Ascend Flow Control Processors for Non-Real Time Services White Paper

Introduction

ATM networks support multiple categories of applications, such as circuit emulation, entertainment-quality video, and a variety of packet data applications including Internet. The first two groups of applications are well defined. They have been efficiently supported by Constant Bit Rate (CBR) and Real-Time Variable Bit Rate (rt-VBR) service categories in asynchronous transfer mode (ATM) networks built of CBX 500 ATM switches from Ascend Communications, Inc. The challenge for ATM network providers now is to offer improved and cost-effective support for packet data applications for which traffic patterns and quality of service (QoS) requirements are not as well defined. In general, such applications can tolerate delay relatively well. However, they are less tolerant of cell loss, especially when cell loss occurs indiscriminately under congestion. Network users require non-real time services, such as the ATM Forum-defined Non-real Time Variable Bit Rate (nrt-VBR), Available Bit Rate (ABR), or Unspecified Bit Rate (UBR) service categories, to support these applications.

From the network provider perspective, the bursty nature and relaxed delay requirements of packet data services like Internet or Intranet lead to the economically justified need for oversubscribed networks. (See Figure 1.) As a result, proactive congestion avoidance measures cannot be fully relied on and reactive methods, specifically traffic flow control, are called for to eliminate congestion and ensure high goodput for end-user applications.

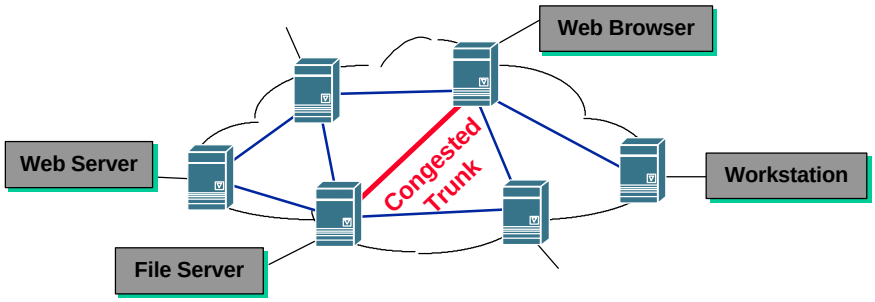


Figure 1.
Packet Data Services Lead to Oversubscription.

Substantial research and debate on the best support of packet data services in ATM networks resulted in the publication of the ATM Forum Traffic Management TM 4.0 standard in 1996. The standard defines adaptive rate-based flow control mechanism designed to support packet data applications over ATM networks. The standard also defines packet-level discard methods devised to enhance performance of higher-level protocols, such as TCP/IP over ATM networks.

Ascend™ has been in the forefront of implementing the TM 4.0 standard and now offers a family of flow control processor modules for the CBX 500 switch.

The algorithms implemented in the Ascend Flow Control (FC) Processors are fully compliant with the TM 4.0 standard and work in concert with CBX 500 queuing architecture and call processing and routing software. Together, they ensure effective traffic management in practical ATM networks supporting cost-efficient quality services for present-day and near-future applications. Specifically, the FC processors support such services as nrt-VBR, ABR, and UBR. The processors also allow network providers to offer an enhanced UBR+ service.

Background

Effective traffic management in ATM networks requires three network-wide mechanisms: proactive congestion avoidance through traffic isolation, QoS-aware routing, and reactive congestion removal through traffic flow control. (See Figure 2.) Traffic isolation is ensured in CBX 500 by its Quad-Plane queuing architecture. Ascend's Virtual Network Navigator (VNN) and Connection Admission Control (CAC) software supports QoS-aware routing. With the FC processors, Ascend now provides the complete set of building blocks required to create the ATM networks needed today to efficiently support customers' real-time and non-real time applications.

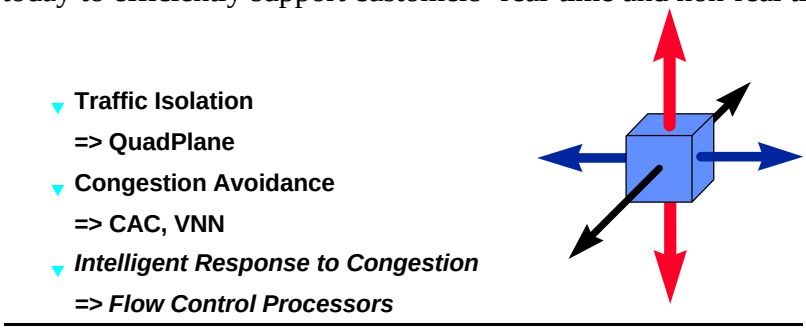


Figure 2.
Effective Traffic Management Elements.

Traffic flow control has been used effectively for many years in data networks. The principal higher-layer protocol supporting packet data applications is the TCP/IP protocol. TCP employs an end-to-end flow control algorithm and adapts the transmission of packets by the source to the congestion conditions along the source-to-destination route. Only endstations participate in TCP flow control, and congestion is detected by observed packet loss.

Even though flow control has been proven effective in supporting packet data applications in pre-ATM networks, first generation ATM switches did not have flow control capability. Soon, it

became evident that flow control was necessary to ensure good resource utilization and quality service for data applications. With the FC processors, the CBX 500 ATM switch now offers flow control at the ATM layer.

Ascend Flow Control

The ATM Forum TM 4.0 standard calls for interaction between the end devices and the network and defines several optional flow control algorithms. The Flow Control Processors implement the binary feedback method, a cost-effective and practical option in the standard and make it possible for a network provider to offer per-virtual circuit (per-VC) flow control today. At the edge of ATM networks, the FC processors allow transparent virtual source/virtual destination (VS/VD) interface to ATM devices capable of explicit rate (ER) flow control. Efficient and fair support for packet data services requires three mechanisms: buffering, flow control, and packet level discards in case of unavoidable congestion. The Ascend Flow Control Processors support all three.

Ascend implemented the FC processors in micro-coded hardware engines built as daughter cards for the CBX 500 ATM line cards. (See Figure 3.) The FC processors work at the full range of T1/E1, DS3/E3, OC-3c/STM-1, and OC-12c/STM-4 rates. The FC processors allow ATM network providers to offer superior packet data services.

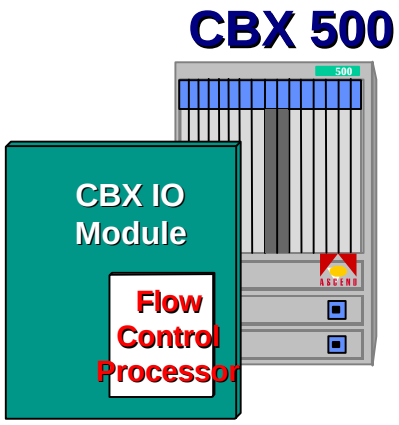


Figure 3.
Flow Control Processor Card for the CBX 500.

The Flow Control Processors are fully compliant with the TM 4.0 standard and implement a practical subset of options defined in the standard. The FC processors ensure effective traffic management through hop-by-hop, rate-based VS/VD flow control with binary feedback and per-VC queuing. At the edge of the network, the FC processors enable a transparent VS/VD interface

to ATM end devices capable of ER flow control. The FC processors also provide Early Packet Discard (EPD), enhancing goodput of higher level protocols like TCP/IP over ATM networks.

Example Network Application

The Ascend Flow Control Processors can be effectively utilized in ATM networks assuring quality services for packet data applications. The Internet/Intranet Service Provider (ISP) ATM network shown in Figure 4 is a good example. An ISP network typically has a large number of users connected to the network through service access multiplexers or remote access concentrators. Access multiplexers and concentrators are, in turn, connected to the ATM network trough T1 or DS3 access links. To maintain low costs, these access links are fully subscribed or even somewhat oversubscribed. At the edge of the ATM network CBX 500 ATM switches aggregate and concentrate access traffic, resulting in a fivefold bandwidth oversubscription on the OC-3c links connecting the edge switches to the backbone network. Traffic entering the network from the edge switches is further concentrated in the backbone network. As a result, OC-12c links between the CBX 500 switches on the backbone are oversubscribed up to twenty times. Such traffic concentration and bandwidth oversubscription is economically justified as it allows the network operator to offer low-cost services to Internet users who can tolerate occasional delay in obtaining images or stock quotes but demand low service cost. However, flow control is needed to assure quality services for packet data applications at high bandwidth oversubscription rates that are common in ISP networks.

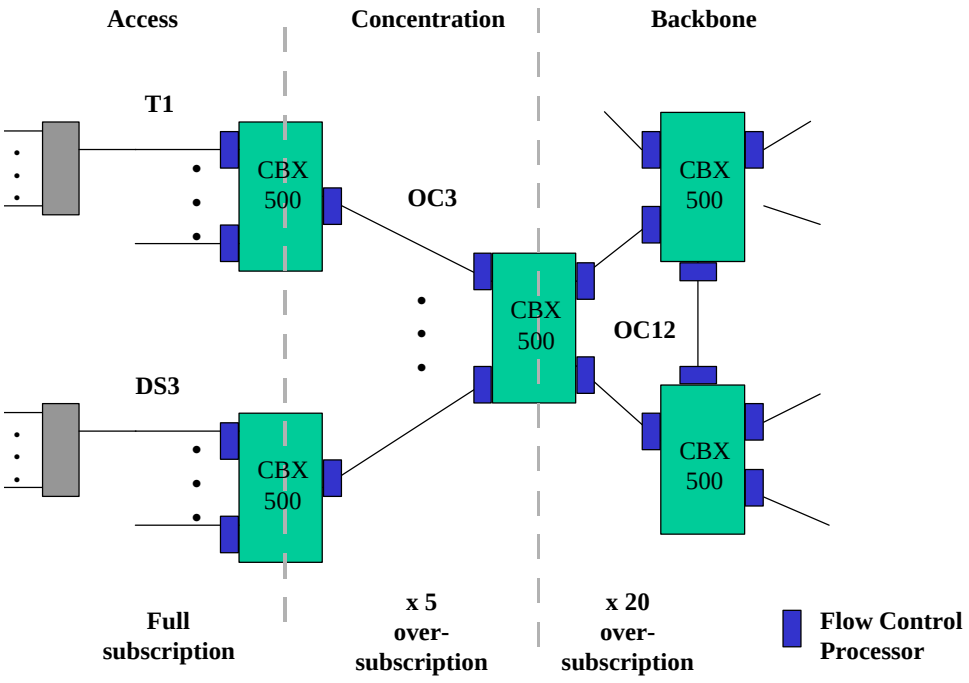


Figure 4.
A Typical ISP Network.

FC processors - Principles of Operation

Ascend Flow Control Processors implement rate-based, hop-by-hop flow control. The FC processors do not require endstation participation in the flow control process. However, the FC processors provide a transparent VS/VD interface to ATM devices capable of ER flow control. The FC processors assure service quality and fairness for all user connections. Hop-by-hop rate control effectively pushes congestion out to the network edge. EPD, an important feature offered by Ascend Flow Control Processors, enhances performance of higher layer protocols like TCP/IP over ATM networks built of CBX switches. The rate control capabilities of the FC processors also allow egress traffic shaping. In short, Ascend Flow Control Processors provide a wide range of features, described in detail below, enabling the ATM network provider to offer cost effective services to packet data users.

Service Quality

The FC processors allow each connection to request a guaranteed minimum cell rate (MCR). The connection also defines its peak cell rate (PCR), the maximum rate at which it will transmit cells. The effect of flow control is that, depending on congestion conditions, the cell rate of the connection varies between its MCR and its PCR. The connection can always transmit at its MCR

rate, even when heavy congestion conditions exist. In mild congestion, the connection transmits at the allowable cell rate (ACR), adjusted to avoid cell losses and to ensure fairness concerning other coexisting connections. In the absence of congestion, the connection can transmit at its PCR rate. In short, the FC processors ensure non-real time connections guaranteed bandwidth and buffering in heavy congestion, fair share of network resources in mild congestion, and high throughput when congestion is removed from the network. (See Figure 5.)

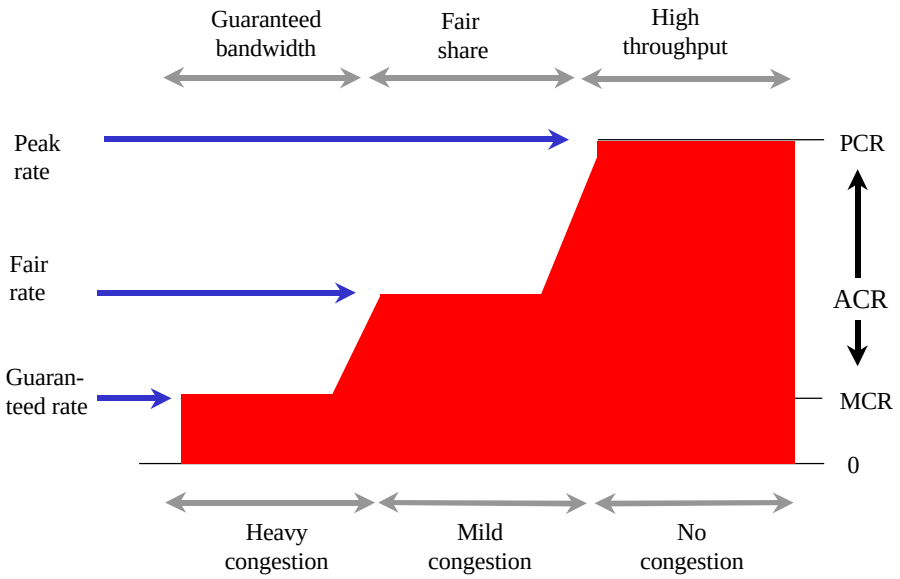


Figure 5.
Flow Control Assures Service Quality.

Fairness

Fair access to network resources is of prime interest to network users. The FC processors ensure fair distribution of buffering capacity and transmission bandwidth between non-real time connections. At the same time, no resources are wasted by setting them aside. Bandwidth and buffering fairness are relevant only under congestion conditions. In the absence of congestion, the FC processors allow each connection to use as much bandwidth and buffering space as it needs. The amount of buffering space and the bandwidth available to a connection in congestion are proportional to the connection's MCR. (See Figure 6.)

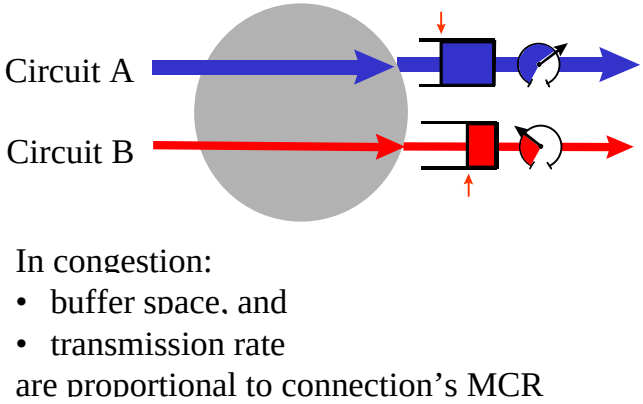


Figure 6.
Rate and Buffer Fairness.

The following example illustrates the fairness ensured by the TM 4.0 standard. Assume that Circuit A arrives from a DS3 source, Circuit B arrives from a T1 source, and both are routed together over an OC-12c trunk. Also assume that Circuit A has a contract for 10 Mbps MCR, and Circuit B has a contract for 1 Mbps MCR. Without congestion in the trunk, both circuits can transmit at their line rates, DS3 and T1, respectively. However, should the trunk become congested, their rates will be limited but will never be below the 10 Mbps and 1 Mbps MCR rates, respectively. In this case of mild congestion, the rates of both circuits will be changing, adapting to the temporary congestion conditions. This ensures that both circuits can transmit at rates proportional to their contracted MCR rates (for example, circuit A at 20 Mbps and circuit B at 2 Mbps).

Congestion Pushed Out of the Network

The principle of rate-based flow control calls for adaptive adjustment of the rate at which an upstream node transmit cells, depending on the congestion state of the downstream node. If the downstream node experiences congestion, the upstream node should reduce its transmission rate. When the congestion in the downstream node abates, the upstream node can again increase its transmission rate. The network-wide effect is to push congestion out to the network edge. (See Figure 7.) The mechanism efficiently reduces cell loss in network nodes for bursty packet data connections.

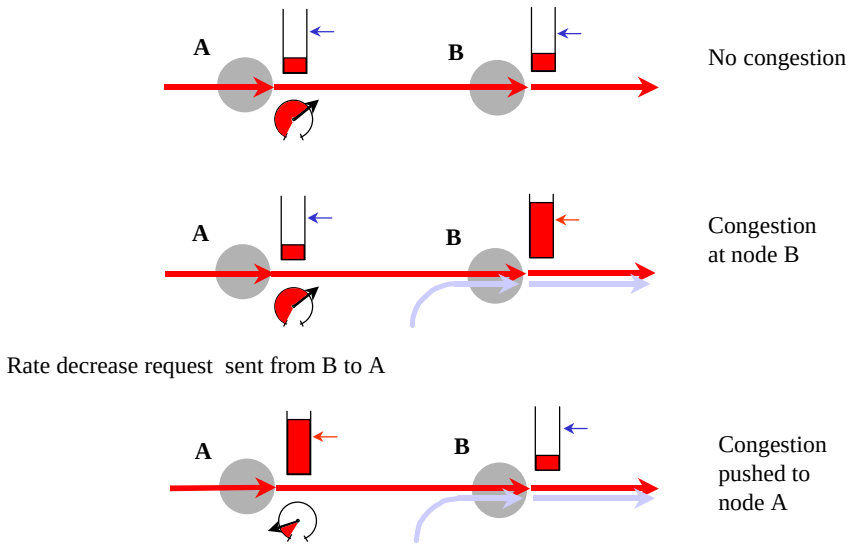


Figure 7. Flow Control Pushes Congestion Out to Network Edge.

End-to-end vs. Hop-by-hop Flow Control

The TM 4.0 standard defines two options for flow control. The rate control principle is the basis of both. The end-to-end option involves end-user devices and calls for exchange of Resource Management (RM) cells between the source and destination devices. Information about possible congestion along the connection path is inserted to the RM cells as they cross the switches. Information contained in the RM cells arriving from the downstream direction is used to adjust the transmission rates at the output ports of the upstream switches. (See Figure 8.)

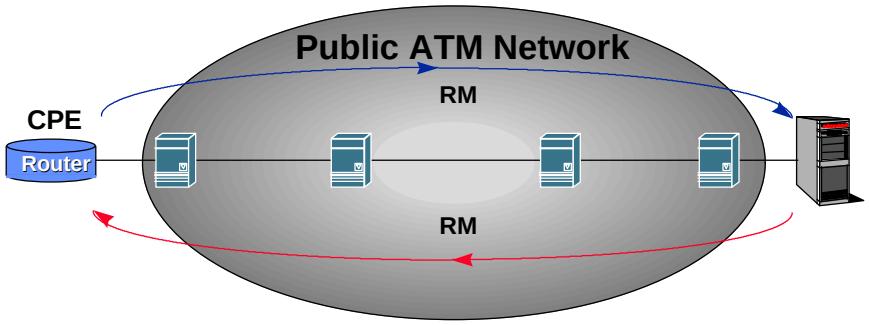


Figure 8. End-to-end Flow Control.

The hop-by-hop option does not require that the end devices participate in the flow control process. Therefore, network providers can offer it today. In the hop-by-hop implementation of the TM 4.0 standard, end-to-end connection paths are partitioned to form a chain of one-hop segments. The upstream node in a segment acts as the VS and the downstream node in the segment acts as the VD, hence the term VS/VD flow control. A node in the middle of a path acts as both a virtual source, for the downstream segment, and a virtual destination, for the upstream segment. Virtual sources adjust their rates based on the feedback they receive directly from the corresponding virtual destinations. (See Figure 9.)

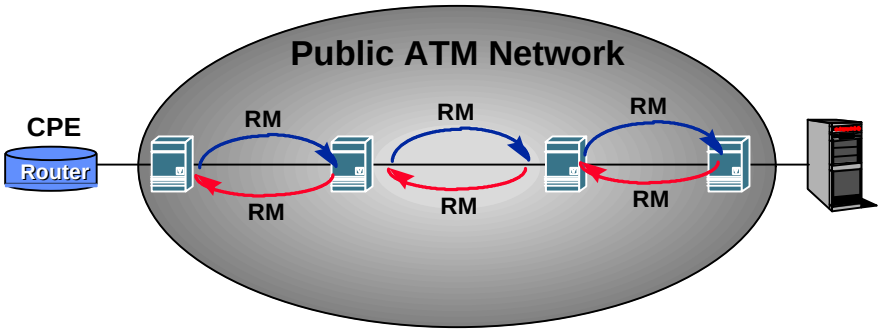


Figure 9.
Hop by Hop Flow Control.

Is a Flow Control Processor Needed in Every Switch?

Both ends of each flow control segment must be capable of participating in the flow control mechanisms supported by the FC processors. However, it is possible to establish a flow control segment across one or more ATM switches that are not capable of flow control. (See Figure 10.) In this case, the flow control loop is transparent to the non-participating switches.

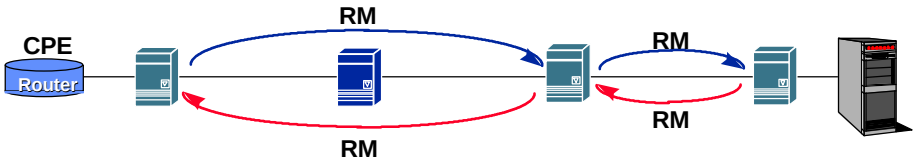


Figure 10.
Establishing Flow Control Across ATM Switches Not Capable of Flow Control.

Is Flow Control Needed at Network Edge?

The Ascend Flow Control Processors allow transparent interface**,** at the network edge**,** to ATM devices capable of ER flow control. However, flow control-capable end-devices are not required to allow implementation of flow control across the network. FC processors can maintain effective flow control along the VC, from the network edge to the network edge, through a sequence of hop-by-hop VS/VD segments. This way, congestion is pushed out of the network even without end-device participation. When ER flow control-enable devices are connected to a CBX 500 switch at the network edge, the FC processor at the edge maps the explicit rate RM cells used between the end-device and the switch to the binary feedback RM cells used within the network. In effect, end-to-end flow control is ensured. (See Figure 11.)

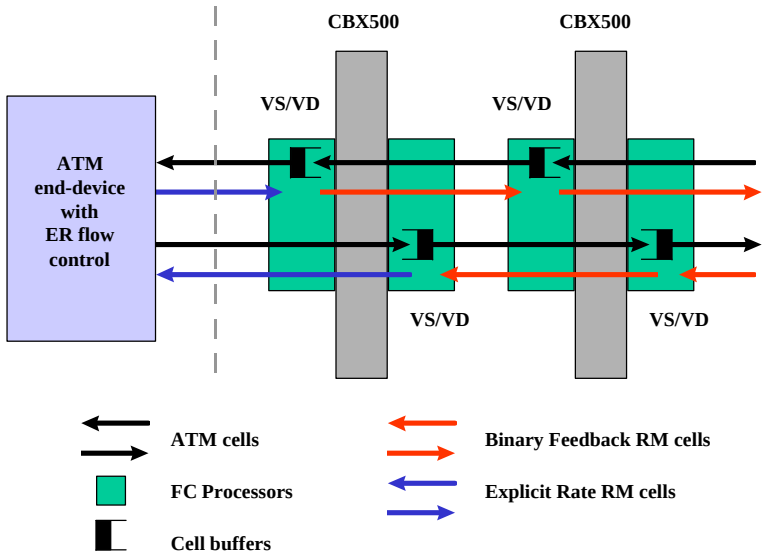


Figure 11.
Ensuring End-to-end Flow Control Through Hop-by-hop VS/VD Segments.

Early Packet Discard

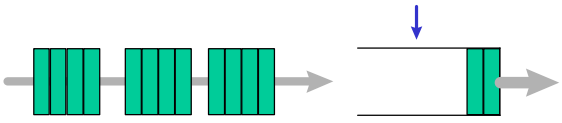
The main difficulty in supporting higher level protocols (like TCP/IP) traffic over ATM networks lies in the problem of indiscriminate cell loss in congestion. Even if a single cell from a packet is lost, the higher layer protocol declares the entire multi-cell packet lost. When two cells are lost from the same packet, the protocol declares one packet lost. But, if the two discarded cells belong to two different packets, the protocol declares two packets lost. The problem is compounded because the cells from different connections can be intermixed in the cell stream and, when buffer

overflows, the number of connections that may need to retransmit packets may be equal to the number of cells lost.

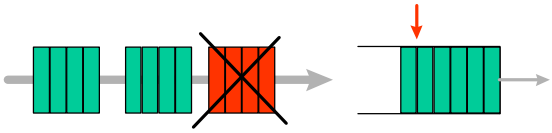
Higher level data transport protocols, like TCP/IP deal well with packet loss. Data networks, to be economically justifiable, must allow bandwidth oversubscription and, therefore, data losses are unavoidable. TCP implements end-to-end flow control and a TCP source can adjust its packet transmissions in case of congestion. To ensure good support of TCP connections by ATM networks, a method is needed to avoid indiscriminate cell loss to eliminate the scatter of cell loss over multiple packets and connections.

A good measure assuring this objective is the EPD technique, standardized by ATM Forum in the TM 4.0 standard. At the onset of congestion, all cells of the current packet are admitted to the queue, but all cells of the next packet are discarded. The discard is called the “early discard” because it happens before actual buffer overflow occurs. The reason that the cells of the current packet are not discarded is that the congestion may be detected when some cells of this packet have already been admitted to the queue. If the remaining cells were discarded, the partial packet would be of no use anyway, and bandwidth would be wasted.

The EPD method has the beneficial effect of preventing possible loss of cells from multiple packets and/or multiple connections. In effect, cell loss in congestion happens in a controlled manner, improving the throughput of higher layer protocols. The FC processors implement EPD on a per-VC basis. This implementation has the additional benefit of focusing cell losses in connections that are trying to use more than their fair bandwidth share. (See Figure 12.)



No congestion, all cells passed through



After the onset of congestion, all cells of the next packet are discar

Early Packet Discard
- **eliminates scattering of cell loss over multiple connections**
- **enhances performance of higher layer protocols like TCP/IP**

Figure 12.
Early Packet Discard.

Egress Traffic Shaping

Rate control capabilities of the FC processors can be very useful whenever a need exists to shape the ATM egress traffic at the network edge. One example is a simple device, an xDSL multiplexer not equipped with sophisticated buffering capabilities, employed between an ATM edge switch and xDSL end-users. In this case, many VCs share a high-speed link between the ATM switch and the multiplexer. Without traffic shaping, the egress peak rate of a VC can reach the link rate and exceed the rate of the multiplexer tributary over which the VC is demultiplexed to its destination. As a result, without traffic shaping, cell losses are unavoidable.

The FC processors can readily solve the problem. The FC processors ensure per-VC queuing and per-VC cell transmission scheduling. The ACR parameter limits the egress cell rate for each VC. This parameter can be provisioned to be a constant and different VCs can have different ACR constants. In effect, the egress traffic of each VC can be shaped by limiting its peak cell rate as needed and cell losses at the xDSL multiplexer can be avoided. The FC processors allow traffic shaping on a per-VC or per-virtual path (per-VP) basis. (See Figure 13.)

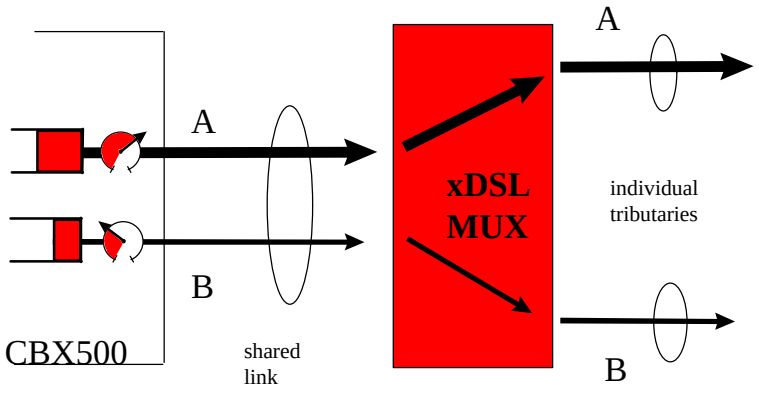


Figure 13.
Egress Traffic Shaping - Provisionable ACR for each VC.

Congestion Detection and RM Cells

A congestion detection mechanism and RM cells support flow control. The FC processors detect congestion based on buffer occupancy thresholds. The RM cells exchanged between the switches

carry the information about desired rate changes. The FC processors implement congestion detection and RM cells on a per VC basis, in compliance with the TM 4.0 standard.

Buffer Thresholds

The FC processors implement fair and efficient buffer allocation across VCs. They provide multiple congestion thresholds, both per-VC thresholds and global thresholds. A VC is declared to be in congestion when both its own “Congestion” threshold and the “Global Congestion” threshold have been exceeded. This way, the buffering space is available to VCs without restrictions in the absence of congestion. However, when congestion occurs, the VCs that have used more than their fair share of the buffering space will be declared congested. The mechanism is fair and allows good utilization of network resources under varying congestion conditions. (See Figure 14.)

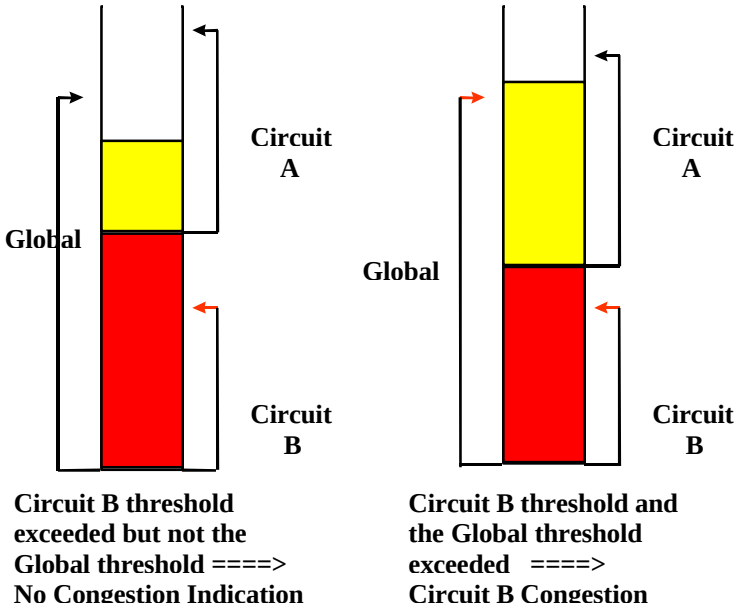


Figure 14.
Efficient Resource Use Through Buffer Sharing.

Discards are effected when buffer content exceeds discard thresholds. Again, the FC processors provide both per-VC and global discard thresholds. Each VC may be set up in one of two modes of operation: the CLP1 discard mode or the EPD discard mode. For a VC in the CLP1 discard mode, the FC processors will discard the incoming CLP1 cells whenever both the Discard and Global Discard thresholds are exceeded. For a VC in the EPD mode, the FC processors will

discard all cells belonging to the next packet arriving after both the Discard and Global Discard thresholds are exceeded. Finally, the FC processors discard all incoming cells unconditionally when the Global CLP0+1 threshold is crossed. This threshold corresponds to the total buffer capacity assigned to a port. (See Figure 15.)

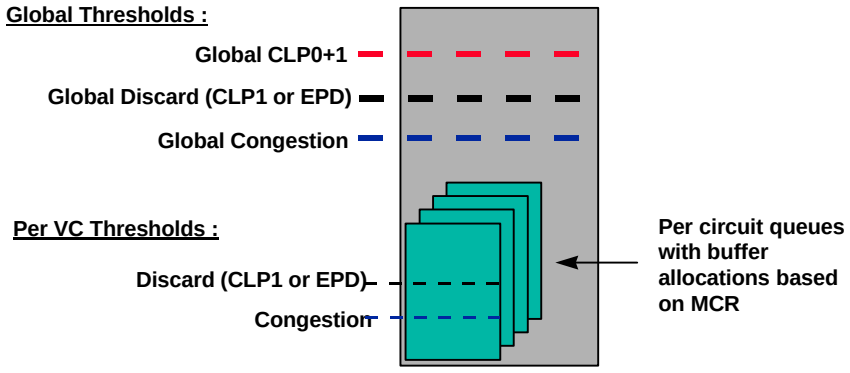


Figure 15.
Buffer Allocation and Thresholds.

The FC processors have large, 128K cell buffers. This total buffering capacity can be statically allocated between ports served by an FC processor. The amount of buffering capacity allocated to a port can range from 1k to 128k cells.

Binary Feedback vs. Explicit Rate

The TM 4.0 standard provides two optional feedback methods, the binary feedback method and the explicit rate method. In the binary feedback method, the downstream node can send two requests to the upstream node: the “decrease rate” request or the “increase rate” request. Additionally, when periodic RM cells are implemented, a “do nothing” request is also possible. In the explicit rate method, the downstream node should compute the exact rate that it can support without cell loss and send this information, encoded in an RM cell, to the upstream node. Of the two optional feedback methods, the more complicated and more expensive explicit rate calls for participation of endstations in the end-to-end flow control. Few endstations will have this capability in the near future as the desktop ATM paradigm is not expected to become prevalent for many years. The FC processors implement the more practical, cost-effective, binary feedback method.

RM Cells

The FC processors implement two types of RM cells—Cascade Resource Management (CCRM) cells and Backward Congestion Message (BCM) cells. The CCRM cells are a subset of TM 4.0 standard RM cell definition. (See Table 1.)

Table 1.
Virtual Destination/Virtual Source Activity.

Destination behavior	Source behavior

▼ Objective: notify upstream switch of congestion
▼ Generate RM cells, considering:
  ▼ local queue congestion state
  ▼ downstream switch RM indicators
▼ RM cells indicate rate increase, decrease, or no change
▼ RM cells generated at regular intervals
  ▼ Configurable
  ▼ Default = 100 msec
  ▼ Allowable range: 30 msec +
	▼ Objective: adjust the VC cell rate (ACR)
	▼ Adjust rates based on:
	  ▼ backward RM cells received
	  ▼ bandwidth utilization
	▼ Monitor VCs for “idle state”
	  ▼ Defined as 1-8 cell arrival intervals (configurable)
	  ▼ If idle, reset ACR to ICR, stop genrating RM cells
	▼ Terminate backward RM cell associated with its control loop
	▼ Update and forward RM cells not associated with its control loop

CCRM Cells

The CCRM cells are generated periodically in the backwards direction and provide positive information as to the virtual source action desired by the virtual destination. The CCRM cell generation period can range from 30 to 100 milliseconds. (The default is 100 milliseconds.) The RM cell generation period is provisionable and the number of VCs supported relates to the selected RM cell generation period. (See Table 2.) The CCRM cells for all active VCs are generated within the provisioned RM cell interval.

Table 2.
RM Interval and Number of VC Supported.

RM Interval	# of VC supported
100 ms	12K
50 ms	6K
30 ms	4K

BCM Cells

The FC processors can optionally generate and process BCM cells to ensure interoperability with switches that do not implement CCRM cells. In this case, BCM cells are generated by a virtual destination only in case of congestion. The virtual source will reduce its rate when it receives a BCM cell. If a BCM cell is not received by the source within a prescribed period of time, the

source will increase its rate. The BCM cells are not subject to generation period restrictions as are the CCRM cells.

Rate Adjustment Algorithm

Rate reduction is necessary, as a final resort, when congestion occurs despite application of congestion avoidance measures, such as traffic separation and CAC. Congestion occurrences are as unavoidable as oversubscription is desirable. When congestion abates, rates can be increased again. Fairness is of paramount importance in rate adjustments. The Ascend flow control rate adjustment algorithm (described below) ensures fair allocation of bandwidth between virtual circuits (VCs).

For each VC, the FC processor in the virtual source switch will adjust the VC's ACR in response to CCRM or BCM cells received from the virtual destination switch. The source maintains the ACR according to the algorithm described below.

1. The FC processor changes the ACR under the following conditions:

    - CCRM cell is received
    - BCM cell is received
    - BCM cell is not received within BCM timeout interval.
    - VC goes idle

2. The FC processor decreases the ACR when a CCRM cell carrying a request to reduce VC rate is received. Alternatively, the FC processor decreases the ACR when a BCM cell indicating downstream congestion is received.
3. The FC processor increases the ACR when a CCRM cell carrying a permission to increase VC rate is received. Alternatively, the FC processor increases the ACR on BCM timeout.
4. The ACR is decreased by the Rate Decrease Factor (RDF) and increased by the Rate Increase Factor (RIF). The RDF and RIF values are provisionable. The ACR value is never reduced below MCR and never increased above the PCR. New values of ACR are computed as follows:

    In case of rate decrease: *New ACR = max [current ACR *(1 - RDF), MCR]*

    In case of rate increase: *New ACR = min [current ACR + RIF * PCR, PCR]*

5. Initial value of ACR is set to be equal to Initial Cell Rate (ICR). The ICR is proportional to the VC's MCR (SCR for nrt-VBR VCs) through a provisionable constant.
6. When a VC is declared idle, the ACR is reduced to the VC's MCR, or SCR, rate. Thus, when VC activity resumes, the virtual source will start transmitting at the MCR rate.

Summary

Addition of the Ascend Flow Control Processors to the Ascend CBX 500 ATM product line allows network providers to build ATM networks, efficiently supporting full range of ATM services today. The synergy between Quad-Plane queuing architecture, VNN and CAC software, and the FC processors ensures ATM network users of quality services, optimized for their real-time and non-real time applications.

The Ascend Flow Control Processors, fully compliant with the ATM Forum TM 4.0 standard, ensure effective traffic management through a rate-based, hop-by-hop VS/VD flow control with binary feedback and per-VC queuing. At the edge of the network, the FC processors enable transparent interface to ATM end-devices capable of ER flow control. They also provide EPD, enhancing goodput of higher level protocols like TCP/IP over ATM networks.

With the FC processors, ATM network providers can now build networks optimized to support non-real time services, such as nrt-VBR, ABR, UBR, and UBR+. The demand for such services is clear, evidenced by the rapid growth of Internet and Intranet applications.

Glossary

ABR - Available Bit Rate (service category)

ACR - Allowed Cell Rate

ATM - Asynchronous Transfer Mode

BCM - Backward Congestion Message (cell)

CAC - Connection Admission Control

CBR - Constant Bit Rate (service category)

CCRM - Cascade Resource Management (cell)

CLP - Cell Loss Priority

EFCI - Explicit Forward Congestion Indication

EPD - Early Packet Discard

ER - Explicit Rate

FC - Flow Control

ICR - Initial Cell Rate

ISP - Internet/Intranet Service Provider

MCR - Minimum Cell Rate

NRTS - Non-Real Time Services

nrt-VBR - Non-real Time Variable Bit Rate

PCR - Peak Cell Rate

QoS - Quality of Service

RDF - Rate Decrease Factor

RIF - Rate Increase Factor

RM - Resource Management

rt-VBR - Real-Time Variable Bit Rate (service category)

SCR- Sustainable Cell Rate

TCP/IP - Transmission Control Protocol/Internet Protocol

TM 4.0 - ATM Forum Traffic Management TM 4.0 standard

UBR - Unspecified Bit Rate (service category)

UBR+ - Enhanced Unspecified Bit Rate

VC - Virtual Circuit

VD - Virtual Destination

VNN - Virtual Network Navigator

VP - Virtual Path

VS - Virtual Source

VS/VD - Virtual Source/Virtual Destination (a hop-by-hop flow control using virtual sources and destinations)

xDSL - one of DSL, Digital Subscriber Loop, technologies: ADSL, HDSL, VDSL