Vergleich von Datenverkehrsrichtlinien und Datenverkehrsform zur Bandbreitenbegrenzung

Inhalt

Einleitung

In diesem Dokument werden die Funktionsunterschiede zwischen Traffic Shaping und Traffic Policing beschrieben, die die Ausgaberate begrenzen.

Voraussetzungen

Anforderungen

Es gibt keine spezifischen Anforderungen für dieses Dokument.

Verwendete Komponenten

Dieses Dokument ist nicht auf bestimmte Software- und Hardware-Versionen beschränkt.

Die Informationen in diesem Dokument beziehen sich auf Geräte in einer speziell eingerichteten Testumgebung. Alle Geräte, die in diesem Dokument benutzt wurden, begannen mit einer gelöschten (Nichterfüllungs) Konfiguration. Wenn Ihr Netzwerk in Betrieb ist, stellen Sie sicher, dass Sie die möglichen Auswirkungen aller Befehle kennen.

Konventionen

Weitere Informationen zu Dokumentkonventionen finden Sie unter Cisco Technical Tips

Conventions (Technische Tipps von Cisco zu Konventionen).

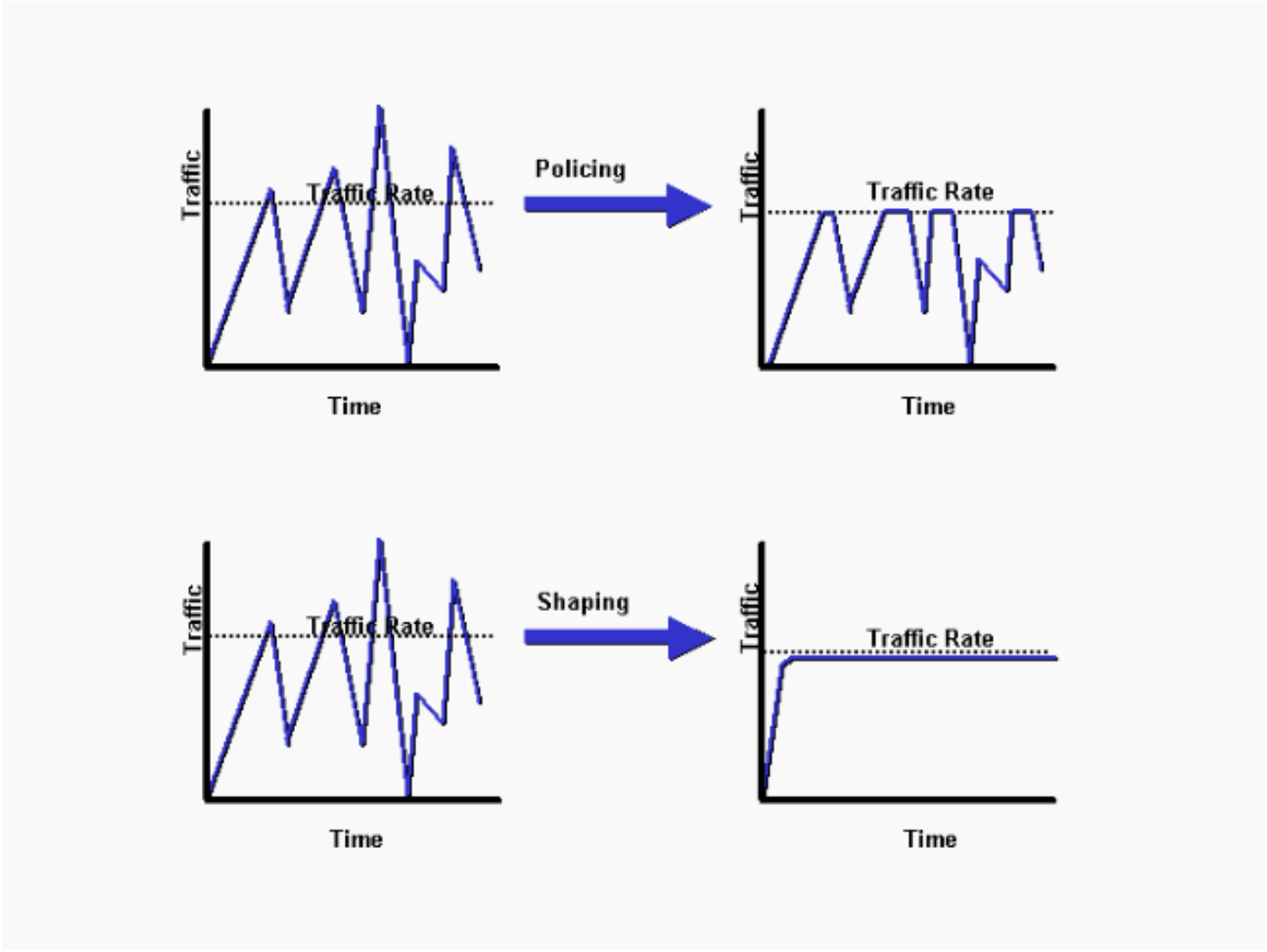
Hintergrundinformationen

In diesem Dokument werden die funktionalen Unterschiede zwischen Traffic Shaping und Richtlinienzuweisung erläutert. Beide schränken die Datenverkehrsausgabe ein. Beide Mechanismen verwenden einen Token-Bucket als Datenverkehrsmesser, um die Paketrate zu messen. Weitere Informationen zu Tokenbuckets finden Sie unter Was ist ein Tokenbucket?

Richtliniendurchsetzung im Vergleich zu Shaping

Traffic Policing propagiert Bursts. Wenn die Datenverkehrsrate die konfigurierte maximale Rate erreicht, wird überschüssiger Datenverkehr verworfen (oder markiert). Das Ergebnis ist eine Produktionsrate, die wie ein Sägezahn mit Kämmen und Trögen erscheint. Im Gegensatz zum Policing behält Traffic Shaping überzählige Pakete in einer Warteschlange bei und plant die überzähligen Pakete dann für eine spätere Übertragung über einen bestimmten Zeitraum. Das Ergebnis des Traffic Shaping ist eine geglättete Paketausgangsrate.

Das nächste Diagramm veranschaulicht die wichtigsten Unterschiede zwischen den beiden Datenverkehrsoptionen.



Shaping impliziert das Vorhandensein einer Warteschlange und eines ausreichenden Arbeitsspeichers, um verzögerte Pakete zu puffern, während das Policing dies nicht tut. Warteschlangen sind ein Konzept für ausgehenden Datenverkehr. Pakete, die eine Schnittstelle verlassen, werden in die Warteschlange eingereiht und können angepasst werden. Auf eingehenden Datenverkehr über eine Schnittstelle kann nur die Richtlinienzuweisung angewendet werden. Stellen Sie sicher, dass Sie beim Aktivieren des Shaping über ausreichend Speicher verfügen. Darüber hinaus erfordert das Shaping eine Funktion, die die spätere Übertragung von verzögerten Paketen plant. Mit dieser Zeitplanfunktion können Sie die Shaping-Warteschlange in verschiedene Warteschlangen organisieren. Beispiele für diese Funktion sind Class Based Weighted Fair Queuing (CBWFQ) und Low Latency Queuing (LLQ).

Auswahlkriterien

In der nächsten Tabelle werden die Unterschiede zwischen Shaping und Richtlinienzuweisung aufgelistet, um Ihnen bei der Auswahl der richtigen Datenverkehrslösung zu helfen.

	Formgebung	Richtlinien
Ziel	Puffert überschüssige Pakete über die zugesicherten Raten und stellt sie in die Warteschlange.	Überschüssige Pakete können über die festgelegten Raten verworfen (oder kommentiert) werden. Puffert nicht.*
Token-Aktualisierungsrate	Erhöht zu Beginn eines Zeitintervalls. (Die Mindestanzahl von Intervallen ist erforderlich.)	Kontinuierlich basierend auf Formel: `1 / Committed Information Rate`
Tokenwerte	Konfiguriert in Bits pro Sekunde.	Konfiguriert in Byte.
Konfigurationsoptionen	• Shape-Befehl in der modularen QoS-Befehlszeilenschnittstelle (MQC) zur Implementierung von klassenbasiertem Shaping • Frame-Relay Traffic-Shape-Befehl zur Implementierung von Frame Relay Traffic Shaping (FRTS). • Traffic-shape-Befehl zum Implementieren von Generic Traffic Shaping (GTS).	• Richtlinienbefehl im MQC zum Implementieren der klassenbasierten Richtlinienzuweisung. • Rate-Limit-Befehl zur Implementierung der Committed Access Rate (CAR).
Bei eingehendem Datenverkehr anwendbar	Nein	Ja
Bei ausgehendem Datenverkehr anwendbar	Ja	Ja

Spitzen	Steuert Bursts und glättet die Ausgaberate über mindestens acht Zeitintervalle. Verwendet einen undichten Bucket, um den Datenverkehr zu verzögern, wodurch eine Glättung erreicht wird.	Verteilt Ausbrüche. Keine Glättung.
Vorteile	Geringere Wahrscheinlichkeit, dass überzählige Pakete verworfen werden, da überzählige Pakete gepuffert werden (Puffert Pakete bis zur Länge der Warteschlange. Wenn der überschüssige Datenverkehr mit hohen Übertragungsraten anhält, kann es zu Verlusten kommen.) In der Regel werden Neuübertragungen aufgrund verlorener Pakete vermieden.	Steuert die Ausgaberate bei Paketverlusten. Vermeidet Verzögerungen aufgrund von queuing.
Nachteile	Mögliche Verzögerung aufgrund von queuing, insbesondere langen Warteschlangen.	Überschüssige Pakete werden bei entsprechender Konfiguration verworfen, TCP-Fenstergrößen werden gedrosselt, und die Gesamtausgabe der betroffenen Datenverkehrsströme wird reduziert. Übermäßig aggressive Burst-Größen können zu übermäßigen Paketverlusten führen und die Gesamtausgabe drosseln, besonders bei TCP-basierten Datenflüssen.
Optionale Paketkennzeichnung	Nein	Ja (mit älterer CAR-Funktion).

* Obwohl die Richtlinienvergabe keinen Puffer anwendet, wird ein konfigurierter queuing Mechanismus auf konforme Pakete angewendet, die in die Warteschlange gestellt werden müssen, während sie auf die Serialisierung an der physischen Schnittstelle warten.

Token-Aktualisierungsrate

Ein wesentlicher Unterschied zwischen Shaping und Policing ist die Rate, mit der Token aufgefüllt werden. Sowohl das Shaping als auch das Policing verwenden die Metapher des Token-Buckets. Ein Token-Bucket selbst hat keine Verwerfungs- oder Prioritätsrichtlinie.

Mit Token-Bucket-Funktionalität:

- Token werden mit einer bestimmten Geschwindigkeit in den Eimer gelegt.

- Jedes Token gibt der Quelle die Berechtigung, eine bestimmte Anzahl von Bits in das Netzwerk zu senden.

- Um ein Paket zu senden, muss der Verkehrsregler in der Lage sein, eine Anzahl von Token, die der Paketgröße entspricht, aus dem Bucket zu entfernen.

- Wenn sich nicht genügend Token im Bucket befinden, um ein Paket zu senden, wartet das Paket entweder, bis der Bucket genügend Token hat (im Fall eines Shapers), oder das Paket wird verworfen oder markiert (im Fall eines Policers).

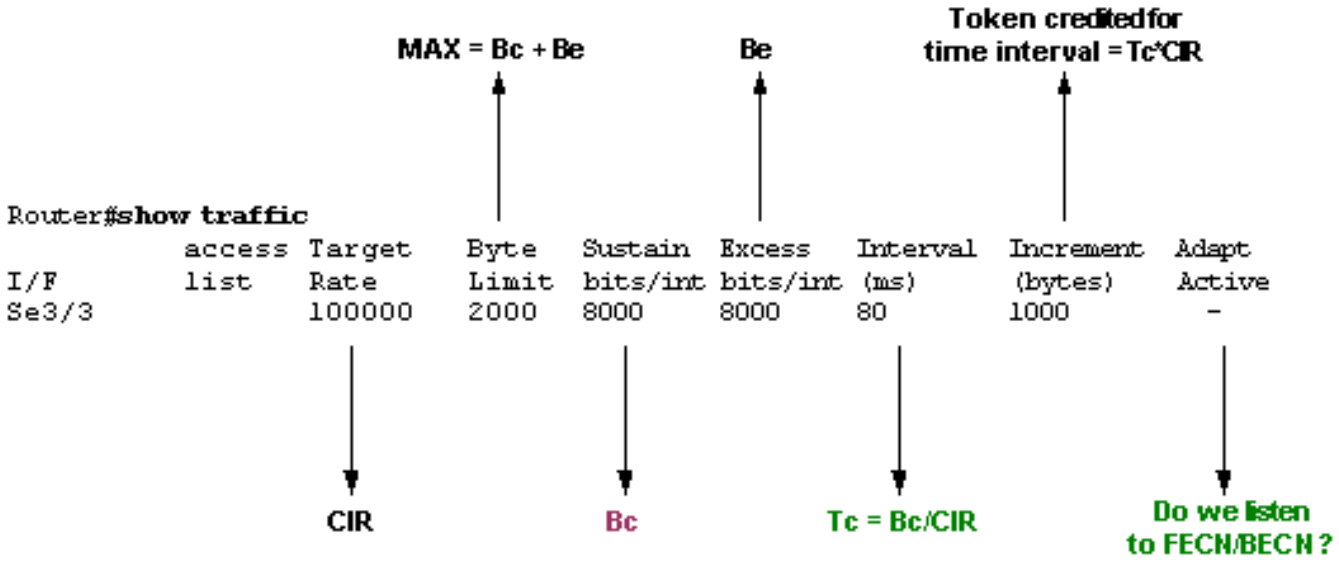
- Der Eimer selbst hat eine bestimmte Kapazität. Wenn der Puffer voll ist, werden neue Token, die eintreffen, verworfen und stehen zukünftigen Paketen nicht zur Verfügung. Somit ist der größte Burst, den eine Quelle in das Netzwerk senden kann, zu jeder Zeit in etwa proportional zur Größe des Buckets. Ein Token-Eimer erlaubt Geschmeidigkeit, begrenzt sie aber.

Beim Shaping wird der Token-Bucket in zeitlichen Intervallen inkrementiert, die einen Bit-pro-Sekunde (Bit/s)-Wert verwenden. Ein Shaper verwendet die folgende Formel:

```
Tc = Bc/CIR (in seconds)
```

In dieser Gleichung steht Bc für den Committed Burst und CIR für Committed Information Rate. (Weitere Informationen finden Sie unter Konfigurieren des Frame Relay-Traffic-Shaping.) Der Wert von Tc definiert das Zeitintervall, in dem die Bc-Bits gesendet werden, um die durchschnittliche CIR-Rate in Sekunden beizubehalten.

Der Bereich für Tc liegt zwischen 10 ms und 125 ms. Beim Distributed Traffic Shaping (DTS) der Cisco Serie 7500 beträgt die Mindestgeschwindigkeit 4 ms. Der Router berechnet diesen Wert intern auf Basis der CIR- und BC-Werte. Wenn Bc/CIR kleiner als 125 ms ist, wird die aus dieser Gleichung berechnete Tc verwendet. Wenn Bc/CIR größer als oder gleich 125 ms ist, wird ein interner Tc-Wert verwendet, wenn Cisco IOS® feststellt, dass der Datenverkehrsfluss mit einem kleineren Intervall stabiler sein kann. Verwenden Sie den Befehl **show traffic-shape (Verkehrsform anzeigen),** um zu bestimmen, ob Ihr Router einen internen Wert für Tc oder den Wert verwendet, den Sie in der Befehlszeile konfiguriert haben. Die nächste Beispielausgabe des Befehls **show traffic-shape** wird unter Show Commands for Frame Relay Traffic Shaping erläutert.



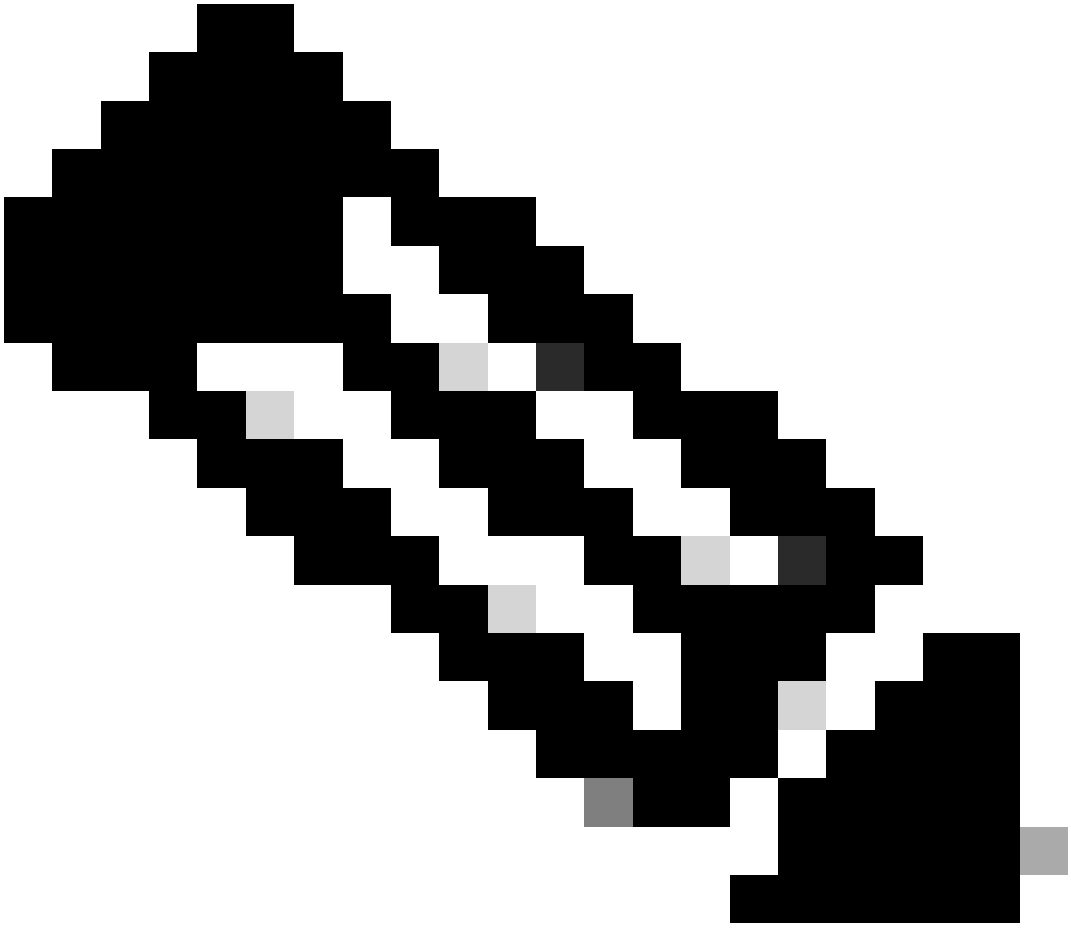
Ausgabe von Datenverkehr anzeigen

Wenn der überschüssige Burst (Be) auf einen Wert ungleich 0 eingestellt ist, erlaubt der Shaper die Speicherung von Token im Eimer bis Bc + Be. Der größte Wert, den der Token-Bucket jemals erreichen kann, ist Bc + Be, und Überlauftokens werden verworfen. Die einzige Möglichkeit, mehr als Bc-Token im Bucket zu haben, besteht darin, nicht alle Bc-Token während eines oder mehrerer Tc zu verwenden. Da der Token-Eimer jedes Tc mit Bc-Token aufgefüllt wird, können Sie ungenutzte Token für die spätere Verwendung bis Bc + Be ansammeln.

Im Gegensatz dazu werden Token durch klassenbasiertes Policing und Rate-limiting Add kontinuierlich zum Bucket hinzugefügt. Im Einzelnen wird die Token-Ankunftsrate wie folgt berechnet:

```
(time between packets<which is equal to t-t1>* policer rate)/8 bits per byte
```

Mit anderen Worten, wenn die vorherige Ankunft des Pakets bei t1 war und die aktuelle Zeit t ist, wird der Bucket mit t-t1 Byte auf der Basis der Token-Ankunftsrate aktualisiert.



Hinweis: Eine Datenverkehrsüberwachung verwendet Burst-Werte, die in Byte angegeben sind, und die vorherige Formel wandelt Bits in Byte um.

Dieses Beispiel verwendet eine CIR (oder Policer-Rate) von 8.000 bps und einen normalen Burst von 1.000 Byte:

<#root>

```
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
```

Die Token-Buckets beginnen bei 1000 Bytes voll. Wenn ein 450-Byte-Paket eingeht, entspricht das Paket den Vorgaben, da genügend Bytes im Token-Bucket verfügbar sind. Die konforme Aktion (Senden) wird vom Paket ausgeführt, und 450 Byte werden aus dem Token-Bucket entfernt (und belassen Sie 550 Byte). Wenn das nächste Paket 0,25 Sekunden später eingeht, werden dem Token-Bucket 250 Byte gemäß der folgenden Formel hinzugefügt:

```
(0.25 * 8000)/8
```

Bei der Berechnung bleiben 700 Byte im Token-Bucket. Wenn das nächste Paket 800 Byte umfasst, überschreitet das Paket und die Aktion zum Überschreiten (Verwerfen) wird ausgeführt. Aus dem Token-Bucket werden keine Bytes entnommen.

Traffic Shaping

Cisco® IOS unterstützt die folgenden Traffic Shaping-Methoden:

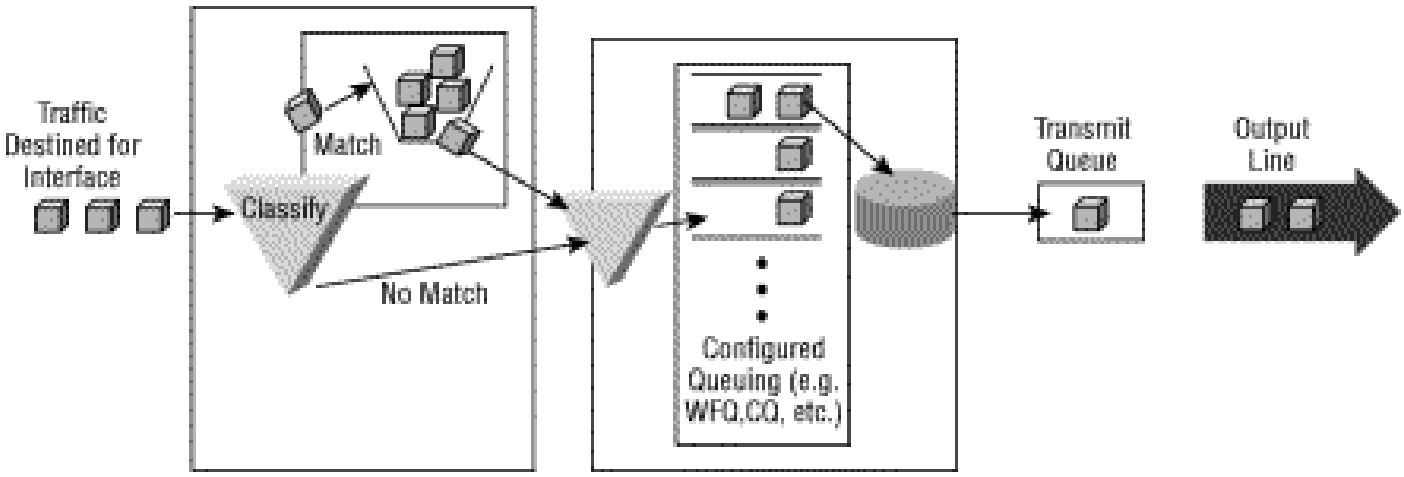
- Generisches Traffic Shaping

- Frame-Relay-Traffic-Shaping

- Klassenbasiertes Shaping und verteiltes klassenbasiertes Shaping

Alle Traffic Shaping-Methoden sind in der Implementierung ähnlich, auch wenn sich ihre Befehlszeilenschnittstellen (CLIs) in gewissem Maße unterscheiden, und sie verwenden verschiedene Arten von Warteschlangen, um zurückgestellten Datenverkehr einzudämmen und zu steuern. Cisco empfiehlt ein klassenbasiertes Shaping und verteiltes Shaping, die mit der modularen QoS-CLI konfiguriert werden.

Das nächste Diagramm zeigt, wie eine QoS-Richtlinie Datenverkehr in Klassen einteilt und Pakete in die Warteschlange einordnet, die die konfigurierten Shaping-Raten überschreiten.



Datenverkehrs-Policing

Cisco IOS unterstützt die folgenden Methoden der Datenverkehrsüberwachung:

- Committed Access Rate

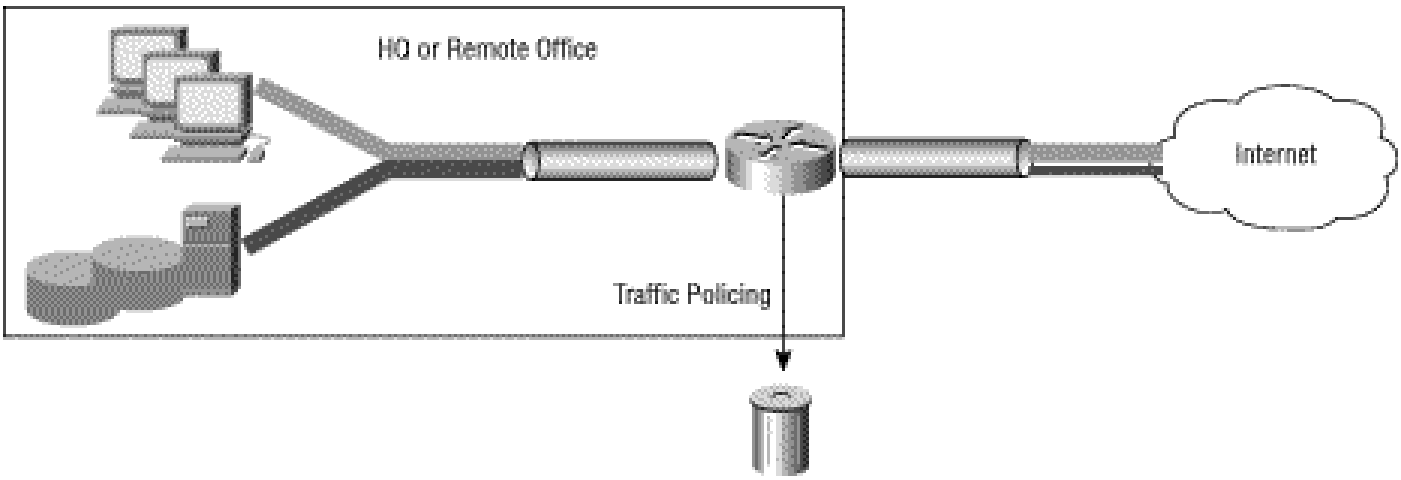
- Klassenbasiertes Policing

Die beiden Mechanismen weisen wichtige Funktionsunterschiede auf, wie unter Compare Class Based Policing und Committed Access Rate erläutert. Cisco empfiehlt eine klassenbasierte Richtlinienvergabe sowie andere Funktionen der modularen QoS-CLI, wenn QoS-Richtlinien angewendet werden.

Verwenden Sie den Befehl **police**, um anzugeben, dass für eine Datenverkehrsklasse eine Höchstgeschwindigkeit festgelegt werden muss. Wenn diese Geschwindigkeit überschritten wird, müssen sofort Maßnahmen ergriffen werden. Mit anderen Worten, mit dem **Befehl "police"** ist es keine Option, das Paket zu puffern und später zu senden, wie es beim Befehl "**shape"** der Fall ist.

Darüber hinaus bestimmt der Token-Bucket bei der Richtlinienvergabe, ob ein Paket die angewendete Rate überschreitet oder mit ihr übereinstimmt. In beiden Fällen wird durch die Richtlinienvergabe eine konfigurierbare Aktion implementiert, die die IP-Rangfolge oder den Differentiated Services Code Point (DSCP) umfasst.

Das nächste Diagramm zeigt eine häufige Anwendung der Datenverkehrssteuerung an einem Überlastungspunkt, an dem QoS-Funktionen im Allgemeinen angewendet werden.



Vergleich zwischen minimaler und maximaler Bandbreitensteuerung

Sowohl der Befehl **shape** als auch der Befehl **police** beschränken die Ausgaberate auf einen maximalen Kbit/s-Wert. Wichtig ist, dass keiner der beiden Mechanismen eine garantierte Mindestbandbreite in Zeiten von Engpässen bietet. Verwenden Sie den Befehl **bandwidth** oder **priority** (**Bandbreite** oder **Priorität**), um solche Garantien bereitzustellen.

Eine hierarchische Richtlinie verwendet zwei Dienstrichtlinien: eine übergeordnete Richtlinie, um einen QoS-Mechanismus auf ein Datenverkehrsaggregat anzuwenden, und eine untergeordnete Richtlinie, um einen QoS-Mechanismus auf einen Fluss oder eine Teilmenge des Aggregats anzuwenden. Logische Schnittstellen, z. B. Subschnittstellen und Tunnelschnittstellen, erfordern eine hierarchische Richtlinie mit der Datenverkehrsfunktion limiting auf der übergeordneten Ebene und der Warteschlangenverwaltung auf niedrigeren Ebenen. Die Traffic-limiting Funktion reduziert die Ausgaberate und erzeugt (vermutlich) eine Überlastung, wie man an queuing überzähligen Paketen sieht.

Die nächste Konfiguration ist nicht optimal und veranschaulicht den Unterschied zwischen dem Befehl "**Police"** und dem Befehl "**shape",** wenn limiting ein Datenverkehr auf eine maximale Rate aggregiert wird (in diesem Fall "class-default"). In dieser Konfiguration sendet der **Befehl "policy"** Pakete aus den untergeordneten Klassen, basierend auf der Größe des Pakets und der Anzahl der Bytes, die in der Übereinstimmungsklasse verbleiben und Token-Buckets überschreiten. (Siehe Traffic Policing.) Dies führt dazu, dass die für die VoIP- (Voice over IP) und IP-Klassen (Internet Protocol) festgelegten Tarife nicht garantiert werden können, da die Polizeifunktion die Garantien der Prioritätsfunktion außer Kraft setzt.

Wird jedoch der Befehl **shape** verwendet, ergibt sich ein hierarchisches Warteschlangensystem, und es werden alle Garantien gegeben. Mit anderen Worten, wenn die angebotene Last die Formrate überschreitet, wird die Geschwindigkeit der VoIP- und IP-Klassen garantiert, und der Standardklassenverkehr (auf untergeordneter Ebene) geht verloren.

Vorsicht: Diese Konfiguration wird nicht empfohlen und veranschaulicht den Unterschied zwischen dem **Befehl** "**Police**" und dem Befehl "**shape",** wenn dadurch die Gesamtzahl des Datenverkehrs begrenzt wird.

```
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
```

Damit die vorherige Konfiguration sinnvoll ist, muss die Richtlinienzuweisung durch Shaping ersetzt werden.

Beispiele:

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

Hinweis: Weitere Informationen zu übergeordneten und untergeordneten Richtlinien finden Sie unter QoS Child Service Policy for Priority Class (QoS-Richtlinie für untergeordnete Dienste für Prioritätsklasse).

Zugehörige Informationen

- **Technologischer Support für Quality of Services (QoS)**
- **Technischer Support und Dokumentation für Cisco Systeme**

Informationen zu dieser Übersetzung

Cisco hat dieses Dokument maschinell übersetzen und von einem menschlichen Übersetzer editieren und korrigieren lassen, um unseren Benutzern auf der ganzen Welt Support-Inhalte in ihrer eigenen Sprache zu bieten. Bitte beachten Sie, dass selbst die beste maschinelle Übersetzung nicht so genau ist wie eine von einem professionellen Übersetzer angefertigte. Cisco Systems, Inc. übernimmt keine Haftung für die Richtigkeit dieser Übersetzungen und empfiehlt, immer das englische Originaldokument (siehe bereitgestellter Link) heranzuziehen.