



Regulating Packet Flow Using Traffic Shaping

Last Updated: September 06, 2011

This module contains overview information about regulating the packet flow on a network. Regulating the packet flow (that is, the flow of traffic) on the network is also known as traffic shaping. Traffic shaping allows you to control the speed of traffic leaving an interface. This way, you can match the flow of the traffic to the speed of the interface receiving the packet. Cisco provides three mechanisms for regulating or shaping traffic: Class-Based Traffic Shaping, Generic Traffic Shaping (GTS), and Frame Relay Traffic Shaping (FRTS). Before configuring any of these mechanisms, it is important that you understand the overview information presented in this module.

- [Finding Feature Information, page 1](#)
- [Information About Traffic Shaping, page 1](#)
- [Where to Go Next, page 6](#)
- [Additional References, page 6](#)

Finding Feature Information

Your software release may not support all the features documented in this module. For the latest feature information and caveats, see the release notes for your platform and software release. To find information about the features documented in this module, and to see a list of the releases in which each feature is supported, see the Feature Information Table at the end of this document.

Use Cisco Feature Navigator to find information about platform support and Cisco software image support. To access Cisco Feature Navigator, go to www.cisco.com/go/cfn. An account on Cisco.com is not required.

Information About Traffic Shaping

- [Benefits of Shaping Traffic on a Network, page 2](#)
- [Cisco Traffic Shaping Mechanisms, page 2](#)
- [Token Bucket and Traffic Shaping, page 3](#)

- [Traffic Shaping and Rate of Transfer, page 4](#)
- [How Traffic Shaping Regulates Traffic, page 4](#)
- [Traffic Shaping versus Traffic Policing, page 5](#)

Benefits of Shaping Traffic on a Network

The benefits of shaping traffic on the network include the following:

- It allows you to control the traffic going out an interface, matching the traffic flow to the speed of the interface.
- It ensures that traffic conforms to the policies contracted for it.
- Traffic shaping helps to ensure that a packet adheres to a stipulated contract and determines the appropriate quality of service to apply to the packet.
- It avoids bottlenecks and data-rate mismatches. For instance, central-to-remote site data speed mismatches.
- Traffic shaping prevents packet loss.

Here are some scenarios for which you would use traffic shaping:

- Control access to bandwidth when, for example, policy dictates that the rate of a given interface should not on the average exceed a certain rate even though the access rate exceeds the speed.
- Configure traffic shaping on an interface if you have a network with differing access rates. Suppose that one end of the link in a Frame Relay network runs at 256 kbps and the other end of the link runs at 128 kbps. Sending packets at 256 kbps could cause failure of the applications using the link.

A similar, more complicated case would be a link-layer network giving indications of congestion that has differing access rates on different attached data terminal equipment (DTE); the network may be able to deliver more transit speed to a given DTE device at one time than another. (This scenario warrants that the token bucket be derived, and then its rate maintained.)

- If you offer a substrate service. In this case, traffic shaping enables you to use the router to partition your T1 or T3 links into smaller channels.
- Traffic shaping is especially important in Frame Relay networks because the switch cannot determine which packets take precedence, and therefore which packets should be dropped when congestion occurs. Moreover, it is of critical importance for real-time traffic such as Voice over Frame Relay (VoFR) that latency be bounded, thereby bounding the amount of traffic and traffic loss in the data link network at any given time by keeping the data in the router that is making the guarantees. Retaining the data in the router allows the router to prioritize traffic according to the guarantees it is making. (Packet loss can result in detrimental consequences for real-time and interactive applications.)

Cisco Traffic Shaping Mechanisms

Cisco provides three traffic shaping mechanisms: Class-Based Traffic Shaping, GTS, and FRTS.

All three mechanisms are similar in implementation, though their command-line interfaces (CLIs) differ somewhat and they use different types of queues to contain and shape traffic that is deferred. In particular, the underlying code that determines whether a packet is sent or delayed is common to all three mechanisms, and all three mechanism use a token bucket metaphor (see the [Token Bucket and Traffic Shaping, page 3](#)).

The table below lists the differences between traffic shaping mechanisms.

Table 1 *Differences Between Traffic Shaping Mechanisms*

Traffic Shaping Mechanism	Class-Based	GTS	FRTS
Command-Line Interface	<ul style="list-style-type: none"> Applies configuration on a per-class basis 	<ul style="list-style-type: none"> Applies configuration on a per interface or subinterface basis traffic group command supported 	<ul style="list-style-type: none"> Classes of parameters Applies configuration to all virtual circuits (VCs) on an interface through inheritance mechanism No traffic group
Queues Supported	<ul style="list-style-type: none"> Class-based WFQ (CBWFQ) 	<ul style="list-style-type: none"> Weighted Fair Queueing (WFQ) per interface or subinterface 	<ul style="list-style-type: none"> WFQ, strict priority queue with WFQ, custom queue (CQ), priority queue (PQ), first-in first-out (FIFO) per VC
For More Details, See The . . .	"Regulating Packet Flow on a Per-Class Basis -- Using Class-Based Traffic Shaping" module	"Regulating Packet Flow on a Per-Interface Basis -- Using Generic Traffic Shaping" module	"MQC-Based Frame Relay Traffic Shaping" module

Token Bucket and Traffic Shaping

Traffic shaping uses a token bucket metaphor to shape traffic. A token bucket is a formal definition of a rate of transfer. It has three components: a burst size, a mean rate, and a time interval (Tc). Although the mean rate is generally represented as bits per second, any two values may be derived from the third by the relation shown as follows:

$$\text{mean rate} = \text{burst size} / \text{time interval}$$

Here are some definitions of these terms:

- Mean rate--Also called the committed information rate (CIR), it specifies how much data can be sent or forwarded per unit time on average.
- Burst size--Also called the committed burst (Bc) size, it specifies in bits (or bytes) per burst how much traffic can be sent within a given unit of time to not create scheduling concerns. (For a traffic shaper, it specifies bits per burst.)
- Time interval--Also called the measurement interval, it specifies the time quantum in seconds per burst.

By definition, over any integral multiple of the interval, the bit rate of the interface will not exceed the mean rate. The bit rate, however, may be arbitrarily fast within the interval.

A token bucket is used to manage a device that regulates the data in a flow. For example, the regulator might be a traffic shaper. A token bucket itself has no discard or priority policy. Rather, a token bucket discards tokens and leaves to the flow the problem of managing its transmission queue if the flow overdrives the regulator.

In the token bucket metaphor, tokens are put into the bucket at a certain rate. The bucket itself has a specified capacity. If the bucket fills to capacity, newly arriving tokens are discarded. Each token is

permission for the source to send a certain number of bits into the network. To send a packet, the regulator must remove from the bucket a number of tokens equal in representation to the packet size.

If not enough tokens are in the bucket to send a packet, the packet waits until the bucket has enough tokens. If the bucket is already full of tokens, incoming tokens overflow and are not available to future packets. Thus, at any time, the largest burst a source can send into the network is roughly proportional to the size of the bucket.

Note that the token bucket mechanism used for traffic shaping has both a token bucket and a data buffer, or queue; if it did not have a data buffer, it would be a traffic policer. For traffic shaping, packets that arrive that cannot be sent immediately are delayed in the data buffer.

For traffic shaping, a token bucket permits burstiness but bounds it. It guarantees that the burstiness is bounded so that the flow will never send faster than the capacity of the token bucket plus the time interval multiplied by the established rate at which tokens are placed in the bucket. It also guarantees that the long-term transmission rate will not exceed the established rate at which tokens are placed in the bucket.

Traffic Shaping and Rate of Transfer

Traffic shaping limits the rate of transmission of data. You can limit the data transfer to one of the following:

- A specific configured rate
- A derived rate based on the level of congestion

As mentioned, the rate of transfer depends on these three components that constitute the token bucket: burst size, mean rate, time (measurement) interval. The mean rate is equal to the burst size divided by the interval.

When traffic shaping is enabled, the bit rate of the interface will not exceed the mean rate over any integral multiple of the interval. In other words, during every interval, a maximum of burst size can be sent. Within the interval, however, the bit rate may be faster than the mean rate at any given time.

One additional variable applies to traffic shaping: excess burst (B_e) size. The B_e size corresponds to the number of noncommitted bits--those outside the CIR--that are still accepted by the Frame Relay switch but marked as discard eligible (DE).

In other words, the B_e size allows more than the burst size to be sent during a time interval in certain situations. The switch will allow the packets belonging to the excess burst to go through but it will mark them by setting the DE bit. Whether the packets are sent depends on how the switch is configured.

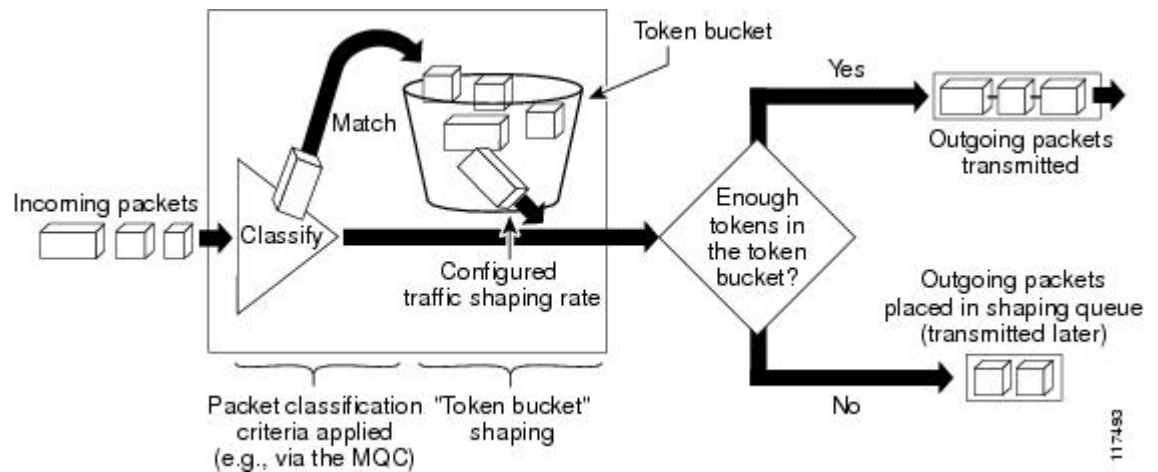
When the B_e size equals 0, the interface sends no more than the burst size every interval, achieving an average rate no higher than the mean rate. However, when the B_e size is greater than 0, the interface can send as many as B_c plus B_e bits in a burst, if in a previous time period the maximum amount was not sent. Whenever less than the burst size is sent during an interval, the remaining number of bits, up to the B_e size, can be used to send more than the burst size in a later interval.

How Traffic Shaping Regulates Traffic

As mentioned previously, Cisco provides three mechanisms for shaping traffic: Class-Based Traffic Shaping, GTS, and FRTS. All three mechanisms are similar in implementation, though their CLIs differ somewhat and they use different types of queues to contain and shape traffic that is deferred.

The figure below illustrates how a traffic shaping mechanism regulates traffic.

Figure 1



In the figure above, incoming packets arrive at an interface. The packets are classified using a "classification engine," such as an access control list (ACL) or the Modular Quality of Service Command-Line Interface (MQC). If the packet matches the specified classification, the traffic shaping mechanism continues. Otherwise, no further action is taken.

Packets matching the specified criteria are placed in the token bucket. The maximum size of the token bucket is the B_c size plus the B_e size. The token bucket is filled at a constant rate of B_c worth of tokens at every T_c . This is the configured traffic shaping rate.

If the traffic shaping mechanism is active (that is, packets exceeding the configured traffic shaping rate already exist in a transmission queue), at every T_c , the traffic shaper checks to see if the transmission queue contains enough packets to send (that is, up to either B_c (or B_c plus B_e) worth of traffic).

If the traffic shaper is not active (that is, there are no packets exceeding the configured traffic shaping rate in the transmission queue), the traffic shaper checks the number of tokens in the token bucket. One of the following occurs:

- If there are enough tokens in the token bucket, the packet is sent (transmitted).
- If there are not enough tokens in the token bucket, the packet is placed in a shaping queue for transmission at a later time.

Traffic Shaping versus Traffic Policing

Although traffic shaping and traffic policing can be implemented together on the same network, there are distinct differences between them, as shown in the table below.

Table 2 *Differences Between Traffic Shaping and Traffic Policing*

	Traffic Shaping	Traffic Policing
Triggering Event	<ul style="list-style-type: none"> Occurs automatically at regular intervals (Tc). <p>or</p> <p>Occurs whenever a packet arrives at an interface.</p>	<ul style="list-style-type: none"> Occurs whenever a packet arrives at an interface.
What it Does	<ul style="list-style-type: none"> Classifies packets. If packet does not meet match criteria, no further action is taken. Packets meeting match criteria are sent (if there are enough tokens in the token bucket) <p>or</p> <p>Packets are placed in a queue for transmission later.</p> <ul style="list-style-type: none"> If the number of packets in the queue exceed the queue limit, the packets are dropped. 	<ul style="list-style-type: none"> Classifies packets. If packet does not meet match criteria, no further action is taken. Packets meeting match criteria and conforming to, exceeding, or violating a specified rate, receive the configured policing action (for example, drop, send, mark then send). Packets are not placed in queue for transmission later.

Where to Go Next

To configure Class-Based Traffic Shaping, see the "Regulating Packet Flow on a Per-Class Basis -- Using Class-Based Traffic Shaping" module.

To configure GTS, see the "Regulating Packet Flow on a Per-Interface Basis -- Using Generic Traffic Shaping" module.

To configure FRTS, see the "MQC-Based Frame Relay Traffic Shaping" module.

Additional References

Related Documents

Related Topic	Document Title
Cisco IOS commands	Cisco IOS Master Commands List, All Releases
QoS commands: complete command syntax, command modes, command history, defaults, usage guidelines, and examples	<i>Cisco IOS Quality of Service Solutions Command Reference</i>

Related Topic	Document Title
Packet classification	"Classifying Network Traffic" module
MQC, policy maps, class maps, and hierarchical policy maps	"Applying QoS Features Using the MQC" module
WFQ, CBWFQ, PQ, CQ, FIFO and other queueing mechanisms	"Congestion Management Overview" module
Class-Based Traffic Shaping	"Regulating Packet Flow on a Per-Class Basis -- Using Class-Based Traffic Shaping" module
GTS	"Regulating Packet Flow on a Per-Interface Basis -- Using Generic Traffic Shaping" module
FRTS	"MQC-Based Frame Relay Traffic Shaping" module

Standards

Standard	Title
No new or modified standards are supported, and support for existing standards has not been modified.	--

MIBs

MIBs	MIBs Link
No new or modified MIBs are supported, and support for existing MIBs has not been modified.	To locate and download MIBs for selected platforms, Cisco IOS releases, and feature sets, use Cisco MIB Locator found at the following URL: http://www.cisco.com/go/mibs

RFCs

RFCs	Title
No new or modified RFCs are supported, and support for existing RFCs has not been modified.	--

Technical Assistance

Description	Link
The Cisco Support and Documentation website provides online resources to download documentation, software, and tools. Use these resources to install and configure the software and to troubleshoot and resolve technical issues with Cisco products and technologies. Access to most tools on the Cisco Support and Documentation website requires a Cisco.com user ID and password.	http://www.cisco.com/cisco/web/support/index.html

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1005R)

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.