



QoS et Multimédia SIR / RTS

Les principes fondamentaux de la qualité
de service

Isabelle Guérin Lassous

Isabelle.Guerin-Lassous@ens-lyon.fr

<http://perso.ens-lyon.fr/isabelle.guerin-lassous>

Introduction

- Systèmes communicants multimédia dans l'Internet tel qu'il est aujourd'hui
 - Pas de garanties sur les paramètres importants
 - Mécanismes locaux
- Comment mettre en place de la qualité de service dans le réseau
 - Pour améliorer ces paramètres
 - Et éventuellement offrir des garanties

Introduction

- Réseaux téléphoniques
 - Commutation par circuits
 - Approche avec connexion
 - Pour le transport de la voix seulement
 - Coûteux et peu flexible
- Réseaux de données
 - Commutation par paquets
 - Approche sans connexion
 - Pour le transport des données initialement
 - Optimiser l'utilisation des ressources du réseau

Introduction

- Convergence des services
 - Différents types de trafic
 - Profils différents
 - Besoins différents
 - Utilisation d'une infrastructure réseau physique commune
 - Réduction des coûts
 - Comment faire cohabiter ces applications ?
- Surdimensionnement du réseau
 - Problème de coût
 - Attention à l'évolution du trafic (en termes de profil et de volume)
- QoS

Qu'est-ce que la QoS ?

- Plein de définitions
- ETSI – QoS
 - « performance d'un service qui détermine le degré de satisfaction d'un utilisateur de ce service »
- IETF – QoS
 - « la capacité à séparer le trafic ou à différencier différents types de trafic afin de traiter certains flux différemment d'autres flux »

Qu'est-ce que la QoS ?

- Ne crée pas de nouvelles ressources réseaux
 - Ressources ?
- Réalise un partage des ressources
 - Non nécessairement équitable

Qu'est-ce que la QoS ?

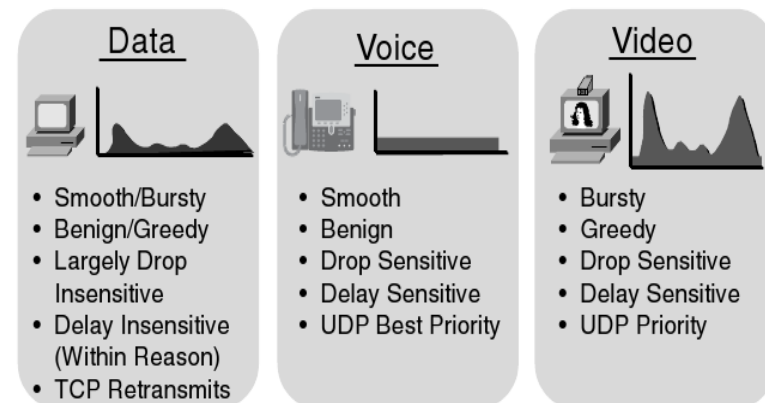
- QoS intrinsèque
 - Directement fournie par le réseau
 - Décrite en termes de paramètres objectifs
 - Taux de perte, délai, gigue, etc.
- QoS perçue
 - Qualité perçue par les utilisateurs
 - Peut être par ex. mesurée par le Mean Opinion Score (MOS)
 - Opinion moyenne entre les différents utilisateurs
 - Beaucoup plus subjectif
 - Paramètres externes
 - Prix, marketing, etc.

Flux et classe

- Flux
 - Ensemble de paquets répondant à un même service/application
- Classe
 - Groupe de flux avec un objectif commun
 - Appartient au même type d'application
 - A les mêmes contraintes
 - A les mêmes profils

Hétérogénéité

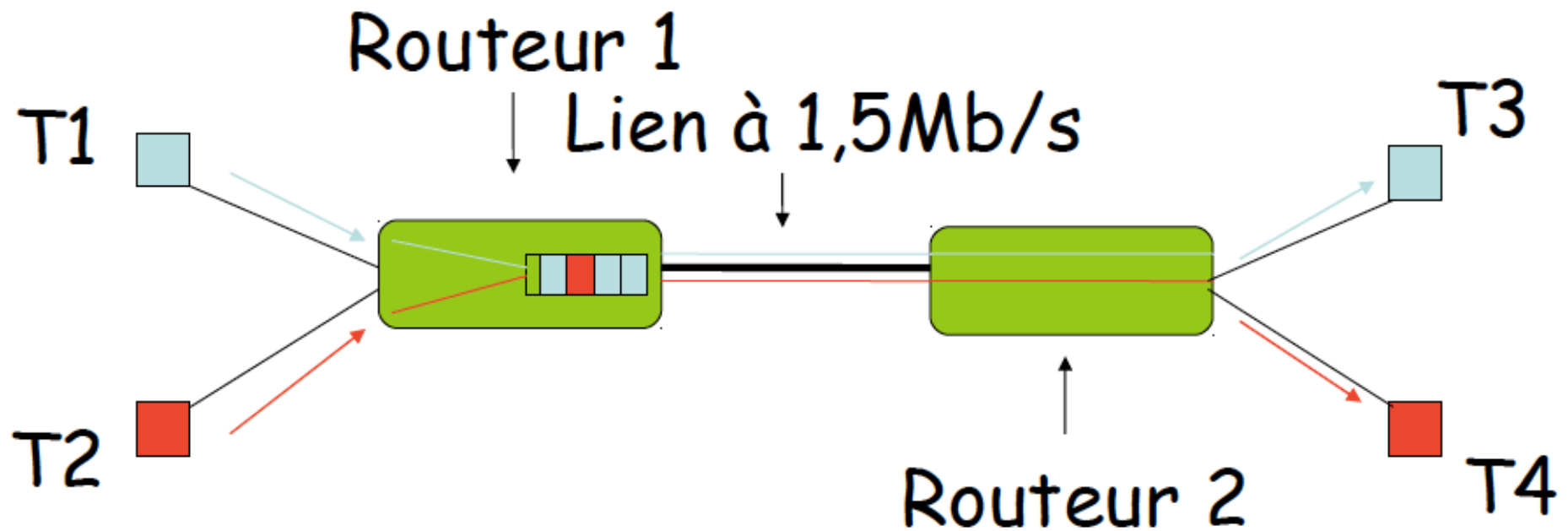
- Les réseaux sont hétérogènes
- Portions de réseaux peuvent être
 - Gérées par différents opérateurs/ fournisseurs de service
 - Utiliser différents types de transmission
 - Câble, optique, radio, satellite
 - Implémenter différents protocoles
 - Ipv4, v6, MPLS, ATM, etc.
- Différentes applications
 - Différents profils de trafic
 - Différents besoins
- Différents utilisateurs
 - Demandent des services différents
 - Sont prêts à mettre des prix différents
- La QoS doit s'appliquer de bout-en-bout



Book « End-to-End QoS Network Design » - 2005

Les principes fondamentaux de la QoS

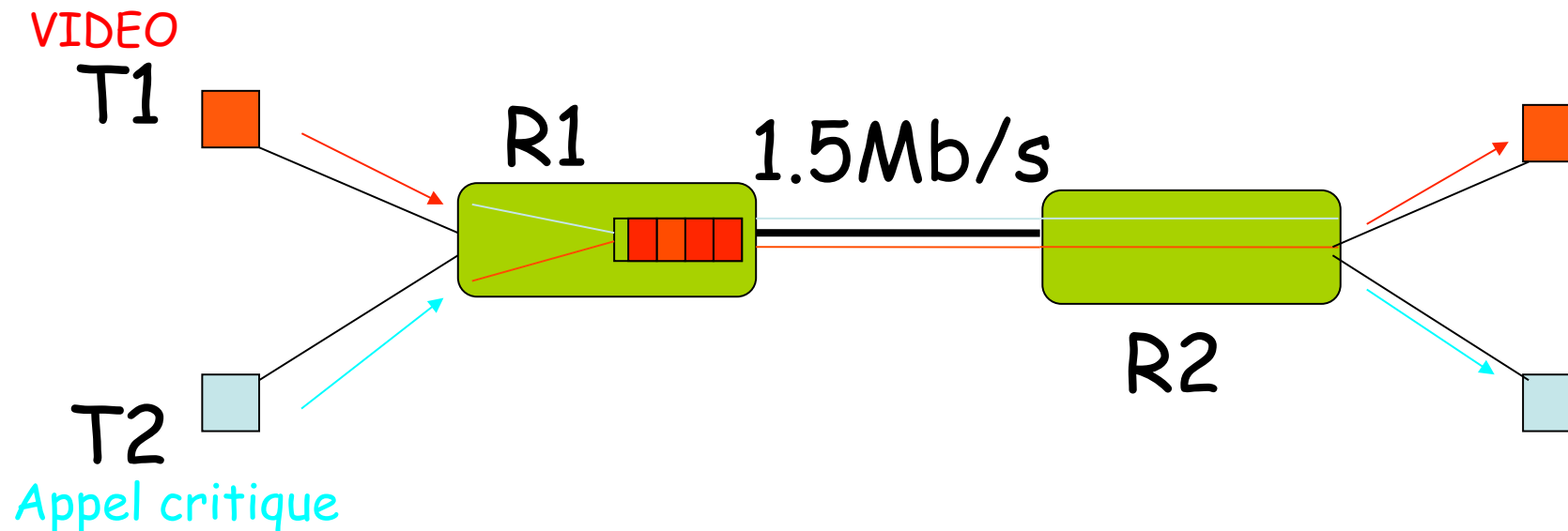
Un modèle d'étude simple



2 applications (bleue et rouge)

1^{er} cas

- Application vidéo à 2 Mb/s
- Appel critique (e.g. pompiers) à 64 kb/s
- **Que peut-il se passer ?**

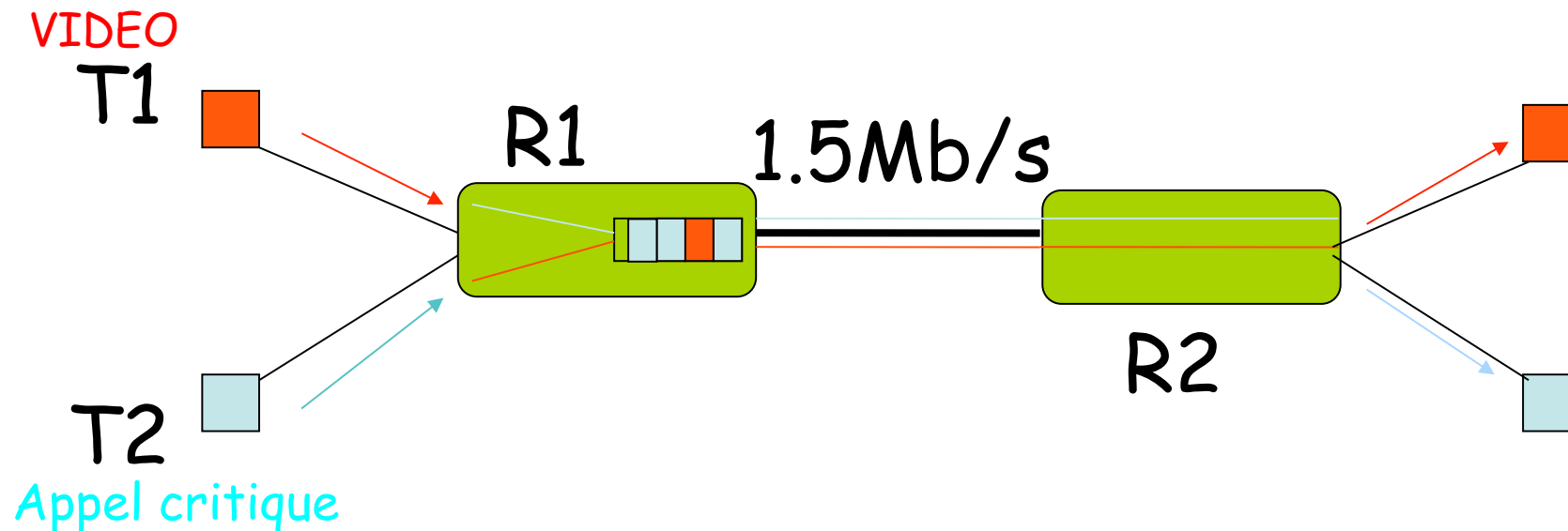


1^{er} principe

- Pour donner la **priorité** à certains paquets
 - Il faut savoir que ce sont des paquets prioritaires
- Approches
 - **Marquage** des paquets

2^e cas

- Priorité à l'appel qui a besoin de 64 kb/s
 - Envoi des paquets à un débit supérieur, e.g. à 1 Mb/s (pour faire transiter autre chose)
 - **Est-ce juste ?**

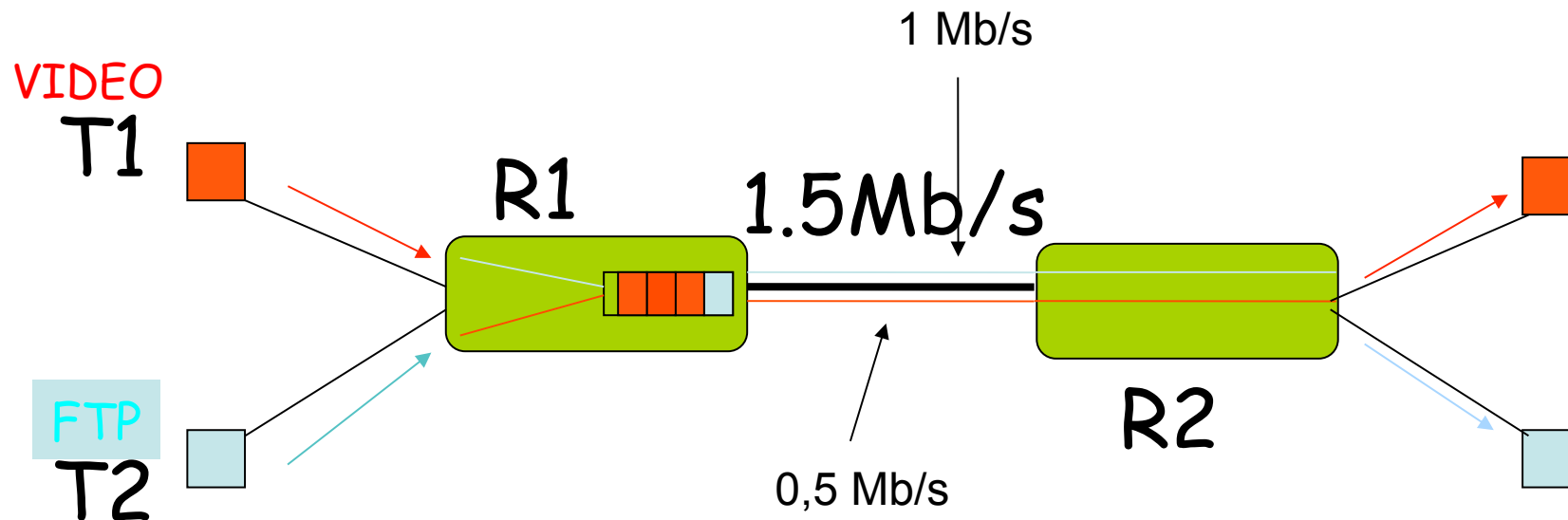


2^e principe

- Il faut pouvoir **isoler (protéger)** certains flux des autres
- Approches
 - Contrôle de débit à l'entrée du réseau
 - **Lissage** du trafic
 - Allocation d'une proportion de bande passante à chaque application
 - **Ordonnancement**

3^e cas

- Isolation des flux
 - 1Mb/s est donné au flux vidéo
 - Le flux vidéo est mis en pause
- Est-ce efficace ?

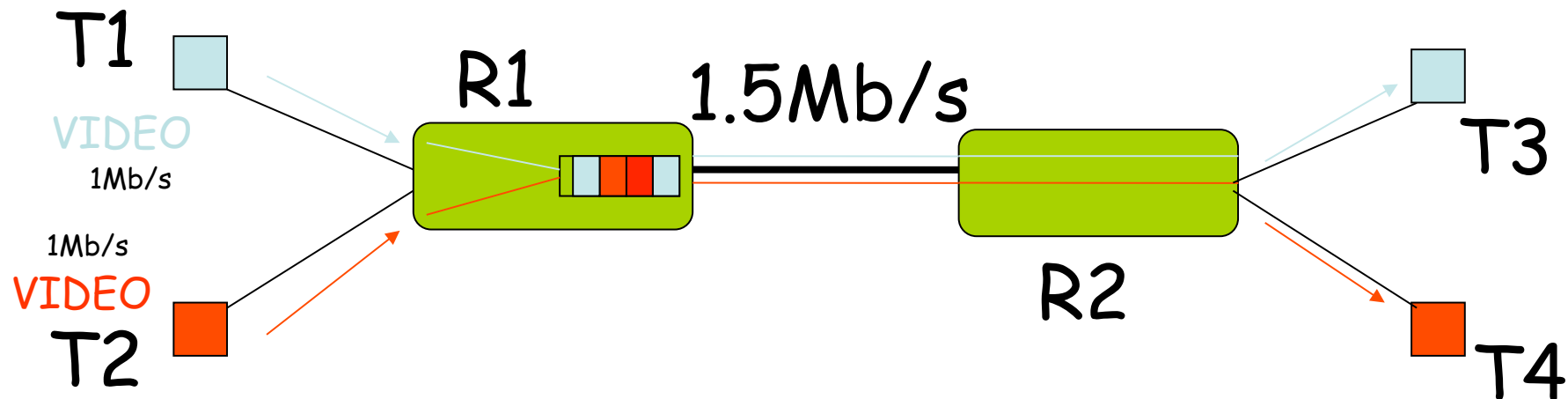


3^e principe

- Utilisation efficace des ressources
- Approches
 - Solutions dynamiques
 - e.g. libération explicite de ressources, mesure de ressources disponibles

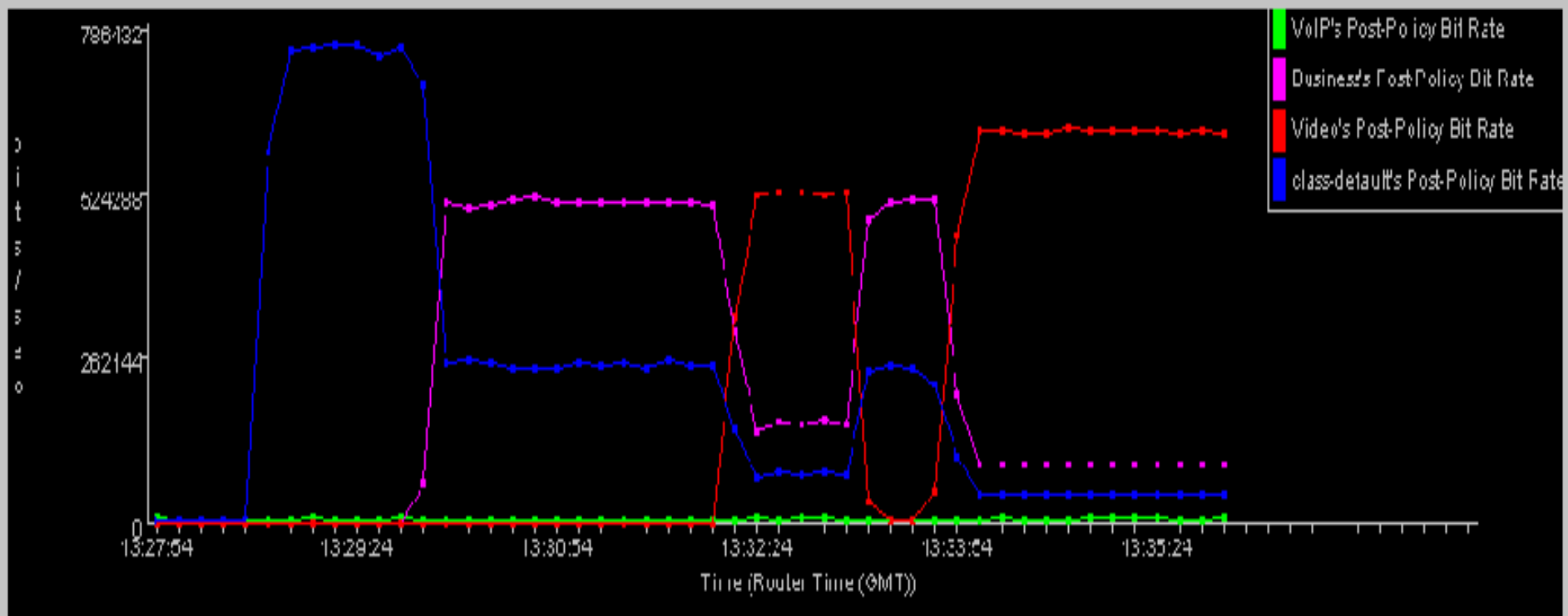
4^e cas

- 2 flux vidéo de 1 Mb/s
 - demandent plus que la capacité du lien partagé
- Que se passe-t-il ?



4^e principe

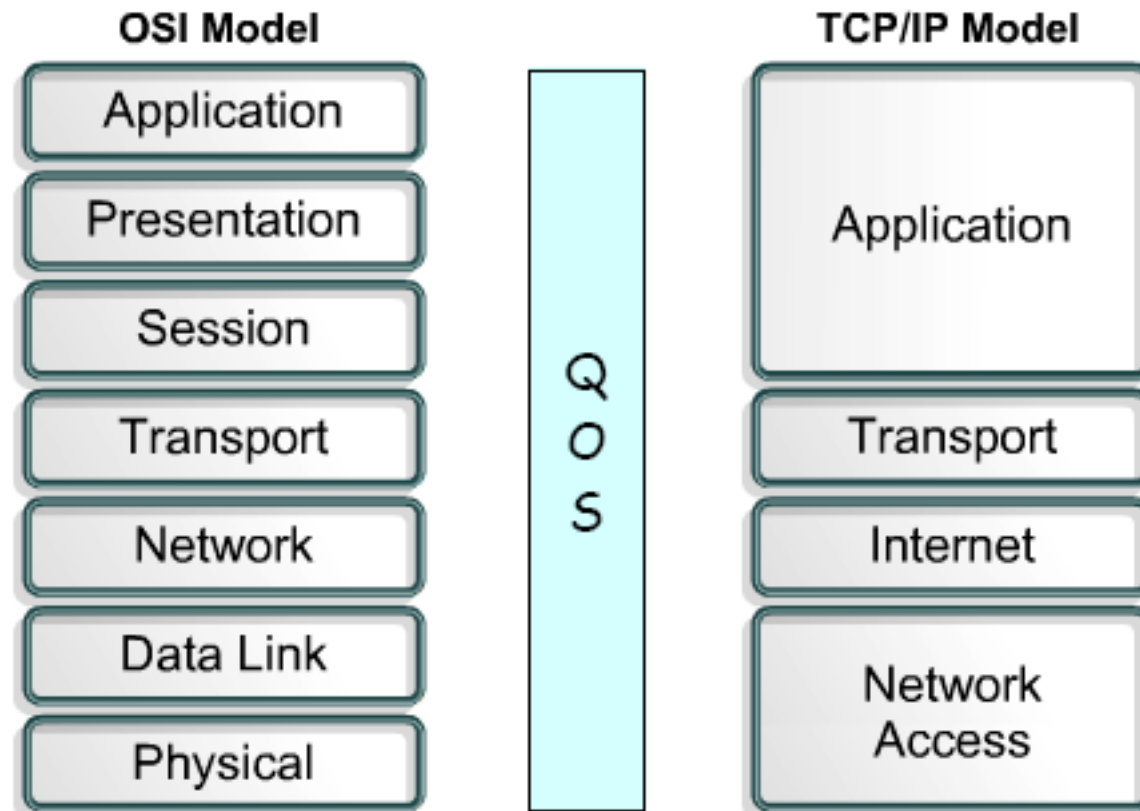
- **Contrôle** des flux en entrée
- Approches
 - **Contrôle d'admission**
 - Admission ou refus de l'entrée d'un flux dans le réseau
 - **Gestion des files d'attente**
 - Rejet de paquets en cas de congestion
 - **Régulation des débits des flux**
 - Pour atteindre un partage donné sans saturer le lien



Serial1/1, Output Policy DS_PP3, Per-class Post-Policy Bit Rate

Mise en œuvre des principes

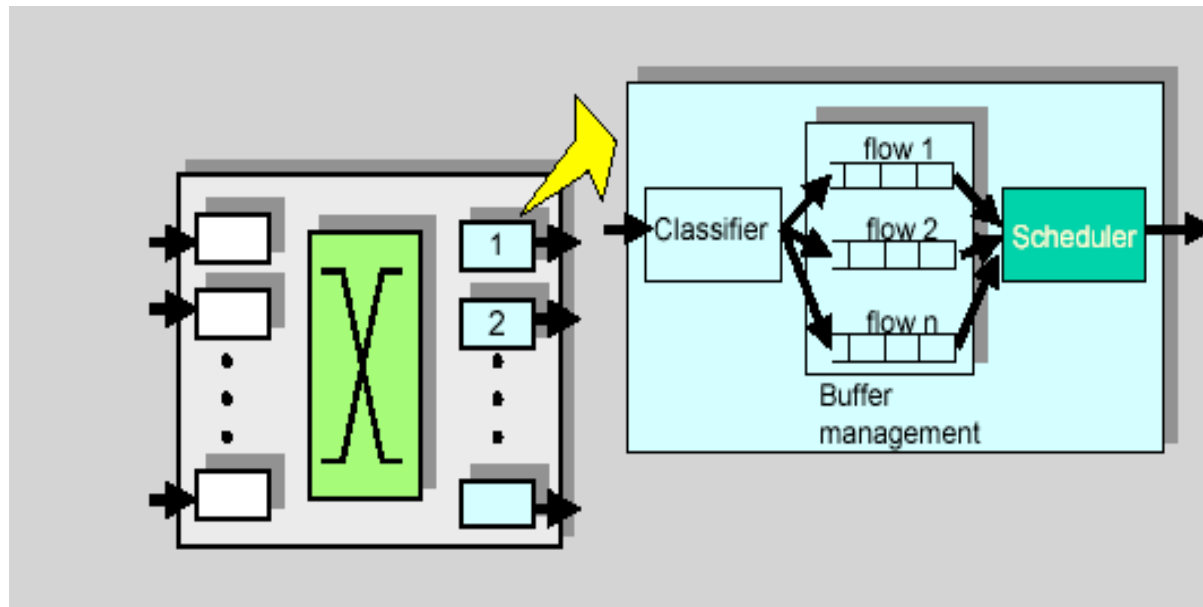
- Pas de solution magique / unique
- Variété de solutions
 - Combinaison de blocs de base
 - Compromis entre garanties et extensibilité



Synthèse sur les blocs de base pour la QoS

Quels blocs ?

Modèle de routeur utilisé



Classificateur

- Classifier
- Inspecte le paquet entrant, l'identifie, puis le classe
 - Lui associe une classe de service
 - Marquage
- Classe de service
 - Définit les traitements à appliquer dans le réseau
- Différentes règles de classification possibles
- En entrée du réseau, mais
 - changement de marquage en cours de route possible

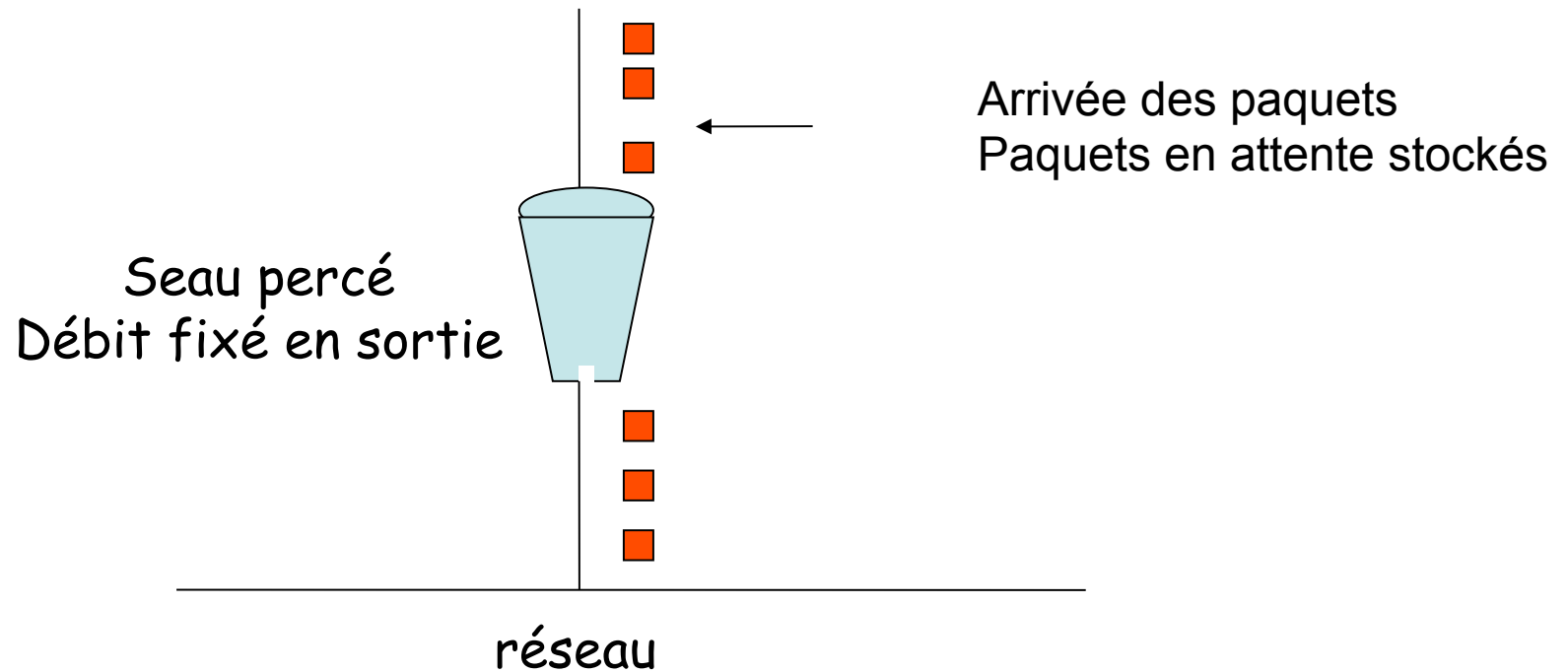
Régulateur

- Policer
- Assure que le trafic est conforme à un débit fixé
 - Limite de bande passante
- Trafic qui dépasse cette limite est rejeté
 - Régulateur strict (hard policing)
- Régulateur plus souple (soft policing)
 - Marquage différent du trafic qui dépasse la limite
- Peut être couplé à un outil de mesure

Lisseur

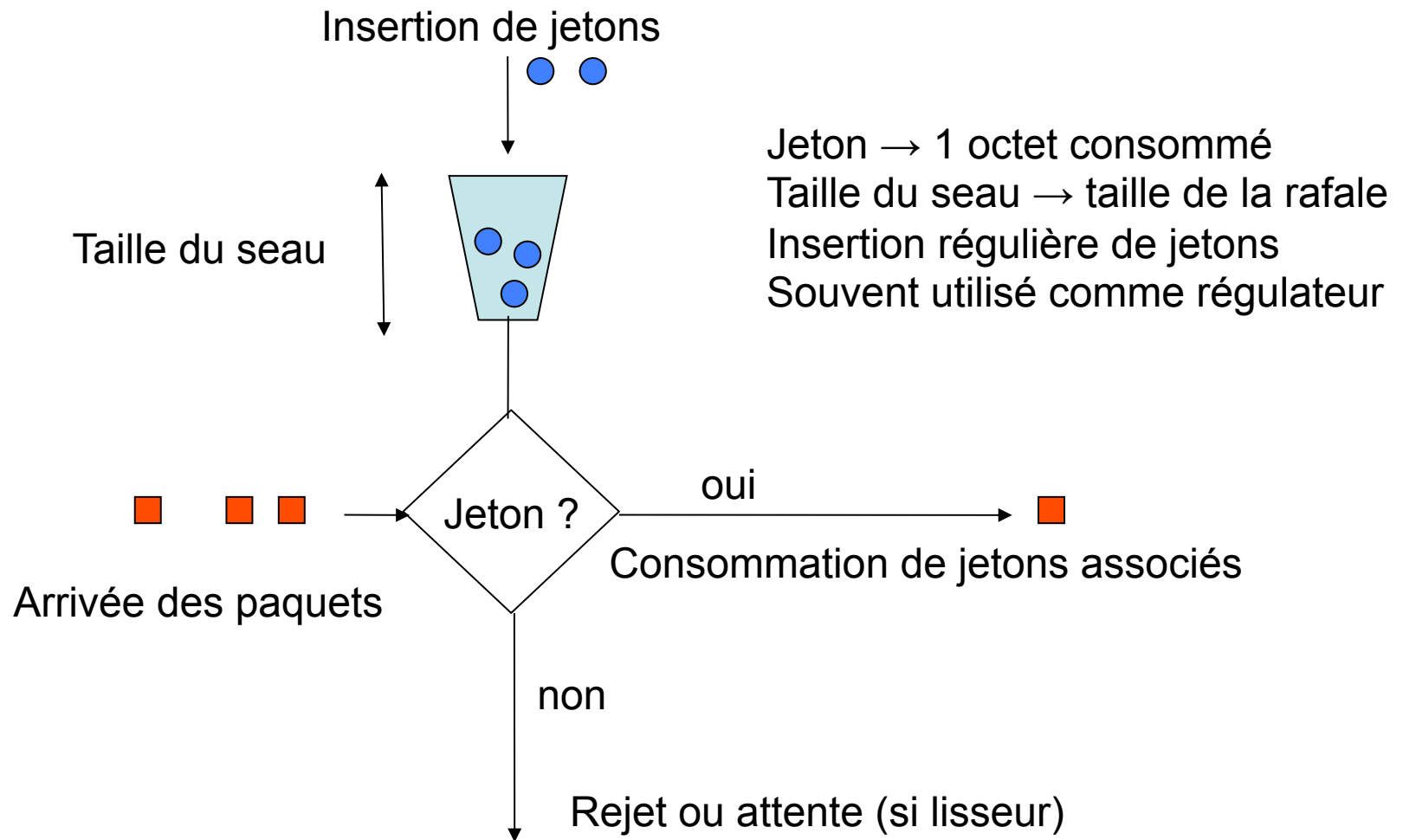
- Shaper
- Fait en sorte que le trafic soit conforme à la limite de bande passante
 - Stockage du trafic non conforme
 - Délai introduit sur les paquets du trafic non conforme
- « TCP-friendly »

Seau percé



Leaky bucket
Souvent utilisé comme lisseur
Certains paquets peuvent être rejetés
Délai d'attente qui peut être ajouté

Seau à jetons



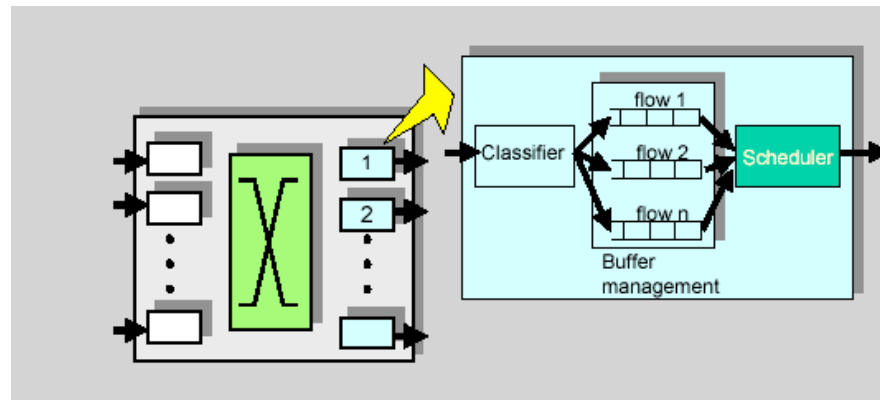
Gestion des files d'attente

- Queue management
- Paquet arrivant
 - D'abord stocké
 - Puis question : est-ce qu'on le garde dans la file ?
 - Accepté ou rejeté

Cf cours Contrôle de congestion

Ordonnancement

- Scheduler
- Opération de multiplexage
 - n entrées \rightarrow 1 sortie



FIFO

- 1 seule file d'attente
- Avantages
 - Simple
 - Performances correctes si on est loin de la saturation
 - Adapté aux applications basées sur TCP non sensibles au délai
- Inconvénients
 - Pas de différenciation de service en cas de saturation
 - Pas de contrôle du délai et de la gigue
 - Les flux goutons peuvent occuper tout le buffer

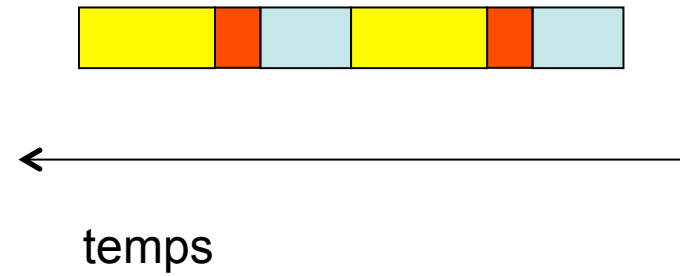
Round-Robin Queueing

- File d'attente par flux
 - FIFO par file
- Ordonnancement Round-Robin entre les différentes files
- Inconvénient ?

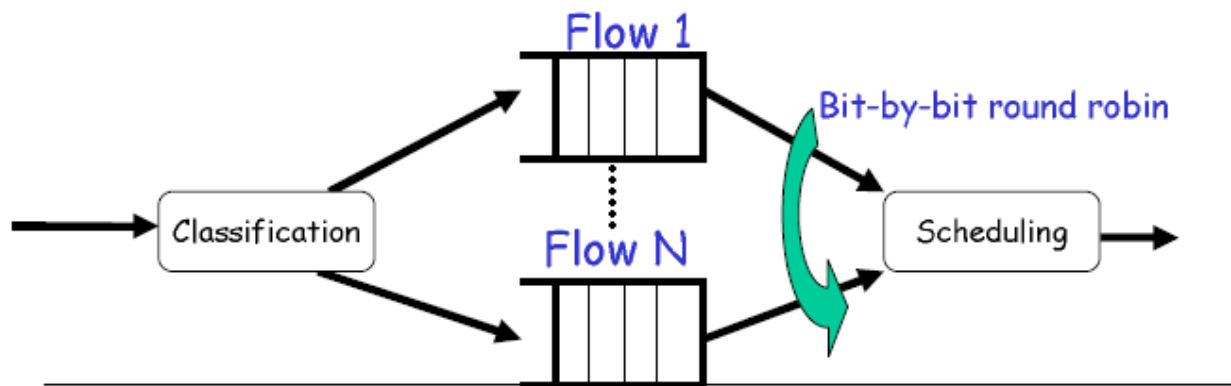
Round-Robin Queueing



→
R-R Q



Fair Queueing



Inspiré des slides Nick McKeown

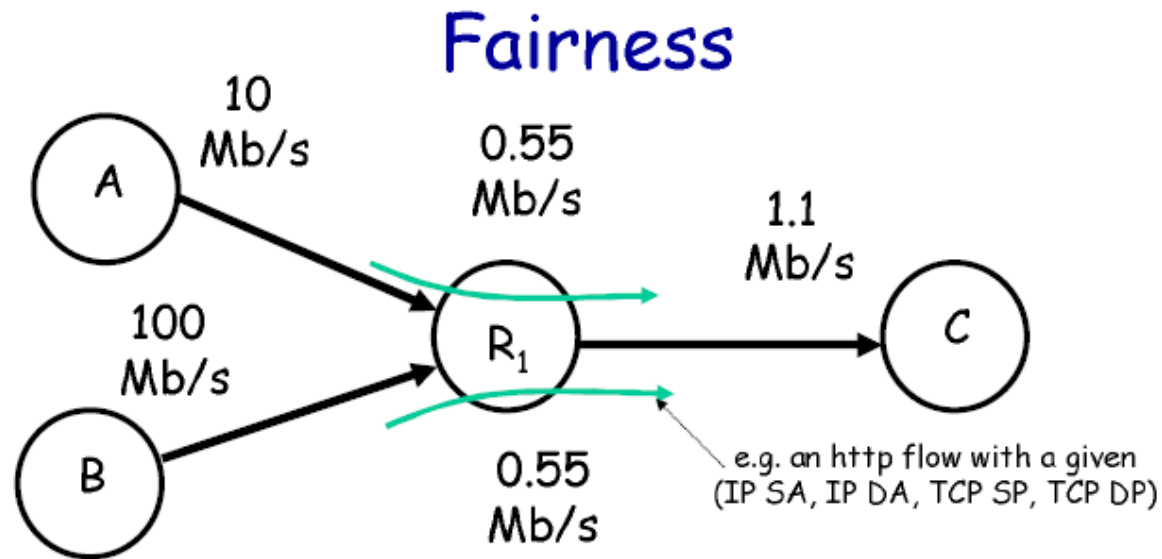
- Emulation d'un round-robin bit à bit
 - * Marquage bit à bit
 - * Sortie du paquet quand on marque le dernier bit

Fair Queueing

- **Avantage**
 - Isolation des flux
- **Inconvénient**
 - Compliqué à réaliser

Equité

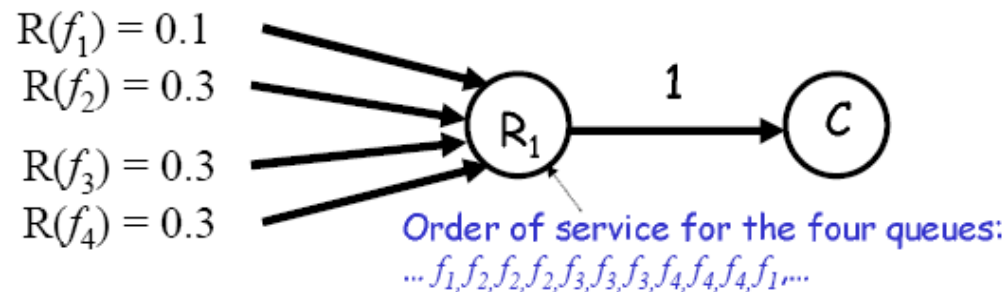
Même débit pour tout le monde ?



What is the "fair" allocation:
(0.55Mb/s, 0.55Mb/s) or (0.1Mb/s, 1Mb/s)?

Weighted Fair Queueing

- Poids au niveau de chaque file
- Fair queueing basé sur les poids



Also called "Generalized Processor Sharing (GPS)"

Inspiré des slides Nick McKeown

Weighted Fair Queueing

- Avantage
 - Permet de fournir différents débits aux flux
- Inconvénients
 - Difficile à implémenter
 - Quels schémas d'équité veut-on obtenir ?

Priority Queueing

- Ordre de priorités sur les différentes files d'attente
- Les paquets de la file la plus prioritaire sortent toujours en premier
- Une file est servie s'il n'y a aucun paquet dans les files d'attente plus prioritaires
- Avantages
 - Simple
 - Vraie différenciation de service
- Inconvénient
 - Famine possible sur les flux moins prioritaires

Weighted Round Robin

- Poids sur chaque file d'attente
- Nombre de paquets servis dans une file dépend du poids
- Politique de round robin appliquée entre les files
- Avantages
 - Simple
 - Différenciation de service sans famine d'une file
- Inconvénient
 - Taille des paquets non prise en compte

Deficit Weighted Round Robin

- Quantum sur chaque file d'attente
 - Correspond à un poids exprimé en octets
 - Valeur minimum autorisée à chaque passage
- Crédits
 - Indique ce qui n'a pas été utilisé après le service de la file
- Compteur de déficit
 - = quantum + crédits
 - Correspond au nombre max d'octets qui peut être servi dans la file à chaque passage
- Examen round-robin des files
- Possibilité d'utiliser les quantaux des files vides
- Avantages
 - Taille des paquets prise en compte
 - Meilleure équité
- Inconvénient
 - Pas de file totalement prioritaire

Priority-based Deficit Weighted Round Robin

- File d'attente prioritaire
 - Servie en 1er si elle a des paquets
- Files moins prioritaires sont servies selon la politique DWRR
- Plusieurs files d'attente prioritaires
 - Servies selon DWRR
- Souvent implanté dans les routeurs
 - Mais pas de standardisation

Utilisation des classes de services

- Approches précédentes
 - Autant de files d'attente que de flux
 - Problème d'extensibilité
- Une file d'attente par classe de service (CoS)
 - Application des techniques précédentes sur les différentes files

Conservatif – Non conservatif

- Ordonnancement conservatif
 - S'il y a au moins un paquet en attente dans le routeur, alors le lien est utilisé et un paquet est envoyé
- Ordonnancement non conservatif
 - Le lien peut ne pas être utilisé même s'il y a un paquet dans le routeur
 - Délai plus grand mais gigue réduite

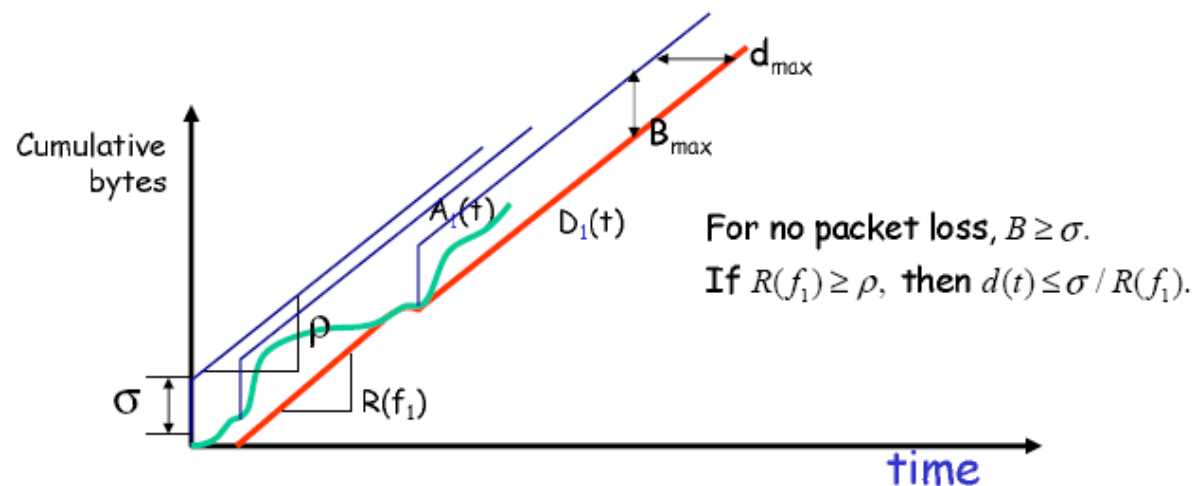
Garantie de débit

- Possible de garantir un débit sur un lien
 - WFQ / DWRR
 - Si pas de saturation du lien
- Possible de garantir un débit sur un chemin
 - Si le débit cherché n'excède pas la capacité des liens
 - Si pas de saturation du chemin

Garantie sur le delai

- On peut offrir des garanties sur le délai de bout-en-bout

(σ, ρ) Constrained Arrivals and Minimum Service Rate



Theorem [Parekh, Gallager '93]: If flows are leaky-bucket constrained, and routers use WFQ, then end-to-end delay guarantees are possible.