

Planification de la sortie QoS sur les commutateurs des gammes Catalyst 6500/6000 exécutant le logiciel système CatOS

Contenu

[Introduction](#)

[Conditions préalables](#)

[Conditions requises](#)

[Components Used](#)

[Conventions](#)

[Informations générales](#)

[Pertes de la file d'attente de sortie](#)

[Types de mise en file d'attente impliqués dans la planification des sorties sur Catalyst 6500/6000](#)

[Dépose de queue](#)

[Détection précoce aléatoire et détection précoce pondérée aléatoire](#)

[Rotation circulaire pondérée](#)

[File d'attente de priorité stricte](#)

[Fonctionnalité de mise en file d'attente de sortie de différentes cartes de ligne sur le Catalyst 6000](#)

[Fonctionnalités de la commande show port](#)

[Comprendre la fonctionnalité de mise en file d'attente d'un port](#)

[Créer une qualité de service sur Catalyst 6500/6000](#)

[Mécanisme de planification des sorties sur Catalyst 6500/6000](#)

[Configuration, surveillance et planification des sorties sur Catalyst 6500/6000](#)

[Configuration par défaut pour QoS sur Catalyst 6500/6000](#)

[Configuration](#)

[Surveillance de la planification des sorties et vérification de la configuration](#)

[Utiliser la planification des sorties pour réduire les délais et la gigue](#)

[Réduire le délai](#)

[Réduire la gigue](#)

[Informations connexes](#)

Introduction

La planification des sorties garantit que le trafic important n'est pas abandonné en cas de surabonnement important. Ce document traite de toutes les techniques et de tous les algorithmes impliqués dans la planification des sorties sur les commutateurs de la gamme Cisco Catalyst 6500/6000 qui exécutent le logiciel système Catalyst OS (CatOS). Ce document fournit également un bref aperçu de la fonctionnalité de mise en file d'attente des commutateurs Catalyst 6500/6000 et de la façon de configurer les différents paramètres de planification de sortie.

Remarque : si vous exécutez le logiciel Cisco IOS® sur vos commutateurs Catalyst 6500/6000, référez-vous à [Planification des sorties QoS sur les commutateurs de la gamme Catalyst 6500/6000 exécutant le logiciel système Cisco IOS](#) pour plus d'informations.

Conditions préalables

Conditions requises

Aucune spécification déterminée n'est requise pour ce document.

Components Used

Les exemples de ce document ont été créés à partir d'un Catalyst 6000 avec un Supervisor Engine 1A et une carte PFC (Policy Feature Card). Mais les exemples sont également valables pour un Supervisor Engine 2 avec une carte PFC2 ou pour un Supervisor Engine 720 avec une carte PFC3.

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. If your network is live, make sure that you understand the potential impact of any command.

Conventions

Pour plus d'informations sur les conventions utilisées dans ce document, reportez-vous à [Conventions relatives aux conseils techniques Cisco](#).

Informations générales

Pertes de la file d'attente de sortie

Les pertes en sortie sont provoquées par une interface congestionnée. Une cause courante de cette situation pourrait être le trafic provenant d'une liaison à bande passante élevée qui est commutée vers une liaison à bande passante plus faible ou le trafic provenant de plusieurs liaisons entrantes qui sont commutées vers une liaison sortante unique.

Par exemple, si une grande quantité de trafic en rafale entre sur une interface gigabit et est commutée vers une interface de 100 Mbits/s, cela peut entraîner une incrémentation des pertes de sortie sur l'interface de 100 Mbits/s. En effet, la file d'attente de sortie sur cette interface est submergée par l'excès de trafic dû à l'incompatibilité de vitesse entre les bandes passantes entrantes et sortantes. Le débit de trafic sur l'interface sortante ne peut pas accepter tous les paquets qui doivent être envoyés.

La solution ultime pour résoudre le problème est d'augmenter la vitesse de la ligne. Cependant, il y a des façons d'empêcher, de diminuer ou de contrôler les pertes en sortie quand vous ne voulez pas augmenter la vitesse de la ligne. Vous pouvez empêcher les pertes en sortie seulement si elles sont une conséquence de brèves rafales de données. Si les pertes en sortie sont provoquées par un flux constant à haut débit, vous ne pouvez pas empêcher les pertes. Cependant, vous pouvez les contrôler.

Types de mise en file d'attente impliqués dans la planification des sorties sur Catalyst 6500/6000

Dépose de queue

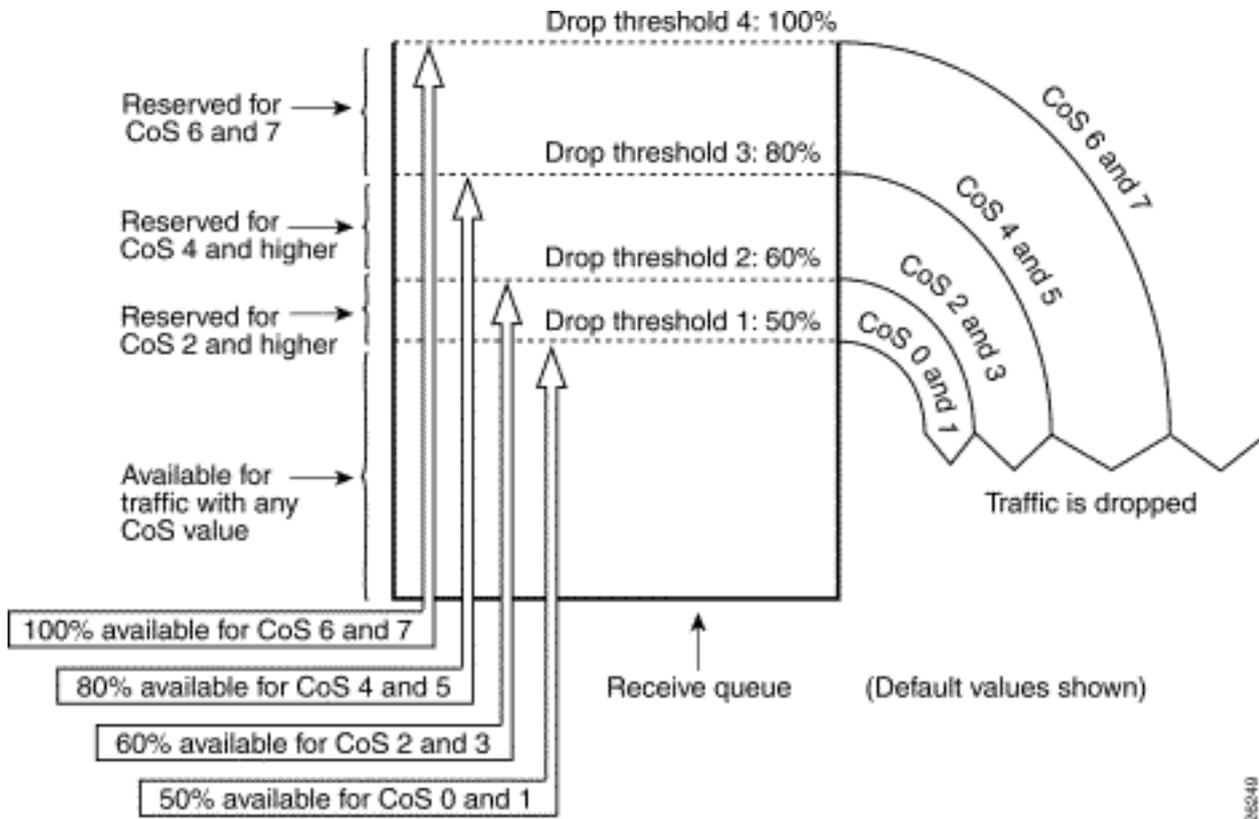
La perte de queue est un mécanisme de base d'évitement de congestion. La perte de queue traite tout le trafic de manière égale et ne fait pas de différence entre les classes de service (CoS) lorsque les files d'attente commencent à se remplir pendant les périodes de congestion. Lorsque la file d'attente de sortie est pleine et que la file d'attente est en vigueur, les paquets sont abandonnés jusqu'à ce que la congestion soit éliminée et que la file d'attente ne soit plus pleine. La perte de queue est le type le plus basique d'évitement de congestion et ne prend en compte aucun paramètre QoS.

Le Catalyst 6000 a mis en oeuvre une version avancée de l'évitement de congestion de la déconnexion de la queue qui supprime tous les paquets avec une certaine CoS lorsqu'un certain pourcentage de remplissage de la mémoire tampon est atteint. Avec la perte de queue pondérée, vous pouvez définir un ensemble de seuils et associer une CoS à chaque seuil. Dans l'exemple de cette section, il existe quatre seuils possibles. Les définitions de chaque seuil sont les suivantes :

- Le seuil 1 est atteint lorsque 50 % du tampon est rempli. CoS 0 et 1 sont affectés à ce seuil.
- Le seuil 2 est atteint lorsque 60 % de la mémoire tampon est remplie. Les CoS 2 et 3 sont affectés à ce seuil.
- Le seuil 3 est atteint lorsque 80 % du tampon est rempli. Les CoS 4 et 5 sont affectées à ce seuil.
- Le seuil 4 est atteint lorsque 100 % du tampon est rempli. Les CoS 6 et 7 sont affectées à ce seuil.

Dans le diagramme de la [Figure 1](#), tous les paquets avec une CoS de 0 ou 1 sont abandonnés si la mémoire tampon est remplie à 50 %. Tous les paquets avec une CoS de 0, 1, 2 ou 3 sont abandonnés si les tampons sont remplis à 60 %. Les paquets avec une CoS de 6 ou 7 sont abandonnés lorsque les tampons sont complètement remplis.

Figure 1



Remarque : Dès que le remplissage de la mémoire tampon tombe en dessous d'un certain seuil, les paquets avec la CoS associée ne sont plus abandonnés.

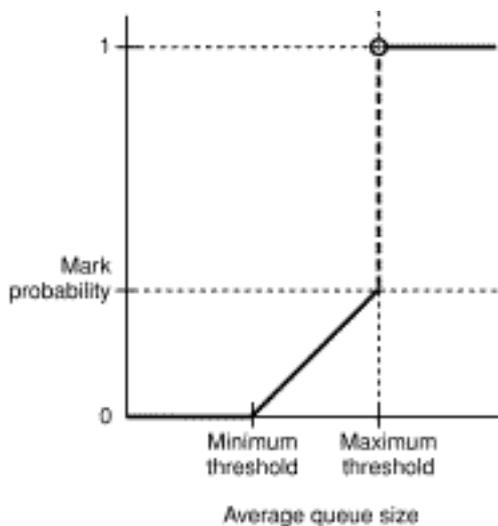
Détection précoce aléatoire et détection précoce pondérée aléatoire

La détection WRED (Weighted Random Early Detection) est un mécanisme d'évitement de congestion qui supprime aléatoirement les paquets avec une certaine priorité IP lorsque les tampons atteignent un seuil de remplissage défini. WRED est une combinaison de ces deux fonctionnalités :

- Déplacement de la queue
- Détection précoce aléatoire (RED)

RED n'est pas sensible à la priorité ou à la CoS. RED utilise l'un des seuils uniques lorsque la valeur de seuil de la mémoire tampon est remplie. RED commence à supprimer des paquets aléatoirement (mais pas tous les paquets, comme dans la chute de queue) jusqu'à ce que le seuil maximum (max) soit atteint. Une fois le seuil maximal atteint, tous les paquets sont abandonnés. La probabilité qu'un paquet soit abandonné augmente linéairement avec l'augmentation du remplissage de la mémoire tampon au-dessus du seuil. Le diagramme de la [Figure 2](#) montre la probabilité de perte de paquets :

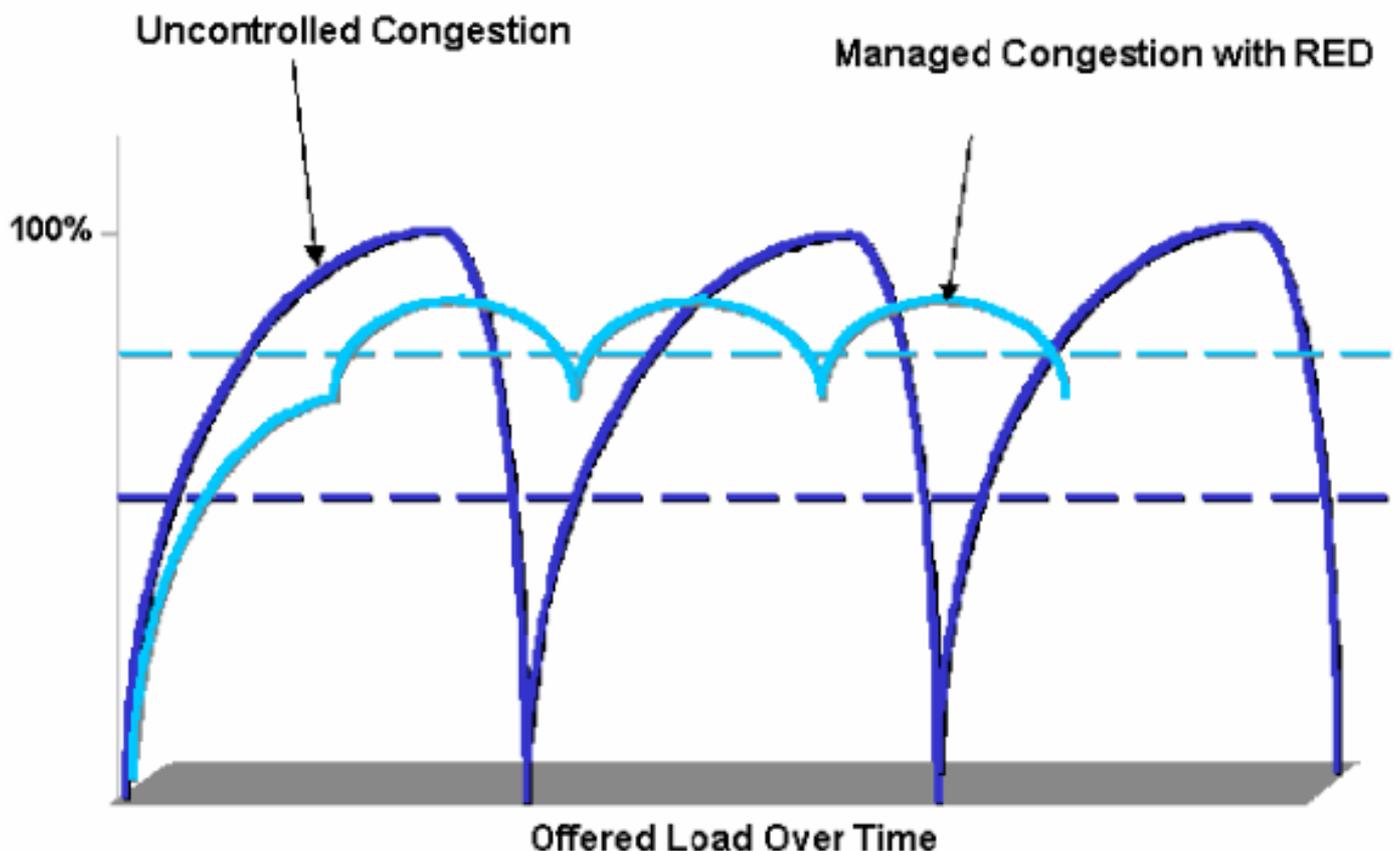
Figure 2 - Probabilité d'abandon de paquets



Note : La probabilité de la marque dans ce diagramme est réglable en ROUGE, ce qui signifie que la pente de la probabilité de chute linéaire est réglable.

RED et WRED sont des mécanismes très utiles d'évitement de congestion pour le trafic basé sur TCP. Pour les autres types de trafic, RED n'est pas très efficace. En effet, RED tire parti du mécanisme de fenêtrage utilisé par TCP pour gérer l'encombrement. RED évite l'encombrement typique qui se produit sur un routeur lorsque plusieurs sessions TCP passent par le même port de routeur. Ce mécanisme est appelé synchronisation réseau globale. Le diagramme de la [Figure 3](#) montre comment le ROUGE a un effet de lissage sur la charge :

Figure 3 - ROUGE pour éviter les encombrements



Pour plus d'informations sur la façon dont le routeur RED peut réduire la congestion et lisser le trafic à travers le routeur, référez-vous à la section [Comment le routeur interagit avec le protocole TCP](#) du document [Aperçu de l'évitement de congestion](#).

WRED est similaire à RED en ce que les deux définissent des seuils minimum (min) et, lorsque ces seuils minimum sont atteints, les paquets sont abandonnés de manière aléatoire. WRED définit également certains seuils max et, lorsque ces seuils max sont atteints, tous les paquets sont abandonnés. WRED est également compatible CoS, ce qui signifie qu'une ou plusieurs valeurs CoS sont ajoutées à chaque paire seuil minimum/seuil max.. Lorsque le seuil minimal est dépassé, les paquets sont abandonnés aléatoirement avec la CoS affectée. Considérez cet exemple avec deux seuils dans la file d'attente :

- Les CoS 0 et 1 sont affectés au seuil minimal 1 et au seuil maximal 1. Le seuil minimal 1 est défini à 50 % du remplissage de la mémoire tampon et le seuil maximal 1 à 80 %.
- Les CoS 2 et 3 sont affectés au seuil minimal 2 et au seuil maximal 2. Le seuil minimal 2 est défini sur 70 % du remplissage de la mémoire tampon et le seuil maximal 2 sur 100 %.

Dès que le tampon dépasse le seuil minimal 1 (50 %), les paquets avec CoS 0 et 1 commencent à être abandonnés de manière aléatoire. Plus de paquets sont abandonnés à mesure que l'utilisation de la mémoire tampon augmente. Si le seuil minimum 2 (70 %) est atteint, les paquets avec CoS 2 et 3 commencent à être abandonnés de manière aléatoire.

Remarque : À ce stade, la probabilité de perte pour les paquets avec CoS 0 et 1 est beaucoup plus élevée que la probabilité de perte pour les paquets avec CoS 2 ou CoS 3.

Chaque fois que le seuil maximal 2 est atteint, les paquets avec CoS 0 et 1 sont tous abandonnés, tandis que les paquets avec CoS 2 et 3 continuent d'être abandonnés de manière aléatoire. Enfin, lorsque 100 % est atteint (seuil max. 2), tous les paquets avec CoS 2 et 3 sont abandonnés.

Les diagrammes des [figures 4](#) et [5](#) illustrent un exemple de ces seuils :

Figure 4 - WRED avec deux ensembles de seuils minimum et de seuils maximum (deux services)

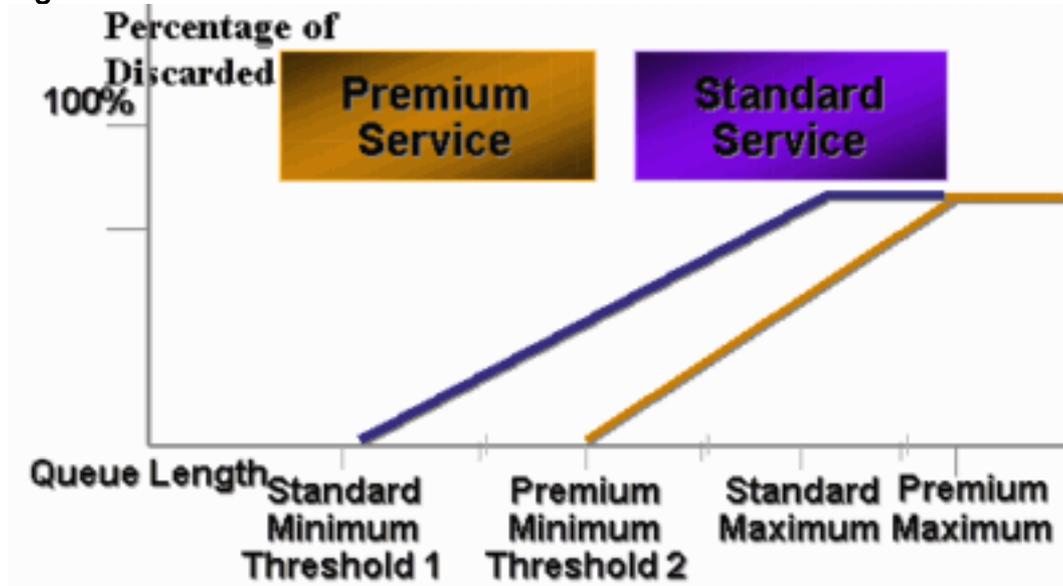
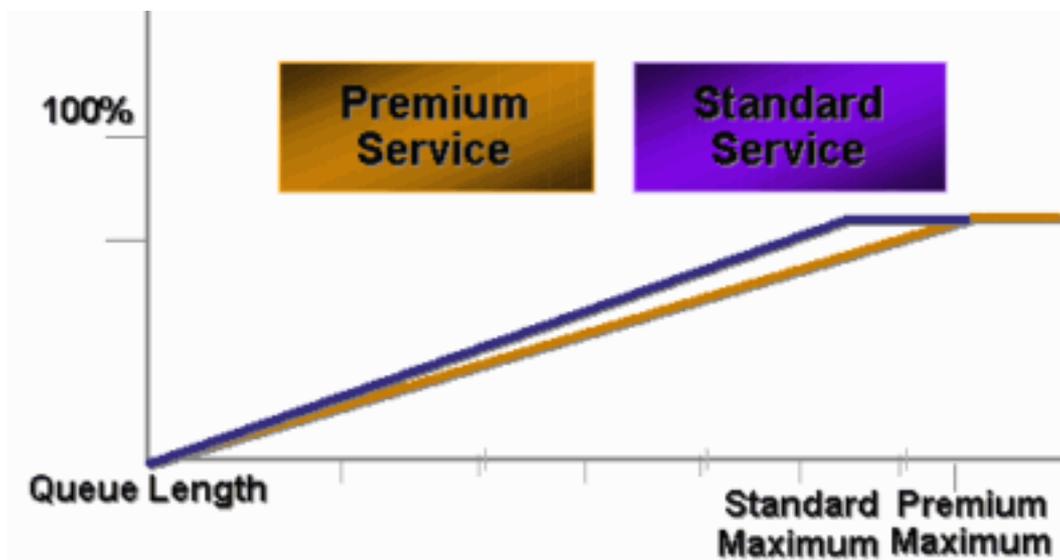


Figure 5 - WRED avec deux ensembles de services, mais les deux seuils minimaux sont égaux à 0



La première mise en oeuvre de CatOS de WRED a uniquement défini le seuil max, alors que le seuil min a été codé en dur à 0 %. La partie inférieure du diagramme de la [Figure 5](#) met en évidence le comportement qui en résulte.

Remarque : La probabilité de perte d'un paquet est toujours non nulle, car cette probabilité est toujours supérieure au seuil minimum. Ce comportement a été corrigé dans les versions 6.2 et ultérieures du logiciel.

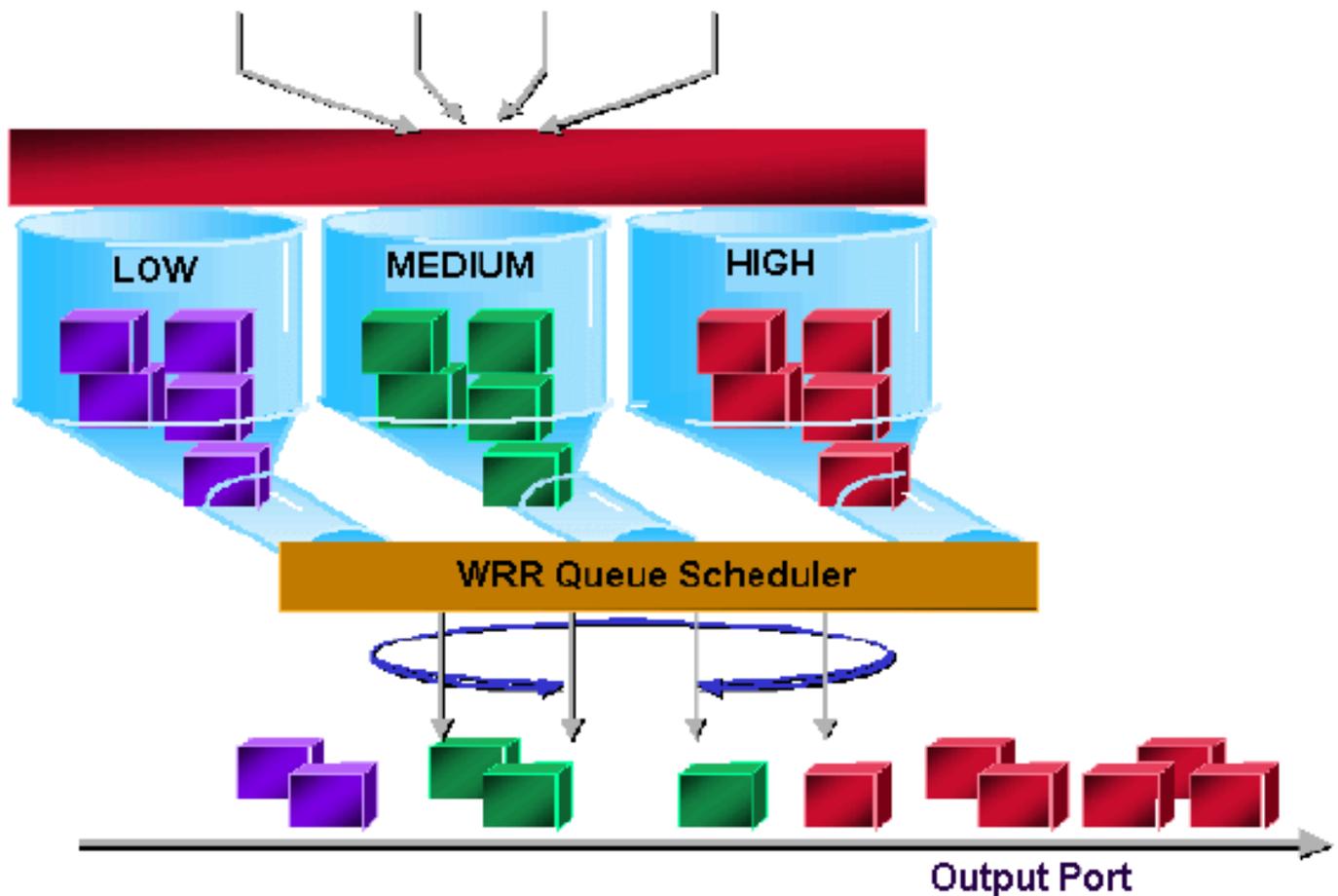
Rotation circulaire pondérée

Le WRR (Weighted Round-Robin) est un autre mécanisme de planification des sorties sur le Catalyst 6000. WRR fonctionne entre deux files d'attente ou plus. Les files d'attente pour WRR sont vidées de manière circulaire et vous pouvez configurer le poids de chaque file d'attente. Par défaut, les ports ont deux files d'attente WRR sur le Catalyst 6000. La valeur par défaut est :

- Servir la file d'attente WRR hautement prioritaire 70 % du temps
- Pour traiter la file d'attente WRR de faible priorité 30 % du temps

Le diagramme de la [Figure 6](#) montre un WRR qui comporte trois files d'attente desservies de manière WRR. La file d'attente de priorité élevée (paquets rouges) envoie plus de paquets que les deux autres files d'attente :

Figure 6 - Planification des sorties : WRR



Remarque : la plupart des cartes de ligne 6500 implémentent le WRR par bande passante. Cette implémentation du WRR par bande passante signifie que chaque fois que le planificateur autorise une file d'attente à transmettre des paquets, un certain nombre d'octets sont autorisés à être transmis. Ce nombre d'octets peut représenter plusieurs paquets. Par exemple, si vous envoyez 5 120 octets en un tour, vous pouvez envoyer trois paquets de 1 518 octets, pour un total de 4 554 octets. Les octets en excès sont perdus ($5120 - 4554 = 566$ octets). Par conséquent, avec un poids extrême (comme 1 % pour la file d'attente 1 et 99 % pour la file d'attente 2), le poids exact configuré peut ne pas être atteint. Cette incapacité à atteindre le poids exact est souvent le cas pour les paquets de plus grande taille.

Certaines cartes de ligne de nouvelle génération, telles que le 6548-RJ-45, surmontent cette limite par la mise en oeuvre de la méthode DWRR. Le DWRR transmet à partir des files d'attente, mais il n'affame pas la file d'attente de faible priorité. Le DWRR suit la file d'attente de faible priorité en cours de transmission et compense le prochain cycle.

[File d'attente de priorité stricte](#)

Un autre type de file d'attente dans le Catalyst 6000, une file d'attente de priorité stricte, est toujours vidé en premier. Dès qu'un paquet est dans la file d'attente de priorité stricte, il est envoyé.

Les files d'attente WRR ou WRED ne sont vérifiées qu'après vidage de la file d'attente de priorité stricte. Une fois chaque paquet transmis à partir de la file d'attente WRR ou WRED, la file d'attente de priorité stricte est vérifiée et vidée, si nécessaire.

Remarque : toutes les cartes de ligne avec un type de file d'attente similaire à 1p2q1t, 1p3q8t et 1p7q8t utilisent DWRR. D'autres cartes de ligne utilisent le WRR standard.

Fonctionnalité de mise en file d'attente de sortie de différentes cartes de ligne sur le Catalyst 6000

Fonctionnalités de la commande show port

Si vous n'êtes pas sûr de la fonctionnalité de mise en file d'attente d'un port, vous pouvez émettre la commande **show port ability**. Voici la sortie de la commande sur une carte de ligne WS-X6408-GBIC :

```
Model                WS-X6408-GBIC
Port                 4/1
Type                 No GBIC
Speed                1000
Duplex               full
Trunk encap type     802.1Q, ISL
Trunk mode           on, off, desirable, auto, nonegotiate
Channe                yes
Broadcast suppression percentage(0-100)
Flow control         receive- (off, on, desired) , send- (off, on, desired)
Security              yes
MembersshIP          static, dynamic
Fast start           yes
QOS scheduling        rx- (1q4t) , tx- (2q2t)
CoS rewrite           yes
ToS rewrite           DSCP
UDLD                 yes
SPAN                 source, destination
COPS port group      none
```

Ce port possède un type de sortie de mise en file d'attente appelé 2q2t.

Comprendre la fonctionnalité de mise en file d'attente d'un port

Plusieurs types de files d'attente sont disponibles sur les commutateurs Catalyst 6500/6000. Les tableaux de cette section peuvent devenir incomplets lorsque de nouvelles cartes de ligne sont publiées. Les nouvelles cartes de ligne peuvent introduire de nouvelles combinaisons de files d'attente. Pour obtenir une description actuelle de toutes les files d'attente disponibles pour les modules de commutation Catalyst 6500/6000, reportez-vous à la section *Configuration de la qualité de service* pour votre version CatOS de [la documentation logicielle de la gamme Catalyst 6500](#).

Remarque : Cisco Communication Media Module (CMM) ne prend pas en charge toutes les fonctions QoS. Vérifiez les notes de version de votre version logicielle spécifique afin de déterminer les fonctionnalités prises en charge.

Ce tableau explique la notation de l'architecture QoS du port :

Tx 1/ Rx 2 ide	Notation de file d'attente	Nombre de files d'attente	File d'attente prioritaire	Nombre de files d'attente WRR	Nombre et type de seuil pour les files d'attente WRR
Tx	2q2t	2	—	2	2 lignes de

					queue configurables
Tx	1p2q2t	3	1	2	2 WRED configurables
Tx	1p3q1t	4	1	3	1 WRED configurable
Tx	1p2q1t	3	1	2	1 WRED configurable
Rx	1q4t	1	—	1	4 lignes de queue configurables
Rx	1p1q4t	2	1	1	4 lignes de queue configurables
Rx	1p1q0t	2	1	1	Non configurable
Rx	1p1q8t	2	1	1	8 WRED configurables

¹ Tx = transmission.

² Rx = réception.

Ce tableau répertorie tous les modules et les types de file d'attente du côté Rx et Tx de l'interface ou du port :

module	Files d'attente Rx	Files d'attente Tx
WS-X6K-S2-PFC2	1p1q4t	1p2q2t
WS-X6K-SUP1A-2GE	1p1q4t	1p2q2t
WS-X6K-SUP1-2GE	1q4t	2q2t
WS-X6501-10GEX4	1p1q8t	1p2q1t
WS-X6502-10GE	1p1q8t	1p2q1t
WS-X6516-GBIC	1p1q4t	1p2q2t
WS-X6516-GE-TX	1p1q4t	1p2q2t
WS-X6416-GBIC	1p1q4t	1p2q2t
WS-X6416-GE-MT	1p1q4t	1p2q2t
WS-X6316-GE-TX	1p1q4t	1p2q2t
WS-X6408A-GBIC	1p1q4t	1p2q2t
WS-X6408-GBIC	1q4t	2q2t
WS-X6524-100FX-MM	1p1q0t	1p3q1t
WS-X6324-100FX-SM	1q4t	2q2t
WS-X6324-100FX-MM	1q4t	2q2t
WS-X6224-100FX-MT	1q4t	2q2t
WS-X6548-RJ-21	1p1q0t	1p3q1t
WS-X6548-RJ-45	1p1q0t	1p3q1t

WS-X6348-RJ-21	1q4t	2q2t
WS-X6348-RJ21V	1q4t	2q2t
WS-X6348-RJ-45	1q4t	2q2t
WS-X6348-RJ-45V	1q4t	2q2t
WS-X6148-RJ-45V	1q4t	2q2t
WS-X6148-RJ21V	1q4t	2q2t
WS-X6248-RJ-45	1q4t	2q2t
WS-X6248A-TEL	1q4t	2q2t
WS-X6248-TEL	1q4t	2q2t
WS-X6024-10FL-MT	1q4t	2q2t

[Créer une qualité de service sur Catalyst 6500/6000](#)

Trois champs du Catalyst 6500/6000 sont utilisés pour définir la qualité de service :

- **Priorité IP** : les trois premiers bits du champ ToS (Type of Service) de l'en-tête IP
- **Le point de code de services différenciés (DSCP)** : les six premiers bits du champ ToS de l'en-tête IP
- **CoS** : les trois bits utilisés au niveau de la couche 2 (L2) Ces trois bits font partie de l'en-tête ISL (Inter-Switch Link) ou se trouvent à l'intérieur de la balise IEEE 802.1Q (dot1q). Il n'y a pas de CoS dans un paquet Ethernet non étiqueté.

[Mécanisme de planification des sorties sur Catalyst 6500/6000](#)

Lorsqu'une trame est envoyée à partir du bus de données à transmettre, la CoS du paquet est le seul paramètre pris en compte. Le paquet passe ensuite par un planificateur, qui choisit la file d'attente dans laquelle le paquet est placé. Par conséquent, n'oubliez pas que la planification des sorties et tous les mécanismes abordés dans ce document sont uniquement compatibles CoS.

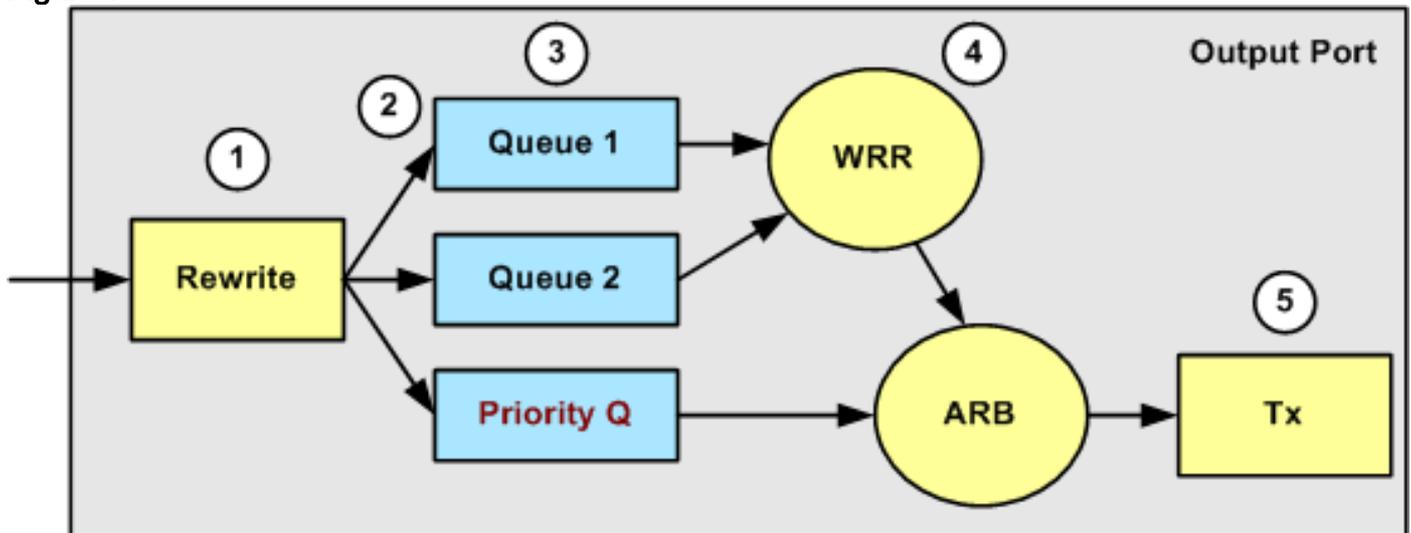
Le Catalyst 6500/6000 avec une carte MSFC (Multilayer Switch Feature Card) utilise un DSCP interne afin de classer le paquet. Le Catalyst 6500/6000 configuré avec QoS activé attribue une valeur DSCP lorsque la décision de transfert est prise au niveau PFC. Ce DSCP est affecté à tout paquet, qui inclut des paquets non IP, et est mappé à la CoS afin d'activer la planification de sortie. Vous pouvez configurer le mappage entre les valeurs DSCP et CoS sur Catalyst 6500/6000. Si vous laissez la valeur par défaut, vous pouvez dériver la CoS du DSCP. La formule est :

$$DSCP_value / 8$$

En outre, la valeur DSCP est mappée dans la CoS du paquet sortant, si le paquet est un paquet IP étiqueté ISL ou dot1q (VLAN non natif). La valeur DSCP est également inscrite dans le champ ToS de l'en-tête IP.

Le diagramme de la [Figure 7](#) montre une file d'attente 1p2q2t. Les files d'attente WRR sont vidées à l'aide du planificateur WRR. Il existe également un arbitre qui vérifie entre chaque paquet des files d'attente WRR afin de déterminer s'il y a quelque chose dans la file d'attente de priorité stricte.

Figure 7



1. Le champ ToS est réécrit dans l'en-tête IP et le champ CoS 802.1p/ISL.
2. La file d'attente de planification et le seuil sont sélectionnés sur la base de la CoS, via une carte configurable.
3. Chaque file d'attente a une taille et des seuils configurables, et certaines files d'attente ont WRED.
4. La mise en file d'attente utilise le WRR entre deux files d'attente.
5. L'encapsulation sortante peut être dot1q, ISL ou none.

[Configuration, surveillance et planification des sorties sur Catalyst 6500/6000](#)

[Configuration par défaut pour QoS sur Catalyst 6500/6000](#)

Cette section fournit un exemple de sortie de la configuration QoS par défaut sur un Catalyst 6500/6000, en plus des informations sur la signification de ces valeurs et sur la manière dont vous pouvez les régler.

La QoS est désactivée par défaut lorsque vous émettez cette commande :

```
set qos disable
```

Les commandes de cette liste indiquent l'affectation par défaut pour chaque CoS dans un port 2q2t. La CoS 0 et 1 de la file d'attente 1 sont affectés à son premier seuil et la CoS 2 et 3 à son deuxième seuil. Les CoS 4 et 5 de la file d'attente 2 sont affectés à son premier seuil et les CoS 6 et 7 à son deuxième seuil :

```
set qos map 2q2t tx 1 1 cos 0
```

```
set qos map 2q2t tx 1 1 cos 1
```

```
set qos map 2q2t tx 1 2 cos 2
```

```
set qos map 2q2t tx 1 2 cos 3
set qos map 2q2t tx 2 1 cos 4
set qos map 2q2t tx 2 1 cos 5
set qos map 2q2t tx 2 2 cos 6
set qos map 2q2t tx 2 2 cos 7
```

Ces commandes affichent le niveau de seuil par défaut sur un port 2q2t pour chaque file d'attente :

```
set qos drop-threshold 2q2t tx queue 1 80 100
set qos drop-threshold 2q2t tx queue 2 80 100
```

Vous pouvez affecter le poids par défaut à chacune des files d'attente WRR. Émettez cette commande afin d'attribuer les poids par défaut pour la file d'attente 1 et la file d'attente 2 :

Remarque : La file d'attente de priorité faible est traitée 5/260 % du temps et la file d'attente de priorité élevée est traitée 255/260 % du temps.

```
set qos wrr 2q2t 5 255
```

La disponibilité totale de la mémoire tampon est divisée entre les deux files d'attente. La file d'attente de faible priorité est correctement affectée à 80 % des tampons disponibles, car il s'agit de la file d'attente qui est la plus susceptible d'avoir des paquets mis en mémoire tampon et qui est en attente pendant un certain temps. Émettez cette commande afin de définir la disponibilité :

```
set qos txq-ratio 2q2t 80 20
```

Vous pouvez afficher des paramètres similaires pour le port 1p2q2t dans cette configuration :

```
set qos map 1p2q2t tx 1 1 cos 0
set qos map 1p2q2t tx 1 1 cos 1
set qos map 1p2q2t tx 1 2 cos 2
set qos map 1p2q2t tx 1 2 cos 3
set qos map 1p2q2t tx 2 1 cos 4
set qos map 1p2q2t tx 3 1 cos 5
set qos map 1p2q2t tx 2 1 cos 6
set qos map 1p2q2t tx 2 2 cos 7
```

```
set qos wrr 1p2q2t 5 255
```

```
set qos txq-ratio 1p2q2t 70 15 15
```

```
set qos wred 1p2q2t tx queue 1 80 100
```

```
set qos wred 1p2q2t tx queue 2 80 100
```

Remarque : Par défaut, CoS 5 (trafic vocal) est affecté à la file d'attente de priorité stricte.

Configuration

La première étape de configuration consiste à activer la QoS. N'oubliez pas que la QoS est désactivée par défaut. Lorsque QoS est désactivé, le mappage CoS n'est pas pertinent. Il y a une file d'attente unique qui est servie en tant que FIFO, et tous les paquets y sont abandonnés.

```
bratan> (enable) set qos enable
```

```
QoS is enabled
```

```
bratan> (enable) show qos status
```

```
QoS is enabled on this switch
```

La valeur CoS doit être affectée à la file d'attente ou au seuil pour tous les types de file d'attente. Le mappage défini pour un type de port 2q2t n'est appliqué à aucun port 1p2q2t. En outre, le mappage qui est fait pour 2q2t est appliqué à tous les ports qui ont un mécanisme de mise en file d'attente 2q2t. Émettez la commande suivante :

```
set qos map queue_type tx Q_number threshold_number cos value
```

Remarque : Les files d'attente sont toujours numérotées pour commencer par la file d'attente de priorité la plus basse possible et se terminer par la file d'attente de priorité stricte qui est disponible. Voici un exemple :

- La file d'attente 1 est la file d'attente WRR de faible priorité
- La file d'attente 2 est la file d'attente WRR de haute priorité
- La file d'attente 3 est la file d'attente de priorité stricte

Vous devez répéter cette opération pour tous les types de files d'attente. Sinon, conservez l'affectation CoS par défaut. Voici un exemple pour 1p2q2t :

Configuration

```
set qos map 1p2q2t tx 1 1 cos 0
```

```
!--- This is the low-priority WRR queue threshold 1, CoS 0 and 1. set qos map 1p2q2t tx 1 1 cos 1 and 1
```

```
set qos map 1p2q2t tx 1 2 cos 2
```

```
!--- This is the low-priority WRR queue threshold 2, CoS 2 and 3. set qos map 1p2q2t tx 1 2 cos 3 and 3
```

```
set qos map 1p2q2t tx 2 1 cos 4
!--- This is the high-priority WRR queue threshold 1, CoS 4. set qos map 1p2q2t tx 3 1 cos 5
!--- This is the strict priority queue, CoS 5. set qos map 1p2q2t tx 2 1 cos 6
!--- This is the high-priority WRR queue threshold 2, CoS 6. set qos map 1p2q2t tx 2 2 cos 7 and
7
```

Sortie de console

```
tamer (enable) set qos map 1p2q2t tx 1 1 cos 0
```

QoS tx priority queue and threshold mapped to cos successfully

Vous devez configurer le poids WRR pour les deux files d'attente WRR. Émettez la commande suivante :

```
set qos wrr Q_type weight_1 weight_2
```

Weight_1 se rapporte à la file d'attente 1, qui doit être la file d'attente WRR de faible priorité. Le *poids_1* doit toujours être inférieur au *poids_2*. Le poids peut prendre n'importe quelle valeur comprise entre 1 et 255. Vous pouvez affecter le pourcentage avec ces formules :

- File d'attente 1 :

$$\text{weight_1} / (\text{weight_1} + \text{weight_2})$$

- File d'attente 2 :

$$\text{weight_2} / (\text{weight_1} + \text{weight_2})$$

Vous devez également définir le poids des différents types de files d'attente. Le poids n'a pas besoin d'être le même. Par exemple, pour 2q2t, où la file d'attente 1 est desservie 30 % du temps et la file d'attente 2 70 % du temps, vous pouvez émettre cette commande afin de définir le poids :

```
set qos wrr 2q2t 30 70
!--- This ensures that the high-priority WRR queue is served 70 percent of the time !--- and
that the low-priority WRR queue is served 30 percent of the time.
```

Sortie de console

```
tamer (enable) set qos wrr 2q2t 30 70
```

QoS wrr ratio is set successfully

Vous devez également définir le ratio de file d'attente de transmission, qui fait référence à la façon dont les tampons sont répartis entre les différentes files d'attente. Émettez la commande suivante :

```
set qos txq-ratio port_type queue1_val queue2_val ... queueN_val
```

Remarque : Si vous avez trois files d'attente (1p2q2t), vous devez définir la file d'attente WRR de priorité élevée et la file d'attente de priorité stricte au même niveau pour des raisons matérielles.

Configuration

```
set qos txq-ratio 1p2q2t 70 15 15
```

!--- This gives 70 percent of the buffer of all 1p2q2t ports to the low-priority WRR !--- queue and gives 15 percent to each of the other two queues. set qos txq-ratio 2q2t 80 20 !--- This gives 80 percent of the buffer to the low-priority queue, !--- and gives 20 percent of the buffer to the high-priority queue.

Sortie de console

```
tamer (enable) set qos txq-ratio 1p2q2t 70 15 20
```

Queue ratio values must be in range of 1-99 and add up to 100

```
Example: set qos txq-ratio 2q2t 20 80
```

```
tamer (enable) set qos txq-ratio 1p2q2t 70 30 30
```

Queue ratio values must be in range of 1-99 and add up to 100

```
Example: set qos txq-ratio 2q2t 20 80
```

```
tamer (enable) set qos txq-ratio 1p2q2t 80 10 10
```

```
QoS txq-ratio is set successfully
```

Comme le montre cette sortie de console, la somme des valeurs de file d'attente doit être égale à 100. Laissez la plus grande partie des tampons pour la file d'attente WRR de faible priorité, car cette file d'attente a besoin de la plus grande mise en mémoire tampon. Les autres files d'attente sont traitées avec une priorité plus élevée.

La dernière étape consiste à configurer le niveau de seuil pour la file d'attente WRED ou pour la file d'attente de suppression de queue. Émettez les commandes suivantes :

```
set qos wred port_type [tx] queue q_num thr1 thr2 ... thrn
```

```
set qos drop-threshold port_type tx queue q_num thr1 ... thr2
```

Configuration

```
set qos drop-threshold 2q2t tx queue 1 50 80
```

!--- For low-priority queues in the 2q2t port, the first threshold is defined at 50 !--- percent and the second threshold is defined at 80 percent of buffer filling. set qos drop-threshold 2q2t tx queue 2 40 80

!--- For high-priority queues in the 2q2t port, the first threshold is defined at 40 !--- percent and the second threshold is defined at 80 percent of buffer filling. set qos wred 1p2q2t tx queue 1 50 90

!--- The commands for the 1p2q2t port are identical. set qos wred 1p2q2t tx queue 2 40 80

Sortie de console

```
tamer (enable) set qos drop-threshold 2q2t tx queue 1 50 80
```

Transmit drop thresholds for queue 1 set at 50% 80%

```
tamer (enable) set qos drop-threshold 2q2t tx queue 2 40 80
```

Transmit drop thresholds for queue 2 set at 40% 80%

```
tamer (enable) set qos wred 1p2q2t tx queue 1 50 90
```

WRED thresholds for queue 1 set to 50 and 90 on all WRED-capable 1p2q2t ports

```
tamer (enable) set qos wred 1p2q2t tx queue 2 40 80
```

WRED thresholds for queue 2 set to 40 and 80 on all WRED-capable 1p2q2t ports

La commande **set qos wred 1p2q2t tx queue 2 40 80** fonctionne conjointement avec la CoS pour le mappage de seuil. Par exemple, lorsque vous émettez les commandes dans la liste ci-dessous, vous assurez que, sur le port 1p2q2t dans la direction de transmission, les paquets avec CoS 0, 1, 2 et 3 sont envoyés dans la première file d'attente (la file d'attente WRR basse). Lorsque les tampons de cette file d'attente sont remplis à 50 %, WRED commence à abandonner les paquets avec CoS 0 et 1. Les paquets avec CoS 2 et 3 ne sont supprimés que lorsque les tampons de la file d'attente sont remplis à 90 %.

```
set qos map 1p2q2t tx 1 1 cos 0
```

```
set qos map 1p2q2t tx 1 1 cos 1
```

```
set qos map 1p2q2t tx 1 2 cos 2
```

```
set qos map 1p2q2t tx 1 2 cos 3
```

```
set qos wred 1p2q2t tx queue 1 50 90
```

[Surveillance de la planification des sorties et vérification de la configuration](#)

Une commande simple à utiliser afin de vérifier la configuration d'exécution actuelle pour la planification de sortie d'un port est **show qos info runtime mod/port**. La commande affiche les informations suivantes :

- Type de mise en file d'attente sur le port
- Mappage de CoS aux différentes files d'attente et aux différents seuils
- Le partage de tampon
- Poids WRR

Dans cet exemple, les valeurs sont de 20 % WRR pour la file d'attente 1 et de 80 % WRR pour la file d'attente 2 :

```
tamer (enable) show qos info runtime 1/1
```

```
Run time setting of QoS:
```

```
QoS is enabled
```

```
Policy Source of port 1/1: Local
```

```
Tx port type of port 1/1 : 1p2q2t
```

```
Rx port type of port 1/1 : 1p1q4t
```

```
Interface type: port-based
```

```
ACL attached:
```

```
The qos trust type is set to untrusted
```

Default CoS = 0

Queue and Threshold Mapping for 1p2q2t (tx):

Queue	Threshold	CoS
1	1	0 1
1	2	2 3
2	1	4 6
2	2	7
3	1	5

Queue and Threshold Mapping for 1p1q4t (rx):

All packets are mapped to a single queue

Rx drop thresholds:

Rx drop thresholds are disabled

Tx drop thresholds:

Tx drop-thresholds feature is not supported for this port type

Tx WRED thresholds:

Queue #	Thresholds - percentage (* abs values)
1	80% (249088 bytes) 100% (311168 bytes)
2	80% (52480 bytes) 100% (61440 bytes)

Queue Sizes:

Queue #	Sizes - percentage (* abs values)
1	70% (311296 bytes)
2	15% (65536 bytes)
3	15% (65536 bytes)

WRR Configuration of ports with speed 1000Mbps:

Queue #	Ratios (* abs values)
1	20 (5120 bytes)
2	80 (20480 bytes)

(*) Runtime information may differ from user configured setting due to hardware granularity.

tamer (enable)

Dans l'exemple suivant, notez que les pondérations WRR ne sont pas la valeur par défaut de 1. Les poids ont été définis sur les valeurs 20 pour la file d'attente 1 et 80 pour la file d'attente 2. Cet exemple utilise un générateur de trafic pour envoyer 2 Go de trafic à un Catalyst 6000. Ces 2 Gbit de trafic doivent sortir par le port 1/1. Comme le port 1/1 est surabonné, de nombreux paquets sont abandonnés (1 Gbit/s). La commande **show mac** montre qu'il y a beaucoup de pertes de sortie :

tamer (enable) **show mac 1/1**

Port	Rcv-Unicast	Rcv-Multicast	Rcv-Broadcast
1/1	0	1239	0

Port	Xmit-Unicast	Xmit-Multicast	Xmit-Broadcast
1/1	73193601	421	0

Port	Rcv-Octet	Xmit-Octet
1/1	761993	100650803690

MAC	Dely-Exced	MTU-Exced	In-Discard	Out-Discard
1/1	0	-	0	120065264

Last-Time-Cleared

Considérez les paquets abandonnés. Voici comment le modèle de trafic suggéré est fractionné :

- 1 Go de trafic avec priorité IP 0
- 250 Mo de trafic avec priorité IP 4
- 250 Mo de trafic avec priorité IP 5
- 250 Mo de trafic avec priorité IP 6
- 250 Mo de trafic avec priorité IP 7

Selon le mappage CoS, ce trafic est envoyé :

- 1 Gb de trafic vers la file d'attente 1 seuil 1
- 0 Mo de trafic vers la file d'attente 1 seuil 2
- 500 Mo de trafic vers la file d'attente 2 seuil 1
- 250 Mo de trafic vers la file d'attente 2 seuil 2
- 250 Mo de trafic vers la file d'attente 3 (file d'attente de priorité stricte)

Le commutateur doit faire confiance au trafic reçu afin que la priorité IP entrante soit préservée dans le commutateur et utilisée pour mapper la valeur CoS pour la planification de sortie.

Remarque : La priorité IP par défaut au mappage CoS est la priorité IP égale la CoS.

Émettez la commande **show qos stat 1/1** afin de voir les paquets qui ont été abandonnés et le pourcentage approximatif :

- À ce stade, aucun paquet n'est abandonné dans la file d'attente 3 (CoS 5).
- 91,85 % des paquets abandonnés sont des paquets CoS 0 dans la file d'attente 1.
- 8 % des paquets abandonnés sont CoS 4 et 6 dans la file d'attente 2, seuil 1.
- 0,15 % des paquets abandonnés sont CoS 7 dans la file d'attente 2, seuil 2.

Cette sortie illustre l'utilisation de la commande :

```
tamer (enable) show qos stat 1/1
```

```
Tx port type of port 1/1 : 1p2q2t
```

```
Q3T1 statistics are covered by Q2T2.
```

```
Q #      Threshold #:Packets dropped
```

```
----
```

```
1      1:110249298 pkts, 2:0 pkts
```

```
2      1:9752805 pkts, 2:297134 pkts
```

```
3      1:0 pkts
```

```
Rx port type of port 1/1 : 1p1q4t
```

```
Rx drop threshold counters are disabled for untrusted ports
```

```
Q #      Threshold #:Packets dropped
```

```
----
```

```
1      1:0 pkts, 2:0 pkts, 3:0 pkts, 4:0 pkts
```

```
2      1:0 pkts
```

Si vous redéfinissez le poids du WRR sur la valeur par défaut une fois les compteurs effacés, seulement 1 % des paquets abandonnés se trouvent dans la file d'attente 2 au lieu des 8 % qui apparaissaient précédemment :

Remarque : la valeur par défaut est 5 pour la file d'attente 1 et 255 pour la file d'attente 2.

```
tamer (enable) show qos stat 1/1
```

```

TX port type of port 1/1 : 1p2q2t
Q3T1 statistics are covered by Q2T2
Q #                Threshold #:Packets dropped
---                -
1                  1:2733942 pkts, 2:0 pkts
2                  1:28890 pkts, 2:6503 pkts
3                  1:0 pkts
Rx port type of port 1/1 : 1p1q4t
Rx drop threshold counters are disabled for untrusted ports
Q #                Threshold #:Packets dropped
---                -
1                  1:0 pkts, 2:0 pkts, 3:0 pkts, 4:0 pkts
2                  1:0 pkts

```

Utiliser la planification des sorties pour réduire les délais et la gigue

L'exemple de la section [Surveillance de la planification de sortie et vérification de la configuration](#) montre l'avantage de la mise en oeuvre de la planification de sortie, qui évite une perte de trafic VoIP ou critique en cas de surabonnement au port de sortie. La sursouscription se produit rarement dans un réseau normal, en particulier sur une liaison Gigabit. En général, la sursouscription ne se produit que pendant les heures de pointe du trafic ou pendant les rafales de trafic dans un laps de temps très court.

Même sans surabonnement, la planification des sorties peut être très utile dans un réseau où la QoS est mise en oeuvre de bout en bout. La planification des sorties permet de réduire les délais et la gigue. Cette section fournit des exemples de la façon dont la planification des sorties peut aider à réduire les délais et la gigue.

Réduire le délai

Le délai d'un paquet est augmenté du temps "perte de" dans la mémoire tampon de chaque commutateur pendant l'attente de transmission. Par exemple, un petit paquet vocal avec une CoS de 5 est envoyé depuis un port lors d'une sauvegarde ou d'un transfert de fichiers volumineux. Si vous n'avez pas de QoS pour le port de sortie et si vous supposez que le petit paquet vocal est mis en file d'attente après 10 gros paquets de 1 500 octets, vous pouvez facilement calculer le temps de vitesse Gigabit nécessaire pour transmettre les 10 grands paquets :

$(10 \times 1500 \times 8) = 120,000$ bits that are transmitted in 120 microseconds

Si ce paquet doit traverser huit ou neuf commutateurs pendant qu'il traverse le réseau, un délai d'environ 1 ms peut en résulter. Ce montant compte uniquement les retards dans la file d'attente de sortie du commutateur qui est croisée dans le réseau.

Remarque : si vous devez mettre en file d'attente les mêmes 10 paquets de grande taille sur une interface de 10 Mbits/s (par exemple, avec un téléphone IP et un PC connecté), le délai introduit est le suivant :

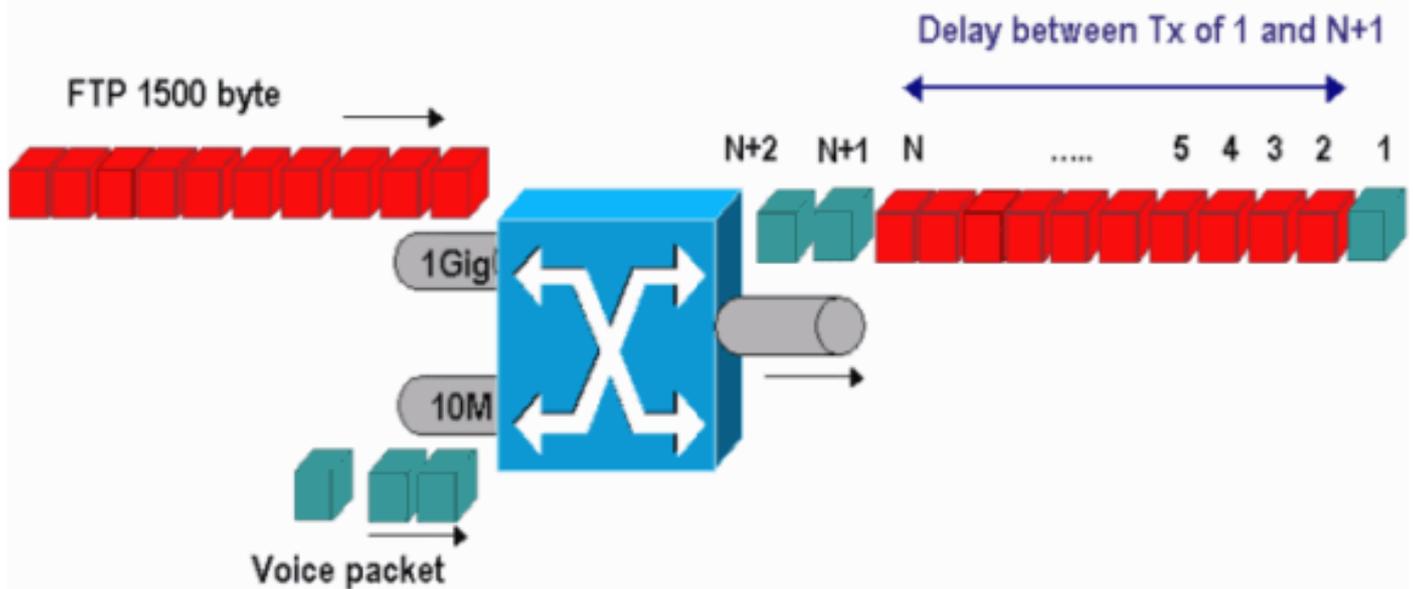
$(10 \times 1500 \times 8) = 120,000$ bits that are transmitted in 12 ms

La mise en oeuvre de la planification de sortie garantit que les paquets vocaux avec une CoS de 5 sont placés dans la file d'attente de priorité stricte. Ce placement garantit que ces paquets sont envoyés avant tout paquet avec une CoS inférieure à 5, ce qui réduit les délais.

Réduire la gigue

Un autre avantage important de la mise en oeuvre de la planification des sorties est qu'elle réduit la gigue. La gigue est la variation de délai observée pour les paquets dans le même flux. Le diagramme de la [Figure 8](#) montre un exemple de scénario de la façon dont la planification des sorties peut réduire la gigue :

Figure 8



Dans ce scénario, un port de sortie unique doit envoyer deux flux :

- Un flux vocal entrant sur un port Ethernet 10 Mbits/s
- Un flux FTP entrant sur une liaison ascendante Ethernet 1 Gbit/s

Les deux flux quittent le commutateur par le même port de sortie. Cet exemple montre ce qui peut se produire sans l'utilisation de la planification de sortie. Tous les paquets de données volumineux peuvent être entrelacés entre deux paquets vocaux, ce qui crée de la gigue dans la réception du paquet vocal à partir du même flux. Il y a un délai plus important entre la réception du paquet n et du paquet $n+1$ lorsque le commutateur transmet le paquet de données volumineux. Cependant, le délai entre $n+1$ et $n+2$ est négligeable. Cela provoque une gigue dans le flux de trafic vocal. Vous pouvez facilement éviter ce problème avec l'utilisation d'une file d'attente de priorité stricte. Assurez-vous que la valeur CoS des paquets voix est mappée à la file d'attente de priorité stricte.

Informations connexes

- [Planification de la sortie QoS sur les commutateurs des gammes Catalyst 6500/6000 exécutant le logiciel système Cisco IOS](#)
- [Présentation de Qos \(Qualité de service\) sur les commutateurs de la gamme Catalyst 6000](#)
- [Pages de support pour les produits LAN](#)
- [Page de support sur la commutation LAN](#)
- [Support et documentation techniques - Cisco Systems](#)