CHAPTER 4

Traffic Management

The LightStream 2020 multiservice ATM switch (LS2020 switch) traffic management facility, called ControlStream, allows network administrators to maximize available network resources. ControlStream provides the capability to control network resource allocation and ensure efficient use of resources that have not been explicitly allocated.

This chapter is recommended reading if you plan to perform the following tasks:

- Configure an LS2020 switch, particularly if you plan to configure expert mode attributes
- Troubleshoot the LS2020 network

The ControlStream traffic management facility controls two key aspects of the quality of service (QoS) provided to every virtual channel connection (VCC) in an LS2020 network:

- **Transmit priority**—Controls delays experienced by traffic on a VCC
- **Bandwidth availability**—Efficiently allocates network transmission resources

The first section of this chapter discusses transmit priority. Delay-sensitive traffic can be given preferential treatment in an LS2020 network by assigning it a higher transmit priority.

The remainder of this chapter discusses support for allocating network bandwidth, which is controlled by four complementary mechanisms that operate at different levels in the network, as indicated below:

- **Bandwidth allocation**—Keeps track of the bandwidth reserved for each VCC.
- **Connection admission control (CAC)**—Prevents network users from allocating more bandwidth than the network can provide.
- **Traffic policing**—Operates at the edges of the network to ensure that a VCC, once established, does not attempt to use more bandwidth than the network currently has available.
- **Selective cell discard**—Deals with momentary oversubscription of a trunk or edge port. When a traffic surge exceeds the buffer capacity of an output port, the selective cell discard mechanism selectively discards cells, giving preference to different traffic classes according to parameters established by the network administrator.

These bandwidth allocation mechanisms are supported by two additional traffic management facilities:

- **Rate-based congestion-avoidance mechanism**—Keeps the traffic policers on edge modules informed about how much bandwidth is currently available in the network. The traffic policers admit only network traffic that has a high probability of being delivered.
- **Traffic shaping**—Refers to a mechanism used to achieve certain traffic characteristics or to modify certain traffic characteristics in order to match the desired quality of service (QoS).

For example, incoming packet traffic can be metered to reduce surges that might otherwise exceed the buffer capacity of an output port.

Transmit Priority Levels

Delay-sensitive traffic (such as SNA traffic) must traverse the network quickly. By setting the transmit priority attribute (also known as forwarding priority or transfer priority), an LS2020 network administrator can control delays experienced by traffic on a VCC.

When more than one cell or packet is waiting to be forwarded through a switch, trunk, or edge port, cells or packets on VCCs with a higher transmit priority are always forwarded before cells or packets on VCCs with a lower priority. Consequently, traffic on a higher priority VCC experiences consistently less delay than traffic on a lower priority VCC traversing the same path.

Transmit priority levels in an LS2020 network are defined as follows:

- Three transmit priority levels are assigned for user data traffic
- A fourth (and higher) priority level is assigned to internal control traffic (such as VCC connection setup messages and congestion-avoidance updates)
- A fifth (and highest) priority level is assigned to user constant bit rate (CBR) traffic

The fourth and fifth priority levels ensure that the network remains responsive under all traffic conditions.

For details on setting the transmit priority attributes, see the *LightStream 2020 Configuration Guide.*

Bandwidth Allocation

To efficiently use network resources, an LS2020 network keeps track of the bandwidth available at each trunk and edge port. Two categories of bandwidth are defined for purposes of allocating transmission capacity in an LS2020 network:

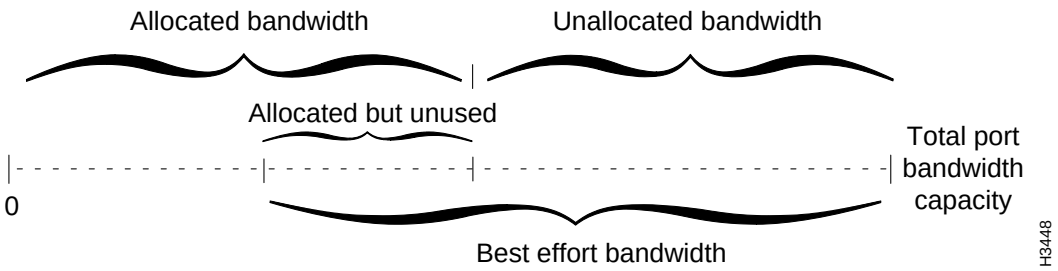
- Allocated bandwidth
- Best effort bandwidth

You use allocated bandwidth for traffic that must be passed through the network under all circumstances. This bandwidth is explicitly reserved for use along the path of a VCC.

Best effort bandwidth is that which is available on a trunk or edge port after allocated bandwidth needs have been met. You use best effort bandwidth for traffic that can be dropped under conditions of network congestion.

Figure 4-1 shows the relationship between allocated and best effort bandwidth on a trunk or edge port.

Figure 4-1 Allocated and Best Effort Bandwidth on Trunk or Edge Port



The allocated bandwidth is the total amount of bandwidth currently reserved by VCCs passing through the port. Allocated bandwidth rises and falls as VCCs are added, removed, or modified.

Best effort bandwidth is the sum of the following:

- The amount of unallocated bandwidth (the difference between the total port transmission capacity and the currently allocated bandwidth)
- The amount of allocated bandwidth not currently in use

The availability of unallocated bandwidth is tracked by the global information distribution (GID) service described in the chapter entitled "ATM Technology." Availability of best effort bandwidth is tracked by the rate-based congestion-avoidance system, which is discussed later in this chapter.

Connection Admission Control

The connection admission control (CAC) mechanism determines whether the network can support a requested VCC by establishing a potential path between the two endpoints for the VCC and deciding whether enough bandwidth exists along that path to support the new VCC.

When a new VCC is created, its bandwidth requirements are determined by configuration parameters set by the network administrator. The CAC mechanism uses two of these parameters:

- ***Insured rate***—Specifies the amount of bandwidth explicitly reserved for a VCC

  This explicitly-reserved bandwidth contributes to the total allocated bandwidth in the network. Traffic that uses allocated bandwidth is referred to as *insured* traffic.

- ***Maximum rate***—Specifies the highest rate at which a VCC is allowed to carry sustained traffic

  Beyond the maximum rate, all traffic on a VCC is discarded.

The network rejects a VCC if no path exists with the capacity to accept the full insured traffic rate. However, the network permits a VCC to be established through the use of trunk or edge ports that do not have the capacity to accept the full maximum rate.

The LS2020 network reserves 100 percent of the insured rate for each connection it accepts. As a result, insured traffic is never dropped under conditions of network congestion. The network reserves only a fraction of the difference between the maximum rate and the insured rate, thereby ensuring that consumers of best effort bandwidth are distributed evenly across available trunks.

When the network reserves bandwidth for a packet interface, the bandwidth reservation on each trunk and edge port traversed by the VCC is adjusted upward to account for the fragmentation that occurs when variable-length packets are segmented into fixed-length ATM cells.

When a VCC is removed from the network, the bandwidth reserved for the VCC becomes available for reallocation.

Traffic Policing

Traffic policing refers to network mechanisms used to detect, discard, or modify ATM cells that violate the quality of service (QoS) parameters agreed to during connection setup. Although applicable to both public and private networks, traffic policing is most often done by public ATM service providers who base their tariff practices on a guaranteed QoS contract.

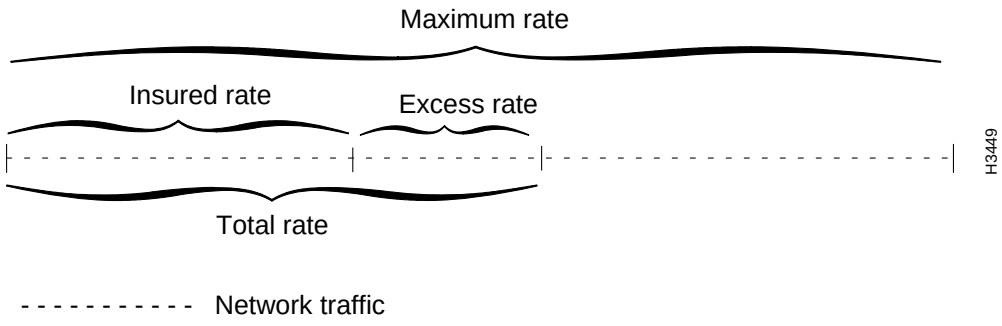
In an LS2020 network, traffic policing is accomplished at the edges of the network for both frame-based and cell-based traffic. The traffic policing mechanism decides whether to accept a unit of incoming traffic (packet or cell), and whether that traffic is to be carried using allocated or best effort bandwidth.

Every VCC in an LS2020 network is controlled by a traffic policer at the input edge port. The operation of the policer is governed by the insured and maximum rates discussed in the previous section, plus two additional parameters:

- Total tate
- Excess rate

The relationships among the VCC traffic policing parameters are illustrated in Figure 4-2.

Figure 4-2 Relationship among VCC Traffic Policing Parameters



The *total rate* is the aggregate bandwidth that the LS2020 network is currently accepting for a given VCC. This rate varies over time depending on information received from the rate-based congestion-avoidance system. The total rate is never lower than the insured rate, nor is it ever higher than the maximum rate. The *excess rate* is the difference between the total rate and the insured rate.

The operation of the traffic policer is also influenced by two parameters not shown in Figure 4-2:

- Insured burst
- Maximum burst

These parameters are per-VCC configuration elements set by the network administrator. They determine how much traffic can be buffered instantaneously for an individual VCC.

As traffic arrives for transmission on a VCC, the LS2020 network uses the total rate and maximum burst parameters to determine which traffic, if any, should be dropped before it enters the network.

The insured rate and insured burst parameters are used to distinguish between insured traffic (using allocated bandwidth) and best effort traffic (using best effort bandwidth). Best effort traffic can be dropped within the network under conditions of traffic congestion.

Leaky Bucket Algorithm

LS2020 traffic policers use the leaky bucket algorithm required by the ATM Forum UNI specification.

The leaky bucket algorithm behaves like a bucket with a hole in the bottom. If data flows into the bucket faster than data flows out through the hole, the bucket eventually "overflows," causing data to be discarded until enough volume again exists in the bucket to accept new data.

The leaky bucket algorithm uses two parameters to control traffic flow:

- **Average rate**—The average number of cells per second that "leak" from the hole in the bottom of the bucket and enter the network.
- **Burst rate**—The rate at which cells are allowed to accumulate in the bucket, expressed in cells per second. For example, if the average burst rate is 10 cells per second, a burst of 10 seconds allows 100 cells to accumulate in the bucket.

The leaky bucket algorithm also uses two state variables:

- **Current time**—The current wall clock time.
- **Virtual time**—A measure of how much data has accumulated in the bucket, expressed in seconds.

  For example, if the average rate is 10 cells per second and 100 cells have accumulated in the bucket, then the virtual time is 10 seconds ahead of the current time.

The leaky bucket algorithm operates on each incoming cell, as indicated in the following formula:

```
virtual time = max (virtual time, current time)
if (virtual time + 1/average rate > current time + burst)
drop the incoming cell
else
put the cell in the bucket
virtual time = virtual time + 1/average rate
```

If, for example, the average rate is 10 cells per second, and the burst is 50 cells, the virtual time and current time remain the same as long as the input rate remains at or below 10 cells per second.

If an instantaneous burst of 25 cells is received, the virtual time moves ahead of the current time by 2.5 seconds. If this is followed immediately by a second burst of 30 cells, the virtual time moves ahead of the current time by 5 seconds, and the last 5 of the 30 cells are dropped.

For packet traffic, the unit of incoming data is larger than a single ATM cell. For packet interfaces, the leaky bucket algorithm takes the packet size into account, as shown in the following formula:

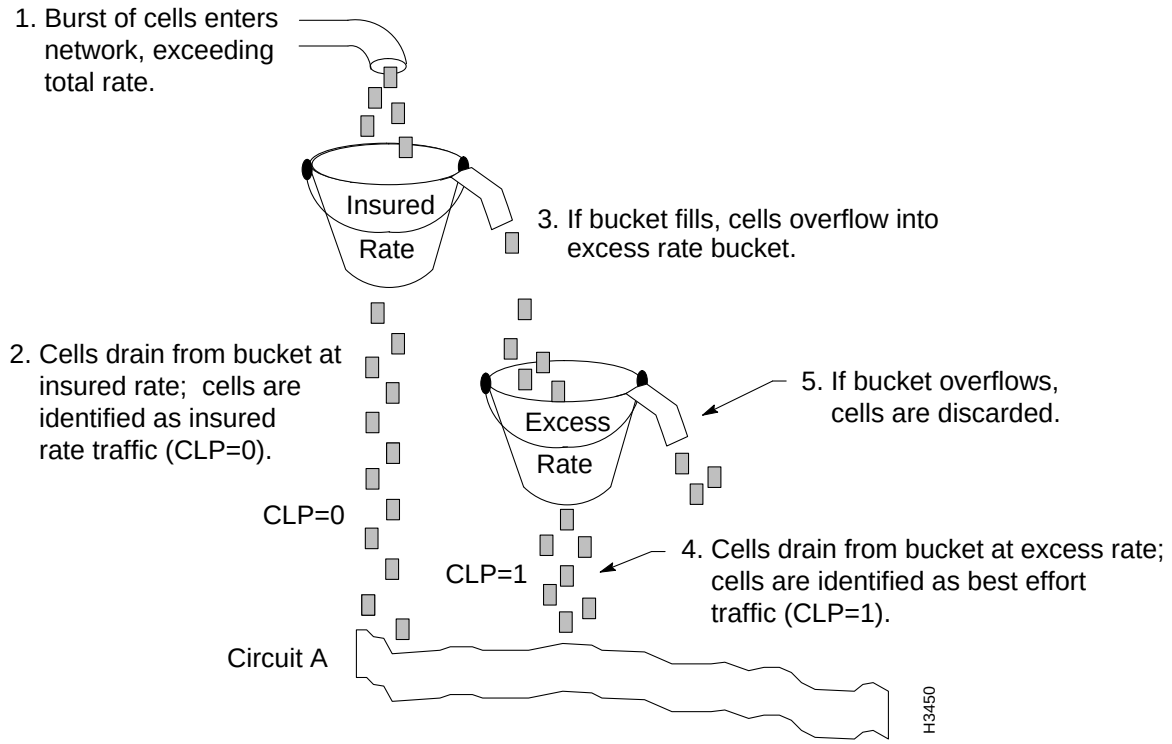
```
virtual time = max (virtual time, current time)
if (virtual time + (packet size / average rate) >   current time + burst)
drop the incoming packet
else
segment the packet into cells
put the cells in the bucket
virtual time = virtual time + (packet size/average rate)
```

In this version of the algorithm, packet size is the number of cells required to transport the packet across the network, including the cell overhead imposed by the processing that occurs in the ATM adaptation layer (AAL).

The algorithm drops the entire packet if it does not fit into the volume available in the leaky bucket. Therefore, it is important to make sure that the burst size on a packet interface is large enough to accommodate at least one or two maximum-size packets.

The relationships of the leaky buckets in an LS2020 traffic policer are shown in Figure 4-3.

Figure 4-3 Operation of Dual Leaky Bucket Traffic Policer



The insured rate bucket in Figure 4-3 determines whether an incoming unit of data (packet or cell) can be accommodated by the insured bandwidth for the VCC. The applicable parameters for this leaky bucket are the insured rate and the insured burst for the VCC. If the test succeeds, the unit of data is segmented into cells (if it is a packet) and prepared for transmission by the LS2020 switch.

The excess rate bucket determines whether enough best effort bandwidth is available to accommodate the incoming unit of data. The applicable parameters for this leaky bucket are the excess rate and maximum burst for the VCC. If the test succeeds, the unit of data is segmented into cells (if it is a packet) and prepared for transmission by the LS2020 switch.

All traffic entering the network through the excess rate bucket is tagged by having its cell loss priority (CLP) bit in the cell header set to "1." This allows the selective cell discard mechanism to distinguish between traffic using *best effort* bandwidth and traffic using *allocated* bandwidth.

Note A special case exists that is not shown in Figure 4-3. On an ATM UNI, the user device can explicitly tag cells by setting the CLP bit in the ATM header (see Figure 1-5). Because these cells are treated as best effort traffic, they are passed directly to the excess rate bucket. Ordinarily, the user device sends enough CLP=0 traffic to consume the bandwidth reserved for the VCC, and the CLP=1 cells are regulated by means of the excess rate bucket, along with any CLP=0 traffic that exceeds the reserved bandwidth. In the unusual case where the user does not send enough CLP=0 traffic to consume the reserved bandwidth, *and at the same time* sends more CLP=1 traffic than the network is currently admitting, the traffic policer assigns the unused reserved bandwidth to CLP=1 traffic.

Traffic Policing Examples

The following examples illustrate how the LS2020 traffic policers operate in typical network situations.

- **Reliable delivery and predictable flow**—For applications that require reliable delivery of a predictable traffic flow, it is best to reserve enough bandwidth to carry the maximum expected data rate.

  To reserve enough bandwidth for this purpose, the network administrator sets both the insured rate and the maximum rate to the maximum expected data rate. As long as the data rate stays within the insured rate and insured burst, all traffic is forwarded. Any traffic exceeding this rate is dropped. All cells flow into the network through the upper leaky bucket (see Figure 4-3), and the setting of the CLP bit to "0" indicates that they are using allocated bandwidth.

  This mode of operation is similar (but not identical) to that of a time division multiplexing (TDM) switch, in which a fixed amount of bandwidth is reserved for each user.

- **File transfer applications**—For file transfer applications in which the user wants to use available bandwidth between two network endpoints on an irregular basis, it is best to use only best effort bandwidth.

  To do so, the network administrator sets the insured rate to zero and the maximum rate to the highest expected data rate. In this case, the amount of bandwidth available to the VCC is regulated by the rate-based congestion-avoidance mechanism. All cells flow into the network through the lower leaky bucket (see Figure 4-3), and the setting of the CLP bit to "1" indicates that they are using best effort bandwidth.

  This mode of operation is similar to that of a packet switch or router, in which the user has access to all the available bandwidth, but no guaranteed bandwidth.

- **Bandwidth reservation**—For some applications, it is useful to reserve enough bandwidth to accommodate routine traffic and to provide best effort bandwidth during periods of peak usage.

  In this case, the network administrator sets the insured rate to accommodate the largest traffic rate expected under routine conditions and sets the maximum rate to accommodate the largest non-routine traffic rate. With these settings, all traffic within the insured and burst rates uses allocated bandwidth, and all traffic between the insured and maximum rates uses best effort bandwidth.

  This mode of operation combines the best of both TDM and packet switching technologies, because each user has access to all the available bandwidth that the system affords, yet a minimum amount of bandwidth is reserved at all times.

Selective Cell Discard Mechanism

For the most part, traffic policers admit only as much traffic as the network can accommodate. Occasionally, simultaneous traffic surges by several different sources overload trunk or output ports to the point that cells must be discarded. In such cases, cells are selected for discard according to the cell drop eligibility level assigned to them at the edge of the network.

A cell can be assigned one of three levels of drop eligibility:

- Best effort
- Best effort plus
- Insured (or guaranteed)

These cell drop eligibility levels are described in Table 4-1.

Because insured cells use allocated bandwidth, they are never selected for discard when traffic congestion occurs. Best effort and best effort plus cells consume unused bandwidth and, therefore, can be dropped under conditions of network congestion. These two levels of cell drop eligibility are assigned on a per-VCC basis by setting a configuration parameter.

Table 4-1 Cell Drop Eligibility

Type of Service	Drop Eligibility	Cell Action
Best effort	Most eligible to be dropped	Dropped first when network congestion occurs.
Best effort plus	Less eligible to be dropped	Dropped next after attempt at best effort transmission when network congestion occurs
Insured (also known as guaranteed)	Least eligible to be dropped	Never dropped when network congestion occurs

Rate-Based Congestion Avoidance Mechanism

The LS2020 rate-based congestion-avoidance mechanism monitors resource utilization within the network and periodically updates traffic policers to admit only as much best effort traffic as the network can accommodate.

This mechanism provides real-time control for preventing congestion and reacting to congestion when it occurs. Network congestion occurs when the traffic load exceeds the capacity of a network transmission resource. Congestion results in increased traffic delays and reduced network throughput.

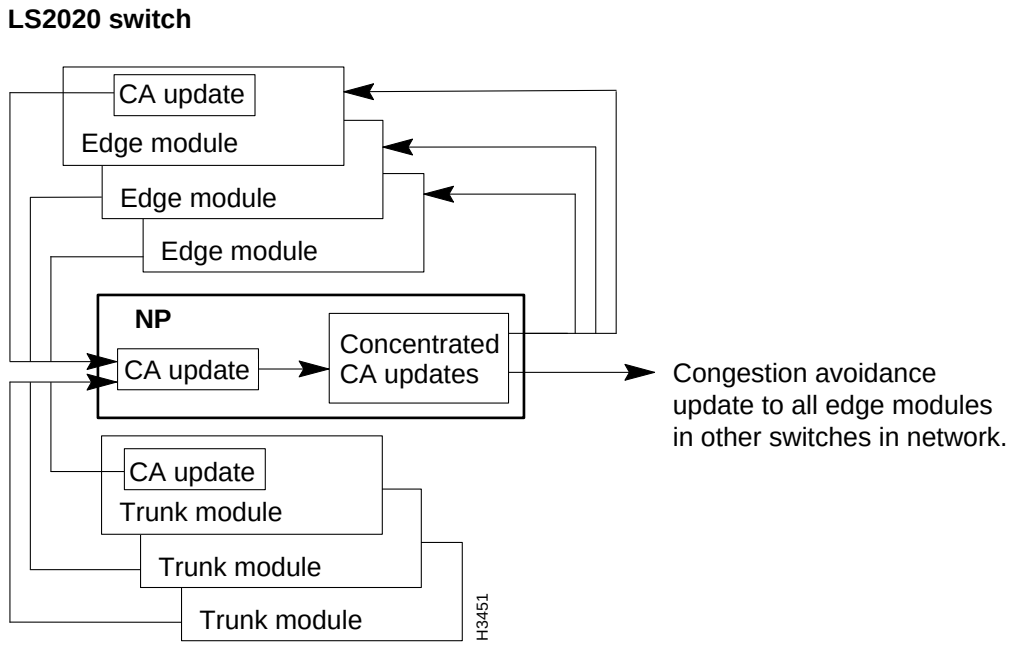
Because the LS2020 network does not permit over-allocation of insured bandwidth, insured traffic is not affected by traffic congestion. Therefore, the LS2020 rate-based congestion-avoidance mechanism regulates only best effort and best effort plus traffic.

Congestion occurs primarily because resource allocation for best effort traffic relies on the statistical nature of such traffic. Since traffic sources are not likely to generate bursts at the same time, networks are generally designed to have less bandwidth capacity than the aggregate input/output capacity of attached hosts. This characteristic contributes to the economic advantage that a network has over a collection of comparatively costly, dedicated lines.

When short traffic bursts occur that exceed the capacity of the network, the selective cell discard mechanism drops best effort traffic. However, this solution works only for short-lived congestion situations. If traffic continues to exceed the capacity of the network, it is more efficient to drop traffic at the edge of the network, since doing so allows more bandwidth to be used by traffic that will reach its destination.

The rate-based congestion-avoidance mechanism operates as a continuous loop (see Figure 4-4). Trunk and edge modules periodically generate congestion-avoidance updates and pass the updates to associated NPs. Each NP then concentrates this information into a larger update and sends it out to every edge module in the network.

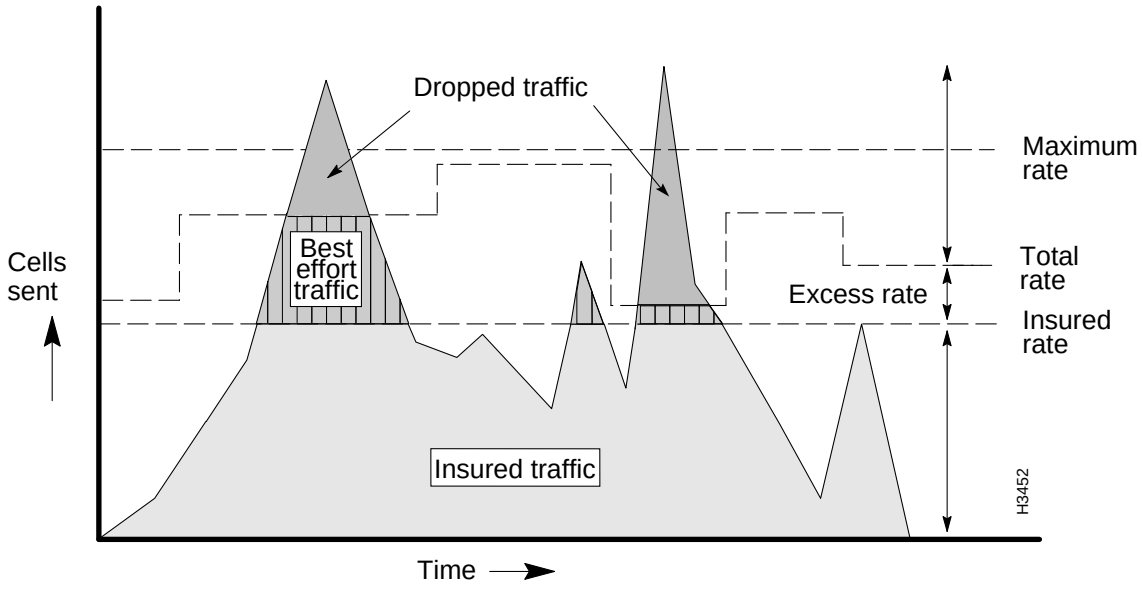
Figure 4-4 Rate-Based Congestion Avoidance Feedback Loop



The traffic policer for every VCC is continually updated and admits only as much best effort traffic as the network can accommodate. When a surge of traffic impacts a trunk or output port, all the VCCs traversing the port are quickly throttled down. When the traffic surge abates, all the VCCs are allowed to send at higher data rates.

Figure 4-5 shows the effect that the rate-based congestion-avoidance mechanism has on traffic policers for a VCC carrying both best effort and insured traffic.

Figure 4-5 Thresholds of Congestion Avoidance Mechanism



The important characteristics of a rate-based congestion-avoidance mechanism include the following:

- Whenever the insured traffic on a trunk or output edge port does not consume all the bandwidth reserved for it, the congestion-avoidance mechanism makes the remaining bandwidth available to best effort traffic, along with any unreserved bandwidth. This is a key difference between an LS2020 switch and a TDM switch, which cannot dynamically reallocate reserved bandwidth.
- The estimates in a congestion-avoidance calculation indicate the total available best effort bandwidth per VCC. Thus, the estimates take into account the number of VCCs traversing the trunk or output line, in addition to the amount of traffic.
- For packet traffic, all the cells in a packet are either dropped or sent into the network. This behavior (unlike random cell dropping) maximizes the throughput of TCP/IP traffic under conditions of network congestion.

Traffic Shaping

Traffic shaping minimizes the occurrence of traffic bursts on the network. You can shape traffic by segmenting it, placing it into buffers, and delaying its propagation into the network. These actions ensure a more constant flow of network traffic.

In an LS2020 network, traffic shaping is performed at all packet interfaces (see the incoming packet traffic in Figure 4-6). Traffic entering ATM UNI interfaces, however, does not need to be shaped, since such traffic obeys the traffic policing parameters established for each VCC.

Figure 4-6 Traffic Shaping