# Comparaison de la régulation et du formatage du trafic pour limiter la bande passante

## Table des matières


Introduction

Conditions préalables

- Exigences
- Composants utilisés

Informations générales

Réglementation et mise en forme

Critères de sélection

Taux d'actualisation des jetons

Modélisation du trafic

Contrôle du trafic

Contrôles de bande passante minimum et maximum

Informations connexes


## Introduction

Le présent document décrit les différences fonctionnelles entre la régulation du trafic et la réglementation du trafic, qui limitent tous deux le débit de sortie.

## Conditions préalables

### Exigences

Aucune exigence spécifique n'est associée à ce document.

### Composants utilisés

Ce document n'est pas limité à des versions de matériel et de logiciel spécifiques.

The information in this document was created from the devices in a specific lab environment. All of

the devices used in this document started with a cleared (default) configuration. Si votre réseau est en ligne, assurez-vous de bien comprendre l’incidence possible des commandes.

## Informations générales

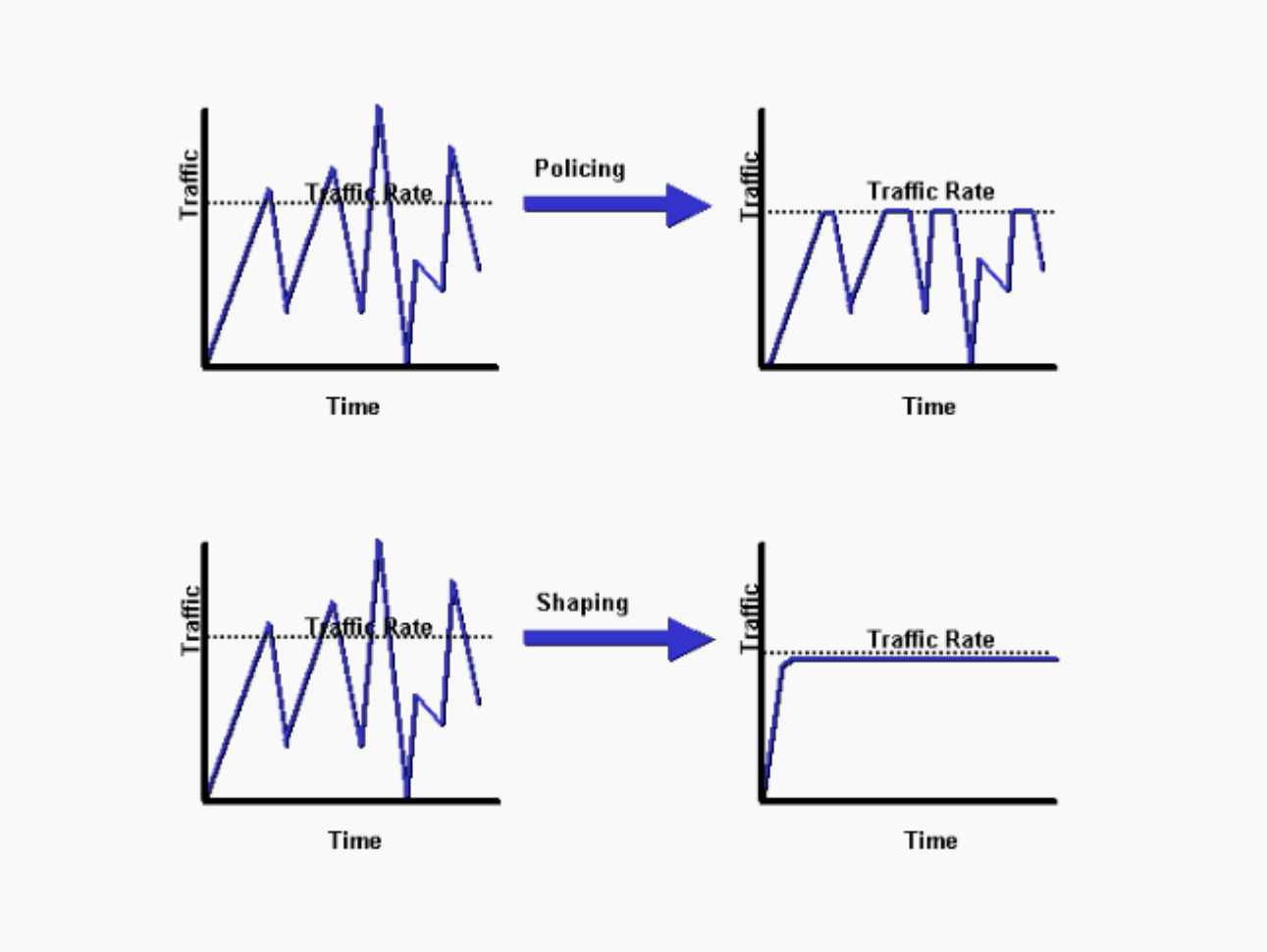
Ce document clarifie les différences fonctionnelles entre le formatage du trafic et la réglementation. Les deux limitent le débit de sortie du trafic. Les deux mécanismes utilisent un saut de jeton comme indicateur de trafic pour mesurer le débit de paquets. Pour plus d'informations sur les compartiments à jetons, consultez Qu'est-ce qu'un compartiment à jetons ?

Les concepts présentés dans ce document restent pertinents pour comprendre les conceptions de qualité de service Cisco. Le formatage du trafic et la réglementation du trafic continuent d'être utilisés pour appliquer des débits de trafic maximum et gérer la congestion. Pour les nouveaux déploiements, Cisco recommande l'utilisation de la mise en forme basée sur les classes et de la réglementation basée sur les classes de la ligne de commande modulaire QoS (MQC), si elle est prise en charge par la plate-forme. Des mécanismes hérités, tels que Generic Traffic Shaping (GTS), Frame Relay Traffic Shaping (FRTS), Committed Access Rate (CAR) et la commande rate-limit, sont inclus pour le contexte historique ou pour prendre en charge les discussions spécifiques à la plate-forme.

## Réglementation et mise en forme

La réglementation du trafic propage les rafales. Lorsque le débit de trafic atteint le débit maximal configuré, le trafic excédentaire est abandonné (ou signalé). Le résultat est une vitesse de sortie qui apparaît comme une dent de scie avec des crêtes et des creux. Contrairement à la réglementation, le formatage du trafic conserve les paquets en excès dans une file d'attente, puis planifie l'excès pour une transmission ultérieure par incréments de temps. Le résultat du formatage du trafic est un débit de sortie de paquets lissé.

Le schéma suivant illustre les principales différences entre les deux options de trafic.




Réglementation et mise en forme

La mise en forme implique l'existence d'une file d'attente et d'une mémoire suffisante pour mettre en mémoire tampon les paquets retardés, contrairement à la réglementation. Les files d'attente sont un concept sortant ; les paquets qui quittent une interface sont mis en file d'attente et peuvent être formatés. Seule la réglementation peut être appliquée au trafic entrant sur une interface. Assurez-vous que vous disposez de suffisamment de mémoire lorsque vous activez le formatage. En outre, la mise en forme nécessite une fonction qui planifie la transmission ultérieure de tout paquet retardé. Cette fonctionnalité de planification vous permet d'organiser la file d'attente de mise en forme en différentes files d'attente. Des exemples de cette fonctionnalité sont CBWFQ Queuing (Class Based Weighted Fair) et LLQ Queuing (Low Latency).

---

Remarque : Le formatage du trafic et la réglementation du trafic appliquent tous deux un débit de trafic configuré. La mise en forme est un mécanisme de mise en file. La réglementation peut être appliquée en entrée ou en sortie, en fonction de la plate-forme et de la prise en charge des fonctionnalités.

---

## Critères de sélection

Le tableau suivant répertorie les différences entre la mise en forme et la réglementation pour vous aider à choisir la solution de trafic appropriée.

| | Modélisation du trafic | Contrôle du trafic |
|---|---|---|
| Comportement principal | Mettre en mémoire tampon et mettre en file d'attente les paquets excédentaires sur les débits validés. | Abandonne (ou remarque) les paquets en excès par rapport aux débits garantis. Ne met pas en mémoire tampon.* |
| Taux d'actualisation des jetons | Les jetons sont réapprovisionnés périodiquement à chaque intervalle Tc. Au début de chaque intervalle, le shaper ajoute des jetons Bc, jusqu'à la limite de bucket configurée. Tc est limité par des limites spécifiques à la plate-forme. | Les jetons sont réapprovisionnés en fonction du temps écoulé entre les arrivées de paquets. Le régulateur calcule le nombre de jetons à ajouter avec : ((t - t1) * débit_régulateur_bps) / 8, où t - t1 est le temps écoulé depuis le paquet précédent. |
| Valeurs de jeton | Configuré en bits par seconde (bits/s). | Configuré en octets. |
| Options de configuration | • Commande Shape dans MQC pour implémenter la mise en forme basée sur les classes.<br>• Frame-relay traffic-shape pour implémenter FRTS.<br>• Traffic-shape pour implémenter GTS. | • Commande de police dans le MQC pour mettre en oeuvre la police basée sur les classes.<br>• Rate-limit pour implémenter CAR. |
| Direction type | Sortant | Entrant ou sortant, selon la plate-forme |
| Rafales | Contrôle les rafales et lisse le débit de sortie sur au moins huit intervalles de temps. Il utilise le comptage de compartiments à jetons et met en file d'attente les paquets excédentaires. Les paquets mis en file d'attente sont libérés au fil du temps, ce qui crée un effet de lissage similaire à celui d'un seau percé. | Autorise les rafales dans les limites de rafales configurées. Il ne met pas les paquets en mémoire tampon pour lisser le débit de transmission. |
| Avantages | Moins susceptible de rejeter les paquets en excès puisque les paquets en excès sont mis en mémoire tampon. (Met en mémoire tampon les paquets jusqu'à la longueur de la file d'attente. Des abandons peuvent se produire si l'excès de trafic est maintenu à des taux élevés.) Évite généralement les retransmissions dues aux paquets abandonnés. | Contrôle le débit de sortie via les abandons de paquets. Évite les retards dus à queuing. |
| Inconvénients | Peut entraîner des retards dus à des | Abandonne les paquets |

| | | |
|---|---|---|
| | files d'attente queuing, particulièrement profondes. | excédentaires (lorsqu'ils sont configurés), limite la taille des fenêtres TCP et réduit le taux de sortie global des flux de trafic affectés. Des tailles de salves trop agressives peuvent entraîner des pertes de paquets excessives et limiter le débit de sortie global, en particulier avec les flux basés sur TCP. |
| Remarques facultatives sur les paquets | Non | Oui, il peut éventuellement transmettre, supprimer ou marquer des paquets, y compris la priorité IP ou les actions DSCP, en fonction des actions du régulateur MQC configuré et de la prise en charge de la plate-forme. |



Remarque : Bien que la réglementation n'applique pas de tampon, un mécanisme de mise en file d'attente configuré s'applique aux paquets conformes qui doivent être mis en file d'attente pendant qu'ils attendent d'être sérialisés au niveau de l'interface physique.

## Taux d'actualisation des jetons

Une différence clé entre la mise en forme et la réglementation est la vitesse à laquelle les jetons sont réapprovisionnés. La mise en forme et la réglementation utilisent toutes deux la métaphore du saut de jeton. Un compartiment de jetons n'a pas de stratégie de suppression ou de priorité.

Avec la fonctionnalité token bucket :

- Les jetons sont placés dans le compartiment à un certain débit.
- Chaque jeton autorise la source à envoyer un certain nombre de bits dans le réseau.
- Pour envoyer un paquet, le régulateur de trafic doit pouvoir retirer du compartiment un nombre de jetons égal en représentation à la taille du paquet.
- S'il n'y a pas assez de jetons dans le compartiment pour envoyer un paquet, le paquet attend que le compartiment ait assez de jetons (dans le cas d'un formateur), ou le paquet est rejeté ou marqué vers le bas (dans le cas d'un régulateur).
- Le compartiment lui-même a une capacité spécifiée. Si le compartiment atteint sa capacité maximale, les nouveaux jetons qui arrivent sont rejetés et ne sont pas disponibles pour les

paquets à venir. Ainsi, à tout moment, la plus grande rafale qu'une source peut envoyer dans le réseau est à peu près proportionnelle à la taille du compartiment. Un saut de jeton permet la rafale mais la limite.

La mise en forme incrémente le compartiment de jeton à des intervalles temporisés qui utilisent une valeur de bits par seconde (bits/s). Un outil de mise en forme utilise la formule suivante :

<#root>

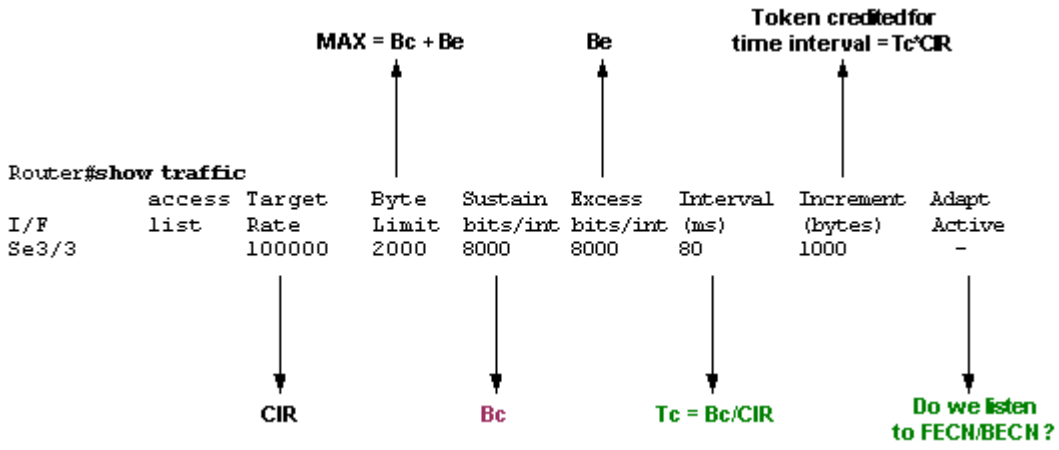
**Tc = Bc/CIR**

(in seconds)

Dans cette équation :

- Bc : rafale validée. La quantité de trafic autorisée à être envoyée pendant un intervalle de temps.
- CIR : débit d'informations garanti. Débit de trafic moyen appliqué par le formateur ou le régulateur.
- Tc : intervalle de temps validé. Intervalle de temps utilisé par un shaper pour envoyer la rafale validée tout en conservant le CIR configuré en secondes.

La plage de Tc est comprise entre 10 ms et 125 ms. Sur certaines plates-formes existantes avec DTS (Distributed Traffic Shaping), la durée de vie minimale est de 4 ms. Le routeur calcule cette valeur en interne en fonction des valeurs CIR et Bc. Si Bc/CIR est inférieur à 125 ms, il utilise le Tc calculé à partir de cette équation. Si Bc/CIR est supérieur ou égal à 125 ms, il utilise une valeur Tc interne si Cisco IOS détermine que le flux de trafic peut être plus stable avec un intervalle plus petit. Utilisez la commande show traffic-shape pour déterminer si votre routeur utilise une valeur interne pour Tc ou la valeur que vous avez configurée à la ligne de commande.




show traffic output

Lorsque la rafale en excès (Be) est configurée sur une valeur différente de 0, le modélisateur permet de stocker les jetons dans le compartiment, jusqu'à Bc + Be. La plus grande valeur que le saut de jetons peut atteindre est Bc + Be, et les jetons de débordement sont abandonnés. La seule façon d'avoir plus de jetons Bc dans le bucket est de ne pas utiliser tous les jetons Bc pendant un ou plusieurs Tc. Puisque le compartiment à jetons est réapprovisionné chaque Tc avec des jetons Bc, vous pouvez accumuler les jetons inutilisés pour une utilisation ultérieure jusqu'à Bc + Be.

En revanche, la réglementation basée sur les classes et la limitation du débit réapprovisionnent les jetons en fonction du temps écoulé entre les arrivées de paquets. Lorsqu'un paquet arrive, le régulateur calcule le nombre de jetons à ajouter à partir du temps écoulé depuis le paquet précédent et du débit du régulateur configuré. Plus précisément, le taux d'arrivée de jeton est calculé comme suit :

Tokens added = ((t - t1) * policer_rate_bps) / 8 bytes

En d'autres termes, si l'arrivée précédente du paquet était à t1 et que l'heure actuelle est t, le saut est mis à jour avec la valeur t-t1 des octets en fonction du taux d'arrivée du jeton.

---

Remarque : Un régulateur de trafic utilise des valeurs de rafale spécifiées en octets, et la formule précédente convertit les bits en octets.

---

Voici un exemple qui utilise un débit de données garanti (CIR) de 8 000 bits/s et une rafale normale de 1 000 octets :

```
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
```

Les compartiments de jeton commencent à 1 000 octets. Si un paquet de 450 octets arrive, le paquet est conforme car suffisamment d'octets sont disponibles dans le compartiment à jetons. L'action de conformité (transmission) est effectuée par le paquet et 450 octets sont supprimés du compartiment à jetons (et laissent 550 octets). Si le paquet suivant arrive 0,25 seconde plus tard, 250 octets sont ajoutés au saut de jeton selon la formule suivante :

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

1. Le compartiment commence plein :
    - Rafale normale = 1000 octets
    - Jetons initiaux = 1000 octets
2. Le premier paquet arrive :
    - Taille du paquet = 450 octets
    - Le paquet est conforme car le compartiment contient suffisamment de jetons.
    - Le régulateur supprime l'équivalent en jetons de 450 octets du compartiment.

    Calcul : 1 000 - 450 = 550 octets restants
3. Le paquet suivant arrive 0,25 seconde plus tard :
    - Débit du contrôleur = 8 000 bits/s
    - Jetons ajoutés pendant 0,25 seconde :

    (0,25 × 8 000) / 8 = 250 octets
4. Regroupement après actualisation :
    - Jetons existants restants = 550 octets
    - Nouveaux jetons ajoutés = 250 octets

Calcul : 550 + 250 = 800 octets

Le calcul laisse 800 octets dans le compartiment à jetons (jetons restants dans le compartiment du régulateur). Si le paquet suivant comporte 900 octets, le paquet dépasse et l'action de dépassement (abandon) est effectuée. Aucun octet n'est prélevé dans le compartiment de jeton.

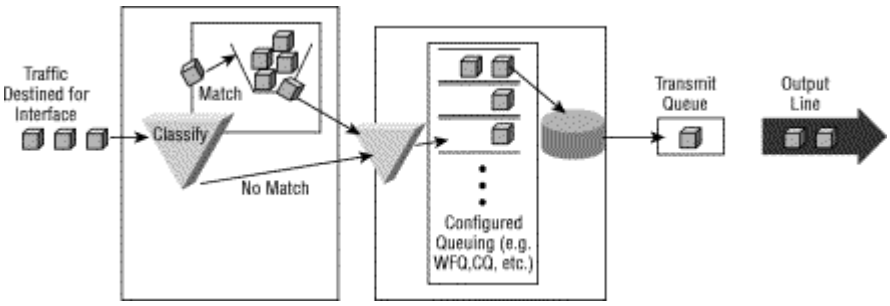
## Modélisation du trafic

Cisco IOS® prend en charge les méthodes suivantes de formatage du trafic :

- Formatage de trafic générique
- Formatage du trafic Frame Relay
- Mise en forme basée sur les classes et mise en forme distribuée basée sur les classes

Toutes les méthodes de formatage du trafic sont similaires dans leur implémentation, bien que leurs interfaces de ligne de commande (CLI) diffèrent quelque peu, et elles utilisent différents types de files d'attente pour contenir et formater le trafic différé. Cisco recommande le formatage basé sur les classes et le formatage distribué, qui sont configurés avec l'interface de ligne de commande modulaire QoS.

Le schéma suivant illustre comment une stratégie QoS organise le trafic en classes et met en file d'attente les paquets qui dépassent les débits de mise en forme configurés.




## Contrôle du trafic

Cisco IOS prend en charge les méthodes suivantes de réglementation du trafic :

- Débit d'accès garanti

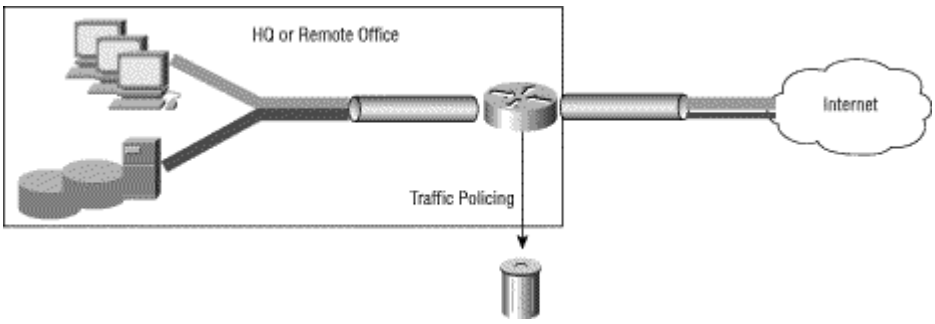
- Contrôle basé sur les classes

Les deux mécanismes présentent des différences fonctionnelles importantes, comme expliqué dans Comment configurer la réglementation basée sur les classes. Cisco recommande la réglementation basée sur les classes et d'autres fonctionnalités de l'interface de ligne de commande QoS modulaire lorsque des politiques QoS sont appliquées.

Utilisez la commande police pour spécifier qu'une classe de trafic doit avoir un débit maximal qui lui est imposé, et si ce débit est dépassé, une action immédiate doit être prise. En d'autres termes, avec la commande police, ce n'est pas une option pour mettre le paquet en mémoire tampon et l'envoyer plus tard, comme c'est le cas pour la commande shape.

En outre, avec la réglementation, le saut de jeton détermine si un paquet dépasse ou est conforme au débit appliqué. Dans les deux cas, la réglementation implémente une action configurable, qui inclut la priorité IP ou le DSCP (Differentiated Services Code Point).

Le schéma suivant illustre une application courante de la réglementation du trafic au niveau d'un point de congestion, où les fonctionnalités QoS s'appliquent généralement.




## Contrôles de bande passante minimum et maximum

Les commandes shape et police limitent le trafic à un débit maximum configuré. Dans MQC, le débit configuré pour la moyenne de forme et la police est spécifié en bits/s, sauf indication contraire de la syntaxe spécifique à la plate-forme. Il est important de noter qu'aucun de ces mécanismes ne garantit une bande passante minimale pendant les périodes d'encombrement. Utilisez la commande bandwidth ou priority pour fournir de telles garanties.

Une stratégie hiérarchique utilise deux stratégies de service : une stratégie parent pour appliquer un mécanisme QoS à un agrégat de trafic et une stratégie enfant pour appliquer un mécanisme QoS à un flux ou à un sous-ensemble de l'agrégat. Les interfaces logiques, telles que les sous-interfaces et les interfaces de tunnel, nécessitent une stratégie hiérarchique avec la fonctionnalité de traficlimiting au niveau parent et la mise en file d'attente aux niveaux inférieurs. La fonctionnalité

de traficlimiting réduit le débit de sortie et (vraisemblablement) crée un encombrement, comme le montrent les paquets en queuing excès.

La configuration suivante est sous-optimale et est montrée pour illustrer la différence entre la commande police et la commande shape quand limiting un trafic s'agrège, dans ce cas class-default, à un débit maximum. Dans cette configuration, la commande police envoie des paquets à partir des classes enfants en fonction de la taille du paquet et du nombre d'octets qui restent dans les compartiments de jeton conformes et excèdent. (Voir Réglementation du trafic.) Il en résulte que les débits donnés aux classes Voix sur IP (VoIP) et Protocole Internet (IP) ne peuvent pas être garantis puisque la fonctionnalité de police prime sur les garanties apportées par la fonctionnalité de priorité.

Cependant, si la commande shape est utilisée, le résultat est un système de mise en file d'attente hiérarchique, et toutes les garanties sont faites. En d'autres termes, lorsque la charge offerte dépasse le débit de forme, les classes VoIP et IP sont assurées de leur débit, et le trafic par défaut de la classe (au niveau enfant) subit des pertes.

---

⚠ Mise en garde : Cette configuration n'est pas recommandée, et est montrée pour illustrer la différence entre la commande police et la commande shape quand elle limite un agrégat de trafic.

---

<#root>

```
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
```

Pour que la configuration précédente ait un sens, la réglementation doit être remplacée par une mise en forme. Exemple :

<#root>

```
policy-map parent
    class class-default

     shape average 3300000 103000 0

     service-policy child
```

Utilisez la commande show policy-map interface <interface> pour vérifier les stratégies configurées, les compteurs de classes, le taux offert, les compteurs de pertes, le taux de formatage, la profondeur de file d'attente et les compteurs de conformité/dépassement/violation du régulateur.

<#root>

Router#

```
show policy-map interface gigabitEthernet 0/0/2

 GigabitEthernet0/0/2


 Service-policy output: parent

    Class-map: class-default (match-any)
      0 packets, 0 bytes
      5 minute offered rate 0000 bps, drop rate 0000 bps
      Match: any
      Queueing
      queue limit 64 packets
      (queue depth/total drops/no-buffer drops) 0/0/0
      (pkts output/bytes output) 0/0

 shape (average) cir 3300000, bc 103000, be 0 target shape rate 3300000


 Service-policy : child

        queue stats for all priority classes:
          Queueing
          queue limit 512 packets
          (queue depth/total drops/no-buffer drops) 0/0/0
          (pkts output/bytes output) 0/0

        Class-map: VoIP (match-all)
          0 packets, 0 bytes
          5 minute offered rate 0000 bps, drop rate 0000 bps
          Match: ip precedence 5

 Priority: 128 kbps, burst bytes 3200, b/w exceed drops: 0
```

```
        Class-map: IP (match-all)
          0 packets, 0 bytes
          5 minute offered rate 0000 bps, drop rate 0000 bps
          Match: ip precedence 3

 Priority: 1000 kbps, burst bytes 25000, b/w exceed drops: 0



        Class-map: class-default (match-any)
          0 packets, 0 bytes
          5 minute offered rate 0000 bps, drop rate 0000 bps
          Match: any

          queue limit 64 packets
          (queue depth/total drops/no-buffer drops) 0/0/0
          (pkts output/bytes output) 0/0
```

 Remarque : Pour en savoir plus sur les stratégies parent et enfant, reportez-vous à QoS : Stratégie de service enfant pour la classe de priorité.

L'exemple précédent est intentionnellement sous-optimal et n'est fourni que pour illustrer la différence entre la réglementation et la mise en forme dans une politique de QoS hiérarchique. Dans une conception QoS recommandée, la commande priority est généralement réservée au trafic en temps réel, sensible à la latence, tel que la voix. Les classes en temps non réel qui nécessitent une garantie de bande passante minimale utilisent généralement la commande bandwidth avec CBWFQ.

Cette approche s'aligne sur les conseils de conception QoS de Cisco pour LLQ et CBWFQ. LLQ fournit un traitement de priorité strict pour le trafic sensible aux retards, tandis que CBWFQ fournit des garanties de bande passante pour d'autres classes de trafic pendant l'encombrement. Par conséquent, dans cet exemple, la classe IP est mieux représentée par la bande passante plutôt que par la priorité, à moins que cette classe ne transporte du trafic qui nécessite explicitement un traitement prioritaire strict.

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

## Informations connexes

- Assistance technologique pour la qualité de service (QoS)
- Assistance et documentation techniques - Cisco Systems

À propos de cette traduction
Cisco a traduit ce document en traduction automatisée vérifiée par une personne dans le cadre d'un service mondial permettant à nos utilisateurs d'obtenir le contenu d'assistance dans leur propre langue.

Il convient cependant de noter que même la meilleure traduction automatisée ne sera pas aussi précise que celle fournie par un traducteur professionnel.