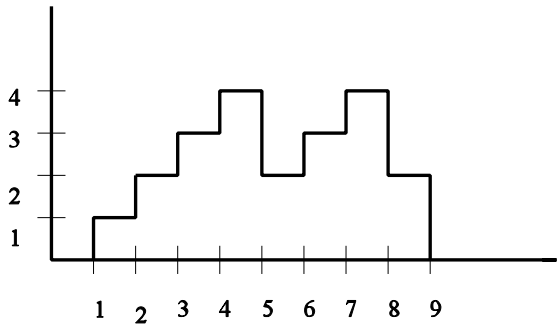

figure 3

This scheme has a number of shortcomings not the least of which is that if all types of traffic were being carried by the switch, and the switch were congested to the point that the reduce Tx rate cell could not get through, C1 and C2 would continue to increase their Tx rate until the switch could no longer maintain a cell loss probability of 10^{-8}. When cells are dropped, many data applications consider the data to be corrupt and will therefore request that the data be retransmitted. These additional request would come at a time when the switch was already congested to the point that it could no longer maintain its negotiated QoS contracts. This represents a network storm, and would bring the network to its knees.

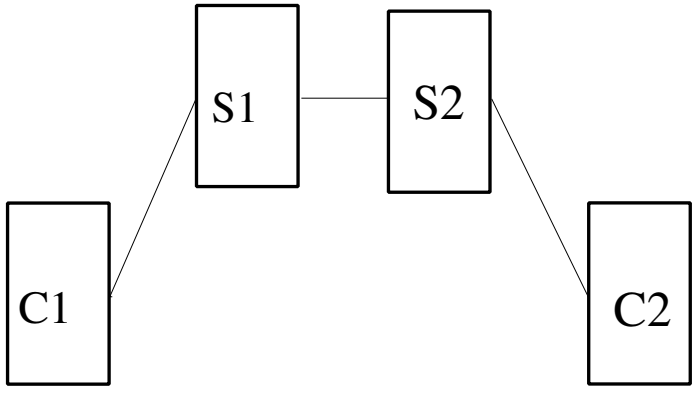

figure 4

Another shortcoming is that the scheme proved to have a fairness problem. Thus, the probability that a call will be asked to reduce its rate (P) increases exponentially with the number of hops (N) the call must make. A hop being the connection between switches. Two switches yields one hop. (fig 4)

$$P^N$$

This was termed the beatdown problem.[7]

6 Evolution of the Rate Based Flow Control Scheme

In a recent article by Bonomi and Fendick [11] the evolution of the RBFC scheme was revealed to the public. The following summarizes information gleaned from the article.

Ramakrishnan - Jain

DECnet protocol to adjust window sizes implemented in ANSI Frame Relay standard.

Newman

1. Proposed RBFC for the support of ABR Service.
2. Backward Congestion Notification
3. Resource Management cell generated only under congested conditions.

Hluchyj and Yin

1. Destination would periodically check to see if the EFCI bit was set. If not, it would send a RM cell back to the source giving permission to increase the rate.
2. If the RM cell is not received by the source it will automatically decrease its transmission rate.
3. Linear increase / exponential decrease.

ATM Forum RBFC Group

Modified Hluchyj and Yin scheme with:

1. Initial Cell Rate
2. Negative Feedback
3. Segmentation of control loop

Explicit Rate Group

1. Network would calculate a fairshare of the available BW and provide the source with an explicit rate.
2. Source will periodically generate a RM cell.
3. Any switch along the path (from source to destination) that does not have the capacity to support the current rate can reduce the sources Tx rate.

Barnhart

Proportional Rate Control Algorithm (PRCA)

1. Positive Feedback
2. Bi-directional Feedback (Forward / Backward)
3. Segmentation of control loop
4. Exponential Ramp-up of Tx rate at a "frequency roughly proportional to the rate of the source".
5. Destination will send a RM cell. The period being determined by the number of data cells sent from the source. (Number of cells / RM).
6. Upon receipt of a RM cell the source will check the EFCI bit. If the EFCI bit is not set the source will increase its rate. Else the source will decrease its rate.
9. Use it or lose it. If the source is not taking advantage of its allowed BW its BW will be reduced to a pre-assigned level.

Fedorkow and A. Jain

Enhanced Proportional Rate Control Algorithm (EPRCA)

Combined PRCA and Explicit Rate Schemes

This allowed for increased flexibility of the switch architecture because the switch can control congestion through EFCI, or an explicit rate, or both.

1. Explicit Rate interpreted as a dynamic upper bound on the rate calculated by PRCA.

7 Proportional Rate Control Algorithm

Significant enhancements were made to the RBFC scheme by Anna Charny as she worked on her thesis at MIT [8]. Before considering Charny's scheme, let us consider the following requirements that were placed upon ABR traffic.

1 ABR traffic should never compromise the cell loss probability of CBR or VBR traffic.

2. The switch should never refuse a guaranteed traffic (CBR, VBR) connection because of the ABR traffic it is supporting.

3. All users of the ABR service should have equal access to the BW that is available.

One way to guarantee equal access is to have the switches compute a fairshare of the BW according to the number of calls that are requesting ABR service. In the Charny's scheme, the clients would send a cell containing it's optimum (desired) Tx rate. An ABR call requesting less than the fairshare of bandwidth would be granted its request. If on the other hand, the caller requested more than the fairshare, the fairshare would be placed in the Resource Management (RM) cell and a bit would be set notifying the source that its rate had been reduced. The RM cell would then be forwarded to the destination. The destination would then return the RM cell to the source. The source could then specify the fairshare as its desired rate. This process is repeated until all calls get their desired (albeit possibly not their initially desired) rate. Charny showed that by having the sources specify an explicit rate, the fairshare could be complied with in a relatively few number of round trips (approximately 4 times the number of bottlenecks). A bottleneck being a switch that could not grant the initial desired rate to the source.

The Charny's scheme used positive feedback as opposed to negative feedback so as to remove the possibility of a network storm. Positive feedback requires that the source reduce it's Tx rate with every cell that is transmitted, and that the source can increase it's rate only when it has transmitted and received an RM cell wherein it's Tx rate is specified according to the fairshare.

This fairshare is computed as follows [7]:

$$Fairshare = \frac{LinkBandwidth - \Sigma BandwidthofUnderloadingVCs}{NumberofVCs - NumberofUnderloadingVCs}$$

The scheme required what has been termed per-VC accounting. Meaning, the switch had to be capable of sampling the rate of all of it's guaranteed traffic connections so as to determine the available BW. (The assumption being that the switch will have to police the rate of each Virtual Connection (VC) in order to verify that it is within the bounds of its QoS, and will therefore already be aware of these rates.) The available BW was then used to compute the fairshare. This

process had to be repeated each time a connection sent an RM cell. The network was also required to recompute the fairshare whenever a caller disconnected from the system, or when a new caller was added to the system. The computational power required at the switch for this scheme is substantial. The scheme was subsequently enhanced.

8 Enhanced Proportional Rate Control Algorithm

EPRCA was first defined as a means of keeping a running exponentially weighted average that was used to compute and assign the Explicit Rate (ER) to each Virtual Circuit (VC). The formula is:

$$MACR = (1-\alpha) * MACR + \alpha * ACR$$

where $\alpha = 1/16$, MACR is the Mean Allowed Cell Rate, and ACR is the Allowed Cell Rate. The ACR is specified by the source in each RM cell, and the MACR is computer by the switch. The formula has the function of weighting the previous MACR much more than the ACR in its computation of the ER. This scheme used a queue threshold to determine when the switch was congested. Subsequent schemes involve using the derivative (differential) of the queue length as a method of detecting an early indication of pending congestion thereby reducing the required size of the queue.

A number of parameters were defined for use in this control algorithm including the following:

Peak Cell Rate	PCR The upper bound of the source Tx rate. The source may exceed the PCR but the cells shall be marked as out of bound, and will therefore be dropped at any switch along the route that is congested.
Minimum Cell Rate	MCR The lower bound for the source Tx rate.
Initial Cell Rate	ICR Because of the use it or lose it aspect of the control system, this field is defined for when a source has been inactive long enough to have lost its rate allocation. The source must therefore send at this initial rate until feedbeck is received from the network.
Additive Increase Rate	AIR The maximum rate at which a source may increase per RM cell.
Rate Decrease Factor	RDF = 2^N Used to decrease the Allowed Cell Rate. (ACR/RDF) The Rate is decreased for each interval of Nrm/ACR cell.

Number of cells/RM	Nrm = 2^N where Nrm - 1 cells can be transmitted between RM cells.
Allowed Cell Rate	ACR Sources shall not transmit in band cells at a rate higher than the ACR.

The basic operation is that the source begins by requesting a PCR and a MCR from the network. Once these parameters have been negotiated the network assigns values for parameters such as AIR, Nrm and RDF. The source begins by transmitting a RM cell followed by its data cells. When the RM cell is returned if Congestion Indication (CI) = 0 and No Increase (NI) = 0 the source must compute a new ACR:

ACR = ACR + Nrm*AIR
ACR = min(ACR, ER, PCR)
ACR = max(ACR, MCR)

If CI = 1 the source shall reduce its rate by at least ACR*Nrm/RDF. A number of other parameters are currently being evaluated.

The ATM Forum spec allows switch manufactuers to choose the method of allocating bandwidth (BW). A number of other methods of implementation are possible. The spec also allows for the network to be divided into virtual source and destinations thereby enabling hop by hop flow control similar to the credit based proposal. Switches may also send the RM cell directly back to the source (BECN). In an ER switch environment this is not desireable because, the total information available in the network is not taken into account (a switch further down may be more congested).

9 Summary

The RBFC schemes allow the vendors a great deal of flexibility in the design and optimization of their switches. A few of the parameters vendors can choose to optimize would be cell transfer delay, buffer size, or profit. The amount of control over the ABR traffic can also be varied. For example in a local environment, end to end flow control may be sufficient. Where as in a WAN it may be necessary to segment the flow control into virtual sources and destinations. Thus exercising a greater amount of control over traffic shaping in the network by providing quicker response.

A great deal of research has been accomplished in the RBFC technique. Unfortunately, most of the work is embargoed by the ATM Forum. In all likelihood, when the traffic committee releases the ATM Forum Traffic Management Specification, much of the work will be published.

A few vendors have already implemented RBFC in their switches, providing dramatic improvement to the transportation of all types of traffic when ABR service is provided by the network. A great deal of research and development in more efficient methods of congestion control may yet be accomplished. Thus, congestion control in ATM networks is likely to be an active area of research for some time.

References

[1] Handel, Huber, Schroder. 1994. "ATM Networks: Concepts, Protocols, Applications", Addison Wesley.

[2] McDysan, D. 1995. "ATM: Theory & Applications", McGraw-Hill.

[3] Partridge, C. 1993. "Gigabit Networking", Addison Wesley.

[4] ATM Forum. 1993. "ATM User Networks Interface Specification 3.0", Prentice Hall.

[5] Fischer, W. Aug 1994 "Traffic Control for Statistical Multiplexing of Data Traffic", IEEE Communications.

[6] Newman, P. Aug 1994 "Traffic Management for ATM Local Area Networks", IEEE Communications.

[7] Jain, R. 1995. "Congestion Control and Traffic management in ATM Networks: Recent Advances and A Survey", Computer Networks and ISDN Systems.

[8] Charny, A. 1994. "An Algorithm for Rate Allocation in a Packet-Switching Network with Feedback", MIT.

[9] Sui and Tzeng, 1995. "Intelligent Congestion Control for ABR Service in ATM Networks", U of C, Irvine.

[10] Ebrahim, Z. 1992. "A Brief Tutorial on ATM" Cell Relay Retreat http://www-ipg.umds.ac.uk/~dlgh/teaching/atm_tutorial.html

[11] Bonomi and Fendick, 1995. "The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service", IEEE Network (March/April)