FortiOS™ Handbook
Traffic Shaping for FortiOS 5.0

Traffic Shaping for FortiOS 5.0

March 5, 2014

01-502-120097-20130429

Contents

Change Log

Date	Change Description
2014-02-13	-Added Change Log. -Edited the document throughout to bring content up to date for patch 6. -Updated document with new content from "What's new in FortiOS 5.0 Patch 3"(IPv6 support for traffic shaping, reverse shared traffic shaping, and per-IP traffic shaping). -Added section: Adding Per-IP traffic shapers to a security policy.

Introduction

Welcome and thank you for selecting Fortinet products for your network protection.

With the ever-increasing demands on network systems for a number of protocols, including email, HTTP traffic both internally and externally to the internet, voice over IP, FTP, and more, slow traffic is becoming a reality. Important traffic may even be dropped or slowed to an unusable speed. Web traffic delays can result in a loss of revenue for businesses.

Traffic shaping attempts to normalize traffic peaks and bursts to prioritize certain flows over others. There is a physical limitation to the amount of data which can be buffered and to the length of time it can be buffered.

FortiGate units provide Quality of Service (QoS) by applying bandwidth limits and prioritization. Using traffic shaping, you can adjust how your FortiGate unit allocates resources to different traffic types to improve performance and stability of latency sensitive or bandwidth intensive network applications.

This document describes Quality of Service (QoS), traffic shaping, FortiGate traffic shaping algorithms, and includes configuration procedures for traffic shaping on FortiGate units.

This guide contains the following sections:

The purpose of traffic shaping describes traffic shaping theories and quality of service.

Traffic shaping methods lists different methods of applying traffic shaping within FortiOS, and explains how to use TOS and Differentiated Services.

Examples provides basic application scenarios for shapers.

Troubleshooting traffic shaping lists diagnose commands to use for determining if traffic shapers are working correctly.

The purpose of traffic shaping

Traffic shaping, or traffic management, controls the bandwidth available and sets the priority of traffic processed by the policy to control the volume of traffic for a specific period (bandwidth throttling) or rate the traffic is sent (rate limiting).

Traffic shaping attempts to normalize traffic peaks and bursts to prioritize certain flows over others. But there is a physical limitation to the amount of data which can be buffered and to the length of time. Once these thresholds have been surpassed, frames and packets will be dropped, and sessions will be affected in other ways.

A basic traffic shaping approach is to prioritize certain traffic flows over other traffic whose potential loss is less disadvantageous. This would mean that you accept certain sacrifices in performance and stability on low-priority traffic, to increase or guarantee performance and stability to high-priority traffic.

If, for example, you are applying bandwidth limitations to certain flows, you must accept the fact that these sessions can be limited and therefore negatively impacted.

Note that traffic shaping is effective for normal IP traffic at normal traffic rates. Traffic shaping is not effective during periods when traffic exceeds the capacity of the FortiGate unit. Because packets must be received by the FortiGate unit before they are subject to traffic shaping, if the FortiGate unit cannot process all of the traffic it receives, then dropped packets, delays, and latency are likely to occur.

To ensure that traffic shaping is working at its best, make sure that the interface Ethernet statistics show no errors, collisions or buffer overruns.

Accelerated interfaces (NPx network processors and CE) affect traffic shaping. For more information, see the *FortiGate Hardware* Guide.

Quality of Service

Quality of Service (QoS) is the capability to adjust some quality aspects of your overall network traffic. This can include such techniques as priority-based queuing and traffic policing. Because bandwidth is finite and because some types of traffic are slow, jitter or packet loss sensitive, bandwidth intensive, or operation critical, QoS can be a useful tool for optimizing the performance of the various applications on your network.

Before implementing QoS, organizations should first identify the types of traffic that are important to the organization, the types of traffic that use high amounts of bandwidth, and the types of traffic that are sensitive to latency or packet loss.

For example, a company might want to guarantee sufficient bandwidth for revenue producing e-commerce traffic. They need to ensure that transactions can be completed and that clients do not experience service delays and interruptions. At the same time, the company may need to ensure low latency for voice over IP (VoIP) traffic used by sales and customer support, while traffic latency and bursts may be less critical to the success of other network applications such as long term, resumable file transfers. Many organizations discover that QoS is especially important for managing their voice and streaming multi-media traffic. These types of traffic can rapidly consume bandwidth and are sensitive to latency.

Discovering the needs and relative importance of each traffic type on your network will help you to design an appropriate overall approach, including how you will configure each available QoS component technique. Some organizations discover that they only need to configure bandwidth limits for some services. Other organizations determine that they need to fully configure

interface and security policy bandwidth limits for all services, and prioritize queuing of critical services relative to traffic rate.

You can implement QoS on FortiGate units using the following techniques:

Traffic policing	Drops packets that do not conform to bandwidth limitations.
Traffic shaping	Ensures that the traffic may consume bandwidth at least at the guaranteed rate by assigning a greater priority queue if the guarantee is not being met. Also ensures that the traffic cannot consume bandwidth greater than the maximum at any given instant in time. Flows greater than the maximum rate are subject to traffic policing.
Queuing	Transmits packets in order of their assigned priority queue for that physical interface. All traffic in a higher priority traffic queue must be completely transmitted before traffic in lower priority queues will be transmitted.

When deciding how to configure QoS techniques, it can be helpful to know when FortiGate units employ each technique in the overall traffic processing flow, and the considerations that arise from those mechanisms.

Traffic policing

The FortiGate unit begins to process traffic as it arrives (ingress) and departs (egress) on an interface. In later phases of the network processing, such as enforcing maximum bandwidth use on sessions handled by a security policy, if the current rate for the destination interface or traffic regulated by that security policy is too high, the FortiGate unit may drop the packet. Time spent on prior processing, such as web filtering, decryption or IPS, is often wasted on packets that are not forwarded. This applies to VLAN interfaces and physical interfaces.

You can prevent this wasted effort on ingress by configuring the FortiGate unit to preemptively drop excess packets when they are received at the source interface, before most other traffic processing is performed:

```
config system interface
  edit <interface_name>
    set inbandwidth <rate_int>
  next
end
```

where `<rate_int>` is the bandwidth limit in Kb/s. Excess packets will be dropped. If `inbandwidth` is `0`, the rate is not limited.

A similar command is available that can be performed on egress as well using the CLI commands:

```
config system interface
  edit <interface_name>
    set outbandwidth <rate_int>
  next
end
```

As with ingress, setting the rate to 0 (zero) sets the rate to unlimited.

Rate limiting traffic accepted by the interface enables you to restrict incoming traffic to rates that, while no longer the full capacity of the interface, at the traffic shaping point in the processing are more likely to result in acceptable rates of outgoing traffic per destination

Traffic Shaping for FortiOS 5.0

interface or all security policies. This conserves FortiGate processing resources for those packets that are more likely to be viable completely to the point of egress.

Excessive traffic policing can degrade network performance rather than improve it. For details on factors you may want to consider when configuring traffic policing, see "Important considerations" on page 14.

Bandwidth guarantee, limit, and priority interactions

After packet acceptance, the FortiGate unit classifies traffic and may apply traffic policing at additional points during processing. It may also apply additional QoS techniques, such as prioritization and traffic shaping. Traffic shaping consists of a mixture of traffic policing to enforce bandwidth limits, and priority queue adjustment to assist packets in achieving the guaranteed rate.

If you have configured prioritization, the FortiGate unit prioritizes egressing packets by distributing them among FIFO (first in, first out) queues associated with each possible priority number. Each physical interface has six priority queues. Virtual interfaces do not have their own queues, and instead use the priority queues of the physical interface to which they are bound.

Each physical interface's six queues are queue 0 to queue 5, where queue 0 is the highest priority queue. However, for the reasons described below, you may observe that your traffic uses only a subset of those six queues. Some traffic may always use a certain queue number. Some queuing may vary by the packet rate or mixture of services. Some queue numbers may be used only by through traffic for which you have configured traffic shaping in the security policy that applies to that traffic session. For example:

- Administrative access traffic will always use queue 0.
- Traffic matching security policies *without* traffic shaping may use queue 0, queue 1, or queue 2. Which queue will be used depends on the priority value you have configured for packets with that ToS (type of service) bit value, if you have configured ToS-based priorities.
- Traffic matching security policies *with* traffic shaping may use any queue. Which queue will be used depends on whether the packet rate is currently below the guaranteed bandwidth (queue 0), or above the guaranteed bandwidth. Packets at rates greater than the maximum bandwidth limit are dropped.
- If the global tos-based-priority is low (3), the priority in a traffic-shaper is medium (2) and a packet flows though a policy that refers to the shaper, the packet will be assigned the priority defined by the shaper, in this case medium (2).

Prioritization and traffic shaping behavior varies by your configuration, the service types and traffic volumes, and by whether the traffic is through traffic, or the traffic originates from or terminates at the FortiGate unit itself.

FortiGate traffic

Administrative access to the FortiGate through HTTPS or SSH, or IPsec tunnel negotiations, security policies do not apply, and therefore FortiGate units do not apply traffic shaping. Such traffic also uses the highest priority queue, queue 0. In other words:

packet priority = 0

Exceptions to this rule include traffic types that are connections related to a session governed by a security policy.

For example, if you have enabled scanning by FortiGuard antivirus, traffic from the sender technically terminates at the FortiGate proxy that scans that traffic type; the FortiGate unit initiates a second connection that transmits scanned content to its destination. Because the second connection's traffic is technically originating from the FortiGate proxy and therefore the

FortiGate unit itself, it uses the highest priority queue, queue 0. However, this connection is logically associated with through traffic, and is therefore subject to possible bandwidth enforcement and guarantees in its governing security policy. In this way, it behaves partly like other through traffic.

Through traffic

For traffic passing through the FortiGate unit, the method a FortiGate unit uses to determine the priority queue varies by whether you have enabled Traffic Shaping. Packets may or may not use a priority queue directly or indirectly derived from the type of service (ToS) bit — sometimes used instead with differentiated services — in the packet's IP header.

If Traffic Shaping is not enabled in the security policy, the FortiGate unit neither limits nor guarantees bandwidth, and traffic for that session uses the priority queue determined directly by matching the ToS bit in its header with your configured values:

```
config system global
   set tos-based-priority {high | low | medium}
end
```

or, if you have configured a priority specifically for that TOS bit value:

```
config system tos-based-priority
   edit <id_int>
      set tos [0-15]
      set priority {high | low | medium}
   next
end
```

where `tos` is the value of the ToS bit in the packet's IP header, and `high` has a priority value of 0 and `low` is 2. Priority values configured in the second location will override the global ToS-based priority. In other words:

packet priority = ToS-based priority

For example, you might specify that packets with a ToS bit value of 2 should use queue 0, the highest priority queue:

```
config system tos-based-priority
   edit 15
      set tos 2
      set priority high
   next
end
```

If Traffic Shaping is enabled in the security policy using shared traffic shapers, the FortiGate unit may instead or also subject packets to traffic policing, or priority queue increase in an effort to meet bandwidth guarantees configured in the shaper:

```
config firewall shaper traffic-shaper
   edit <shaper_name>
      ...
      set priority {high | medium | low}
      set maximum-bandwidth <rate>
      set guaranteed-bandwidth <rate>
end
```

where `high` has a priority value of 1 and `low` is 3, and `<rate>` is the bandwidth limit in kilobits per second.

Traffic Shaping for FortiOS 5.0

Figure 1: Traffic queuing as packet rate increases
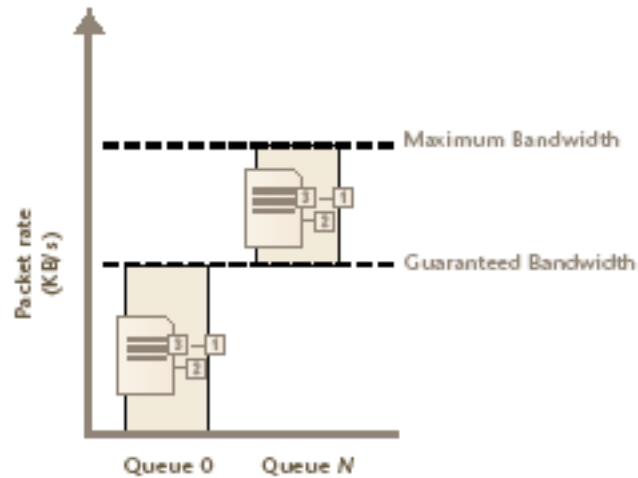
- If the current packet rate is less than Guaranteed Bandwidth, packets use priority queue 0. In other words:

  packet priority = 0

- If the current packet rate is greater than Guaranteed Bandwidth but less than Maximum Bandwidth, the FortiGate unit assigns a priority queue by adding the numerical value of the security policy-based priority, where the value of High is 1, and Low is 3, with the numerical value of the ToS-based priority, where `high` has a priority value of 0 and `low` is 2. Because the two values are added, depending on the your configured ToS-based priorities, packets in this category could use queues from queue 1 to queue 5. In other words:

  packet priority = ToS-based priority + security policy-based priority

  For example, if you have enabled Traffic Shaping in the security policy, and the security policy's Traffic Priority is Low (value 3), and the priority normally applied to packets with that ToS bit is `medium` (value 1), then packets have a total packet priority of 4, and use priority queue 4.

- If the current packet rate exceeds Maximum Bandwidth, excess packets are dropped.

Calculation and regulation of packet rates

Packet rates specified for Maximum Bandwidth or Guaranteed Bandwidth are:

    rate = amount / time

where rate is expressed in kilobits per second (Kb/s).

Burst size at any given instant cannot exceed the amount configured in Maximum Bandwidth. Packets in excess are dropped. Packets deduct from the amount of bandwidth available to subsequent packets and available bandwidth regenerates at a fixed rate. As a result, bandwidth available to a given packet may be less than the configured rate, down to a minimum of 0 Kb/s.

Rate calculation and behavior can alternatively be described using the token bucket metaphor, where:

- a traffic flow has an associated bucket, which represents burst size bounds, and is the size of your configured bandwidth limit
- the bucket receives tokens, which represent available bandwidth, at the fixed configured rate
- as time passes, tokens are added to the bucket, up to the capacity of the bucket; excess tokens are discarded
- when a packet arrives, the packet must deduct bandwidth tokens from the bucket equal to its packet size in order to egress
- packets cannot egress if there are insufficient tokens to pay for its egress; these nonconforming packets are dropped

Bursts are not redistributed over a longer interval, so bursts are propagated rather than smoothed, although their peak size is limited.

Maximum burst size is the capacity of the bucket (the configured bandwidth limit); actual size varies by the current number of tokens in the bucket, which may be less than bucket capacity, due to deductions from previous packets and the fixed rate at which tokens accumulate. A depleted bucket refills at the rate of your configured bandwidth limit. Bursts cannot borrow tokens from other time intervals. This behavior is illustrated in Figure 2 on page 12.

Figure 2: Bursts and bandwidth limits over time
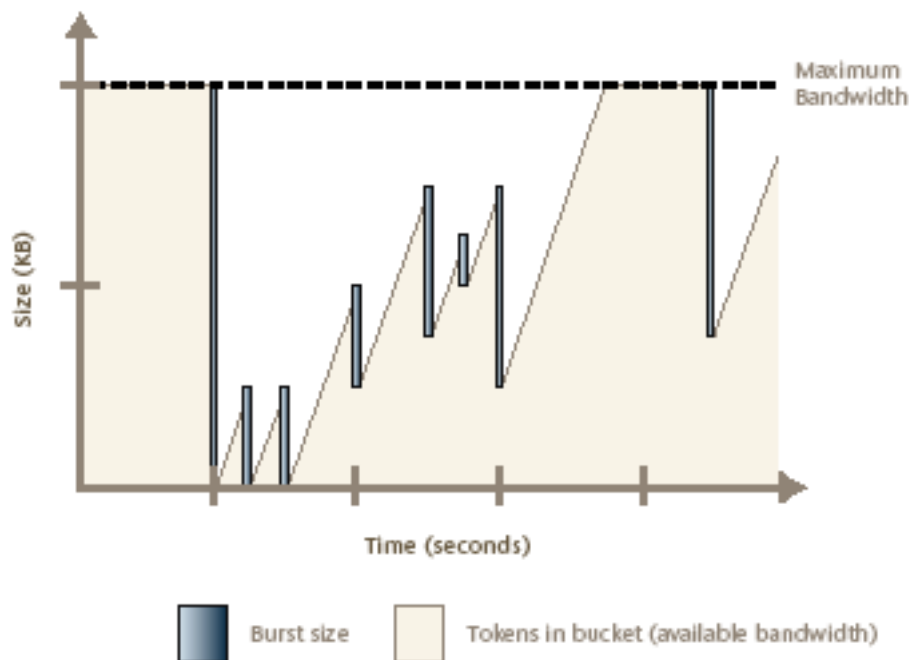
By limiting traffic peaks and token regeneration in this way, the available bandwidth at a given moment may be less than bucket capacity, but your limit on the total amount per time interval is ensured. That is, total bandwidth use during each interval of 1 second is at most the integral of your configured rate.

You may observe that external clients, such as FTP or BitTorrent clients, initially report rates between Maximum Bandwidth and twice that of Maximum Bandwidth, depending on the size of their initial burst. This is notably so when a connection is initiated following a period of no network activity.The apparent discrepancy in rates is caused by a difference of perspective when delimiting time intervals. A burst from the client may initially consume all tokens in the

Traffic Shaping for FortiOS 5.0

bucket, and before the end of 1 second, as the bucket regenerates, be allowed to consume almost another bucket's worth of bandwidth. From the perspective of the client, this constitutes one time interval. From the perspective of the FortiGate unit, however, the bucket cannot accumulate tokens while full; therefore, the time interval for token regeneration begins **after** the initial burst, and does not contain the burst. These different points of reference result in an initial discrepancy equal to the size of the burst — the client's rate contains it, but the FortiGate unit's rate does not. If the connection is sustained to its limit and time progresses over an increasing number of intervals, however, this discrepancy decreases in importance relative to the bandwidth total, and the client's reported rate will eventually approach that of the FortiGate unit's configured rate limit.

For example, your Maximum Bandwidth might be 50 Kb/s and there has been no network activity for one or more seconds. The bucket is full. A burst from an FTP client immediately consumes 50 Kb. Because the bucket completely regenerates over 1 second, by the time almost another 1 second has elapsed from the initial burst, traffic can consume another 49.999 Kb, for a total of 99.999 Kb between the two points in time. From the vantage point of an external FTP client regulated by this bandwidth limit, it therefore initially appears that the bandwidth limit is 99.999 Kb/s, almost twice the configured limit of 50 Kb/s. However, bucket capacity only regenerates at your configured rate of 50 Kb/s, and so the connection can only consume a maximum of 50 Kb during each second thereafter. The result is that as bandwidth consumption is averaged over an increasing number of time intervals, each of which are limited to 50 Kb/s, the effects of the first interval's doubled bandwidth size diminishes proportionately, and the client's reported rate eventually approach your configured rate limit. This effect is illustrated in Table 1 on page 13.

Table 1: Effects of a 50 Kb/s limit on client reported rates

Total size transferred (Kb)	Time (s)	Rate reported by client (Kb/s)
99.999 (50 + 49.999)	1	99.999
149.999	2	74.999
199.999	3	66.666
249.999	4	62.499
299.999	5	59.998
349.999	6	58.333
...	...	...

Guaranteed Bandwidth can also be described using a token bucket metaphor. However, because this feature attempts to achieve or exceed a rate rather than limit it, the FortiGate unit does not discard non-conforming packets, as it does for Maximum Bandwidth; instead, when the flow does not achieve the rate, the FortiGate unit increases the packets' priority queue, in an effort to increase the rate.

Guaranteed and maximum bandwidth rates apply to the bidirectional total for all sessions controlled by the security policy. For example, an FTP connection may entail two separate connections for the data and control portion of the session; some packets may be reply traffic rather than initiating traffic. All packets for both connections are counted when calculating the packet rate for comparison with the guaranteed and maximum bandwidth rate.

Important considerations

In essence, by implementing QoS, you trade some performance and/or stability from traffic X by discarding packets or introducing latency in order to improve performance and stability of traffic Y. The best traffic shaping configuration for your network will appropriately balance the needs of each traffic flow by considering not only the needs of your particular organization, but also the resiliency and other characteristics of each particular service.

For example, you may find that web browsing traffic is both more resistant to interruptions or latency and less business critical than UDP or VoIP traffic, and so you might implement less restrictive QoS measures on UDP or VoIP traffic than on HTTP traffic.

An appropriate QoS configuration will also take into account the physical limits of your network devices, and the interactions of the aforementioned QoS mechanisms, described in "Bandwidth guarantee, limit, and priority interactions" on page 9.

You may choose to configure QoS differently based upon the hardware limits of your network and FortiGate unit. Traffic shaping may be less beneficial in extremely high-volume situations where traffic exceeds a network interface's or your FortiGate model's overall physical capacity. A FortiGate unit must have sufficient resources, such as memory and processing power, to process all traffic it receives, and to process it at the required rate; if it does not have this capacity, then dropped packets and increased latency are likely to occur. For example, if the total amount of memory available for queuing on a physical interface is frequently exceeded by your network's typical packet rates, frames and packets must be dropped. In such a situation, you might choose to implement QoS using a higher model FortiGate unit, or to configure an incoming bandwidth limit on each interface.

Incorrect traffic shaping configurations can actually further degrade certain network flows, because excessive discarding of packets or increased latency beyond points that can be gracefully handled by that protocol can create additional overhead at upper layers of the network, which may be attempting to recover from these errors. For example, a configuration might be too restrictive on the bandwidth accepted by an interface, and may therefore drop too many packets, resulting in the inability to complete or maintain a SIP call.

To optimize traffic shaping performance, first ensure that the network interface's Ethernet statistics are clean of errors, collisions, or buffer overruns. To check the interface, enter the following diagnose command to see the traffic statistics:

```
diagnose hardware deviceinfo nic <port_name>
```

If these are not clean, adjust FortiGate unit and settings of routers or other network devices that are connected to the FortiGate unit. For additional information, see "Troubleshooting traffic shaping" on page 39.

Once Ethernet statistics are clean, you may want to use only some of the available FortiGate QoS techniques, or configure them differently, based upon the nature of FortiGate QoS mechanisms described in "Bandwidth guarantee, limit, and priority interactions" on page 9. Configuration considerations include:

- For maximum bandwidth limits, ensure that bandwidth limits at the source interface and/or the security policy are not too low, which can cause the FortiGate unit to discard an excessive number of packets.

- For prioritization, consider the ratios of how packets are distributed between available queues, and which queue is used by which types of services. If you assign most packets to the same priority queue, it negates the effects of configuring prioritization. If you assign many high bandwidth services to high priority queues, lower priority queues may be starved for bandwidth and experience increased or indefinite latency. For example, you may want to prioritize a latency-sensitive service such as SIP over a bandwidth-intensive service such as

FTP. Consider also that bandwidth guarantees can affect the queue distribution, assigning packets to queue 0 instead of their typical queue in high-volume situations.

- You may or may not want to guarantee bandwidth, because it causes the FortiGate unit to assign packets to queue 0 if the guaranteed packet rate is not currently being met. Comparing queuing behavior for lower-bandwidth and higher-bandwidth situations, this would mean that effects of prioritization only become visible as traffic volumes rise and exceed their guarantees. Because of this, you might want only some services to use bandwidth guarantees, to avoid the possibility that in high-volume situations all traffic uses the same queue, thereby negating the effects of configuring prioritization.

- For prioritization, configure prioritization for all through traffic. You may want to configure prioritization by either ToS-based priority or security policy priority, but not both. This simplifies analysis and troubleshooting.

  Traffic subject to both security policy and ToS-based priorities will use a combined priority from both of those parts of the configuration, while traffic subject to only one of the prioritization methods will use only that priority. If you configure both methods, or if you configure either method for only a subset of your traffic, packets for which a combined priority applies will frequently receive a lower priority queue than packets for which you have only configured one priority method, or for which you have not configured prioritization.

  For example, if both ToS-based priority and security policy priority both dictate that a packet should receive a "medium" priority, in the absence of bandwidth guarantees, a packet will use queue 3, while if only ToS-based priority had been configured, the packet would have used queue 1, and if only security policy-based priority had been configured, the packet would have used queue 2. If no prioritization had been configured at all, the packet would have used queue 0.

For example alternative QoS implementations that illustrate these considerations, see "Examples" on page 32

Traffic shaping methods

In FortiOS, there are three types of traffic shaping configuration. Each has a specific function, and all can be used together in varying configurations. Policy shaping enables you to define the maximum bandwidth and guaranteed bandwidth set for a security policy. Per-IP shaping enables you to define traffic control on a more granular level. Application traffic shaping goes further, enabling traffic controls on specific applications or application groupings.

This chapter describes the types of traffic shapers and how to configure them in the web-based manager and the CLI.

Traffic shaping options

When configuring traffic shaping for your network, there are three different methods to control the flow of network traffic to ensure that the desired traffic gets through while also limiting the bandwidth that users use for other less important or bandwidth consuming traffic. The three shaping options are:

- shared policy shaping - bandwidth management by security policies
- per-IP shaping - bandwidth management by user IP addresses
- application control shaping - bandwidth management by application

Shared policy shaping and per IP shaping are enabled within the security policy, while the application control shaping is configured in *Security Profiles > Application Control > Application Sensors,* and enabled in the security policy by enabling *Application Control* in the *Security Profiles* section.

The FortiGate unit offers three different traffic shaping options, all of which can be enabled at the same time within the same security policy. Generally speaking, the hierarchy for shapers in FortiOS is:

- Application Control shaper
- Security policy shaper
- Per-IP shaper

With this hierarchy, if an application control list has a traffic shaper defined, it will have precedence always over any other security policy shaper. For example, with the example above creating an application control for Facebook, the shaper defined for Facebook will supersede any security policy enabled traffic shapers. While the Facebook application may reach its maximum bandwidth, the user can still have the bandwidth room available from the shared shaper and, if enabled, the per-IP shaper.

Equally, any security policy shared shaper will have precedence over any per-IP shaper. However, traffic that exceeds any of these shapers will be dropped. For example, the policy shaper will take effect first, however, if the per-IP shaper limit is reached first, then traffic for that user will be dropped even if the shared shaper limit for the policy has not been exceeded.

Shared policy shaping

Traffic shaping by security policy enables you to control the maximum and/or guaranteed throughput for a selected security policy. When configuring a shaper, you can select to apply the bandwidth shaping per policy or for all policies. Depending on your selection, the FortiGate unit will apply the shaping rules differently.

Per policy

When selecting a shaper to be *per policy*, the FortiGate unit will apply the shaping rules defined to each security policy individually.

For example, the shaper is set to be per policy with a maximum bandwidth of 1000 Kb/s. There are four security policies monitoring traffic through the FortiGate unit. Three of these have the shaper enabled. Each security policy has the same maximum bandwidth of 1000 Kb/s.

Per policy traffic shaping is compatible with client/server (active-passive) transparent mode WAN optimization rules. Traffic shaping is ignored for peer-to-peer WAN optimization and for client/server WAN optimization not operating in transparent mode.

All policies

When selecting a shaper to be for all policies - *For All Policies Using This Shaper* - the FortiGate unit applies the shaping rules to all policies using the same shaper. For example, the shaper is set to be per policy with a maximum bandwidth of 1000 Kb/s. There are four security policies monitoring traffic through the FortiGate unit. All four have the shaper enabled. Each security policy must share the defined 1000 Kb/s, and is set on a first come, first served basis. For example, if policy 1 uses 800 Kb/s, the remaining three must share 200 Kb/s. As policy 1 uses less bandwidth, it is opened up to the other policies to use as required. Once used, any other policies will encounter latency until free bandwidth opens from a policy currently in use.

Maximum and guaranteed bandwidth

The maximum bandwidth instructs the security policy what the largest amount of traffic allowed using the policy. Depending on the service or the users included for the security policy, this number can provide a larger or smaller throughput depending on the priority you set for the shaper.

The *Maximum Bandwidth* can be set to a value of between 1 and 16776000 kbit/s. The Web-Based Manager gives an error if any value outside of this range is used, but in the CLI a value of 0 can be entered. Setting `maximum-bandwidth` to 0 (zero) prevents any traffic from going through the policy.

The guaranteed bandwidth ensures there is a consistent reserved bandwidth available for a given service or user. When setting the guaranteed bandwidth, ensure that the value is significantly less than the bandwidth capacity of the interface, otherwise no other traffic will pass through the interface or very little an potentially causing unwanted latency.

Traffic priority

Select a Traffic Priority of high, medium or low, so the FortiGate unit manages the relative priorities of different types of traffic. For example, a policy for connecting to a secure web server needed to support e-commerce traffic should be assigned a high traffic priority. Less important services should be assigned a low priority. The firewall provides bandwidth to low-priority connections only when bandwidth is not needed for high-priority connections.

Traffic Shaping for FortiOS 5.0

Be sure to enable traffic shaping on all security policies. If you do not apply any traffic shaping rule to a policy, the policy is set to high priority by default. Distribute security policies over all three priority queues.

VLAN, VDOM and virtual interfaces

Policy-based traffic shaping does not use queues directly. It shapes the traffic and if the packet is allowed by the security policy, then a priority is assigned. That priority controls what queue the packet will be put in upon egress. VLANs, VDOMs, aggregate ports and other virtual devices do not have queues and as such, traffic is sent directly to the underlying physical device where it is queued and affected by the physical ports. This is also the case with IPsec connections.

Shared traffic shaper configuration settings

To configure a shared traffic shaper go to *Firewall Objects* > *Traffic Shaper* > *Shared* and select *Create New*.

Name	Enter a name for the traffic shaper.
Apply Shaper	When selecting a shaper to be *Per Policy*, the FortiGate unit will apply the shaping rules defined to each security policy individually. For example, the shaper is set to be per policy with a maximum bandwidth of 1000 Kb/s. There are four security policies monitoring traffic through the FortiGate unit. Three of these have the shaper enabled. Each security policy has the same maximum bandwidth of 1000 Kb/s.
	Per policy traffic shaping is compatible with client/server (active-passive) transparent mode WAN optimization rules. Traffic shaping is ignored for peer-to-peer WAN optimization and for client/server WAN optimization not operating in transparent mode.
	When selecting a shaper to be for all policies - *For All Policies Using This Shaper* - the FortiGate unit applies the shaping rules to all policies using the same shaper. For example, the shaper is set to be per policy with a maximum bandwidth of 1000 Kb/s. There are four security policies monitoring traffic through the FortiGate unit. All four have the shaper enabled. Each security policy must share the defined 1000 Kb/s, and is set on a first come, first served basis. For example, if policy 1 uses 800 Kb/s, the remaining three must share 200 Kb/s. As policy 1 uses less bandwidth, it is opened up to the other policies to use as required. Once used, any other policies will encounter latency until free bandwidth opens from a policy currently in use.
Traffic Priority	Select level of importance *Priority* so the FortiGate unit manages the relative priorities of different types of traffic. For example, a policy for connecting to a secure web server needed to support e-commerce traffic should be assigned a high traffic priority. Less important services should be assigned a low priority.
	If you do not apply any traffic shaping priority, the priority is set to high priority by default.

Maximum Bandwidth	The maximum bandwidth instructs the security policy what the largest amount of traffic allowed using the policy. Depending on the service or the users included for the security policy, this number can provide a larger or smaller throughput depending on the priority you set for the shaper.

Setting *Maximum Bandwidth* to 0 (zero) provides unlimited bandwidth. |
| **Guaranteed Bandwidth** | The guaranteed bandwidth ensures there is a consistent reserved bandwidth available for a given service or user. When setting the guaranteed bandwidth, ensure that the value is significantly less than the bandwidth capacity of the interface, otherwise no other traffic will pass through the interface or very little and potentially causing unwanted latency.

Setting *Guaranteed Bandwidth* to 0 (zero) provides unlimited bandwidth. |
| **DSCP** | Enter the number for the DSCP value. You can use the FortiGate Differentiated Services feature to change the DSCP (Differentiated Services Code Point) value for all packets accepted by a policy. The network can use these DSCP values to classify, mark, shape, and police traffic, and to perform intelligent queuing. DSCP features are applied to traffic by configuring the routers on your network to apply different service levels to packets depending on the DSCP value of the packet. For more information, see "Differentiated Services". |

Example

The following steps creates a Per Policy traffic shaper called "Throughput" with a maximum traffic amount of 720,000 Kb/s, and a guaranteed traffic of 150,000 Kb/s with a high traffic priority.

To create the shared shaper - web-based manager

1. Go to *Firewall Objects > Traffic Shaper > Shared* and select *Create New*.
2. Enter the *Name* Throughput.
3. Select *Per Policy*.
4. Select the *Maximum Bandwidth* check box and enter the value 120000.
5. Select the *Guaranteed Bandwidth* check box and enter the value 150000.
6. Set the *Traffic Priority* to *High*.
7. Select *OK*.

To create the shared shaper - CLI

```
config firewall shaper traffic-shaper
  edit Throughput
    set per-policy enable
    set maximum-bandwidth 720000
    set guaranteed-bandwidth 150000
    set priority high
  end
```

Per-IP shaping

Traffic shaping by IP enables you to apply traffic shaping to all source IP addresses in the security policy. As well as controlling the maximum bandwidth users of a selected policy, you can also define the maximum number of concurrent sessions.

Per-IP traffic shaping enables you limit the behavior of every member of a policy to avoid one user from using all the available bandwidth - it now is shared within a group equally. Using a per-IP shaper avoids having to create multiple policies for every user you want to apply a shaper. Per-IP traffic shaping is not supported over NP2 interfaces.

Per-IP traffic shaping configuration settings

To configure per-IP traffic shaping go to *Firewall Objects > Traffic Shaper > Per-IP*. and select *Create New*.

Name	Enter a name for the per-IP traffic shaper.
Maximum Bandwidth	The maximum bandwidth instructs the security policy what the largest amount of traffic allowed using the policy. Depending on the service or the users included for the security policy, this number can provide a larger or smaller throughput depending on the priority you set for the shaper. Setting *Maximum Bandwidth* to 0 (zero) provides unlimited bandwidth.
Maximum Concurrent Connections	Enter the maximum allowed concurrent connection.
Forward DSCP Reverse DSCP	Enter the number for the DSCP value. You can use the FortiGate Differentiated Services feature to change the DSCP (Differentiated Services Code Point) value for all packets accepted by a policy. The network can use these DSCP values to classify, mark, shape, and police traffic, and to perform intelligent queuing. DSCP features are applied to traffic by configuring the routers on your network to apply different service levels to packets depending on the DSCP value of the packet. For more information, see "Differentiated Services".

Example

The following steps create a Per-IP traffic shaper called "Accounting" with a maximum traffic amount of 720,000 Kb/s, and the number of concurrent sessions of 200.

To create the shared shaper - web-based manager

1. Go to *Firewall Objects > Traffic Shaper > Per-IP*.
2. Select *Create New*.
3. Enter the *Name* `Accounting`.
4. Select the *Maximum Bandwidth* check box and enter the value `720000`.
5. Select the *Maximum Concurrent Sessions* check box and enter the value `200`.
6. Select *OK*.

To create the shared shaper - CLI

```
config firewall shaper per-ip-shaper
  edit Accounting
    set max-bandwidth 720000
    set max-concurrent-sessions 200
end
```

Adding Per-IP traffic shapers to a security policy

Per-IP traffic shaping is supported by IPv6 security policies. You can add any Per-IP traffic shaper to an IPv6 security policy in the CLI.5.020142014

To add a Per-IP traffic shaper to an IPv6 security policy - CLI

```
config firewall policy6
  edit 0
    set per-ip-shaper 'new-perip-shaper'
  end
```

Application control shaping

Traffic shaping is also possible for specific applications for both shared and per IP shaping. Through the *Security Profiles > Application Control* feature, you can configure a specific application's maximum bandwidth. When configuring the application control features, if the application is set to pass, you can set the traffic shaping options. The shapers available are those set up in the *Firewall Objects > Traffic Shaping* menu.

For more information on configuring application control shapers, see the *Security Profiles Guide*.

Example

This example sets the traffic shaping definition for Facebook to a medium priority, a default traffic shaper.

To add traffic shaping for Facebook - web-based manager

1. Go to *Security Profiles > Application Control > Application Sensors*.
2. Select the *Create New* "Plus" icon in the upper right corner of the screen to create a new application group, and enter the name `Web`.
3. Select *OK*.
4. Select *Create New*.
5. Deselect the check box for Category to unselect all categories and select *Social.Networking*.
6. Select *Traffic Shaping* and select medium-priority from the drop-down list.
7. Select *OK*.

To add traffic shaping for Facebook - CLI

```
config application list
  edit web
    config entries
      edit 1
        set category 23
        set application 17735
        set action pass
        set shaper medium-priority
      end
  end
end
```

Enabling in the security policy

All traffic shapers are enabled within a security policy, including the Application Control shapers. As such, the shapers are in effect after any DoS detection policies, and before any routing or packet scanning occurs.

Traffic shaping is also supported for IPv6 policies.

To enable traffic shaping - web-based manager

1. Go to *Policy > Policy > Policy*.
2. Select *Create New* or select an existing policy and select *Edit*.
3. Select *Traffic Shaping*.
4. Select the shaping option and select the shaper from the drop-down list.
5. Select *OK*.

Shapers applied in the security policy affect outbound or traffic to a destination. To affect inbound, or download, traffic, select *Shared Traffic Shaper Reverse Direction*. For more information, see .

To enable traffic shaping - CLI

```
config firewall policy
  edit <policy_number>
    ...
    set traffic-shaper <shaper_name>
    set per-ip-shaper <shaper_name>
  end
```

Reverse direction traffic shaping

The shaper you select for the security policy (shared shaper) will affect the traffic in the direction defined in the policy. For example, if the source port is port 1 and the destination is port 3, the shaping affects the flow in this direction only, that is, the upload or outbound direction. By selecting *Shared Traffic Shaper Reverse Direction*, you can define the traffic shaper for the policy in the opposite direction, that is, the download or inbound direction. In this example, from port 3 to port 1.

To add a reverse shaper

1. Go to *Policy > Policy > Policy*.

Traffic Shaping for FortiOS 5.0

2. Select *Traffic Shaping*.

3. Select *Shared Traffic Shaper Reverse Direction* and select the shaper from the list.

4. Select *OK*.

Setting the reverse direction only

There may be instances where you only need to have the traffic shaping for incoming connections. That is, the "reverse" direction to the typical traffic shaper.

To add a reverse shaper - web-based manager

1. Go to *Policy > Policy > Policy*.

2. Select *Traffic Shaping*.

3. Select *Shared Traffic Shaper Reverse Direction* and select the shaper from the list.

4. Select *OK*.

To configure a reverse-only shaper - CLI

```
config firewall policy
  edit <policy_number>
     ...
     set traffic-shaper-reverse <shaper_name>
end
```

Application control shaper

Application control shapers are in effect within the application control profile. Within the security policy options, select *Application Control* and select the application from the list.

Type of Service priority

Type of service (ToS) is an 8-bit field in the IP header that enables you to determine how the IP datagram should be delivered, using criteria of Delay, Throughput, Priority, Reliability, and Cost. Each quality helps gateways determine the best way to route datagrams. A router maintains a ToS value for each route in its routing table. The lowest priority ToS is 0, the highest is 7 when bits 3, 4, and 5 are all set to 1. There are 4 other bits that are seldom used or reserved that are not included here.

Together these bits are the tos variable of the `tos-based-priority` command. The router tries to match the ToS of the datagram to the ToS on one of the possible routes to the destination. If there is no match, the datagram is sent over a zero ToS route. Using increased quality may increase the cost of delivery because better performance may consume limited network resources.

Each bit represents the priority as per RFC 1349:

• 1000 - minimize delay

• 0100 - maximize throughput

• 0010 - maximize reliability

• 0001 - minimize monetary cost

The TOS value is set in the CLI using the commands:

```
config system tos-based-priority
  edit <sequence_number>
    set tos [0-15]
    set priority [high | medium | low]
end
```

Where `tos` is the value of the type of service bit in the IP datagram header with a value between 0 and 15, and `priority` is the priority of this type of service priority. These priority levels conform to the firewall traffic shaping priorities, as defined in RFC 1349.

For example, if you want to configure the FortiGate unit so that reliability is the first priority, set the tos value to 4.

```
config system tos-based-priority
  edit 1
    set tos 4
    set priority high
end
```

For a list of ToS values and their DSCP equivalents see "Tos and DSCP mapping" on page 30.

Example

```
config system tos-based-priority
  edit 1
    set tos 1
    set priority low
  next
  edit 4
    set tos 4
    set priority medium
  next
  edit 6
    set tos 6
    set priority high
  next
end
```

TOS in FortiOS

Traffic shaping and TOS follow the following sequence:

- The CLI command `tos-based-priority` acts as a `tos-to-priority` mapping. FortiOS maps the TOS to a priority when it receives a packet.
- Traffic shaping settings adjust the packet's priority according the traffic.
- Deliver the packet based on its priority.

Differentiated Services

Differentiated Services describes a set of end-to-end Quality of Service (QoS) capabilities. End-to-end QoS is the ability of a network to deliver service required by specific network traffic from one end of the network to another. By configuring differentiated services, you configure

your network to deliver particular levels of service for different packets based on the QoS specified by each packet.

Differentiated Services (also called DiffServ) is defined by RFC 2474 and 2475 as enhancements to IP networking to enable scalable service discrimination in the IP network without the need for per-flow state and signaling at every hop. Routers that can understand differentiated services sort IP traffic into classes by inspecting the DS field in IPv4 header or the Traffic Class field in the IPv6 header.

You can use the FortiGate Differentiated Services feature to change the DSCP (Differentiated Services Code Point) value for all packets accepted by a policy. The network can use these DSCP values to classify, mark, shape, and police traffic, and to perform intelligent queuing. DSCP features are applied to traffic by configuring the routers on your network to apply different service levels to packets depending on the DSCP value of the packet.

If the differentiated services feature is not enabled, the FortiGate unit treats traffic as if the DSCP value is set to the default (00), and will not change IP packets' DSCP field. DSCP values are also not applied to traffic if the traffic originates from a FortiGate unit itself.

The FortiGate unit applies the DSCP value and IPsec encryption to the differentiated services (formerly TOS) field in the first word of the IP header. The typical first word of an IP header, with the default DSCP value, is 4500:

- 4 for IPv4
- 5 for a length of five words
- 00 for the default DSCP value

You can change the packet's DSCP field for traffic initiating a session (forward) or for reply traffic (reverse) and enable each direction separately and configure it in the security policy.

Changes to DSCP values in a security policy effect new sessions. If traffic must use the new DSCP values immediately, clear all existing sessions.

DSCP is enabled using the CLI command:

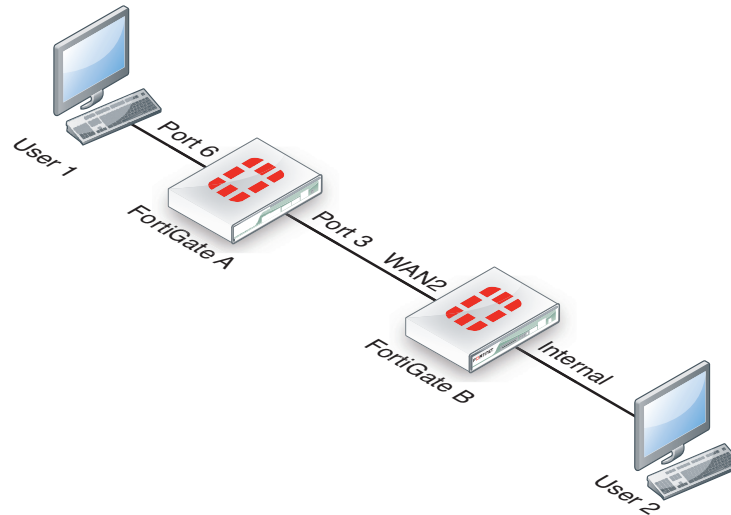```
config firewall policy
  edit <policy_number>
    ...
    set diffserv-forward enable
    set diffservcode-forward <binary_integer>
    set diffserv-reverse enable
    set diffservcode-rev <binary_integer>
end
```

For more information on the different DCSP commands, see the examples below and the *CLI Reference*. If you only set `diffserv-forward` and `diffserv-reverse` without setting the corresponding `diffvercode` values, the FortiGate unit will reset the bits to zero.

For a list of DSCP values and their ToS equivalents see "Tos and DSCP mapping" on page 30. DSCP values can also be defined within a shared shaper as a single value, and per-IP shaper for forward and reverse directions.

## DSCP examples

For all the following DSCP examples, the FortiGate and client PC configuration is the following diagram and used firewall-based DSCP configurations.



### Example

In this example, an ICMP ping is executed between User 1and FortiGate B, through a FortiGate unit. DSCP is disabled on FortiGate B, and FortiGate A contains the following configuration:

```
config firewall policy
  edit 2
     set srcintf port6
     set dstintf port3
     set src addr all
     set dstaddr all
     set action accept
     set schedule always
     set service ANY
     set diffserv-forward enable
     set diffservcode-forward 101110
  end
```
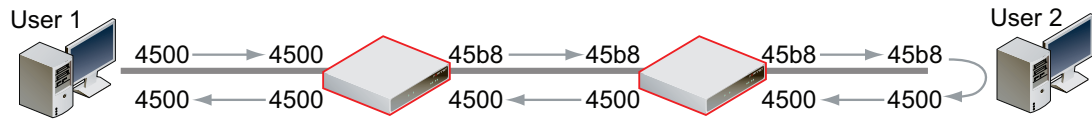
As a result, FortiGate A changes the DSCP field for outgoing traffic, but not to its reply traffic. The binary DSCP values used map to the following hexadecimal

TOS field values, which are observable by a sniffer (also known as a packet tracer):

- DSCP 000000 is TOS field 0x00

- DSCP 101110 is TOS field 0xb8, the recommended DSCP value for expedited forwarding (EF)

If you performed an ICMP ping between User 1 and User 2, the following output illustrates the IP headers for the request and the reply by sniffers on each of FortiGate unit's network

interfaces. The right-most two digits of each IP header are the TOS field, which contains the DSCP value.
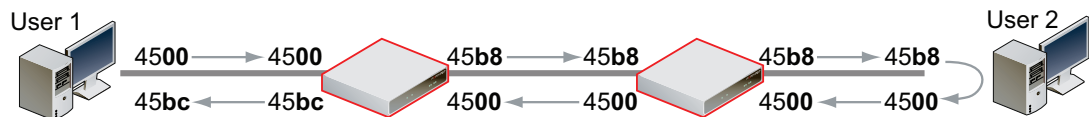


## Example

In this example, an ICMP ping is executed between User 1 and FortiGate B, through FortiGate A. DSCP is disabled on FortiGate B, and FortiGate A contains the following configuration:

```
config firewall policy
    edit 2
        set srcintf port6
        set dstintf port3
        set src addr all
        set dstaddr all
        set action accept
        set schedule always
        set service ANY"
        set diffserv-forward enable
        set diffserv-rev enable
        set diffservcode-forward 101110
        set diffservcode-rev 101111
    end
```

As a result, FortiGate A changes the DSCP field for both outgoing traffic and its reply traffic. The binary DSCP values in map to the following hexadecimal TOS field values, which are observable by a sniffer (also known as a packet tracer):

• DSCP 000000 is TOS field 0x**00**

• DSCP 101110 is TOS field 0x**b8**, the recommended DSCP value for expedited forwarding (EF)

• DSCP 101111 is TOS field 0x**bc**

If you performed an ICMP ping between User 1 and User 2, the output below illustrates the IP headers observed for the request and the reply by sniffers on each of FortiGate A's and FortiGate B's network interfaces. The right-most two digits of each IP header are the TOS field, which contains the DSCP value.

Example

In this example, an ICMP ping is executed between User 1 and FortiGate B, through FortiGate A. DSCP is enabled for both traffic directions on FortiGate A, and enabled only for reply traffic on FortiGate B. FortiGate A contains the following configuration:

```
config firewall policy
   edit 2
      set srcintf port6
      set dstintf port3
      set src addr all
      set dstaddr all
      set action accept
      set schedule always
      set service ANY
      set diffserv-forward enable
      set diffserv-rev enable
      set diffservcode-forward 101110
      set diffservcode-rev 101111
   end
```
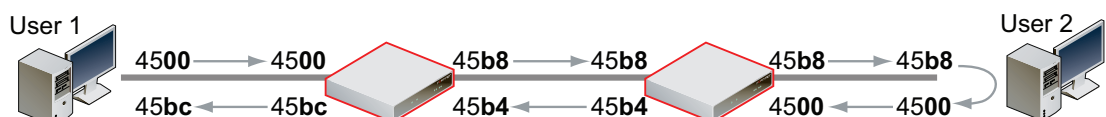
FortiGate B contains the following configuration:

```
config firewall policy
   edit 2
      set srcintf wan2
      set dstintf internal
      set src addr all
      set dstaddr all
      set action accept
      set schedule always
      set service ANY
      set diffserv-rev enable
      set diffservcode-rev 101101
   end
```

As a result, FortiGate A changes the DSCP field for both outgoing traffic and its reply traffic, and FortiGate B changes the DSCP field only for reply traffic. The binary DSCP values in this configuration map to the following hexadecimal TOS field values:

- DSCP 000000 is TOS field 0**x00**
- DSCP 101101 is TOS field 0**xb4**
- DSCP 101110 is TOS field 0**xb8**, the recommended DSCP value for expedited forwarding (EF)
- DSCP 101111 is TOS field 0**xbc**

If you performed an ICMP ping between User 1 and User 2, the output below illustrates the IP headers observed for the request and the reply by sniffers on each of FortiGate A's and FortiGate B's network interfaces. The right-most two digits of each IP header are the TOS field, which contains the DSCP value.

Traffic Shaping for FortiOS 5.0

Example

In this example, HTTPS and DNS traffic is sent from User 1 to FortiGate B, through FortiGate A. DSCP is enabled for both traffic directions on FortiGate A, and enabled only for reply traffic on FortiGate B. FortiGate A contains the following configuration:

```
config firewall policy
  edit 2
      set srcintf port6
      set dstintf port3
      set src addr all
      set dstaddr all
      set action accept
      set schedule always
      set service ANY
      set diffserv-forward enable
      set diffserv-rev enable
      set diffservcode-forward 101110
      set diffservcode-rev 101111
  end
```

FortiGate B contains the following configuration:

```
config firewall policy
  edit 2
      set srcintf wan2
      set dstintf internal
      set src addr all
      set dstaddr all
      set action accept
      set schedule always
      set service ANY
      set diffserv-rev enable
      set diffservcode-rev 101101
  end
```

As a result, FortiGate A changes the DSCP field for both outgoing traffic and its reply traffic, but FortiGate B changes the DSCP field only for reply traffic which passes through its internal interface. Since the example traffic does not pass through the internal interface, FortiGate B does not mark the packets. The binary DSCP values in this configuration map to the following hexadecimal TOS field values:

- DSCP 000000 is TOS field 0**x00**

- DSCP 101101 is TOS field 0**xb4**, which is configured on FortiGate B but not observed by the sniffer because the example traffic originates from the FortiGate unit itself, and therefore does not match that security policy.

- DSCP 101110 is TOS field 0**xb8**, the recommended DSCP value for expedited forwarding (EF)

- DSCP 101111 is TOS field 0**xbc**

If you sent HTTPS or DNS traffic from User 1 to FortiGate B, the following would illustrate the IP headers observed for the request and the reply by sniffers on each of FortiGate A's and

FortiGate B's network interfaces. The right-most two digits of each IP header are the TOS field, which contains the DSCP value.
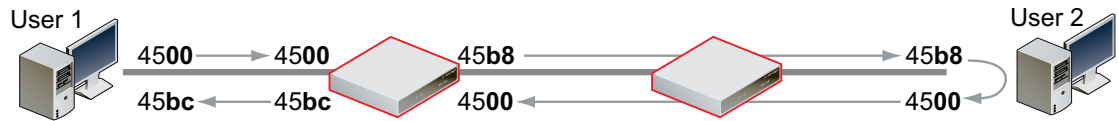
Tos and DSCP mapping

The table below lists the mapping of DSCP and ToS hexadecimal values for QoS.

Table 2: ToS to DSCP mappings

Service Class	DSCP Bits	DSCP Value	ToS Value	ToS Hexidecimal
Network Control	111000	56-63	224	0xE0
Internetwork Control	110000	48-55	192	0xC0
Critical - Voice Data (RTP)	101110	46	184	0xB8
	101000	40	160	0xA0
Flash Override Video Data	100010	34	136	0x88
	100100	36	144	0x90
	100110	38	152	0x98
	100000	32	128	0x80
Flash Voice Control	011010	26	104	0x68
	011100	28	112	0x70
	011110	30	120	0x78
	011000	24	96	0x60
Immediate Deterministic (SNA)	010010	18	72	0x48
	010100	20	80	0x50
	010110	22	88	0x58
	010000	16	64	0x40
Priority Controlled Load	001010	10	40	0x28
	001100	12	48	0x30
	001110	14	56	0x38
	001000	8	32	0x20
Routine - Best Effort	000000	0	0	0x00
Routine - Penalty Box	000010	2	8	0x08

Traffic Shaper Monitor

You can view statistical information about traffic shapers and their bandwidth from *Firewall Objects > Monitor > Traffic Shaper Monitor*.

Refresh	Select to refresh the information on the page.
Reset	Select to reset the information to clear the current information from the page. New information is included on the page.
Report By	Select to display dropped packets or current bandwidth. The chart changes its name so you know what current information is displayed.
Traffic Shaper Usage Dropped packets	The bar chart displays the packets that were dropped by traffic shaper.
Traffic Shaper Usage Current Bandwidth	The bar chart displays the current bandwidth of traffic shapers.

Examples

While it is possible to configure QoS using a combination of security policies and in ToS-based priorities, and to distribute traffic over all six of the possible queues for each physical interface, the results of those configurations can be more difficult to analyze due to their complexity. In those cases, prioritization behavior can vary by several factors, including traffic volume, ToS (type of service) or differentiated services markings, and correlation of session to a security policy.

The following simple examples illustrate QoS configurations using either prioritization by security policy, or prioritization by ToS bit, but not both. The examples also assume you are not configuring traffic shaping for interfaces that receive hardware acceleration from network processing units (NPU).

QoS using priority from security policies

Configurations implementing QoS using the priority values defined in security policies are capable of applying bandwidth limits and guarantees.

In addition to configuring traffic shaping, you may also choose to limit bandwidth accepted by each interface. This can be useful in scenarios where bandwidth being received on source interfaces frequently exceeds the maximum bandwidth limit defined in the security policy. In this case, rather than wasting processing power on packets that will only be dropped later in the processing to enforce those limits, you may choose to preemptively police the traffic.

Note that if you implement QoS using security policies rather than ToS bit, the FortiGate unit applies QoS to all packets controlled by the policy. Control is less granular than prioritization by ToS bit, but has the benefits of correlating quality of service to a security policy, enabling you to distribute traffic over up to four of the possible 6 priority queues (queue 0 to queue 3), not requiring other devices in your network to set or respect the ToS bit, and of enabling you to configure bandwidth limits and guarantees.

In this example, we limit the bandwidth accepted by each source interface, limit the bandwidth used by sessions controlled by the security policy, and then configure prioritized queuing on the destination interface based upon the priority in the security policy, subject to alternative assignment to queue 0 when necessary to achieve the guaranteed packet rate.

To limit bandwidth accepted by an interface

In the CLI, enter the following commands:

```
config system interface
  edit <name_str>
    set inbandwidth <rate_int>
  next
end
```

where `<rate_int>` is the bandwidth limit in Kb/s. Excess packets will be dropped.

To configure bandwidth guarantees, limits, and priorities

1. Go to *Firewall Objects > Traffic Shaper > Shared*, and select *Create New*.
2. Enter a name for the shaper.

3. Enter the *Guaranteed Bandwidth*, if any.

   Bandwidth guarantees affect prioritization. While packet rates are less than this rate, they use priority queue 0. If this is not the effect you intend, consider entering a small guaranteed rate, or enter 0 to effectively disable bandwidth guarantees.

4. Enter a *Maximum Bandwidth*.

   Packets greater than this rate will be discarded.

5. Select the *Traffic Priority*.

   High has a priority value of 1, while Low is 3. While the current packet rate is below Guaranteed Bandwidth, the FortiGate unit will disregard this setting, and instead use priority queue 0.

6. Select *OK*.

Sample configuration

This sample configuration limits ingressing bandwidth to 500 Kb/s. It also applies separate traffic shapers to FTP and HTTP traffic. In addition to the interface bandwidth limit, HTTP traffic is subject to a security policy bandwidth limit of 200 Kb/s.

All egressing FTP traffic greater than 10 Kb/s is subject to a low priority queue (queue 3), while all egressing HTTP traffic greater than 100 Kb/s is subject to a medium priority queue (queue 2). That is, unless FTP traffic rates are lower than their guaranteed rate, and web traffic rates are greater than their guaranteed rate, FTP traffic is lower priority than web traffic.

Traffic less than these guaranteed bandwidth rates use the highest priority queue (queue 0).

Set the inbandwidth limits. This setting is only available in the CLI:

```
config system interface
  edit wan1
    set inbandwidth 500
  next
end
```

Create the traffic shapers or FTP and HTTP.

To configure the shapers - web-based manager

1. Go to *Firewall Objects > Traffic Shaper > Shared*, and select *Create New*.
2. Enter FTP for the name of the shaper.
3. Enter the *Guaranteed Bandwidth*, of 10 Kbps.
4. Enter a *Maximum Bandwidth* of 500 Kbps.
5. Select the *Traffic Priority* of *Low*.
6. Select *OK*.
7. Select *Create New*.
8. Enter HTTP for the name of the shaper.
9. Enter the *Guaranteed Bandwidth*, of 100 Kbps.
10. Enter a *Maximum Bandwidth* of 200 Kbps.
11. Select the *Traffic Priority of Medium*.
12. Select *OK*.

Traffic Shaping for FortiOS 5.0

To configure the shapers - CLI

```
config firewall shaper traffic-shaper
  edit FTP
    set maximum-bandwidth 500
    set guaranteed-bandwidth 10
    set per-policy enable
    set priority low
    end
  next
  edit HTTP
    set maximum-bandwidth 200
    set guaranteed-bandwidth 100
    set per-policy enable
    set priority medium
end
```

QoS using priority from ToS or differentiated services

Configurations implementing QoS using the priority values defined in either global or specific ToS bit values are not capable of applying bandwidth limits and guarantees, but are capable of prioritizing traffic at per-packet levels, rather than uniformly to all services matched by the security policy.

In addition to configuring traffic prioritization, you may also choose to limit bandwidth being received by each interface. This can sometimes be useful in scenarios where you want to limit traffic levels, but do not want to configure traffic shaping within a security policy. This has the benefit of policing traffic at a point before the FortiGate unit performs most processing.

Note that if you implement QoS using ToS octet rather than security policies, the FortiGate unit applies QoS on a packet by packet basis, and priorities may be different for packets and services controlled by the same security policy. This is more granular control than prioritization by security policies, but has the drawbacks that quality of service is may not be uniform for multiple services controlled by the same security policy, packets will only use up to three of the six possible queues (queue 0 to queue 2), and bandwidth cannot be guaranteed. Other devices in your network must also be able to set or preserve ToS bits.

In this example, we limit the bandwidth accepted by each source interface, and then configure prioritized queuing on the destination interface based upon the value of the ToS bit located in the IP header of each accepted packet.

To limit bandwidth accepted by an interface, in the CLI, enter the following commands:

```
config system interface
  edit <name_str>
    set inbandwidth <rate_int>
  next
end
```

where `<rate_int>` is the bandwidth limit in Kb/s. Excess packets will be dropped.

To configure priorities, in the CLI, configure the global priority value using the following commands:

```
config system global
    set tos-based-priority {high | low | medium}
        end
```

where `high` has a priority value of 0 and `low` is 2.

If you want to prioritize some ToS bit values differently than the global ToS-based priority, configure the priority for packets with that ToS bit value using the following commands:

```
config system tos-based-priority
    edit <id_int>
        set tos [0-15]
        set priority {high | low | medium}
    next
end
```

where and `tos` is the value of the ToS bit in the packet's IP header, and `high` has a priority value of 0 and `low` is 2. Priority values configured in this location will override the global ToS-based priority.

Sample configuration

This sample configuration limits ingressing bandwidth to 500 Kb/s. It also queues egressing traffic based upon the ToS bit in the IP header of ingressing packets.

Unless specified for the packet's ToS bit value, packets use the low priority queue (queue 2). For ToS bit values 4 and 15, the priorities are specified as medium (value 1) and high (value 0), respectively.

```
config system interface
    edit wan1
        set inbandwidth 500
    next
end
config system global
    set tos-based-priority low
end
config system tos-based-priority
    edit 4
        set tos 4
        set priority medium
    next
    edit 15
        set tos 15
        set priority high
    next
end
```

Example setup for VoIP

In this example, there are three traffic shaping requirements for a network:

- Voice over IP (VoIP) requires a guaranteed, high-priority for bandwidth for telephone communications.
- FTP bursts must be contained so as not to consume any available bandwidth. As such this traffic needs to be throttled to a smaller amount.
- A consistent bandwidth requirement is needed for all other email and web-based traffic.

To enable this requirement, you need to create three separate shapers and three security policies for each traffic type.

For this example, the actual values are not actual values, they are used for the simplicity of the example.

Creating the traffic shapers

First create the traffic shapers that define the maximum and guaranteed bandwidth. The shared shapers will be used, some with per-policy and some all policies as shown in the table, to better control traffic.

VoIP shaper

The VoIP functionary is a key component to the business as a communication tool and as such requires a guaranteed bandwidth.

To create a VoIP shaper - web-based manager

1. Go to *Firewall Objects > Traffic Shaping > Shared*.
2. Enter the *Name* `voip`.
3. Select *Per Policy*.
4. Enter the *Maximum Bandwidth* of `1000` Kb/s
5. Enter the *Guaranteed Bandwidth* of `800` Kb/s.
6. Select a *Traffic Priority* of *High*.
7. Select *OK*.

To create a VoIP shaper - CLI

```
config firewall shaper traffic-shaper
  edit voip
    set maximum-bandwidth 1000
    set guaranteed-bandwidth 800
    set per-policy enable
    set priority high
  end
```

This ensures that whatever number of policies use this shaper, the defined bandwidth will always be the same. At the same time, the bandwidth is continually guaranteed at 800 Kb/s but if available can be as much as 1000 Kb/s. Setting the priority to high ensures that the FortiGate unit always considers VoIP traffic as the most important.

FTP shaper

The FTP shaper sets the maximum bandwidth to use to avoid sudden spikes by sudden uploading or downloading of large files, and interfering with other more important traffic.

To create a FTP shaper - web-based manager

1. Go to *Firewall Objects > Traffic Shaping > Shared*.

2. Enter the *Name* `ftp`.

3. Select *For all Policies Using This Shaper*.

4. Enter the *Maximum Bandwidth* of `200` Kb/s

5. Enter the *Guaranteed Bandwidth* of `200` Kb/s.

6. Select a *Traffic Priority* of *Low*.

7. Select *OK*.

To create a FTP shaper - CLI

```
config firewall shaper traffic-shaper
  edit ftp
    set maximum-bandwidth 200
    set guaranteed-bandwidth 200
    set priority low
end
```

For this shaper, the maximum and guaranteed bandwidth are set low and to the same value. In this case, the bandwidth is restricted to a specific amount. By also setting the traffic priority low ensures more important traffic will be able to pass before FTP traffic.

Regular traffic shaper

The regular shaper sets the maximum bandwidth and guaranteed bandwidth for everyday business traffic such as web and email traffic.

To create a regular shaper - web-based manager

1. Go to *Firewall Objects > Traffic Shaping > Shared*.

2. Enter the *Name* `daily_traffic`.

3. Select *Per Policy*.

4. Enter the *Maximum Bandwidth* of `600` Kb/s

5. Enter the *Guaranteed Bandwidth* of `600` Kb/s.

6. Select a *Traffic Priority* of *Medium*.

7. Select *OK*.

To create a regular shaper - CLI

```
config firewall shaper traffic-shaper
  edit daily_traffic
    set maximum-bandwidth 600
    set guaranteed-bandwidth 600
    set per-policy enable
    set priority medium
end
```

For this shaper, the maximum and guaranteed bandwidth are set to a moderate value of 600 Kb/s. It is also set for per policy, which ensures each security policy for day-to-day business traffic has the same distribution of bandwidth.

Creating security policies

To employ the shaper, create security policies that use the shapers within the policies. Create a separate policy for each service and enable traffic shaping. For example, a policy for FTP traffic, a policy for SIP and so on.

For the following steps the VoIP traffic shaper is enabled as well as the reverse direction option. This ensures that return traffic for a VoIP call has the same guaranteed bandwidth as the outgoing call.

To enable traffic shaping in the security policy - web-based manager

1. Go to *Policy > Policy > Policy* and select *Create New*.
2. Leave the *Policy Type* as *Firewall* and leave the *Policy Subtype* as *Address*.
3. Enter the following:

Incoming interface	Internal
Source address	All
Outgoing interface	WAN1
Destination address	All
Schedule	always
Service	SIP
Action	ACCEPT

4. Select *Traffic Shaping*.
5. From the drop-down list, select the voip shaper created in the previous steps.
6. Select *Reverse Direction Traffic Shaping*.
7. Select *OK*.

To enable traffic shaping in the security policy - CLI

```
config firewall policy
  edit 6
    set srcintf internal
    set scraddr all
    set dstintf wan1
    set dstaddr all
    set action accept
    set schedule always
    set service sip
    set traffic-shaper voip
    set reverse-traffic-shaper voip
  end
```

Troubleshooting traffic shaping

This chapter outlines some troubleshooting tips and steps to diagnose the shapers and whether they are working correctly. These diagnose commands include:

- `diagnose system tos-based-priority`
- `diagnose firewall shaper traffic-shaper`
- `diagnose firewall per-ip-shaper`
- `diagnose debug flow`

Interface diagnosis

To optimize traffic shaping performance, first ensure that the network interface's Ethernet statistics are clean of errors, collisions, or buffer overruns. To check the interface, enter the following diagnose command to see the traffic statistics:

```
diagnose hardware deviceinfo nic <port_name>
```

Shaper diagnose commands

There are specific diagnose commands you can use to verify the configuration and flow of traffic, including packet loss due to the employed shaper.

All of these diagnose troubleshooting commands are supported in both IPv4 and IPv6.

TOS command

Use the following command to list command to view information of the TOS lists and traffic.

```
diagnose system tos-based-priority
```

This example displays the priority value currently correlated with each possible TOS bit value. Priority values are displayed in order of their corresponding TOS bit values, which can range between 0 and 15, from lowest TOS bit value to highest.

For example, if you have not configured TOS-based priorities, the following appears...

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

...reflecting that all packets are currently using the same default priority, high (value 0).

If you have configured a TOS-based priority of `low` (value 2) for packets with a ToS bit value of 3, the following appears...

```
0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0
```

...reflecting that most packets are using the default priority value, except those with a ToS bit value of 3.

Shared shaper

To view information for the shared traffic shaper for security policies enter the command

```
diagnose firewall shaper traffic-shaper list
```

The resultant output displays the information on all available shapers. The more shapers available the longer the list. For example:

```
name Throughput
maximum-bandwidth 1200000 Kb/sec
guaranteed-bandwidth 50000 Kb/sec
current-bandwidth 0 B/sec
priority 1
packets dropped 0
```

Additional commands include:

`diagnose firewall shaper traffic-shaper state` - provides the total number of traffic shapers on the FortiGate unit.

`diagnose firewall shaper traffic-shaper stats` - provides summary statistics on the shapers. Sample output looks like the following:

```
shapers 9 ipv4 0 ipv6 0 drops 0
```

Per-IP shaper

To view information for the per-IP shaper for security policies enter the command

```
diagnose firewall shaper per-ip-shaper list
```

The resultant output displays the information on all available per-IP shapers. The more shapers available the longer the list. For example:

```
name accounting_group
maximum-bandwidth 200000 Kb/sec
maximum-concurrent-session 55
packet dropped 0
```

Additional commands include:

`diagnose firewall shaper per-ip-shaper state` - provides the total number of per-ip shapers on the FortiGate unit.

`diagnose firewall shaper per-ip-shaper stats` - provides summary statistics on the shapers. Sample output looks like the following:

```
memory allocated 3 packet dropped: 0
```

You can also clear the per-ip statistical data to begin a fresh diagnoses using:

```
diagnose firewall shaper per-ip-shaper clear
```

Packet loss with statistics on shapers

For each shaper there are counters that allow to verify if packets have been discarded. To view this information, in the CLI, enter the command diagnose firewall shaper. The results will look similar to the following output:

```
diagnose firewall shaper traffic-shaper list
name limit_GB_25_MB_50_LQ
maximum-bandwidth 50 Kb/sec
```

```
guaranteed-bandwidth 25 Kb/sec
current-bandwidth 51 Kb/sec
priority 3
dropped 1291985
```

The diagnose command output is different if the shapers are configured either per-policy or shared between policies.

For per-IP the output would be:

```
diagnose firewall shaper per-ip-shaper list

name accounting_group
maximum-bandwidth 200000 Kb/sec
maximum-concurrent-session 55
packet dropped 3264220
```

Packet lost with the debug flow

When using the debug flow diagnostic command, there is a specific message information that a packet has exceed the shaper limits and therefor discarded:

```
diagnose debug flow show console enable
diagnose debug flow filter addr 10.143.0.5
diagnose debug flow trace start 1000


id=20085 trace_id=11 msg="vd-root received a packet(proto=17,
    10.141.0.11:3735->10.143.0.5:5001) from port5."
id=20085 trace_id=11 msg="Find an existing session, id-0000eabc,
    original direction"
id=20085 trace_id=11 msg="exceeded shaper limit, drop"
```

Session list details with dual traffic shaper

When a Security Policy has a different traffic shaper for each direction, it is reflected in the session list output from the CLI:

```
diagnose system session list

session info: proto=6 proto_state=02 expire=115 timeout=3600
    flags=00000000 sock
flag=00000000 sockport=0 av_idx=0 use=4
origin-shaper=Limit_25Mbps prio=1 guarantee 25600/sec max 204800/sec
    traffic 48/sec
reply-shaper=Limit_100Mbps prio=1 guarantee 102400/sec max 204800/sec
    traffic 0/sec
ha_id=0 hakey=44020
policy_dir=0 tunnel=/
state=may_dirty rem os rs
statistic(bits/packets/allow_err): org=96/2/1 reply=0/0/0 tuples=2
orgin->sink: org pre->post, reply pre->post dev=2->3/3->2
    gwy=10.160.0.1/0.0.0.0
```

```
hook=pre dir=org act=dnat
    192.168.171.243:2538->192.168.182.110:80(10.160.0.1:80)
hook=post dir=reply act=snat
    10.160.0.1:80->192.168.171.243:2538(192.168.182.110:80)
pos/(before,after) 0/(0,0), 0/(0,0)
misc=0 policy_id=2 auth_info=0 chk_client_info=0 vd=0 serial=00011e81
    tos=ff/ff app=0 dd_type=0 dd_rule_id=0
```

Additional Information

- Packets discarded by the shaper impact flow-control mechanisms like TCP. For more accurate testing results prefer UDP protocol.

- Traffic shaping accuracy is optimum for security policies without a protection profile where no FortiGate content inspection is processed.

- Do not oversubscribe an outbandwith throughput. For example, sum[guaranteed BW] < outbandwith. For accuracy in bandwidth calculation, it is required to set the "outbandwidth" parameter on the interfaces. For more information see "Bandwidth guarantee, limit, and priority interactions" on page 9.

- The FortiGate unit is not prioritizing traffic based on the DSCP marking configured in the security policy. However, TOS based prioritizing can be made at ingress. For more information see "Differentiated Services" on page 24.

Index