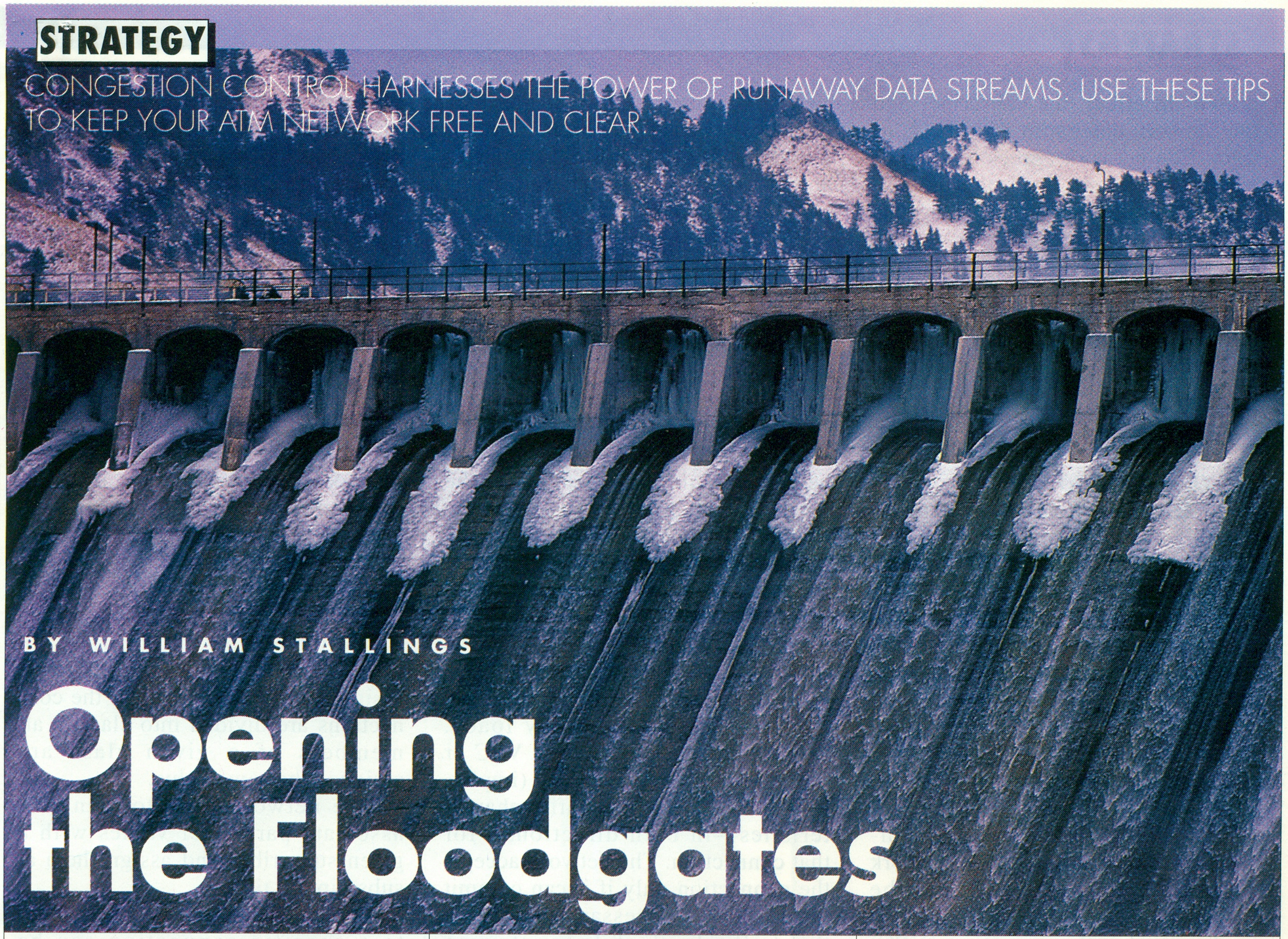

CONGESTION CONTROL HARNESSES THE POWER OF RUNAWAY DATA STREAMS. USE THESE TIPS TO KEEP YOUR ATM NETWORK FREE AND CLEAR.

BY WILLIAM STALLINGS

Opening the Floodgates

Long ago mankind grew weary of existing at the whim of the elements. Among the most nagging of nature's creations was water. It dropped from the sky at unpredictable moments, showered the planet's inhabitants indiscriminately, and sometimes turned into frigid blizzards.

While nothing much could be done about rain, lakes and rivers were a different story. Through an innovative engineering triumph, the dam was born—and with it one of the most basic forms of congestion control on Earth.

A different form of congestion control is equally crucial to the ATM-based network's success. Without it, traffic from user nodes may exceed network capacity, causing the ATM switches' memory buffers to overflow. This, in turn, leads to data losses.

ATM networks present congestion control difficulties not found in other types of networks, including frame relay systems. ATM's high data rates and switching speeds mean that hefty chunks of information can easily be lost. And for data traffic, the loss of even one cell could require retransmission of thousands. This problem is compounded by the limited number of overhead bits available for exerting control over the flow of user cells.

This is currently the subject of intense research, but no consensus has emerged for a full-blown traffic and congestion control strategy. The International Telecommunications Union-Telecommunications Standards Section (ITU-TSS) has defined a restricted, initial set of traffic and congestion control capabilities targeted toward simple mechanisms and realistic network efficiency levels, specified in I.371. The ATM Forum has published a more advanced version in the ATM user-network interface specification 3.0.

The sheer quantity of congestion control techniques, some of which can be used alone or in combination with others, makes this function the most perplexing aspect of ATM technology. In this article, we'll examine the congestion control problem and techniques used to solve it.

TIMING IS EVERYTHING

ITU-TSS and the ATM Forum have defined a set of traffic and congestion control functions that operate across a spectrum of timing intervals.

Four timing levels are addressed:

- *Cell insertion time*—Functions at this level react immediately to cells as they are transmitted.
- *Round-trip propagation time*—At this level, the network responds within the lifetime of a cell in the network and may provide feedback indications to the source.
- *Connection duration*—At this level, the network determines whether a new connection at a given quality of service (QOS) can be accommodated and what performance levels will be negotiated.
- *Long term*—These are controls that affect more than one ATM connection and are established for long-term use.

The essence of traffic control involves determination of whether a given, new ATM connection can be accommodated and negotiation with the subscriber concerning the performance parameters that will be supported. In effect, the subscriber and the network enter into a traffic contract: The network agrees to support traffic at a certain level on this connection, and the subscriber agrees not to exceed this level.

Traffic control functions are intended to establish and enforce these traffic parameters, a task that includes congestion avoidance. If traffic control fails, congestion control functions are invoked to rectify the situation.

STAYING AFLOAT

ATM traffic control refers to the actions the network performs to avoid congestion conditions or to minimize congestion effects. These functions include network resource management, connection admission control, usage parameter control, priority control, and traffic shaping.

In network resource management, resources are allocated so that traffic flows are separated according to service characteristics. So far, the use of virtual paths is the only specific traffic control function based on network resource management to be addressed.

A virtual path connection (VPC) provides a convenient means of grouping similar virtual channel connections (VCC). A VCC is similar to an X.25 virtual circuit; it provides a logical connection between two end users or between an end user and a network service. The network provides aggregate capacity and performance characteristics on the virtual path that the virtual connections share.

Each VCC has an associated set of QOS parameters. The parameters that are of primary concern in network resource management are cell loss ratio, cell transfer delay, and cell delay variation—all of which are affected by the amount of resources the network devotes to the VPC.

For a VCC extending through multiple virtual path connections, performance depends on that of the consecutive VPCs and on the handling of the connection by any node that performs VCC-related functions. This node may be a switch, a concentrator, or another piece of network equipment.

Each VPC's performance depends on the capacity of that connection and the traffic characteristics of the VCCs contained within it. The performance of each VCC-related function depends on the node's switching and processing speed and on the priority in which cells are handled.

VCCs vary in performance, and there are many ways to group these connections. If all virtual channel connections in a VPC are handled the same way, they should deliver similar performance in terms of cell loss ratio, cell transfer delay, and cell delay variation. When different VCCs in the same virtual path connection require different QOSs, the VPC performance objective that the network and subscriber negotiate should be set according to the most demanding VCC required.

EBB AND FLOW

Connection admission control is the network's first line of defense against excessively heavy loads. When a user requests a new VPC or VCC, he or she must specify (implicitly or explicitly) the traffic characteristics—in both directions—for that connection. The network accepts the connection only if it can commit the resources necessary to support that traffic level while maintaining the specified QOS of existing connections. By accepting the connection, the network forms a "traffic contract" with the user and provides the specified QOS as long as the user complies with the contract.

The traffic contract addresses peak cell rate (PCR), cell delay variation (CDV), sustainable cell rate (SCR), and burst tolerance. Let's examine each of these in detail.

Peak cell rate is the maximum rate at which cells are generated by the source on the connection. To accurately determine this rate, we must take cell delay variation into account. Although a source may generate cells at a constant peak rate, cell delay variations affect timing, causing cells to "clump up" and gaps to form. Thus, the source may temporarily exceed the peak cell rate.

To properly allocate resources to this connection, the network must know the peak cell rate and the CDV.

The user must specify the PCR and CDV for every connection. As an option for variable bit-rate sources, the user can also specify a sustainable cell rate and burst tolerance. These parameters are analogous to PCR and CDV (respectively) but apply to an average rate of cell generation rather than a peak rate.

The user can describe future cell flow in more detail using the SCR and burst tolerance as well as the PCR and CDV. This may allow the network to more efficiently utilize its resources. For example, when a number of VCCs are statistically multiplexed over a virtual path connection, knowledge of average and peak cell rates lets the network allocate buffers large enough to handle the traffic efficiently without cell loss.

The traffic parameters for any given connection (VPC or VCC) can be specified in several ways. First, the network operator can implicitly define them using default rules. In this case, all connections are assigned the same values; if the connections are divided into classes, all members of a given class are assigned the same value.

The network operator can also associate parameter values with a given subscriber and assign them at subscription time.

Finally, parameter values tailored to a specific connection can be assigned at connection time. In a permanent virtual connection, the network assigns these values when the connection is set up. For a switched virtual connection, the user and the network negotiate the parameters via signaling protocol.

Another aspect of QOS that may be requested or assigned for a connection is cell loss priority. A user may request two priority levels for an ATM connection, indicating the priority of an individual cell through the cell loss priority (CLP) bit in the cell header.

When two priority levels are used, the traffic parameters for both cell flows must be specified. This is typically done by specifying one set of traffic parameters for high-priority traffic (CLP=0) and one set for all traffic (CLP=0 or 1). This scheme can result in more efficient allocation of network resources.

COAST GUARD

Once the connection-admission control function accepts a connection, the network's usage parameter control (UPC) function monitors it to determine whether traffic is con-

GENERIC CELL RATE ALGORITHM

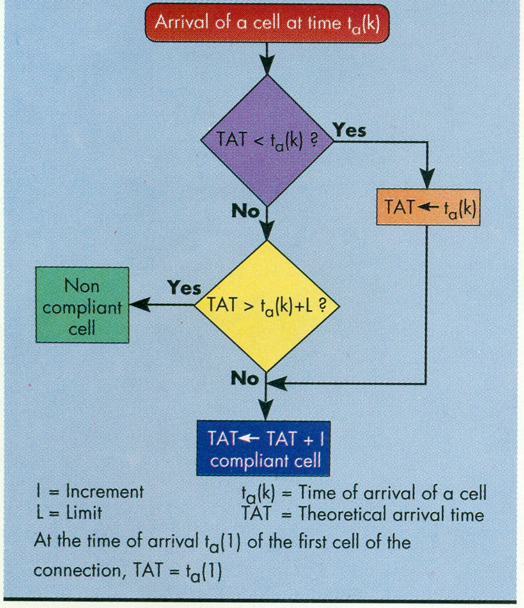
Arrival of a cell at time $t_a(k)$
TAT < $t_a(k)$?
Yes
TAT ← $t_a(k)$
No
Non compliant cell
Yes
TAT > $t_a(k)$+L ?
No
TAT ← TAT + I
compliant cell
I = Increment
L = Limit
$t_a(k)$ = Time of arrival of a cell
TAT = Theoretical arrival time
At the time of arrival $t_a(1)$ of the first cell of the connection, TAT = $t_a(1)$

FIGURE 1: *This generic cell rate algorithm, also used for the sustainable cell rate, is expressed as GCRA(I,L).*

forming to the contract. The main goal is to protect network resources from an overload on one connection that could impair the QOS on other connections. UPC does this by detecting violations of assigned parameters and taking appropriate actions to address them.

In basic terms, traffic flow is compliant if the cell-transmission peak rate doesn't exceed the specified peak cell rate, subject to possible cell delay variation within the specified bounds. An algorithm provided by I.371 serves as an operational definition of the relationship between peak cell rate and CDV, and can be used for usage parameter control to monitor compliance with the traffic contract.

Figure 1 illustrates this generic cell rate algorithm (GCRA), so named because it's also used for the sustainable cell rate. The algorithm takes two arguments, an increment *I* and a limit *L*, and is expressed as GCRA(*I*,*L*).

Suppose you had a specified peak cell rate *R* and a CDV tolerance limit *t*. The arrival time between cells, in the absence of CDV, would be *T*=1/*R*. With CDV, *T* would be the average interarrival time at the peak rate. Thus, the peak cell rate algorithm is expressed as GCRA(*T*,*t*).

The algorithm is initialized with the arrival of the first cell on a connection at time tA(1). The algorithm updates a theoretical arrival time (TAT), which is the target time for the next cell arrival. If the cell arrives later than the TAT, it's compliant, and the TAT is updated to the arrival time plus *I*. If the cell arrives earlier than the TAT but within *t* time units of TAT, then it's still considered compliant, and TAT is incremented by *T*. In the latter case, the cell may arrive early because it does so within the CDV tolerance.

Finally, if the cell arrives too early (before TAT-*t*), then it is outside the CDV tolerance bounds and is declared noncompliant. In this case, the TAT remains unchanged.

An example of this algorithm is shown in Figure 2. (The errors in the original ATM Forum specification have been corrected for this figure.) The time to insert a single 53-octet cell is *d*, and *T*=4.5*d*. Thus, the peak cell rate is equal to the data rate at the user-network interface divided by 4.5.

Let's look at another example. If the data rate is 150Mbps, the peak cell rate is 150/4.5 =26.67Mbps. Part 1 of the figure allows the minimum CDV tolerance *t*=*d*/2. This is just enough to accommodate the fact that because data is transmitted in cells, each arrival time will be an integer multiple of *d*, whereas the increment value is on a 0.5 mark. This tight tolerance means the cell arrival time can never drift very far from the TAT.

As the CDV tolerance *t* increases, cell arrivals can drift further and further from the TAT. More important, however, the potential for cell clumping also increases. The highest volume of clumping occurs when a source can transmit multiple cells back to back (at the full link rate), which is possible when *t* exceeds *T-d*. Parts 3 and 4 of Figure 2 illustrate back-to-back cell clumping.

Another look at Figure 1 shows that it's impossible to build up "credit." If a cell arrives late, meaning there's been an idle period on the connection, the next value of TAT is set relative to the current arrival rather than the current TAT value. If TAT were simply incremented by *T* after each cell arrival, then long idle periods would let sources send long strings of cells at the full link rate. This would create a surge not accommodated in network resource allocation.

The network uses the GCRA (or a similar one) to ensure compliance with the negotiated traffic contract. In the simplest strategy, compliant cells are passed and noncompliant ones are discarded at UPC.

At the network's option, cell tagging may also be used for noncompliant cells. In this case, a noncompliant cell tagged CLP=0 may be tagged CLP=1 and passed. This cell would then be subject to being discarded further on down the network.

If the user has negotiated two lev-

CELL ARRIVAL AT THE PUBLIC USER-NETWORK INTERFACE (T = 4.5 X 6)

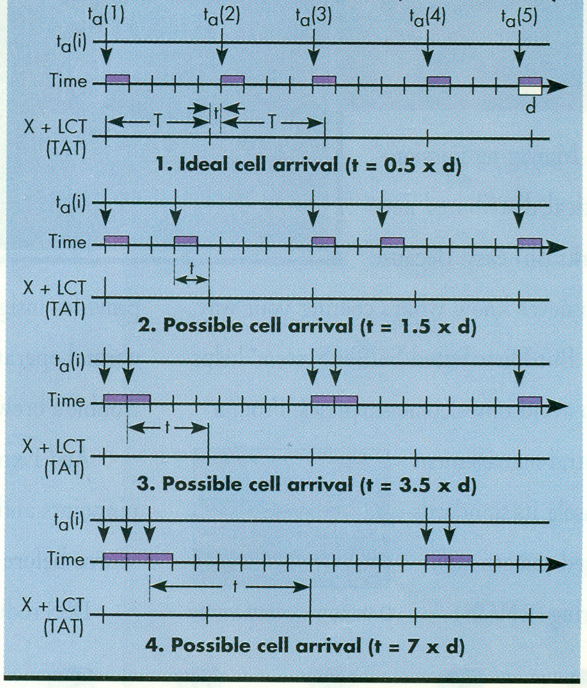
$t_a(1)$
$t_a(2)$
$t_a(3)$
$t_a(4)$
$t_a(5)$
$t_a(i)$
Time
d
T
t
T
X + LCT (TAT)
1. Ideal cell arrival (t = 0.5 x d)
$t_a(i)$
Time
t
X + LCT (TAT)
2. Possible cell arrival (t = 1.5 x d)
$t_a(i)$
Time
t
X + LCT (TAT)
3. Possible cell arrival (t = 3.5 x d)
$t_a(i)$
Time
t
X + LCT (TAT)
4. Possible cell arrival (t = 7 x d)

FIGURE 2: *The peak cell rate in this algorithm is equal to the data rate at the user-network interface divided by 4.5.*

els of cell loss priority for the network, the situation is more complex. In this case, the following rules apply:

- A cell labeled CLP=0 that conforms to the traffic contract for CLP=0 passes.
- A cell labeled CLP=0 that is noncompliant for CLP=0 traffic but compliant for CLP=0 or 1 traffic is tagged and passed.
- A cell labeled CLP=0 that is noncompliant for CLP=0 traffic and noncompliant for CLP=0 or 1 traffic is discarded.
- A cell labeled CLP=1 that is compliant for CLP=1 traffic is passed.
- A cell labeled CLP=1 that is noncompliant for CLP=0 or 1 traffic is discarded.

The UPC function first tests the CLP=0 flow for compliance and then the combined CLP=0 or 1 flow. If the tagging option is used, a noncompliant CLP=0 cell is tagged but is still considered part of the CLP=0 or 1 flow and subjected to the second test.

NAVIGATING THE BUOYS

Priority control comes into play when the network, at some point beyond the UPC function, discards CLP=1 cells. The objective is to discard lower-priority cells to conserve performance for higher-priority ones. The network has no way to discriminate between cells that the source has labeled lower priority and those the UPC function tagged.

In traffic policing, data flow is regulated so that cells, frames, or packets that exceed a certain performance level are discarded or tagged. The GCRA represents one form of traffic policing.

It may be desirable to supplement a traffic policy with a traffic shaping policy. The latter, which is used to smooth out traffic flow and reduce cell clumping, can result in more equitable allocation of resources and shorter average delay times.

A simple approach to traffic shaping is to use a form of the leaky bucket algorithm known as token bucket (Figure 3). In this scheme, a token generator produces tokens at a rate of X per second and places them in the token bucket, which has a maximum capacity of X tokens. Arriving cells are placed in a buffer with a maximum capacity of *K* cells.

Before a cell can be transmitted through the server, one token must be removed from it. If the token bucket is empty, the cell is queued waiting for the next token. Thus, if there's a backlog of cells and an empty bucket, then cells are emitted at the rate of X per second, with no delay variation until the backlog is cleared. In this manner, the token bucket smoothes out bursts of cells.

AN OUNCE OF PREVENTION

ATM congestion control refers to the set of actions the network takes to minimize the intensity, spread, and duration of congestion. These actions are triggered by congestion in one or more network elements. The two primary mechanisms are selective cell discarding and explicit forward congestion indication.

Selective cell discarding is similar to priority control. In the priority control function CLP=1, excess cells are discarded to avoid congestion. In this context, excess cells are limited so the performance objectives for the CLP=0 and CLP=1 flows are still met.

Once congestion occurs, the network is no longer bound to meet all performance objectives. To recover from this situation, it's free to discard any CLP=1 cell and may even discard CLP=0 cells on ATM connections that aren't complying with their traffic contract.

Explicit forward congestion notification for ATM networks operates in basically the same manner as it does in frame relay networks. Any ATM network node experiencing congestion may set an explicit forward congestion indication in the header of cells on connections passing through the node. This tells the user that congestion avoidance procedures should be initiated for traffic flowing in the same direction as the received cell. The user can then invoke actions in higher-layer protocols to adaptively reduce the connection's cell rate.

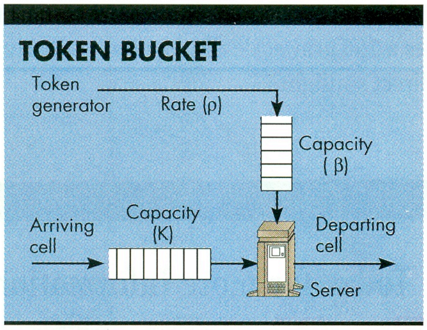

FIGURE 3: *This traffic shaping mechanism, the token bucket algorithm, smooths out bursts of cells.*

The mechanisms described thus far have focused on control schemes for delay-sensitive traffic such as voice and video. But these techniques aren't suited for handling data traffic.

Data traffic is the subject of ongoing research and standardization efforts. Because most of this traffic is much burstier than voice or video traffic, a constant or near-constant delivery rate isn't required. The user's concern is throughput; from a network standpoint, the issue is the possibility that simultaneous bursts of traffic from numerous users could overwhelm switches, causing cells to be dropped.

Until recently, there were three proposals for variable bit-rate congestion control on the table at the ATM Forum. In October 1994, the forum adopted rate-based congestion control, where the network monitors cell flow rate on individual virtual paths and virtual connections, and is responsible for informing source stations of the maximum cell rate that can be tolerated on each path at any given time.

RIDE THE WAVE

Networks need congestion control to protect their switches and buffers from overloads. This ultimately benefits subscribers, who are vulnerable to cell loss in the absence of effective congestion control measures.

With the demand for ATM-based products and services, users are faced with a mixed bag. Some products have no congestion control measures in place, while others use proprietary schemes.

With congestion control specifications solidifying, you should become familiar with the features of ATM services and switches. Expect—and demand—to see manufacturers in this arena move toward standardized solutions. ■

William Stallings is an author and president of Comp-Comm Consulting of Brewster, MA. This article is based on his new book ISDN and Broadband ISDN, With Frame Relay and ATM. *He can be reached on the Internet at stallings@acm.org.*