
Qualité de Service (QoS) et gestion du trafic dans les réseaux

La gestion de la QoS dans les réseaux

- La problématique de la QoS
- Les principes de l'ATM
- La gestion de la QoS et du trafic en ATM
- La boîte à outils de la gestion de la QoS dans les réseaux IP
- Les mécanismes de la gestion de la QoS et du trafic dans les réseaux IP
- MPLS et la gestion de la QoS dans les réseaux Ethernet

Partie 1 Problématique de la Qualité de Service (QoS) et de la gestion du trafic

Problématique

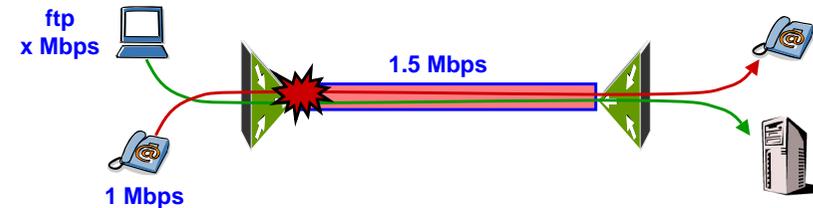
- Pourquoi la QoS ?
- Qu'est-ce que la QoS ?
- La gestion du trafic dans un réseau asynchrone
- Les paramètres de QoS et de la performance des réseaux

A-t-on besoin de QoS ?

- **NON !**
 - Loi de Moore appliquée aux ports des routeurs
 - explosion de la bande passante des LANs (10 Mbps → 100 Mbps → 1 Gbps → 10 Gbps)
 - explosion de la bande passante des WANs (DWDM avec 32 couleurs)
- **OUI !**
 - il y a des zones dans le réseau où la bande passante est plus rare
 - les applications multimédias sont de plus en plus gourmandes
 - certaines applications ont des besoins forts de QoS (VoIP, IPTV)
 - mélange d'applications aux besoins différents
 - résister aux surintensités de trafic (attaques de type DDOS, trafics d'urgence liés aux catastrophes)
- **SANS DOUTE !**
 - dans les portions du réseau où la bande passante est rare

Coexistence d'applications de nature différente

- **Dans un contexte de bande passante limitée, cohabitation :**
 - d'applications temps réel (bornées en bande passante, contraintes temporelles strictes)
 - avec des applications de transfert de données (non bornées en bande passante, contraintes temporelles lâches)



- **Fonctions à mettre en place :**
 - marquer les paquets pour distinguer plusieurs classes de QoS
 - isoler les classes de QoS entre elles
 - optimiser les ressources du réseau
 - processus d'admission d'appels

Deux approches différentes !

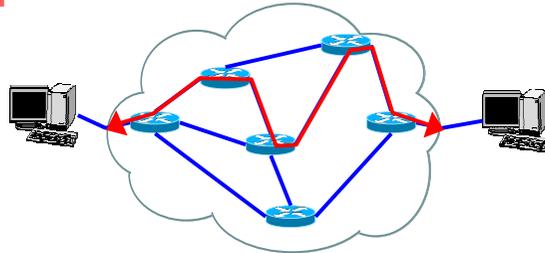
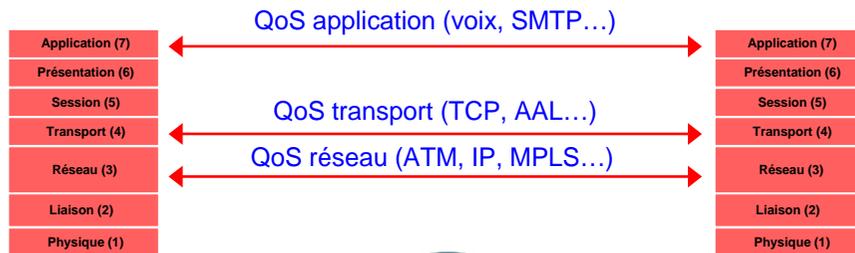
- **Approche « opérateur de réseau » (ATM)**
 - le réseau fait presque tout (gestion du trafic, QoS, maintenance,...)
 - le client ne fait presque rien (sauf AALs)
- **Approche « universitaire » (IP)**
 - le réseau ne fait presque rien (essentiellement du routage)
 - le client fait beaucoup (contrôle de flux par TCP,...)



Problématique

- **Pourquoi la QoS ?**
- **Qu'est-ce que la QoS ?**
 - QoS : laquelle?
 - la QoS dans un SLA technique
 - les éléments de QoS dans un SLA réseau
- **La gestion du trafic dans un réseau asynchrone**
- **Les paramètres de QoS et de la performance des réseaux**

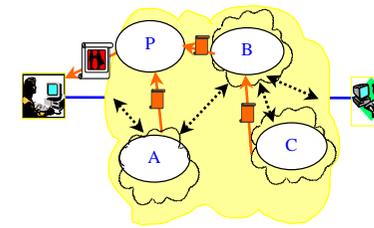
La QoS et les couches ISO



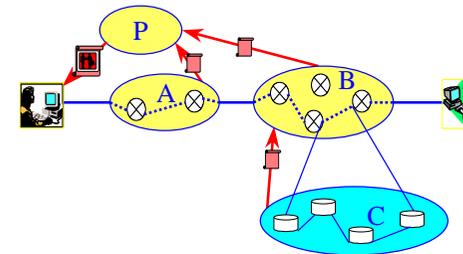
Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 9 - 12 décembre 2007

QoS et SLA (Service Level Agreement)



- La QoS du service final s'appuie sur :
 - une relation entre le client et son fournisseur (l'opérateur de réseau) : le SLA
 - une récursivité entre les différents fournisseurs



- Vocabulaire
 - **partage (« apportionnement »)** : répartition de la QoS entre les différents réseaux/fournisseurs
 - **traduction (« mapping »)** : traduction de la QoS de couche N (ATM) en une QoS de couche N-1 (SDH)

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 10 - 12 décembre 2007

QoS absolue ou relative

- QoS relative
 - mode **non connecté**
 - classification à l'entrée (DiffServ)
 - différentes technologies possibles :
 - IP (DiffServ)
 - Ethernet (802.1p et 802.1Q)
 - MPLS (EXP, label de LSP)
- QoS absolue
 - mode **connecté**
 - réservation effectuée le long d'un chemin
 - mise en œuvre d'une politique d'acceptation des nouveaux appels (CAC), à la main ou à l'aide d'un protocole de signalisation
 - ATM : PNNI
 - MPLS : RSVP-TE, CR-LDP

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 11 - 12 décembre 2007

Problématique

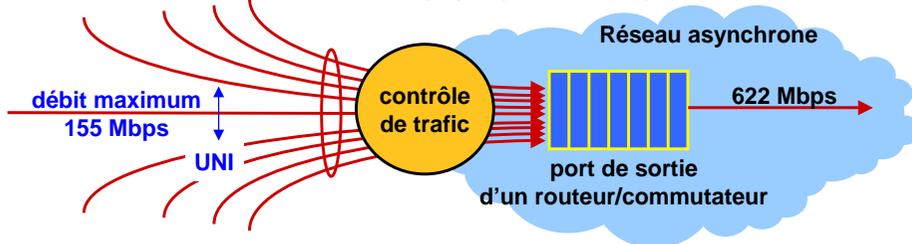
- Pourquoi la QoS ?
- Qu'est-ce que la QoS ?
- Les contraintes des réseaux asynchrones sur la gestion du trafic (ATM, IP ou Ethernet)
- Les paramètres de QoS et de la performance des réseaux

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 12 - 12 décembre 2007

L'accès client

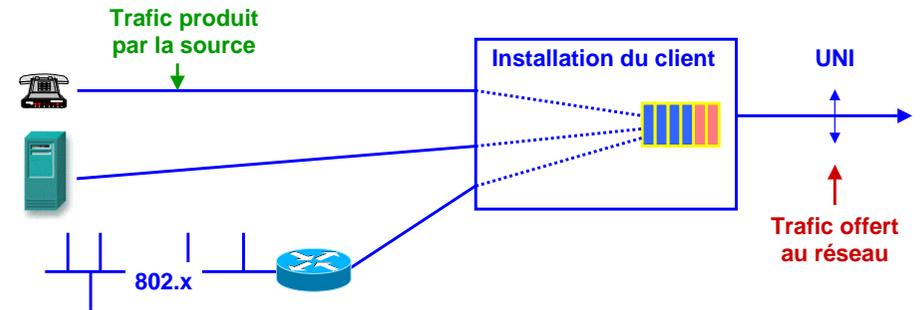
- A l'accès des clients, le débit physique peut valoir : PDH (2 Mbps, 34 Mbps), SDH (155 Mbps, 622 Mbps), GE (1 Gbps)
- Dans le réseau, le débit physique commuté par port peut valoir : 155 Mbps, 622 Mbps, 2.5 Gbps, 10 Gbps
- **Le débit est « à la carte »** : sur le lien d'accès au réseau, le rythme d'insertion de l'information à transmettre est à l'initiative de la source de trafic dans la limite du débit physique du multiplex



- Si on veut garantir une QoS dans le réseau, il faut mettre en place un **contrôle de trafic à l'accès**

Le multiplexage

- Plusieurs étages de multiplexage séparent le point d'émission de l'information de la source et le point d'entrée dans le réseau
- Plusieurs affluents peuvent être associés à la connexion : OAM (maintenance) et RM (gestion des ressources)



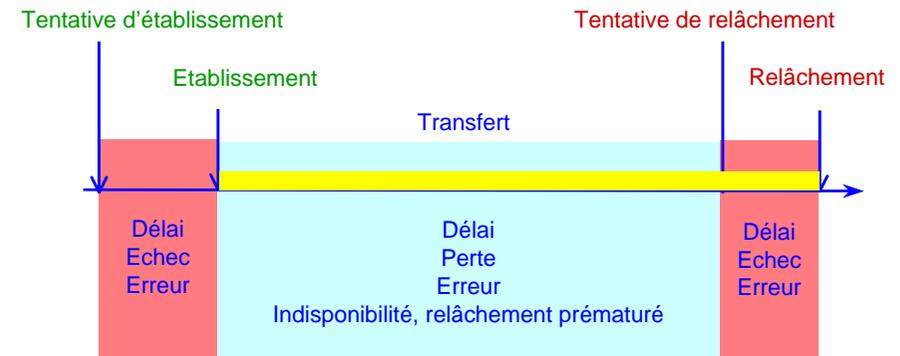
- Il existe une **gigue de multiplexage inhérente** : le contrôle de trafic doit donc être **tolérant** !

Problématique

- Pourquoi la QoS ?
- Qu'est-ce que la QoS ?
- La gestion du trafic dans un réseau asynchrone
- **Les paramètres de QoS et de la performance des réseaux**
 - les paramètres clés de QoS
 - comment gérer la QoS ?
 - QoS absolue ou relative ?

Définition des paramètres de QoS

- Pour définir un paramètre de QoS, on utilise :
 - des événements de référence
 - une réalisation de référence
 - une observation statistique (taux, moyenne, quantile,...)



Classification des paramètres de QoS

- **Modèle 3x3 de l'ITU-T**

uniquement pour les techniques
en mode connecté

	Etablissement	Transfert	Relâchement
Réalisation	taux de blocage	taux de perte	taux d'échec au relâchement
Exactitude	taux d'établissements erronés	taux d'erreurs	probabilité de relâchement prématuré
Délais	délai d'établissement	délai de transfert, gigue	délai de relâchement

- **Paramètres dérivés : les réalisations sont des franchissements de seuils observés sur d'autres paramètres de QoS (ex disponibilité)**

Les dégradations de QoS et leurs remèdes

Dégradation de QoS	Solution dans le terminal	Solution dans le réseau
Information corrompue (erreurs, pertes)	- redondance du codage à l'émission - insertion d'infos de bourrage	- contrôle du trafic dans le réseau (rejet sélectif, contrôle de flux)
Délai de transfert variable	- horodatage - ajout d'un délai - restitution d'une période fixe	- contrôle de la variation du délai - espacement
Délai de bout-en-bout élevé	- augmentation de la fréquence d'émission	- sélection de chemins plus rapides - accélération de la commutation

Conclusion

- **On a besoin de gérer la QoS et le trafic**
 - dans les portions du réseau où la bande passante est rare (il y en a toujours !)
 - si on gère des applications temps réel
- **Les applications**
 - de données peuvent s'adapter en débit et en délai
 - temps-réel demandent un niveau de QoS élevé
- **Les paramètres clés de la QoS sont :**
 - le délai (et sa variation)
 - la perte
 - le débit
 - la disponibilité

Partie 2 ATM (Asynchronous Transfer Mode)

ATM

ATM est une technique élaborée dans les années 80 par les opérateurs de réseau (technique « **carrier class** ») et destinée à supporter de façon intégrée (« Réseau Numérique à Intégration de Services Large Bande ») les nouveaux services de télécommunication

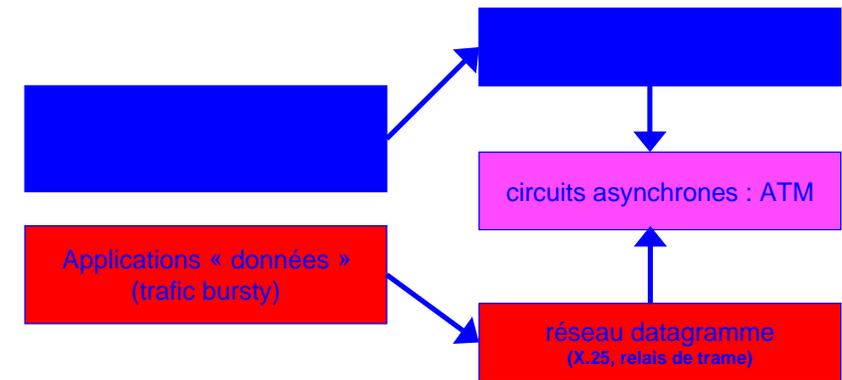
● Principes fondamentaux

- Les applications
- ATM : une technique fédératrice
- Le modèle fonctionnel

● La couche ATM

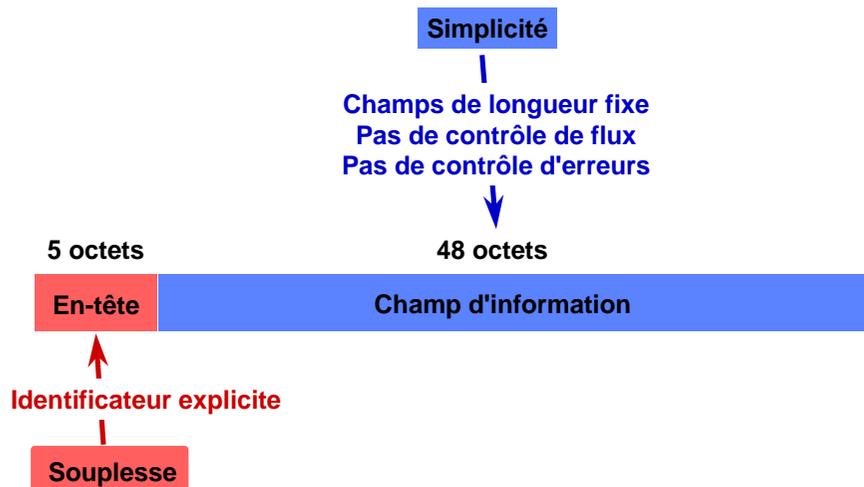
- La couche d'adaptation aux supports physiques
- La couche d'adaptation aux applications

Objectifs de l'ATM : faire un réseau multi-service



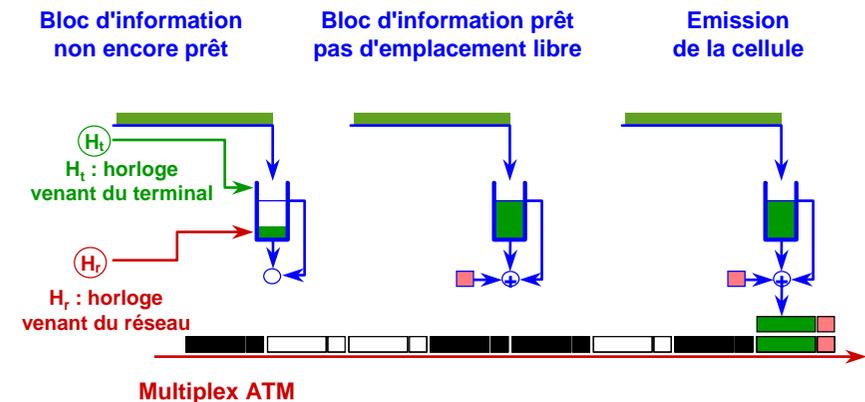
- Un unique mode de transfert pour toutes les applications
- Des couches d'adaptation pour porter toutes les applications
- Des « services de couche ATM » différents pour supporter des contraintes (trafic/QoS) différentes

L'ATM : un mode connecté basé sur la cellule



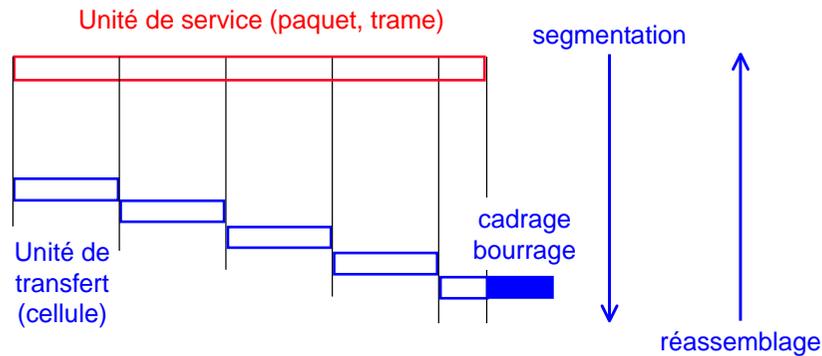
1^{er} principe fondamental de l'ATM : indépendance temporelle

- Source et réseau fonctionnent de manière asynchrone



2ème principe fondamental de l'ATM : indépendance sémantique service - réseau

- Segmentation et réassemblage (se fait dans une couche d'adaptation)

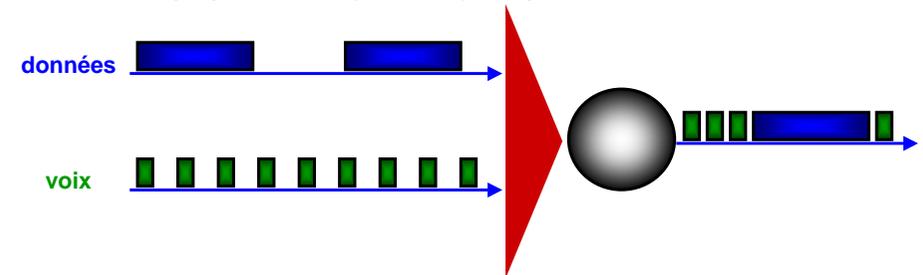


- Il n'y a pas de protocole spécifique dans le réseau (sauf signalisation)

Entrelaçage des flux

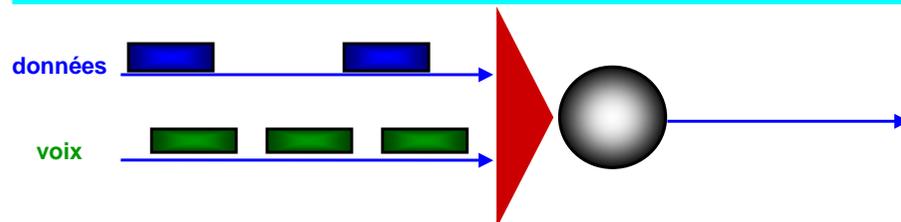
- Multiplexage

- données : paquets longs (1500 octets), sans contrainte temps réel
- voix : paquets courts (50 octets), temps réel



- Pb : à chaque étage de multiplexage, les petits paquets doivent attendre que les grands paquets passent (délai, gigue)
- ex : sur un lien E3, l'émission de 1500 octets nécessite 353 µs !
- Il faut des cellules de taille identique pour tous les services ! C'est que ce fait Cisco avec la fonction LFI (Link Fragmentation Interface)

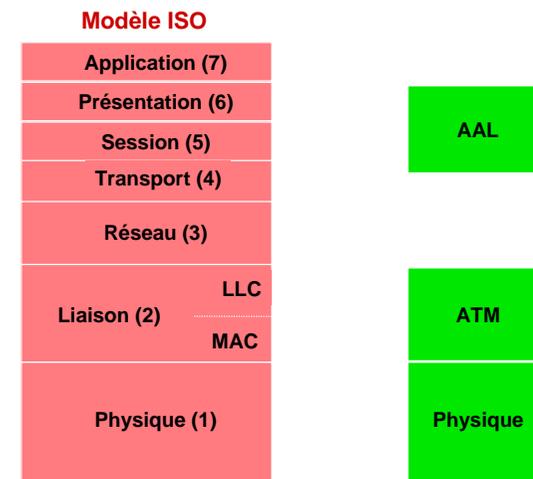
Entrelaçage des flux



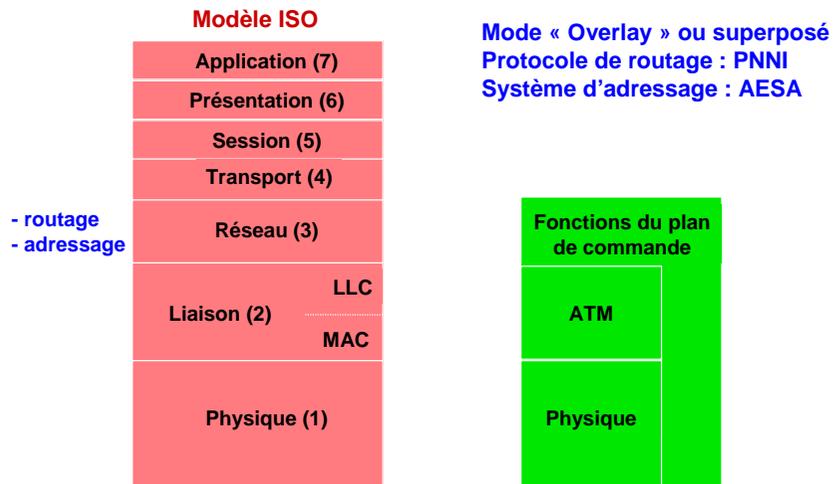
Taille cellule	E1 (2 Mbps)	E3 (34 Mbps)	STM-1 (155 Mbps)	STM-4 (622 Mbps)	STM-16 (2,5 Gbps)
50 octets	208 µs	11,8 µs	2,7 µs	0,668 µs	0,167 µs
1500 octets	6250 µs	353 µs	80 µs	20 µs	5 µs

- Pb : le temps d'émission d'une cellule dépend du débit du lien et de la taille de la cellule (sérialisation). D'où délai, gigue
- Pour garantir le temps réel, il faut des cellules de petite taille !

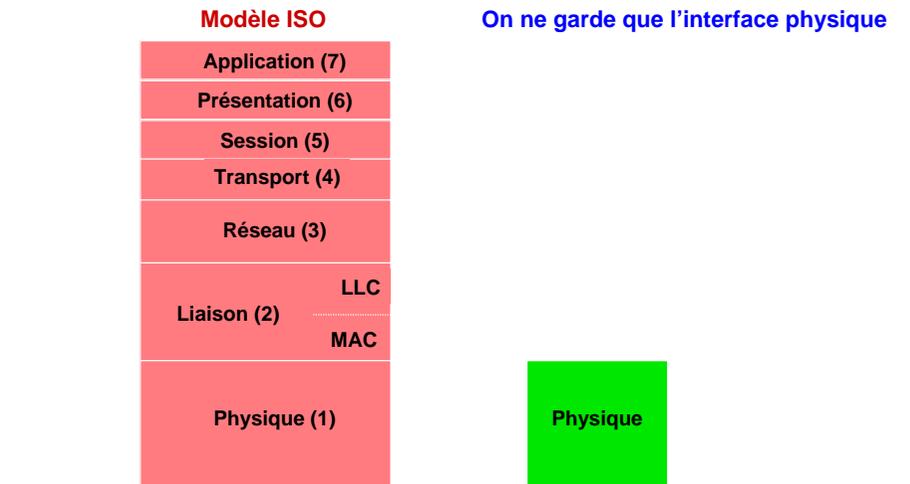
L'ATM : une technique de niveau 2



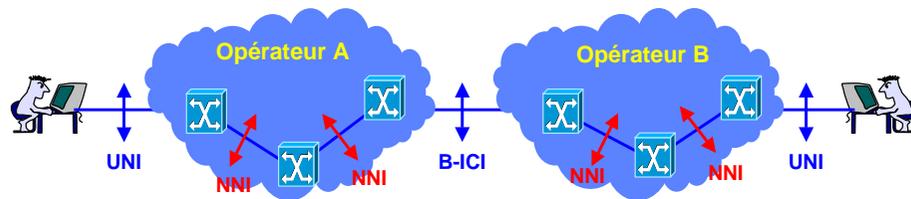
Mais aussi de niveau 3...



... ou de niveau 1



Les interfaces

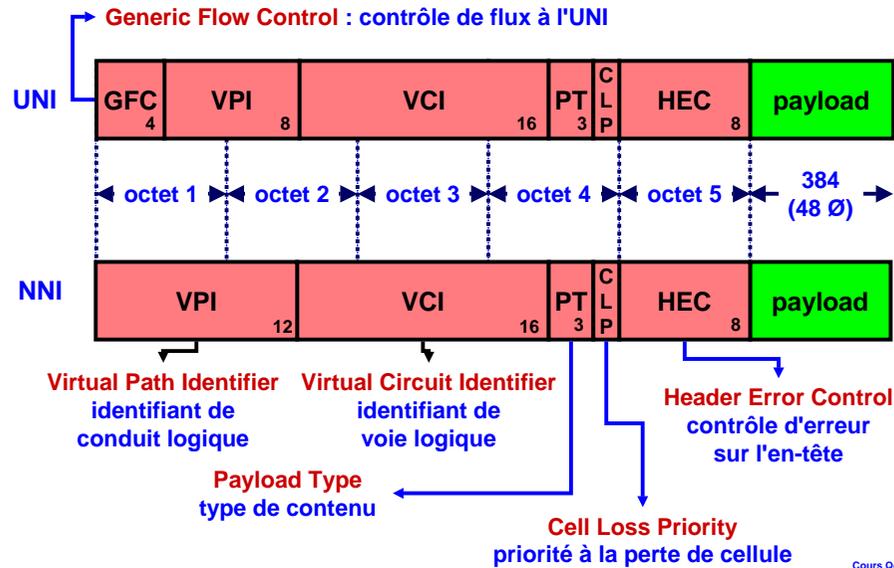


UNI : User Network Interface
 NNI : Network to Network Interface
 B-ICI : BISDN Inter-Carrier Interface (ATM Forum)
 INI : Inter Network Interface (ITU-T)

ATM

- Principes fondamentaux
- La couche ATM
 - la structure de la cellule et les différents champs de l'en-tête
 - l'acheminement
- La couche d'adaptation aux supports physiques
- La couche d'adaptation aux applications

En-tête : structure générale



Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 33 - 12 décembre 2007

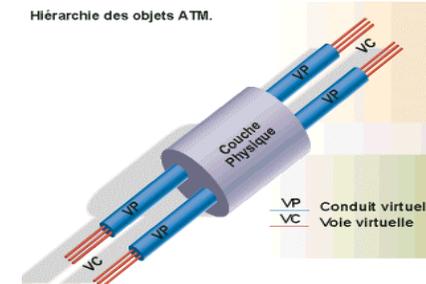
En-tête : champs VPI et VCI (Virtual Path et Virtual Circuit Identifiers)



Un identifiant logique

- double
 - 4 096 VP (Virtual Path ou Conduit virtuel) au NNI (et 256 à l'UNI)
 - 65 536 VC (Virtual Circuit ou Voie logique)
- indépendant du débit du lien

Hierarchie des objets ATM.



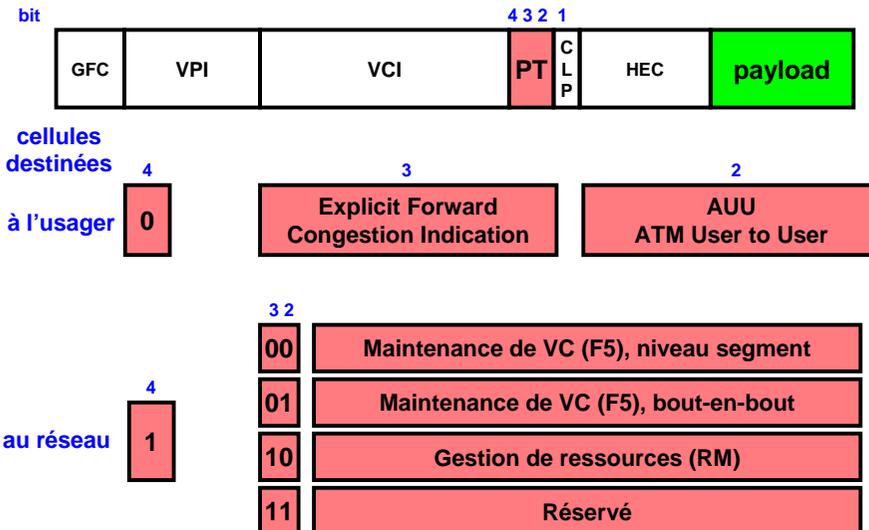
Pourquoi une hiérarchie VP-VC ?

- simplification des procédures d'établissement des connexions
- simplification de la gestion du trafic (débits constants et débits variables)

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 34 - 12 décembre 2007

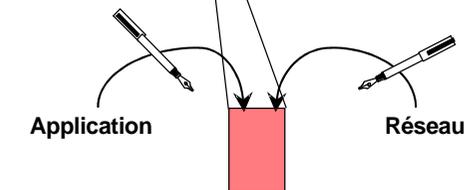
En-tête : champ PT (Payload Type)



Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 35 - 12 décembre 2007

En-tête : champ CLP (Cell Loss Priority)



bit CLP : 0 ou 1

- en cas de congestion, on détruit prioritairement les cellules CLP=1

bit positionné par l'application

- permet de distinguer 2 types de cellules

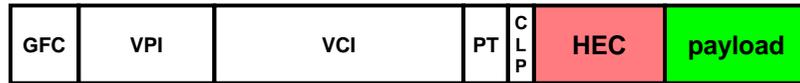
bit positionné par le réseau

- permet de déclasser les cellules non-conformes à un contrat de trafic (tagging ou marquage)

Copyright © 2007 - Jean-Raymond LOUVION

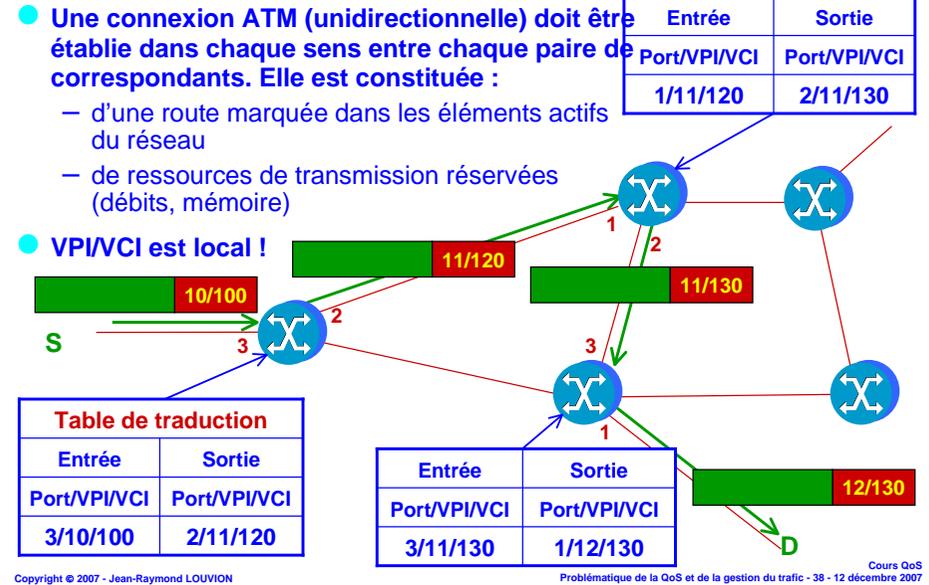
Cours QoS
Problématique de la QoS et de la gestion du trafic - 36 - 12 décembre 2007

En-tête : champ HEC (Header Error Control)



- **Fonction de la couche physique**
- **Calcul**
 - à partir des 4 octets de l'en-tête, calcule 1 octet (méthode du reste de la division par un polynôme générateur)
 - corrige une erreur simple
 - détecte certaines erreurs multiples
- **Délimitation des cellules**
 - vérification bit à bit (recherche du cadrage)
 - puis vérification cellule par cellule (confirmation du cadrage)

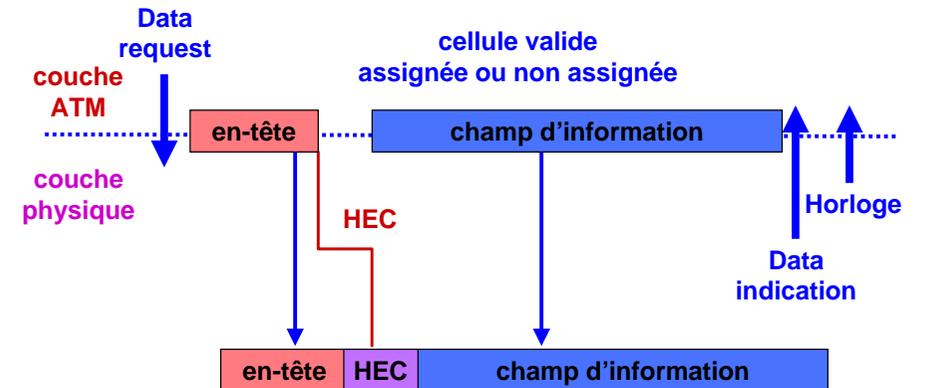
L'acheminement en ATM (1/2)



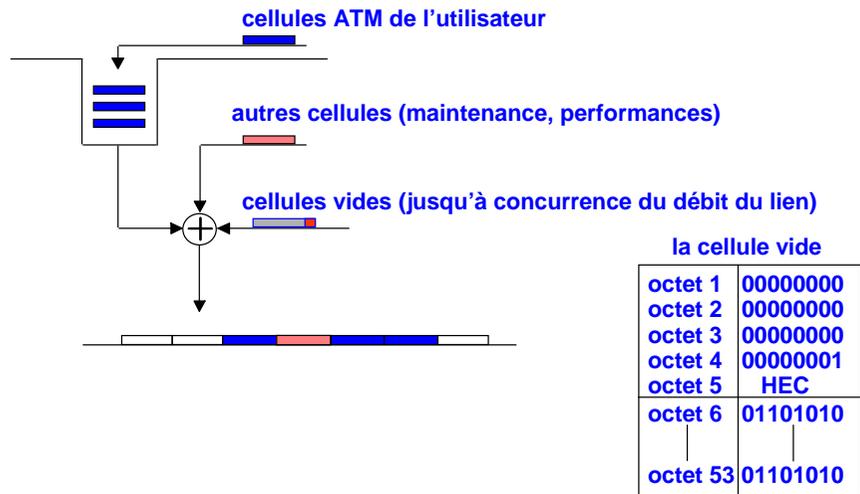
ATM

- Introduction
- La couche ATM
- La couche d'adaptation aux supports physiques
- La couche d'adaptation aux applications

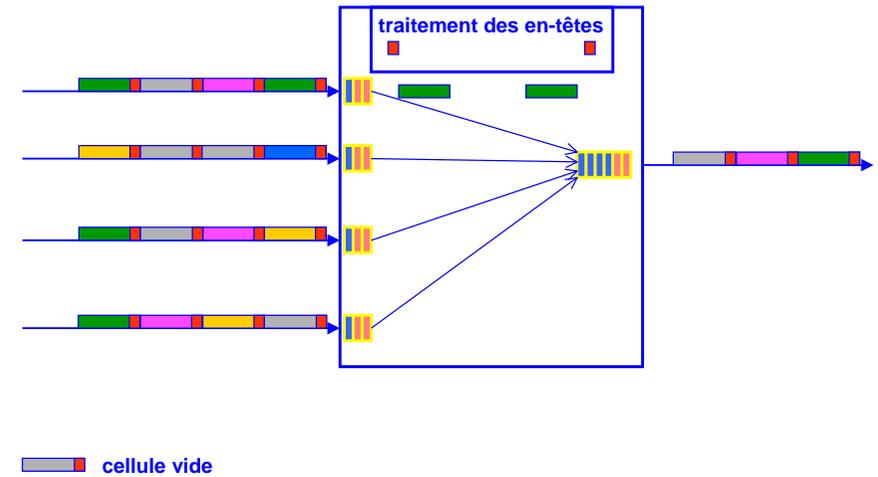
Primitives échangées entre couche physique et couche ATM



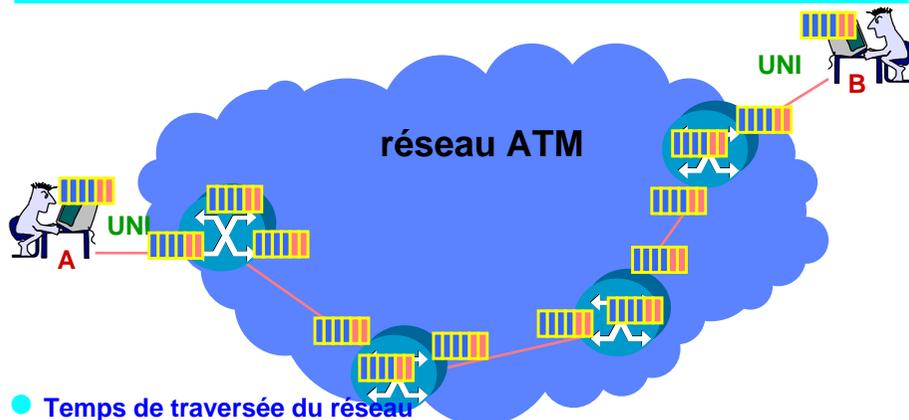
Adaptation des débits



Fonctionnement schématique d'un commutateur ATM



Un réseau ATM est un système de files d'attente

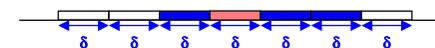


● Temps de traversée du réseau

- partie fixe :
 - temps de propagation,
 - temps de sérialisation (temps-cellule ou temps d'émission de la cellule sur un lien donné)
- partie variable : temps de traversée des files d'attente

Le temps-cellule δ

- C'est le temps de "séréalisation", i.e. le temps nécessaire pour émettre une cellule au débit du lien physique
- C'est aussi le temps séparant deux cellules consécutives (utiles ou vides) sur le lien physique
- δ (en μs) = $424 / \mu$ (en Mbps), où μ est le débit du lien physique
- Exemples numériques :
 - E3 (33,920 Mbps) : 12,5 μs
 - STM-1 (149,760 Mbps) : 2,831 μs
 - STM-4 (599,040 Mbps) : 0,708 μs



ATM

- Introduction
- La couche ATM
- La couche d'adaptation aux supports physiques
- La couche d'adaptation aux applications
 - les fonctions de la couche AAL : la segmentation-réassemblage et la convergence (le reste)
 - AAL1 : AAL pour services temps réel
 - AAL5 : AAL pour services de données
 - AAL2 : AAL pour services temps réel à débit réduit

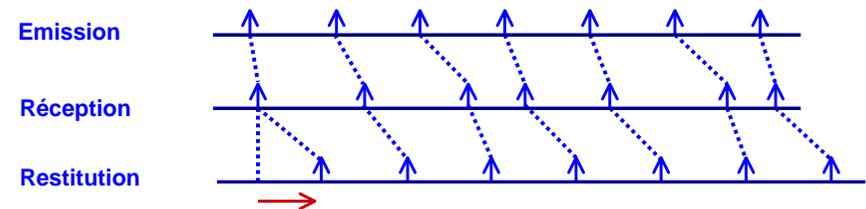
Typologie des AALs

	AAL	1	2	X	5 ↑ 3/4	X
<i>Contrainte temporelle</i>	stricte	stricte	lâche	lâche		
<i>Débit</i>	constant	variable	variable	variable		
<i>Mode de connexion</i>	connecté	connecté	connecté	non connecté		
<i>Type de service</i>	temps réel	temps réel bas débit	données			

AAL1

- AAL1 : services temps réel à débit constant
- Fonctions :
 - respect des contraintes temporelles
 - méthode adaptative
 - méthode SRTS (Synchronous Residual Time Stamp)
 - protection en avant des flux sensibles (vidéo)
 - entrelaçage : ne supprime pas les erreurs, mais étale leur conséquences, facilitant leur traitement par un code correcteur
 - correction des cellules perdues : corrige jusqu'à 4 cellules perdues dans un bloc de 128 pour une redondance relativement faible (3,8%)
- Champ particulier :
 - 1 octet qui permet la numérotation des cellules de 1 à 8
 - il reste donc 47 octets disponibles !

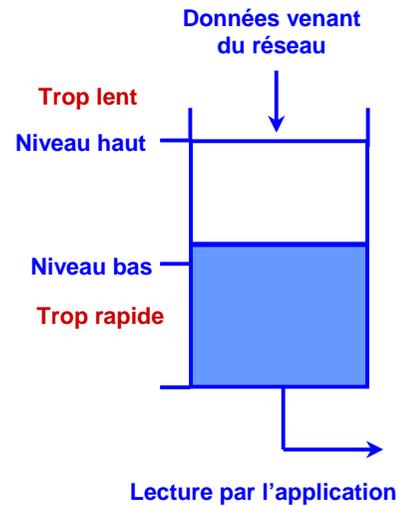
La gigue cellule : le problème



- Les cellules, émises régulièrement par l'application, arrivent irrégulièrement dans le récepteur : **gigue cellule ou CDV (Cell Delay Variation)**
- Effets possibles : famine ou débordement du récepteur
- Solution : la méthode adaptative
 - retarder la première cellule
 - buffer de réception

La gigue cellule : la méthode adaptative

- on démarre avec une file remplie à moitié
- on asservit la fréquence de lecture sur le niveau de remplissage de la file
 - si la file se remplit, la lecture est trop lente
 - si elle se vide, la lecture est trop rapide
- La file peut être utilisée pour ajouter le retard initial de restitution



AAL5

- AAL5 : transport des services de données
- L'AAL5 de base fournit un mode non assuré (un message détecté incomplet ou erroné est éliminé, aucune retransmission n'est prévue)
- Options :
 - mode assuré (ce sont les couches supérieures qui s'en chargent !)
 - applications autres que le transfert de données (ex : VoD : transmission d'un message incomplet ou erroné)

AAL5 : la structure

● Sous-couche Convergence



- le **bourrage** ajouté a pour but d'aligner la taille totale sur celle des cellules ATM
- le champ **Fin** est constitué, entre autres de :
 - Longueur : taille du message (0 ... 64 koctets)
 - CRC 32 : détection d'erreurs sur l'ensemble du message

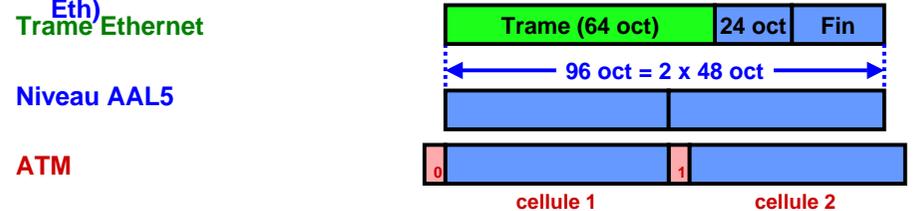
● Sous-couche SAR

- le champ SAR est vide
- on ne garde qu'une seule fonction : délimitation des messages (AUU)

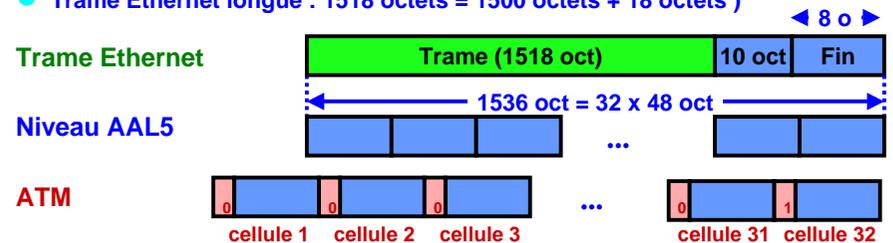


Transport d'Ethernet sur ATM : cas simplifié

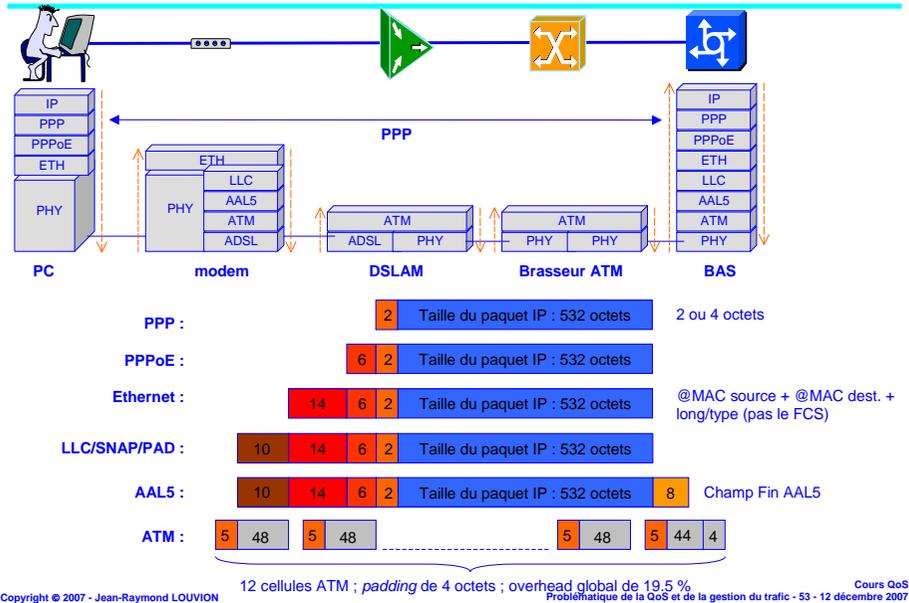
- Trame Ethernet courte : 64 octets = 46 octets (data) + 18 octets (en-tête Eth)



- Trame Ethernet longue : 1518 octets = 1500 octets + 18 octets)



Transport d'Ethernet sur ATM : collecte ADSL



En résumé (1)

- ATM combine simplicité (comme le réseau téléphonique) et souplesse (comme les réseaux de données)
- Deux principes fondamentaux :
 - indépendance temporelle source-réseau et réseau-réseau
 - indépendance sémantique source-réseau
- Modélisation ISO : ATM est une technique de niveau 2 (il peut également être considéré comme de niveau 1 ou 3)
- La taille de la cellule (5+48 octets) résulte d'un compromis
- L'en-tête comprend les champs :
 - VPI et VCI : identifiant des connexions
 - PT : 3 bits pour diverses fonctions (maintenance, gestion des ressources)
 - CLP : priorité à la perte
 - HEC : correction des erreurs
 - GFC : contrôle de flux, uniquement à l'accès du réseau

En résumé (2)

- L'ATM est une technique en mode connecté :
 - on établit un chemin avant d'échanger des informations
 - les connexions sont pt-pt ou pt-mpt
- L'adaptation aux supports physiques recourt à des fonctions originales :
 - découplage entre le débit ATM et le débit du support
 - cadrage cellule permettant de se passer d'une structure de trame
- L'ATM s'adapte à tout type de support physique existant (PDH, SDH, ADSL, 25.6 Mbit/s ATM Forum)
- La couche AAL gère la QoS exigée par les applications :
 - aspects temporels : absorption de la gigue cellule,...
 - aspects sémantiques : pertes et erreurs de cellules
 - 3 types d'AALs sont utilisées :
 - AAL1 : applications temps réel
 - AAL5 : applications de données avec protocole simplifié
 - AAL2 : applications temps réel à bas débit

Partie 3 Gestion du trafic et de la QoS en ATM

Gestion du trafic et la QoS en ATM

- La définition et le contrôle des débits
 - l'effet du profil de trafic sur un réseau
 - la notion de débit-crête
 - les algorithmes :
 - Algorithme d'Espacement Virtuel (VSA)
 - Leaky Bucket
- La performance
- Les contrats de trafic
- La boîte à outils de la gestion de trafic
- Le routage à la QoS
- Quelques applications

Les différentes formes d'un trafic ATM ...

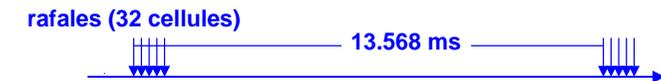
- 1 Mbps émis régulièrement (émulation de circuit à la sortie de la source de trafic)



- 1 Mbps perturbé (multiplexage de la zone privée)

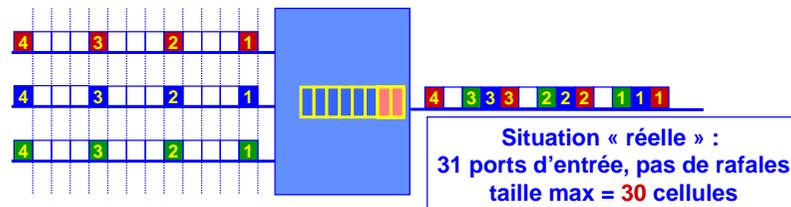


- 1 Mbps émis en rafales dont la taille dépend de l'application (par exemple à la sortie d'un routeur)

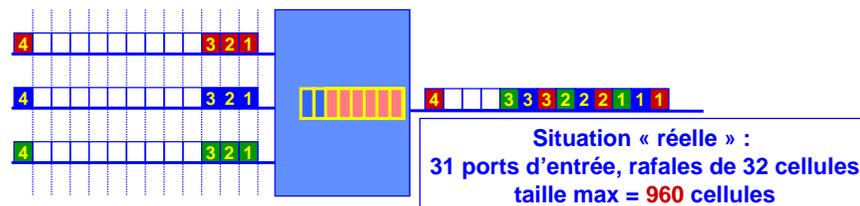


... et leur effet sur le réseau

Flux espacés

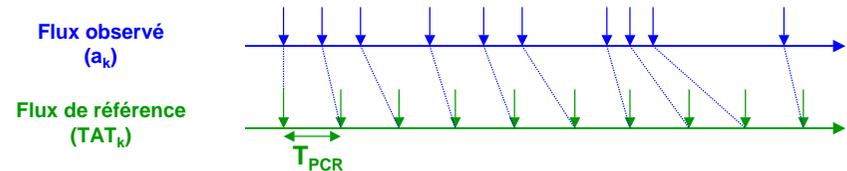


Flux en rafales



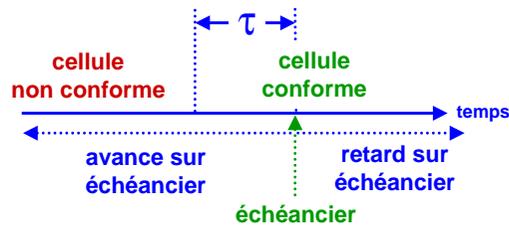
Définitions : débit-crête (PCR) et Cell Delay Variation (CDV)

- Idée :
 - on choisit un flux de référence :
 - à débit constant (ou parfaitement espacé)
 - défini par son **débit PCR (en bps)** ou par son **intervalle inter-cellule $T_{PCR} = 424 / \text{PCR}$ (en s)**
 - on mesure les écarts (CDV) entre le flux observé et le flux de référence



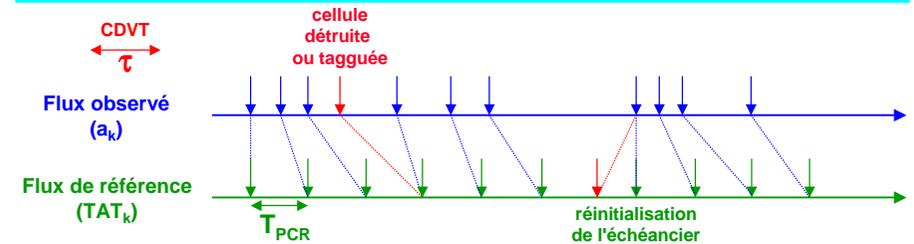
- Remarques :
 - le choix d'un débit-crête n'est pas unique !
 - si le débit de référence est trop petit (T_{PCR} trop grand), la CDV diverge !
 - de la même façon, un lien sera représenté par son débit p (en bps) ou son temps-cellule δ (en s)

Contrôle du débit-crête, algorithme #1 : l'algorithme d'espacement virtuel VSA (T_{PCR} , τ)



- τ est appelée tolérance de gigue de multiplexage ou CDVT (Cell Delay Variation Tolerance)
- L'algorithme VSA va être utilisé dans les contrats de trafic

Règles complémentaires

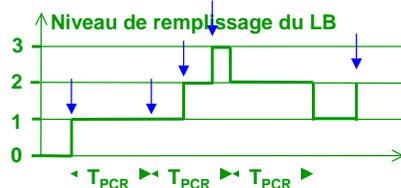
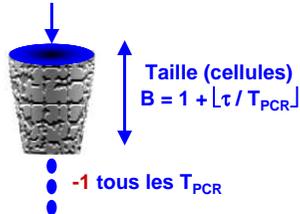


- Si une cellule est non-conforme, l'échéancier ne progresse pas !
– la cellule non-conforme doit être éliminée du comptage !
- Si une cellule est « en retard », l'échéancier est resynchronisé sur la date d'arrivée de cette cellule !
– un flux de trafic silencieux n'acquiert pas de droit à la parole !

Contrôle du débit-crête, algorithme #2 : le Leaky Bucket LB(T_{PCR} , τ)

LB à temps discret (T_{PCR} , τ)

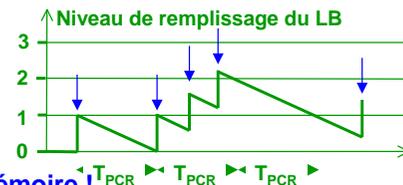
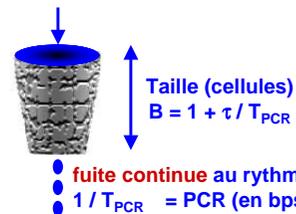
+1 à chaque cellule conforme qui arrive



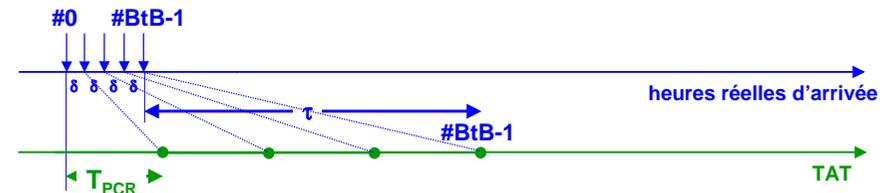
- Les cellules ne sont pas mises en mémoire !
- Le LB à temps continu est équivalent au VSA

LB à temps continu (T_{PCR} , τ)

+1 à chaque cellule conforme qui arrive



Worst Case Traffic (WCT)

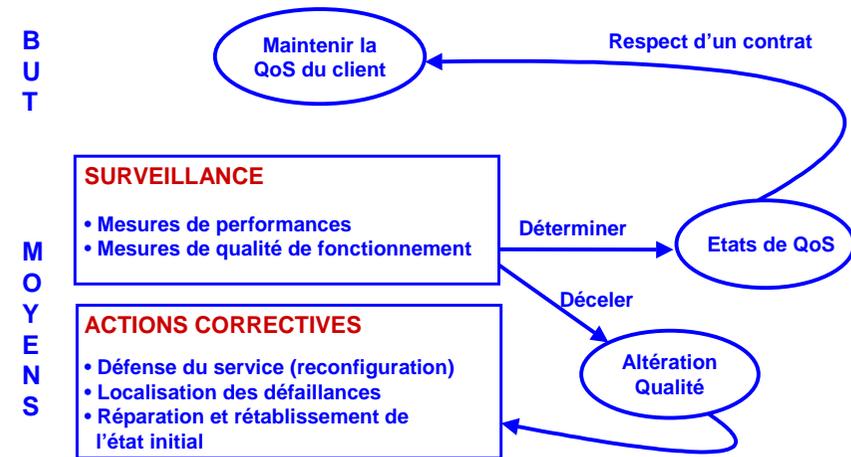


- C'est le trafic conforme à un VSA (PCR , τ) qui correspond à la pire rafale qui peut être émise au débit du lien. Elle est composée de BtB (Back-to-Back) cellules.
- On trouve : $\tau = (BtB - 1) * (T_{PCR} - \delta)$
- Et donc : $BtB = 1 + \lfloor \tau / (T_{PCR} - \delta) \rfloor$
- A.N. :
– lien STM-1 ($\delta = 2.831 \mu s$), $PCR = 1$ Mbps, $BtB = 32$, alors $\tau = 13.056$ ms
– lien STM-1, $PCR = 10$ Mbps, $\tau = 200 \mu s$, alors $BtB = 6$

Gestion du trafic et la QoS en ATM

- La définition et le contrôle des débits
- **La performance**
 - les paramètres de performance en ATM
 - les objectifs de performance
 - partage des objectifs de performance entre opérateurs ATM
- Les contrats de trafic
- La boîte à outils de la gestion de trafic
- Le routage à la QoS
- Quelques applications

Fonctions des flux OAM (Operation, Administration & Maintenance)

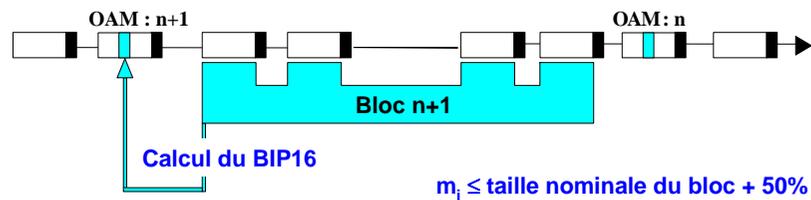


Comment estime-t-on la performance ?

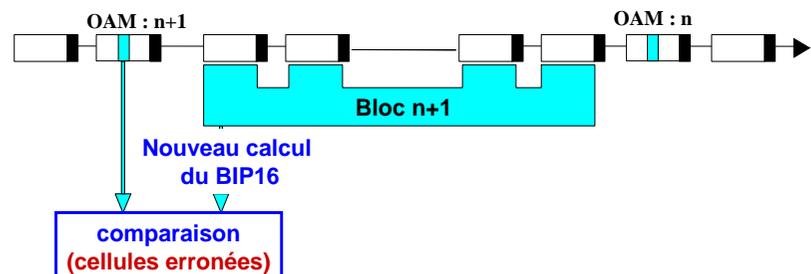


I.610

Emission



Réception

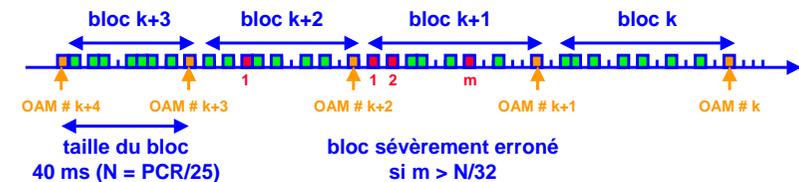


Paramètres de performance sémantiques



I.356

- **Bloc** : séquence de N cellules consécutives (de la connexion) compris entre deux cellules OAM de performance consécutives



- Paramètres

- SECBR (Severely Errored Cell Block Ratio)
- hors SECBs :
 - CER (Cell Error Ratio)
 - CLR (Cell Loss Ratio)
 - CMR (Cell Misinsertion Rate)

Paramètres de performance temporels



I.356

● Moyenne des délais d'un ensemble de cellules (CTD : Cell Transfer delay)

● Cell Delay Variation (CDV 2-points)

- pour chaque cellule, on calcule la CDV : délai - délai d'une cellule de référence (le délai minimum)
- On détermine le quantile à 10^{-3} de la distribution



● Attention : ne pas confondre avec la CDV 1-point, utilisée pour la conformité !



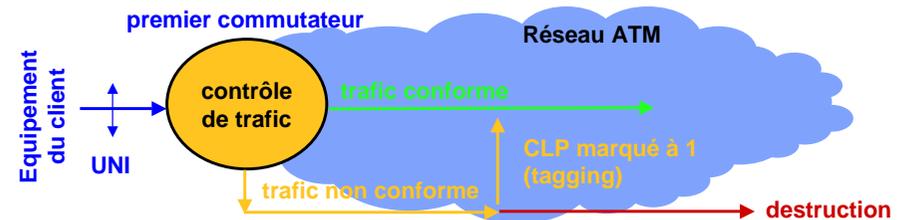
Objectifs de QoS à l'ITU-T

	contrat de trafic	CTD	CDV	CLR _{0+v}	CLR ₀	CMR	CER	SECBR
QoS 1 sévère	DBR SBR1 ABT	400 ms	3 ms	$3 \cdot 10^{-7}$	NA	1/jour (défaut)	10^{-6} (défaut)	10^{-4} (défaut)
QoS 2 tolérante	DBR SBR1 ABT	U	U	10^{-5}	NA	1/jour (défaut)	10^{-6} (défaut)	10^{-4} (défaut)
QoS 3 2 niveaux	SBR2/3 ABR	U	U	NA	10^{-5}	1/jour (défaut)	10^{-6} (défaut)	10^{-4} (défaut)
QoS 4 non spécifiée	tous	U	U	U	U	U	U	U
QoS 5 sévère 2 niveaux	SBR2/3 ABR	400 ms	6 ms	NA	$3 \cdot 10^{-7}$	1/jour (défaut)	10^{-6} (défaut)	10^{-4} (défaut)

Gestion du trafic et la QoS en ATM

- Définition et contrôle des débits en ATM
- La performance
- Les contrats de trafic
 - contrats pour flux à débit constant
 - contrats pour flux à débit variable
 - contrats adaptatifs
- La boîte à outils de la gestion de trafic
- Le routage à la QoS
- Quelques applications

Le contrat de trafic



Contrat de trafic = Description du trafic offert + QoS requise



I.371

ATC (ATM Transfer Capability)
descripteur trafic source
tolérances de gigue (CDVT)
mécanismes de conformité

Classe de QoS



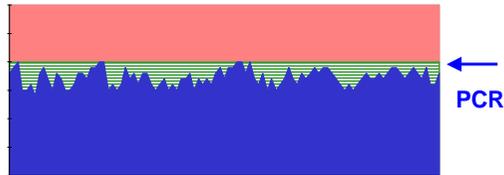
UNI 4.0

Catégorie de service
descripteur trafic source
tolérances de gigue (CDVT)
mécanismes de conformité

Paramètres individuels de QoS

Contrat DBR - CBR

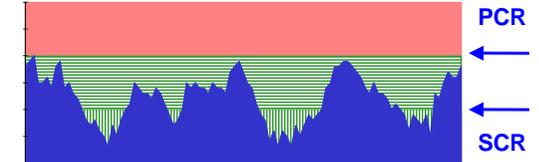
- ITU-T :
DBR (Deterministic Bit Rate)
- ATM Forum :
CBR (Constant Bit Rate)



- Applications
 - flux à débit constant
 - mode quasi-circuit (pas d'utilisation statistique des ressources)
- Descripteurs de trafic
 - PCR associé à une tolérance de gigue CDVT
- 2 options
 - DBR-1 : déclaration séparée données 0+1 et OAM
 - DBR-2 : déclaration agrégée
- Ingénierie simple : on additionne les PCR

Contrat SBR - VBR

- ITU-T :
SBR (Statistical Bit Rate)
- ATM Forum :
VBR (Variable Bit Rate)
VBR-rt et VBR-nrt



- Applications : multiplexage à l'aveugle de sources à débit variable
- Descripteurs de trafic
 - PCR associé à une tolérance de gigue CDVT
 - SCR (Sustainable Cell Rate) associé à une taille maximum de rafale (MBS : Maximum Burst Size) et une tolérance
- 3 options
 - VBR1 : PCR(0+1) et SCR(0+1) ; les cellules non-conformes sont détruites
 - VBR2 : PCR(0+1) et SCR(0) ; les cellules non-conformes sont détruites
 - VBR3 : idem VBR2, mais les cellules non-conformes au SCR(0) sont déclassées

Contrats UBR et UBR+

- ITU-T et ATM Forum :
UBR (Unspecified Bit Rate)



- Applications : uniquement données peu sensibles
- Pas de police
- Pas de garantie de Qualité de service
- Mécanismes de prévention de congestion (EPD, PPD)
- Descripteurs de trafic
 - PCR (facultatif et à titre indicatif)
 - et MCR (cas de UBR+)
- Ingénierie ultra simple
 - UBR : pas de fonction d'acceptation des appels (une nouvelle connexion est acceptée dans tous les cas)

Contrats «gigue forte» et contrats «gigue faible»

- Contrats « gigue faible »
 - Gigue minimum résultant du multiplexage de flux déterministes
 - La valeur de la gigue dépend de la charge et du débit du lien de sortie
 - Calcul :
borne supérieure du temps de traversée d'une file $nD / D / 1$ (n grand)
 $\tau_{PCR} / \delta = \max(T_{PCR} / \delta, \alpha \cdot (1 - \delta / T_{PCR}))$
où $\alpha = 80$ pour une charge de 85 %
 - Connexion à 2 Mbps sur un lien STM-1 : $\tau_{PCR} = 79$. $\delta = 223 \mu s$
- Contrats « gigue forte »
 - Valeur de gigue permettant de prendre en compte des rafales de taille réaliste (32 ou 200 cellules)
 - Nécessité de mettre en place un espacement (réduction de la gigue)
 - Rafales de 32 cellules sur une connexion à 2 Mbps sur un lien STM-1
 $\tau_{PCR} = 31 \cdot (T_{PCR} - \delta) = 6\ 481 \mu s$

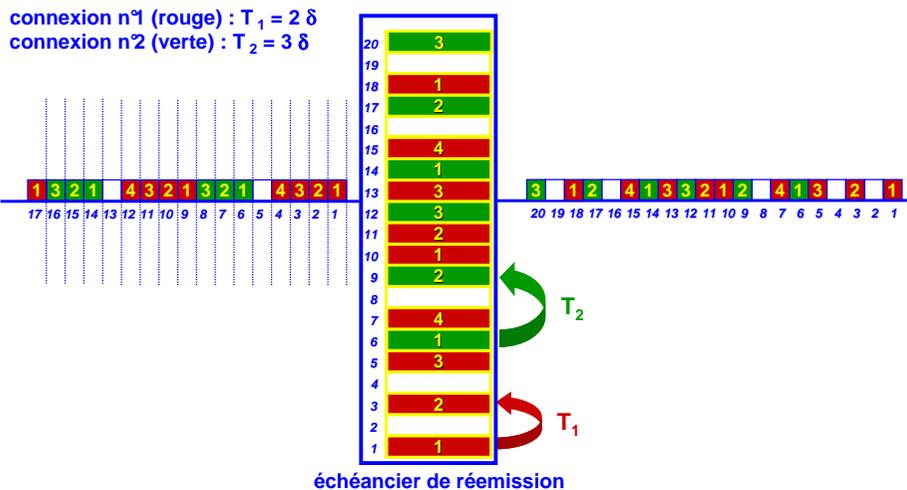
Gestion du trafic et la QoS en ATM

- La définition et le contrôle des débits
- La performance
- Les contrats de trafic
- **La boîte à outils de la gestion de trafic**
 - **L'espaceur des cellules**
 - **Les mécanismes EPD/PPD**
- Le routage à la QoS
- Quelques applications

L'espaceur : le concept

- **Constat**
 - la taille des mémoires nécessaires dans un réseau dépend beaucoup des caractéristiques des flux de trafic
 - pour garantir un $CLR = 10^{-10}$ sur des artères chargées à 85%, la mémoire de multiplexage doit contenir environ 64 rafales
 - et ce, dans tous les commutateurs du réseau
- **Idee : rendre les flux de trafic les plus « lisses » possible**
- **L'espaceur (shaping) s'appuie sur un des algorithmes de contrôle (VSA ou Leaky Bucket)**
- **Intérêt**
 - la mise en mémoire ne se fait qu'une seule fois (à l'entrée du réseau)
 - le dimensionnement est alors simple (on est ramené au cas du CBR)

L'espaceur : la réalisation

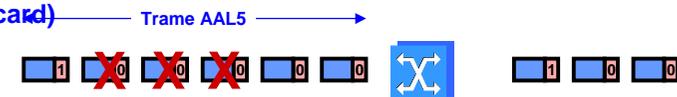


- **Il faut une grande mémoire : 256 Kcellules**

Mécanismes d'évitement de congestion au niveau trame AAL5 : PPD et EPD

- En cas de congestion, la perte d'une seule cellule appartenant à une trame AAL5 rend la trame inutilisable par l'application : le **goodput** (débit applicatif réellement transmis) peut alors devenir nul !
 - paquets de 1500 octets (trame AAL5 de 32 cellules), $CLR = 3\%$, sans EPD : toutes les trames sont touchées et goodput = 0 !
 - avec EPD : 3% des trames sont perdues et goodput = 97% !

- **PPD (Partial Packet Discard)**



- **EPD (Early Packet Discard)**



Gestion du trafic et la QoS en ATM

- Définition et contrôle des débits
- La performance
- Les contrats de trafic
- La boîte à outils
- **Le routage à la QoS**
 - contrôle du trafic ATM dans les commutateurs
 - modes de gestion des connexions ATM
 - le protocole PNNI
- Quelques applications

Principes fondamentaux de PNNI

- **Permettre l'établissement de VC et VP ATM à la demande**
 - savoir router ces VC et VP à travers tout le réseau
 - respecter les contraintes spécifiées par le client
 - Qualité de Service (CTD, CDV et CLR)
 - Contrat de trafic (CBR, rt-VBR, nrt-VBR, ABR, UBR)
- **Rendre l'ATM aussi « Plug & Play » que possible**
 - minimiser le travail de configuration et de maintenance du réseau ATM
 - gérer l'évolution des ressources du réseau en temps réel
- **Passer facilement à l'échelle !**



2 protocoles : **PNNI Routing** et **PNNI Signalling**

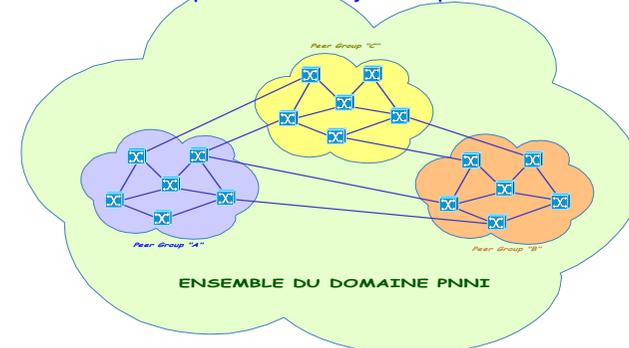
PNNI : 2 protocoles en un



- **PNNI signalling**
- **PNNI routing**
 - protocole à état des liens (comme OSPF)
 - échange 3 types de messages, dans le but de :
 - découvrir les nœuds voisins (paquets « Hello »)
 - échanger des informations topologiques (paquets PTSE : PNNI Topology State Element)
 - déclarer les adresses des équipements raccordés (paquets « Database Summary »)

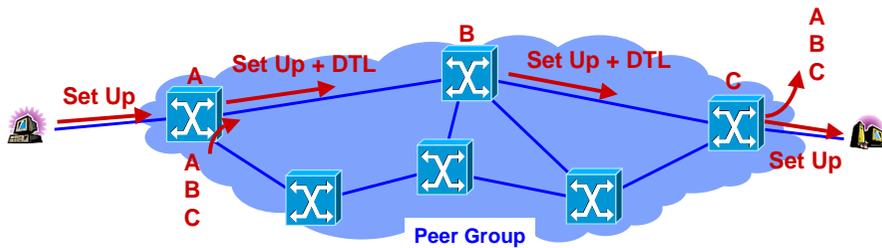
Hiérarchie : la notion de Peer Group

- Les grands réseaux sont difficile à gérer : complexité en n^2 (dans la pratique, au-delà de 20 nœuds)
- On partitionne le réseau en sous-ensembles, appelés **Peer Groups** :
 - dans un même Peer Group, chaque nœud a une vision détaillée de tous les autres
 - mais un nœud n'a qu'une vision synthétique des autres Peer Groups

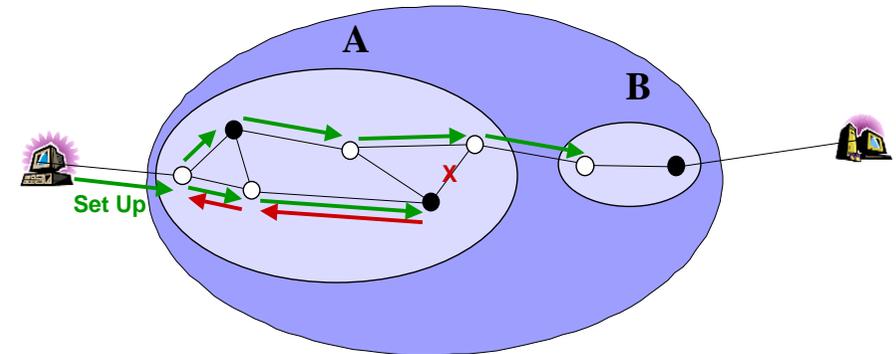


Routage d'un appel : le « source routing »

- Les tables de routage sont créées et mises à jour dynamiquement
- Si les décisions de routage se prenaient « Hop by Hop », il y aurait risque de bouclage (cf. réseaux IP et champ TTL)
- Le routage PNNI utilise le **source routing** : le nœud par lequel l'appel entre dans le réseau détermine l'ensemble du chemin (DTL : **Designated Transit List**) que la connexion devra emprunter pour le traverser
- Une DTL peut contenir jusqu'à 20 identifiants de nœuds

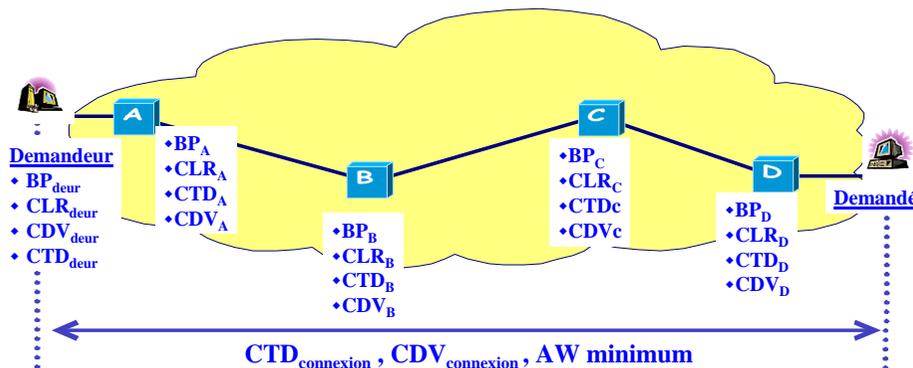


Erreur de routage : la procédure de «crankback»



- La « source route » est établie à partir de la vision du réseau qu'a le « Border Node », mais l'information peut être erronée (bande passante insuffisante, lien tombé)
- La procédure de **crankback** permet un retour au créateur de la DTL, qui est seul habilité à en calculer une autre !

Métriques et Attributs

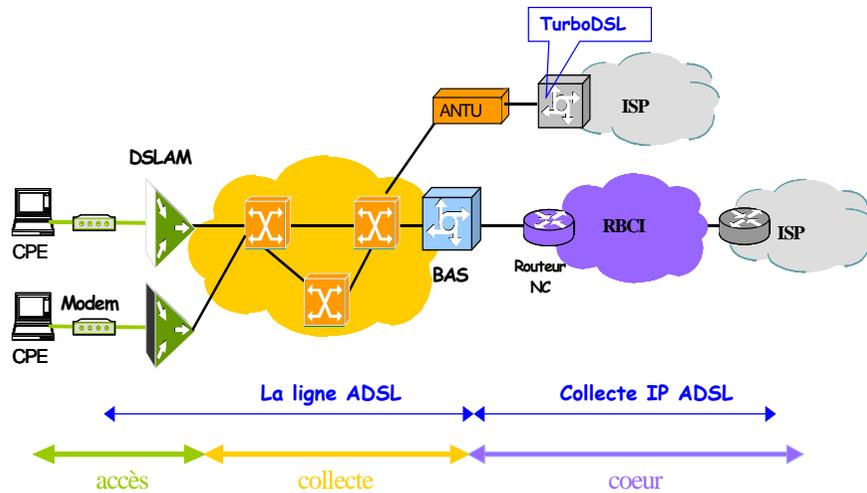


- Métrique = contrainte de bout-en-bout :**
AW (Administrative Weight), CTD max , CDV
calculée pour un chemin (on effectue la somme sur le chemin)
- Attribut = contrainte locale :** débit, CLR,...
décision pour un nœud ou un lien (décision locale)

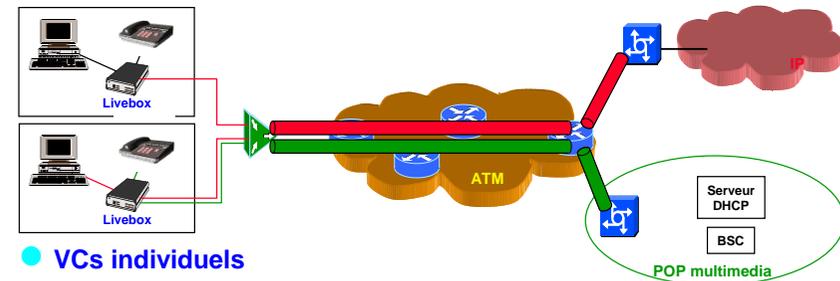
Gestion du trafic et la QoS en ATM

- Définition et contrôle des débits
- La performance
- Les contrats de trafic
- La boîte à outils
- Le routage à la QoS
- Quelques applications**

Architecture générale des services sur ADSL



VC conversationnels



VCs individuels

- VC internet UBR
- VC conversationnels CBR symétriques (PCR=60kbps, 160 kbps ou 320 kbps ; CDVT=10 ms)
 - Sens montant : limité par la Livebox
 - Sens descendant : limité par le BAS

VPs de collecte

- VP internet UBR+
- VP conversationnel d'abord CBR puis rt-VBR3 (gain statistique)

PCR	CDVT	SCR	MBS	Débit équivalent réservé
15	15 ms	1 Mbps	26	1 437 kbps / 1 142 kbps
Mbps	15	15 ms	2 Mbps	2 689 kbps / 2 238 kbps
Mbps	15	15 ms	3 Mbps	3 821 kbps / 3 297 kbps

cellules

En résumé

- Il est nécessaire de contrôler les débits à l'accès au réseau. On utilise pour cela 2 algorithmes équivalents : l'Algorithme d'Espacement Virtuel et le Leaky Bucket à temps continu
- La QoS est définie en cohérence avec les mécanismes d'estimation de la performance (CTD, 2-pt CDV, SECBR, CLR, CER, CMR)
- On définit plusieurs contrats de trafic : CBR, VBR (rt et nrt), ABR, UBR
- On dispose d'une « boîte à outils » très riche :
 - espacement
 - gestion dynamique des ressources (cellules RM)
 - mécanismes de destruction sélective de cellules (EPD, PPD)
- On dispose d'un protocole de routage qui peut prendre en compte les contraintes de QoS (PNNI)

Partie 4 La boîte à outils de la gestion de la QoS dans les réseaux IP

La boîte à outils

- **La classification des paquets et le comportement des routeurs**
 - Pourquoi et comment classer les paquets ?
 - politiques de gestion des files d'attente
- Le marquage (associé à une classification)
- L'évitement de la congestion
- Le contrôle de la bande passante
- Les limites du routage IP classique

Classification des paquets IP

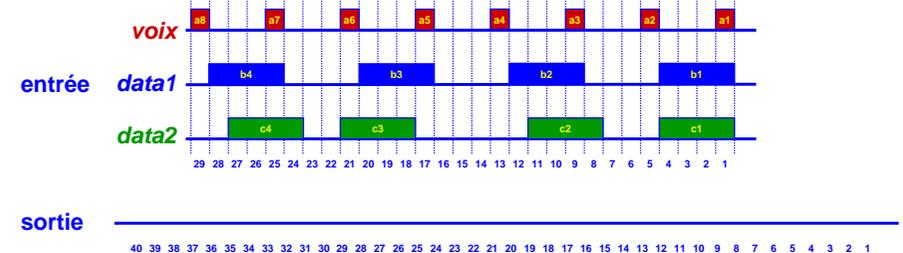
- Dans une technologie en mode "non connecté", il faut remplacer la notion de connexion par celle de flux (ou de flot) de paquets.
- La classification peut se faire par flux élémentaire (identifiés par le quintuplet : @ IP source / @ IP destination, n° port source / n° port destination, type de protocole de niveau 3) : IntServ
- ou par flux agrégés, sur tout autre champ de niveau transport ou applicatif : DiffServ, MPLS
 - au niveau 2 : @ MAC source / @ MAC destination
 - au niveau 3 : champ ToS
 - au niveau 4 : type de protocole de niveau 4 (tcp, udp, rtp,...)
 - aux niveaux supérieurs : Deep Packet Inspectors (DPI)
- On fait cette classification une seule fois à l'entrée du réseau !
- En MPLS, un tel flux sera appelé Forwarding Equivalent Class (FEC). L'ensemble des paquets IP qui le constituent sont acheminés de la même façon (même destination, même traitement dans les routeurs)

Gestion du trafic dans un routeur

- Le comportement du routeur est modélisé, de façon à pouvoir prédire son impact sur les paquets qui le traversent (délai, perte)
- **Elements à prendre en compte :**
 - les files d'attente et leur mode de gestion : taille, seuils
 - Tail Drop : destruction des paquets qui parviennent à une file déjà remplie
 - Partial Buffer Sharing : admission sélective des paquets dans la file
 - RED (Random Early Detection) : destruction sélective des paquets
 - l'échéancement (scheduling) : la façon dont les files d'attente sont servies
 - PQ (Priority Queuing)
 - PRR (Packet Round Robin)
 - DRR (Deficit Round Robin) ou WRR (Weighted Round Robin)
 - ordonnancement préemptif / non préemptif
- Ce comportement est décrit par la notion de Per Hop Behaviour (PHB) dans les services DiffServ.

Etude des politiques de gestion des files

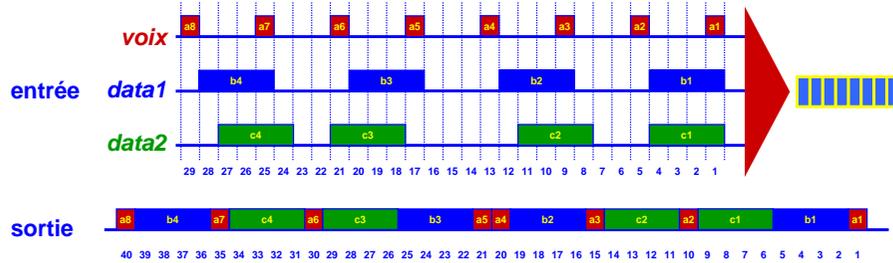
- On modélise le multiplexage de 3 flux :
 - 1 flux de voix, composé de paquets de taille 1 (la taille des paquets de voix est utilisée comme unité)
 - 2 flux de données (data1 et data2), composés de paquets de taille 4
- On suppose que les débits de toutes les interfaces sont identiques



Politique FIFO

- File d'attente unique servie en mode FIFO (First In First Out) ou FCFS (First Come First Served)

- la plus simple
- fonctionne bien en l'absence de congestion
- en cas de congestion, rejet de paquets (tail drop)

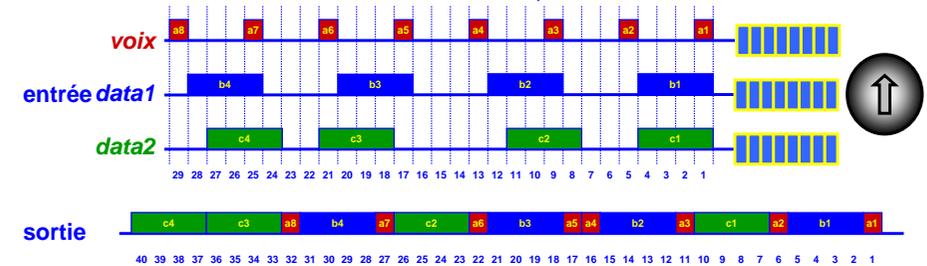


- **Avantage : simplicité, rapidité et coût**
- **Inconvénient : les flux temps réel sont perturbés par les flux de données**

Politique à « Priorités »

- Ensemble de files d'attente avec des priorités

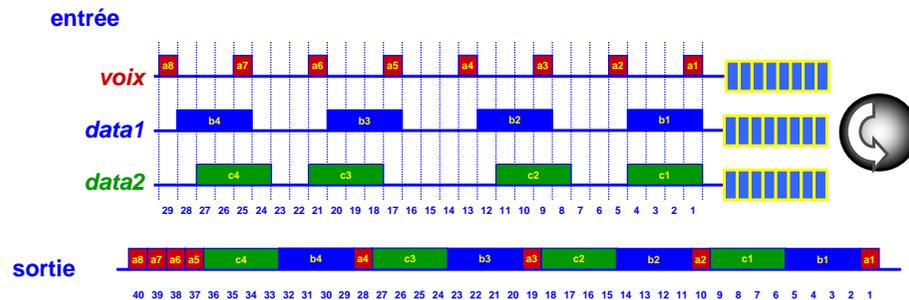
- files d'attente différentes selon les classes de trafic
- les files sont traitées dans l'ordre des priorités



- **Avantages :**
 - simplicité, rapidité et coût
 - privilégier de manière absolue un trafic par rapport à un autre
 - **Inconvénient :**
 - un type de trafic peut dominer les autres (si une file d'attente ne se vide jamais)
- Copyright © 2007 - Jean-Raymond LOUVION Cours QoS
Problématique de la QoS et de la gestion du trafic - 98 - 12 décembre 2007

Politique du « Round Robin »

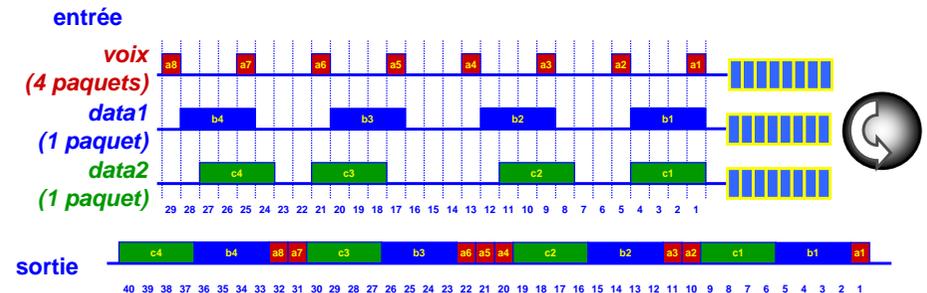
- On sert un paquet de chaque file non vide à tour de rôle



- **Inconvénient : les flux comportant des grands paquets sont favorisés, et les flux comportant des petits paquets sont défavorisés**

Politique du Weighted « Round Robin »

- On peut servir plusieurs paquets de chaque file non vide à tour de rôle



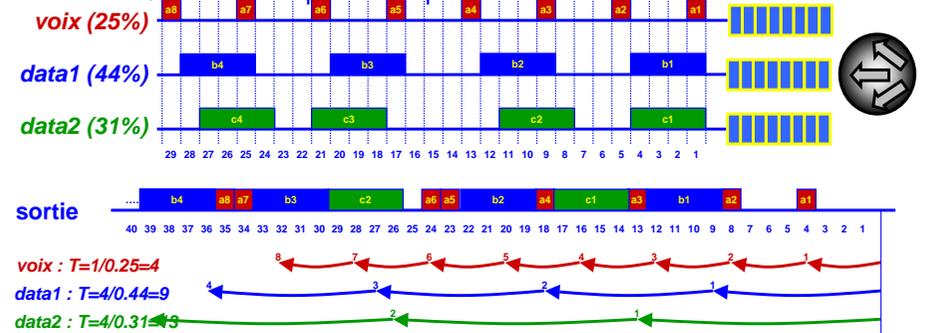
- **Avantage : gomme les effets liés à la disparité de la taille des paquets**
- **Inconvénient : il peut subsister une gigue importante pour les flux comportant des petits paquets**

Evolutions du WRR vers le WFQ

- **Fair Queuing**
 - on simule un service cyclique bit par bit
 - on sert les paquets dans l'ordre correspondant au temps auquel leur dernier bit aurait été transmis en émission bit à bit
- **Weighted Fair Queuing**
 - on accorde un poids différent aux files, en servant plus d'un bit par visite
- **Avantage**
 - c'est un mécanisme plus équitable (fair), notamment dans le cas des paquets de tailles différentes
 - mais, on ne peut éviter une certaine gigue de multiplexage (comme dans l'espacement en ATM)

Politique du « Weighted Fair Queuing »

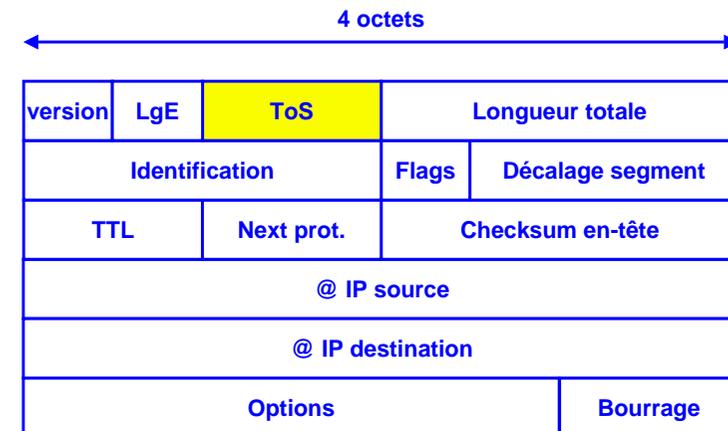
- **Multiplexage temporel : on calcule, pour chaque paquet, son heure théorique de ré-émission (en fonction de sa taille)**
 - les flux à faible volume ne sont pas asphyxiés, et les flux à fort volume ne monopolisent pas la bande passante
 - les poids permettent d'attribuer plus de bande passante aux trafics temps-réel
 - WFQ réalise un espacement par classe de trafic



La boîte à outils

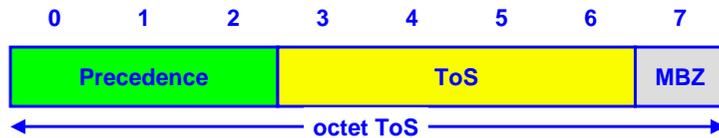
- La classification des paquets et le comportement des routeurs
- **Le marquage (associé à une classification)**
- L'évitement de la congestion
- Le contrôle de la bande passante
- Les limites du routage IP classique

Paquet IPv4



Champ ToS

- Dès l'origine de l'Internet, le champ ToS était censé définir une certaine différenciation de services au niveau du traitement des paquets IP
- Elle n'a jamais été mise en œuvre !



Mesure l'importance du paquet :

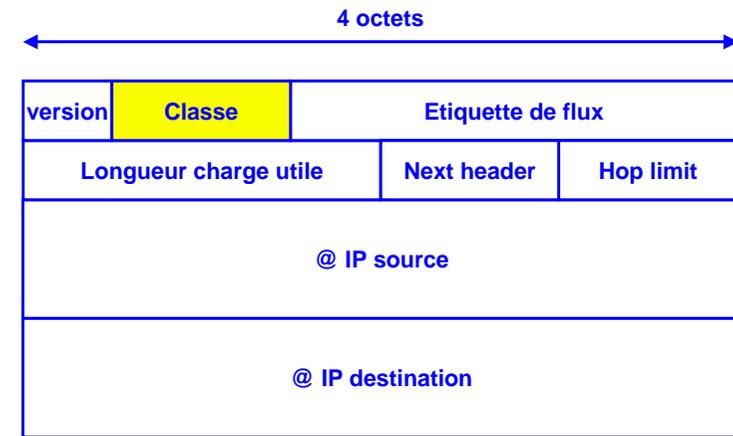
- 111 – network control
- 110 – internetwork control
- 101 – critic/ECP
- 100 – flash override
- 011 – flash
- 010 – immediate
- 001 – priority
- 000 – routine

Champ Type of Service :

- 1000 – minimize delay
- 0100 – maximize throughput
- 0010 – maximize reliability
- 0001 – minimize monetary cost
- 0000 – normal service

Must Be Zero

Paquet IPv6



- Champ Classe de 8 bits
- Utilisé par DiffServ comme champ DS

La boîte à outils

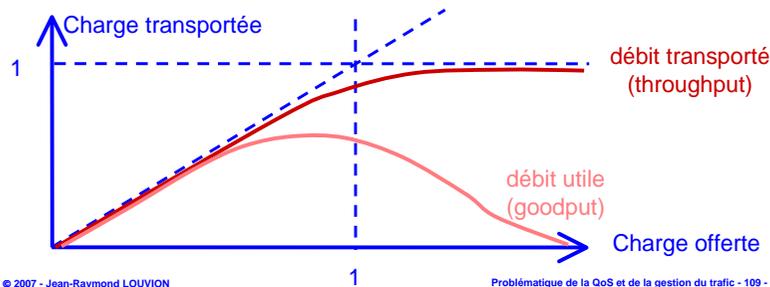
- La classification des paquets et le comportement des routeurs
- Le marquage (associé à une classification)
- L'évitement de la congestion
- Le contrôle de la bande passante
- Les limites du routage IP classique

La boîte à outils

- La classification des paquets et le comportement des routeurs
- Le marquage (associé à une classification)
- L'évitement de la congestion
 - TCP (niveau 4)
 - RED et WRED (niveau 2)
- Le contrôle de la bande passante
- Les limites du routage IP classique

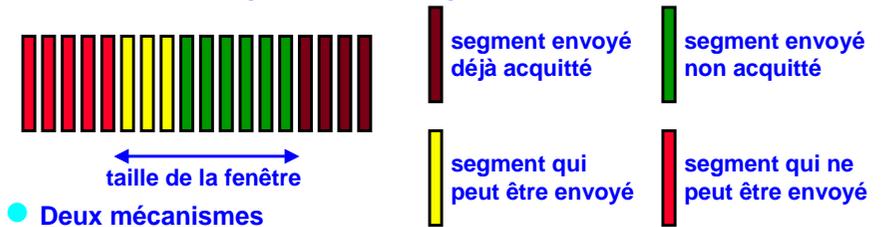
Les outils de traitement de la congestion

- **Impact de la congestion sur la QoS :**
 - perte de paquets
 - augmentation du délai et de la variation de délai
- **3 outils de traitement :**
 - destruction préventive par police/espacement (contrôle à l'accès)
 - destruction préventive par RED et WRED dans les équipements du réseau (niveau 2). Le routage n'est pas pris en compte ici !
 - récupération de bout-en-bout par TCP (niveau 4)



TCP : contrôle de flux au niveau 4

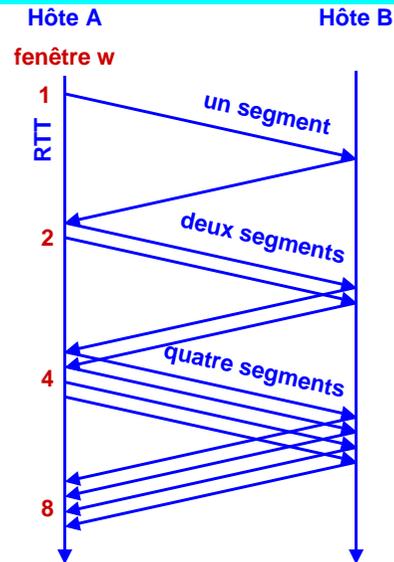
- **Mécanisme de contrôle de congestion de bout-en-bout**
 - contrôle de bout-en-bout (RTT)
 - fenêtre de congestion sur les segments



- **Deux mécanismes**
 - Slow Start
 - évitement de congestion (congestion avoidance)
 - **Débit max : $D_{max} = F / RTT = (w_{max} \cdot MSS) / RTT$**
 - où la fenêtre maximum F est la quantité d'information (en octets) qui peut être envoyée pendant une durée RTT (c-à-d w_{max} segments de taille MSS)
 - ex : si RTT=10 ms, F = 16 koctets, MSS=1460 octets, $w_{max} = 11$ et $D_{max} = 12,8$ Mbits/s
- Copyright © 2007 - Jean-Raymond LOUVION
Problématique de la QoS et de la gestion du trafic - 110 - 12 décembre 2007

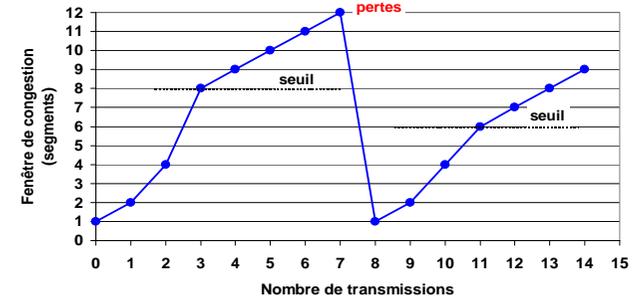
TCP : mécanisme de Slow Start

- **Mécanisme de Slow Start**
 - initialisation : fenêtre = 1
 - pour chaque segment acquitté : fenêtre = fenêtre + 1 jusqu'à (perte ou fenêtre $w >$ seuil)
- **Accroissement exponentiel de la fenêtre (doublement tous les RTT)**
- **Identification d'une perte par :**
 - expiration d'un time-out
 - duplication d'un acquittement (numéro du prochain segment attendu par le récepteur)



TCP : mécanisme d'évitement de congestion

- **Evitement de congestion**
 - Mécanisme qui se succède au Slow Start, lorsque la fenêtre $w >$ seuil
 - jusqu'à (perte)
 - { tous les w segments acquittés : fenêtre = fenêtre + 1 }
 - seuil = fenêtre / 2
 - fenêtre = 1
 - mécanisme de Slow Start

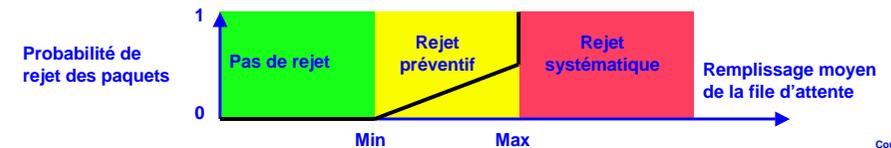


Propriétés de TCP

- **TCP est équitable :**
 - partage de la bande passante entre les connexions TCP
 - les connexions TCP obtiennent chacune la même proportion de la bande passante du goulet d'étranglement
 - ex. : 1000 flux TCP se partagent un lien STM-1
chaque flux recevra un débit moyen de 1,5 Mbps
 - l'**espacement** permet de limiter la bande passante vue par les connexions TCP, et donc de réguler, **sans perte**, le débit de ces connexions (en augmentant le RTT)
- **Pb : les applications UDP ne collaborent pas !**

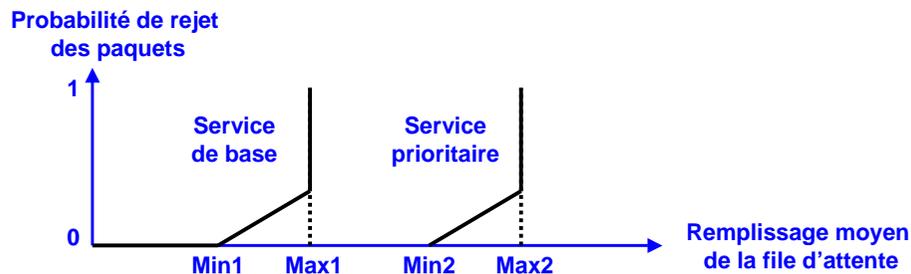
RED (Random Early Discard)

- **Avec des files d'attente gérées en « tail drop » :**
 - TCP traite les congestions trop tard !
 - les connexions TCP se synchronisent !
 - latence importante (files d'attente de grande taille)
- **RED fait un traitement préventif :**
 - un paquet perdu est interprété par TCP comme un signal de congestion, et provoque donc un ralentissement de la connexion TCP
 - idée : on rejette certains paquets avant apparition de la congestion
 - le rejet est effectué de manière aléatoire entre les différentes connexions TCP : on évite ainsi la synchronisation globale
 - les paquets sont éliminés plus tôt (latence réduite)
 - les paramètres ne sont pas faciles à régler !



WRED (Weighted RED)

- **Les seuils à partir desquels on commence à rejeter des paquets sont différents selon les priorités**
 - ex : 25% pour la priorité basse, 50% pour la priorité haute
- **Il n'y a pas d'impact sur le temps de transfert du paquet ! (on ne retarde pas les paquets)**



La boîte à outils

- **La classification des paquets et le comportement des routeurs**
- **L'évitement de la congestion**
- **Le marquage (associé à une classification)**
- **Le contrôle de la bande passante**
 - la police
 - le shaping
- **Les limites du routage IP classique**

Le Token Bucket

- Mécanisme basé sur la notion de jeton :

1 jeton = droit à émettre 1 octet

- Deux paramètres :

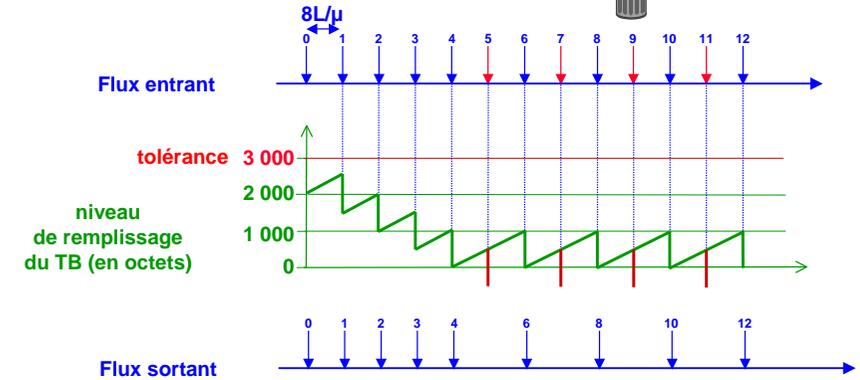
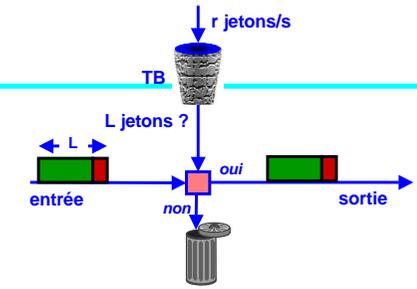
- débit de référence : r (en bps)
- tolérance : taille du bucket b (en octets)

- Dual du Leaky Bucket !



Police : un exemple

débit du lien physique : μ (en bps)
 débit de référence : $r = \mu/2$ (en bps)
 paquets de taille L : 1000 octets



Shaping et spacing

- Mise en mémoire des paquets qui arrivent et ré-émission à une date conforme à l'échéancier construit sur le rythme (débit) négocié

- Effet :

- le trafic en entrée est irrégulier (par exemple, il arrive en rafales, dont la taille est cependant bornée par la taille du Token Bucket)
- le trafic en sortie est régulièrement espacé

- Shaping (mise en forme) :

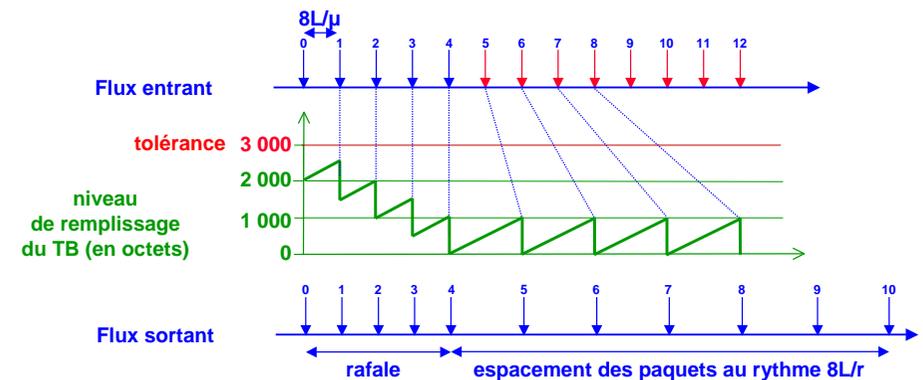
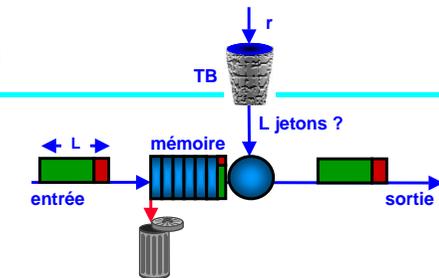
- fonctionne comme la police, mais les paquets sont mis en mémoire avant émission
- le trafic passe de façon transparente, jusqu'à la tolérance maximum, puis il est espacé au débit de référence

- Spacing (espacement) :

- le Token Bucket est réglé à la valeur minimum
- la totalité du trafic est espacé au débit de référence

Shaping (mise en forme)

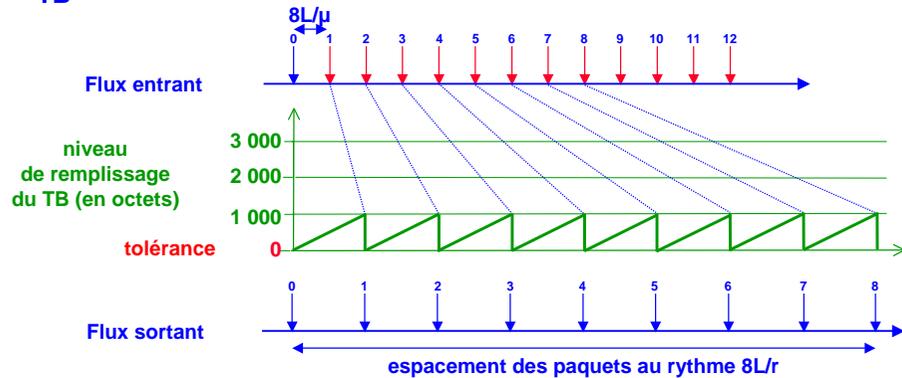
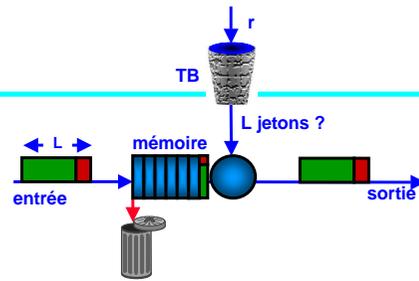
débit du lien physique : μ (en bps)
 débit de référence : $r = \mu/2$ (en bps)
 paquets de taille L : 1000 octets



Spacing (espacement)

débit du lien physique : μ (en bps)
 débit de référence : $r = \mu/2$ (en bps)
 paquets de taille L : 1000 octets

- Le premier paquet d'une rafale (mémoire vide) vide le TB

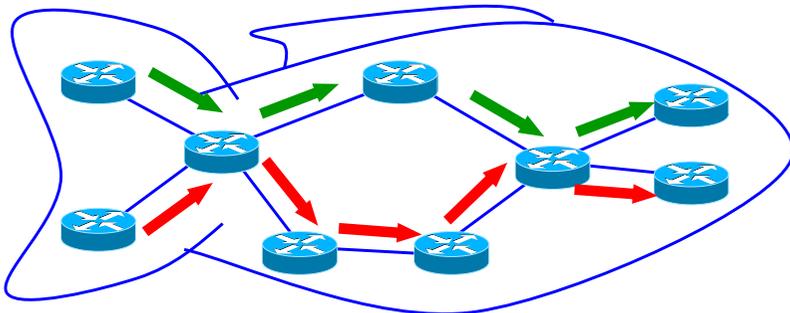


La boîte à outils

- La classification des paquets et le comportement des routeurs
- L'évitement de la congestion
- Le marquage (associé à une classification)
- Le contrôle de la bande passante
- Les limites du routage IP classique

Le routage et la QoS

- Dans les réseaux IP traditionnels, routage et QoS ont été traités de façon indépendante
 - routage : trouver un chemin qui optimise une métrique (poids administratif,...)
 - le contrôle d'admission peut refuser des trafics qu'il serait possible d'écouler sur une route alternative (le problème du poisson)
 - le routage s'appuie uniquement sur @IP destination



Solution : le routage contraint

- Idée : optimiser une métrique en respectant des contraintes
 - de performance (bande passante minimum, taux de perte maximum,...)
 - administratives (exclure ou forcer certains liens)
 - déterminées par la source ou par l'administrateur
- Moyen :
 - améliorer les protocoles de routage IGP de type « link state », en ajoutant des extensions TE (Traffic Engineering) : OSPF-TE, IS-IS-TE
 - ajouter un protocole de signalisation dédié (RSVP, CR-LDP)
- C'est ce que fait PNNI en ATM ;-)

En résumé

- Qui dit asynchrone (ATM, IP, Ethernet) dit file d'attente !
- Il y a beaucoup de mécanismes de gestion des files d'attente (FIFO, à priorités, Round Robin, WRR, WFQ, LLQ, CBWFQ)
- Mécanismes de traitement de la congestion : TCP, RED et WRED
- Il faut contrôler la bande passante (police et shaping)
- En mode non connecté, il faut classer les flux !
- Les protocoles de routage doivent prendre en compte les contraintes de QoS (extensions TE) !

Partie 5

La gestion du trafic et de la QoS dans les réseaux IP

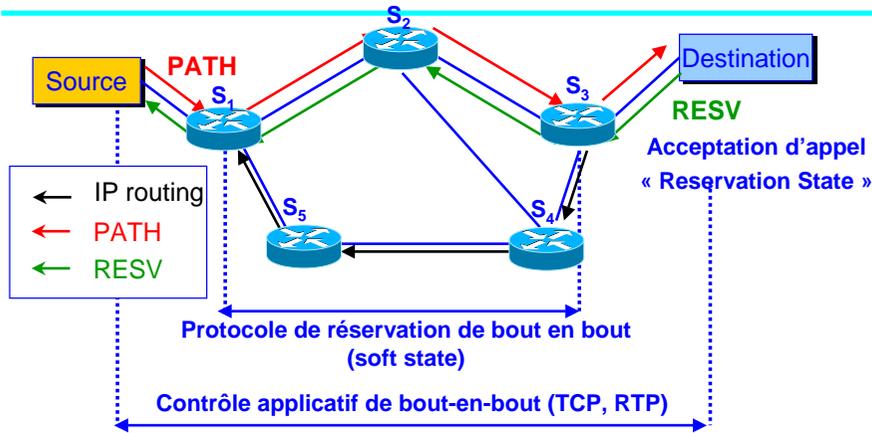
La gestion du trafic et de la QoS en IP

- **Réservation de ressources : RSVP**
 - le principe
 - le fonctionnement
 - signalisation soft state/hard state
- Services intégrés : Intserv
- Services différenciés : Diffserv

Principes de RSVP

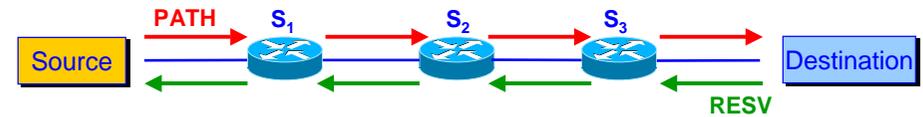
- RSVP (Resource reSerVation Protocol) a été créé pour véhiculer des informations permettant de réserver des ressources dans un réseau IP.
Ce n'est pas un protocole de routage, mais un protocole de **signalisation**
- Réservation de ressources sur le « plus court chemin » entre émetteur (source) et récepteur (destination)
 - RSVP s'interface avec un protocole de routage classique
 - chaque nœud participe au processus de réservation
 - environnement dynamique (les chemins peuvent changer en cours de session, les niveaux de réservation peuvent être modifiés)
 - conçu pour les flux unicast et multicast
- Réservation initiée par le récepteur (≠ ATM)
- Protocole à « soft state »

Le fonctionnement de RSVP (1)



- La route est déterminée par un protocole de routage classique
- Elle est ensuite « épinglée » par les messages PATH et les messages RESV suivent en retour le même chemin

Le fonctionnement de RSVP (2)



- PATH : source → destination RESV : source ← destination
- Paramètres de trafic de la source (TSpec)
 - Informations sur les ressources du réseau (AdSpec)
 - Découverte de route, annonce de ses caractéristiques et marquage du chemin inverse
 - Choix du service IntServ par le(s) récepteur(s)
 - Paramètres de trafic de la réservation faite par le(s) récepteur(s)
 - Suivi du chemin défini par le PATH
 - Réservation des ressources (RSpec)

NB : si un routeur ne traite pas RSVP, il est transparent

Les attributs des messages PATH et RESV

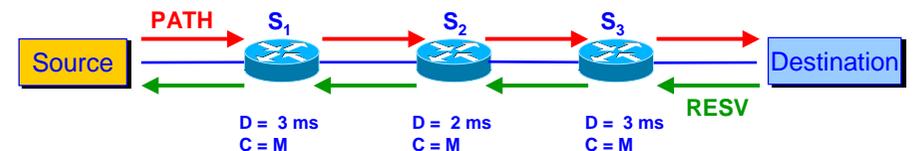


- **Message PATH**
 - champ Sender_Tspec : caractéristiques de trafic du flux
 - champ Adspec : les messages PATH collectent les informations qui vont servir au récepteur à estimer la QoS de bout-en-bout (modèle OPWA : One Pass With Advertising)
 - les messages PATH ont les mêmes adresses IP source et destination que les paquets de données (le routage se fait de bout-en-bout)
- **Message RESV**
 - champ Rspec contient les caractéristiques de la réservation à effectuer. Il peut être différent de Tspec pour deux raisons :
 - Le contrôle de trafic peut être amené à réserver des ressources différentes de celles indiquées dans Tspec
 - Dans le cas d'un flux multicast, il faut agglomérer les caractéristiques des flux allant vers les différentes feuilles
 - le routage se fait hop-by-hop (puisque la route a été marquée !)

Le fonctionnement de RSVP (3)

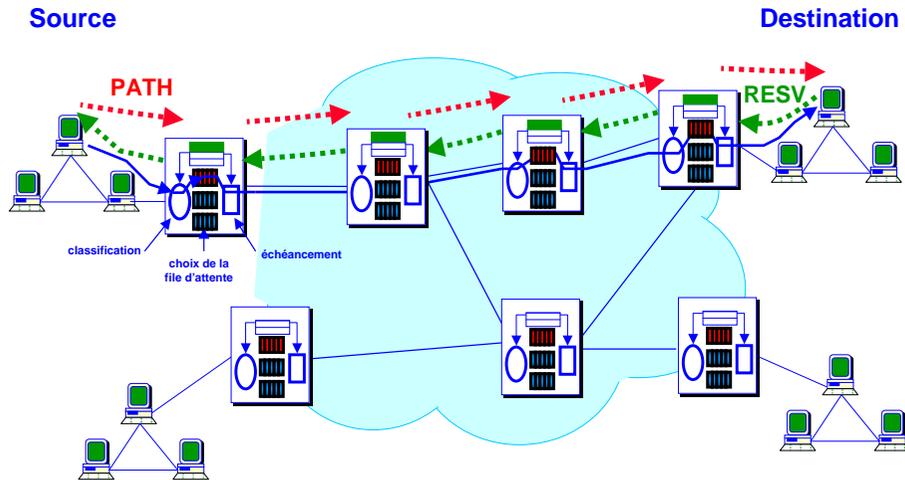
- Flux VoIP : codec G.711, débit IP 72 kbps, paquets de 360 octets

D : temps d'attente dans le routeur
C : temps d'émission du paquet de longueur M (au débit R)



Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : ∅	Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : D = 3 ms C = M	Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : D = 5 ms C = 2.M	Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : D = 8 ms C = 3.M
			T-spec + R-spec : R = 72 kbps S = 0
			T-spec + R-spec : ou R-spec : R = 100 kbps S = 19 ms

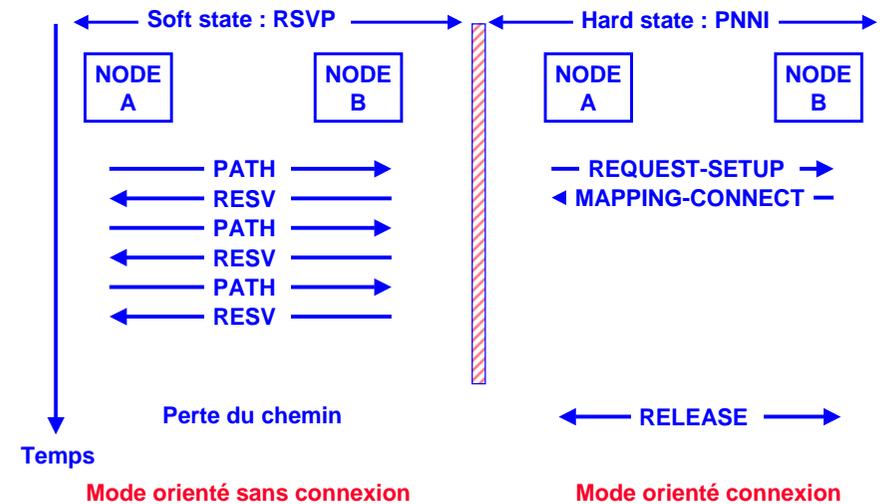
Mise en oeuvre de RSVP



Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 133 - 12 décembre 2007

Signalisation soft state/hard state



Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 134 - 12 décembre 2007

P. Ashwood-Smith, 1999

La gestion du trafic et de la QoS en IP

- **Réservation de ressources : RSVP**
- **Services intégrés : IntServ**
 - les objectifs de IntServ
 - les profils de trafic IP : objet TSPEC
 - les services intégrés : « controlled load » et « guaranteed service »
 - utilisation de RSVP pour IntServ
 - les limitations de IntServ/RSVP
- **Services différenciés : DiffServ**

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 135 - 12 décembre 2007

Objectifs de IntServ (Integrated Services)

- Permettre à l'utilisateur d'une application sur IP de requérir un service de niveau IP différent du service best effort. Deux services sont offerts :
 - service à charge contrôlée (Controlled Load Service)
 - service garanti (Guaranteed Service)
- Ne pas pénaliser indûment les utilisateurs du service IP best effort
- Ne pas modifier le routage IP
- Pouvoir fonctionner pour des flots unicast et des flots multicast
- IntServ définit des services, mais pas les moyens de la négociation
- IntServ s'appuie sur les protocoles de routage traditionnels (OSPF), et la QoS n'est pas prise en compte dans le routage
- Définition d'un modèle de comportement de chaque routeur : le PHB (Per Hop Behaviour)
 - classification
 - mécanismes de gestion des files d'attente
 - arbitrage entre les files d'attente

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 136 - 12 décembre 2007

Cours QoS

IntServ et RSVP

- IntServ nécessite de mettre en œuvre un mécanisme de signalisation explicite des flux (RSVP, par ex) pour :
 - décrire les caractéristiques des flux de données
 - réserver les ressources adéquates pour le flux

IntServ définit des services

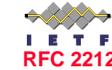
La signalisation IntServ peut être réalisée par RSVP ou par d'autres protocoles

RSVP est un protocole de signalisation

Il peut servir à la signalisation IntServ ou à la signalisation d'autres réservations IP (MPLS)

Les services IntServ (en plus du best effort)

Guaranteed Service



- **Garanties**
 - bande passante minimale
 - délai maximal de bout-en-bout
 - pas de garantie sur la gigue
- **Engagement de la source**
 - respecter le débit déclaré T_{spec} (vérifié par un token bucket)
- **Mécanismes dans les routeurs**
 - gestion des files d'attente (WFQ)
- **Ingénierie du réseau**
 - réservation de ressources (R_{spec})
 - contrôle d'admission des nouveaux flux basé sur les Worst Case Traffics

Controlled Load Service



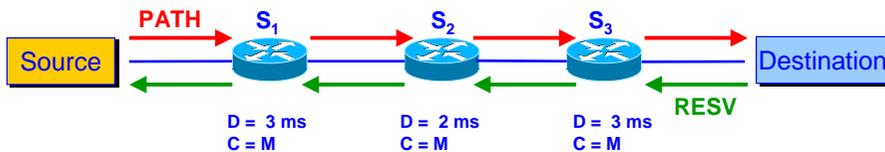
- **Garanties**
 - service correspondant au « service best effort sur réseau non chargé »
 - pas de garantie ni sur le délai, ni sur la gigue
- **Engagement de la source**
 - respecter le débit déclaré T_{spec} (vérifié par un token bucket)
- **Pas de mécanismes dans les routeurs**
- **Ingénierie du réseau**
 - contrôle d'admission réseau basé sur la mesure d'agrégats de trafic

Réservation des ressources (1)

- Flux VoIP : codec G.711, débit IP 72 kbps, paquets de 360 octets, délai de bout-en-bout D_{e-t-e} souhaité 200 ms, service garanti

D : temps d'attente dans le routeur

C : temps d'émission du paquet de longueur M (au débit R)



Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : ∅	Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : D = 3 ms C = M	Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : D = 5 ms C = 2.M	Sender T-spec : r = 72 kbps b = 720 octets p = ∞ Adspec : D = 8 ms C = 3.M
----------------------------------------------------------------------------	--------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------

T-spec + R-spec :
R = 72 kbps
S = 0

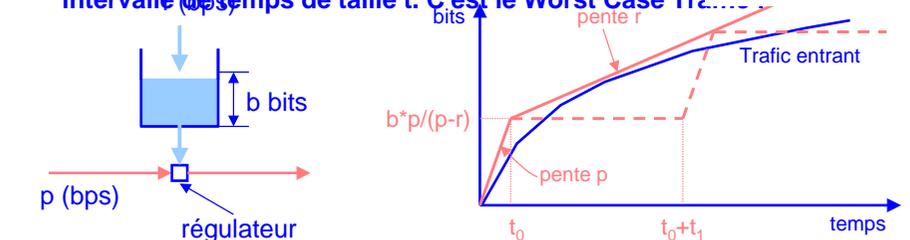
ou

T-spec + R-spec :
R = 100 kbps
S = 19 ms

Les routeurs cumulent les informations dans Adspec

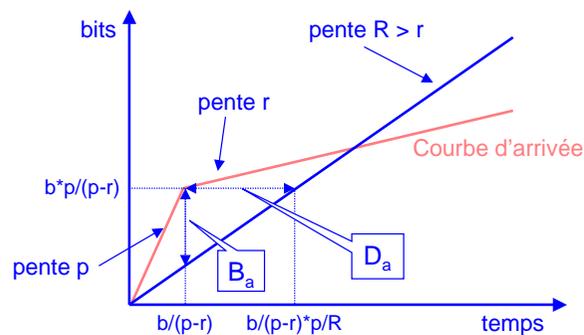
Le Token Bucket et la courbe d'arrivée

- Le Token Bucket peut être utilisé dans une fonction de police ou de shaping
- Paramètres
 - r : débit moyen négocié, c-à-d débit de régénération des jetons
 - b : taille du bucket
 - p : débit-crête ou débit maximum d'émission de la source
- Un bit est transmis, seulement s'il y a un jeton disponible
- Courbe d'arrivée : nombre maximum de bits transmis pendant un intervalle de temps de taille t. C'est le Worst Case Traffic



Réservation des ressources (2)

- Dans l'approche IntServ/RSVP, chaque routeur :
 - alloue au flux une bande passante R , au moins égale à r , définie dans Rspec du message RESV
 - réserve une taille mémoire B_a de telle façon que :
 - aucun paquet ne soit détruit
 - aucun paquet ne subisse un délai supérieur à D_a



p : débit du lien de sortie

$$D_a = b/R * (p-R)/(p-r)$$

$$B_a = b * (p-R)/(p-r)$$

si $r < R \ll p$, alors :
 $D_a \sim b/R$
 $B_a \sim b$

Réservation des ressources (3)

- Le délai D_{e-t-e} d'un paquet de taille M est alors la somme des 3 termes suivants :
 - le délai d'espacement (temps de traversée d'un TB(r, b, p)) :
 $D_1 = b/R * (p-R)/(p-r) \sim b/R$, si p est grand devant r et R
 - le temps d'émission cumulée des paquets (obtenu à partir des paramètres C de Adspec) :
 $D_2 = (M + C_{tot}) / R$
 - le délai d'attente cumulée dans les routeurs (obtenu à partir des paramètres D de Adspec) :
 $D_3 = D_{tot}$
- C'est la destination qui calcule une estimation de D_{e-t-e} , en fonction des indications fournies par tous les routeurs du chemin
- Dans notre exemple, on a :
 - $D_1 = b/R$, $D_2 = 4 * M/R$ et $D_3 = 8$ ms
 - si $R = r = 72$ kbps, alors $D_1 = 80$ ms, $D_2 = 160$ ms, d'où $D_{e-t-e} = 247$ ms
 - si $R = 100$ kbps, alors $D_1 = 58$ ms, $D_2 = 115$ ms, d'où $D_{e-t-e} = 181$ ms on aura donc une marge S de 19 ms !

Problèmes non résolus avec IntServ/RSVP

- Généralisation impossible à grande échelle :
 - l'agrégation n'est pas possible
 - les méthodes d'aiguillage IP ne supportent pas la multiplication des états dans les routeurs
 - le coût de la signalisation (soft state)
- Quelle QoS ?
 - la notion de QoS supportée par Intserv est très floue, et n'est pas nécessairement adaptée aux besoins des applications
 - il n'est pas possible de choisir des routes en fonction des besoins de l'application, ni de fixer les chemins des flots IntServ, et IntServ ne permet pas l'agrégation de flots

La gestion du trafic et de la QoS en IP

- Réservation de ressources : RSVP
- Services intégrés : IntServ
- Services différenciés : DiffServ
 - principes de DiffServ
 - architecture pour DiffServ
 - marquage des paquets : le DSCP
 - les limitations de DiffServ

Principes de DiffServ (Differentiated Service)

Objectifs

- permettre à des domaines IP d'offrir, à grande échelle, d'autres services que le service best effort
- ne pas offrir les services aux flots individuels, mais à des agrégats de trafic (simplifier IntServ)
- simplifier les traitements dans les routeurs
- simplifier la commande : configuration plutôt que signalisation RSVP

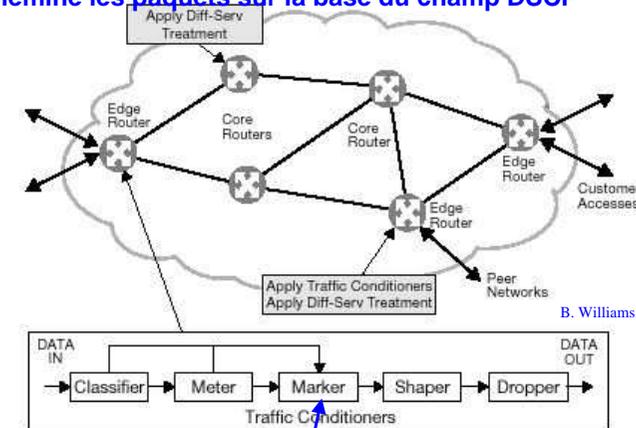
Approche

- en bordure du domaine, classer chaque flux élémentaire dans une classe de service prédéfinie
- dans le domaine, traiter chaque classe de service de façon adaptée
- 14 classes au lieu de 100 000 flux
- mais QoS relative !

Architecture DiffServ (1/2)

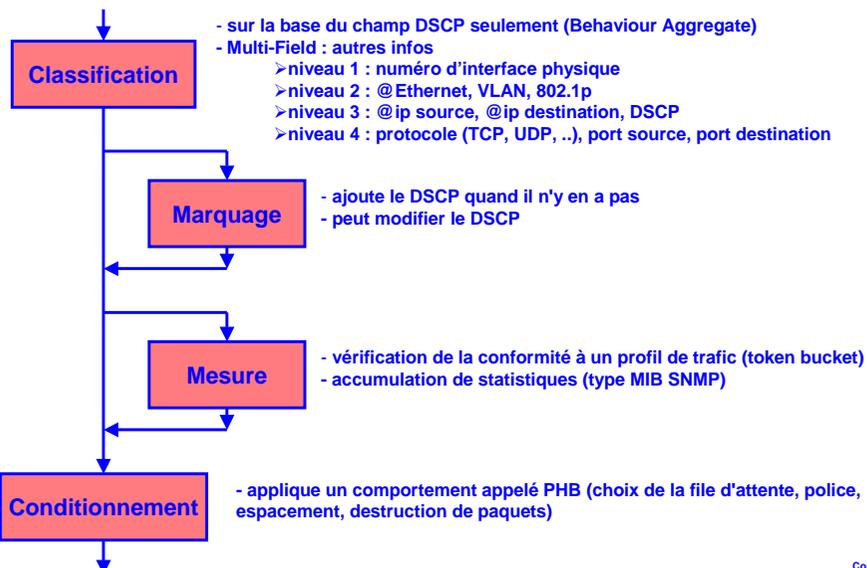
Chaque routeur d'un domaine DiffServ :

- implémente un PHB (Per Hop Behaviour) sur la base du champ DSCP
- achemine les paquets sur la base du champ DSCP



B. Williams, Ericsson, Mars 2000

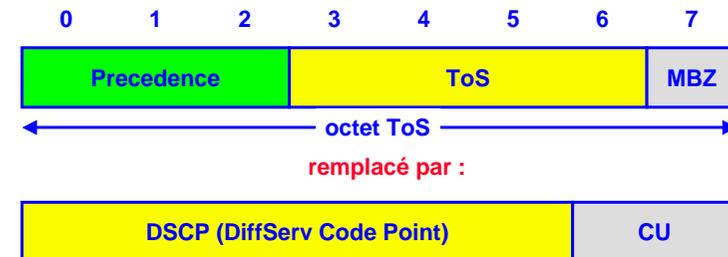
Architecture DiffServ (2/2)



Le marquage des paquets : champ DSCP

Champ DSCP sur 6 bits :
64 valeurs possibles

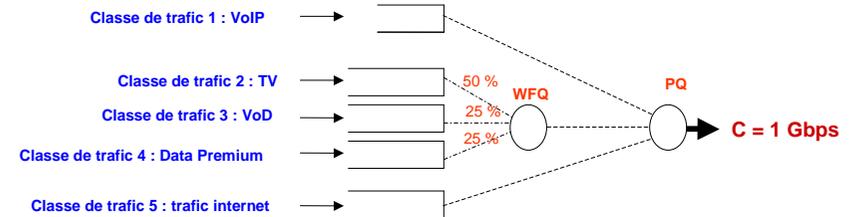
Ver.	LgE	ToS	Longueur totale	
Identification		Flag	offset	
TTL	Next prot.	Checksum en-tête		
@ IP source				
@ IP destination				
Options			Bourrage	



Per Hop Behaviour (PHB)

- Un PHB s'appuie sur deux mécanismes :
- Gestion de file d'attente
 - Tail Drop : destruction des paquets qui parviennent à une file déjà remplie
 - Partial Buffer Sharing : admission sélective des paquets dans la file
 - RED (Random Early Detection) : destruction sélective des paquets
- Ordonnancement
 - Objectif : choisir le paquet suivant à transmettre
 - Algorithme
 - PQ (Priority Queuing) : Garantie « forte » sur la disponibilité du débit et un faible délai et une faible gigue pour la file « priorité 1 »
Garanties moindres pour la file la moins prioritaire
 - PRR (Packet Round Robin) : Partage inéquitable du débit
 - DRR (Deficit Round Robin) ou WRR (Weighted Round Robin) : Chaque file i obtient un débit au moins égal à $W_i / \sum W_j$
Garantit la disponibilité d'un débit minimum

Comportement des nœuds : un exemple de PHB



- Ordonnancement hiérarchique
 - Class Based Queuing (CBQ)
 - La VoIP aura tout le débit requis
 - Risque de famine => débit de la VoIP à contrôler/limiter
 - La TV + VoD + Data Premium se partageront le débit restant
 - La TV prendra la moitié du débit laissé par la VoIP
 - La VoD prendra le quart du débit laissé par la VoIP
 - La Data Premium prendra le quart du débit laissé par la VoIP
 - Le trafic « internet » disposera du débit restant ...

Les valeurs de DSCP déjà attribuées

- 000 000 : PHB par défaut
- xxx 000 : PHB « class selector »
cohérence avec le champ precedence de l'octet ToS
- 101 110 : PHB EF (Expedited Forwarding)
équivalent à un service « liaison louée »

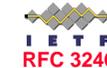
Valeurs de CS :	
111	– network control
110	– internetwork control
101	– critic/ECP
100	– flash override
011	– flash
010	– immediate
001	– priority
000	– routine

- PHB AF (Assured Forwarding)
services « olympiques »
4 classes
3 niveaux de priorité

Drop precedence	Classe 1	Classe 2	Classe 3	Classe 4
low	001 010 (AF11)	010 010 (AF21)	011 010 (AF31)	100 010 (AF41)
medium	001 100 (AF12)	010 100 (AF22)	011 100 (AF32)	100 100 (AF42)
high	001 110 (AF13)	010 110 (AF23)	011 110 (AF33)	100 110 (AF43)

Les services DiffServ existants

Expedited Forwarding EF



- Garanties
 - émule une « liaison louée IP » entre deux routeurs
 - analogue au CBR ATM
 - taux de perte, délai et gigue faibles

Engagement de la source

- respecter le débit déclaré (vérifié par un token bucket)
- le trafic en excès est espacé (retardé)

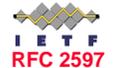
Mécanismes dans le routeur

- files strictement prioritaires

Ingénierie du réseau

- CAC sur l'agrégat effectuée par un Bandwidth Broker

Assured Forwarding AF



- « Garanties »
 - émule un réseau peu chargé
 - 4 classes de service et 3 niveaux de priorité à la perte (gold, silver et bronze)
 - analogue au nrt-VBR ATM
 - CLR plus faible que celui du best effort

Engagement de la source

- respecter le débit déclaré (vérifié par un token bucket)
- le trafic en excès est converti en trafic best effort

Mécanismes dans le routeur

- files plus prioritaires que le best effort

Problèmes non résolus par DiffServ

- **Contrôle de la QoS offerte par un domaine**
 - les PHB ont une signification uniquement **locale à un routeur**
 - on ne sait pas comment allouer les ressources dans un domaine (sauf à utiliser un serveur externe : Bandwidth Broker)
- **Il n'est pas possible d'offrir une QoS de bout-en-bout aux applications car les flots individuels ne sont pas reconnus**
- **En fait, DiffServ permet seulement de réaliser plusieurs services « best effort » avec des niveaux de QoS différents**
- **DiffServ doit s'appuyer sur d'autres outils permettant l'ingénierie de trafic (MPLS)**

En résumé

- **Les architectures opérationnelles actuelles utilisent majoritairement un réseau de transport ATM (IP/ATM) , ou surdimensionnent les réseaux (IP/WDM)**
- **IP ne propose pas aujourd'hui un support, même partiel, à la QoS de bout-en-bout**
- **De multiples architectures IP tentent de résoudre les problèmes ouverts : RSVP, IntServ, DiffServ fournissent des solutions partielles**
- **Des solutions propriétaires mixent toutes ces architectures pour offrir une QoS en environnement intradomaine (mais pas de bout en bout !)**

DiffServ	IntServ	ATM
EF	service garanti	CBR
AF	service à charge contrôlée	VBR

MPLS (Multi Protocol Label Switching) et évolutions d'Ethernet

MPLS et Ethernet : plan

- **Les origines de MPLS**
- **Les grands principes de MPLS**
- **Les protocoles mis en œuvre dans MPLS**
- **Les applications de MPLS**
- **De MPLS à GMPLS**
- **Rappels sur Ethernet et besoins d'évolution**
- **Les ajouts récents dans Ethernet**

Méthode classique d'acheminement dans un réseau IP

- Dans un routeur IP, la décision d'acheminement doit être prise pour chaque paquet IP. Elle peut dépendre :
 - du préfixe d'adresse destination
 - de la valeur de la FEC
 - du champ ToS
- Dans chaque routeur, pour chaque paquet IP, la FEC doit être évaluée à partir son en-tête
- Dans un réseau IP « stable », les paquets appartenant à la même FEC empruntent la même route
 - cette route est obtenue à partir du protocole de routage SPF
- Idée : créer un circuit virtuel pour les paquets IP appartenant à la même FEC, de façon à ne pas ré-évaluer la FEC dans chaque routeur (mode connecté)

MPLS et Ethernet : plan

- Les origines de MPLS
- Les grands principes de MPLS
- Les protocoles mis en œuvre
- Les applications
- De MPLS à GMPLS
- Rappels sur Ethernet et besoins d'évolution
- Les ajouts récents dans Ethernet

Fonctionnement de MPLS

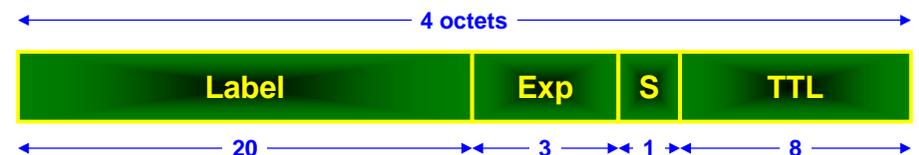


- MPLS est une technologie en mode **connecté** qui s'appuie sur une technique de niveau 2
- MPLS peut utiliser soit le label de niveau 2 (s'il existe), soit un label propre (shim label)
 - ATM : label = VCI/VPI
 - Frame Relay : label = DLCI
 - PPP/LAN/PoS : "shim header" inséré entre les en-têtes L2 et L3
- Il est possible d'empiler les labels !
- Contrôle et acheminement sont des fonctions séparées (comme en ATM). MPLS ne remplace pas le routage de niveau 3 : il y a toujours un routage à l'entrée et à la sortie du réseau MPLS
- MPLS apparaît ainsi comme une sous-couche (niveau 2,5)

"Shim Headers" MPLS (1 à n)



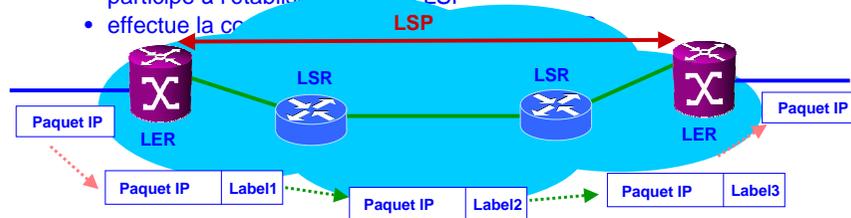
Un label spécifique : le « shim header »



- Label : valeur de label (0 à 16 réservés)
- Exp : Experimental (anciennement "Class of Service")
- S : fin de pile, (1 = dernière entrée dans la pile de labels) puisqu'on peut empiler les labels
- TTL : Time to Live

Architecture d'un réseau MPLS

- La "connexion" MPLS s'appelle un LSP (Label Switch Path)
- Chaque noeud d'un réseau MPLS est défini par sa position et ses fonctions
 - en périphérie : LER (Label Edge Router)
 - en entrée : détermine à quelle classe (FEC : Forwarding Equivalent Class) le paquet appartient, et ajoute le label
 - en sortie : enlève le label et achemine le paquet IP
 - en coeur : LSR (Label Switch Router)
 - participe à l'établissement des LSP
 - effectue la commutation des paquets



Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 161 - 12 décembre 2007

MPLS et Ethernet : plan

- Les origines de MPLS
- Les grands principes de MPLS
- Les protocoles mis en œuvre
- Les applications
- De MPLS à GMPLS
- Rappels sur Ethernet et besoins d'évolution
- Les ajouts récents dans Ethernet

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 162 - 12 décembre 2007

Comment sont distribués les labels ?

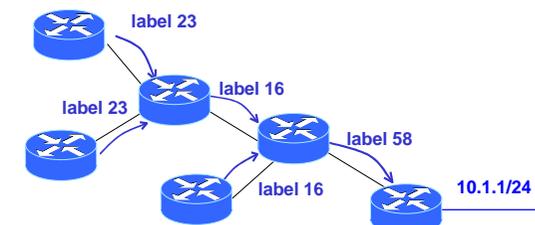
- Les routes sont fournies par les protocoles de routage IP classique
- Il y a plusieurs façons de construire un protocole de distribution des labels :
 - faire porter l'information par un protocole de routage existant (piggy-backing)
 - assure la cohérence entre routage et distribution de label
 - exemple : BGP
 - étendre un protocole existant
 - RSVP (ajout d'objets)
 - créer un nouveau protocole
 - LDP (Label Distribution Protocol)
- Les labels sont attribués dans l'ordre inverse (de la destination vers la source)
 - le dernier noeud du chemin attribue le label en premier
 - pour un même chemin, chaque interface de chaque noeud reçoit un label différent (identifiant local comme en ATM)

Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 163 - 12 décembre 2007

Utilisation du protocole LDP

- LDP établit des LSPs mono-directionnels, d'une façon distribuée, en se basant sur des informations locales aux LSRs
- Les LSPs établis par LDP suivent la route IP classique
 - pour une FEC donnée, chaque LSR utilise le label envoyé par le LSR aval, qui est le next-hop (dans la table de routage) pour cette FEC
 - le LSP est construit bond-par-bond
- Le label peut être unique par interface ou par LSR
- Un LDP-LSP est donc un arbre multipoint-à-point, pour chaque FEC



Copyright © 2007 - Jean-Raymond LOUVION

Cours QoS
Problématique de la QoS et de la gestion du trafic - 164 - 12 décembre 2007

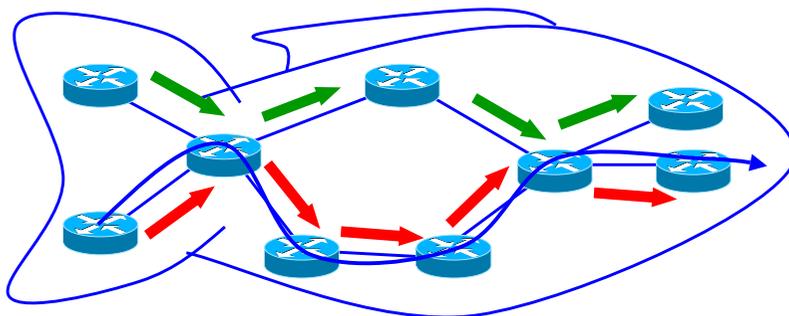
Méthode de routage explicite (ER)

- ER (Explicit Route) est une liste d'identifiants de LSR le long du LSP
 - strict : tous les LSRs sont spécifiés
 - lâche : quelques LSRs seulement sont spécifiés
- C'est le LER d'entrée qui est responsable de la détermination de la route explicite ER et de l'initialisation de la signalisation
- Les routes sont établies en utilisant
 - soit RSVP étendu pour gérer la distribution des labels (nouvel objet ERO - Explicit Route Object qui permet de définir une route explicite)
 - soit LDP étendu (CR-LDP) pour prendre en compte les routes explicites et réserver des ressources

MPLS et Ethernet : plan

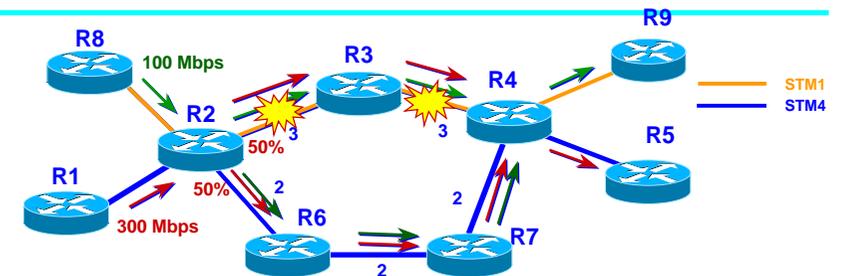
- Les origines de MPLS
- Les grands principes de MPLS
- Les protocoles mis en œuvre
- Les applications
- De MPLS à GMPLS
- Rappels sur Ethernet et besoins d'évolution
- Les ajouts récents dans Ethernet

Retour sur la problématique du Poisson



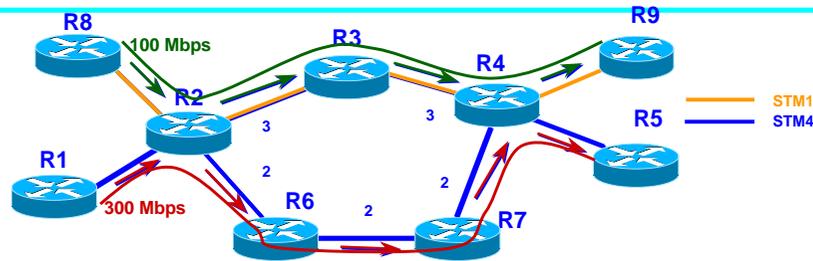
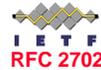
- Le chemin vert est calculé par les protocoles de routage comme étant le moins "cher"
- Tous les paquets emprunteront ce chemin, même s'il est congestionné, alors que le chemin rouge (calculé comme plus cher) peut être vide !

Equilibrage de charges en IP



- Solution au pb du Poisson :
 - on peut équilibrer les charges en jouant avec les métriques et en utilisant le protocole ECMP (Equal Cost Multi Path) de partage de charge
 - équilibrage de charge en R2 : même débit sur les 2 chemins
- Mais le routage IP ne sait pas prendre en compte des liens de capacité différentes, sans surcharger les liens de plus faible débit
 - pb : capacité de R2-R6-R7-R4 > capacité de R2-R3-R4
- Ingénierie statique

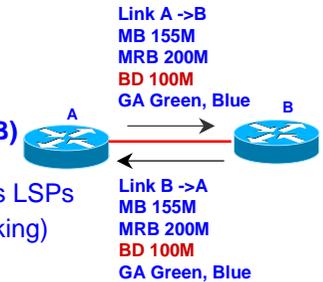
Solution MPLS-TE (Traffic Engineering)



- **Etablir un "tunnel MPLS-TE" sur un chemin calculé**
 - non pas par les protocoles de routage IP, mais en tenant compte des contraintes de trafic des sources et des ressources disponibles dans le réseau
 - c'est un LSP unidirectionnel point-à-point auquel est associé un ensemble de paramètres TE
- **Cela requiert un protocole de routage par contraintes, capable de prendre en compte les ressources disponibles**
- **Exemple : on construit deux tunnels**
 - R1-R2-R6-R7-R4-R5 d'un débit de 300 Mbps

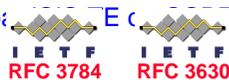
Paramètres TE

- **Bande passante maximum (MB)**
 - bande passante maximum qui peut être réservée par un LSP
- **Bande passante maximum réservable (MRB)**
 - bande passante maximum totale qui peut être réservée par l'ensemble des LSPs
 - elle peut être supérieure à MB (overbooking)
- **Bande passante disponible (BD)**
 - bande passante résiduelle réservable
 - elle est modifiée dynamiquement par les réservations
- **Groupes administratifs (GA)**
 - groupes administratifs (couleurs) (32 maximum)
 - utilisé pour inclure ou exclure certains liens



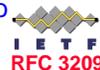
Routage par contraintes MPLS-TE

- **Fonction de découverte de la topologie TE**
 - permet à tous les routeurs d'avoir une vision actualisée de la topologie TE
 - réalisée par : IETF RFC 3784 et IETF RFC 3630 -TE (extension avec des champs TE)



- **Fonction de placement (ou de calcul) du tunnel**
 - algorithme de routage par contrainte
 - peut être réalisé par le routeur de tête, ou de façon centralisée

- **Fonction d'établissement (ou de signalisation) du tunnel**
 - réservation des ressources dans chaque nœud et distribution labels
 - contrôle d'admission
 - réalisé par RSVP-TE (extension pour le routage explicite et la distribution des labels)



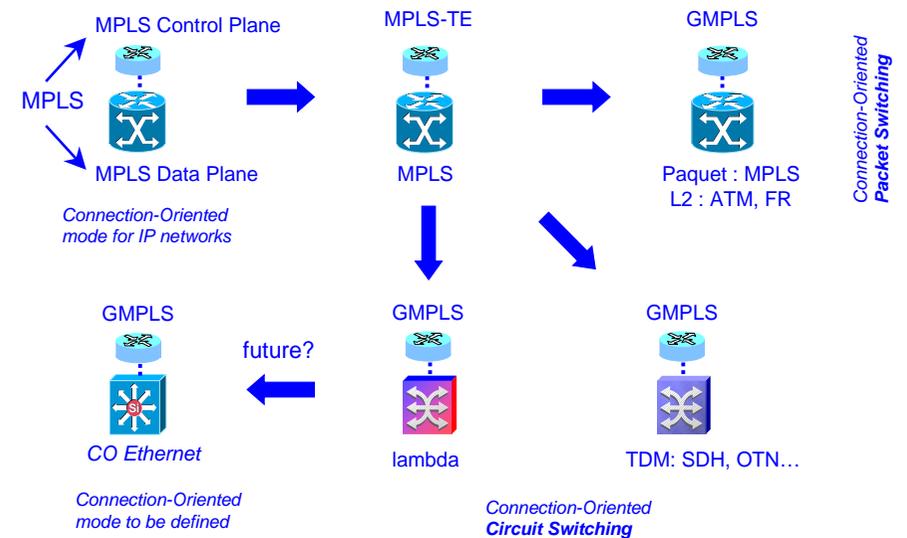
MPLS : en résumé

- **MPLS est une technique permettant d'intégrer, dans les mêmes machines, IP et les techniques de niveau 2 (ATM, Frame Relay)**
- **MPLS établit des circuits virtuels. Il transforme les routeurs IP en commutateurs de labels**
- **MPLS peut prendre en compte des contraintes de trafic et de QoS dans l'établissement de ses circuits !**
- **Applications nombreuses :**
 - routage fonction de la QoS
 - mécanismes de protection (Fast ReRoute)

MPLS et Ethernet : plan

- Les origines de MPLS
- Les grands principes de MPLS
- Les protocoles mis en œuvre
- Les applications
- **De MPLS à GMPLS**
- Rappels sur Ethernet et besoins d'évolution
- Les ajouts récents dans Ethernet

De MPLS à GMPLS : un plan de contrôle généralisé

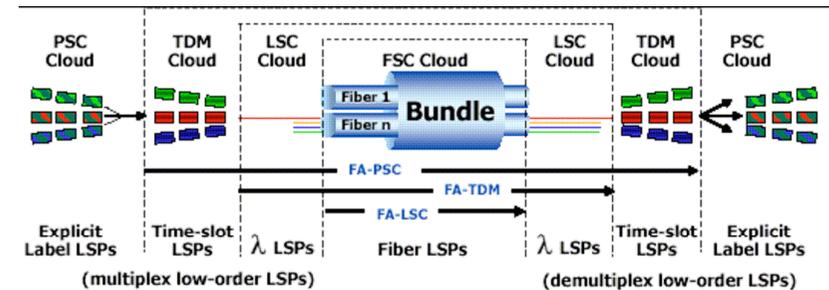


Motivations de GMPLS

- **Constat :**
 - la technique MPLS comporte à la fois
 - un plan de données commuté
 - un plan de contrôle adapté à la commutation (routage avec contraintes, signalisation pour l'établissement des circuits)
 - les réseaux de paquets et les réseaux de transmission partagent les mêmes principes de base
- **Idée : réutiliser les mécanismes de contrôle de MPLS-TE (Traffic Engineering) à la fois :**
 - pour les réseaux de paquet (PSC = Packet Switch Capable) et les réseaux de circuit (non-PSC)
 - pour tous les types de nœuds : routeurs, commutateurs, routeurs optiques, systèmes WDM
- **IETF : Common Control And Measurement Plane (CCAMP) group**

Hiérarchie des labels

- **GMPLS étend le concept de label**
 - Labels explicites : en-tête MPLS, VPI/VCI ATM, DLCI Frame Relay,...
 - Labels implicites : timeslot (#VC), longueur d'onde, port,...



Concepts de GMPLS

● Concept 2 : agrégation de liens

- Des nœuds adjacents peuvent être connectés par un large nombre de liens physiques (ex: des longueurs d'onde)
- De multiples liens parallèles peuvent être signalés comme un lien unique par le protocole de routage
- Moins d'informations à transmettre via le protocole de routage, qui voit l'agrégation mais pas les composants du lien
- Gestion des liens physiques : LMP (Link Management Protocol)

● Concept 3 : liens non-numérotés

- Comment attribuer une adresse IP à chaque lien (canal TDM, lambda ou fibre) d'un réseau optique ?
- Le concept de lien numéroté permet de configurer des chemins sans utiliser une adresse IP sur chaque interface physique ou logique
- Identification des liens :
 - un ID unique pour chaque interface (32 bits)
 - un ID additionnel pour chaque lien (32 bits)

Le plan de contrôle GMPLS

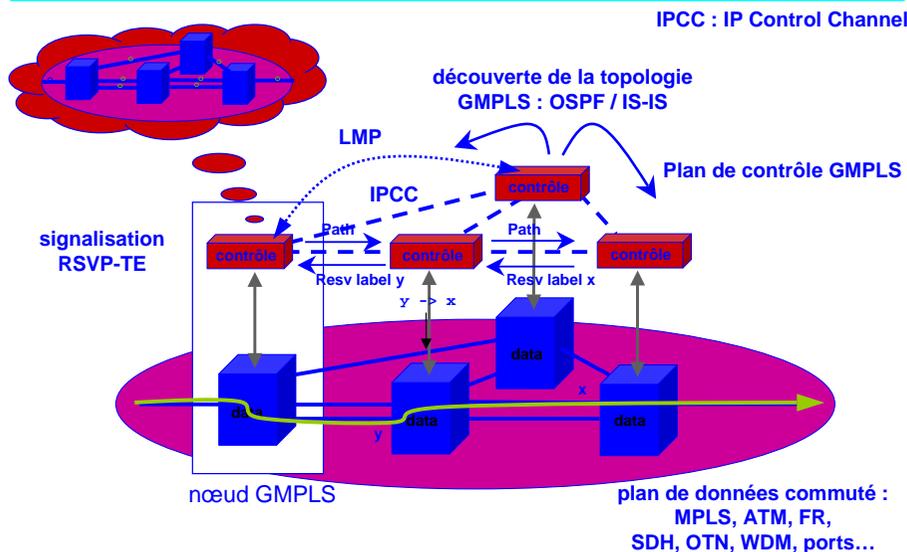
● routage

- similaire aux réseaux MPLS-TE
- ajout des disponibilités des ressources :
 - Type de chemin que l'on peut établir sur un lien donné
 - Bande passante utilisée
 - Le nombre maximal de LSP que l'on peut établir sur un lien
 - Type de protection de lien
- topologie TE : chaque nœud NE connaît la topologie de son domaine
- chaque NE calcule les routes pour les requêtes qu'il reçoit (cf. PNNI)

● signalisation

- établissement/relâchement des connexions
- modification des paramètres des connexions

Le fonctionnement de GMPLS

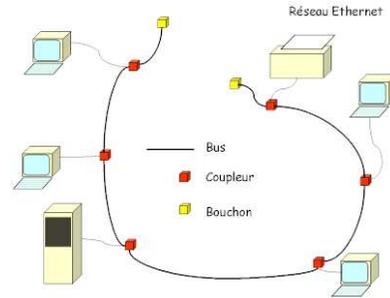


MPLS et Ethernet : plan

- Les origines de MPLS
- Les grands principes de MPLS
- Les protocoles mis en œuvre dans MPLS
- Les applications de MPLS
- De MPLS à GMPLS
- **Rappels sur Ethernet et besoins d'évolution**
- Les ajouts récents

Rappels sur Ethernet (1)

- Ethernet existe depuis longtemps (1975)
- C'est une technique rustique conçue pour les réseaux locaux
 - plusieurs sources se raccordent (par des prises) au même support physique (bus)
- Elle met en œuvre un protocole de contention et de résolution de congestion CSMA/CD
- Le débit physique peut être :
 - 10 Mbps : Ethernet classique (câble coaxial, paires torsadées)
 - 100 Mbps : Fast Ethernet (paires torsadées, fibres optiques)
 - 1 Gbps : Gigabit Ethernet (fibres optiques, GBIC)
 - 10 Gbps (fibres optiques monomode ou multimode)



Rappels sur Ethernet (2)

- Caractéristiques d'Ethernet
 - le débit utile est plus faible que le débit physique (overhead et collisions)
 - le mode de transmission est de type bidirectionnel alterné : les signaux transitent dans les deux sens, mais pas simultanément
- Avantages d'Ethernet
 - Ethernet est « Plug&Play »
 - facile à dépanner et à gérer
 - les interfaces sont bon marché

Evolution d'Ethernet

- Les opérateurs mettent Ethernet dans le réseau d'accès. Ethernet doit donc évoluer :
 - il faut assurer la connectivité de plus de 4000 groupes de clients
 - les tables d'adresses MAC sont trop grandes
 - le protocole STP converge trop lentement
 - il n'y a pas d'OAM (nécessaire pour les grands réseaux)
 - il faut offrir une bonne QoS
- Beaucoup d'activité en normalisation :
 - IEEE : extension des protocoles natifs
 - IETF : interfonctionnement avec MPLS
 - ITU-T : spécifications du transport
 - MEF (Metro Ethernet Forum) : spécifications des services
- Arrivée des services Triple Play (Internet, video et voix)

La trame Ethernet



802.3

- La couche 2 est découpée en :
 - une sous-couche LLC (Logical Link Control) : IEEE 802.2
 - contrôle de flux
 - récupération d'erreurs (par retransmission)
 - une sous-couche MAC (Medium Access Control) : IEEE 802.3
 - adressage / méthode d'accès
 - détection d'erreurs (de type CRC)



Le champ < Longueur / Type >

- Champ de 2 octets
- Si < Longueur / Type > est compris entre 46 et 1500, il représente la longueur utile de la trame (en octets) : format 802.3
- Si < Longueur / Type > est supérieur à 1500 (0x05DC), il représente la nature du protocole client : format Ethernet II
- En particulier, si < Longueur / Type > est égal à 0x8100, il annonce un QTag : format 802.1Q

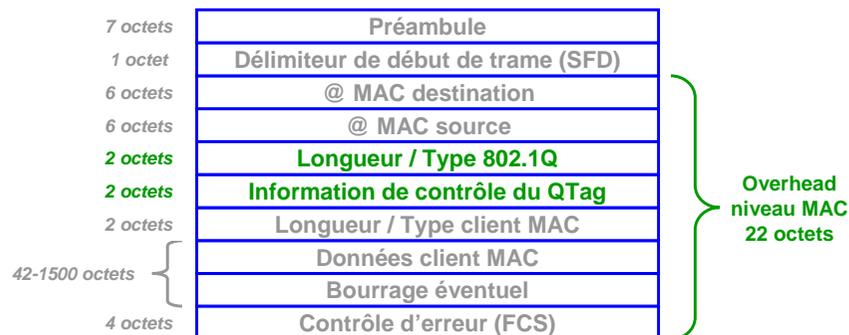
MPLS et Ethernet : plan

- Les origines de MPLS
- Les grands principes de MPLS
- Les protocoles mis en œuvre dans MPLS
- Les applications de MPLS
- De MPLS à GMPLS
- Rappels sur Ethernet et besoins d'évolution
- **Les ajouts récents**
 - la notion de VLAN
 - Q-in-Q
 - MAC-in-MAC

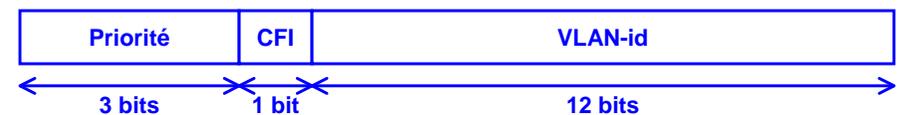
La norme 802.1Q



- La trame Ethernet classique ne permet pas de transporter des indicateurs de qualité
- La norme 802.1Q permet d'ajouter un préfixe (QTag)



Le champ QTag



- **Priorité (« User Priority »)**
 - autorise 8 niveaux de priorité (cf. ToS de IP, Exp de MPLS)
 - permet de définir des classes de trafic
- **CFI (« Canonical Format Indicator »)**
 - indique la présence d'un champ d'information de routage
- **VLAN-id (Virtual LAN identifier)**
 - identifie l'appartenance à un LAN virtuel
 - 4096 VLANs

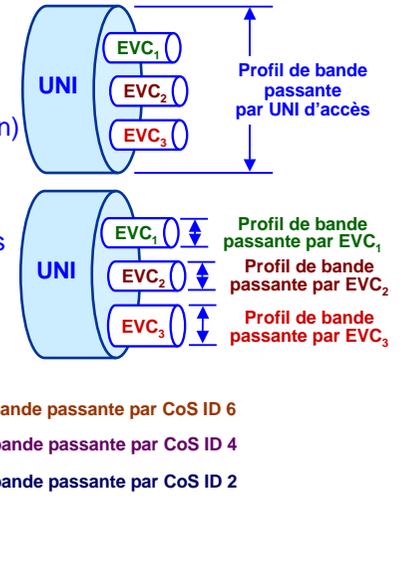
Classification des flux selon les priorités

Priorité	CFI	VLAN-id
----------	-----	---------

Priorité	Type de trafic	Signification
0	Best Effort	Trafic LAN actuel
1	Background	Gros transfert non urgent
2		Réservé
3	Excellent Effort	Service best effort pour clients importants
4	Controlled Load	Applications sujettes à la CAC
5	Vidéo	Délai < 100 ms
6	Audio	Délai < 10 ms
7	Network Control	Administration

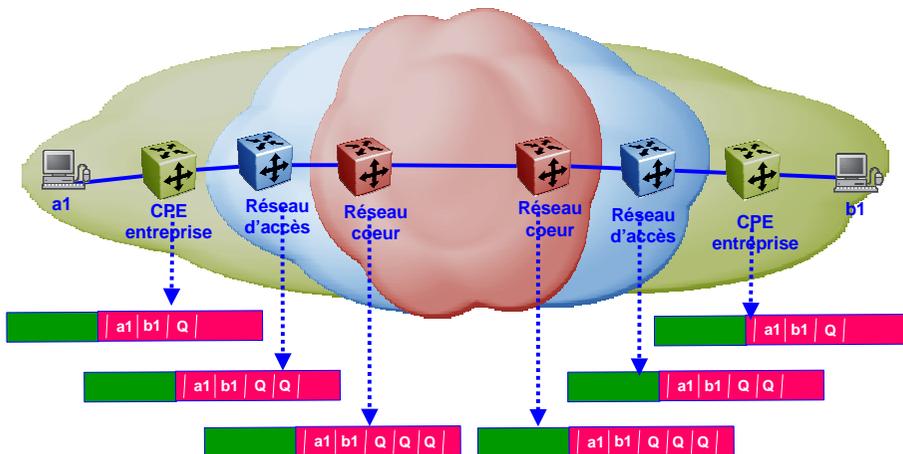
Notion de bande passante Ethernet

- Le MEF a défini 3 types de profils de bande passante :
 - par UNI d'accès
 - par EVC (Ethernet Virtual Connexion)
 - par CoS ID
- Il y a 4 paramètres principaux :
 - CIR/CBS : débit compatible avec les objectifs de QoS
 - EIR/EBS : débit en excès autorisé



Q-in-Q : principes

- Dans le réseau on a besoin de plus de 4096 VLAN id
- Il faut isoler le domaine opérateur du domaine client

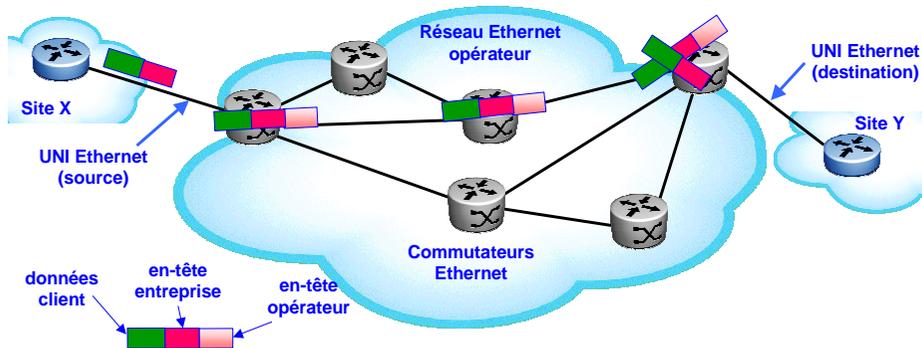


Q-in-Q : format de la trame

@ MAC destination client
@ MAC source client
Longueur / Type 802.1Q opérateur
QTag opérateur
Longueur / Type 802.1Q client
QTag client
Longueur / Type client
Données client
Bourrage éventuel
Contrôle d'erreur (FCS)

MAC-in-MAC : principes

- Les équipements de cœur ne doivent apprendre que les adresses MAC des bridges de périphérie !



MAC-in-MAC : format de la trame

@ MAC destination opérateur
@ MAC source opérateur
Longueur / Type 802.1Q opérateur
QTag opérateur
Longueur / Type MAC-in-MAC opérateur
Service Tag
@ MAC destination client
@ MAC source client
Longueur / Type 802.1Q client
QTag client
Longueur / Type client
Données client
Bourrage éventuel
Contrôle d'erreur (FCS)
Contrôle d'erreur (FCS) opérateur

Ethernet : en résumé

- La technique Ethernet, conçue pour les réseaux locaux, est maintenant introduite dans les réseaux d'accès, à cause de son coût et de sa souplesse
- La structure de la trame Ethernet est adaptée à ces nouveaux objectifs par l'introduction d'un champ supplémentaire : le **QTag**
- Il est ainsi possible de créer des LANs virtuels, qui peuvent prendre en compte les contraintes de QoS
- Pour rendre indépendantes, dans les réseaux, la partie accès de la partie cœur, les organismes de normalisation ont défini des approches nouvelles consistant à emboîter les champs de la trame Ethernet : **Q-in-Q** et **MAC-in-MAC**