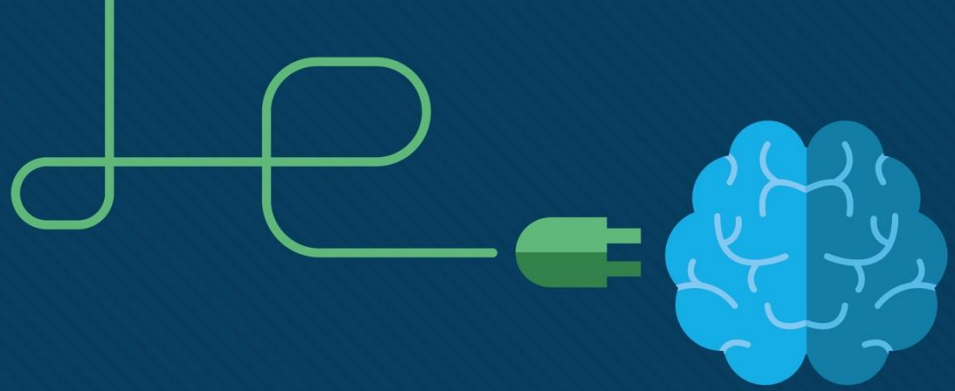


Module 9: Concepts QoS

Réseau, Sécurité et Automatisation D'entreprise v7.0 (ENSA)

Objectifs de ce module

Titre du module: Concepts QoS

Module Objective: Expliquer comment les périphériques réseau mettent en œuvre la QoS.

Titre du rubrique	Objectif du rubrique
Qualité des transmissions réseau	Expliquer l'impact sur la qualité des caractéristiques des transmissions réseau.
Caractéristiques du trafic	Décrire la configuration réseau minimale requise pour la voix, la vidéo et le trafic de données.
Algorithmes de file d'attente	Décrire les algorithmes de file d'attente utilisés par les périphériques réseau.
Modèles de QoS	Décrire les différents modèles de QoS.
Techniques de mise en œuvre de la QoS	Expliquer comment la QoS utilise des mécanismes pour garantir la qualité des transmissions.

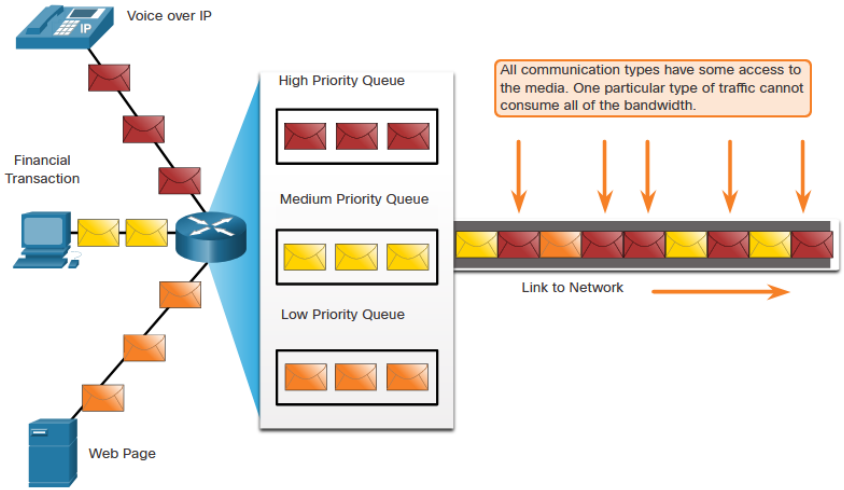
9.1 Qualité de transmission réseau

Vidéo - L'importance de la qualité de service

Cette vidéo explique la qualité de service (QoS) et pourquoi elle est nécessaire.

Hiérarchisation du trafic

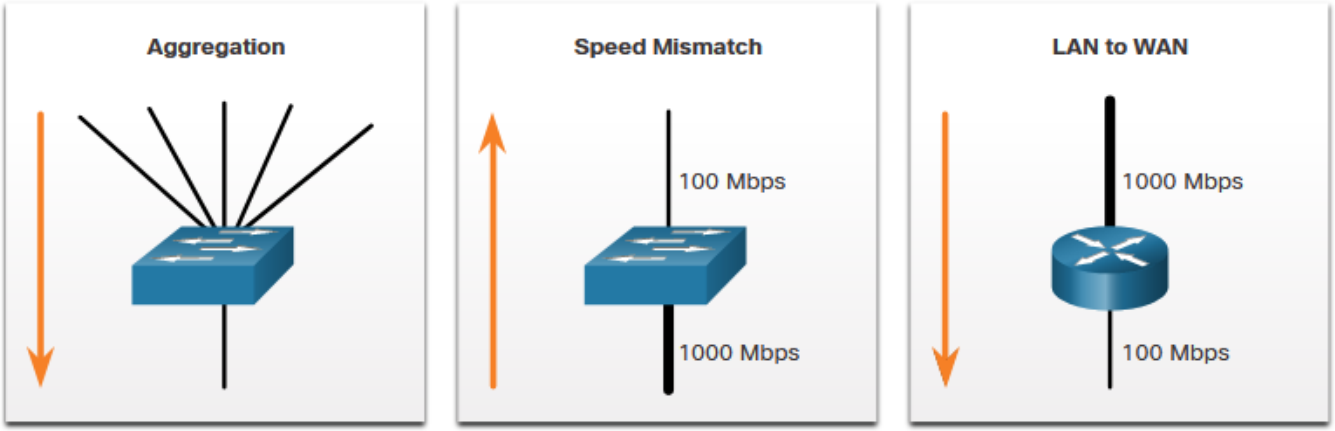
- Lorsque le volume de trafic est supérieur au volume pouvant être transporté sur le réseau, les périphériques placent les paquets en file d'attente dans la mémoire en attendant que des ressources se libèrent.
- Cette mise en file d'attente peut provoquer des retards, car les nouveaux paquets ne peuvent pas être transmis avant le traitement des paquets précédents.
- Si le nombre de paquets dans la file d'attente continue à augmenter, la mémoire du périphérique se remplit et les paquets sont détruits.
- Pour résoudre ce problème, il est possible de mettre en œuvre une technique QoS qui consiste à répartir les données dans plusieurs files d'attente, comme indiqué sur la figure.



Remarque: Un appareil implémente la qualité de service uniquement en cas de congestion.

Bande passante, encombrement, délai et gigue

- La bande passante réseau est mesurée en bits pouvant être transmis en une seconde, soit en «bits par seconde» (bits/s).
- L'encombrement d'un réseau entraîne des délais. Une interface est encombrée lorsqu'elle reçoit plus de trafic que le volume qu'elle peut prendre en charge. Les points de congestion d'un réseau sont idéals pour l'implémentation d'un mécanisme de QoS.
- Les points de congestion typiques sont l'agrégation, la disparité de débit et la liaison du LAN vers le WAN.



Bande passante, encombrement, délai et gigue (Suite)

Le délai ou la latence désigne le temps nécessaire à un paquet pour passer de la source à la destination.

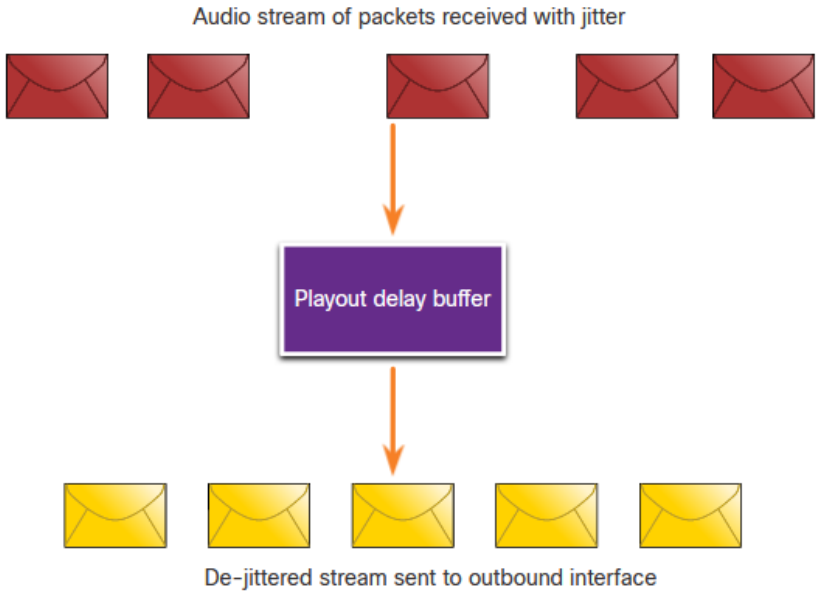
- Le délai fixe est le laps de temps nécessaire à l'exécution d'un processus donné, par exemple la durée requise pour placer un bit sur le support de transmission.
- Le délai variable, qui peut être plus ou moins long, dépend de plusieurs facteurs tels que le volume de trafic à traiter.
- La gigue est la variation de délai entre les paquets reçus.

Délai	Description
Délai lié au code	Durée fixe nécessaire à la compression des données au niveau de la source avant la transmission au premier appareil d'interconnexion des réseaux, généralement un commutateur
Délai du groupage par paquets	Durée fixe nécessaire à l'encapsulation d'un paquet avec toutes les informations d'en-tête requises
Délai de mise en file d'attente	Durée variable d'attente d'une trame ou d'un paquet avant d'être transmis sur la liaison
Délai de sérialisation	Délai fixe nécessaire à la transmission d'une trame vers le câble.
Délai de propagation	Durée variable nécessaire au passage de la trame entre la source et la destination.
Délai de gigue	Durée fixe nécessaire au stockage en mémoire tampon d'un flux de paquets, puis à son envoi à intervalles réguliers

Perte de paquets

Sans mécanismes de QoS, soumis à une contrainte temporelle, tels que la vidéo en temps réel et la voix, seront abandonnés à la même fréquence que les données non soumises à cette contrainte

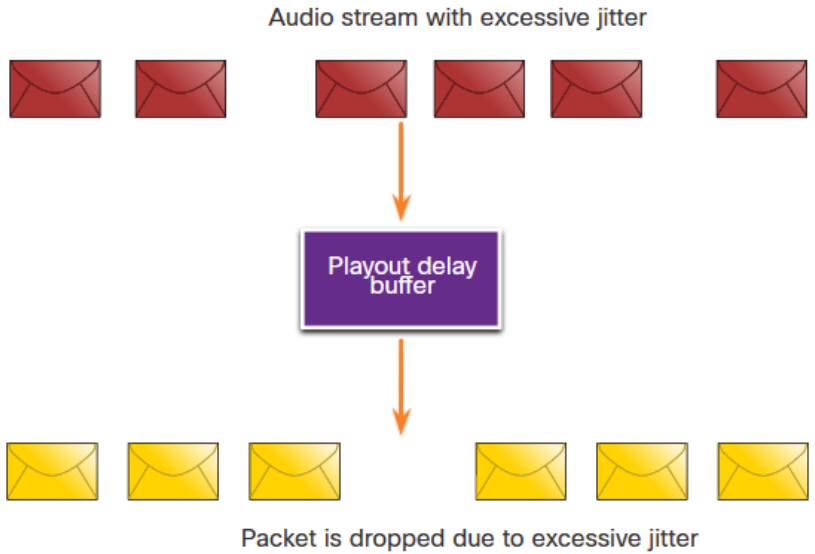
- Lorsqu'un routeur reçoit un flux de données audio numérique RTP (Real-Time Protocol) pour la voix sur IP (VoIP), il doit compenser la gigue générée en utilisant la mise en mémoire tampon.
- La mise en mémoire tampon consiste à mettre les paquets dans un tampon pour ensuite les diffuser en flux régulier.



Perte de paquets (Suite)

Si la gigue est forte au point que certains paquets arrivent en dehors de la mise de ce tampon, ces paquets sont ignorés et les coupures s'entendent dans l'enregistrement sonore.

- Si un seul paquet est perdu, le processeur de signal numérique (DSP) rajoute ce qu'il pense correspondre à l'enregistrement sonore manquant et l'utilisateur n'entendra rien.
- Cependant, lorsque la gigue est trop importante pour que le DSP compense les paquets manquants, des problèmes de son surviennent.



Remarque: Dans un réseau correctement conçu, ce phénomène doit être proche de zéro.

9.2 - Caractéristiques du trafic

Vidéo - Caractéristiques du trafic

Cette vidéo expliquera les caractéristiques du trafic voix, vidéo et données.

Tendances du trafic réseau

Au début des années 2000, la voix et les données constituaient les types de trafics IP prédominants.

- Le trafic voix nécessite un volume de bande passante prévisible et des délais de transmission des paquets constants.
- Le trafic de données n'est pas en temps réel et ses besoins en bande passante sont imprévisibles.
- Il peut temporairement se faire par salves, par exemple en cas de téléchargement d'un fichier volumineux, ce qui peut entraîner la consommation de la totalité de la bande passante d'une liaison.

Plus récemment, le trafic vidéo a acquis une importance de plus en plus cruciale pour les communications et les opérations professionnelles.

- Selon l'indice VNI (Virtual Networking Index) Cisco, le trafic vidéo représentait 70% du trafic en 2017.
- En 2022, le vidéo représentera 82% de tous les trafics.
- Le trafic vidéo mobile atteindra 60,9 exaoctets par mois en 2022.

Les différents types de trafics (voix, vidéo et données) impliquent des besoins extrêmement variés en termes de réseau.

Voix

Le trafic vocal est prévisible et fluide et très sensible aux délais et aux paquets abandonnés.

- Les paquets vocaux doivent bénéficier d'une priorité plus élevée que le reste du trafic.
- Les produits Cisco utilisent la plage de ports RTP comprise entre 16384 et 32767 afin de placer ce trafic en priorité maximale.

La voix peut tolérer un certain degré de latence, de gigue et de perte sans effets notables.

La latence ne peut pas dépasser 150 ms.

- la gigue ne doit pas dépasser 30 ms; et la perte de paquets ne doit pas dépasser 1%.
- Le trafic voix nécessite au moins 30 Kbit/s de bande passante.

Caractéristiques du trafic voix	Requêtes unidirectionnelles
• Fluide • Minimal • Sensible aux pertes • Sensible aux retards • Priorité UPD	• Latence ≤ 150ms • Gigue ≤ 30 ms • Perte ≤ 1% bande passante (30-128 Kbit/s)

Vidéo

Le trafic peut être imprévisible, incohérent et en salve. Contrairement à la voix, la vidéo récupère moins bien en cas de perte et comporte un plus grand volume de données par paquet.

- Le nombre et la taille des paquets vidéo varient toutes les 33 ms selon le contenu de la vidéo.
- Les ports UDP, par exemple le port 554 utilisé pour le Real-Time Streaming Protocol (RSTP), doivent être prioritaires par rapport au trafic réseau moins soumis à des contraintes temporelles.
- La latence ne doit pas dépasser 400 ms. la gigue ne doit pas dépasser 50 ms; la perte de paquets vidéo ne doit pas dépasser 1%. Le trafic vidéo nécessite au moins 384 Kbit/s de bande passante.

Caractéristiques du trafic vidéo	Requêtes unidirectionnelles
• En salves • Gourmand • Sensible aux pertes • Sensible aux retards • Priorité UPD	• Latence ≤ 200-400 ms • Gigue ≤ 30-50 ms • Loss ≤ 0.1 – 1% • Bande passante (384 Kbps - 20 Mbps)

Données

Les applications de données qui ne tolèrent pas la perte de données, comme les e-mails et les pages web, utilisent le protocole TCP pour garantir que les éventuels paquets perdus lors du transit seront renvoyés.

- Le trafic de données peut être fluide ou en salve.
- Le trafic du réseau est généralement fluide et prévisible.

Certaines applications TCP peuvent être très extrêmement gourmandes et consomme une grande partie de la capacité du réseau. Lorsque vous téléchargez un fichier volumineux, par exemple un film ou un jeu, le protocole FTP consomme autant de bande passante qu'il peut.

Caractéristiques du trafic de données
• Fluide/en salves (burst) • Minimal/gourmand • Insensible aux pertes • Insensible aux retards • Retransmission TCP

Données (Suite)

Par rapport à la voix et à la vidéo, le trafic de données est relativement peu sensible aux pertes et aux retards. La qualité de l'expérience ou la QoE est importante à considérer avec le trafic de données.

- Les données proviennent-elles d'une application interactive?
- Les données sont-elles essentielles?

Facteur	Essentiel	Non essentiel
Interactif	Accordez la priorité au délai le plus bas de l'ensemble du trafic de données et essayez d'obtenir un délai de réponse de 1 à 2 secondes.	Les applications peuvent bénéficier de ce délai inférieur.
Non interactif	Tant que la bande passante minimale nécessaire est assurée, le délai peut varier considérablement.	Obtient la bande passante restante une fois que les besoins du trafic voix, vidéo et de données ont été satisfaits.

9.3 Algorithmes de mise en file d'attente

Vidéo - Algorithmes de QoS

Cette vidéo couvrira les éléments suivants:

- La mise en file d'attente FIFO
- File d'attente équitable pondérée (WFQ)
- File d'attente équitable pondérée basée sur la classe (CBWFQ)
- File d'attente à faible latence (LLQ)

Présentation de la mise en file d'attente

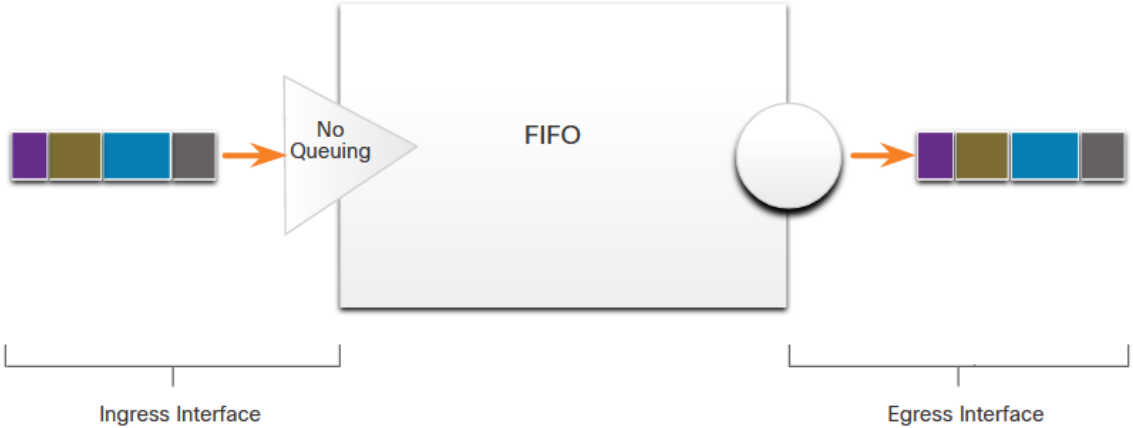
La politique QoS que l'administrateur réseau a implémenté devient active en cas d'encombrement sur la liaison. La mise en file d'attente est un outil de gestion des congestions qui permet de stocker en mémoire tampon, de hiérarchiser et, si nécessaire, de réorganiser les paquets avant leur transmission à la destination.

Différents algorithmes de mise en file d'attente sont disponibles:

- FIFO (First-In, First-Out)
- File d'attente équitable pondérée (WFQ)
- File d'attente équitable pondérée basée sur la classe (CBWFQ)
- File d'attente à faible latence (LLQ)

Premier entrant premier sorti (FIFO)

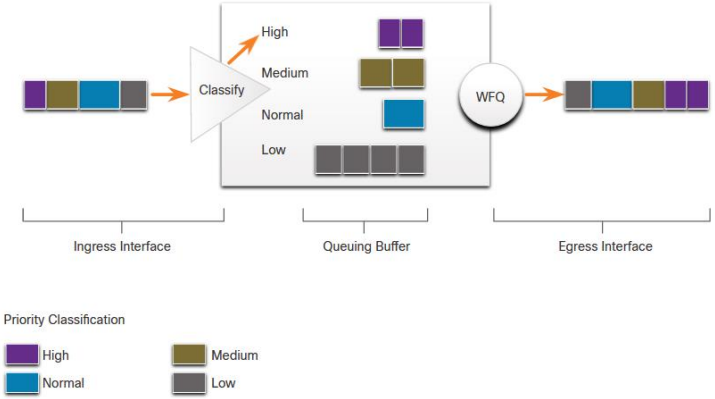
- L'algorithme FIFO met en file d'attente les paquets et les transfère dans l'ordre de leur arrivée.
- Le FIFO n'a pas de concept de priorité ou de classe de trafic et de ce fait, ne prend pas de décision sur la priorité des paquets.
  - Il n'y a qu'une seule file d'attente et tous les paquets sont traités de la même manière.
  - Les paquets sont envoyés par une interface dans l'ordre dans lequel ils arrivent.



Mise en file d'attente équitable pondérée (WFQ)

WFQ (Weighted Fair Queuing) est une méthode de programmation automatisée grâce à laquelle la bande passante est allouée au trafic réseau de façon équitable.

- WFQ applique la priorité, ou les pondérations, au trafic identifié, le classe en conversations ou en flux, puis détermine la quantité de bande passante autorisée par chaque flux par rapport aux autres flux.
- WFQ permet de classer le trafic en différents flux selon l'adressage des adresses IP source et de destination, les adresses MAC, les numéros de port, le protocole, et la valeur du type de service (ToS).
- La méthode WFQ n'est pas prise en charge en cas de tunnélisation et de chiffrement, car ces fonctionnalités entraînent la modification des informations de contenu des paquets dont cette méthode a besoin à des fins de classification.



Mise en file d'attente pondérée basée sur les classes (CBWFQ)

CBWFQ étend la fonctionnalité de mise en file d'attente pondérée (WFQ) standard afin de fournir la prise en charge des classes de trafic définies par l'utilisateur.

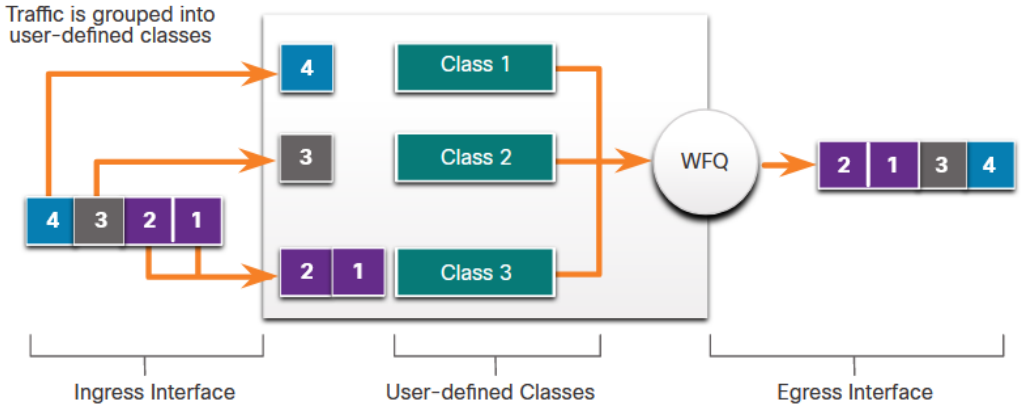
- Les classes de trafic sont définies en fonction de critères de correspondance incluant les protocoles, les listes de contrôle d'accès (ACL) et les interfaces d'entrée.
- Les paquets qui répondent à ces critères pour une classe constituent le trafic pour cette classe.
- Une file d'attente FIFO est réservée à chaque classe et le trafic appartenant à une classe est dirigé dans la file d'attente correspondant à cette classe.

- Une classe peut être attribuée à des caractéristiques telles que la bande passante, le poids et la limite maximale de paquets. La bande passante attribuée à une classe sera la bande passante garantie à cette classe lors d'un encombrement.
- Les paquets appartenant à une classe sont soumis aux limites de bande passante et de file d'attente, qui est le nombre maximum de paquets autorisés à s'accumuler dans la file d'attente, qui caractérisent la classe.

Mise en file d'attente pondérée basée sur les classes (CBWFQ) (Suite)

Une fois qu'une file d'attente a atteint sa limite configurée, l'ajout d'autres paquets à la classe entraîne la perte de queues (Tail drop) ou de paquets, selon la configuration de la politique de classe.

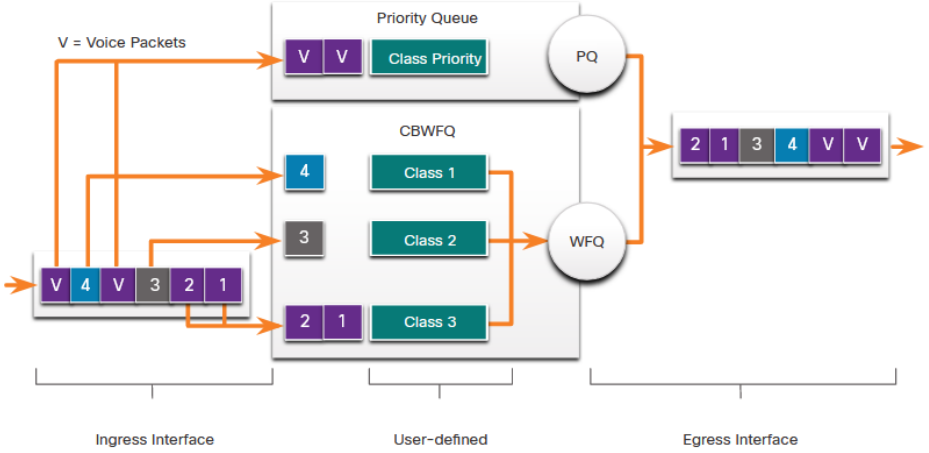
- Le « Tail drop » détruira les nouveaux paquets arrivant sur la file d'attente ayant complètement épuisé ses ressources de mise en file d'attente de paquets.
- C'est la réponse par défaut de mise en file d'attente en cas de congestion. Le «Tail drop» d'attente s'applique de manière identique à tout le trafic et à toutes les classes de service.



Mise en file d'attente à faible latence (LLQ)

Avec la fonctionnalité LLQ, la stratégie CBWFQ bénéficie d'une capacité de mise en file d'attente à priorité stricte.

- La priorité stricte permet aux paquets soumises à des contraintes temporelles, par exemple la voix, d'être envoyées avant les paquets présents dans d'autres files d'attente.
- Ainsi, les paquets soumises à des contraintes temporelles comme la voix sont envoyées en premier (avant les paquets présents dans d'autres files d'attente), leur offrant un traitement préférentiel par rapport au reste du trafic.
- Cisco recommande que seul le trafic vocal soit dirigé vers la file d'attente prioritaire.



9.4 Modèles de QoS

Vidéo - Modèles de QoS

Cette vidéo couvrira les éléments suivants:

- Remise au mieux (Best effort)
- Services intégrés (IntServ)
- Services différenciés (DiffServ)

Sélection d'un modèle de politique de QoS approprié

Il existe trois modèles d'implémentation QoS. La QoS est implémentée dans un réseau à l'aide du modèle IntServ ou DiffServ.

- Le modèle IntServ offre un meilleur niveau de qualité de service, il est extrêmement gourmand en ressources et donc limité en matière d'évolutivité.
- Le modèle DiffServ mobilise moins de ressources et est plus évolutif.
- Les deux modèles IntServ et DiffServ sont parfois déployés conjointement dans les implémentations QoS du réseau.

Modèle	Description
Remise au mieux (Best effort)	• Il ne s'agit pas d'une implémentation dans la mesure où la stratégie QoS n'est pas explicitement configurée. • Ce modèle est utilisé lorsque la qualité de service n'est pas nécessaire.
Services intégrés (IntServ)	• Ce modèle propose une qualité de service très élevée aux paquets IP, avec remise garantie. • Il définit un processus de signalisation pour que les applications puissent indiquer au réseau qu'elles nécessitent une qualité de service spéciale pendant une certaine période et qu'il faut réserver de la bande passante. • Le modèle IntServ peut considérablement limiter l'évolutivité d'un réseau.
Services différenciés (DiffServ)	• Ce modèle offre une implémentation QoS très flexible et évolutive. • Les périphériques réseau détectent des classes de trafic et appliquent des niveaux de qualité de service spécifiques aux différentes classes de trafic.

Remise au mieux

La conception de base d'Internet prévoit la remise des paquets «Remise au mieux» et n'offre aucune garantie.

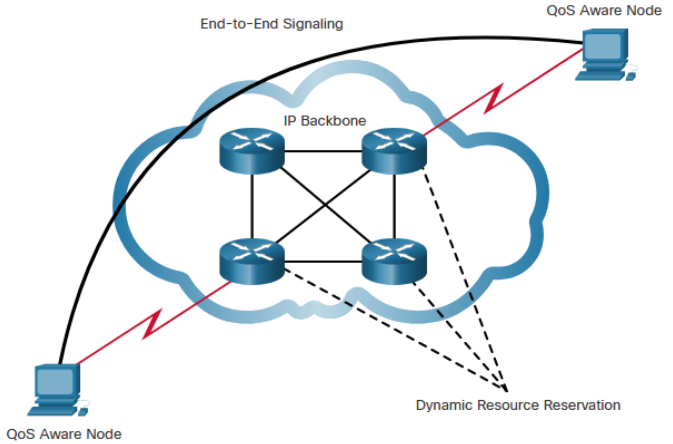
- Comme ce modèle applique un traitement identique à tous les paquets réseau, le message vocal d'urgence est traité de la même manière qu'une photo numérique jointe à un e-mail.
- Bénéfices et inconvénients du modèle Remise au mieux:

Bénéfices	Inconvénients
Il s'agit du modèle le plus évolutif.	Il n'offre aucune garantie de remise.
L'évolutivité est uniquement limitée par la bande passante disponible, laquelle affecte alors l'ensemble du trafic	Le délai et l'ordre de remise des paquets sont aléatoires et rien ne garantit leur arrivée.
Aucun mécanisme QoS spécial ne doit être implémenté.	Aucun paquet ne bénéficie d'un traitement préférentiel.
C'est le modèle le plus simple et rapide à déployer.	Les données essentielles sont traitées de la même façon que les e-mails normaux.

Services intégrés (IntServ)

IntServ offre la qualité de service de bout en bout dont les applications en temps réel ont besoin.

- IntServ gère explicitement les ressources réseau, ce qui permet de garantir la qualité de service requise aux flux de paquets de certains utilisateurs, parfois appelés microflux.
- Met en œuvre des mécanismes de contrôle d'admission et de réservation des ressources sous forme de composants pour établir et préserver la qualité de service.
- Il repose sur l'utilisation d'une approche prenant en compte le type de connexion. Chaque communication individuelle doit spécifier explicitement son descripteur de trafic et les ressources demandées au réseau.
- Le routeur de périphérie responsable du contrôle d'admission s'assure qu'il existe suffisamment de ressources disponibles sur le réseau.



Services intégrés (IntServ) (suite)

Dans le modèle IntServ, l'application demande un type de service spécifique au réseau avant de transmettre les données.

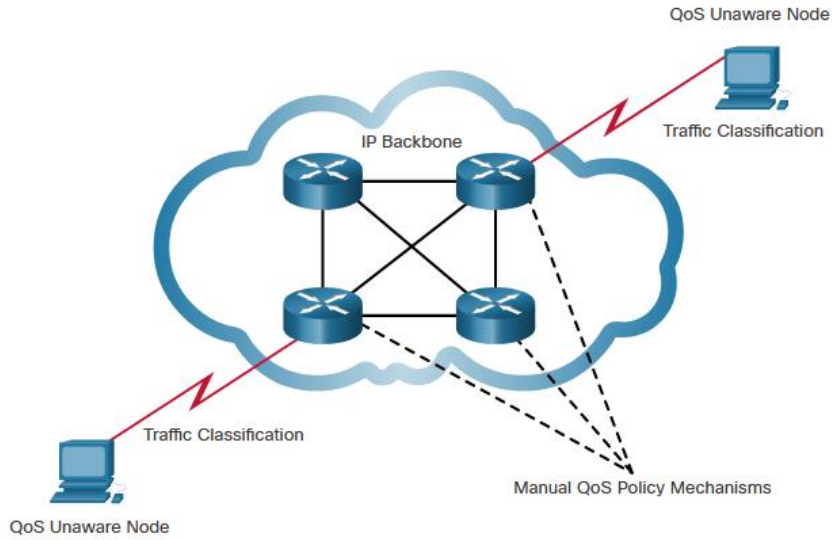
- L'application avertit le réseau de son profil de trafic et demande un type particulier de service, dont ses exigences en termes de bande passante et de latence.
- Le modèle IntServ utilise le protocole RSVP (Resource Reservation Protocol) pour communiquer les besoins QoS du trafic d'une application à tous les périphériques du chemin du réseau.
- Si les périphériques du chemin sont en mesure de réserver la bande passante nécessaire, l'application source peut commencer à transmettre les données. Si la réservation demandée échoue à un emplacement quelconque du chemin, l'application source n'envoie pas de données.

Bénéfices	Inconvénients
• Contrôle d'admission des ressources explicite, de bout en bout. • Contrôle d'admission de la stratégie par demande. • Signalisation des numéros de port dynamiques.	• Consommation importante de ressources due aux exigences de signalisation continue de l'architecture dynamique. • Approche basée sur les flux, inadaptée aux implémentations de grande taille, par exemple l'internet

Services différenciés (DiffServ)

Le modèle QoS de services différenciés (DiffServ) propose un mécanisme simple et évolutif pour classer et gérer le trafic réseau.

- N'est pas une stratégie QoS de bout en bout car il ne peut pas appliquer les garanties de bout en bout.
- Les hôtes transmettent du trafic à un routeur, ce dernier organise (agrège) les flux en classes et propose la stratégie QoS appropriée aux classes.
- Met en œuvre et applique des mécanismes QoS saut par saut, en appliquant des critères globaux à chaque classe de trafic pour garantir la flexibilité et l'évolutivité.



Services différenciés (DiffServ) (suite)

- DiffServ divise le trafic réseau en classes selon les besoins métier. Un niveau de service différent peut être ensuite affecté à chaque classe.
- Lorsque les paquets transitent sur un réseau, les différents périphériques réseau identifient la classe du paquet et lui appliquent un niveau de service spécifique à cette classe.
- Le modèle DiffServ propose un large choix de niveaux de service.

Bénéfices	Inconvénients
• Haute évolutivité • Large choix de niveaux de qualité	• Aucune garantie stricte de la qualité de service • Nécessite le fonctionnement conjoint de mécanismes complexes sur l'ensemble du réseau

9.5 Techniques d'implémentation QoS

Vidéo - Techniques d'implémentation QoS

Cette vidéo couvrira les éléments suivants:

- Outils de mise en œuvre (classification et marquage, prévention de la congestion et gestion des encombrements)
- Marquage du trafic

Prévention des pertes de paquets

La perte de paquets est généralement due à une congestion au niveau d'une interface. La plupart des applications qui utilisent TCP enregistrent un ralentissement car TCP s'adapte automatiquement à l'encombrement du réseau. La perte de segments TCP entraîne une réduction de la taille de fenêtre des sessions TCP. Certaines applications n'utilisent pas le protocole TCP et ne sont pas en mesure de gérer les abandons de paquets (flux fragiles).

Plusieurs mesures permettent d'éviter la perte de paquets pour les applications à risques:

- Augmenter la capacité des liaisons pour réduire ou éviter la congestion.
- Garantir suffisamment de bande passante et augmenter l'espace de la mémoire tampon pour pouvoir prendre en charge les pics de trafic des flux fragiles. WFQ, CBWFQ et LLQ peuvent garantir la bande passante et fournir un transfert hiérarchisé vers des applications sensibles aux pertes de données.
- Rejeter les paquets moins prioritaires avant que la congestion ne se produise. Les fonctions QoS du logiciel Cisco IOS proposent des mécanismes de mise en file d'attente, tel que WRED (weighted random early detection), qui commencent à supprimer les paquets à priorité plus faible avant que la congestion se produise.

Outils QoS

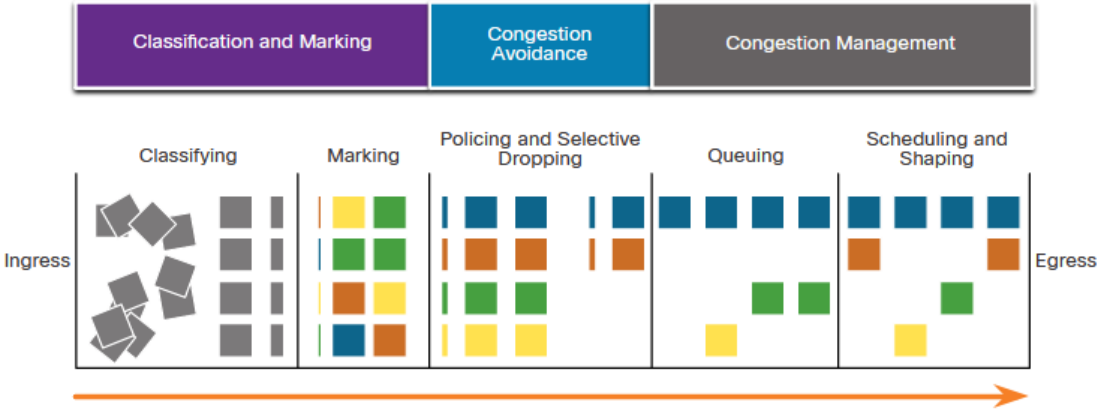
Il existe trois catégories d'outils QoS, décrites dans le tableau.

Outils QoS	Description
Outils de classification et de marquage	• Les sessions, ou les flux, sont analysées afin de déterminer la classe de trafic à laquelle elles appartiennent. • Une fois la classe identifiée, les paquets sont marqués.
Outils de prévention de l'encombrement	• Les classes de trafic représentent des ressources réseau allouées, l'allocation étant définie dans la stratégie QoS. • La stratégie QoS identifie également le traitement appliqué au trafic (suppression sélective d'une partie du trafic, délai ou nouveau marquage) pour éviter la congestion du réseau. • Principal outil de prévention d'encombrement, WRED permet de réguler le trafic de données TCP en utilisant efficacement la bande passante avant que des dépassements de file d'attente n'entraînent des abandons de paquets.
Outils de gestion de l'encombrement	• Lorsque le trafic dépasse les ressources réseau disponibles, il est placé en file d'attente en attendant que des ressources se libèrent. • Le système Cisco IOS propose plusieurs outils de gestion de l'encombrement, dont les algorithmes CBWFQ et LLQ.

Outils QoS (Suite)

La figure montre la séquence des outils de QoS utilisés lorsqu'ils sont appliqués aux flux de paquets.

- Les paquets en entrée sont classés et leur en-tête IP est marqué.
- Pour éviter l'encombrement du réseau, des ressources sont allouées aux paquets sur la base des stratégies définies.
- Ils sont ensuite placés en file d'attente puis transmis vers l'interface de sortie en fonction de la stratégie de surveillance et de régulation QoS définie.



Remarque: La classification et le marquage peuvent être effectués à l'entrée ou à la sortie tandis que d'autres tâches QoS, notamment la mise en file d'attente et la régulation, sont généralement effectuées à la sortie.

Classification et marquage

Avant de pouvoir appliquer une stratégie QoS à un paquet, ce dernier doit être classé.

La classification détermine la classe de trafic à laquelle les paquets ou les trames appartiennent. Les politiques ne peuvent être appliquées qu'une fois le trafic marqué.

La classification des paquets varie selon l'implémentation QoS choisie.

- Pour classer le trafic des couches 2 et 3, plusieurs méthodes sont possibles, dont les interfaces, les listes de contrôle d'accès et les mappages de classes.
- Il peut également être classé au niveau des couches 4 à 7 à l'aide de la fonction NBAR (Network Based Application Recognition).

Classification et marquage (Suite)

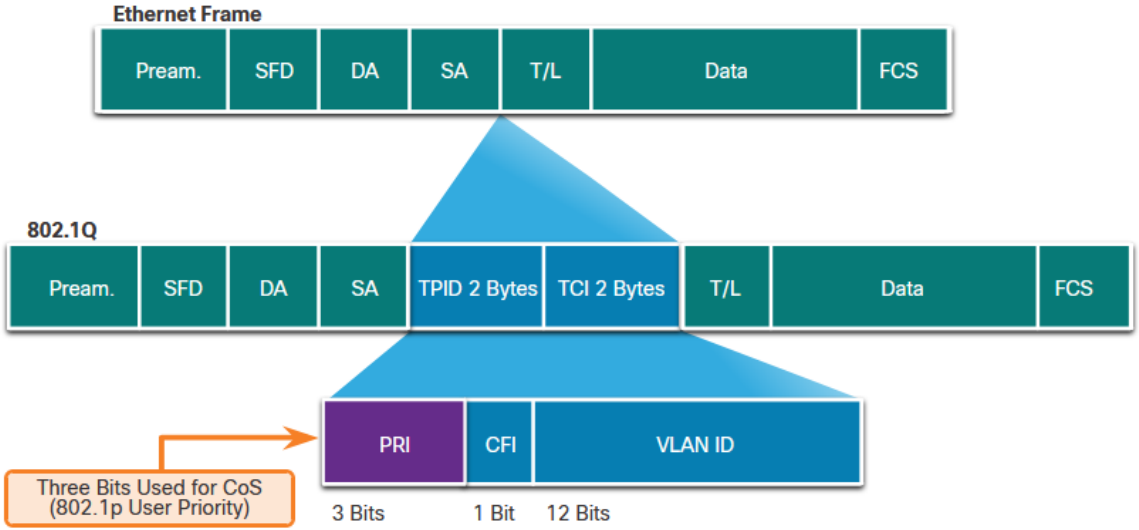
Le marquage appliqué au trafic dépend de la technologie. La décision de marquer le trafic au niveau de la couche 2 ou 3 (ou des deux) n'est pas anodine. Voici quelques points à prendre en compte avant de choisir:

- Le marquage des trames au niveau de la couche 2 peut être effectué pour le trafic non-IP.
- Le marquage des trames au niveau de la couche 2 est la seule option QoS disponible pour les commutateurs qui ne prennent pas en charge le trafic IP.
- Le marquage de la couche 3 porte les informations QoS de bout en bout.

Outils QoS	Couche	Champ de marquage	Largeur en bits
Ethernet (802.1q, 802.1p)	2	Classe de service (CoS)	3
802.11 (WiFiFi)	2	Identifiant de trafic (TID) Wi-Fi	3
MPLS	2	Expérimental (EXP)	3
IPv4 et IPv6	3	Priorité IP (IPP)	3
IPv4 et IPv6	3	Marquage DSCP (Differentiated Services Code Point)	6

Marquage de couche 2

802.1Q est le standard IEEE qui prend en charge l'étiquetage (Tag) des VLAN au niveau de la couche 2 des réseaux Ethernet. Lorsque 802.1Q est mis en œuvre, deux champs sont ajoutés à la trame Ethernet et insérés après le champ d'adresse MAC source.



Marquage de couche 2 (Suite)

Le standard 802.1Q inclut également le schéma de hiérarchisation QoS plus connu sous le terme IEEE 802.1p. Le standard 802.1p utilise les trois premiers bits du champ Données de contrôle des balises (TCI). Ce champ de 3 bits, ou champ de Priorité (PRI), contient le marquage de la classe de service.

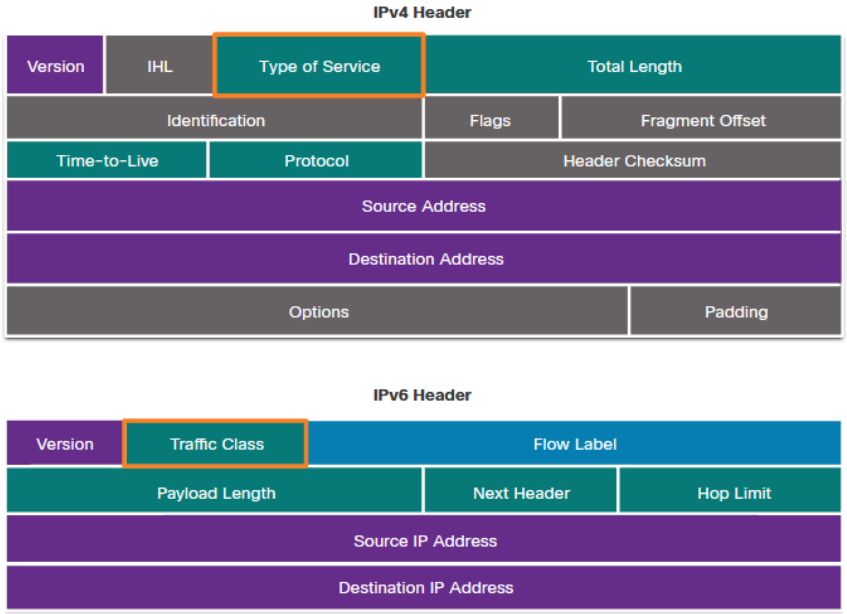
Trois bits signifie qu'une trame Ethernet de couche 2 peut être marquée avec l'un des huit niveaux de priorité (valeurs 0-7).

Valeur CoS	Valeur CoS binaire	Description
0	000	Données de Remise au mieux
1	001	Données de priorité moyenne
2	010	Données de priorité forte
3	011	Signalisation d'appels
4	100	Vidéoconférence
5	101	Support voix (trafic voix)
6	110	Réservé
7	111	Réservé

Marquage de couche 3

Le marquage des paquets avec IPv4 et IPv6 s'effectue à l'aide d'un champ de 8 bits situé au niveau des en-têtes de paquet.

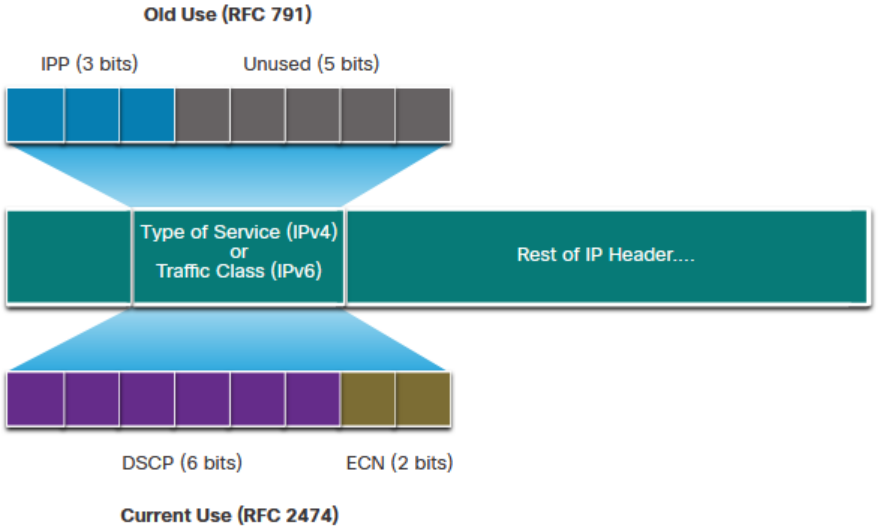
Pv4 et IPv6 prennent en charge un champ de 8 bits pour le marquage, le champ Type de service (ToS) pour IPv4 et le champ Classe de trafic pour IPv6.



Type de service et champ de classe de trafic

Le type de service (IPv4) et la classe de trafic (IPv6) portent le marquage des paquets tel qu'attribué par les outils de classification QoS.

- La RFC 791 a spécifié le champ IPP (3-bit IP Precedence) à utiliser pour le marquage de la qualité de service.
- RFC 2474 remplace RFC 791 et redéfinit le champ ToS en renommant et en élargissant le champ IPP par 6 bits.
- Ces six bits s'appellent le champ DSCP (Differentiated Services Code Point) et proposent jusqu'à 64 classes de service.
- Les deux bits IP ECN (Extended Congestion Notification) restants peuvent être utilisés par les routeurs compatibles ECN pour marquer les paquets au lieu de les abandonner.



Valeurs DSCP

Les 64 valeurs DSCP sont réparties en trois catégories:

- **Remise au mieux (Best effort)** - Il s'agit de la catégorie par défaut pour l'ensemble des paquets IP. La valeur du champ DSCP est égale à 0. Un routage normal est appliqué au niveau de chaque saut. En cas de congestion sur un routeur, ces paquets seront abandonnés. Aucune politique de QoS n'a été implémentée.
- **Expedited Forwarding (EF)**: La RFC 3246 définit que le flux EF sera identifié comme un champ DSCP à 46 (binaire **101**110). Les trois premiers bits (101) correspondent à la valeur CoS 5, qui est utilisé à la couche 2 pour le trafic voix. Pour la couche 3, Cisco conseille d'utiliser les valeurs EF uniquement pour marquer les paquets voix.
- **Assured Forwarding (AF)** - La norme RFC 2597 définit l'AF comme l'utilisation des 5 bits DSCP les plus significatifs pour indiquer les files d'attente et la préférence de suppression.

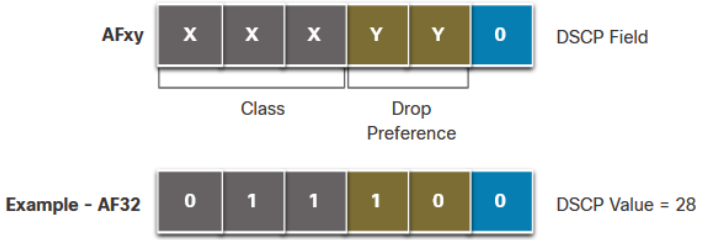
Valeurs DSCP (Suite)

Les valeurs de transfert assuré sont indiquées dans la figure.

La formule **AFXy** est spécifiée comme suit:

- Les 3 bits les plus pertinents sont utilisés pour spécifier la classe. La classe 4 correspond à la meilleure file d'attente, et la classe 1 à la file d'attente la moins bonne.
- Les 4e et 5e bits, les plus pertinents, sont utilisés pour indiquer la probabilité de perte.
- Le 6e bit est mis à zéro.

Assured Forwarding Values			
	Low Drop	**Medium Drop**	**High Drop**
Class 4	AF41 (34)	AF42 (36)	AF43 (38)
Class 3	AF31 (26)	AF32 (28)	AF33 (30)
Class 2	AF21 (18)	AF22 (20)	AF23 (22)
Class 1	AF11 (10)	AF12 (12)	AF13 (14)

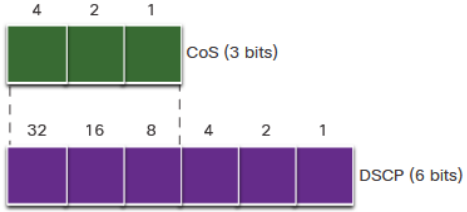


Par exemple, AF32 appartient à la classe 3 (binaire 011), avec une probabilité de perte moyenne (binaire 10). La valeur du champ DSCP est égale à 28, car elle intègre 6e bit qui à zéro, soit en binaire 011100.

Bits sélecteurs de classe (CS)

Bits sélecteurs de classe (CS):

- Les 3 premiers bits les plus significatifs du champ DSCP et indique la classe.
- Mappez directement aux 3 bits du champ CoS et du champ IPP pour maintenir la compatibilité avec 802.1p et RFC 791.



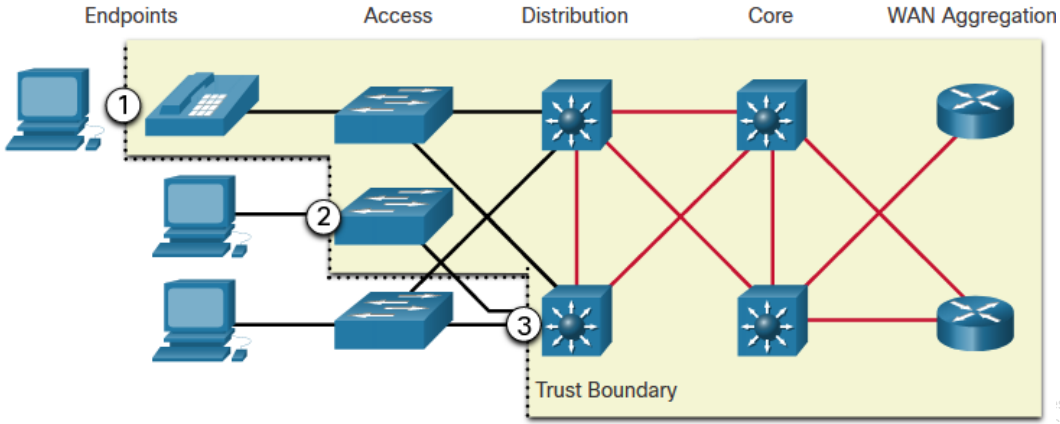
CoS values, Class Selectors, and corresponding DSCP 6-bit value

CoS Value	CoS Binary Value	Class Selector (CS)	CS Binary	DSCP Decimal Value
0	000	CS0*/DF	000 000	0
1	001	CS1	001 000	8
2	010	CS2	010 000	16
3	011	CS3	011 000	24
4	100	CS4	100 000	32
5	101	CS5	101 000	40
6	110	CS6	110 000	48
7	111	CS7	111 000	56

Limites de confiance

Le classement et le marquage du trafic doivent s'effectuer le plus près possible, techniquement et administrativement, de la source. Cela permet de définir la limite de confiance.

1. Les terminaux sécurisés disposent de fonctionnalités et de renseignements qui leur permettent de marquer le trafic des applications à l'aide de valeurs CoS de couche 2 et/ou DSCP de couche 3.
2. Pour les terminaux sécurisés, le trafic peut être marqué au niveau du commutateur de couche 2.
3. Le trafic peut également être marqué au niveau des commutateurs/routeurs de couche 3.



Prévention de la congestion

Les outils de prévention de congestion permettent de surveiller les charges de trafic sur les réseaux afin d'anticiper et d'éviter les encombrements au niveau des congestions du réseau commun et de l'internet avant que les encombrements ne deviennent un problème.

- Ils surveillent les charges de trafic réseau pour anticiper et éviter tout encombrement au niveau des goulots d'étranglement réseau et inter-réseau habituels avant que la situation ne devienne problématique.
- Ils surveillent la profondeur moyenne de la file d'attente. Lorsque la file d'attente est inférieure au seuil minimal, il n'y a pas de suppressions. Au fur et à mesure que la file d'attente se remplit pour atteindre le seuil maximal, un faible pourcentage des paquets est supprimé. Une fois le seuil maximal atteint, tous les paquets sont supprimés.

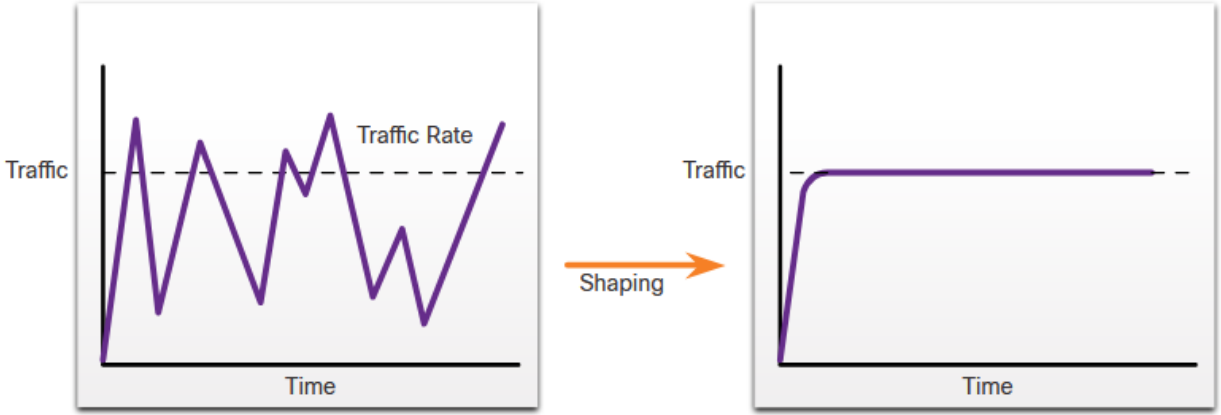
Certaines techniques de prévention de la congestion appliquent un traitement préférentiel au moment de choisir les paquets à d'être abandonner.

- WRED permet d'éviter l'encombrement des interfaces réseau en gérant les tampons et en autorisant la réduction ou la diminution du trafic TCP avant que ceux-ci soient remplis.
- Grâce à WRED, il est possible d'éviter les pertes de paquet tout en optimisant l'utilisation du réseau et les performances des applications utilisant TCP.

Mise en forme et régulation

La régulation et la limitation du trafic sont deux mécanismes proposés par la solution QoS de Cisco IOS pour éviter l'encombrement.

- La régulation du trafic maintient les paquets excédentaires dans une file d'attente et planifie une transmission ultérieure, répartie graduellement dans le temps. La mise en forme du trafic entraîne un débit de sortie de paquets lissé.
- La mise en forme s'applique au contenu sortant. Les paquets envoyés à partir d'une interface sont mis en file d'attente, puis éventuellement mis en forme. La régulation (policing) s'applique quant à elle au trafic entrant d'une interface.



Mise en forme et régulation (Suite)

La régulation est appliquée au trafic entrant sur une interface. La régulation est généralement mise en œuvre par les prestataires de services dans le cadre d'un contrat avec débit garanti (CIR). Le prestataire de services peut toutefois autoriser le dépassement du débit garanti lorsque son réseau n'est pas encombré.

Conseils de régulation QoS

Les stratégies QoS doivent prendre en compte le chemin complet de la source à la destination.

Voici quelques conseils qui aident à garantir la meilleure expérience pour les utilisateurs finaux:

- Activez la mise en file d'attente sur chaque périphérique dans le chemin entre la source et la destination.
- Classifier et marquer le trafic le plus près possible de la source.
- Mise en forme et régulation des flux de trafic le plus près possible de leurs sources.

9.6 Module pratique et questionnaire

Qu'est-ce que j'ai appris dans ce module?

- Les transmissions vocales et vidéo en direct créent des attentes plus élevées en matière de qualité parmi les utilisateurs et créent un besoin de qualité de service (QoS).
- En l'absence de mécanismes de QoS, les paquets sont traités dans l'ordre dans lequel ils arrivent. En cas d'encombrement, il se peut que les périphériques réseau comme les routeurs et les commutateurs abandonnent les paquets.
- En l'absence des paquets soumis à une contrainte temporelle de QOS, tels que la vidéo en temps réel et la voix, seront abandonnés à la même fréquence que les données non soumises à cette contrainte, par exemple les e-mails et la navigation web.
- Cette mise en file d'attente peut provoquer des retards, car les nouveaux paquets ne peuvent pas être transmis avant le traitement des paquets précédents.
- On distingue le délai fixe et le délai variable.
- Les sources de délai sont délai de code, délai de paquétisation, délai de mise en file d'attente, délai de sérialisation, délai de propagation, délai de gigue
- La gigue est la variation de délai entre les paquets reçus.
- La voix et le trafic vidéo sont deux des principales raisons de la qualité de service.
- La circulation vocale est fluide et bénigne, mais elle est sensible aux abandons et aux délais.

Qu'est-ce que j'ai appris dans ce module? (Suite)

- La voix peut tolérer un certain degré de latence, de gigue et de perte sans effets notables.
- Le trafic vidéo est plus exigeant que le trafic vocal car la taille des paquets qu'il envoie sur le réseau est plus importante.
- Le trafic vidéo est en salve, consommateur de ressources, sensible aux abandons de paquets et aux délais.
- Le trafic de données n'est pas aussi exigeant que le trafic vocal et vidéo. Les paquets de données utilisent souvent des applications TCP qui peuvent retransmettre des données et, par conséquent, ne sont pas sensibles aux abandons et aux délais.
- La regulation QoS que l'administrateur réseau a implémentée devient active en cas d'encombrement sur la liaison.
- La mise en file d'attente est un outil de gestion des congestions qui permet de stocker en mémoire tampon, de hiérarchiser et, si nécessaire, de réorganiser les paquets avant leur transmission à la destination.
- L'algorithme FIFO met en file d'attente les paquets et les transfère dans l'ordre de leur arrivée. Le FIFO n'a pas de concept de priorité ou de classe de trafic et de ce fait, ne prend pas de décision sur la priorité des paquets.
- WFQ est une méthode de programmation automatisée grâce à laquelle la bande passante est allouée au trafic réseau de façon équitable. Elle applique des priorités ou des pondérations au trafic identifié et le classe en conversations ou flux.

Qu'est-ce que j'ai appris dans ce module? (Suite)

- CBWFQ étend la fonctionnalité de mise en file d'attente pondérée (WFQ) standard afin de fournir la prise en charge des classes de trafic définies par l'utilisateur. Avec le CBWFQ, vous définissiez des classes de trafic en fonction de critères de correspondance incluant les protocoles, les listes de contrôle d'accès (ACL) et les interfaces d'entrée. Avec la fonctionnalité LLQ, la stratégie CBWFQ bénéficie d'une capacité de mise en file d'attente à priorité stricte (PQ).
- Il existe trois modèles d'implémentation de QoS: Remise au mieux (Best effort), les Services intégrés (IntServ) et les Services différenciés (DiffServ).
- Le modèle d'architecture IntServ a été développé pour répondre aux besoins des applications en temps réel, telles que la vidéo à distance, les conférences multimédia, les applications de visualisation de données et la réalité virtuelle.
- Le modèle de QOS DiffServ spécifie un mécanisme simple et évolutif pour la classification et la gestion du trafic réseau. La conception du modèle DiffServ s'affranchit des limites associées aux modèles de remise au mieux et des services intégrés.
- Il existe trois catégories d'outils de qualité de service : les outils de classification et de marquage, les outils de prévention des encombrements et les outils de gestion de l'encombrement.
- La classification détermine la classe de trafic à laquelle les paquets ou les trames appartiennent.

Qu'est-ce que j'ai appris dans ce module? (Suite)

- Pour classer le trafic des couches 2 et 3, plusieurs méthodes sont possibles, dont les interfaces, les listes de contrôle d'accès et les mappages de classes. Il peut également être classé au niveau des couches 4 à 7 à l'aide de la fonction NBAR (Network Based Application Recognition).
- La gestion de la congestion repose sur des méthodes de mise en file d'attente et de planification, selon lesquelles le trafic excédentaire est mis en mémoire tampon ou en attente (et parfois abandonné) lorsqu'il attend d'être envoyé sur une interface de sortie.
- Les outils de prévention des encombrements permettent de surveiller les charges de trafic sur les réseaux afin d'anticiper et d'éviter les encombrements au niveau des congestions du réseau commun et de l'internet avant que les encombrements ne deviennent un problème.
- Cisco IOS inclut une solution de prévention de l'encombrement basée sur la détection anticipée aléatoire pondérée (WRED).


CISCO