

用于限制带宽的通信管制与通信整形的比较

目录

[简介](#)

[开始使用前](#)

[规则](#)

[先决条件](#)

[使用的组件](#)

[管制与整形](#)

[选择标准](#)

[令牌刷新速率](#)

[流量整形](#)

[流量策略](#)

[最小带宽与最大带宽控制](#)

[相关信息](#)

简介

本文档阐释整形和策略之间的功能差异，它们都可以限制输出速率。虽然两种机制将令牌桶用作数据流计量仪，测量数据包速率，但它们具有重要的功能区别。（[什么是令牌桶？](#)中描述了令牌桶。）

开始使用前

规则

有关文档规则的详细信息，请参阅 [Cisco 技术提示规则](#)。

先决条件

本文档没有任何特定的前提条件。

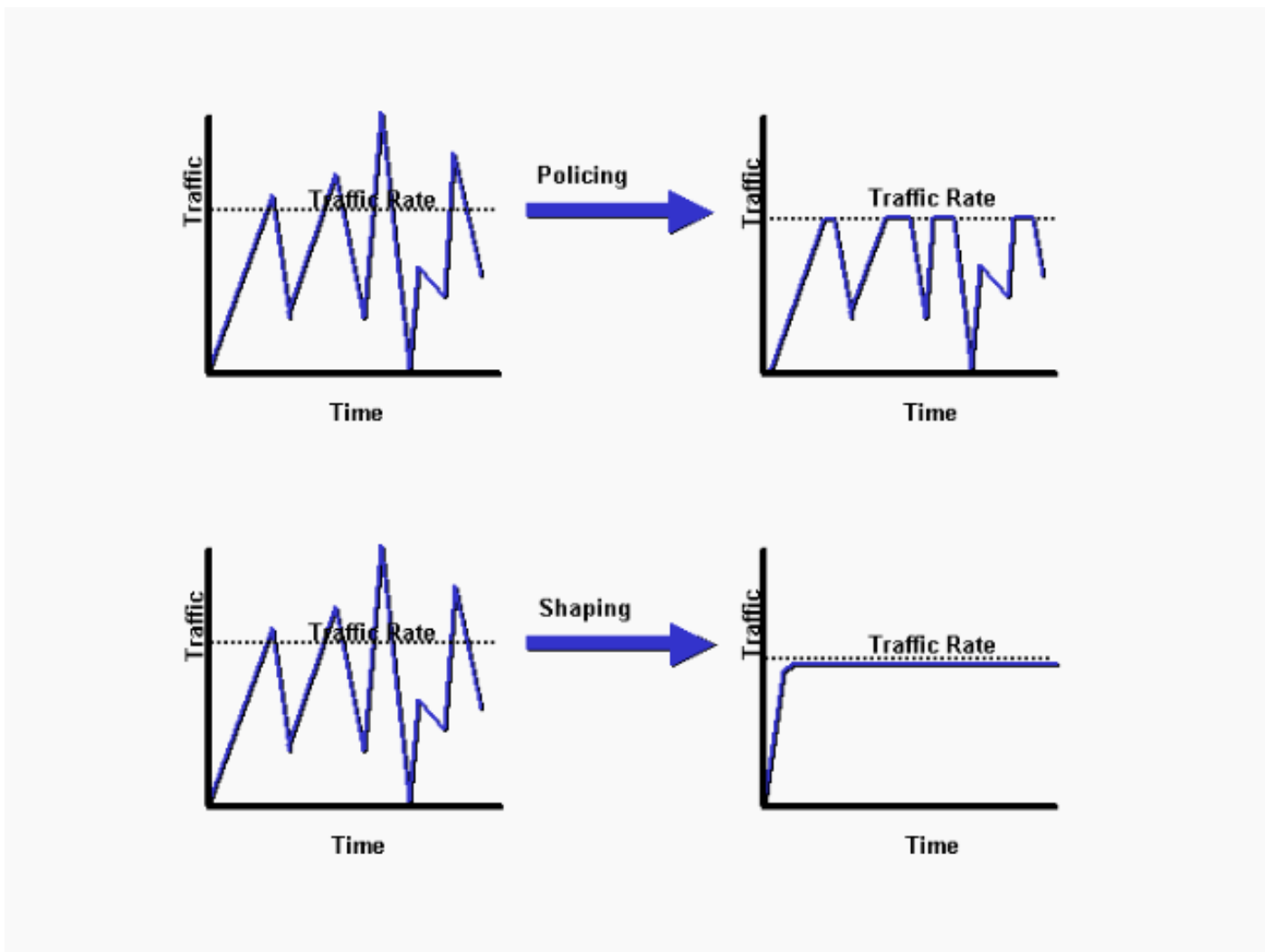
使用的组件

本文档不限于特定的软件和硬件版本。

本文档中的信息都是基于特定实验室环境中的设备创建的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您是在真实网络上操作，请确保您在使用任何命令前已经了解其潜在影响。

管制与整形

下图揭示了两者之间的关键区别。流量策略会传播突发流量。当流量速率达到所配置的最大速率时，将丢弃（或重新标记）超额流量。结果显示为带有波峰和波谷的锯齿形输出速率。与策略相比，流量整形在队列中保留超额的数据包，然后安排超额数据包在以后增加的时间里进行传输。流量整形的结果是一个平滑的数据包输出速率。



整形暗示队列和用于缓冲延迟数据包的充足内存的存在，而策略却不是。排队是一个出站概念；传出接口的数据包将进行排队并且可以整形。对接口上的入站流量只能实施策略。启用整形时，请确保有足够内存。此外，整形要求具有一个针对任何延迟数据包的后期传输的计划功能。该计划功能允许您将整形队列安排到不同队列中。日程安排功能示例是基于类的加权公平排队(CBWFQ)和低延迟排队(LLQ)。

选择标准

下表列出了整形和策略之间的区别，帮助您选择最佳解决方案。

	整形	修正
目标	对高出承诺速率的超额数据包进行缓冲和排队。	丢弃（或重新标记）高出承诺速率的超额数据包。不进行缓冲。*
令牌刷	在时间间隔的开始处递增。（需要最小间隔数。）	持续基于公式： $1 /$

新速率		
令牌值	按比特/秒配置。	按字节配置。
配置选项	<ul style="list-style-type: none"> • 服务质量模块命令行界面 (MQC) 中的 shape 命令可实现基于类的整形。 • frame-relay traffic-shape 命令实现帧中继流量整形 (FRTS)。 • traffic-shape 命令实现通用流量整形 (GTS)。 	<ul style="list-style-type: none"> • MQC 中的 police 命令可实现基于类的策略。 • rate-limit 命令实现承诺接入速率 (CAR)。
适用于入站	否	是
适用于出站	是	是
突发流量	通过在至少八个时间间隔上保持平滑的输出速率来控制突发流量。使用漏桶来延迟流量以达到平滑效果。	传播突发流量。不进行平滑处理。
优点	由于对超额数据包进行了缓冲，因而降低了丢弃超额数据包的可能性。(数据包最多可缓冲到队列的长度。如果超额流量持续保持较高速率，则可能会发生掉包情况。)通常可避免因掉包而进行重新传输。	通过掉包控制输出速率。避免因排队导致延迟。
缺点	可能因排队而导致延迟，尤其是在队列较长的情况下。	丢弃多余数据包 (当配置) 时，节流的 TCP 窗口尺寸和减少受影响数据流的整体输出速率。非常激进的突发流量大小可能导致超额数据包丢弃，控制整体输出速率，特别基于 TCP 的数据流。
可选数据	否	是 (使用传统的 CAR 功能)。

包 重 新 标 记		
-----------------------	--	--

*虽然策略不适用缓冲，配置的排队机制适用于一致的数据包，该数据包在等待被连载到物理接口时需要排队。

令牌刷新速率

整形和策略之间的主要区别就是补充令牌的速率。此部分将说明这一区别。

简而言之，整形和策略都使用了令牌桶这一比喻性概念。令牌桶本身没有丢弃或优先级策略。下面我们来看看令牌桶的工作原理：

- 令牌以一定速率放到桶中。
- 每个令牌允许信息源将一定数量的位发送到网络。
- 要发送数据包，数据流调整器必须能够从令牌桶中去除等于数据包大小的令牌数量。
- 如果桶中没有足够的令牌来发送数据包，数据包要么等待有足够的令牌(在整形的情况下)，要么被丢弃或被降级(在警策略情况下)。
- 桶本身具有指定的容量。如果桶达到了容量，最近到达的令牌被丢弃并且不能容纳将来的数据包。因此在任何时间，信息源能够发送到网络的最大突发传输都是令牌桶大小的组成部分。令牌桶允许突变流量，但会对其进行限制。

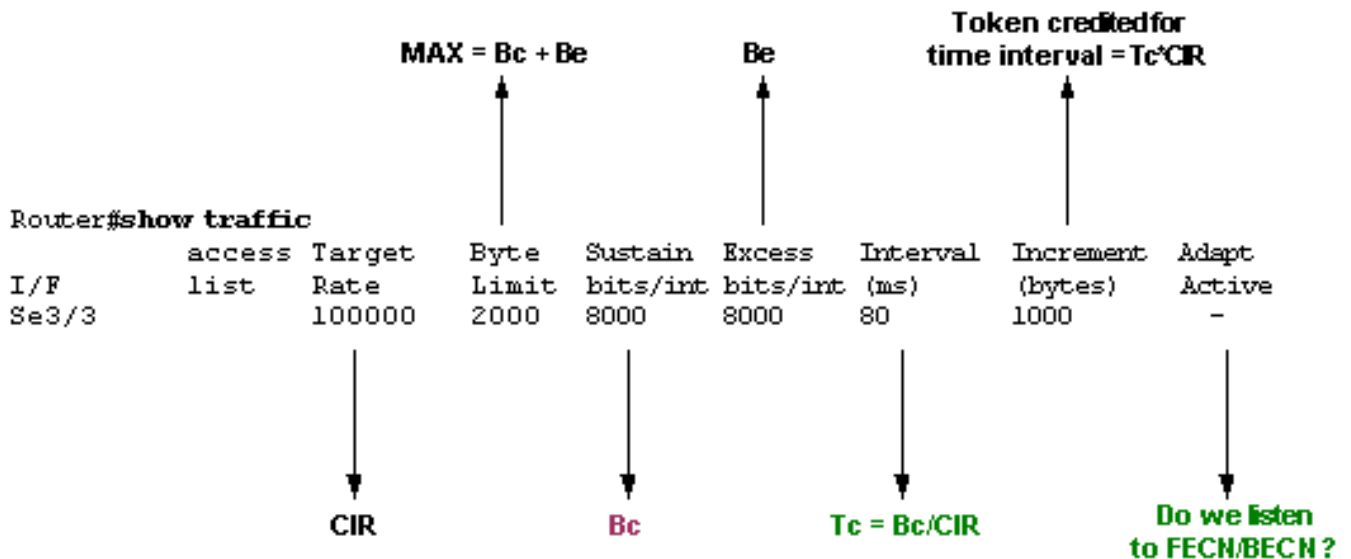
记住令牌桶比喻，让我们查看整形和策略管理如何将令牌添加到桶。

使用比特/秒(位/秒)值，整形增加令牌桶在计时的间隔。整形器使用以下公式：

$$Tc = Bc / CIR \text{ (in seconds)}$$

在此等式中，Bc 表示承诺突发流量，CIR 表示承诺信息速率。(有关详细信息，请参阅[配置帧中继流量整形](#)。) Tc 值定义了为维持 CIR 的平均速率，要在其间发送 Bc 比特流量的时间间隔(以秒计)。

Tc 的范围在 10 毫秒到 125 毫秒之间。使用在 Cisco 7500 系列的分布式流量整形，最低 Tc 是 4 毫秒。路由器会基于 CIR 和 Bc 值在内部计算此值。如果 Bc/CIR 少于 125 毫秒，它将使用该等式计算出的 Tc。如果 Bc/CIR 大于或等于 125 毫秒，并且 Cisco IOS 认为较小的时间间隔会使数据流更稳定，则会使用内部 Tc 值。使用 **show traffic-shape** 命令确定路由器是使用 Tc 的内部值还是使用在命令行配置的值。下面是 **show traffic-shape** 命令的示例输出，有关其说明，请参阅[帧中继流量整形的 show 命令](#)。



当超额突发传输(Be)配置为非0值时，成型机则允许令牌存储在桶中，高至Bc+Be。令牌桶可以达到的最大值是 Bc + Be，溢出令牌被丢弃。在桶中具有超过 BC 令牌数的唯一方式是在一个或多个 Tc 期间不要使用所有 BC 令牌。因为令牌桶通过每 Tc 与 BC 令牌被重新补充，您能累计未使用的令牌用于 BC +Be 的最新应用。

相反，基于类的策略和速率限制会持续将令牌添加到桶中。具体而言，令牌到达速率的计算如下：

$$(time\ between\ packets < \text{which is equal to } t-t1 > * policer\ rate) / 8\ bits\ per\ byte$$

换句话说，如果数据包以前的到达时间是t1并且当前时间是t，那么根据令牌到达的速率，桶被更新，字节值为 t-t1。注意信息数据流策略器指定使用字节表示突发值，上述公式应从位转换到字节。

查看使用 8000 bps 的 CIR (或策略速率) 和 1000 个字节的正常突发传输的例子。

```
Router(config)# policy-map police-setting Router(config-pmap)# class access-match Router(config-pmap-c)# police 8000 1000 conform-action transmit exceed-action drop
```

令牌桶在达到 1000 字节时将充满。如果 450 字节数据包已经到达，那么数据包符合条件，因为这时令牌桶中已经有足够的字节可以使用。一致行动(传输)是由数据包采取的，450 个字节从令牌桶中清除(留下 550 个字节)。如果下一个数据包 25 秒钟后到达，250 个字节将根据以下公式添加到令牌桶上：

$$(0.25 * 8000) / 8$$

计算结果将 700 字节留在令牌桶中。如果下一个数据包是 800 个字节，数据包将超出并且将采取超出行动(丢弃)。这时不会从令牌桶中取走任何字节。

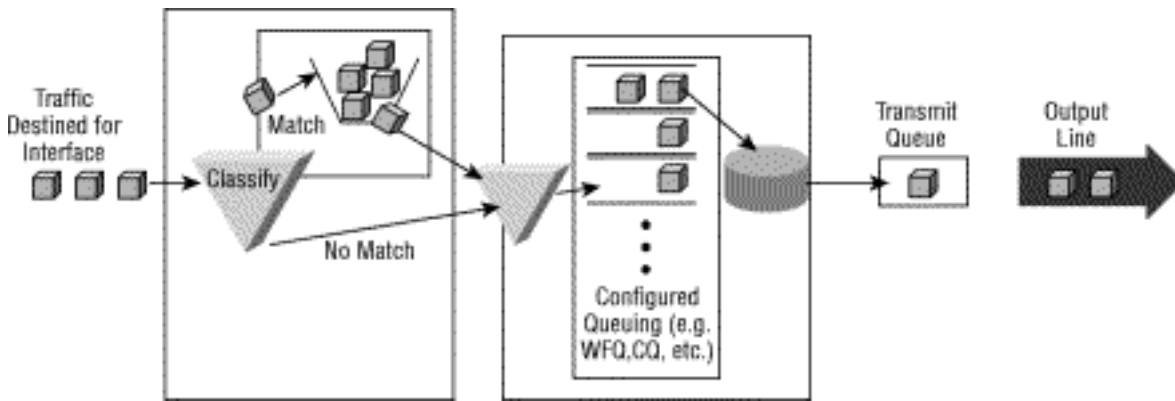
流量整形

Cisco IOS 支持以下流量整形方法：

- [一般流量整形](#)
- [帧中继流量整形](#)
- [基于类的整形](#)和[分布式基于类的整形](#)

所有流量整形的实施方法是相似的，虽然他们的命令行接口 (CLI) 有些不同，但是他们使用不同类型的队列，来包含和整形被延迟的数据流。Cisco 建议使用基于类的整形和分布式整形，可以使用 QoS CLI 配置这类整形。

以下图表说明了 QoS 策略如何将数据流分类成超出所配置的整形速率的数据包级别和队列。



流量策略

Cisco IOS 支持以下流量策略方法：

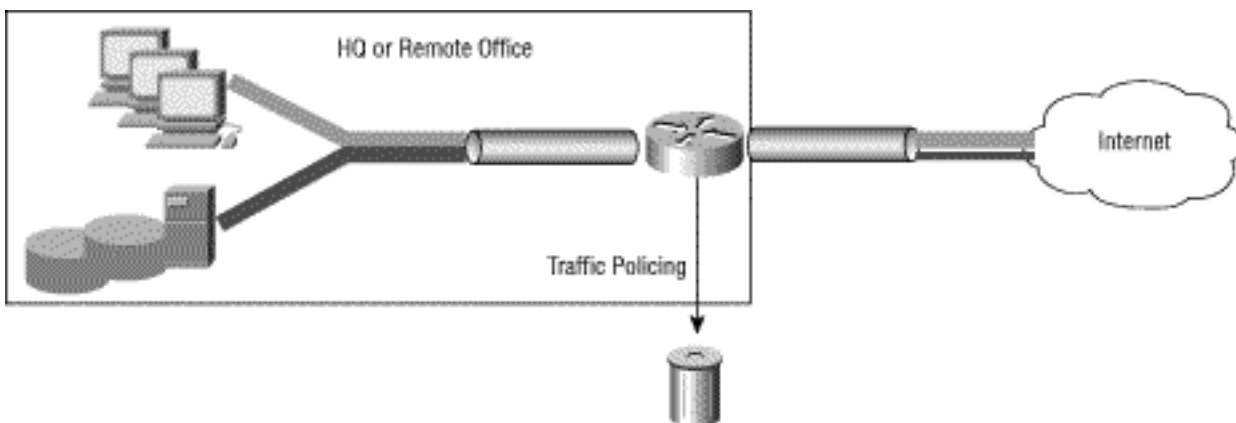
- [承诺接入速率](#)
- [基于类的策略](#)

这两种机制在功能上具有重要区别，[比较基于类的策略和承诺接入速率](#)。当应用 QoS 策略时，Cisco 建议使用基于类的策略和模块 QoS CLI 的其他功能。

使用 **police** 命令，指定流量等级应该有强加的最大速率，如果超出该速率，必须立即采取行动。换句话说，在 **police** 命令下，不能选择缓冲数据包后将其发出，而 **shape** 命令则可以这样做。

另外在使用警管的情况下，令牌桶将确定数据包是否超出或是否与应用的速率一致。无论如何，监管实现一个可配置行为，包括设置IP优先级或差分服务码点(DSCP)。

以下图表显示了拥塞点数据流策略的常见应用，其中通常应用 QoS 功能。



最小带宽与最大带宽控制

shape 和 **police** 命令都会将输出速率限制为最大 kbps 值。重要的是，这两种机制在拥塞期间都不会提供最低带宽保证。可以使用 **bandwidth** 或 **priority** 命令提供这种保证。

分级策略使用 2 个服务策略 – 父策略将 QoS 机制应用到数据流聚合中；子策略将 QoS 机制应用到流或聚合子集中。逻辑接口(例如子接口和隧道接口)需要一个分级策略，在父级上提供流量限制功能、在低级别上提供排队功能。流量限制功能可以减少输出速率并(据推测)创建拥塞，正如排列超额数据包所见。

以下配置不是最理想的，目的是显示在限制流量聚合（在此种情况下级别默认）低于最大速率时，**police** 与 **shape** 命令之间的区别。在此配置中，**police** 命令根据数据包的大小和符合及超额令牌桶中剩余的字节数量从子类中发送数据包。（请参阅[流量策略](#)。）结果是速率给对VoIP和网络协议(IP)类不可以保证，因为**监控功能覆盖优先级功能做的保证**。

然而，如果使用 **shape** 命令，则结果为分层队列系统，并做出所有保证。换句话说，当流入负载超出整形速率的时候，会保证 VoIP 和 IP 类的速率，并且类默认数据流(在子级上)会发生丢包。

警告： 建议不使用此配置，并显示此设置以说明在限制流量聚合时 **police** 和 **shape** 命令之间的区别。

```
class-map match-all IP
  match ip precedence 3
class-map match-all VoIP
  match ip precedence 5

policy-map child
  class VoIP
    priority 128
  class IP
    priority 1000

policy-map parent
  class class-default
    police 3300000 103000 103000 conform-action transmit exceed-action drop
  service-policy child
```

为了让上述配置有意义，应该用整形替换策略。例如：

```
policy-map parent
  class class-default
    shape average 3300000 103000 0
  service-policy child
```

为了得知更多parent和子策略，请参考[QoS优先级的儿童服务策略](#)。

[相关信息](#)

- [QoS 技术支持](#)
- [技术支持 - Cisco Systems](#)