Université Nice Sophia Antipolis
Membre de UNIVERSITÉ CÔTE D'AZUR

iut réseaux & télécoms Sophia Antipolis

Module M3102-03
Réseaux d'opérateurs, d'accès, étendus

L. Sassatelli

sassatelli@unice.fr

http://www.i3s.unice.fr/~sassatelli

Situation dans la formation DUT R&T

Semestre 3							
UE 31: Approfondissement métiers						Volume horaire étudiant en formation	
Référence	Nom module	Coeff	CM	TD	TP	encadrée	dirigée
M3101	Infrastructure sans fil d'entreprise	2	6	6	18	30	
M3102	Technologies de réseaux opérateurs	3	15	12	33	60	
M3103	Technologies d'accès	1,5	9	6	15	30	
M3104	Gestion d'annuaires unifiés	1,5	6	6	18	30	
M3105	Services réseaux avancés	2	6	6	18	30	
M3106	Transmission large bande	1,5	6	12	12	30	
M3107	Réseaux cellulaires	2	9	6	15	30	
M3108 C	Supervision des réseaux	1,5	9	9	12	30	
M3109	PT :Gestion de projet	1		15		15	90
Total UE31		**16**	66	78	141	**285**	90
UE 32: Renforcement des compétences transversales et scientifiques						Volume horaire étudiant en formation	
Référence	Intitulé	Coeff	CM	TD	TP	encadrée	dirigée
M3201	Anglais: Le monde du travail	3		15	30	45	
M3202	EC: S'insérer dans le milieu professionnel	2		9	21	30	
M3203	PPP : Savoir collaborer	1		12	18	30	
M3204	Matrices et graphes	2	6	21	3	30	
M3205	Transmissions guidées en hyperfréquence et optique	2	9	12	9	30	
M3206	Automatisation des tâches d'administration	2	6	6	18	30	
M3207 C	Sécurité et performance	2	9	9	12	30	
Total UE 32		**14**	30	84	111	**225**	0
Total semestre 3		**30**	96	162	252	**510**	90

Notation

- DS 1 (écrit 1h30), semaine 47 : 0.3
- DS 2 (écrit 1h30), semaine 3 : 0.3
- QCM (surprises en séance) : 0.1
- TP (note rapport+séance) : 0.3

Plan général du cours

I. **Organisation des opérateurs de l'Internet**
II. **TCP et Qualité de service**
III. **Commutation par circuits virtuels**
IV. **Commutation par circuits virtuels dans le mode IP : MPLS**

Plan détaillé du cours

I. Organisation des opérateurs de l'Internet

I.1. Introduction aux réseaux étendus (WAN)

I.2. Hiérarchie et relations entre opérateurs (ISP)

I.3. Routage entre AS : le protocole BGP

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

II.1.a. Principe du contrôle de congestion

II.1.b. Rappels : transfert fiable et contrôle de flux

II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

II.2.a. Définition de la QoS et classes d'applications

II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS

II.2.c. Stratégies possible pour garantir un niveau de QoS

II.3. Techniques de traitement de la QoS

II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing

II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets

II.3.c. Gestion de buffer : comment abandoner les paquets en excès

II.3.d. Marquage du trafic : DiffServ dans IP

Plan détaillé du cours

III. Commutation par Circuit Virtuel (VC)

III.1. Commutation par circuit et Commutation par paquet

III.2. Commutation par circuit virtuel : une combinaison des deux

III.3. Avantage de la commutation par VC pour la QoS

IV. Commutation par VC dans le monde IP : MPLS

IV.1. Fonctionnement de MPLS

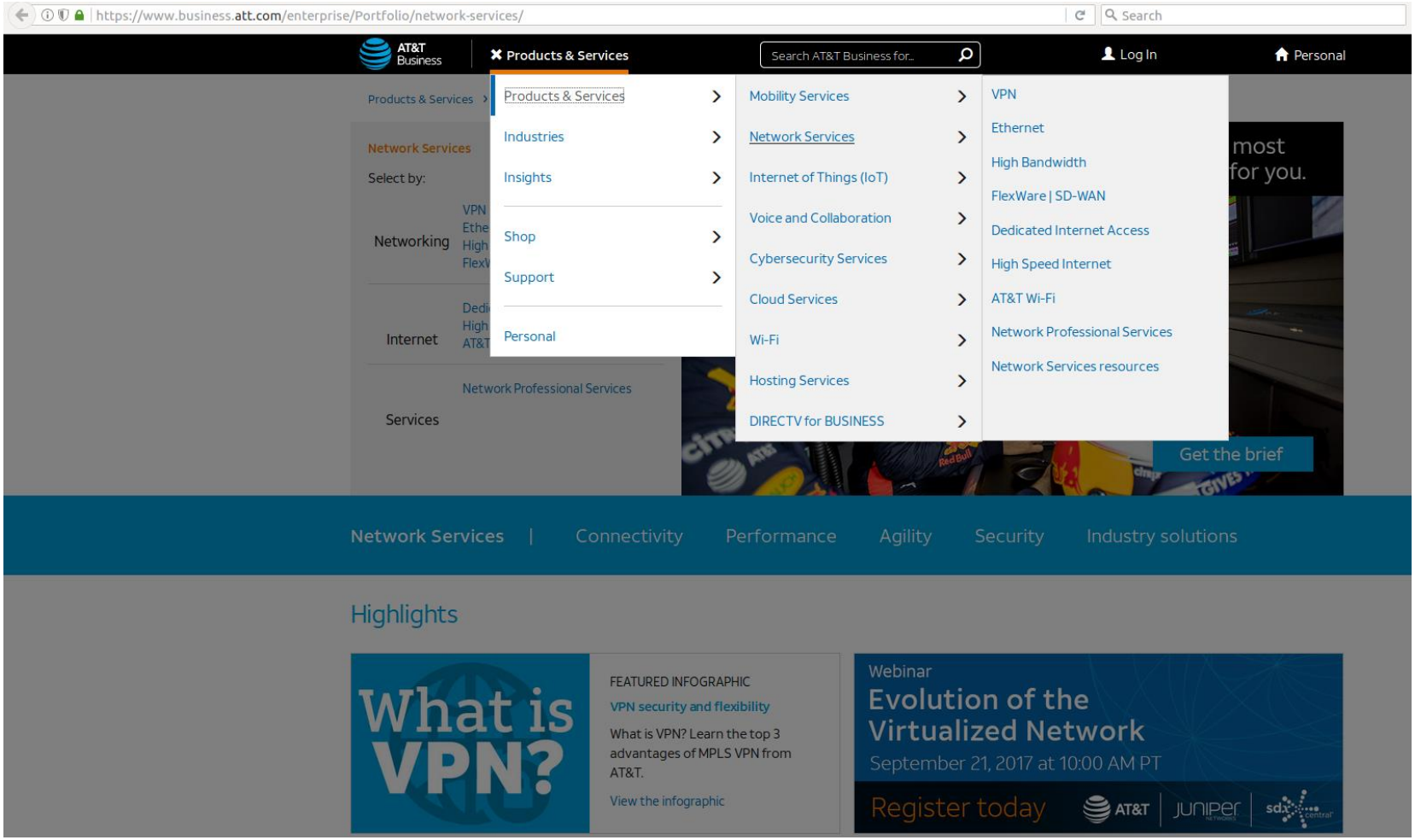
IV.2. Ingénierie de trafic avec MPLS : MPLS-TE

IV.3. Offres de service MPLS : les VPN basés sur MPLS

IV.3.a. IP-VPN

IV.3.b. Ethernet-VPN : VPLS

Les offres d'interconnexion de LAN



Références principales

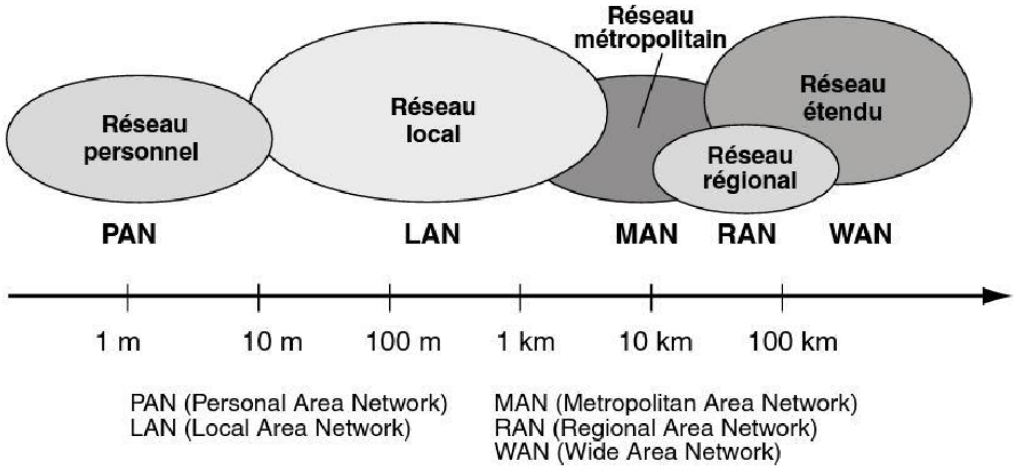
- G. Pujolle, « *Les réseaux édition 2008 »*, ed. Eyrolles
- http://www.cisco.com/en/US/docs/internetworking/technology/handbook/ito_doc.html
- A. S. Tanenbaum, « *Computer Networks, Fourth edition »,* ed. Prentice Hall
- J. F. Kurose and K. W. Ross, « *Computer Networking, a top-down approach, Fifth edition »*, ed. Pearson education
- J. F. Kurose and K. W. Ross, slides of chapter 3, online
- Cours de J. Drouot (ESIL)
- AT&T website: http://www.business.att.com/enterprise/business-solutions/
- Cours de E. Bost (Freescale Semiconductors)
- C. Servin, “Réseaux et Télécoms”, 3e édition, ed. Dunod

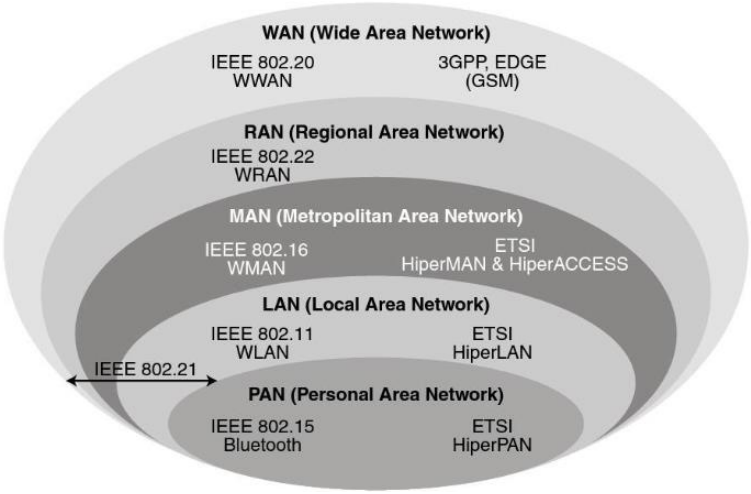
Plan

I. Organisation des opérateurs de l'Internet
   - I.1. Introduction aux réseaux étendus (WAN)
   - I.2. Hiérarchie et relations entre opérateurs (ISP)
   - I.3. Routage entre AS : le protocole BGP

II. TCP et Qualité de service
III. Commutation par circuits virtuels
IV. Commutation par circuits virtuels dans le mode IP : MPLS

Classification des réseaux informatiques

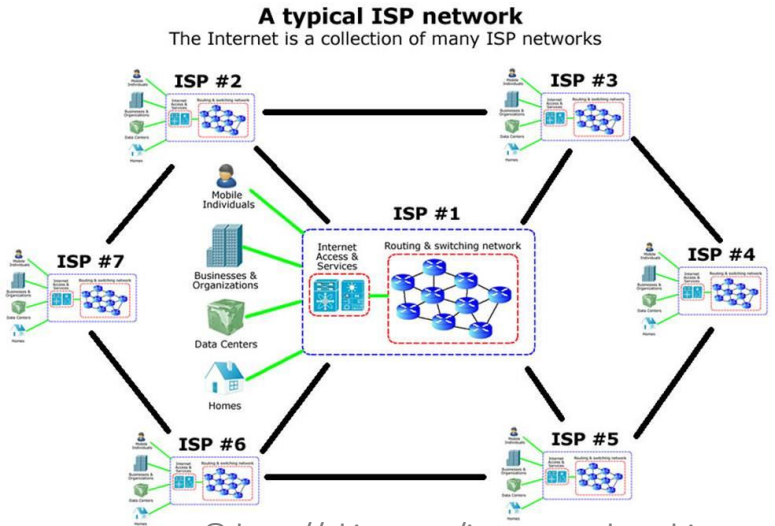




-Personal Area Network: quelques mètres, pour équipements personnels
- Local Area Network: réseaux intra-entreprises, jusqu'à plusieurs Mbps
- Metropolitan Area Networks: interconnexion des entreprises sur un réseau spécialisé à haut débit
-Regional Area Network: 50km pour le sans-fil, beaucoup d'utilisateurs par antenne
- Wide Area Networks: pays ou plusieurs continents

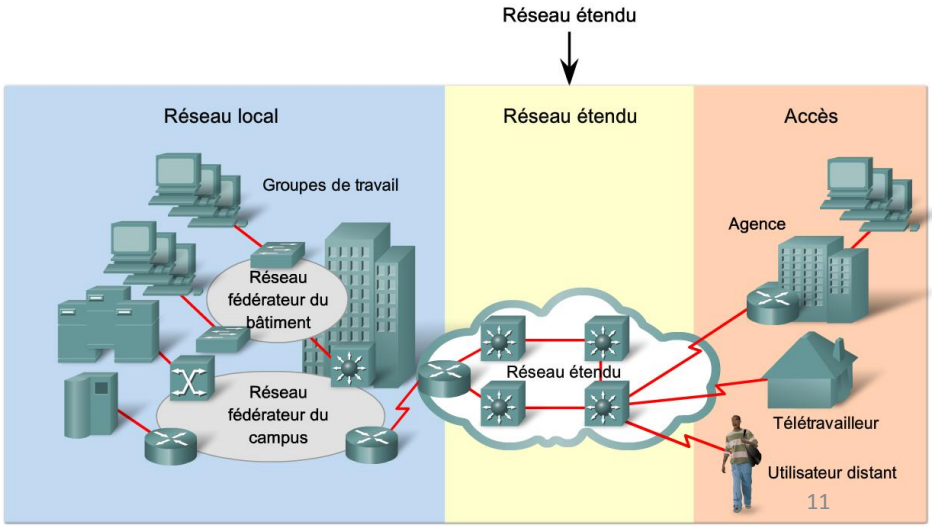
Réseau d'opérateurs et Réseaux étendus: définition

Réseau d'opérateur (ISP): Infrastructure physique appartenant à une entreprise de télécoms donnée, supportant l'Internet public.

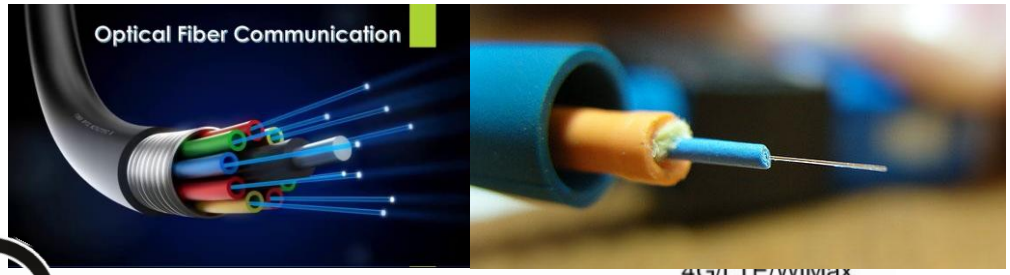


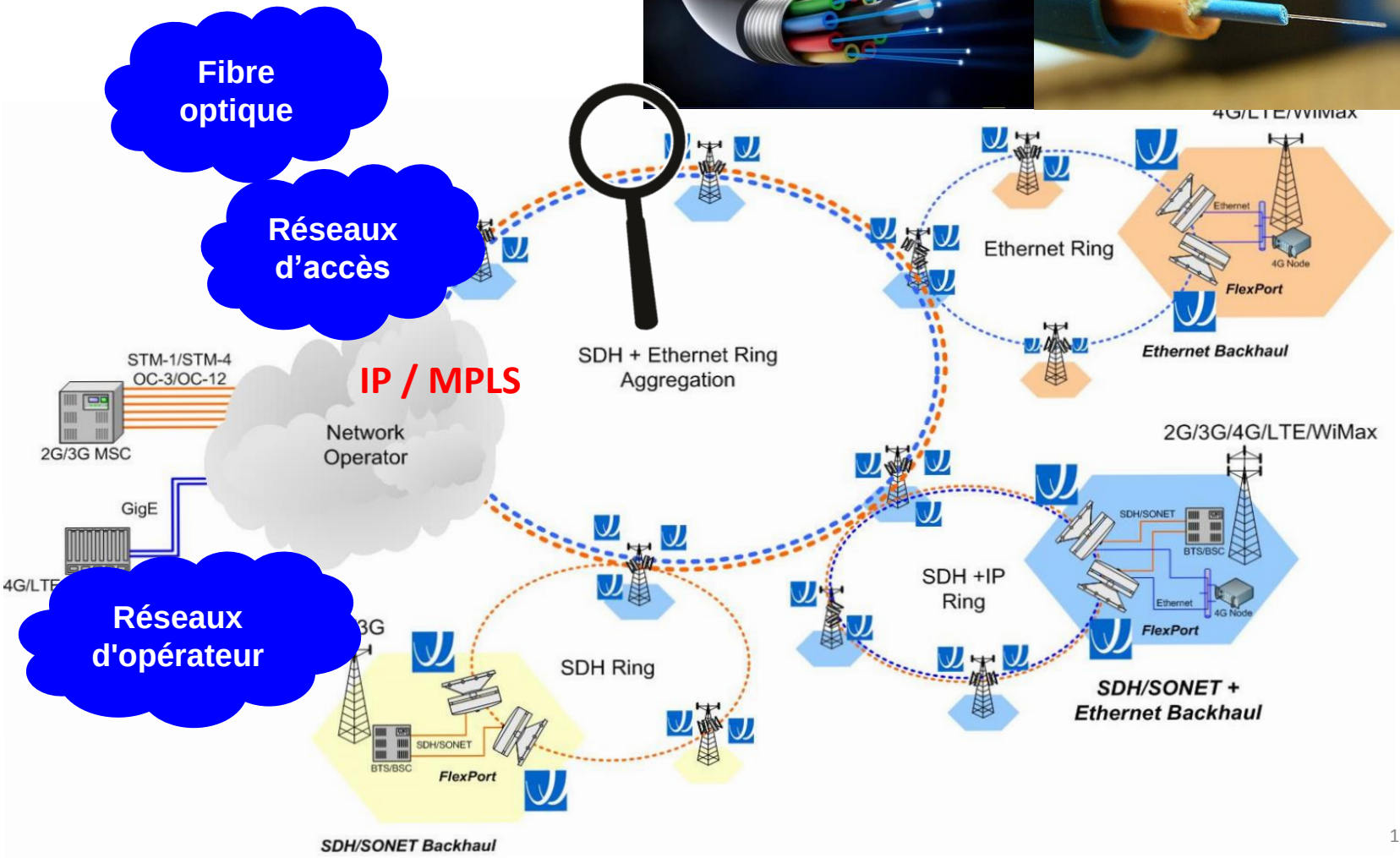
© http://akioz.com/isp-network-architecture/

Réseau étendu (WAN): Interconnexion de réseaux locaux (d'entreprise) distants, pour constituer un réseau d'entreprise sur une large aire géographique.



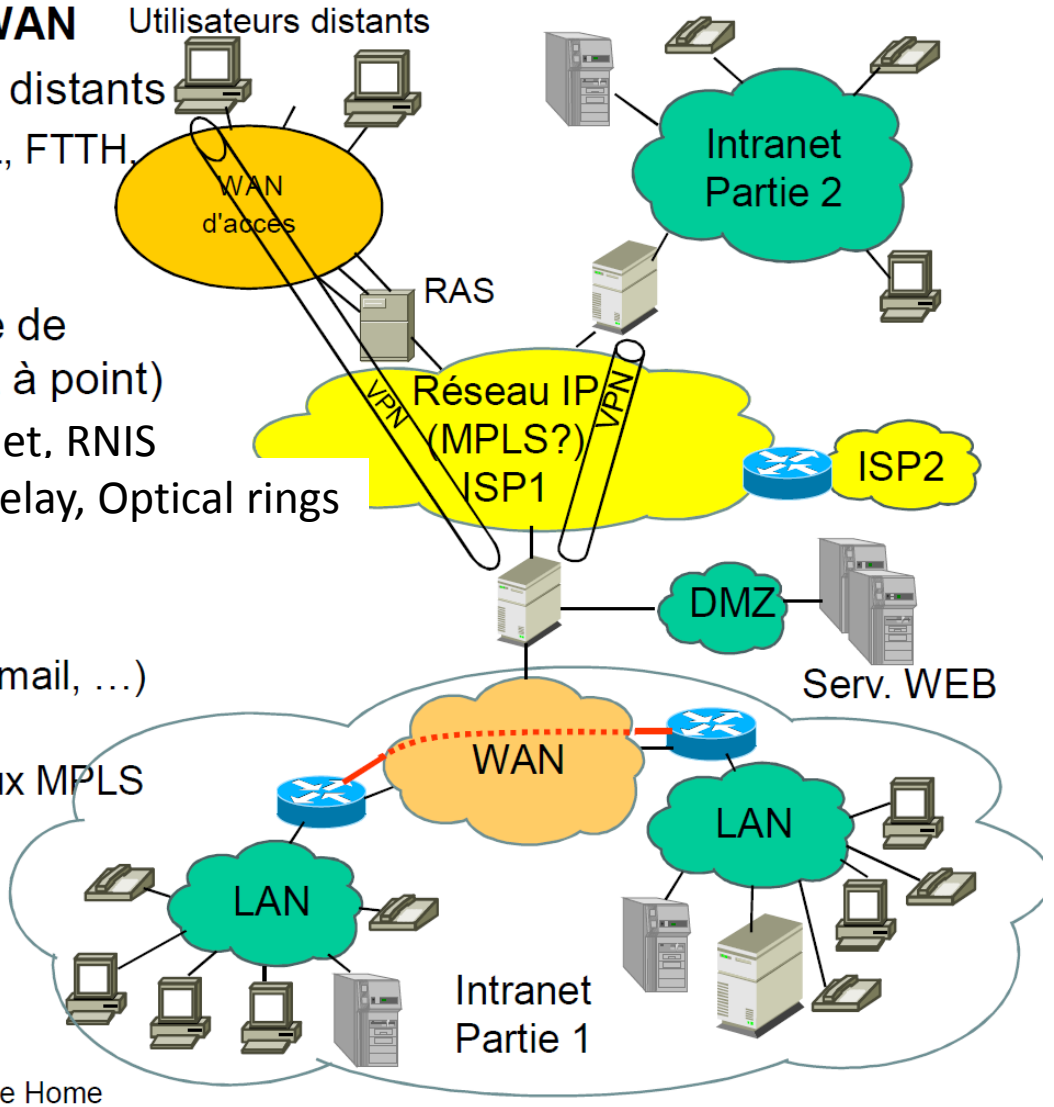
Exemple de réseau d'opérateur mobile





Les différents aspects des WAN

- **Trois aspects des réseaux WAN**
  - Raccorder des utilisateurs distants
    - RTC, RNIS, câble, ADSL, FTTH, UMTS, …
  - Raccorder les LAN d'une entreprise sous le contrôle de l'entreprise (services point à point)
    - Liaisons louées, Ethernet, RNIS
    - IP-VPN, VPLS, Frame Relay, Optical rings
  - Réseaux WAN des ISP, interconnectés entre eux
    - Services Internet (WEB, mail, …) sur réseau IP pur
    - Services VPN sur réseaux MPLS

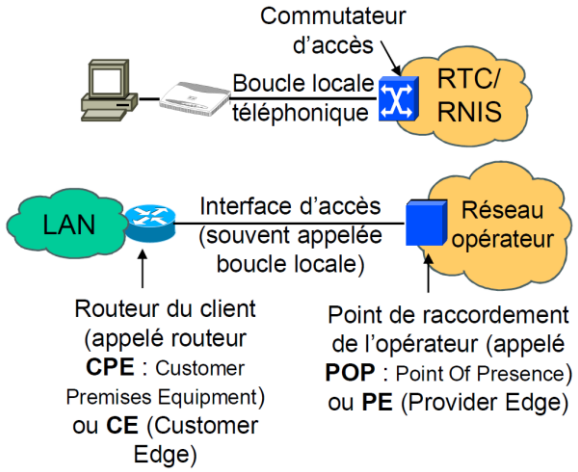


RAS : Remote Access Server
DMZ : De-Militarized Zone
ISP : Internet Service Provider
VPN : Virtual Private Network FTTH : Fiber To The Home

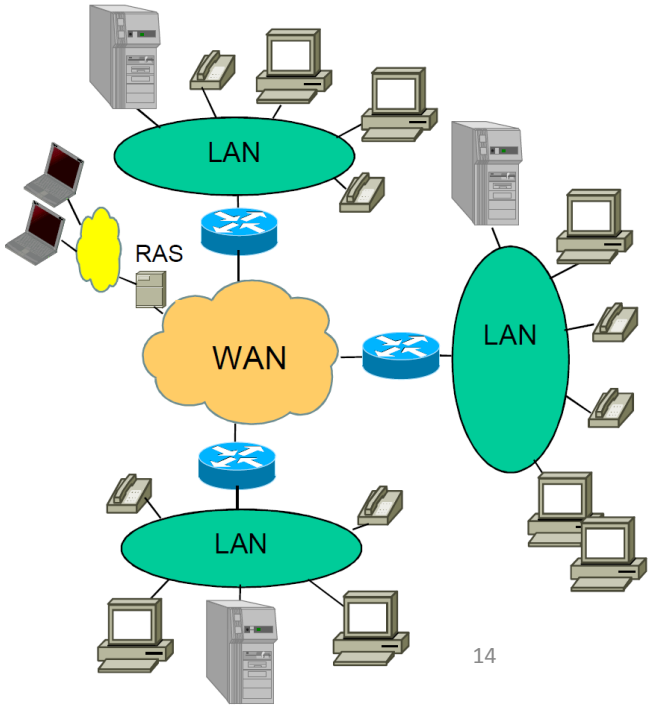
Réseaux étendus: présentation

- Caractéristiques de performance des WAN:

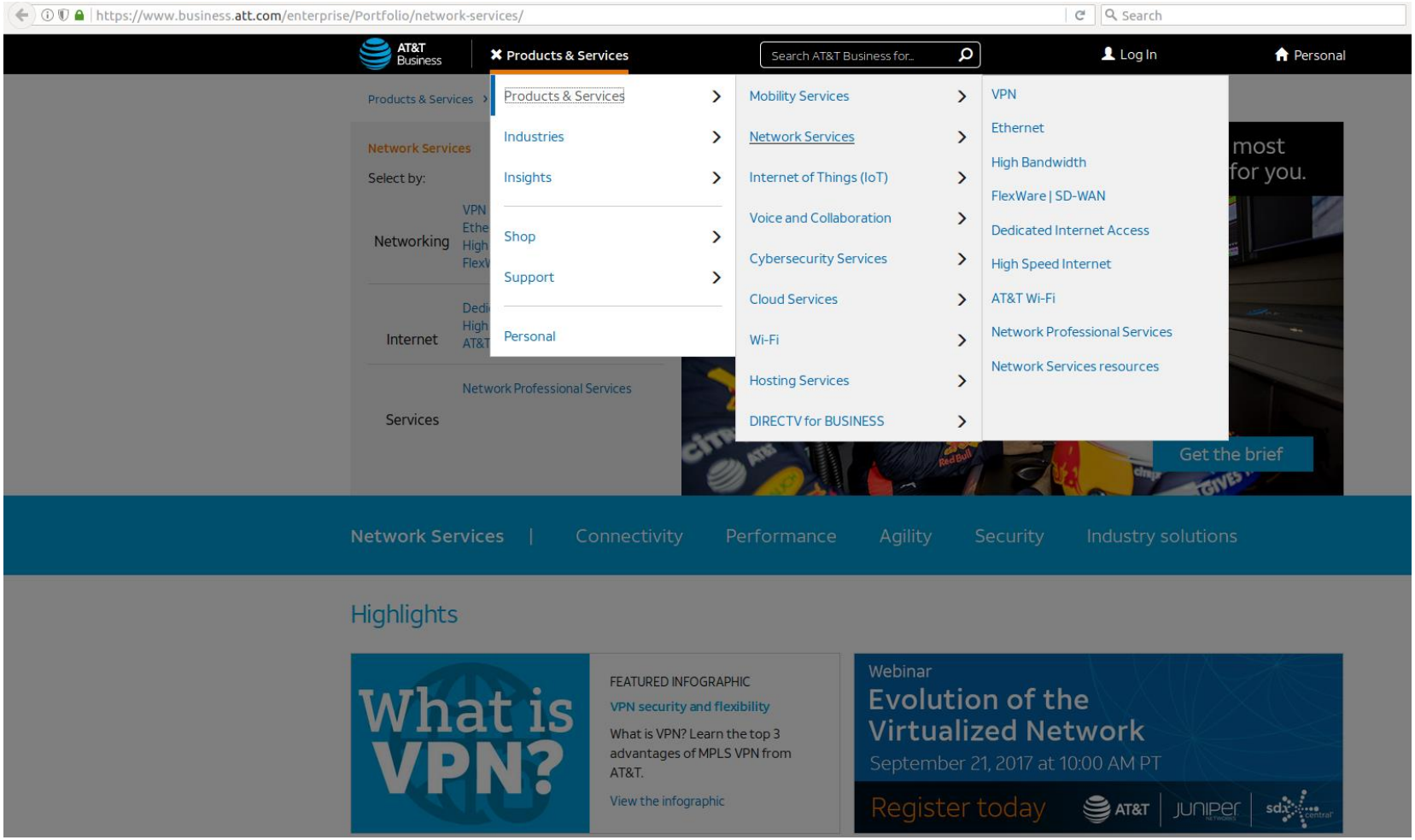
  - Prix beaucoup plus élevés que ceux des LAN
    - Même si les prix baissent rapidement
  - Délais de traversée du réseau plus élevés
    - Plusieurs dizaines ou centaines de ms



- Besoin croissant de convergence et de qualité de service:
  - Un seul réseau pour les trafics de données et les trafics télécoms
    - Au niveau LAN
    - Au niveau WAN
  - Les trafics doivent être répartis selon des classes de service et traités en conséquence



Les offres d’interconnexion de LAN

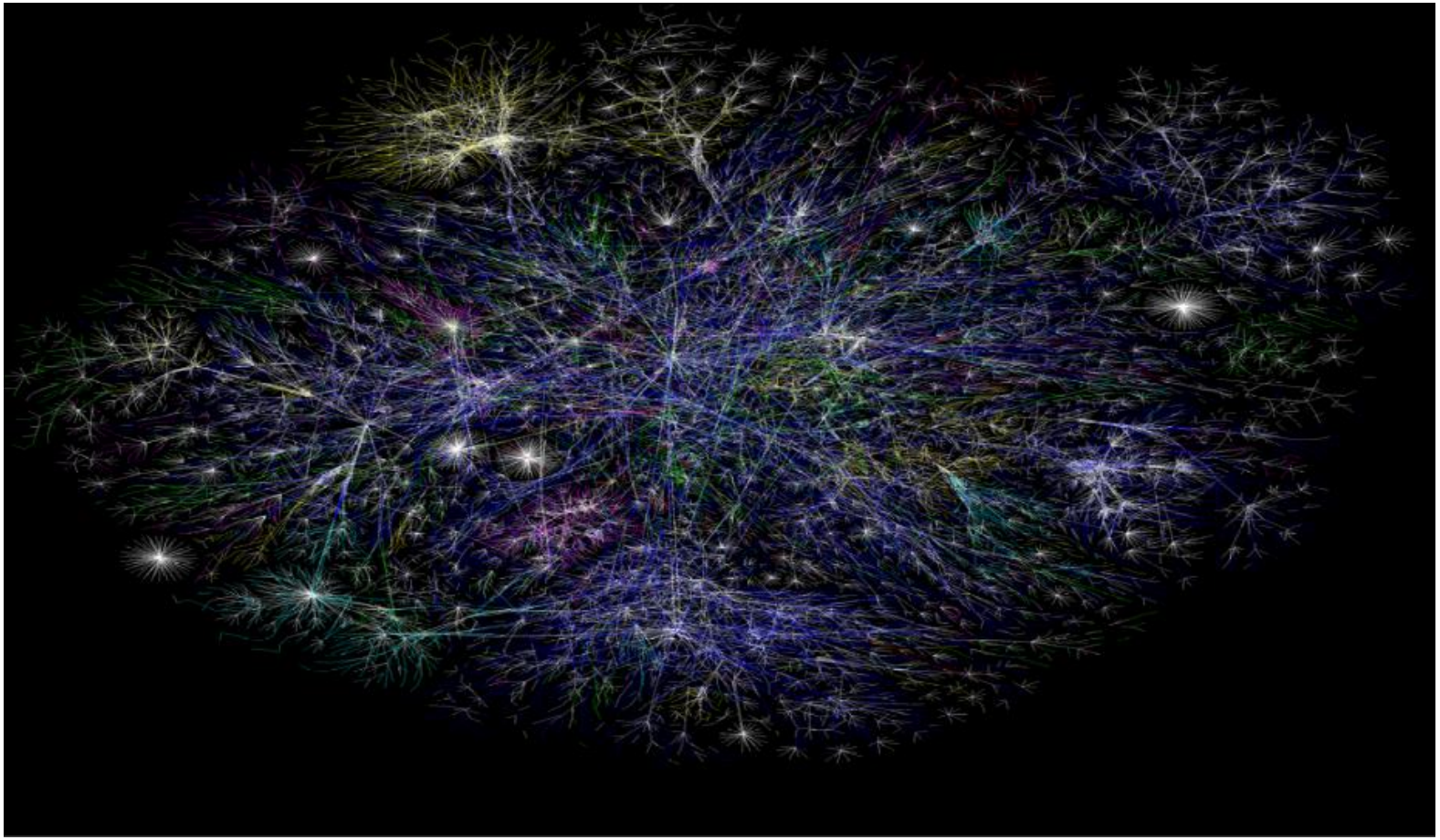


Plan

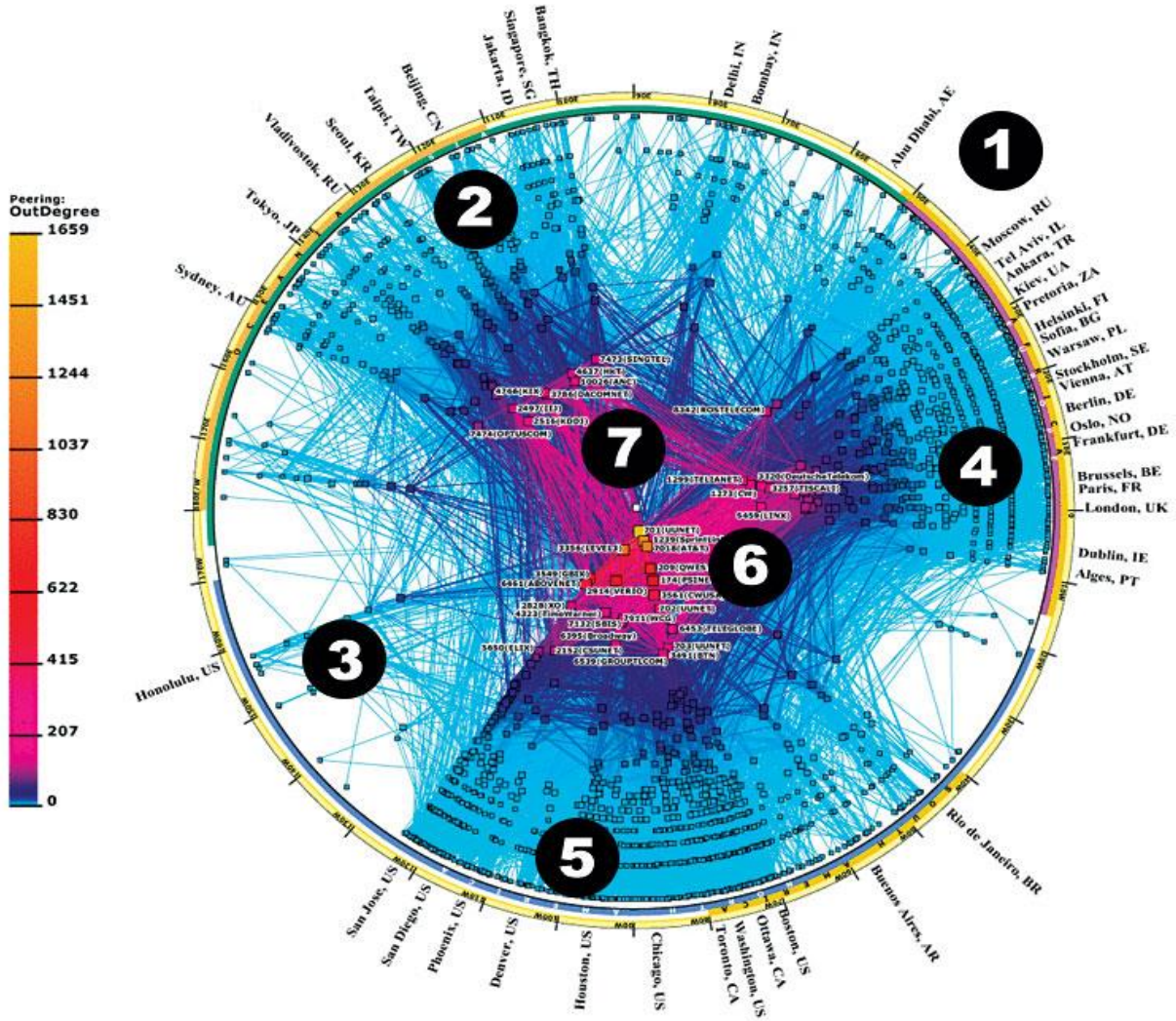
I. Organisation des opérateurs de l'Internet
   - I.1. Introduction aux réseaux étendus (WAN)
   - I.2. Hiérarchie et relations entre opérateurs (ISP)
   - I.3. Routage entre AS : le protocole BGP

II. TCP et Qualité de service
III. Commutation par circuits virtuels
IV. Commutation par circuits virtuels dans le mode IP : MPLS

L’Internet à l’échelle globale



L’Internet à l’échelle globale



Vue rapide de la structure de l'Internet

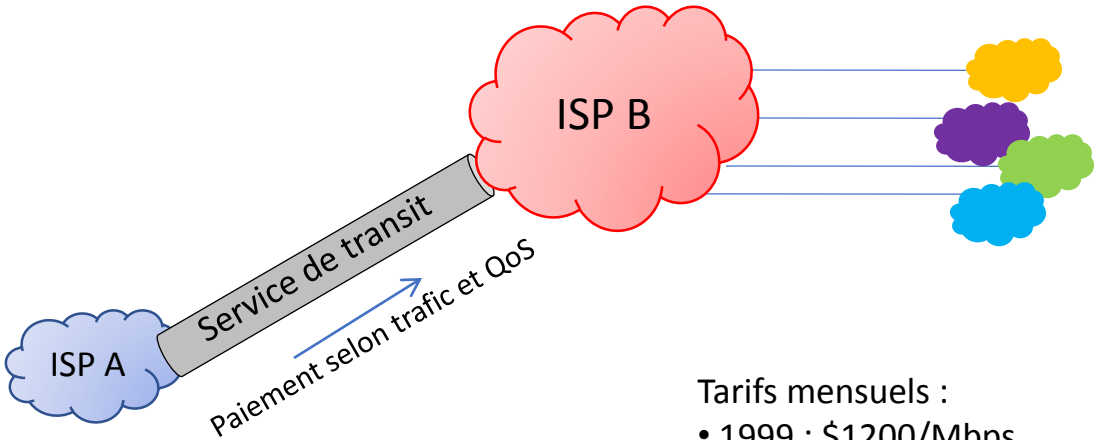
- On va voir rapidement comme l'Internet est interconnecté : comment le sont les ISPs
- Pour cela, on va expliquer les termes :
  - Internet transit
  - Internet peering
  - Internet Peering Ecosystem, Tier 1, Tier 2, Tier 3 ISPs, leurs positions et motivations, et le rôle des Internet Exchange Points (IXPs).

Relation de Transit entre 2 ISPs

- Def : Internet est un réseau de réseaux
- Def : un ISP vend de l'accès à l'Internet mondial

-> Un ISP doit être lui-même connecté à un ISP déjà connecté à l'Internet mondial

- 2 méthodes pour l'interconnexion d'ISPs : transit et peering
- Def : Transit est le service vendu par un ISP pour donner accès à Internet



Tarifs mensuels :
- 1999 : $1200/Mbps
- 2004 : $120/Mbps
- 2008 : $12/Mbps
- 2013 : $1/Mbps

Relation de Peering entre 2 ISPs

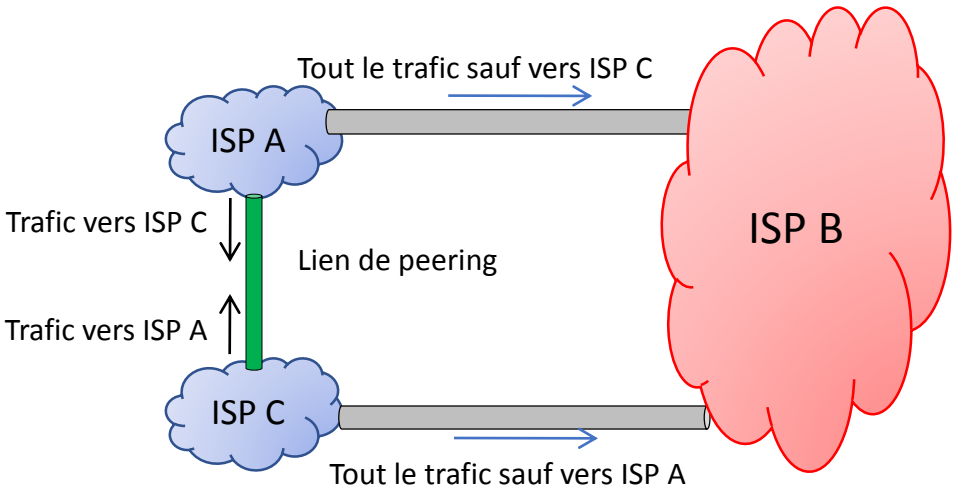
- Le transit semble facile et bon marché : pourquoi besoin d'un autre type d'intercx ?
  - $1/Mbps, si en moy 100Gbps -> $100K/mois : pas négligeable !
  - > d'où le peering

- Def : peering est l'échange réciproque d'accès aux clients de chacun des ISPs engagés

- Si 2 ISPs réalisent qu'ils s'échangent beaucoup de trafic et passent par le même ISP de niv sup pour transit :
  - Ils peuvent court-circuiter ISP B
  - Peut faire économiser

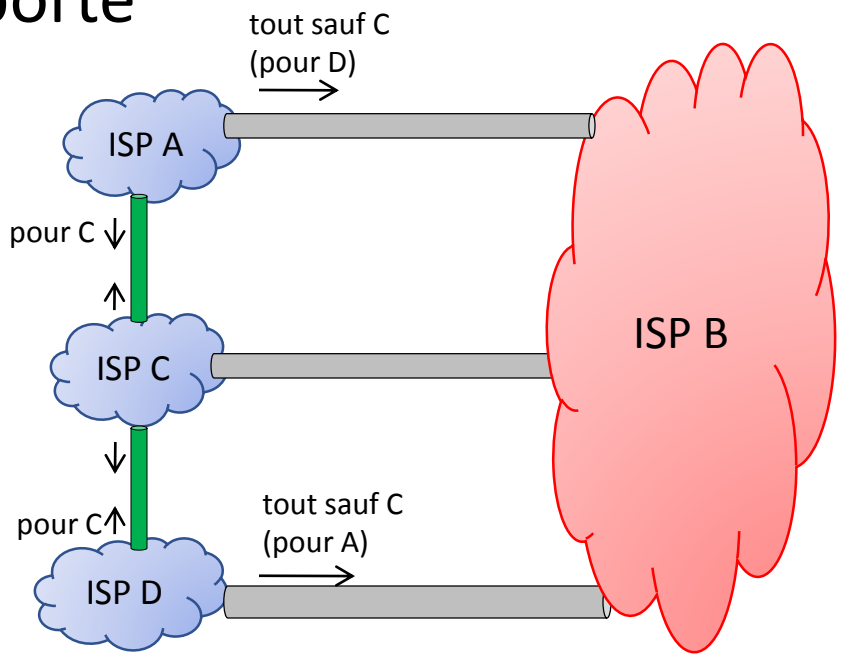
  selon le coût de l'installation de peering
  - Bénéfices en perf : shortest

  path plus court, délai plus faible

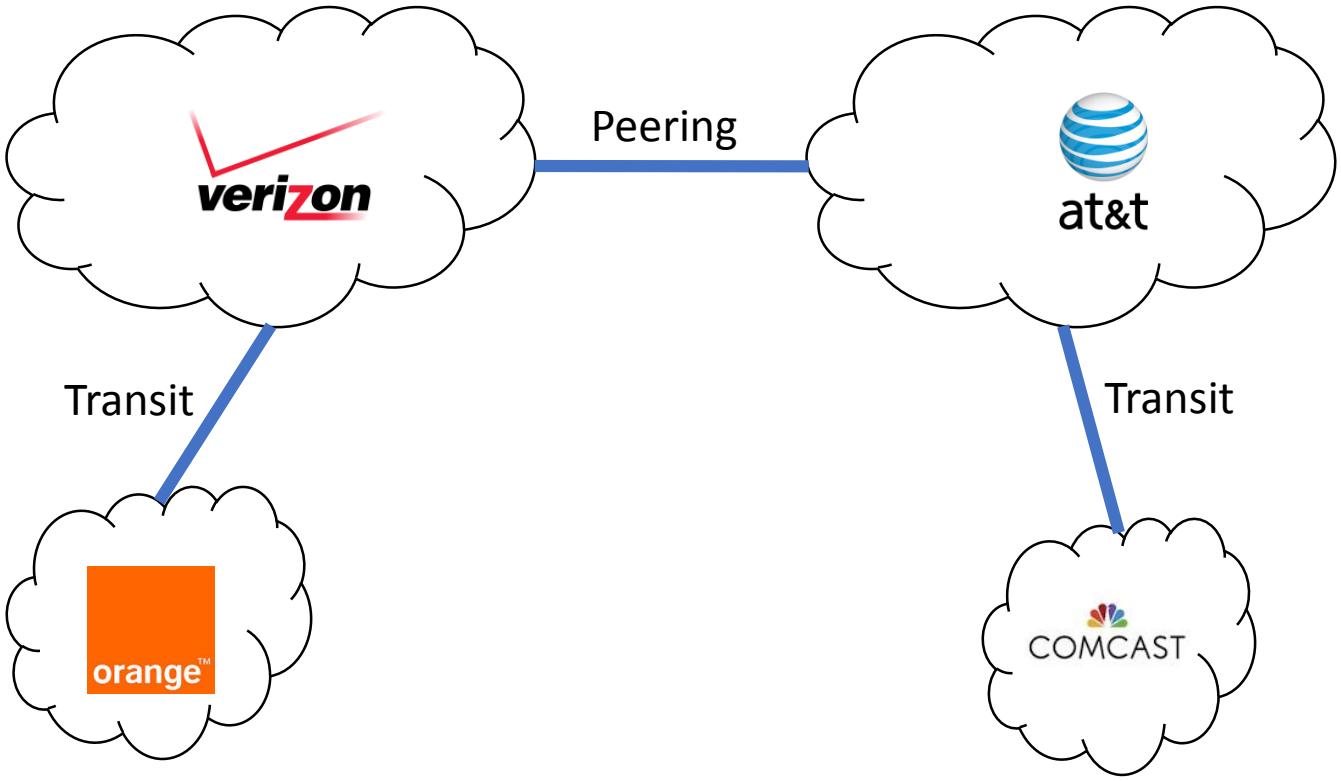


2 points importants sur le Peering

- La relation de peering n'est pas transitive
- La relation de peering n'apporte pas l'accès à tous les réseaux de l'ISP connecté, comme la relation de transit le fait.



Transit et peering

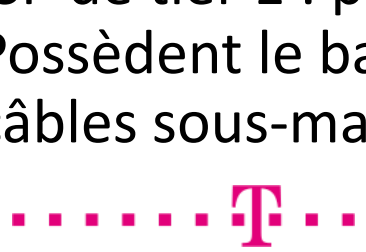


Hiérarchie des ISP

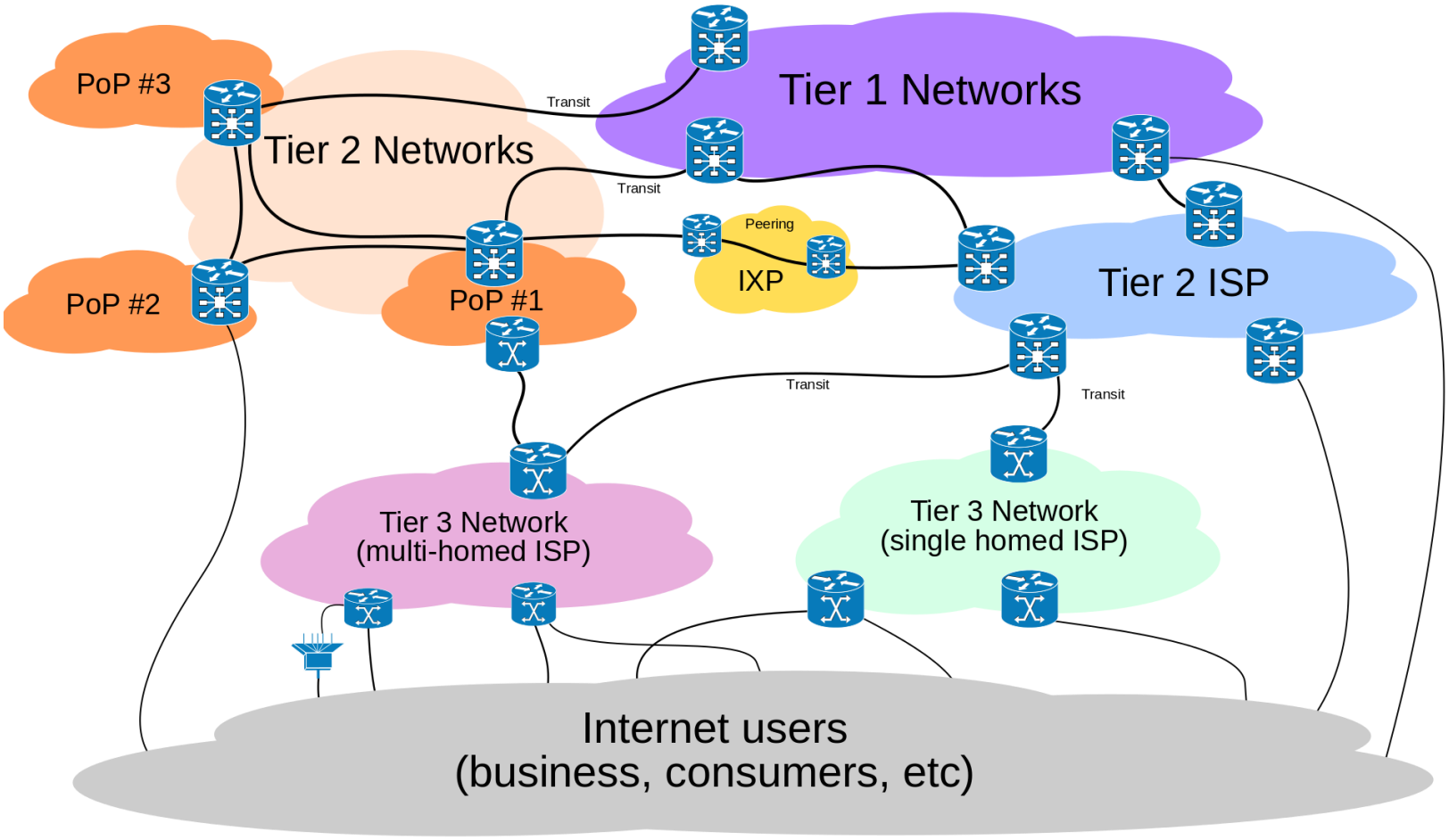
- ISP de tiers 3 : plus bas niveau. Ne fournissent pas de transit.

- ISP de tier 2 : fournissent et achètent du transit.

- ISP de tier 1 : plus haut niveau. N'achètent pas de transit. Possèdent le backbone de l'Internet (Tbps), possèdent câbles sous-marins, etc.



Hiérarchie des ISP



Rappel: Switch et Routeur

- On dit que 2 interfaces sont sur le même réseau local si :
  -
  -
- Un switch est un équipement interconnectant des interfaces appartenant à
- Un routeur est un équipement interconnectant des interfaces appartenant à

Internet eXchange Points (IXP)

- IXP : infrastructure pour l'échange de trafic entre des AS, et qui opère à la couche 1 ou 2, = point de peering

- Peut être : simple comme un switch, un immeuble avec switchs stackés et segmentés, ou une plateforme complexe comme le DE-CIX Apollon à Francfort, distribué sur 4 data centers, ou SFINX à Paris sur 2 PoPs.
- La grande majorité des IXP reposent sur du switching Ethernet.
- En France: explorer les infos sur https://www.sfinx.fr/ ou https://www.franceix.net/fr/

Figure 2: Number of IXPs per country (data from PCH).

Main building of the
London Internet Exchange (LINX)

A 19-inch rack used
for switches at the DE-CIX
in Frankfurt, Germany

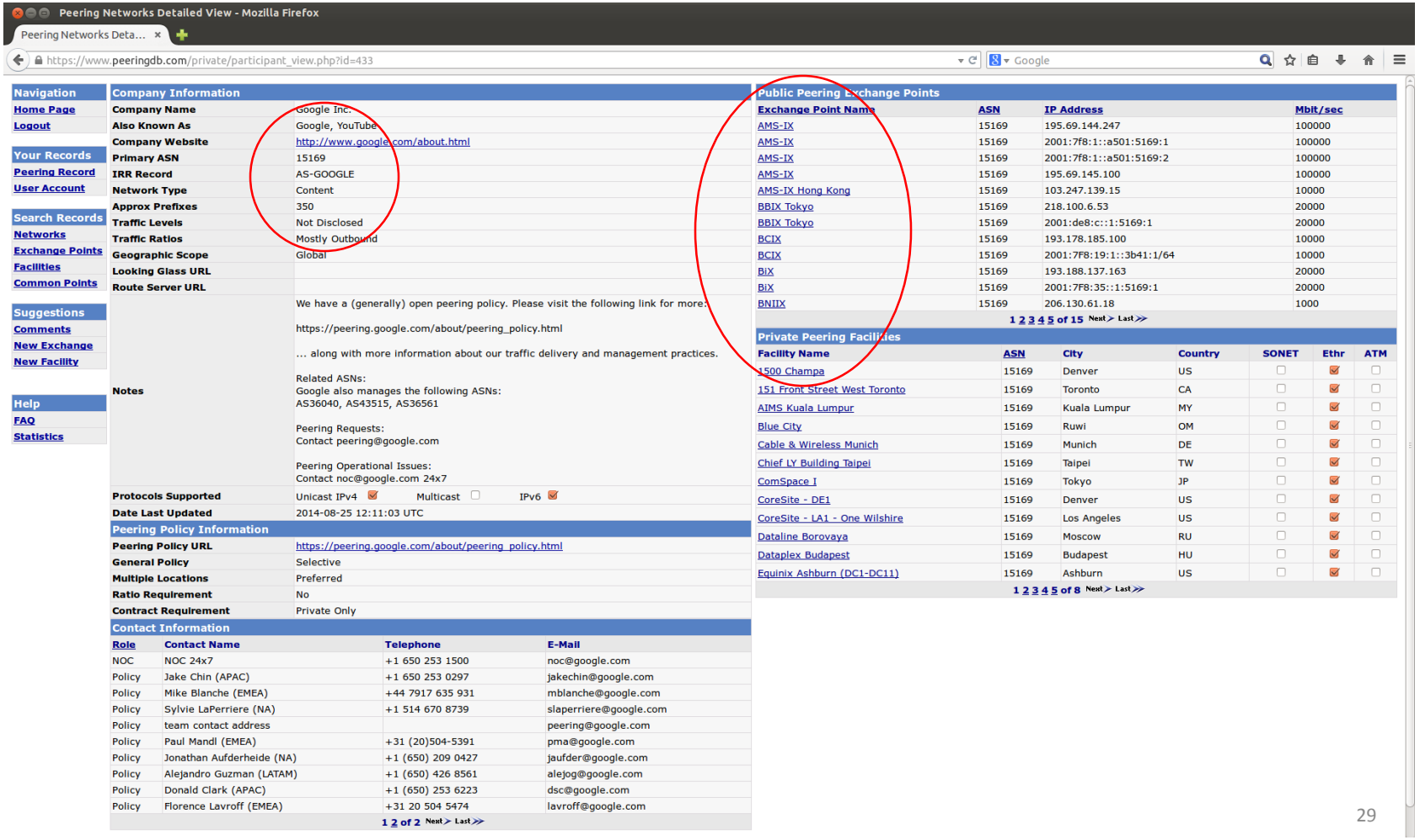
Internet eXchange Points (IXP)

- Chaque réseau participant doit :
  1. avoir un ASN public
  2. apporter un routeur dans l'IXP et connecter un de ses ports Ethernet au switch de l'IXP (et un de ses ports WAN vers le cœur de son propre réseau)
  3. le routeur doit faire tourner BGP car l'échange de routes à travers l'IXP se fait uniquement avec BGP
  4. chaque participant doit se conformer à la politique d'usage de l'IXP

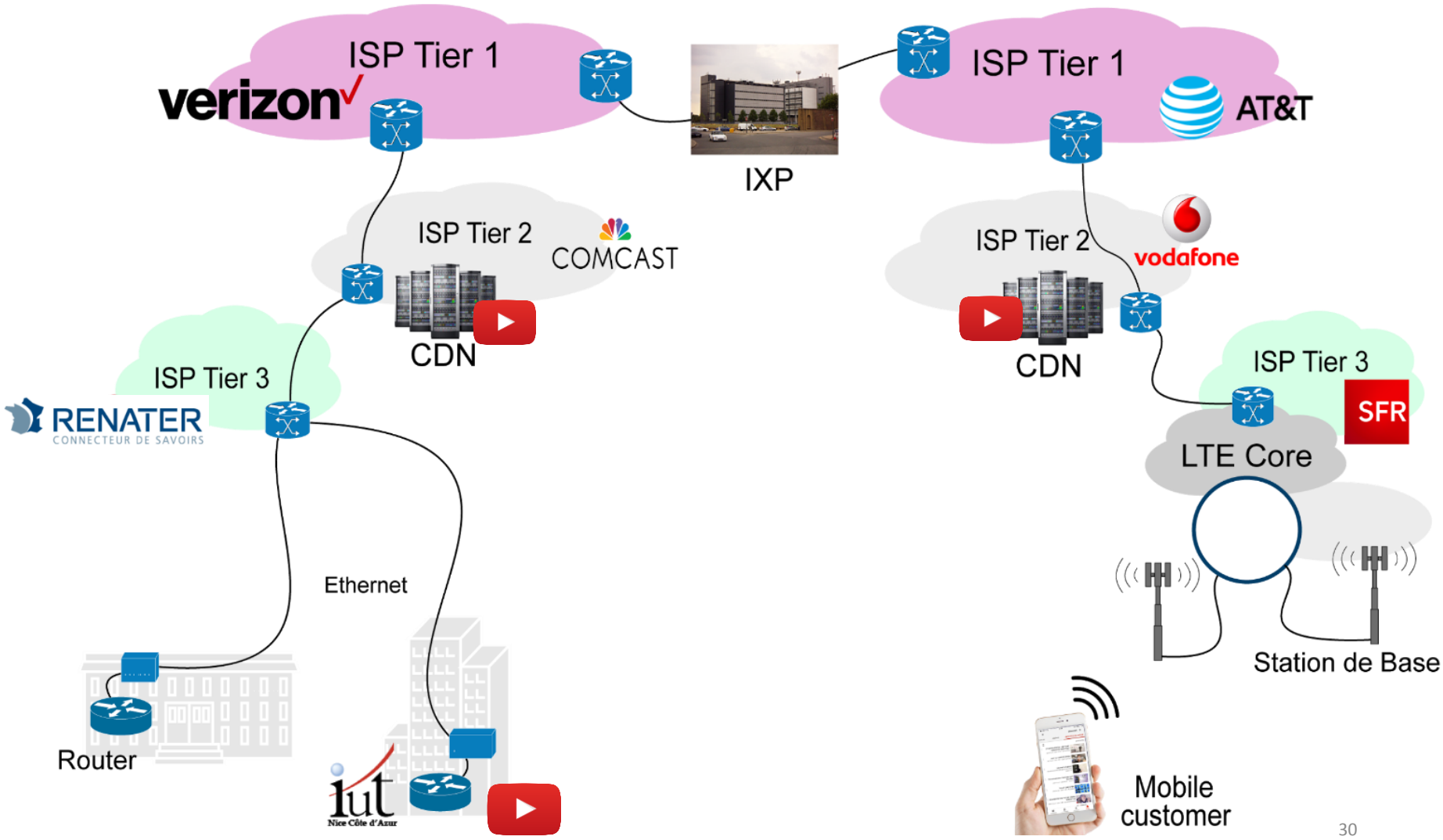
- Donc 2 réseaux qui veulent établir un peering entre eux au sein d'un IXP :
  - coût unique pour établir un circuit depuis leur point de présence
  - coût mensuel pour utiliser un des port Ethernet de l'IXP
  - possiblement une cotisation annuelle pour être membre de l'association possédant l'IXP (non-profit en Europe)
  - L'établissement d'un lien de **peering public** à un IXP n'implique en principe aucun paiement entre les 2 parties

Internet eXchange Points (IXP)

- Des acteurs majeurs tels que Google ou Netflix, incitent les autres réseaux, notamment les petits ISPs, à s'appairer directement en étant présent à beaucoup d'IXP : Google Inc. 2017: https://www.peeringdb.com/net/433



Distribution de contenu vidéo dans l'Internet



Internet eXchange Points (IXP)

- Private peering : interconnexion directe entre 2 ISPs, à travers un medium de couche 1 ou 2, qui offre une capacité dédiée.
  - Les seules entités connectées au switch sont celles en peering avec l'ISP qui opère le switch.
  - Le peering privé de Free : https://www.peeringdb.com/net/120
- Remote peering : Les IXPs ont maintenant des accords avec des réseaux de transport, pour pratiquer le "peering distant" :
  - le routeur du peer distant peut être n'importe où dans le monde, et connecté via Ethernet-over-MPLS à l'IXP
  - exemple : 20% des participants de AMS-IX ainsi connectés, en croisssance

- Free à Palo Alto: https://www.peeringdb.com/net/120
- DE-CIX: https://www.de-cix.net/locations/france/marseille
- https://www.peeringdb.com/search?q=de-cix

Source : *There is More to IXPs than Meets the Eye*, ACM SIGCOMM Computer Communication Review, Oct. 2013

A retenir

- Définitions réseaux étendus, d'opérateurs :

- Hiérarchie des ISP :

- IXP :

Plan

I. Organisation des opérateurs de l'Internet
   - I.1. Introduction aux réseaux étendus (WAN)
   - I.2. Hiérarchie et relations entre opérateurs (ISP)
   - I.3. Routage entre AS : le protocole BGP

II. TCP et Qualité de service
III. Commutation par circuits virtuels
IV. Commutation par circuits virtuels dans le mode IP : MPLS

- Sources principales
  - Cours BGP de Aiman Atta (makeiteasiest.com)
  - Cours de Nomi Beo (en ligne)

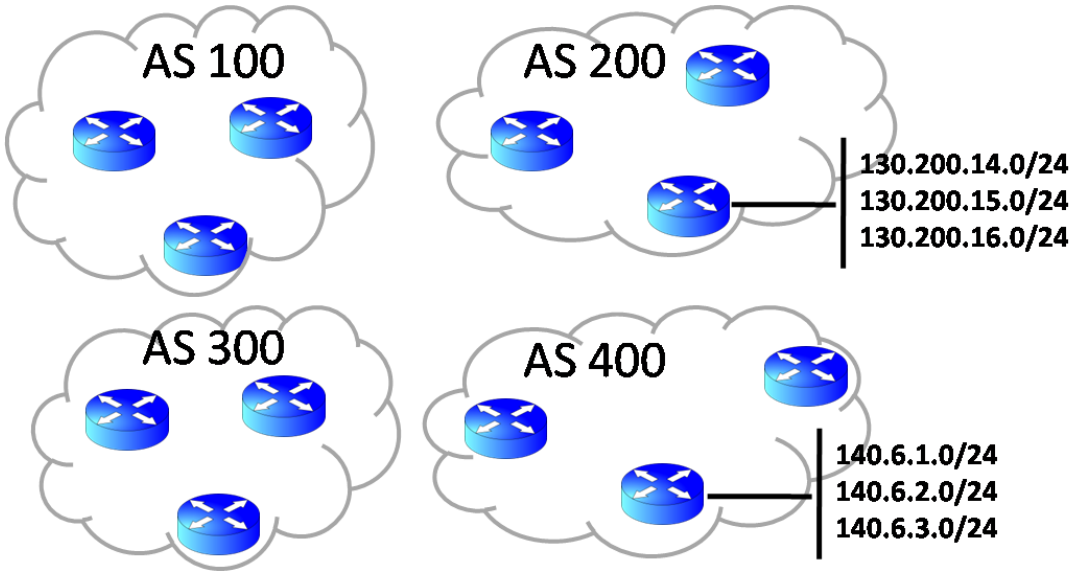
Plan

- Introduction de BGP
  - Tables BGP réelles de l'Internet
- Options de connexion à Internet : quand utiliser BGP ?
- Quantités d'updates : comment utiliser BGP ?
- BGP en action
  - Messages BGP
  - Etats du voisinage
  - Configuration
    - Configuration de voisins
    - Injection de routes dans BGP
  - Vérification et résolution des problèmes
- Attributs importants pour le choix des routes

Introduction à Border Gateway Protocol (BGP)

- ***Autonomous System :*** groupe de réseaux IP différents, connectés entre eux et qui sont gérés par une unique instance de protocole de routage, différente des AS voisins

- -> correspond en général à un ISP
- -> les annonces de réseaux entre 2 AS ne peuvent se faire qu'avec BGP



- BGP est un *Exterior Gateway Protocol* (**EGP**) : classe de protocoles de routage entre ***Autonomous Systems*** **(AS)**

Introduction à BGP

- Principe: 1 table de routage générale, 1 table BGP, entrées BGP inclues dans la table Générale sous certaines conditions

- BGP utilise le port **TCP** 179 et maintient une **relation BGP** entre **voisins** ou **pairs** (*neighbors or peers*)
  - BGP est un protocole de couche appli, pour remplir la table de routage
  - Une session est établie entre 2 voisins BGP, avec fiabilité
  - Repose sur les message *keepalive* de TCP pour maintenir la connexion ouverte

- Par défaut, pour atteindre un réseau, BGP choisit le chemin qui minimise le nombre d'AS traversés (RIP minimise le nb de routeurs traversés)

- Ce critère de décision peut être changé avec tous les **attributs de chemins** possibles (facteurs de décision pour choisir le chemin)

- BGP est crucial pour l'Internet: https://www.cnet.com/news/how-pakistan-knocked-youtube-offline-and-how-to-make-sure-it-never-happens-again/

Introduction à BGP

- Pourquoi BGP et pas juste un IGP (OSPF, etc.) ?
  - Internet c'est gros comment ? 200K préfixes dans une table de routage d'Internet
  - Est-ce qu'un IGP peut gérer une telle quantité d'updates ? NON
  - OSPF utilise son algo dès que quoique ce soit se passe dans les réseaux
- BGP est le protocole de routage le plus lent, à dessein :
  - 1 fois toutes les 30s entre peers eBGP
  - (Updates envoyées 1 fois toutes les 5s entre peers iBGP)
  - Donc si une route tombe, alors BGP pourrait prendre plusieurs heures avant d'en informer tous les routeurs de l'Internet
  - Imaginons que dès qu'un réseau tombe, ça déclenche une update -> beaucoup de BW consommée par ces updates ET beaucoup de réseau tombent et remontent par seconde sur l'Internet -> on ne peut pas se permettre de reporter à tout le monde un changement d'état temporaire
  - BGP a bcp d'outils pour gérer les problèmes liés à un très grand nombre de réseaux.

Introduction à BGP

- Buts de BGP: scalabilité et fiabilité
- Pas de mécanisme de sécurité: considérés non-nécessaires en 1989, et aujourd'hui considérés rajouter trop de surcharge
- Les routeurs BGP:
  - ne peuvent pas vérifier qu'un AS est autorisé à annoncé un certain préfixe
  - ne peuvent pas vérifier l'identité d'un AS distant
  - ne peuvent pas vérifier les routes et/ou les attributs

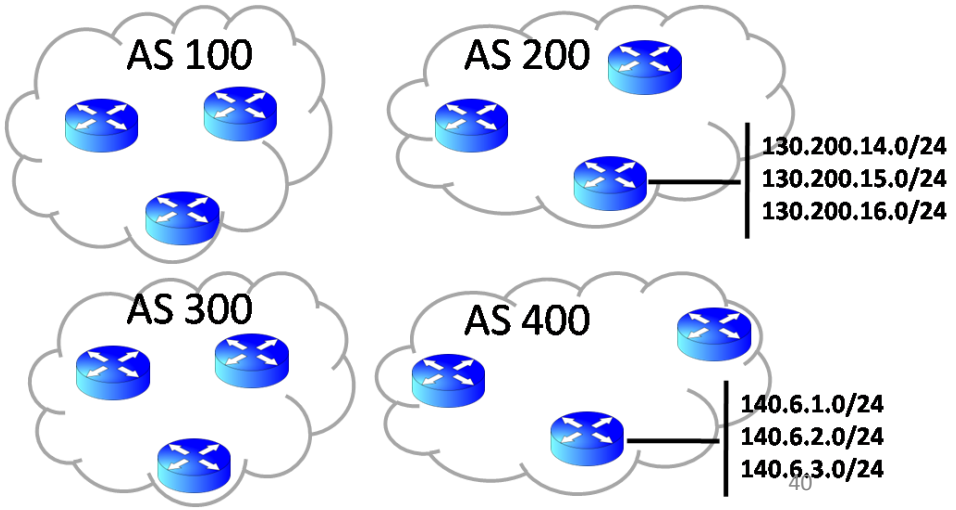
-> implique une confiance mutuelle implicite.

- Principe: 1 table de routage générale, 1 table BGP, entrées BGP inclues dans la table Générale sous certaines conditions

Introduction à BGP

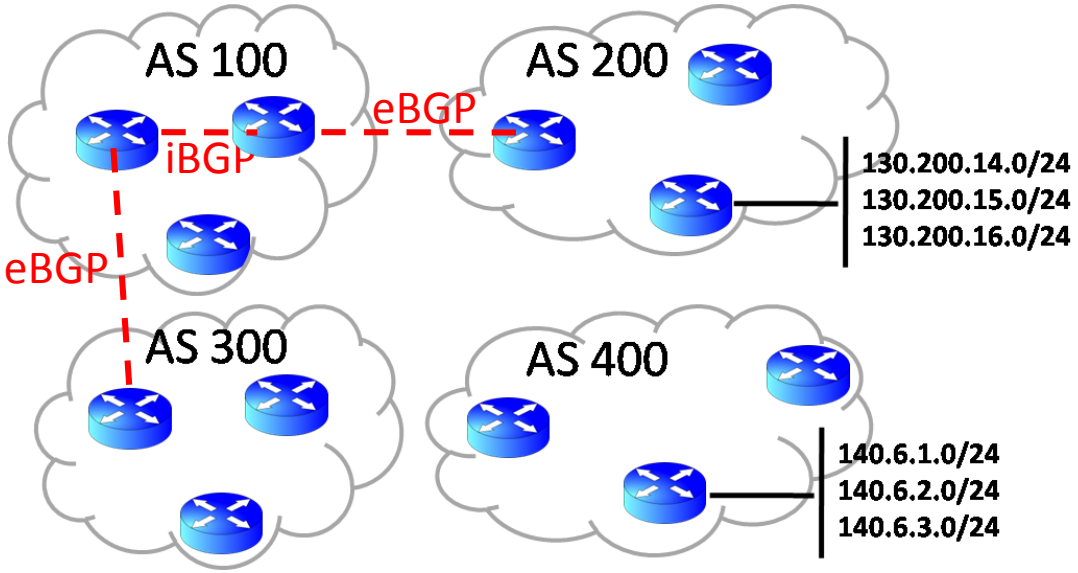
- ISPs utilisent BGPv4 pour échanger MaJ de routage sur l'Internet
- *Interior Gateway Protocol* (IGP) : RIP, OSPF, IS-IS, EIGRP
- Besoin de configurer les voisins explicitement (statiques) – pas de découverte automatique
- BGP : pas de concept de multipaths et load balancing (contrairement aux IGPs)
- Même esprit que distance vector, mais la distance est en terme de nombre d'AS traversés, pas de routeurs

- Numéros d'AS : 32 bits
  - Public : 1 – 64495
  - Private : 64512 - 65534



Introduction à BGP

- 2 versions de BGP : iBGP et eBGP
  - iBGP (internal) : BGP entre 2 routeurs BGP du même AS
  - eBGP (external) : BGP entre 2 routeurs de 2 AS différents



Le problème : faire fonctionner Internet

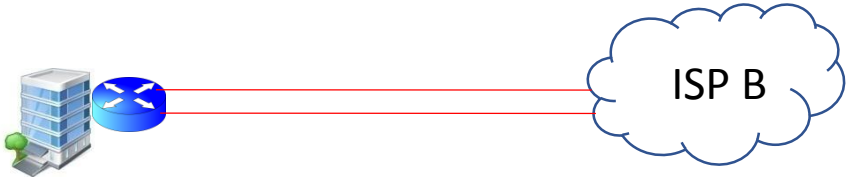
- Sur Internet : @IP publiques
- Attribuées à des entités, dans différents AS
- Visibles à: http://whois.domaintools.com
- Ces caractéristiques de votre connexion visible à : https://bgp.he.net/

Quand utiliser BGP ?

En tant que client, utiliser BGP que dans des cas bien particuliers :

- 1 ISP et 1 cx : une default route suffit

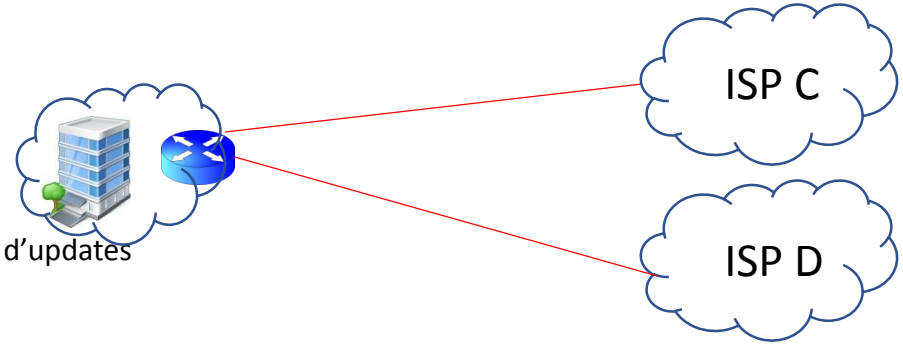
- 1 ISP et plus d'1 cx : cx redondantes. Load sharing ou backup cx



- Plus d'1 ISP : cie pour qui l'accès permanent est critique. Si les 2 cx sont actives (pas une en backup uniquement) : ensemble de routes différents des 2 ISPs. Le client peut optimiser le choix de route vers la destination.

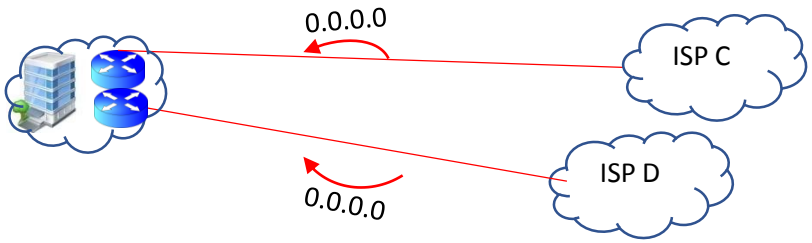
BGP intéressant

- Attention à la quantité d'updates

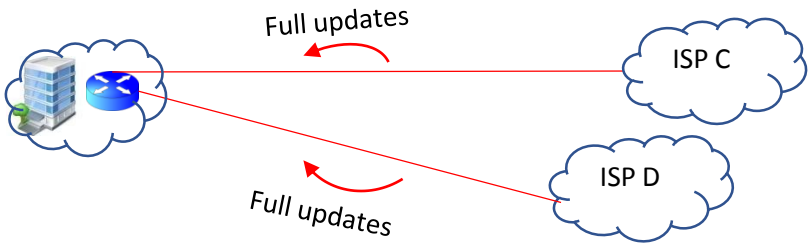


Quantité d'updates avec BGP

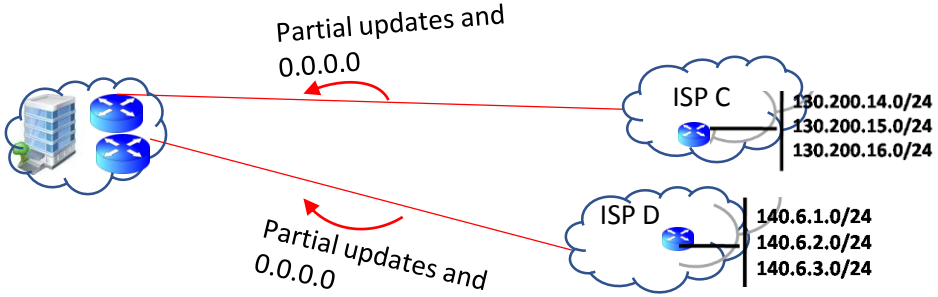
- Default route : on peut utiliser BGP et ne recevoir que la default route (mais on publie nos réseaux) – pas de mémoire supplémentaire requise aux CE



- Toutes les updates



- Compromis : seulement certaines routes annoncées au client par ses ISPs



BGP messages

- **Open** : session TCP entre 2 routeurs, échange de ASN et MD5 (authentification)
- **Keepalive** : comme « Hello » dans OSPF pour garder la session ouverte entre 2 voisins BGP
- **Update** : échange de préfixe et attributs
- **Notification** : notifie erreur puis la relation entre 2 peers est ré-initalisée

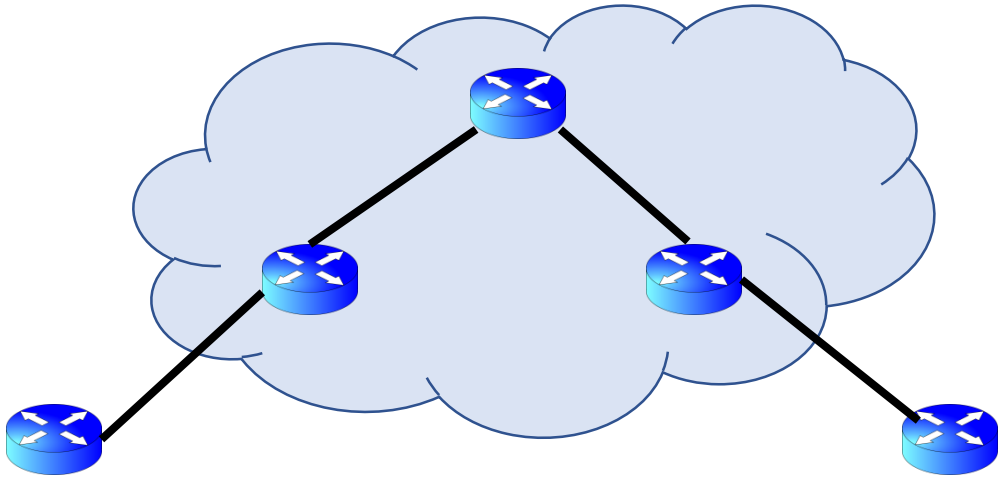
Etats des voisins

1. **Idle** : il y a un problème (comme Active)
- **Active** : cx TCP tentée, et pas établie
2. **Opensent** : cx TCP ok, mais pas de message BGP reçu
3. **Openconfirm** : Open message a été envoyé et reçu
4. **Established** : la relation entre les 2 routeurs fonctionne et ils peuvent s'échanger des messages d'updates

- Authentification MD5 : encrypte les messages BGP envoyés entre routeurs ; compromettre BGP est un des plus gros risque encouru par un ISP, car ça peut lui faire perdre toute la connectivité de son réseau avec les autres.
- Pas généralisée et ne résout pas les problème d'annonce de routes fausses.

Règle 1 de BGP : règle de synchronisation

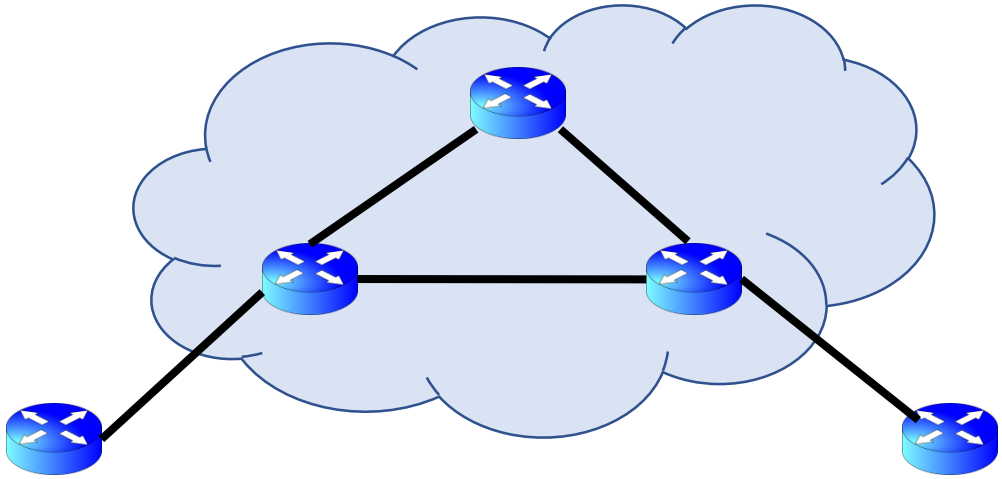
- Les routes apprises via IBGP doivent être également apprises par un IGP (OSPF,…) avant d'être annoncées à d'autres peers d'AS différents.



- La synchronisation peut être désactivée que si les routeurs IBGP sont directement connectés.

Règle 2 de BGP : Split-horizon

- En IGP : route apprise par une interface jamais renvoyée sur cette même interface.
- En BGP : route apprise par IBGP jamais envoyée à un autre voisin IBGP



- Car pas de mécanisme pour détecter des boucles d'updates en IBGP.
- Soit on suppose que tous les voisins IBGP sont fully-meshed.
- Si ça n'est pas le cas car il y a trop de voisins IBGP : route-reflector.

Catégories d'attributs de chemin (PA)

- BGP n'a pas de métrique unique (pour déterminer le meilleur chemin d'AS vers un réseau), mais un ensemble de PA hiérarchisés
- 2 catégories de PA :

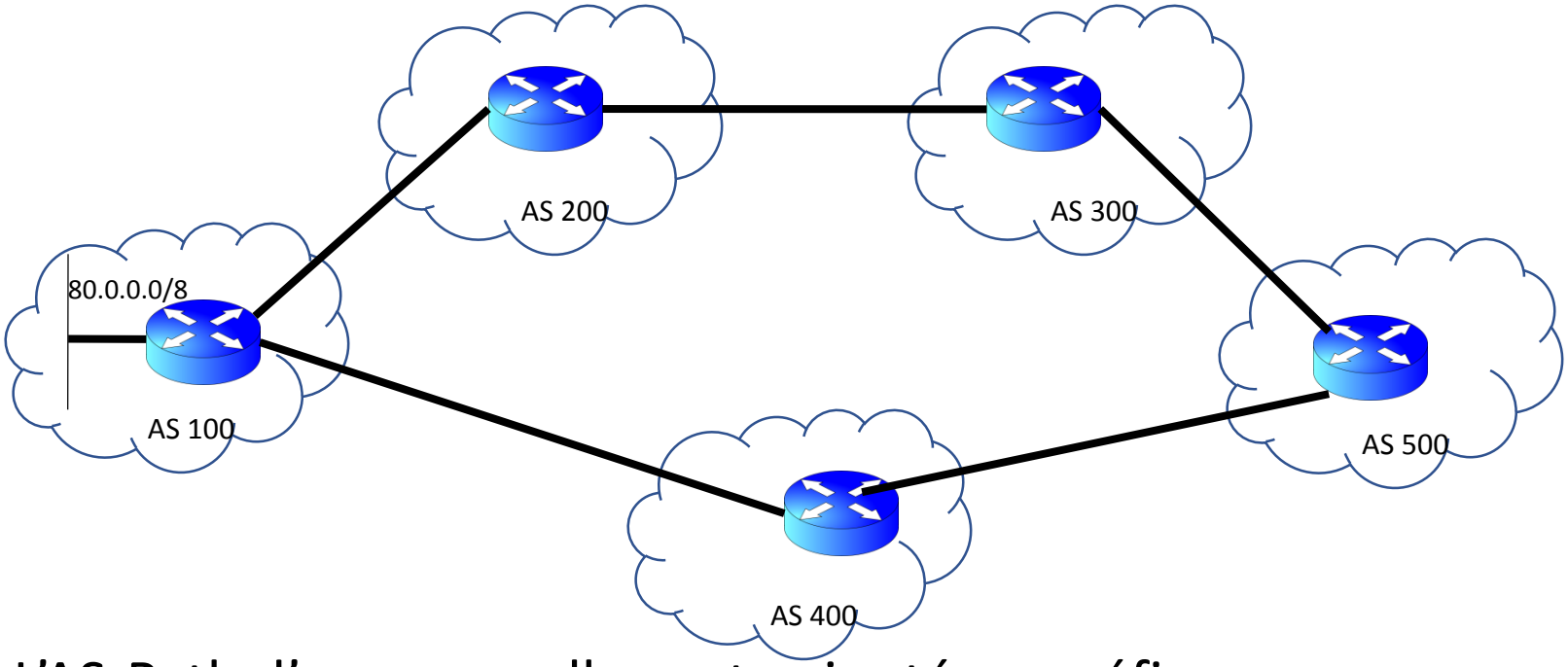
```
R11#sh ip bgp
BGP table version is 842, local router ID is 10.0.114.11
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
* i10.0.12.0/24    10.0.113.1               0    100      0 ?
*>                 10.0.112.12          35840             32768 ?
*> 10.0.23.0/24    10.0.112.12          58880             32768 ?
*                  10.0.114.4                             0 400 300 200 ?
* i10.0.112.0/24   10.0.113.13          33280    100      0 ?
*>                 0.0.0.0                  0             32768 ?
* i10.0.113.0/24   10.0.113.1               0    100      0 ?
*>                 10.0.112.12          33280             32768 ?
* i10.0.114.0/24   10.0.113.13          35840    100      0 ?
*>                 0.0.0.0                  0             32768 ?
* i10.0.123.0/24   10.0.113.13          30720    100      0 ?
*>                 10.0.112.12          30720             32768 ?
*> 10.0.221.0/24   10.0.112.12          58880             32768 ?
*                  10.0.114.4                             0 400 300 200 ?
```

Supportés	Optionnels
Origin (oblig)	Aggregator
AS-Path (oblig)	Community
Next-Hop (oblig)	Multi-Exit Discriminator
Local-Preference	
Atomic Aggregate	

Attribut AS-Path

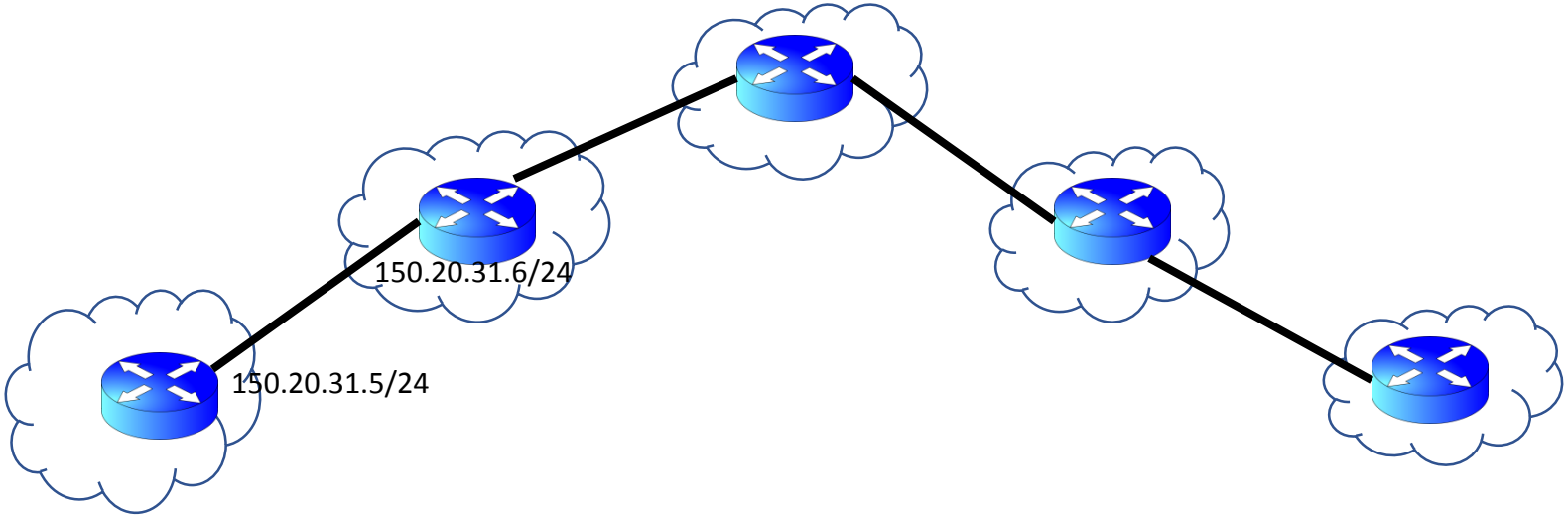
- C'est la suite des AS traversés
- Obligatoire dans chaque update: chaque routeur annonçant un réseau inclut la liste des AS à traverser dans l'update.



- L'AS-Path d'une nouvelle route ajouté en préfixe
- On ajoute l'ASN à chaque passage de frontière d'AS
- Utilisé pour la prévention de boucle

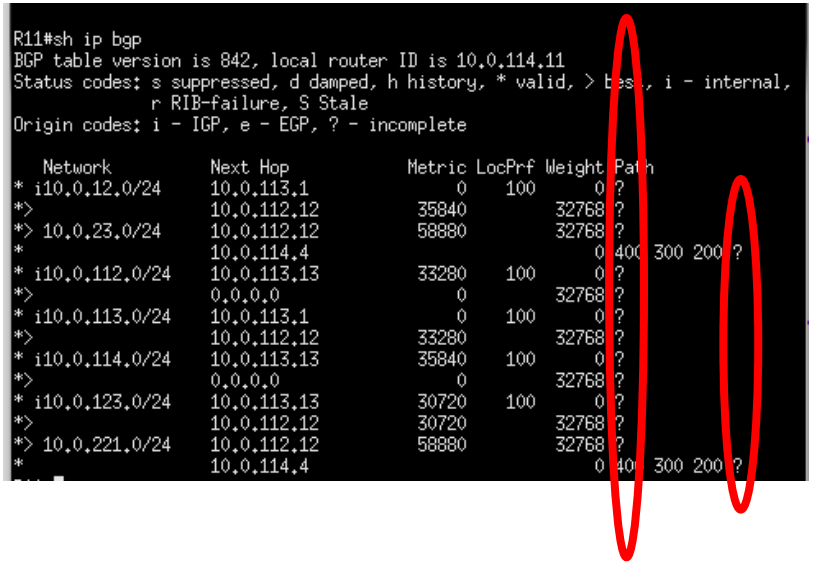
Attribut Next-Hop

- Obligatoire dans chaque update évidemment
- En général l'@IP du routeur dont on reçoit l'update



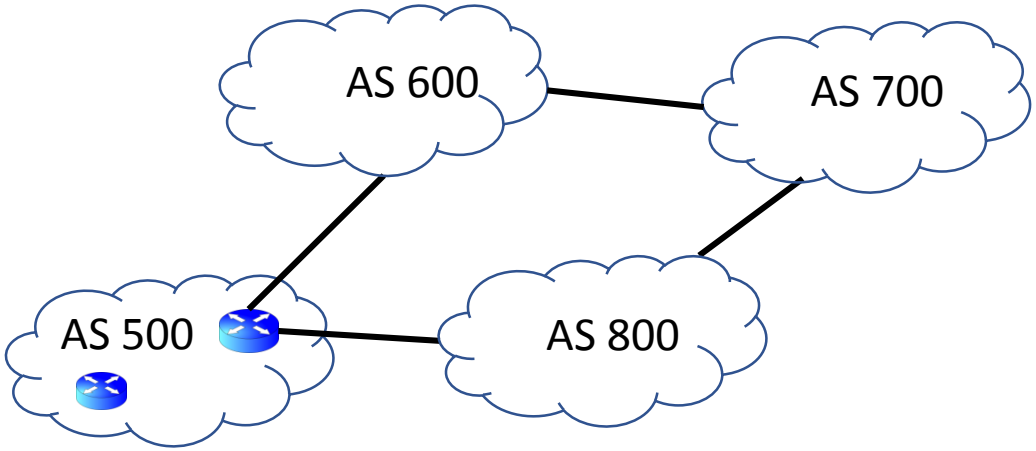
Attribut Origin

- Comment la route a été introduite dans BGP
- Les valeurs possibles sont :
  - IGP (i) : annonce déclenchée par la commande network
  - EGP (e) : (Exterior Gateway Protocol) vieux BGP, plus du tout utilisé
  - Incomplete (?) : ce réseau installé dans BGP par re-distribution (automatique, par exemple de EIGRP dans BGP, pas par la commande network)



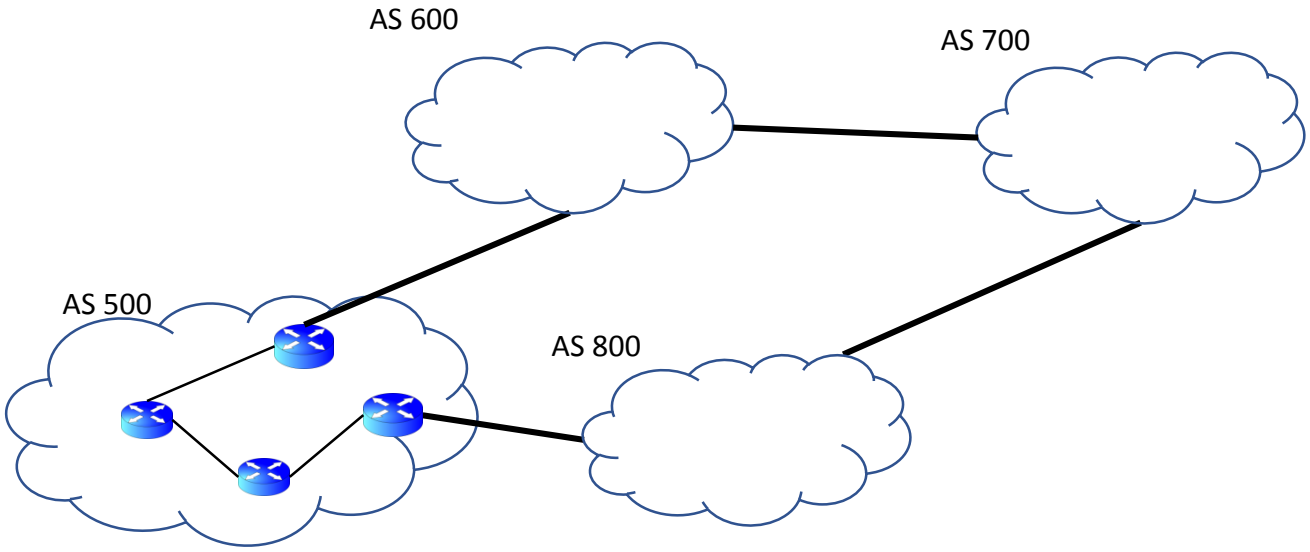
Attribut Weight

- Propriétaire Cisco
- Dit comment sortir de l'AS, en donnant plus de poids à routes reçues d'un certain voisin BGP
- attribut local -> pas annoncé aux autres routeurs
- Attribué à un neighbor, pour préferer ses routes



Attribut Local Preference

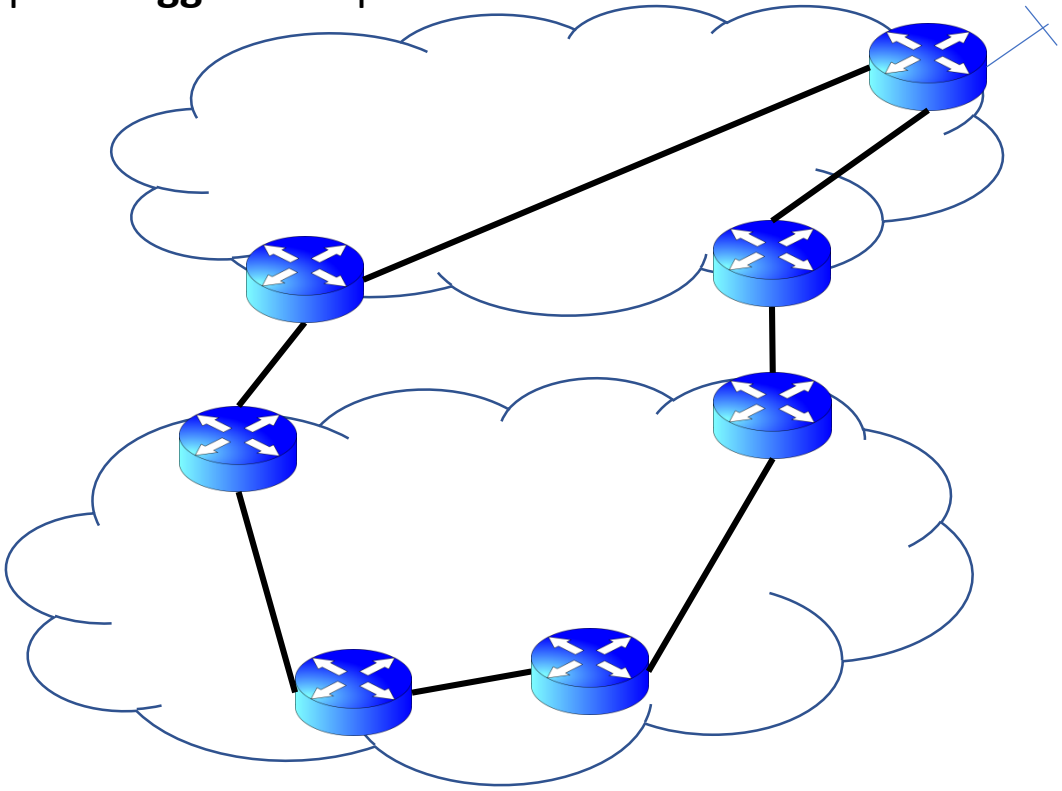
- Permet d'indiquer par quelle route on veut sortir de l'AS
- Annoncé au sein de l'AS (avec iBGP)
- C'est la LP la plus haute qui gagne.



Attribut Multi-Exit Discriminator (MED)

- Principe de routage au sein d'un AS: faire sortir le trafic le plus vite = *Hot potato routing* car
  - Entre 2 chemins IBGP, préfère celui vers voisin le plus proche selon IGP
- Or, règle d'or de BGP : un AS ne peut jamais dire à un autre AS comment router son trafic

→ MED est là pour **suggérer** un point d'entrée dans votre AS



Attribut Community

- Pour filtrer les routes à re-distribuer dans/en dehors de notre AS

Critères de sélection de la meilleure route par BGP

- Exclut routes avec Next-hop inaccessible
- Plus grand Weight (local au routeur)
- Plus grande Local Preference (globale dans l'AS)
- Routes introduites dans BGP par le routeur lui-même
- Plus court AS-path (nombre d'AS à traverser)
- Plus petit Origin code : IGP<EGP<Incomplete
- Plus petit MED
- Routes apprises par EBGP à celle apprises par IBGP
- Entre 2 chemins IBGP, préfère celui vers voisin le plus proche selon IGP
- Entre 2 chemins EBGP, préfère le plus vieux (stable)
- Chemin depuis routeur avec plus petit ID

Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

- II.1.a. Principe du contrôle de congestion
- II.1.b. Rappels : transfert fiable et contrôle de flux
- II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

- II.2.a. Définition de la QoS et classes d'applications
- II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS
- II.2.c. Stratégies possible pour garantir un niveau de QoS

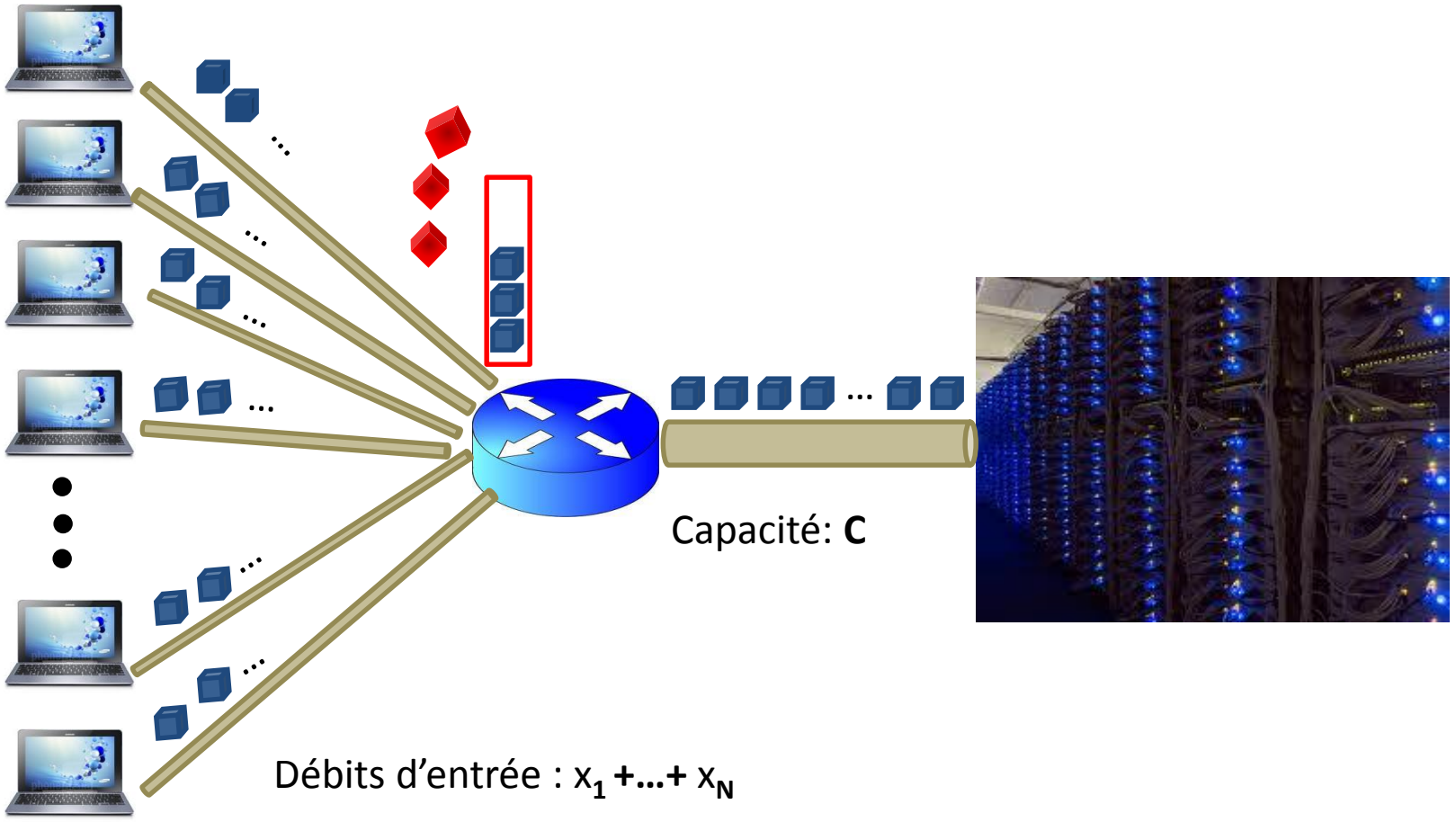
II.3. Techniques de traitement de la QoS

- II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing
- II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets
- II.3.c. Gestion de buffer : comment abandoner les paquets en excès
- II.3.d. Marquage du trafic : DiffServ dans IP

III. Commutation par circuits virtuels

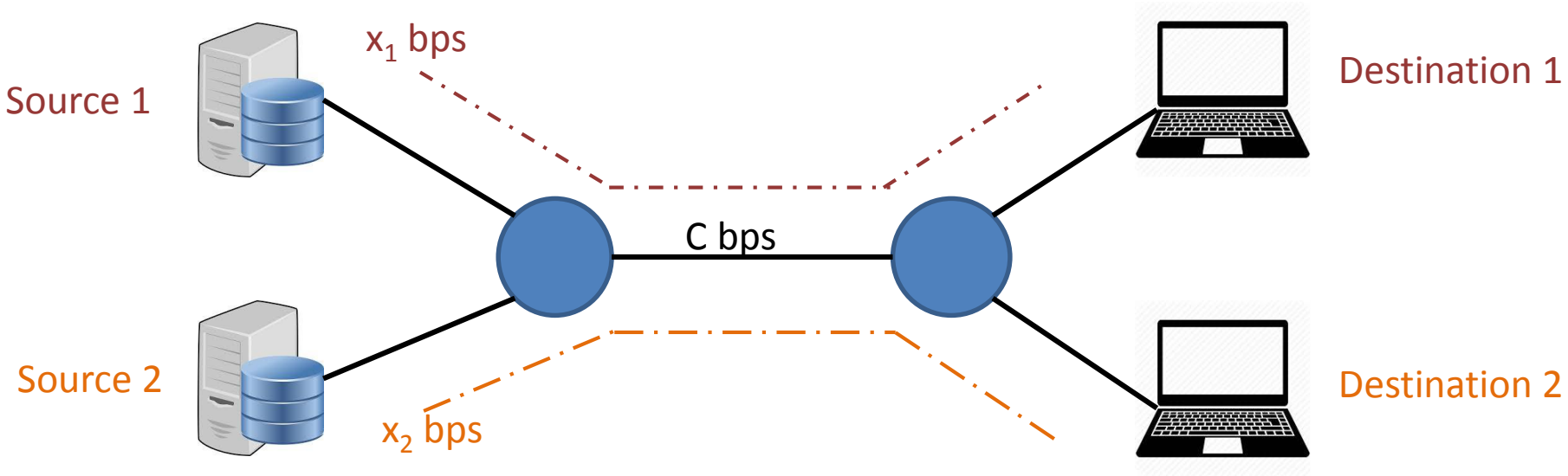
IV. Commutation par circuits virtuels dans le mode IP : MPLS

Le buffer d'un routeur



Problème de congestion quand $\sum_{i=1}^{N} x_i > C$

Comment partager la capacité ? Le contrôle de débit



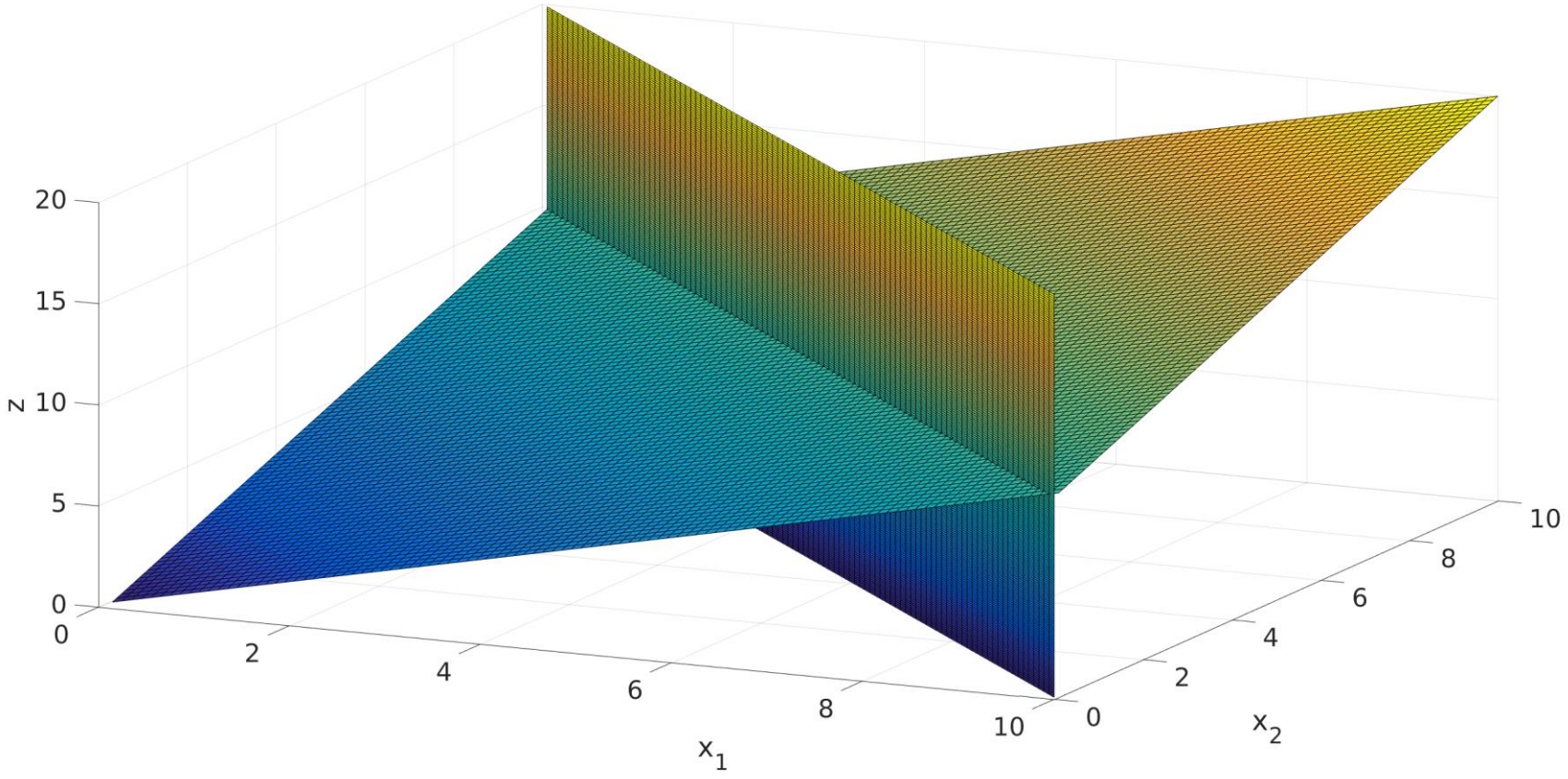
- Trouver x_1 et x_2 solutions au problème

$$\max_{x_1,x_2} x_1+x_2$$

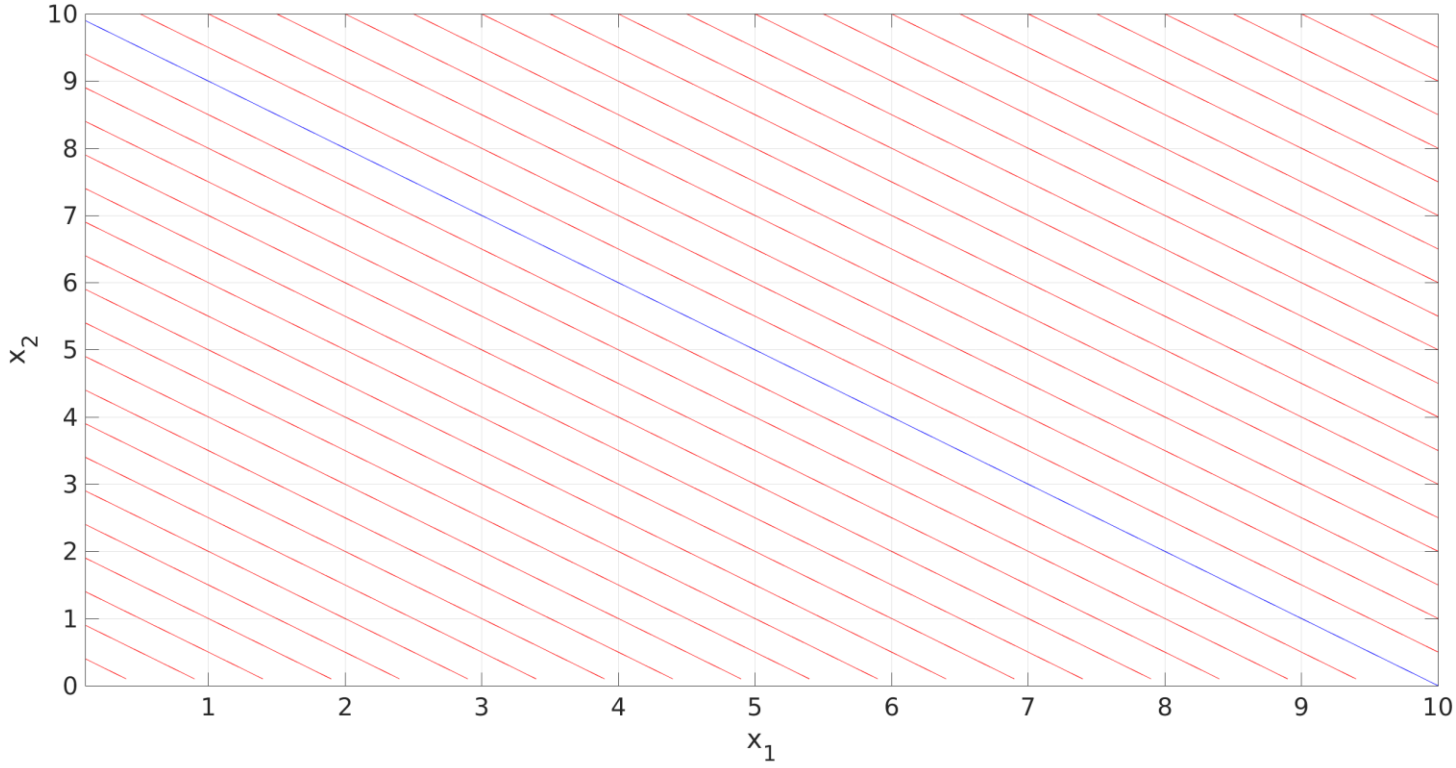
Tel que

$$x_1+x_2 = C$$

Comment partager la capacité ? Le contrôle de débit

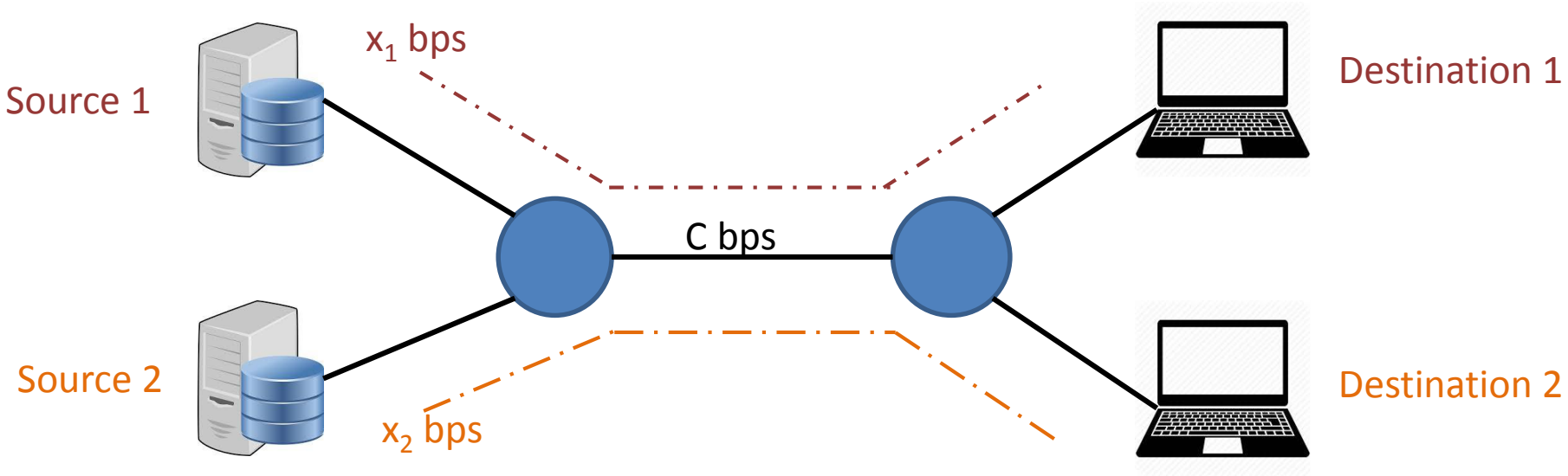


Comment partager la capacité ? Le contrôle de débit



-> pas de contrainte d'équité entre les sessions

Comment partager la capacité ? Le contrôle de débit



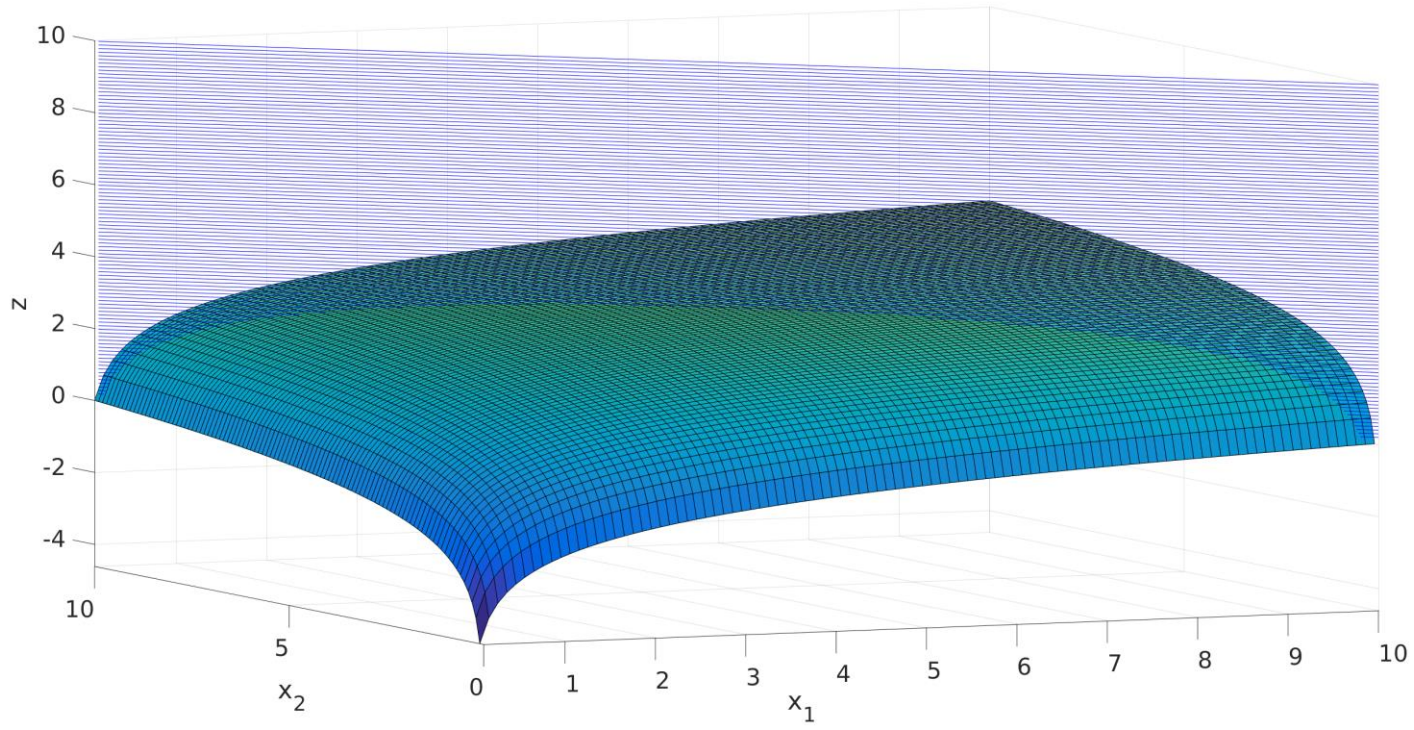
- Equité:

$$\max_{x_1,x_2} \log(x_1)+\log(x_2)$$

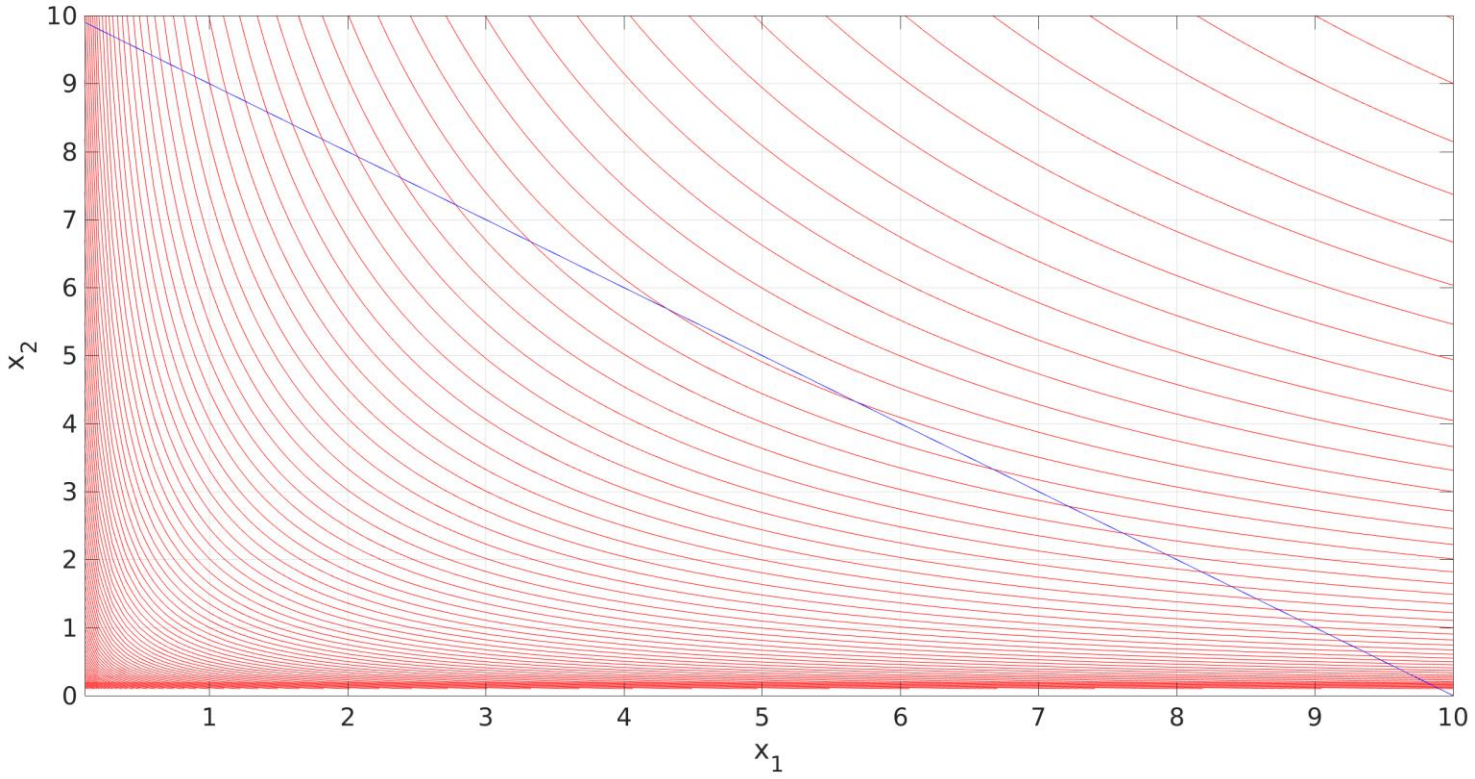
Tel que

$$x_1+x_2 = C$$

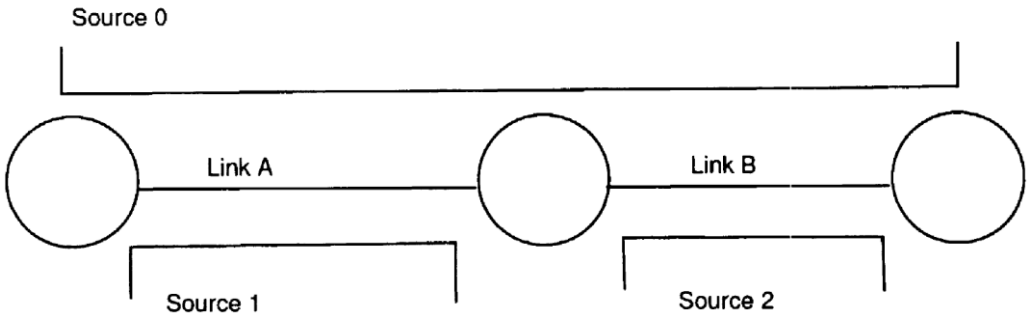
Comment partager la capacité ? Le contrôle de débit



Comment partager la capacité ? Le contrôle de débit



Cas général



$$\max_{\{x_r\} \in \mathcal{S}} \sum_r U_r(x_r)$$

Tel que :

$$\sum_{r:l \in r} x_r \leq c_l, \qquad l \in \mathcal{L},$$

$$x_r \geq 0, \qquad r \in \mathcal{S},$$

Vos applications Web: besoin de TCP

Question : quel débit doit-on utiliser pour les transferts sur Internet ?

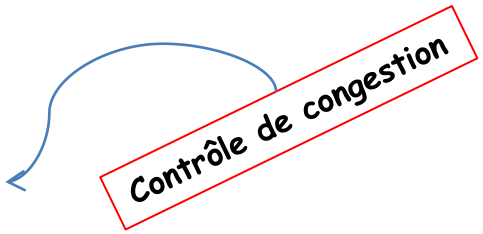
- Nécessité d'un contrôle de l'utilisation des ressources du réseau
- Pour éviter les problèmes de stabilité
- Le contrôle doit être distribué

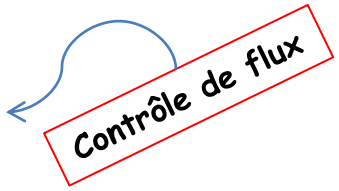
--> TCP : 90% des bytes sur Internet, 80% des paquets

Optimisation: le cas de TCP

- TCP cherche à résoudre, sur chaque machine:
- Trouver débit source maximal x_i^{opt}

- Tel que:
  - stabilité du réseau
  - équité entre les sessions
  - non-saturation du buffer de réception





Le problème de la congestion dans le réseau

- La congestion apparaît quand des flows demandent plus de ressources que les liens et équipements du réseau ne peuvent fournir:

  Sending rate > capacity of the bottleneck

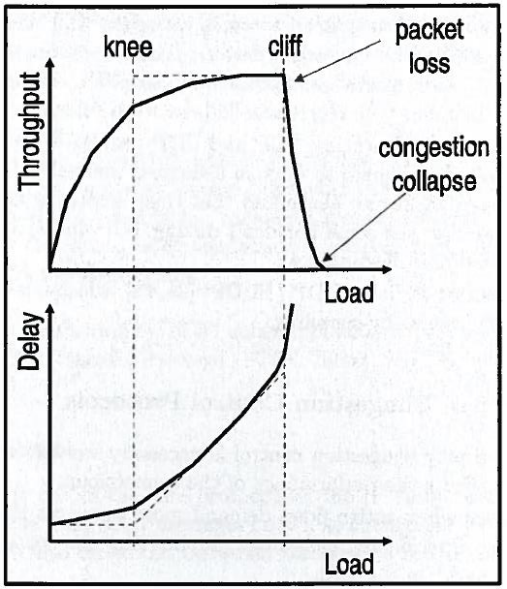
-> D'abord mise en file d'attente des paquets

-> Puis perte si le débit ne décroit pas

=> Le contrôle de congestion permet d'éviter le *congestion collapse*: effondrement du débit causé par un grand nombre d'abandons de paquets

Importance des protocoles de contrôle de congestion

- Reprise sur erreur sans contrôle de congestion: TCP sans CC avec correction d'erreur:
- Plusieurs sessions (flows) TCP
- Le débit total aggrégé est variable, autour de la moyenne *load*
- Plus la moyenne *load* augmente, plus souvent on dépasse C
- --> la mise en file d'attente commence
- ---> quand le buffer est complètement plein, des paquets sont perdus



- TCP sans CC ne diminue pas le débit qu'il met en entrée du réseau, il se contente de faire de la correction: au bout d'un timeout, supposé > temps de congestion, on retransmet
- La durée de la congestion augmente avec le nombre de sessions
- si le timeout est trop court, le débit nominal d'entrée ne diminue pas, on ne fait que re-transmettre, le débit utile de sortie s'effondre

Buts du contrôle de congestion

- Problème : moduler le débit appliqué en entrée du réseau par le protocole de transport
- 2 buts principaux de TCP avec CC:
  - Maximiser l'utilisation de la capacité des liens tout en évitant la congestion dans le réseau (i.e., maintenir la charge en dessous du *knee*)
  - Résoudre rapidement les situations de congestion pour éviter le congestion collapse (i.e., éviter impérativement le *cliff*)

- ET partager équitablement les ressources entre tous les utilisateurs
- ET transport fiable: basé ACKs

Approches pour le contrôle de congestion

2 grandes classes d'approches du contrôle de congestion:

Contrôle de congestion de bout-en-bout (*end-to-end*):

- pas de retour explicite du réseau
- congestion inferrée à partir des pertes et du délai vu par l'hôte d'extrêmité (*end host*)
- C'est l'approche prise par TCP

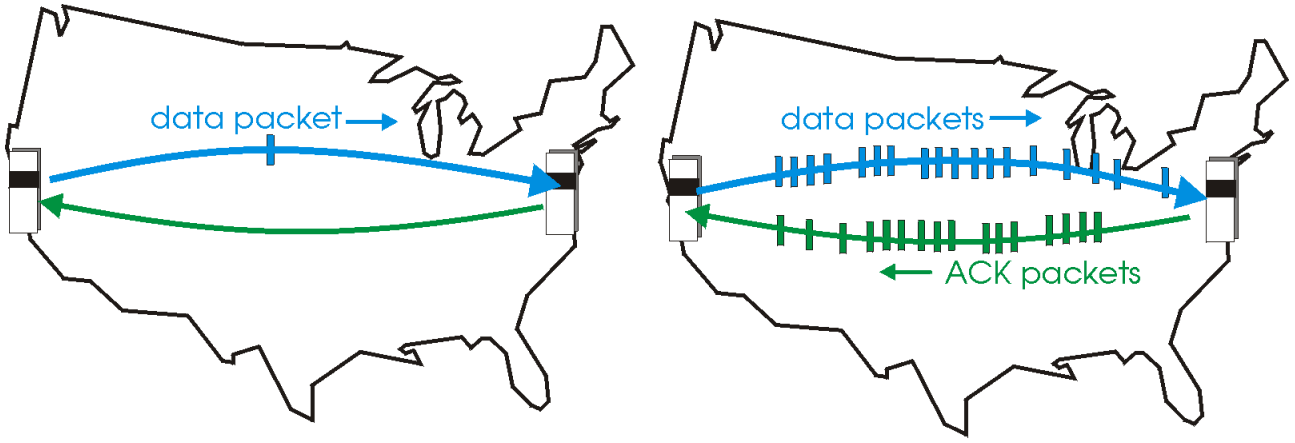
Contrôle de congestion assisté par le réseau:

- les routeurs fournissent un retour aux end hosts
  - Un bit indiquant la congestion (FR, ATM, TCP ECN)
  - Routeurs ERN

La classe de protocoles à laquelle TCP appartient: les protocoles en pipeline

Pipelining: l'émetteur envoie plusieurs paquets sans attendre l'ACK du premier

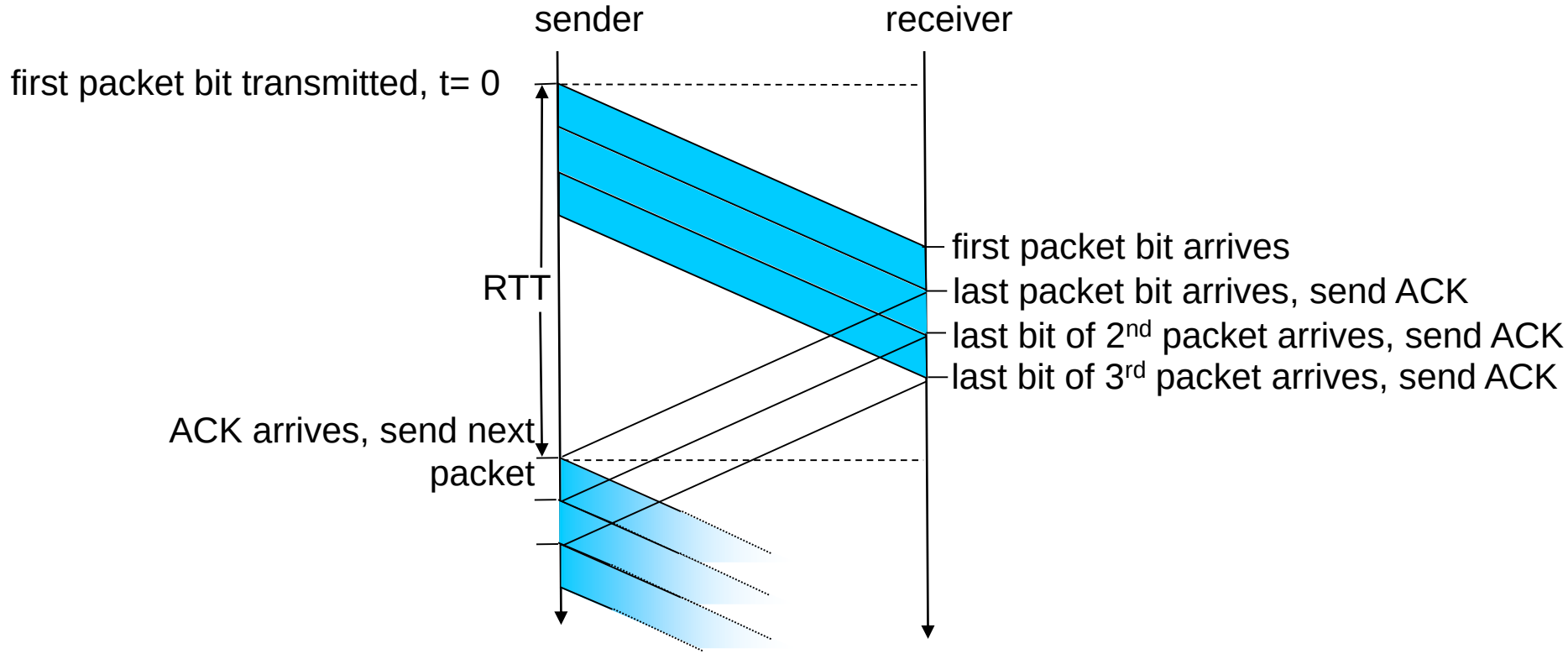
- La plage de numéros de séquence doit être augmentée
- Mise en mémoire à l'émetteur et au récepteur



(a) a stop-and-wait protocol in operation

(b) a pipelined protocol in operation

Pipelining: augmente l'utilisation



$$\text{Utilisation} = \frac{\text{débit utilisé}}{\text{débit dispo}} = \frac{\frac{NL}{RTT}}{D}$$

Increase utilization by a factor of N=3!

Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

- II.1.a. Principe du contrôle de congestion
- II.1.b. Rappels : transfert fiable et contrôle de flux
- II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

- II.2.a. Définition de la QoS et classes d'applications
- II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS
- II.2.c. Stratégies possible pour garantir un niveau de QoS

II.3. Techniques de traitement de la QoS

- II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing
- II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets
- II.3.c. Gestion de buffer : comment abandoner les paquets en excès
- II.3.d. Marquage du trafic : DiffServ dans IP

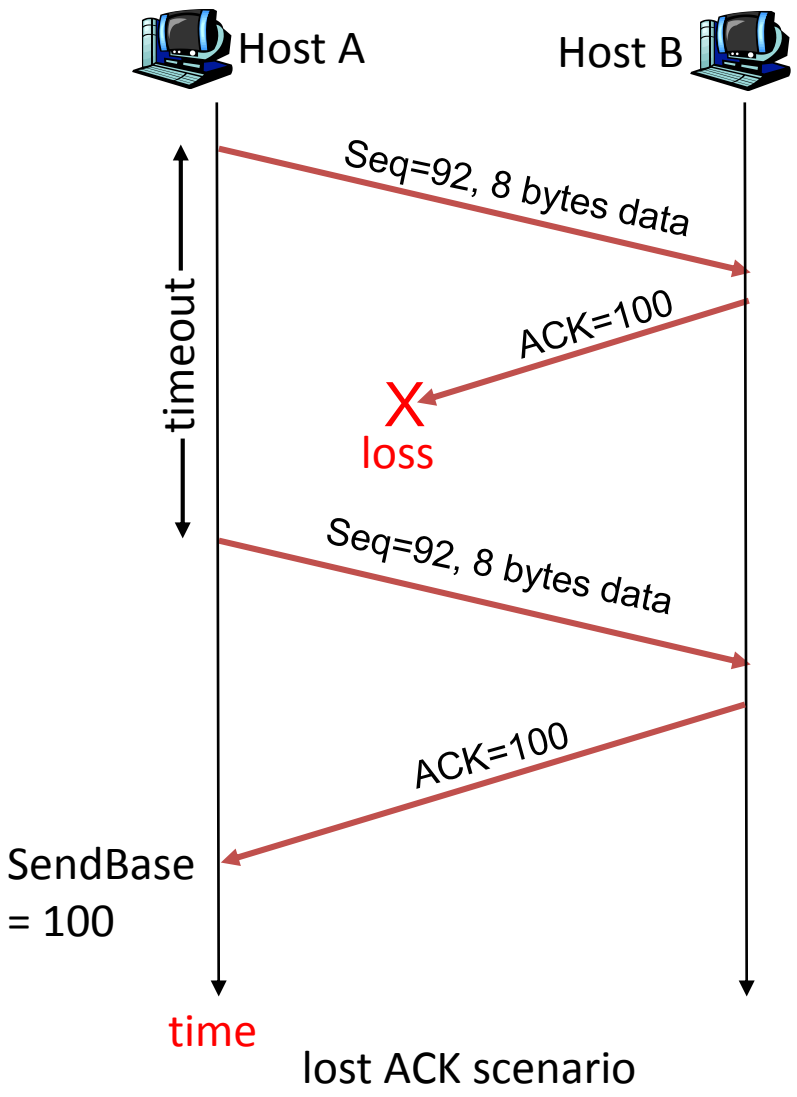
III. Commutation par circuits virtuels

IV. Commutation par circuits virtuels dans le mode IP : MPLS

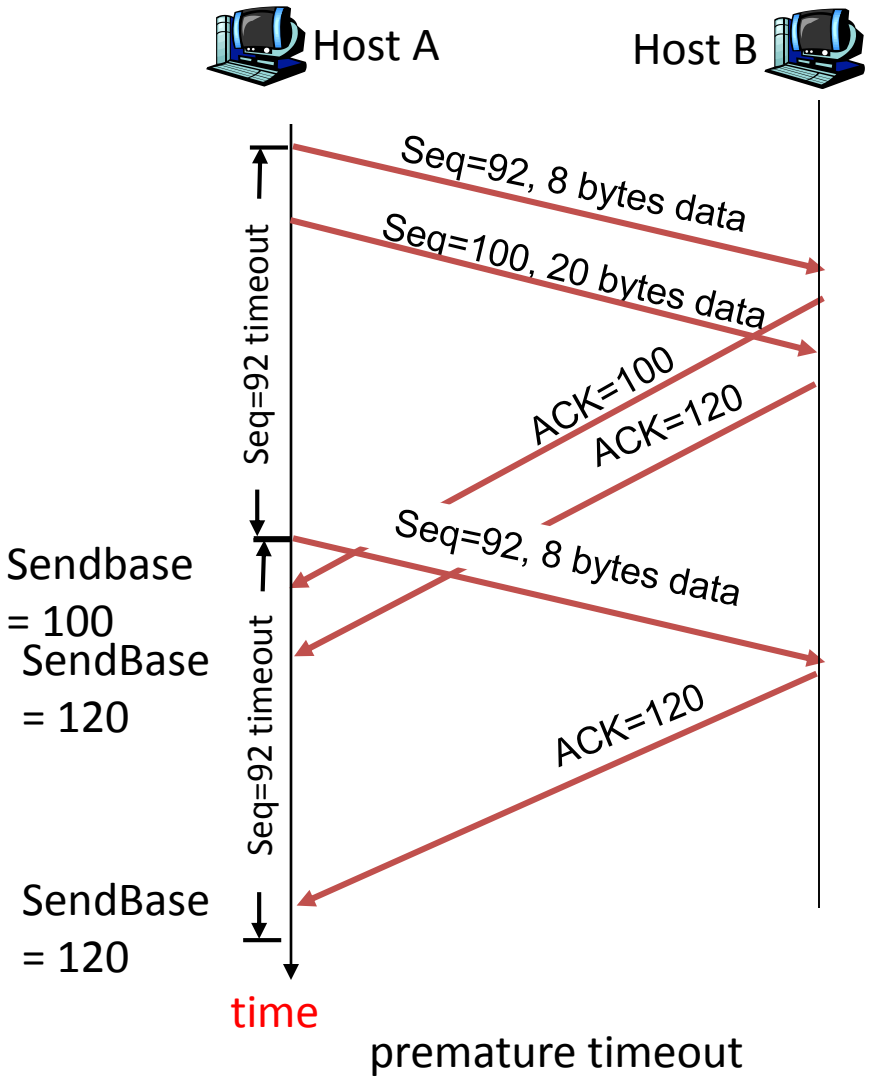
Transfert fiable TCP

- TCP crée un service de tf au dessus du service non-fiable de IP. Fiabilité= _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
- => retransmettre si paquets perdu
- => comment inférer que paquet perdu?
- Pertes détectées et retransmissions déclenchées par :
  - Evénement de timeout
  - Acks dupliqués

Scenarios de retransmission TCP



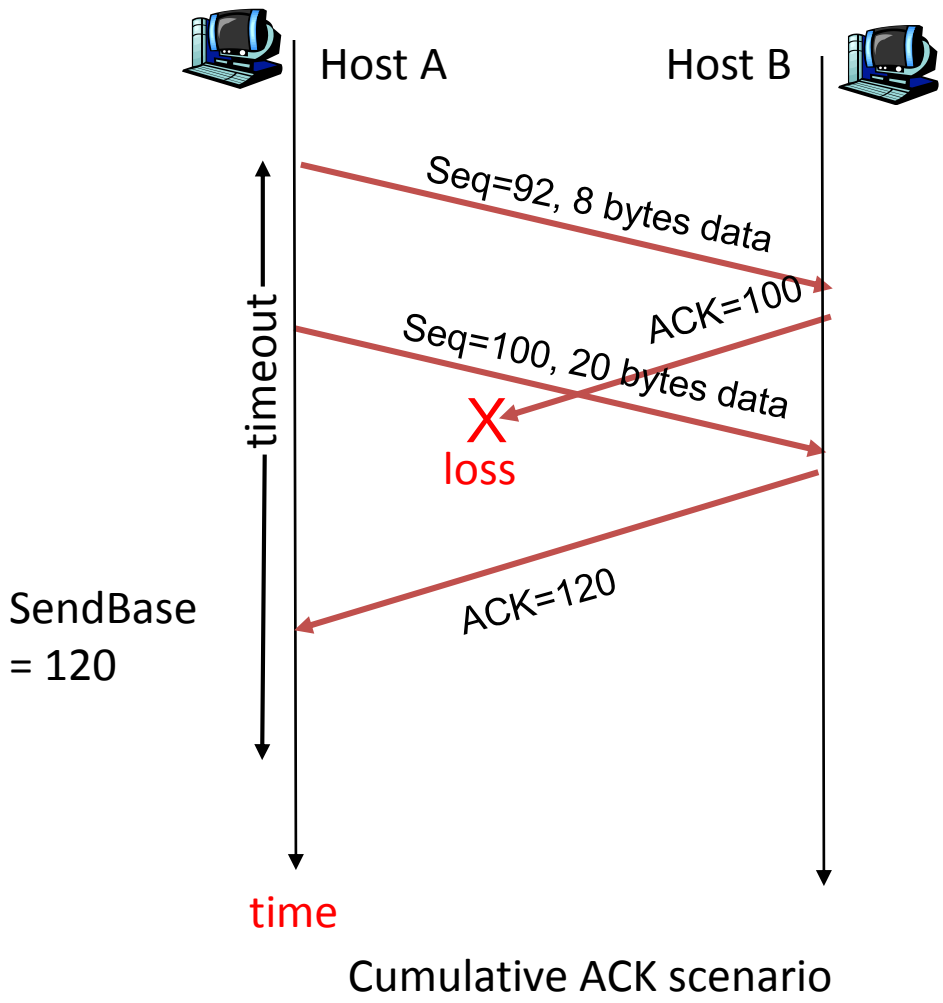
lost ACK scenario



premature timeout

Jim Kurose, Keith Ross. *Computer Networking: A Top Down Approach*, 4th edition. Addison-Wesley, July 2007.

Scenarios de retransmission TCP



Cumulative ACK scenario

Fast Retransmit

- Période de timeout souvent assez longue :
  - long delai avant ré-émission d'un paquet perdu
- Detection de segments perdus via ACKs dupliqués.
  - L'émetteur envoie souvent plusieurs segments à la suite.
  - Si un segment est perdu, il y aura probablement des ACKs dupliqués.
- Si l'émetteur reçoit 4 ACKs demandant le même paquet, il suppose que le segment a été perdu:
  - fast retransmit: renvoyer le segment avant que le timer de timeout expire

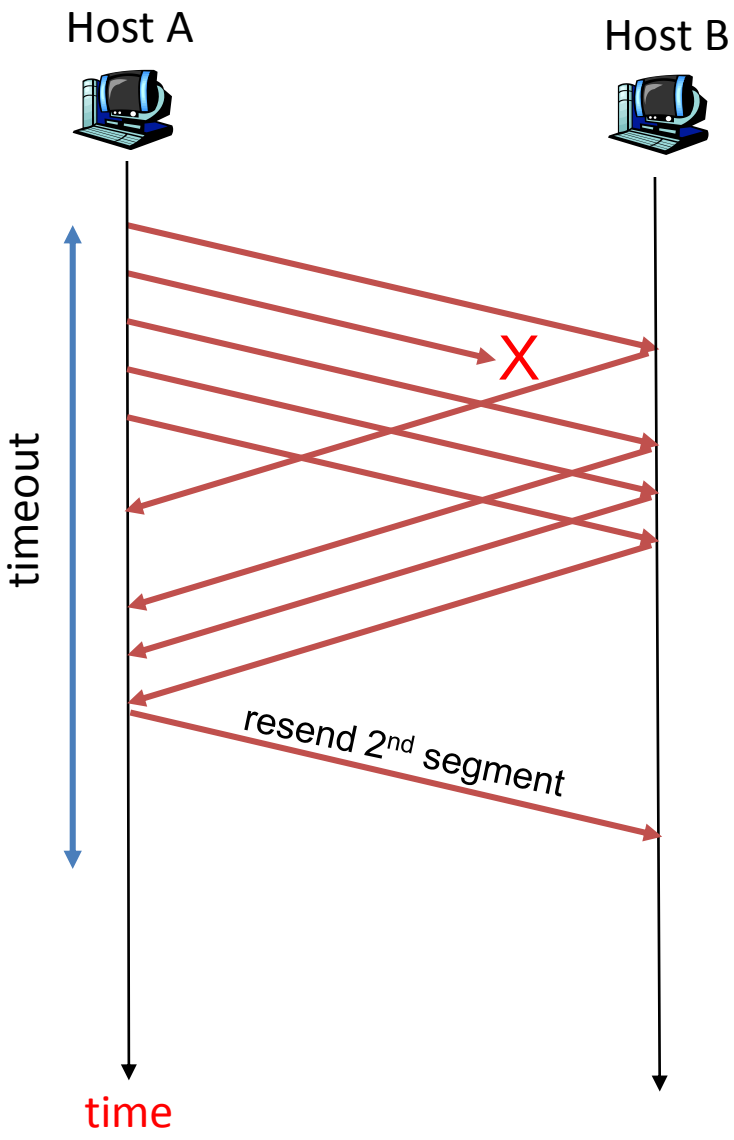
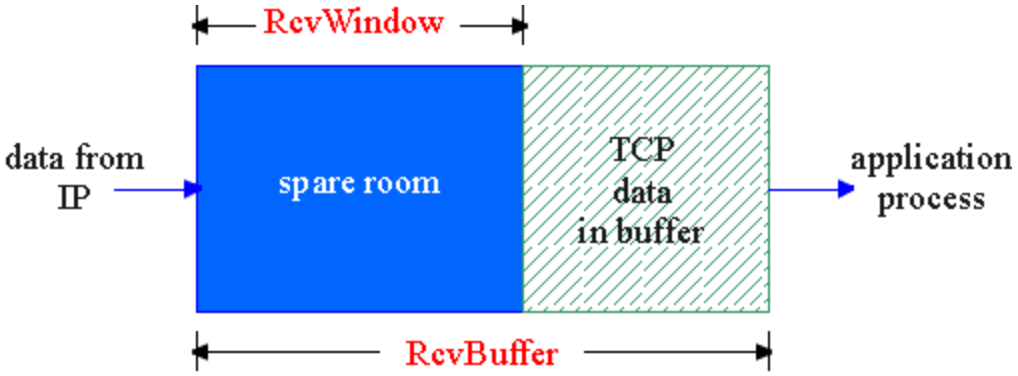


Figure 3.37 Resending a segment after triple duplicate ACK

Contrôle de flux TCP

- Le récepteur TCP a un buffer de réception:



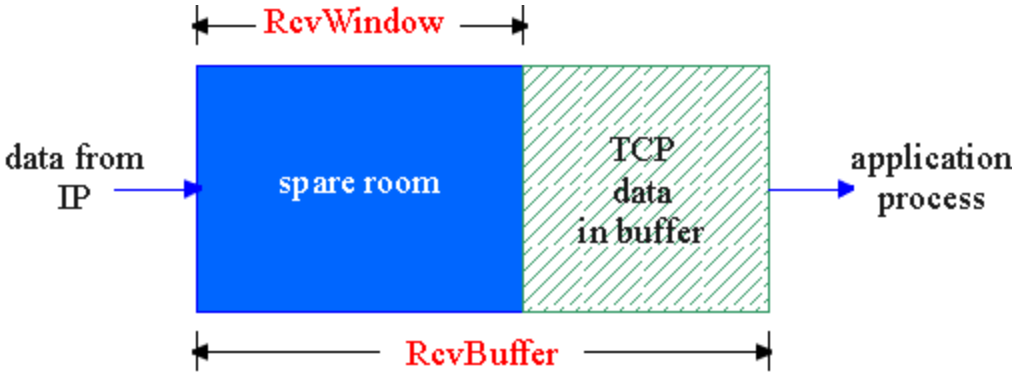
- Le processus application peut lire les données du buffer lentement

Contrôle de flux

L'émetteur ne va pas submerger le buffer du récepteur en envoyant trop et trop vite

- Il faut faire correspondre le débit d'émission avec le débit auquel l'appli peut lire les données

TCP Flow control: how it works



- Rcvr annonce la valeur **`RcvWindow`** dans l'entête TCP des segments renvoyés vers l'émetteur
- L'émetteur limite le nombre de paquets envoyés non encore accusés (ACKés) à **`RcvWindow`**
  - Garantie que le buffer de réception ne sature pas

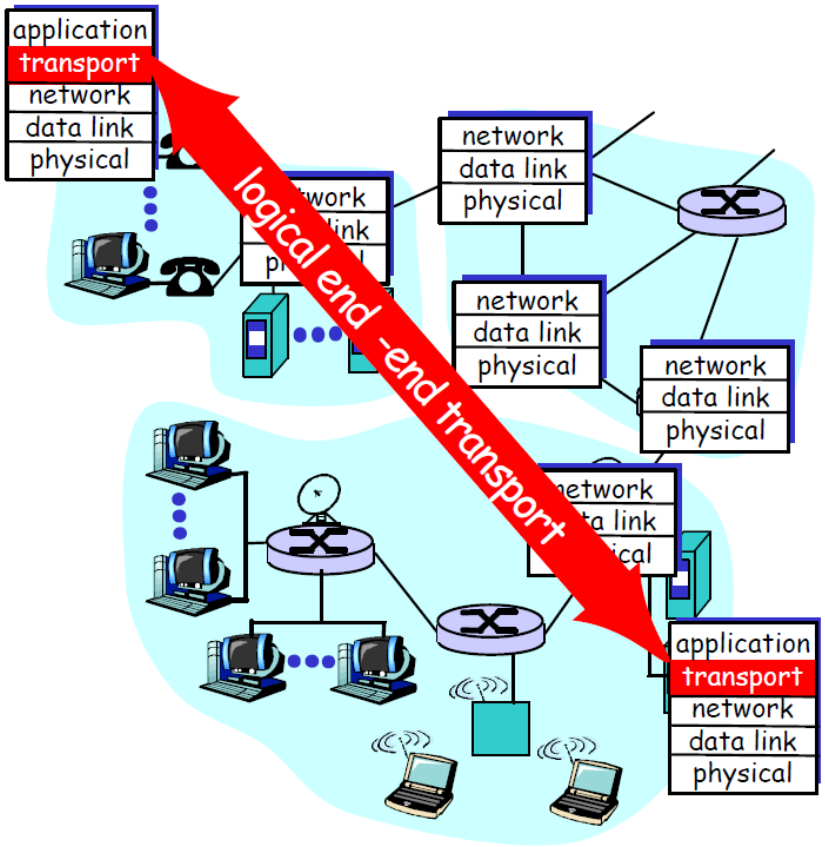
Plan général du cours

Protocole de contrôle de congestion standard: TCP et ses premières versions

- L'étude du contrôle de congestion dans les réseaux IP a débuté en 1986, avec les premiers *congestion collapses* sur Internet (dû à un nombre croissant de flows)

- 1988: Van Jacobson propose le 1er protocole de contrôle et évitement de congestion: TCP Tahoe

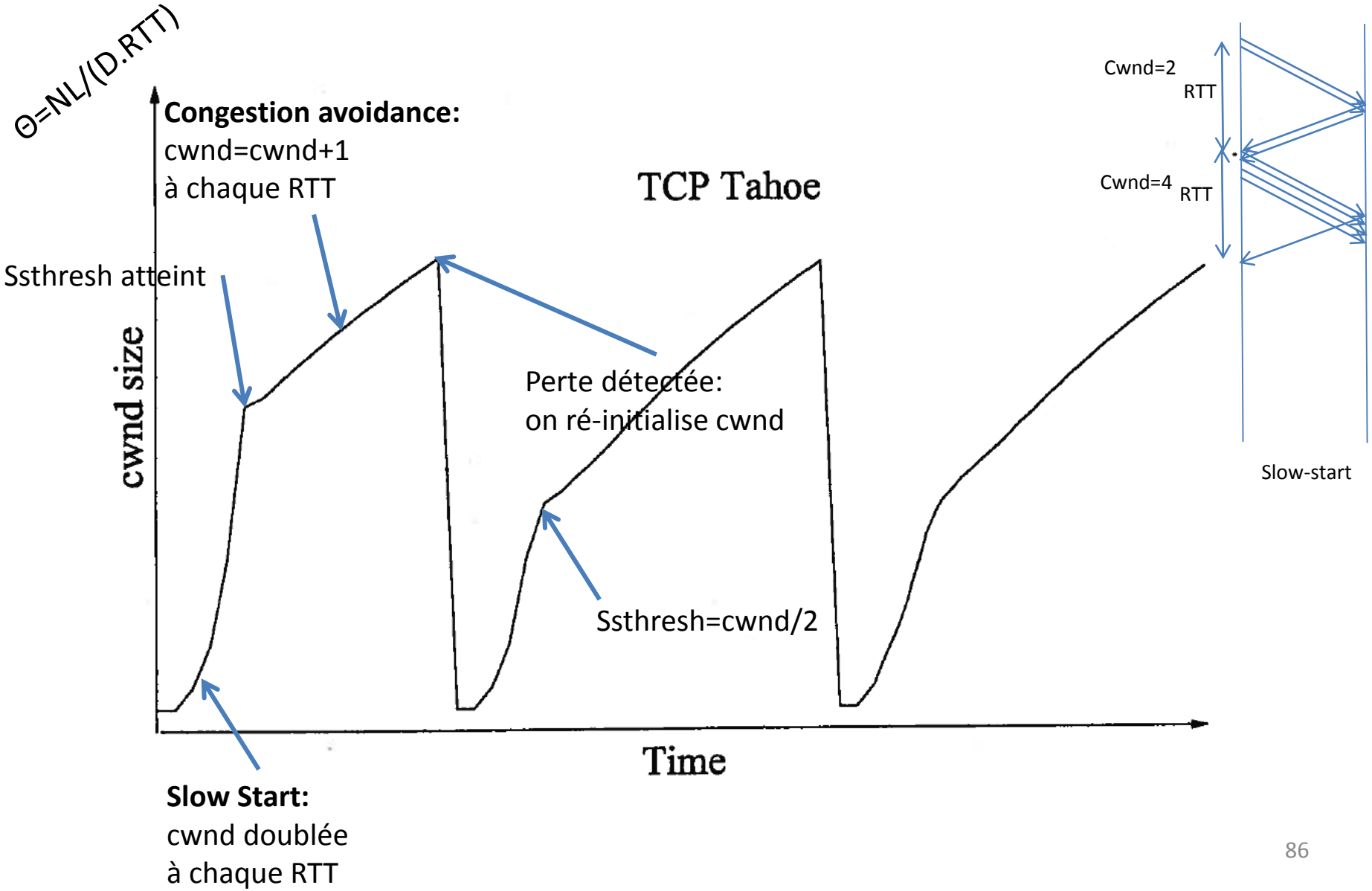
- Fonctionnement End-to-End (E2E): exécution du protocole qu'aux extrêmités



Principe de TCP

- Principe: pour chaque connexion, TCP maintient une ***congestion window (cwnd)*** **= nombre de segments TCP envoyés non encore acknowledgés**.
- But: atteindre une débit élevé tout en évitant le plus longtemps possible la congestion
- Paramètres importants :
  - RTT : temps d'aller-retour
  - RTO : temps au bout duquel, si pas de retour d'ACK, on déclare le segment perdu
  - ssthresh : seuil fixé à ~64KB initialement
  - > Ces paramètres sont constamment ré-évalués au cours du temps

TCP Tahoe: 1ère version de TCP



Contrôle de congestion TCP : détails

- L'émetteur limite le débit de transmission par la limitation du nombre d'octets envoyés à **`cwnd`**
- A peu près,

$$\text{Débit} = \frac{\text{Cwnd x taille paquet}}{\text{RTT}} \text{ Bytes/sec}$$

- **`cwnd`** est dynamique, fonction de la congestion du réseau perçue

Comment l'émetteur détecte la congestion ?

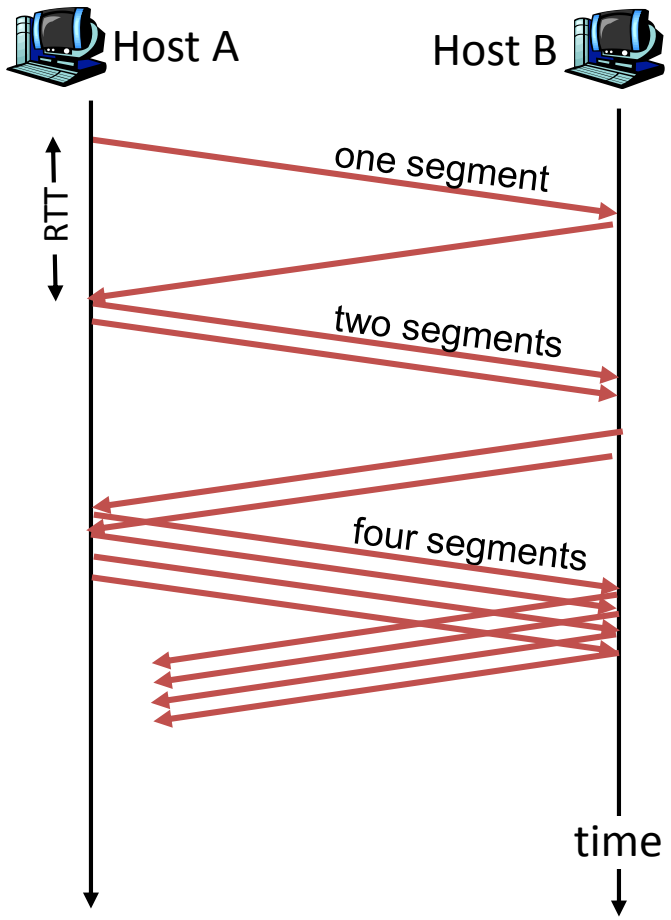
- Événement de perte = timeout *ou* 3 acks dupliqués
- L'émetteur TCP reduit le débit (**`cwnd`**) après une perte

3 mécanismes:

- AIMD
- slow start
- prudent après timeout

Slow Start TCP (more)

- Quand la connexion commence, augmenter le débit exponentiellement vite jusqu'à ce que la première perte (due à la congestion) se produise :
  - doubler **`cwnd`** chaque RTT
  - fait en augmentant **`cwnd`** d'un MSS à chaque ACK reçu
- En résumé: le débit initial est faible mais il augmente exponentiellement vite



Raffinement: slow-start + congestion avoidance + détection de perte améliorée

- But: atteindre une débit élevé tout en évitant le plus longtemps possible la congestion

- Après 3 dup ACKs :
  - **`cwnd`** diminuée de moitié (*Fast recovery*)
  - recommence à croître linéairement

- Mais après un événement de timeout :
  - **`cwnd`** ré-initialisée à 1 MSS;
  - slow-start (croissance exp)
  - jusqu'à un seuil, puis croissance linéaire

- La phase de croissance linéaire s'appelle la phase de *congestion avoidance*

Philosophie :

- 3 dup ACKs indiquent un réseau capable de délivrer des segments
- un timeout indique un problème de congestion plus grave

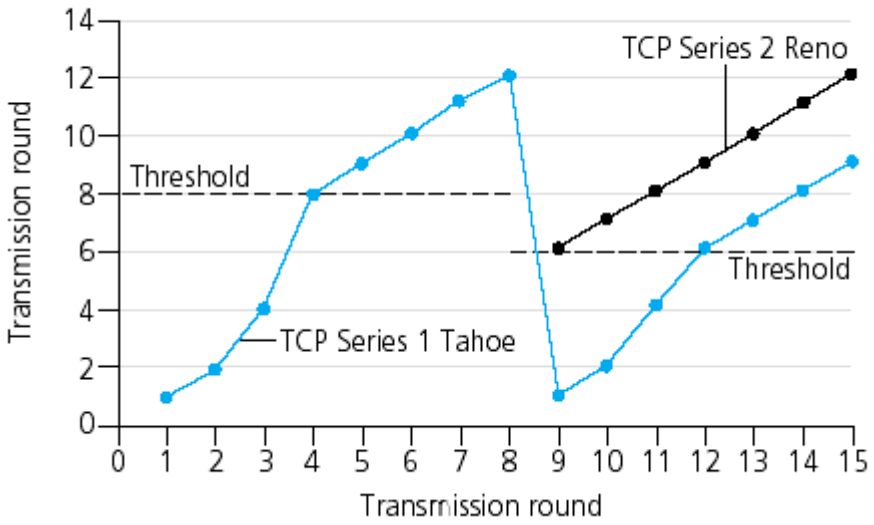
Raffinement

Q: Quand l'augmentation exponentielle doit basculer en linéaire ?

A: Quand **cwnd** atteint 1/2 de sa valeur avant dernière perte.

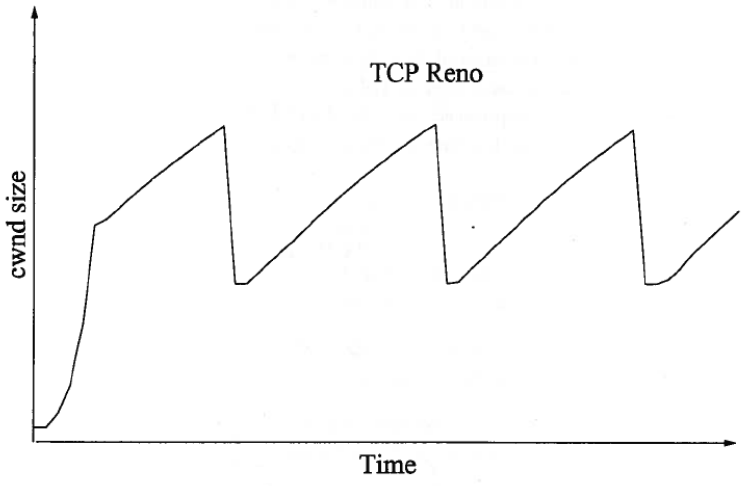
Q: Comment diminuer **cwnd** sur détection de perte ?

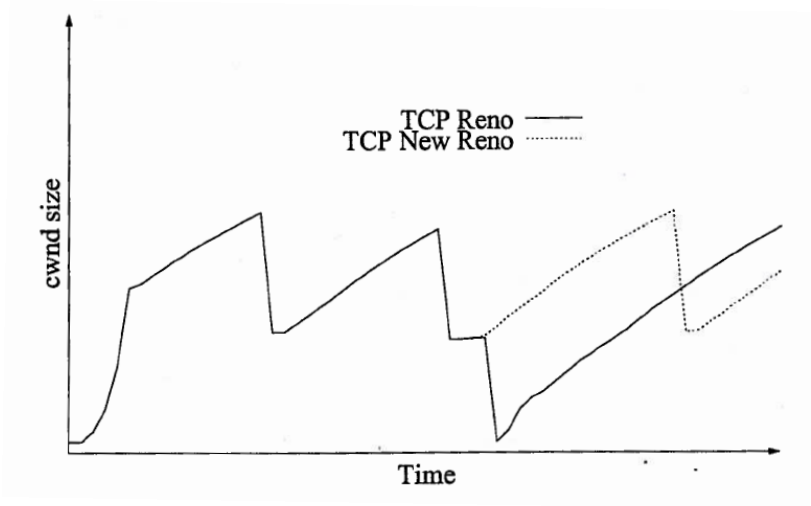
A: =1 si timeout, **cwnd** =/2 sinon



Versions de TCP

- Evolutions:
  1. **TCP Tahoe** : seulement RTO et cwnd remise à 1 à chaque perte
  2. Evolution de TCP Tahoe : **Fast retransmit** utilisé en plus de RTO
  3. **TCP Reno**: **Fast recovery**: cwnd=cwnd/2 au lieu de cwnd=1 après perte
  4. **TCP New Reno**: seulement 1 FR/FR après 2 pertes de paquets de la même fenêtre





Résumé : Controle de congestion TCP

- Quand **`cwnd`** est en dessous de **`ssthresh`**, l'émetteur est dans la phase de slow-start, la fenêtre croît exponentiellement.
- Quand **`cwnd`** est au dessus de **`ssthresh`**, l'émetteur est dans la phase de congestion-avoidance, la fenêtre croît linéairement.
- Quand un triple duplicate ACK se produit, **`ssthresh`** est mis à **`cwnd/2`** et **`cwnd`** est mise à **`ssthresh`**.
- Quand un timeout se produit, **`ssthresh`** est mis à **`cwnd/2`** et **`cwnd`** est mise à 1 MSS.

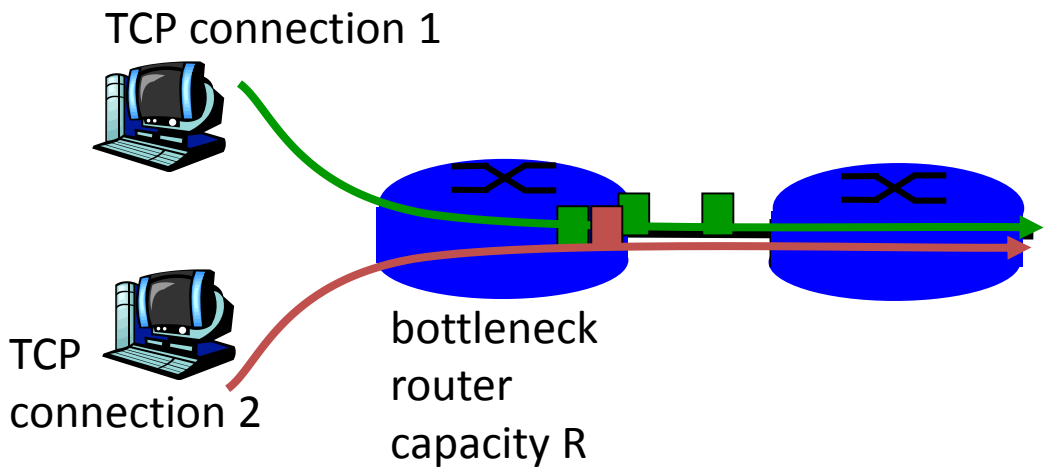
Jim Kurose, Keith Ross. *Computer Networking: A Top Down Approach*, 4th edition. Addison-Wesley, July 2007.

TCP sender congestion control

State	Event	TCP Sender Action	Commentary
Slow Start (SS)	ACK receipt for previously unacked data	cwnd = cwnd + MSS, If (cwnd > ssthresh) set state to "Congestion Avoidance"	Resulting in a doubling of cwnd every RTT
Congestion Avoidance (CA)	ACK receipt for previously unacked data	cwnd = cwnd + MSS * (MSS/cwnd)	Additive increase, resulting in increase of cwnd by 1 MSS every RTT
SS or CA	Loss event detected by triple duplicate ACK	ssthresh = cwnd / 2, cwnd = ssthresh, Set state to "Congestion Avoidance"	Fast recovery
SS or CA	Timeout	ssthresh = cwnd / 2, cwnd = 1 MSS, Set state to "Slow Start"	Enter slow start
SS or CA	Duplicate ACK	Increment duplicate ACK count for segment being acked	cwnd and ssthresh not changed

L’équité de TCP

But de l’équité : si K sessions TCP partagent le même goulot d’étranglement de BP R, chacune doit obtenir un débit de R/K



Equité

Equité et UDP

- Les appli multimedia souvent n'utilisent pas TCP
  - Ne veulent pas être bridées ou étouffées par le CC
- Utilisent UDP à la place :
  - injectent audio/video à un débit constant, tolèrent des pertes de paquets

Equité et connexions TCP parallèles

- rien n'empêche les appli d'ouvrir plusieurs connexions en parallèle entre 2 hôtes
- les Web browsers font ça
- Exemple: lien de BP R supportant 9 connexions;
  - une nouvelle appli demande 1 TCP, obtient un débit de R/10
  - une nouvelle appli demande 10 TCPs, obtient R/2 !

Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

- II.1.a. Principe du contrôle de congestion
- II.1.b. Rappels : transfert fiable et contrôle de flux
- II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

- II.2.a. Définition de la QoS et classes d'applications
- II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS
- II.2.c. Stratégies possible pour garantir un niveau de QoS

II.3. Techniques de traitement de la QoS

- II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing
- II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets
- II.3.c. Gestion de buffer : comment abandoner les paquets en excès
- II.3.d. Marquage du trafic : DiffServ dans IP

III. Commutation par circuits virtuels

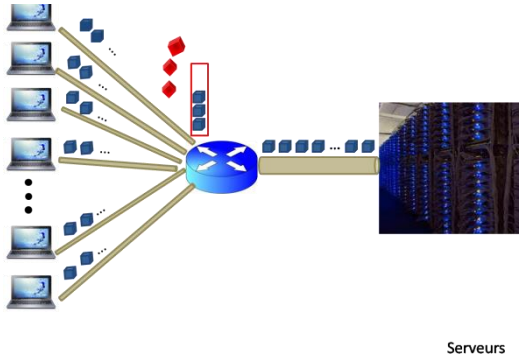
IV. Commutation par circuits virtuels dans le mode IP : MPLS

Plan général du cours

Qualité de service : définition

- Performances du service fourni par le réseau à l'application.
- Performances mesurées par les métriques de :
  - Débit
  - Délai
  - Variation du délai
  - Taux de perte
  - Taux d'indisponibilité

Le problème de QoS

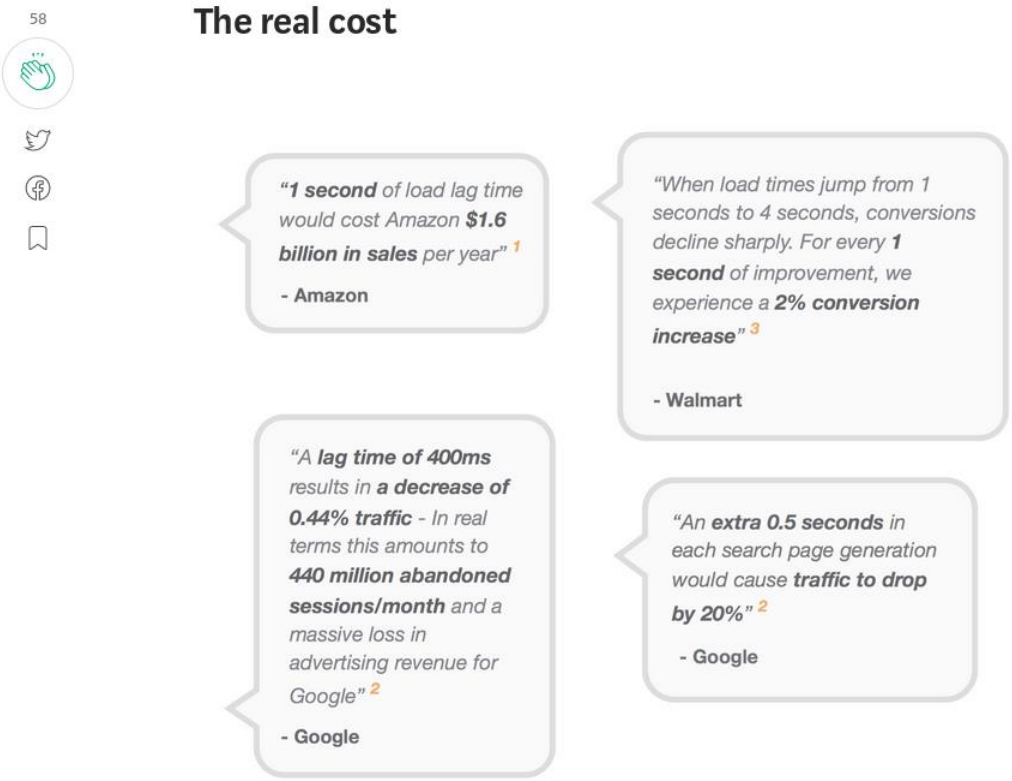


- Rappel congestion : mise en file d'attente des paquets à un éqt intermédiaire si débit d'entrée > débit de sortie
- TCP, pour utiliser le débit maximum qu'il ne connaît pas, provoque de la congestion quand il dépasse le débit, et réduit le débit quand il la détecte
- -> il remplit donc le buffer fréquemment
- -> augmente le temps de transfert des paquets
- -> pas grave si paquets de téléchargement, grave si paquet de Voix sur IP (VoIP), ou HTTP web browsing
- -> le problème de la QoS est là : éviter que les paquets ayant des contraintes (par exemple de délai) ne séjournent derrière des paquets moins pressés dans les buffers pleins des routeurs ou switchs.

Classification des applications

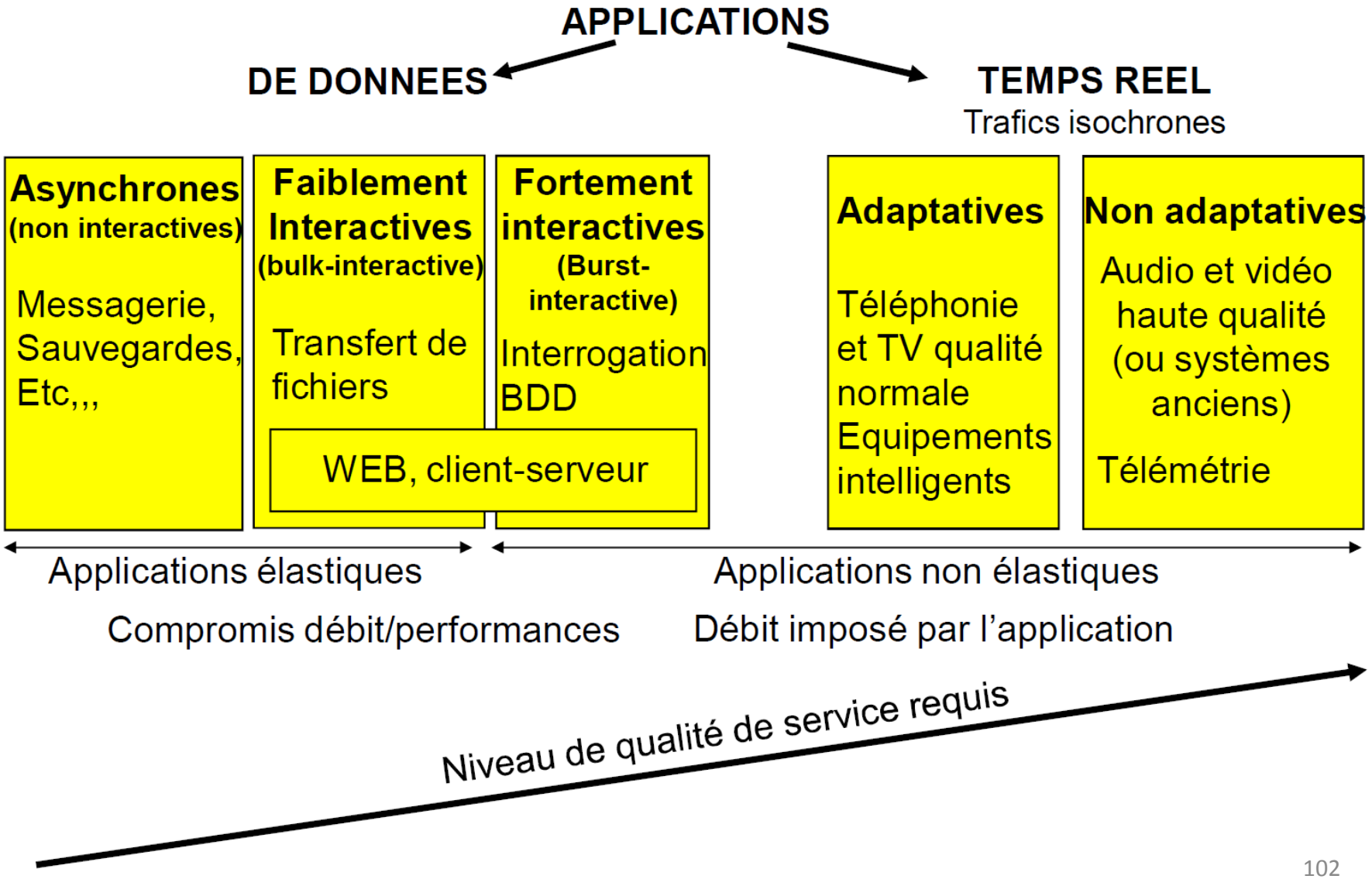
Application	Pertes	Débit	Sensibilité au délai
file transfer	no loss	elastic	no
e-mail	no loss	elastic	no
Web documents	no loss	elastic	no
real-time audio/video	loss-tolerant	audio: 5kbps-1Mbps video:10kbps-5Mbps	yes, 100's msec
stored audio/video	loss-tolerant	same as above	yes, few secs
interactive games	loss-tolerant	few kbps up	yes, 100's msec
instant messaging	no loss	elastic	yes and no

Impact of slow page load time on website performance



- https://medium.com/@vikigreen/impact-of-slow-page-load-time-on-website-performance-40d5c9ce568a (2016)

Classification des applications



Le trafic de l'Internet

- Types et partage des applications
- Types et partage des accès
- Répartition géographique
- Etat courant et prévisions futures : 2017-2022

→ Référence: Cisco Visual Networking Index Forecast

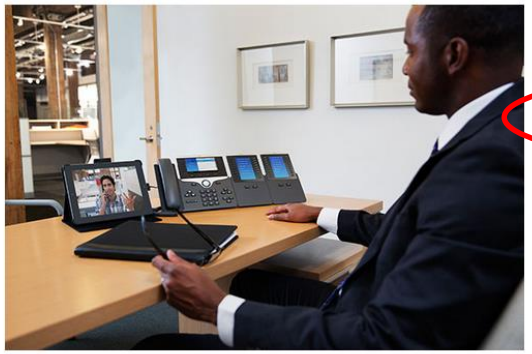
Cisco VNI

https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html#~stickynav=1

Search

Solutions / Service Provider /

VNI Global Fixed and Mobile Internet Traffic Forecasts



Complete Visual Networking Index (VNI) Forecast

12th annual complete VNI forecast. Explore global fixed/mobile traffic projections and Internet trends (2016 – 2021).

Watch the latest VNI webcast for in-depth forecast insights and trends analysis. Choose AMER/EMEAR for most regions or APJC for Asia Pacific and China.

AMER/EMEAR APJC

Complete Forecast | Mobile Forecast | Cloud Forecast | Contact Us

Complete Visual Networking Index (VNI) Forecast

The complete VNI report forecasts global IP traffic growth for mobile and fixed networks. By the year 2021:

4.6B Global Internet users

White paper: VNI Forecast and Methodology, 2016-2021

27.1B Networked devices and connections

Infographic: 2017 Cisco Complete VNI Forecast

82% Of all IP traffic will be video

White paper: The Zettabyte Era

AMER/EMEAR Webcast APJC Webcast

Executive summary

- Annual global IP traffic will reach 3.3 ZB (10^21 B) by 2021.
- **Global IP traffic will increase nearly threefold over the next 5 years, and will have increased 127-fold from 2005 to 2021.**
- Busy-hour Internet traffic is growing more rapidly than average Internet traffic.
- Smartphone traffic will exceed PC traffic by 2021.
- **Traffic from wireless and mobile devices will account for more than 63 percent of total IP traffic by 2021.**
- Global Internet traffic in 2021 will be equivalent to 127 times the volume of the entire global Internet in 2005.
- **The number of devices connected to IP networks will be three times as high as the global population in 2021.**
- Broadband speeds will nearly double by 2021.

Video highlights

- It would take an individual more than 5 million years to watch the amount of video that will cross global IP networks each month in 2021.
- **Globally, IP video traffic will be 82 percent of all consumer Internet traffic by 2021, up from 73 percent in 2016.**
- **Live Internet video** will account for 13 percent of Internet video traffic by 2021.
- **Internet video surveillance traffic increased 72 percent in 2016, from 516 Petabytes (PB) per month at the end of 2015 to 883 PB per month in 2016.**
- **Virtual reality and augmented reality traffic** will increase 20-fold between 2016 and 2021, at **a CAGR of 82 percent**.
- **Internet video to TV grew 50 percent in 2016.**
- **Consumer Video-on-Demand (VoD) traffic will nearly double by 2021.** The amount of VoD traffic in 2021 will be equivalent to 7.2 billion DVDs per month.
- **Content Delivery Network (CDN) traffic will carry 71 percent of all Internet traffic by 2021.** Up from 52 percent in 2016.

Applications non-adaptatives

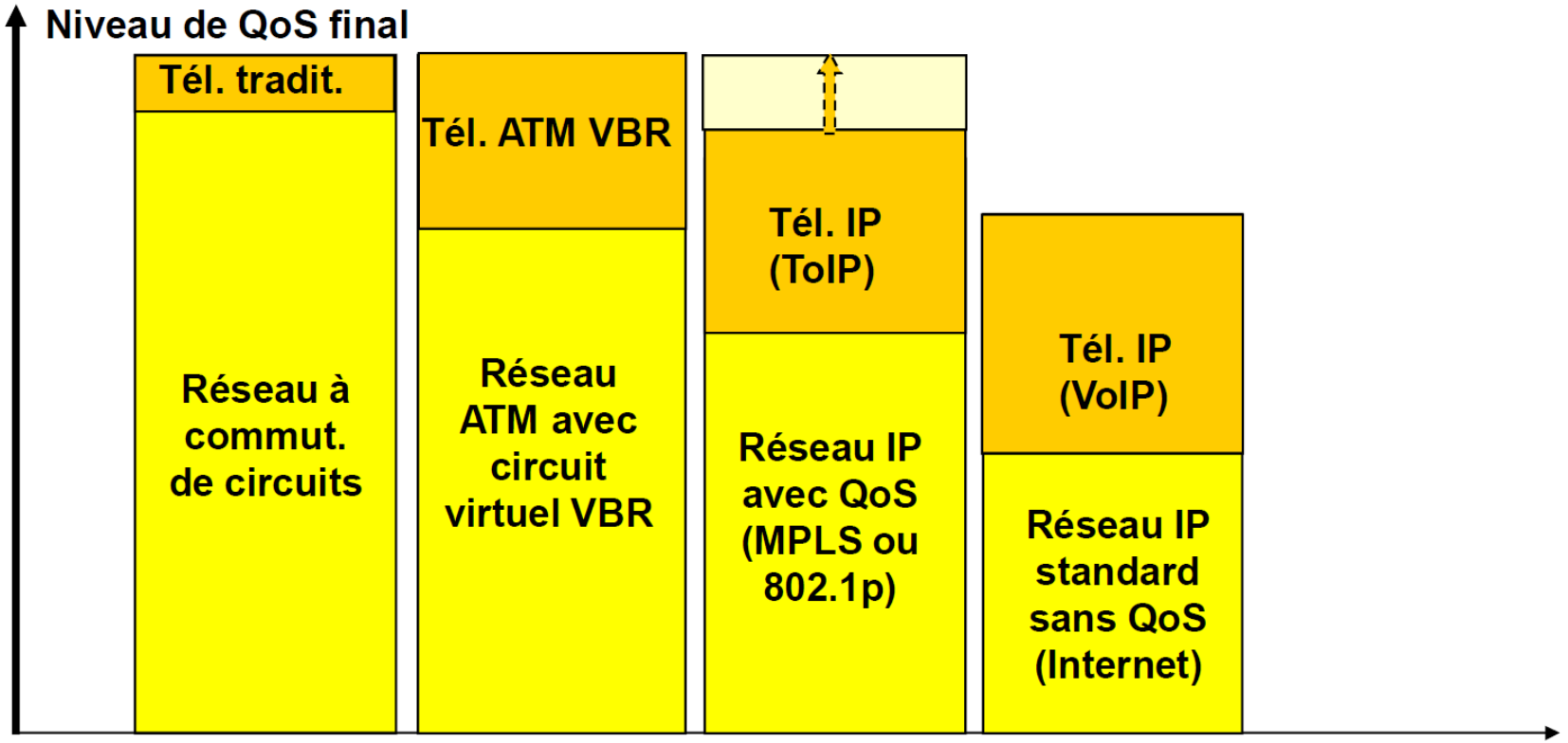
- **Les applications temps réel (voix, vidéo) ont un rythme imposé (*isochrone*) qui doit se maintenir de la source au destinataire**
  - Elles s'adaptent beaucoup plus mal à l'état de congestion du réseau
  - Le débit spécifié doit être assuré par le réseau sous peine de non fonctionnement
    - Ce débit était constant dans le cas des codecs sans compression
    - Le débit est variable quand les techniques de compression sont plus évoluées
  - Le délai doit être faible pour les applications temps réel interactives
    - Téléphonie ou vidéoconférence
  - Le délai doit en tout cas être le plus constant possible
    - La variation de délai (la ***gigue*** ou ***jitter***) doit être faible
  - Les pertes de paquet ne sont pas récupérées par retransmission
  - Ces applications utilisent UDP, qui ne tient pas compte de l'état de congestion du réseau
    - Ces applications doivent donc être protégées

Applications adaptatives

- **Les applications de données sur TCP s'adaptent à l'état de congestion du réseau**
  - Contrôle de flux et récupération de paquets perdus par TCP
  - L'allongement du délai n'est en général pas un problème
    - Les applications bulk sont peu interactives
- **Les applications de données de type burst ne s'adaptent pas à l'allongement du délai**
  - Mécontentement des utilisateurs
- **Les applications temps réel traditionnelles n'avaient pas à s'adapter**
  - Elles utilisaient un réseau à commutation de circuits
    - Délai très court, gigue très faible, débit dédié
- **Les applications temps réel modernes deviennent adaptatives**
  - Pour s'adapter aux fluctuations d'un réseau à commutation de paquets
  - Buffer de stockage à l'arrivée pour supprimer la gigue
    - Si la gigue est inférieure à un plafond, et au détriment du délai
  - Interpolation en cas de paquets perdus

Applications adaptatives

- **Les équipements d'extrémité sont de plus en plus intelligents**
  - Ils participent de plus en plus au niveau de QoS final obtenu
  - Et demandent moins de QoS au réseau
- **Tendance vers un réseau plus simple donc plus rapide et des stations plus intelligentes**



Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

- II.1.a. Principe du contrôle de congestion
- II.1.b. Rappels : transfert fiable et contrôle de flux
- II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

- II.2.a. Définition de la QoS et classes d'applications
- II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS
- II.2.c. Stratégies possible pour garantir un niveau de QoS

II.3. Techniques de traitement de la QoS

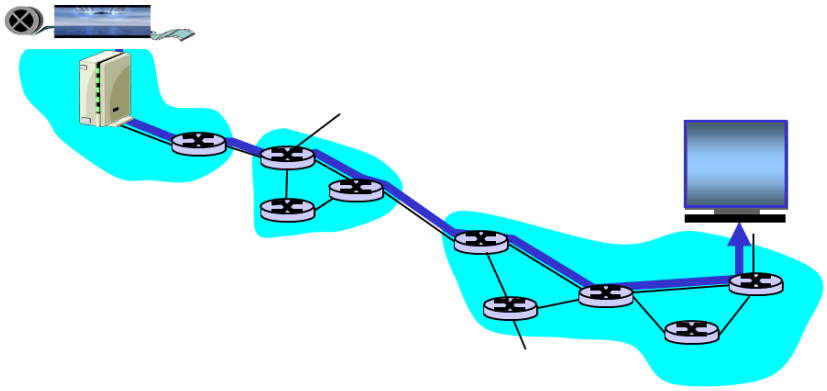
- II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing
- II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets
- II.3.c. Gestion de buffer : comment abandoner les paquets en excès
- II.3.d. Marquage du trafic : DiffServ dans IP

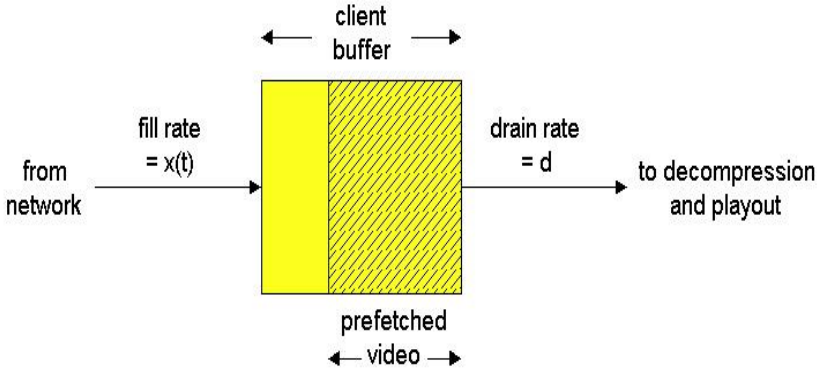
III. Commutation par circuits virtuels

IV. Commutation par circuits virtuels dans le mode IP : MPLS

Streaming: définition

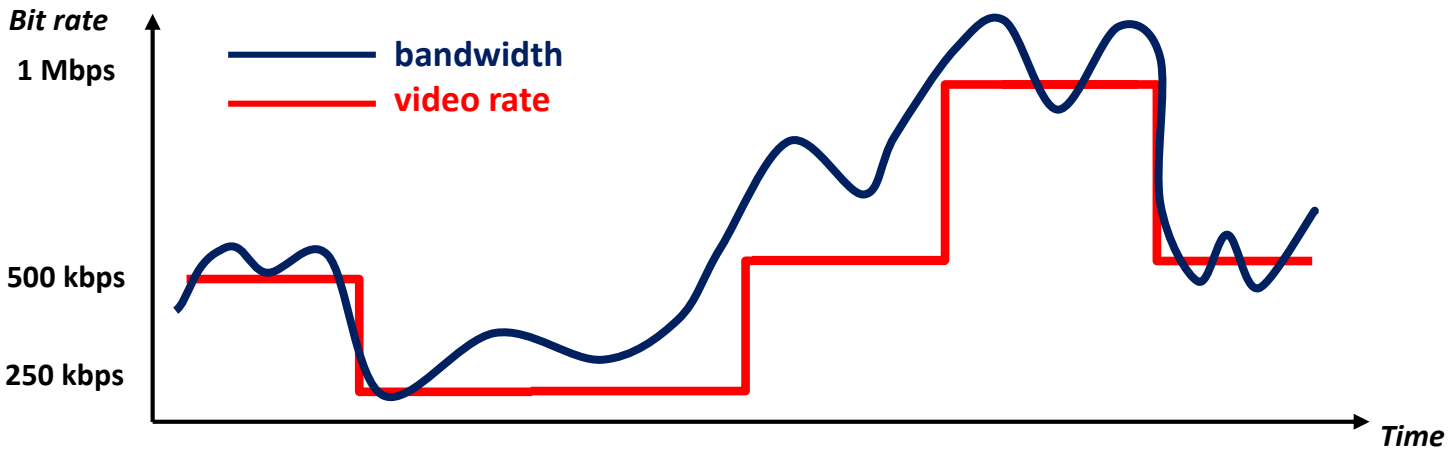
- Streaming : la destination commence à consommer le contenu avant qu'il ne soit complètement reçu
- Terminologie pour le streaming vidéo:
  - *video rate*: débit auquel la video est encodée (en bps), donc débit de lecture en sortie du buffer
  - *bandwidth*: bande passante disponible sur le chemin src-dst, donc débit de remplissage du buffer





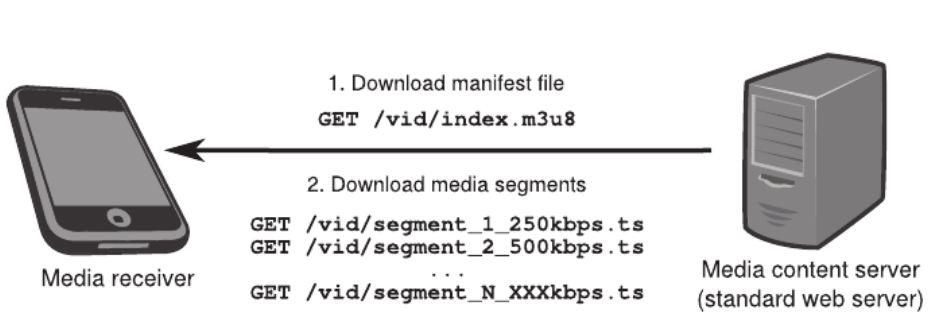
Streaming adaptatif: principe

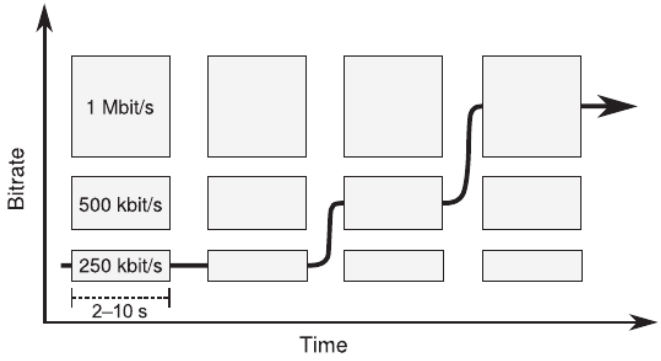
- L'appli cliente adapte au cours du temps le *video rate* demandé pour ne pas dépasser la BP disponible qui varie, et éviter que le buffer ne se vide.



Dynamic Adaptive Streaming over HTTP: MPEG DASH

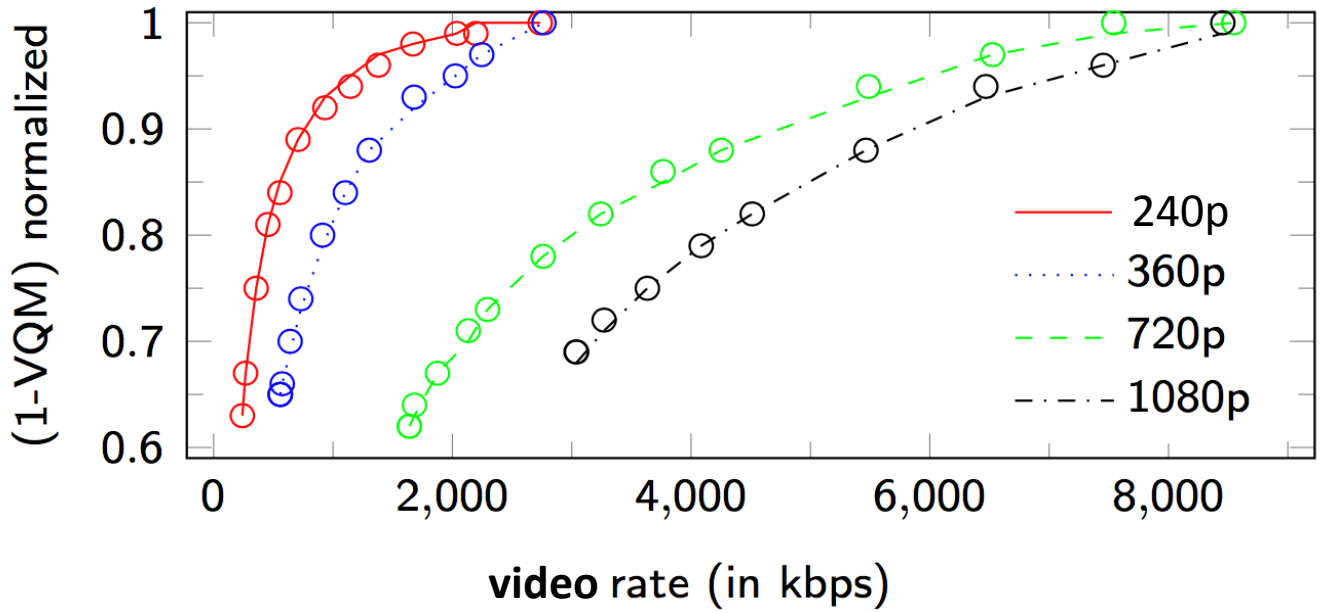
- Plusieurs versions de la même video sont disponibles au serveur, avec des qualités (debit d'encodage, *video rates*) différentes
- Les requêtes HTTP se font par segments (2s ou 5s), à chaque segment demandé on choisit le debit d'encodage en function de la BP (débit de téléchargement) et de la taille du buffer



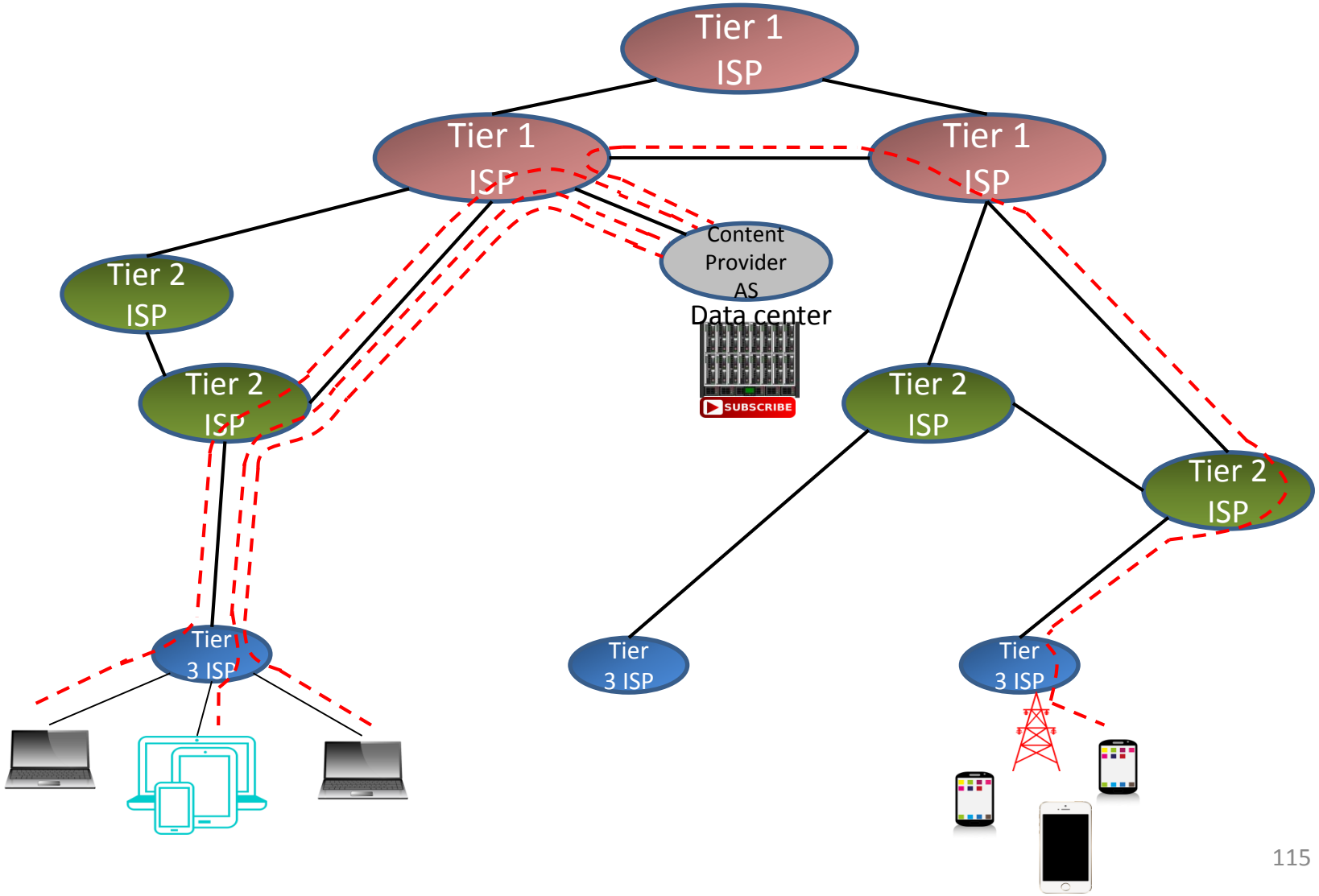


La perception humaine est une fonction complexe de la QoS

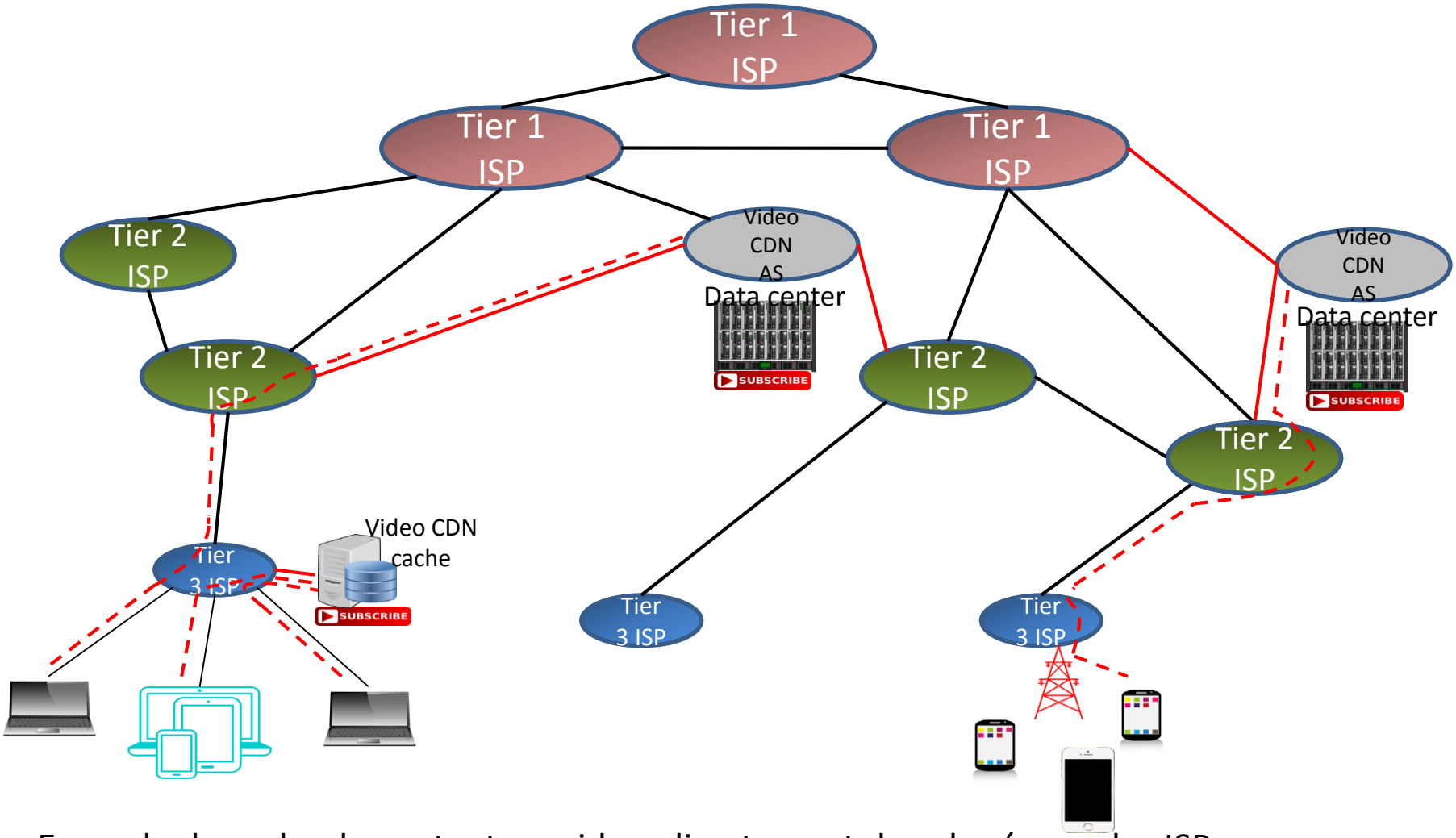
- On peut relier le débit du réseau à la qualité visuelle perçue pour chaque résolution :



Un autre levier pour le streaming : les caches



Un autre levier pour le streaming : les caches



Exemple de cache de content providers directement dans le réseau des ISP: https://openconnect.netflix.com/en/

Plan général du cours

A-t-on vraiment besoin de traiter la QoS ?

- Les débits des accès augmentent :
  - Fixe: fibre se déploie (Cisco VNI: moyenne à ~50Mbps en 2020)
  - Mobile: 4G vise 100Mbps, bientôt 5G qui vise 1Gbps
- Les applications deviennent adaptatives (HTTP adaptive streaming: video live et VoD distribuée avec MPEG DASH)
- L'infrastructure CDN se généralise pour répliquer le contenu proche des utilisateurs pour lesquels il est populaire.

Oui, le traitement de la QoS est nécessaire pour gérer la compétition entre flux

- **Les différents types d'applications induisent des flux qui se contrarient**
  - Les transferts de gros volumes (par exemple les transferts de fichiers) sont élastiques, et prennent toute la bande passante disponible
    - Que ce soit 64 Kbps ou 100Mbps!
  - Les applications interactives (question/réponse) consomment une bande passante prévisible, mais exigent des délais courts
    - Elles sont pénalisées par les transferts de fichiers
  - Les applications temps réel (téléphonie ou vidéo) ne sont pas élastiques (besoin borné en bande passante), mais elles doivent avoir un débit garanti, et un délai court et fixe
    - Elles sont pénalisées par les applications de gros transferts
- **Avoir une bande passante surdimensionnée ne suffit pas**
  - Il faut empêcher les applications élastiques de prendre toute la bande passante
- **Avoir beaucoup de bande passante permet juste de simplifier le traitement de la qualité de service**
  - Besoin de mécanismes de QoS simples

Paramètres-clé de la QoS

Paramètres	Métriques	Applications concernées
• Latence : délai total de traversée du réseau	Délai Round Trip Time (délai AR)	Applications interactives
• Gigue (jitter) : variation du délai	Différence entre délai max et délai min	Applications temps réel (audio, vidéo)
• Bande passante disponible	Débit minimum Débit moyen Débit en pointe	Applications interactives Audio et vidéo
• Perte de paquets	Taux de perte Taux d'erreurs	Applications temps réel
• Disponibilité	Disponibilité = % du temps de bon fonctionnement	Applications critiques pour l'entreprise

C'est la congestion qui est responsable de la plupart des dégradations de QoS

Stratégies de traitement de la qualité de service

- **Cinq moyens complémentaires d'assurer la qualité de service**
- **1. Surdimensionnement de la capacité du réseau**
- **2. Utiliser des applications adaptatives**
  - Interpolation de données manquantes (temps réel)
  - Émettre à débit variable en fonction de la congestion
  - Buffers de réception pour compenser la gigue, …
- **3. Traitement sélectif du trafic sans réservation préalable**
  - Le trafic est classifié, et le traitement différencié (DiffServ)
  - Chaque classe a son traitement spécifique
    - Files d'attente séparées
    - Traitement spécifique en cas de congestion
  - Les trafics à privilégier sont prioritaires
- **4. Réservation dynamique de ressources**
  - Frame Relay, ATM, IntServ - RSVP
- **5. Ingénierie de trafic**
  - Répartir le trafic dans le réseau en fonction de la bande passante disponible

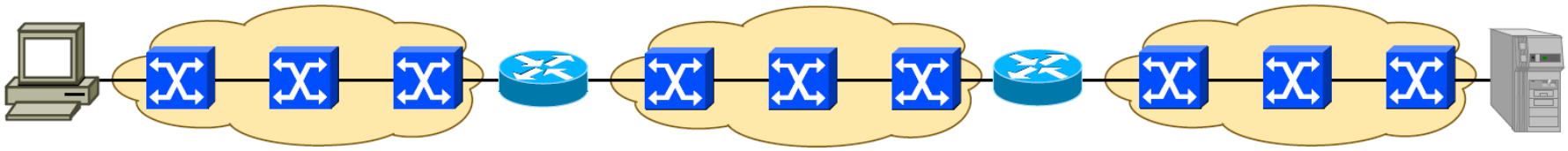
-> MPLS-TE

Plan de données

Plan de routage

Stratégies de traitement de la qualité de service

- **Dans quelle couche OSI doit-on traiter la qualité de service?**
- **Traitement au niveau des couches supérieures?**
  - Oui si on veut compenser les imperfections résiduelles de QoS
  - Par TCP (applications élastiques) ou par les applications adaptatives (VoIP)
- **Traitement au niveau 3?**
  - Oui si on veut conserver une QoS sur l'ensemble du parcours
  - Traitement dans les routeurs
- **Traitement au niveau 2?**
  - Oui si on veut que la QoS soit homogène de bout en bout
  - Traitement dans chaque commutateur de chaque sous-réseau multipoint
- **Le traitement de la qualité de service est réparti dans toutes les couches**
  - La répartition des rôles est variable selon les modèles, c'est le résultat final qui compte!
    - Exemples : RTC, VoIP sur Internet, ToIP en entreprise



RTC : Réseau Téléphonique commuté VoIP : Voice over IP ToIP : Telephony over IP

Optimisation: le cas d'un ISP

- Un ISP cherche à maximiser son nombre de clients:
- Trouver $\mathbf{r}_{opt}$ = argmax nombre clients

- Avec $\mathbf{r}_{e,i}$ le débit affecté au client i sur le lien e
- Sous contrainte de $\mathbf{r}$
  - tous les contrats clients (SLAs) satisfaits par l'ISP

→ En d'autres termes, un ISP veut :

Faire passer tous les flux demandés et de divers types (vidéo live ou à la demande, applications asynchrones, téléphonie et vidéo-conférence, etc.), sans gaspiller les ressources: utiliser au maximum les câbles payés et installés

Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

- II.1.a. Principe du contrôle de congestion
- II.1.b. Rappels : transfert fiable et contrôle de flux
- II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

- II.2.a. Définition de la QoS et classes d'applications
- II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS
- II.2.c. Stratégies possible pour garantir un niveau de QoS

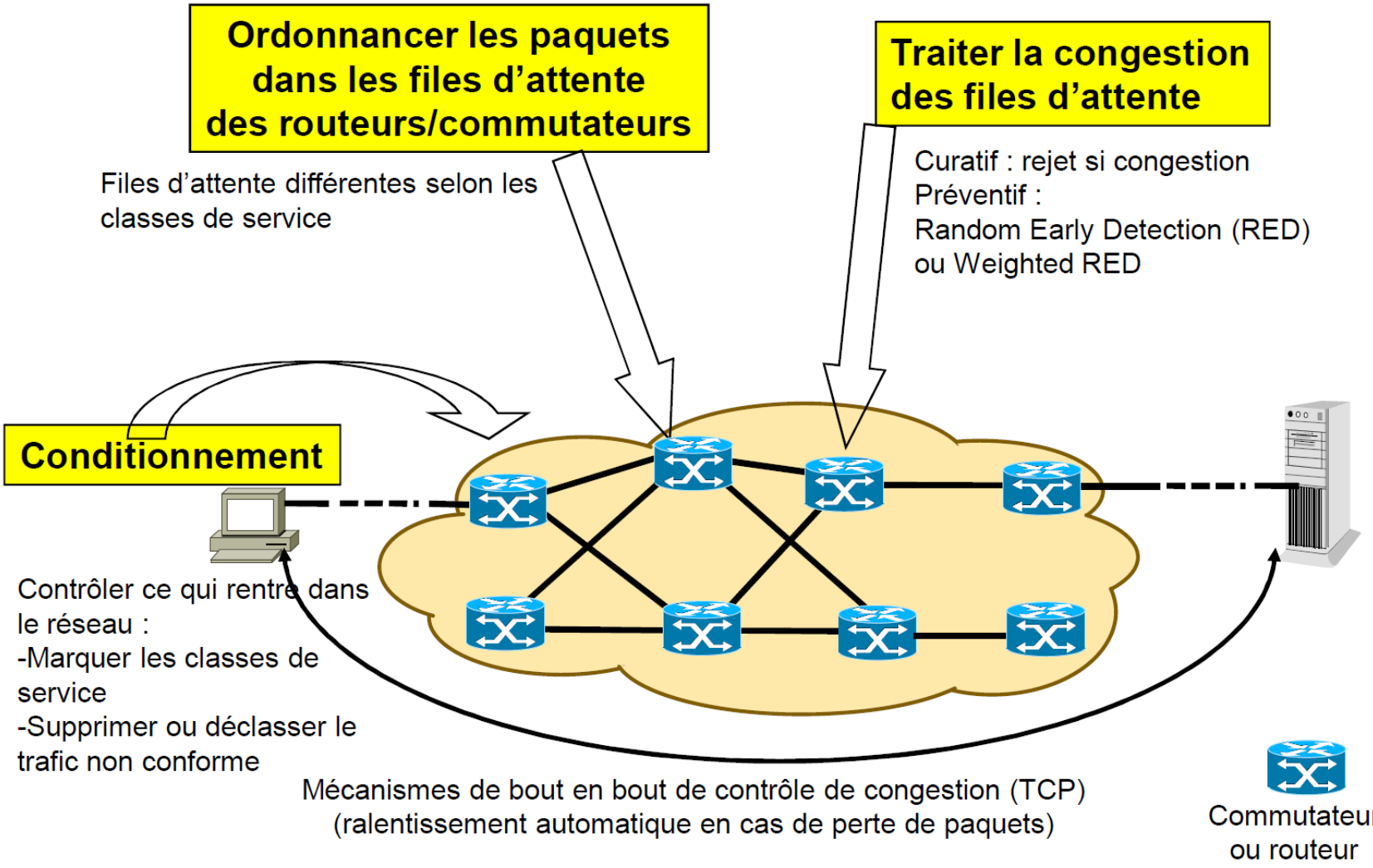
II.3. Techniques de traitement de la QoS

- II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing
- II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets
- II.3.c. Gestion de buffer : comment abandoner les paquets en excès
- II.3.d. Marquage du trafic : DiffServ dans IP

III. Commutation par circuits virtuels

IV. Commutation par circuits virtuels dans le mode IP : MPLS

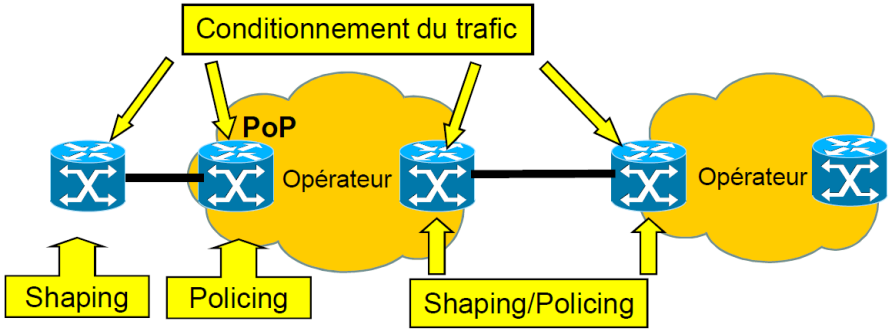
Plan de données: actions sur les données



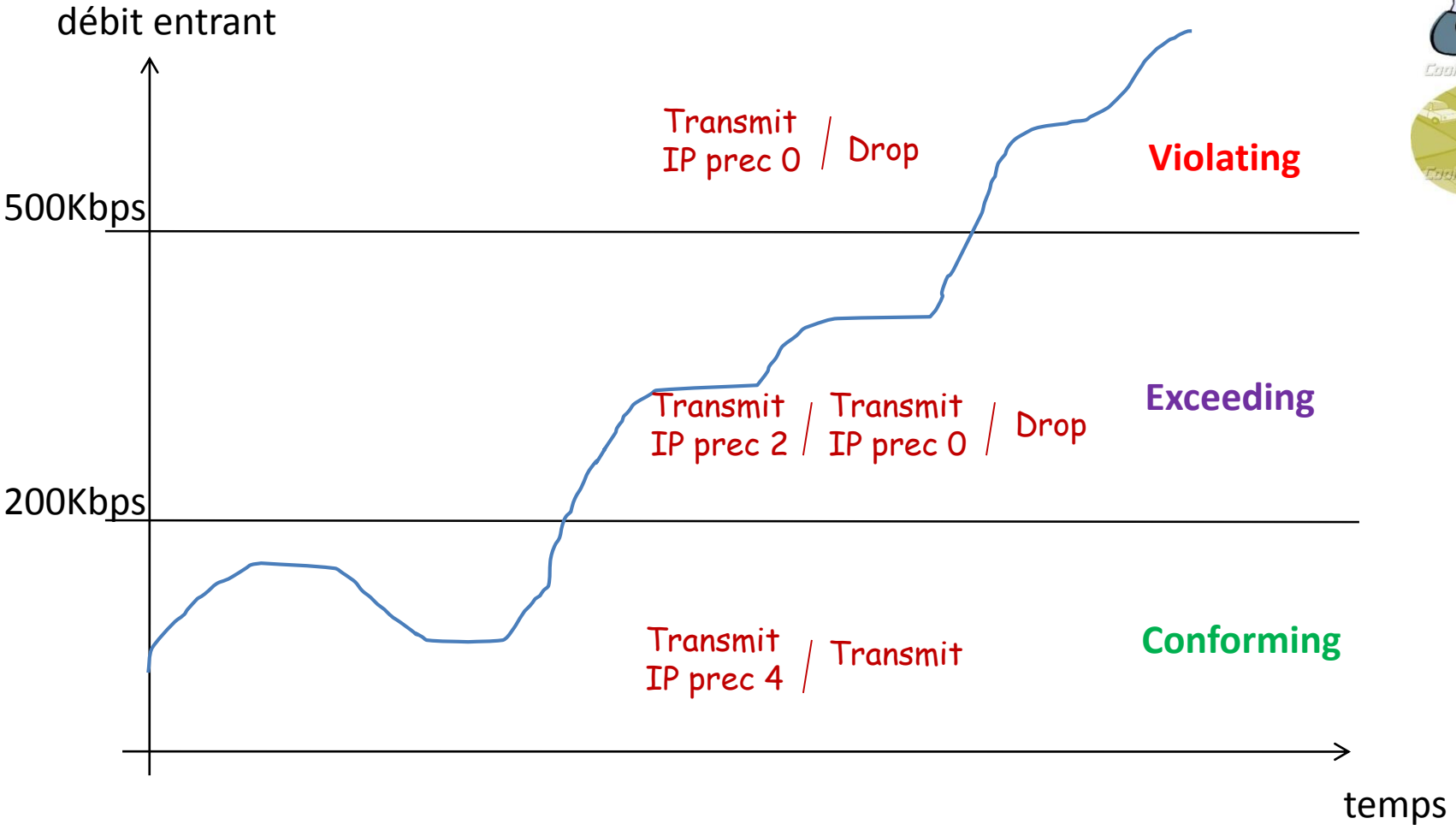
Policing et Shaping

- **Shaping = limitation**
- Implémenté côté client
- Pour le trafic sortant
- Pour limiter le débit moyen ou pic injecté dans le réseau de l'ISP (respecter SLA)
- Le trafic en excès est mis en attente

- **Policing = régulation**
- Implémenté côté ISP
- Pour le trafic entrant et sortant
- Pour appliquer contrat (SLA):
- marquage ou abandon des paquets quand en excès

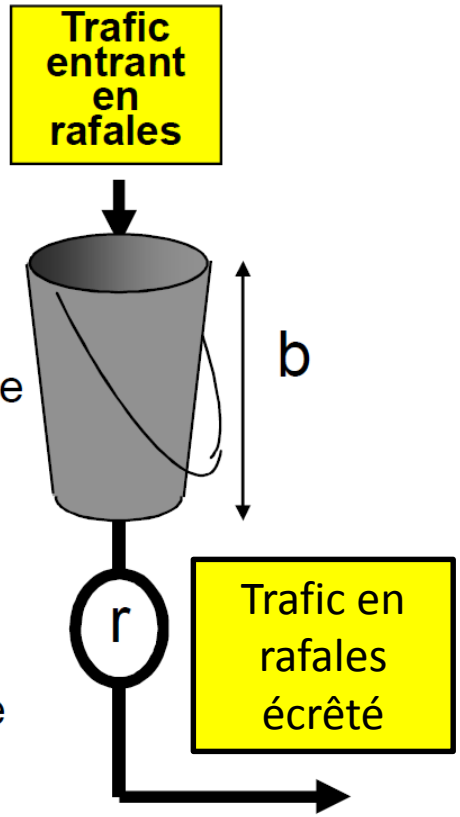


Policing et Shaping : les termes



Leaky bucket (seau percé)

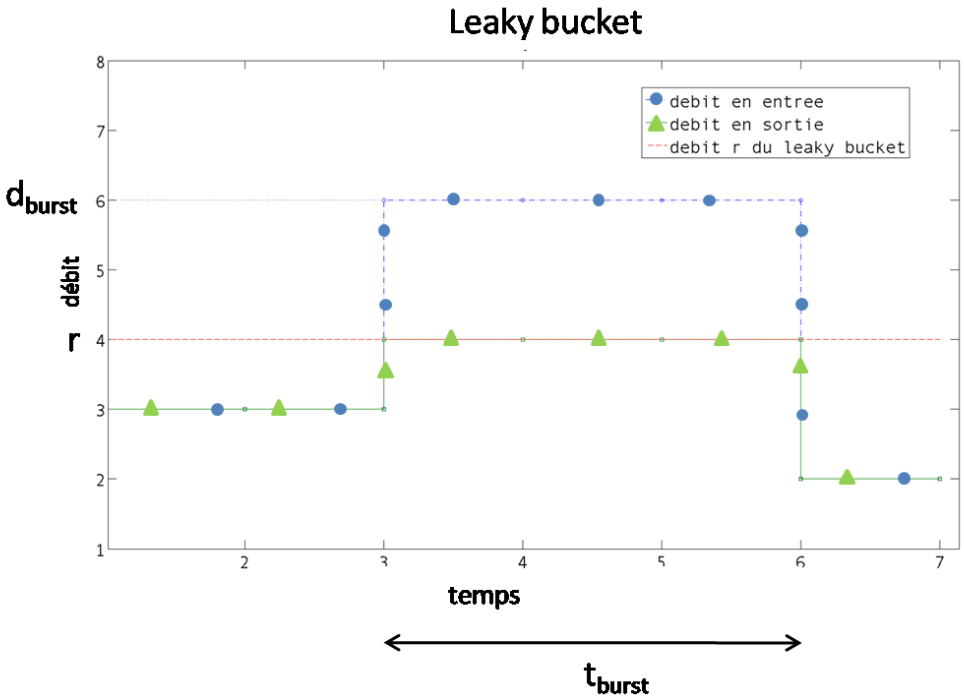
- **Un seau par flux ou classe de trafic**
  - Taille du seau : b octets
  - Le seau fuit au débit de r octets/s
  - Les octets d'un paquet entrant sont placés dans le seau
  - Un paquet qui ferait déborder le seau est rejeté
  - Quand le paquet de taille d est au fond du seau, on le retient pendant d / r avant de l'envoyer
  - La taille du seau b détermine la taille maximale de la rafale
- **Utilisation pour le shaping**
  - Le trafic en entrée est en rafales (bornées par la taille du seau), le trafic en sortie est régulier
- **Utilisation pour le policing**
  - Permet de surveiller le débit moyen (r) et la taille de rafale (b) du trafic entrant
    - Pour chaque paquet conforme, on ajoute d octets dans le seau
    - Un paquet qui ferait déborder le seau n'est pas conforme



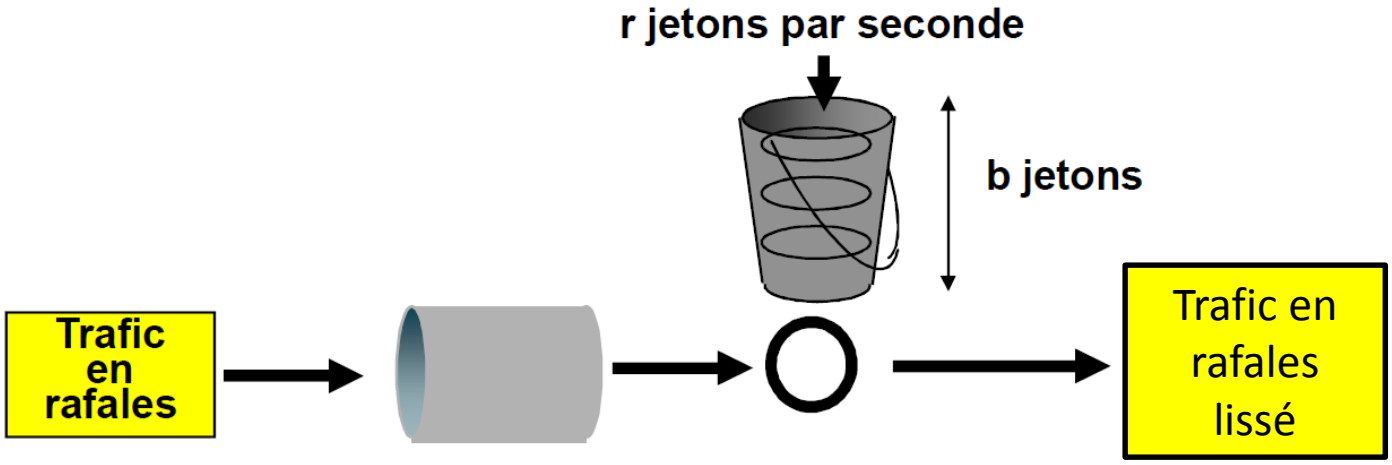
==> **Débit d'entrée écrêté à r.** Permet d'autoriser rafale de taille (debit_raf-r).tps_raf ≤ b bits, ou encore Δr ≤ b/Δt.

Leaky bucket

- ==> **Débit d'entrée écrêté à r.** Permet d'autoriser rafale de taille (debit_raf-r).tps_raf ≤ b bits,
- ou encore Δr ≤ b/Δt.



Token bucket (seau à jetons)



- **On place des jetons dans le seau avec un débit r**
  - La capacité du seau est de **b** jetons
  - Le trafic de données ne passe pas dans le seau
  - Les jetons sont rejetés si le seau est plein
  - Pour émettre un paquet de taille **d** octets, on doit enlever **d** jetons du seau

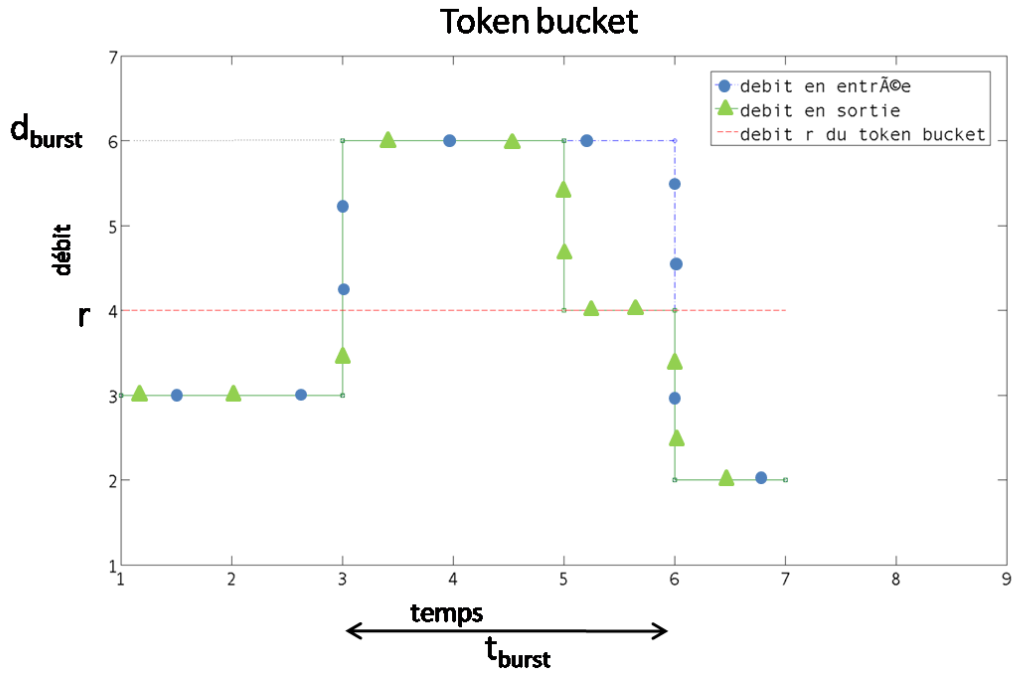
==> Débit de sortie peut dépasser r de Δr pendant Δt pourvu que: $(r+\Delta r)\Delta t \leq r\Delta t+b$, ou encore $\Delta r \leq b/\Delta t$.

Pour recharger le seau de b jetons, il faut redescendre à un débit de (r-x) pendant au moins t sec, (t,x) tel que: $(r-x)t \geq b$

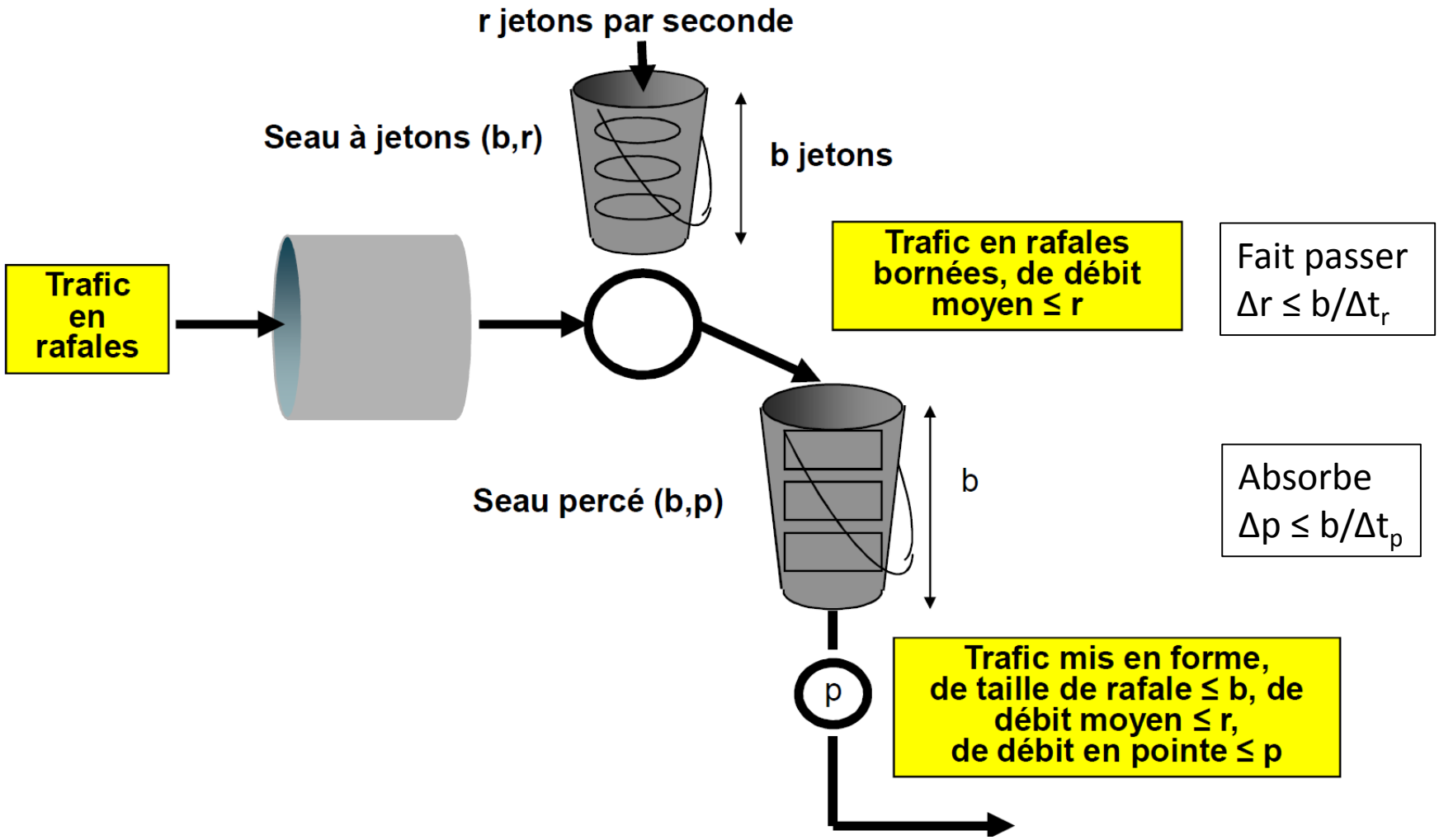
--> **Le débit moyen sur une longue période est toujours ≤ r**.

Token bucket

- ==> Débit de sortie peut dépasser r de Δr pendant Δt pourvu que: (r+Δr)Δt ≤ rΔt+b, ou encore Δr ≤ b/Δt.
- Pour recharger le saut de b jetons, il faut redescendre à un débit de (r-x) pendant au moins t sec, (t,x) tel que: (r-x)t≥b
- --> **Le débit moyen sur une longue période est toujours ≤ r**.



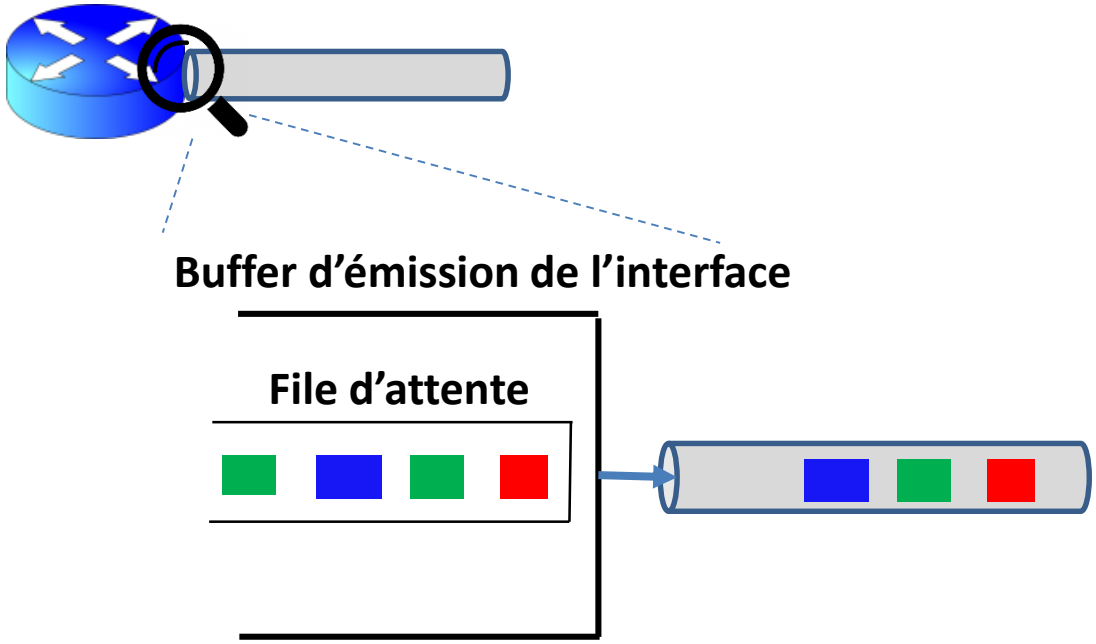
Utilisation combinée de *leaky* et *token buckets*



Plan général du cours

Gestion de file First In First Out (FIFO)

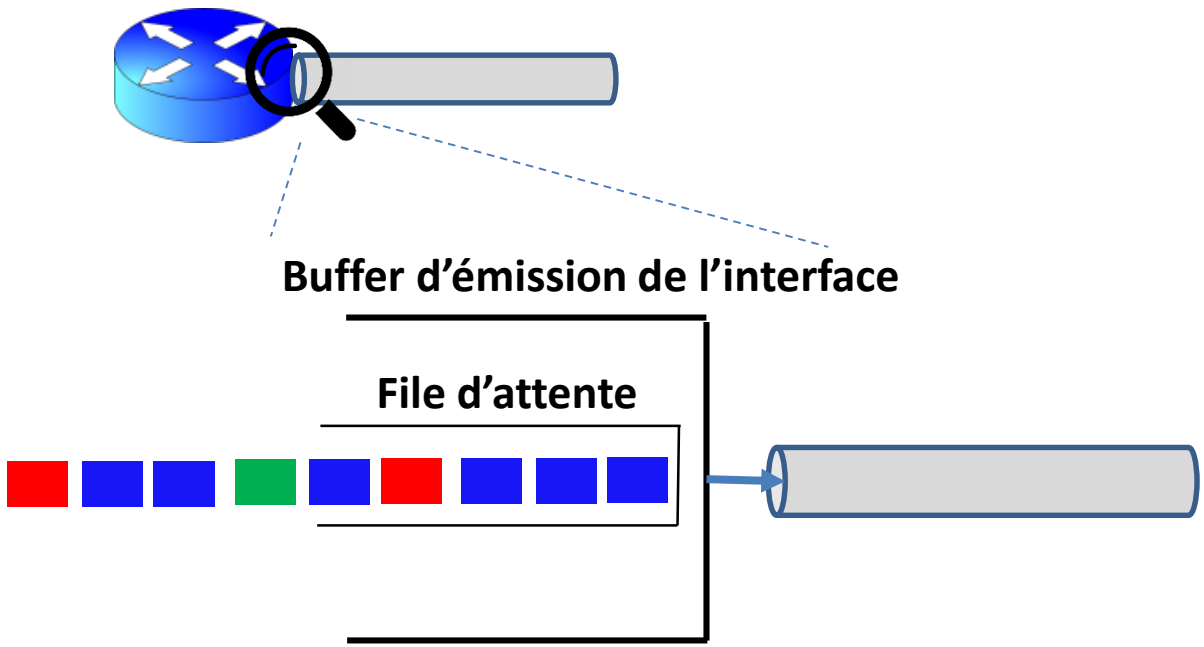
- Premier paquet arrivé, premier paquet sorti



- Fonctionne bien si tous les flux partageant la file ont environ les mêmes débits et contraintes en temps.

Gestion de file First In First Out (FIFO)

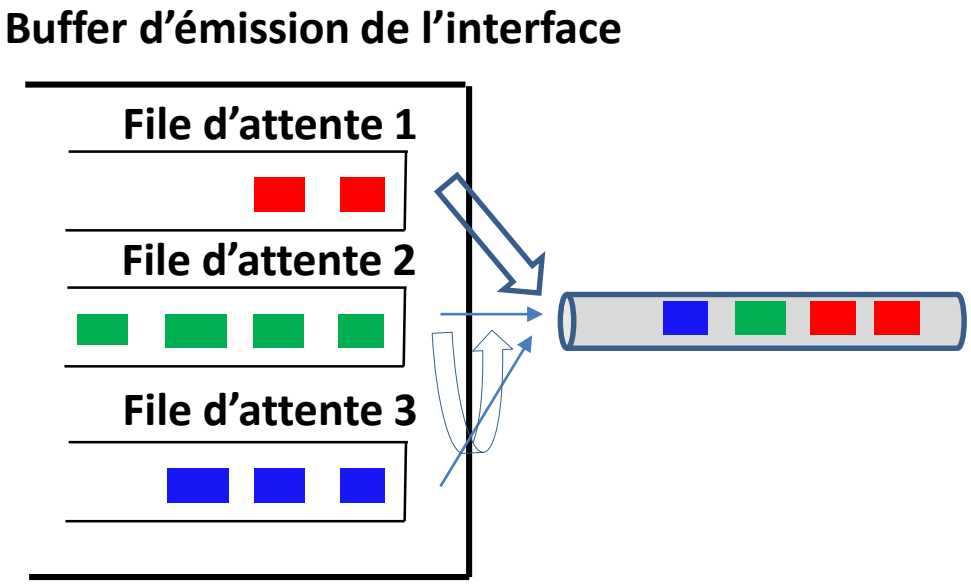
- **MAIS** si certains flux consomment plus de débit et d'autres ont besoin de délais très faibles: problème



- Un transfert de fichier peut monopoliser une grosse partie de la BP, et faire patienter des paquets pressés

Gestion de file à Priorité Stricte

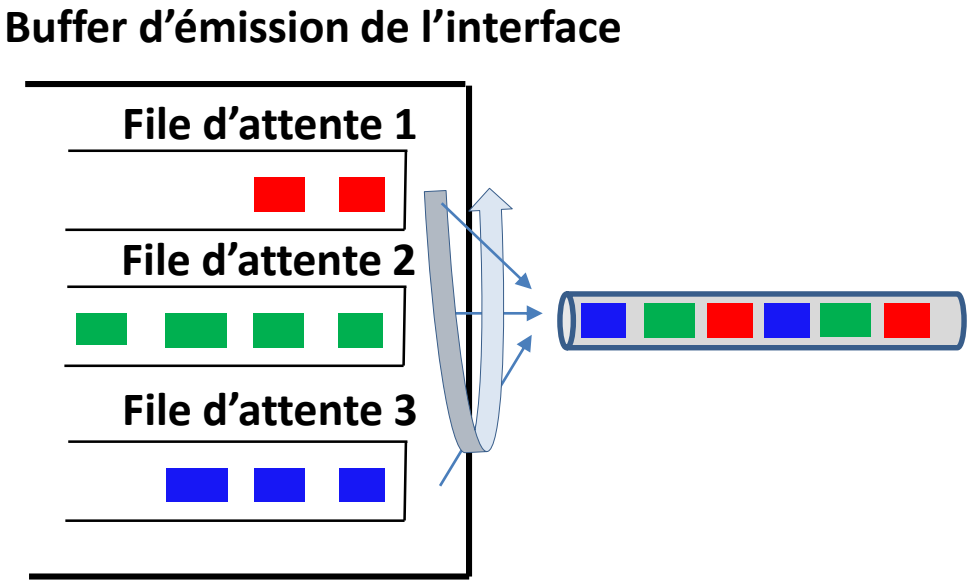
- Paquets orientés dans chaque file grâce à leur marquage de classe (dans entête IP ou Eth)
- Une file n'est servie que si les files plus prioritaires sont vides



- Problème: un flux peut bloquer les autres si une file prioritaire ne se vide pas

Gestion de file *Round Robin (RR)*

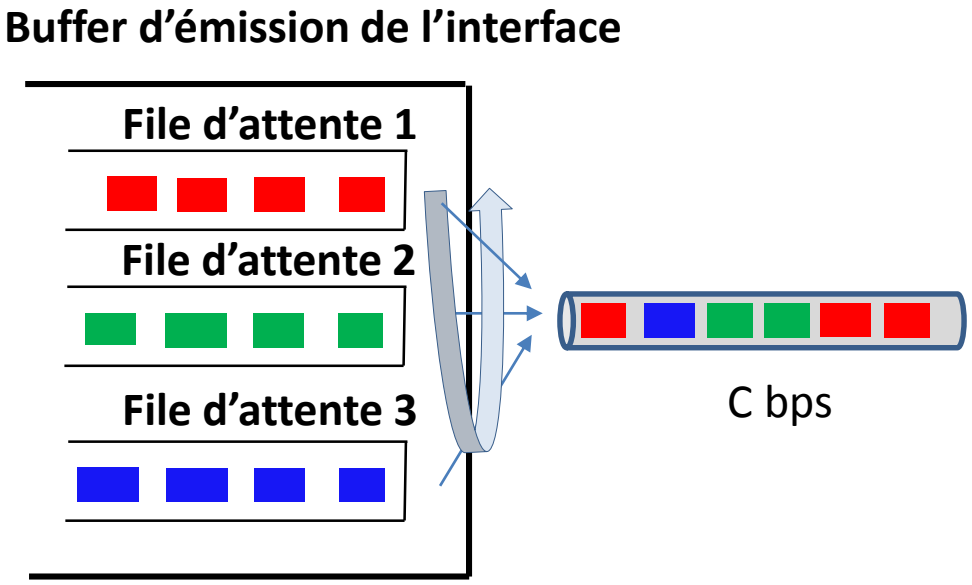
- On sert un paquet de chaque file non-vide à tour de role



- Empêche un flux d'en bloquer un autre

Gestion de file *Weighted Round Robin*

- Si on veut donner plus de priorité (débit) aux paquets d'une file, on sert plus de 1 paquet par cycle :



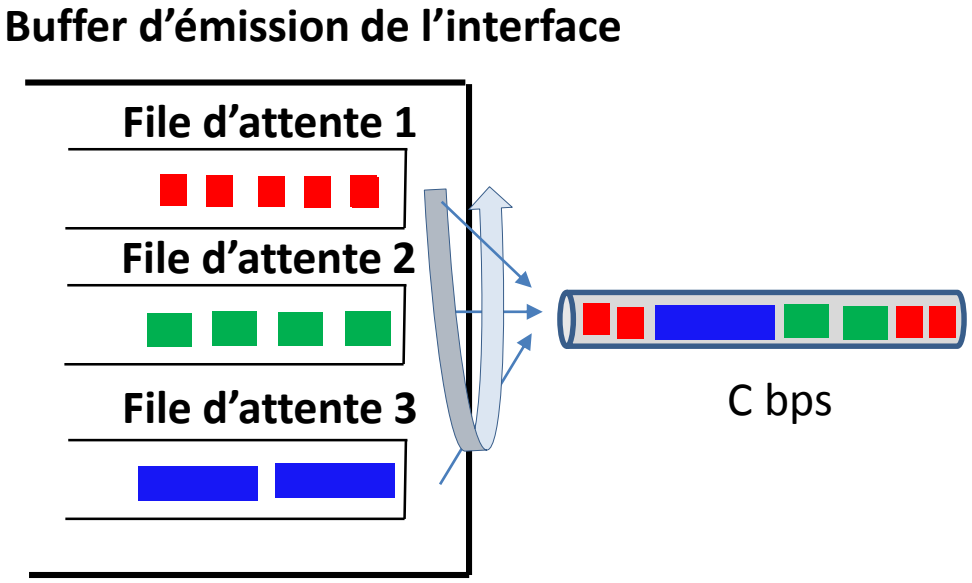
- On détermine le nombre de paquets à sortir directement en fonction du débit affecté à la file i : $D_i = F_i^{debit} C$

→ Fraction de paquets à sortir pour file i : $F_i^{paquet} = F_i^{debit}$

Gestion de file *Weighted Round Robin (WRR)*

- **MAIS** si les paquets n'ont pas la même taille ?

→ Le partage suivant D_i n'est plus valide, les flux ayant des paquets longs reçoivent plus de débit

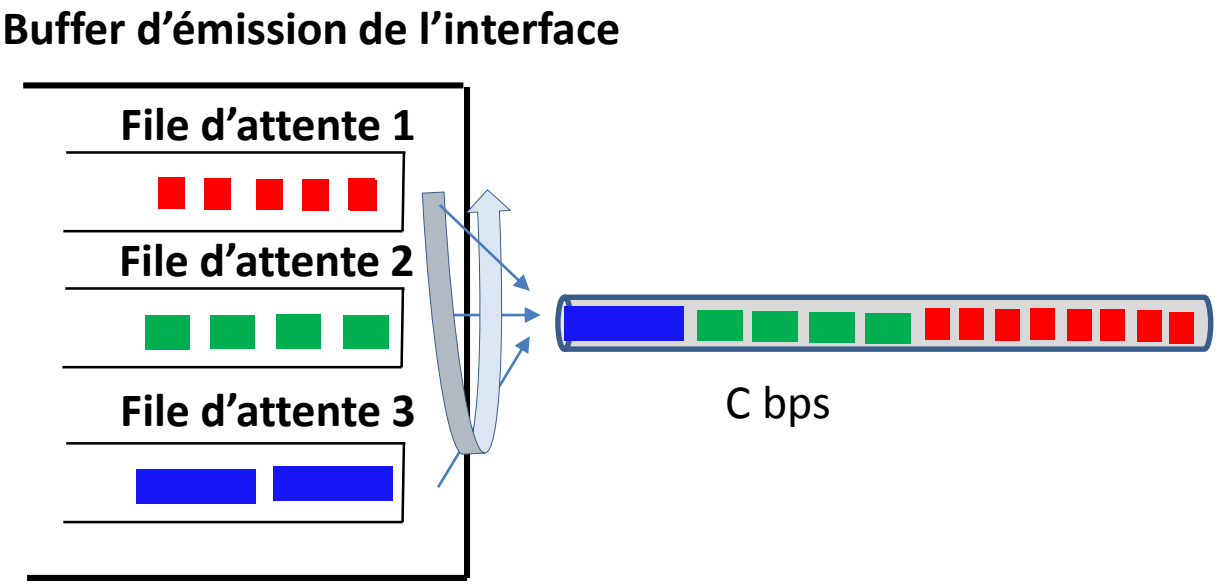


- Débit obtenu par file i = $(F_i^{paquet} S_i)/(F_1^{paquet} S_1 + F_2^{paquet} S_2 + F_3^{paquet} S_3)$ C

≠ Débit visé $D_i = F_i^{paquet}$ C

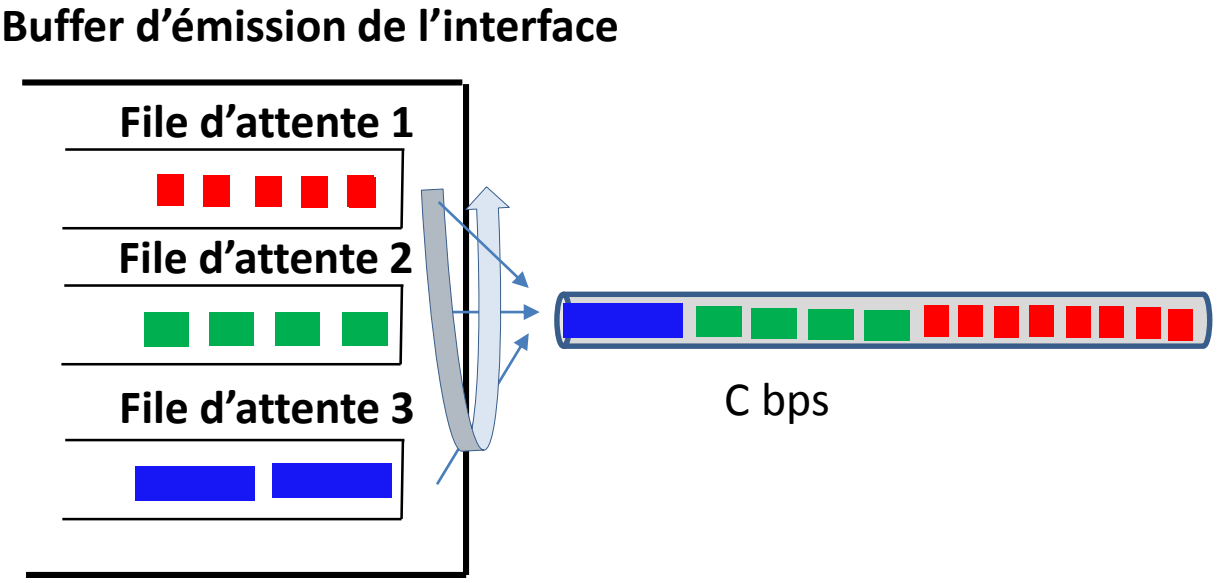
Gestion de file Class Based Weighted Fair Queuing *(CBWFQ)*

- On résout le problème en calculant le nombre de paquets à sortir pour obtenir le débit visé avec : $P_i^{paquet} = F_i^{debit}/S_i \times P_{total}$
- puis $F_i^{paquet} = P_i^{paquet} / (P_1^{paquet} + P_2^{paquet} + P_3^{paquet})$



Gestion de file Content Based Weighted Fair Queuing *(CBWFQ)*

- Ainsi les flux à petits paquets (souvent temps réel) ont une BP effective
- Les flux à gros paquets (souvent transfert/bulk) ne monopolisent pas la BP
- Si on connaît la taille des paquets, on peut aussi contrôler les temps d'attente



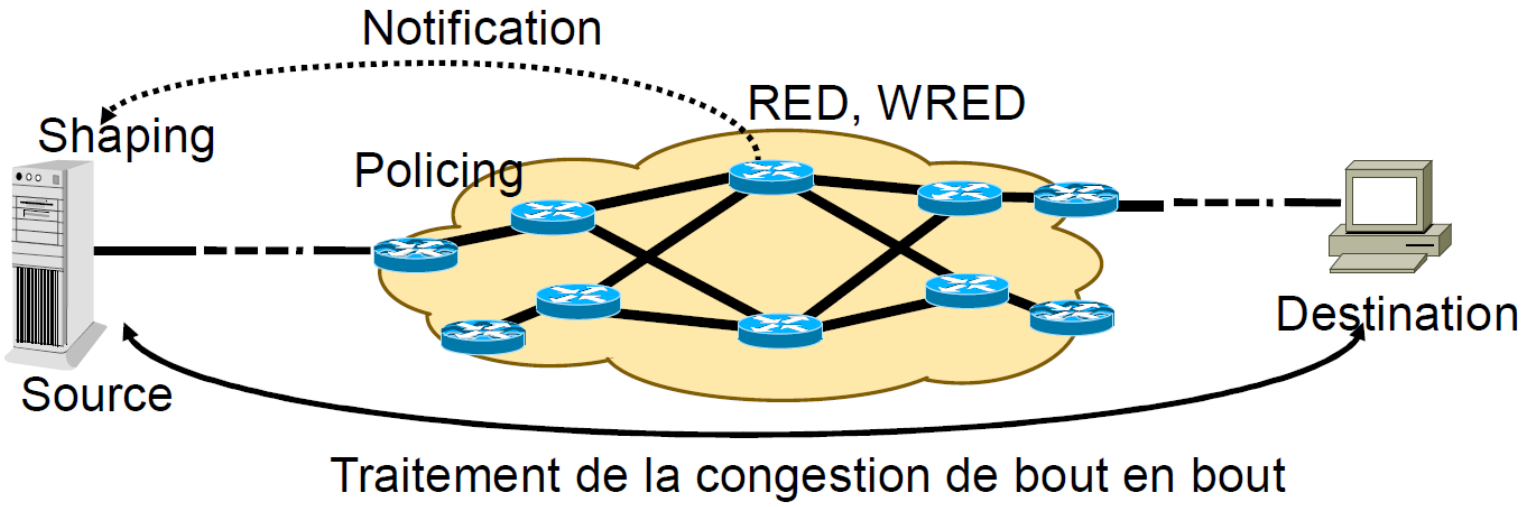
WFQ garantit une borne max du délai

- Dans un réseau où toutes les sessions sont conditionnées aux entrées par des leaky et token buckets et traversent des nœuds avec des files WFQ, alors on peut assurer que le délai de traversée ne dépasse pas une borne fonction de ces paramètres.
- Résultat de 1994 de Parekh et Gallager :
  - Parekh, A. K. and Gallager, R. G., "A Generalized Processor Sharing Approach to flow control in Integrated Services Networks - The Multiple Node Case," IEEE/ACM Transactions on Networking, Vol 2 Issue 1, pp 137-150, April 1994.

Plan général du cours

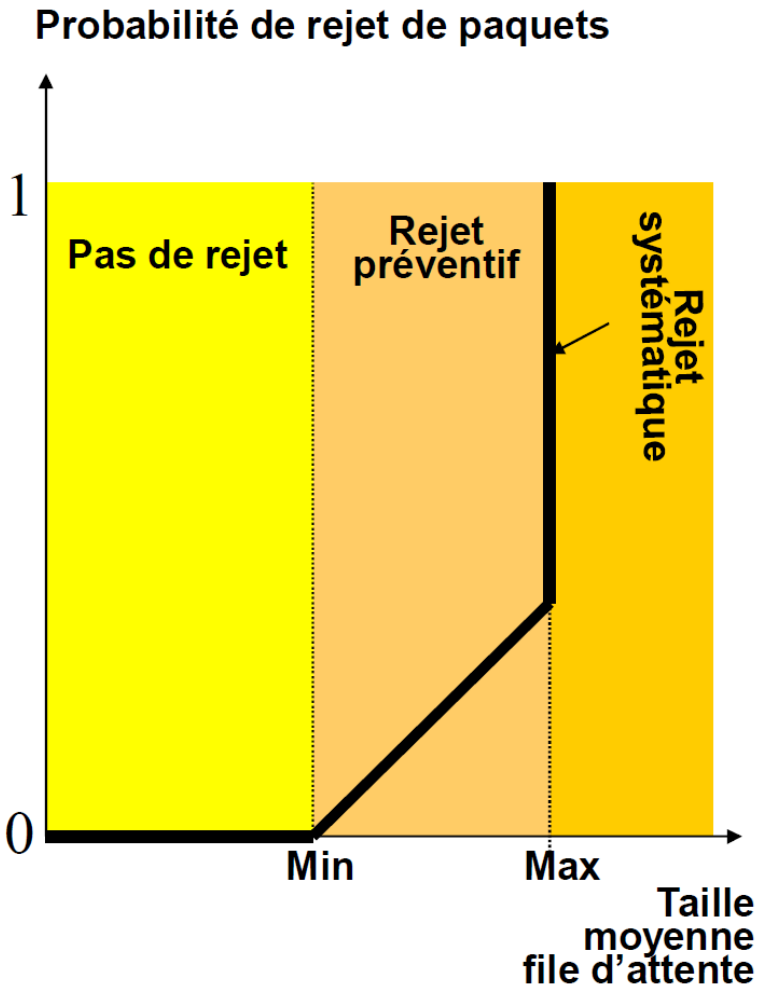
Les outils de traitement de la congestion

- **La congestion a un impact important sur des paramètres clés de la QoS**
  - Le taux de pertes de paquets
  - Le délai et la variation de délai
- **3 moyens de traiter la congestion**
  - Préventivement par *shaping* / *policing* (voir précédemment)
  - De façon curative de bout en bout (par TCP)
  - Préventivement par les équipements réseau : RED et WRED
    - Avec ou sans notification explicite à la source



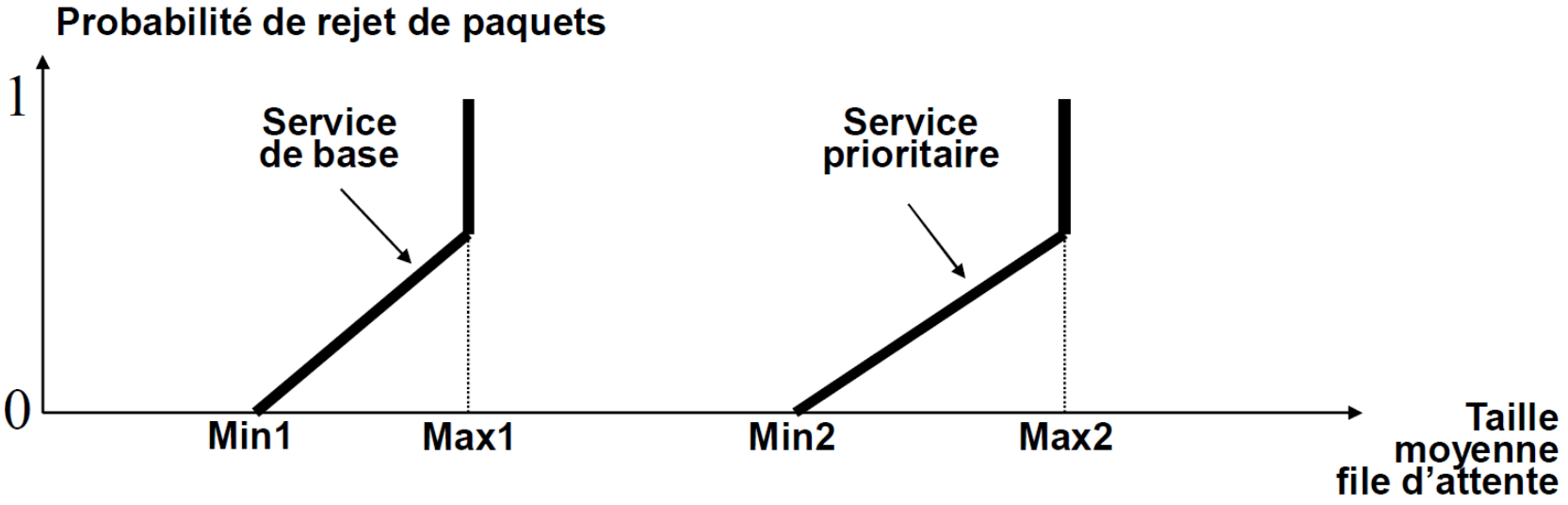
Random Early Detection (RED)

- **Le traitement de la congestion par TCP est curatif**
  - Il provoque la synchronisation des connexions TCP
    - *Slow-start* synchronisés des différentes connexions TCP
- **RED permet un traitement préventif**
  - Un paquet perdu est un signal de congestion pour TCP, qui provoque un ralentissement (*congestion avoidance*)
  - On rejette des paquets après un seuil mais avant la congestion
    - Rejet aléatoire entre les différentes connexions
    - Ce qui lisse la taille de la file d'attente tout en évitant la synchronisation globale
  - RFC 2309



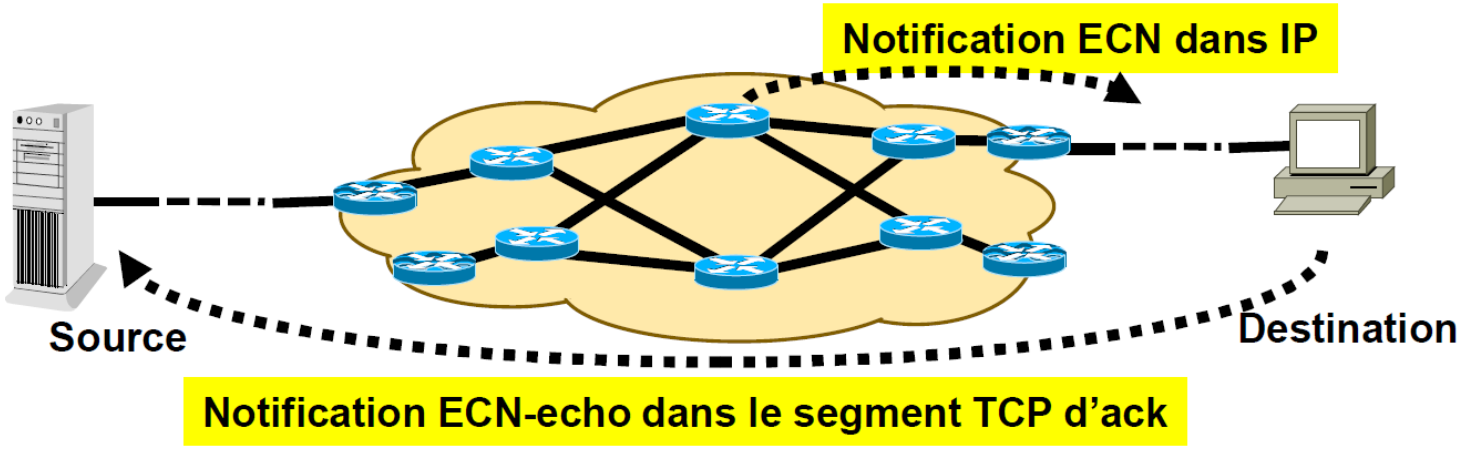
Weighted RED

- **Les seuils à partir desquels on commence à rejeter des paquets peuvent être différents selon les classes de service**
  - Exemple : 30% de la taille du buffer pour le service de base, et 60% pour le service prioritaire (priorité en cas de congestion)
  - La priorité peut être marquée dans le champ ToS /DS
- **RED est équitable et simple, mais efficace seulement avec les trafics TCP**



Notification explicite de congestion

- **Utilisation conjointe de RED et ECN (Explicit Congestion Notification)**
  - RFC 2481
  - Un routeur RED positionne le bit ECN au lieu de jeter le paquet
    - Le bit ECN est recopié dans l'entête TCP d'ack (ECN-echo)
    - Un seul paquet marqué ECN provoque une réaction à la congestion
  - Un 2e bit peut indiquer si TCP est compatible ECN
    - Selon ce bit le marquage RED est un rejet ou le bit ECN



Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

II.1. TCP et le contrôle de congestion dans le réseau

- II.1.a. Principe du contrôle de congestion
- II.1.b. Rappels : transfert fiable et contrôle de flux
- II.1.c. Le contrôle de congestion par TCP

II.2. Classification des applications et besoin de QoS

- II.2.a. Définition de la QoS et classes d'applications
- II.2.b. Streaming video sur HTTP : s'adapter en Best Effort, sans garantie de QoS
- II.2.c. Stratégies possible pour garantir un niveau de QoS

II.3. Techniques de traitement de la QoS

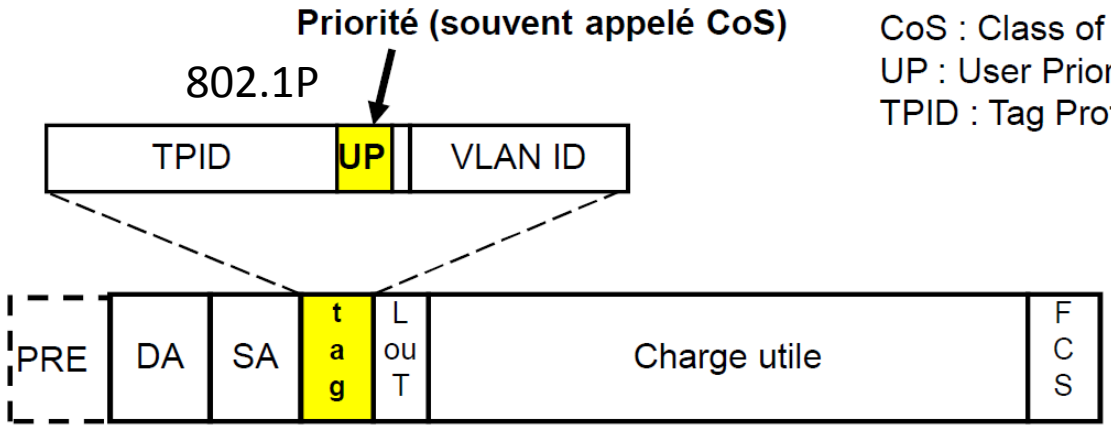
- II.3.a. Conditionnement du trafic à l'entrée dans l'ISP : shaping et policing
- II.3.b. Gestion de file d'attente : ordonnancement pour faire sortir les paquets
- II.3.c. Gestion de buffer : comment abandoner les paquets en excès
- II.3.d. Marquage du trafic : DiffServ dans IP

III. Commutation par circuits virtuels

IV. Commutation par circuits virtuels dans le mode IP : MPLS

802.1p : Priorité des trames Ethernet

- **La gestion des priorités du commutateur nécessite que les trames véhiculent une information de priorité**
- **Les trames Ethernet natives n'ont pas d' indication de priorité**
- **Une étiquette (tag) 802.1Q doit être ajoutée**
  - Cette étiquette comporte un champ priorité (3 bits)
  - Le tag peut être inséré par le premier commutateur selon des critères prédéfinis, ou par une station digne de confiance
    - En général une file d'attente de priorité stricte : téléphonie et vidéo
    - Et des files d'attente standard (souvent de type RRQ) qui empêchent les flux de se cannibaliser



CoS : Class of Service
UP : User Priority
TPID : Tag Protocol ID

802.1p : Priorité des trames Ethernet

- **Le tableau suivant donne un exemple d'utilisation du champ CoS**
  - Les commutateurs sont souvent limités à quatre files d'attente de sortie
  - Files d'attente de type Priority Queuing ou Round Robin Queuing
    - Mais les noms différent selon les constructeurs
    - Exemple SRR (Shaped Round Robin Queuing) chez Cisco

Priorité	Nom	Type de trafic
7	Network control	Administration de réseau
5 et 6	Vidéo et audio	Voix ou multimédia (priorité stricte)
3 et 4	Excellent effort et Controlled load	Niveaux de priorité intermédiaires
2	Reserved	Réservé
1	Background	Trafic bulk non urgent
0	Best effort	Trafic LAN standard

Principe de DiffServ

- Dans l'Internet traditionnel, tous les flows sont traités de la même façon
- Avec ATM, on peut offrir des garanties en terme de QoS au prix de plus de signalisation, et de décisions d'acceptation/rejet de nouveaux flows pour maintenir le service des flows présents.
- En effet, ressources limitées -> besoin de rejeter des flows pour maintenir le niveau de service

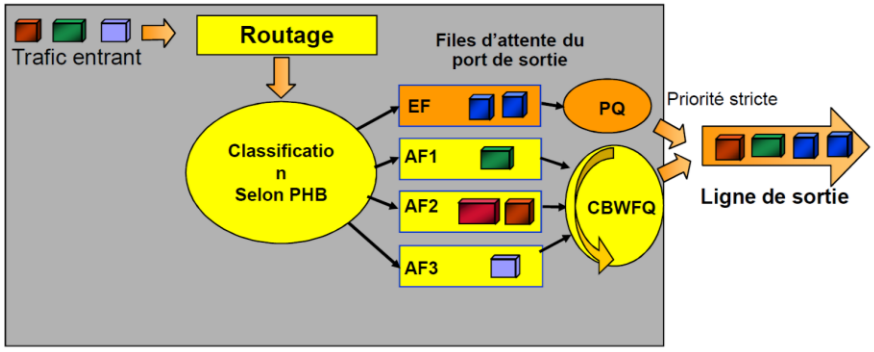
-> contraire à la nature *Best-effort* de l'Internet où il n'y a pas de contrôle d'admission

- Mais important de différencier les classes de flows : les applis le nécessitant pourraient avoir un plus haut débit ou un plus petit délai (appli temps-réel telles que voix, vidéo,…).

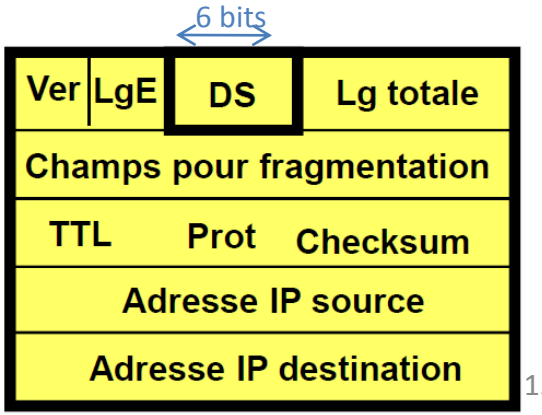
-> introduction de Diffserv : basé sur le marquage des paquets aux bords du réseau selon le niveau de performance que le SP veut leur fournir, et traitement différencié de ces paquets dans le coeur du réseau.

Principe de DiffServ

- 4 classes de flows définies, et les paquets de flows de classes différentes mis dans des files différentes
- Différenciation supplémentaire des paquets dans la même classe : 3 niveaux intra-classe pour les paquets d'un même flow. La combinaison d'une classe et d'un niveau dans la classe est appelée ou *Per-Hop Behavior – PHB*.
- Chacun des 12 PHB correspond à un *code point*, affecté au paquet à son entrée dans le réseau.
- Un niveau intra-classe est aussi appelé *drop precedence* : probabilité d'abandon fonction de ce niveau si saturation de la file affectée à cette classe (par ex les paquets de synchro TCP).



CBWFQ : Class-Based Weighted Fair Queing

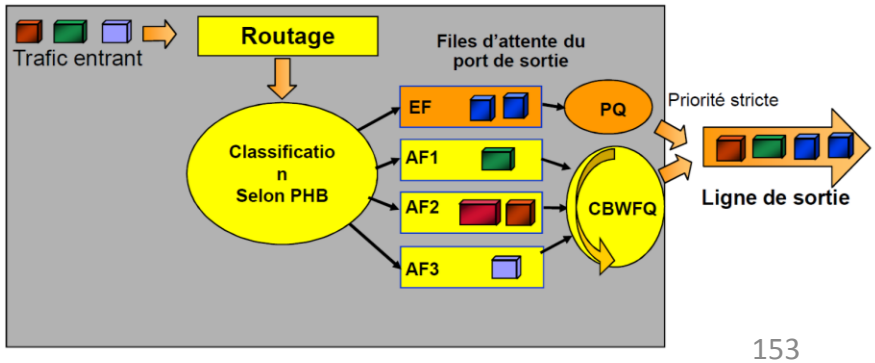


Mode de fonctionnement de DiffServ

L'architecture DiffServ a 2 composants :

1. Routeur de bord de l'ISP :
   - Politiques d'affectation de priorités aux paquets : décidées par le SP sur la base du SLA passé avec le client. Dépend notamment du comportement de la source du flow (quel est son débit courant par rapport à son débit moyen et de crête prévus dans le SLA).
   - Assignation du DiffServ code point (DSCP) à chaque paquet.
2. Routeurs de coeur : appliquer des traitement différents aux paquets en fonction de leurs DSCP.

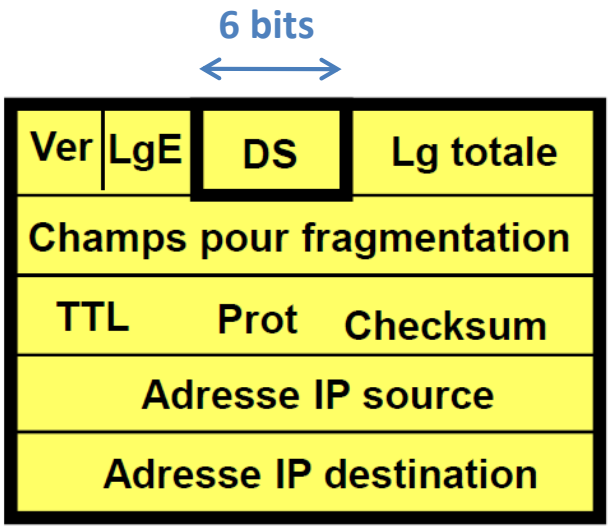
   Traitement différents: en termes de décisions d'ordonnancement et d'abandon.

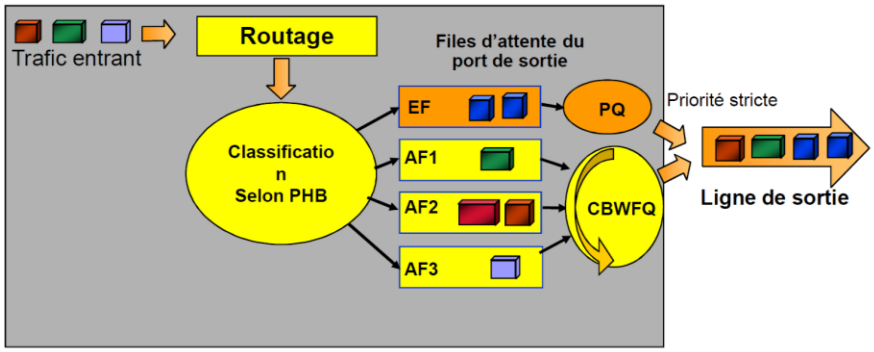


CBWFQ : Class-Based Weighted Fair Queing

Format du champ DS

- DSCP = 0 associé au Best-effort
- 1 valeur de DSCP associée au PHB EF (Expedited Forwarding)
- 9 valeurs de DSCP associées aux PHB AF (Assured Forwarding)





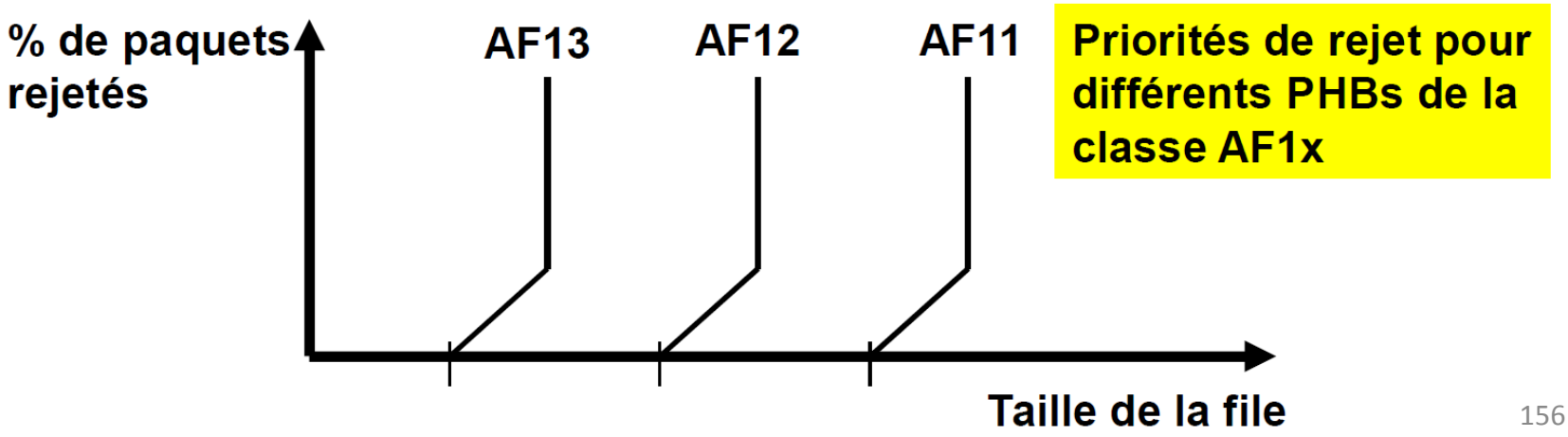
CBWFQ : Class-Based Weighted Fair Queing

Le PHB EF (Expedited Forwarding)

- **Le PHB EF (appelé auparavant premium) est conçu pour des flux avec**
  - Faible taux de perte, faible délai, faible gigue et bande passante garantie
  - Similaire à une liaison spécialisée virtuelle de débit de pointe connu entre deux points extrémité à travers un domaine DS
  - Valeur du code DSCP pour PHB EF = 101110
- **Pour minimiser délai et gigue, les paquets doivent perdre le moins de temps possible dans les files des routeurs**
  - Conditionnement (policing et shaping) du trafic en fonction du débit de pointe en bordure d'un domaine
  - Dimensionnement du réseau de telle sorte que le débit de pointe soit inférieur aux débits des liaisons de sortie de chaque routeur dans le réseau
  - Paquets rangés dans une file de sortie à haute priorité
    - PQ avec priorité fixe préemptive ou WFQ avec un poids important
    - Mais aucun mécanisme particulier n'est imposé
- **RFC 3246 et RFC 3260**

Le PHB AF (Assured Forwarding)

- Pour déterminer les niveaux intra-classe :
  - soit un leaky ou token bucket -> 2 niveaux (paquets in et paquets out)
  - soit CIR et PIR -> 3 niveaux



Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

III. Commutation par circuits virtuels

III.1. Commutation par circuit et Commutation par paquet

III.2. Commutation par circuit virtuel : une combinaison des deux

III.3. Avantage de la commutation par VC pour la QoS

IV. Commutation par VC dans le mode IP : MPLS

2 mondes sur les câbles : le réseau téléphonique et le réseau Internet

- Il n'y avait au début de l'Internet qu'une infrastructure physique : celle déployée au XX^{ème} siècle pour les réseaux téléphoniques.
- Cette infrastructure peut donc transporter à la fois le trafic Internet et le trafic téléphonique.
- Mais le réseau téléphonique et le réseau Internet ont 2 fonctionnements radicalement différents.
- Principe du réseau téléphonique : une qualité constante est assurée par communication =>le nombre de communications simultanées passant par un câble est limité – Contrôle d'accès au réseau
- Principe de l'Internet : la qualité n'est pas garantie, elle dépend du nombre d'utilisateurs – Pas de contrôle d'accès

Rappel : le réseau téléphonique

- Différentes façon de résoudre P2.
- **Définition** : la commutation est un terme général désignant le mode de transfert de l'information, sous forme de signaux ou paquets, entre l'entrée et la sortie d'un équipement traversé.

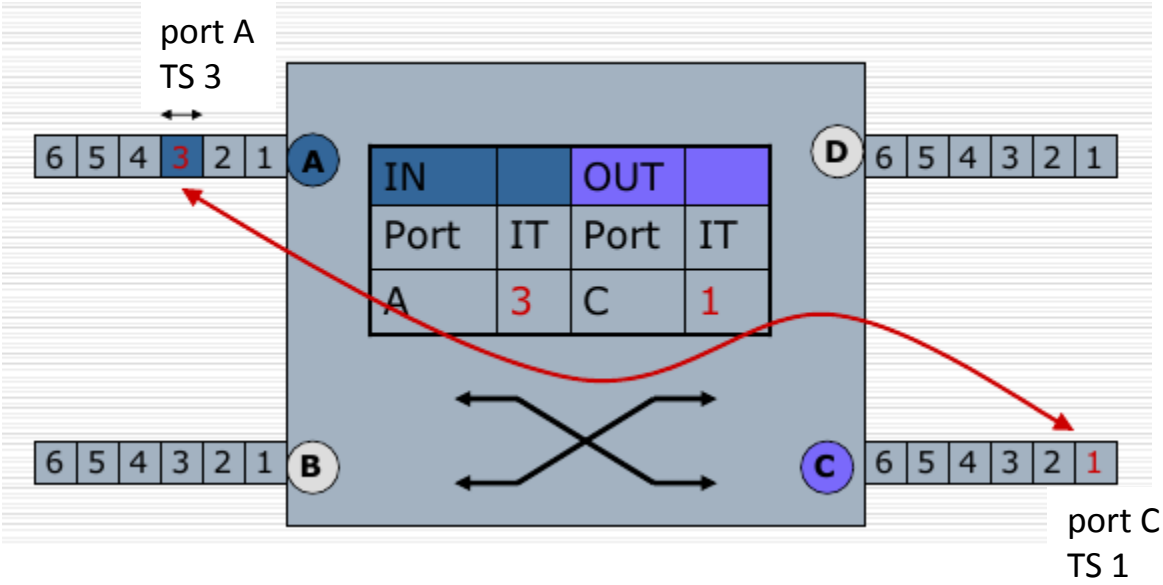
- **Commutation spatiale** : la mise en relation entre l'entrée sur laquelle le signal arrive dans l'équipement et la sortie sur laquelle il est transféré se fait spatialement.

Opérateurs humains

Crossbar switch

La commutation temporelle

- **Commutation temporelle** : la mise en relation entre l'entrée sur laquelle le signal arrive dans l'équipement et la sortie sur laquelle il est transféré se fait temporellement.
- Un octet arrivant sur un certain numéro de TS dans la trame entrante est recopié sur un autre numéro de TS dans la trame sortante



La commutation de circuit

- **Commutation de circuit** : action d'établir l'équivalent d'un circuit électrique physique **dédié** entre 2 abonnés

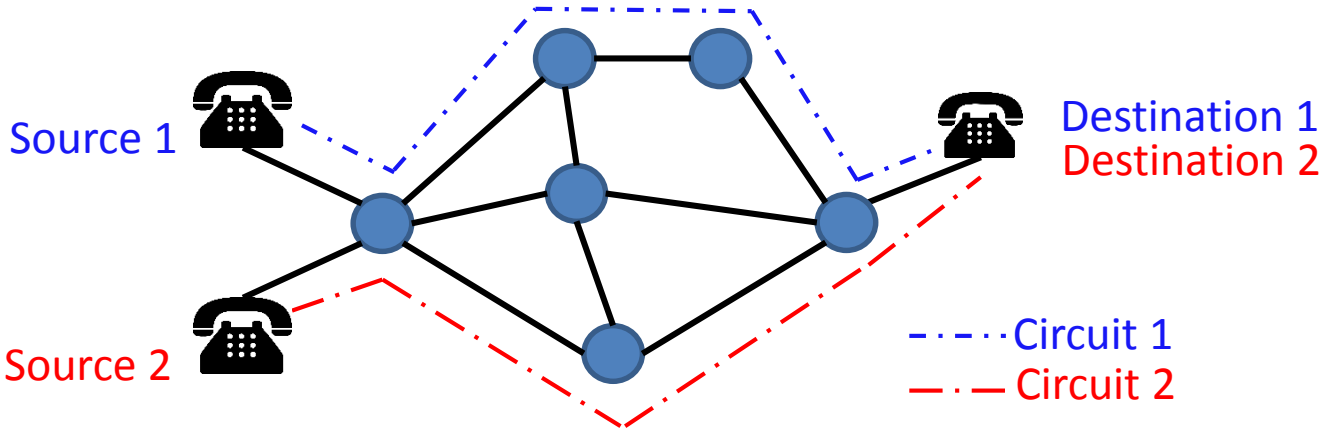
Contraintes d'une com. tél.

- --> le débit disponible est constant
- --> le délai d'attente dans les équipements est nul

- 2 solutions pour la réalisation de la commutation de circuit :
  - par commutation spatiale : une continuité électrique est établie (d'où le nom)
  - par commutation temporelle : avec **réservation** d'un TS dans chaque trame en sortie d'un commutateur pour une communication téléphonique

Commutation de circuit

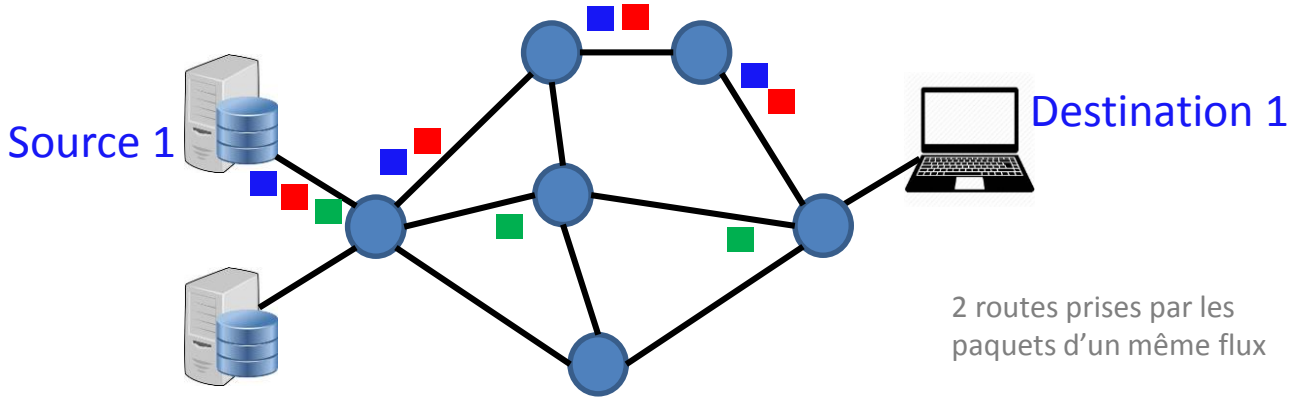
- Commutation de circuit (circuit switching) :
  - chemin unique établi avant la transmission des données, restant fixé pendant la communication
  - ressources réservées à chaque nœud traversé (paire de fils de cuivre, ou time slot)



→ Inefficacité dans le cas des rafales de trafic: cas des données informatiques

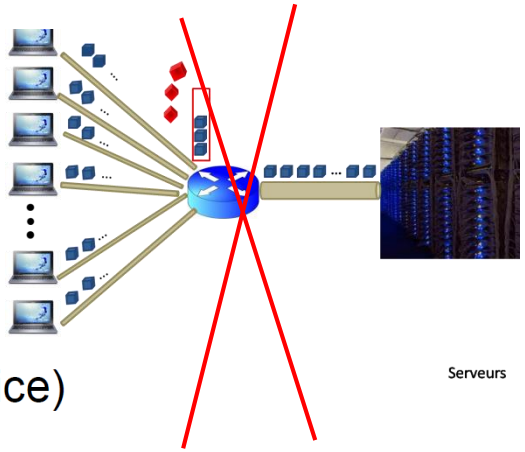
Commutation de paquets

- Paquet (ou *datagram*) : ensemble d'octets transférés ensemble, comme une unité faisant partie d'un flux (*flow*)
- Commutation de paquets:
  - pas d'établissement d'un chemin fixé avant transmission des paquets de données
  - donc pas de réservation de ressources



→ Bien adapté aux trafic variables (le routeur envoie dès que possible le paquet sur le lien de sortie)

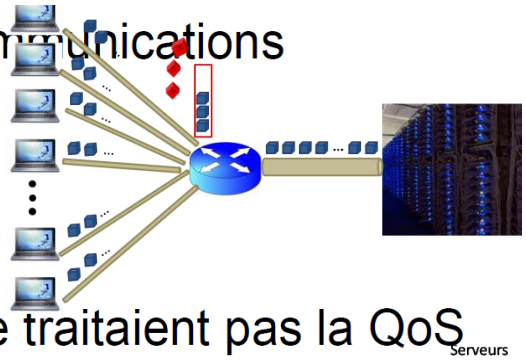
Commutation de circuits ou commutation de paquets ? Avantages et inconvénients



- **Avantages de la commutation de circuits**
  - Simple : pas de congestion et de priorités à gérer
  - Offre le meilleur niveau possible de QoS (Qualité de Service)
    - Réservation statique : bande passante garantie
    - Temps de traversée du réseau très court et presque fixe
    - Adapté de façon native aux trafics isochrones (téléphonie, vidéo)
- **Inconvénients de la commutation de circuits**
  - Mal adaptée aux trafics sporadiques et de débit variable
    - Ressources du réseau mal utilisées
    - Coûts supérieurs pour l 'opérateur donc pour l'utilisateur
    - Les gâchis deviennent prohibitifs à haut débit (services large bande : ex. vidéo)
    - Le trafic de données est le plus souvent sporadique
    - Le trafic vocal devient à débit variable avec la compression des silences
    - Le trafic vidéo avec compression génère naturellement un débit variable
  - Problème d'évolutivité
    - Le débit des circuits n'est pas configurable dynamiquement

Commutation de circuits ou commutation de paquets ?

- **Avantages de la commutation de paquets**
  - Les lignes sont partagées entre les différentes communications
  - Meilleure utilisation du réseau
  - Plus adapté aux trafics sporadiques et variables
- **Inconvénients de la commutation de paquets**
  - Les réseaux initiaux à commutation de paquets ne traitaient pas la QoS
  - Bande passante non garantie (fluctuante selon l'état de congestion)
  - Délais non garantis, plus longs et plus variables



- **Objectif : prendre le meilleur des 2 mondes de la commutation de circuits et de la commutation de paquets**
  - Partager les lignes entre plusieurs trafics
    - Grâce à des réseaux à commutation de paquets
  - Assurer des garanties de qualité de service variables selon les applications, réparties en classes de service
    - Garanties de débit
    - Temps de transit et gigue bornés ou au moins réduits

Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

III. Commutation par circuits virtuels

III.1. Commutation par circuit et Commutation par paquet

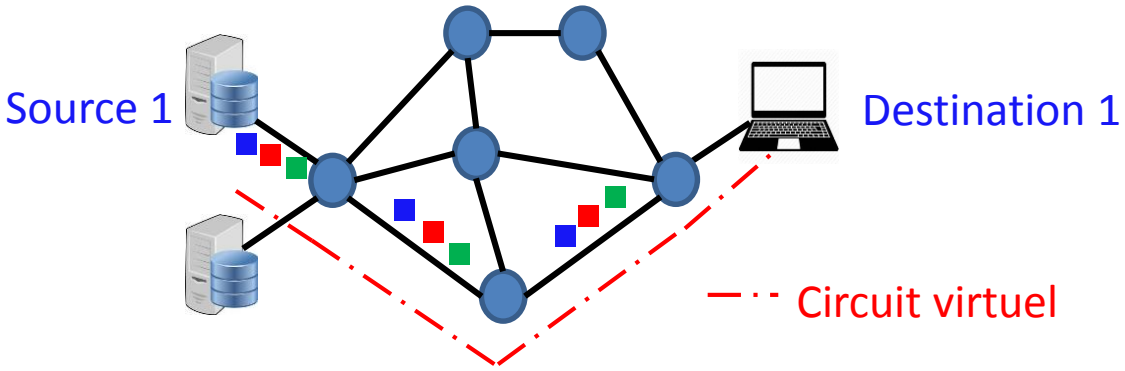
III.2. Commutation par circuit virtuel : une combinaison des deux

III.3. Avantage de la commutation par VC pour la QoS

IV. Commutation par VC dans le mode IP : MPLS

Solution hybridant commutation de circuit et commutation de paquet :

- Commutation de paquets avec connexion
- = Commutation par **circuit virtuel**:
  - établissement d'un chemin fixé avant transmission des paquets de données
  - Mais pas de réservation de ressources

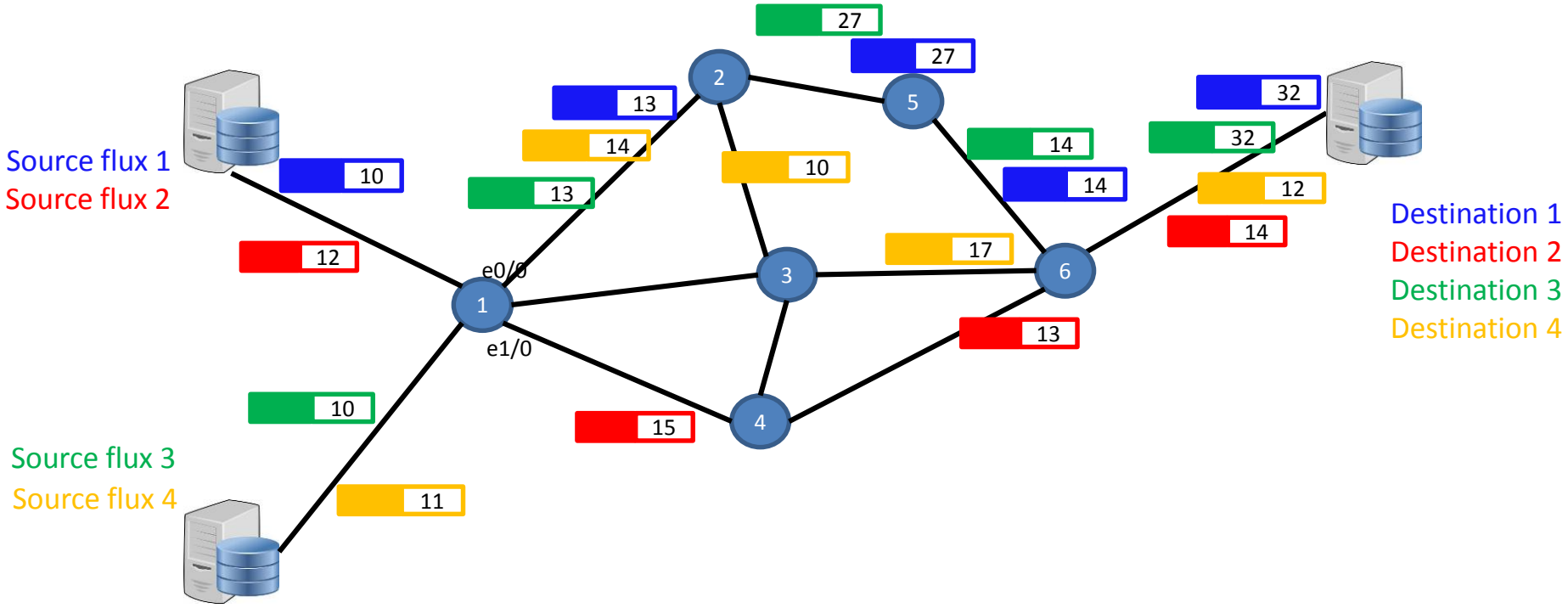


→ Pas de réservation: pas de gaspillage

→ Chemin fixé: possibilité de donner des priorités pour assurer QoS à flux sensibles

Commutation par VC: implémentation

- La commutation se fait par label (opposé à adresse): signification locale et non globale



Switching table of Node 1

Label d'entrée	Port de sortie	Label de sortie
10	e0/0	13
11	e0/0	14
12	e1/0	15

Switching table of Node 2

Label d'entrée	Port de sortie	Label de sortie
13	e1/0	27
14	e2/0	10

Switching table of Node 6

Label d'entrée	Port de sortie	Label de sortie
13	e0/0	14
14	e0/0	32
17	e1/0	12

Plan général du cours

I. Organisation des opérateurs de l'Internet

II. TCP et Qualité de service

III. Commutation par circuits virtuels

III.1. Commutation par circuit et Commutation par paquet

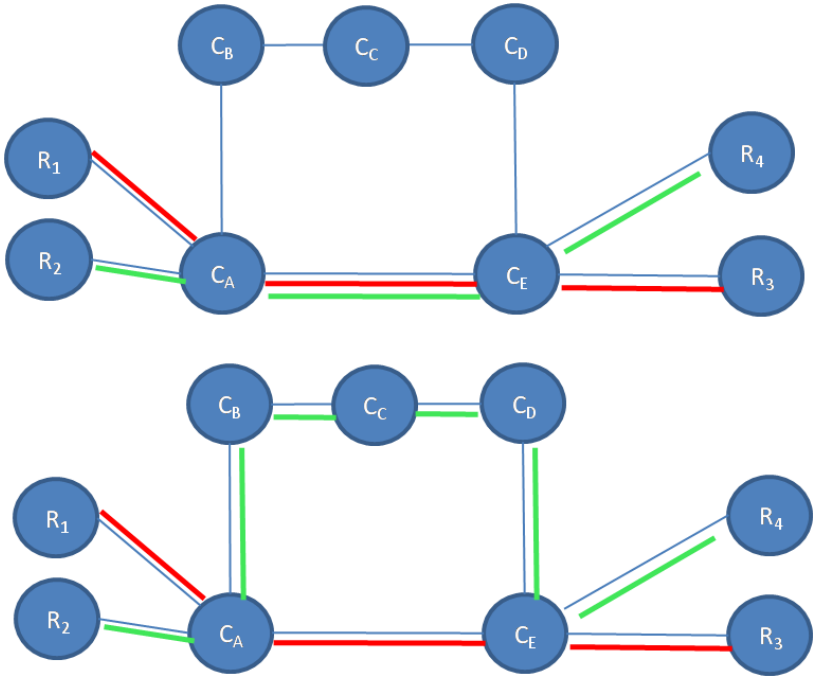
III.2. Commutation par circuit virtuel : une combinaison des deux

III.3. Avantage de la commutation par VC pour la QoS

IV. Commutation par VC dans le mode IP : MPLS

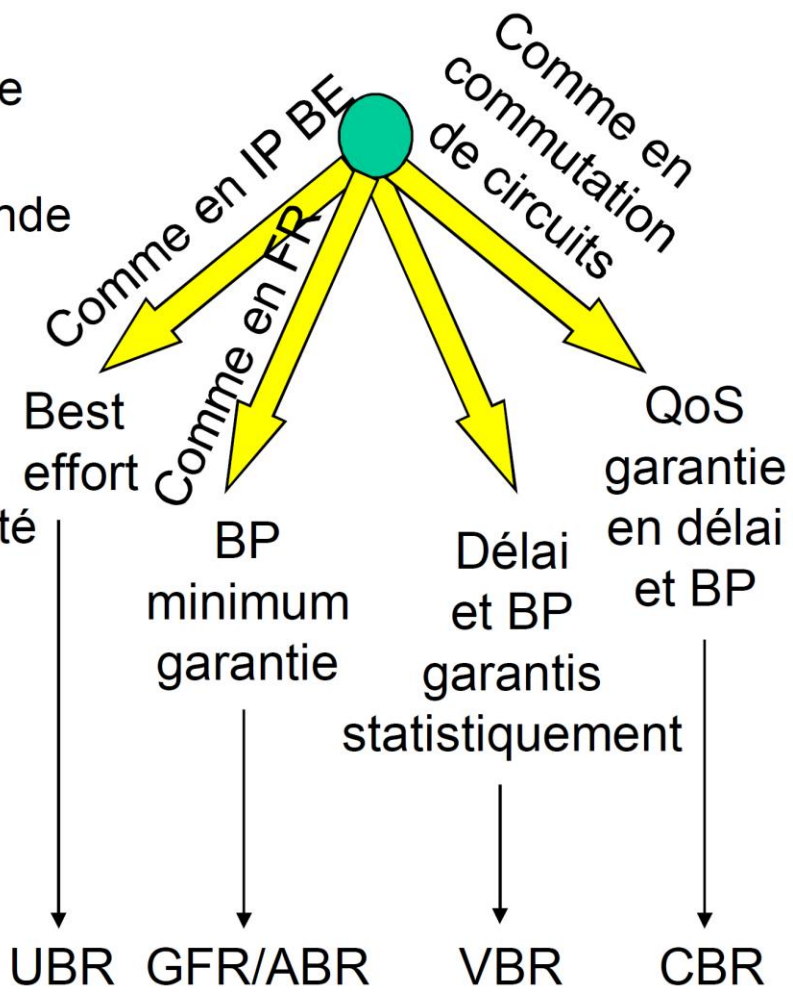
Le lien entre QoS et VC

- Le mode VC permet de choisir une route fixe, que vont emprunter l'ensemble des paquets entre une source et une destination.
- Donc en choisissant les bons commutateurs, on peut assurer un qualité (un débit min, un délai max) de bout en bout.
- Une fois qu'une route est fixée, on peut fixer les autres en fonction pour satisfaire les contraintes de qualité.



Catégories de service

- **Chaque circuit virtuel aura la catégorie de service adaptée aux besoins de l'application qui l'utilisera**
  - Garantie absolue en délai et en bande passante : CBR
  - Garantie statistique en délai et en bande passante : VBR
  - Besoin d 'un minimum de bande passante (applications élastiques) : GFR ou ABR
  - Pas de traitement spécifique de qualité de service (« best effort ») : UBR

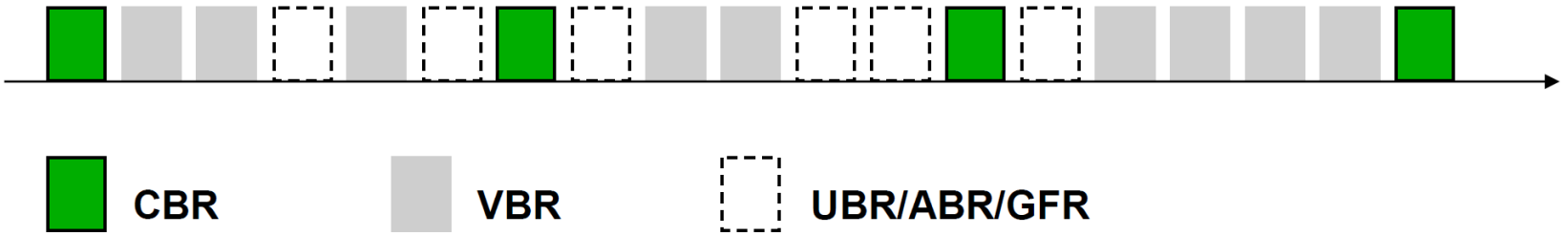


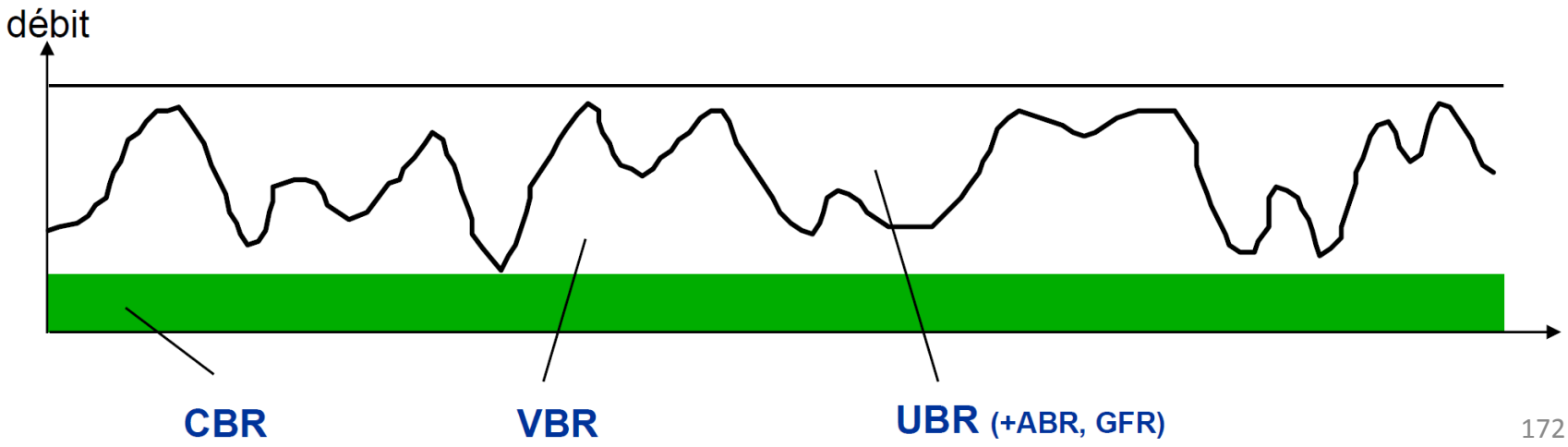
CBR : Constant Bit Rate
VBR : Variable Bit Rate
ABR : Available Bit Rate
GFR : Guaranteed Frame Rate
UBR : Unspecified Bit Rate

ATM-TM: Classes de service

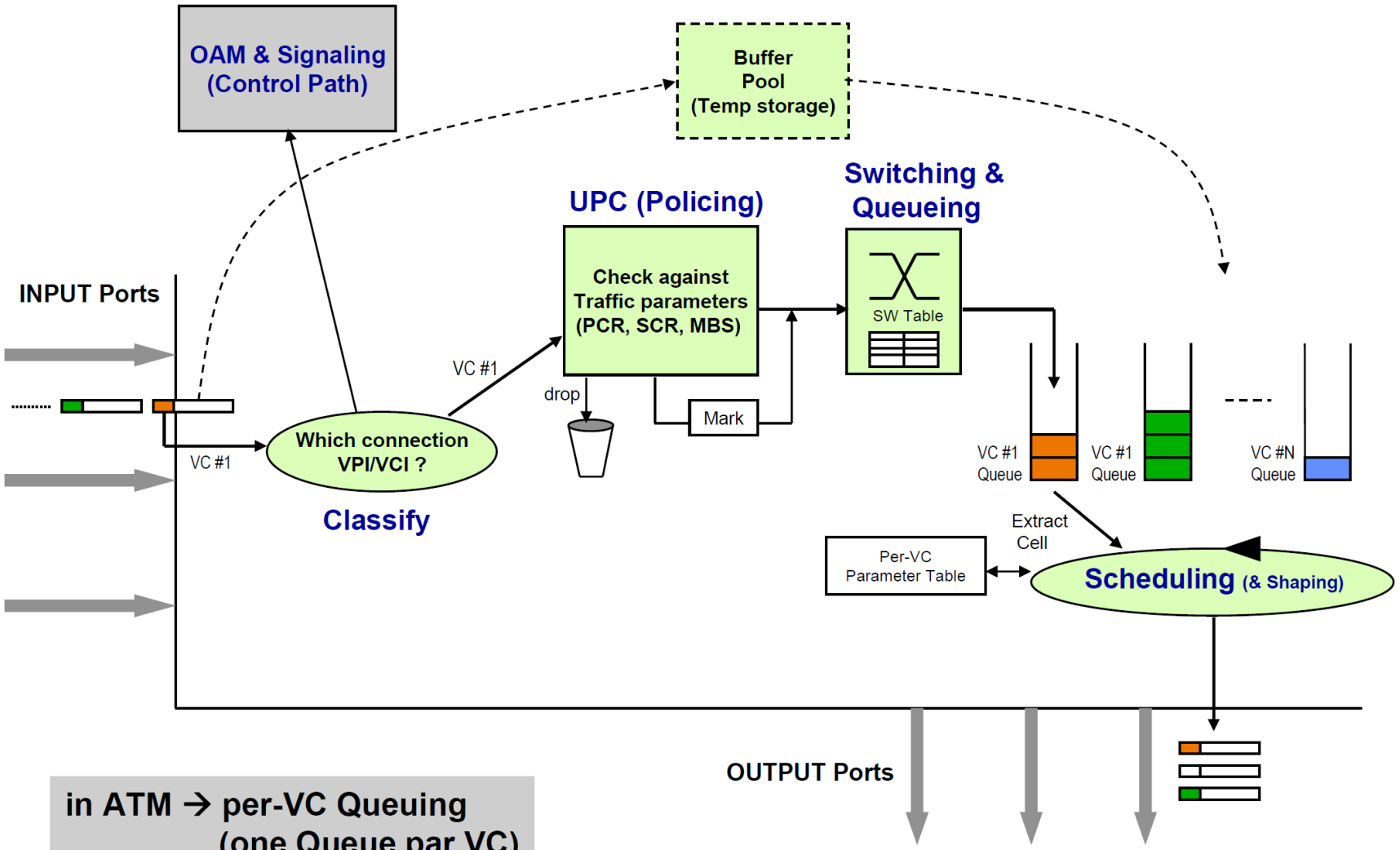
CBR Constant Bit Rate (téléphonie, son, fax, video à débit constant)
VBR Variable Bit Rate (voix compressée, data temps-réel)
UBR (+ABR, GFR) Unspecified Bit Rate (data non-critique)

Chaque connection (VC), selon sa classe de service, sera caractérisée par des **paramètres de traffic** (PCR, SCR, MCR, MBS) et de **QoS** (CDV, Max CTD, CLR)



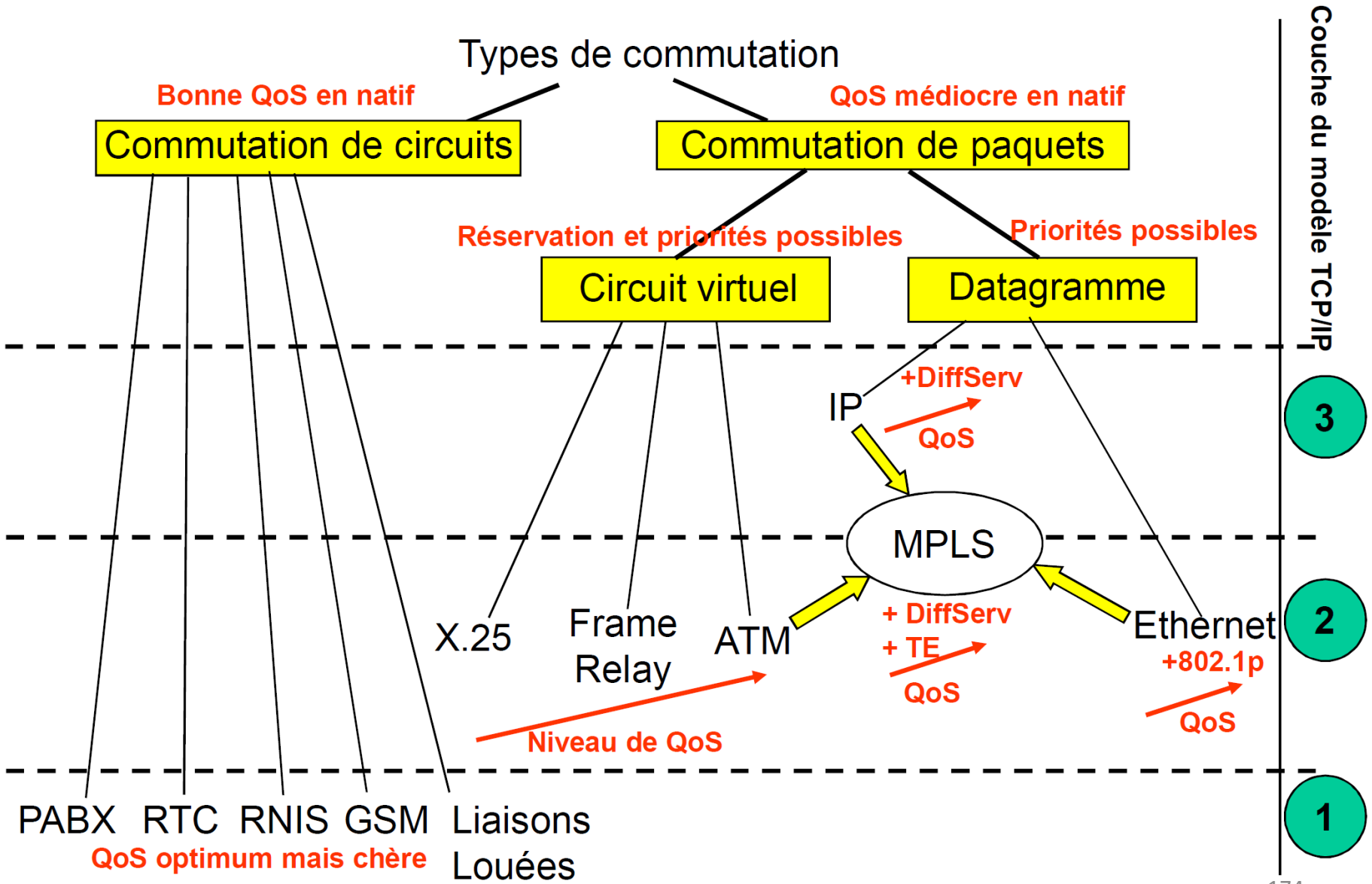


TM/QoS in ATM Switch: Functional description



in ATM → per-VC Queuing
(one Queue par VC)

Evolution de la commutation et de la QoS



ATM: un tentative du “tout-VC”…

- 1988: la technologie ATM est introduite par les opérateurs (ITU) résoudre unifier le transport de la voix et des données (permettre la QoS sans gaspillage des ressources)
- En ayant des switchs ATM comme uniques équipements LAN et WAN
- ATM reposait/repose sur
  - paquets petits: limite temps de remplissage
  - et de taille fixe: permet de prévoir délais d’attente et permet traitements hardware

→ Débit maximum standardisé à 622Mbps

… finalement supplantée par IP et MPLS

- **Dans le contexte actuel, ATM présente des inconvénients**
  - Il cohabite mal avec IP
    - Conversion d'adresses, mode multicast différent
    - Duplication des protocoles de routage
  - Le hardware permet maintenant le traitement de paquets de longueur variable
  - Les petites cellules ont moins d'intérêt à très haut débit
    - Et pénalisent même le traitement (plus de cellules à traiter)
  - ATM va trop loin dans les possibilités de qualité de service
    - Les stations d'extrémité peuvent s'adapter à une qualité de service un peu moins bonne
    - La qualité de service native est meilleure à très haut débit

- **De nombreux constructeurs ont proposé des solutions de commutation IP faisant la synthèse de la commutation ATM et du routage IP**
  - IP switching (Nokia), ARIS (IBM), Tag Switching (Cisco), etc…

- **MPLS (MultiProtocol Label Switching) est la synthèse IETF de ces propositions**

Plan général du cours

I. **Organisation des opérateurs de l'Internet**

II. **TCP et Qualité de service**

III. **Commutation par circuits virtuels**

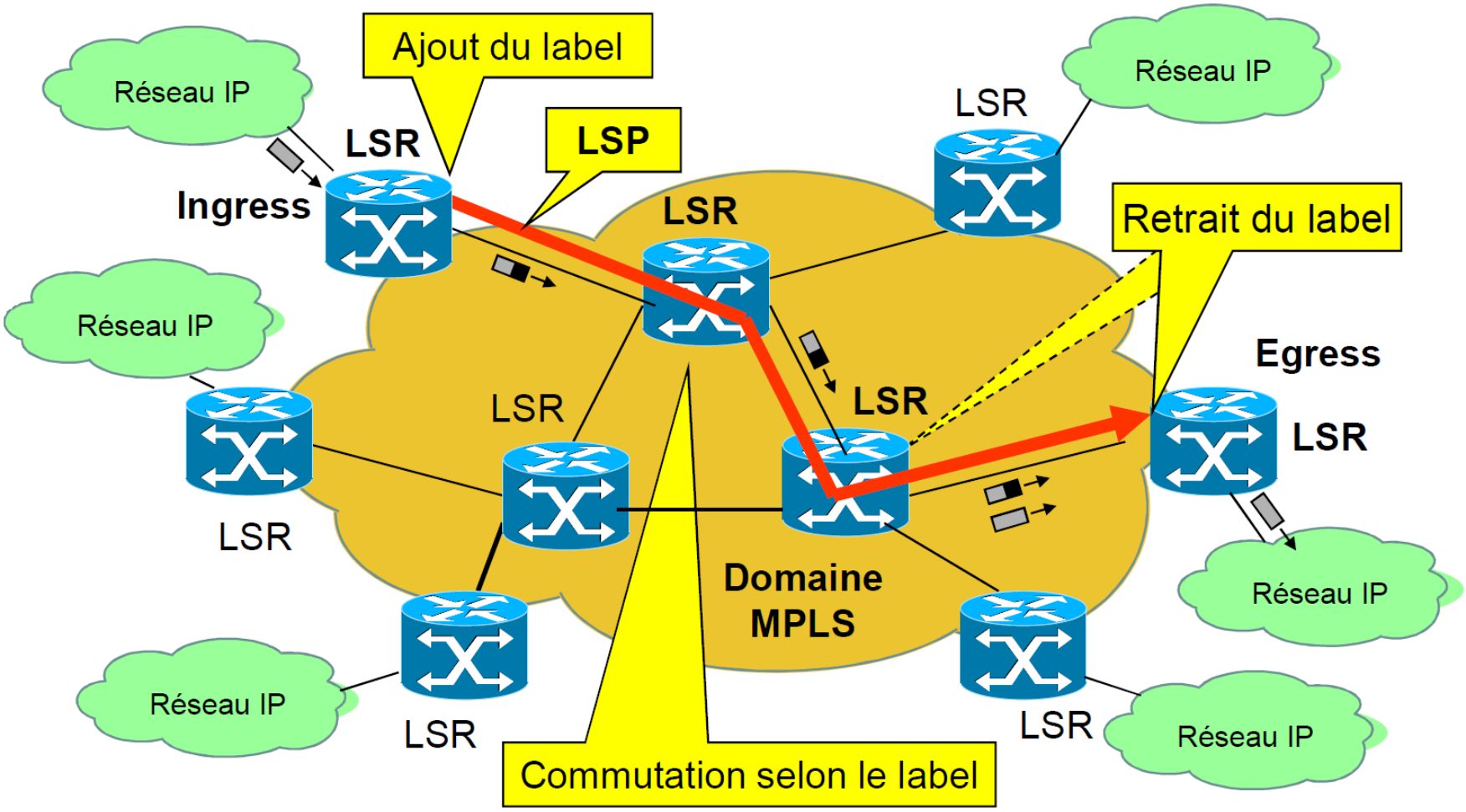
IV. **Commutation par VC dans le monde IP : MPLS**

- **IV.1. Fonctionnement de MPLS**
- **IV.2. Ingénierie de trafic avec MPLS : MPLS-TE**
- **IV.3. Offres de service MPLS : les VPN basés sur MPLS**
  - IV.3.a. IP-VPN
  - IV.3.b. Ethernet-VPN : VPLS

Principes de MPLS

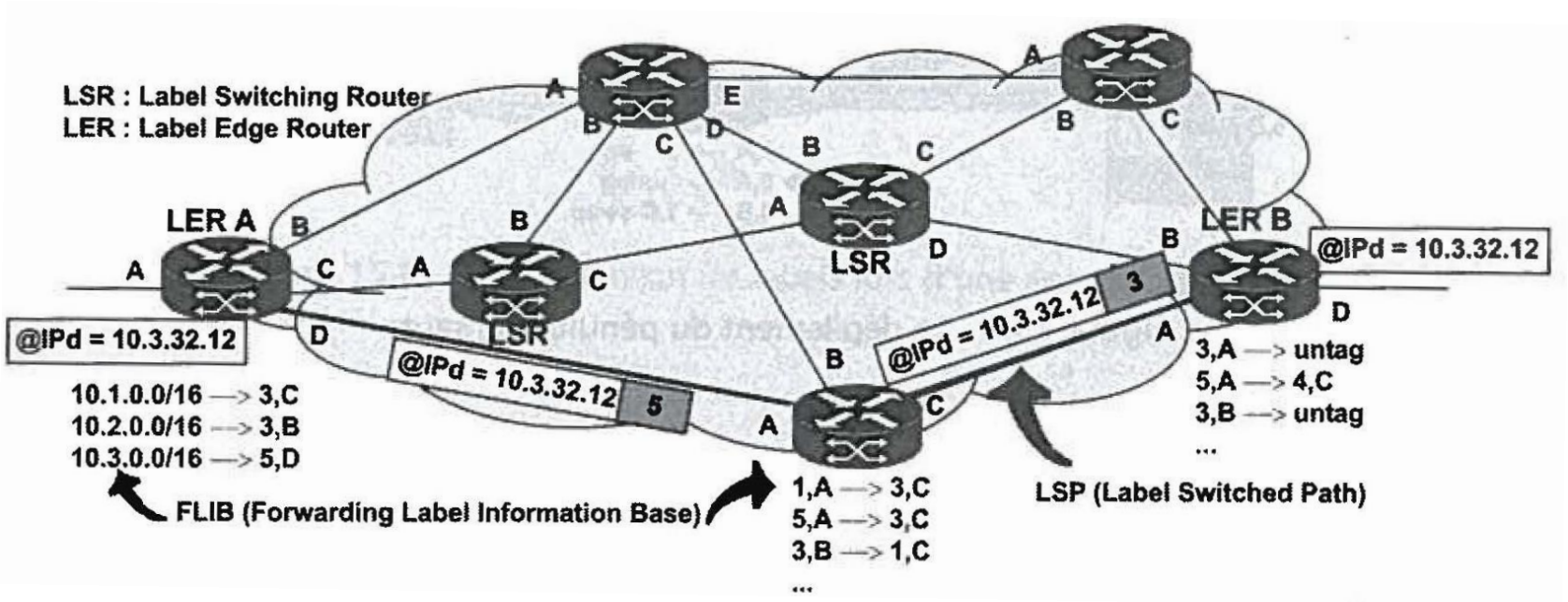
- **Des chemins prédéfinis relient les points d'extrémité du réseau**
  - Les LSP (Label Switched Path)
    - Un LSP est unidirectionnel
  - Les LSP sont établis par un protocole de signalisation en suivant la route déterminée par les protocoles de routage IP
    - Les LSP peuvent être établis à l'initiative de l'administrateur (proche des circuits virtuels permanents)
    - Ou ils sont établis automatiquement à l'initiative d'un point d'extrémité du réseau dés qu'il apprend par les protocoles de routage l'existence d'un nouveau préfixe IP
- **Les équipements MPLS s'appellent des LSR (Label Switch Router)**
- **A l'entrée du réseau, le 1er LSR (« Ingress LSR ») analyse le paquet IP**
  - Il choisit alors le LSP et insère un label devant le paquet IP
- **Les équipements suivants (les LSR du cœur de réseau) relaient le paquet en se basant seulement sur le label**
- **Le LSR de sortie (« Egress LSR) retire le label**
  - Dans certaines implémentations, c'est l'avant dernier LSR qui retire le label
- **A la sortie le paquet est routé selon le fonctionnement IP traditionnel**

Principes de MPLS



Principes de MPLS

Acheminement d'un datagramme IP dans un réseau MPLS:



Dans un réseau MPLS, un même paquet MPLS peut recevoir plusieurs labels (Push tag).
L'empilement de labels permet de définir une **agrégation de routes** en interne dans le réseau et des VPN.
L'opération **Pop tag** permet de supprimer le label de haut de pile, alors que **Untag** supprime le dernier label.

Principes et composantes: la FEC

Forwarding Equivalent Class (FEC)

Définition

- Un ensemble de paquets à traiter de la même façon
- Ils sont tous envoyés au même prochain saut
- Une FEC est identifiée par un label

Exemples

- Paquets unicast dont l'adresse destinataire a le même préfixe
- Paquets unicast dont l'adresse destinataire a le même préfixe et le même champ ToS (ou DS)
- Paquets unicast faisant l'objet d'une décision d'ingénierie de trafic
- Paquets appartenant à un même VPN
- Paquets multicast de même source et mêmes destinataires

Granularité

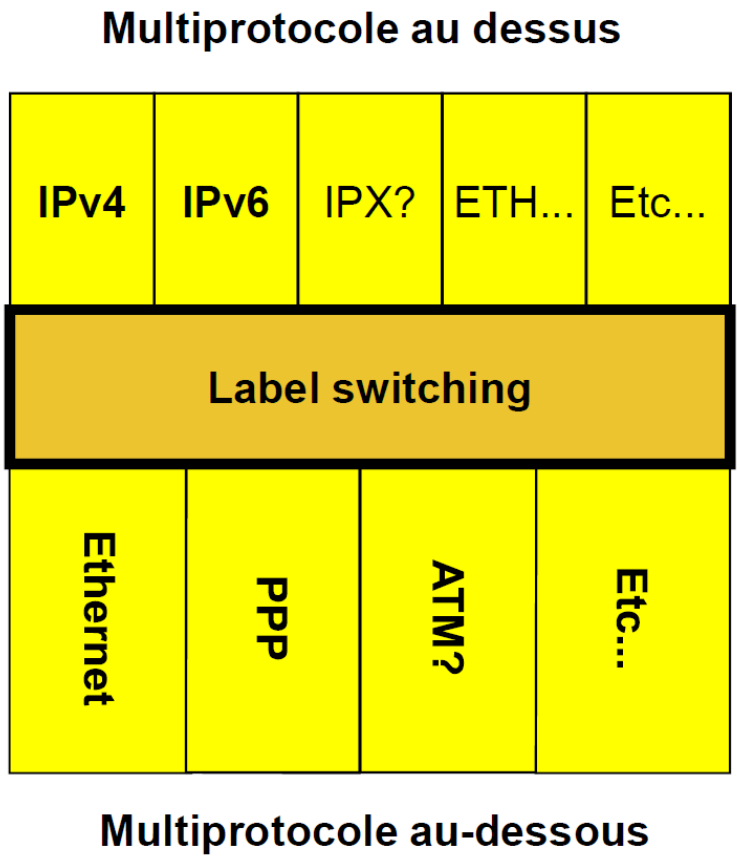
- Peut être quelconque
- Grossière si la FEC regroupe des adresses destinataires de même préfixe
- Fine si la FEC correspond au flux d'une application entre 2 machines

MPLS forwarding

- At ingress LSR:
  - Find the appropriate FEC from packet header
  - Bind label to FEC
  - Encapsulate IP packet in a MPLS packet
- In the core:
  - Perform label switching
  - Send packet on output link
- At egress LSR:
  - De-encapsulate IP packet from MPLS packet
  - Use the FIB to find the next hop

Réseaux d'infrastructure supportés

- **Protocole supérieur quelconque**
  - IPv4 ou IPv6 (niveau 3)
  - Ethernet (pour service VPLS)
- **Protocole de niveau 2 quelconque**
  - ATM (pour migration)
  - PPP (sur liaisons Sonet/SDH)
  - Ethernet 1 ou 10 Gbps
  - Combinaison des approches précédentes
- **MPLS est donc flexible**
  - Peut utiliser l'infrastructure ATM existante, puis migrer vers Ethernet ou autre
  - Peut évoluer facilement vers IPv6
  - Peut transporter n'importe quel trafic
    - Par ex. des trames Ethernet (VPN de niveau 2 (VPLS)



VPLS : Virtual Private LAN Service

Rappel : le routage IP passe à l'échelle

- La hiérarchisation de l'adressage IP permet de ne pas augmenter la taille des tables de routage avec le nombre de machines : ce ne sont pas des adresses machines qui sont stockées, mais des adresses de réseaux de différents niveaux.

- Les adresses de réseau désignent aussi la localisation « spatiale » des équipements IP par rapport aux autres.

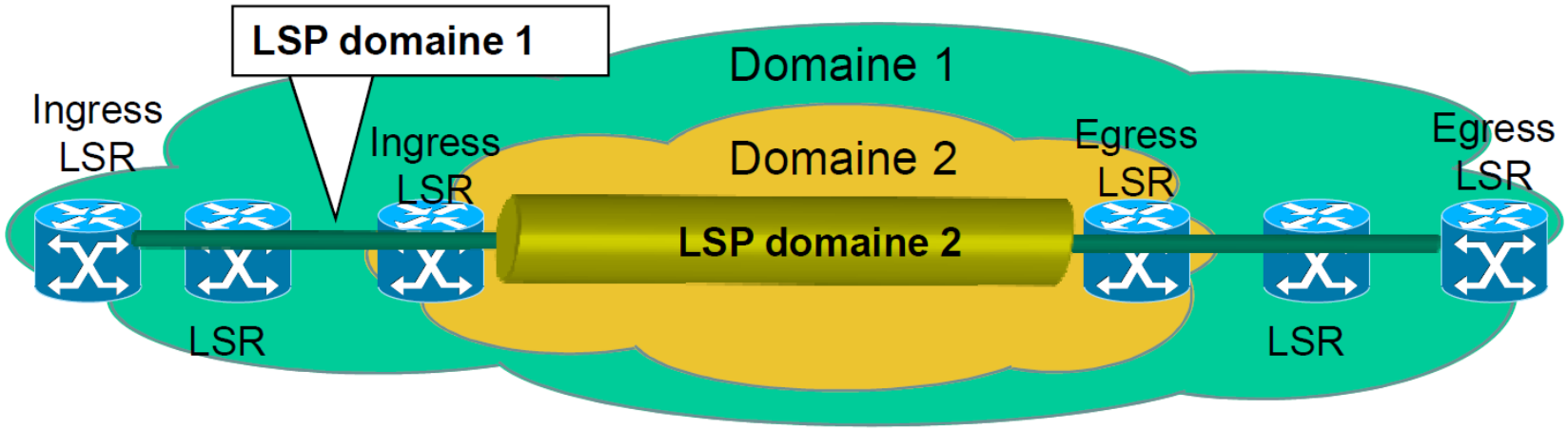
- Ainsi au « cœur » de l'Internet, les adresses réseaux sont sur peu de bits, et au fur-et-à-mesure qu'on approche de la périphérie, on se concentre sur des sous-réseaux de plus en plus fins.

- MPLS (comme ATM avec les VPi) permet aussi « l'agrégation de routes», de VC, grâce à la hiérarchisation des domaines : un LSP peut être défini entre 2 points dans le cœur du réseau pour grouper les LSP partageant une portion commune.

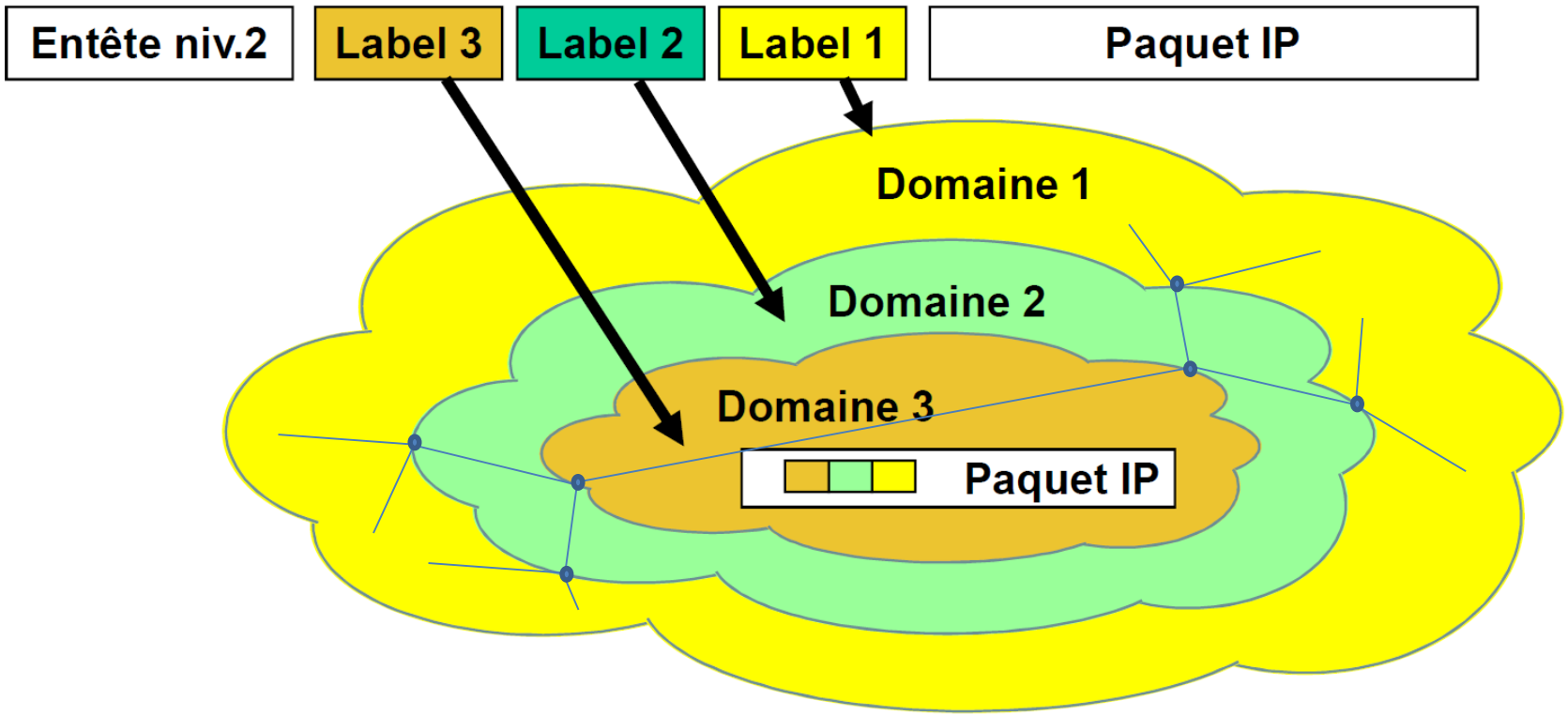
  --> passage à l'échelle de MPLS

Domaines MPLS

- **Le réseau peut être découpé en domaines administratifs**
  - Ces domaines peuvent être hiérarchisés
- **Relais entre les domaines**
  - Entre ses LSR d'extrémité, le LSP du domaine 2 sert de tunnel au LSP du domaine 1
    - Le paquet est alors précédé d'une pile de 2 labels
      - La pile peut contenir un nombre quelconque de labels
      - Similaire aux conduits ATM, mais plus de 2 niveaux de hiérarchie
    - Permet de réaliser des niveaux d 'agrégation dans le cœur de réseau



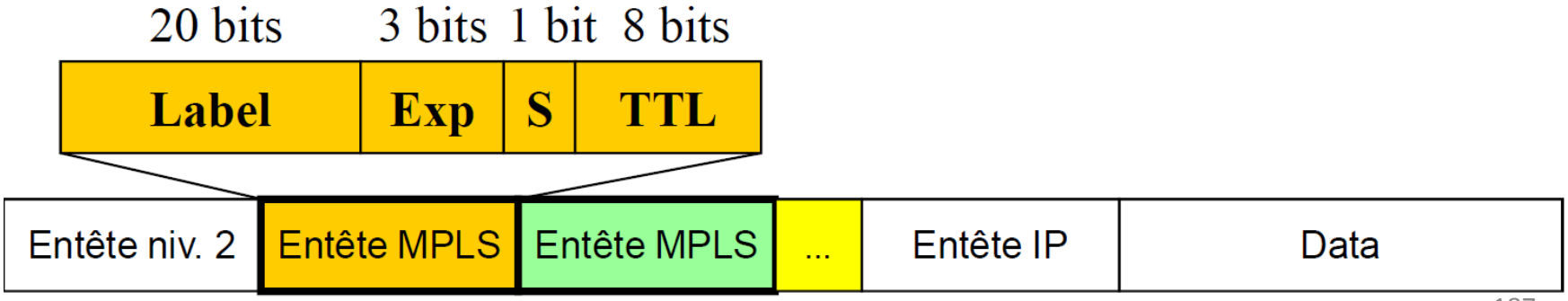
Domaines MPLS hiérarchisés



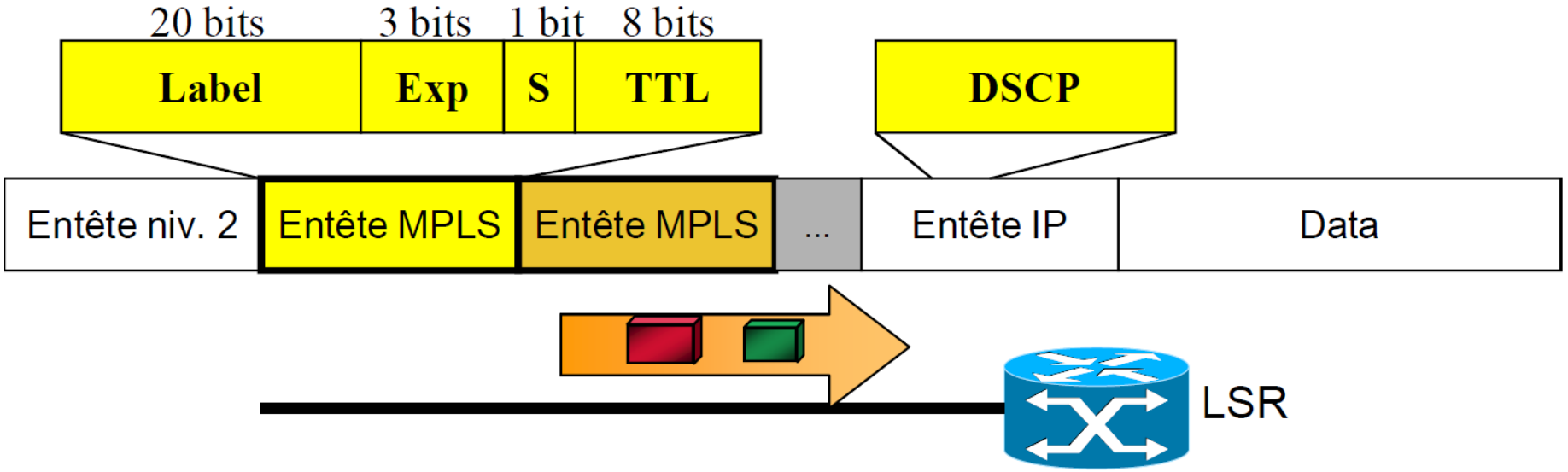
- **Les domaines MPLS peuvent être hiérarchisés grâce aux piles de labels**
  - Pour augmenter les performances au cœur du réseau
  - Pour permettre les interactions entre opérateurs
- **Les piles de labels sont aussi utilisées pour garantir l'étanchéité des VPN**

Labels MPLS

- **L'en-tête MPLS est inséré entre le niveau 2 et l'en-tête IP**
  - Label : numéro sur 20 bits (Valeurs 0 à 16 réservées)
  - S sert à gérer des labels hiérarchisés (Stack)
    - Marque le dernier label avant l'entête IP
  - Exp peut être utilisé pour traiter la QoS : files d'attente et rejet
    - Fonctionnement conforme à DiffServ
  - TTL a le même rôle que dans IP (détection de boucles)
- **Opérations sur les labels**
  - Swap (dans les LSRs), push (dans ingress LSR), pop (dans egress LSR)
- **Le label peut éventuellement être implicite**
  - Par exemple une longueur d'onde

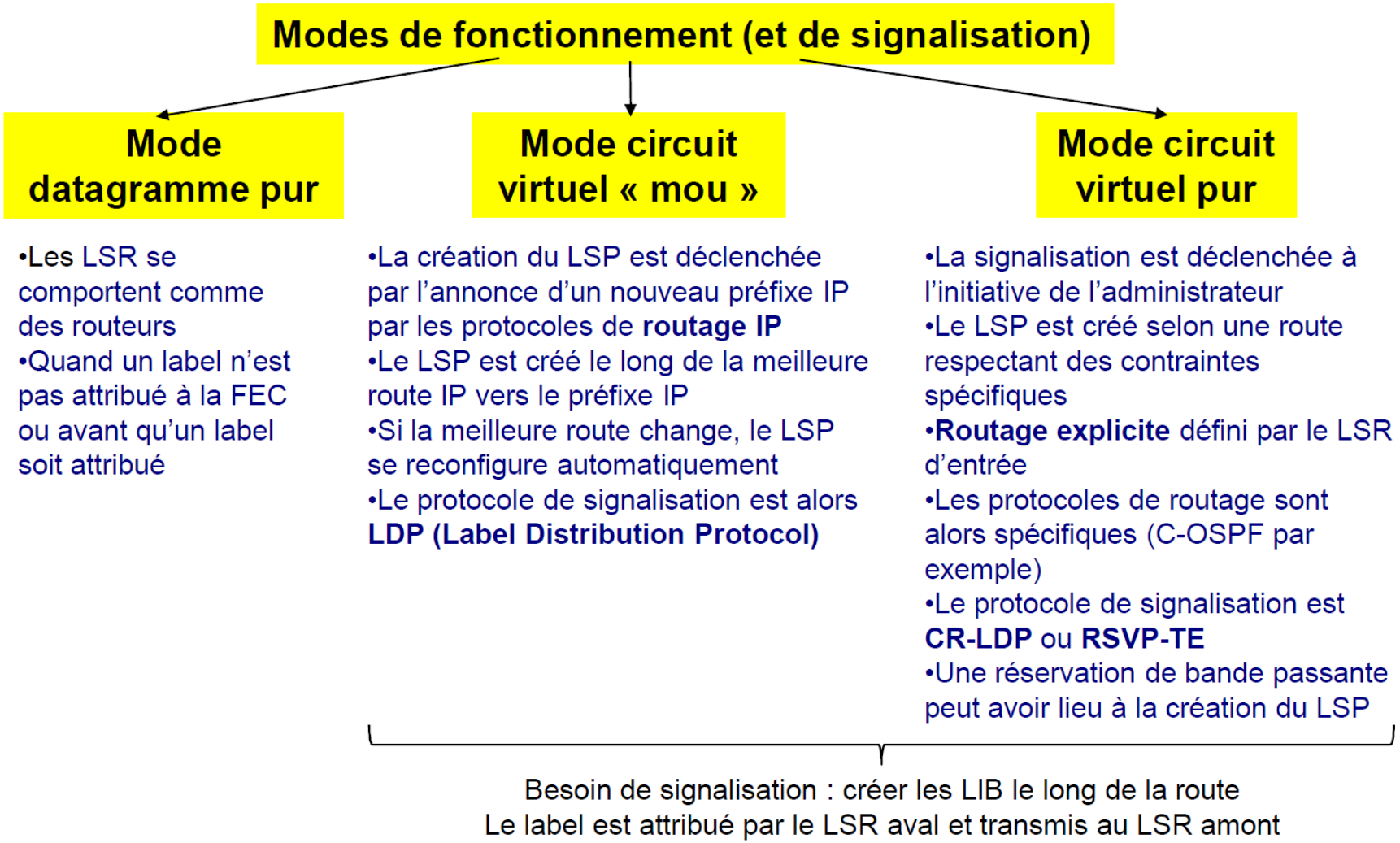


MPLS et DiffServ (RFC 3270)



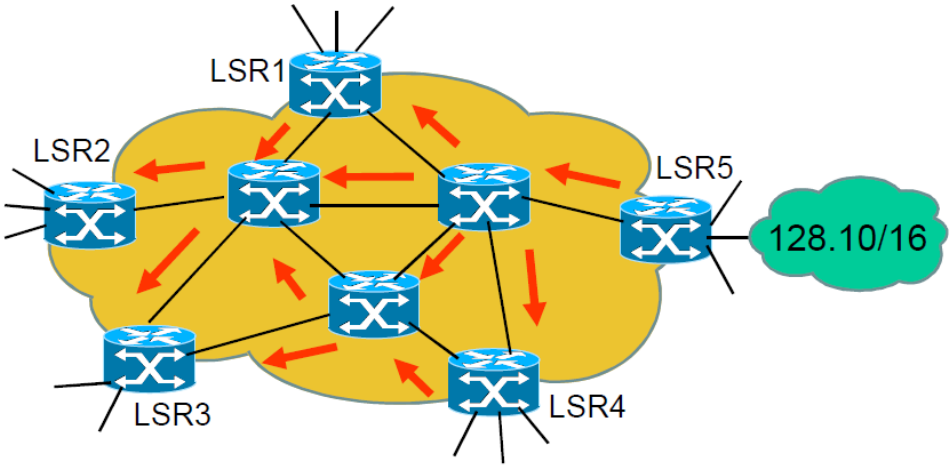
- **Les LSR doivent appliquer le traitement de QoS demandé dans le champ DSCP des paquets IP entrants**
  - Mais le DSCP est dans l'entête IP et ne sera donc plus visible
  - Le LSR Ingress marque le champ Exp (CoS) de l'entête MPLS
    - Il ne peut y avoir que 8 PHB au plus (3 bits)!
- **Les LSR MPLS utilisent DiffServ de la même façon que les routeurs**

Modes de fonctionnement de MPLS

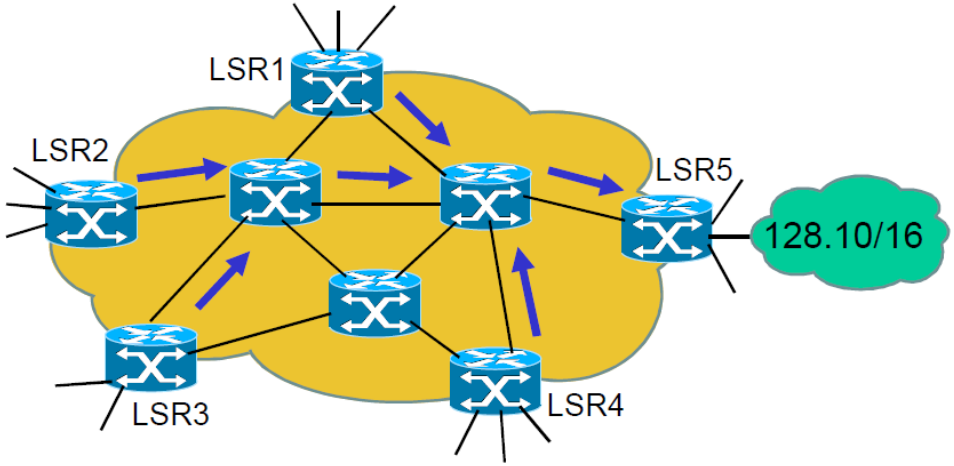


Exemple de signalisation par LDP

- **Le préfixe 128.10/16 est annoncé par le protocole de routage**
  - Par les messages OSPF →
  - Les LSR d'entrée LSR1, LSR2, LSR3, LSR4) apprennent l'existence du préfixe 128.10/16

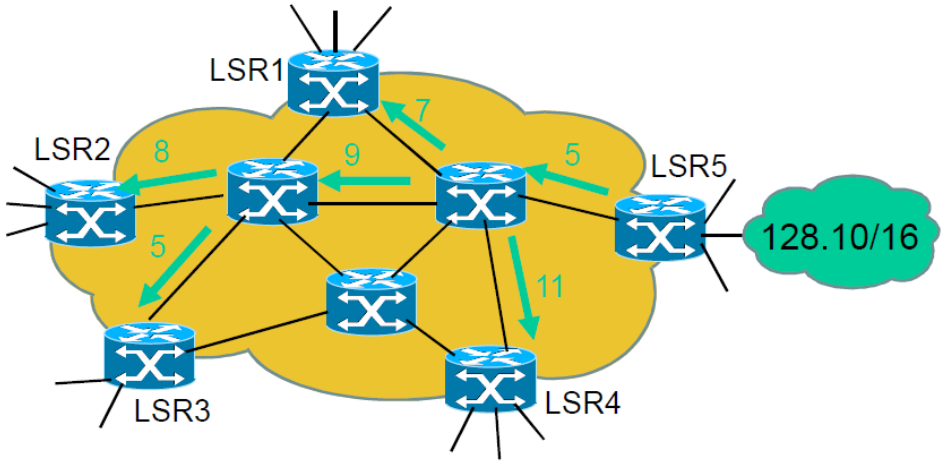


- **Les LSR d'entrée (LSR1, LSR2, LSR3 et LSR4) demandent l'établissement d'un LSP vers 128.10/16**
  - Par des messages *Label Request* de LDP →
  - Ces messages suivent la meilleure route IP vers 128.10/16

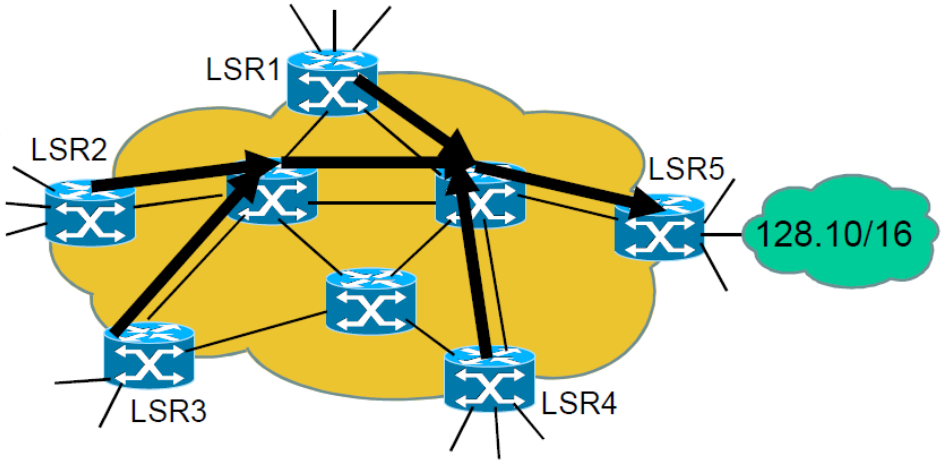


Exemple de signalisation par LDP

- **Les labels sont attribués en commençant par l'aval**
  - En partant de l'egress LSR (LSR5), chaque LSR répond au message *Label Request* par un message label mapping → contenant un n°de label
  - Les Ingress LSR reçoivent leur label



- **Le LSP (multipoint à point) est établi, depuis chaque ingress LSR vers 128.10/16**



Plan général du cours

I. **Organisation des opérateurs de l'Internet**

II. **TCP et Qualité de service**

III. **Commutation par circuits virtuels**

IV. **Commutation par VC dans le monde IP : MPLS**

IV.1. Fonctionnement de MPLS

IV.2. Ingénierie de trafic avec MPLS : MPLS-TE

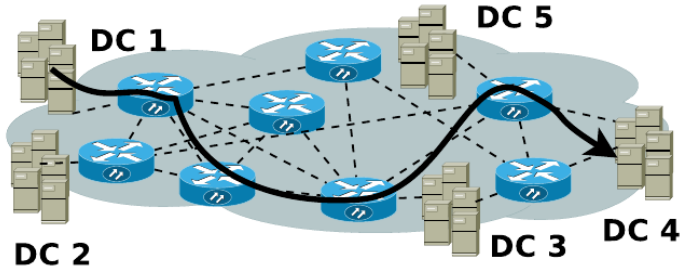
IV.3. Offres de service MPLS : les VPN basés sur MPLS

IV.3.a. IP-VPN

IV.3.b. Ethernet-VPN : VPLS

MPLS Traffic Engineering (MPLS-TE)

- **Ingénierie de trafic (TE) : processus de décision pour router/répartir le trafic à travers le réseau du SP**
- Les mécanismes de TE permettent d'utiliser efficacement les ressources du réseau tout en maintenant de bonnes performances pour le trafic.
- MPLS permet de choisir entièrement le chemin suivi par le trafic, alors que TE basé sur le routage (OSPF) ne permet que d'employer le plus court chemin -> pas souvent optimal
- Le TE basé sur MPLS est le plus répandu aujourd'hui, et supporté par les fabricants majeurs tels que Cisco et Juniper.
- Tous les SP de Tier 1, et la plupart de Tier 2, 3 et 4 utilisent MPLS pour TE et pour les VPN de niveau 3 et 2, et les réseaux d'accès de tél mobile.



A. Pathak, M. Zhang, Y.-C. Hu, R. Mahajan, and D. Maltz. *Latency inflation with MPLS-based traffic engineering*. ACM SIGCOMM conference on Internet measurement (IMC), Nov. 2011, Berlin, Germany,.

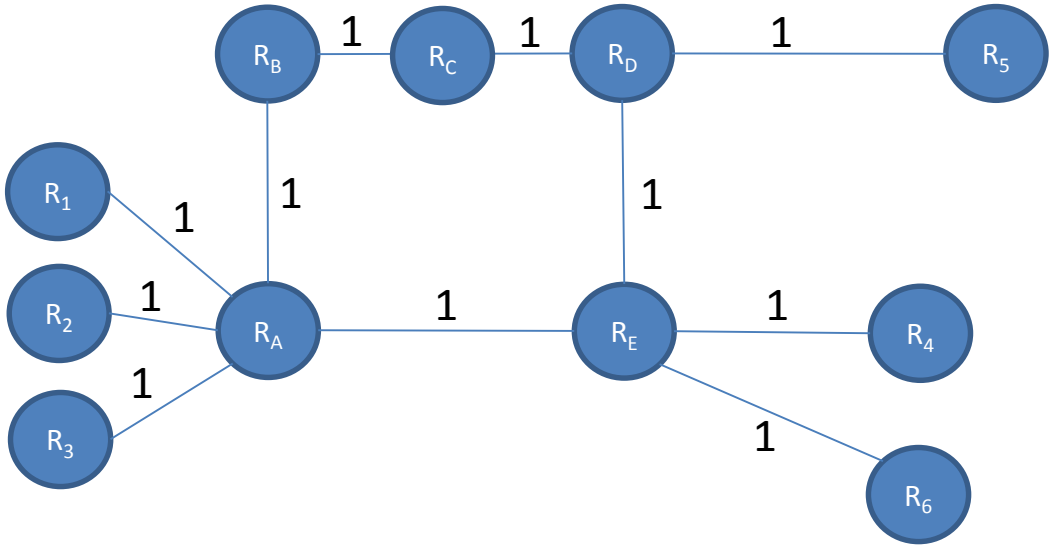
Principe du TE

- Le SP veut configurer 3 LSP : LSP14, LSP25 et LSP26, tels que :

Ingress/Egress	R_4	R_5	R_6
R_1	1 Gbps		
R_2		0.2 Gbps	0.8 Gbps
R_3			

Matrice de trafic

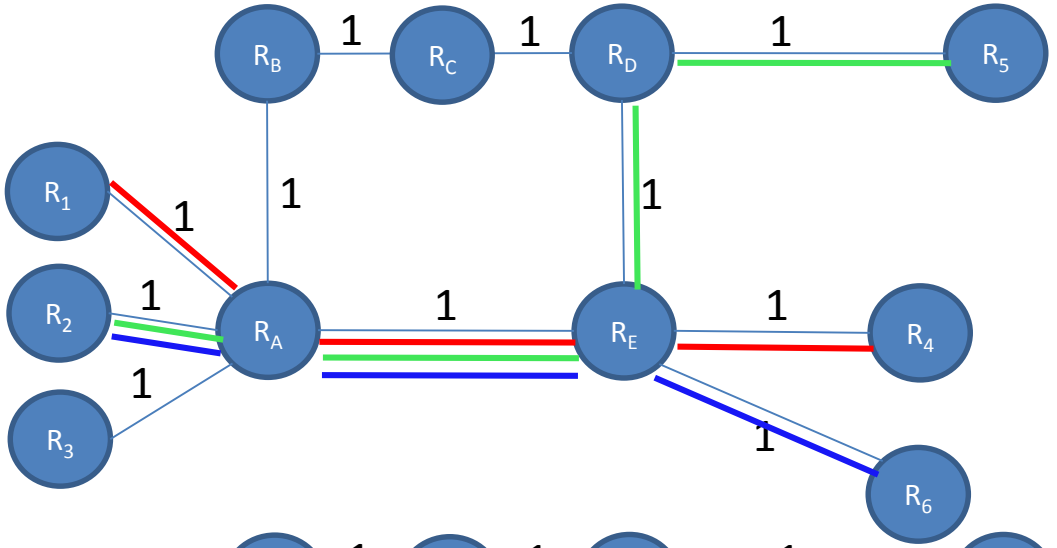
- dans ce réseau :



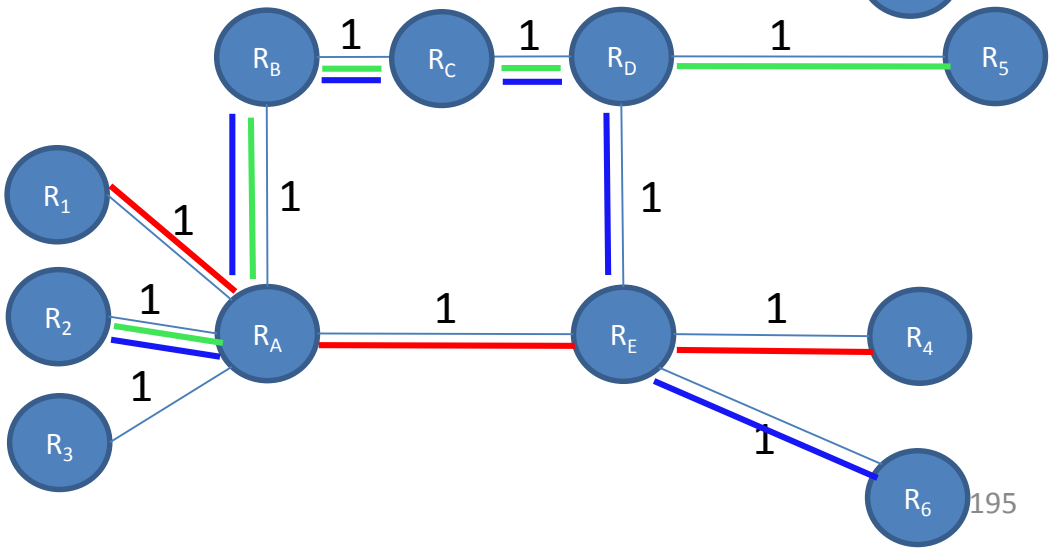
- La matrice de trafic se traduit en :

LSP\Caractéristiques	BP	Priorité	Route
LSP14	1 Gbps	1	? (rouge)
LSP25	0.2 Gbps	2	? (vert)
LSP26	0.8 Gbps	3	? (bleu)

- Solution SPF : pas de prise en compte des contraintes de BP



- Solution CSPF : prise en compte des contraintes de BP



Algorithmes MPLS-TE

- Après qu'un chemin est sélectionné, le LSP décompte la BP requise sur l'interface de sortie de chaque routeur du chemin. Chaque interface de sortie de routeur maintient un compteur pour sa BP courante réservable.

- **L'information de BP réservable et la topologie du réseau (TED) est périodiquement disséminée sur le réseau (=graphe de connexions avec BP réservable sur chaque branche).**

- Priorité et préemption : Chaque LSP est configuré avec 2 valeurs de priorité : priorité d'établissement et priorité de maintien.
  - La priorité d'établissement détermine si un nouvel LSP peut être établi en préemptant un LSP existant.
  - La priorité de maintien détermine dans quelle mesure un LSP existant peut garder sa réservation.
  - Un nouvel LSP avec une haute priorité d'établissement peut préempter un LSP existant avec une basse priorité de maintien si : (a) il n'y a pas assez de BP réservable dans le réseau; ou (b) le nouvel LSP ne peut pas être établi à moins qu'un LSP existant ne soit effacé.

Algorithmes MPLS-TE

- **CSPF** : trie les LSP selon leurs priorités et sélectionne le plus court chemin pour chaque LSP.
  - Commence avec le LSP de plus haute priorité, élague le TED en enlevant les liens qui n'ont pas une BP réservable suffisante,
  - Assigne ensuite le chemin le plus court dans ce TED élagué au LSP et met à jour la BP réservable sur les liens affectés.
  - Ce processus se poursuit jusqu'à ce qu'il ne reste plus de LSP.

- **Ré-optimisation** : CSPF est lancé périodiquement pour ré-assigner à chaque LSP un meilleur chemin si possible = autoBP

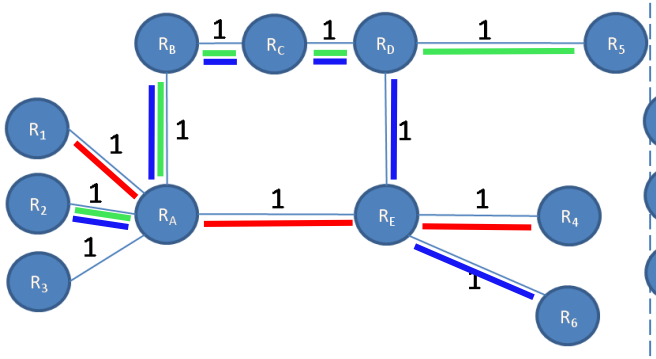
Algorithmes MPLS-TE

- **AutoBP** : MPLS ne contrôle pas le débit (BP) utilisé par le trafic sur un LSP : un LSP peut porter tout débit de trafic indépandemment de sa BP réservée.
  - A la place de ce possible contrôle, un mécanisme d'autoBP permet à un LSP d'ajuster sa BP réservée au débit courant.
  - Pour utiliser autoBP, un LSP a besoin de plusieurs paramètres en plus : seuil d'ajustement, intervalle d'ajustement et intervalle d'échantillonage.
  - Chaque intervalle d'échantillonage (ex : 5 min), un LSP mesure le débit moyen qu'il supporte. Chaque intervalle d'ajustement (ex : 15 min), il calcule le max du débit moyen mesuré sur chaque intervalle d'éch.
  - Si le max du débit utilisé diffère de la BP réservée courante de plus que le seuil d'ajustement, alors le LSP invoque CSPF avec le max du débit comme nouvelle BP à réserver.

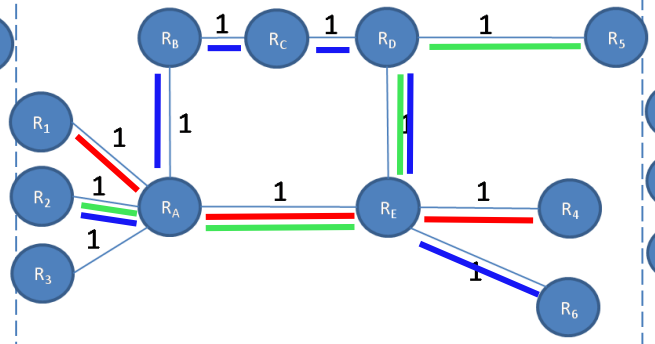
Auto-BP et CSPF

Caractéristiques	BP allouée	BP utilisée	Priorité	BP allouée	BP utilisée	Priorité	BP allouée	BP utilisée	Priorité
LSP14	1 Gbps	0.5	1	0.5	0.9	1	0.9 Gbps	0.9	1
LSP25	0.2 Gbps	0.3	2	0.3	0.2	2	0.2 Gbps	0.2	2
LSP26	0.8 Gbps	0.7	3	0.7	0.7	3	0.7 Gbps	0.7	3

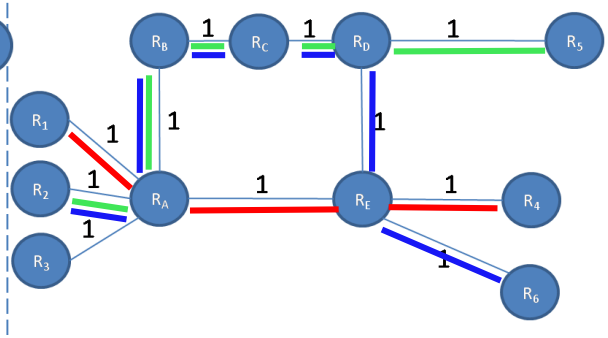
t_1



t_1+15min



t_1+30min



MPLS : quelques chiffres

- 2011 [1]:
  - 90% des LSP avec 7 sauts ou moins, certains avec plus de 15
  - 25% de tous les chemins en 2011 traversent au moins 1 tunnel MPLS, 4% plus d'un
  - Il semble que beaucoup d'AS emploient de la classification de trafic et de l'ingénieurie dans leurs tunnels.
- 2017 [2]:
  - 87% des opérateurs étudiés déploient MPLS

[1] J. Sommers, B. Eriksson and P. Barford. *On the Prevalence and Characteristics of MPLS Deployments in the Open Internet*. ACM Internet Measurement Conference, 2011.
[2] Y. Venaudel, P. Mérindol, J.-J. Pansiot and B. Donnet. Through the Wormhole: Tracking Invisible MPLS Tunnels. ACM Internet Measurement Conference, Nov. 2017.

Révision

- Principe de MPLS :

- Principe de MPLS-TE :

Plan général du cours

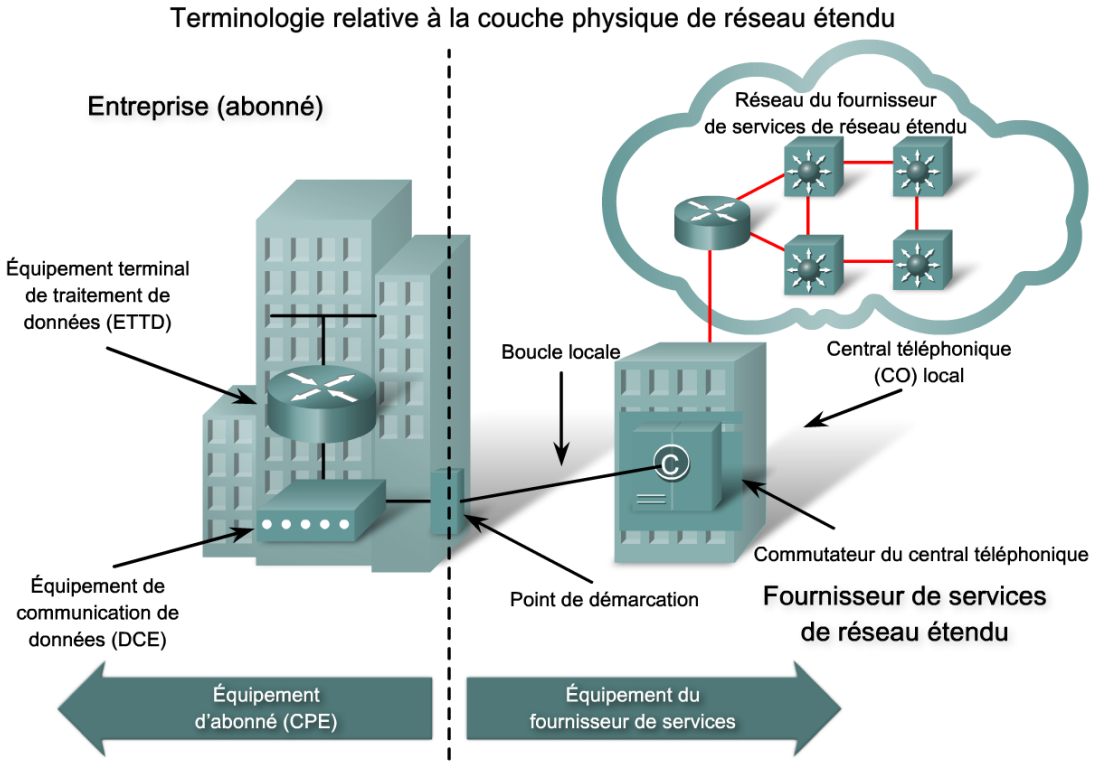
I. **Organisation des opérateurs de l'Internet**
II. **TCP et Qualité de service**
III. **Commutation par circuits virtuels (VC) : le cas d'ATM**
IV. **Commutation par VC dans le mode IP : MPLS**
    IV.1. Fonctionnement de MPLS
    IV.2. Ingénierie de trafic avec MPLS : MPLS-TE
    IV.3. Offres de service MPLS : les VPN basés sur MPLS
        IV.3.a. IP-VPN
        IV.3.b. Ethernet-VPN : VPLS

I. **Accès : Réseaux optiques**
II. **Accès : Technologies xDSL**

Concepts de la couche physique de réseaux étendus (1)

• CPE (Customer Premises Equipment) ou **CE (Customer Edge)**
Équipement de l'abonné directement relié par un port physique à son ISP

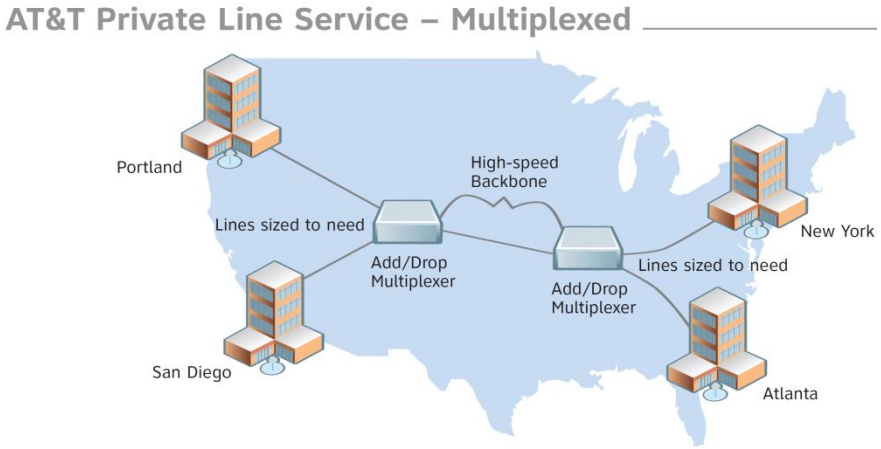
L'abonné est propriétaire de l'équipement ou le loue à son fournisseur de services.



• **PE (Provider Edge)**:
Équipement de l'ISP directement relié au client, au CE

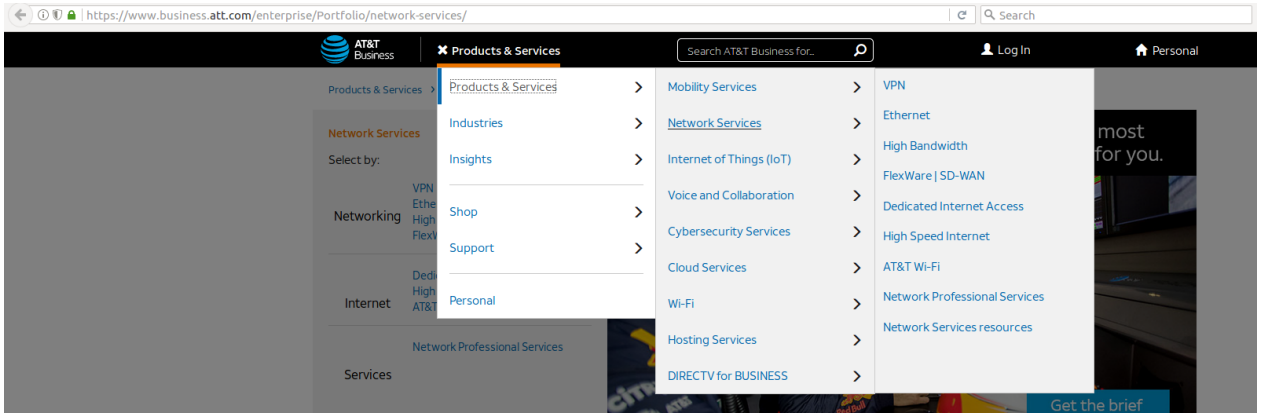
Le problème

- L'interconnexion de LAN distants appartenant à une même entreprise



Identifier sa solution réseau idéale

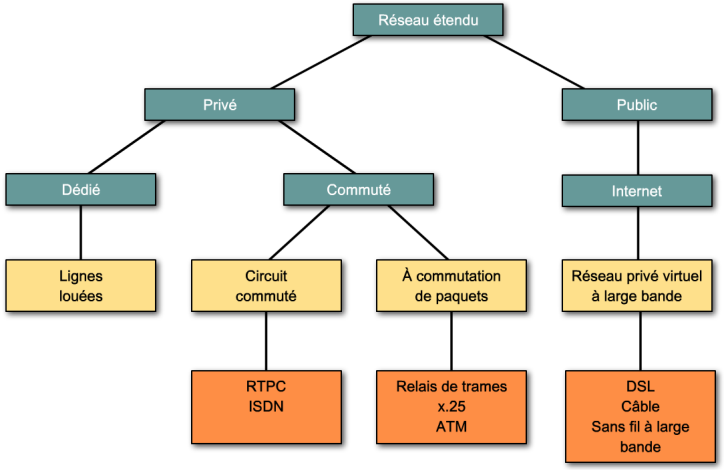
- Solutions technologiques pour les contraintes de l'entreprise



  - **VPN**
  - **Ethernet** metro, ou WAN (à plus grande échelle géographique)
  - **Internet connectivity** : DSL et câble
  - **High Bandwidth** : anneaux optiques pour la fiabilité
  - **Dedicated Internet Access** : lignes louées

Options de connexions de réseaux étendus

- OPTIONS DE CONNEXION DE RÉSEAU ÉTENDU PRIVÉ:



➢ **Liaison dédiée**:

-Un circuit dédié et permanent entre 2 sites du client et une destination distante par l'intermédiaire du réseau du AP.

- Pour connexions dédiées permanentes: lignes point à point de capacités variées et limitées uniquement par les installations physiques sous-jacentes et la volonté des utilisateurs à payer.

Les lignes point à point sont généralement louées à un opérateur et prennent le nom de **lignes louées**.

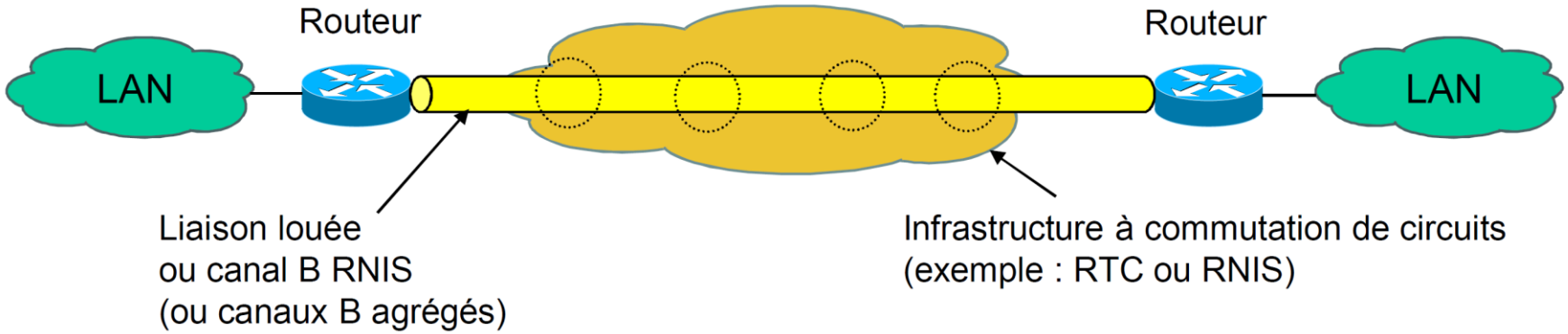
➢ **Liaison commutée :**

- Liaisons à commutation de circuits: établissement dynamique d'un circuit dédié (pour la voix ou les donnée). Ex: connexions commutées analogiques (RTPC) et les lignes RNIS

- Liaisons à commutation de paquets: avec connexion. Paquets avec labels
  Ex: FR, ATM, X.25 et Metro Ethernet + VPN basé réseau (sur MPLS)

Services WAN de niveau 1

Le client voit le réseau de l'opérateur comme un câble le reliant à l'autre LAN.



- **Le réseau WAN est basé sur la commutation de circuits**
  - L'opérateur fournit un « tuyau » : un circuit physique entre 2 routeurs
  - Les commutateurs de circuits sont transparents pour le protocole de niveau 2 (et protocoles de niveau supérieur)
- **Le protocole de niveau 2 entre les routeurs n'est pas imposé**
  - Protocole point à point quelconque
    - Le plus souvent PPP (Point to Point Protocol)
    - Sinon protocole propriétaire
- **Facturation selon la distance et le débit**
- **Solution chère, en général réservée aux très grosses entreprises**

Lignes louées: types de ligne et débits

Type de ligne	Débit binaire
56	56 Kbits/s
64	64 Kbits/s
T1	1,544 Mbits/s
E1	2,048 Mbits/s
J1	2,048 Mbits/s
E3	34,064 Mbits/s
T3	44,736 Mbits/s
OC-1	51,84 Mbits/s
OC-3	155,54 Mbits/s

Type de ligne	Débit binaire
OC-9	466,56 Mbits/s
OC-12	622,08 Mbits/s
OC-18	933,12 Mbits/s
OC-24	1 244,16 Mbits/s
OC-36	1 866,24 Mbits/s
OC-48	2 488,32 Mbits/s
OC-96	4 976,64 Mbits/s
OC-192	9 953,28 Mbits/s
OC-768	39 813,12 Mbits/s

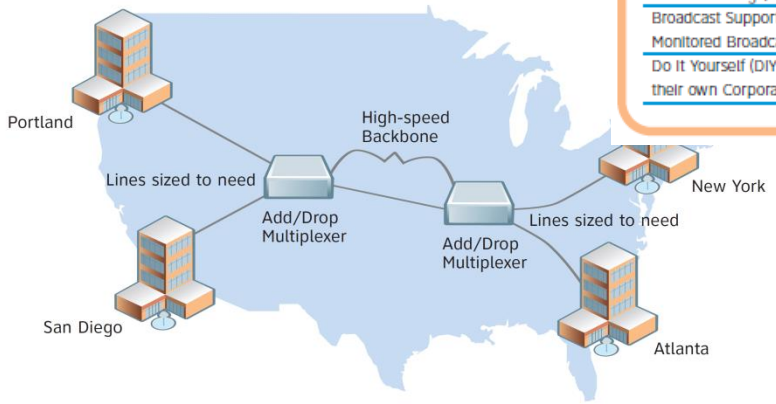
Private/Leased Lines

FEATURES

- Configures private lines over physically diverse POP to POP routes if required
- Uses self-healing SONET Network to route customer traffic
- SLAs that provide credits in the rare event of a service interruption
- Premium Services with Enhanced SLAs
- Multiple types of configurations – Point to Point, Multipoint and Add/Drop Multiplexer
- Cost-effective and scalable speeds up to 9.9 Gbps (OC192)

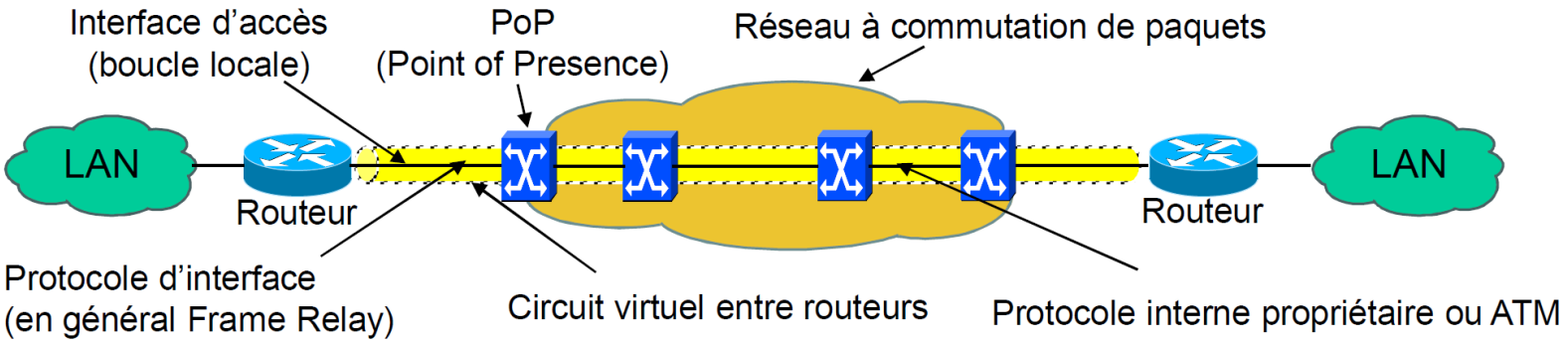
Business Need	Application	SC	T1	T3	Wavelength & SONET Service	Optical Mesh Service
Supply-Chain Management: Lower costs by managing assets and resources	Inventory tracking	•		•	•	•
	Order tracking	•			•	•
	Point of sale	•	•		•	•
	Shipment tracking	•			•	•
	Electronic Data Interchange				•	•
Information Sharing: Maintain a competitive edge by ensuring that employees have access to timely information and can work together to implement corporate goals	Hoot 'n' Hollers	•				
	Automatic Ring Down	•				
	Videoconferencing			•	•	•
	CAD-CAM			•	•	•
	CAD-CAM-CAE			•	•	•
Connectivity: Increase productivity by enabling workgroup communication	Bulk File Transfer	•	•	•	•	•
	Host-to-Host Connectivity		•	•	•	•
	ISP Interconnectivity and Hosting		•	•	•	•
	LAN Interconnection	•	•	•	•	•
	High volume voice, video, data			•	•	•
	Network Consolidation			•	•	•
	Bulk voice, data and video to multiple locations				•	•
Imaging: Overcome the challenges of sharing large, graphic-intensive files	Imaging			•	•	•
	Teleradiology			•	•	•
Automation: Increase efficiency by simplifying processes and automating tasks	Automatic Teller Machines	•			•	•
	Remote telemetry					
	Bulk file transfer & mirroring					
Enhanced Customer Service: Retain customers by employing personalized customer service strategies	Customer Info Systems				•	•
Distance Learning: Overcome the challenges of a dispersed and/or mobile workforce	Distance Learning/ Remote Consultation		•	•	•	•
Media Distribution: Deliver Broadcast Video/ Content Storage/Streaming Video	Distribution of various forms of media				•	•
Broadcast Support: Deliver Managed/ Monitored Broadcast Video	Full-time or occasional use/ news/sports/special events			•	•	•
Do It Yourself (DIY) Networking: Customers desire to build their own Corporate Utility Network	Customer looking to build their own backbone infrastructure				•	•

AT&T Private Line Service – Multiplexed



Services WAN de niveau 2 avec VC

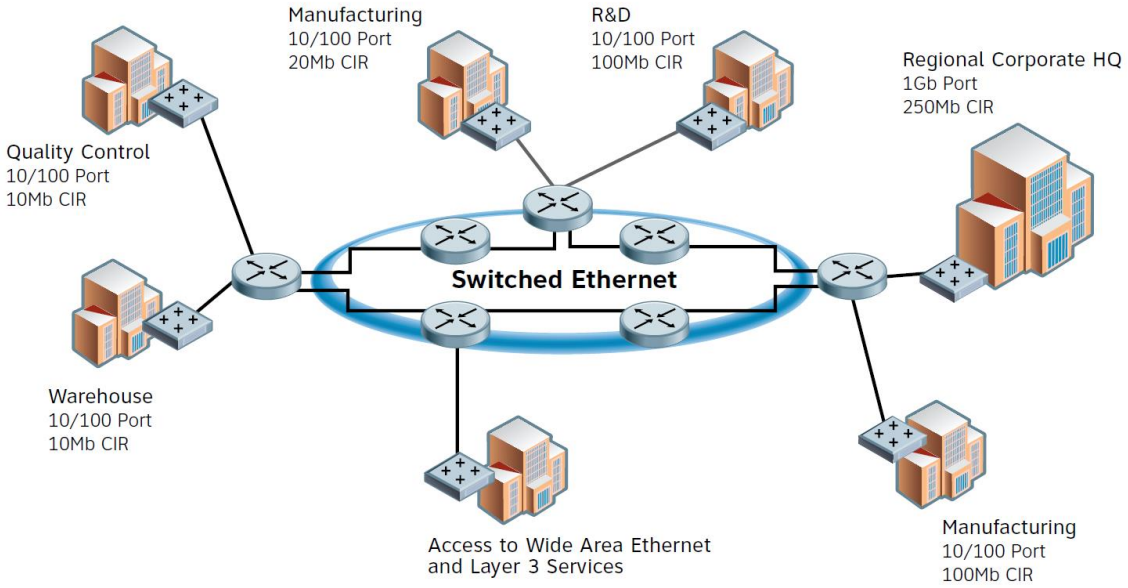
Le client voit le réseau de l'opérateur comme un switch le reliant à l'autre LAN.



- **Le réseau WAN est basé sur la commutation de paquets**
  - Un circuit virtuel entre 2 routeurs CPE
  - Le routeur utilise un protocole d'interface avec le PoP pour créer et exploiter un circuit virtuel le reliant au routeur distant
    - Le protocole d'interface le plus courant à ce niveau était FR, est de plus en plus Ethernet
    - X25 est obsolète (adapté à des besoins périmés), ATM reste en général à l'intérieur du réseau
  - Le POP est un commutateur de paquets
- **Frame Relay encore courant car bien adapté au trafic de données**
  - Mais peu adaptée à la téléphonie sur IP
- **Mais en voie de déclin, car l'opérateur doit gérer plusieurs infrastructures réseau**
  - Une infrastructure pour la téléphonie (commutation de circuits)
  - Une infrastructure pour l'interconnexion des LAN
  - Une infrastructure pour Internet

Switched Ethernet & Virtual Private LAN Service (VPLS or layer-2 VPN)

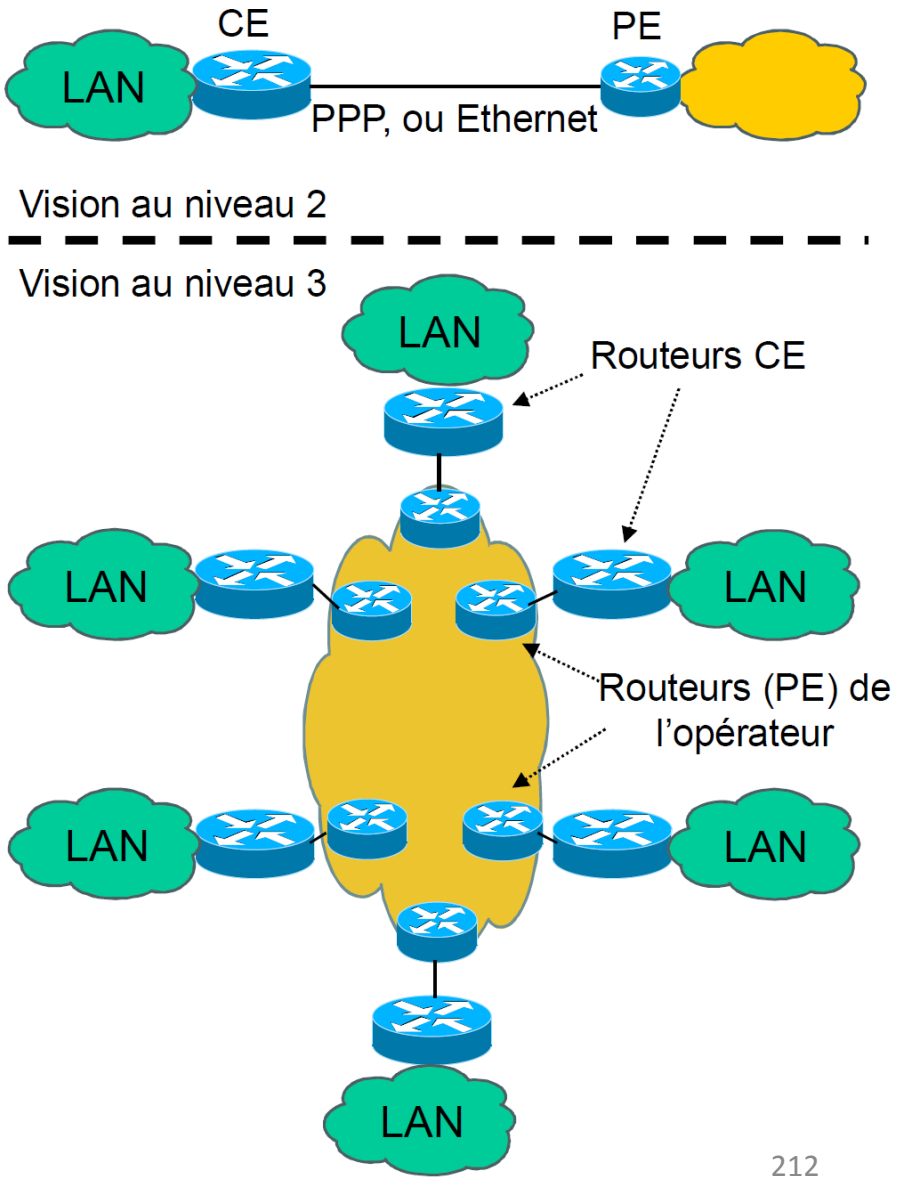
Ethernet switché peut étendre le LAN tout en connectant des sites distants à l'échelle de la ville ou plus.



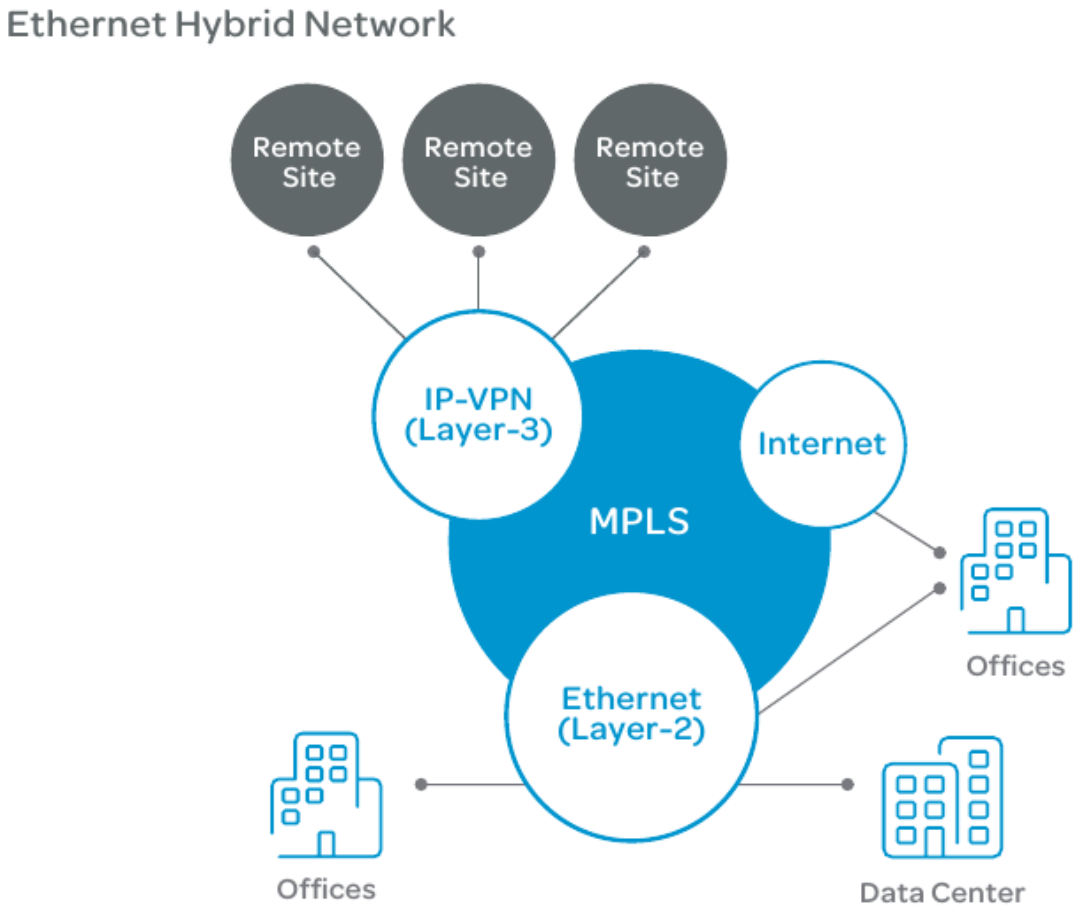
You can use Ethernet Virtual Private Line (EVPL) in your Metro or across a Wide Area Network, nationwide or globally and potentially save since unlike other private line services, you only pay based on the ports or connections to the network and not mileage.

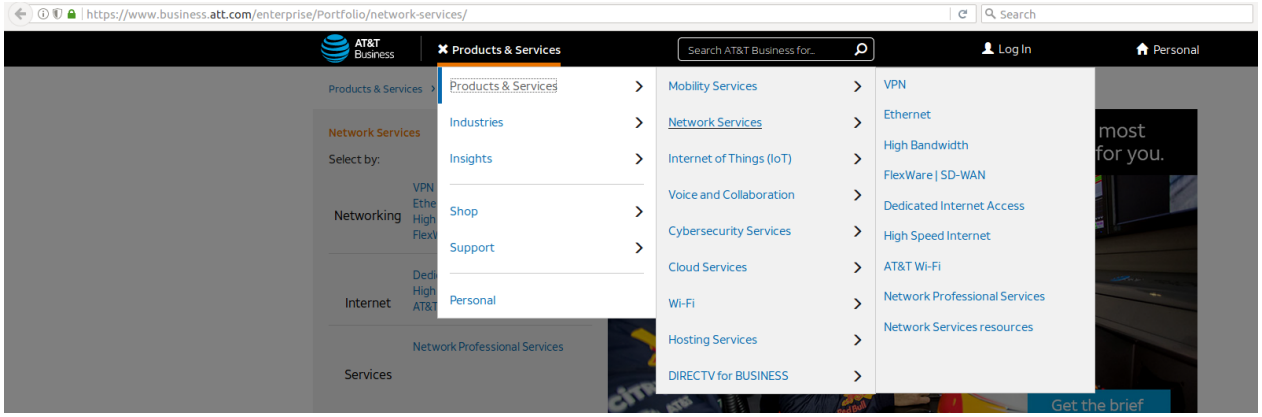
Services WAN de niveau 3

- **Au niveau 2 les routeurs CE ne se voient pas**
  - Le routeur CE voit le routeur PE
  - Le protocole de niveau 2 peut être quelconque
    - PPP, Ethernet, etc…
- **Au niveau 3 les routeurs CE ne se voient pas directement**
  - Le routeur CE voit le routeur PE
  - Les routeurs CE envoient tout le trafic externe ver le routeur PE
    - Par une table statique (la route par défaut)
- **Les équipements du réseau de l'opérateur (PE) sont concernés par le protocole de niveau 3 IP**
  - Le client a externalisé son routage : c'est l'opérateur qui s'en charge
  - Solution appelée VPRN (Virtual Private Routed Network)



Hybrid network with L2- and L3- multi-point VPNs





MPLS ET OFFRES DE SERVICE

Sources:

- *VPLS Technical Tutorial*, Technology White Paper, Alcatel-Lucent, 2010
- *Enabling High-Performance Data Services with Ethernet WAN and IP VPN*, IDC White Paper, 2011
- The CCIE R&S: http://aitaseller.wordpress.com/2012/09/10/mpls-layer-3-vpns/

Plan général du cours

I. **Organisation des opérateurs de l'Internet**

II. **TCP et Qualité de service**

III. **Commutation par circuits virtuels**

IV. **Commutation par VC dans le monde IP : MPLS**

IV.1. Fonctionnement de MPLS

IV.2. Ingénierie de trafic avec MPLS : MPLS-TE

IV.3. Offres de service MPLS : les VPN basés sur MPLS

IV.3.a. IP-VPN

IV.3.b. Ethernet-VPN : VPLS

Préambule: principe des Virtual Private Networks (VPNs)

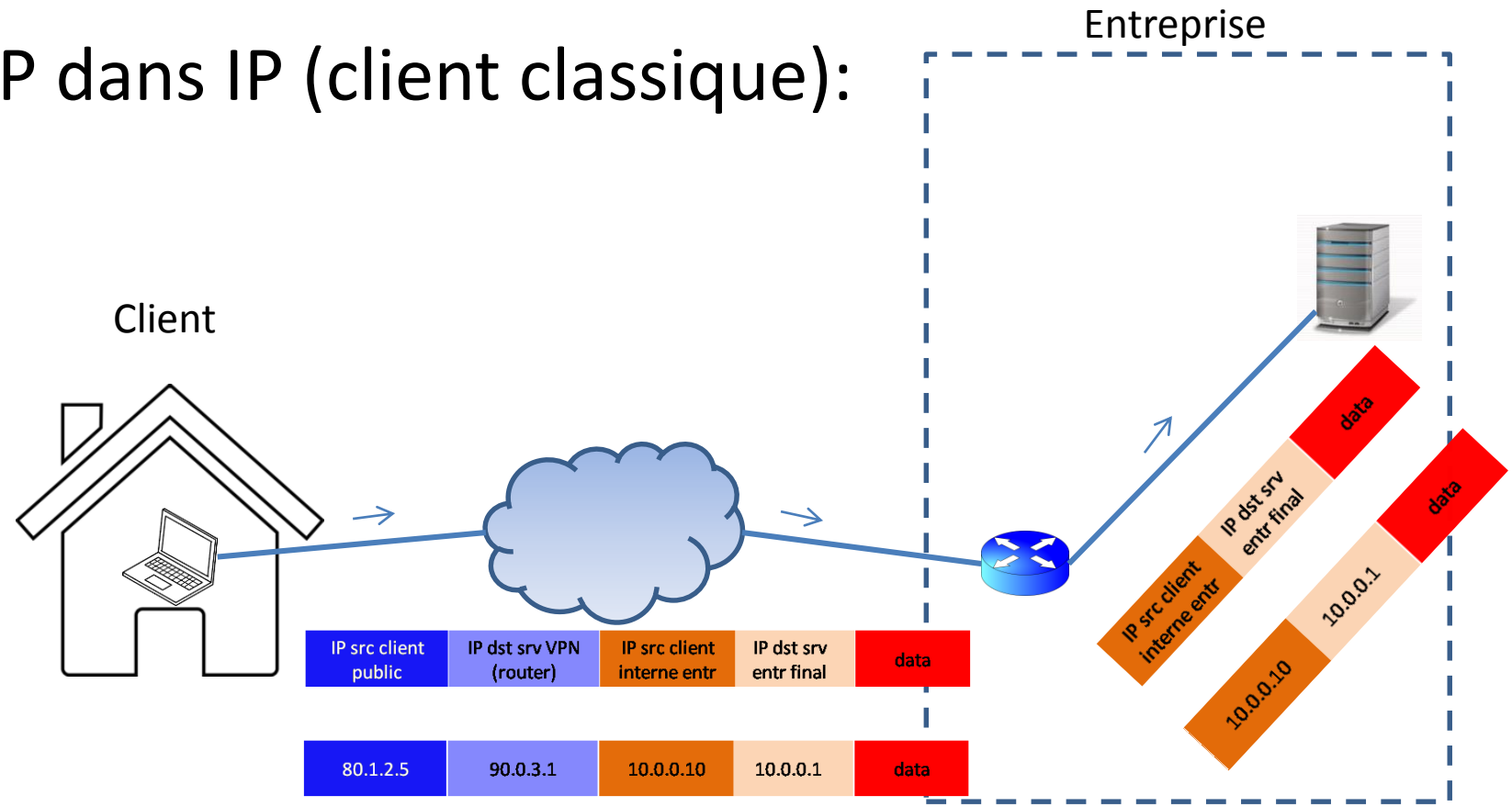
- Principe du ***tunneling*** (tunnelage): un paquet généré à la couche i (1 à 5) est transporté par un protocole de couche j, avec j=i à 5, comme si c'était des données de couche j+1.

- Exemples:

Protocole de couche i	Encapsulé par j	Comme si c'était un paquet de couche j+1
IP (3)	IP (3)	TCP (4)
TCP (4)	TCP (4)	Appli (5)
Ethernet (2)	MPLS (2.5)	IP (3)

Préambule: principe des VPNs

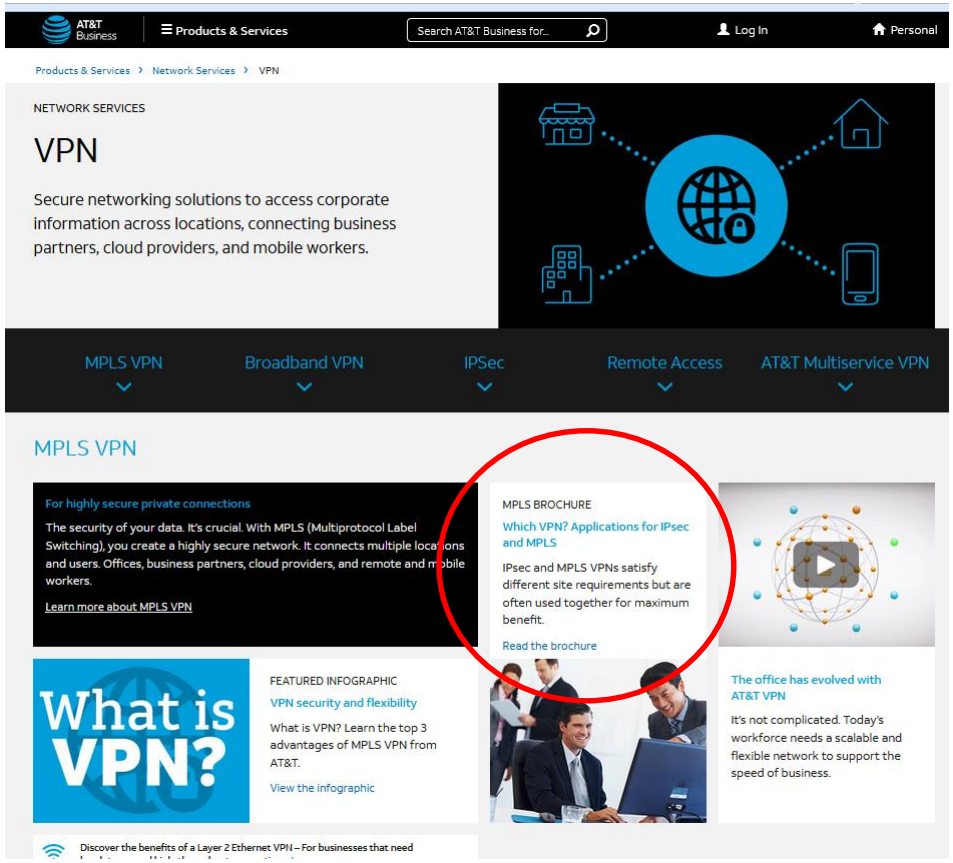
- IP dans IP (client classique):



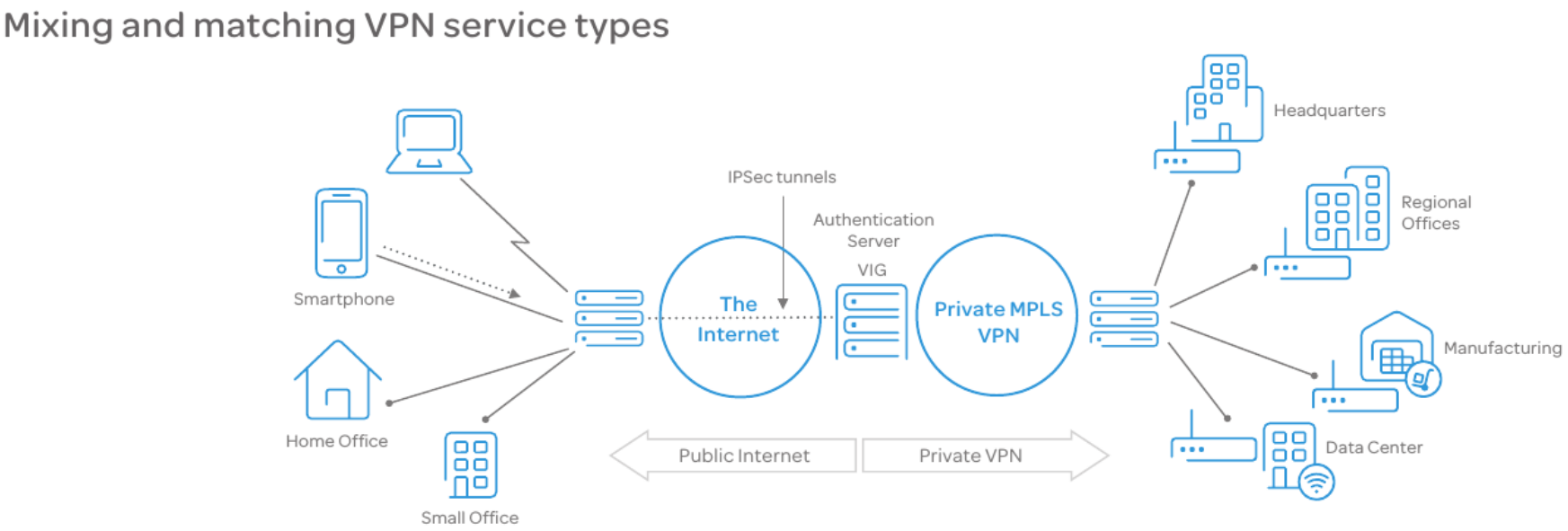
--> Le client peut faire les mêmes opérations que s'il était dans les locaux de l'entreprise.
--> Les ISPs n'ont pas (et ne peuvent pas si IPSec) à lire le paquet destiné au réseau d'entreprise (adr IP privées utilisables).
--> Les ISPs traitent le paquet entreprise comme un payload de couche 4 normal.

Les solutions VPN: IPsec ou MPLS

- https://www.business.att.com/solutions/Family/network-services/vpn/
- https://www.business.att.com/content/productbrochures/vpn-applications-product-brief.pdf



Les solutions VPN: IPsec ou MPLS



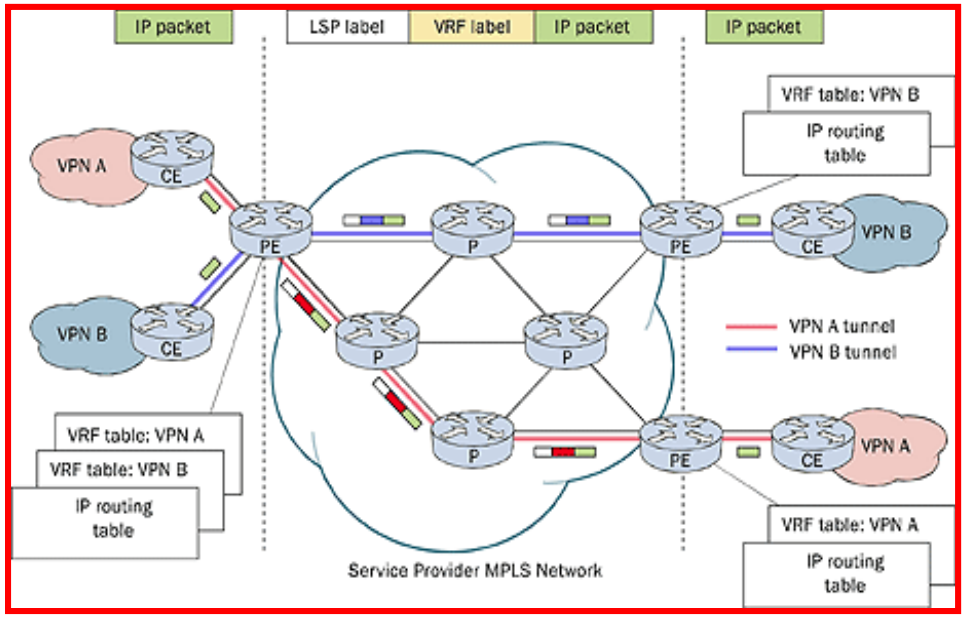
Une application de MPLS: Différents types de VPN basés MPLS

- MPLS permet le support des **VPNs multipoints**: service fournis par les ISPs aux entreprises pour interconnecter plusieurs de leurs LANs distants comme si elles possédaient un routeur ou switch **dédié**:
  - Layer-3 VPN: l'entreprise voit le réseau de l'opérateur comme un routeur qui lui appartient
  - Layer-2 VPN: l'entreprise voit le réseau de l'opérateur comme un switch qui lui appartient

- --> Permet à l'ISP de gérer ses ressources (faire son TE) sans en informer les clients, pourvu que leurs contraintes de QoS dans le contrat soient respectées

Principe de fonctionnement : la séparation des tables de routage ou d'adressage

- Pour la confidentialité entre VPNs: les sites clients (leurs tables) interconnectés par des LSP MPLS différents
- La table dépend du type de VPN:
  - L3 VPN : les tables contiennent les préfixes IP et s'appellent Virtual Routing and Forwarding Tables (VRF). Les VRFs sont simplement des tables de routage dédiées.
  - L2 VPN : Virtual Forwarding Tables (VFT), contiennent les adresses de couches 2, ou les DLCI de FR, etc…
  - VPLS : contiennent les adresses MAC Ethernet, et les VLAN IDs si VLAN, mappées aux LSPs menant aux autres sites. Même rôle que les MAC tables dans les switches Ethernet.



Les avantages des VPN basés MPLS

- Service sans connection : Internet doit son succès à la techno TCP/IP basique: pas d'action nécessaire avant la communication entre 2 hôtes. Un VPN basé MPLS supprime le besoin du d'encryption pour assurer la confidentialité, donc beaucoup moins de complexité.
- Passage à l'échelle : seuls les routeurs PE mémorisent les routes des VPN qu'ils gèrent. Les routeurs P non. Donc pas d'augmentation de complexité dans le coeur du réseau avec l'augmentation de clients.
- Sécurité : Les paquets d'un VPN ne peuvent pas par erreur aller dans un autre VPN:
  - Sur le bord, assure que les paquets d'un client sont placés dans le bon VPN.
  - Dans le coeur, le trafic des VPN reste séparé. Le spoofing (essai d'avoir accès à un routeur PE) quasi-impossible car les paquets reçus des clients sont IP. Ces paquets IP doivent être reçus sur une interface ou sous-interface particulière attachée à un seul label VPN.
- Adressage flexible : beaucoup de clients utilisent des plages d'adresses privées, et ne veulent pas les convertir en public (temps et argent). Les VPN MPLS permettent à ces clients de continer à utiliser ces adresses privées sans besoin de NAT.
- Qualité de service : permet de satisfaire 2 contraintes importantes pour les VPN:
  - Performance prédictible et implémentation de politiques pour SLA
  - Supporte plusieurs niveaux de service dans un VPN MPLS

  Le trafic est classifié et labellisé au bord pour être traité de façon différenciée selon les classes (avec différents délais ou proba d'abandon par exemple).

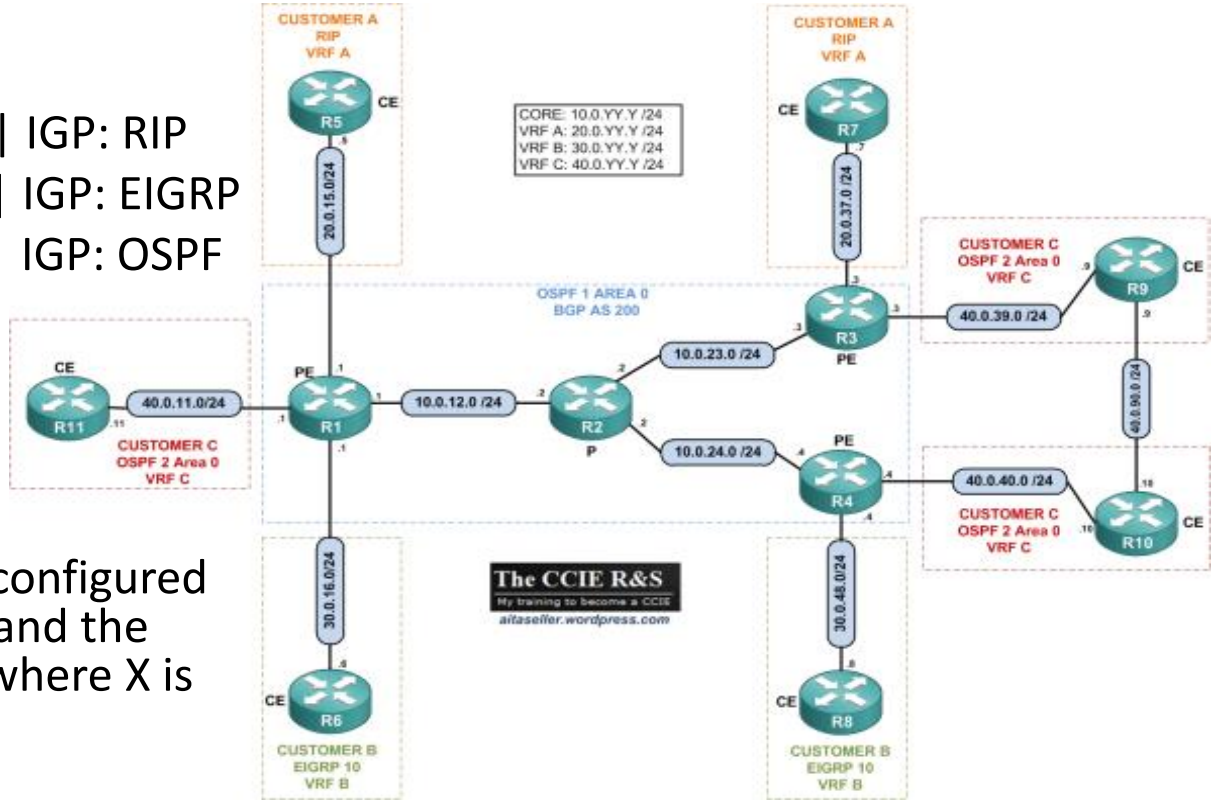
Plan général du cours

I. **Organisation des opérateurs de l'Internet**
II. **TCP et Qualité de service**
III. **Commutation par circuits virtuels (VC) : le cas d'ATM**
IV. **Commutation par VC dans le mode IP : MPLS**
    IV.1. Fonctionnement de MPLS
    IV.2. Ingénierie de trafic avec MPLS : MPLS-TE
    IV.3. Offres de service MPLS : les VPN basés sur MPLS
        IV.3.a. IP-VPN
        IV.3.b. Ethernet-VPN : VPLS

I. **Accès : Réseaux optiques**
II. **Accès : Technologies xDSL**

Configuration d'un IP-VPN basé MPLS sur un exemple Cisco

- Platform/IOS: Cisco 691/12.4(15)
- **VRFs:**
  - Customer A: VRF A | IGP: RIP
  - Customer B: VRF B | IGP: EIGRP
  - Customer C: VRF C | IGP: OSPF
  - ISP: Core IGP: OSPF
    MP-BGP AS 200
- **Addressing:**
  - All the routers are configured with a Loopback IP and the format X.X.X.X /32 where X is the router number.



The CCIE R&S: http://aitaseller.wordpress.com/2012/09/10/mpls-layer-3-vpns/

Les VPN MPLS sont une combinaison de différents protocoles et technologies: ils s'appuient sur MPLS et peuvent gérer différents protocoles de routage pour les clients. Ces VPN s'appuient sur le protocole MP-BGP (multiprotocol BGP) pour échanger les routes VPN. MP-BGP est une évolution de BGP gérant les VRF. MP-BGP gère une nouvelle famille d'adresses nommées "VPNv4" (VPN IPv4).

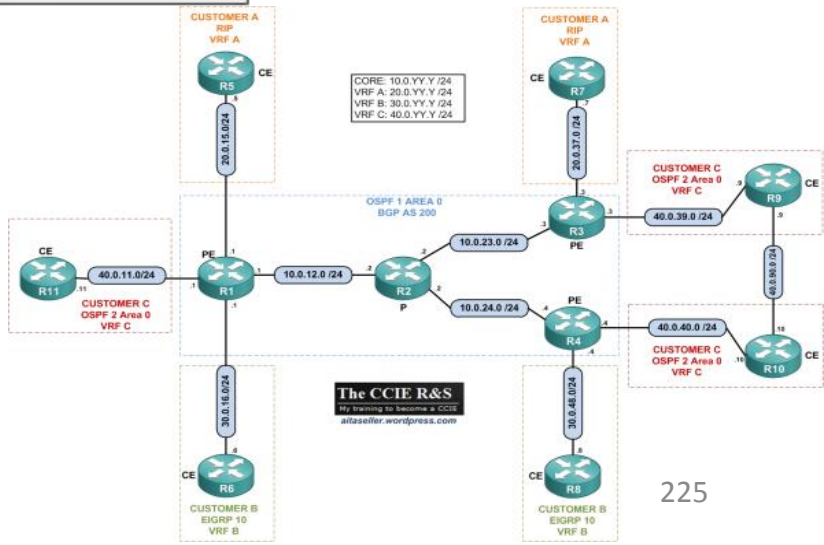
1. Configuration MPLS du coeur de l'ISP

Activer LDP sur tous les routeurs du coeur:

```
R1-R4
conf t
router ospf 1
!!!ENABLE LDP on all OSPF interfaces!!!
mpls ldp autoconfig
```

Les voisins LDP échangent préfixes et labels associés, constituant la FLIB :

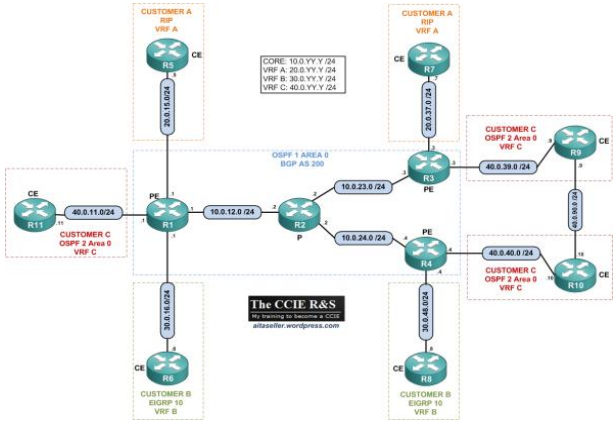
```
R1#sh mpls forwarding-table
Local  Outgoing    Prefix           Bytes tag   Outgoing    Next Hop
tag    tag or VC   or Tunnel Id     switched    interface
16     Pop tag     2.2.2.2/32       0           Fa1/0       10.0.12.2
17     17          3.3.3.3/32       0           Fa1/0       10.0.12.2
18     18          4.4.4.4/32       0           Fa1/0       10.0.12.2
19     Untagged    10.0.24.0/24     0           Fa1/0       10.0.12.2
20     Untagged    10.0.23.0/24     0           Fa1/0       10.0.12.2
```



2. Etablissement des sessions MP-BGP entre les PE

Etablissons les sessions MP-BGP entre les PE (pour ensuite échanger les routes entre les CE).
On établit un maillage complet de sessions MP-BGP entre les 3 PE en désactivant la famille d'@ IPv4 unicast (car ici on échangera seulement des préfixes VPNv4).

```
R1,R3,R4
conf t
router bgp 200
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor X.X.X.X remote-as 200
 neighbor X.X.X.X update-source Loopback0
!
 address-family vpnv4
  neighbor X.X.X.X activate
  neighbor X.X.X.X send-community extended
```



On vérifie sur R1 que les sessions sont établies entre tous les PE:

```
R1#sh bgp vpnv4 unicast all summary
BGP router identifier 1.1.1.1, local AS number 200
BGP table version is 1, main routing table version 1

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
3.3.3.3         4   200      16      17        1    0    0 00:13:33        0
4.4.4.4         4   200      15      15        1    0    0 00:11:51        0
```

Etape suivante: configurer les VRF respectifs sur les différents PE, et les protocoles de routage CE-PE.

3. Configuration des VRF

- Pour créer une nouvelle VRF: commande `ip vrf <VRF_NAME>` pour entrer en mode config de VRF.

- Puis configuration d'un *Route Distinguisher (RD)* pour chaque VRF : identifie le PE et la VRF dont la route est issue.
- RD+IPv4 prefix = VPNv4 prefix

- Ici on prendra : - VRF A: rd=200:1 - VRF B: rd=200:2 - VRF C: rd=200:3
- Donc : la route 20.0.37.0/24 dans VRF A de R3 est annoncée en 200:1: 20.0.37.0/24

- L'attribut *Route Target (RT)* (64 bits) est attaché à chaque route exportée.
- Quand une route VPNv4 est reçue par un PE, c'est le RT associé qui détermine s'il l'insère dans une de ses VRFs. Le RT peut donc être utilisé pour partager des routes entre des VPNs différents, et 2 routes d'un même VPN peuvent avoir des RDs différents.

- Quand un routeur PE reçoit un préfixe VPNv4, le routeur regarde le RT attaché et vérifie s'il a une VRF qui correspond au RT attaché à la route :
  - Si le PE a ce RT dans les RT à importer, le préfix VPNv4 est importé dans la VRF correspondante.
  - Sinon le préfixe est abandonné.

- RT définis dans le mode de config VRF: `route-target export X:Y`

- Dans notre exemple, on utilise les RT suivants pour les VRF:
  VRF A: rt=200:1 pour import et export
  VRF B: rt=200:2 pour import et export
  VRF C: rt=200:3 pour import et export

D'où la configuration des différents VRF sur les PE :

```
R1
!
ip vrf VRFA
 rd 200:1
 route-target export 200:1
 route-target import 200:1
!
ip vrf VRFB
 rd 200:2
 route-target export 200:2
 route-target import 200:2
!
ip vrf VRFC
 rd 200:3
 route-target export 200:3
 route-target import 200:3

R3
!
ip vrf VRFA
 rd 200:1
 route-target export 200:1
 route-target import 200:1
!
ip vrf VRFC
 rd 200:3
 route-target export 200:3
 route-target import 200:3
!
R4
!
ip vrf VRFB
 rd 200:2
 route-target export 200:2
 route-target import 200:2
!
ip vrf VRFC
 rd 200:3
 route-target export 200:3
 route-target import 200:3
```

On associe ensuite chaque VRF à l'interface CE correspondante: `ip vrf forwarding <VRF NAME>`

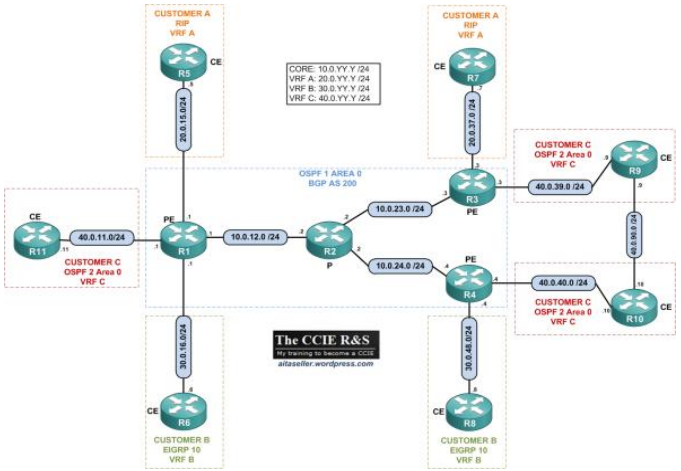
```
R1#sh ip vrf interfaces
Interface              IP-Address      VRF                              Protocol
Fa0/0                  20.0.15.1       VRFA                             up
Fa0/1                  30.0.16.1       VRFB                             up

R3#sh ip vrf interfaces
Interface              IP-Address      VRF                              Protocol
Fa1/0                  20.0.37.3       VRFA                             up
Fa0/1                  40.0.39.3       VRFC                             up

R4#sh ip vrf interfaces
Interface              IP-Address      VRF                              Protocol
Fa0/1                  30.0.48.4       VRFB                             up
Fa1/0                  40.0.40.4       VRFC                             up
```

Chaque routeur CE est placé dans la bonne VRF.

Prochaine étape : Configurer les protocoles de routage PE-CE de chaque site pour que l'ISP et les différents clients puissent échanger leurs préfixes.



4. Configuration des protocoles de routage PE-CE

On utilise ici RIP, EIGRP et OSPF. Chaque CE est déjà configuré et annonce son adresse de loopback.

Commençons avec la config du process RIP tenant compte des VRF :

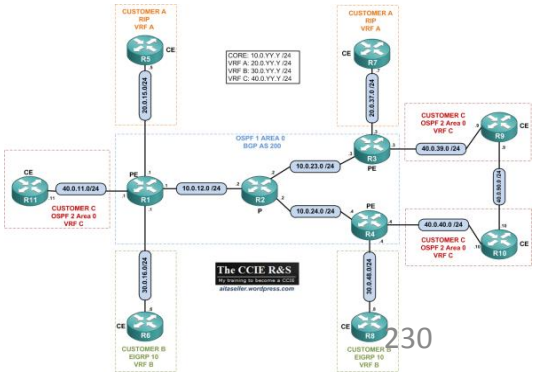
```
R1#sh run | sec rip
router rip
 version 2
 no auto-summary
 !
 address-family ipv4 vrf VRFA
  network 20.0.0.0
  no auto-summary
 exit-address-family
```

On vérifie que R1 a appris l'adresse de loopback de R5 par une mise à jour RIP :

```
R1#sh ip route vrf VRFA 5.5.5.5
Routing entry for 5.5.5.5/32
  Known via "rip", distance 120, metric 1
  Redistributing via rip
  Last update from 20.0.15.5 on FastEthernet0/0, 00:00:01 ago
  Routing Descriptor Blocks:
  * 20.0.15.5, from 20.0.15.5, 00:00:01 ago, via FastEthernet0/0
      Route metric is 1, traffic share count is 1
```

Activation de RIP sur R3 et on vérifie que R3 a appris l'adresse de loopback de R7 :

```
R3#sh ip route vrf VRFA 7.7.7.7
Routing entry for 7.7.7.7/32
  Known via "rip", distance 120, metric 1
  Redistributing via rip
  Last update from 20.0.37.7 on FastEthernet1/0, 00:00:14 ago
  Routing Descriptor Blocks:
  * 20.0.37.7, from 20.0.37.7, 00:00:14 ago, via FastEthernet1/0
      Route metric is 1, traffic share count is 1
```



Pareil pour EIGRP :

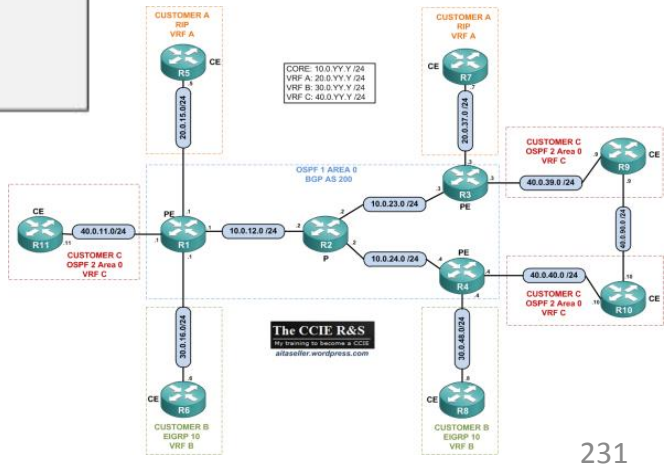
```
R1
router eigrp 1
 no auto-summary
 !
 address-family ipv4 vrf VRFB
  network 30.0.16.1 0.0.0.0
  auto-summary
  autonomous-system 10          -> specify the AS in the routing context
 exit-address-family
```

Et on vérifie que R1 a appris l'adresse de loopback de R6 :

```
R1#sh ip route vrf VRFB 6.6.6.6
Routing entry for 6.6.6.6/32
  Known via "eigrp 10", distance 90, metric 156160, type internal
  Redistributing via eigrp 10
  Last update from 30.0.16.6 on FastEthernet0/1, 00:00:27 ago
  Routing Descriptor Blocks:
  * 30.0.16.6, from 30.0.16.6, 00:00:27 ago, via FastEthernet0/1
!!!OUTPUT OMITTED!!!
```

Idem pour R4 et R8 :

```
R4#sh ip route vrf VRFB 8.8.8.8
Routing entry for 8.8.8.8/32
  Known via "eigrp 10", distance 90, metric 156160, type internal
  Redistributing via eigrp 10
  Last update from 30.0.48.8 on FastEthernet0/1, 00:00:20 ago
  Routing Descriptor Blocks:
  * 30.0.48.8, from 30.0.48.8, 00:00:20 ago, via FastEthernet0/1
!!!OUTPUT OMITTED!!!
```



OSPF: On configure un processus OSPF pour certaines VRF. Pour configurer OSPF comme un protocole de routage PE-CE, il faut un process OSPF pour chaque VRF où on veut de l'OSPF. Coeur de l'ISP utilise OSPF process 1 -> on prend process 2 pour la configuration PE-C.

```
R3
router ospf 2 vrf VRFC
 router-id 3.3.3.3
```

Vérifions que R3 apprend bien l'adresse de loopback de R9 :

```
R3#sh ip route vrf VRFC 9.9.9.9
Routing entry for 9.9.9.9/32
  Known via "ospf 2", distance 110, metric 2, type intra area
  Last update from 40.0.39.9 on FastEthernet0/1, 00:16:33 ago
  Routing Descriptor Blocks:
  * 40.0.39.9, from 40.0.90.9, 00:16:33 ago, via FastEthernet0/1
      Route metric is 2, traffic share count is 1
```

On configure un process OSPF sur R4 et on vérifie que R4 apprend l'adresse de loopback de R10:

```
R4
router ospf 2 vrf VRFC
 router-id 4.4.4.4
```

```
R4#sh ip route vrf VRFC 10.10.10.10
Routing entry for 10.10.10.10/32
  Known via "ospf 2", distance 110, metric 2, type intra area
  Last update from 40.0.40.10 on FastEthernet1/0, 00:17:53 ago
  Routing Descriptor Blocks:
  * 40.0.40.10, from 10.10.10.10, 00:17:53 ago, via FastEthernet1/0
      Route metric is 2, traffic share count is 1
```

Ok, tous ces protocoles de routage PE-CE sont configurés pour les differents sites.
Maintenant il faut configurer l'échange des prefixes VPNv4 entre les PEs dans le but d'échanger les routes des clients entre les différents sites.

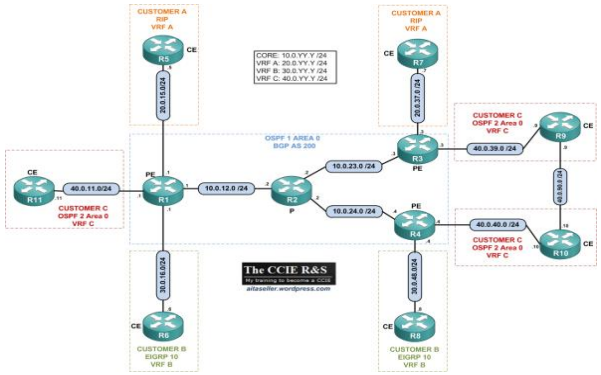
5. Configurer l'échange des préfixes VPNv4 avec MP-BGP

- On a juste besoin de redistribuer chaque protocole de routage CE dans BGP pour le VRF spécifique, et redistribuer BGP dans chaque prot de routage CE pour le VRF spécifique.

- Redistribution mutuelle entre RIP et BGP pour la VRFA :

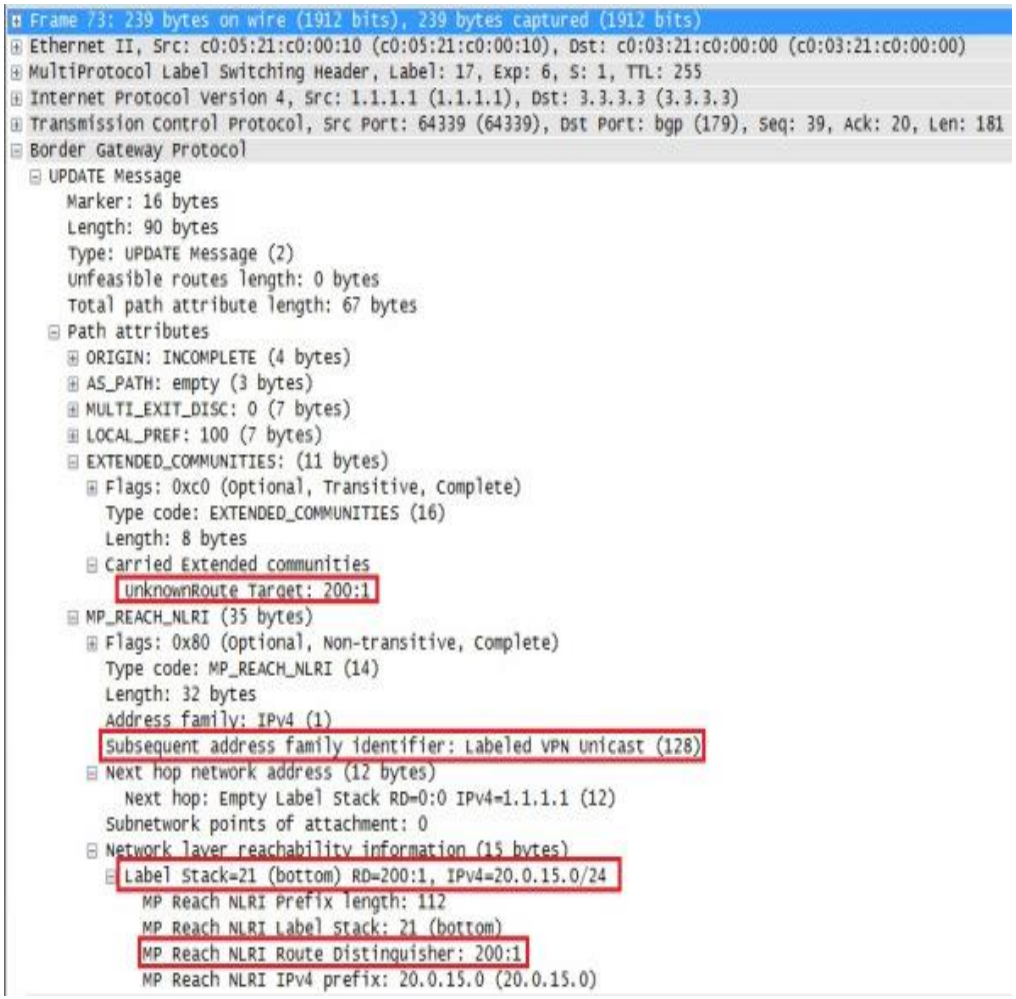
```
R1
router bgp 200
  !
 address-family ipv4 vrf VRFA
  redistribute rip
  no synchronization
 exit-address-family
```

- Dès qu'on redistribue RIP dans le process BGP, R1 annonce les préfixes du client A à tous ses voisins MP-BGP.
- On peut voir ce paquet MP-BGP de mise à jour VPNv4 envoyé par R1 vers R3 pour le préfixe 20.0.15.0/24 :



- Le RT configuré est attaché au préfixe.
- Cette update est envoyée dans le coeur avec MPLS : label 17 correspondant à la loopback de R3

--> Le routeur de coeur P (R2 ici) ne fait que du label switching et pas de routage IPv4.



Regardons si R3 et R4 reçoivent les préfixes annoncés par R1 :

```
R3
41.574: BGP(2): 1.1.1.1 rcvd UPDATE w/ attr: nexthop 1.1.1.1, origin ?,
localpref 100, metric 0, extended community RT:200:1
41.586: BGP(2): 1.1.1.1 rcvd 200:1:20.0.15.0/24
41.602: BGP(2): 1.1.1.1 rcvd UPDATE w/ attr: nexthop 1.1.1.1, origin ?,
localpref 100, metric 1, extended community RT:200:1
41.610: BGP(2): 1.1.1.1 rcvd 200:1:5.5.5.5/32
```

```
R3#sh bgp vpnv4 unicast vrf VRFA
!!!OUTPUT OMITTED!!!

   Network          Next Hop            Metric LocPrf Weight Path
Route Distinguisher: 200:1 (default for vrf VRFA)
*>i5.5.5.5/32       1.1.1.1                  1    100      0 ?
*>i20.0.15.0/24     1.1.1.1                  0    100      0 ?
```

```
R4
02.098: BGP(2): 1.1.1.1 rcvd 200:1:5.5.5.5/32 -- DENIED due to: extended
community not supported;
02.106: BGP(2): 1.1.1.1 rcvd UPDATE w/ attr: nexthop 1.1.1.1, origin ?,
localpref 100, metric 0, extended community RT:200:1
02.114: BGP(2): 1.1.1.1 rcvd 200:1:20.0.15.0/24 -- DENIED due to: extended
community not supported;
```

R4 reçoit les updates mais n'en accepte aucune des 2 puisque on n'a pas configuré de RT d'import en 200:1 sur R4, ce qui est normal car 200:1 correspond au client A, et R4 n'est lié qu'aux clients B et C.

Redistribuons RIP dans BGP sur R3 et vérifions que R1 reçoit les routes et les ajoute à sa table de routage VRF A:

```
R1
27.614: BGP(2): 3.3.3.3 rcvd UPDATE w/ attr: nexthop 3.3.3.3, origin ?,
localpref 100, metric 0, extended community RT:200:1
27.622: BGP(2): 3.3.3.3 rcvd 200:1:20.0.37.0/24
27.638: BGP(2): 3.3.3.3 rcvd UPDATE w/ attr: nexthop 3.3.3.3, origin ?,
localpref 100, metric 1, extended community RT:200:1
27.646: BGP(2): 3.3.3.3 rcvd 200:1:7.7.7.7/32
30.770: BGP(2): Revise route installing 1 of 1 routes for 7.7.7.7/32 ->
3.3.3.3(main) to VRFA IP table
30.770: BGP(2): Revise route installing 1 of 1 routes for 20.0.37.0/24 ->
3.3.3.3(main) to VRFA IP table
```

Redistribution de BGP dans RIP pour que le client A obtienne les routes des 2 sites.
Configurons ça sur R1 et R3 :

```
R1,R3
router rip
!!!OUTPUT OMITTED!!!
!
 address-family ipv4 vrf VRFA
  redistribute bgp 200 metric transparent
!!!OUTPUT OMITTED!!!
```

Vérifions que R5 et R7 reçoivent les routes:

```
R5
43.246: RIP: received v2 update from 20.0.15.1 on FastEthernet0/0
43.246:      7.7.7.7/32 via 0.0.0.0 in 2 hops
43.254:      20.0.37.0/24 via 0.0.0.0 in 1 hops

R5#sh ip route rip
     20.0.0.0/24 is subnetted, 2 subnets
R       20.0.37.0 [120/1] via 20.0.15.1, 00:00:13, FastEthernet0/0
     7.0.0.0/32 is subnetted, 1 subnets
R       7.7.7.7 [120/2] via 20.0.15.1, 00:00:13, FastEthernet0/0
```

```
R7#sh ip route rip
     20.0.0.0/24 is subnetted, 2 subnets
R       20.0.15.0 [120/1] via 20.0.37.3, 00:00:07, FastEthernet0/0
     5.0.0.0/32 is subnetted, 1 subnets
R       5.5.5.5 [120/2] via 20.0.37.3, 00:00:07, FastEthernet0/0
```

Le coeur MPLS du réseau apparaît comme un seul saut. R7 annonce une métrique de 1 pour 7.7.7.7/32 et R5 reçoit 7.7.7.7/32 avec une métrique de 2.
(RIP ajoute une métrique de 1 quand une update et envoyée, et pas quand elle est reçue.)

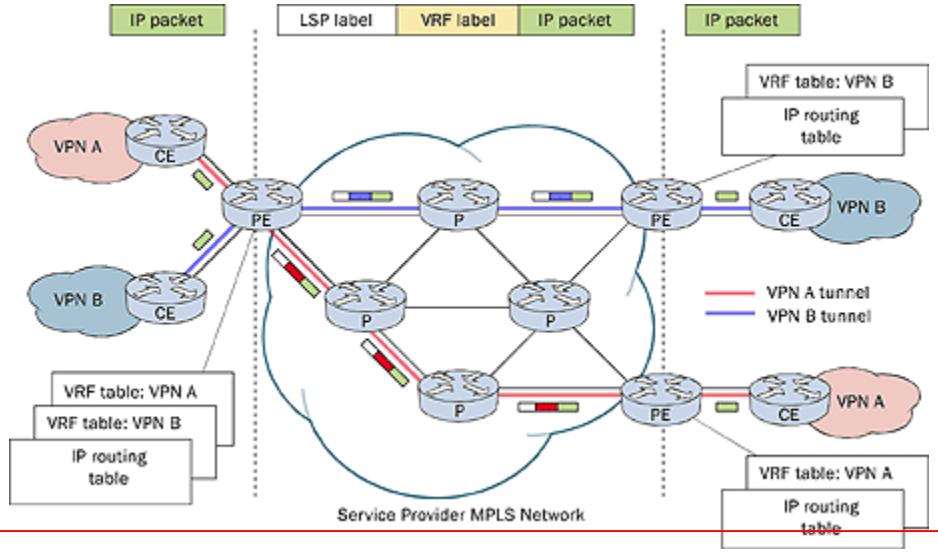
Il semble que les sites du client A peuvent maintenant communiquer. Testons :

```
R5#traceroute 7.7.7.7

Type escape sequence to abort.
Tracing the route to 7.7.7.7

  1 20.0.15.1 36 msec 32 msec 16 msec
  2 10.0.12.2 [MPLS: Labels 17/22 Exp 0] 80 msec 64 msec 76 msec
  3 20.0.37.3 [MPLS: Label 22 Exp 0] 72 msec 52 msec 52 msec
  4 20.0.37.7 116 msec *  128 msec
```

Ok. Label 22 : le label externe 17 est le label de transport, utilisé pour joindre le prochain saut BGP 3.3.3.3, vient de la FLIB. Quand R1 doit joindre l'@ de loopback de R7, il a besoin d'encoder cette info dans un label MPLS pour que R3 sache dans quelle VRF regarder : c'est le label VPN.



Un paquet de client porte donc 2 niveaux de labels dans le coeur :

1. Label externe pour diriger le paquet vers le bon PE dest.
2. Label interne pour indiquer au PE côté dest comment transférer ce paquet au bon routeur CE : dans quelle VRF regarder

```
R1#sh ip bgp vpnv4 vrf VRFA 7.7.7.7
BGP routing table entry for 200:1:7.7.7.7/32, version 11
Paths: (1 available, best #1, table VRFA)
  Not advertised to any peer
  Local
    3.3.3.3 (metric 12) from 3.3.3.3 (3.3.3.3)
      Origin incomplete, metric 1, localpref 100, valid, internal, best
      Extended Community: RT:200:1
      mpls labels in/out nolabel/22
```

Donc quand R2 retire le label de transport 17, R3 reçoit un paquet MPLS avec label 22 et sait qu'il doit chercher le routage à effectuer dans la table de routage VRFA.

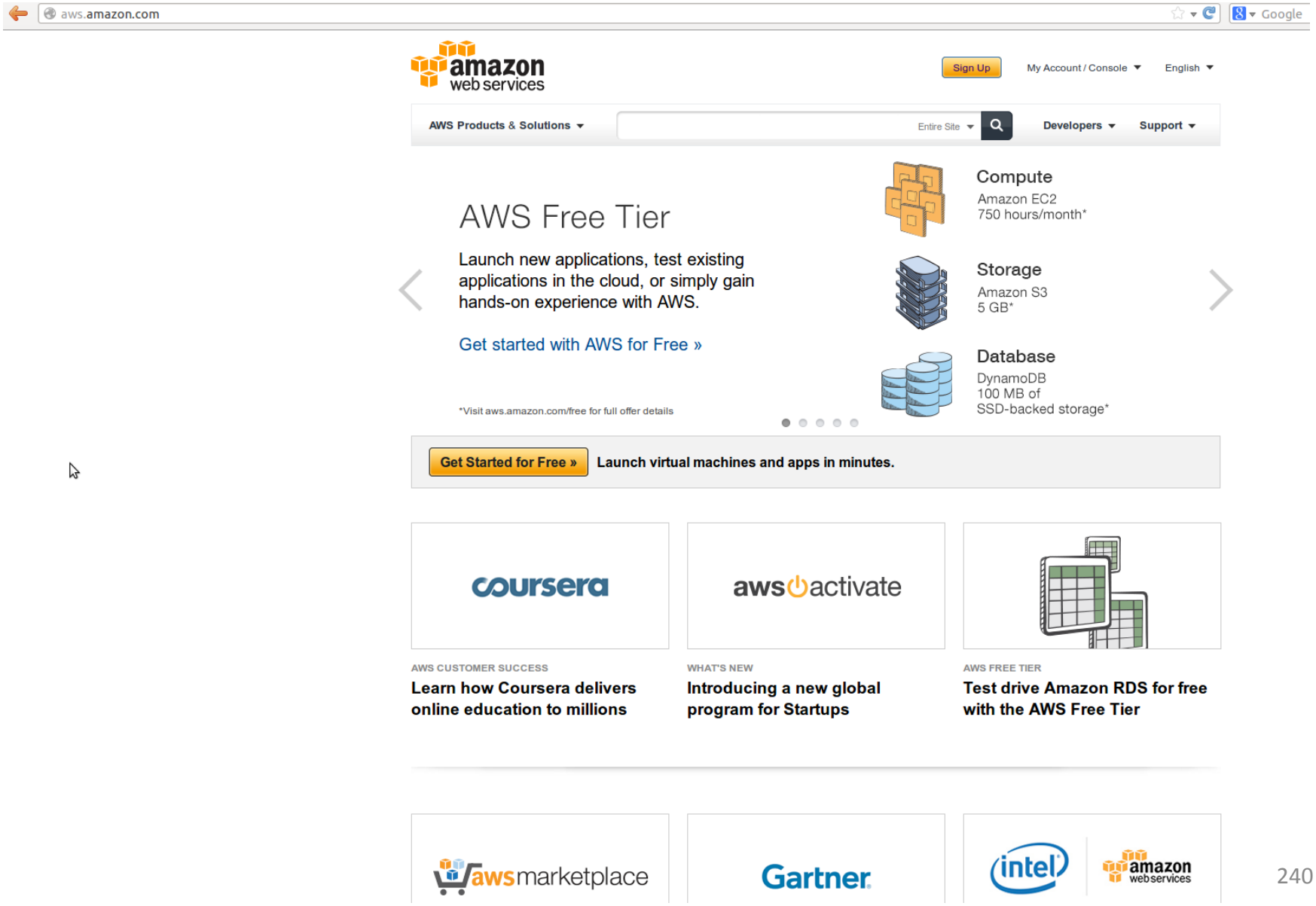
```
R3#sh mpls forwarding-table
Local  Outgoing    Prefix           Bytes tag  Outgoing    Next Hop
tag    tag or VC   or Tunnel Id     switched   interface
22     Untagged    7.7.7.7/32[V]    684        Fa1/0       20.0.37.7
!!!OUTPUT OMITTED!!!
```

Les étapes suivantes correspondent à suivre la même procédure pour les autres clients utilisant d'autres protocoles de routage.:

- Redistribution mutuelle entre EIGRP et BGP pour VRF B
- Redistribution mutuelle entre OSPF et BGP pour VRF C

MPLS, Cloud et Datacenters

MPLS, Cloud et Datacenters



DC consumption

- Worldwide, the digital warehouses use about 30 billion watts of electricity, roughly equivalent to the output of 30 nuclear power plants, according to estimates industry experts compiled for The Times.
- Data centers in the United States account for one-quarter to one-third of that load, the estimates show.

MPLS, Cloud et Datacenters

- Les adresses IPv4 ont été conçues pour 2 fonctions séparées :
  - comme **identificateur** unique des interfaces réseau dans un LAN
  - comme **localisateur** utilisé par le routage, pour identifier l'endroit où se trouve une interface réseau en dehors du LAN
- La hiérarchisation de l'adressage IP permet de ne pas augmenter la taille des tables de routage avec le nombre de machines : adresses réseaux (pas machines) stockées. Au « cœur » de l'Internet, les adresses réseaux sont sur peu de bits, et au fur-et-à-mesure qu'on approche de la périphérie, on se concentre sur des sous-réseaux de plus en plus fins.

- Problème : dans les datacenters, les VMs des clients peuvent migrer de serveurs physiques à d'autres, en fonction des problèmes matériels ou des économies d'énergie, et peuvent également migrer vers un autre DC : c'est le *multihoming*. Ex : serveur Web hébergé dans le Cloud

  -> L'adresse du noeud doit être plus un identifieur qu'un localisateur physique, de façon à ne pas devoir changer quand il y a un déplacement physique de la VM.
- IPv4 ne permet pas cela.

  -> Le protocole LISP est utilisé dans ce but, et basé sur le principe de MPLS : une adresse sert à identifier les noeuds (ex: adresse IPv4 en bout de VPN), une adresse à identifier la localisation (ex: label MPLS).
- Le tunneling permet de séparer les 2 espaces d'adresses, avec adresse localisante comme entête externe de l'adresse identifiante, en entête interne.

Plan général du cours

I. **Organisation des opérateurs de l'Internet**

II. **TCP et Qualité de service**

III. **Commutation par circuits virtuels**

IV. **Commutation par VC dans le monde IP : MPLS**

IV.1. Fonctionnement de MPLS

IV.2. Ingénierie de trafic avec MPLS : MPLS-TE

IV.3. Offres de service MPLS : les VPN basés sur MPLS

IV.3.a. IP-VPN

IV.3.b. Ethernet-VPN : VPLS

Ethernet dans le WAN

- Plusieurs architectures possibles pour porter des trames Ethernet dans le WAN :
  - Extension du protocol Ethernet natif
  - Utilisation de MPLS

Extension du protocol Ethernet natif

- Dans le réseau metro (IEEE 802.1): *Provider Bridged Networks*
  - Le réseau metro comprend des bridges/switches Ethernet

    Un protocol ST est utilisé pour établir un ou plusieurs arbres couvrant.
  - Chaque arbre fournit un chemin entre chaque sites client du même VLAN.
  - Problème : le passage à l'échelle dans le domaine metro :
    - seulement un nombre limité de VLAN est supporté
    - explosion de la taille des tables d'adressage (*MAC address table*)

Extension du protocol Ethernet natif

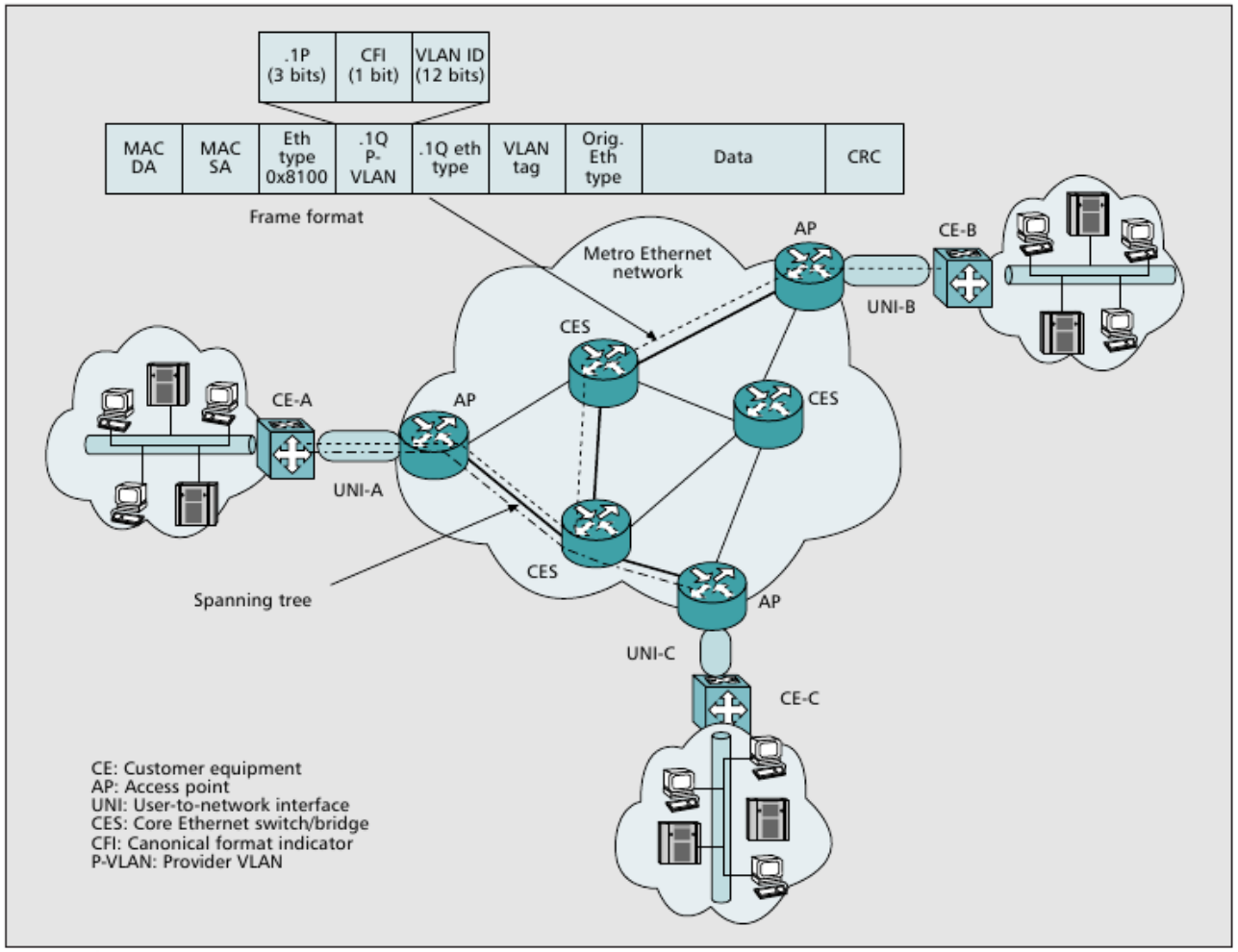


Figure 2. *Network scenario 2: a provider bridged network.*

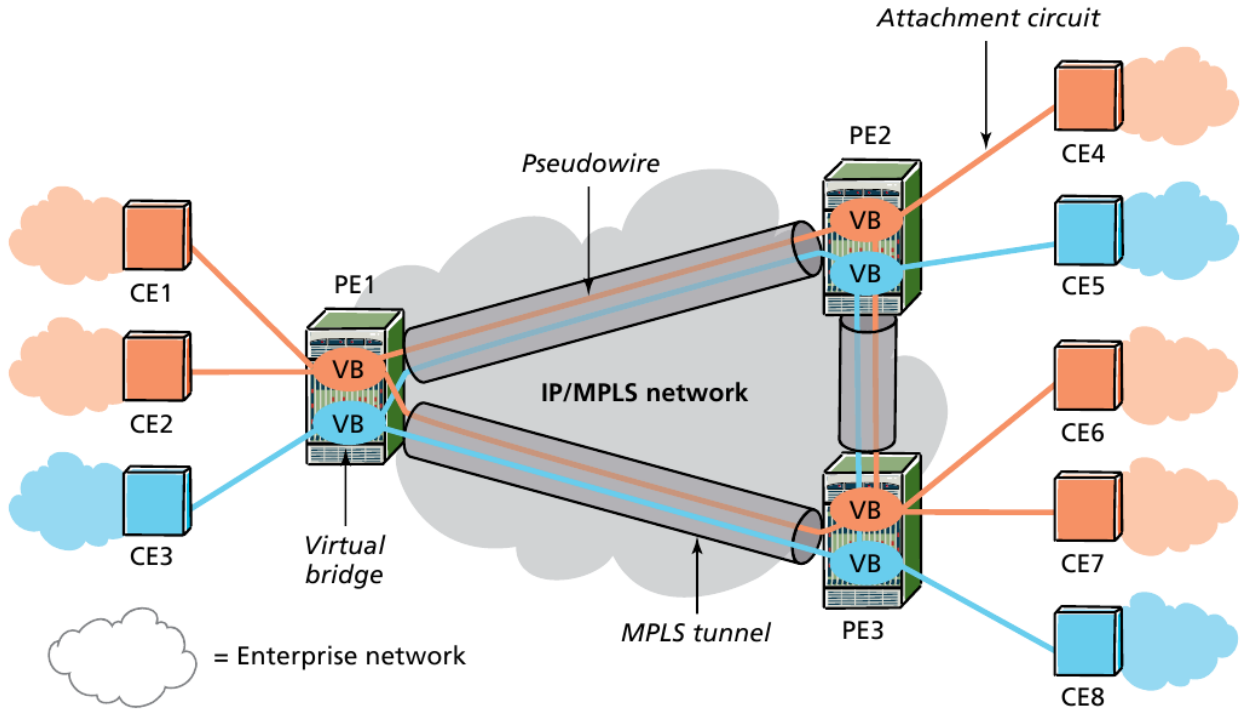
WAN Ethernet basé sur MPLS : Virtual Private LAN Services - VPLS

- https://www.business.att.com/solutions/Family/network-services/ethernet/
- https://www.business.att.com/content/productbrochures/Ethernet_Services_from_ATT.pdf

- VPLS est un type de VPN multi-point de couche 2
- Tous les sites d'un client donné, dans un VPN VPLS, apparaissent comme étant sur le même LAN
- VPLS utilise l'interface Ethernet du client
- Un réseau VPLS comprend des CE, PE et un coeur MPLS:
  - Le CE est routeur ou switch localisé dans les locaux du client. Ethernet est l'interface entre le CE et le PE.
  - Les VPN sont gérés par les PE.
  - Comme VPLS est un service de couche 2 (Ethernet), les PE doivent assurer l'apprentissage des adresses MAC, la commutation et la diffusion par VPN.
  - Le coeur MPLS interconnecte les PE, ne participe pas à la gestion des VPN. Le trafic est juste commuté par les labels MPLS.

WAN Ethernet basé sur MPLS : Virtual Private LAN Services - VPLS

- Un maillage complet de LSP bi-directionnels (appelés *Pseudo-wires*) est créé entre tous les PE d'une instance VPLS.



WAN Ethernet basé sur MPLS : Virtual Private LAN Services - VPLS

- PW = une paire de 2 LSPs dans des directions opposées

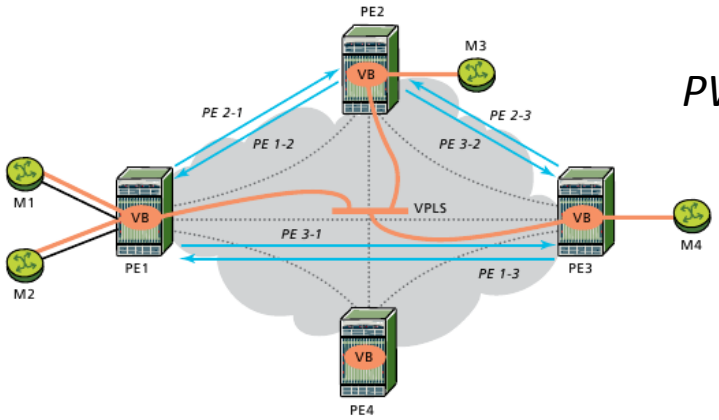
  --> Permet à un PE d'apprendre les @ MAC: quand un PE reçoit une trame Ethernet avec une @MAC source inconnue, le PE sait sur quel PW elle est arrivée, et donc cette adresse sera joignable.

- Le PE implémente un switch pour chaque instance VPLS :
  - Grâce à la FLIB pour chaque instance VPLS,
  - La FLIB est remplie des correspondances @MAC/labels LSP (PW)

VPLS : comment ça marche ?

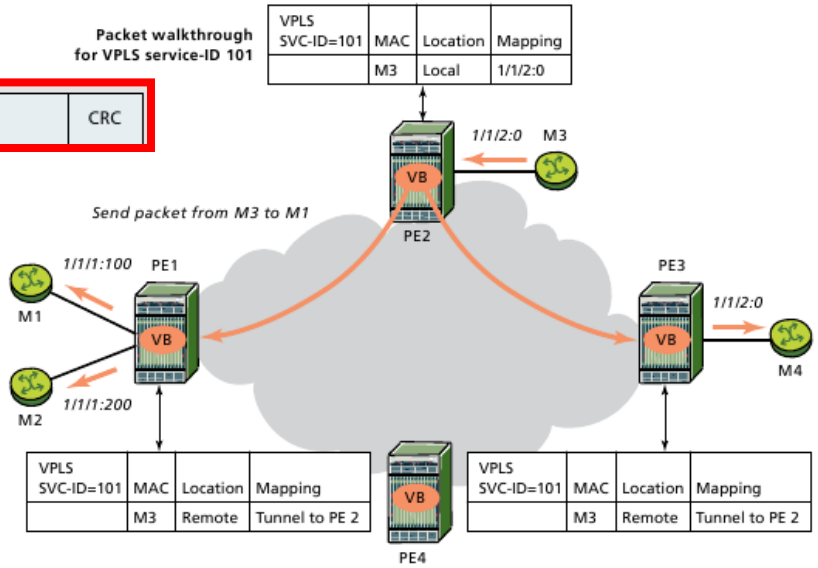
- Création des PWs:

  Une instance VPLS identifiée par Service-identifier 101 (Svc-id 11) -> créer les PW entre PE1, PE2 et PE3



PW signaling

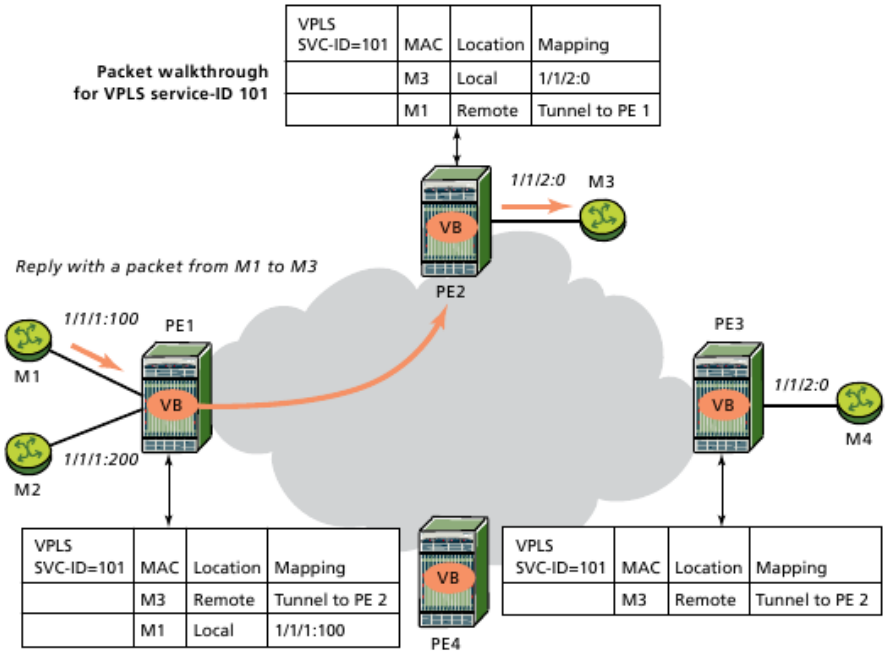
- MAC learning and packet forwarding



VPLS learning

VPLS : comment ça marche ?

- VPLS packet forwarding:



Révision

- Principe d’un VPN :

- Principe d’un L3-VPN :

- Principe d’un L2-VPN :

Comparaison des VPN basés sur MPLS, aux niveaux 2 et 3

Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices

- Bénéfices des IP-VPNs:
  - **Externalisation du control du routage** : le SP applique des priorités aux différents types de trafic en les classant dans des CoS différentes.
  - **Accès flexible**
  - **Passage à l'échelle** : supportent de très grands et nombreux réseaux d'entreprise (avec 100aines ou milliers de sites)
  - **Portée étendue**: offerts par un grand nombre de SP donc couverture géographique étendue
  - **Sécurité**: tables de routages séparées

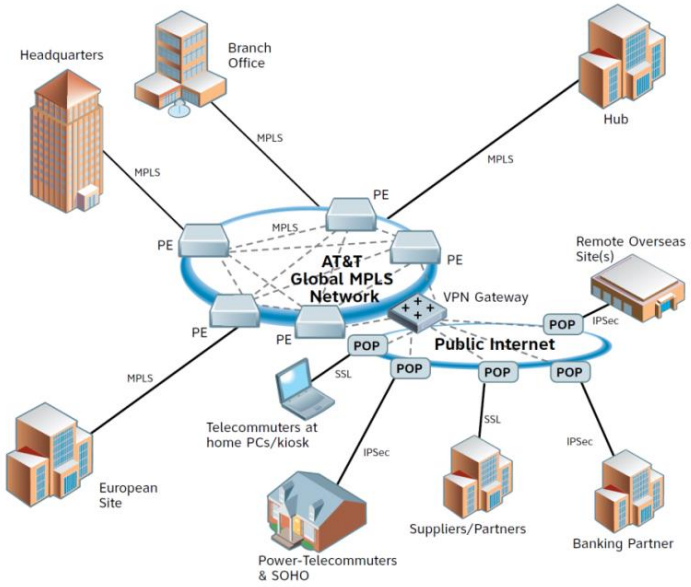
Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices

- Bénéfices des WAN Ethernet :
  - **Contrôle du routage** : quand l'entreprise préfère gérer elle-même son routage
  - **Transparent pour le protocole de couche 3**
  - **Bande passante** plus élevée: de 1MBps à 10GBps

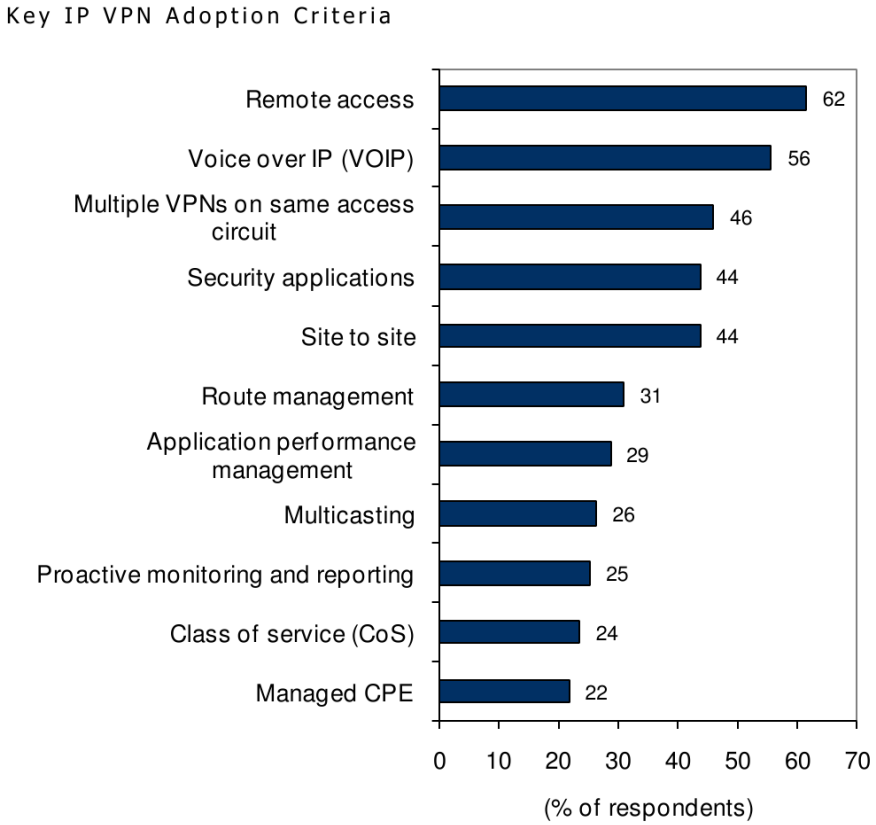
Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices

- 62% des entreprises avec IP VPNs. 56% pour de la VoIP : 25% de plus qu'en 2009
- IP VPNs plutôt pour les "petits" sites, et WAN Ethernet pour les gros sites ou les datacenters avec accès fibre.

Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices



Multiple VPN technology choices working together



U.S. only n = 174

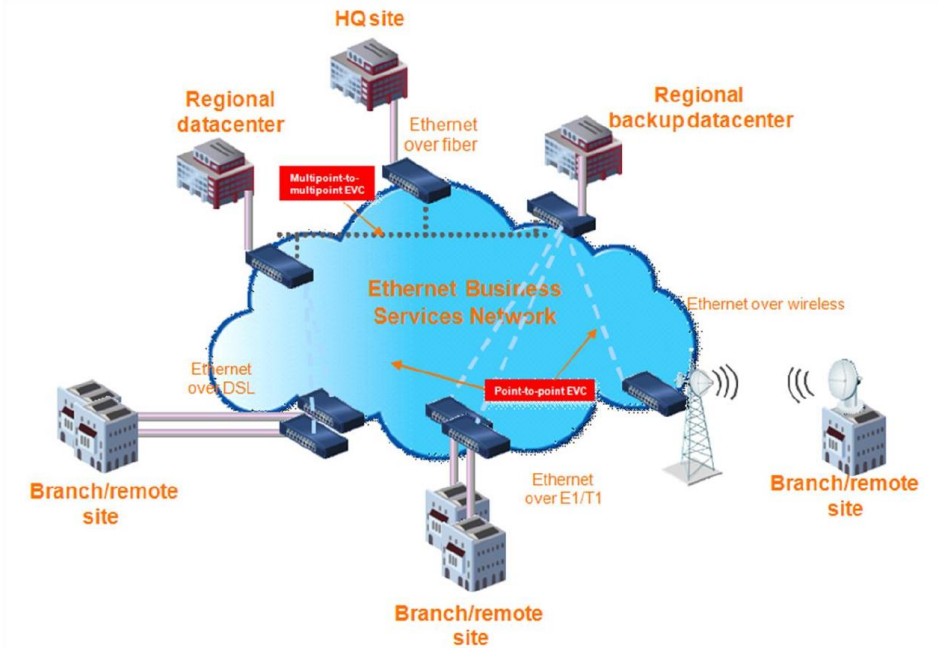
Source: IDC's *WAN Manager Survey*, 2010

Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices

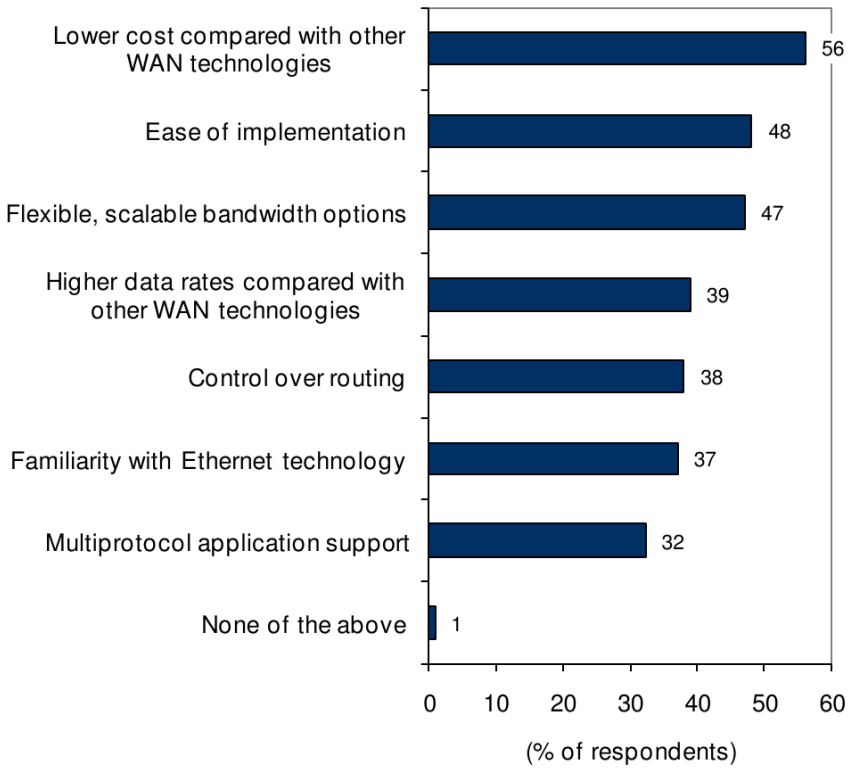
- Adoption de WAN Ethernet par les entreprises pour le coût, la facilité d'implémentation et la familiarité avec Ethernet
- 58% des entr utilisent une solution hybride Ethernet/IP VPN, et augmentation de 33% environ par an.

Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices

Ethernet WAN



Key Ethernet WAN Adoption Criteria



n = 432

Source: IDC's *WAN Manager Survey*, 2010