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 46 : 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 (VC) : le cas d'ATM**
IV. **Commutation par VC dans le mode IP : MPLS**
V. **Accès : Réseaux optiques**
VI. **Accès : Technologies xDSL**

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. Classification

II.2.b. Paramètres-clé de la QoS

II.2.c. Stratégies pour la 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. Ordonnancement des paquets : fonctionnement des files d'attente

II.3.c. Traitement de la congestion : WRED et ECN

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

Plan détaillé du cours

Plan détaillé du cours

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

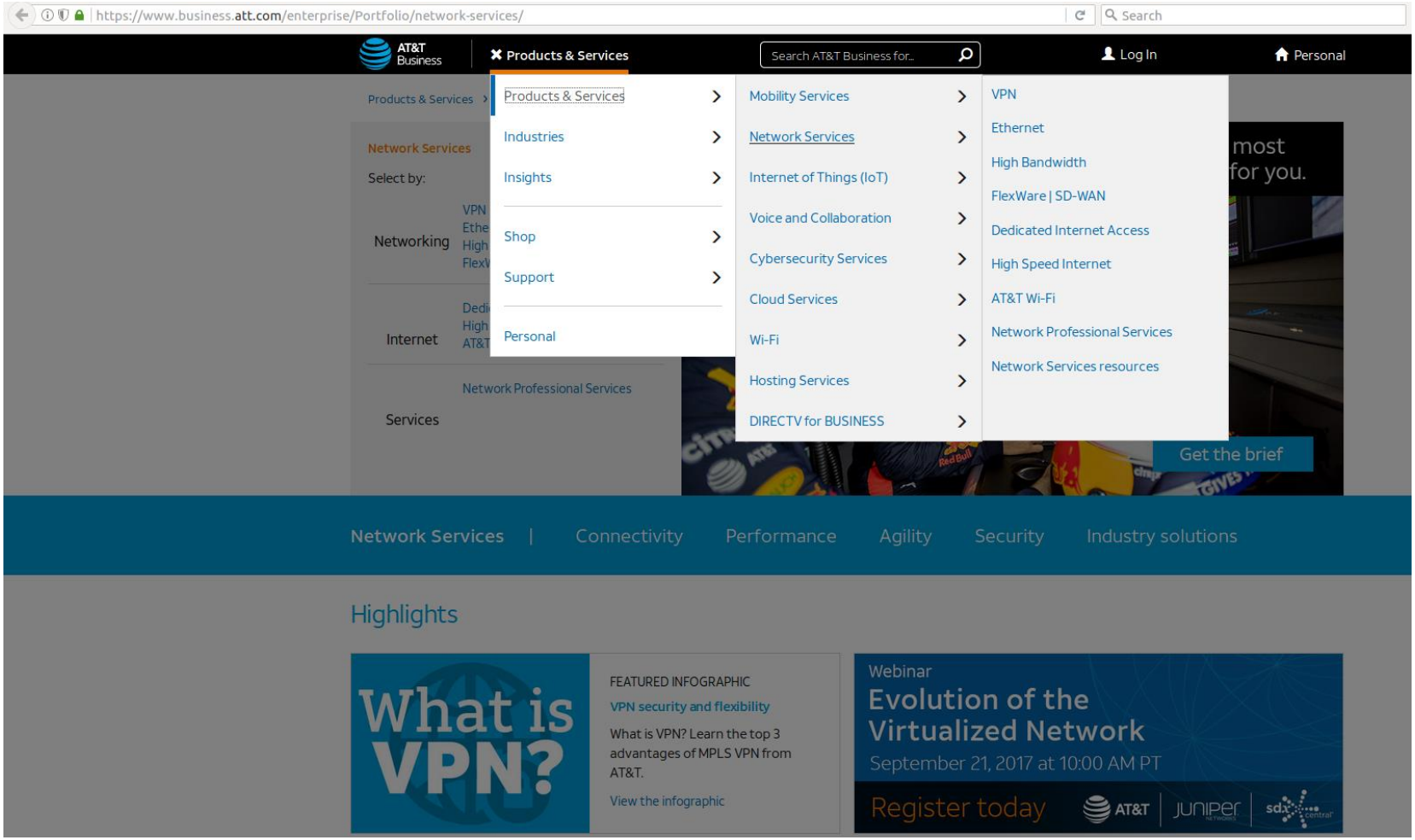
V. Accès : Réseaux optiques

- **V.1. Les différents types de multiplexage**
- **V.2. La hiérarchie numérique synchrone PDH**
- **V.3. La hiérarchie numérique synchrone SDH : réseaux de fibres optiques et redondance pour la fiabilité**

VI. Accès : Technologies xDSL

- **VI.1. Principe : tirer le maximum de la paire de fils de cuivre**
- **VI.2. Technique de TS pour xDSL : l'OFDM**
- **VI.3. Fonctionnement protocolaire entre le FAI et l'abonné (cas dégroupé et non-dégroupé)**

Les offres d'interconnexion de LAN



Références principales

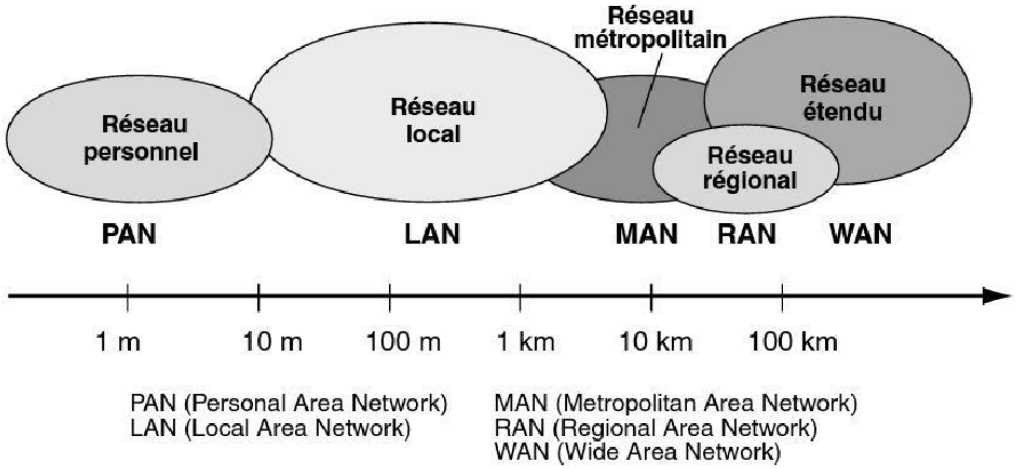
- G. Pujolle, « *Les réseaux édition 2008 »*, ed. Eyrolles
- http://www.cisco.com/en/US/docs/internetworking/technology/handbook/ito_doc.html
- A. S. Tanenbaum, « *Computer Networks, Fourth edition »,* ed. Prentice Hall
- J. F. Kurose and K. W. Ross, « *Computer Networking, a top-down approach, Fifth 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

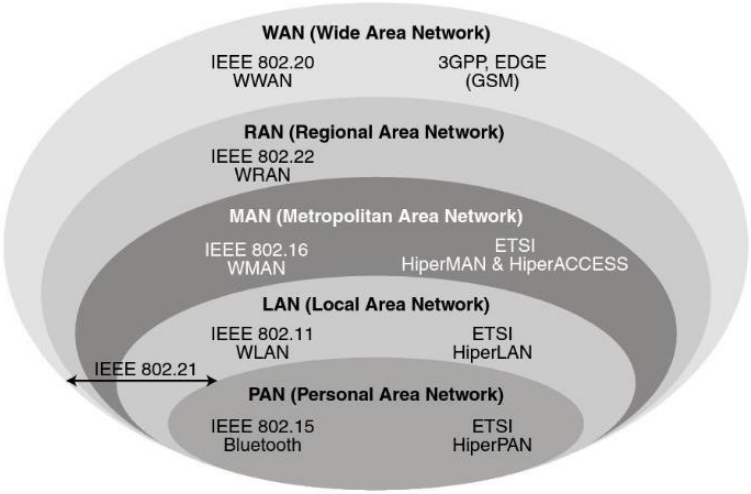
Plan

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 (VC) : le cas d'ATM
IV. Commutation par VC dans le mode IP : MPLS
V. Accès : Réseaux optiques
VI. Accès : Technologies xDSL

Classification des réseaux informatiques

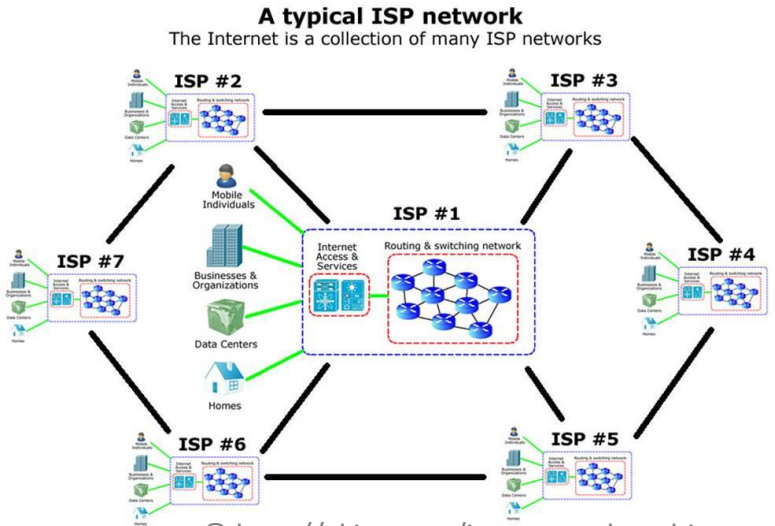




-Personal Area Network: quelques mètres, pour équipements personnels
- Local Area Network: réseaux intra-entreprises, jusqu'à plusieurs Mbps
- Metropolitan Area Networks: interconnexion des entreprises sur un réseau spécialisé à haut débit
-Regional Area Network: 50km pour le sans-fil, beaucoup d'utilisateurs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

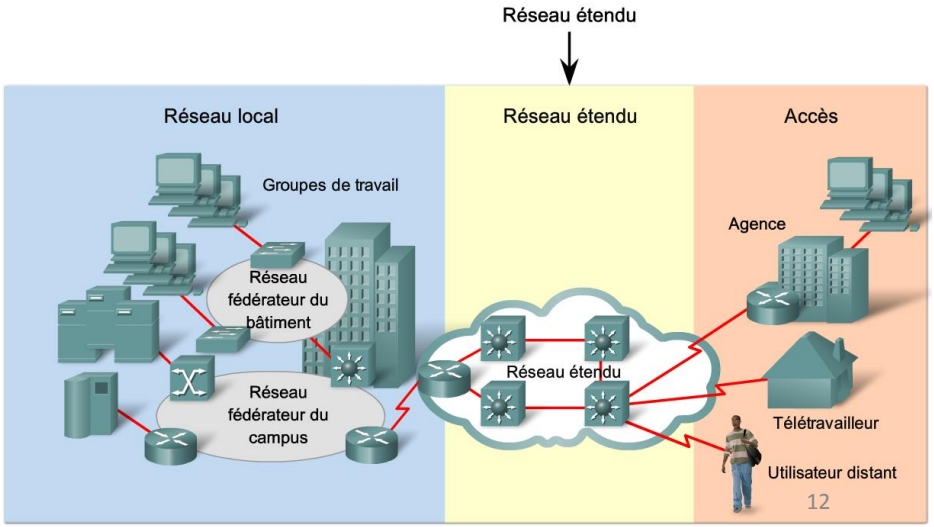
Réseau d'opérateurs et Réseaux étendus: définition

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

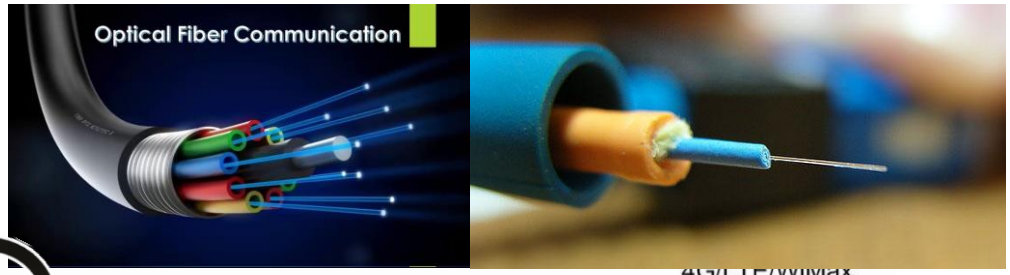


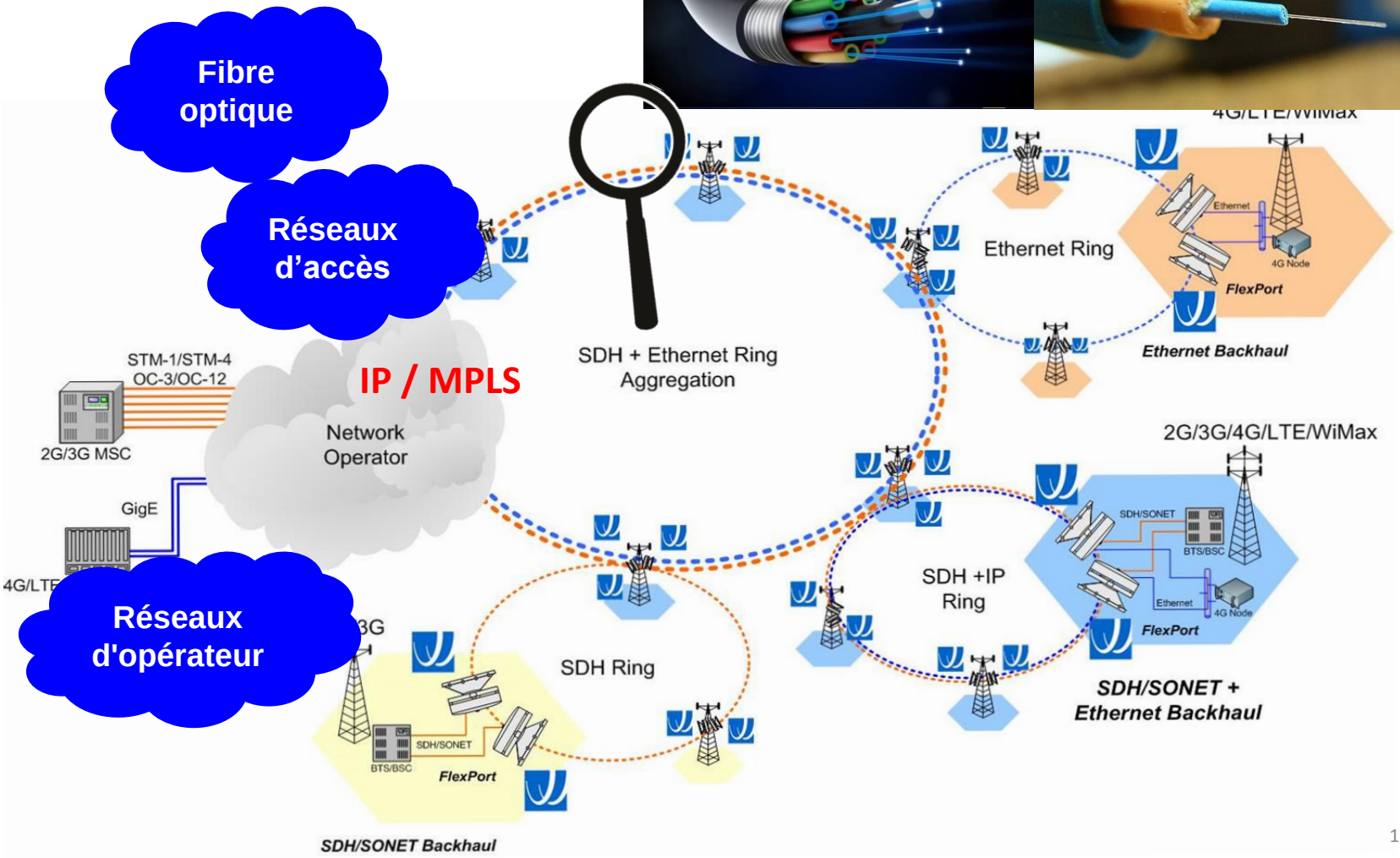
© http://akioz.com/isp-network-architecture/

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



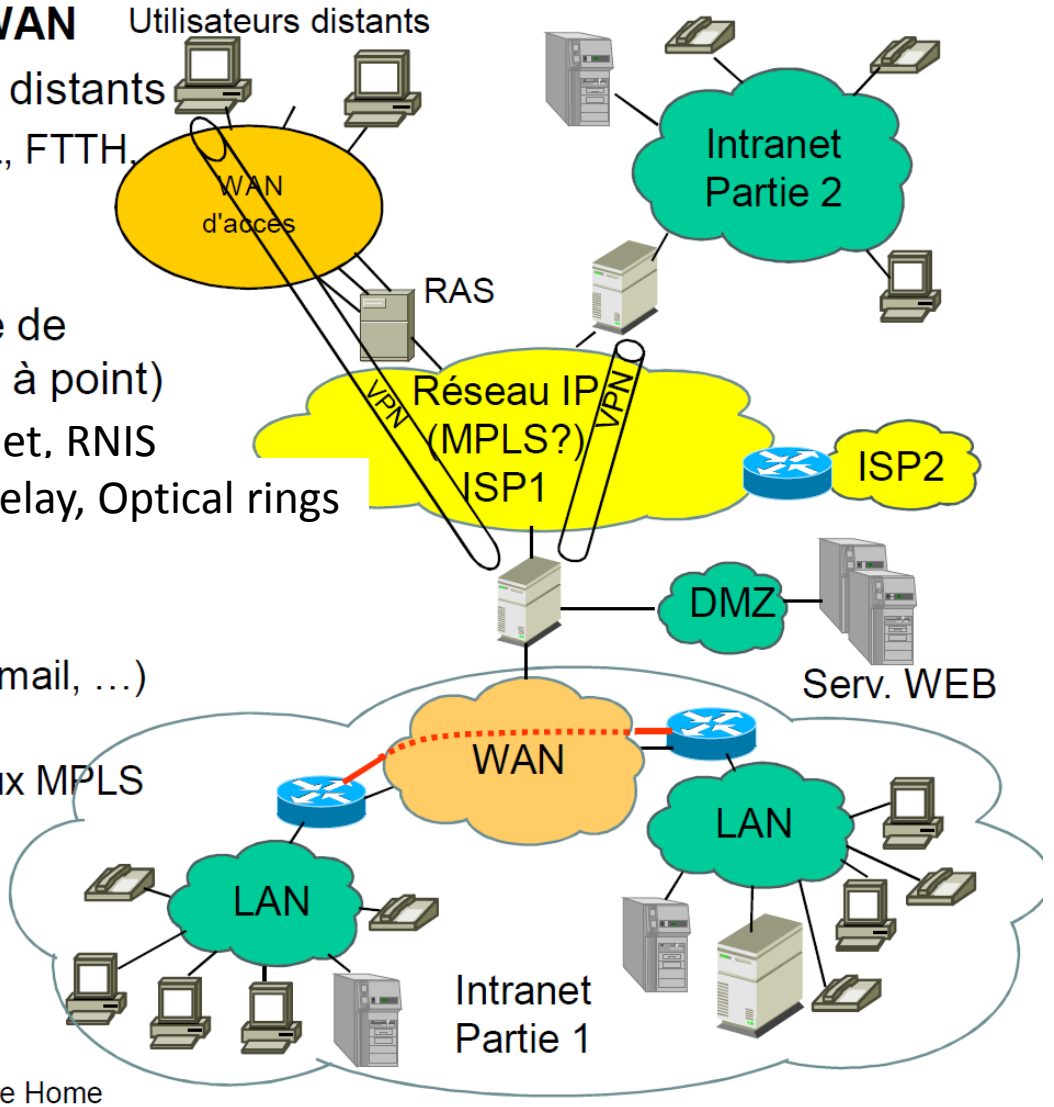
Exemple de réseau d'opérateur mobile





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

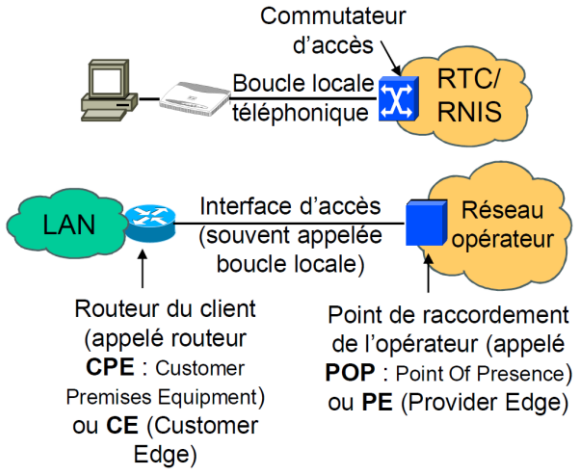


RAS : Remote Access Server
DMZ : De-Militarized Zone
ISP : Internet Service Provider
VPN : Virtual Private Network FTTH : Fiber To The Home

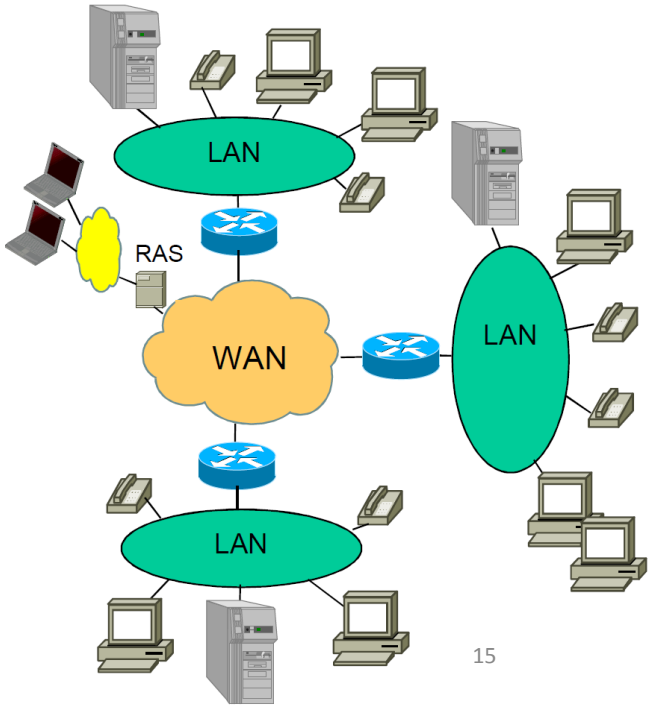
Réseaux étendus: présentation

- Caractéristiques de performance des WAN:

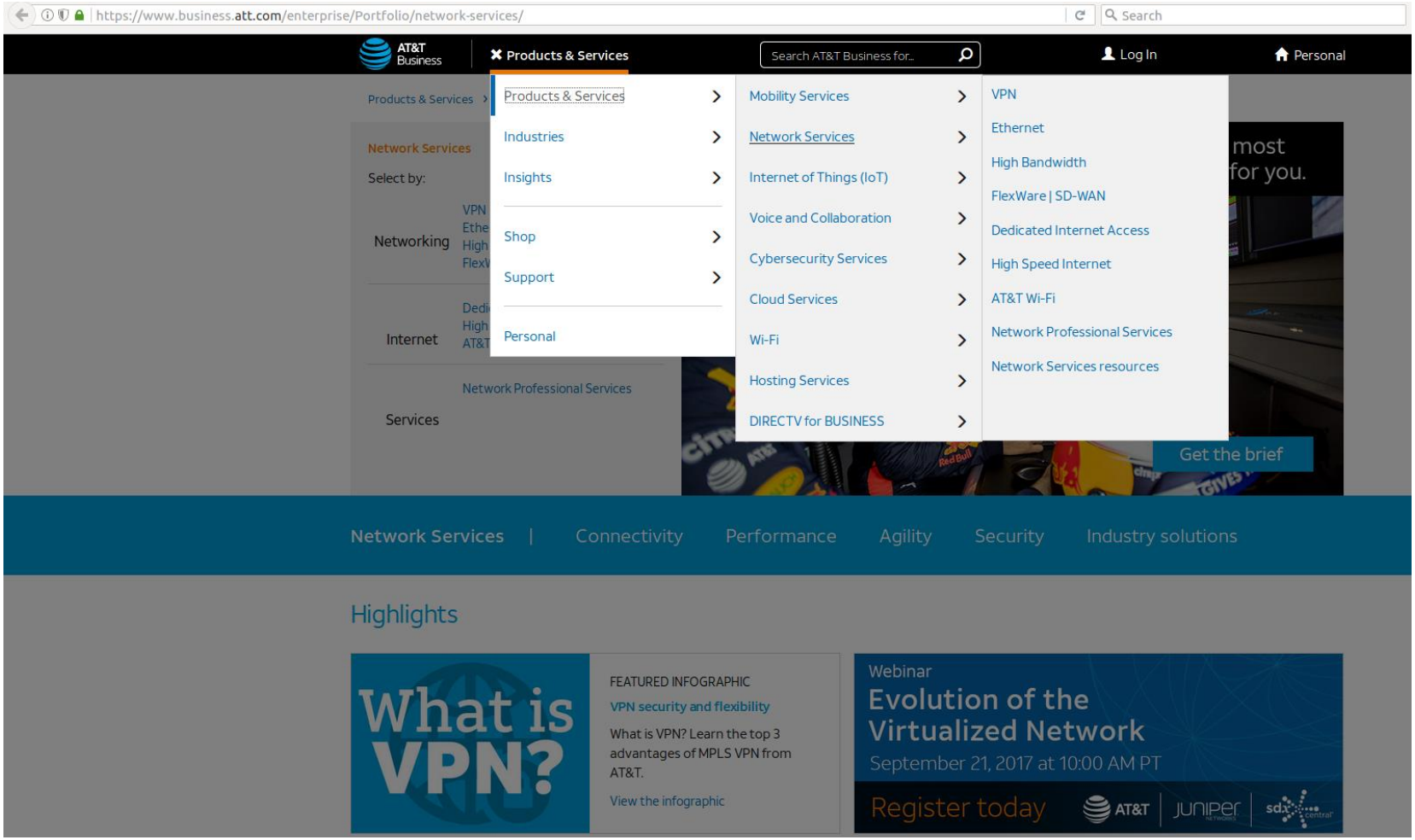
  - Prix beaucoup plus élevés que ceux des LAN
    - Même si les prix baissent rapidement
  - Délais de traversée du réseau plus élevés
    - Plusieurs dizaines ou centaines de ms



- 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

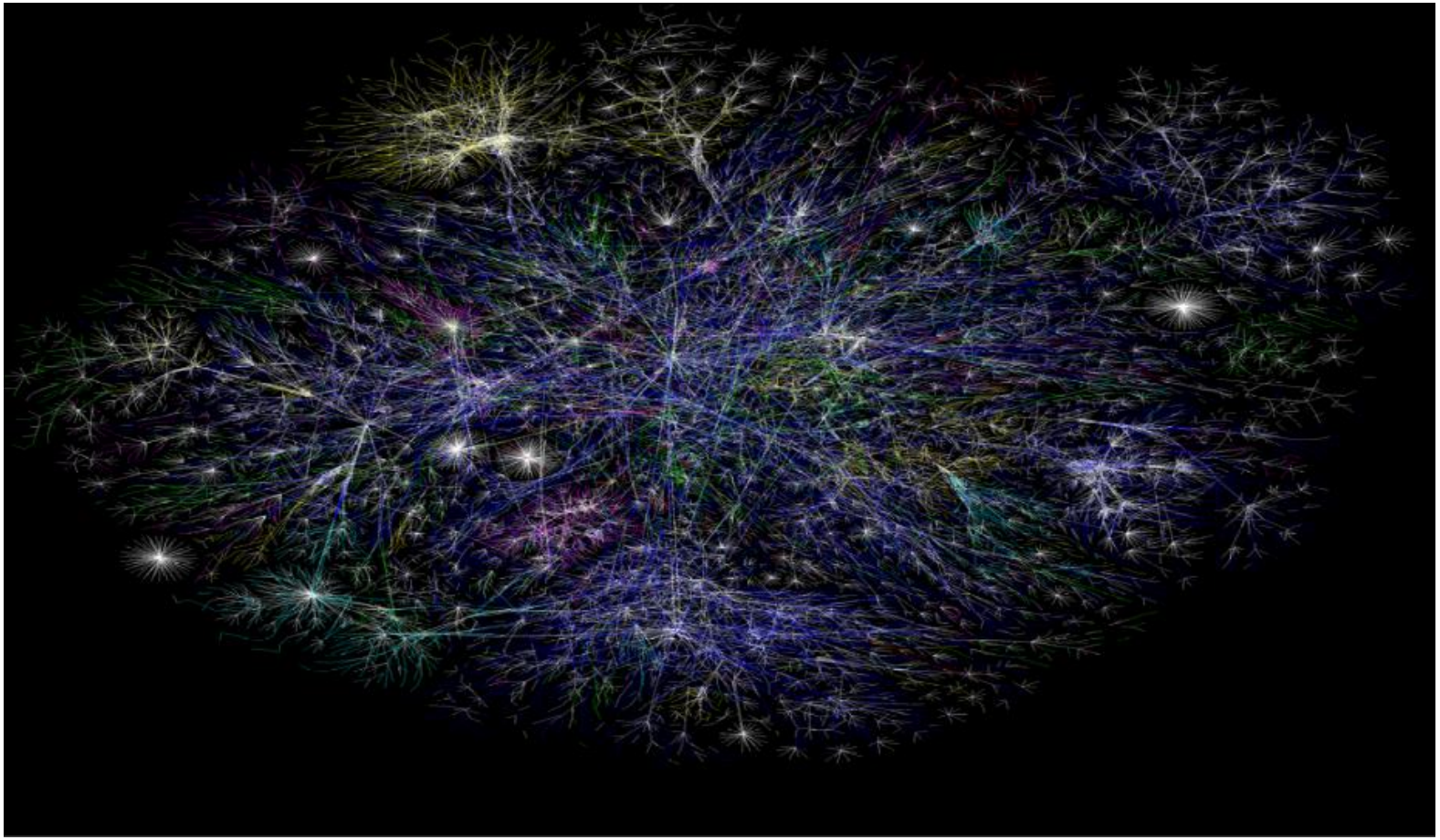


Plan

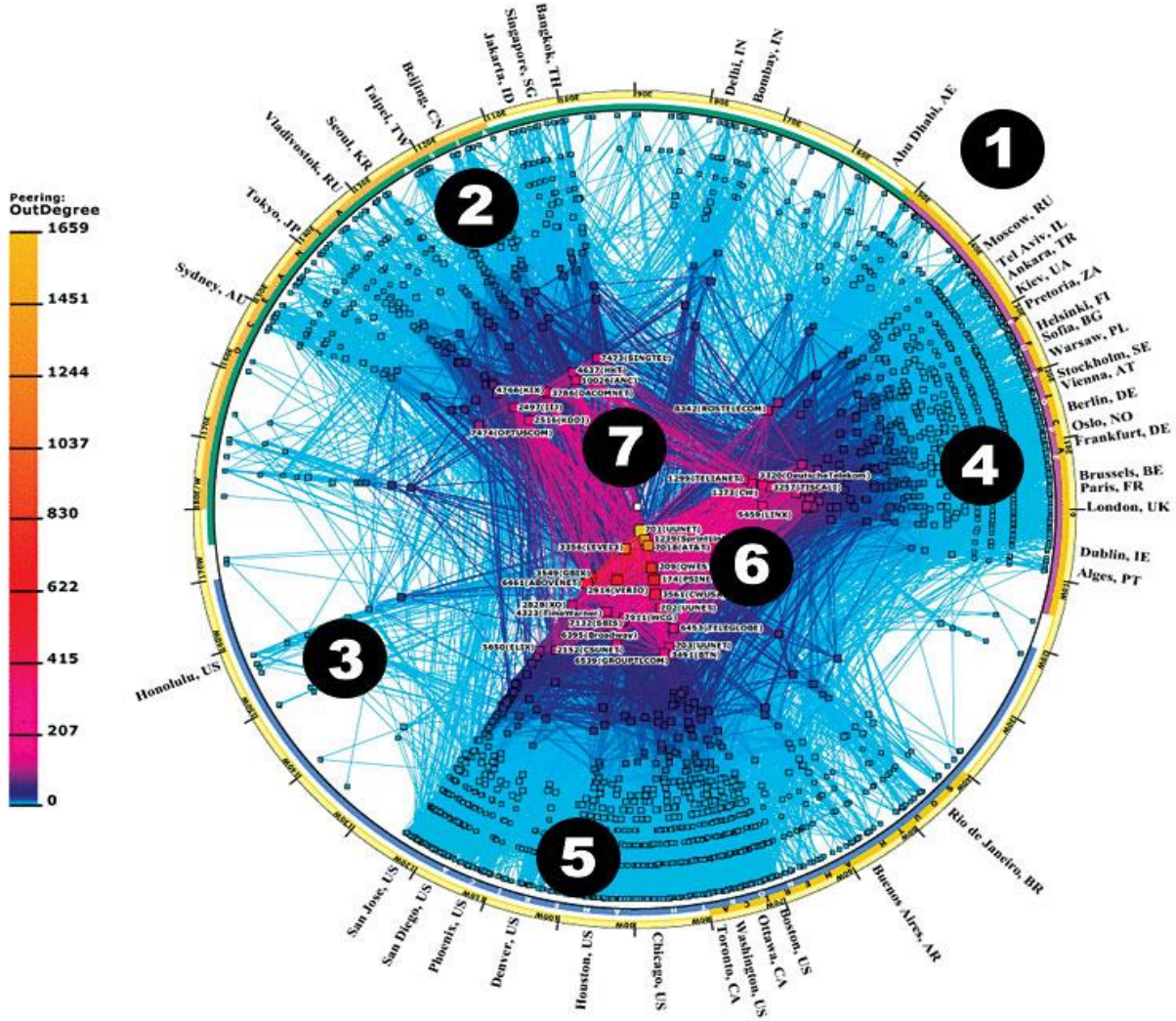
I. Organisation des opérateurs de l'Internet
   - I.1. Introduction aux réseaux étendus (WAN)
   - I.2. Hiérarchie et relations entre opérateurs (ISP)
   - I.3. Routage entre AS : le protocole BGP

II. TCP et Qualité de service
III. Commutation par circuits virtuels (VC) : le cas d'ATM
IV. Commutation par VC dans le mode IP : MPLS
V. Accès : Réseaux optiques
VI. Accès : Technologies xDSL

L’Internet à l’échelle globale



L’Internet à l’échelle globale



Vue rapide de la structure de l'Internet

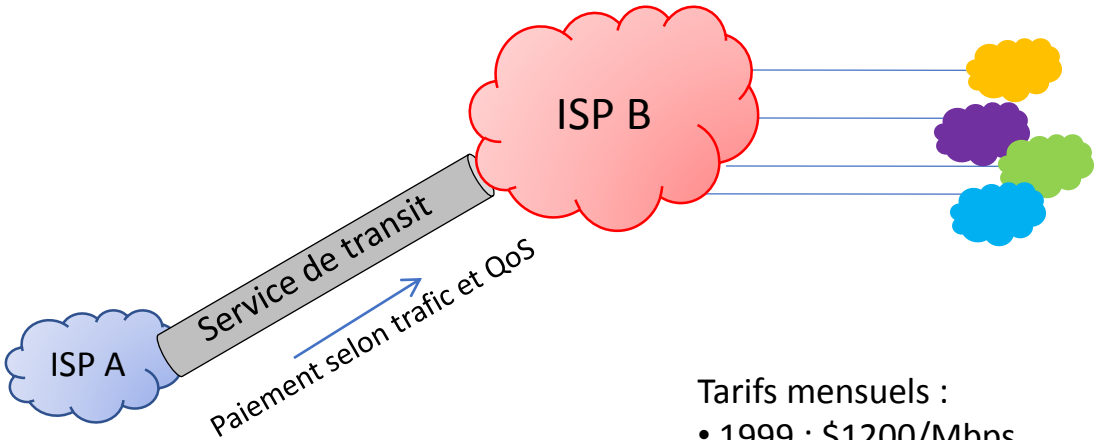
- On va voir rapidement comme l'Internet est interconnecté : comment le sont les ISPs
- Pour cela, on va expliquer les termes :
  - Internet transit
  - Internet peering
  - Internet Peering Ecosystem, Tier 1 ISPs, Tier 2 ISPs, leurs positions et motivations, et le rôle des Internet Exchange Points (IXPs).

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

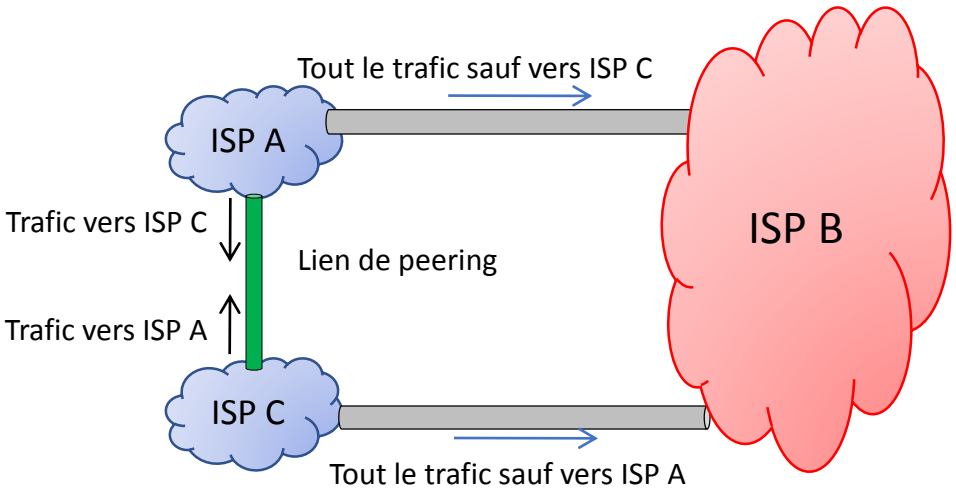
- Le transit semble facile et bon marché : pourquoi besoin d'un autre type d'intercx ?
  - $1/Mbps, si en moy 100Gbps -> $100K/mois : pas négligeable !
  - > d'où le peering

- 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

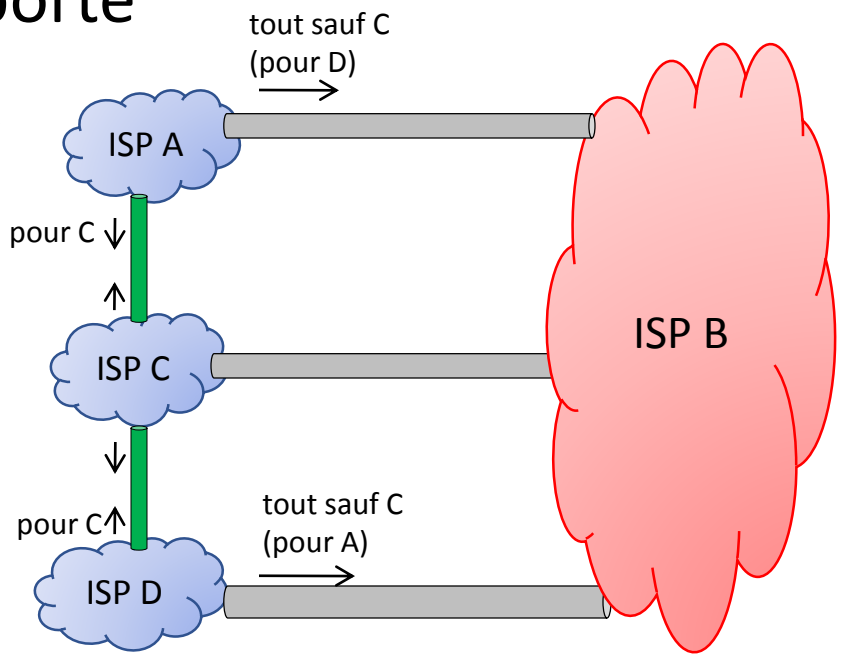
  selon le coût de l'installation de peering
  - Bénéfices en perf : shortest

  path plus court, délai plus faible

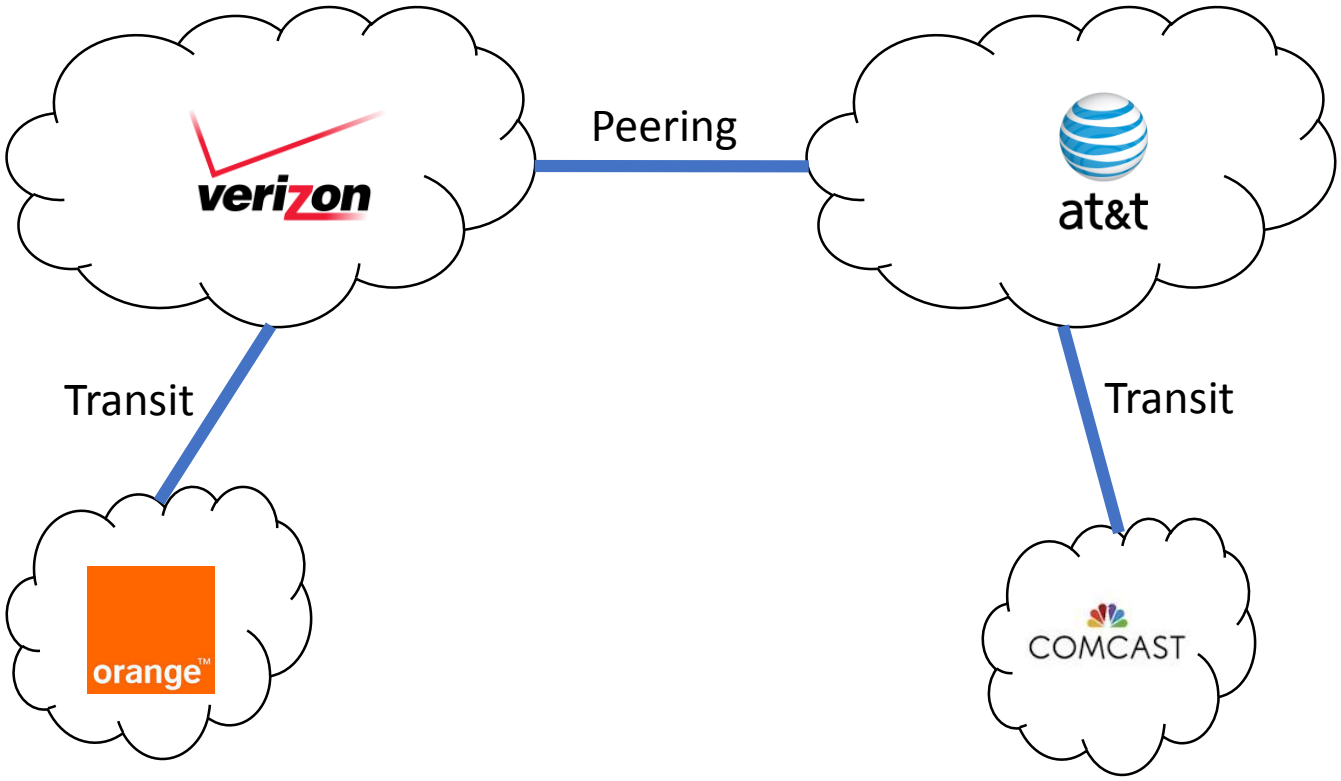


2 points importants sur le Peering

- 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

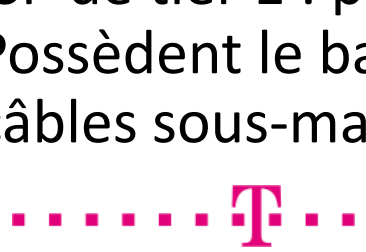


Hiérarchie des ISP

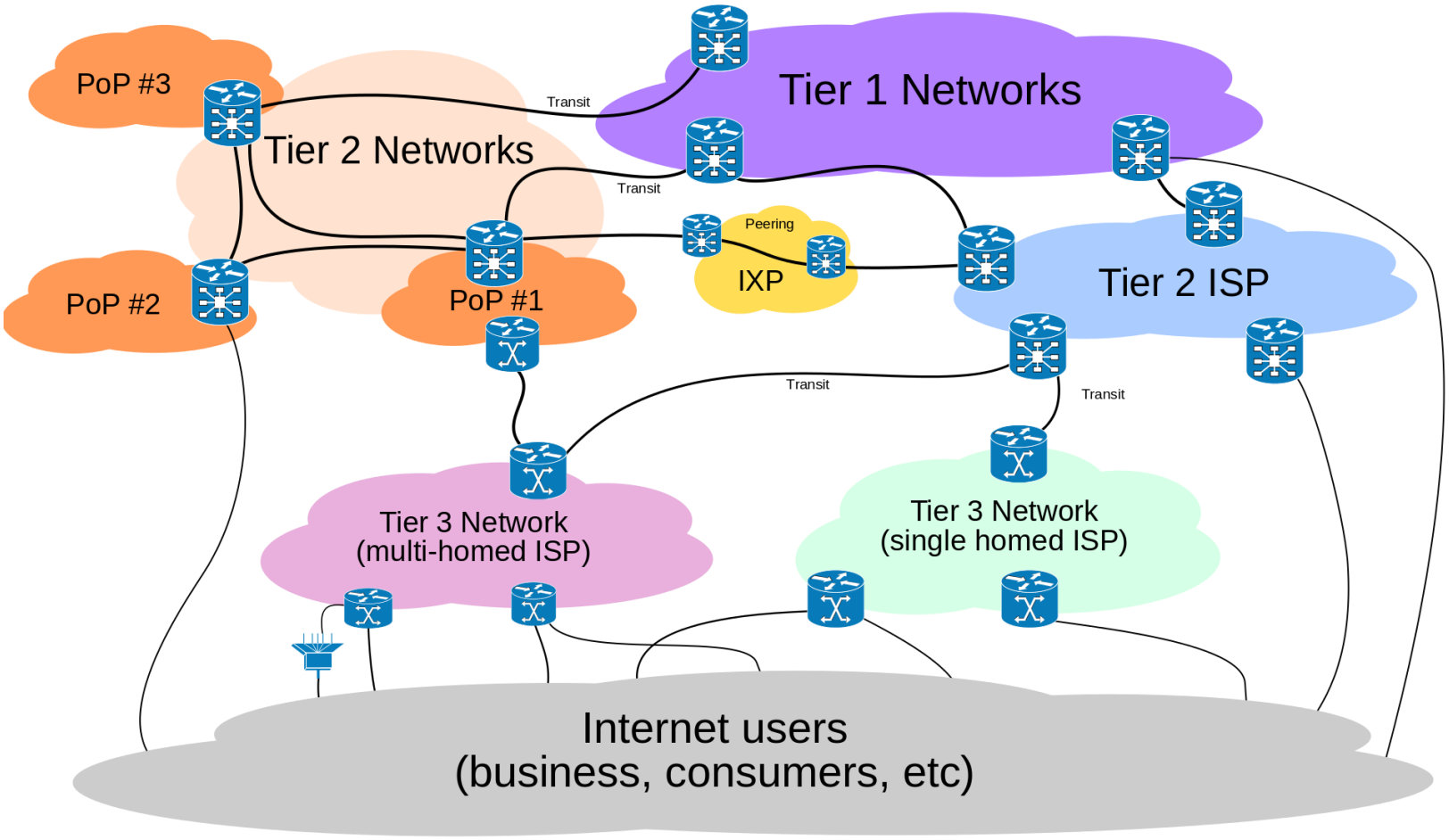
- ISP de tiers 3 : plus bas niveau. Ne fournissent pas de transit.

- 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 de 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 toutes les infos sur https://www.renater.fr/sfinx-ixp?lang=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

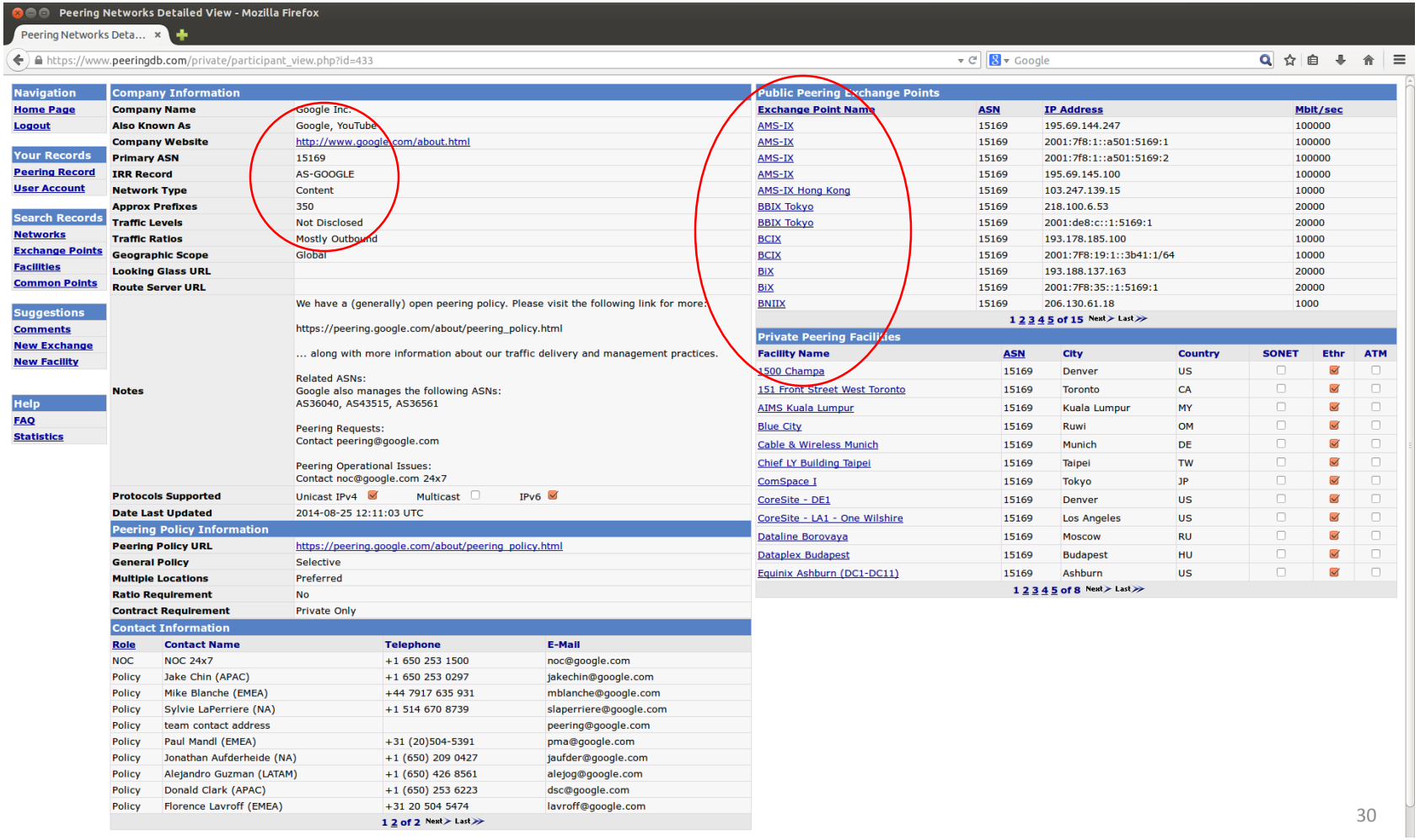
Internet eXchange Points (IXP)

- 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

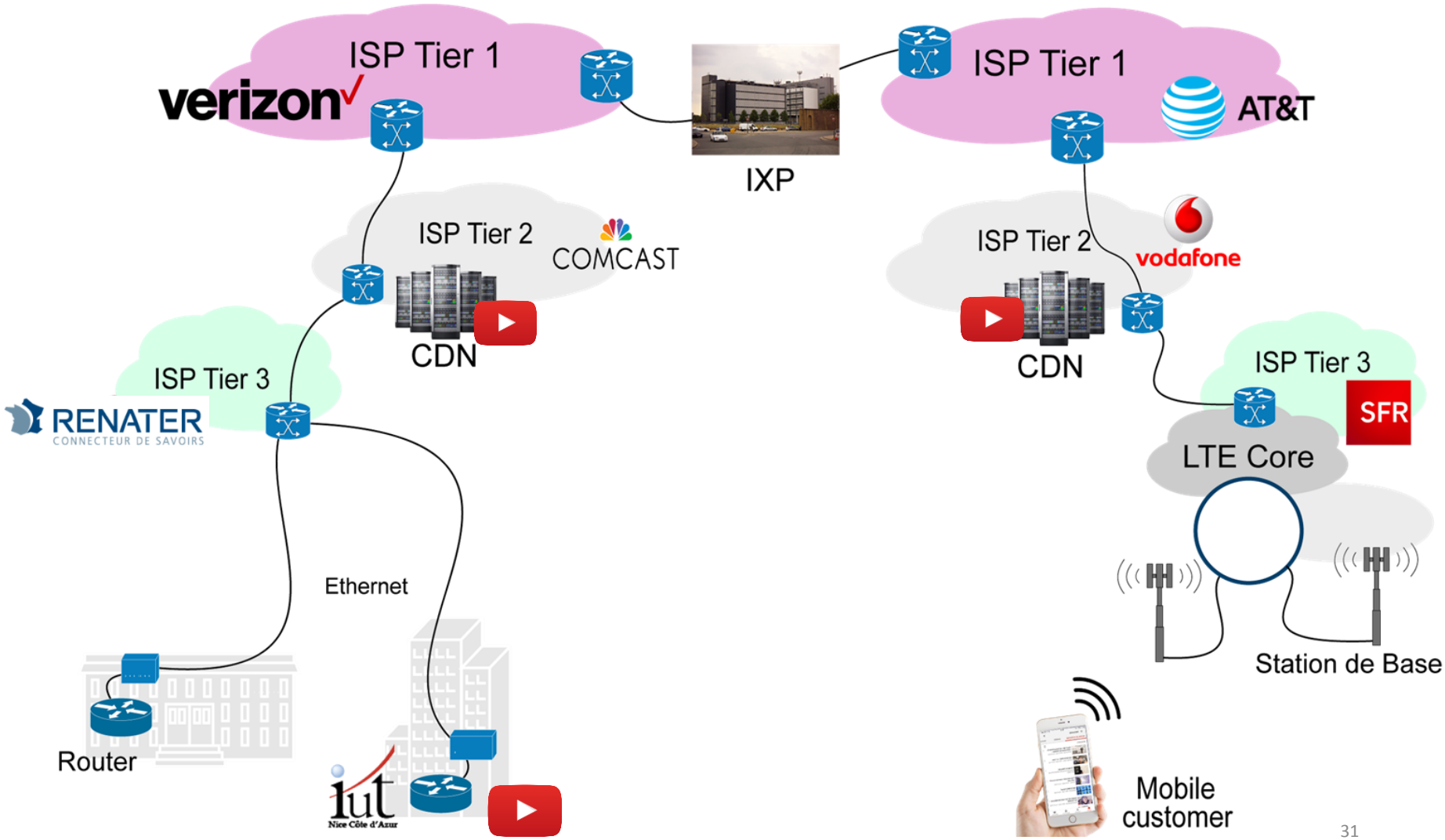
- Donc 2 réseaux qui veulent établir un peering entre eux au sein d'un IXP :
  - coût unique pour établir un circuit depuis leur point de présence
  - coût mensuel pour utiliser un des port Ethernet de l'IXP
  - possiblement une cotisation annuelle pour être membre de l'association possédant l'IXP (non-profit 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.
- 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
- France IX: https://www.peeringdb.com/ix/359
- Lyon IX: https://www.peeringdb.com/ix/69

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 (VC) : le cas d'ATM
IV. Commutation par VC dans le mode IP : MPLS
V. Accès : Réseaux optiques
VI. Accès : Technologies xDSL

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

Plan

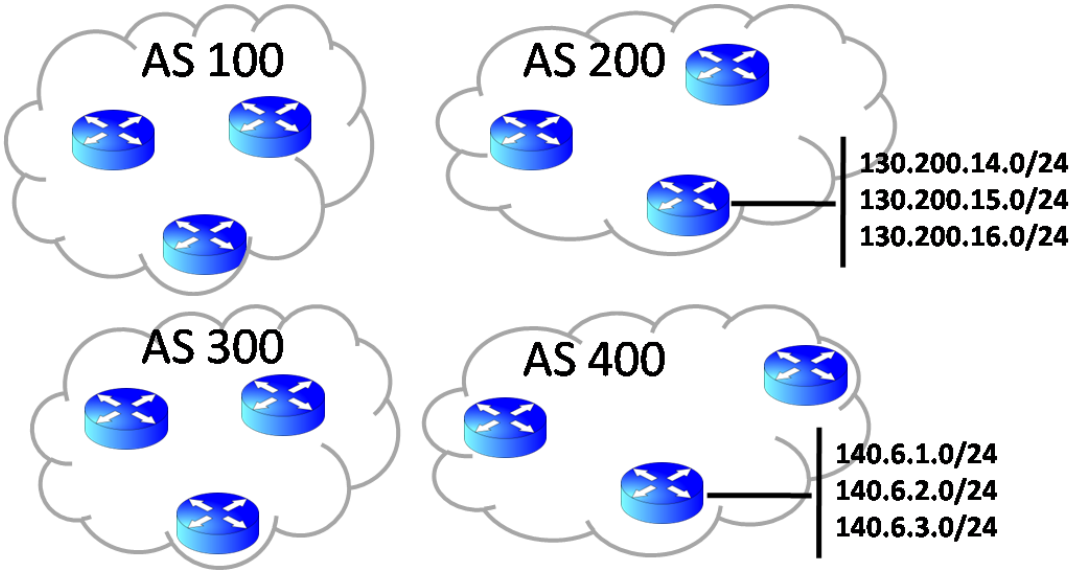
- Introduction de BGP
  - Tables BGP réelles de l'Internet
- Options de connexion à Internet : quand utiliser BGP ?
- Quantités d'updates : comment utiliser BGP ?
- BGP en action
  - Messages BGP
  - Etats du voisinage
  - Configuration
    - Configuration de voisins
    - Injection de routes dans BGP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

Introduction à Border Gateway Protocol (BGP)

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

- ***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



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

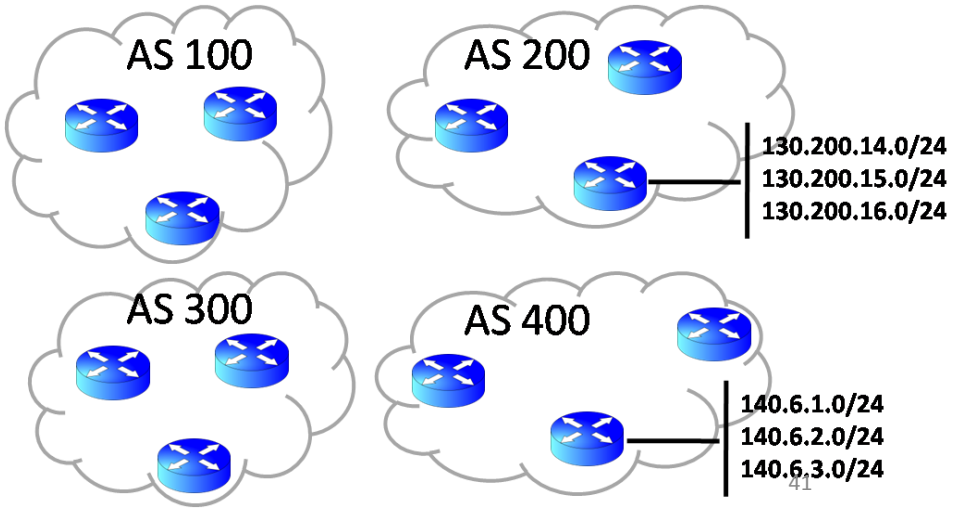
-> implique une confiance mutuelle implicite.

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

Introduction à BGP

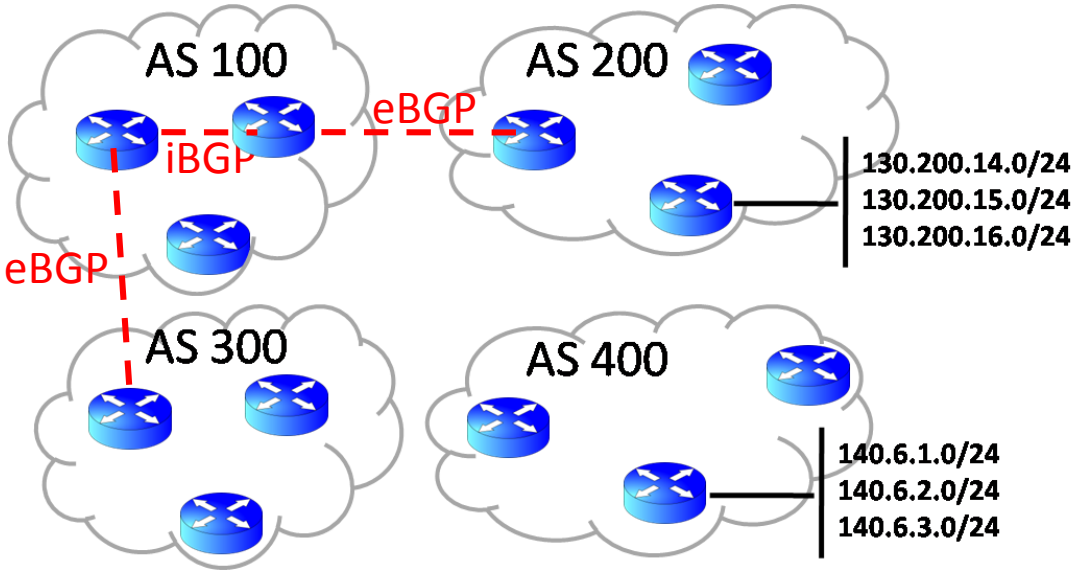
- ISPs utilisent BGPv4 pour échanger MaJ de routage sur l'Internet
- *Interior Gateway Protocol* (IGP) : RIP, OSPF, IS-IS, EIGRP
- Besoin de configurer les voisins explicitement (statiques) – pas de découverte automatique
- BGP : pas de concept de multipaths et load balancing (contrairement aux IGPs)
- 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

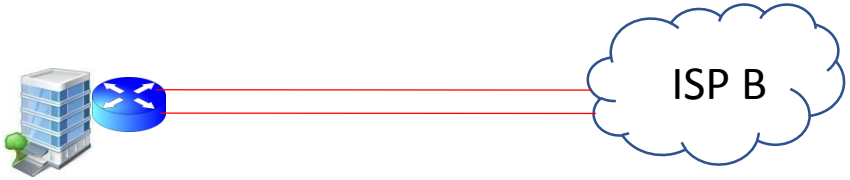
- Sur Internet : @IP publiques
- Attribuées à des entités, dans différents AS
- Visibles à: http://whois.domaintools.com
- Ces caractéristiques de votre connexion visible à : bgp.he.net

Quand utiliser BGP ?

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

- 1 ISP et 1 cx : une default route suffit

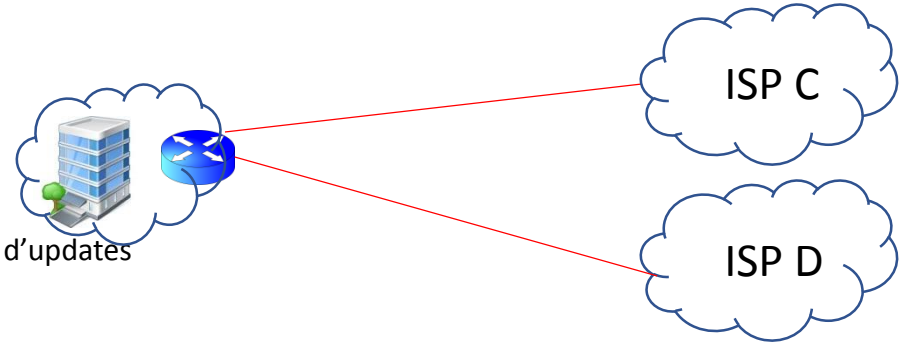
- 1 ISP et plus d'1 cx : cx redondantes. Load sharing ou backup cx



- 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.

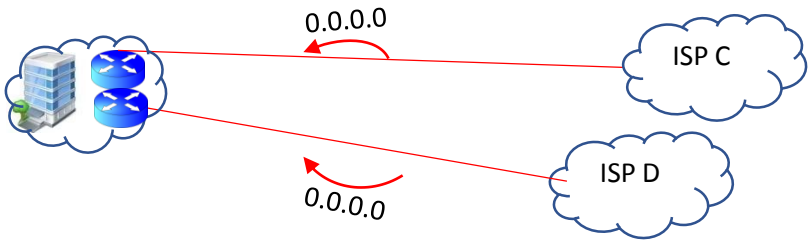
BGP intéressant

- Attention à la quantité d'updates

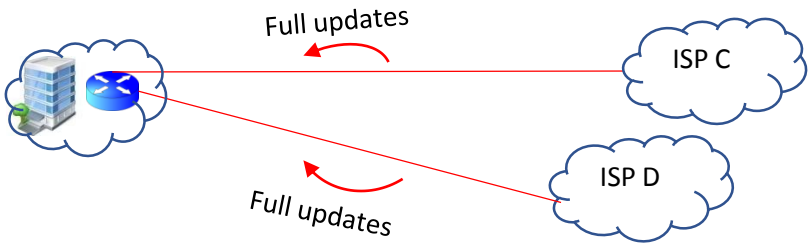


Quantité d'updates avec BGP

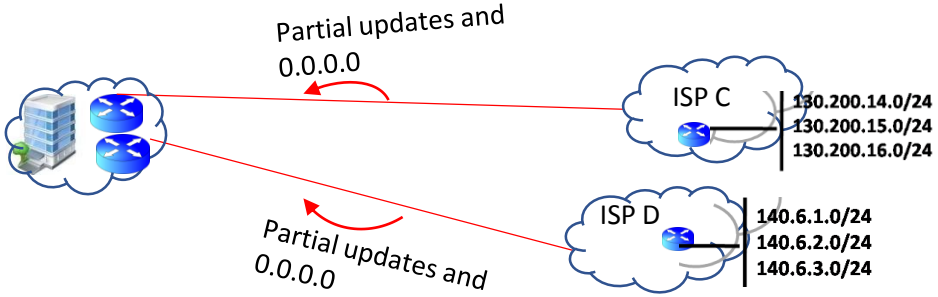
- Default route : on peut utiliser BGP et ne recevoir que la default route (mais on publie nos réseaux) – pas de mémoire supplémentaire requise aux CE



- 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

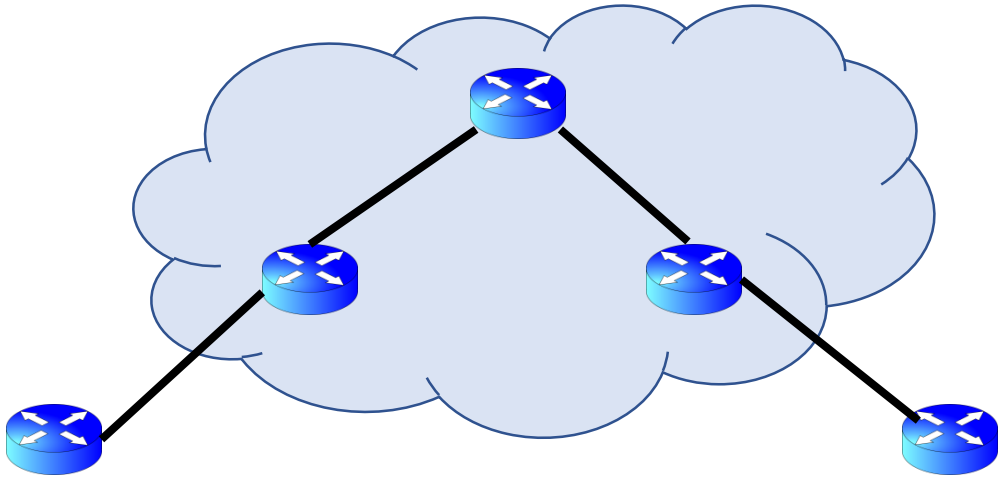
Etats des voisins

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

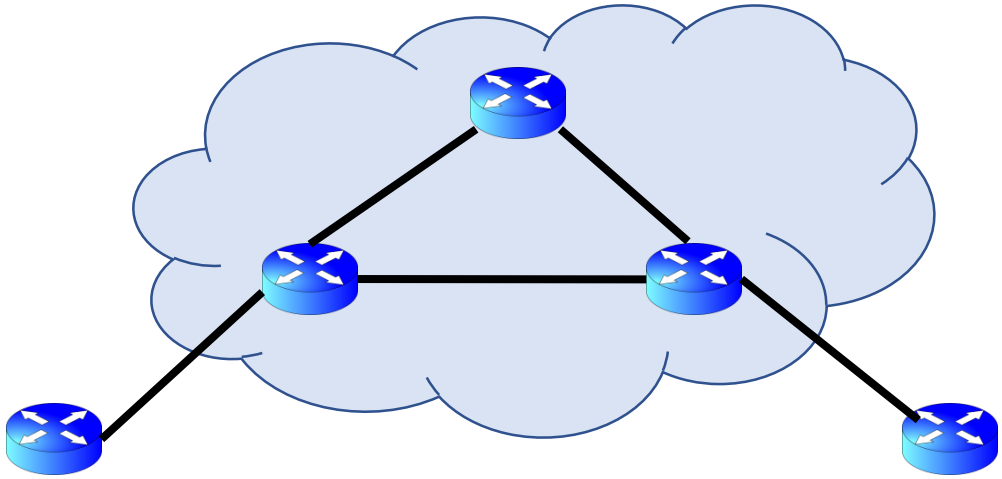
- Les routes apprises via IBGP doivent être également apprises par un IGP (OSPF,…) avant d'être annoncées à d'autres peers d'AS différents.



- 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 :

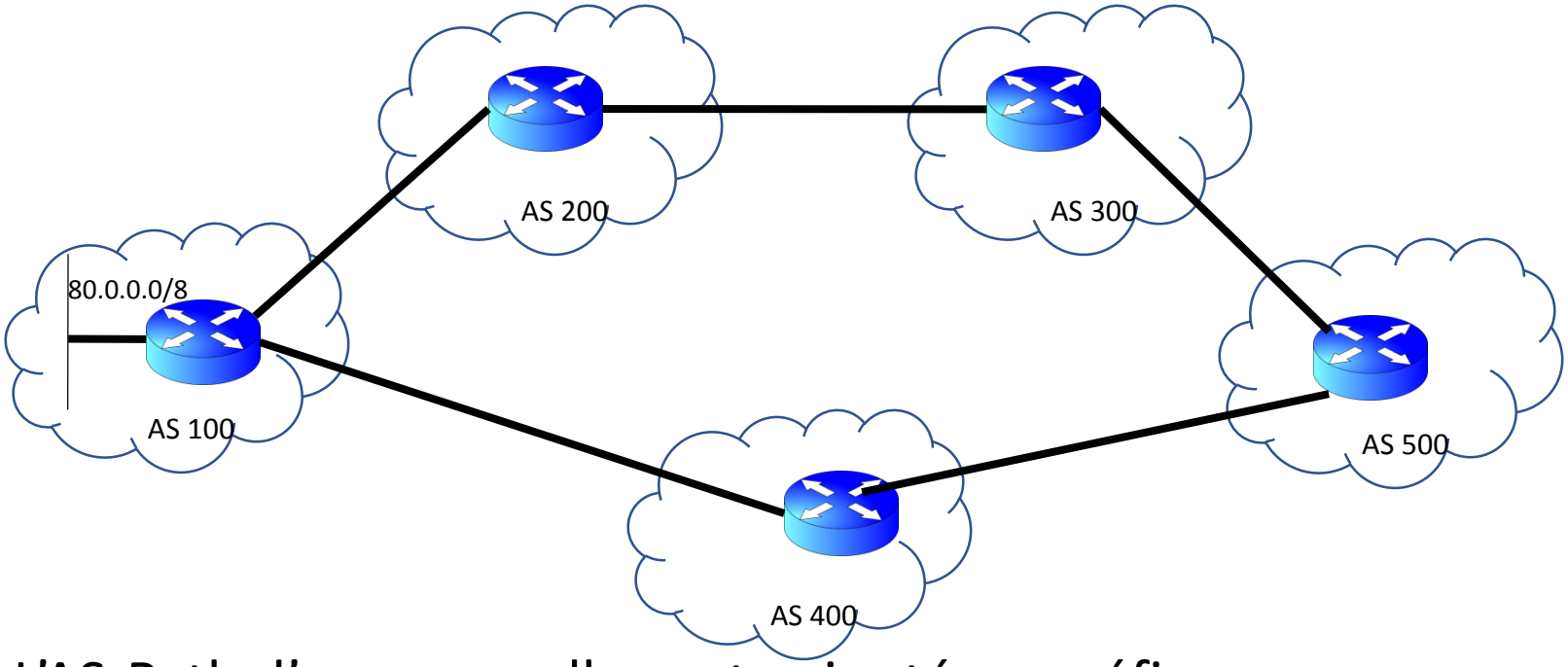
```
R11#sh ip bgp
BGP table version is 842, local router ID is 10.0.114.11
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   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

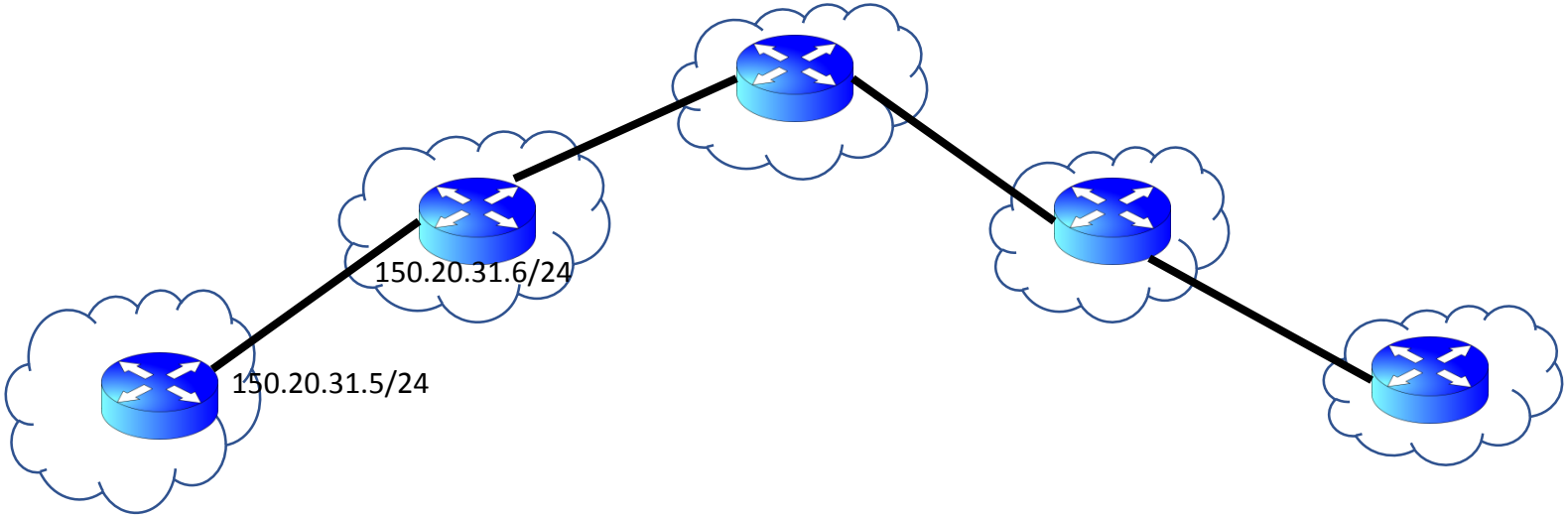
- C'est la suite des AS traversés
- Obligatoire dans chaque update: chaque routeur annonçant un réseau inclut la liste des AS à traverser dans l'update.



- 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

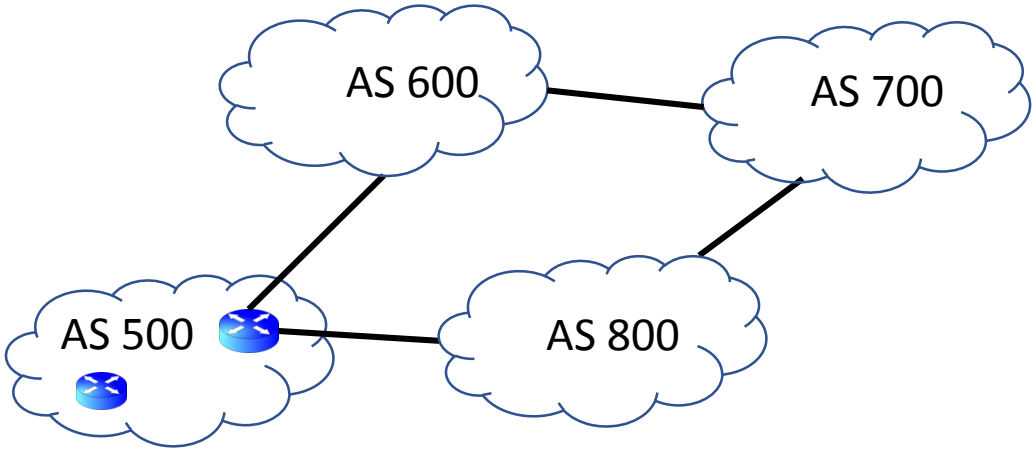


Attribut Origin

- 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)

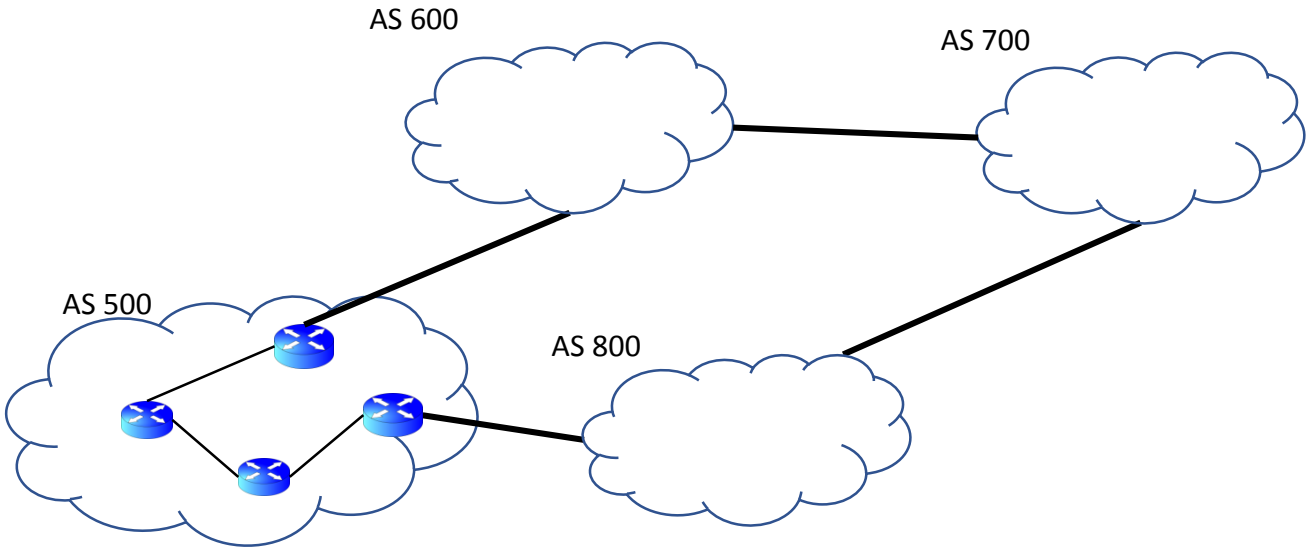
Attribut Weight

- 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



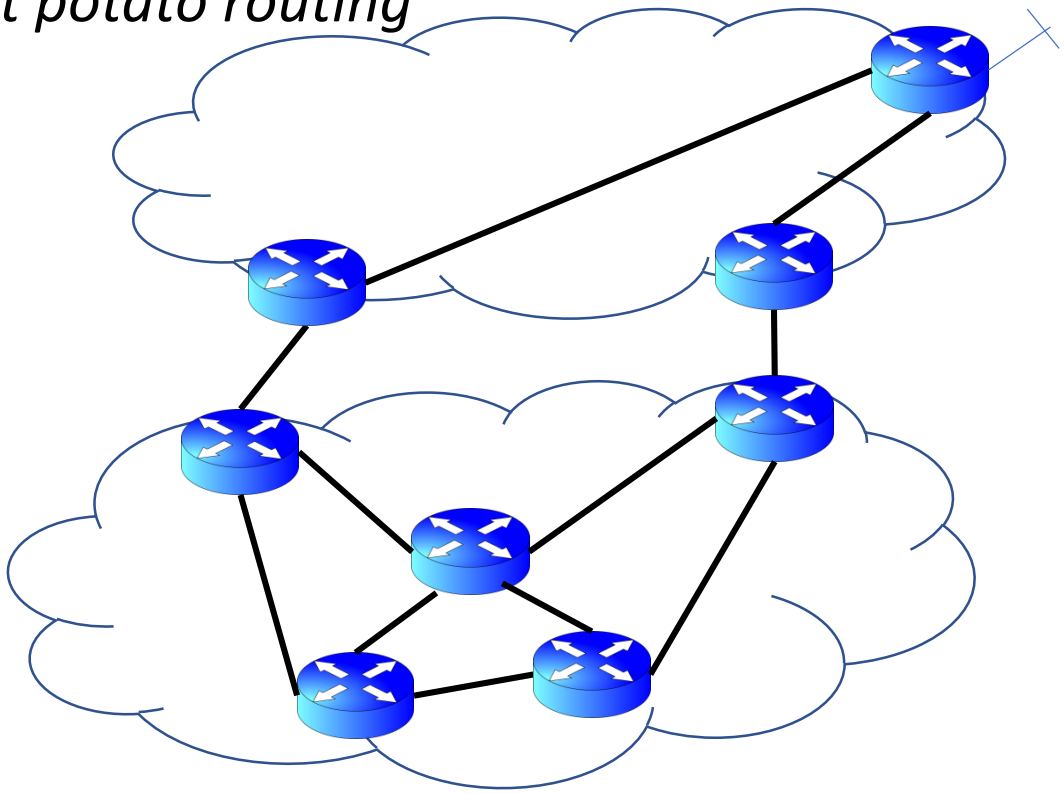
Attribut Local Preference

- 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)

- Règle d'or de BGP : un AS ne peut jamais dire à un autre AS comment router son trafic
- SUGGERE un point d'entrée dans votre AS
- Principe de routage au sein d'un AS: faire sortir le trafic le plus vite = *Hot potato routing*



Attribut Community

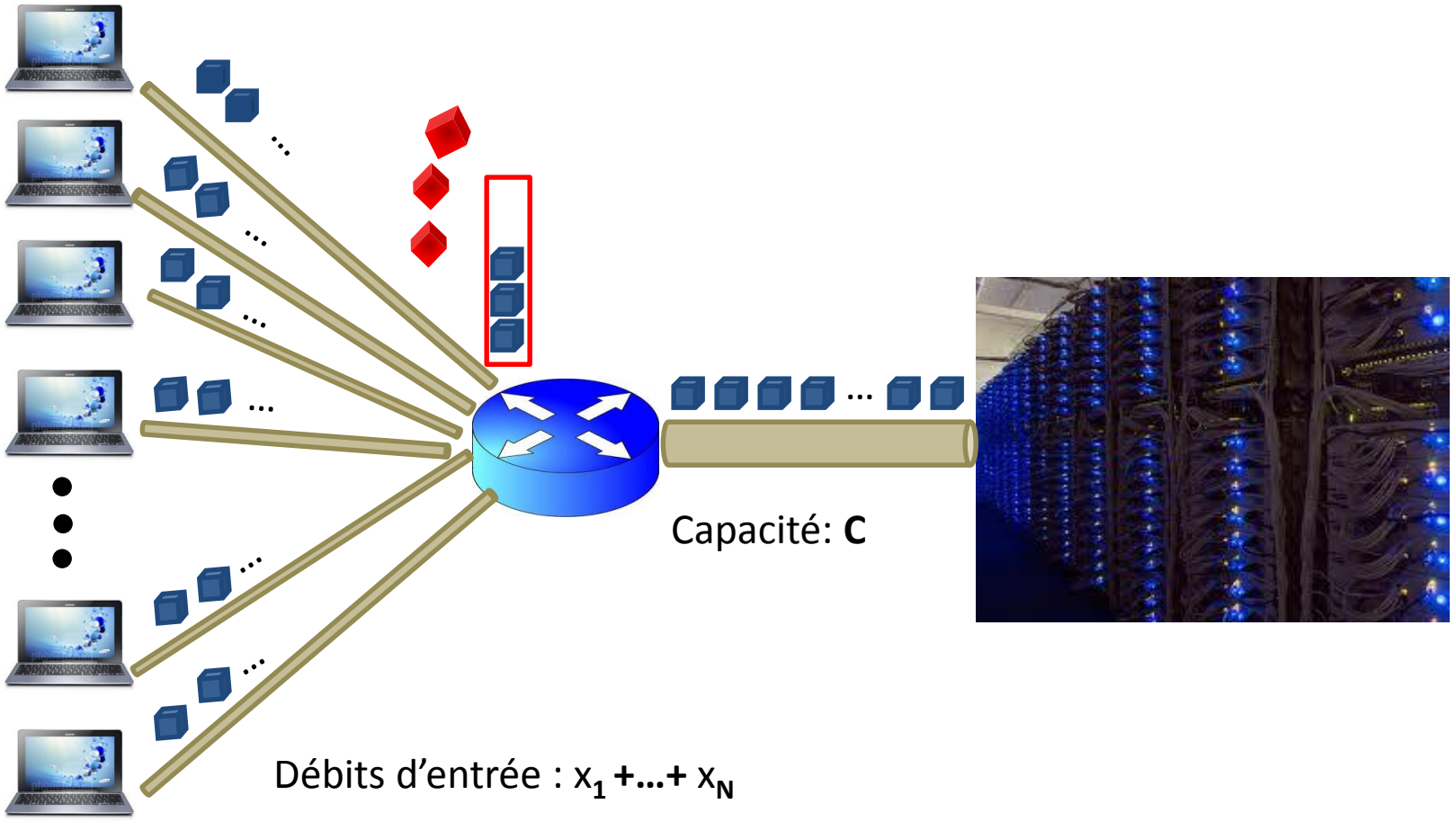
- Pour filtrer les routes à re-distribuer dans/en dehors de notre AS

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

Le buffer d'un routeur

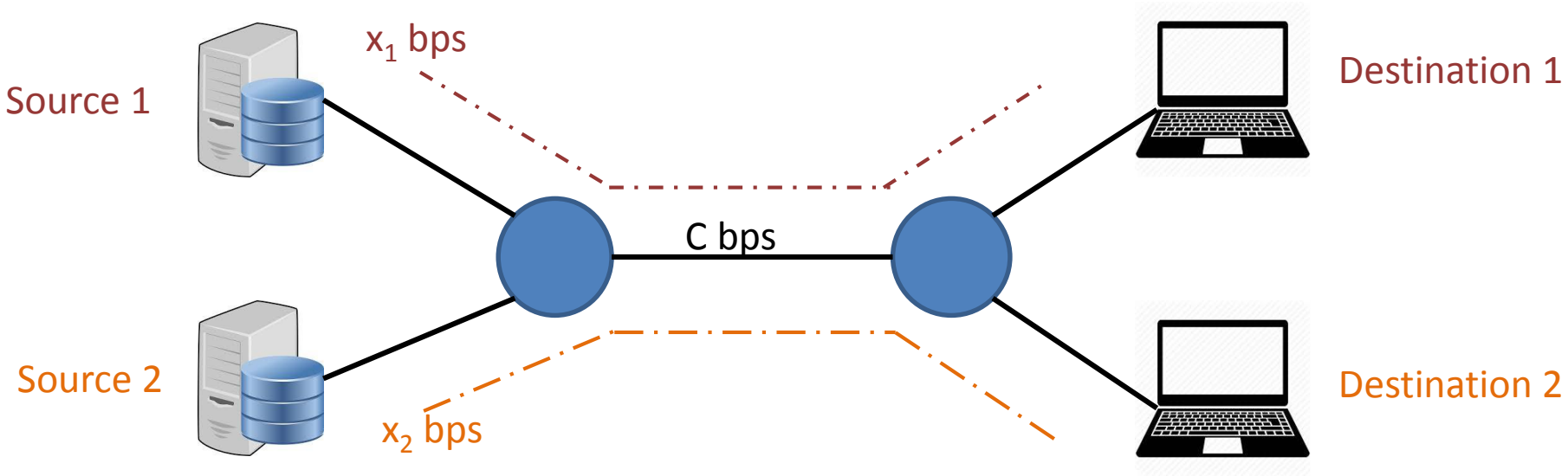


N Clients

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

Serveurs

Comment partager la capacité ? Le contrôle de débit



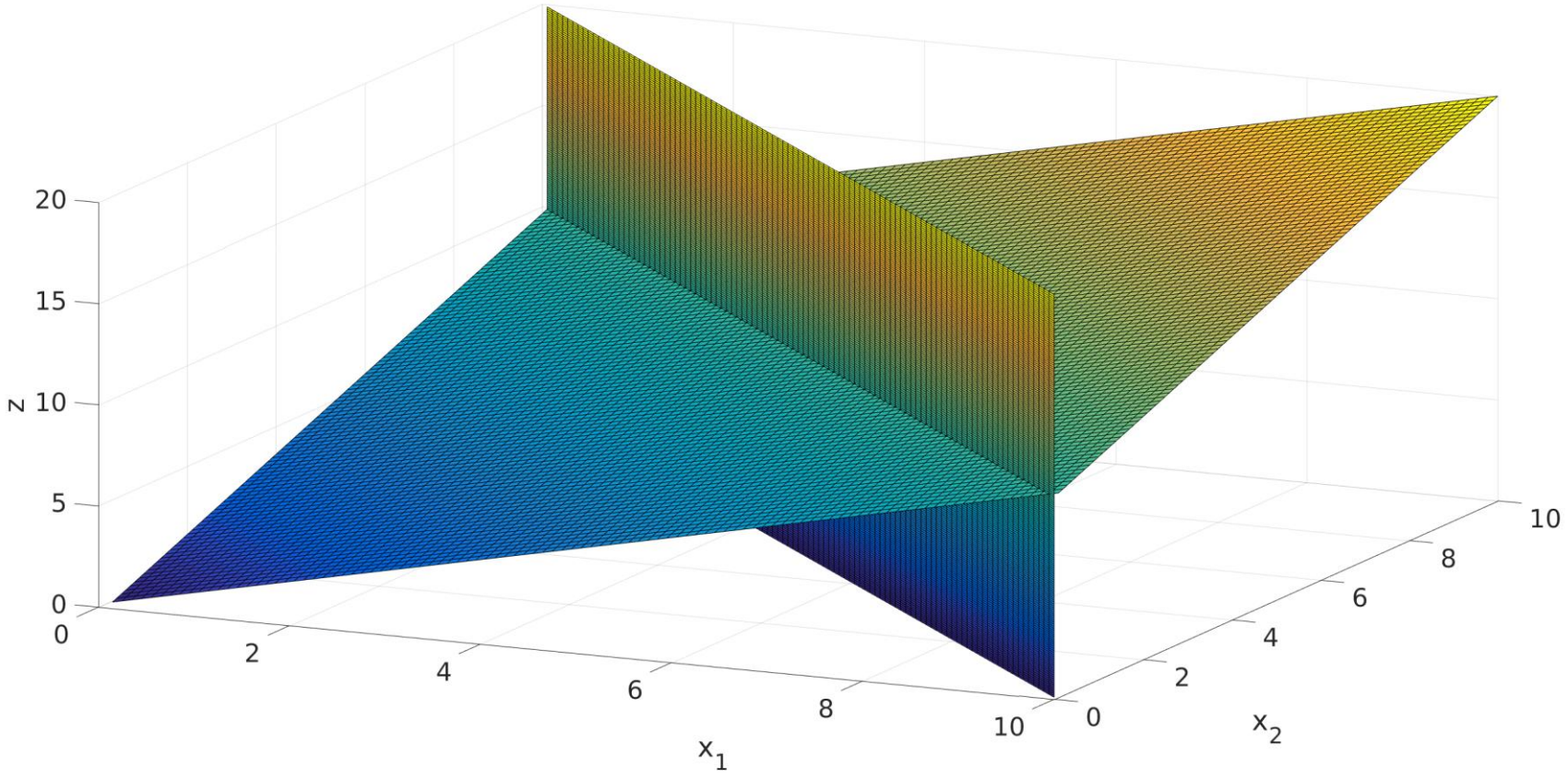
- Trouver x_1 et x_2 solutions au problème

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

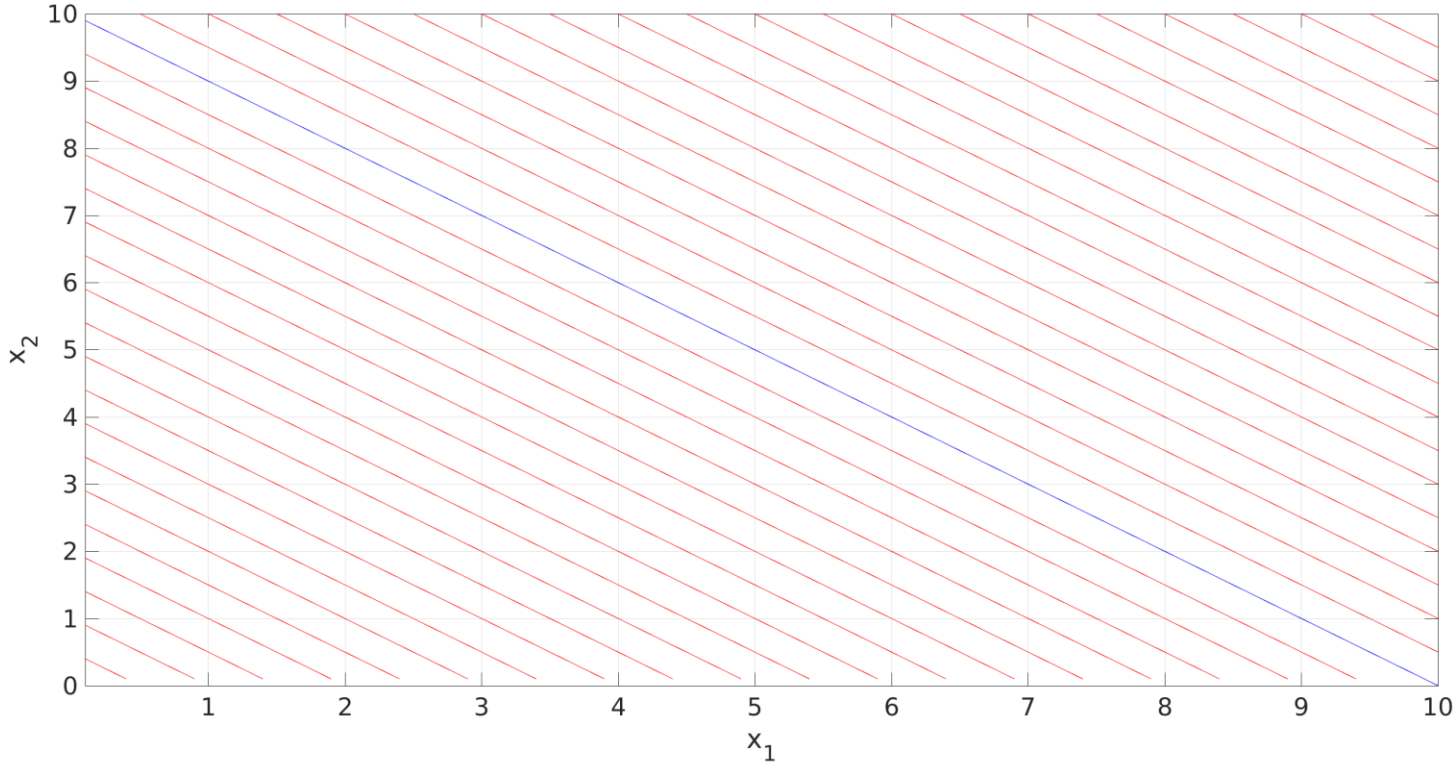
Tel que

$$x_1+x_2 = C$$

Comment partager la capacité ? Le contrôle de débit

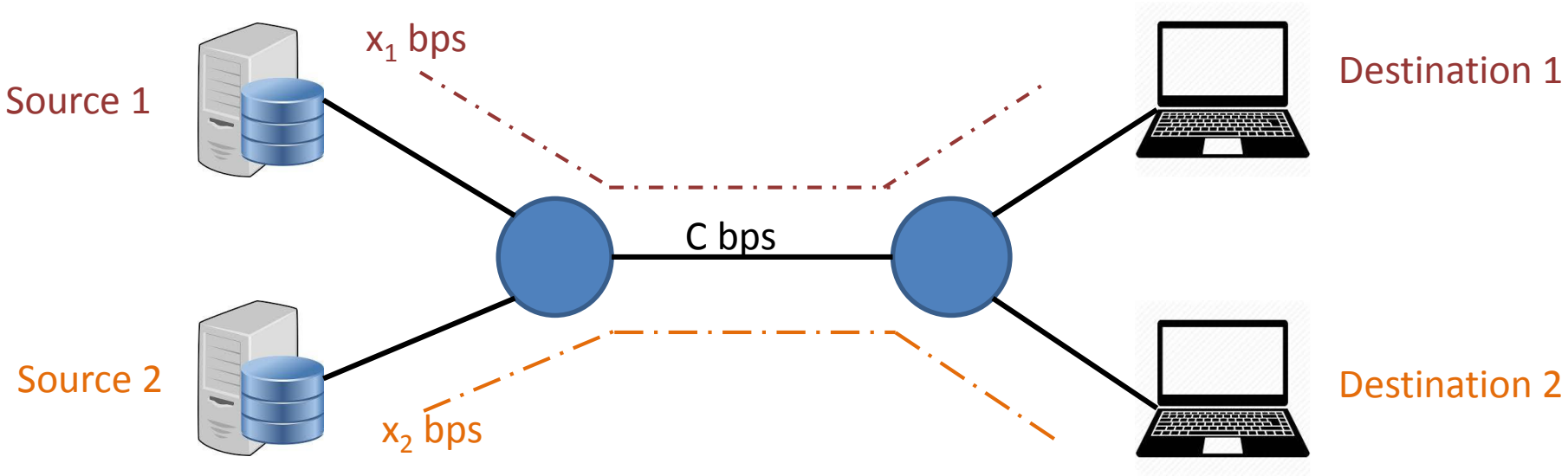


Comment partager la capacité ? Le contrôle de débit



-> pas de contrainte d'équité entre les sessions

Comment partager la capacité ? Le contrôle de débit



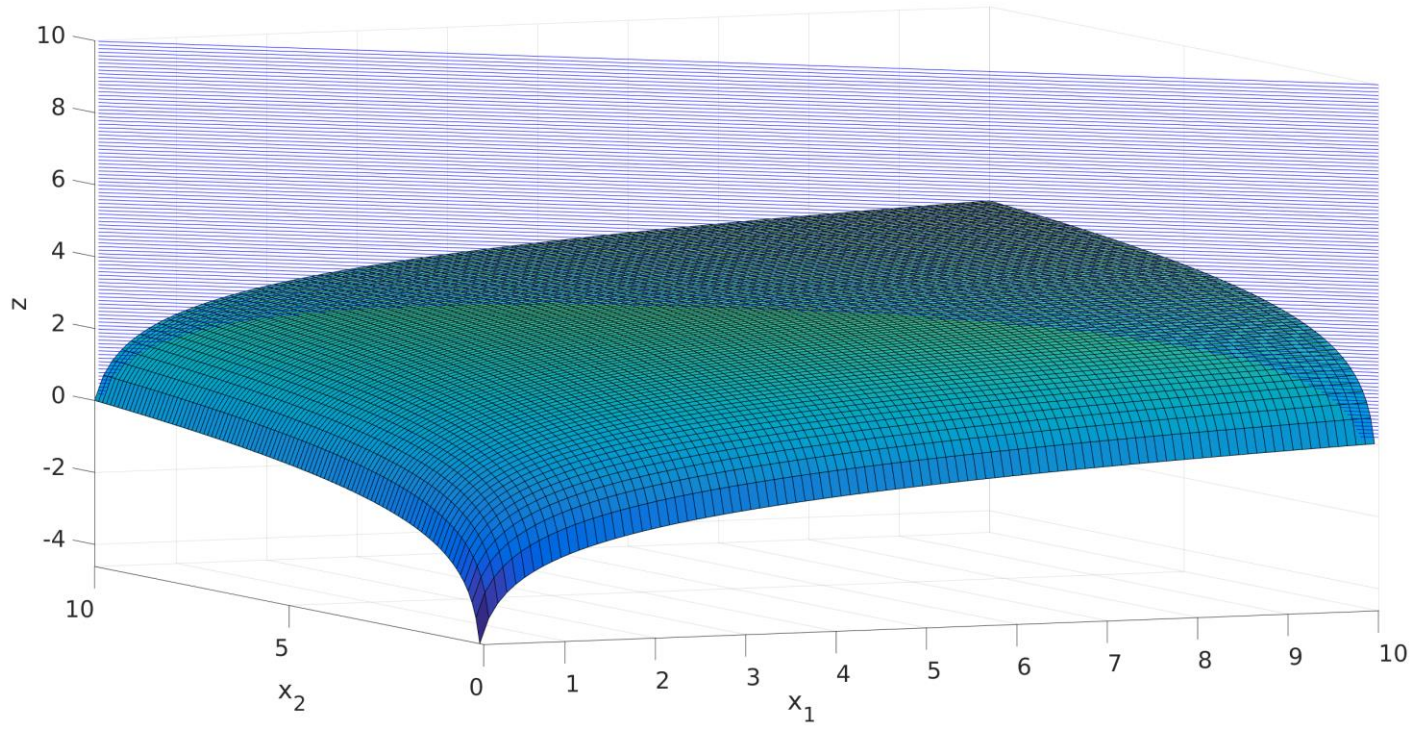
- Equité:

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

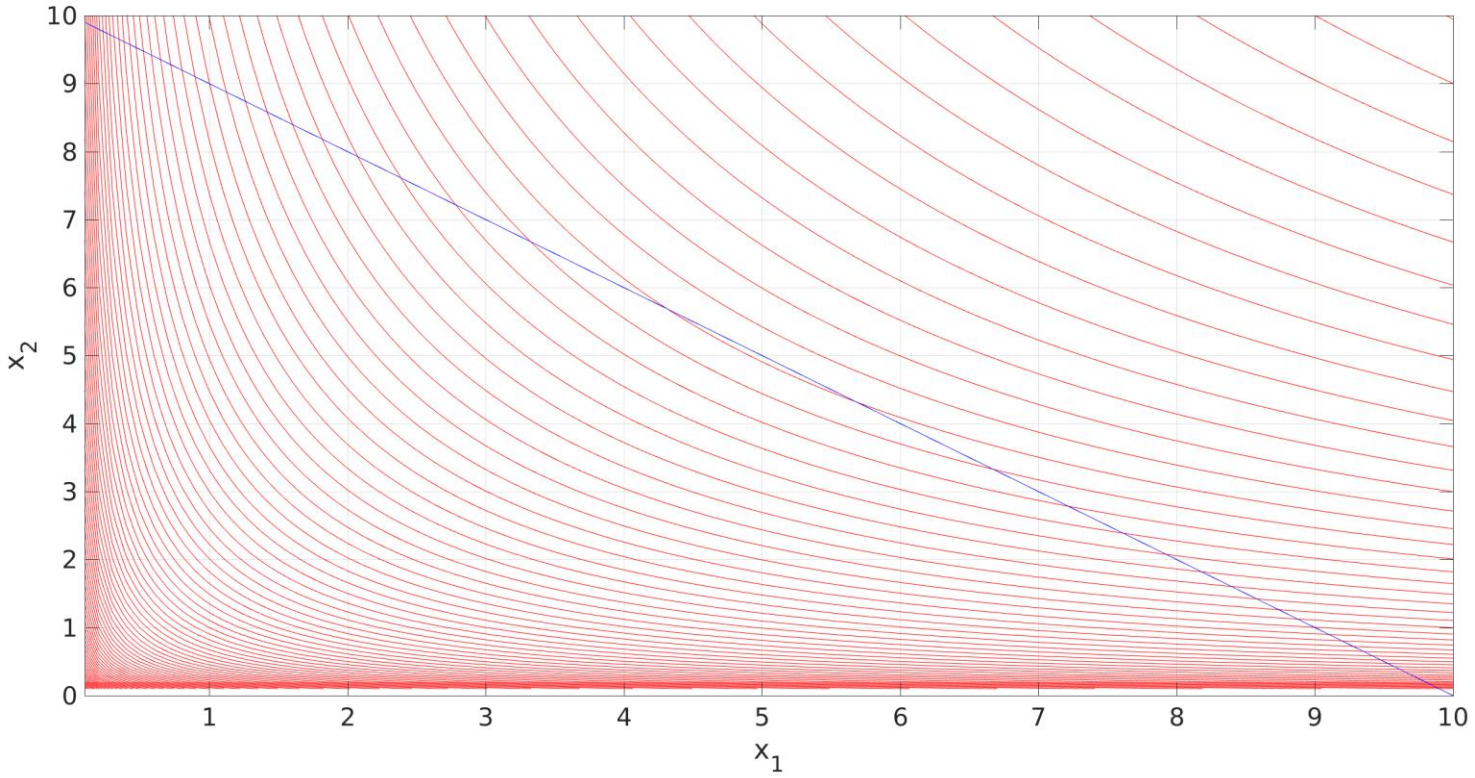
Tel que

$$x_1+x_2 = C$$

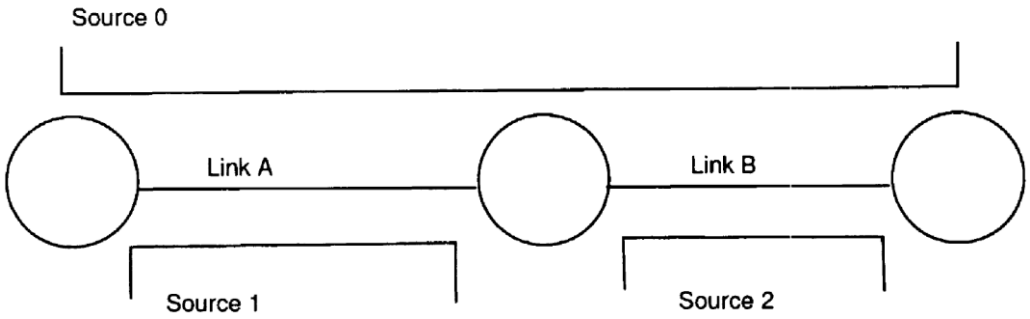
Comment partager la capacité ? Le contrôle de débit



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 ?

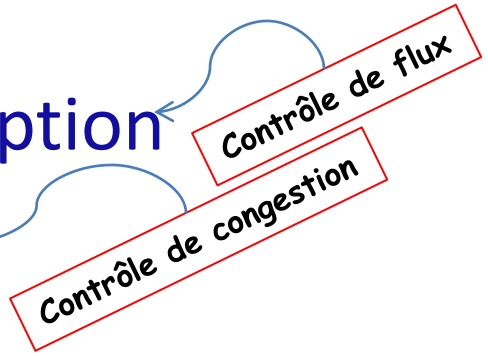
- Nécessité d'un contrôle de l'utilisation des ressources du réseau
- Pour éviter les problèmes de stabilité
- Le contrôle doit être distribué

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



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

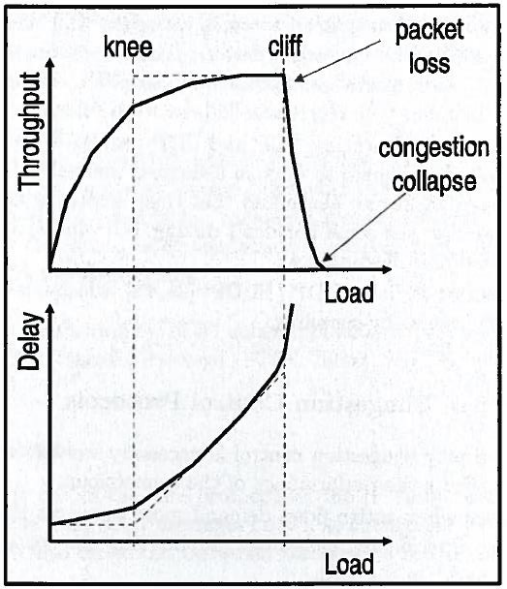
-> D'abord mise en file d'attente des paquets

-> 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 : gérer 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

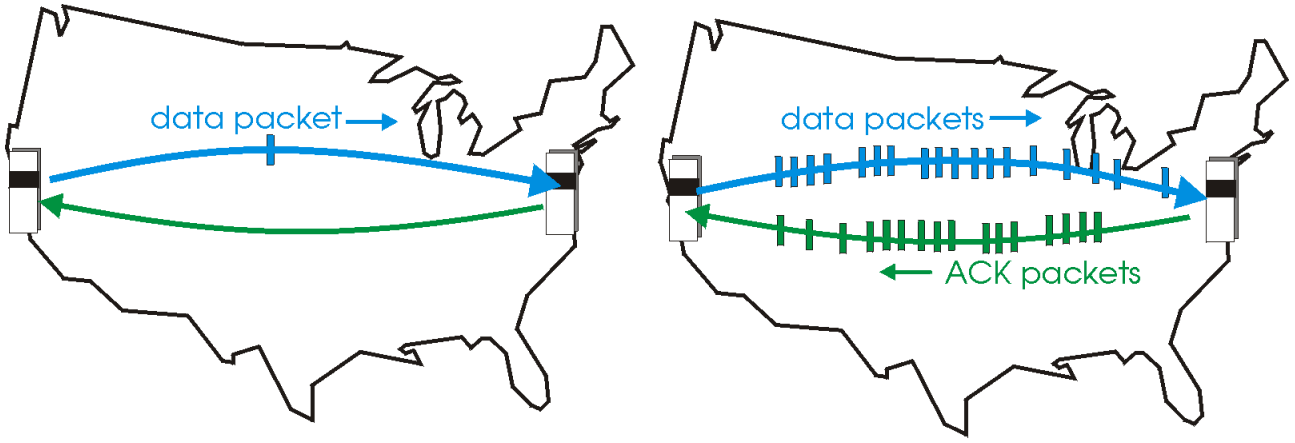
Contrôle de congestion assisté par le réseau:

- 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

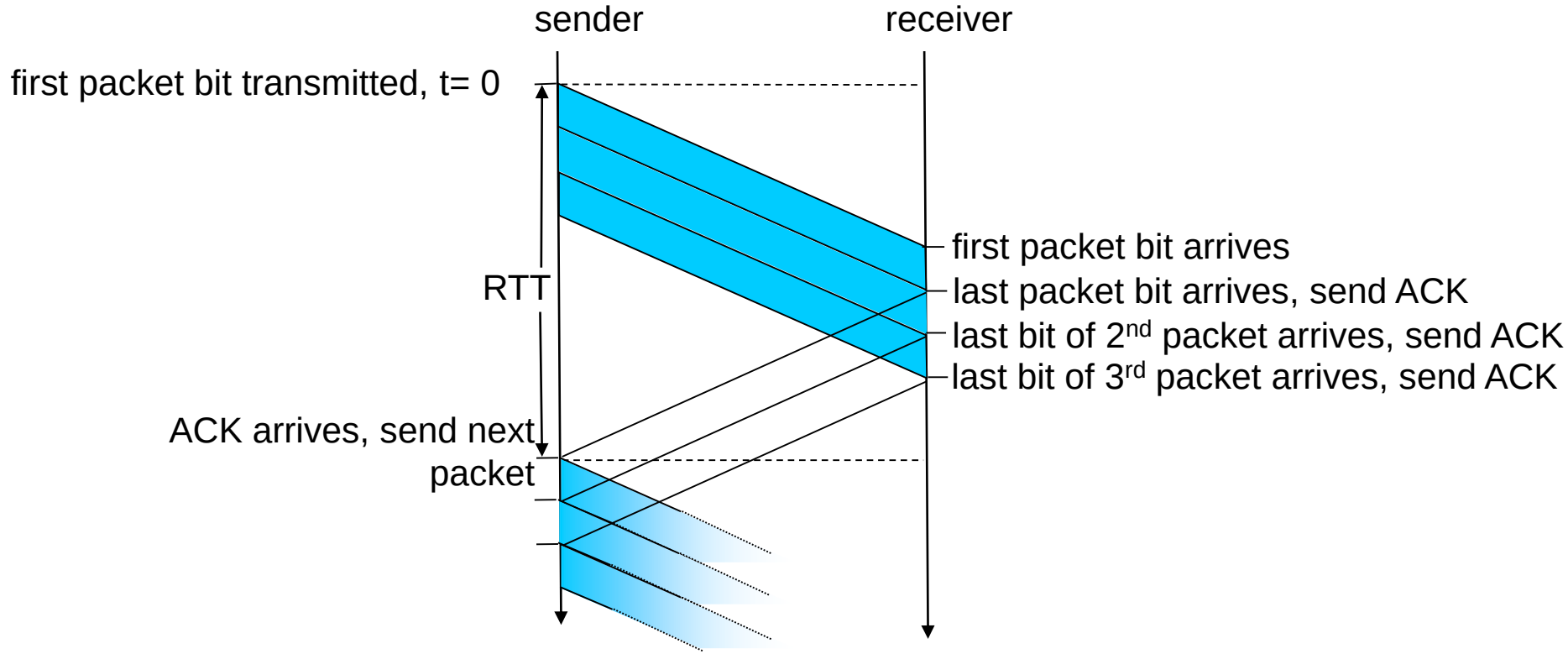
- La plage de numéros de séquence doit être augmentée
- Mise en mémoire à l'émetteur et au récepteur



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

(b) a pipelined protocol in operation

Pipelining: augmente l'utilisation



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

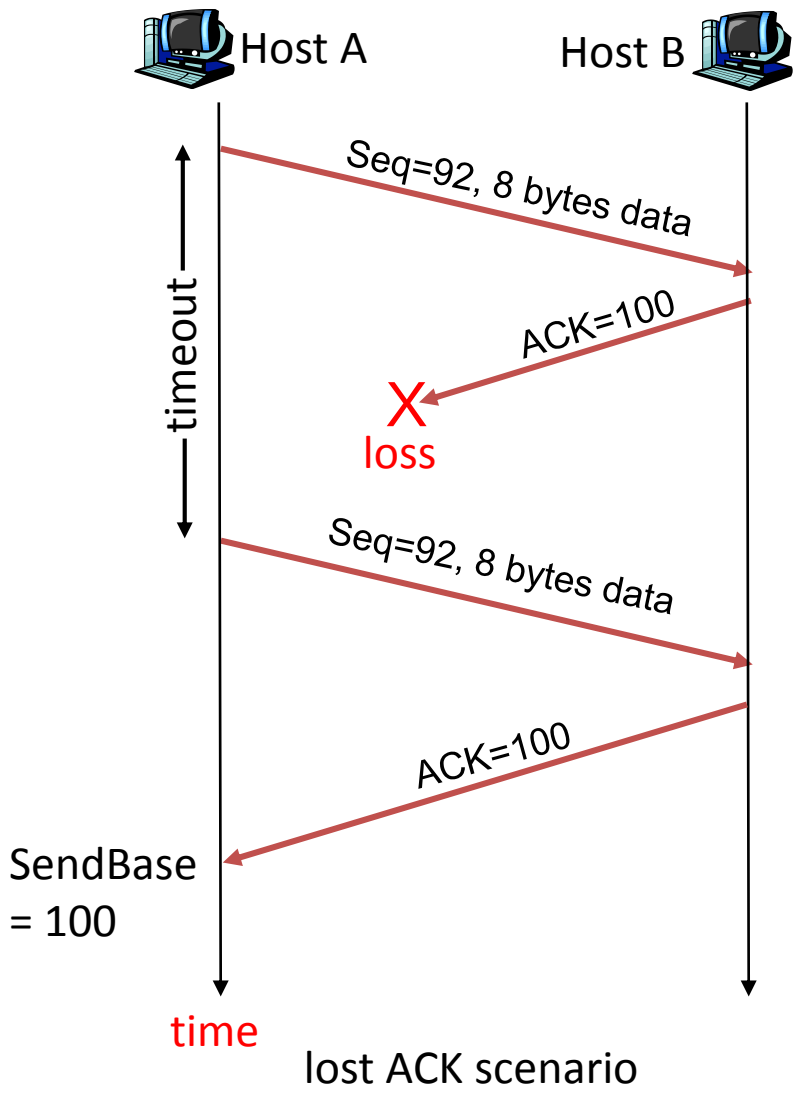
Increase utilization by a factor of N=3!

Plan général du cours

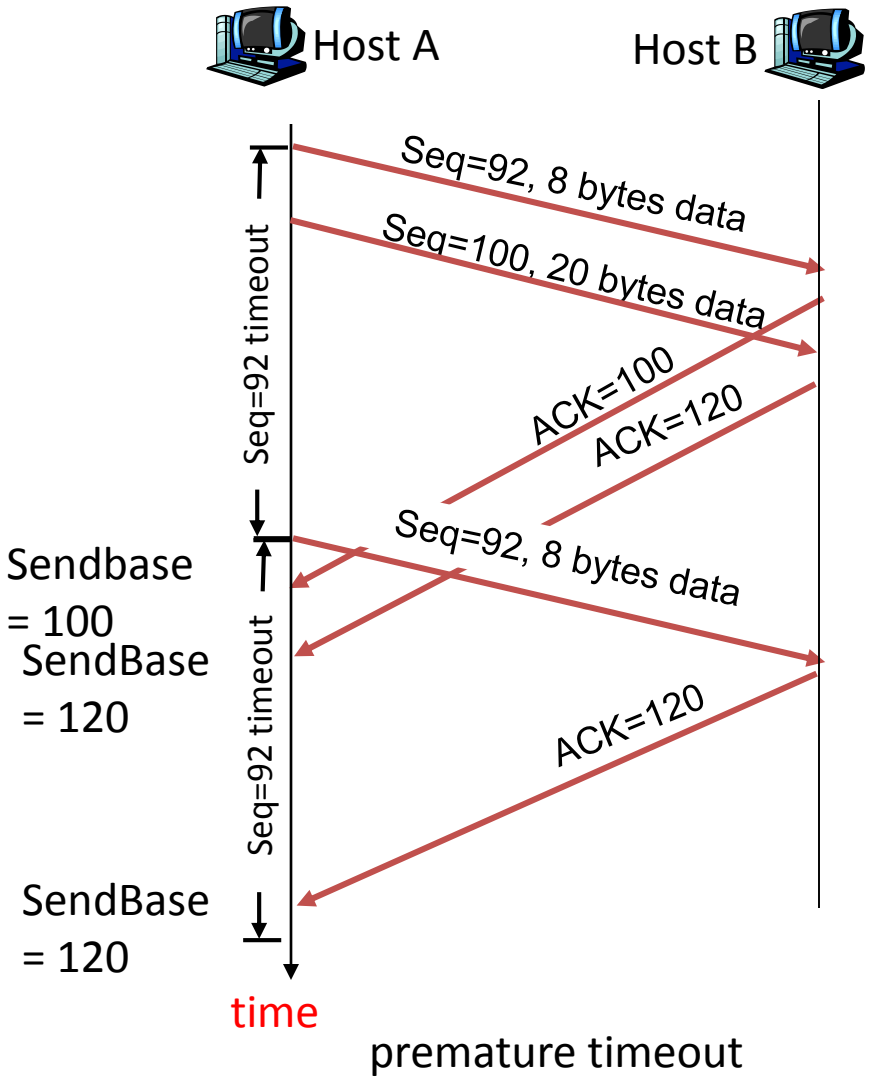
Transfert fiable TCP

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

Scenarios de retransmission TCP



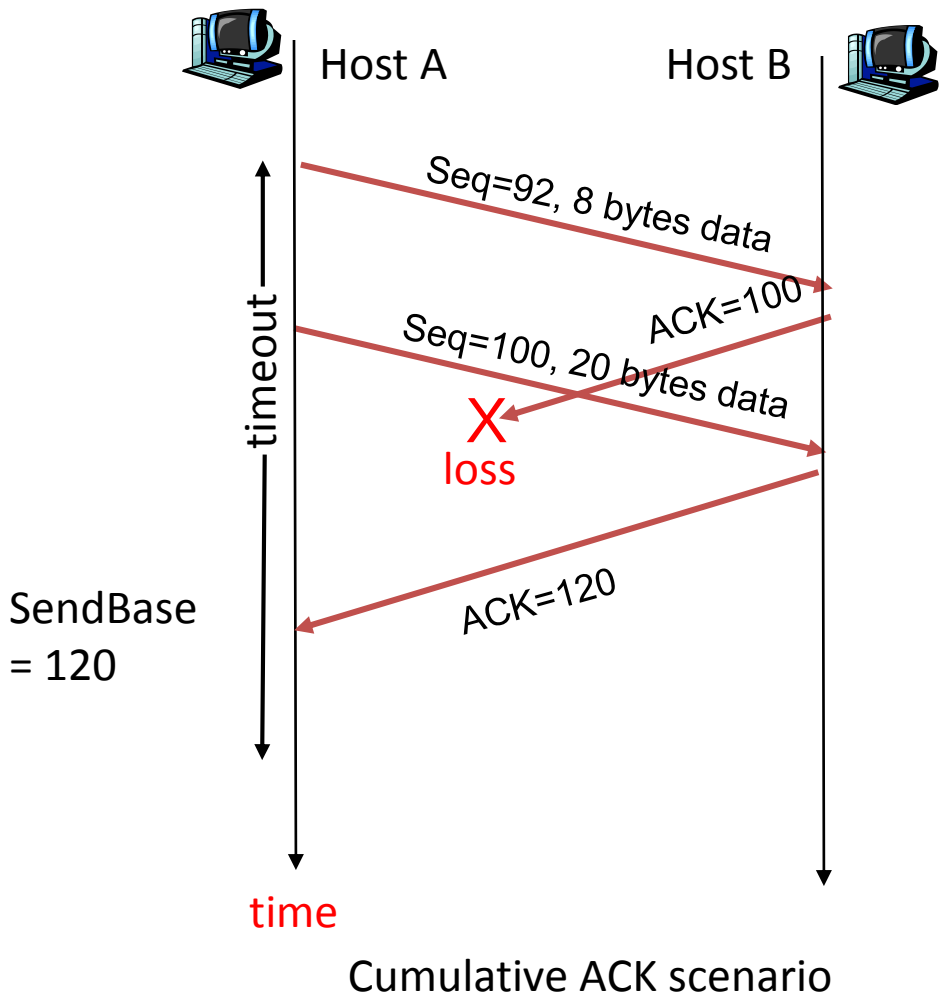
lost ACK scenario



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

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

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

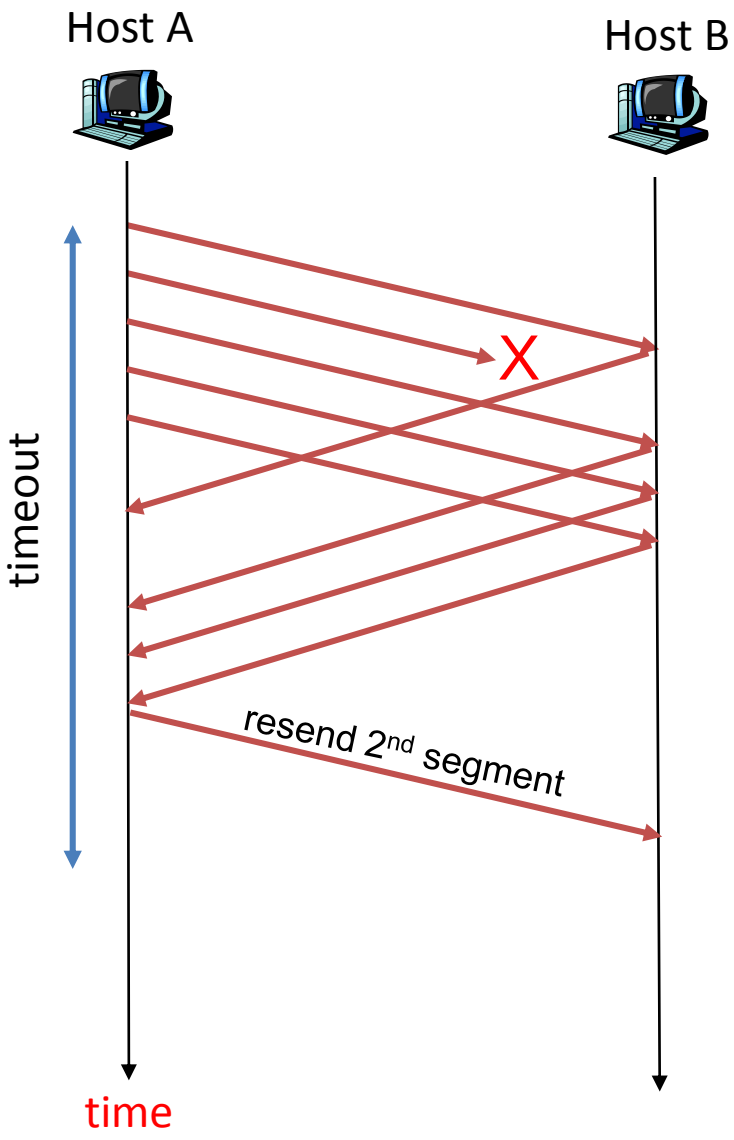
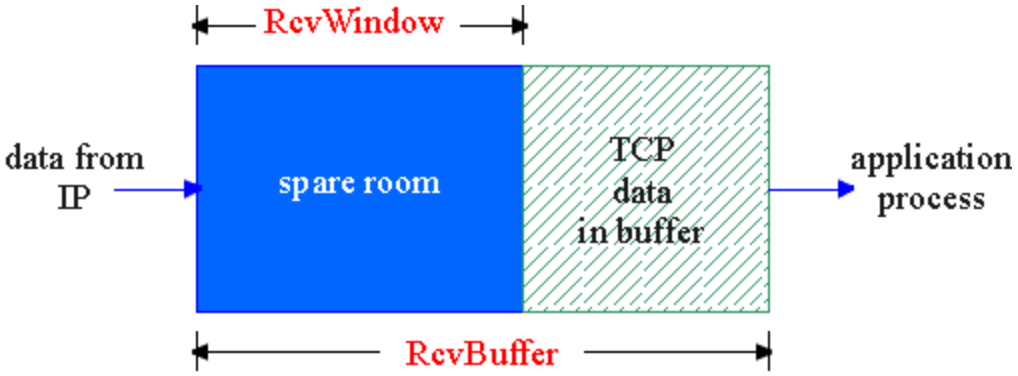


Figure 3.37 Resending a segment after triple duplicate ACK

Contrôle de flux TCP

- Le récepteur TCP a un buffer de réception:



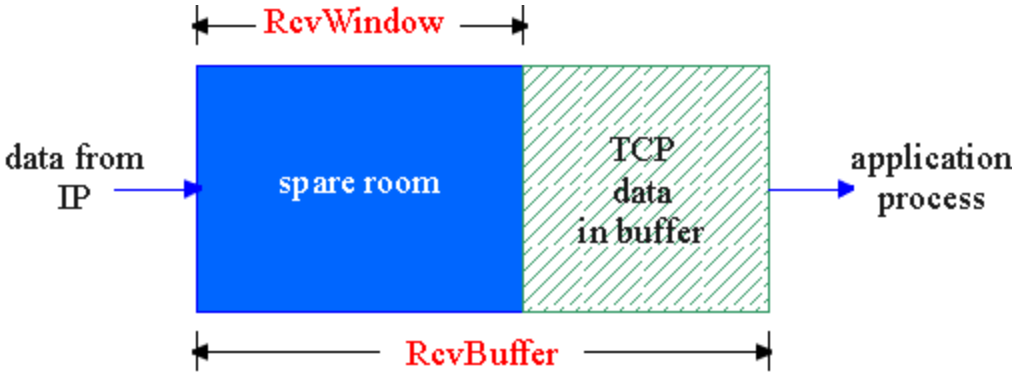
- Le processus application peut lire les données du buffer lentement

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

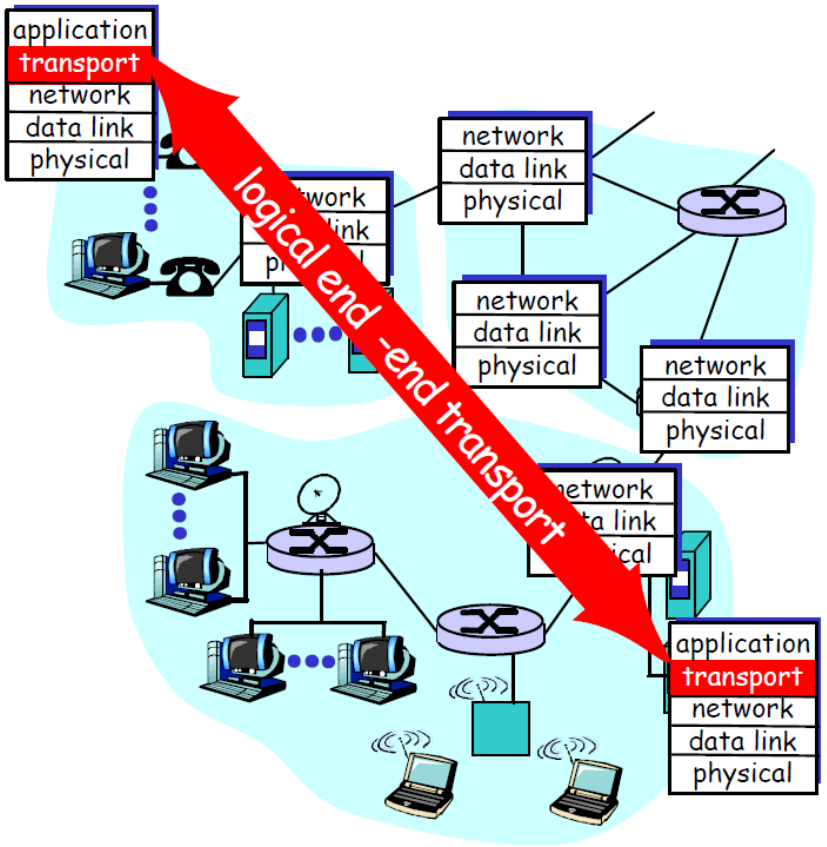
Plan général du cours

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 collapse sur Internet (dû à un nombre croissant de flows)

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

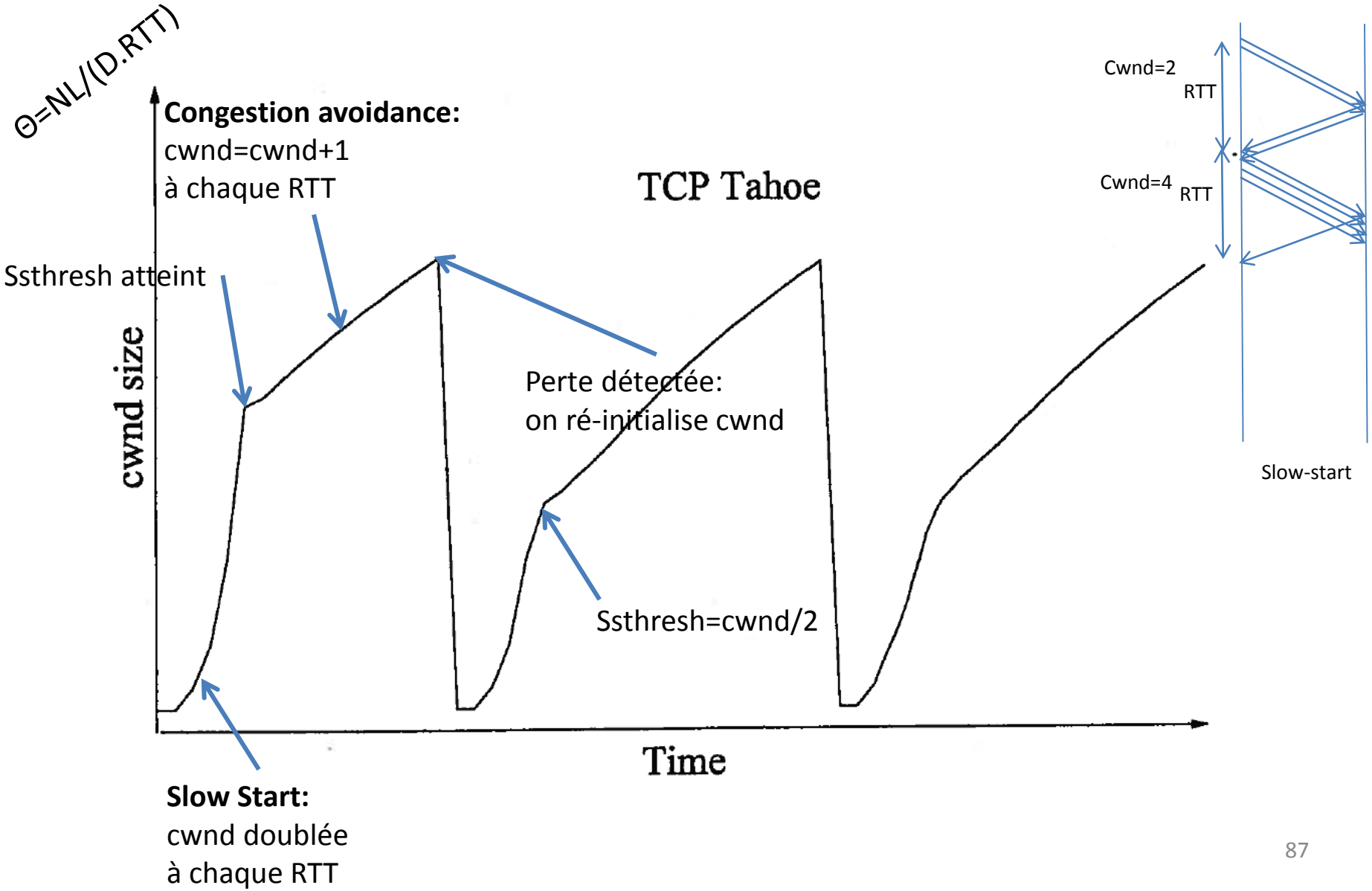
- Fonctionnement End-to-End (E2E): exécution du protocole qu'aux extrêmités



Principe de TCP

- Principe: pour chaque connexion, TCP maintient une ***congestion window (cwnd)*** **= nombre de segments TCP envoyés non encore acknowledgés**.
- 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,

$$\text{Débit} = \frac{\text{cwnd}}{\text{RTT}} \text{ Bytes/sec}$$

- **`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

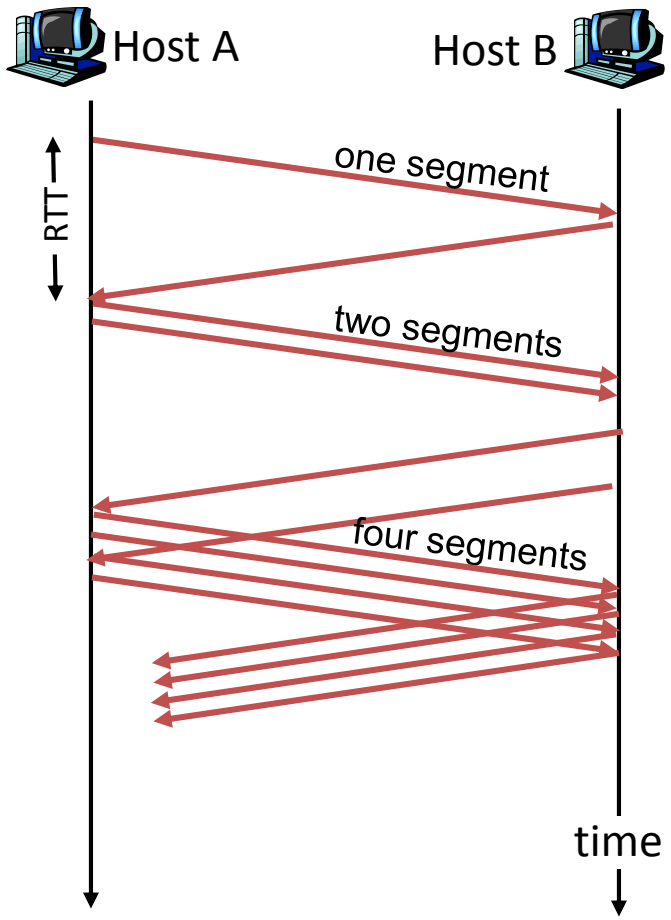
3 mécanismes:

- AIMD
- slow start
- prudent après timeout

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

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



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

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

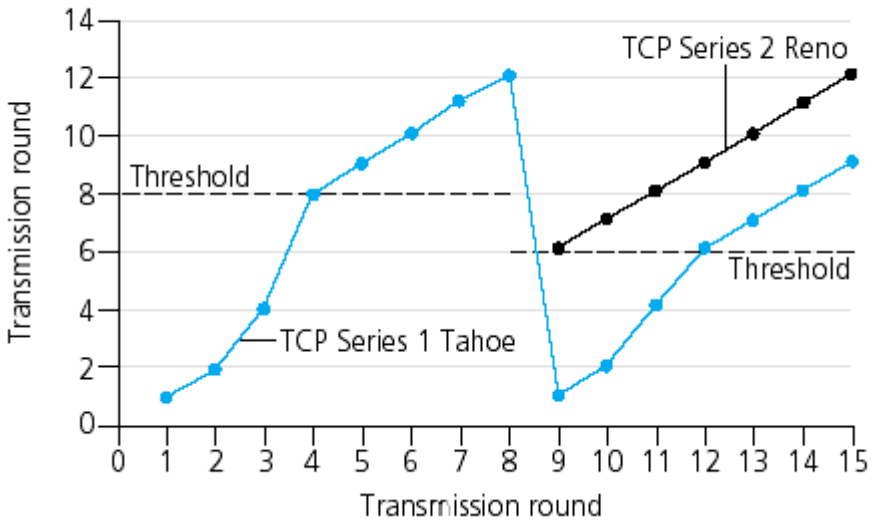
Raffinement

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

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

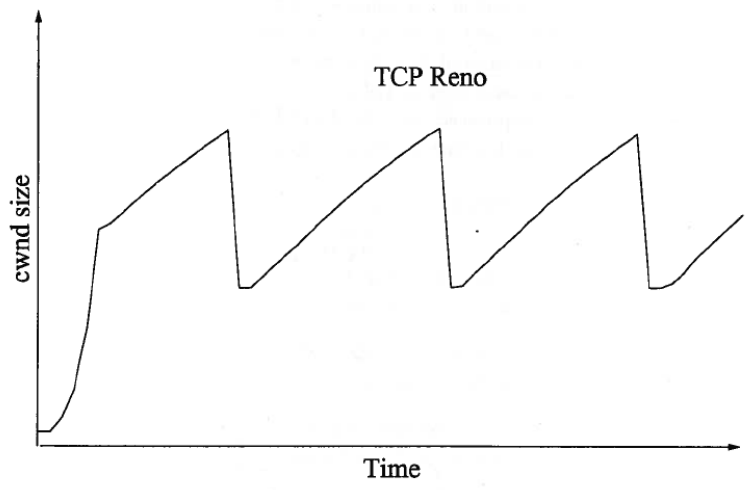
Q: Comment diminuer **`cwnd`** sur détection de perte ?

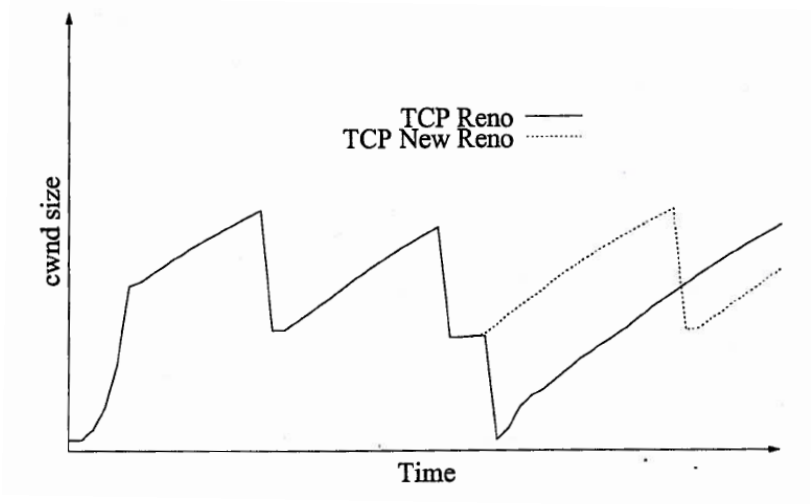
A: =1 si timeout, **`cwnd`** =/2 sinon



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`** and **`cwnd`** est mise à **`ssthresh`**.
- Quand un timeout se produit, **`ssthresh`** est mis à **`cwnd/2`** et **`cwnd`** est mise à 1 MSS.

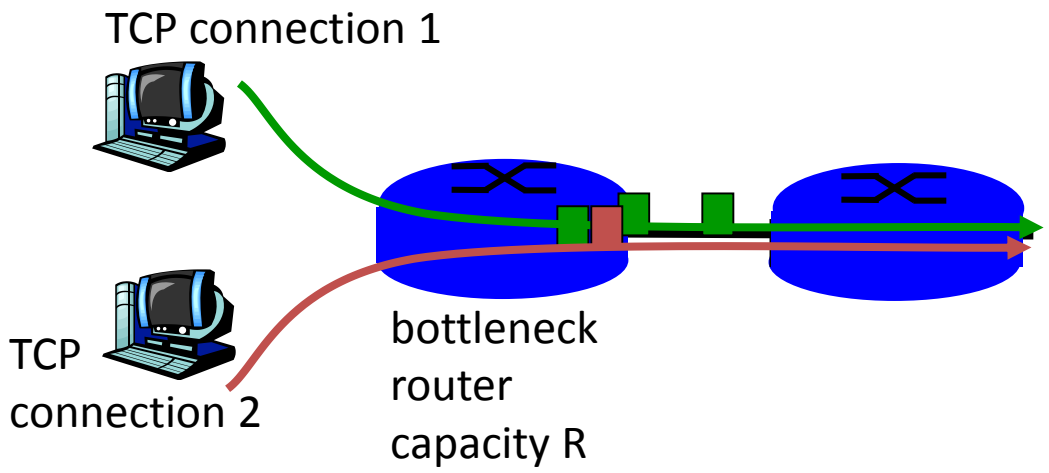
Jim Kurose, Keith Ross. *Computer Networking: A Top Down Approach*, 4th edition. Addison-Wesley, July 2007.

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

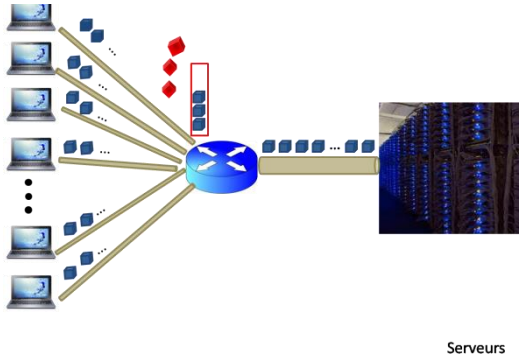
Source principale: Cours de J. Drouot (ESIL)

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

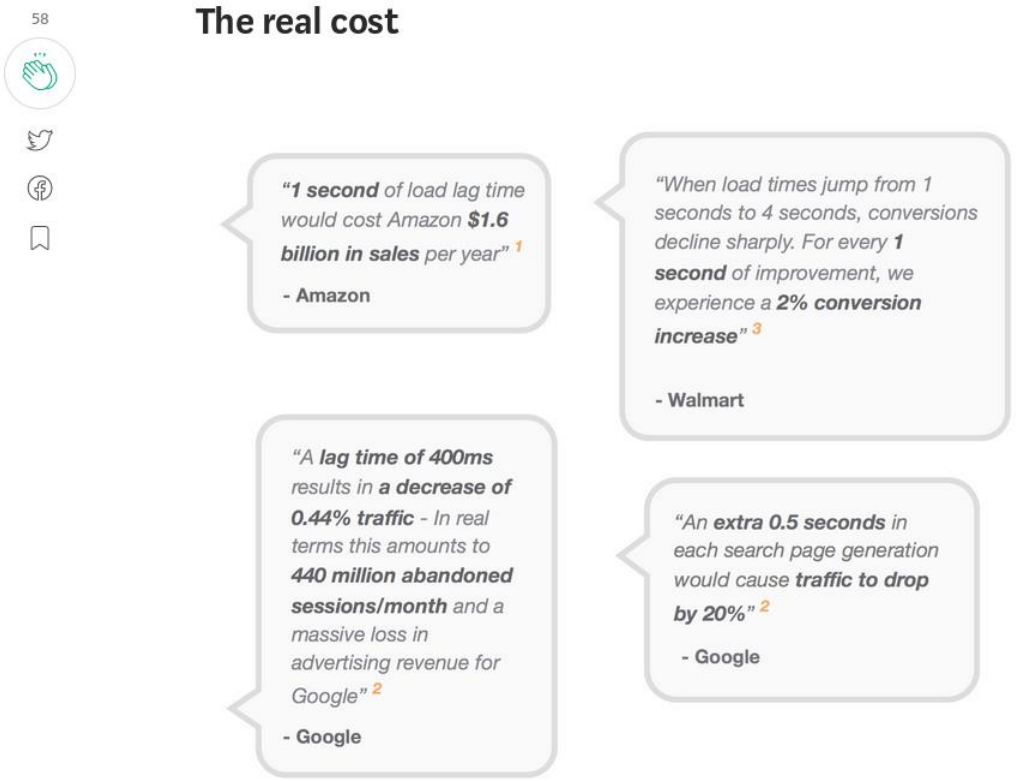


- Rappel congestion : mise en file d'attente des paquets à un éqt intermédiaire si débit d'entrée > débit de sortie
- TCP, pour utiliser le débit maximum qu'il ne connaît pas, provoque de la congestion quand il dépasse le débit, et réduit le débit quand il la détecte
- -> il remplit donc le buffer fréquemment
- -> 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

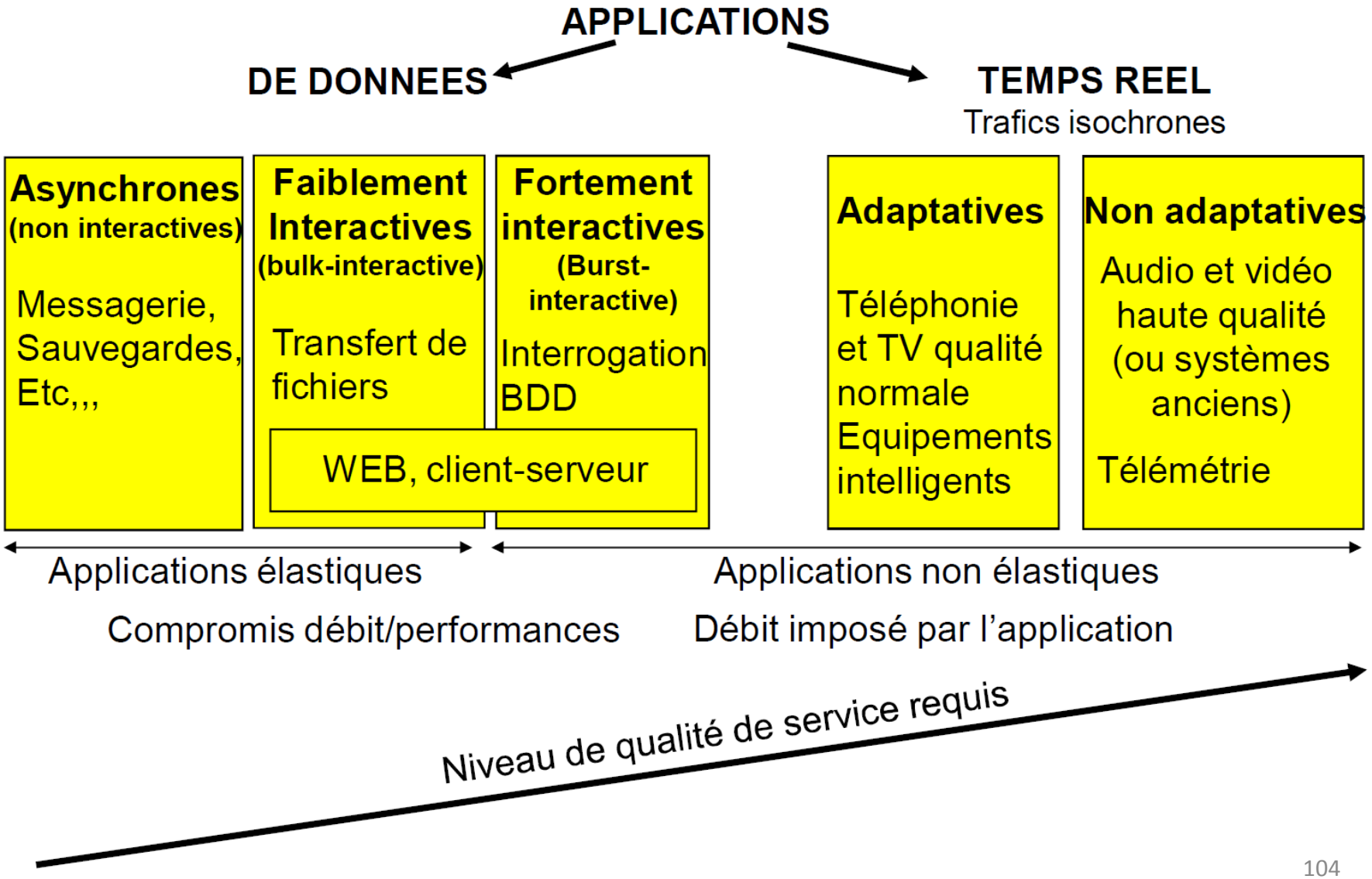
Application	Pertes	Débit	Sensibilité au délai
file transfer	no loss	elastic	no
e-mail	no loss	elastic	no
Web documents	no loss	elastic	no
real-time audio/video	loss-tolerant	audio: 5kbps-1Mbps video:10kbps-5Mbps	yes, 100's msec
stored audio/video	loss-tolerant	same as above	yes, few secs
interactive 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 : 2016-2021

→ Référence: Cisco Visual Networking Index Forecast

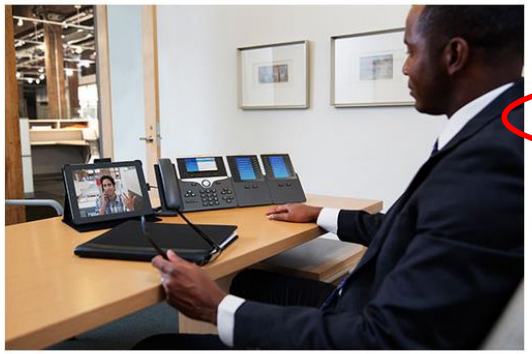
Cisco VNI

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

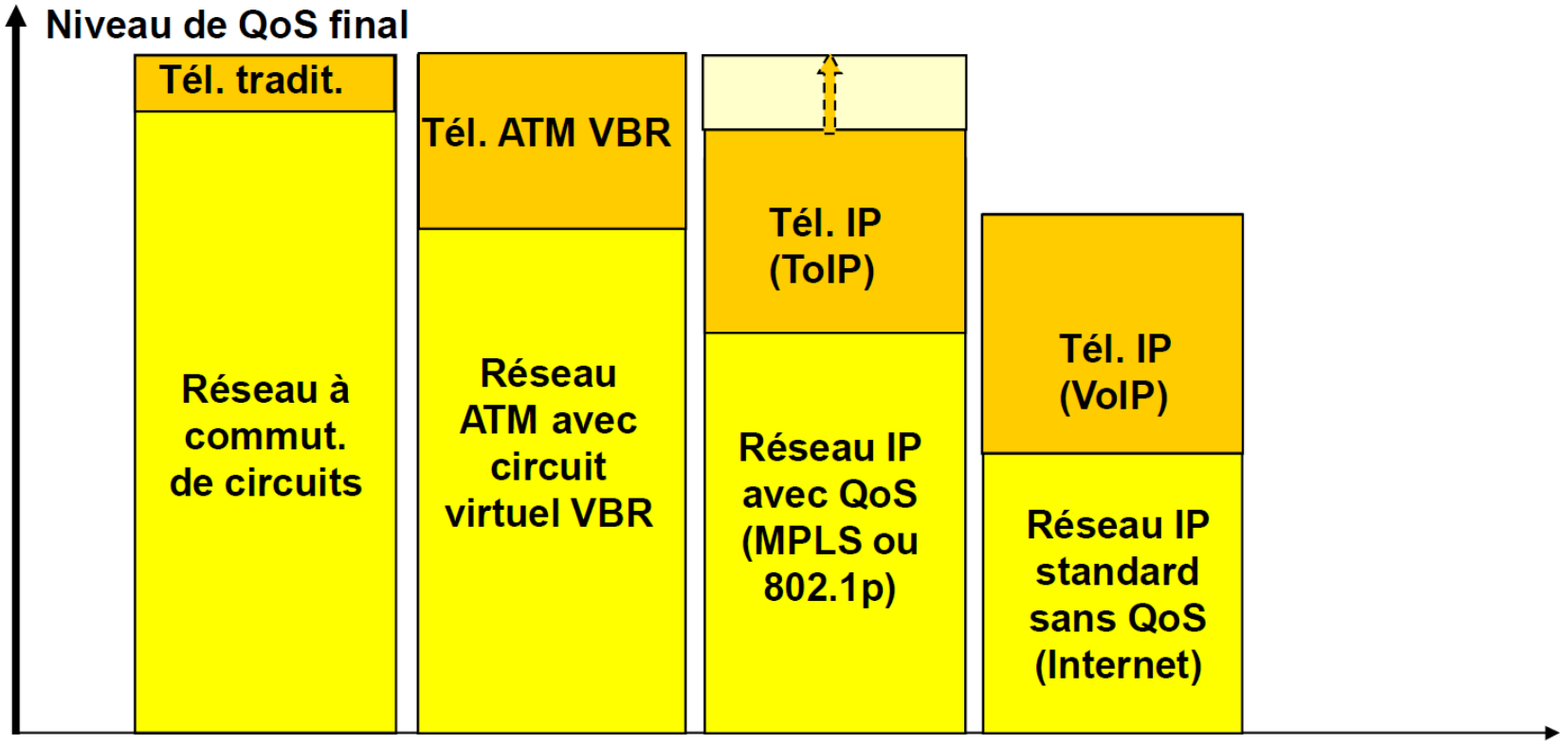
- **Les applications temps réel (voix, vidéo) ont un rythme imposé (*isochrone*) qui doit se maintenir de la source au destinataire**
  - Elles s'adaptent beaucoup plus mal à l'état de congestion du réseau
  - Le débit spécifié doit être assuré par le réseau sous peine de non fonctionnement
    - Ce 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

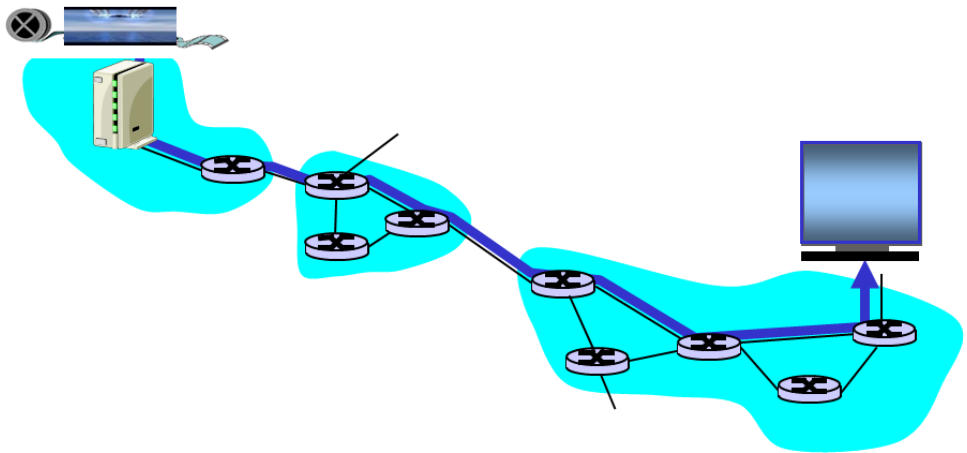
- **Les équipements d'extrémité sont de plus en plus intelligents**
  - Ils participent de plus en plus au niveau de QoS final obtenu
  - Et demandent moins de QoS au réseau
- **Tendance vers un réseau plus simple donc plus rapide et des stations plus intelligentes**



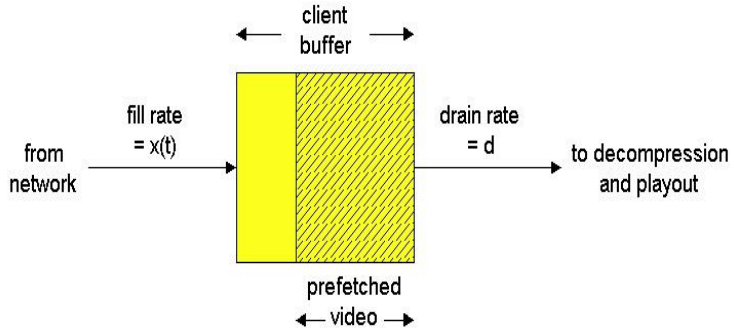
Plan général du cours

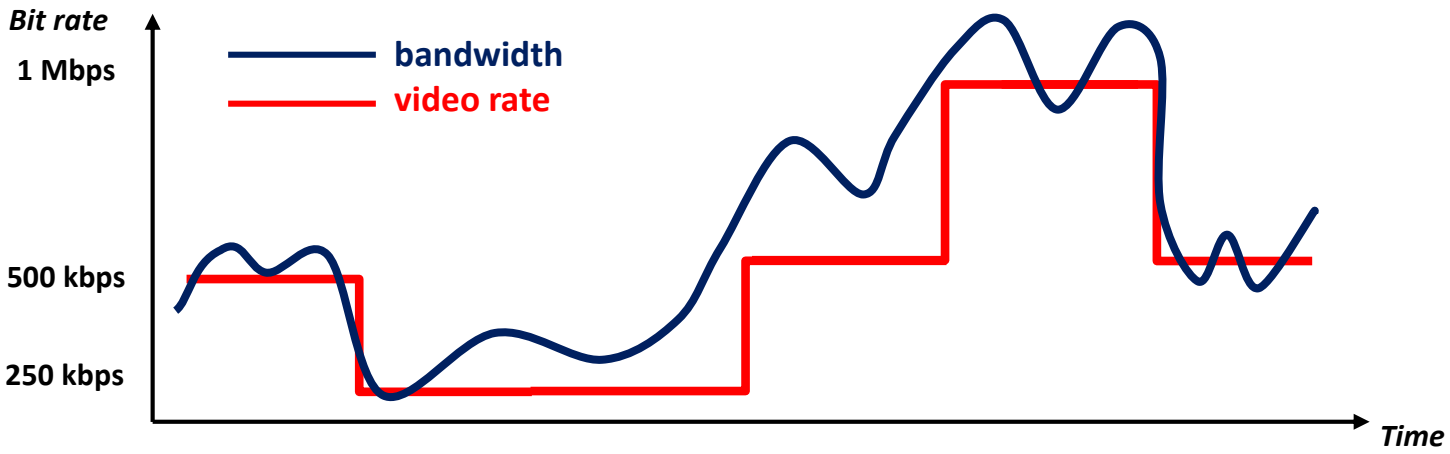
Streaming: définition

- Streaming: The content is sent from the source to the destination, which starts using it before its complete download.
- Terminology:
  - *video rate*: rate (in bps) at which the video has been encoded
  - *bandwidth*: rate available on the path from src to dst

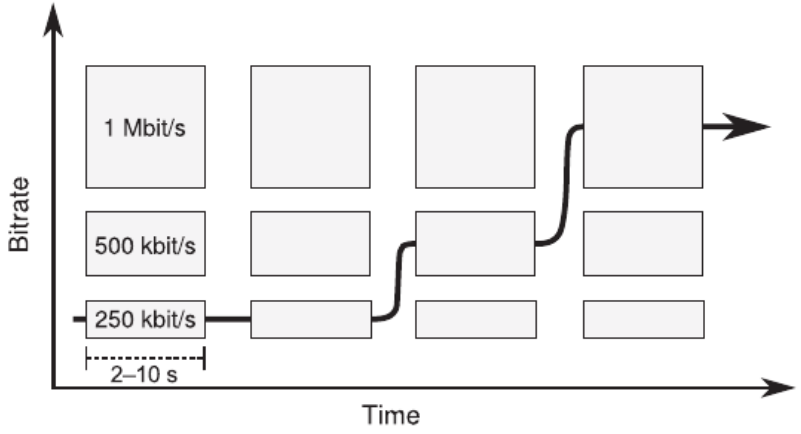


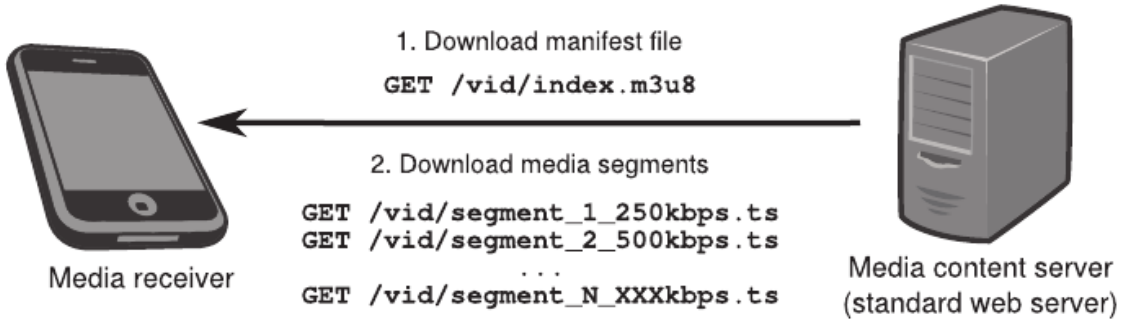
Streaming adaptatif: principe





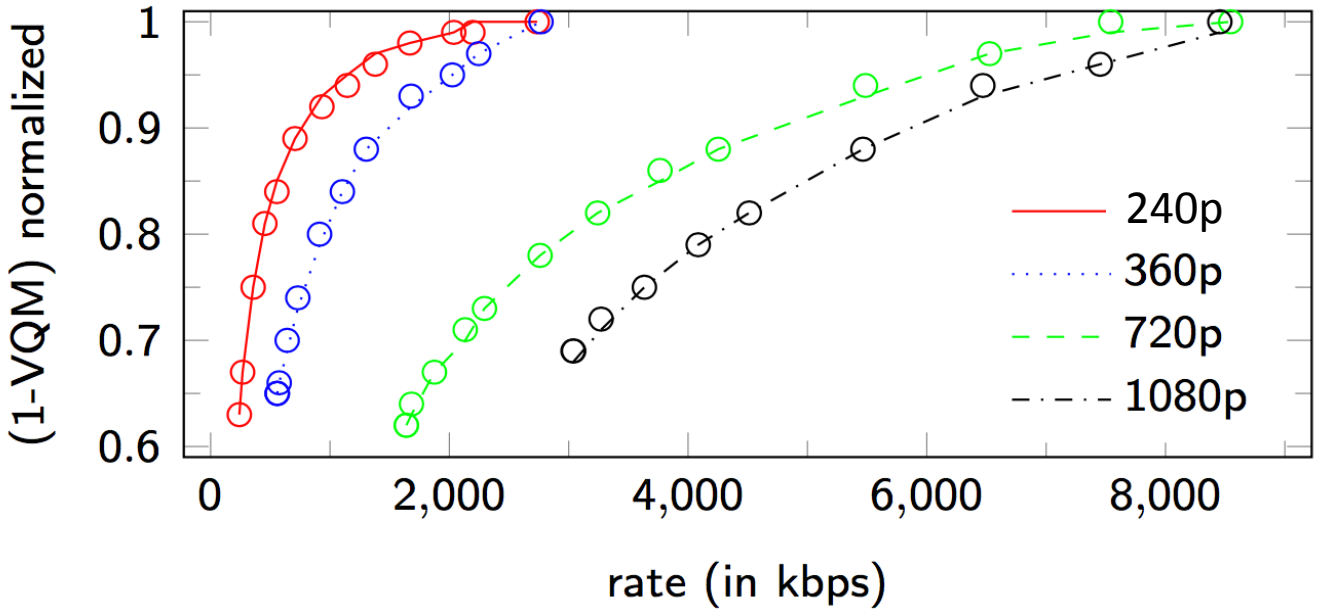
Streaming adaptatif: principe MPEG DASH



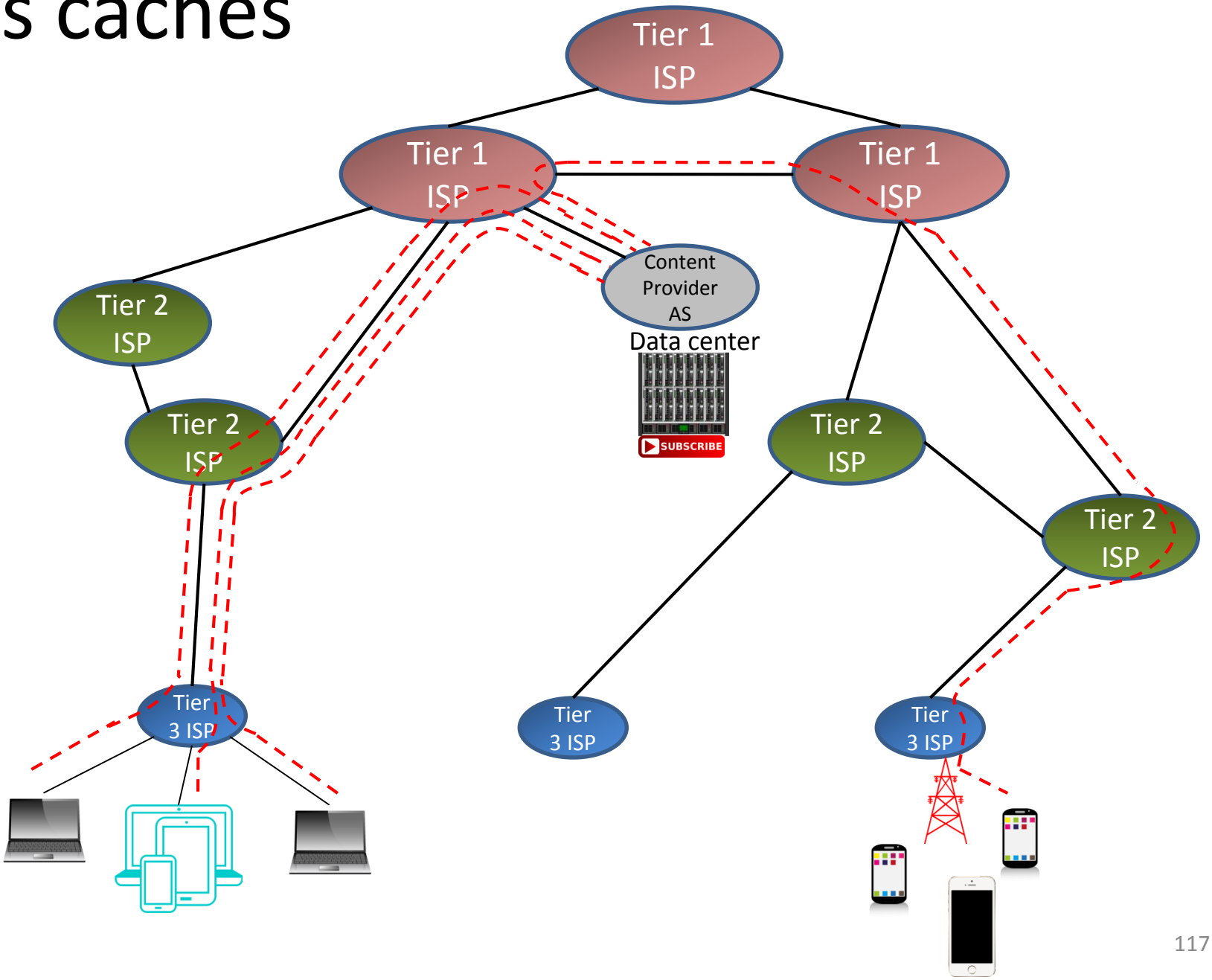


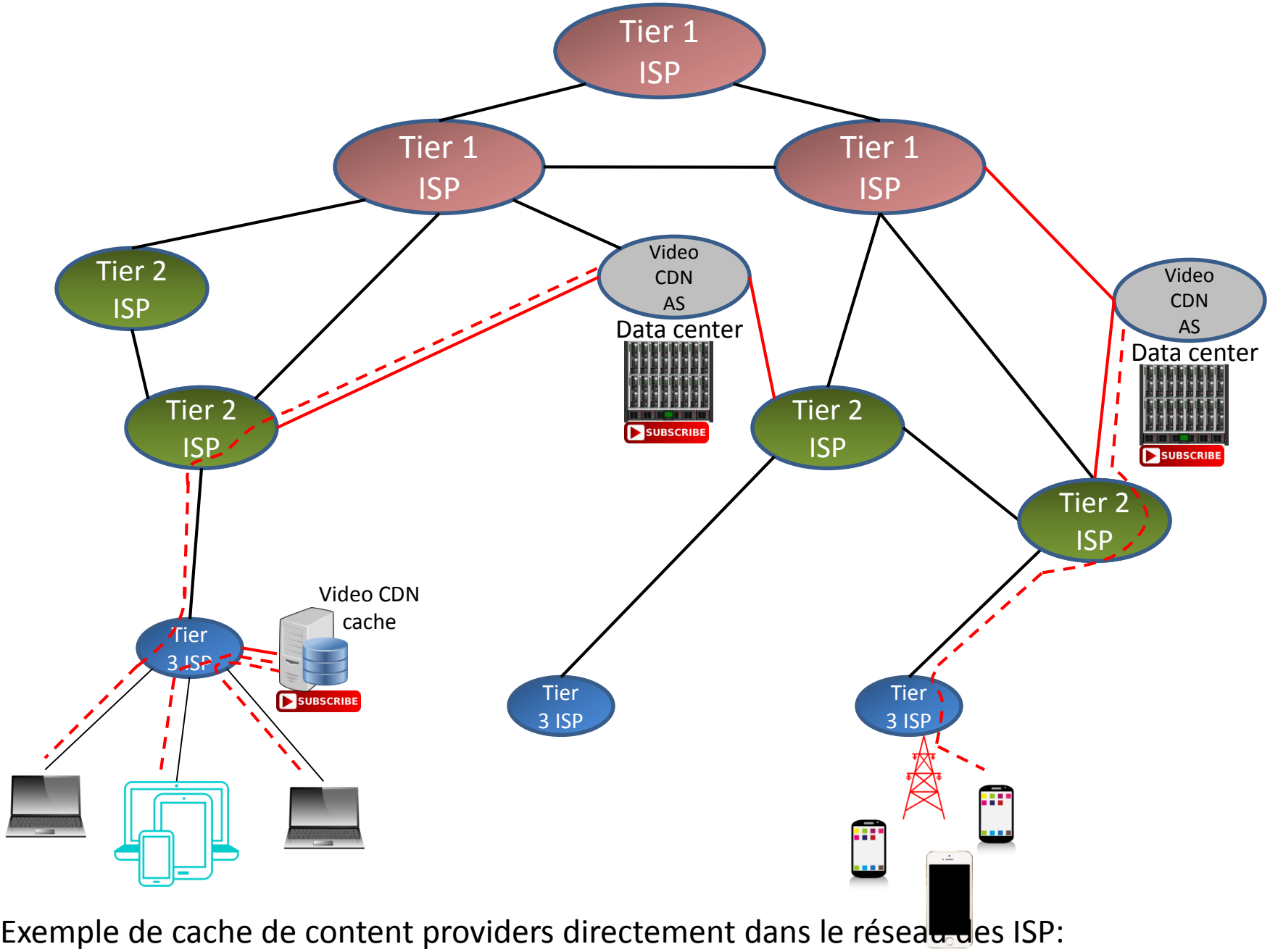
La qualité perçue n'est pas le débit du réseau

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



Les caches


Tier 1
ISP
Tier 1
ISP
Tier 1
ISP
Content
Provider
AS
Data center
SUBSCRIBE
Tier 2
ISP
Tier 2
ISP
Tier 2
ISP
Tier 2
ISP
Tier
3 ISP
Tier
3 ISP
Tier
3 ISP



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 10Gbps
- 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

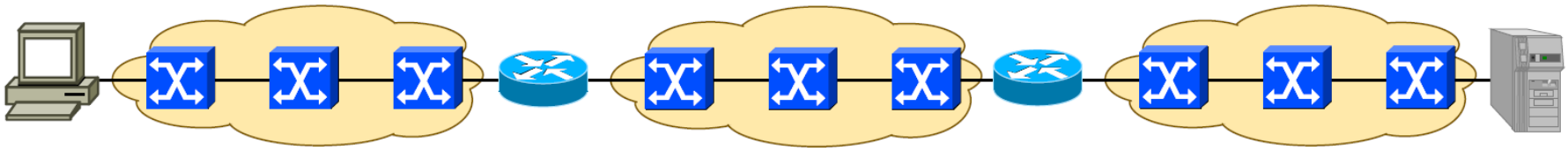
-> **MPLS-TE**

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

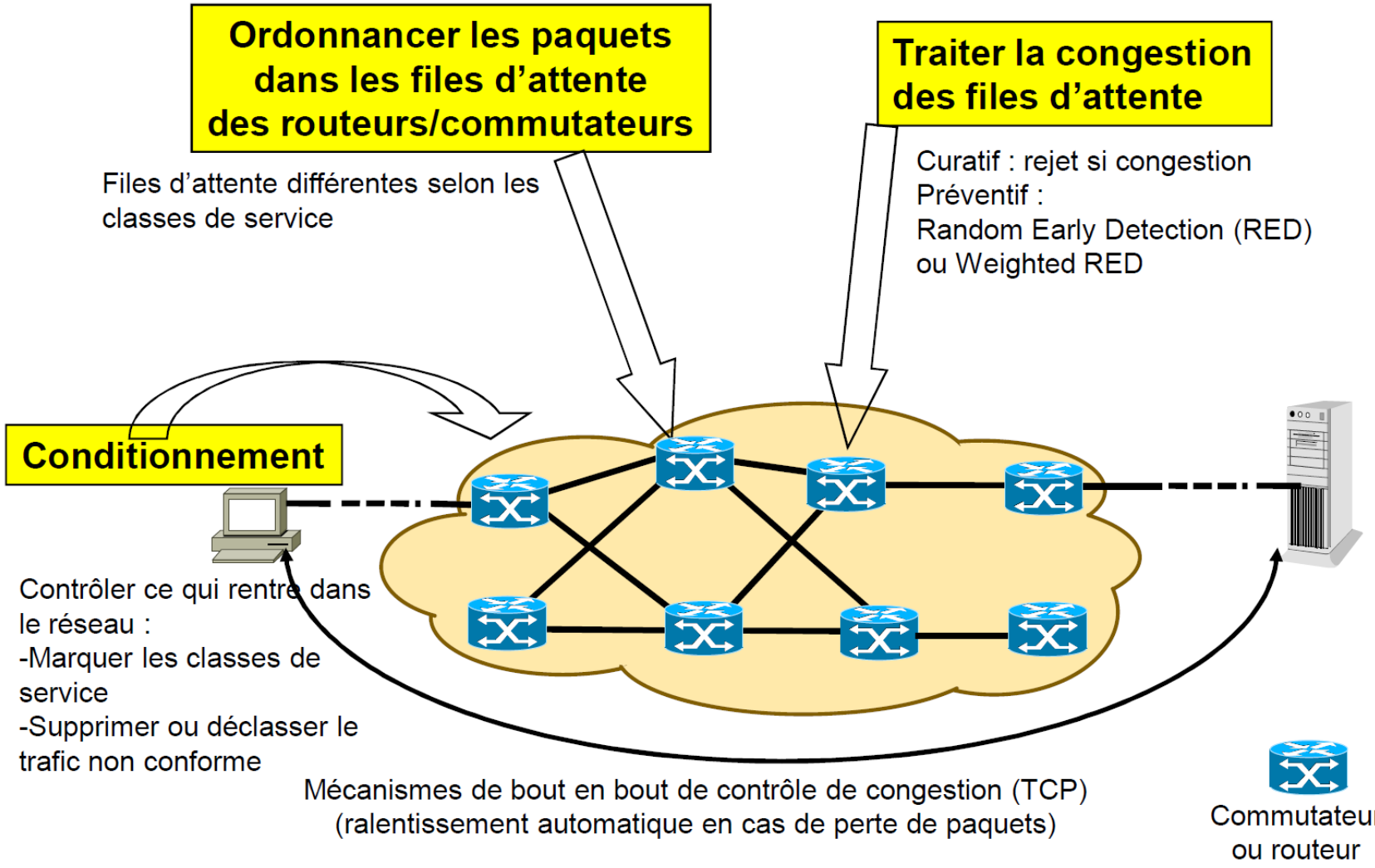
Optimisation: le cas d'un ISP

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

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

Plan général du cours

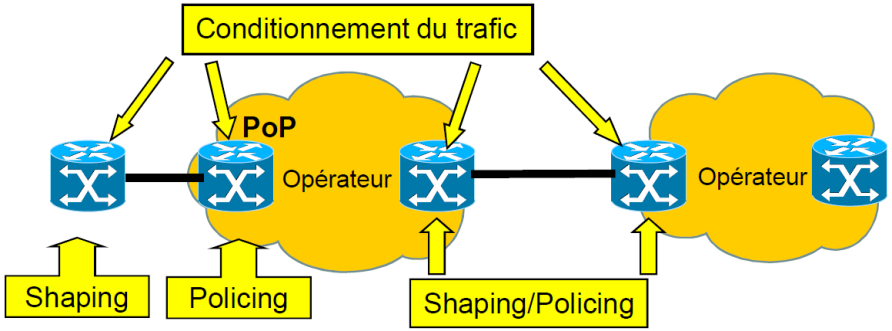
Plan de données: actions sur les données



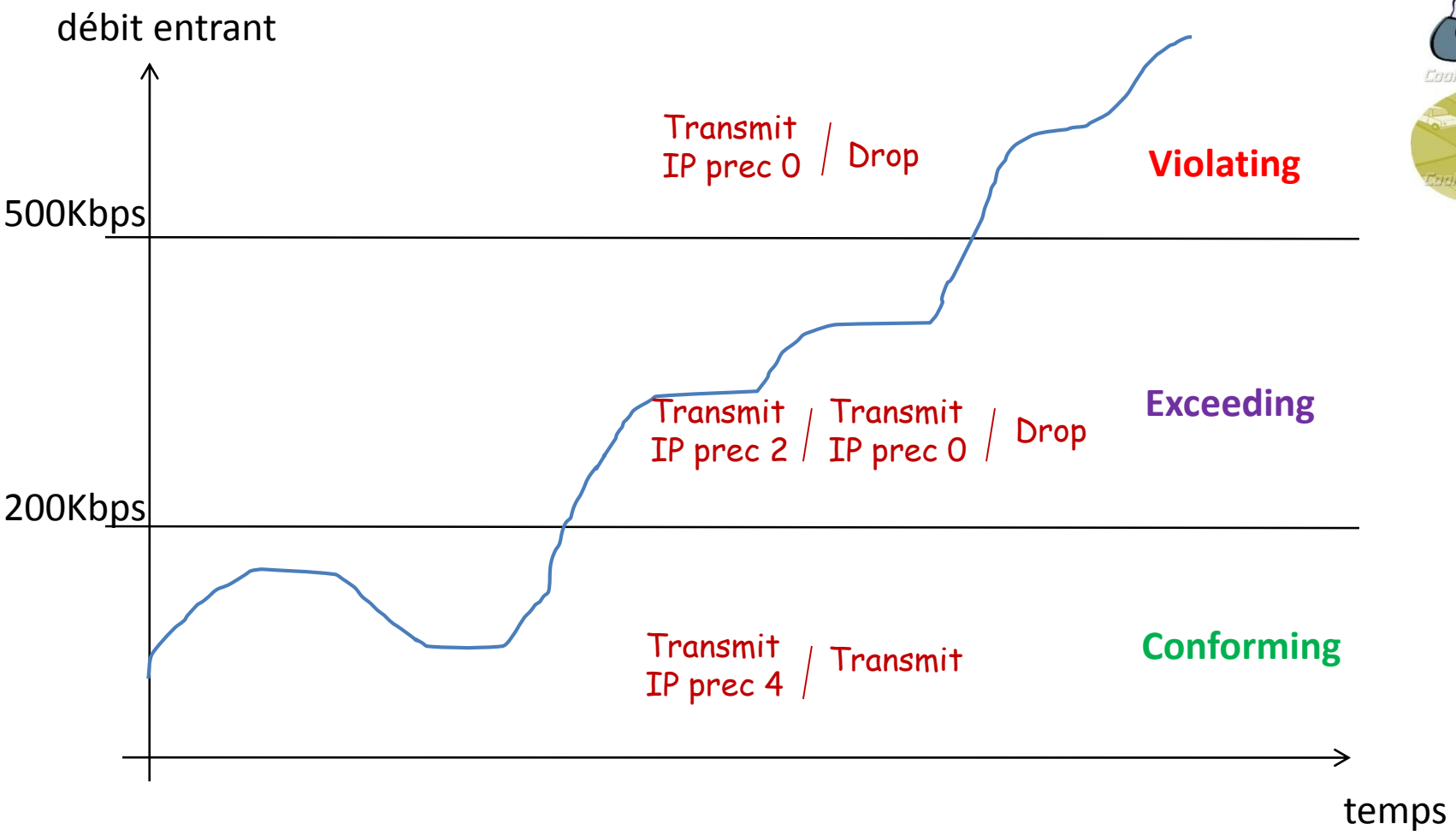
Policing et Shaping

- **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 accord(SLA):
- marquage ou abandon des paquets quand en excès

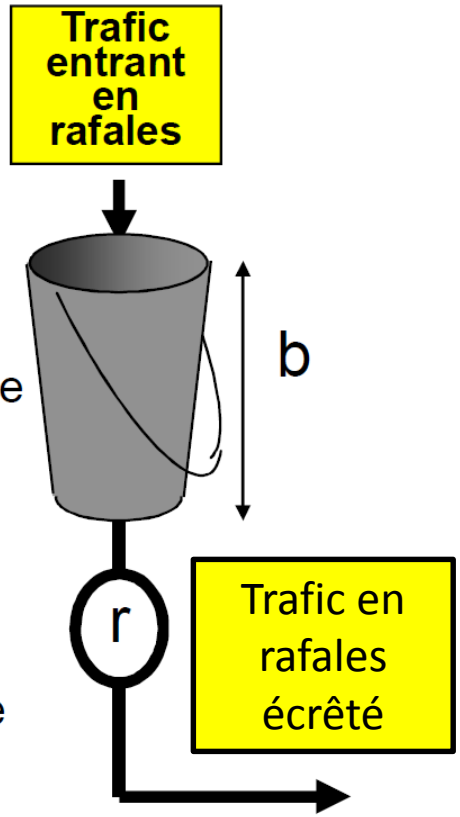


Policing et Shaping : les termes



Leaky bucket (seau percé)

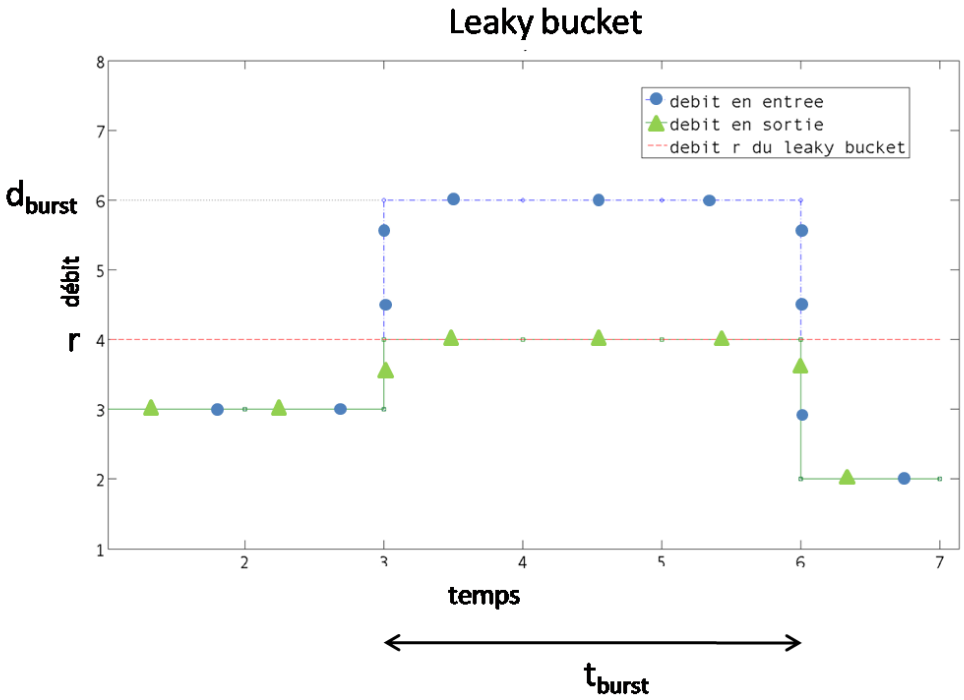
- **Un seau par flux ou classe de trafic**
  - Taille du seau : b octets
  - Le seau fuit au débit de r octets/s
  - Les octets d'un paquet entrant sont placés dans le seau
  - Un paquet qui ferait déborder le seau est rejeté
  - Quand le paquet de taille d est au fond du seau, on le retient pendant 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



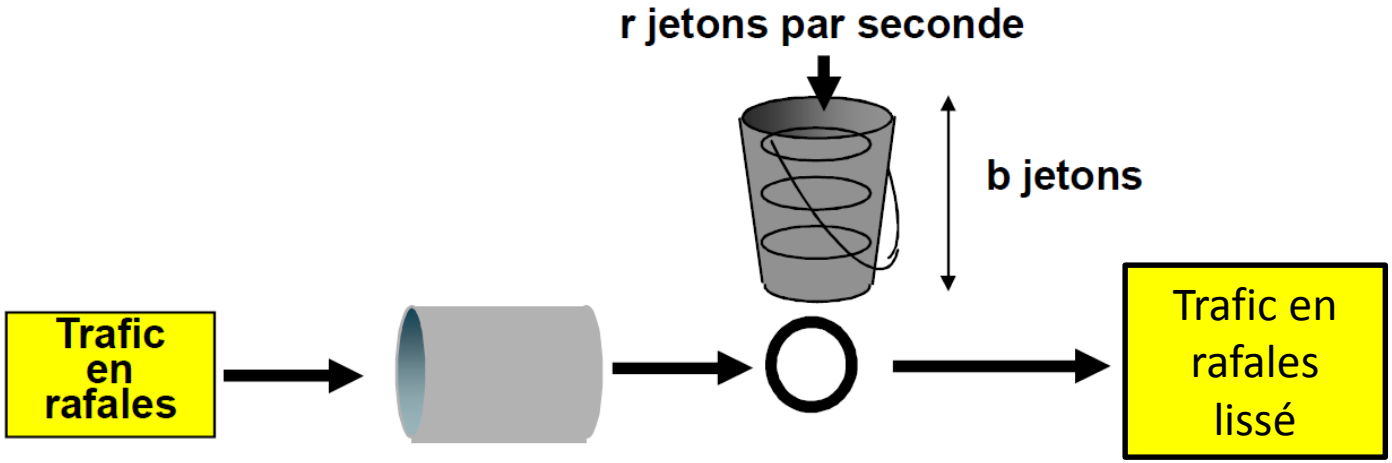
==> **Débit d'entrée écrêté à r.** Permet d'autoriser rafale de taille (debit_raf-r).tps_raf ≤ b bits, ou encore Δr ≤ b/Δt.

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

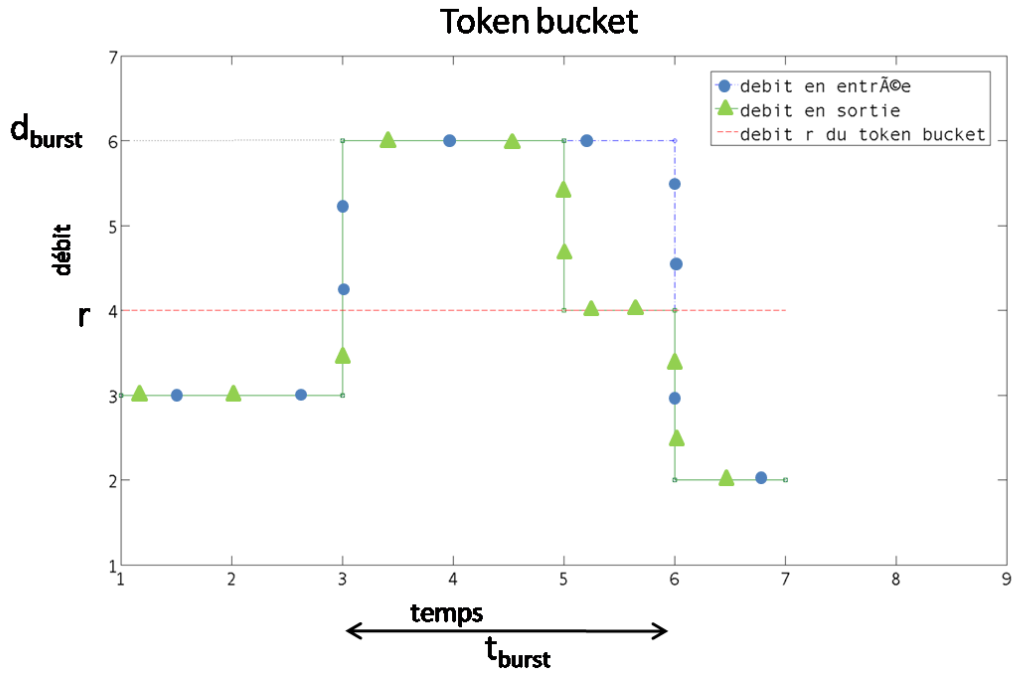
==> Débit de sortie peut dépasser r de Δr pendant Δt pourvu que: $(r+\Delta r)\Delta t \leq r\Delta t+b$, ou encore $\Delta r \leq b/\Delta t$.

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: $(r-x)t \geq b$

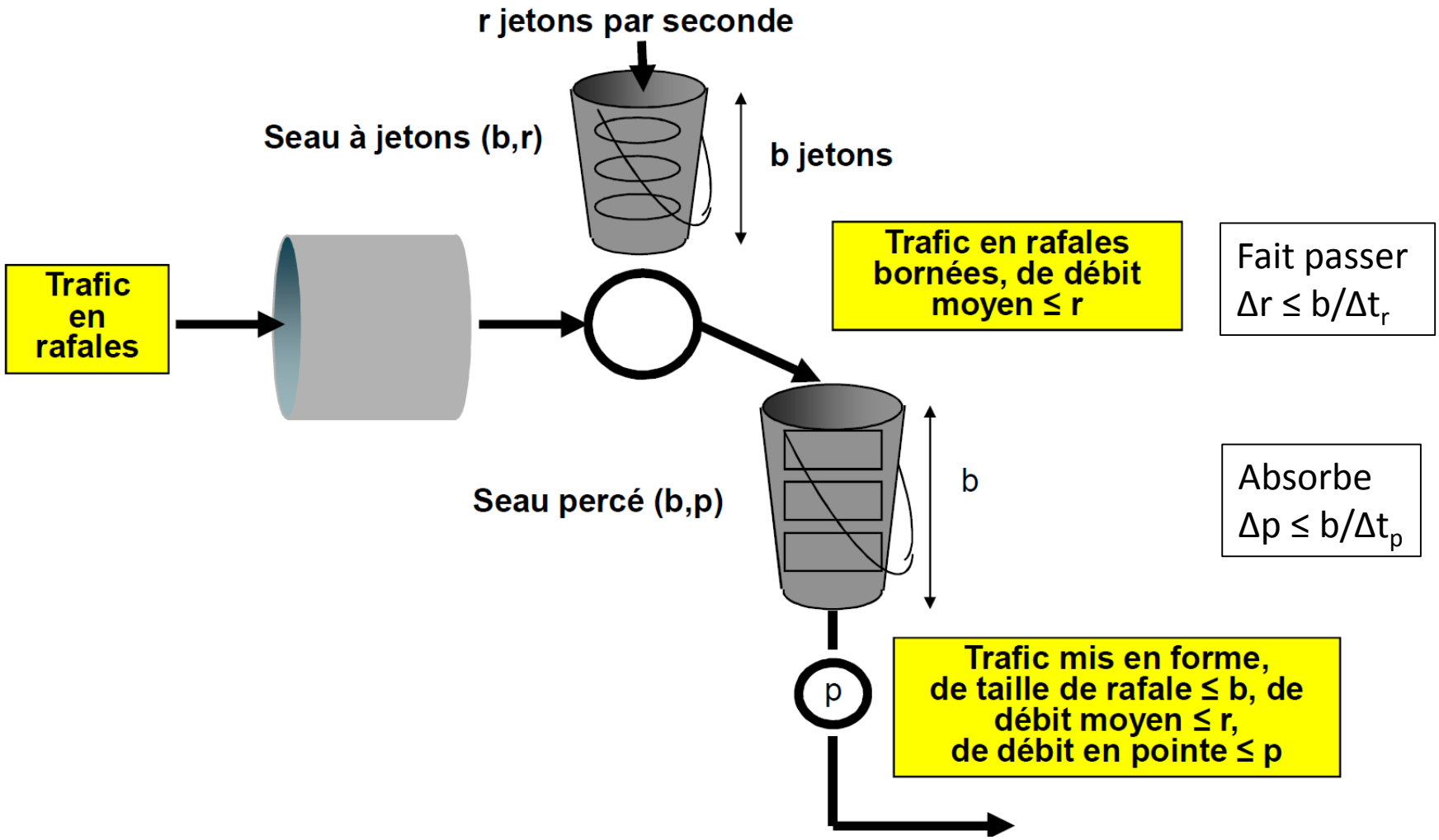
--> **Le débit moyen sur une longue période est toujours ≤ r**.

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**.



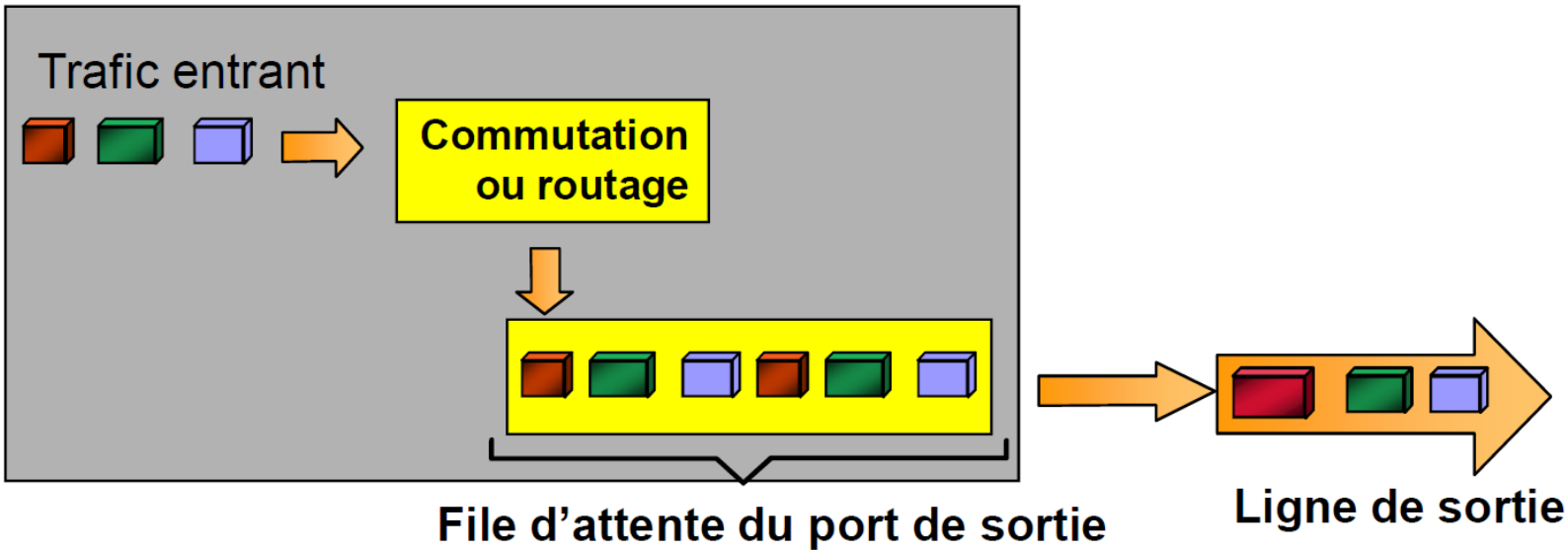
Utilisation combinée de *leaky* et *token buckets*



Plan général du cours

Files d'attente de type FIFO

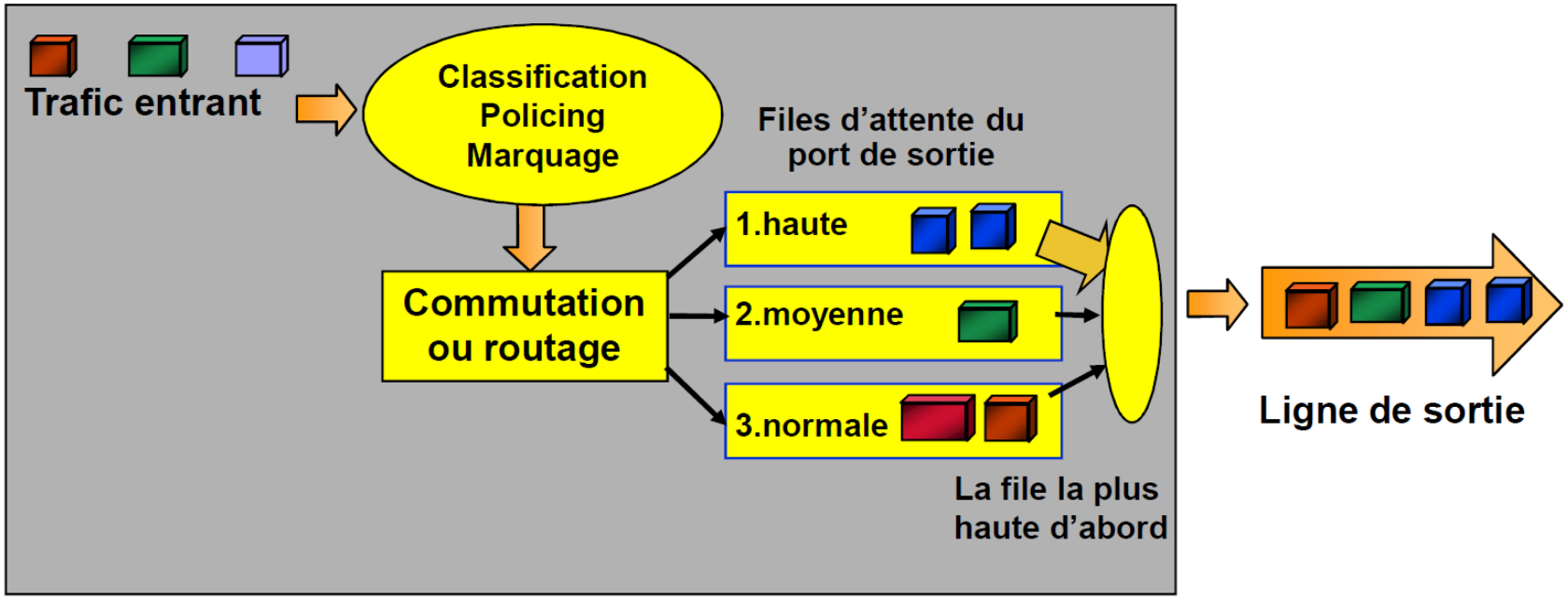
- **FIFO (First-In-First-Out)**
  - La plus simple et la plus courante
  - Fonctionne bien en absence de congestion
    - Rejet de paquets excédentaires en cas de congestion



- **Un transfert de fichiers peut monopoliser toute la bande passante**

Files d'attente *Priority Queuing*

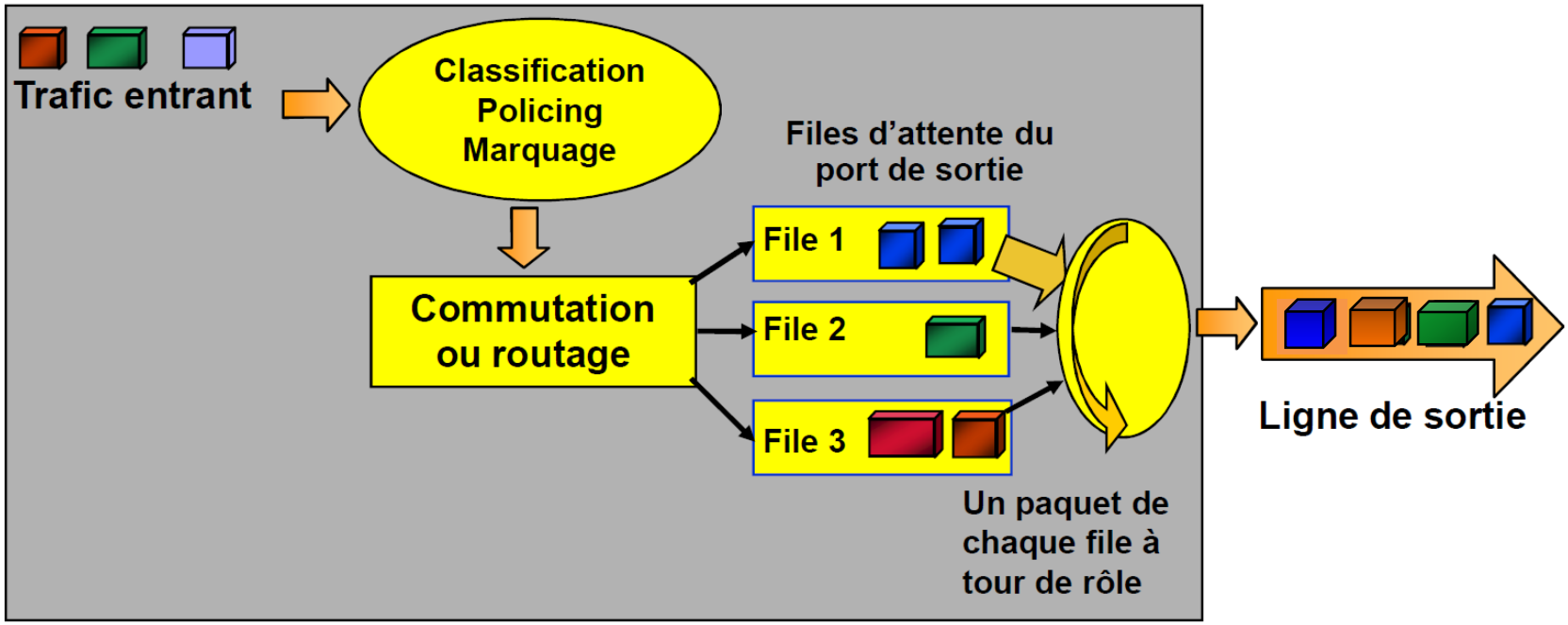
- **Files d'attente par priorités (priority queuing)**
  - Files d'attente différentes selon les classes de trafic
  - Les files d'attente sont traitées dans l'ordre des priorités
    - Une file d'attente n'est servie que si les files de plus haute priorité sont vides



- **Un type de trafic peut dominer les autres**
  - Si une file d'attente prioritaire ne se vide pas

Files d'attente *Round-robin Queuing*

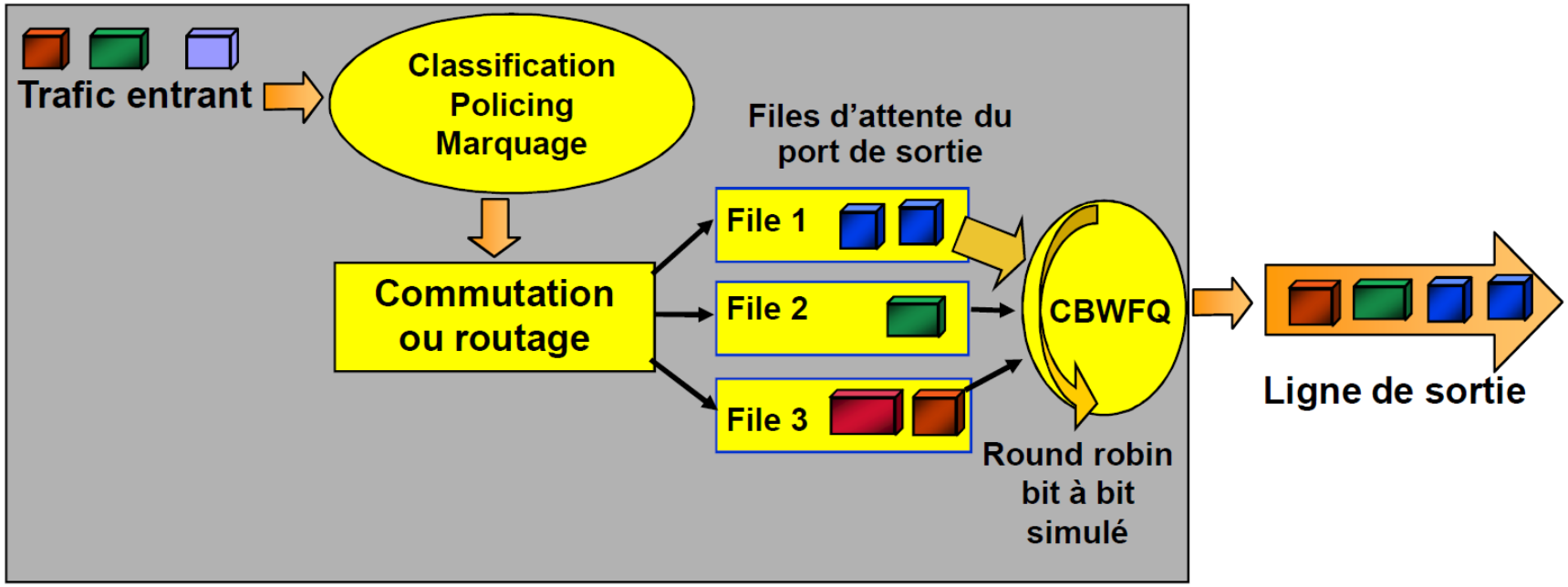
- On sert un paquet de chaque file non vide à tour de rôle
- Version pondérée (Weighted RRQ) : on sert plus d'un paquet par visite



- Empêche les trafics de se « cannibaliser »
- Mais favorise les flux ayant des paquets longs

Class-Based Weighted Fair Queuing

- **Multiplexage temporel (tranches de temps allouées cycliquement)**
  - Les conversations à faible volume ont donc une priorité effective
  - Les conversations à fort volume ne monopolisent pas la bande passante
  - Des poids permettent d'attribuer plus de BP aux trafics sensibles au délai



- **Permet de répartir la bande passante de façon équitable**

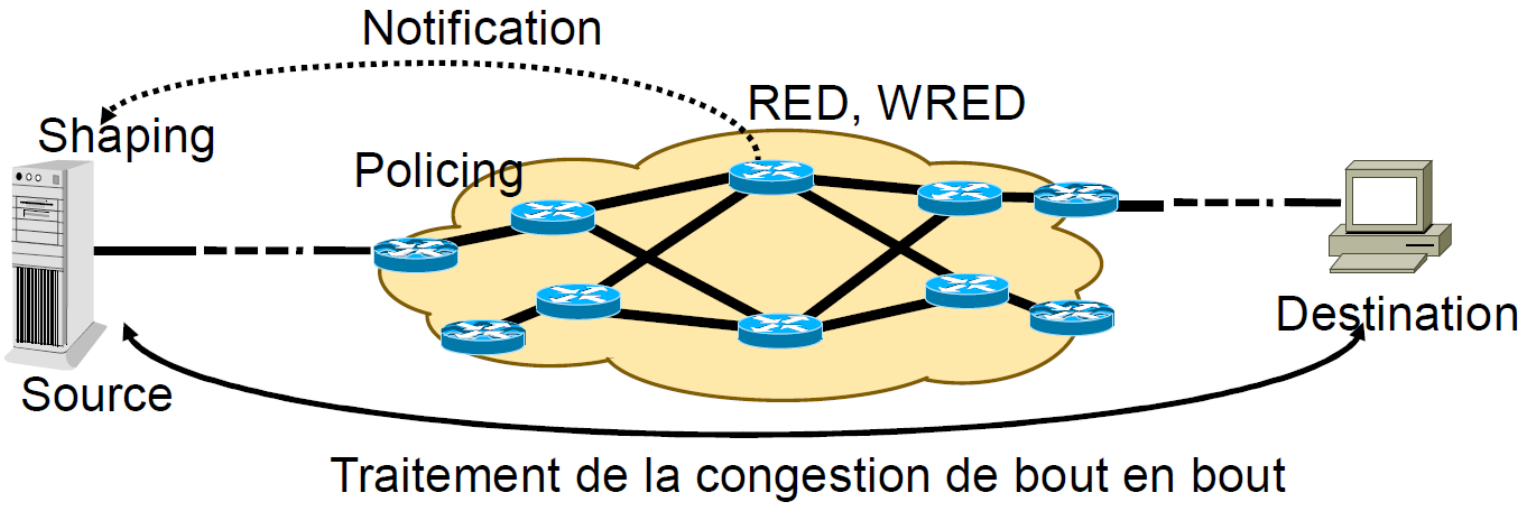
WFQ garantit une borne max du délai

- 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

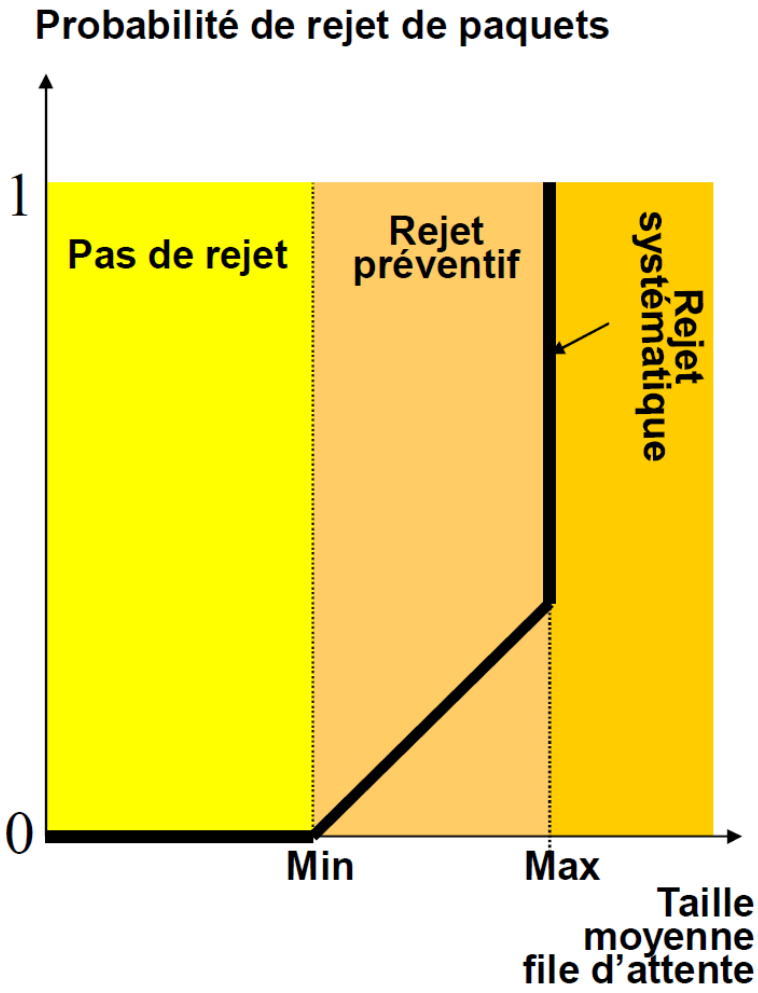
Les outils de traitement de la congestion

- **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



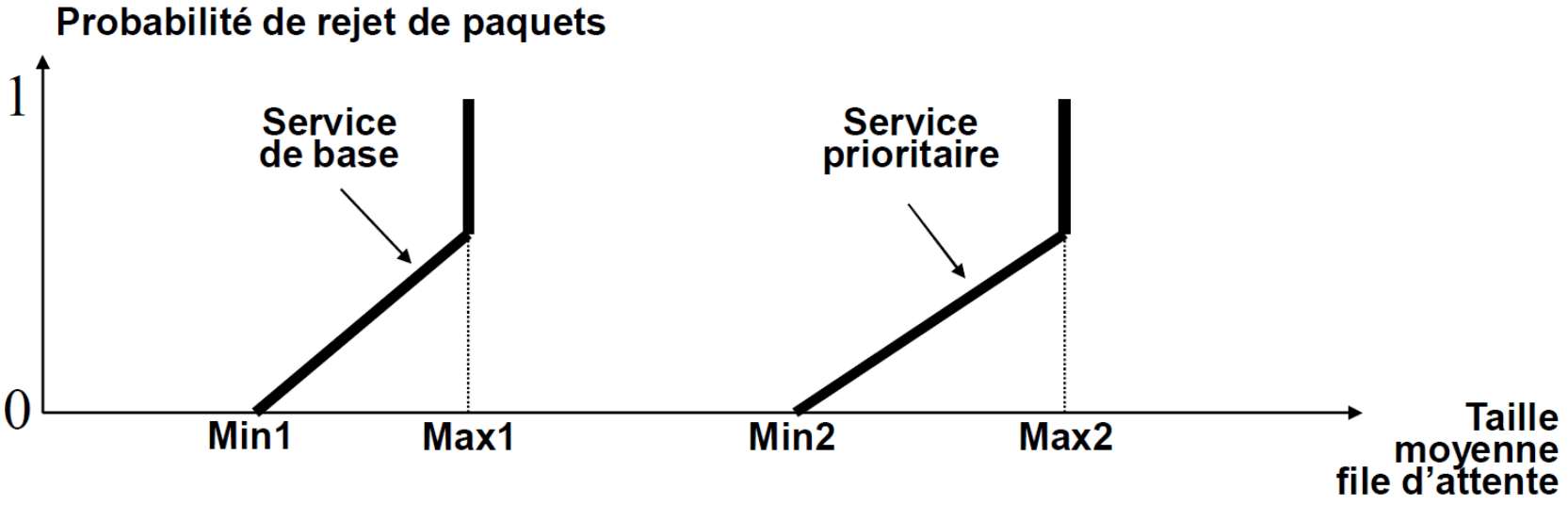
Random Early Detection (RED)

- **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



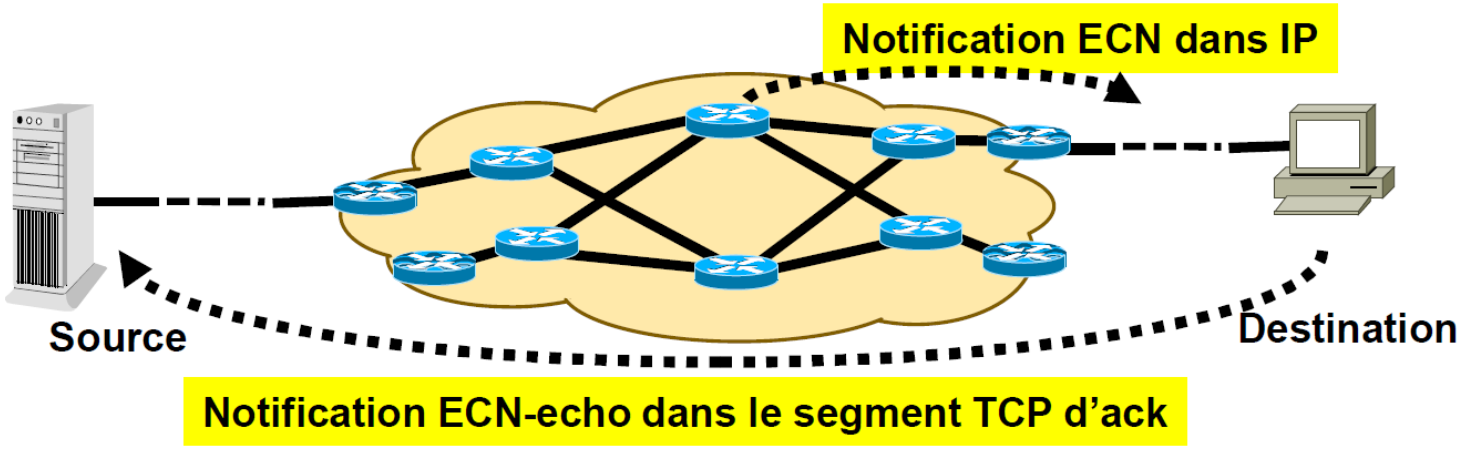
Weighted RED

- **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

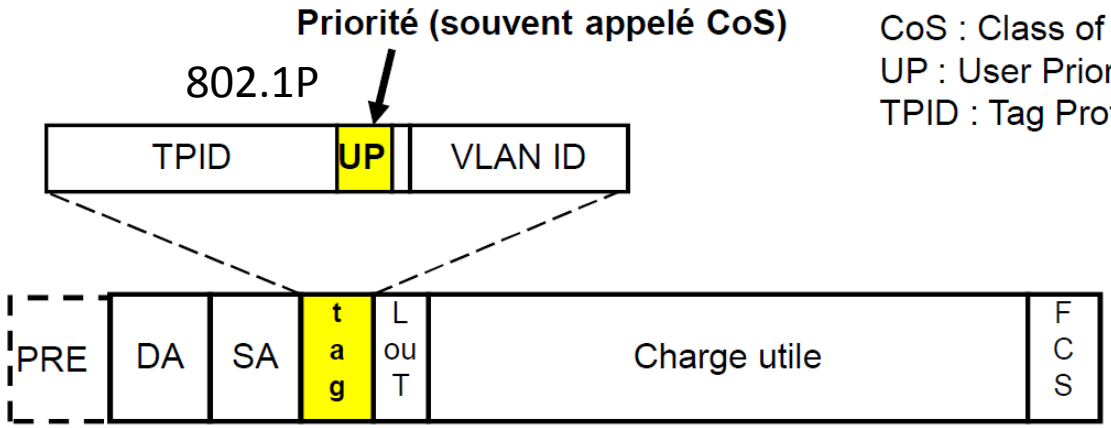
- **Utilisation conjointe de RED et ECN (Explicit Congestion Notification)**
  - RFC 2481
  - Un routeur RED positionne le bit ECN au lieu de jeter le paquet
    - Le bit ECN est recopié dans l'entête TCP d'ack (ECN-echo)
    - Un seul paquet marqué ECN provoque une réaction à la congestion
  - Un 2e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

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

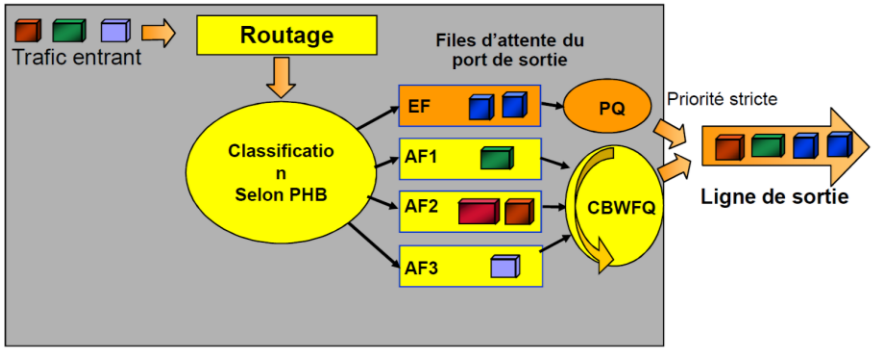
-> contraire à la nature *Best-effort* de l'Internet où il n'y a pas de contrôle d'admission

- 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,…).

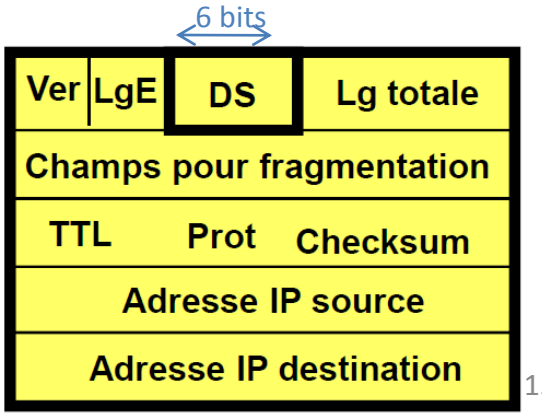
-> introduction de Diffserv : basé sur le marquage des paquets aux bords du réseau selon le niveau de performance que le SP veut leur fournir, et traitement différencié de ces paquets dans le coeur du réseau.

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

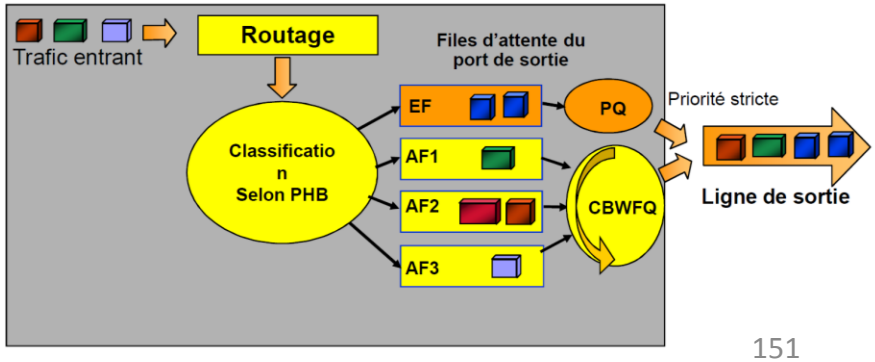


Mode de fonctionnement de DiffServ

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.

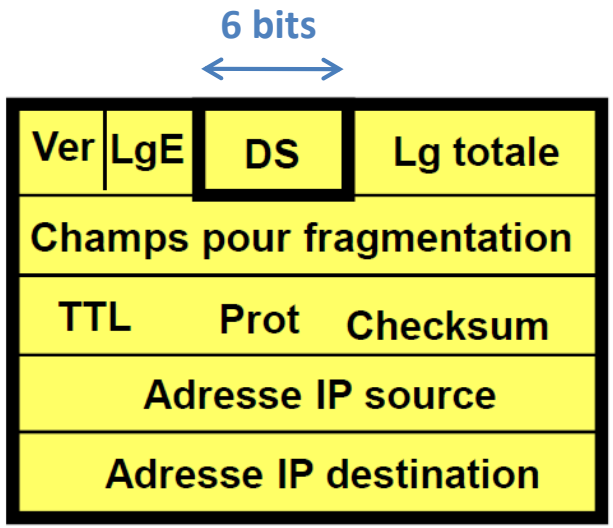
   Traitement différents: en termes de décisions d'ordonnancement et d'abandon.

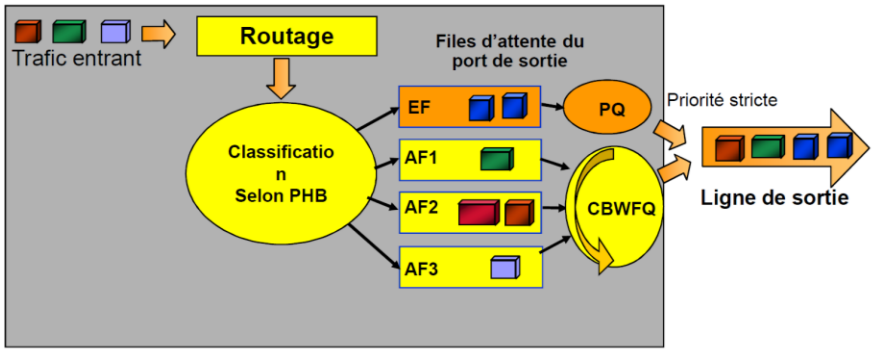


CBWFQ : Class-Based Weighted Fair Queing

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)





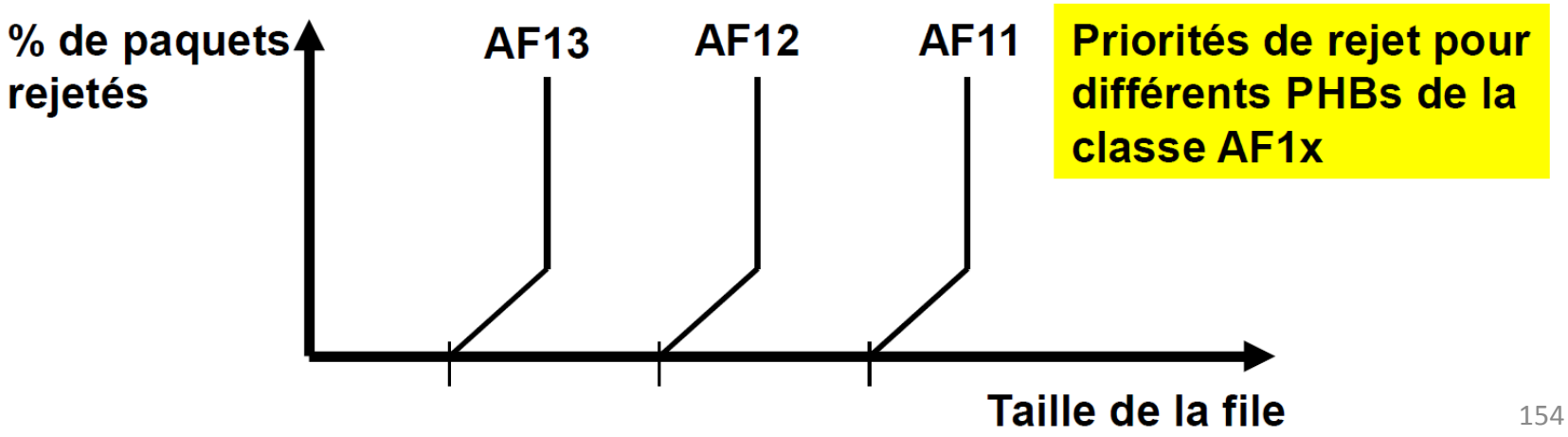
CBWFQ : Class-Based Weighted Fair Queing

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

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.

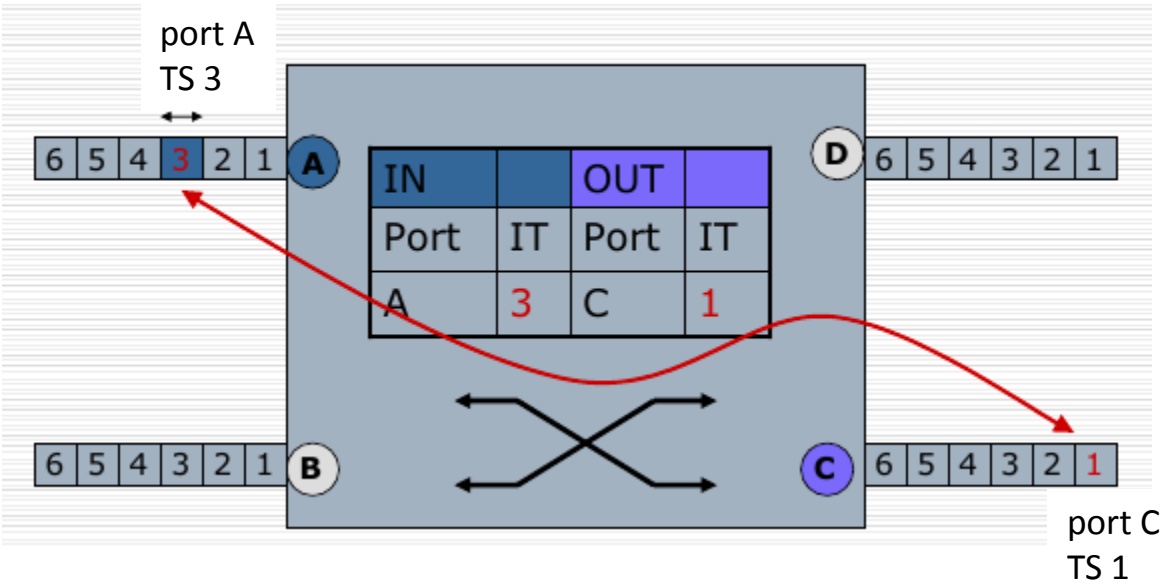
- **Commutation spatiale** : la mise en relation entre l'entrée sur laquelle le signal arrive dans l'équipement et la sortie sur laquelle il est transféré se fait spatialement.

Opérateurs humains

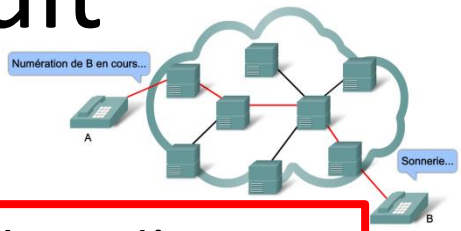
Crossbar switch

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

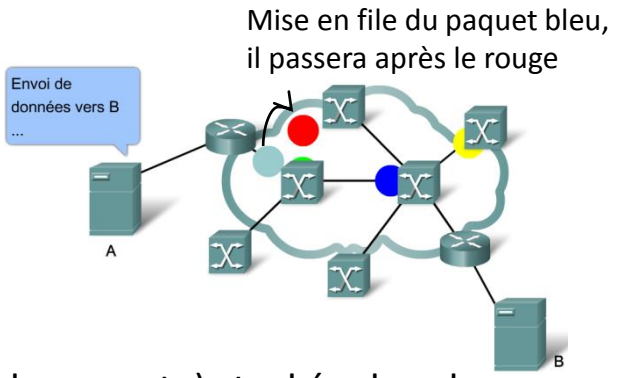
Contraintes d'une com. tél.

  - --> 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

-> Inefficacité dans le cas des rafales d'activité: cas des données informatiques

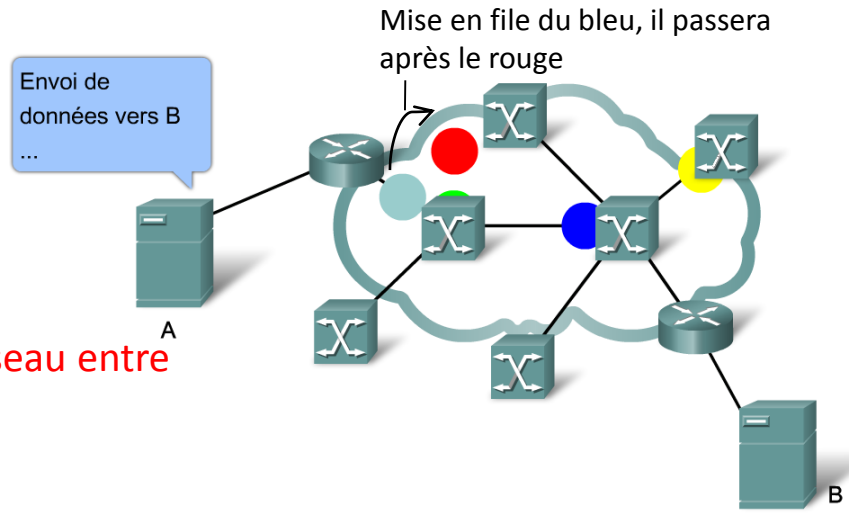
Commutation de paquet avec et sans connexion



- **Paquet** : ensemble de bits transférés ensemble, comme une unité
- **Commutation de paquet** : à partir d'une adresse (possiblement changeante) stockée dans le paquet, il est transféré par les équipements intermédiaires (switchs ou routeurs)
- **Commutation de paquet avec connexion** : la route est calculée et établie (équipts interm. configurés) avant l'envoi des paquets par la source. L'adresse dans le paquet est changée à chaque commutateur, et permet de connaître le suivant (label).
- **Commutation de paquet sans connexion** : les paquets sont envoyés par la source sans établissement préalable de route (pas de changement de configuration des équipts interm.). L'adresse dans le paquet est celle de la destination finale. Ex : le **routage** IP
- Pas de monopolisation de circuit et nombreuses paires de nœuds communiquent sur le même canal car PAS DE RESERVATION.
- Commutation de paquet avec connexion :
  - FR: DLCI, ATM: VPi/VCi, MPLS: labels.
  - La config des commutateurs successifs définit un chemin fixé, un circuit.
  - Si ce circuit n'existe physiquement, n'est occupé, que lorsqu'un paquet se déplace dessus : **circuit virtuel** (VC).

Commutation de paquet avec connexion

• Les réseaux à commutation de paquets établissent des routes via les commutateurs pour des connexions de bout en bout spécifiques. Ces routes sont appelées des circuits virtuels.



Circuit virtuel (VC): circuit logique établi au sein d'un réseau entre deux périphériques réseau.

Il existe deux types de circuits virtuels :

- Circuit virtuel permanent (PVC) : VC établi de façon permanente
  - Pour transmissions de données constantes entre les périphériques.
  - Consomment moins de bande passante lors de l'établissement et de la fermeture des circuits, mais ils augmentent les coûts en raison de leur continuité de service.
  - Généralement configurés par le fournisseur de services suite à une commande de service.

Commutation de paquet avec connexion

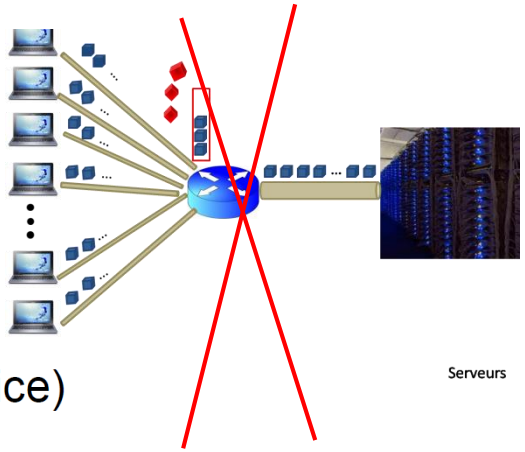
- Circuit virtuel commuté (SVC) : VC établi de façon dynamique sur demande et qui se ferme une fois la transmission terminée.
  - 3 phases pour la communication:
    1. établissement du circuit: création du VC entre les périph d'origine et de destination
    2. transfert des données: transmission des données entre les périphériques via le VC
    3. fermeture du circuit: démantèlement du VC

  - Utilisés pour des transmissions de données intermittentes entre les périph
    -> raisons financières: SVC relâchent le circuit une fois la transmission terminée
    -> réduction des frais de connexion p/r aux PVC

- Pour l'abonné: besoin d'une boucle locale jusqu'à l'emplacement le plus proche auquel le AP met le service à disposition= Point of Présence (POP) du service.

- Exemples de connexions à commutation de paquets avec connexion : X.25, Frame Relay, ATM, MPLS

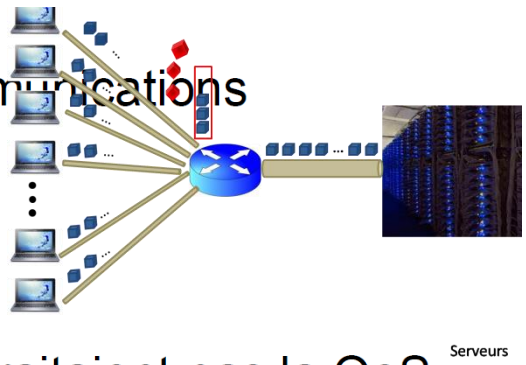
Commutation de circuits ou commutation de paquets ? Avantages et inconvénients



- **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

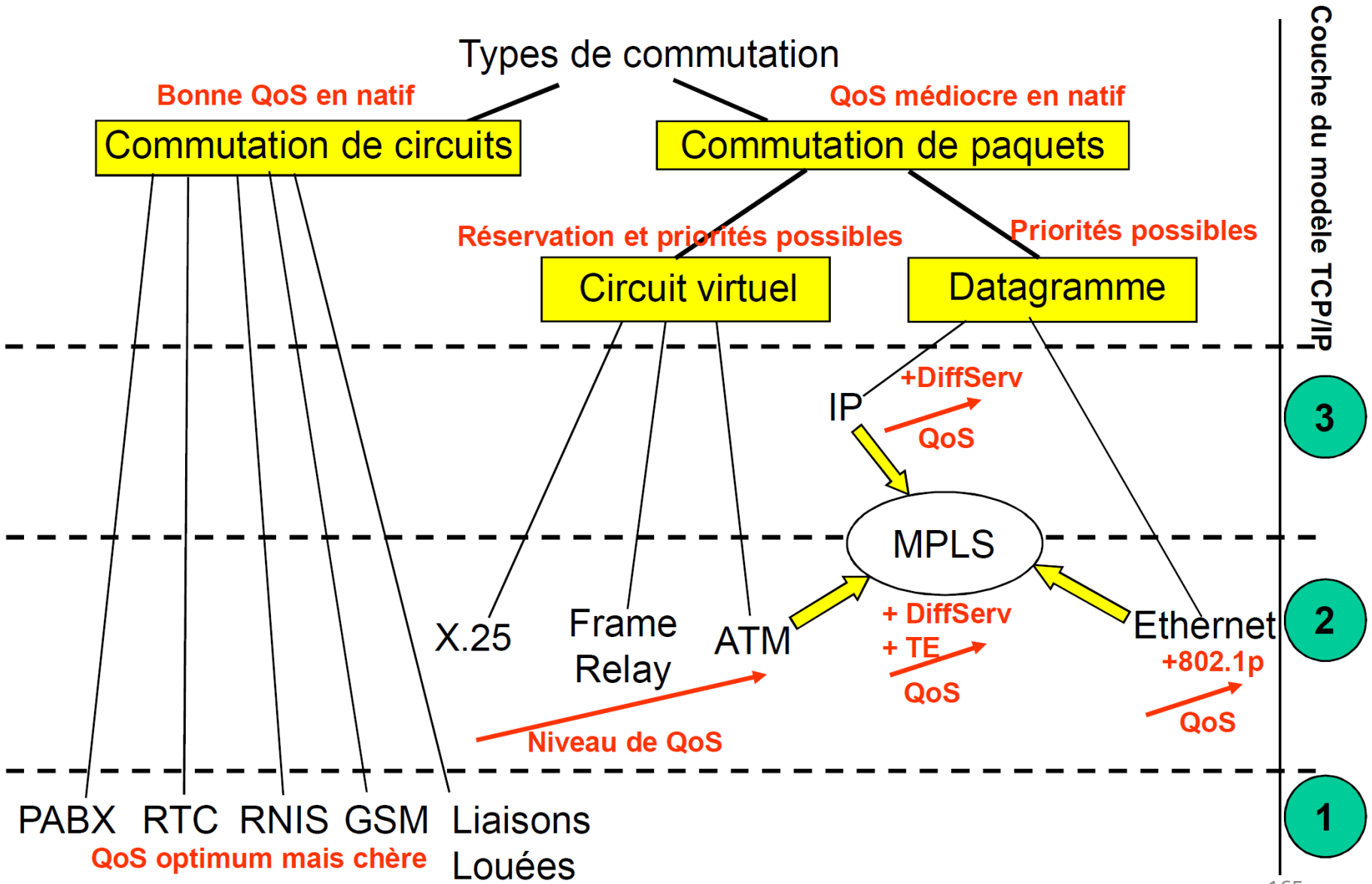
Commutation de circuits ou commutation de paquets ?

- **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

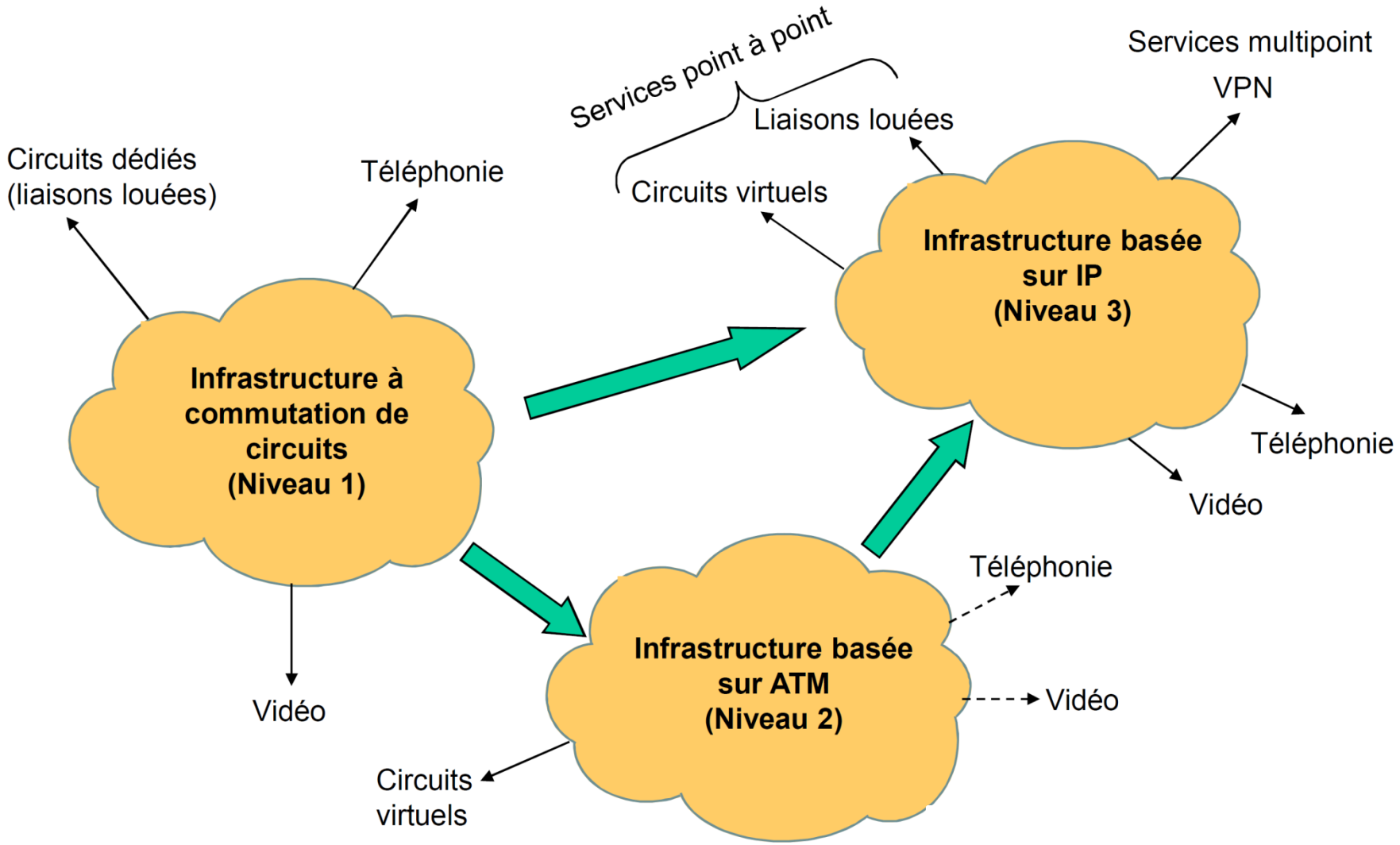
Evolution de la commutation et de la QoS



Les tendances de l'évolution des réseaux d'opérateurs

- **Le choix des opérateurs était de migrer vers un réseau basé sur la technologie ATM**
  - Réseau à commutation de paquets avec traitement de la qualité de service
  - ATM devait alors se généraliser aussi au niveau des entreprises
  - La migration a été très avancée chez les opérateurs historiques
- **Les succès de IP, Internet, et Ethernet ont changé cette orientation**
  - La migration se fait maintenant vers un réseau IP
    - En lui ajoutant des fonctionnalités de traitement de la qualité de service
- **Les trois types d'infrastructures cohabitent actuellement chez les opérateurs**

Les tendances de l'évolution des WAN

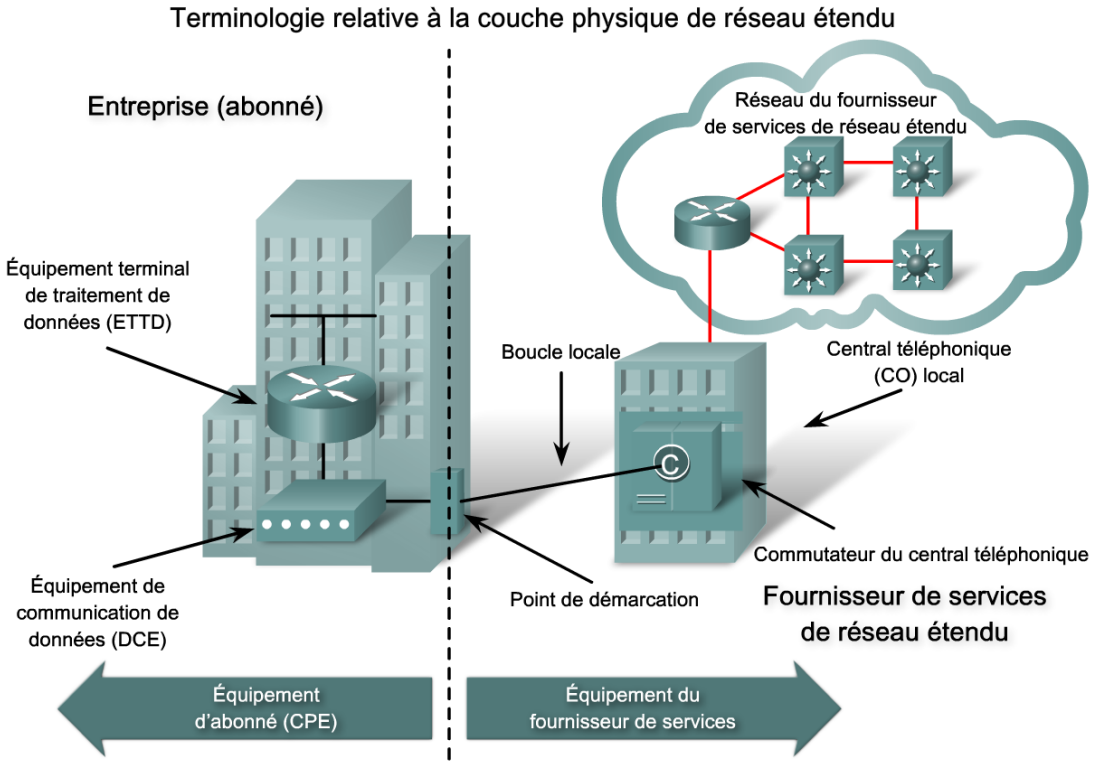


Plan général du cours

Concepts de la couche physique de réseaux étendus (1)

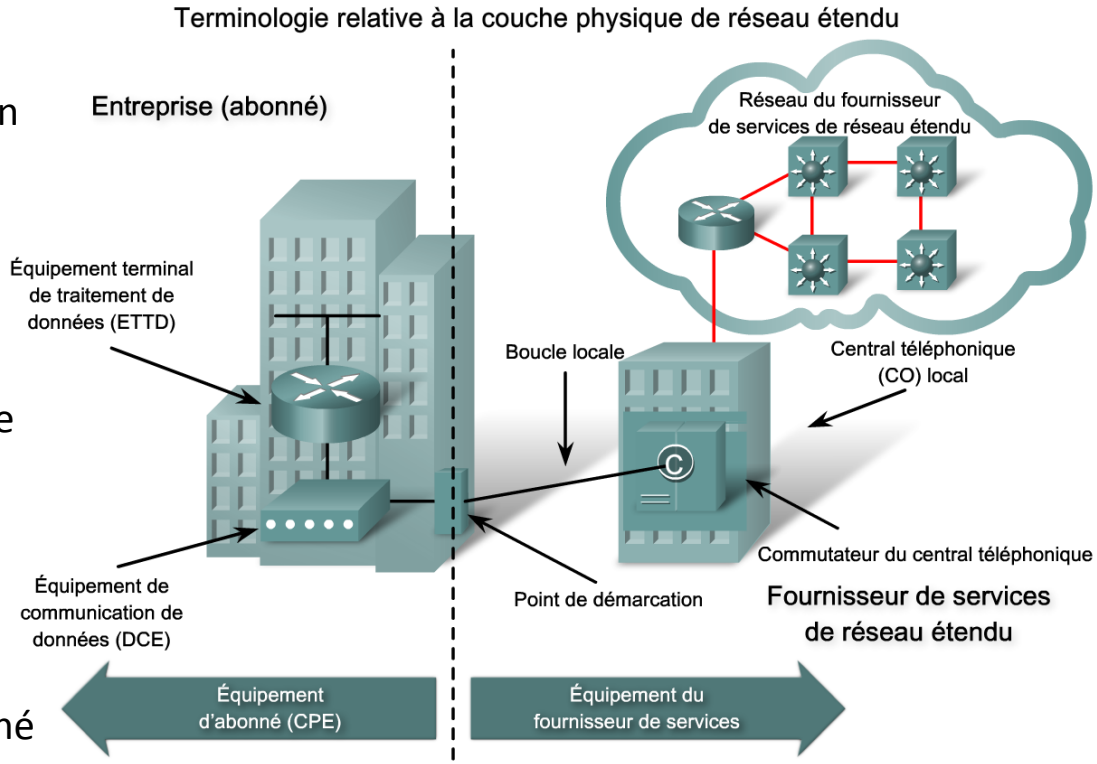
• CPE (Customer Premises Equipment) ou **CE (Customer Edge)**
Équipement de l'abonné directement relié par un port physique à son ISP

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

Concepts de la couche physique de réseaux étendus (2)



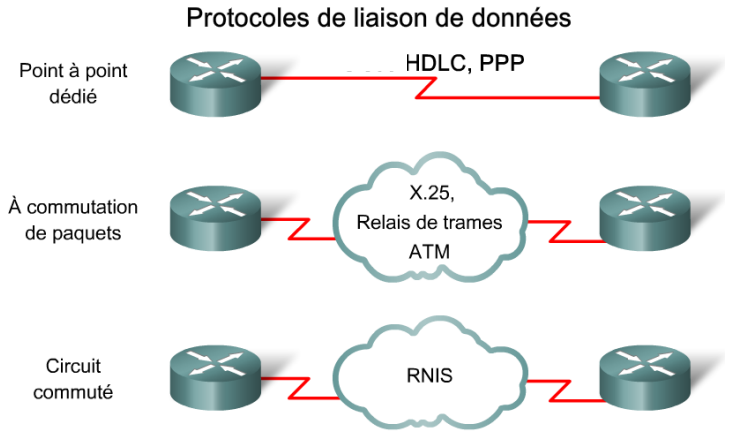
• Point de démarcation : point établi dans un bâtiment ou un complexe pour séparer l'équipement du client de celui du fournisseur de services.

➢Physiquement: boîtier de raccordement de câblage, situé chez le client, qui connecte les câbles de l'équipement d'abonné à la boucle locale.

• Boucle locale : câble téléphonique de cuivre ou à fibre optique qui connecte l'équipement d'abonné sur le site de l'abonné au central téléphonique du fournisseur de services. « Last-mile network».

• Central téléphonique (CO) : installation ou bâtiment de fournisseur de services local dans lequel des câbles téléphoniques locaux relient des lignes de communications grande distance, entièrement numériques et à fibre optique via un système de commutateurs et d'autres équipements.

Concepts de la couche liaison de réseaux étendus



• Protocoles de couche liaison de données pour établir la liaison sur la ligne de communication entre le périphérique d'envoi et de réception

• Différentes technologies: RNIS (ISDN), FR, ou ATM (Asynchronous Transfer Mode).

• Un certain nombre de ces protocoles utilisent le même mécanisme de tramage de base, HDLC, une norme ISO ou l'un de ses sous-ensembles ou variantes.

Protocole	Utilisation
LAPB (from HDLC)	RNIS canaux B, X.25
LAPD	RNIS canal D
LAPF	Frame Relay
PPP	Connexions commutées de réseau étendu série

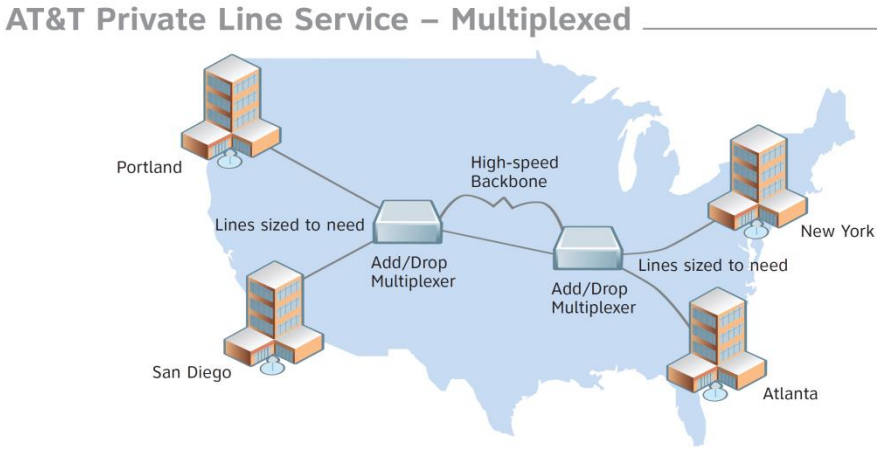
• Le mode ATM diffère des autres, car il utilise des petites cellules de taille fixe de 53 octets (48 octets pour les données), tandis que les autres technologies à commutation de paquets utilisent des paquets de taille variable.

• RNIS et X.25 plus anciens. Par ailleurs, X.25 est encore utilisé dans des PVD pour transactions par carte de crédit depuis des détaillants.

• MPLS

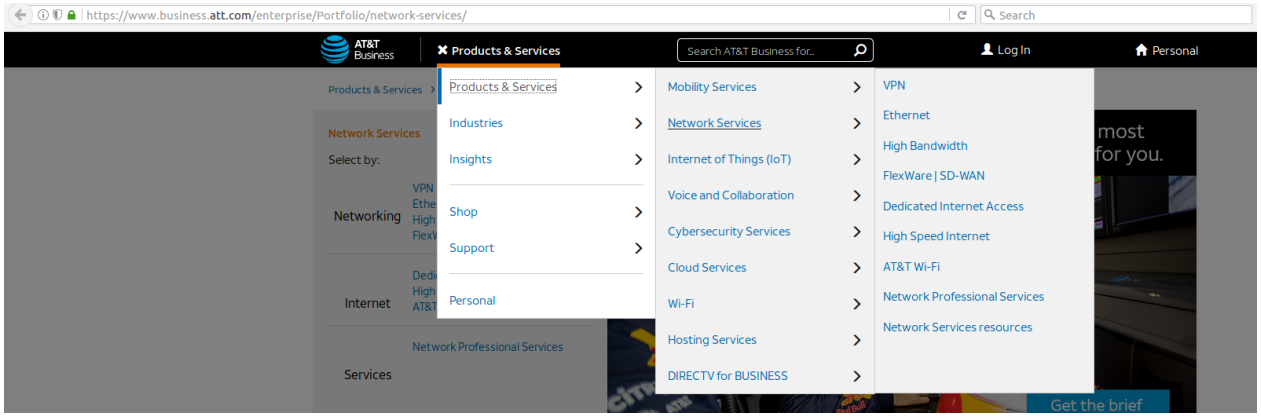
Le problème

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



Identifier sa solution réseau idéale

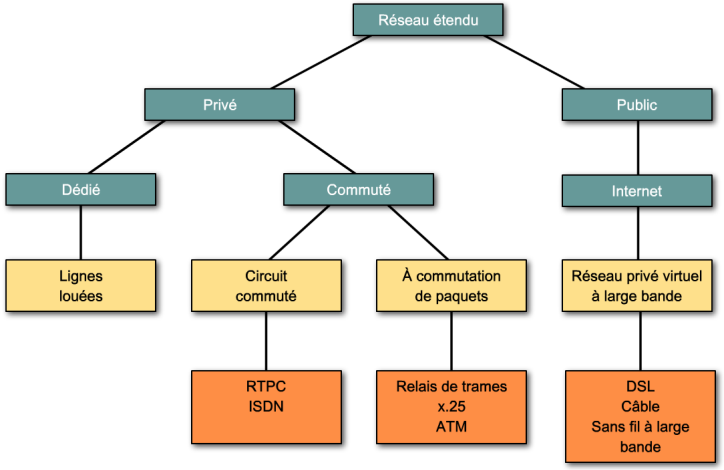
- Solutions technologiques pour les contraintes de l'entreprise



  - **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**.

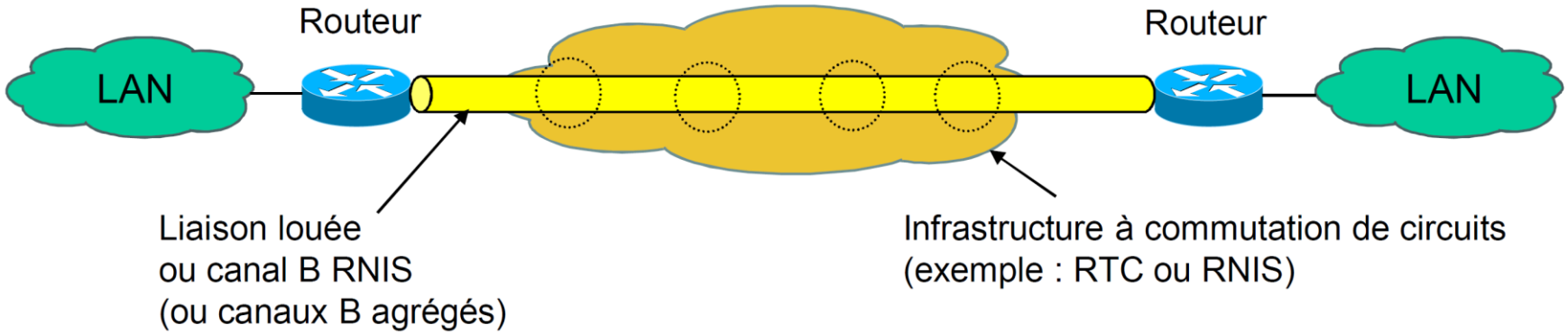
➢ **Liaison commutée :**

- 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

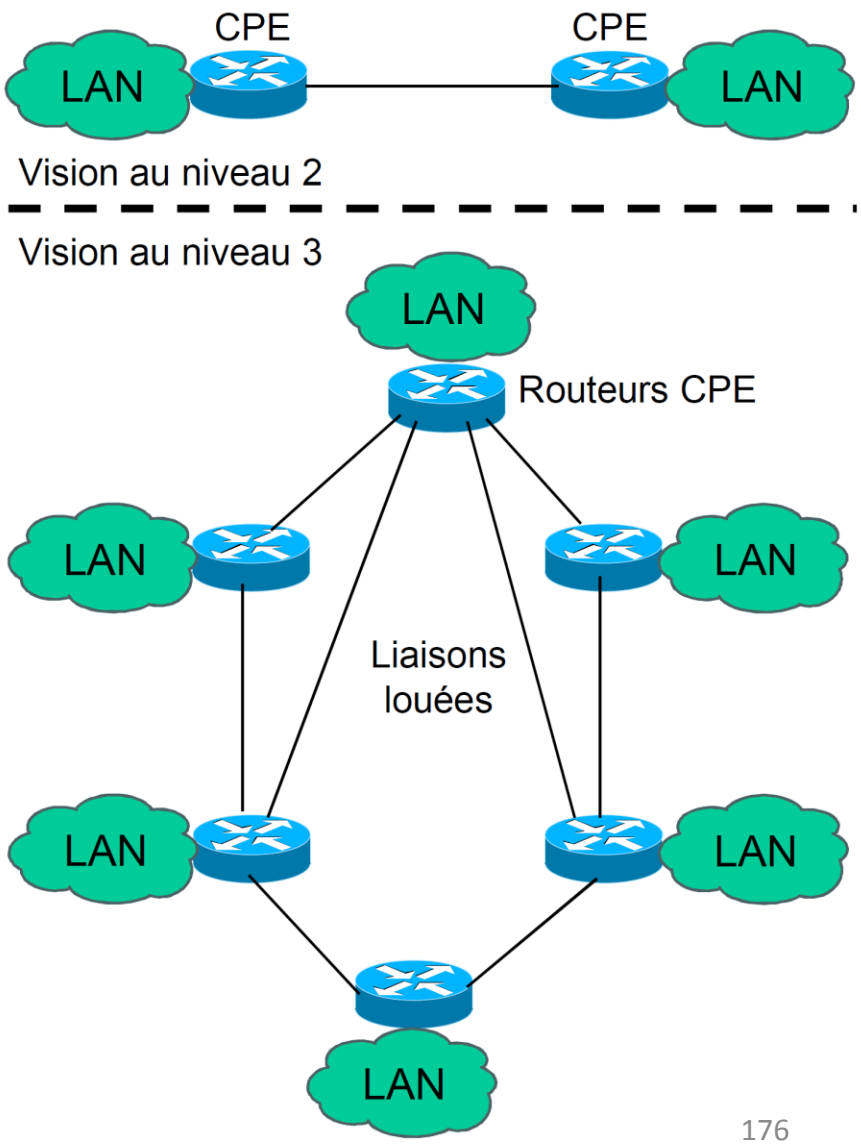
Le client voit le réseau de l'opérateur comme un câble le reliant à l'autre LAN.



- **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**

Services WAN de niveau 1

- **Au niveau 2 les routeurs CPE se voient directement entre eux**
  - Par le protocole PPP par exemple
- **Au niveau 3 les routeurs CPE se voient directement entre eux**
  - Une liaison louée (ou un circuit) est vue comme un sous-réseau IP
    - Avec 2 adresses IP privées
  - Les routeurs CPE choisissent la route en fonction de leurs tables de routage
    - Tables statiques
    - Ou tables dynamiques alimentées par un protocole de routage (RIP, OSPF, …)
- **Les équipements du réseau de l'opérateur (POP et commutateurs internes) ne sont pas concernés par les protocoles de niveau 2 et 3**



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

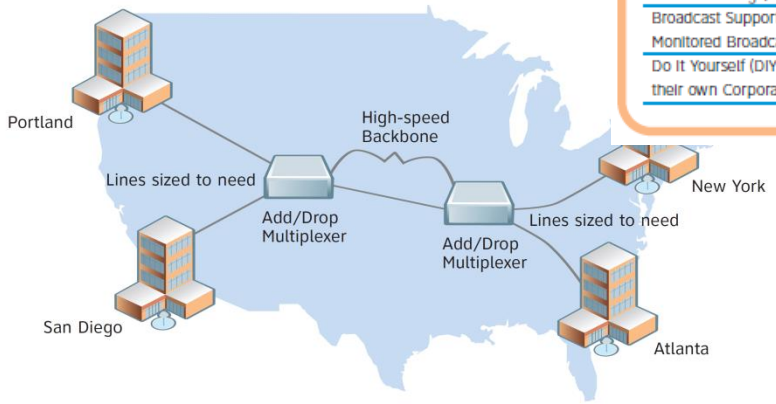
Private/Leased Lines

FEATURES

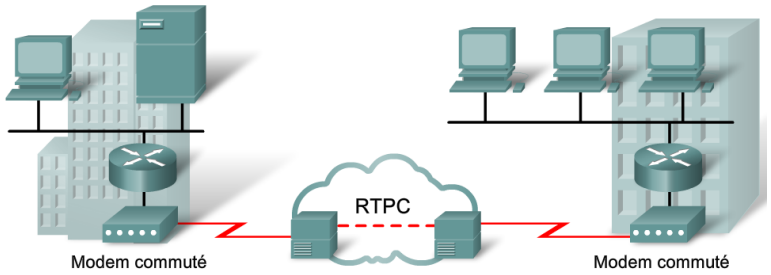
- Configures private lines over physically diverse POP to POP routes if required
- Uses self-healing SONET Network to route customer traffic
- SLAs that provide credits in the rare event of a service interruption
- Premium Services with Enhanced SLAs
- Multiple types of configurations – Point to Point, 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



Options de connexion à commutation de circuits: accès commuté analogique



Historique :

- Transferts de données intermittents de faibles volumes:
  modems et lignes téléphoniques commutées analogiques
  pour connexions commutées de faible capacité et dédiées

- Téléphonie traditionnelle: câble de cuivre, appelé boucle locale,
  pour connecter le combiné téléphonique situé dans les locaux de l'abonné au central téléphonique (CO).

- Boucles locales traditionnelles: transport des données informatiques binaires par le réseau téléphonique vocal au moyen d'un modem.

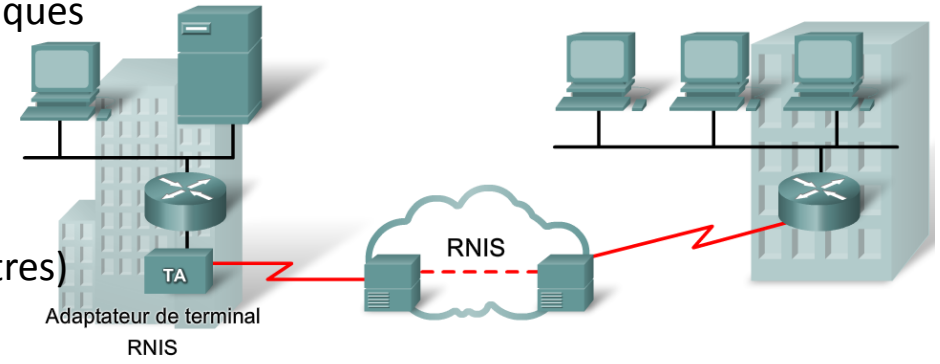
- Les caractéristiques physiques de la boucle locale et sa connexion au RTCP limitent le débit de ce signal à moins de 56 Kbits/s.

  Petites entreprises: pour l'échange de chiffres de vente, de prix, et de courriel.
  Numérotation automatique la nuit ou le week-end -> tarifs heures creuses plus intéressants.
  Tarifs fondés sur la distance entre les points d'extrémité, l'heure du jour et la durée de l'appel.

- Avantages du modem et des lignes analogiques: simplicité, disponibilité et faible coût d'implémentation.
- Inconvénients: faibles débits et temps de connexion relativement long.

Options de connexion à commutation de circuits: RNIS (ISDN)

- Boucle locale d'un RNIS: transport de signaux numériques
-> connexions commutées de plus haute capacité

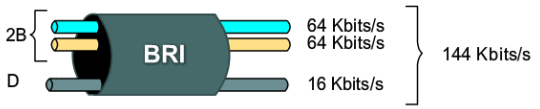


- **canaux Bearer à 64 Kbits/s (B)**: voix ou données
- **canal delta de signalisation (D)**: configuration (et autres)

- 2 types d'interface RNIS:

➢ **Basic Rate Interface (BRI)** :

2 canaux B à 64 Kbits/s + 1 canal D à 16 Kbits/s
Pour utilisateurs individuels et petites entreprises
Canal D (contrôle) souvent sous-utilisé car seulement 2 canaux B à gérer.



-> Certains fournisseurs laissent le canal D transporter des données à bas débit (X.25 à 9,6 Kbps)

➢ **Primary Rate Interface (PRI)** :

USA: 23 canaux B à 64 Kbits/s + 1 canal D à 64 Kbits/s en, débit total jusqu'à 1,544 Mbits/s
Pour des installations de plus grande taille.
Europe, Australie et autres régions: 30 canaux B +1 canal D, débit total jusqu'à 2,048 Mbits/s
Surcharge pour la synchronisation incluse
USA: connexion T1.
Le débit du PRI international correspond à une connexion E1 ou J1.

Options de connexion à commutation de circuits: RNIS (ISDN)

- RNIS BRI pour les petits réseaux étendus: durée d'établissement de la comm du BRI <1s

- Capacité du canal B de 64 Kbits/s > capa de liaison par modem analogique

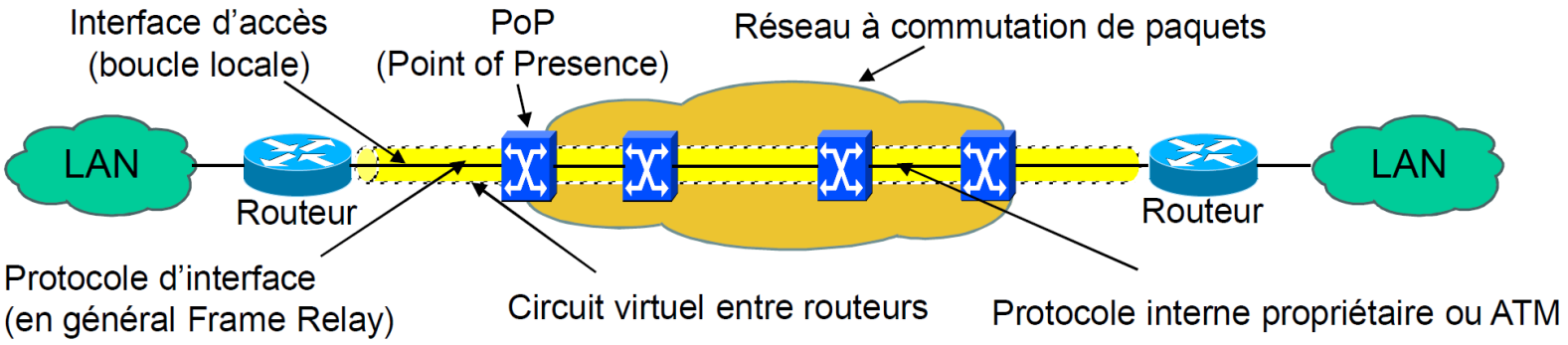
- Si besoin de plus -> activation d'un 2ème canal B -> total de 128 Kbits/s

- Autre application courante du RNIS: fournir des capacités suppl en fct des besoins à une connexion par ligne louée.
- Les tarifs RNIS sont calculés par canal B et sont similaires à ceux des connexions analogiques vocales.

- RNIS PRI: plusieurs canaux B peuvent être connectés entre deux points d'extrémité
  -> vidéoconférences et connexions de données à haut débit sans latence ni gigue

Services WAN de niveau 2 avec VC

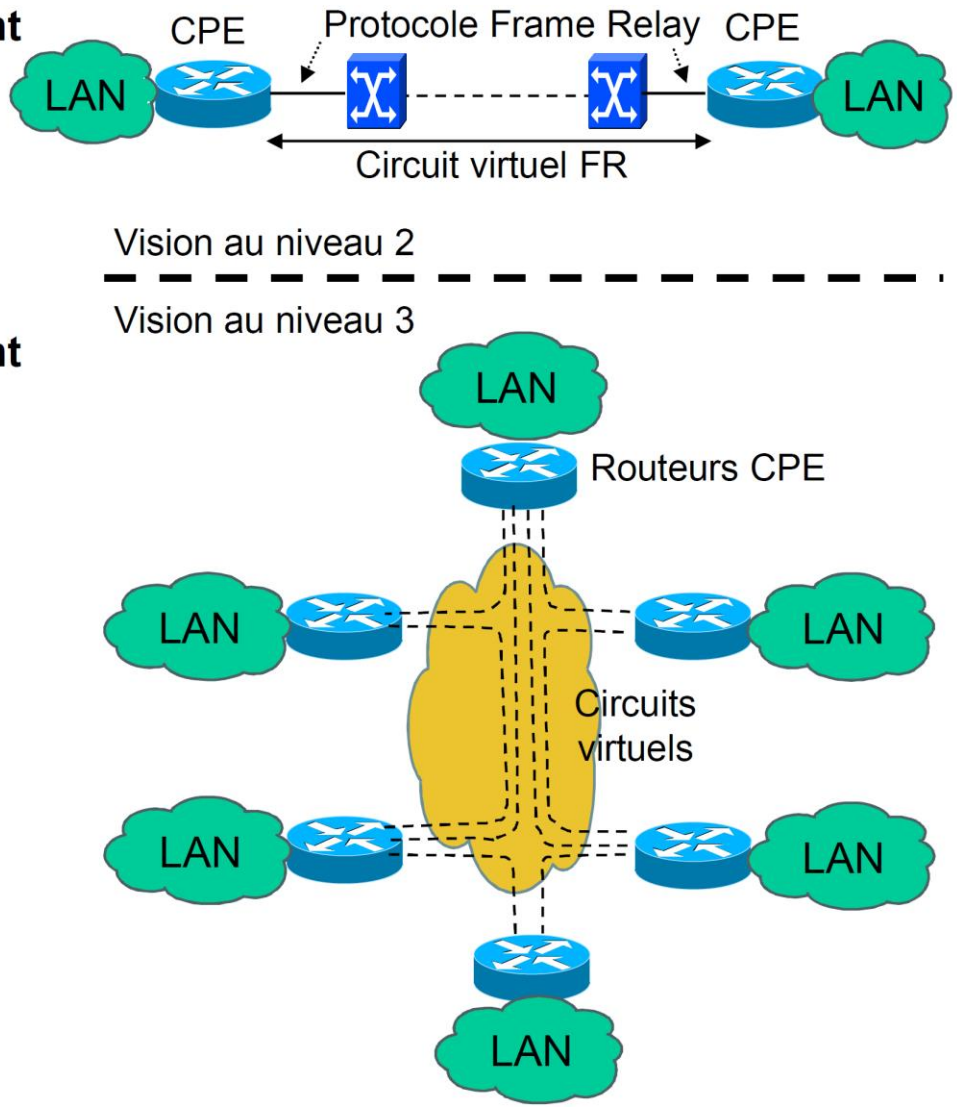
Le client voit le réseau de l'opérateur comme un switch le reliant à l'autre LAN.



- **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

Services WAN de niveau 2 avec VC

- **Au niveau 2 les routeurs CPE se voient indirectement entre eux**
  - Un protocole d'interface (en général Frame Relay) permet d'établir et d'utiliser un circuit virtuel entre routeurs CPE
- **Au niveau 3 les routeurs CPE se voient directement entre eux**
  - Un circuit virtuel est vu comme un sous-réseau IP
    - Avec 2 adresses IP privées
  - Les routeurs CPE choisissent la route en fonction de leurs tables de routage
    - Tables statiques, ou tables dynamiques alimentées par un protocole de routage
- **Les équipements du réseau de l'opérateur (POP) sont concernés par le protocole de niveau 2 mais pas par le protocole de niveau 3**



Options de connexion à commutation de paquets: Frame Relay

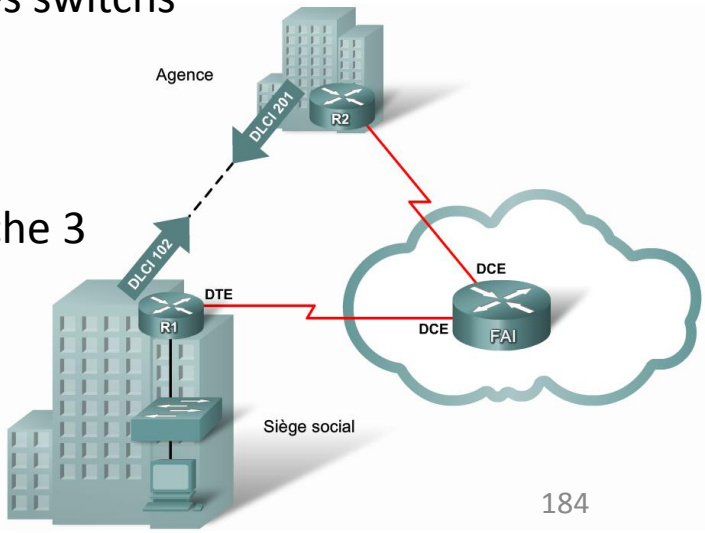
- **VC de FR identifiés de manière unique par un DLCI**, qui garantit une comm bidirectionnelle
- La plupart des connexions de relais de trames sont des PVC et non des SVC.

- Débit partagé moyen pouvant transporter du trafic vocal et de données.
- FR idéal pour connecter les réseaux locaux d'entreprise.
- **Le routeur du réseau local ne nécessite qu'une interface**, même avec plusieurs circuits virtuels.
- **La courte ligne louée vers la périphérie du réseau** de relais de trames autorise des connexions économiques entre des réseaux locaux largement dispersés.
- Mesures prises pour éviter l'accumulation des trames sur les switchs intermédiaires -> phénomènes de gigue réduits

- **Par rapport à X.25**:
  - protocole plus simple, couche 2 au lieu de la couche 3
  - aucun contrôle d'erreur ou de flux

-> réduction de la latence

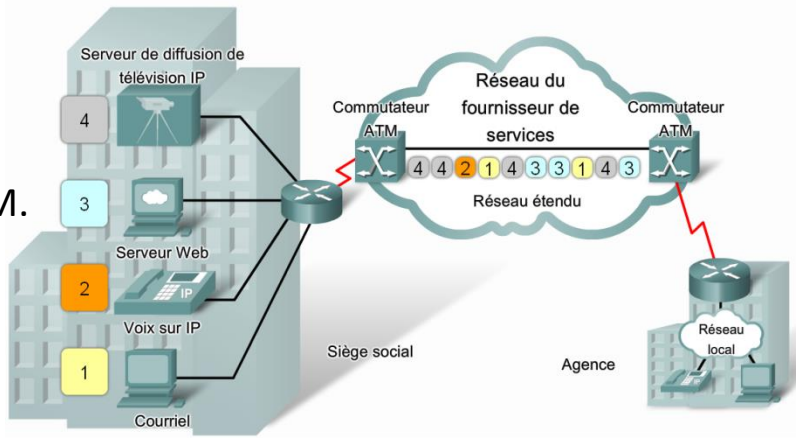
- Débits jusqu'à 4 Mbits/s, certains SP proposant plus



Options de connexion à commutation de paquets: Asynchronous Transfer Mode

- Architecture à cellules, et non à trames.

- **Cellules ATM: toujours longueur fixe =53 octets**, contient:
un en-tête ATM de 5 octets + 48 octets de données utiles ATM.
-> trafic vocal et vidéo, car ce trafic ne tolère pas les délais.
- En effet, le trafic vidéo et vocal n'a pas à attendre la fin de transmission d'un paquet de données de plus grande taille.



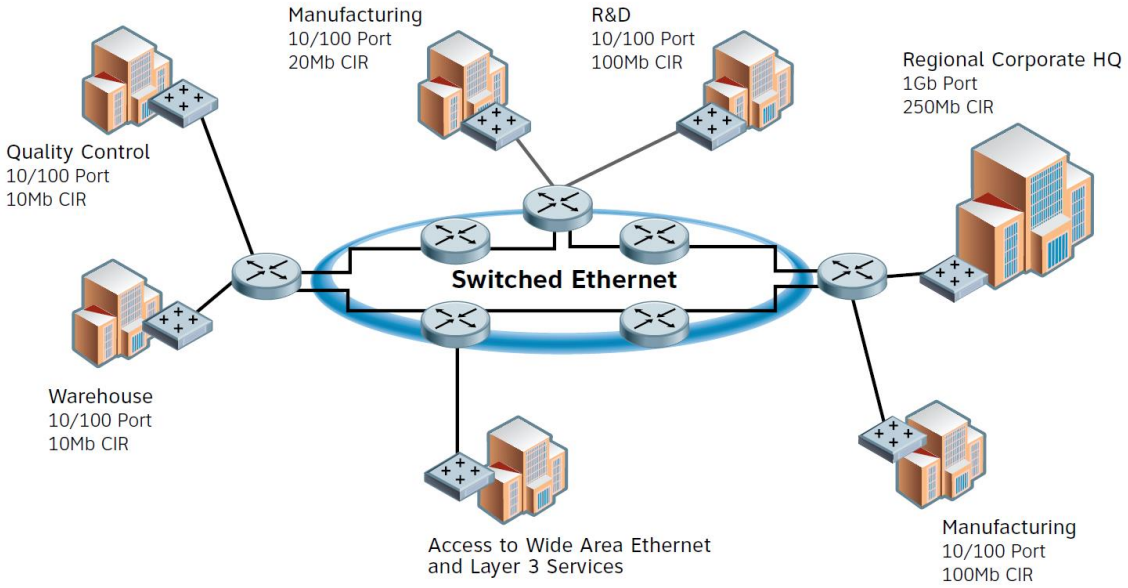
- But: QoS et performance (traitement hardware)

- Cellule ATM de petite taille fixe moins efficace que les trames et paquets de plus grande taille de FR et X.25
  5 octets de surcharge pour chaque ensemble de données utiles de 48 octets
  -> surcharge plus importante (perte de rendement) quand la cellule transporte des paquets de couche réseau segmentés
  + le commutateur ATM doit regrouper les paquets au niveau de la destination.
  -> Une ligne ATM typique nécessite un débit presque 20 % > FR pour transporter au même débit *utile* des données de couche réseau.

- PVC plus courants avec les réseaux étendus.
- Comme les autres technologies partagées, ATM accepte plusieurs VC sur une seule connexion par ligne louée vers la périphérie du réseau.

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

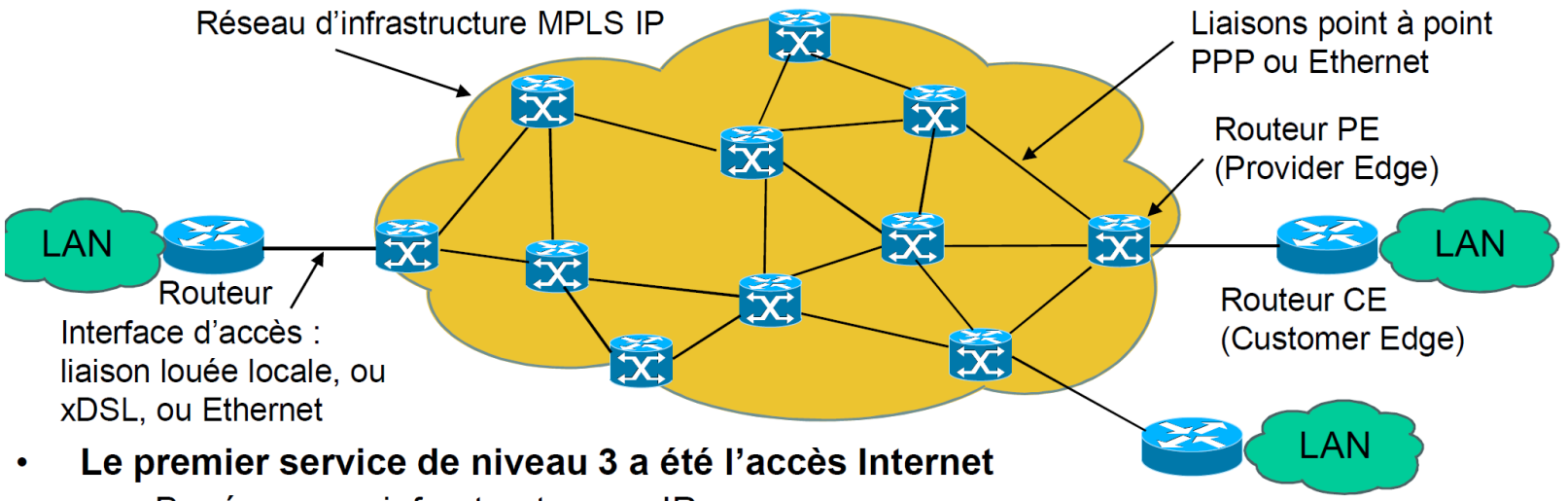
Ethernet switché peut étendre le LAN tout en connectant des sites distants à l'échelle de la ville ou plus.



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

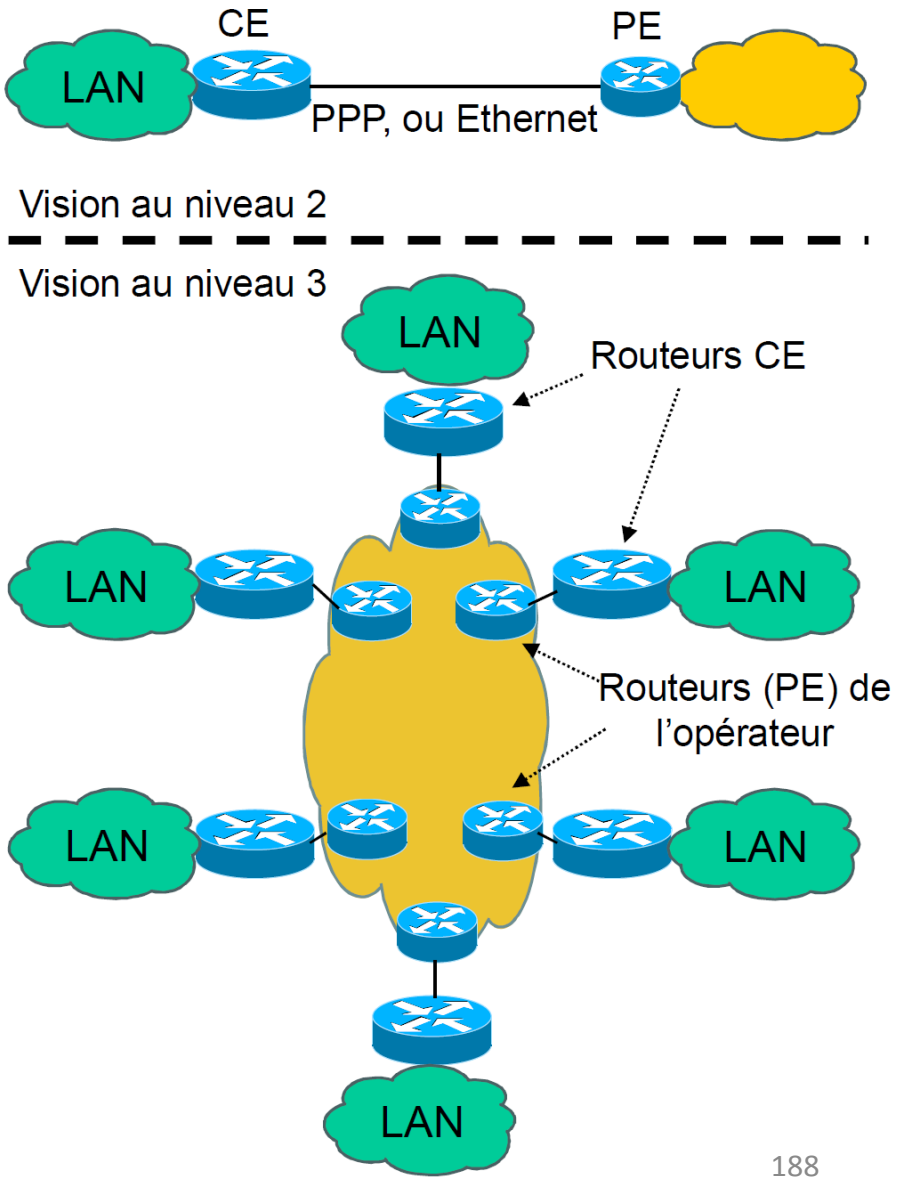
Le client voit le réseau de l'opérateur comme un routeur le reliant à l'autre LAN.



- **Le premier service de niveau 3 a été l'accès Internet**
  - Basé sur une infrastructure en IP pur
- **Puis les opérateurs ont fourni des services VPN IP multipoint**
  - L'opérateur se charge de relier en IP les routeurs CE de l'entreprise
  - L'opérateur se charge de l'étanchéité des trafics entre ses clients
    - Grâce à la technologie MPLS (Multi Protocol Label Switching)
    - Qui permet une forme de circuits virtuels sur un réseau IP
- **Cette solution permettra à terme la convergence des infrastructures réseau**
  - Une infrastructure commune IP pour la téléphonie, l'interconnexion des LAN et Internet
- **Solution de plus en plus souvent proposée par les opérateurs**
  - Cette solution devient dominante avec le développement de MPLS et des solutions de QoS/IP

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)



Options de connexion Internet: Technologie de VPN (1)

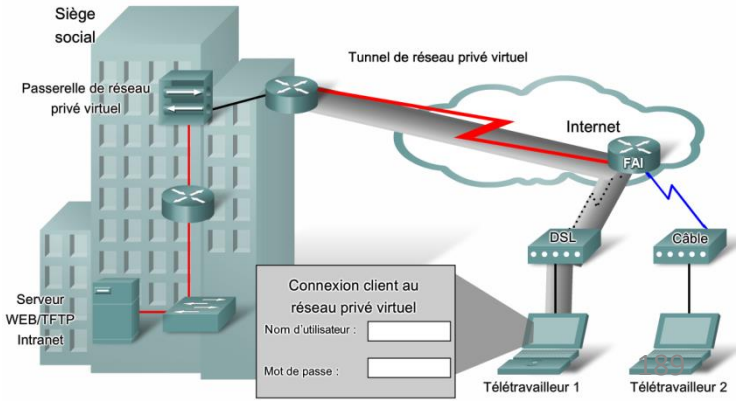
- Utilisation par un télétravailleur ou un bureau distant de services à large bande pour accéder au réseau étendu d'une entreprise -> risques en termes de sécurité.

-> Les services à large bande offrent la possibilité d'utiliser des connexions de réseau privé virtuel (VPN) vers un serveur de réseau privé virtuel (généralement situé dans les locaux de l'entreprise).

- **<u>Réseau privé virtuel</u> (VPN): connexion chiffrée entre des réseaux privés sur un réseau public (Internet).**

- Au lieu d'utiliser une connexion de couche 2 dédiée telle qu'une ligne louée, un VPN utilise des connexions virtuelles appelées tunnels de réseau privé virtuel, qui sont acheminées via Internet depuis le réseau privé de l'entreprise vers le site distant ou l'hôte de l'employé.

→ « VPN à accès à distance » ou « Premises-Based VPNs »



Options de connexion Internet: Technologie de VPN (2)

Avantages du VPN:

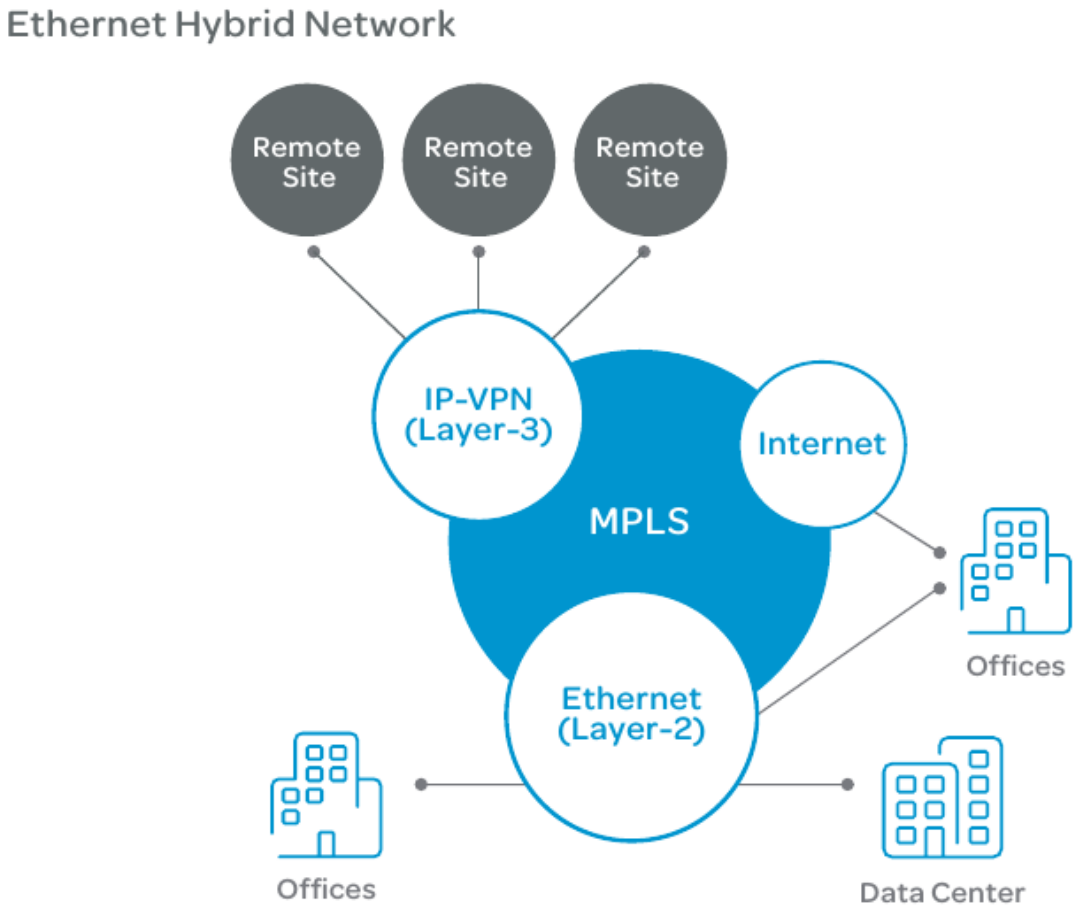
- Économique : les VPN permettent aux organisations d'utiliser le réseau Internet global pour connecter des bureaux et des utilisateurs distants aux locaux principaux d'une entreprise, éliminant ainsi le besoin de liaisons de réseau étendu dédiées et de banques de modems.

- Sécurité : tunneling et/ou protocoles avancés de chiffrement et d'authentification qui protègent les données contre des accès non autorisés.

- Extensibilité : utilisent l'infrastructure Internet au sein des périphériques et des FAI -> il est très facile d'ajouter des nouveaux utilisateurs. Les grandes entreprises peuvent ajouter des volumes importants de capacité sans ajouter d'infrastructure.

- Compatibilité avec la technologie à large bande : VPN pris en charge par AP de services à large bande tels que DSL et câble pour travailleurs mobiles.

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



Options de connexion Internet: Services à large bande (=haut débit)

- Les options de connexion à large bande sont généralement utilisées pour connecter via Internet des télétravailleurs au site de l'entreprise.

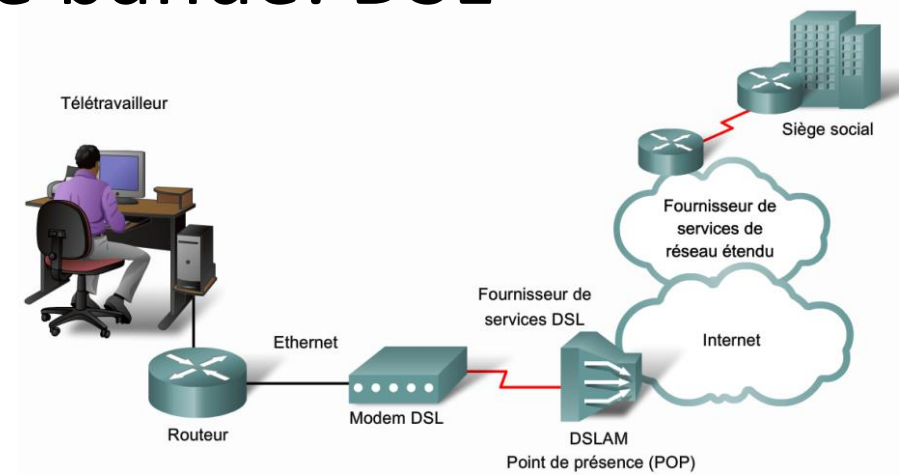
- Les options disponibles sont les suivantes :

  - DSL
  - câble
  - sans fil
  - optique

Options de connexion Internet: Services à large bande: DSL

- DSL: technologie de connexion permanente qui utilise les lignes téléphoniques à paires torsadées existantes pour transporter des données à haut débit et fournir des services IP aux abonnés.

- Modem DSL: signal Ethernet de périph d'utilisateur vers signal DSL, transmis au central téléphonique.

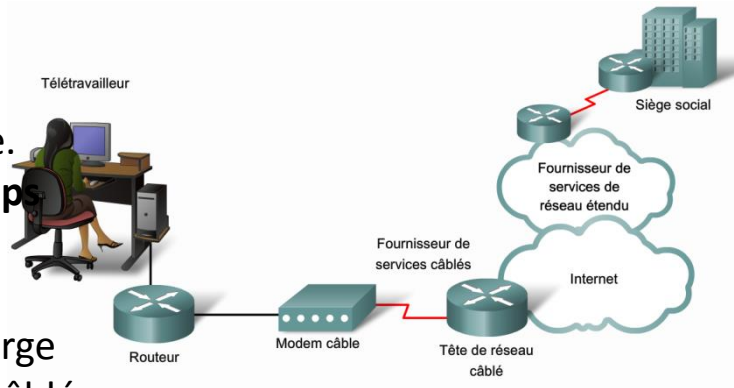


- Plusieurs lignes d'abonnés DSL multiplexées en une liaison unique à haute capacité au moyen d'un multiplexeur d'accès DSL (DSLAM) dans les locaux AP.
- DSLAM: TDM pour agréger lignes d'abonnés vers fibre ou T3/DS3.
- Codage et modulation sophistiqués -> débits pouvant atteindre 20 Mbps.

- Grande diversité de types, normes et normes émergeantes DSL.
- Choix populaire pour les services informatiques d'entreprise desservant des télétravailleurs.
- Les risques encourus en termes de sécurité lors de ce processus peuvent être tempérés par des mesures de sécurité.

Options de connexion Internet: Services à large bande: Câble

- **Câble coaxial**: répandu dans les zones urbaines pour distribuer des signaux de télévision.
- Accès réseau disponible sur certains réseaux de télévision câblée. BW > boucle locale téléphonique conventionnelle -> **30 à 100Mbps**



- Connexion permanente et facile à installer.
- Modem câble: traduit les signaux numériques en fréquences à large bande utilisées pour la transmission sur un réseau de télévision câblée.

- Le bureau local de télévision câblée, appelé **tête de réseau câblé**, comprend le système informatique et les bases de données requises pour fournir l'accès Internet.

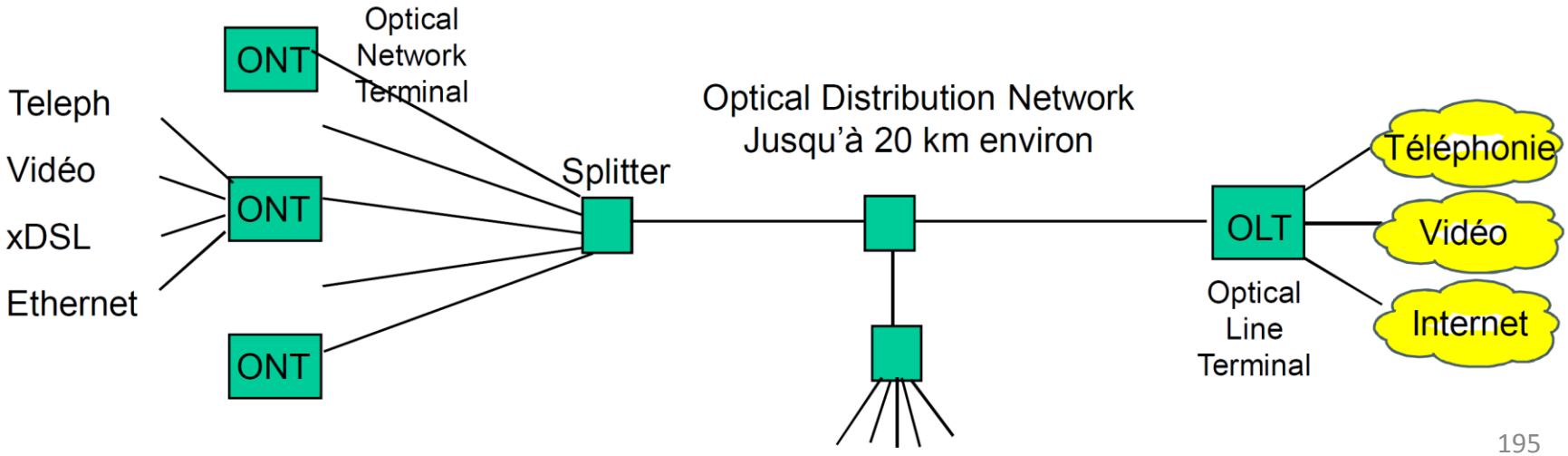
- Le composant le plus important situé au niveau de la tête de réseau est le système de terminaison du modem câble (CMTS): envoie et reçoit des signaux numériques du modem câble sur un réseau câblé et est requis pour fournir des services Internet aux abonnés du câble.

- Les abonnés au modem câble doivent utiliser le FAI associé au fournisseur d'accès TV.
- Tous les abonnés locaux partagent la même bande passante.

-> À mesure que des utilisateurs rejoignent le service, le débit dispo peut être inférieur au débit attendu.

Boucle locale fibre optique

- **Les accès fibre optique doivent succéder aux accès xDSL**
  - FTTC (FTT-Curb), FTTB (FTT-Building), FTTH (Fiber-To-The-Home)
  - C'est surtout FTTH qui se déploie en France
  - Option d'architecture point à point
    - Ethernet actif : chaque client est relié à l'équipement actif par une fibre qui lui est dédiée
    - Plus adapté aux infrastructures nouvellement créées et aux services aux entreprises
  - Option Point à Multipoint ou Passive Optical Network (PON)
    - Le signal optique est divisé en plusieurs branches (jusqu'à 64) dans un splitter
    - La bande passante est partagée

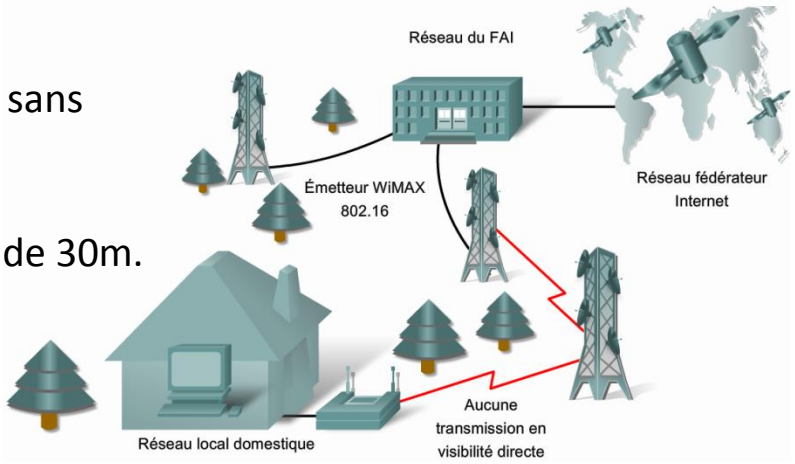


Options de connexion Internet: Services à large bande: Sans-fil (1)

- La technologie sans fil utilise le spectre des radiofréquences sans licence pour envoyer et recevoir des données.

- Jusqu'à une période récente: obligation de portée de moins de 30m.

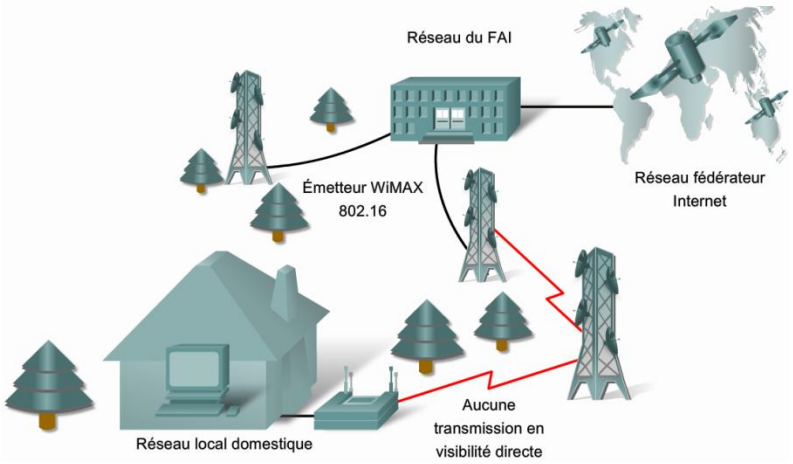
- Récentes améliorations de la techno sans fil à large bande:



- **WiFi municipal** : réseaux sans fil au niveau des municipalités. Accès Internet à haut débit, gratuit ou moins cher qu'autres services à large bande.
  Pour se connecter: modem sans fil avec antenne de radio et directionnelle plus puissante que les adaptateurs sans fil traditionnels.

- **WiMAX** : la nouvelle technologie d'interopératibilité mondiale pour l'accès micro-ondes.
  Norme IEEE 802.16. Haut débit, accès sans fil, couverture similaire au réseau de tel portable.
  Similaire à WiFi, mais vitesses, distance et nb utilisateurs plus élevés.
  Tours WiMAX similaires aux tours de téléphones portables.
  Pour accéder: s'abonner auprès d'un FAI via une tour WiMAX située dans un rayon de 15 km de leur emplacement + ordinateur compatible WiMAX + un code de chiffrement spécifique.

Options de connexion Internet: Services à large bande: Sans-fil (2)



- **Internet par satellite**:
  pour zones rurales où les connexions câblées ou DSL ne sont pas disponibles.
  Antenne parabolique pour communications de données bidirectionnelles.
  La vitesse uplink représente environ 1/10e de la vitesse downlink qui est de 500 Kbits/s.
  Les connexions câble et DSL présentent des vitesses de chargement plus élevées, mais les systèmes par satellite sont dix fois plus rapides qu'un modem analogique.
  Pour accéder: antenne parabolique, deux modems (liaison montante et liaison descendante), et des câbles coaxiaux reliant l'antenne au modem.

Autres boucles locales

- **Accès par la téléphonie cellulaire**
  - GSM avec GPRS (2,5G) : débit <56kbps
  - Edge : débit 100 à 200 Kbps
  - UMTS (3G) : débit 384 kbps
    - 3G+ HSDPA : débit 1,8Mbps puis 3,6 Mbps à terme 7,2 Mbps
    - 3G+ HSUPA : débit montant de 1,4 Mbps
  - [Mobile WiMAX, LTE] sont normes 4G : 1Gbps partagé grâce à MIMO et OFDMA
- **Accès par des hotspots 802.11 (Wi-Fi)**
  - Dans les gares, aéroports, cafés, hôtels, immeubles, etc…

Autres boucles locales

- **Boucle locale radio (BLR)**
  - Accès sans fil point à multipoint en milieu urbain
  - Débuts difficiles
  - Renaissance avec le groupement **WiMAX** : Worldwide Interoperability for Microwave Access
  - Standard 802.16 : bande de 10 à 66Mhz
    - Débit potentiels supérieurs à 100 Mbps, distance 5 km en ligne de vue
  - Amendement 802.16a : extension dans la bande de 2 à 11Ghz
    - Bande 3,5 Ghz ouverte en France pour la BLR
    - Débits moins grands (ex. 75 Mbps), distances supérieures (50km)
    - Pas forcément en ligne de vue
    - Projets en cours de déploiement (mais retards importants)
    - Possibilités de communication point à point
    - Extension vers les mobiles : 802.16e et 802.20
  - Cibles : couverture des zones blanches xDSL

- **Courants porteurs en ligne**
  - Problème de régulation du marché
  - 1ere génération limitée à 2 Mbps par abonné
  - 2e génération en cours de développement
    - Jusqu'à 25 Mbps par abonné

Plan général du cours

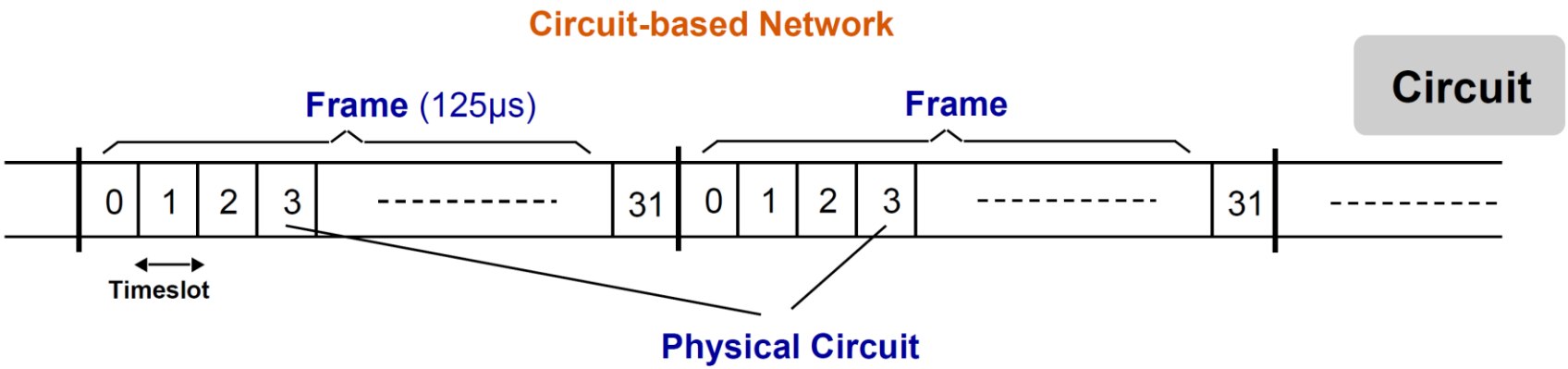
ATM: Introduction

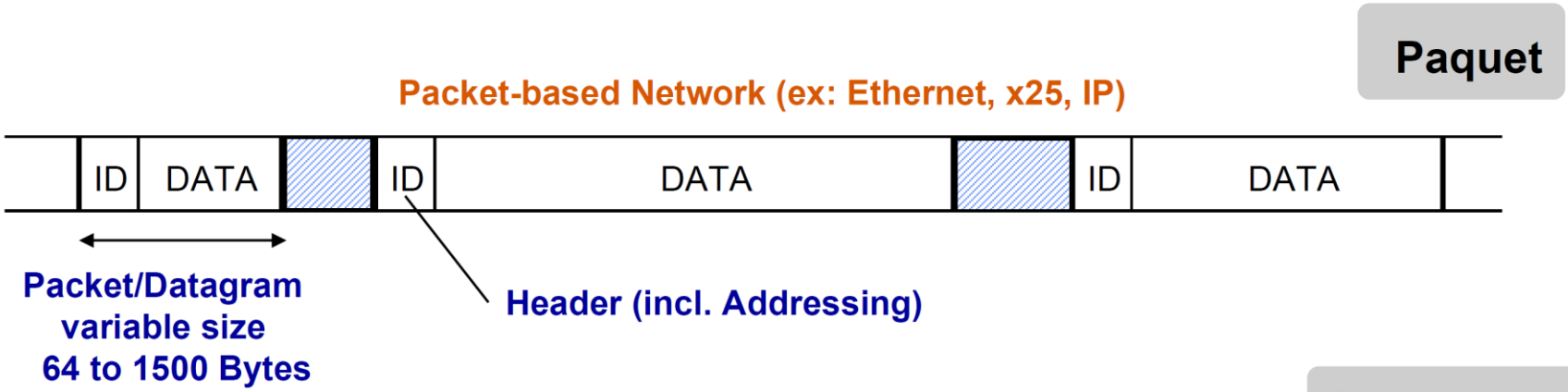
- 2 techniques différentes: routage de paquets et commutation de paquets/trames
- Routage: souple car chaque paquet transporte l'adresse du dest-> paquet pas égaré MAIS QoS difficile
- Commutation de paquets avec connexion: QoS facilement assurée MAIS pb de la mise en place du chemin (VC)
- Chemin ouvert par une procédure de *signalisation*: on signale l'ouverture d'un chemin à « marquer »

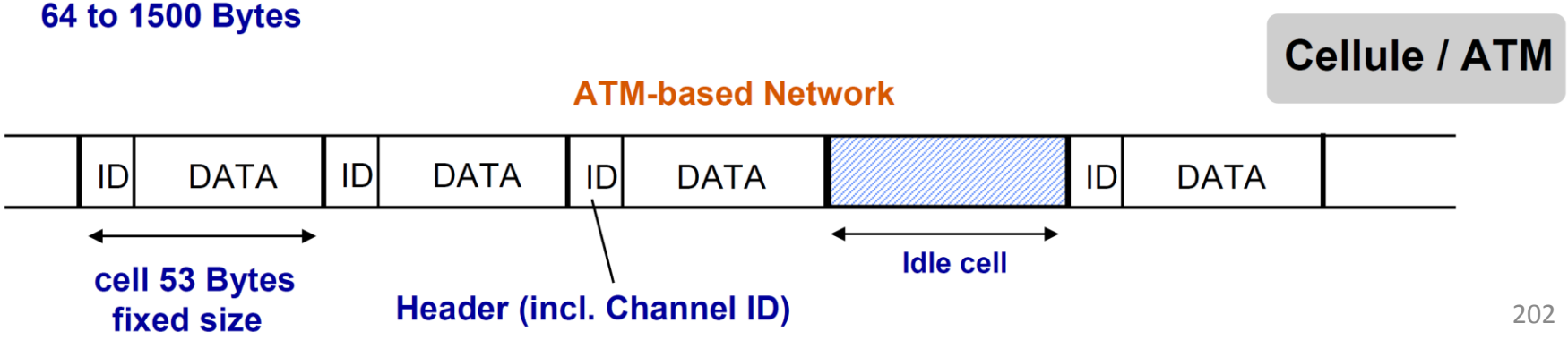
-> la signalisation exige ressources importantes -> réseaux à commutation de paquets plus chers que routage

- Ces 2 catégories se sont développées en parallèle: au départ peu de concurrence
- Avec le temps, routage -> Internet -> étendues au transport d'applis synchrones (téléphonie et video)
- Aujourd'hui, tous 2 en concurrence, puis convergence, pour applis multimédia

Circuits / Paquets / Cellules







ATM: Objectifs de départ

1 - Pouvoir transporter de manière efficace (?) tout type de flux

2 - Repondre aux exigences des services synchrones, donc maitriser débits / delais / gigue

Ex: Lien a 2Mbps, Packet 1500 bytes (12000 bits), delai=6ms

3 - Optimiser la bande passante

DONC ceci a abouti aux caractéristiques suivantes:

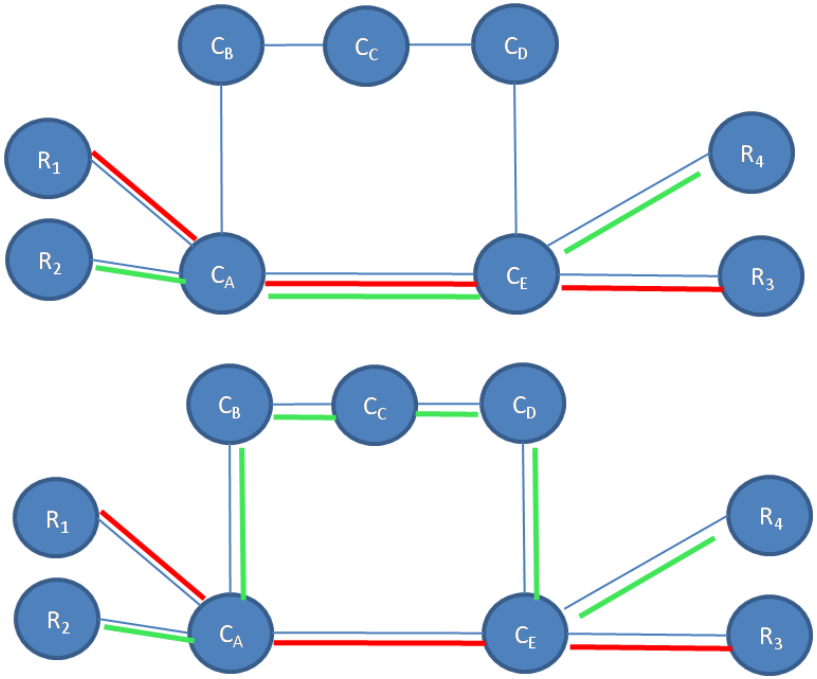
→ Réseau asynchrone (mode circuits mais circuits virtuels)

→ Cellules de taille fixe

→ Traffic Management (QoS)

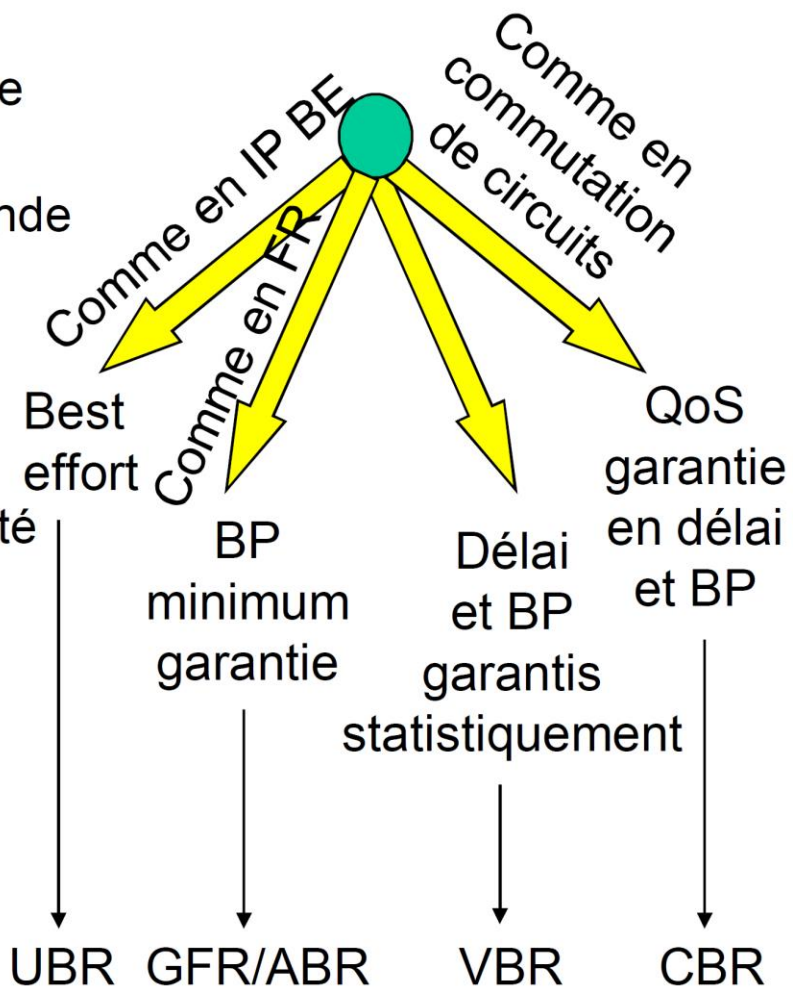
Le lien entre QoS et VC

- 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

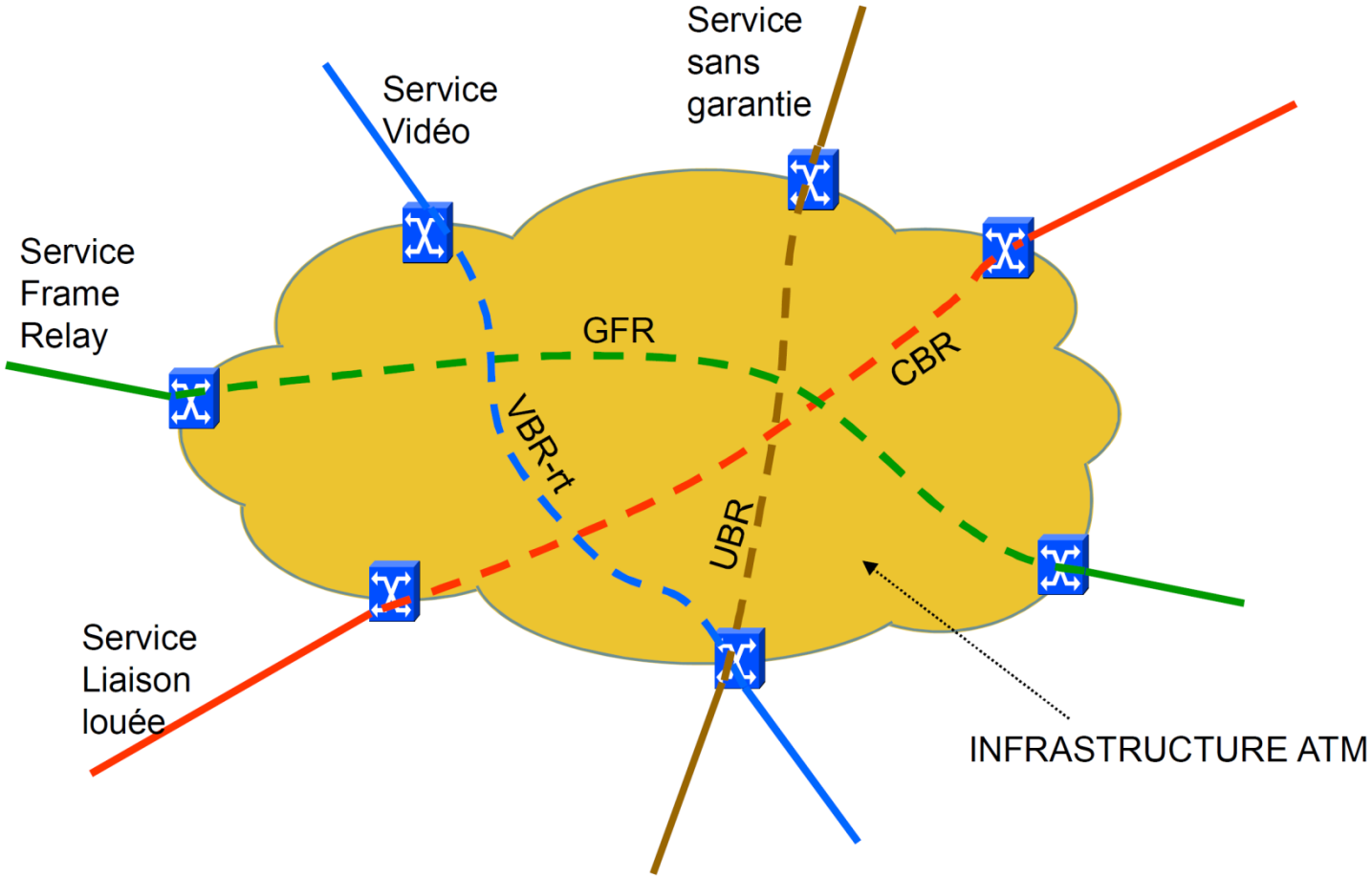
- **Chaque circuit virtuel aura la catégorie de service adaptée aux besoins de l'application qui l'utilisera**
  - Garantie absolue en délai et en bande passante : CBR
  - Garantie statistique en délai et en bande passante : VBR
  - Besoin d 'un minimum de bande passante (applications élastiques) : GFR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



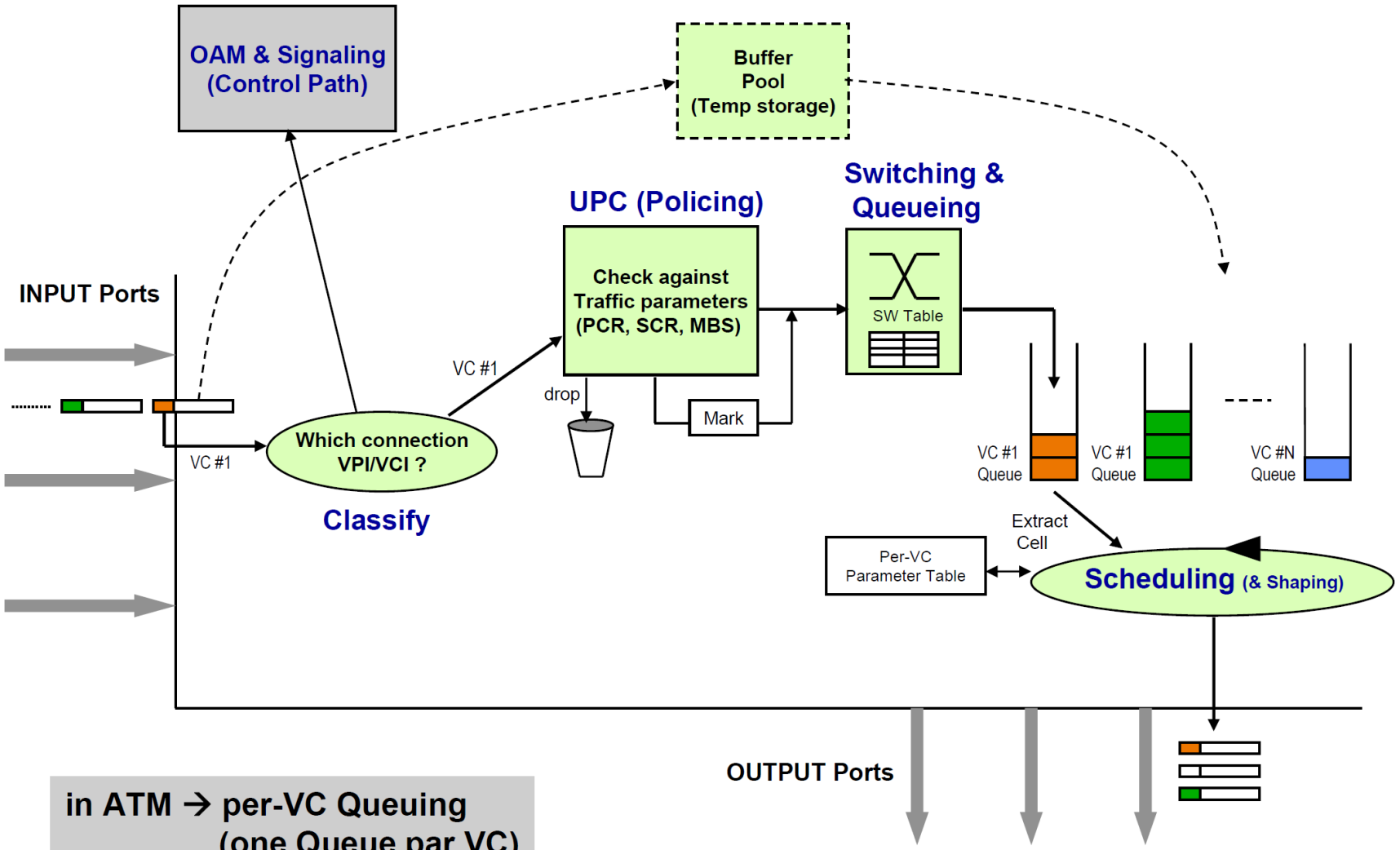
CBR : Constant Bit Rate
VBR : Variable Bit Rate
ABR : Available Bit Rate
GFR : Guaranteed Frame Rate
UBR : Unspecified Bit Rate

Exemple d'utilisation des catégories de service

- **L'utilisation des différentes catégories de service permet à l'opérateur de mutualiser son infrastructure pour offrir des services différents**



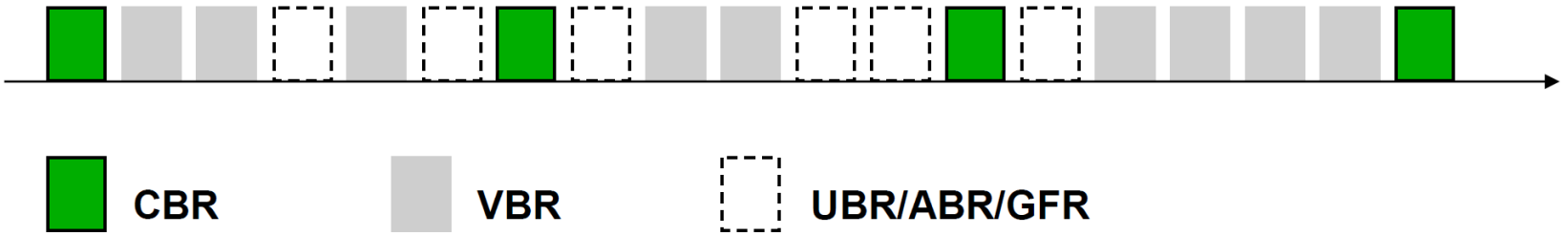
TM/QoS in ATM Switch: Functional description

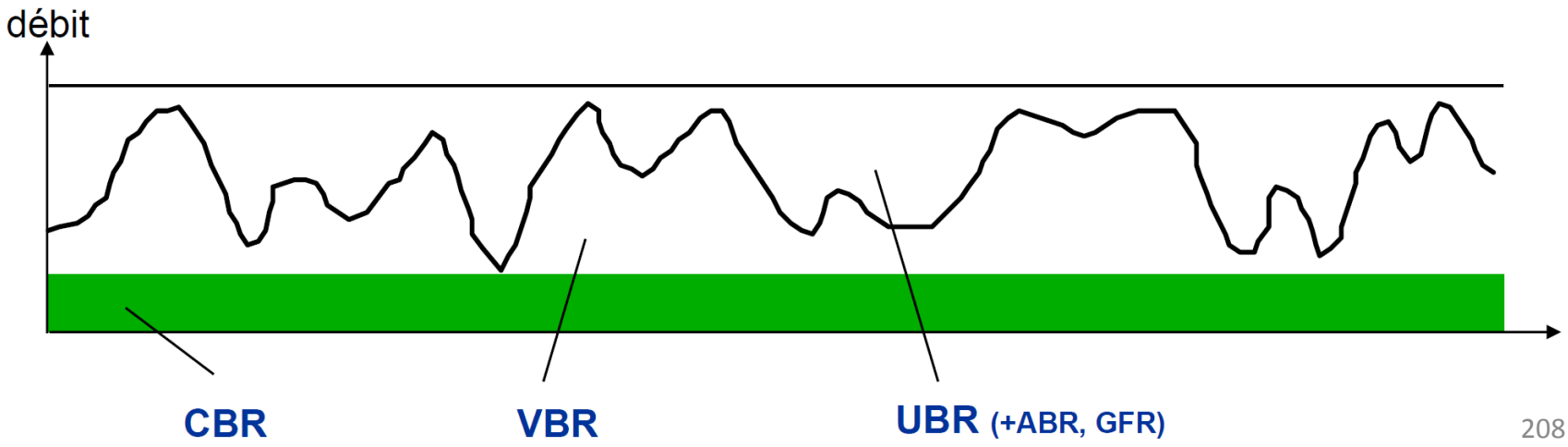


ATM-TM: Classes de service

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)





ATM: Principes

- ATM est **Asynchrone** en référence à la commutation de circuits

- Le format fixe des cellules permet un **multiplexage temporel très précis** de la bande passante entre les applications et une **gestion déterministe des délais**.
→ avantage par rapport à commutation/routage de paquets (de taille variable)

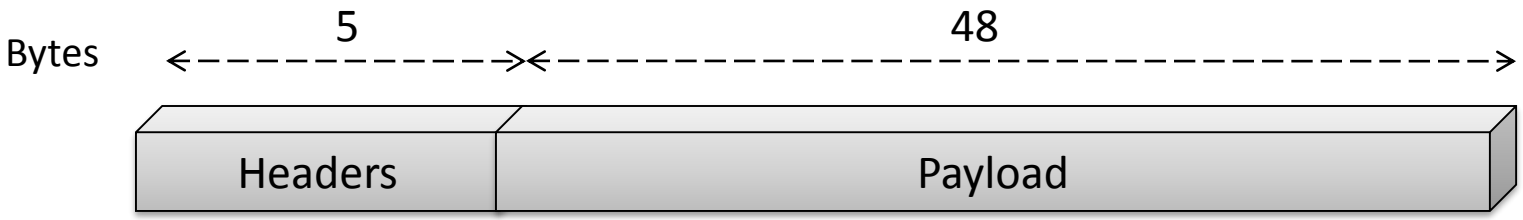
- Contrairement aux circuits synchrones, lorsqu'une source de traffic n'émet pas, elle ne consomme pas de bande passante

- ATM a dès le départ pris en compte tous les aspects de l'Intégration de Services (Notion de RNIS Large Bande) --> différentiation des service, **Traffic Management (TM)**

- ATM est aujourd'hui essentiellement utilisé dans les **réseaux d'accés des technologies multi-services** (Wireless-3G, ADSL), là où les limitations et le coût de la bande passante imposent une allocation fine des ressources du réseau en fonction des applications.

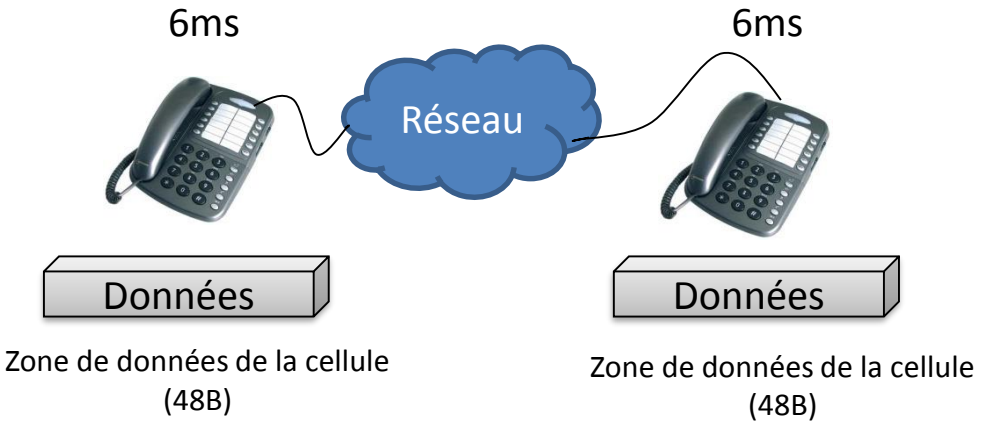
ATM : Principes

- Le flux d'info multiplexé est structuré en petits blocs, ou cellules.
- Assignées à la demande, selon l'activité de la source et les ressources possibles.
- La commutation de cellules est une commutation de trames assez particulière: de petite taille constante.
- 53B, avec 5 d'entête et 43B de données.



Cellule: paquet de taille fixe, et petite

- Remplir 48B à 1B/125µs : 6ms
- En comparaison: remplir 480B: 60ms



X.25 puis Frame Relay puis ATM

- Conçu fin 1980, lancé début 90s avec gros battage: devait résoudre tous les pbs réseaux-télécoms en intégrant voix, données, TV câblée, etc…
- Ca n'a pas été le cas car: pbs de mauvais timing, technologie, implémentation et politique
- Répondait au pb de la QoS sur Internet
- ATM a eu plus gros succès que OSI, et est maintenant largement utilisé en téléphonie, svt pour paquets IP

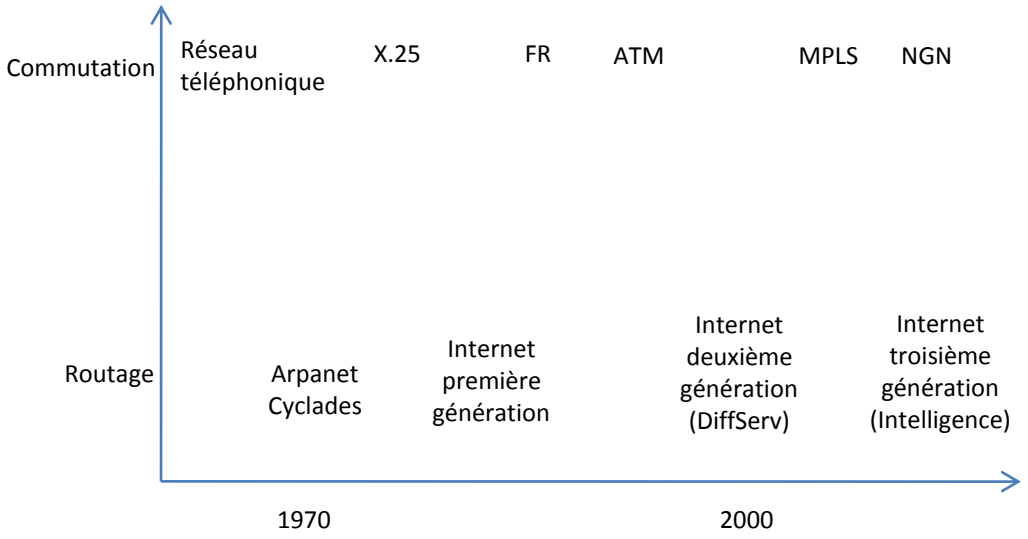
- Parallèlement à la standardisation ATM, grosses compagnies mondiales ont investi en R&D

  -> il y a dans les réseaux énormément d'équipement ATM haute-perf (ATM switchs à xTbits/s)

- Les réseaux dorsaux Internet doivent distribuer le trafic à très haut débit -> plusieurs dorsales d'ISP utilisent beaucoup ATM
- Principalement utilisé pour transport interne des flux, les utilisateurs n'en ont souvent pas conscience, mais ATM est là est encore bien vivant.

ATM: Evolution

- Les techniques de routage n'ont que peu changé, mais la commutation a bcp évolué:
  - 1ère grande norme: X.25 dans les années 1980
  - Puis révolution en 1988 avec adoption du mode de transfert asynchrone: ATM



ATM: Caractéristiques

- Marche avec VC comme FR

- Une partie du header est l'id de connexion -> source, dest et noeuds interm savent à quelle connexion appartient chaque cell -> chaque noeud interm (switch) sait comment commuter la cell

- Commutation de cell en hardware à haute vitesse: plus facile de construire en hardware des switchs pour manipuler des cell courtes de taille fixe = principal argument pour cellules de taille fixe
- Les paquets IP de taille variable doivent être routés en software: process plus long

- Autre plus de l'ATM: hardware peut être configuré pour multiplier les cell en sortie (multi ou broadcast)

- De petites cell ne bloquent pas une ligne pendant longtemps -> QoS plus facile à garantir

- Réseaux ATM organisés comme traditionnels WAN, avec lignes et switch

- La vitesse la plus commune pour des réseaux ATM est 155 Mbps et 622 Mbps

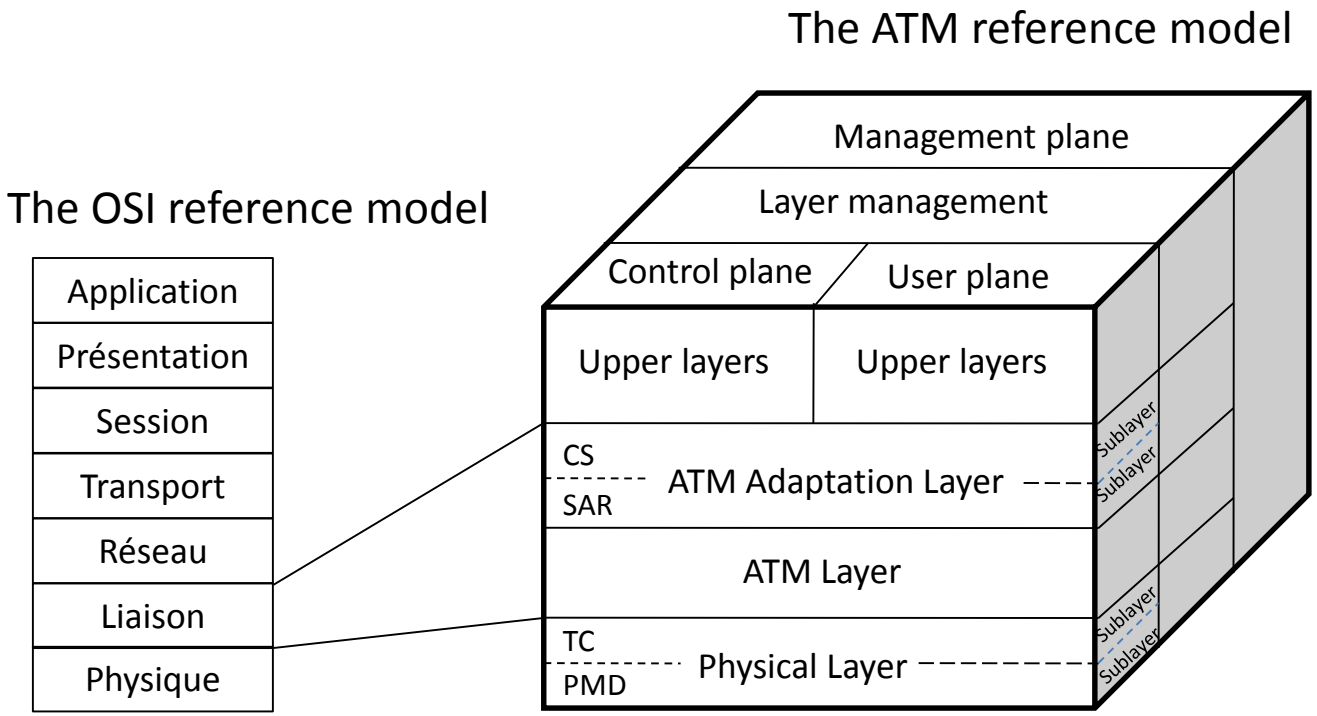
- 155.52 Mbps pour compatibilité avec sys de trans AT&T's SONET

- 622 Mbps = 4 channels of 155.52 Mbps

Le modèle de référence ATM

- **Cette architecture comporte 3 plans: plan utilisateur, plan de contrôle et plan de gestion**.
- Un plan est un réseau. Les 3 réseaux sont multiplexés sur un même réseau physique pour économie, le multiplexage indiquant la simultanéité d'utilisation d'un même composant logiciel ou matériel.

- **Plan utilisateur**: transport de l'information, contrôle de flux, d'erreur, et autres fonctions
- **Plan de contrôle**: signalisation (management de connexion)
- **Plan de management et de couche**: gestion de ressource et coordination entre couches
- Couches PHY et AAL divisés en 2 sous-couches: celle du bas faisant le travail, et au dessus la convergence qui procure une interface à la couche du dessus.



Le modèle de référence ATM

OSI layer	ATM Layer	ATM sublayer	Functionality
	AAL	CS	Providing the standard interface (convergence)
		SAR	Segmentation and Reassembly
2/3	ATM		Flow control Cell header generation/extraction Virtual circuit/path management Cell multiplexing/demultiplexing
2	Physical	TC	Cell rate decoupling Header checksum generation and verification Cell generation Packing/unpacking cells from the enclosing envelope Frame generation
1		PMD	Bit timing Physical network access

ATM Virtual Circuits

- Mode avec connexion: cellule transmise qu'une fois que VC ouvert

- La référence d'un circuit virtuel comporte 2 parties:

    - le numéro VCI (*Virtual Channel Identifier*): identificateur de canal virtuel
    - le numéro VPI (*Virtual Path Identifier*): identificateur de chemin virtuel

- ATM étant en mode connexion, avant toute émission de cell, une connexion doit être mise en place:
  une association entre les références d'entrée et de sortie du réseau doit être définie.

- Routage de la cell de supervision -> met en place le VC, est effectué par des tables de routage

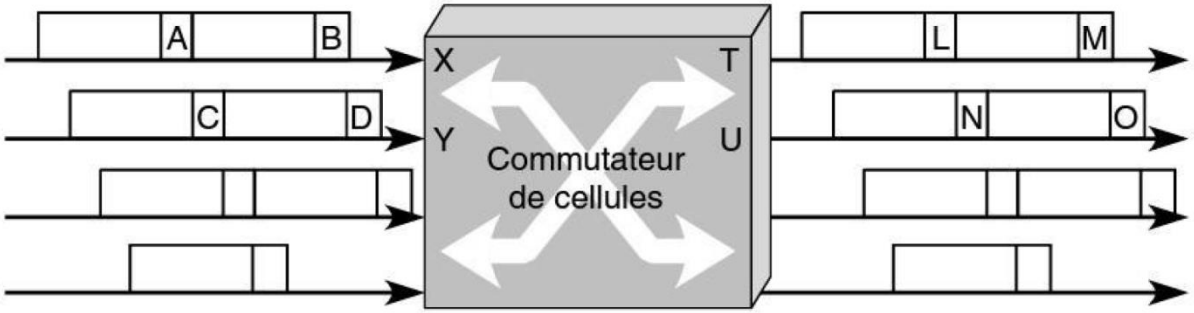
- Cette cell détermine pour chaque nœud l'association entre le ports d'entrée et le port de sortie.

- Ces associations sont regroupées dans la table de commutation.

ATM connections: Virtual Circuits

- Association effectuée entre (référence, port d'entrée) et (référence, port de sortie) dans un switch

- Exemple: si une cell se présente à la porte X avec la ref A, elle est transmise à la sortie T avec la ref L.
  La 2e ligne du tableau de commutation constitue un autre ex: une cell qui entre sur la ligne X avec la ref B est envoyée vers la sortie U, accompagnée de la ref N de sortie



Les trames arrivant sur la porte d'entrée X avec la référence A sont dirigées sur la sortie T avec la référence L.

Ligne d'entrée	Référence d'entrée	Ligne de sortie	Référence de sortie
X	A	T	L
X	B	U	N
Y	C	T	M
Y	D	U	O
.			

ATM connections: Virtual Circuits

- Réseaux ATM orienté connexion => un *virtual channel* doit être établi avant tout transfert de données (*virtual circuit* est *virtual channel* en ATM -> noté VC)

- 2 types de connexion: *virtual path* (VP) identifié par VPI, et *virtual channel* (VC) identifié par VPI+VCI
- La référence d'un circuit virtuel comporte 2 parties:

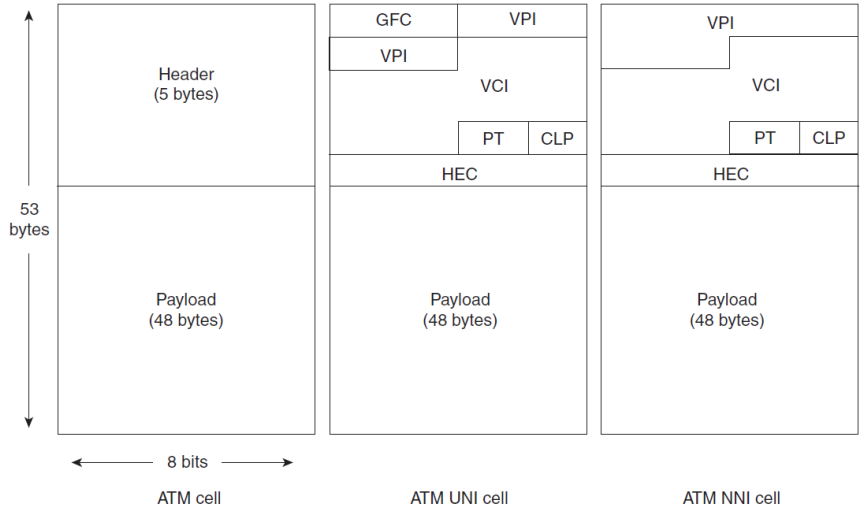
    - le numéro VCI (*Virtual Channel Identifier*): identificateur de canal virtuel
    - le numéro VPI (*Virtual Path Identifier*): identificateur de chemin virtuel

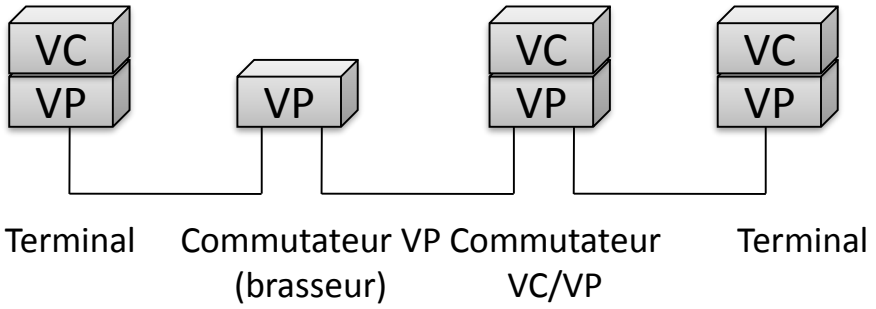
- Un VP est un faisceau de VC, dont tous sont commutés de façon transparente par réseau basé sur VPI
- VPI et VCI ont signification seulement locale (à chaque switch)

ATM connections: Virtual Circuits

-Dans un switch ATM, on commute une cell en utilisant les 2 ref. Dans un brasseur, on ne se sert que d'une ref, celle du conduit.

- Par ex, on peut commuter un ensemble de VC en une seule fois en ne se préoccupant que du VP = brasseur de conduit, ou *cross-connect* (on ne redescend pas au niveau du VC)





- brasseur de VP: tous les VCs du VP commutés simultanément

- on a donc intérêt à regrouper en un VP les VCs qui vont vers la même destination

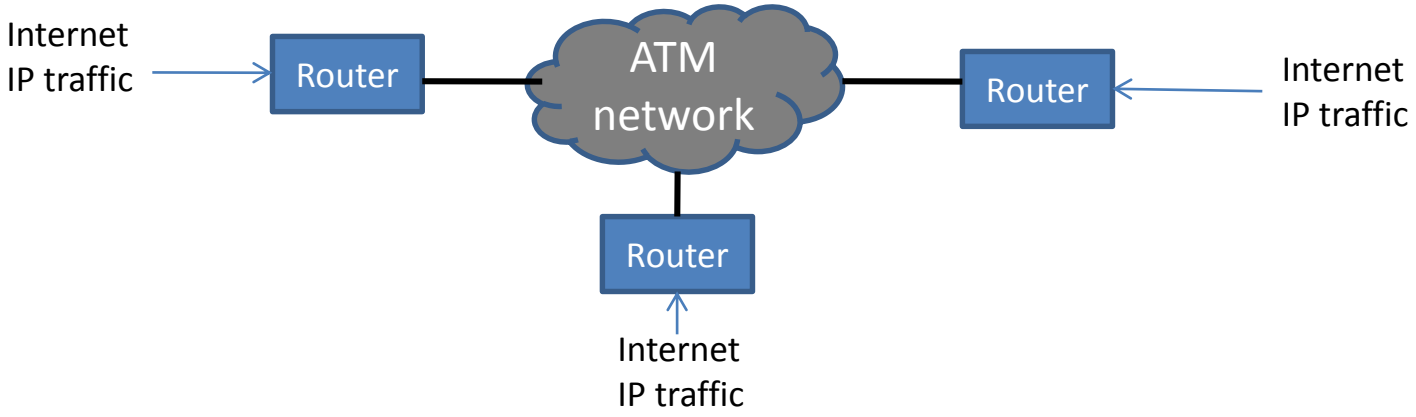
-> diminue la taille des tables de commutation à l'intérieur du réseau

ATM pour encapsuler IP

IP over ATM

- un backbone ATM avec 3 points entrée/sortie (routeur) pour du trafic IP
- un backbone ATM peut s'étendre sur un continent et avoir des 100aines de switches ATM
- En général, les backbones ATM ont un PVC entre chaque paire de points entrée/sortie
  (Recall that ATM uses the jargon "virtual channel" for "virtual circuit".)

- En utilisant des PVCs, les cellules ATM sont commutées d'un point d'entrée à un point de sortie sans avoir à établir et démanteler les VCs.
- Mais PVCs faisables que sur un nombre limité de entry/exit points
  (For n entry points, n(n-1) permanent VCs are necessary.)
- Chaque interface de routeur connectée à un réseau ATM aura 2 adresses: @IP et @ATM (layer 3 and 2)
- L'émetteur maintient table de correspondance entre @ATM et VCIs.



ATM pour encapsuler IP

-Le routeur d'entrée:

1. Examine l'@IP de destination du datagramme.
2. Determine l'@IP du routeur de sortie (i.e., the next router in its route).
3. Le routeur d'entrée voit ATM juste comme un autre protocole de couche 2. (possède table de corresp. @IP/@ATM)
4. IP dans le entry router passe le datagramme à LL (i.e., ATM) avec l'@ATM du exit router.

- Après ces 4 étapes, uniquement ATM se charge de transférer le datagramme à l'@ATM de sortie (Step 3 above):

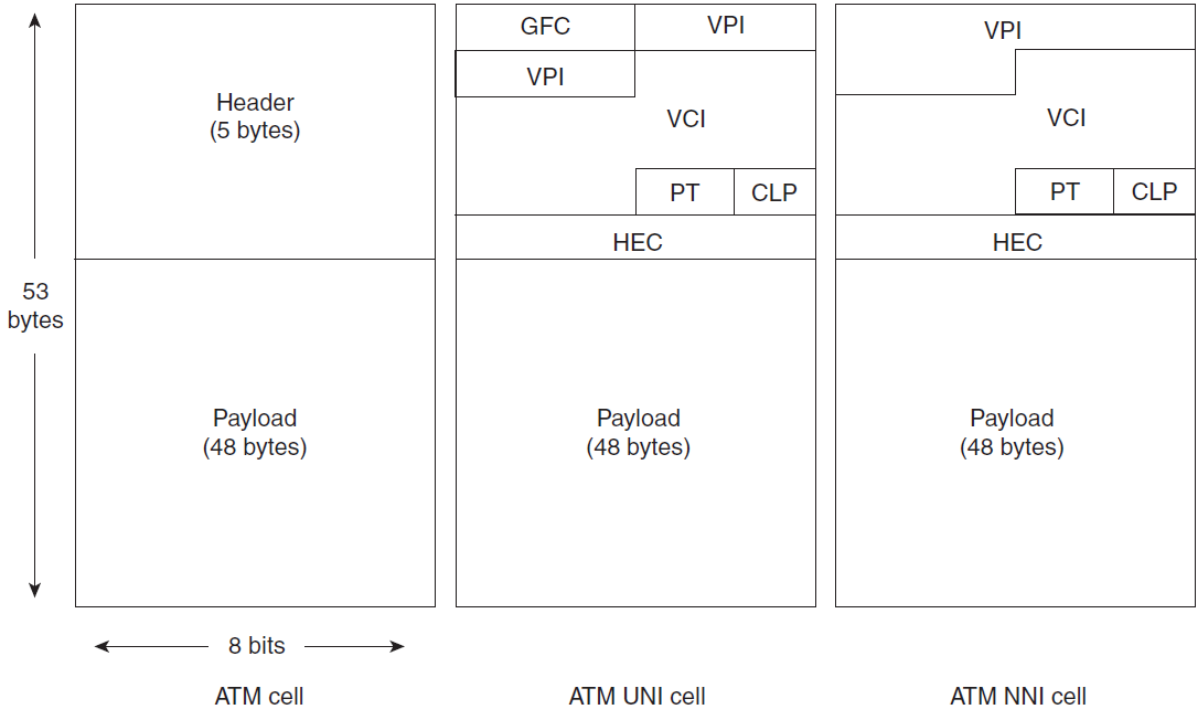
1. Détermine le VCI+VPI pour le VC qui mène à l'@ATM de destination.
2. Segmente le datagramme en cellules côté émetteur du VC (i.e., at the entry router), et ré-assemble les cellules côté récepteur du VC (i.e., at the exit router).

--> introduction to the 3 layers to answer to "How does ATM efficiently segment and reassemble IP datagrams that are sent across an ATM backbone ? ''

ATM: La couche ATM

- 53B = 48B+5B
- zone de données: fragment d'un paquet utilisateur, généralement un fragment d'un paquet IP
- Sur les 48B provenant de la couche sup, jusqu'à 4B peuvent concerner la supervision -> la fragmentation du paquet est faite en blocs de 44B (4B de AAL).

- Entête de la cell ATM différente selon l'interface UNI ou NNI :
  - l'interface NNI (*Network Node Interface*) entre 2 nœuds du réseau
  - l'interface UNI (*User Network Interface*), utilisée pour rentrer dans le réseau ou en sortir



ATM: La couche ATM

PTI (Payload Type Identifier) : Type des données (AAL1. . .) et indication de congestion (Explicit Forward Congestion Indication)

CLP (cell loss priority) : priorité de suppression de cellule

GFC (Generic Flow Control) : non utilisé

HEC (Header Error Control) : CRC sur l'entête.

Le champ CLP (*Cell Loss Priority*)

- indique si la cell peut être perdue (CLP=1), ou si elle est importante

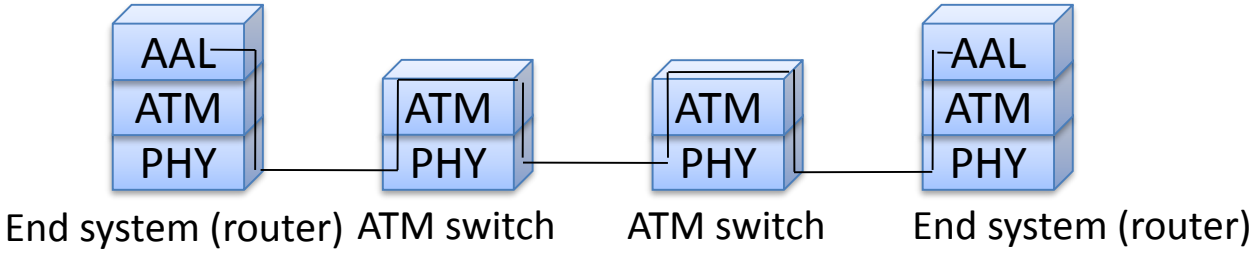
- sa fonction est d'aider au contrôle de congestion: avant d'émettre une cell dans le réseau, il faut respecter un débit d'entrée, négocié au moment de l'ouverture du VC

- toujours possible de faire rentrer cell en surnombre, mais avec indicateur permettant de les perdre pour préserver le transport des infos entrées dans le cadre de l'abonnement

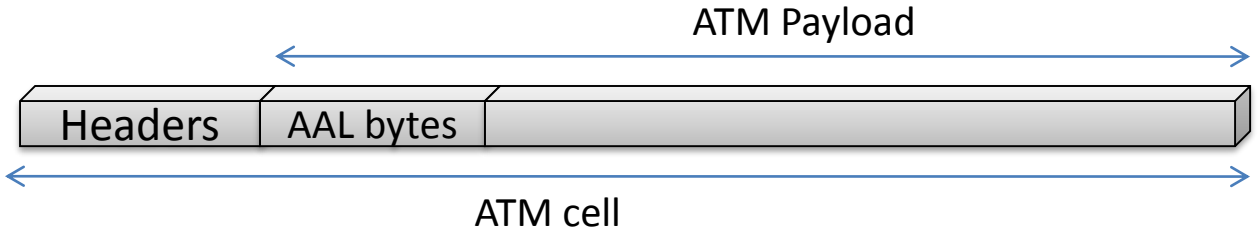
ATM: La couche AAL

- But: permettre aux protocoles (e.g., IP) et applis (e.g., constant bit rate video) existants de tourner au dessus d'ATM

- AAL est implanté dans les terminaux ATM (e.g., entry and exit routers in an Internet backbone), pas dans les switchs ATM intermédiaires -> analogue en ce sens à la couche transport de la pile IP



- La couche AAL a son propre champ d'entête, occupant une partie du payload de la cell ATM:



- Divisée en 2 sous-couches:
  - CS (*Convergence Sublayer*): pour l'interfonctionnement avec la couche du dessus
  - SAR (*Segmentation and Reassembly sublayer*): pour segmentation et ré-assemblage message-cellules

ATM: couche AAL et classes de services

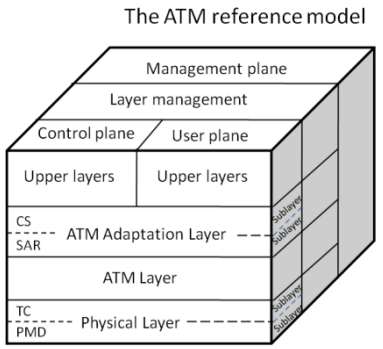
-L'ITU recommande 4 types de protocoles AAL pour supporter ces classes de services :

- **AAL-1**: pour services de la **classe A** -> service d'émulation de circuit
  Le service rendu par l'AAL-1 s'appelle **CBR (*Constant Bit Rate*)**.

- **AAL-2**: défini au départ pour les services de la **classe B** (ex: service vidéo à débit variable)
  Le service rendu par cette classe s'appelle **VBR (*Variable Bit Rate*)**.
  Ce protocole est utilisé par ex sur la partie accès de l'UMTS, pour permettre le multiplexage de plusieurs connexions bas débit sur une connexion ATM de façon à tenir compte au mieux des contraintes temporelles.

- **AAL-5**: aka SEAL (*Simple Efficient Adaptation Layer*): pour trames de données (paquets IP) en mode avec connexion (**classe C**).
  Service élastique et utilise le service **ABR (*Available Bit Rate*)**.

ATM: exemple avec AAL 5

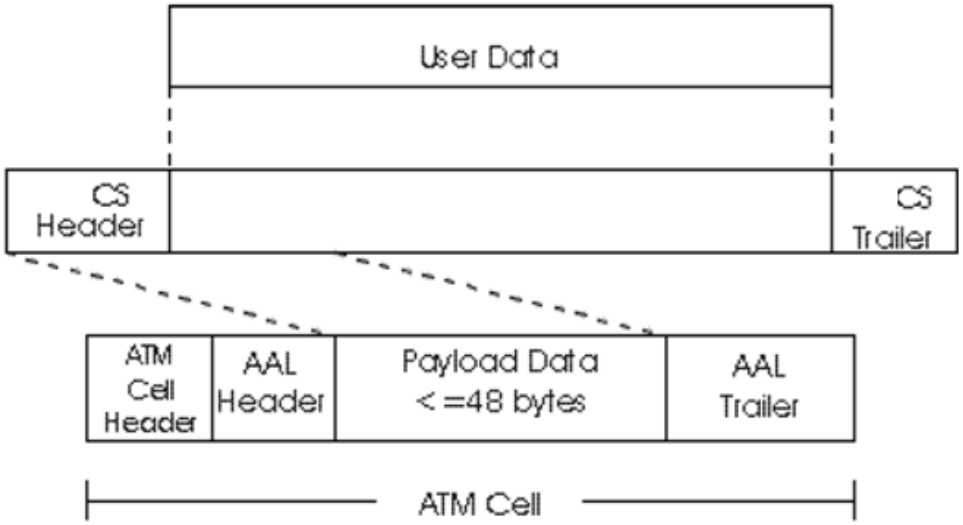


AAL structure

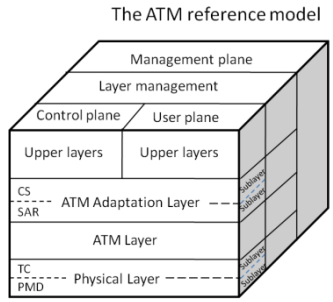
- Les données utilisateur (e.g., an IP datagram) sont d'abord encapsulées dans un CS-PDU avec CS header et CS trailer.
- CS-PDU trop grande pour le payload d'une cell ATM

-> le CS-PDU doit être segmenté à la source ATM et ré-assemblé à la destination ATM.

- La sous-couche SAR segmente le CS-PDU et ajoute un header et un trailer AAL qui sera le payload des cellules ATM.
- (Selon les types AAL, les headers and trailers CS et AAL peuvent être vides.)



ATM: exemple avec AAL 5



AAL 5 (Simple and Efficient Adaptation Layer - SEAL)

- AAL5 est un AAL léger utilisé pour le transport de datagrammes IP sur des réseaux ATM

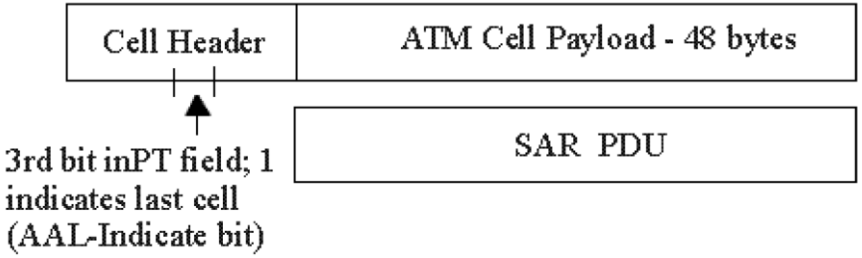
- Un datagramme IP occupe le CS-PDU payload (1 à 65 535 bytes). The AAL5 CS-PDU:

CS-PDU payload	PAD	Length	CRC
0-65535	0-47	2	4

-Le PAD assure que le CS-PDU est un multiple de 48 bytes.

- Le champ longueur identifie la taille du payload CS-PDU, pour pouvoir enlever le PAD au récepteur.

- **A la source ATM**, le SAR de AAL5 découpe le CS-PDU en segments de 48 bytes. Un bit dans le champ PT de l'entête de la cellule ATM, qui est 0 par défaut, est mis à 1 pour la dernière cellulde du CS-PDU.
- **A la destination ATM**, la couche ATM dirige les cellules avec un VCI spécifique vers un buffer de la sous-couche SAR.
- Les entêtes de la cellule ATM sont retirées, et le bit du champ PT est utiliser pour délimiter les CS-PDUs.
- Une fois que le CS-PDU est délimité, il est passé à la sous-couche AAL.
- A la sous-couche CS, le champ *length* est utilisé pour extraire le payload du CS-PDU (ex : un pkt IP)



ATM-TM: Paramètres de trafic et de QoS

Une source de traffic (Application) s'identifie au réseau par un ensemble de paramètres décrivant plus ou moins précisément l'utilisation du réseau ATM que compte faire cette source en terme de traffic et de qualité de service.

Paramètres de traffic

. Peak Cell Rate **PCR**
. Sustainable Cell Rate **SCR**
. Maximum Burst Size **MBS**
. Minimum Cell Rate **MCR**

Paramètres de qualité de service

. Cell Delay Variation **CDV**
. Maximum Cell Transfer Delay **Max CTD**
. Cell Loss Ratio **CLR**

→ Ces paramètres sont échangés à l'interface UNI (User-Network) lors de la phase d'initialisation d'une connection. Ils peuvent être explicites ou implicites (définis par défaut par le réseau).

→ La prise en compte de ces paramètres se fera par des mecanismes appropriés (file d'attente, scheduling, …) mis en oeuvre pour chaque connection (per-VC ou per-flow) dans les equipements ATM du réseau.

		Classes de service					
Paramètres		CBR	VBR-rt	VBR-nrt	ABR	UBR	GFR
Paramètres de QoS							
CTD	*Cell Trandfert Delay*	•	•	•			
CDV	*Cell Delay Variation*	•	•				
CLR	*Cell Loss Ratio*	•	•	•	•		
Paramètres de Trafic							
PCR	*Peak Cell Rate*	•	•	•	•		
SCR	*Sustained Cell Rate*		•	•			
MBS	*Max Burst Size*		•	•			
MCR	*Minimum Cell Rate*				•		•

ATM: Qualité de Service (QoS)

- ATM apporte un gain économique grâce au multiplexage statistique.

- Mais multiplexage statistique de trafic en rafale peut provoquer congestion.
- Travaux de l'ITU et ATM Forum ont visé à minimiser cette congestion et à **maximiser le taux d'utilisation du réseau, tout en garantissant la QoS spécifiée par l'utilisateur.**

-Or c'est la congestion qui est le seul obstable à la QoS

--> ATM gère la congestion avec :

- *Call Admission Control* (CAC): pour garantir QoS, on préfère ne pas admettre de nouveaux clients qui pourraient dégrader la QoS des autres – CAC déterministe et statistique selon la QoS du VC à établir
- On essaye au maximum d'éviter la congestion avec CAC, si elle arrive, CLP pour abandon de cellules et PTI pour notification de congestion

ATM: Conclusion

- Les réseaux de niveau trame sont au nombre de 3: FR, Ethernet et ATM. Si le 1er est en baisse et que le 2e monte, le transfert ATM reste stable, malgré de nombreuses prédictions prévoyant sa chute à partir de 2005.

- On peut expliquer cette stabilité de l'ATM par la grande confiance des operateurs de télécoms à son égard: cette technique est aujourd'hui parfaitement maitrisée, et on est capable de bien la dimensionner.

- De plus, elle offre des garanties de QoS qui lui confèrent un atout supplémentaire.

- Nous n'avons pas décrit la signalisation du réseau ATM, car celle d'origine a été remplacée par une signalisation IP afin de passer à la techno MPLS, décrite après.

- Nous verrons alors que l'ATM reste une technique très importante dans cette nouvelle génération de réseaux.

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 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

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

D'ATM vers MPLS

- **Quelle technologie va servir de base pour l'infrastructure unique de convergence pour les réseaux d'opérateurs?**
- **ATM était le choix naturel**
  - Solution éprouvée
  - Investissements souvent déjà engagés
  - Possibilité de haut débit grâce à son fonctionnement hardware
  - Fonctions riches pour assurer différents niveaux de qualité de service
    - Selon le type de trafic ou le type de service offert au client de l'opérateur
  - Possibilités d'ingénierie de trafic
    - Agrégation de trafics sur des PVC respectant des contraintes
- **Mais l'environnement technique a changé depuis la conception d'ATM**
  - IP est devenu le protocole de niveau 3 incontournable
  - Le hardware est monté en puissance
  - La bande passante est plus abondante et moins chère
  - Les stations d'extrémité sont plus intelligentes

D'ATM vers 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

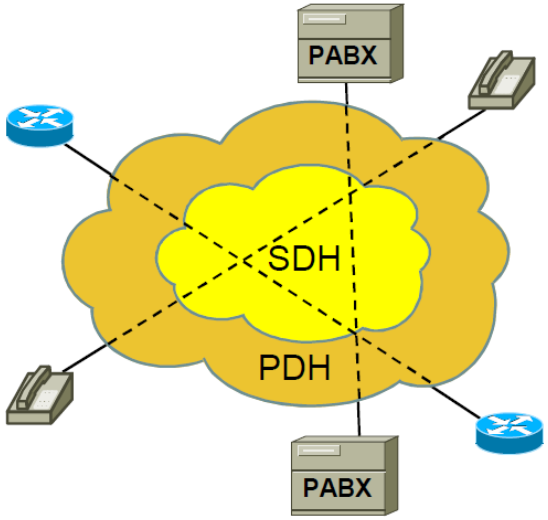
- **De nombreux constructeurs ont proposé des solutions de commutation IP faisant la synthèse de la commutation ATM et du routage IP**
  - IP switching (Nokia), ARIS (IBM), Tag Switching (Cisco), etc…

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

Evolution des infrastructures réseau des opérateurs

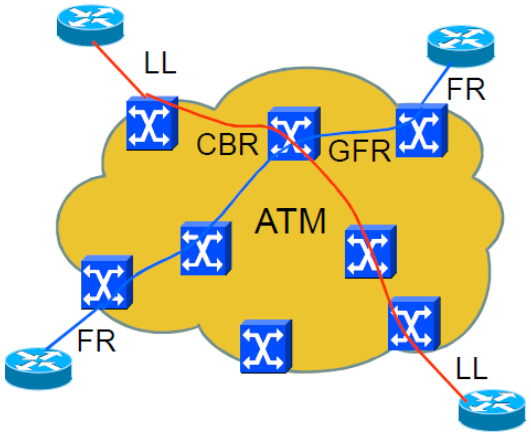
- **L'infrastructure des opérateurs a subi des profondes migrations**

Infrastructure de niveau 1



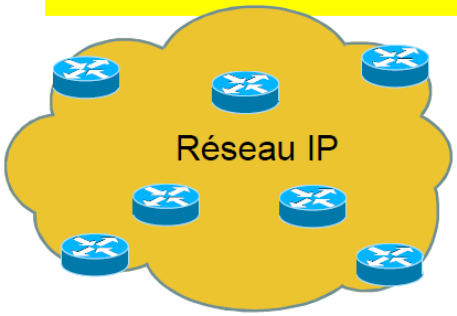
- **Services de niveau 1**
  - Liaisons louées
  - Téléphonie classique
  - Interconnexion de PABX

Infrastructure de niveau 2



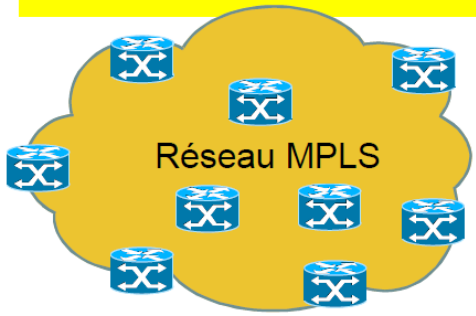
- **Services de niveau 1 et 2**
  - Liaisons louées (CBR)
  - Frame Relay (GFR)
  - Vidéo (VBR-rt)
  - RNIS-Large Bande

Infrastructure de niveau 3



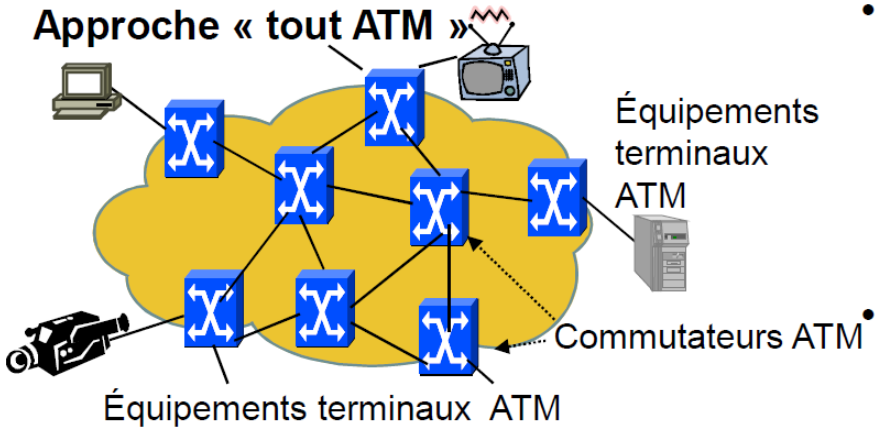
- **Services de niveau 3**
  - Accès Internet
  - VPN IPsec

Infrastructure de niveau 2/3

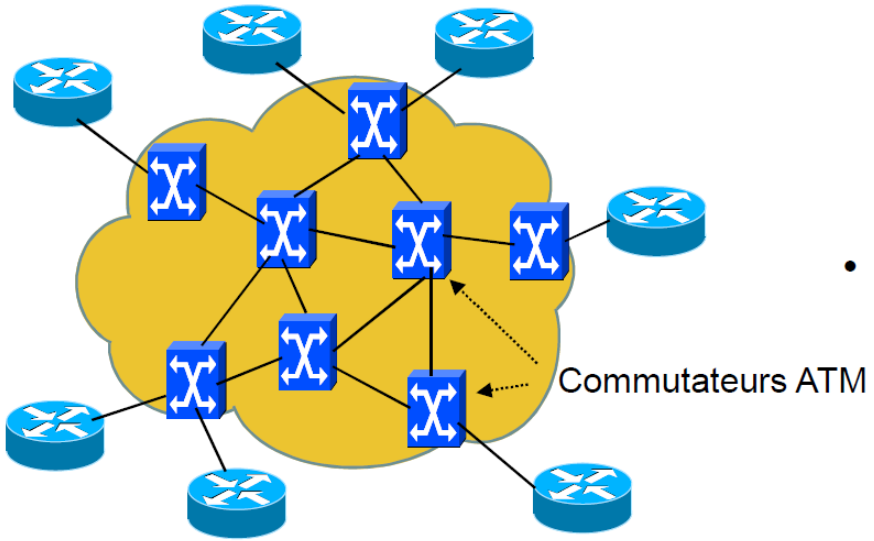


- **Services de niveau 1, 2 et 3**
  - Différents types de VPN

Les approches possibles des opérateurs



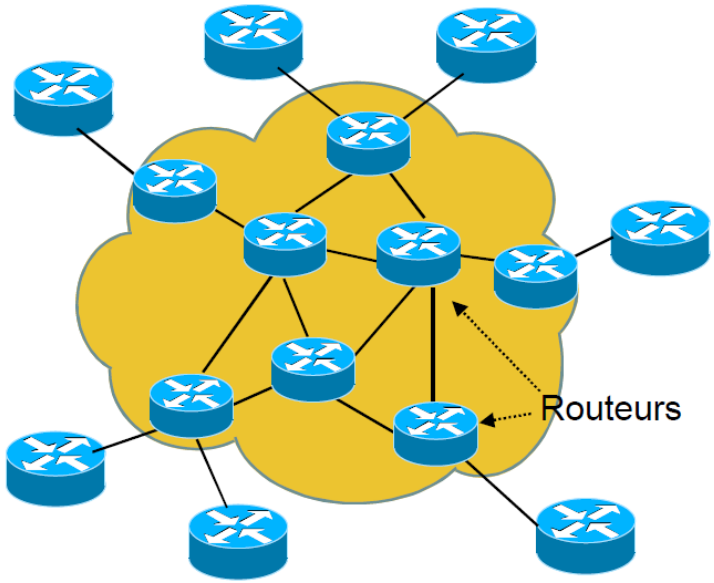
Approche de superposition : IP sur ATM



- **Les opérateurs avaient choisi une approche « tout ATM » sans IP**
  - ATM « sans couture »
  - Le succès d'Ethernet et d'IP a éliminé cette approche
- **Les opérateurs ont alors utilisé leur infrastructure ATM pour transporter les paquets IP**
  - Le mode circuit virtuel d'ATM permet aux opérateurs de faire de l'ingénierie de trafic
    - Choisir sur quel circuit virtuel transporter tel type de trafic
- **Cohabitation de deux technologies**
  - Non conçues pour interopérer
  - Besoin de doubles compétences

Les approches possibles des opérateurs

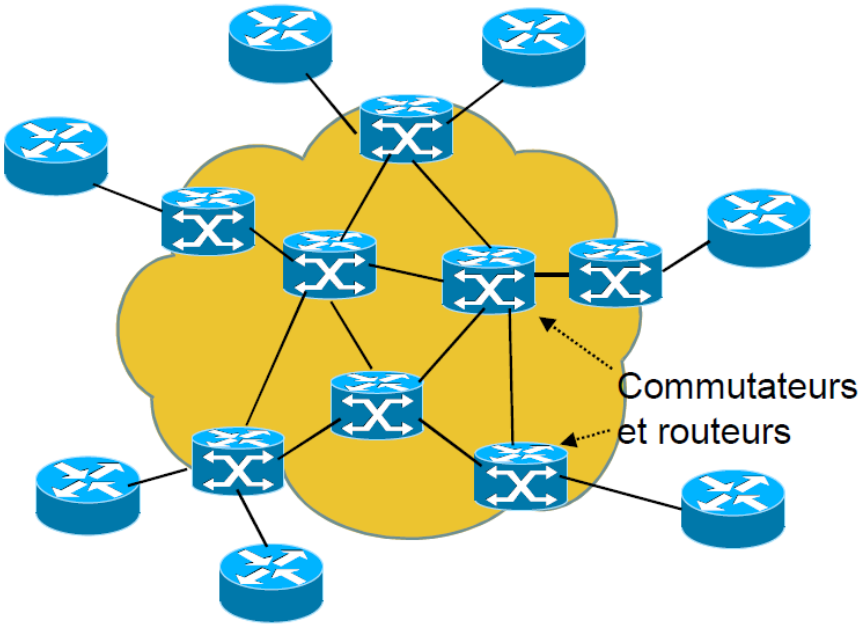
Approche IP pur sans ATM



- **Les opérateurs ont alors envisagé une approche « tout IP »**
  - Une seule compétence
- **Le routage dans le réseau de l'opérateur est celui du mode datagramme IP**
  - Basé sur l'adresse destinataire du paquet IP
  - Tous les paquets sont envoyés sur la meilleure route IP
- **Les routeurs au départ étaient plus lents que les commutateurs ATM**
  - Mais les routeurs rapides ont corrigé ce problème
- **Pas de possibilité d'ingénierie de trafic et d'équilibrage de trafic**

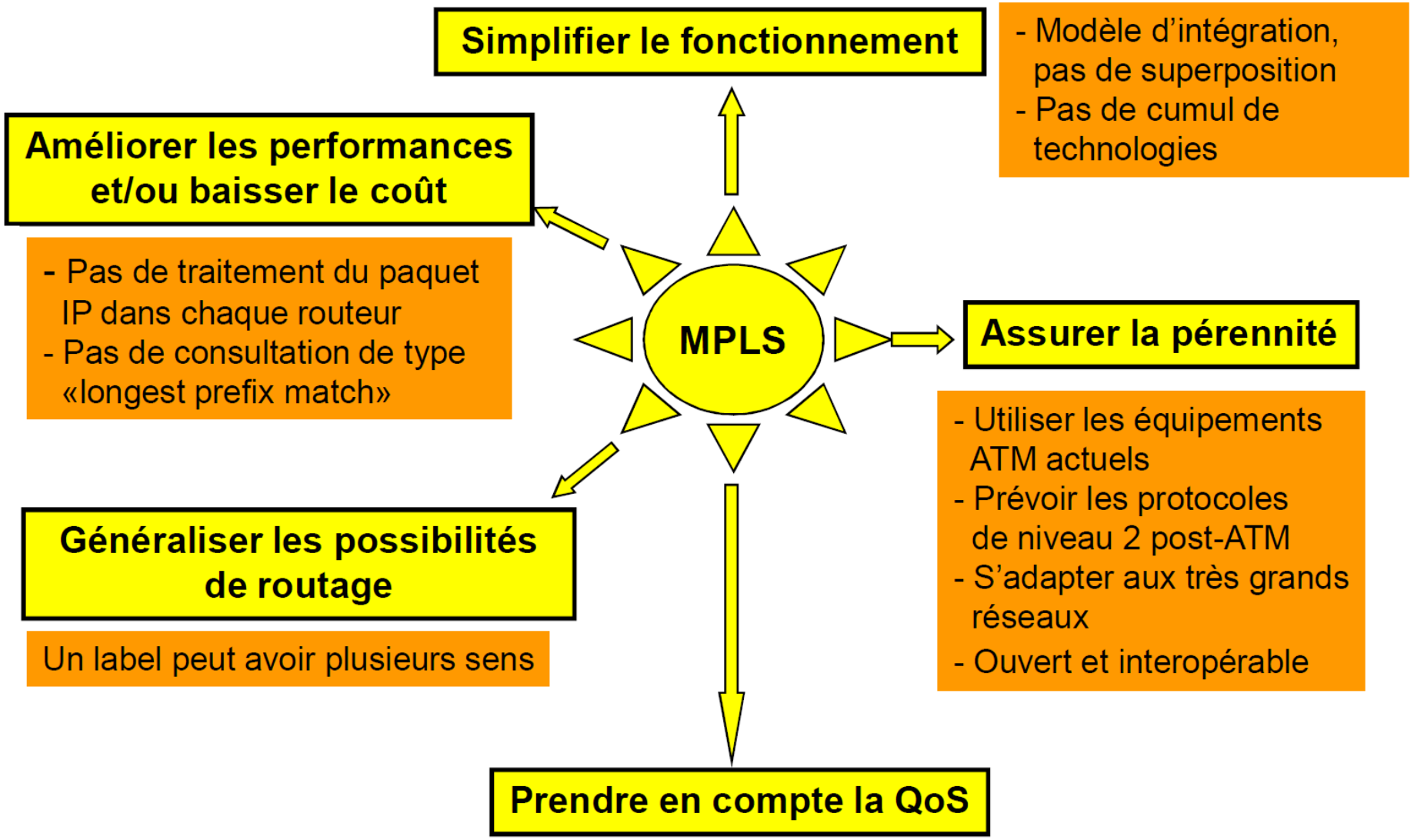
Les approches possibles des opérateurs

Approche d'intégration : MPLS



- **Une approche intermédiaire d'intégration a été proposée**
  - Un seul plan d'adressage IP
  - Les équipements internes de l'opérateur sont des équipements hybrides
    - Ils peuvent fonctionner comme des routeurs IP
    - Ils peuvent aussi fonctionner dans un mode proche du circuit virtuel en commutant selon un numéro de label
- **Le routage n'est plus forcément corrélé à l'adresse destinataire IP**
  - L'ingénierie de trafic est possible
- **IP et MPLS sont des technologies proches conçues pour interopérer**

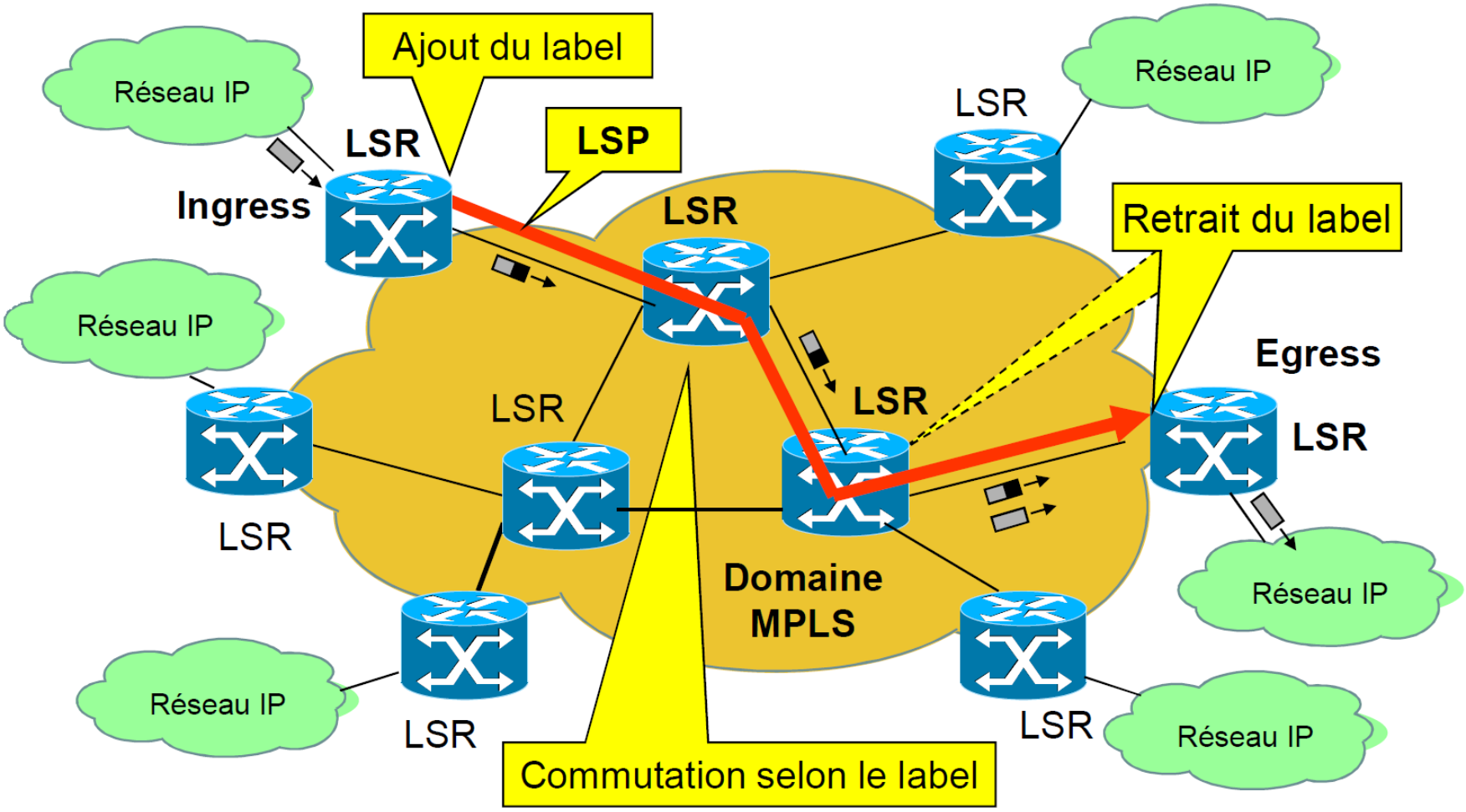
Objectifs de MPLS



Principes de MPLS

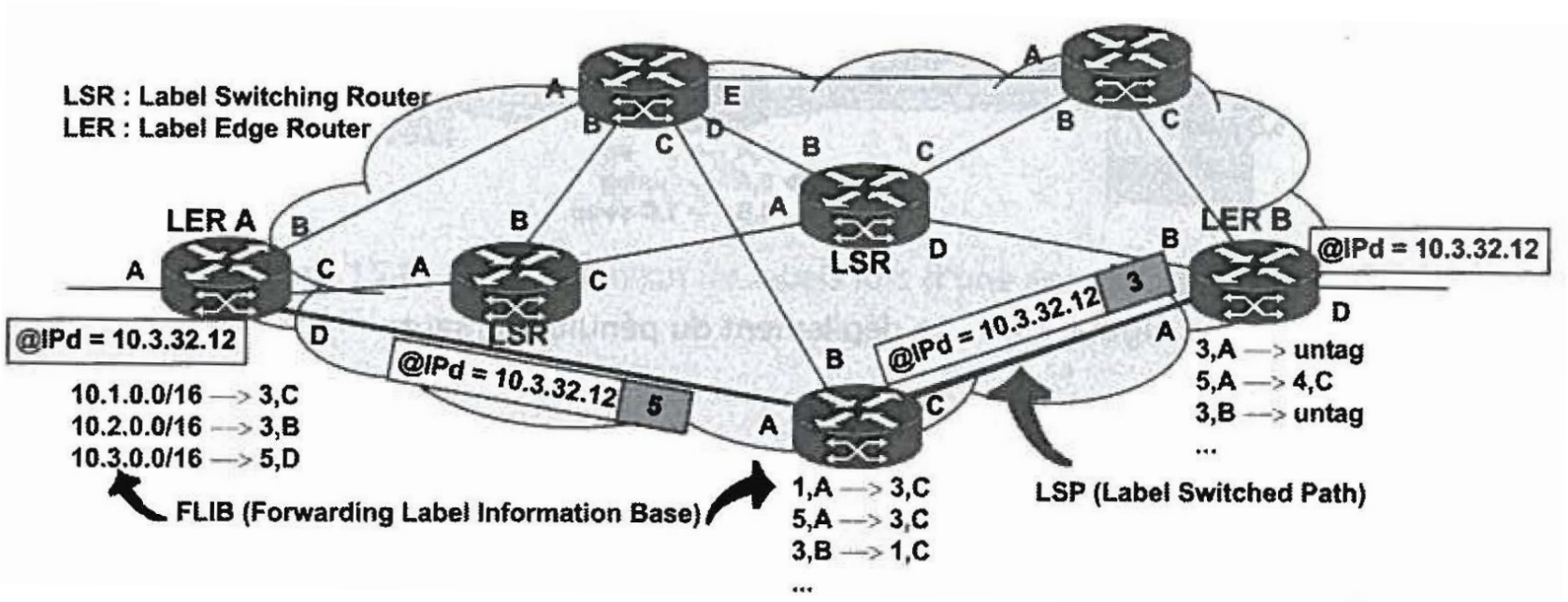
- **Des chemins prédéfinis relient les points d'extrémité du réseau**
  - Les LSP (Label Switched Path)
    - Un LSP est unidirectionnel
  - Les LSP sont établis par un protocole de signalisation en suivant la route déterminée par les protocoles de routage IP
    - Les LSP peuvent être établis à l'initiative 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

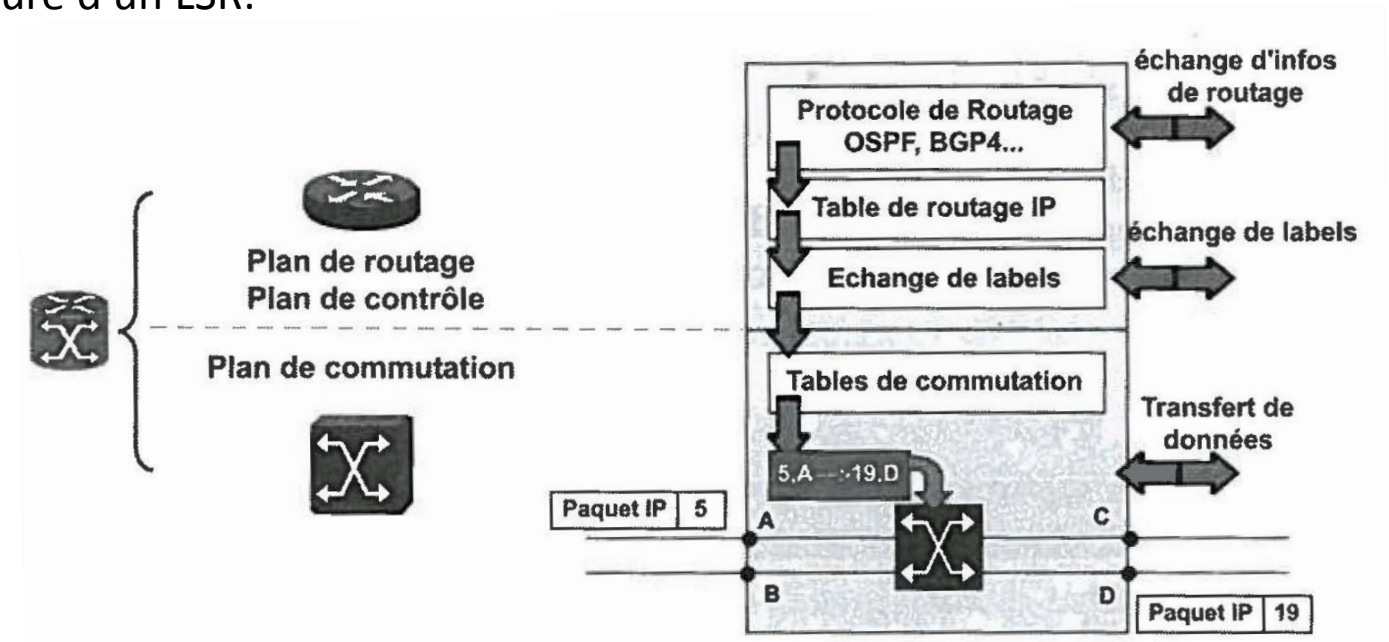
Acheminement d'un datagramme IP dans un réseau MPLS:



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 de MPLS

Structure d'un LSR:



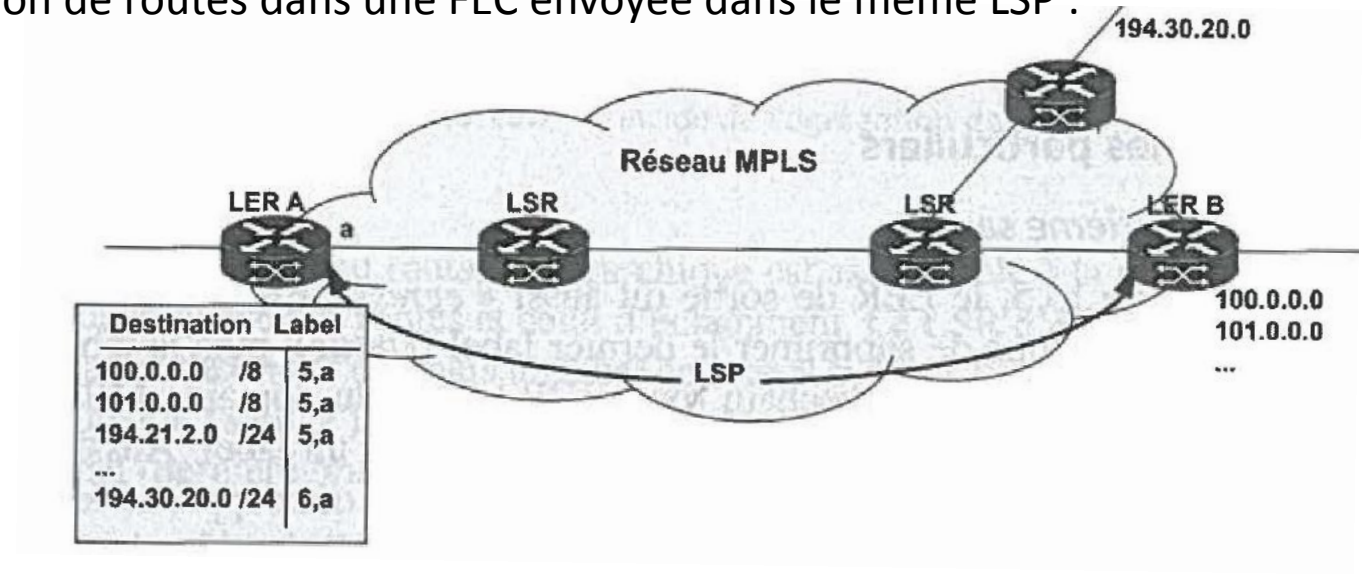
- Un *Label Switching Router (LSR)* correspond à l'association d'un routeur et d'un commutateur.
- En principe un switch FR ou ATM peut évoluer en LSR par une simple mise à jour du logiciel.

- Le plan de routage ou plan de contrôle met en œuvre un protocole de routage IP
- Un protocole spécifique distribue les labels aux LSR voisins : **LDP, *Label Distribution Protocol*.**
- Tous les paquets à destination d'un même nœud de sortie reçoivent le même label et subissent donc le même traitement dans le réseau.

-> **agrégation de routes** à la périphérie: allège les tables et participe à l'amélioration des performances.

Principes de MPLS

Agrégation de routes dans une FEC envoyée dans le même LSP :



- Croissance du réseau facilitée
- Augmentation du nombre de sites transparente au cœur du réseau: toutes les ad réseau connues du LER B reçoivent en entrée le même label

- **L'ensemble des datagrammes qui reçoivent un même label forme une FEC.**
- Le circuit virtuel défini pour une classe d'équivalence est appelé LSP.

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

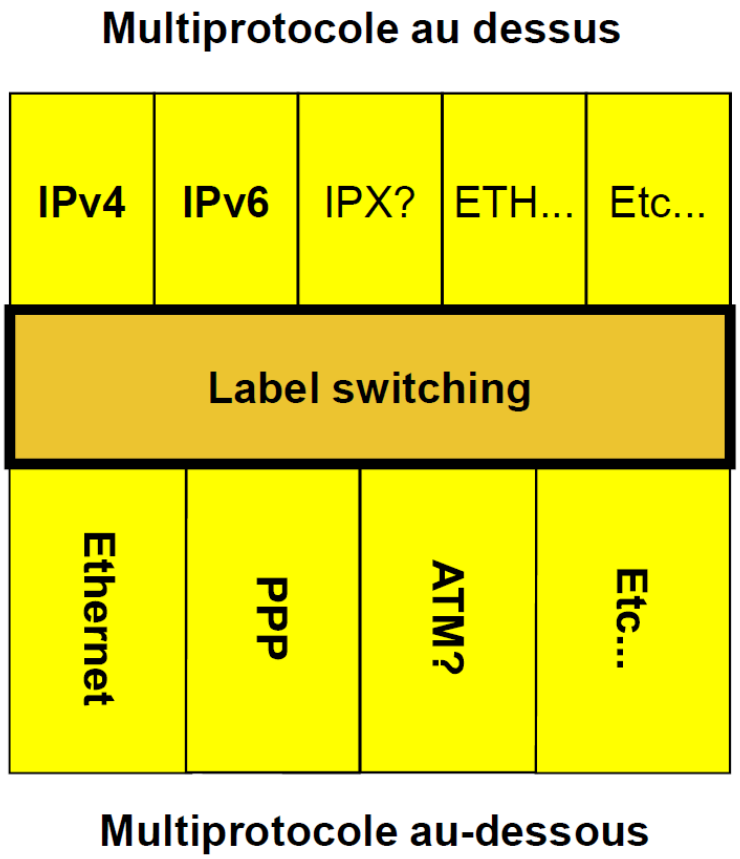
- Peut être quelconque
- Grossière si la FEC regroupe des adresses destinataires de même préfixe
- Fine si la FEC correspond au flux d'une application entre 2 machines

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

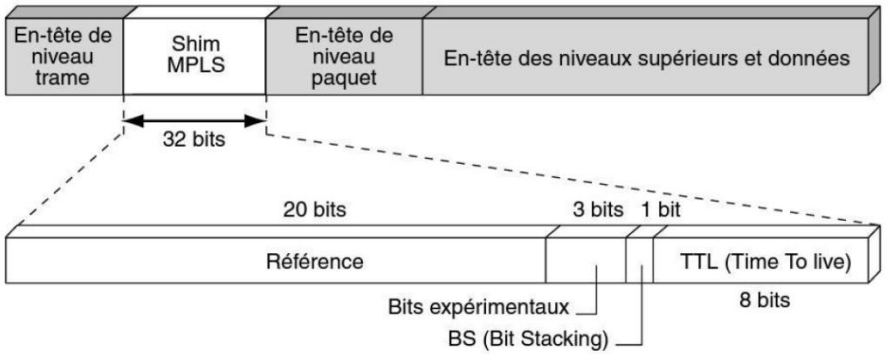
- **Protocole supérieur quelconque**
  - IPv4 ou IPv6 (niveau 3)
  - Ethernet (pour service VPLS)
- **Protocole de niveau 2 quelconque**
  - ATM (pour migration)
  - PPP (sur liaisons Sonet/SDH)
  - Ethernet 1 ou 10 Gbps
  - Combinaison des approches précédentes
- **MPLS est donc flexible**
  - Peut 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)



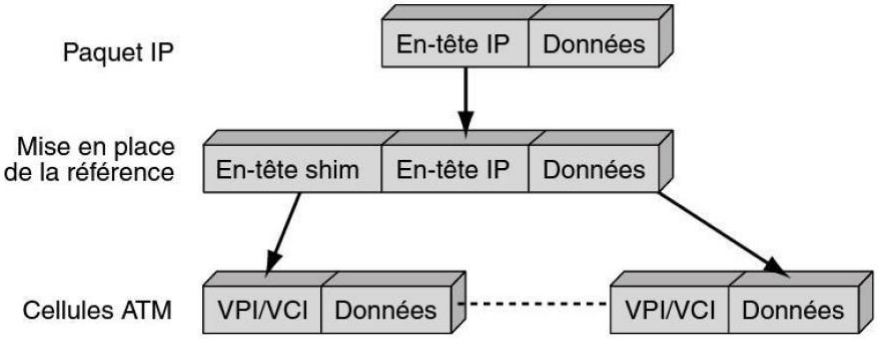
VPLS : Virtual Private LAN Service

MPLS et les infrastructures existantes

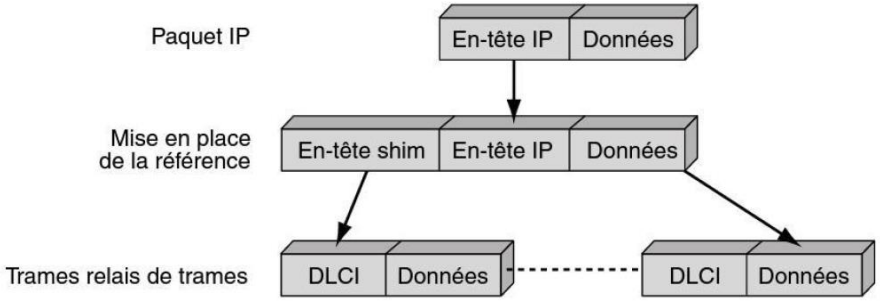
- Quand MPLS est utilisé sur une infrastructure de commutation traditionnelle, il utilise les techniques de la technologie sous-jacente : ceci confère à MPLS une grande facilité d'adaptation puisqu'il ne remet pas en cause les investissements réalisés.
- Dans ces réseaux, la valeur du premier label MPLS affecté est utilisée comme étiquette de commutation : DLCI ou VPi/VCi.
- Un protocole de routage IP traditionnel et un protocole de distribution des labels se substituent à la signalisation d'origine.



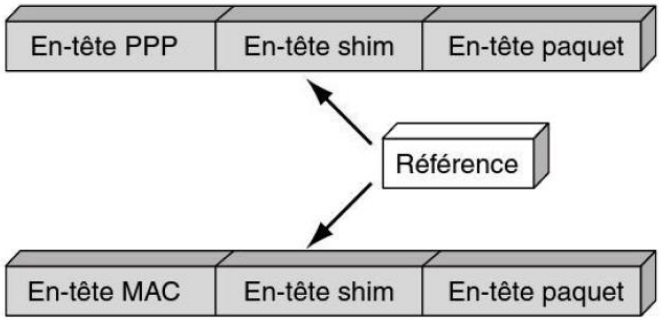
Format générique d'un label (ref) dans MPLS



Mise en place des labels dans ATM



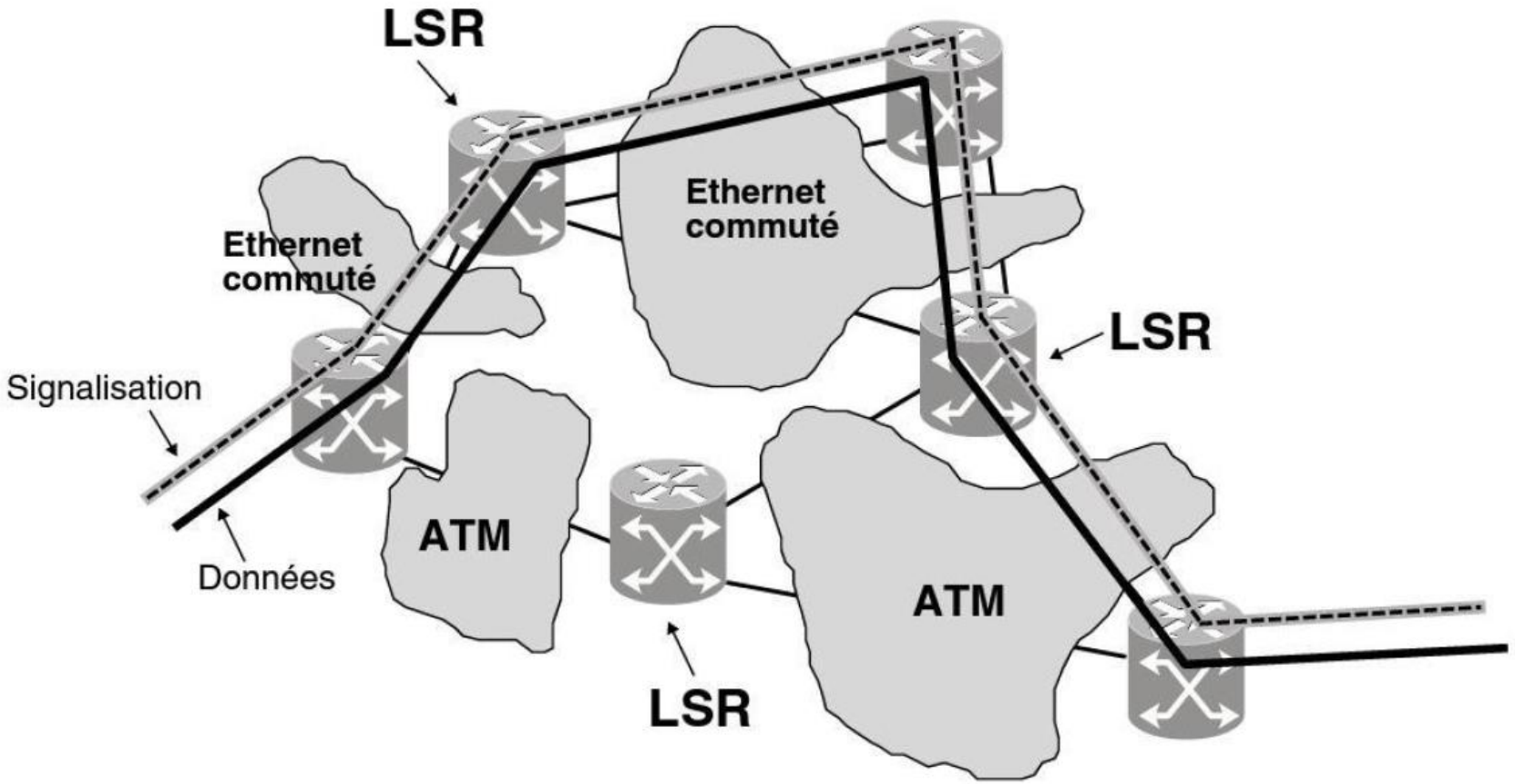
Mise en place des labels dans FR



Mise en place des labels dans PPP et Ethernet

Réseau MPLS de transition

Traversée de plusieurs réseaux spécifiques formant un réseau MPLS afin d'illustrer la transition entre les réseaux de génération ATM et MPLS:



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.

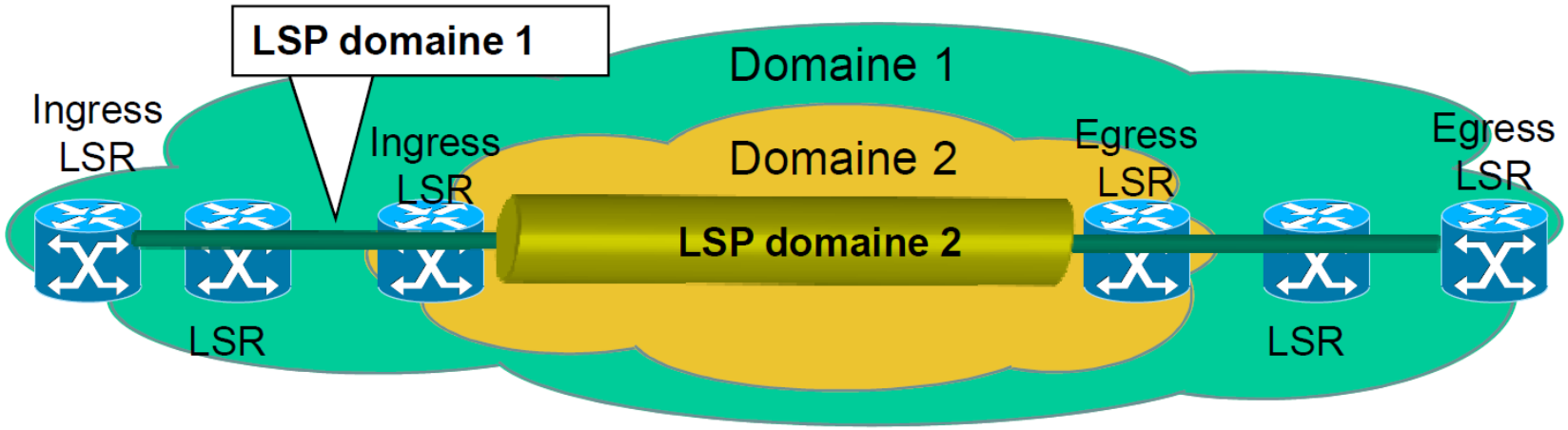
- Ainsi au « cœur » de l'Internet, les adresses réseaux sont sur peu de bits, et au fur-et-à-mesure qu'on approche de la périphérie, on se concentre sur des sous-réseaux de plus en plus fins.

- 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.

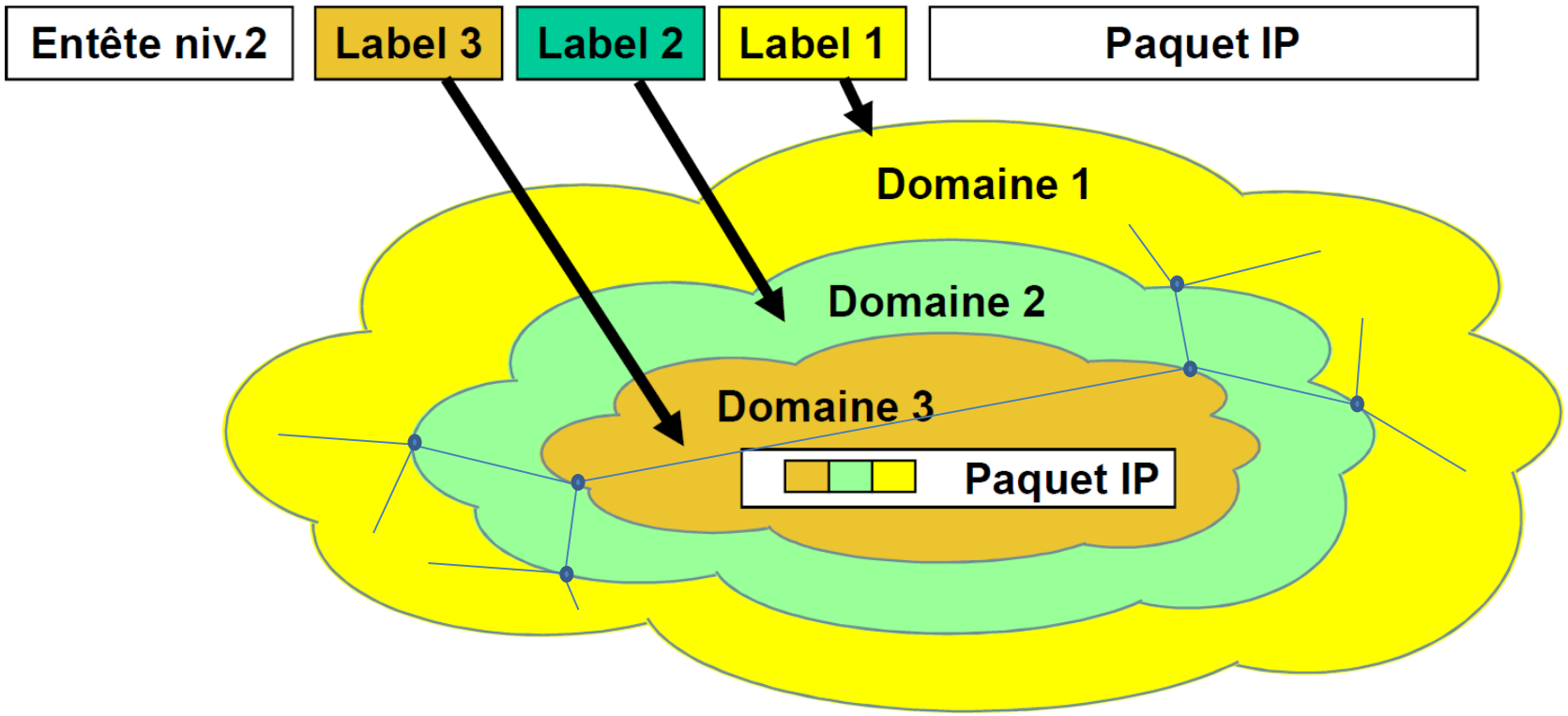
  --> passage à l'échelle de MPLS

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



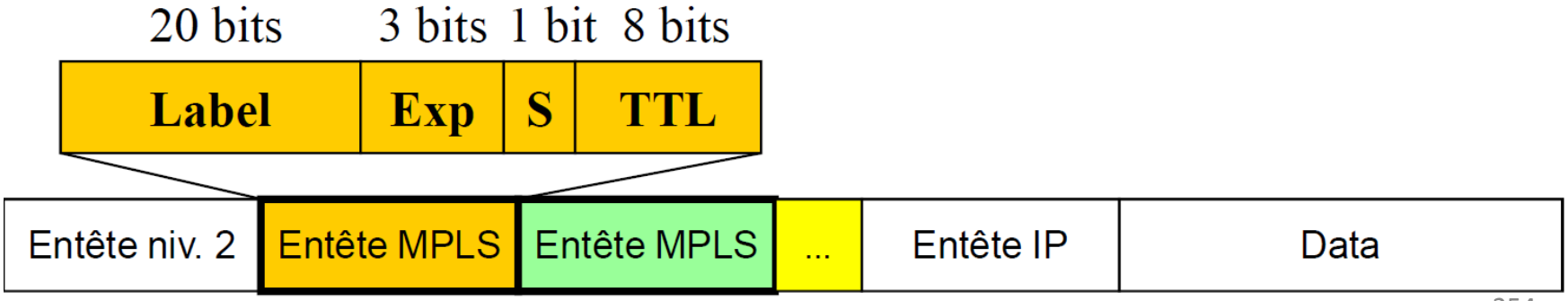
Domaines MPLS hiérarchisés



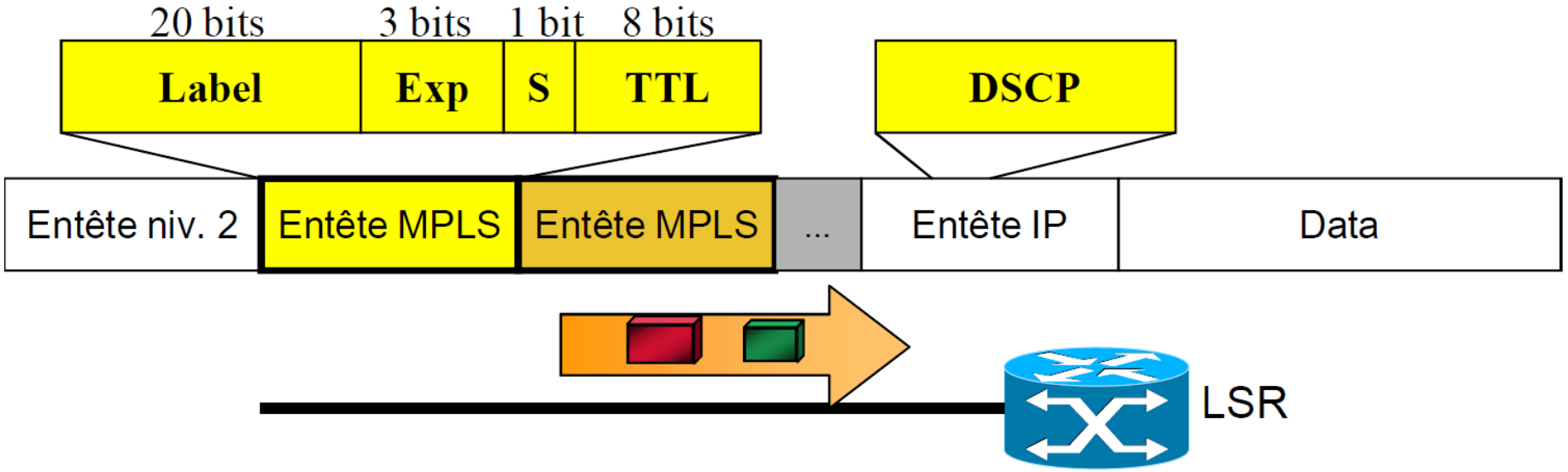
- **Les domaines MPLS peuvent être hiérarchisés grâce aux piles de labels**
  - Pour augmenter les performances au cœur du réseau
  - Pour permettre les interactions entre opérateurs
- **Les piles de labels sont aussi utilisées pour garantir l'étanchéité des VPN**

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

Modes de fonctionnement (et de signalisation)

Mode datagramme pur

- Les LSR se comportent comme des routeurs
- Quand un label n'est pas attribué à la FEC ou avant qu'un label soit attribué

Mode circuit virtuel « mou »

- La création du LSP est déclenchée par l'annonce d'un nouveau préfixe IP par les protocoles de **routage IP**
- Le LSP est créé le long de la meilleure route IP vers le préfixe IP
- Si la meilleure route change, le LSP se reconfigure automatiquement
- Le protocole de signalisation est alors **LDP (Label Distribution Protocol)**

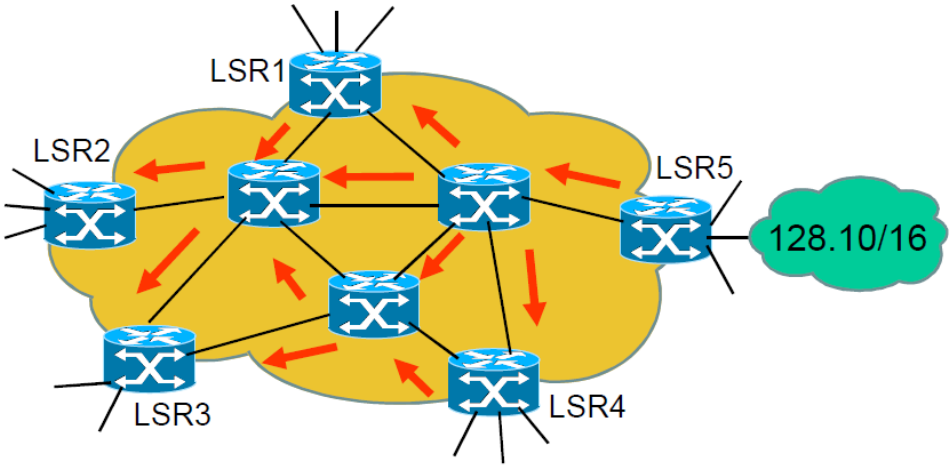
Mode circuit virtuel pur

- La signalisation est déclenchée à l'initiative de l'administrateur
- Le LSP est créé selon une route respectant des contraintes spécifiques
- **Routage explicite** défini par le LSR d'entrée
- Les protocoles de routage sont alors spécifiques (C-OSPF par exemple)
- Le protocole de signalisation est **CR-LDP** ou **RSVP-TE**
- Une réservation de bande passante peut avoir lieu à la création du LSP

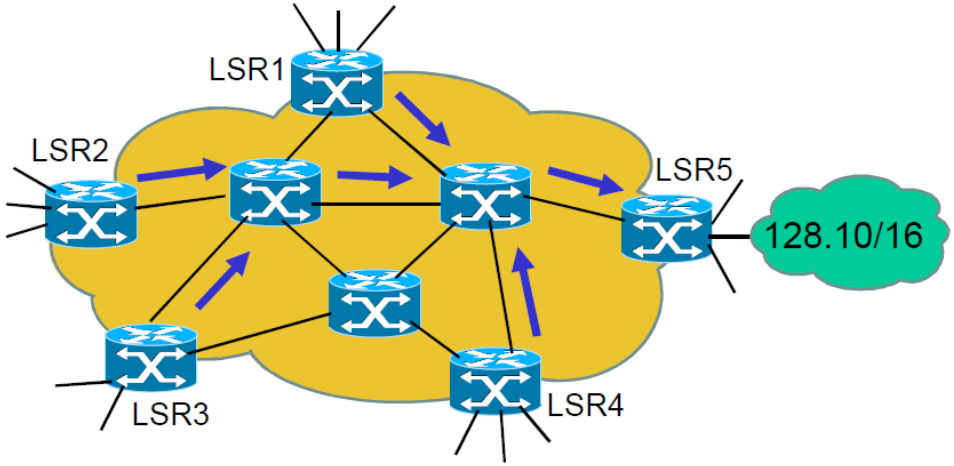
Besoin de signalisation : créer les LIB le long de la route
Le label est attribué par le LSR aval et transmis au LSR amont

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

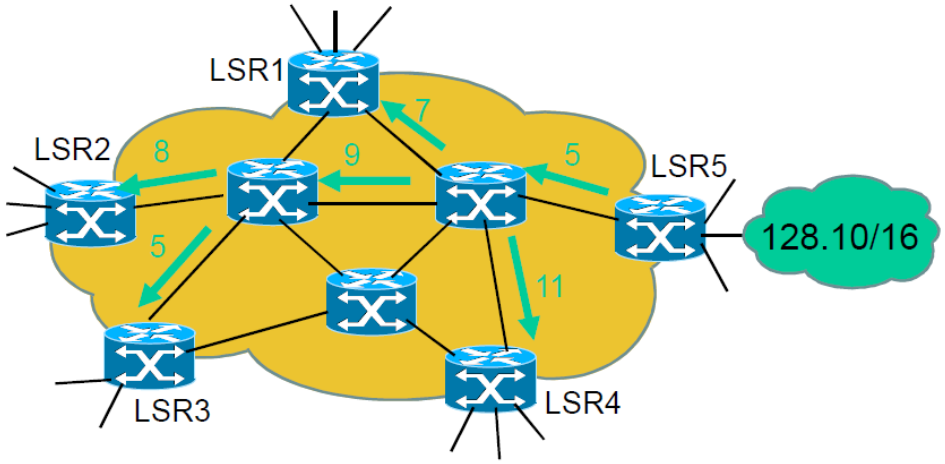


- **Les LSR d'entrée (LSR1, LSR2, LSR3 et LSR4) demandent l'établissement d'un LSP vers 128.10/16**
  - Par des messages *Label Request* de LDP →
  - Ces messages suivent la meilleure route IP vers 128.10/16

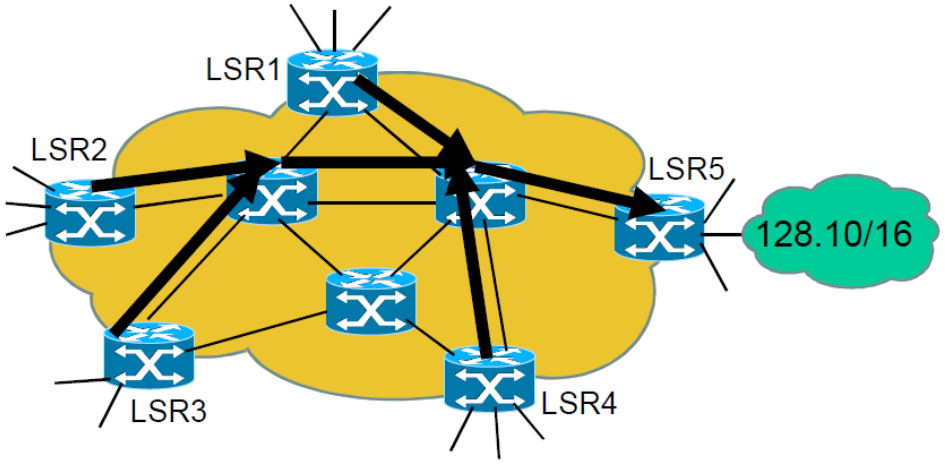


Exemple de signalisation par LDP

- **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**



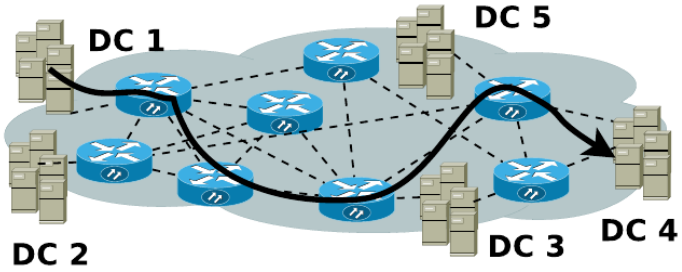
Plan général du cours

I. **Organisation des opérateurs de l'Internet**
II. **TCP et Qualité de service**
III. **Commutation par circuits virtuels (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**

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,.

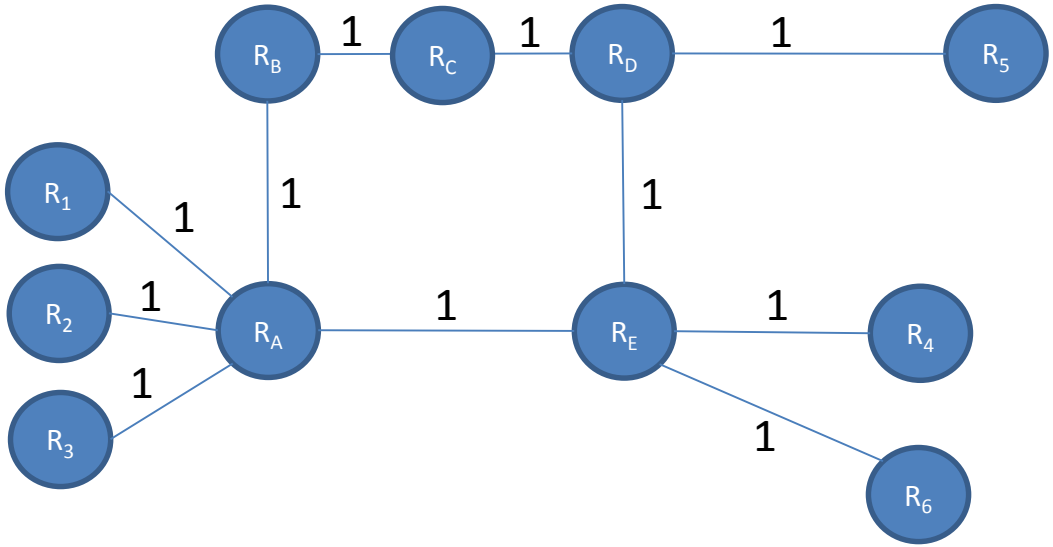
Principe du TE

- 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

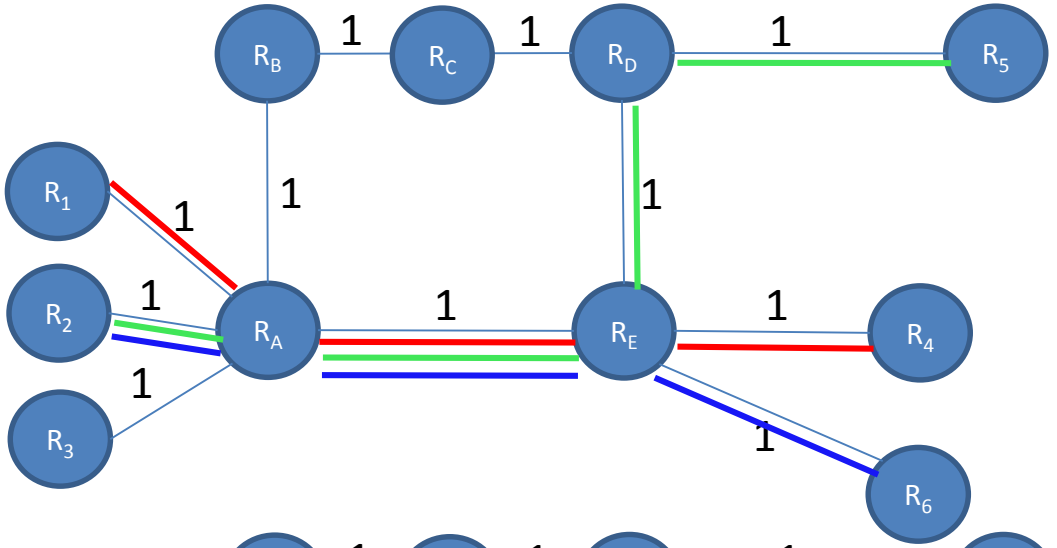
- dans ce réseau :



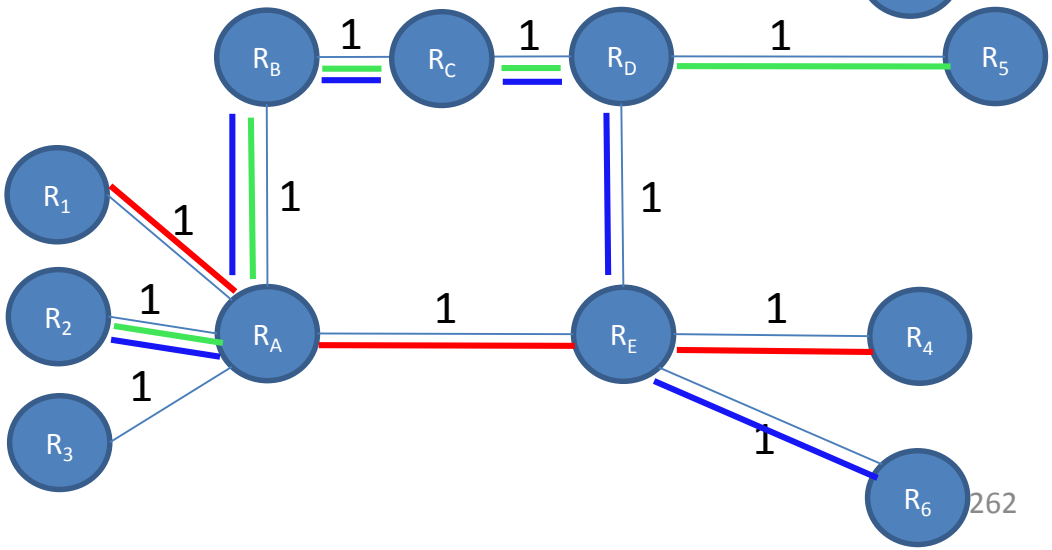
- La matrice de trafic se traduit en :

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

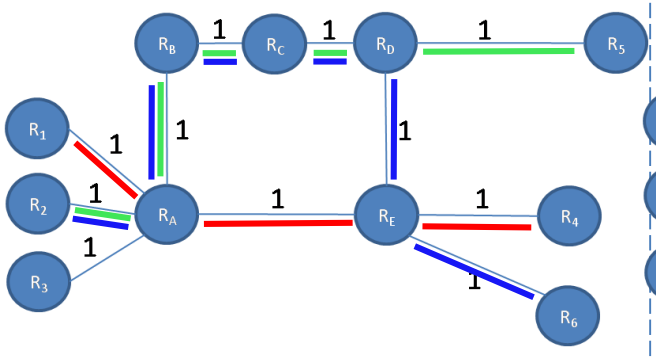
Algorithmes MPLS-TE

- **AutoBP** : MPLS ne contrôle pas le débit (BP) utilisé par le trafic sur un LSP : un LSP peut porter tout débit de trafic indépendemment 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.

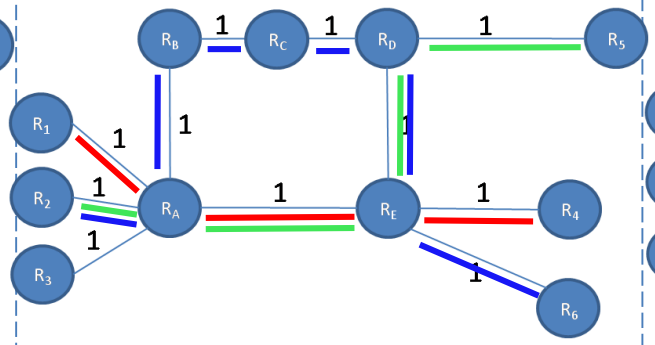
Auto-BP et CSPF

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

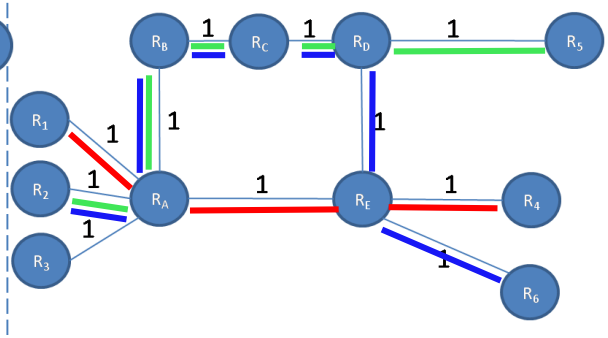
t_1



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

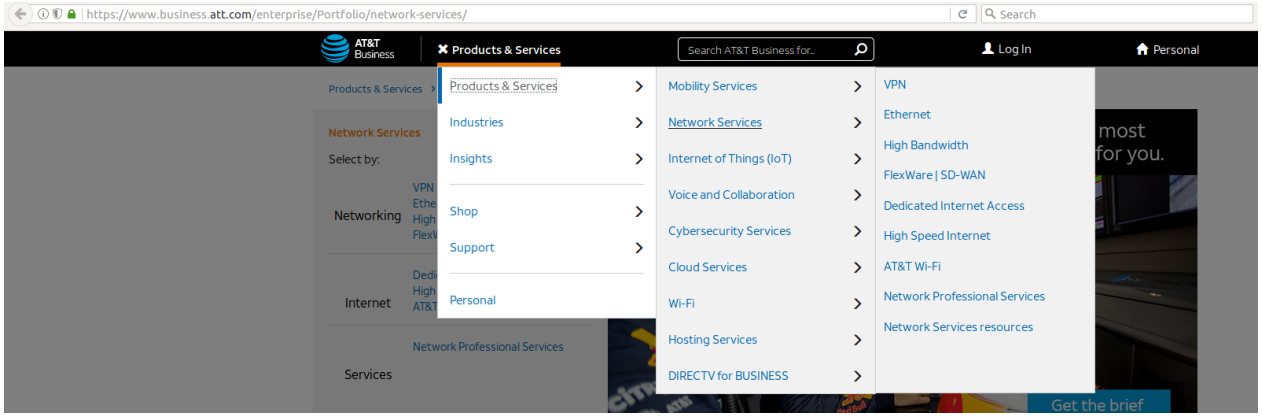
- Principe de MPLS :

- 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**



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/

Préambule: principe des Virtual Private Networks (VPNs)

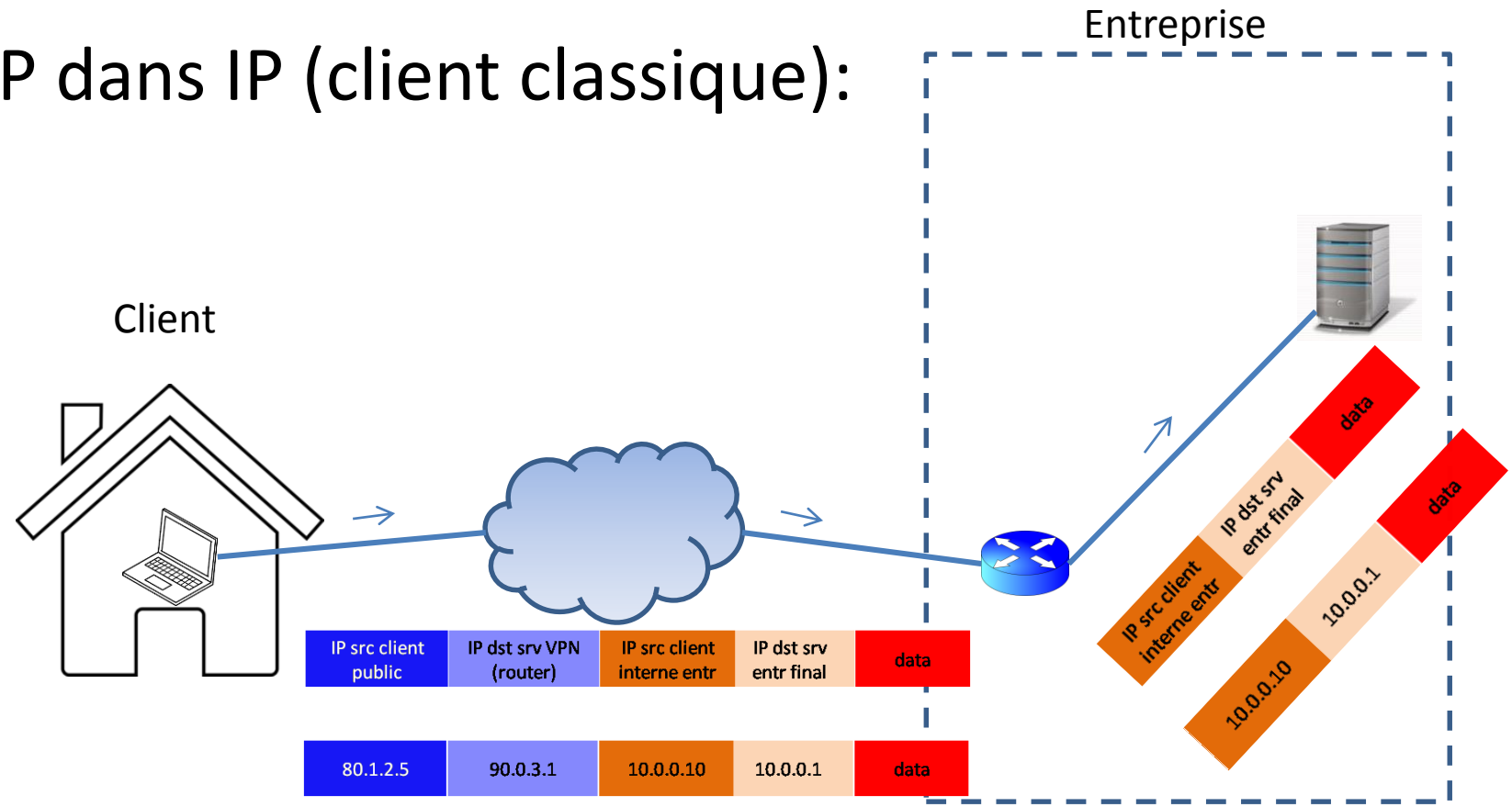
- Principe du ***tunneling*** (tunnelage): un paquet généré à la couche i (1 à 5) est transporté par un protocole de couche j, avec j=i à 5, comme si c'était des données de couche j+1.

- 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

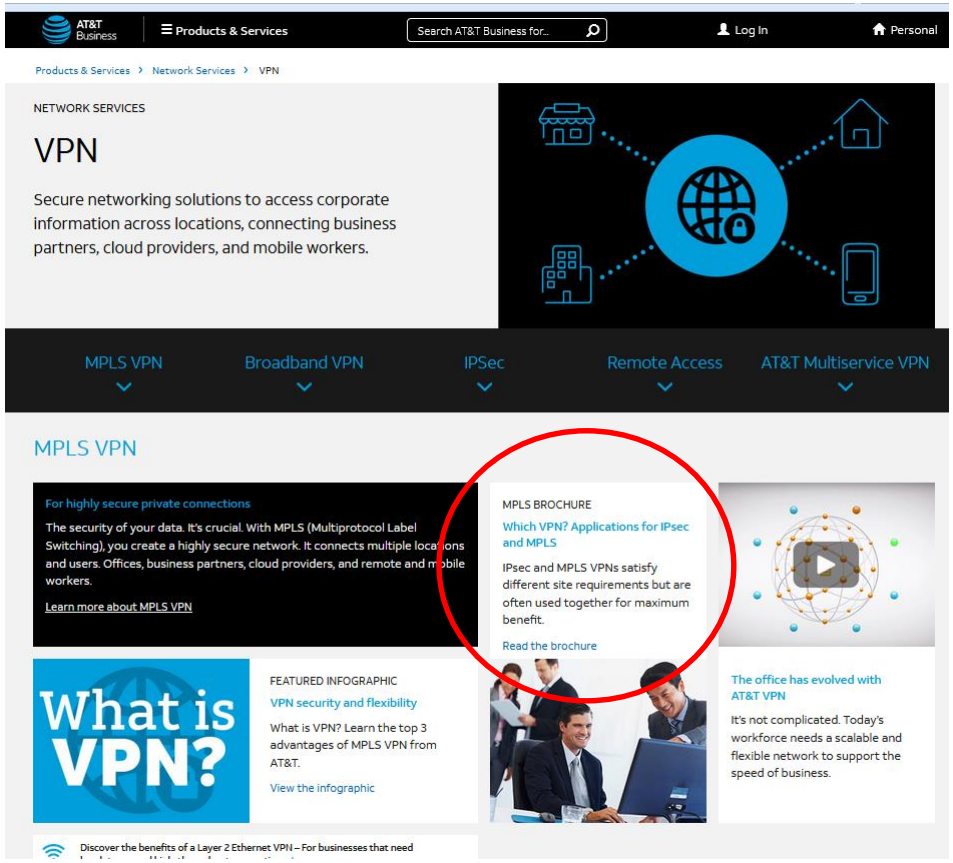
- IP dans IP (client classique):



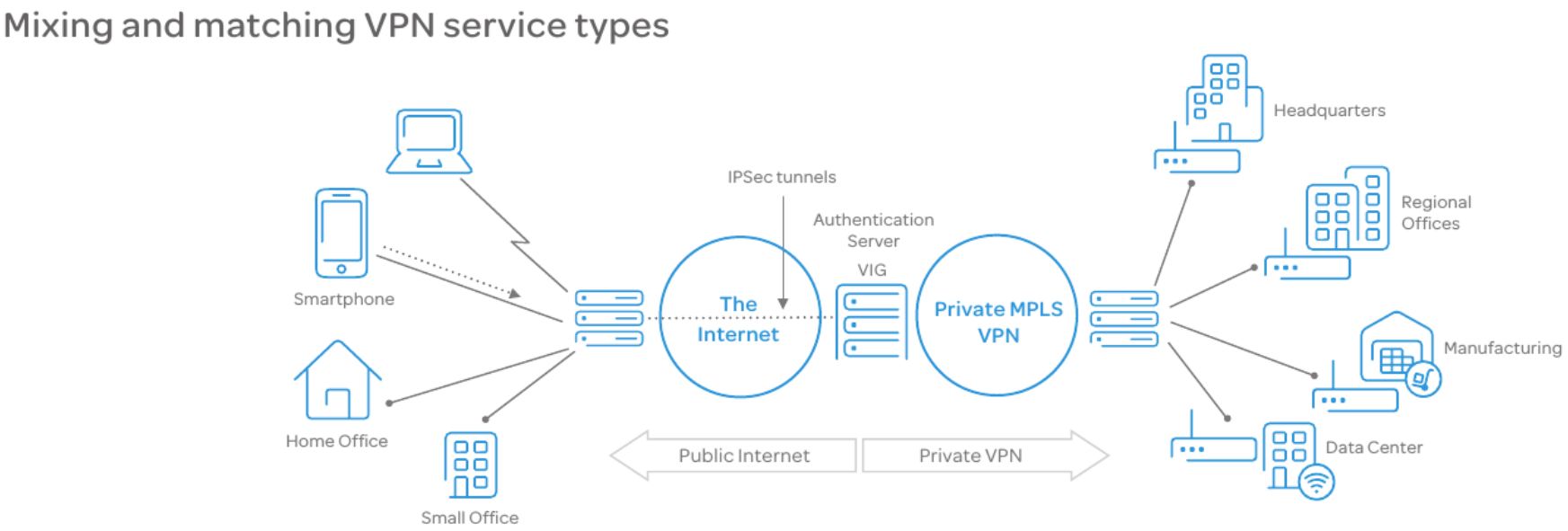
--> Le client peut faire les mêmes opérations que s'il était dans les locaux de l'entreprise.
--> Les ISPs n'ont pas (et ne peuvent pas si IPSec) à lire le paquet destiné au réseau d'entreprise (adr IP privées utilisables).
--> Les ISPs traitent le paquet entreprise comme un payload de couche 4 normal.

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



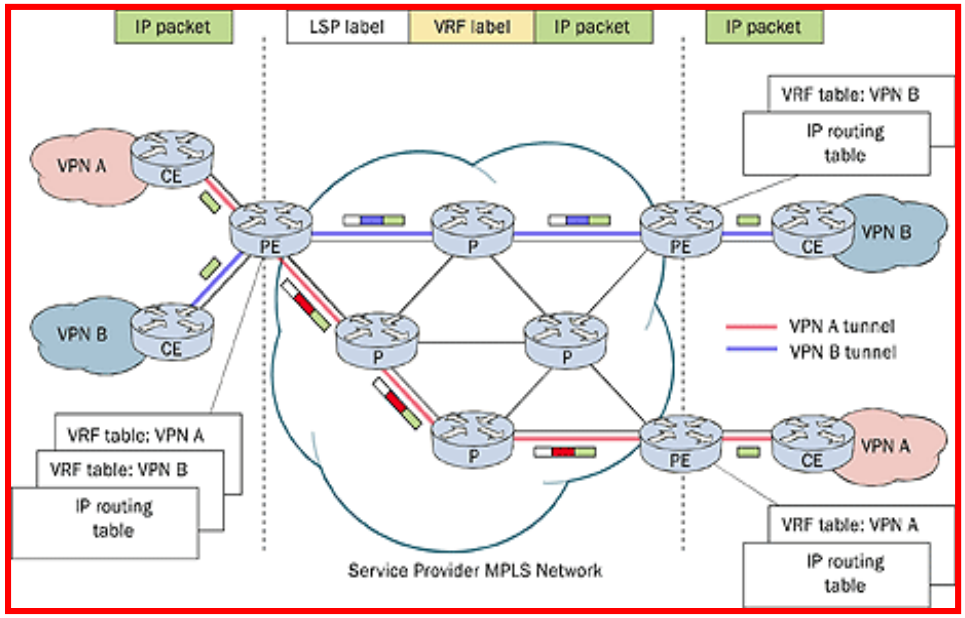
Une application de MPLS: Différents types de VPN basés MPLS

- 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).

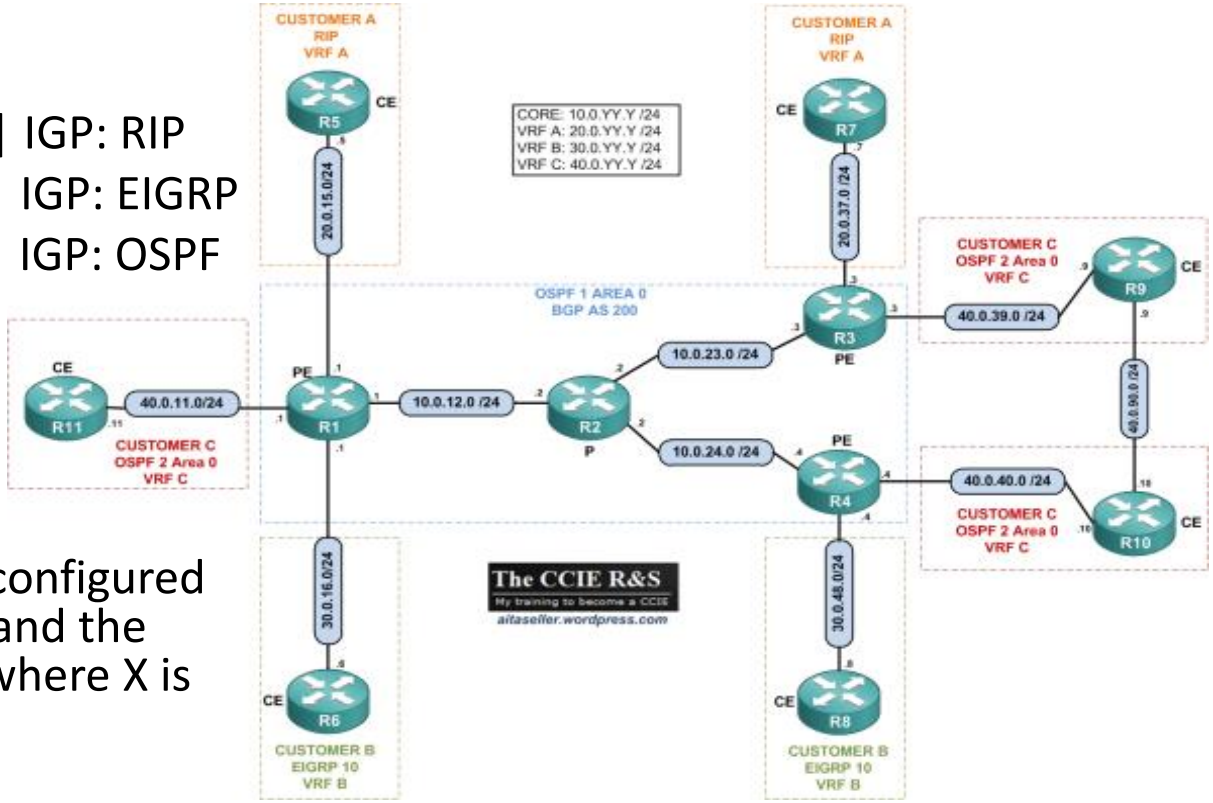
Plan général du cours

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).

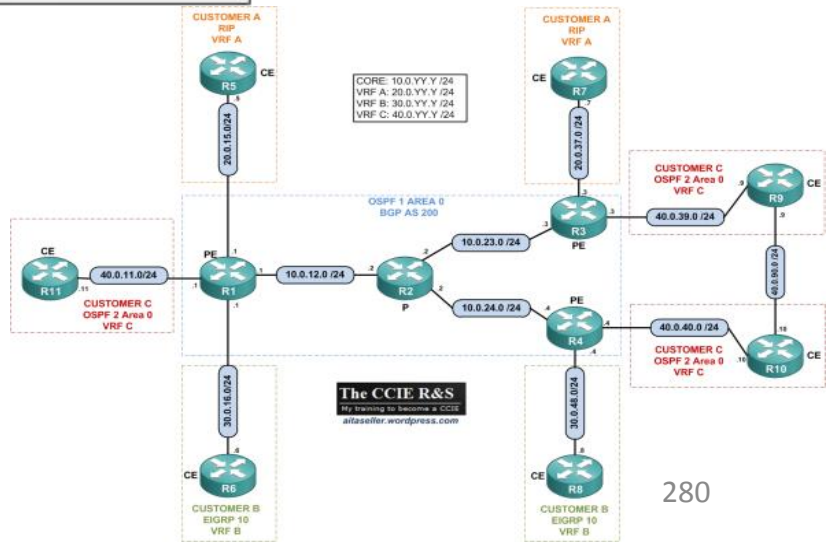
1. Configuration MPLS du coeur de l'ISP

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 :

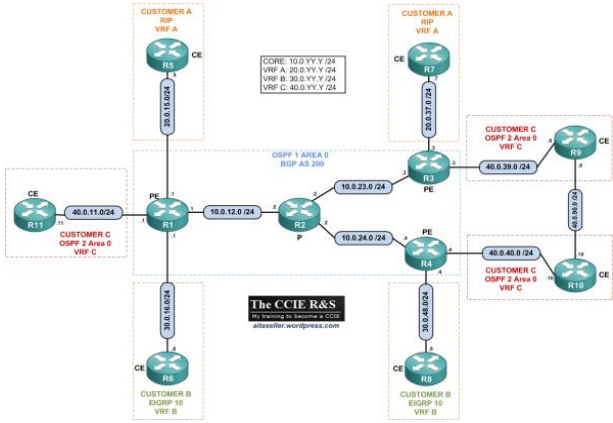
```
R1#sh mpls forwarding-table
Local  Outgoing    Prefix           Bytes tag   Outgoing    Next Hop
tag    tag or VC   or Tunnel Id     switched    interface
16     Pop tag     2.2.2.2/32       0           Fa1/0       10.0.12.2
17     17          3.3.3.3/32       0           Fa1/0       10.0.12.2
18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

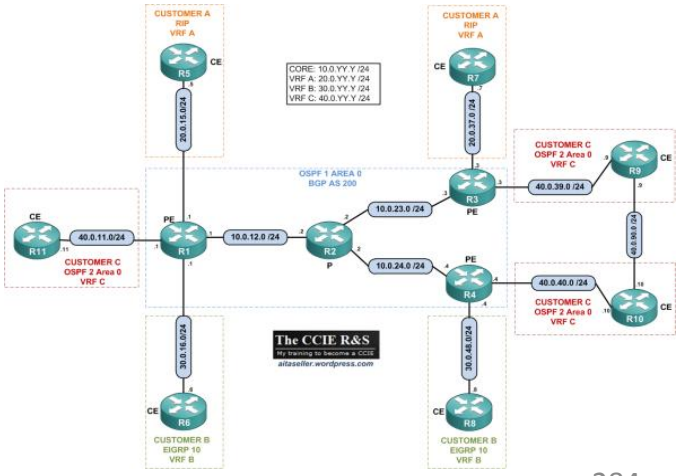
```
R1#sh ip vrf interfaces
Interface              IP-Address      VRF                              Protocol
Fa0/0                  20.0.15.1       VRFA                             up
Fa0/1                  30.0.16.1       VRFB                             up

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 :

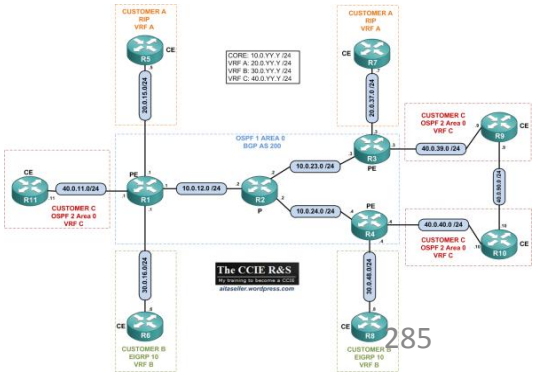
```
R1#sh run | sec rip
router rip
 version 2
 no auto-summary
 !
 address-family ipv4 vrf VRFA
  network 20.0.0.0
  no auto-summary
 exit-address-family
```

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 :

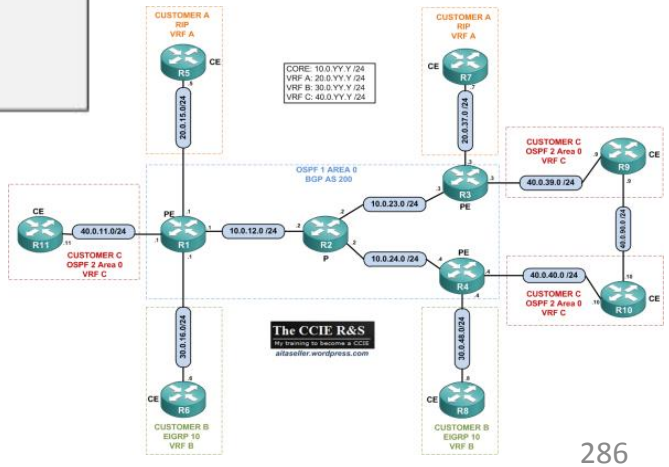
```
R1
router eigrp 1
 no auto-summary
 !
 address-family ipv4 vrf VRFB
  network 30.0.16.1 0.0.0.0
  auto-summary
  autonomous-system 10          -> specify the AS in the routing context
 exit-address-family
```

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.

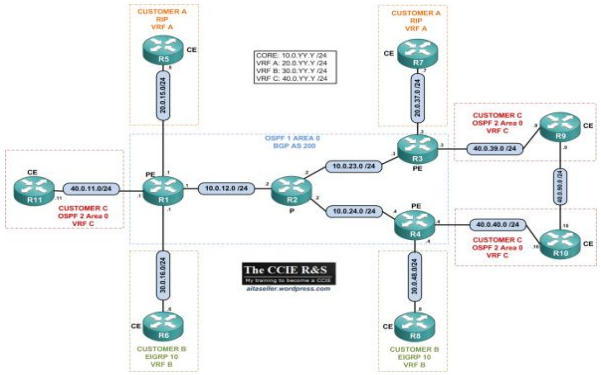
5. Configurer l'échange des préfixes VPNv4 avec MP-BGP

- 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 :

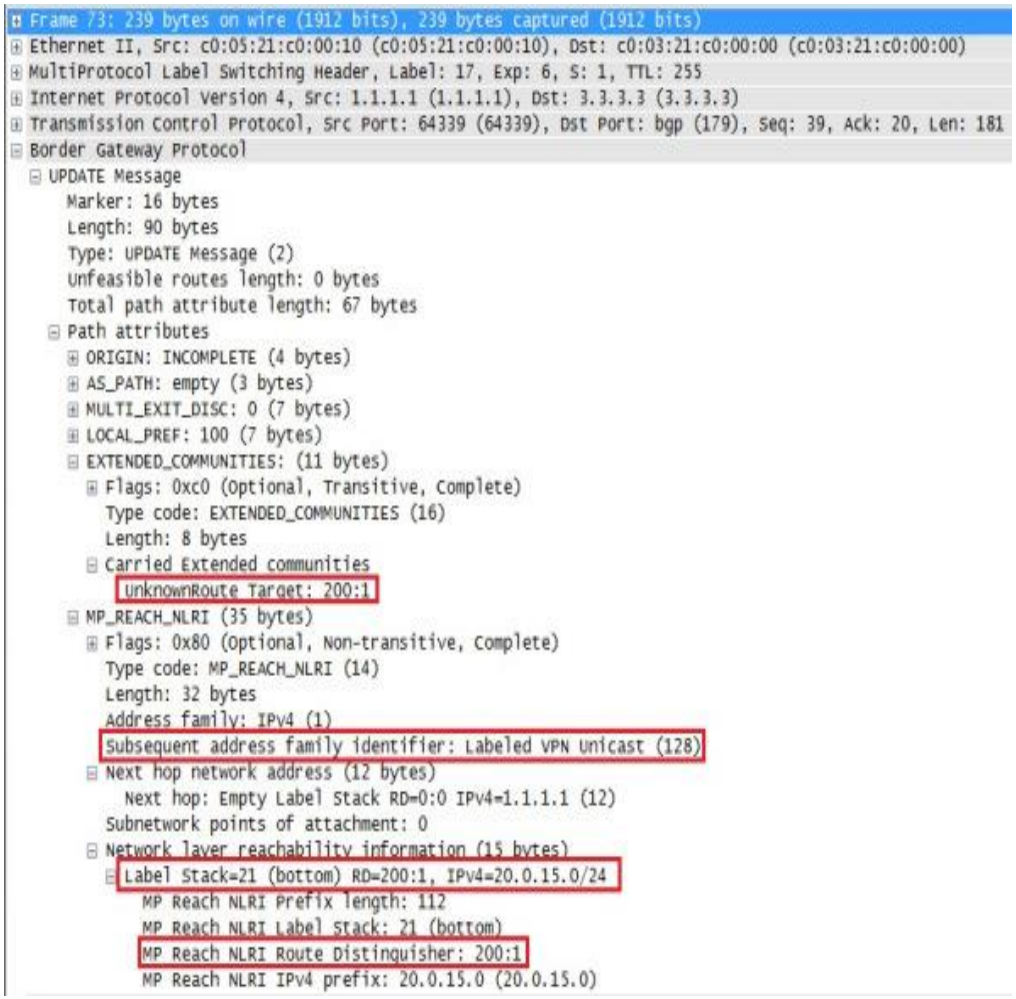
```
R1
router bgp 200
  !
 address-family ipv4 vrf VRFA
  redistribute rip
  no synchronization
 exit-address-family
```

- 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.)

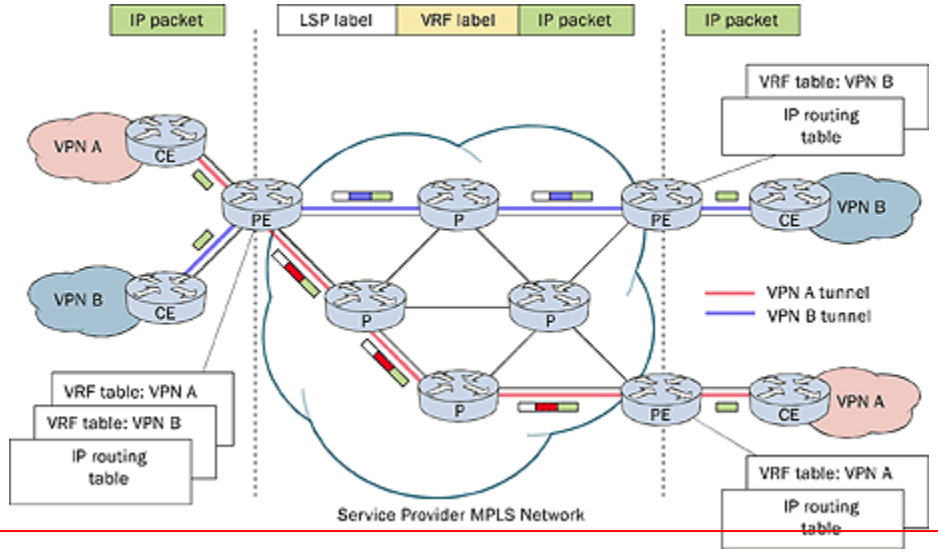
Il semble que les sites du client A peuvent maintenant communiquer. Testons :

```
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.

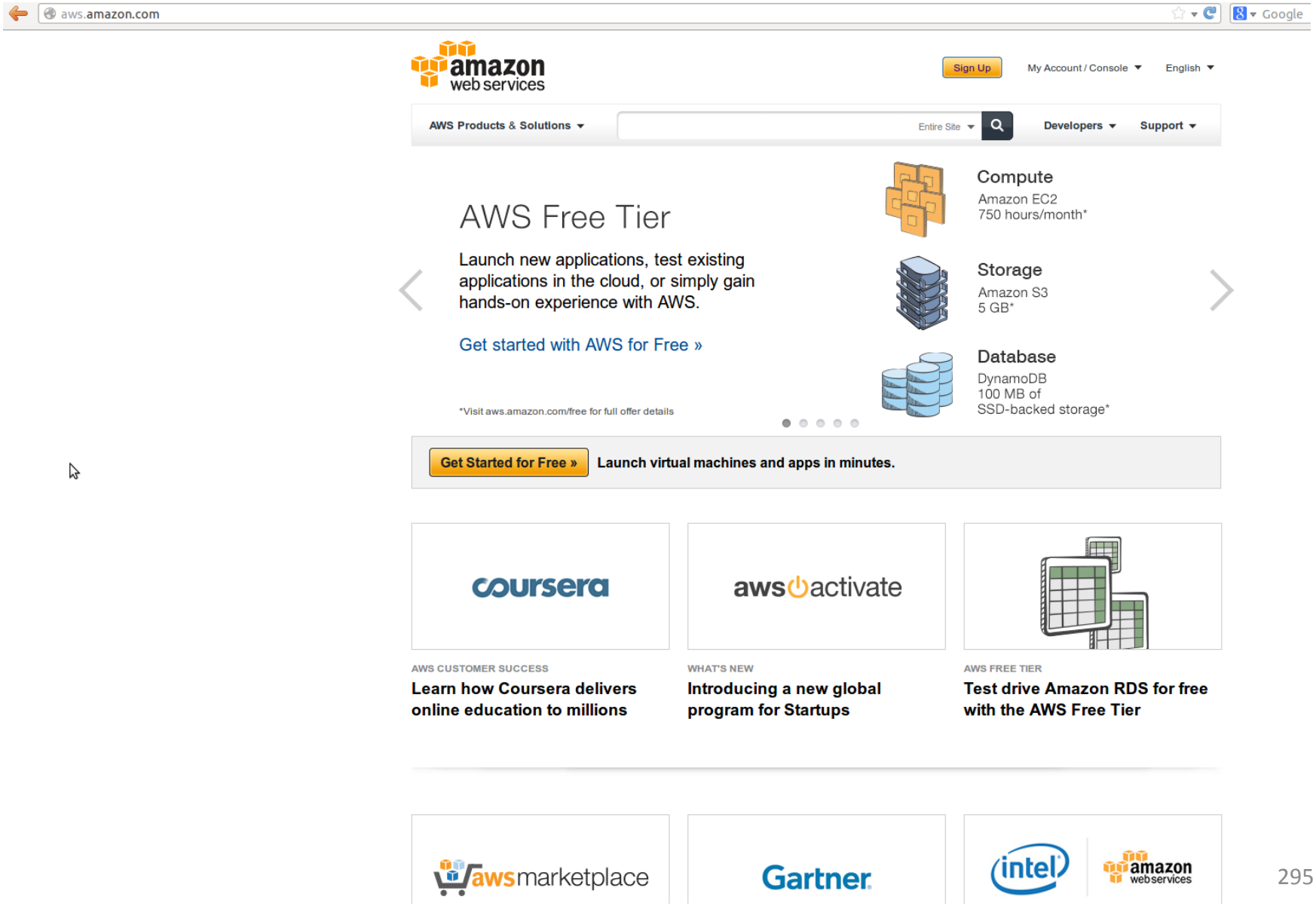
```
R3#sh mpls forwarding-table
Local  Outgoing    Prefix           Bytes tag  Outgoing    Next Hop
tag    tag or VC   or Tunnel Id     switched   interface
22     Untagged    7.7.7.7/32[V]    684        Fa1/0       20.0.37.7
!!!OUTPUT OMITTED!!!
```

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 (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**

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

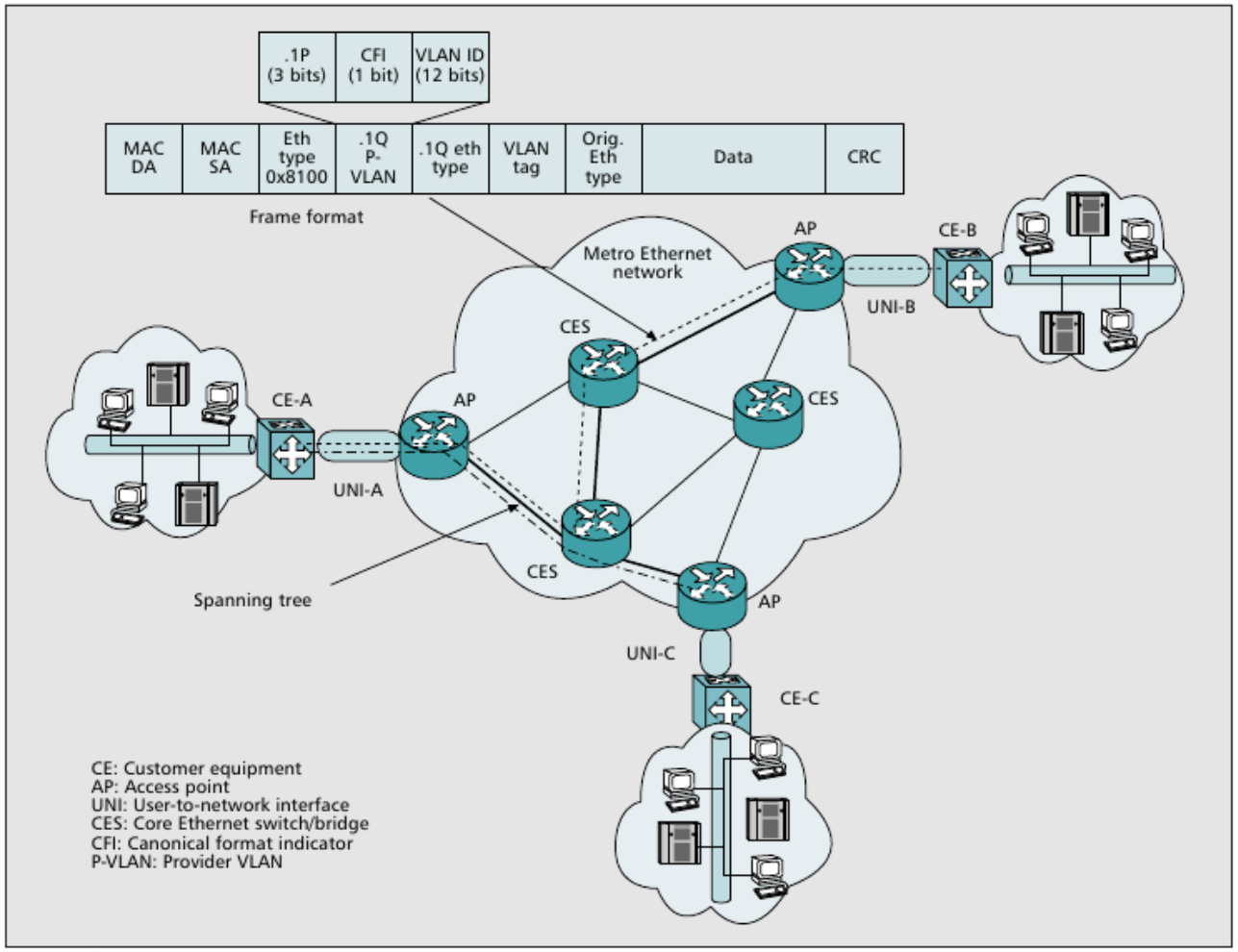


Figure 2. *Network scenario 2: a provider bridged network.*

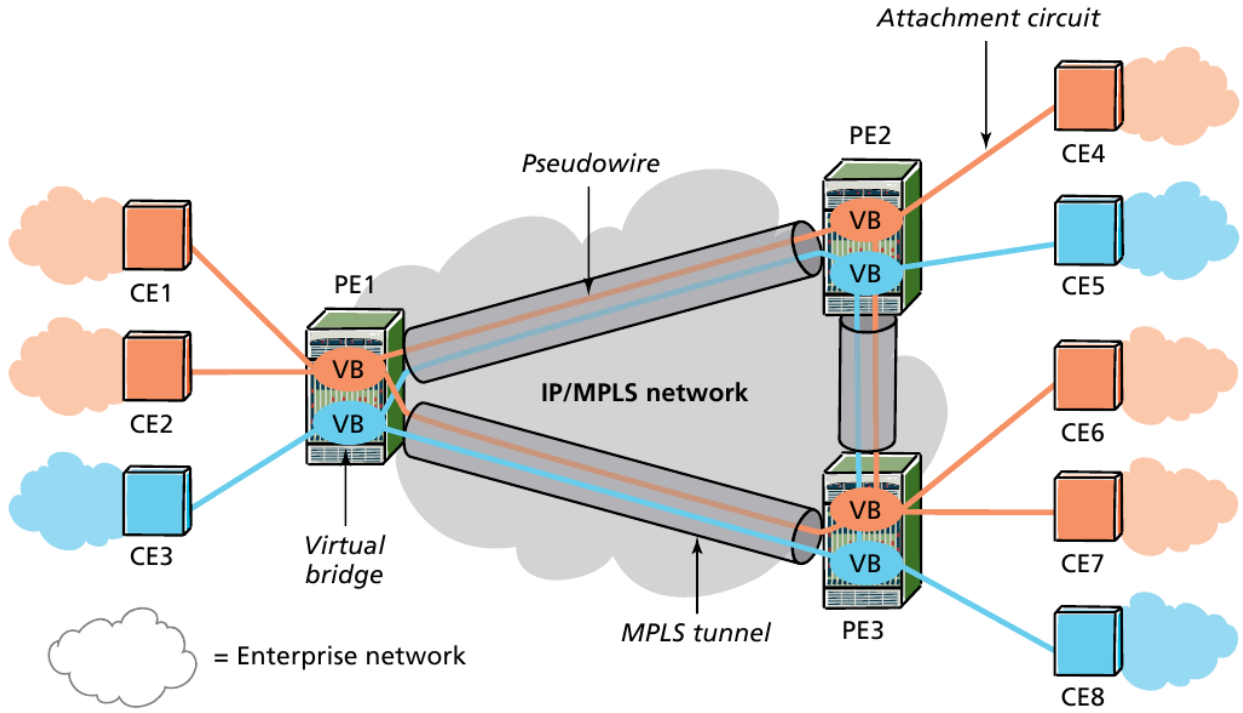
WAN Ethernet basé sur MPLS : Virtual Private LAN Services - VPLS

- 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

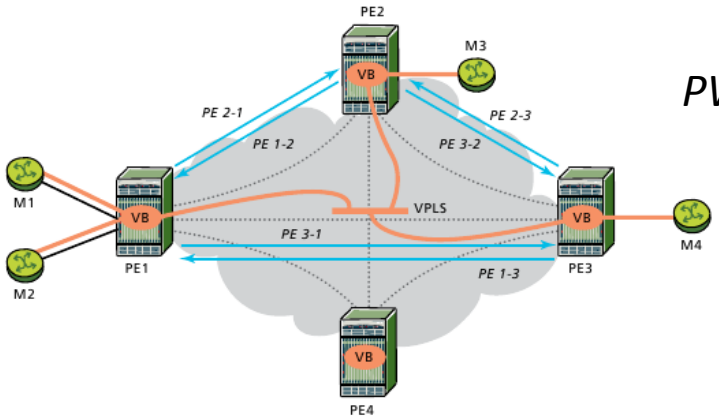
  --> Permet à un PE d'apprendre les @ MAC: quand un PE reçoit une trame Ethernet avec une @MAC source inconnue, le PE sait sur quel PW elle est arrivée, et donc cette adresse sera joignable.

- 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

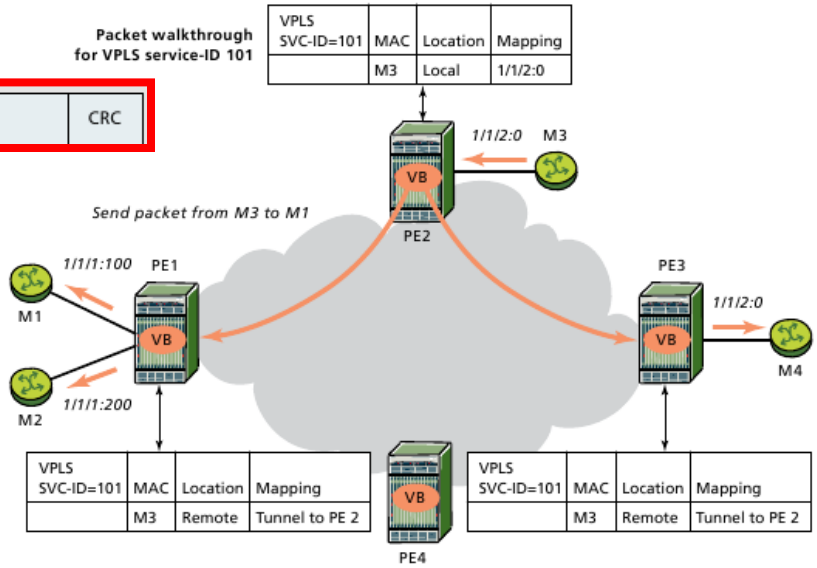


PW signaling

PE1 -> PE2: For SVC-ID 101 use VC label PE 2-1
PE2 -> PE1: For SVC-ID 101 use VC label PE 1-2
PE1 -> PE3: For SVC-ID 101 use VC label PE 3-1
PE3 -> PE1: For SVC-ID 101 use VC label PE 1-3
PE3 -> PE2: For SVC-ID 101 use VC label PE 2-3
PE2 -> PE3: For SVC-ID 101 use VC label PE 3-2

- MAC learning and packet forwarding

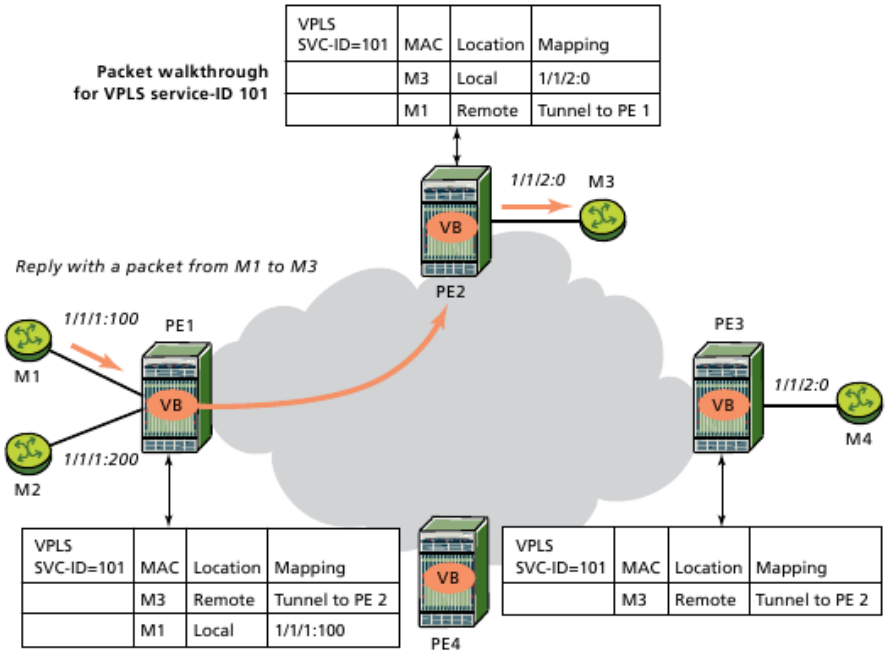
Tunnel label	VC label	MAC DA	MAC SA	VLAN tag	Eth type	Data	CRC



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

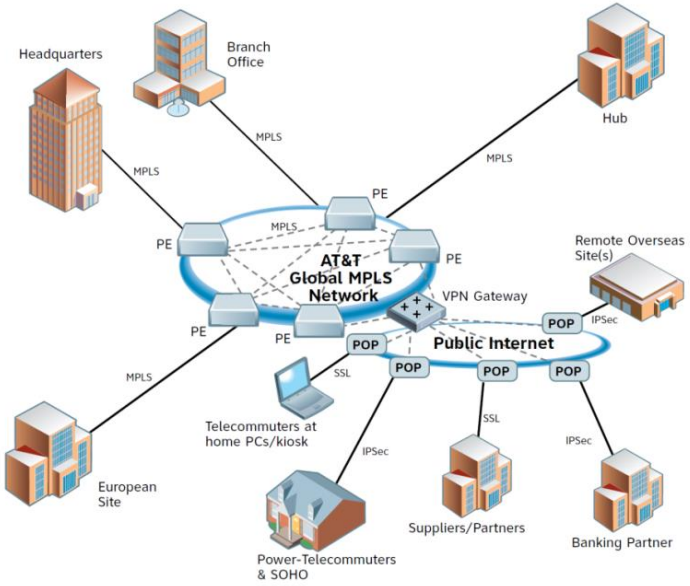
Choisir son service VPN : Ethernet ou IP VPN, ou les 2 ? Critères et bénéfices

- 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

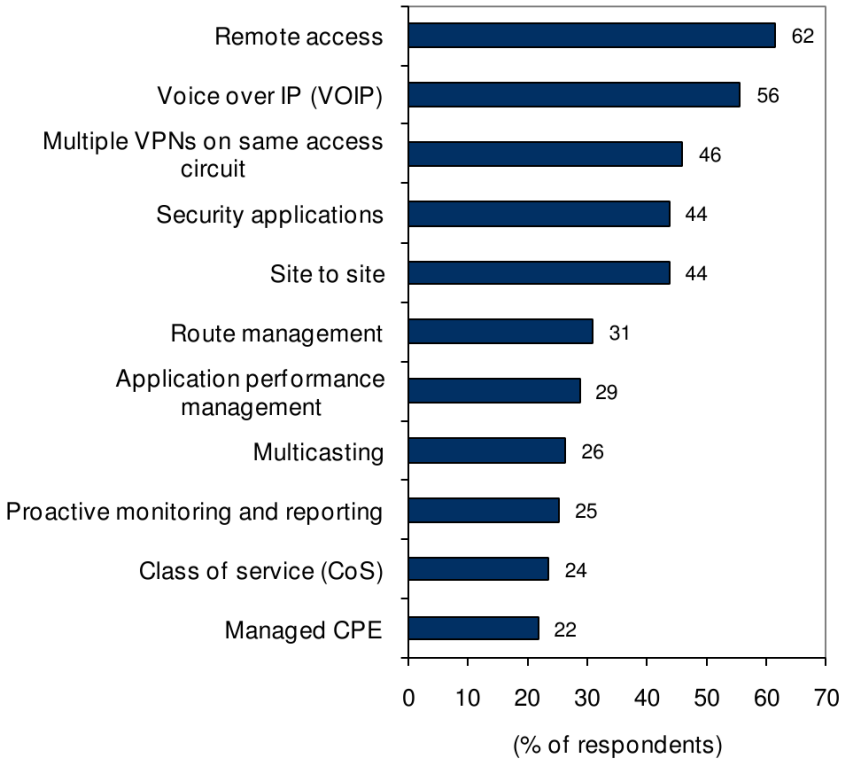
- 62% des entreprises avec IP VPNs. 56% pour de la VoIP : 25% de plus qu'en 2009
- IP VPNs plutôt pour les "petits" sites, et WAN Ethernet pour les gros sites ou les datacenters avec accès fibre.

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



Multiple VPN technology choices working together

Key IP VPN Adoption Criteria



U.S. only n = 174

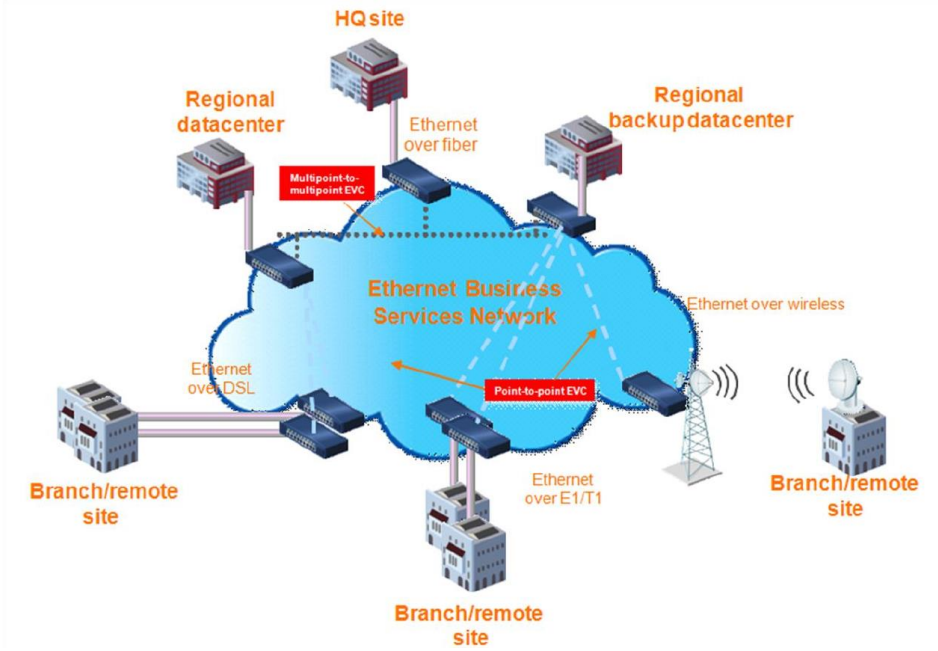
Source: IDC's *WAN Manager Survey*, 2010

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

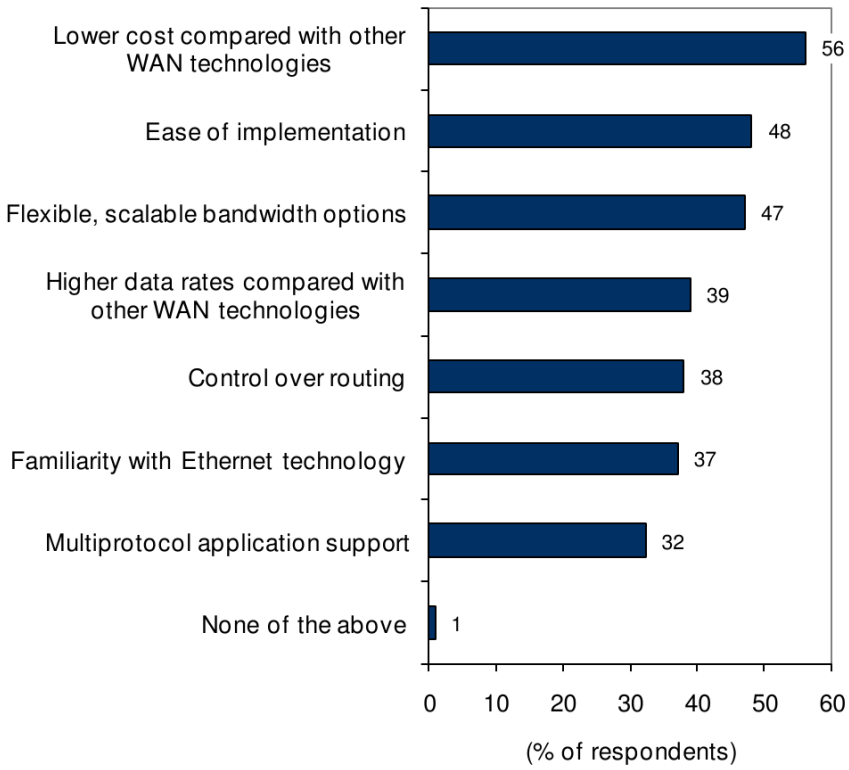
- Adoption de WAN Ethernet par les entreprises pour le coût, la facilité d'implémentation et la familiarité avec Ethernet
- 58% des entr utilisent une solution hybride Ethernet/IP VPN, et augmentation de 33% environ par an.

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

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
V. Accès : Réseaux optiques
   - V.1. Les différents types de multiplexage
   - V.2. La hiérarchie numérique synchrone PDH
   - V.3. La hiérarchie numérique synchrone SDH : réseaux de fibres optiques et redondance pour la fiabilité

VI. Accès : Technologies xDSL

Techniques de multiplexage

- Réseau de télephone : même coût financier pour installer et entretenir une section de câble à large bande passante (haute capacité) qu'une section de câble de faible BP (les coûts correspondent à creuser la tranchée majoritairement).

-> Les compagnies de téléphone ont donc développé des schémas élaborés pour multiplexer plusieurs conversations sur un seul câble physique.

- 2 categories de base : **FDM (Frequency Division Multiplexing)** et **TDM (Time Division Multiplexing)**

- FDM: le spectre de fréquence est divisé en bandes de fréquence, chaque utilisateur ayant l'usage exclusif d'une bande

exemple: radio AM – Le spectre alloué est de 1 MHz, entre 500 et 1500 KHz

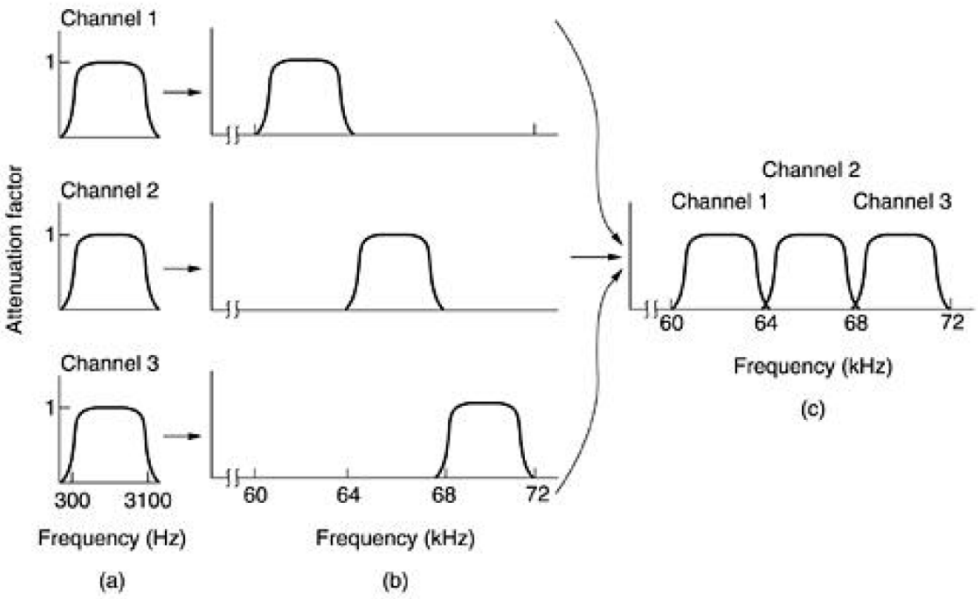
- TDM: les utilisateurs utilisent le câble à tour de rôle (façon *round-robin*), chacun obtenant périodiquement la totalité de la BP pour une courte durée.

- 1. FDM
- 2. FDM appliqué aux fibres optiques (WDM - *wavelength division multiplexing*).
- 3. TDM
- 4. TDM avancé pour les fibres optiques (SONET)

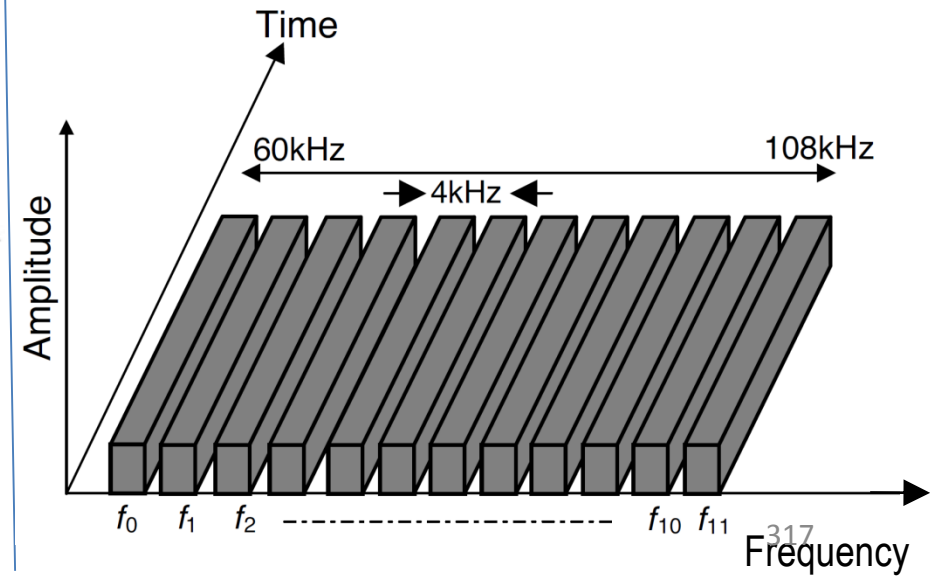
Frequency Division Multiplexing (FDM)

- Figure de gauche: comment 3 canaux téléphoniques sont multiplexés avec FDM
- Le filtres limitent la BP utilisable à 3100 Hz par canal de voix
- 4000 Hz sont alloués à chaque canal pour les garder bien séparés
- Noter que malgré les *guard bands* entre les canaux, il y a du recouvrement entre les canaux adjacents à cause des imperfections des filtres

(a) Spectre initial
(b) Spectre translaté en fréquence
(c) Canaux multiplexés en fréquence



FDM est utilisé en multiplexant 12 abonnés chacun avec 4 KHz de BP dans les fréquences entre 60 et 108 KHz.

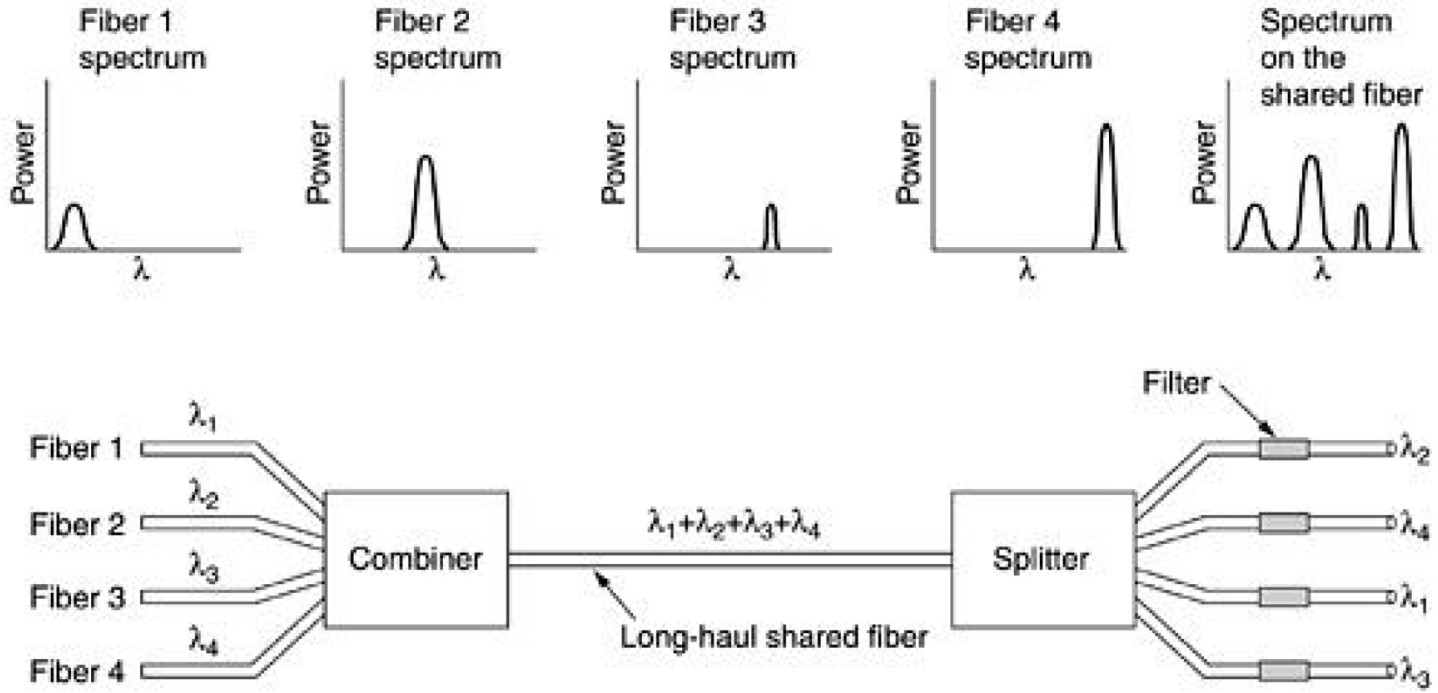


Wavelength Division Multiplexing (WDM)

- Pour les fibres optiques, une variante de FDM est utilisée : **WDM** (**Wavelength Division Multiplexing)**

- Figure: 4 fibres dans un *combiner* optique (chacune avec une longueur d'onde différente)
  Les 4 sont combinées dans une seule fibre partagée pour transmission.
  A la sortie, le signal est séparé sur plusieurs fibres comme à l'entrée (avec un filtre sur les longueurs d'onde).

- **La seule différence avec FDM électrique est que dans un système optique le traitement est passif et fiable.**



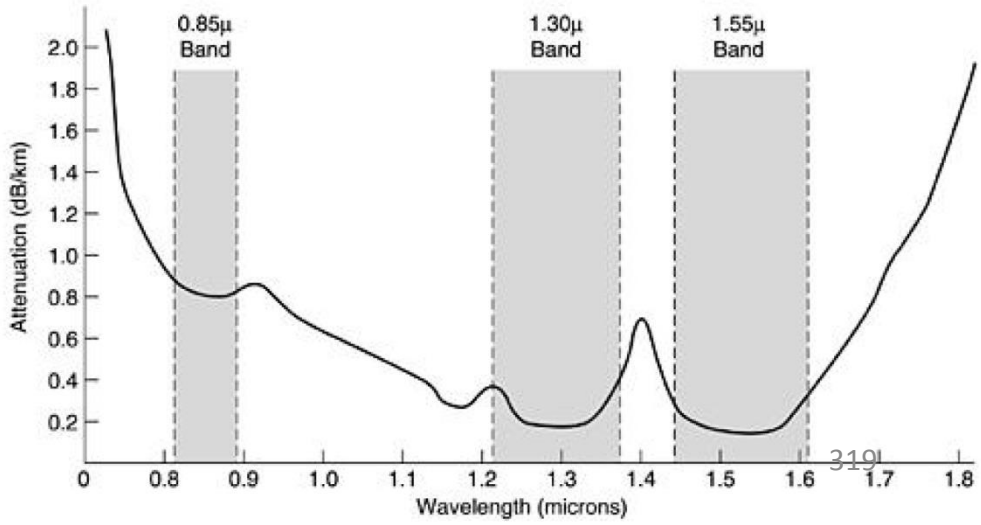
Wavelength Division Multiplexing (WDM)

- WDM inventé dans les années 1990. Fibres mono-mode et multi-mode.
- En 2000, fibres avec 96 canaux de 10 Gbps, soit 960 Gbps au total.

- Quand le nombre de canaux est très grand et les longueurs d'onde très proches (ex : 0.1 nm), c'est du **DWDM (Dense WDM).**
- Depuis que la BP d'une seule fibre est de 25000 GHz, il y a théoriquement la place pour 2500 canaux de 10 Gbps à 1 bit/Hz (et des plus hauts débit sont aussi possibles).

- Une autre avancée réside dans les amplificateurs optiques :
    - Avant, tous les 100 km il était nécessaire de séparer les canaux et de convertir chacun en signal électrique, l'amplifier séparément, le reconvertir en optiques et recombiner.
    - Maintenant, **tous les amplificateurs optiques peuvent regénérer le signal combiné qu'une fois tous les 1000 km sans conversions optiques/électriques.**

- **Switches optiques** basés sur:
  Interféromètres de Fabry-Perot
  ou Mach-Zehnder.



Time Division Multiplexing (TDM)

- FDM: toujours utilisé sur fils de cuivre ou canaux micro-ondes, mais nécessite circuit analogiques et n'est pas adaptés au traitement par les équipement filaires numériques.

- TDM: géré entièrement par de l'électronique numérique, donc plus répandu de loin.

- Cependant, les boucles locales produisent des signaux analogiques -> conversion A/N nécessaire au CO où toutes les lignes d'abonnés sont combinées sur des câbles haute capacité en sortie du CO

- Comment les multiples signaux analogiques de voix sont numérisés et combinés ?:

- Signaux analogiques numérisés au CO en utilisant **PCM (Pulse Code Modulation) :**

- 1 échantillon toutes les 125µs <--> 8000 échantillons par seconde
- 8 bits par échantillon (loi A ou loi µ pour la quantification)

--> Canal de voix à 64Kbits/s

- Le théorème de Shannon-Nyquist dit qu'un échantillonage à 8000Hz est suffisant pour capturer toute l'information d'un canal téléphonique de BP 4KHz. A un taux d'échantillonage plus petit, de l'info serait perdue; à un taux plus grand, pas d'info supplémentaire serait aquise.

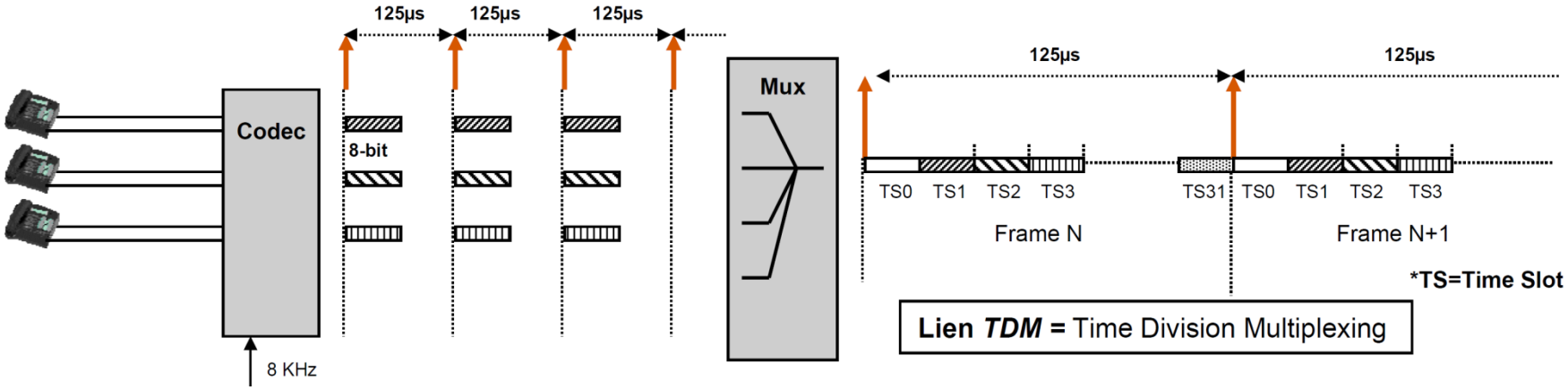
- En conséquence :

Toute trame de couche physique dans un réseau d'opérateur dure 125 µsec.

Principe du réseau numérique synchrone

Tous les principes du réseau numérique synchrone trouvent leur fondement dans les besoins d'acheminement de la voix *(c'est le service attendu du réseau des opérateurs initialement !)*

A l'établissement d'une communication, un circuit synchrone à 64 Kbit/s est établi de bout en bout dans le réseau.



- Chaque circuit se voit allouer une partie fixe de la ressource du réseau pendant toute la durée de la communication.
- L'information est insérée de manière brute dans le circuit physique (échantillons de voix).

-> Réseau extrêmement simple dans son principe, pas de ''couches'' à proprement parler

--> Comment on a utilisé et on utilise ce TDM dans les réseaux de communication de données à haut débit ?

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
V. Accès : Réseaux optiques
   - V.1. Les différents types de multiplexage
   - V.2. La hiérarchie numérique synchrone PDH
   - V.3. La hiérarchie numérique synchrone SDH : réseaux de fibres optiques et redondance pour la fiabilité

VI. Accès : Technologies xDSL

PDH – Plesiochronous Digital Hierarchy

- Technologie de transport développée dans les 60's pour déployer les réseaux téléphoniques numériques nationaux et internationaux basés sur le multiplexage et la commutation de circuits synchrones (TDM).

- Plesiochronous: plesio = presque … presque synchrone, les différents équipements sont synchronisés sur des horloges différentes mais de fréquence et stabilité identiques

- Initialement réservée au transport de la voix, puis dans les 70's s'ouvre au transport de data (RNIS, X.25)

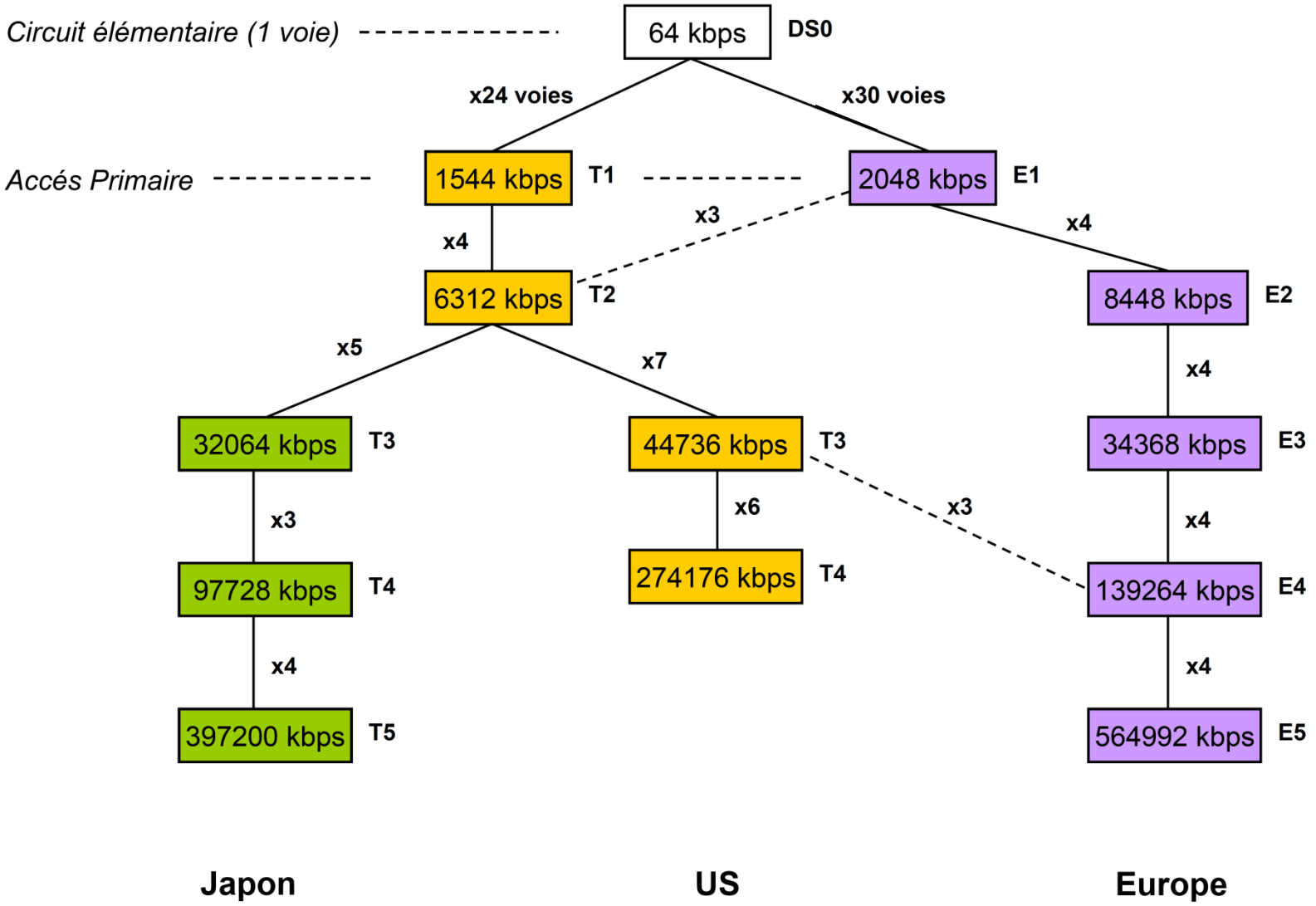
- **1er débit standard T1 adopté par les US: 24 voies a 64 Kbps multiplexées sur un support à 1544 Kbps**
- **suivi par le E1 en Europe: 30 canaux multiplexés sur un support à 2048 Kbps**

- Définit une **hiérarchie standard de débits : chaque niveau de la hiérarchie constituant un multiplex des niveaux inferieurs**.

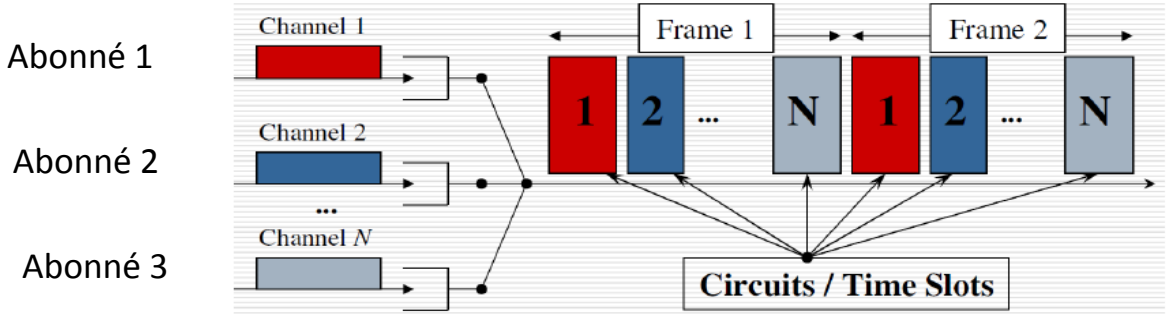
- Le principe d'une trame de base de 125µs (8KHz) reste le point commun à tous les niveaux de la hiérarchie.

- Pour des raisons historiques liées aux infrastructure existantes, la hiérarchie diffère entre Europe, US et Japon.

PDH – la hiérarchie



TDM synchrone dans le réseau téléphonique : partage de débit



- Un canal correspond au signal sonore numérisé provenant d'un abonné : un échantillon de 8 bits est produit toutes les Te=1/Fe=1/8000=125µs.
- Ceci correspond à un débit par canal de 64Kbps.

- **En sortie de CO, agréger N canaux signifie émettre N échantillons en 125µs, donc 8N bits en 125µs.**

-> le débit en sortie doit donc être ≥ N.64Kbps

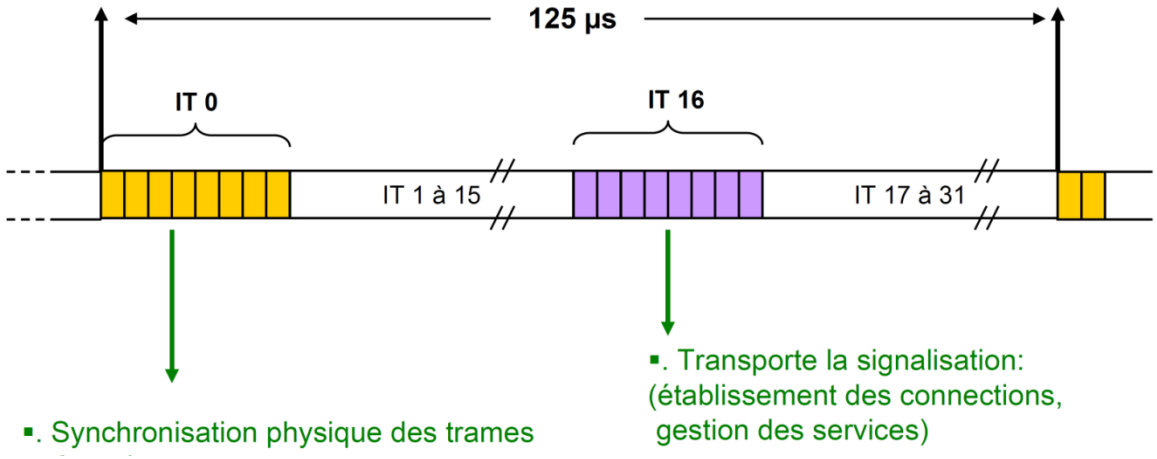
- La durée entre 2 échantillons du même canal ne change jamais (=), le débit de la ligne de sortie détermine le temps mis pour envoyer les 8 bits d'un éch, et donc le nombre d'éch possibles en 125µs, donc de canaux différents qui peuvent être *multiplexés*.

Trame de base E1

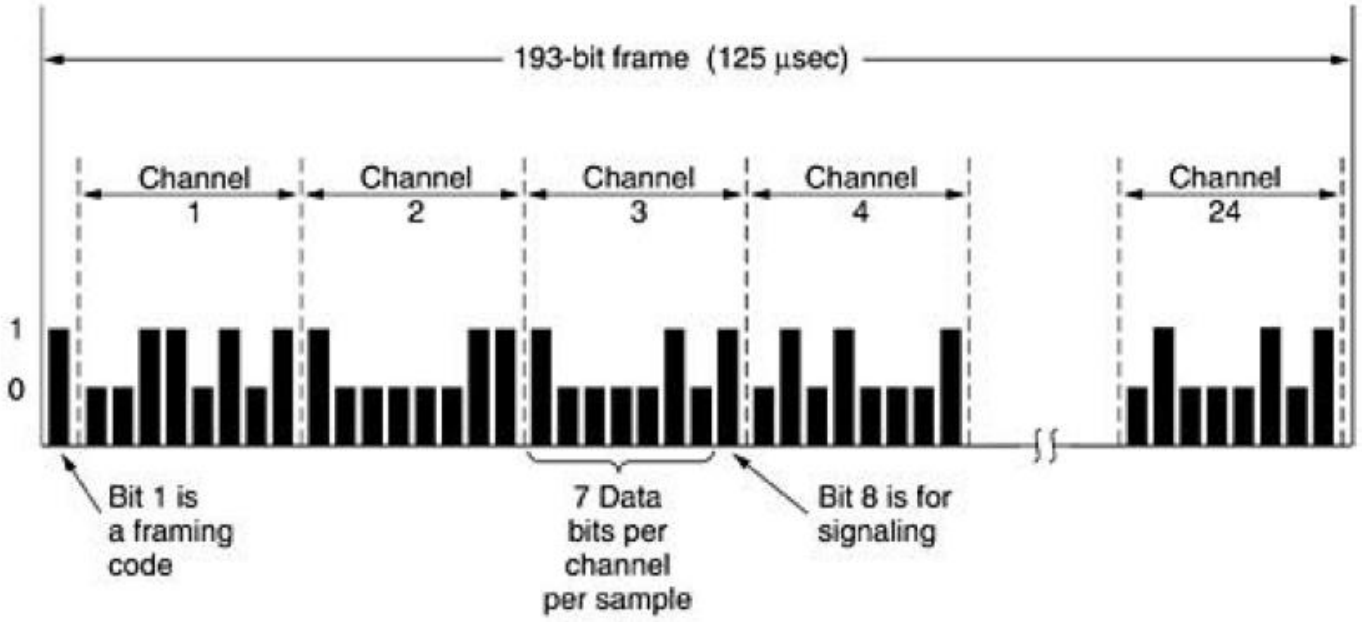
- CCITT -> multiplex E1 à 2048 Kbps : 32 échantillons de 8 bits combinés dans la trame de base durant 125*µsec*.

- 30 des canaux sont utilisés pour porter de l'information et 2 pour la signalisation:
  - 1 canal pour la signalisation des canaux
  - 1 canal pour la synchronisation de trame

Standard ITU-T G.704



Trame T1

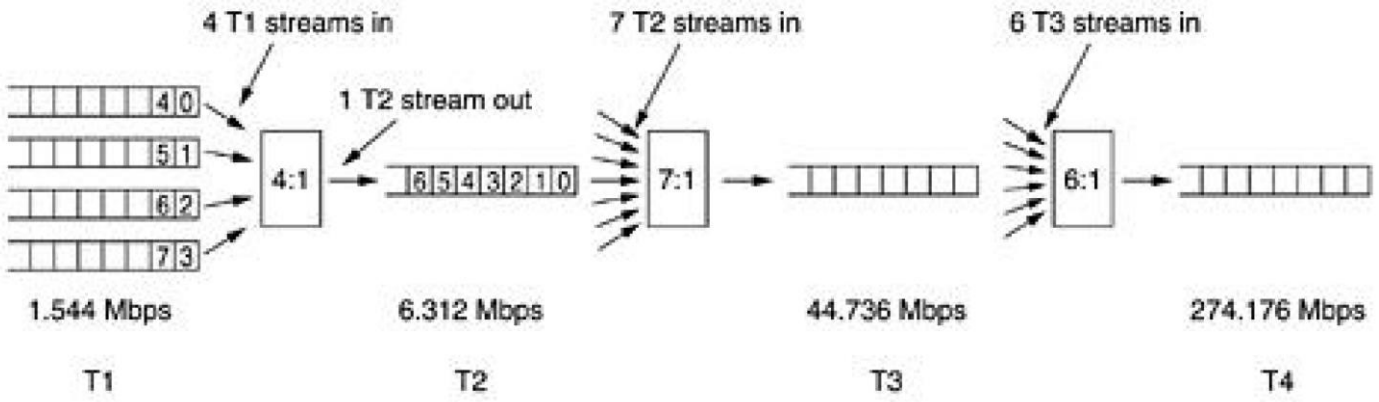

193-bit frame (125 μsec)
Channel 1
Channel 2
Channel 3
Channel 4
Channel 24
1
0
Bit 1 is a framing code
7 Data bits per channel per sample
Bit 8 is for signaling

PDH – Multiplexage de flux T1 dans des multiplexes d'ordres élevés

- TDM permet d'avoir plusieurs flux T1 dans des multiplexes d'ordres supérieurs.

- T1 et T3 sont largement utilisés par les clients, tandis que T2 et T4 sont utilisés seulement dans le coeur du réseau.

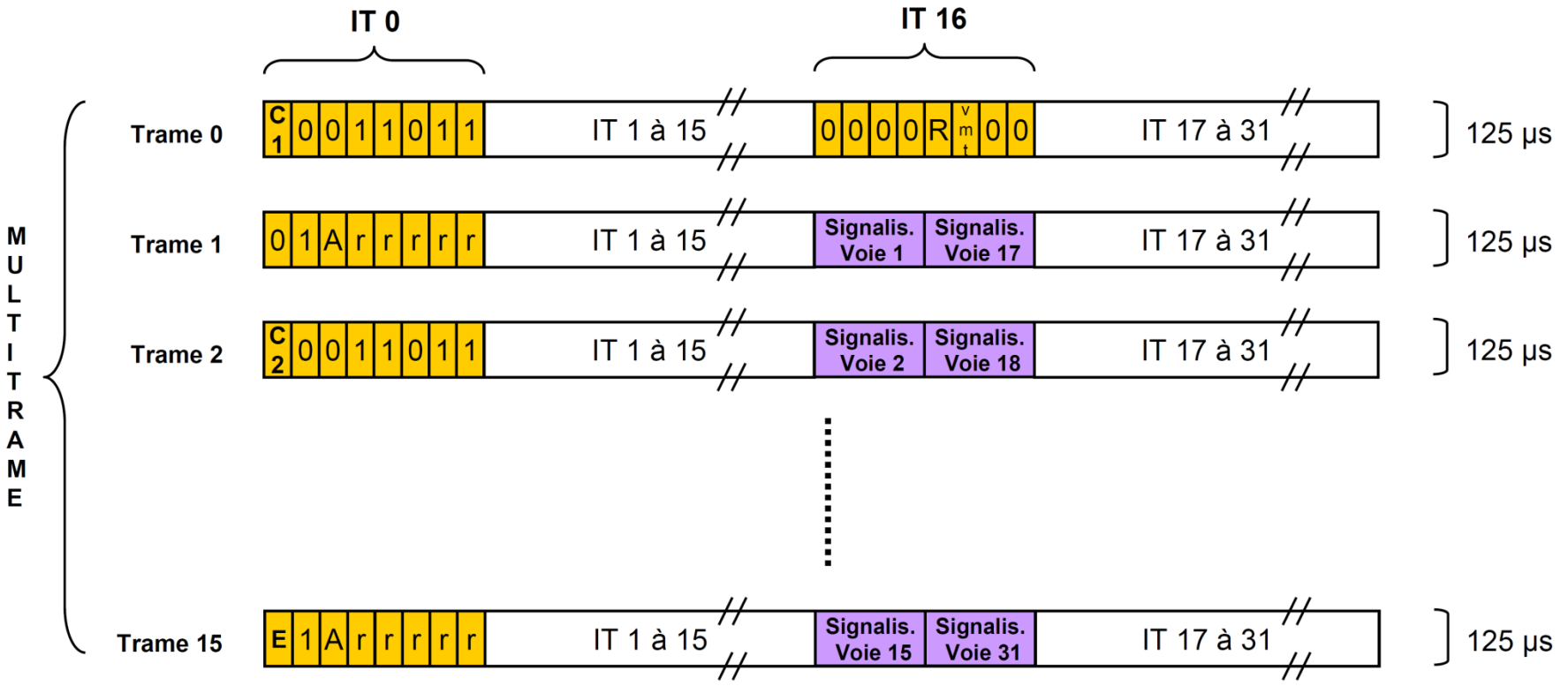
- A chaque étape de multiplexage, un petit *overhead* (surcharge) est ajouté pour décrire la structure du multiplex, et la récupération en cas de perte de synchro entre l'émetteur et le récepteur.



≥ 4 x 1.544 Mbps
Bits suppl. pour récup. de synchro

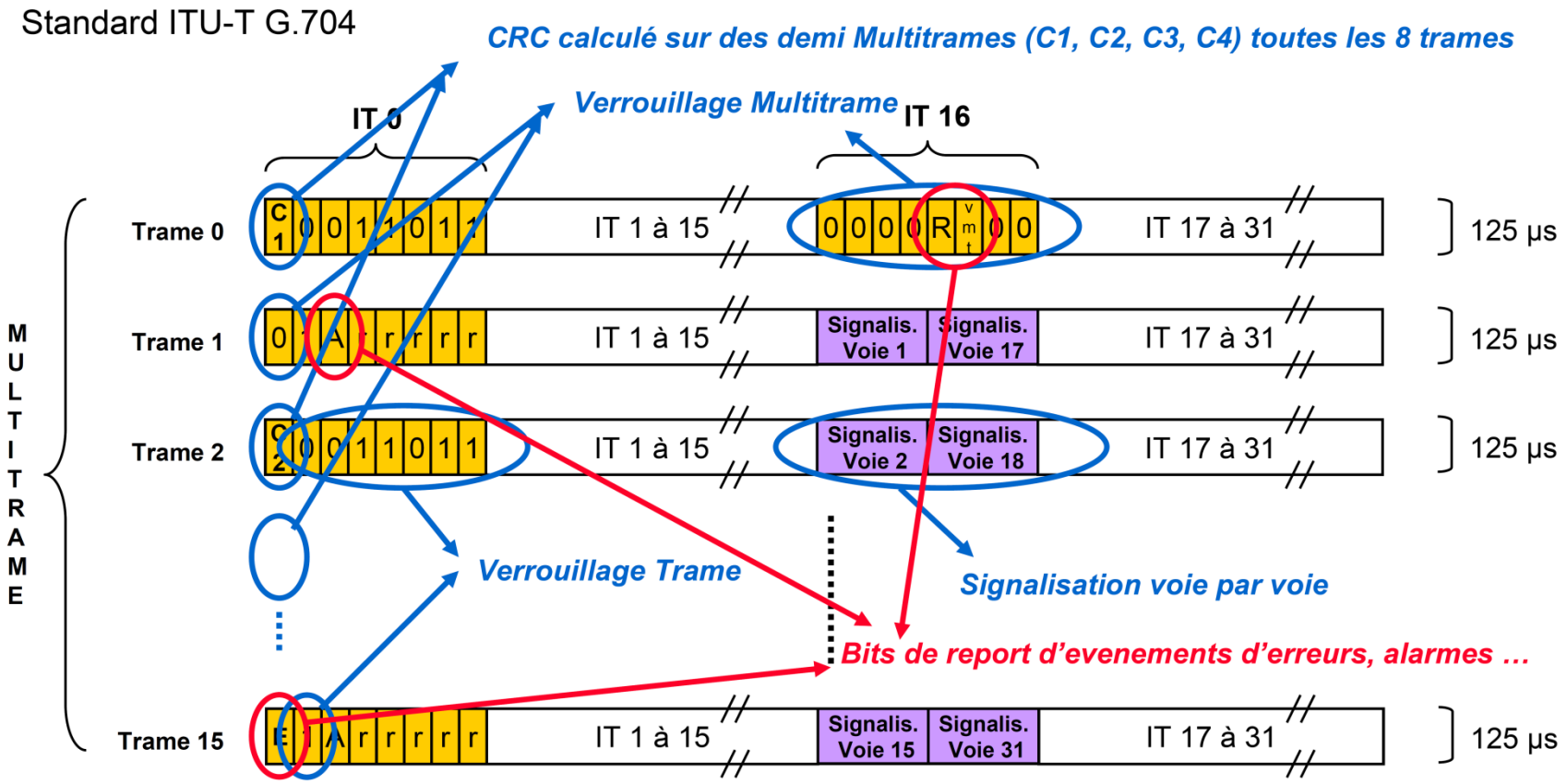
PDH – trame et multi-trame E1

Pour référence



- Des mécanismes de synchronisation résident à plusieurs niveaux: notion de **Trame**, de **Double-Trame** et de **Multi-Trame.**
- L' IT 0 transporte une synchronisation Trame (bit 1), Double-Trame (bits 1 à 7) & Multi-Trame (bit 0 trames impaires) ainsi qu'un contrôle d'erreur par CRC (bit 0 trames paires)
- L' IT 16 de la trame 0 transporte une synchronisation Multi-Trame supplémentaire
- L' IT 16 transporte la signalisation voie par voie (4 bits par voie, 2 voies par trame)

PDH – format de la trame de base E1

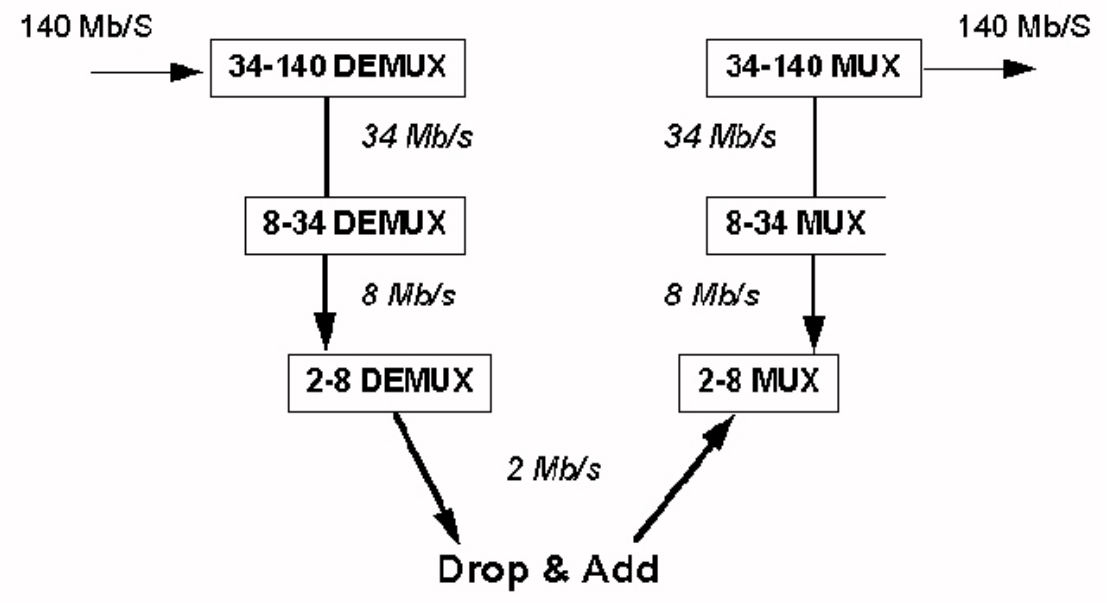


- Des mécanismes de synchronisation résident à plusieurs niveaux: notion de **Trame**, de **Double-Trame** et de **Multi-Trame.**
- L' IT 0 transporte une synchronisation Trame (bit 1), Double-Trame (bits 1 à 7) & Multi-Trame (bit 0 trames impaires) ainsi qu'un contrôle d'erreur par CRC (bit 0 trames paires)
- L' IT 16 de la trame 0 transporte une synchronisation Multi-Trame supplémentaire
- L' IT 16 transporte la signalisation voie par voie (4 bits par voie, 2 voies par trame)

Plesiochronous Digital Hierarchy (PDH)

- Traditionnellemnt, le multiplexage des signaux est plésiochrone (fait **presque** au même rythme car basé sur des horloges très proches).
- De plus, des hiérarchies différentes dans le monde -> problèmes de compatibilité au niveau international

- Pour récupérer un canal à 64 kbit/s d'un signal PDH à 140 Mbit/s, il faut entièrement démultiplexer jusqu'au E1 à 2 Mbit/s avant d'identifier l'emplacement du canal voulu à 64 kbit/s.



PDH multiplexing by steps

Limitations des réseaux PDH

- Incapacité à identifier les canaux individuels dans des multiplexes d'ordres élevés (=flux haut débit) : seulement les topologies point-to-point sont possibles
- Pas assez de possibilité de contrôle et gestion du réseau
- Pas de définition des débits au delà de 565 Mbit/s dans PDH
- Pas les mêmes systèmes au niveau international : besoin d'équipement d'interface

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
V. Accès : Réseaux optiques
   - V.1. Les différents types de multiplexage
   - V.2. La hiérarchie numérique synchrone PDH
   - V.3. La hiérarchie numérique synchrone SDH : réseaux de fibres optiques et redondance pour la fiabilité

VI. Accès : Technologies xDSL

SDH – Synchronous Digital Hierarchy

- SDH, standardisée par l'ITU est une technologie de transport synchrone qui **prolonge les principes de PDH**, conservant notamment la trame de base à 125µs, avec les avantages et **améliorations** suivants:

  - débits plus élevés, jusqu'à 10Gbit/s sur fibres optiques exclusivement

  - inclut des mécanismes de maintenance et de tolérance aux fautes (SDH réserve plus d'overhead que PDH pour ces fonctions)

  - beaucoup mieux architecturée pour supporter dans les équipements les opérations d'insertion/extraction (Add/Drop) entre les différents débits et affluents

  - peut multiplexer et transporter aussi bien des circuits élémentaires 64Kbit/s, des cellules ATM, des trames/paquets

- la technologie équivalente aux US est **SONET** (Synchronous Optical Network) développée par l'ANSI.

- SDH et SONET sont inter-opérables.

- aujourd'hui SDH & SONET sont très utilisées comme technologies de transport dans les coeurs de réseau (Core networks, Backbones)

Plan de la partie SDH

1. SDH Integration of TDM Signals
2. SDH Layers
3. SDH Multiplexing
4. SDH Framing
5. SDH Transport Overhead
6. SDH Topologies
7. SDH Protection Architectures
8. SDH Ring Architectures

Source: Livre – Vivek Alwayn, *Optical Network Design and Implementation*, Cisco Press, Networking technology, 2004

SDH Integration of TDM Signals

- Le format SDH permet différents types de format de signaux sur fibres
- La communication entre différents réseaux est complexe à cause des différences dans les hiérarchies numériques, les techniques d'encodage et de multiplexage.
- Pour résoudre ce pb de conversion de signal, SDH standardise les débits, formats de trames, signalisation, et terminaison entre équipements SDH.

Un multiplexeur SDH accepte différents formats de signaux TDM et multiplexe (*adds*) ces signaux sans conversion. Ces signaux peuvent être démultiplexés (*dropped*) n'importe où.

SDH Integration of TDM Signals

- Horloge: Au début des réseaux, on travaillait avec des horloges locales

  -> pas bien synchronisées, large variations dans la fréquence d'horloge, affectant le débit binaire des signaux.

  -> synchroniser tout ça avec une seule horloge
- Signaux optiques et électriques SDH:
  - Le terme *Synchronous Transport Module-N* (STM-N) correspond à un multiplexe d'ordre N.
  - SONET utilise le terme OC-N pour porteuse optique et STS-N pour le niveau de signal élec.
  - L'unité basique de transmission en SDH est 155.52 Mbps (STM-1).

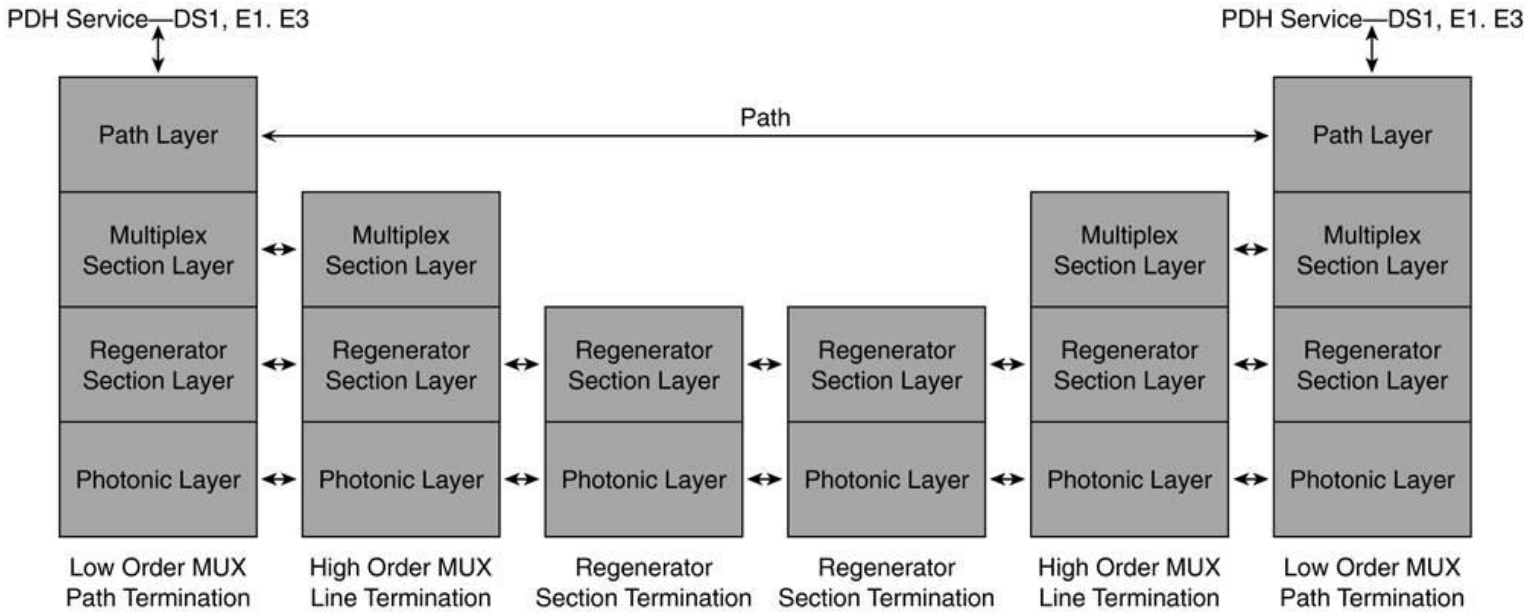
SDH Level	Line Rate (Mbps)	Payload Rate (Mbps)	Overhead Rate (Mbps)	SONET Equivalent
STM-0	51.840	50.112	1.728	STS-1
STM-1	155.520	150.336	5.184	STS-3
STM-3	466.560	451.008	15.552	STS-9
STM-4	622.080	601.344	20.736	STS-12
STM-6	933.120	902.016	31.104	STS-18
STM-8	1244.160	1202.688	41.472	STS-24
STM-13	1866.240	1804.032	62.208	STS-36
STM-16	2488.320	2405.376	82.944	STS-48
STM-32	4976.640	4810.752	165.888	STS-96
STM-64	9953.280	9621.504	331.776	STS-192
STM-256	39813.120	38486.016	1327.104	STS-768

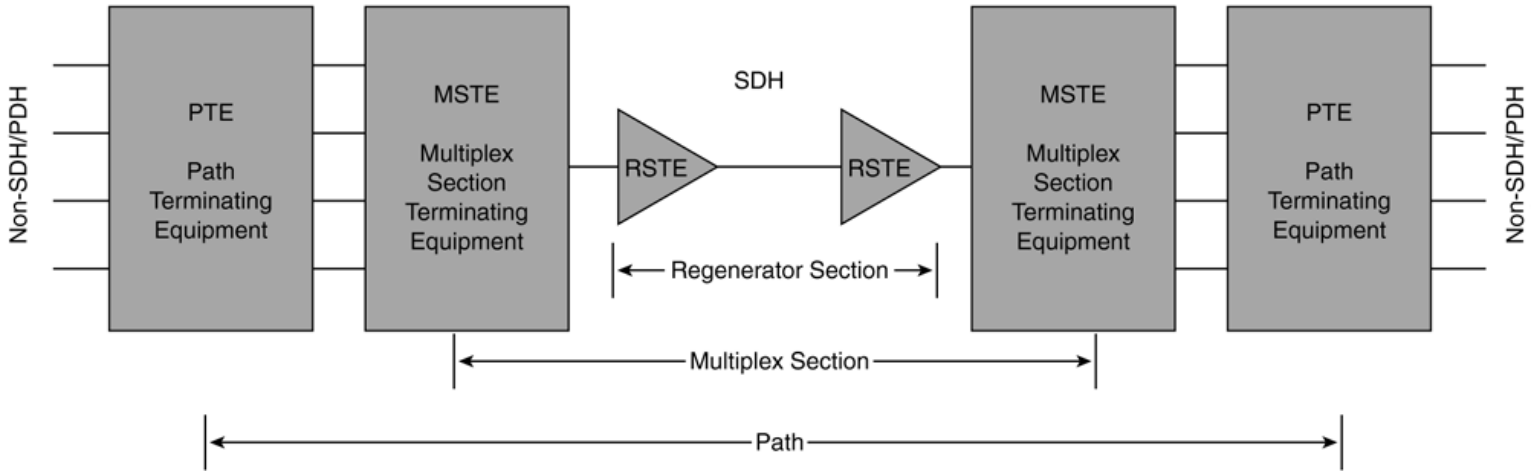
STM-n rates

SDH – la hiérarchie

SONET (US)	Bit rates	SDH
STS-1 OC-1	51.84 Mbps	STM-0
STS-3 OC-3	155.52 Mbps	STM-1
STS-12 OC-12	622.08 Mbps	STM-4
STS-48 OC-48	2488.32 Mbps	STM-16
STS-192 OC-192	9953.28 Mbps	STM-64

SDH layers


PDH Service—DS1, E1. E3
PDH Service—DS1, E1. E3
Path Layer
Path
Path Layer
Multiplex Section Layer
Multiplex Section Layer
Multiplex Section Layer
Multiplex Section Layer
Regenerator Section Layer
Regenerator Section Layer
Regenerator Section Layer
Regenerator Section Layer
Regenerator Section Layer
Regenerator Section Layer
Photonic Layer
Photonic Layer
Photonic Layer
Photonic Layer
Photonic Layer
Photonic Layer
Low Order MUX Path Termination
High Order MUX Line Termination
Regenerator Section Termination
Regenerator Section Termination
High Order MUX Line Termination
Low Order MUX Path Termination

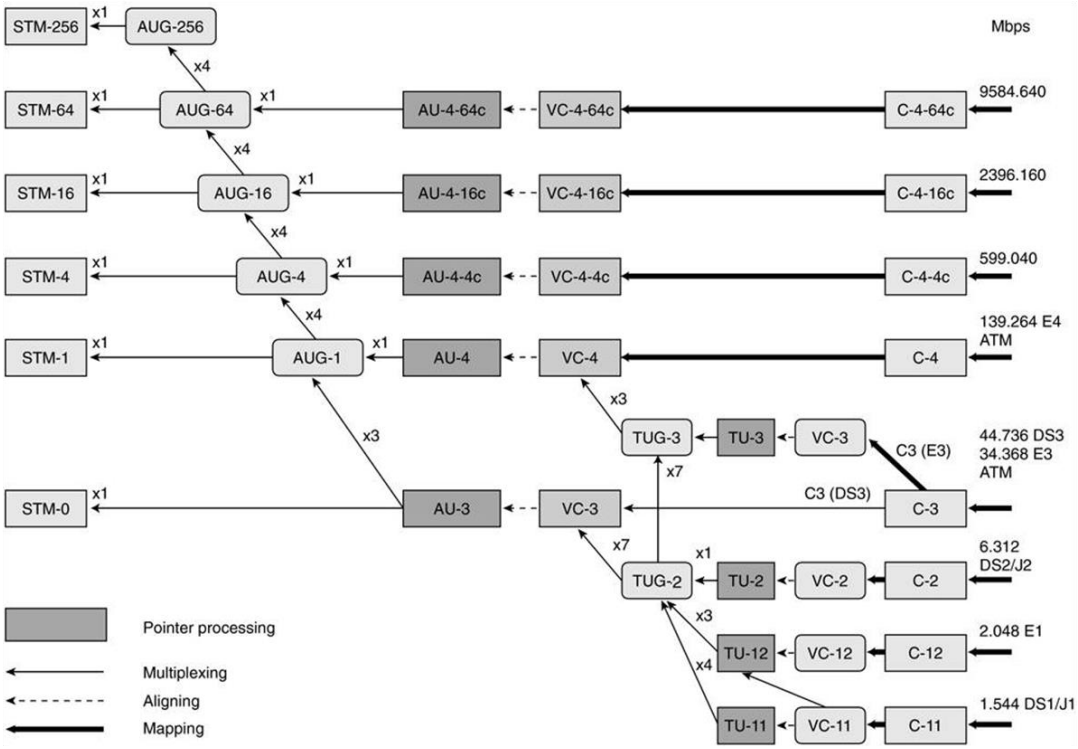

Non-SDH/PDH
PTE
Path Terminating Equipment
MSTE
Multiplex Section Terminating Equipment
SDH
RSTE
RSTE
MSTE
Multiplex Section Terminating Equipment
PTE
Path Terminating Equipment
Non-SDH/PDH
Regenerator Section
Multiplex Section
Path

SDH layers

- ***Path layer***: point auquel le format de trame SDH standard à un débit donné est assemblé.
  - Le *path overhead* (POH) y est ajouté et enlevé, et les infos d'erreur et alarmes sur l'état de bout-en-bout sont gérées à cette couche.
  - PTE: équipt qui multiplexe ou démultiplexe le payload du VC-N.
- ***Multiplex section layer***: Le MSOH pour communication, synchronisation et surveillance des erreurs.
- ***Regenerator section layer***: ajoute le RSOH.

SDH multiplexing

- 4 termes à comprendre pour le principe du multiplexage SDH:
  - Mapping— Quand les affluents sont adaptés dans les VCs par ajoût de bits de *justification* et info POH.
  - Alignement— Quand un pointeur est inclus dans une TU ou une AU, pour localiser le 1er octet du VC.
  - Multiplexage— Quand plusieurs signaux de la lower-order path layer sont adaptés dans un signal de higher-order path.
  - Stuffing— Quand les affluents sont multiplexés et alignés, de la capacité de secours est prévue dans les trames SDH pour donner assez d'espace aux divers débits d'affluents. Donc à certains moments de la hiérarchie, cette capacité en plus est remplie avec des bits de bourrage ne portant pas d'info.



- Les signaux PDH bas niveau (*low-order*) sont mappés dans une entité SDH appelée container (C).
- C est mappé avec octets POH = lower-order virtual container (VC).
- **But: Créer un payload VC uniforme en utilisant des bits de bourrage pour ramener tous les flux d'entrée à un débit binaire commun pour multiplexage synchrone.**
- Lower-order VCs sont alignés avec pointeurs tributary unit (TU) = tributary units (TUs).
- Les TUs sont multiplexés en tributary unit groups (TUGs).
- Les TUGs sont encore multiplexés pour former des higher-order VCs.
- Ces higher-order VCs sont alignés avec byte-stuffing et administration units (AU) = administration units (AUs).
- Les AUs sont encore multiplexés en administrative unit groups (AUGs).
- Les AUGs sont finalement multiplexés avec les octets RSOH and MSOH = STM-N signal.

SDH multiplexing

Low-Order VC Level	VC Bit Rate (Mbps)	PDH Level
VC-11	1.728	DS1/J1/ISDN H11
VC-12	2.304	E1/ISDN H12
VC-2	6.912	DS2/J2
High-Order VC Level		**PD Level**
VC-3	48.960	E3/J3/ISDN H31
VC-3	48.960	DS3/ISDN H32
VC-4	150.336	E4/ISDN H4

- Grâce aux pointeurs et aux valeurs d'offset, les VCs peuvent être portés dans le payload SDH comme des ensembles de donnée indépendants.

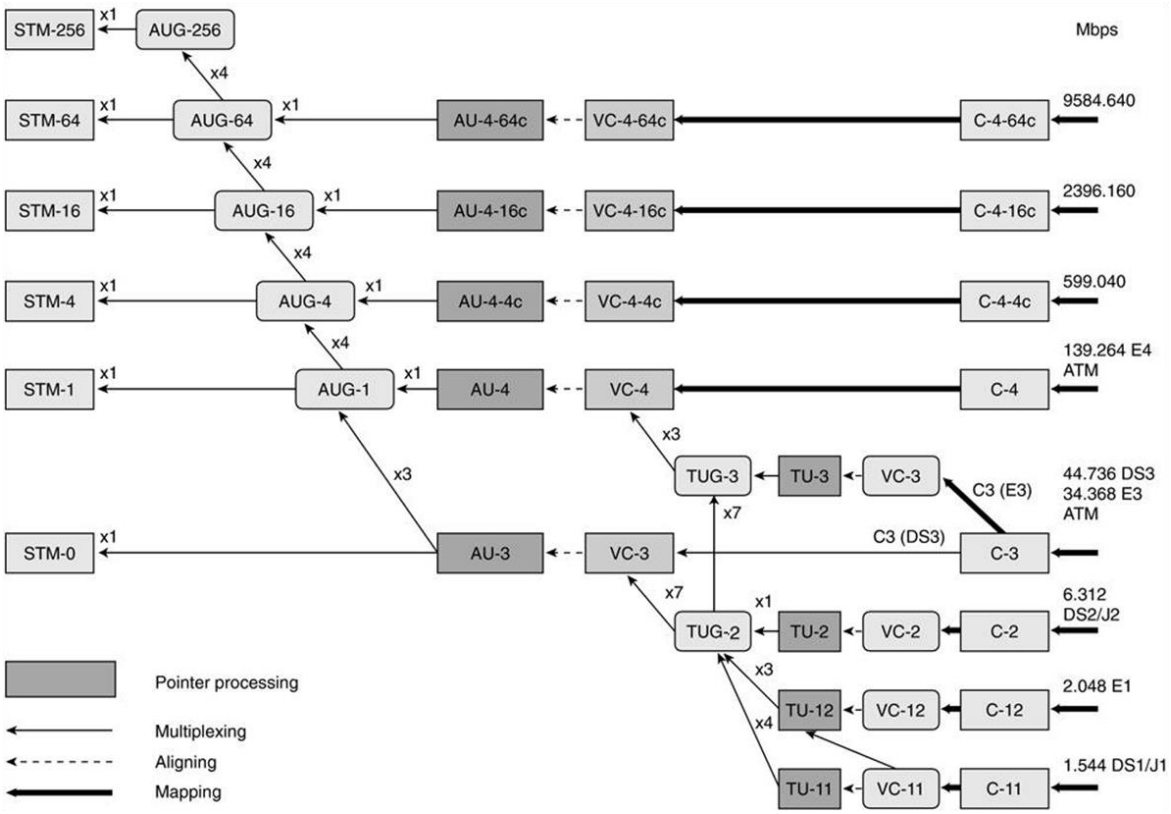
- Les VCs sont utilisés pour le transport des affluents bas débit.

- Les VCs peuvent commencer n'importe où dans la trame STM-1.

- L'endroit de départ est indiqué dans le pointeur TU ou AU.

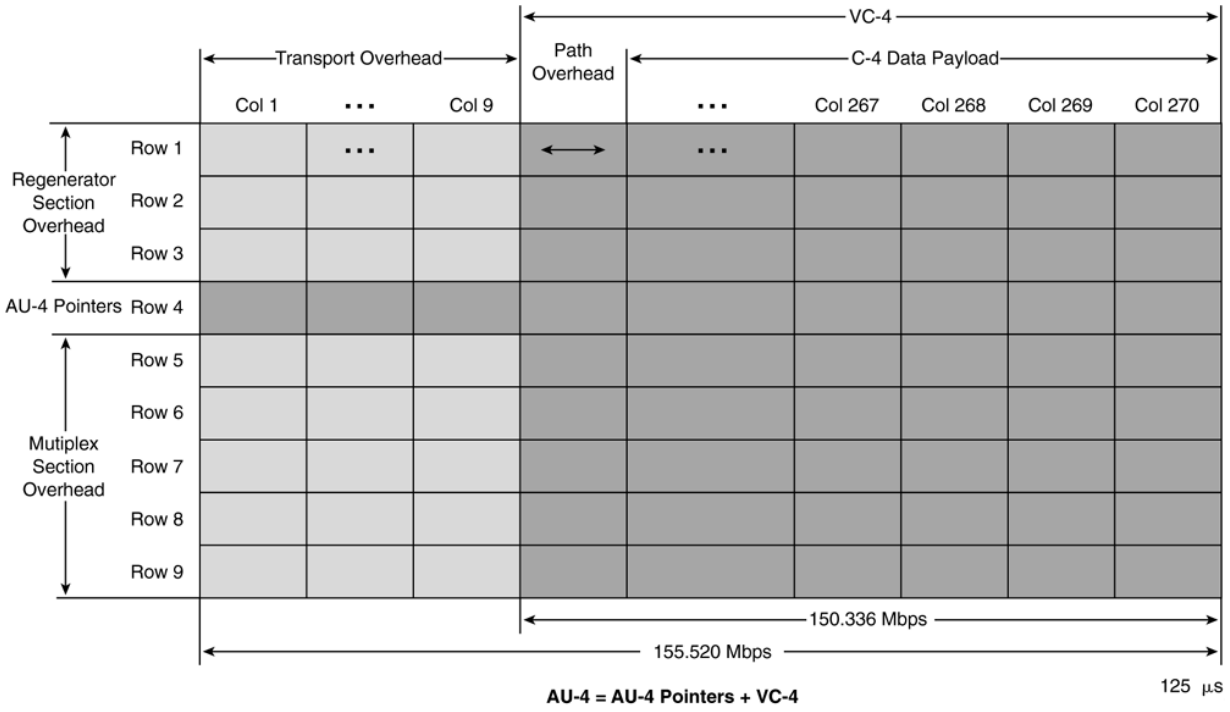
Path	SDH Overhead
C-N to VC-N	C-N + VC-N POH = VC-N (N = 11, 12, 2, 3, or 4)
VC-N to TU-N	VC-N + TU-N pointers = TU-N (N = 11, 12, 2, or 3)
VC-N to AU-N	VC-N + AU-N pointers = AU-N (N = 3 or 4)
AUG-N to STM-N	AUG-N + RSOH + MSOH = STM-N

SDH multiplexing of E1-signals



Les signaux E1 (2.048-Mbps) bas niveau sont mappés vers le container C-12.
Le container C-12 est mappé avec des bytes POH de VC-12 dans un container virtuel VC-12 de bas niveau.
Le VC-12 avec des pointeurs TU-12 est aligné dans une unité d'affluent TU-12 (*tributary unit*).
Le TU-12 est multiplexé (x3) dans un TUG-2.
Le TUG-2 est multiplexé (x7) dans un VC-3 de haut niveau.
Le VC-3 est aligné avec des pointeurs AU-3 pour former un AU-3.
Le AU-3 peut directement former un signal STM-0 ou être multiplexé (x3) pour former un AUG-1.
Le AUG-1 est multiplexé (x1) avec des octets MSOH et des octets RSOH pour former un STM-1.

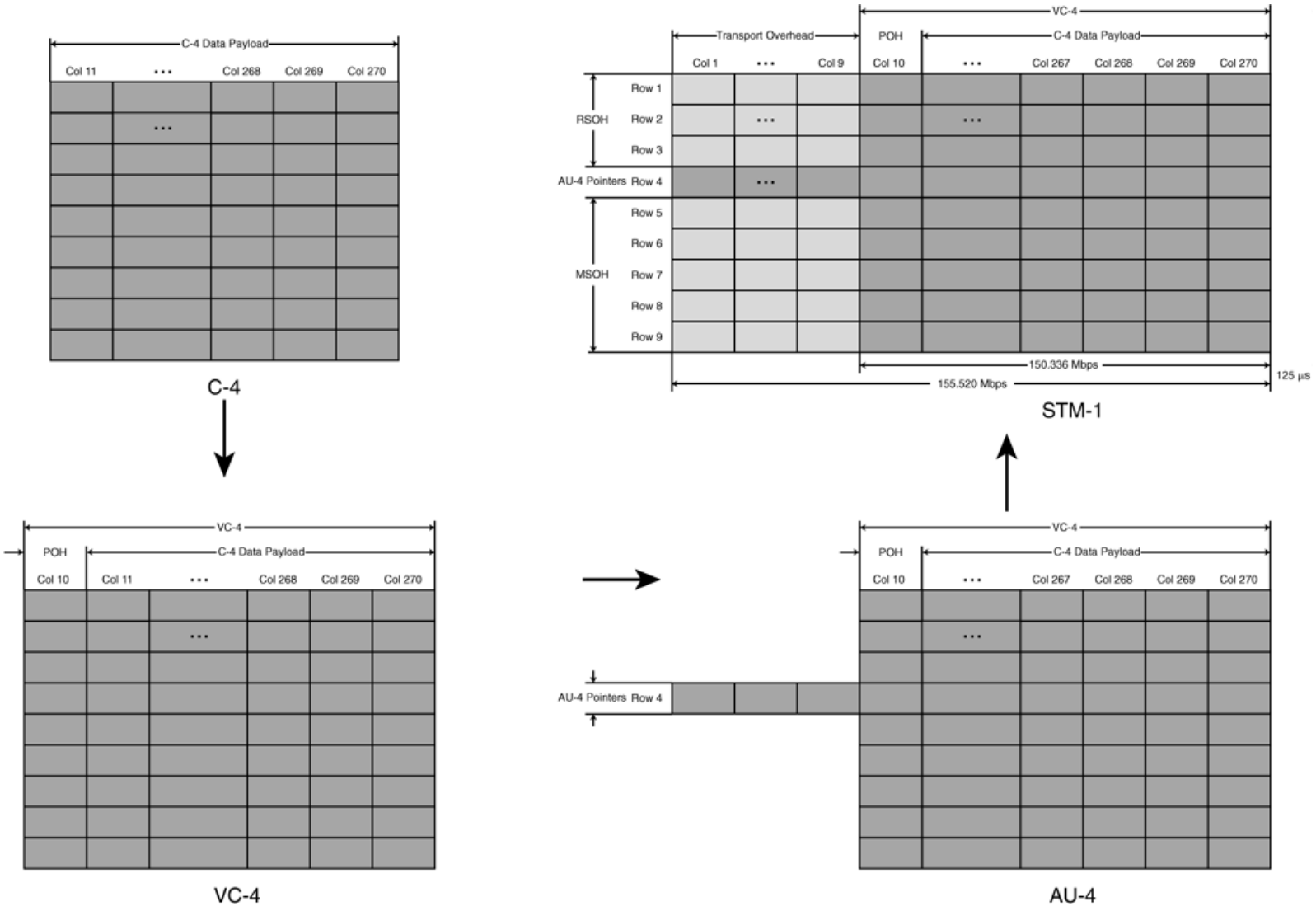
SDH framing



SDH adopte une cadence de 8000 trames par seconde.

- Une trame STM-1 = 9 lignes et 270 colonnes d'octets -> total de 9*270=2430 bytes -> 155.520 Mbps
- Une trame STM-1 = un *transport overhead* + un VC-4
- Les 9 premières colonnes = le *transport overhead*, et les autres 261 colonnes = le VC-4
- Le VC-4 + les 9 bytes de pointeurs (H1 * 3, H2 * 3, H3 * 3) ligne 4 = la AU-4
- Les premières colonnes du VC-4 sont le POH.
- Le VC-4 flotte dans la trame STM-1 et peut occuper n'importe quelle colonne de 10 à 270.
- Les octets du pointeur d'AU-N indiquent la position du POH de VC-4 dans la trame STM-1.
- Dans le transport overhead, les 3 1ères lignes pour le RSOH, les 5 autres pour le MSOH.

STM-1 frame creation



SDH pointers

- SDH fournit des pointeurs sur payload pour autoriser les différences en phase et en fréquence des VCs (VC-N) par rapport à la trame STM-N.
- Dans chaque trame STM-N, un pointeur indique où commence le VC:
  - **les octets H1 et H2 sont la valeur du pointeur = offset en octets du pointeur au premier octet du VC**
- Exemple: si le pointeur de payload de VC-4 est 0, le VC-4 commence à l'octet adjacent à l'octet H3 du TOH.
- **Outil: byte stuffing,** ou **justification** – permet l'alignement dynamique du VC au cas où il glisserait dans le temps.
  - Quand le début du VC n'est pas au tac d'horloge du début de STM-1, des octets non-informatifs sont rajoutés (*stuffed*) juste derrière l'octet H3 = **positive stuffing or positive pointer justification**.
  - Quand le débit du VC est rapide p/r au débit de STM-1, les données sont écrites dans l'octet H3 = **negative stuffing or negative pointer justification**.

SDH Transport Overhead

- Le transport overhead SDH a 2 couches incluant le RSOH et le MSOH.
- Le RSOH est utilisé pour la communication entre régénérateurs.
- Le MSOH est utilisé pour la signalisation entre les multiplexeurs STM-N.
- --> la compréhension de ces octets d'overhead est capitale dans les réseaux optiques

		STM-1 Transport Overhead								
		Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9
RSOH	Row 1	Framing A1	Framing A1	Framing A1	Framing A2	Framing A2	Framing A2	Trace J0	Reserved for National Use	Reserved for National Use
	Row 2	BIP-8 B1	Media Dependent	Media Dependent	Orderwire E1	Media Dependent	Reserved	User F1	Reserved for National Use	Reserved for National Use
	Row 3	Data Com D1	Media Dependent	Media Dependent	Data Com D2	Media Dependent	Reserved	Data Com D3	Reserved	Reserved
AU-4 Pointers		Pointer H1	Pointer H1	Pointer H1	Pointer H2	Pointer H2	Pointer H2	Pointer Action H3	Pointer Action H3	Pointer Action H3
MSOH	Row 5	BIP-8 B2	BIP-8 B2	BIP-8 B2	APS K1	Reserved	Reserved	APS K2	Reserved	Reserved
	Row 6	Data Com D4	Reserved	Reserved	Data Com D5	Reserved	Reserved	Data Com D6	Reserved	Reserved
	Row 7	Data Com D7	Reserved	Reserved	Data Com D8	Reserved	Reserved	Data Com D9	Reserved	Reserved
	Row 8	Data Com D10	Reserved	Reserved	Data Com D11	Reserved	Reserved	Data Com D12	Reserved	Reserved
	Row 9	Sync S1	Reserved	Reserved	Reserved	Reserved	REI-L M1	Orderwire E2	Reserved for National Use	Reserved for National Use

SDH Transport Overhead

- SDH Regenerator Section Overhead:
  - Le RSOH contient uniquement les infos pour les équipts de terminaison de section de régénération : les régénérateurs.
  - Regenerator Section Trace Byte (J0): L'octet J0 est utilisé pour qu'un récepteur de section puisse vérifier la continuité de la connexion vers l'autre bout.
- SDH Multiplex Section Overhead:
  - Le MSOH contient les infos pour les équipts de terminaison de section de multiplex.
  - Automatic Protection Switching Bytes (K1, K2): utilisés pour la signalisation pour *multiplex section protection* (MSP) pour APS, pour *alarm indication signal* (AIS), et *remote defect indication* (RDI).
- SDH High-Order Path Overhead:
  - Le POH est lié au VC du moment de la création du VC par un *path-terminating equipement* (PTE) jusqu'au démultiplexage.
  - Path Status Byte (G1): utilisé pour rapporter le statut du PTE et les performances au PTE origine. Les 2 sens du chemin peuvent être surveillés entièrement.

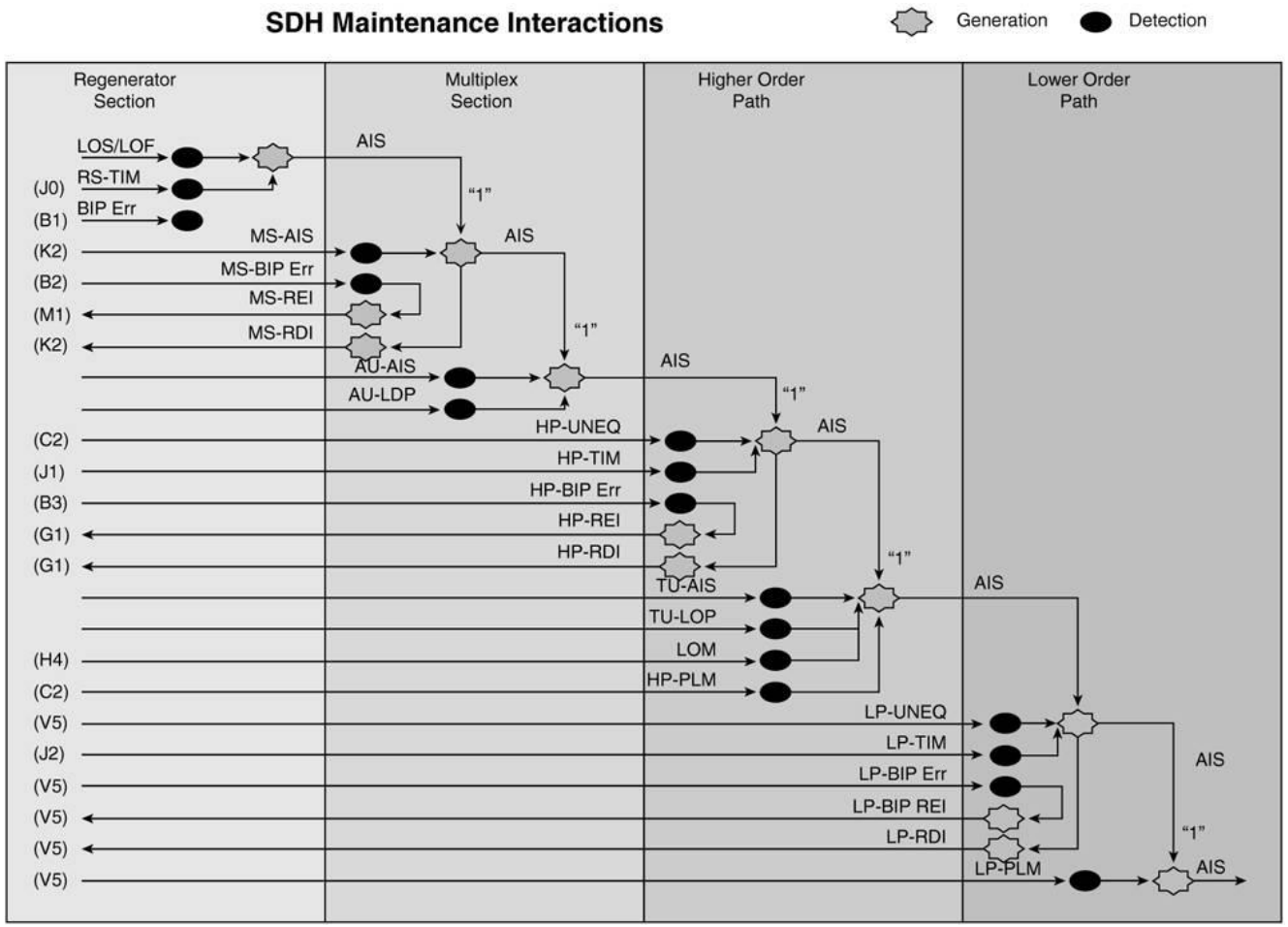
SDH alarms

- L'overhead SDH fournit un nombre d'infos de management et autres fonctions, comme surveillance d'erreur, info de pointeurs, statuts de chemins et sections, problèmes distants, indications de défaillance, messages de synchronisation.

- Les alarmes SDH se définissent en anomalies, défauts et défaillances:
  - Anomalie (*anomaly*): plus petite différence pouvant être observée entre le fonctionnement désiré et effectif. Une seule anomalie ne constitue pas une interruption de la fonction en question.
  - Défaut (*defect*): Si la densité d'anomalies atteint un niveau tel que la fonction est interrompue. Les défauts sont utilisés comme input pour performance monitoring, contrôle des mesures prises et détermination de la cause du défaut.
  - Défaillance (*failure*): Incapacité d'une fonction à effectuer l'action requise au delà d'une durée max.

- La surveillance d'erreur est basée sur *bit-interleaved-parity* (BIP) calculés trame à trame. Ces vérifications BIP sont insérées dans RSOH, MSOH, and POH.
- HO PTE et LO PTE produisent des *remote error indications* (REIs) basées sur des erreurs détectées dans les BIP de HO path et LO path respectivement.

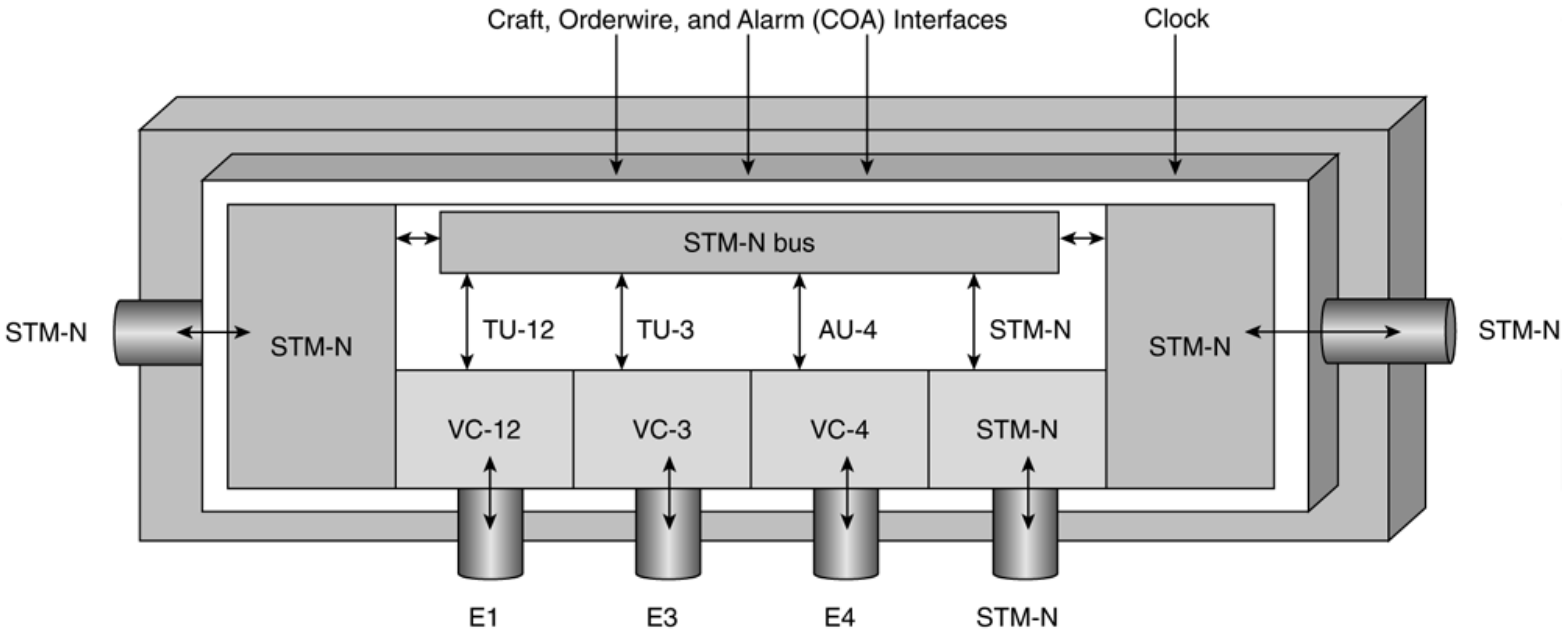
SDH alarms



The various SDH maintenance interactions, alarms and the overhead bytes used to carry the alarms and maintenance information.

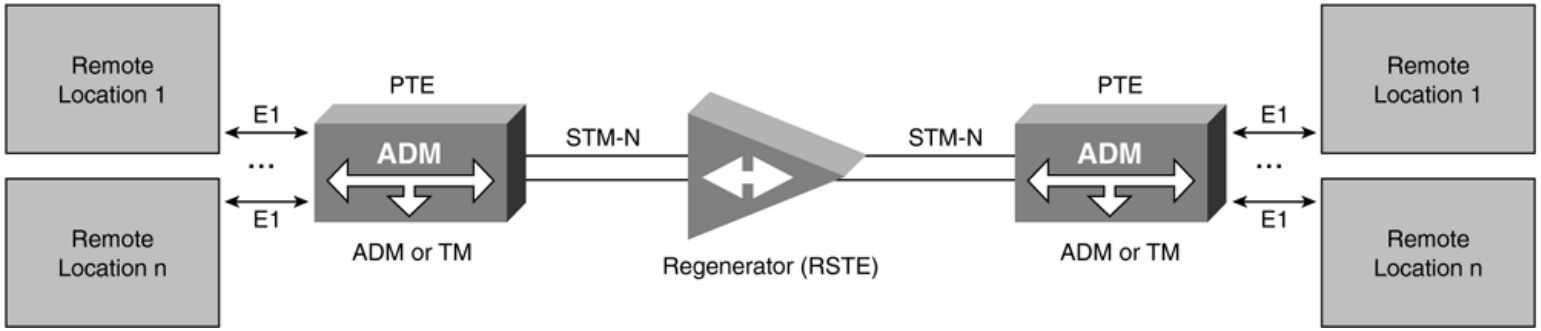
SDH network element

- Equipements effectuant transmission SDH définis some NEs.
- Incluent *regenerator section, multiplex section,* et *path-terminating equipment*.
- **Add/Drop multiplexer**:



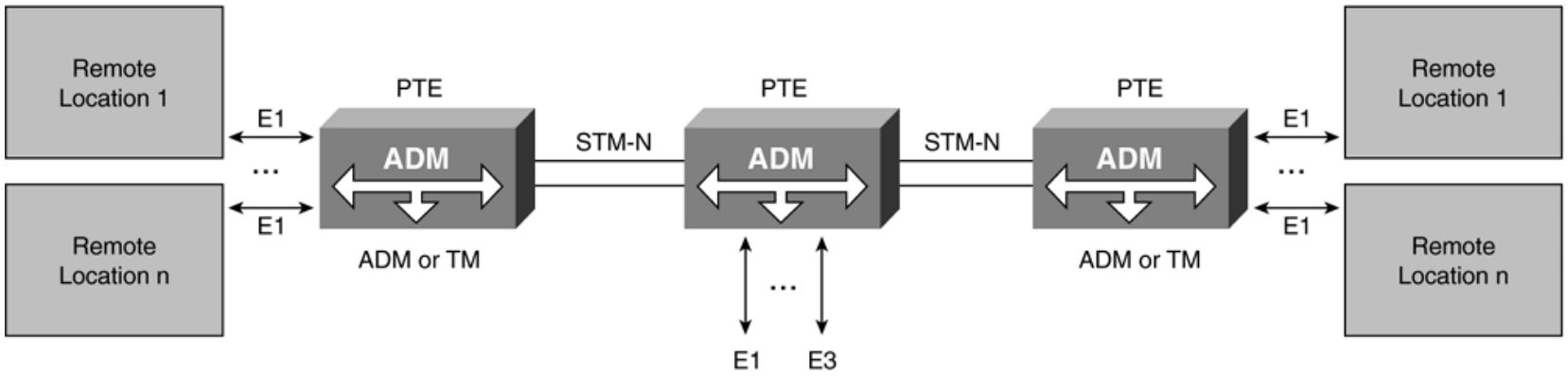
SDH topologies

- Les topologies SDH sont centrées autour de la survivabilité du réseau avec 50 ms ou moins de restoration.
- Différentes topologies peuvent être configurées avec soit des ADMs (*add/drop multiplexers*) soit des DCSs (*digital cross-connect systems*).
- Les topologies SONET ring sont fréquentes aux US, alors qu'en Europe, Asie, et Amérique Latine, c'est plutôt SDH-ring et mesh.
- Protection switching is mainly performed by the intelligent ADMs or DCSs.
- SDH Point-to-Point:
  - très populaires pour interconnecter des datacenters (SAN).
  - utilise des mécanismes de protection 1:N où un lien de protection en réserve (*standby*) pour protéger les N liens actifs (*working*). Protection max avec ratio 1:1
  - Si défaillance, le chemin de protection est utilisé avec un temps de switchover <50 ms
  - Idéalement, le chemin de protection ou fibre doit emprunter un autre chemin physique pour assurer la redondance.



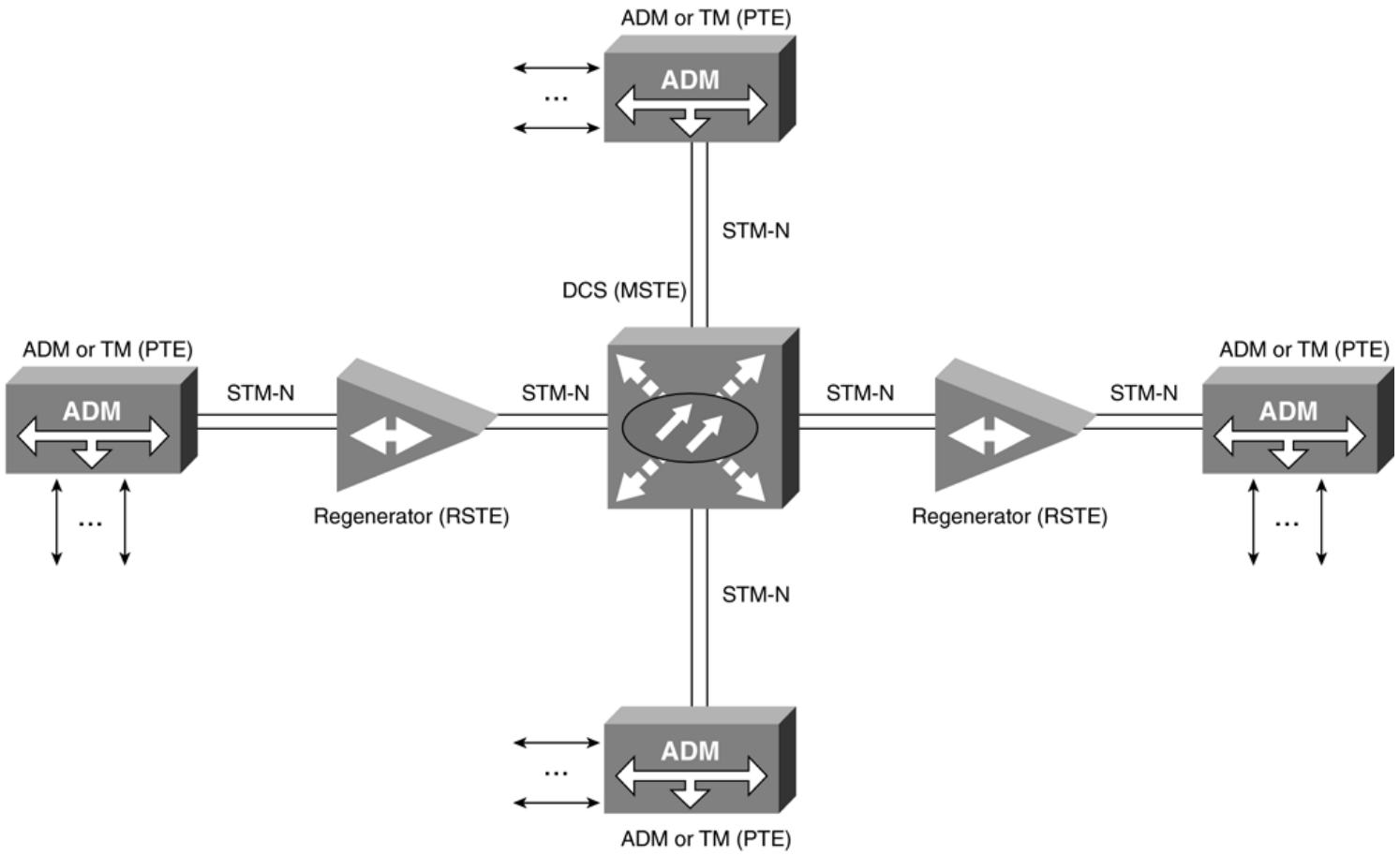
SDH topologies

- Topologie SDH Point-to-Multipoint:
  - implémente l'ajoût (*adding*) et l'abandon (*dropping*) de circuits sur le chemin.
  - Topologie implémentée pour architectures SDH linéaires moyenne et longue distance, peut avoir à ajouter ou retirer des circuits.



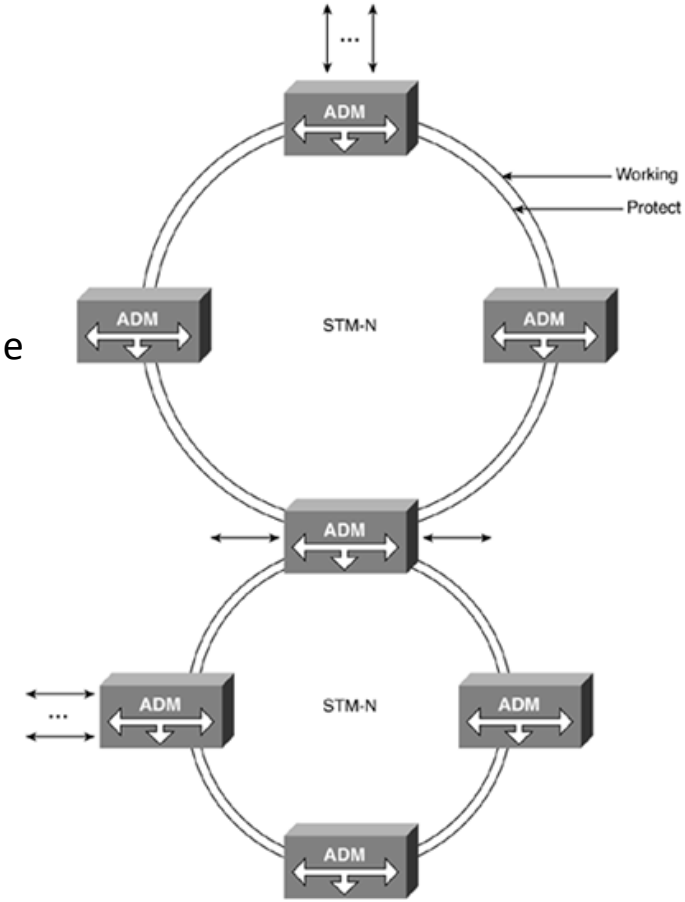
SDH topologies

- SDH Hub Topology:



SDH topologies

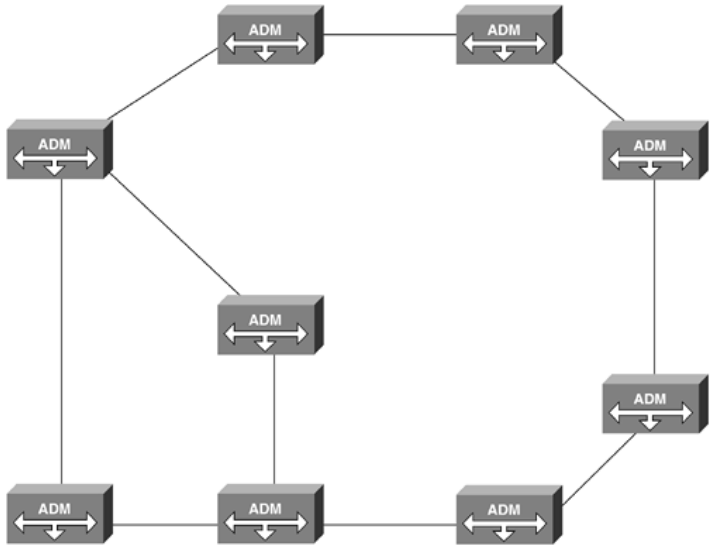
- Topologie SDH Ring: principal avantage est sa survivabilité et rapide réparation.
  - temps de restoration <50ms
  - robust protection mechanisms.
  - Le bloc de base est ADM.
  - Si une fibre est coupée, les ADMs ont l'intelligence de rerouter les services affectés sur un autre chemin sur l'anneau sans interruption.
  - La demande pour des services résistants, la diversité de routage et la flexibilité à ré-arranger des services vers d'autres noeuds
    --> Les anneaux sont l'architecture la plus populaire pour le coeur SDH et les accès metro.
  - Mécanismes de protection avancés, tels que APS, SubNetwork Connection Protection (SNCP) two-fiber, MS-SPRing two-fiber, and MS-SPRing four-fiber.



SDH topologies

- Topologie SDH Mesh : ensemble de sites arbitrairement connectés avec au moins une boucle.
  - Redondance maximum avec options de re-routage multiples.
  - MSPPs de nlle génération profitent de protocoles de protection, tels que *Path Protected Mesh Networking* (PPMN), avec procedures basées Dijkstra pour re-routage

    --> tables de routage construites au démarrage et chgt d'état des liens
  - En qql minutes, chaque noeud a une vision complète de la topologie et de l'état des liens.
  - Créer des circuits TDM ou Ethernet au dessus est alors facile en spécifiant les noeuds source et destination.



SDH Protection Architectures

- APS, linéaires et anneaux
- La spécification stipule 60 ms pour découverte du pb et restoration.
- En général encore plus rapide. (Sur cable, temps de switchover plus grands: 250ms.)
- *SDH protection switching* est invoqué s'il y a LOS, LOF, ou dégradation de signal (BER>limit).

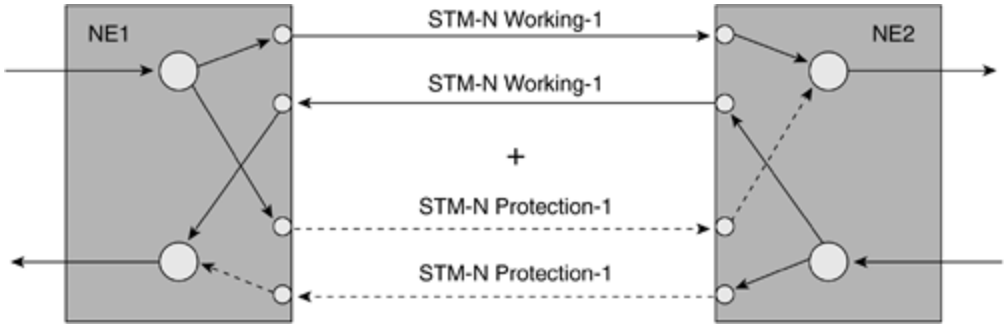
  - peut aussi etre déclenché manuellement, pour maintenance.
- Protection implique qu'une ressource de secours est prévue pour le rétablissement si la ressource primaire défaille, et restoration implique un ré-établissement du chemin de bout-en-bout sur les ressources dispo après défaillance.
- Non-revertive and revertive protection mechanisms.
- With nonrevertive protection, after the original working line has been restored to its proper operating condition, the system will not revert to use the original line as the working path. Instead, it continues to use what was originally the protection line as the working line.
- With revertive protection, the system reverts to the original line after restoration. Revertive protection is most common with 1:N protection schemes.

SDH Protection Architectures

- ***Automatic Protection Switching*** (APS): pour restoration dans le cas d'une défaillance.
  - 2 ensembles de fibres: 1 (*transmit and receive pair*) pour le trafic actif (*working traffic*), et l'autre (*transmit and receive pair*) pour protection
  - Les fibres actives et de protection doivent emprunter des chemins physiques différents pour être efficaces. Les fibres utilisées pour protection peuvent éventuellement porter une copie du trafic (cas 1+1).
  - APS pour topologies linéaires et ring.
  - Seule la *multiplex section* est protégée par APS. Géré par les octets K1 et K2 du MSOH.
  - Path protection gérées par des fonctions réseau de management de plus haut niveau.

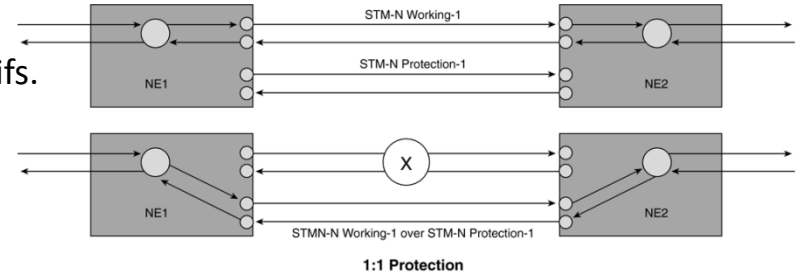
SDH protection architectures

- Automatic Protection Switching:
  - **Protection 1+1**: recopie des signaux STM-n sur les 2 fibres (actif et de protection).
    - ne peut pas transporter de trafic supplémentaire non-protégé
    - signaux constamment et indépendemment surveillés pour failles
    - les équipts sélectionnent entre les 2 selon règles locale
    - commute entre les 2 si faille, en moins de 50 ms
    - pour une vraie efficacité, l'installation de fibre doit avoir de la diversité physique

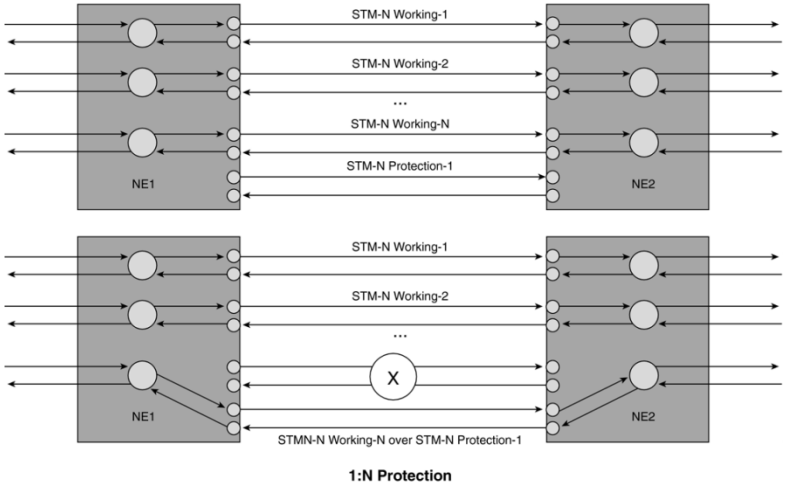


SDH protection architectures

- Automatic Protection Switching:
  - **1:1 and 1:N Protection**: trafic sur la ligne active jusqu'à ce qu'une faille arrive.
    - Ligne de protection utilisée si faille sur l'active.
    - 1:N: 1 chemin de protection pour N chemins actifs.
    - N'importe lequel, mais un seul chemin peut être protégé à la fois. Si faille sur plus d'un chemin à la fois, besoin de priorités.
    - Comme la ligne de protection ne porte pas de trafic en temps normal, possible de porter d'autre trafic non-prioritaire appelé extra traffic, abandonné si la ligne de protection est invoquée.



1:1 Protection



1:N Protection

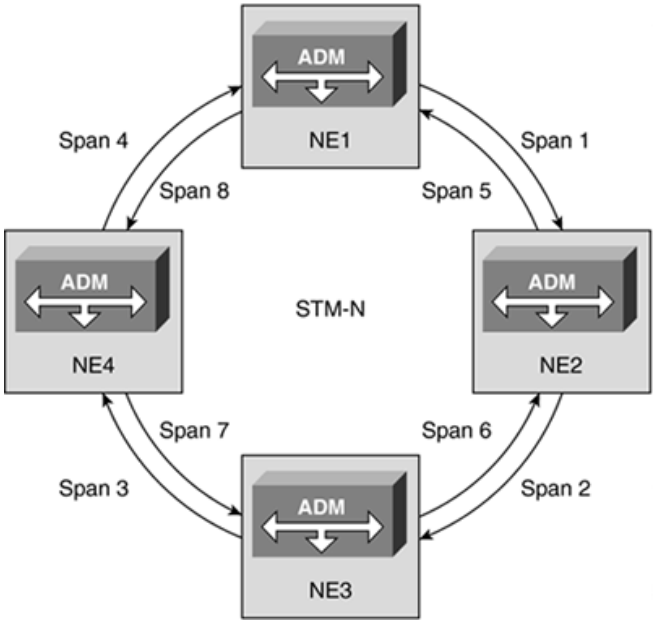
SDH Ring Architectures

SDH Attribute	Value
Fibers per link	2-fiber 4-fiber
Signal direction	Unidirectional Bidirectional
Protection switching	Multiplex section switching Path switching

- Différentes config d'anneau SDH avec caractéristiques précises.
- 3 types majoritaires:
  - Two-fiber subnetwork connection protection ring (two-fiber SNCP)
  - Two-fiber multiplex section-shared protection ring (two-fiber MS-SPRing)
  - Four-fiber multiplex section-shared protection ring (four-fiber MS-SPRing)

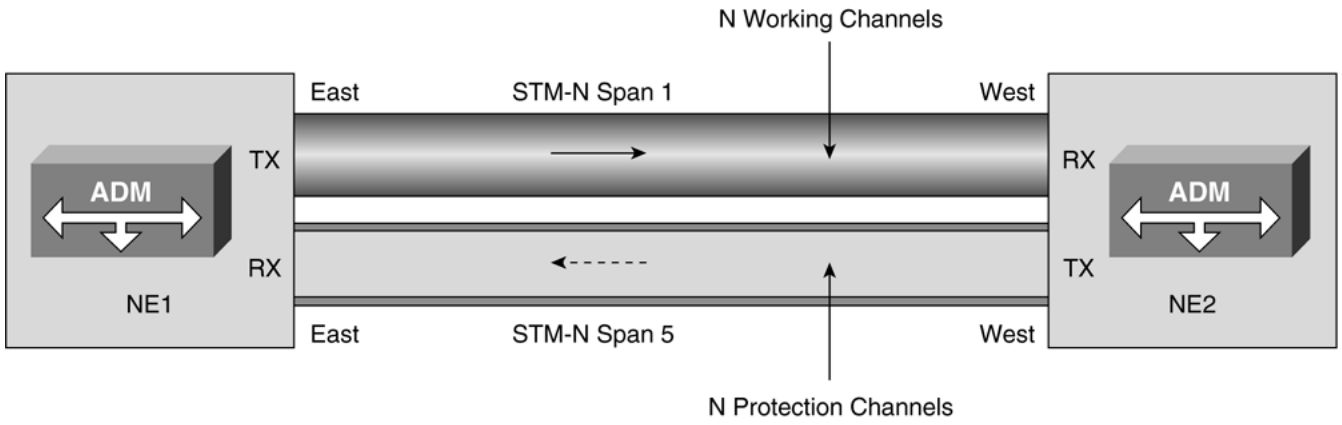
SDH Ring Architectures

- **Unidirectional Versus Bidirectional Rings**:
  - Unidirectional ring: working traffic dans une seul direction entre 2 noeuds
  - Bidirectional ring: working traffic dans les 2 directions entre 2 noeuds



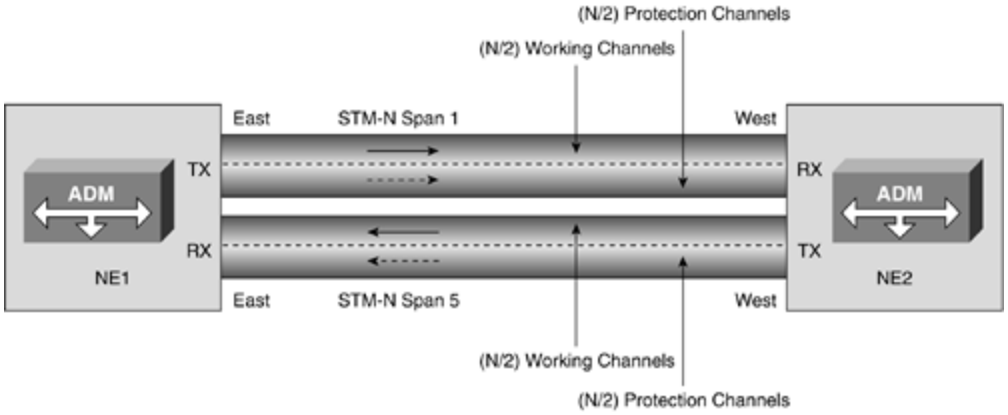
SDH Ring Architectures

- **Two-Fiber Versus Four-Fiber Rings**:
  - Pour systèmes unidirectionnels et bidirectionnels
  - Systèmes unidirectionnels commercialisés, tel que SNCP, sont de type two-fiber, les bidirectionnels, tel que MS-SPRing, sont des 2 types.
  - Un système STM-N unidirectionnel two-fiber avec 2 noeuds:
    - Fibre 1 porte N canaux actifs vers NE2, et fibre 5 N canaux de protection vers NE1.
    - (For example, an STM-16 system would carry 16 working VC-4s eastbound from NE1 to NE2, while carrying 16 separate protection VC-4s westbound from NE2 to NE1.)



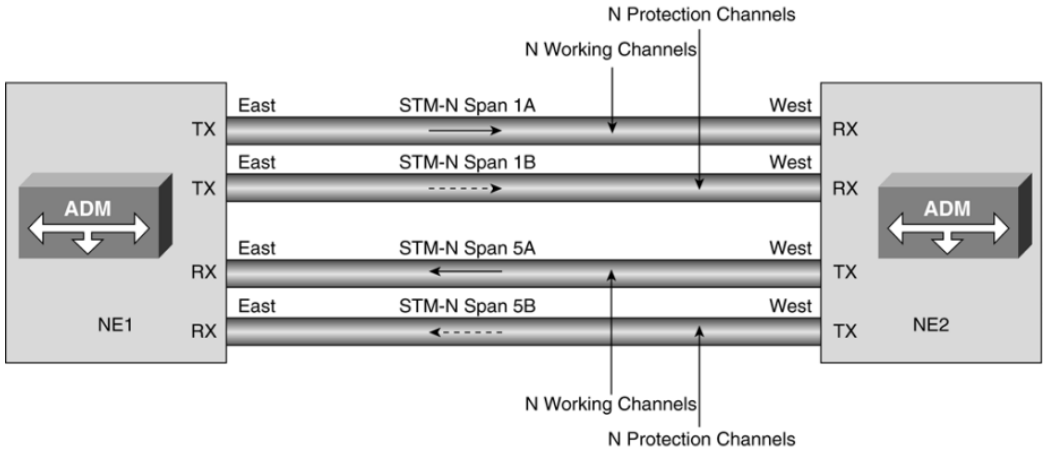
SDH Ring Architectures

– Un système STM-N bidirectionnel two-fiber avec 2 noeuds:
  - Sur chaque fibre, un maximum de la moitié des canaux est défini comme canaux actifs, l'autre moitié comme canaux de protection.
  - Fiber span 1 carries (N/2) working channels and (N/2) protection channels eastbound, and fiber span 5 carries (N/2) working channels and (N/2) protection channels westbound.



SDH Ring Architectures

– Un système STM-N bidirectionnel four-fiber avec 2 noeuds:



-> Two-fiber SNCP fournit N canaux actifs dans une direction, tandis qu'un two-fiber MS-SPRing fournit (N/2) canaux actifs dans chaque direction.

-> Four-fiber MS-SPRing fournit N canaux dans chaque direction.

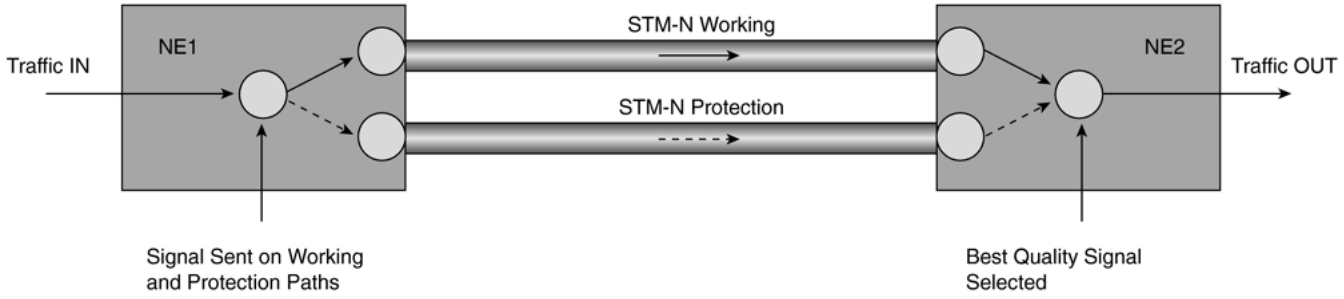
-> 1 défaillance sur chaque segment cause une défaillance réseau sur un two-fiber ring de chaque type.

-> Un four-fiber avec diversité de routage peut supporter de multiples failles.

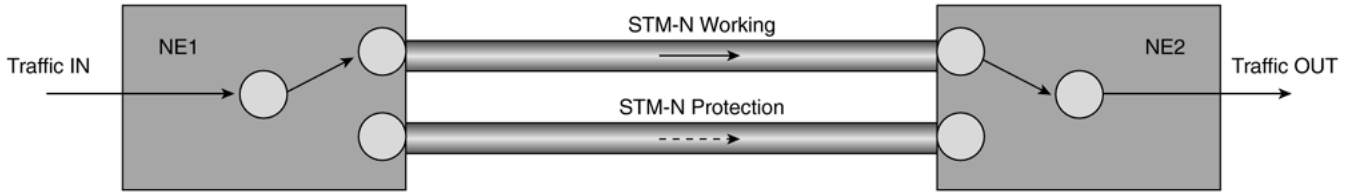
-> Les four-fiber sont très utilisés pour les anneaux sur de grandes distances ou au trafic critique.

SDH Ring Architectures

- **Path and Multiplex Section Switching**:
  - Path switching: une copie du trafic protégé est aussi transmise sur la fibre de protection. Path switching avec two-fiber SNCP rings.

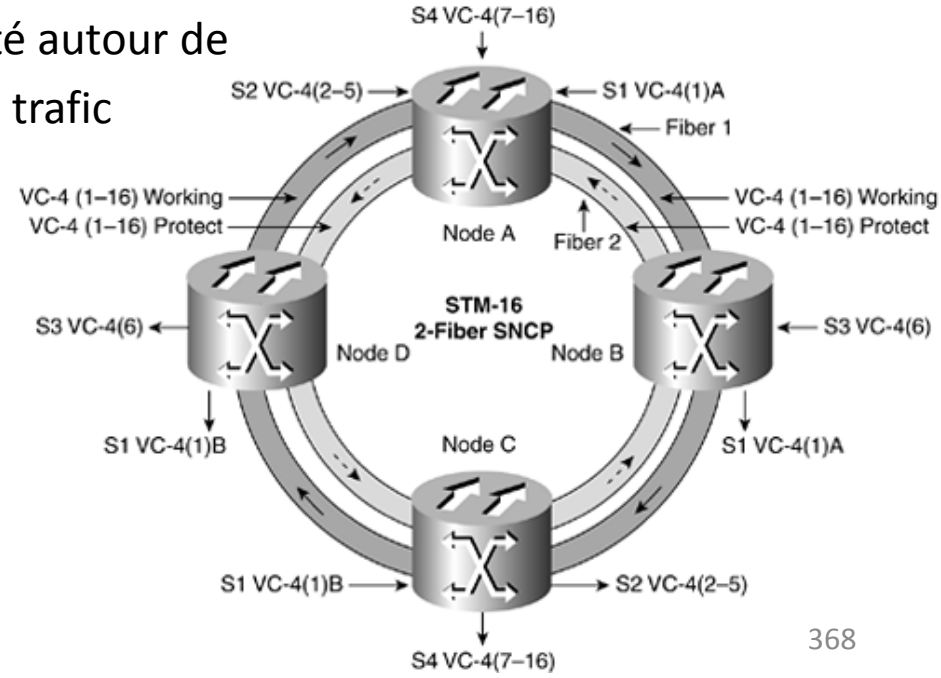


  - MS switching: Si défaillance de fibre ou de noeud, le trafic protégé est switché sur canal de protection, tout trafic présent sur ce canal est abandonné.



SDH Ring Architectures

- **Subnetwork Connection Protection Rings (SNCP)**:
  - Un anneau SNCP est une architecture de transport résiliente, en boucle fermée, qui protège contre les coupures de fibre en fournissant des chemins dupliqués divers géographqt pour chaque circuit.
  - SNCP fournit un chemin double de fibres sur l'anneau.
  - Trafic actif dans un sens, protection dans l'autre : recopie.
  - Comme chaque trafic est transporté autour de l'anneau, SNCPs plutôt pour cas où trafic vient de 2 noeuds uniquement.
  - Cas d'une faille: figure.

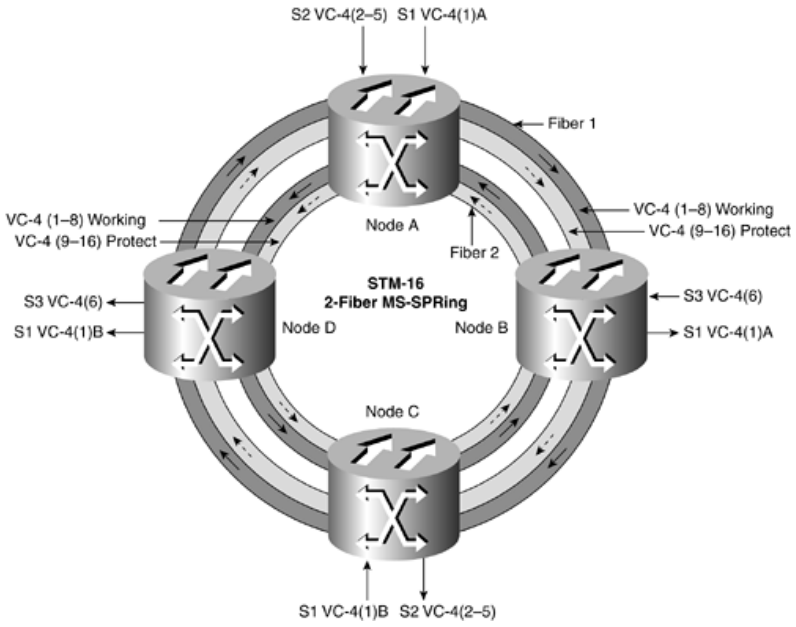


SDH Ring Architectures

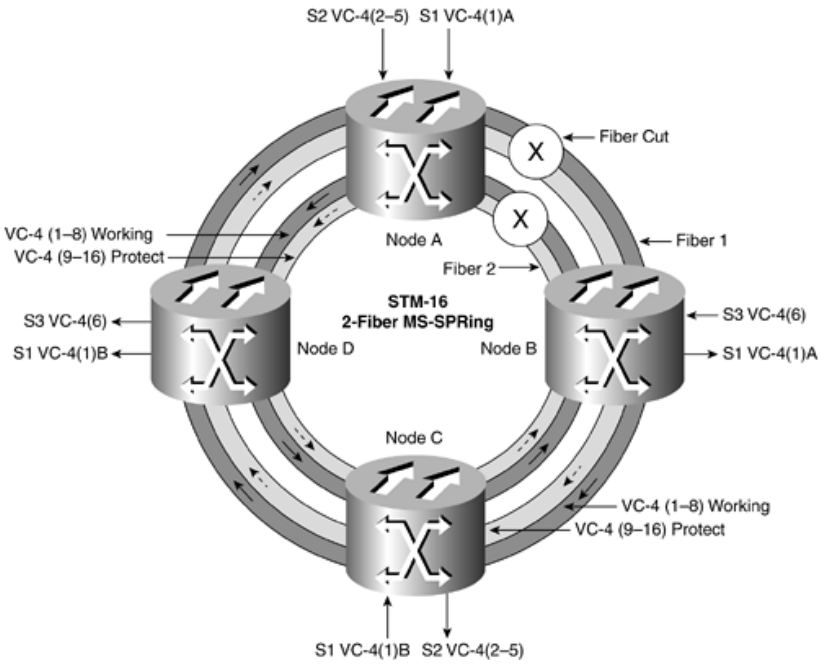
- **Multiplex Section-Shared Protection Rings**:
  - MS-SPRing avec multiplex section bidirectionelle – système two-fiber ou four-fiber.
  - Adapté aux trafic distribués, tel que réseau d'interconnexion entre sites.
  - Permet ré-utilisation de BW autour de l'anneau.
  - supports non-revertive and revertive protection mechanisms.

SDH Ring Architectures

– **Two-Fiber MS-SPRing**: 1 fiber per direction, working and protection traffic in each.
  - With careful bandwidth-capacity planning, MS-SPRing could be quite efficient.



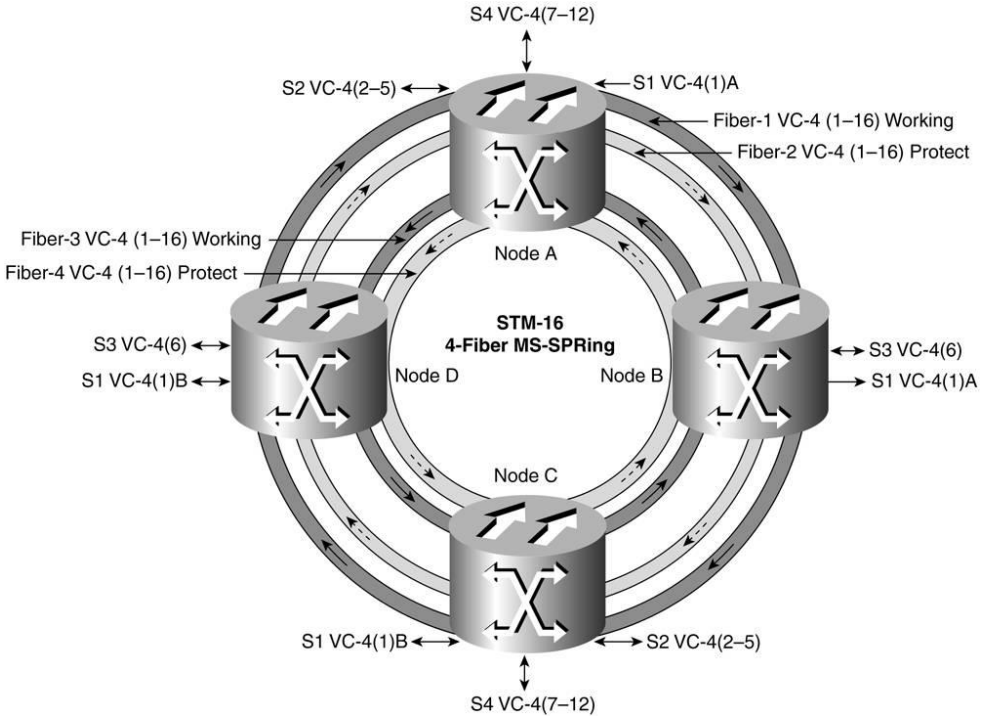
Two-Fiber MS-SPRing



Two-Fiber MS-SPRing Protection

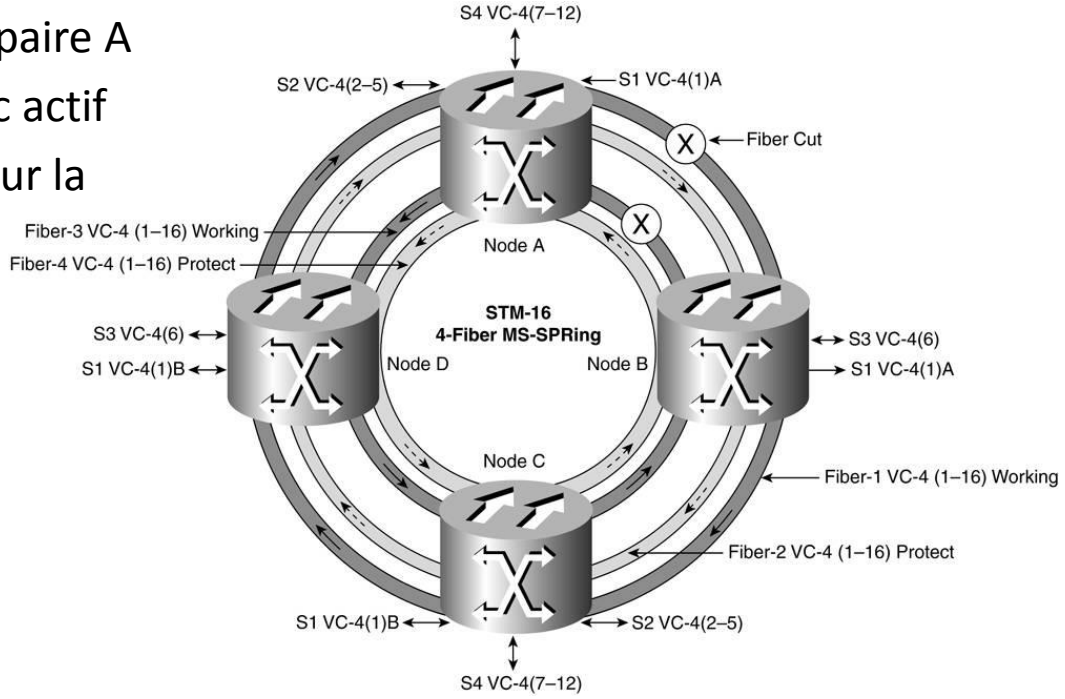
SDH Ring Architectures

– **Four-Fiber MS-SPRing**:
  - double the bandwidth of two-fiber MS-SPRings
  - two fibers are allocated for working traffic and two fibers are allocated for protection



SDH Ring Architectures

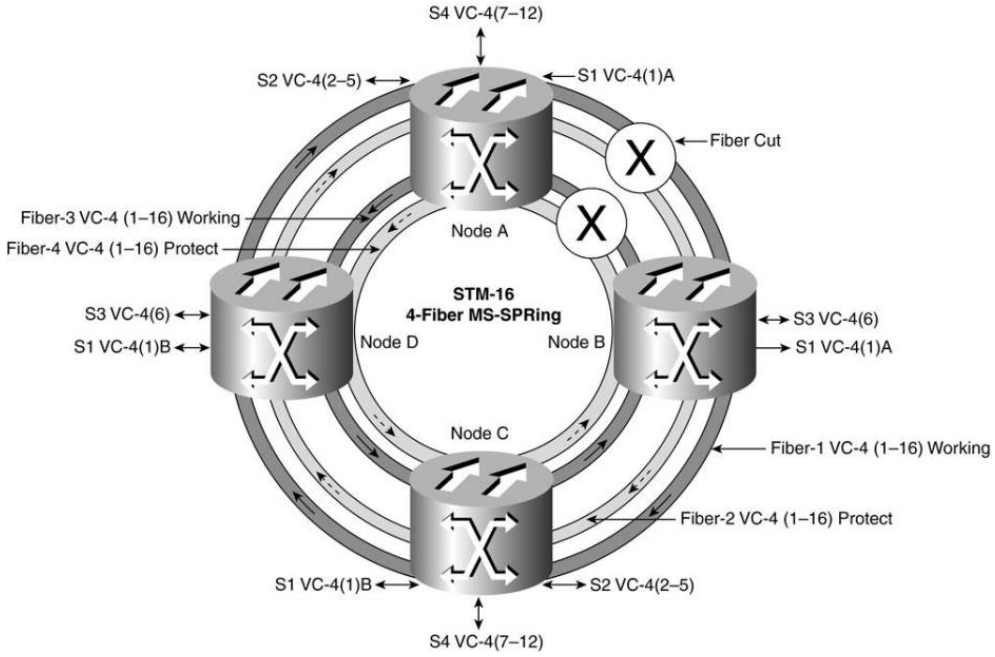
- Four-fiber MS-SPRing permet Path Switching et MS Switching
  --> plus de flexibilité et fiabilité pour trafic protégé.
  - Quand une fibre active défaillit, le trafic switche vers les fibres de protection entre les noeuds, puis retourne vers les fibres actives.
  - Si la fibre active entre la paire A et B défaillit, tout le trafic actif entre ces noeuds passe sur la fibre de protection. Tout trafic non-protégé entre 2 autres noeuds n'est pas affecté.



Four-Fiber MS-SPRing Span Switch

SDH Ring Architectures

- Four-fiber MS-SPRing ring:
  - Si les 2 paires de fibres entre A et B défaillissent, tout le trafic actif entre ces noeuds est passé sur les paires de protection.
  - Tout trafic non protégé entre d'autres noeuds peut etre droppé.



Four-Fiber MS-SPRing Ring Switch

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
V. Accès : Réseaux optiques
VI. Accès : Technologies xDSL
    - VI.1. Principe : tirer le maximum de la paire de fils de cuivre
    - VI.2. Technique de TS pour xDSL : l'OFDM
    - VI.3. Fonctionnement protocolaire entre le FAI et l'abonné (cas dégroupé et non-dégroupé)

Références

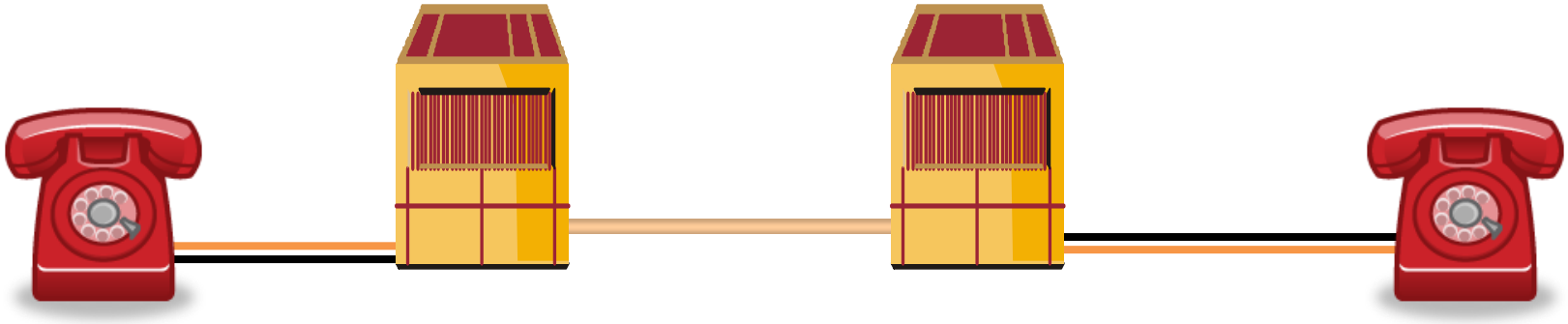
- Cours basé sur les transparents de J. Stein (*RAD data communications*)
- Cours OFDM de P. Ciblat (Télécom ParisTech)
- Site de Christian Caleca

Original PSTN

Manual switching directly connected two *local loops*

Due to microphone technology, audio BW was 4 kHz

Analog switched PSTN

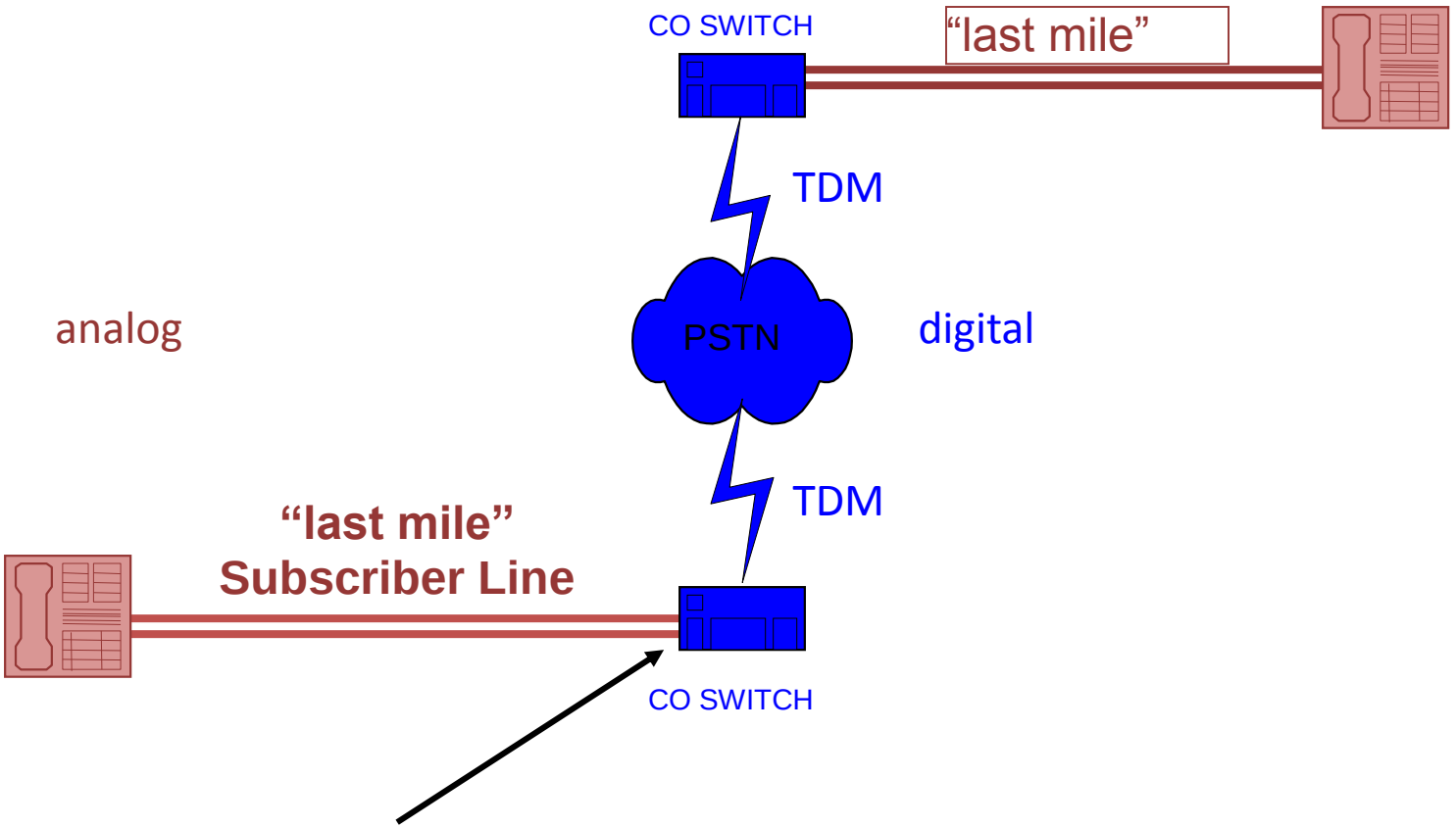


Invention of tube amplifier enabled *long distance*

Between central offices used FDM spaced at 4 kHz
(each cable carrying 1 group = 12 channels)

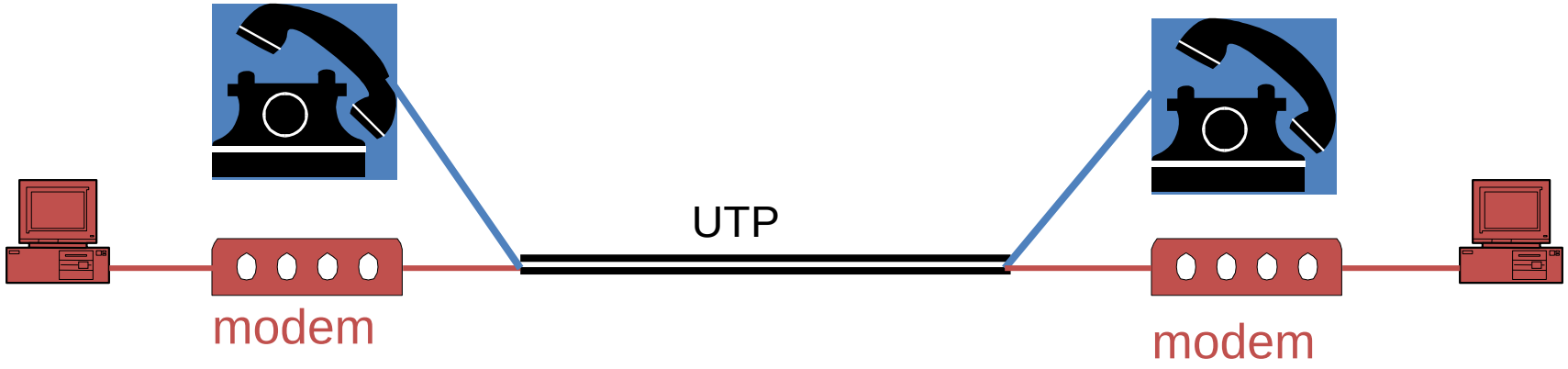
Developed into hierarchical network of automatic switches
(with supergroups, master groups, supermaster groups)

Digital PSTN



LP filter to 4 kHz at input to CO switch (before A/D converter)

Data supported via voice-grade modems



To send data, it is converted into 4 kHz audio (modem)

Data rate is determined by *Shannon's capacity theorem*

- *there is a maximum data rate (bps) called the "capacity" that can be reliably sent through the communication channel*
- *the capacity depends on the BW and SNR*

Where is the limitation ?

A 4 kHz analog channel is presented to subscriber

The 4 kHz limitation is enforced by LP filter

at input to CO switch (before 8 kHz sampling)

The actual subscriber line is not limited to 4 kHz

Indeed, the maximum data rate depends on the BW and SNR

Is there a better way

to use the **s**ubscriber **l**ine for **d**igital transmissions ?

What is UTP?

The achievable data rate is limited by physics of the subscriber line

The subscriber line is an Unshielded Twisted Pair of copper wires

- Two plastic insulated copper wires (two directions over single pair)
- Twisted to reduce crosstalk
- Supplies DC power and audio signal
- Physically, UTP is
  - distributed resistances in series
  - distributed inductances in series
  - distributed capacitances in parallel

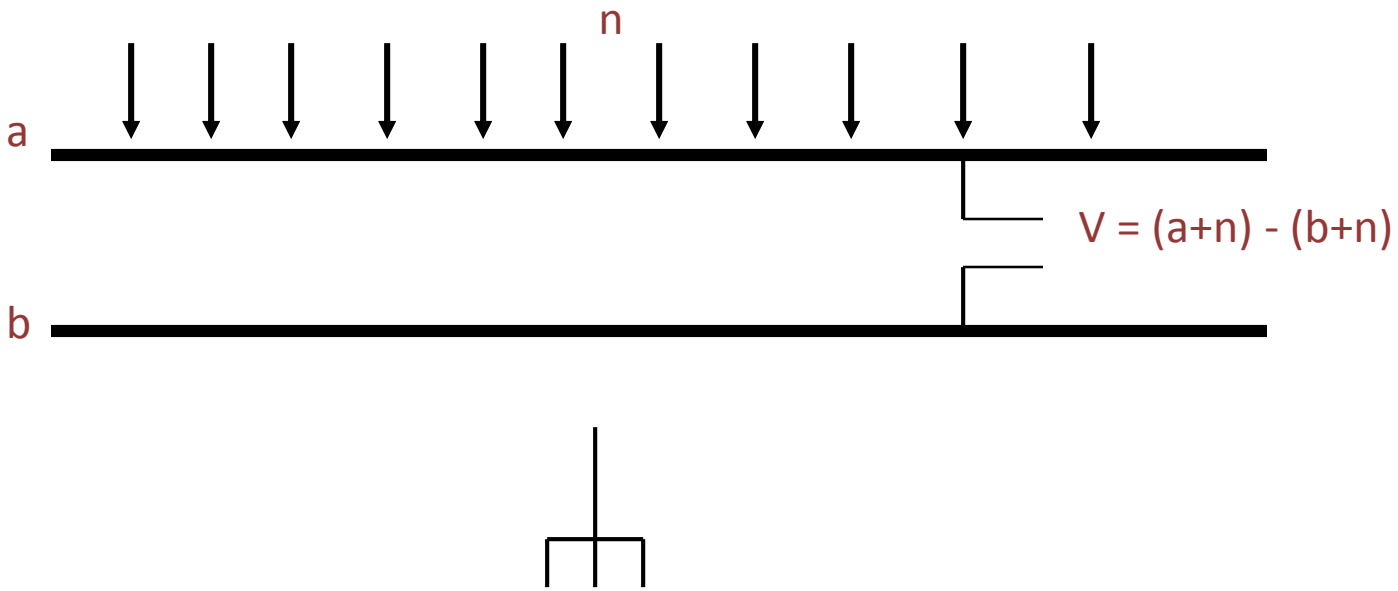
  so the attenuation increases with frequency
- **Various other problems exist (splices, loading coils, etc.)**

Twisting: to be more resilient to outside and inside noises

from Alexander Graham Bell's 1881 patent

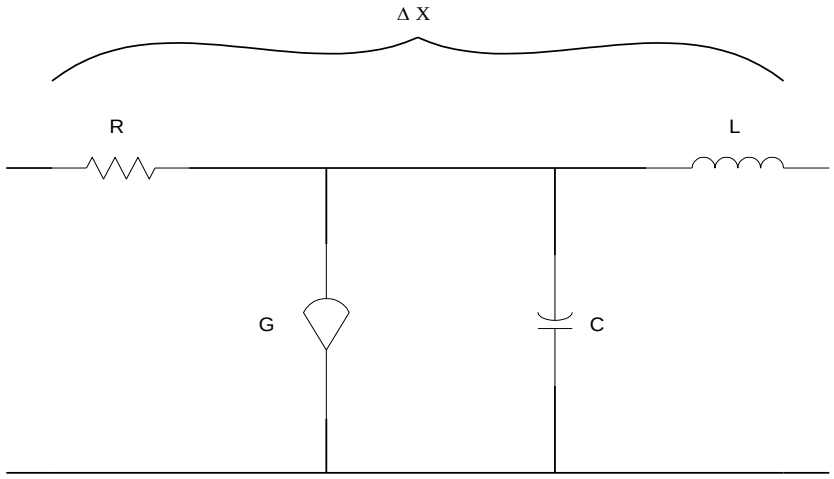
To place the direct and return lines close together.

To twist the direct and return lines around one another so that they should be absolutely equidistant from the disturbing wires



UTP characteristics

- Resistance per unit distance
- Capacitance per unit distance
- Inductance per unit distance
- Cross-admittance (assume pure reactive) per unit distance



Propagation loss

Voltage decreases as travel along cable

Each new section of cable reduces voltage by a factor

So the decrease is exponential

$$V_a / V_b = e^{-\gamma x} = H(f,x)$$

where x is distance between points a and b

We can calculate γ, and hence the loss, directly from the RCLG model

=> SNR decreases with distance

Why is DSL better than a voice-grade modem?

Analog telephony modems are limited to 4 KHz bandwidth

Shannon's channel capacity theorem gives the maximum transfer rate

$$C = Bw \log_2 (SNR + 1)$$

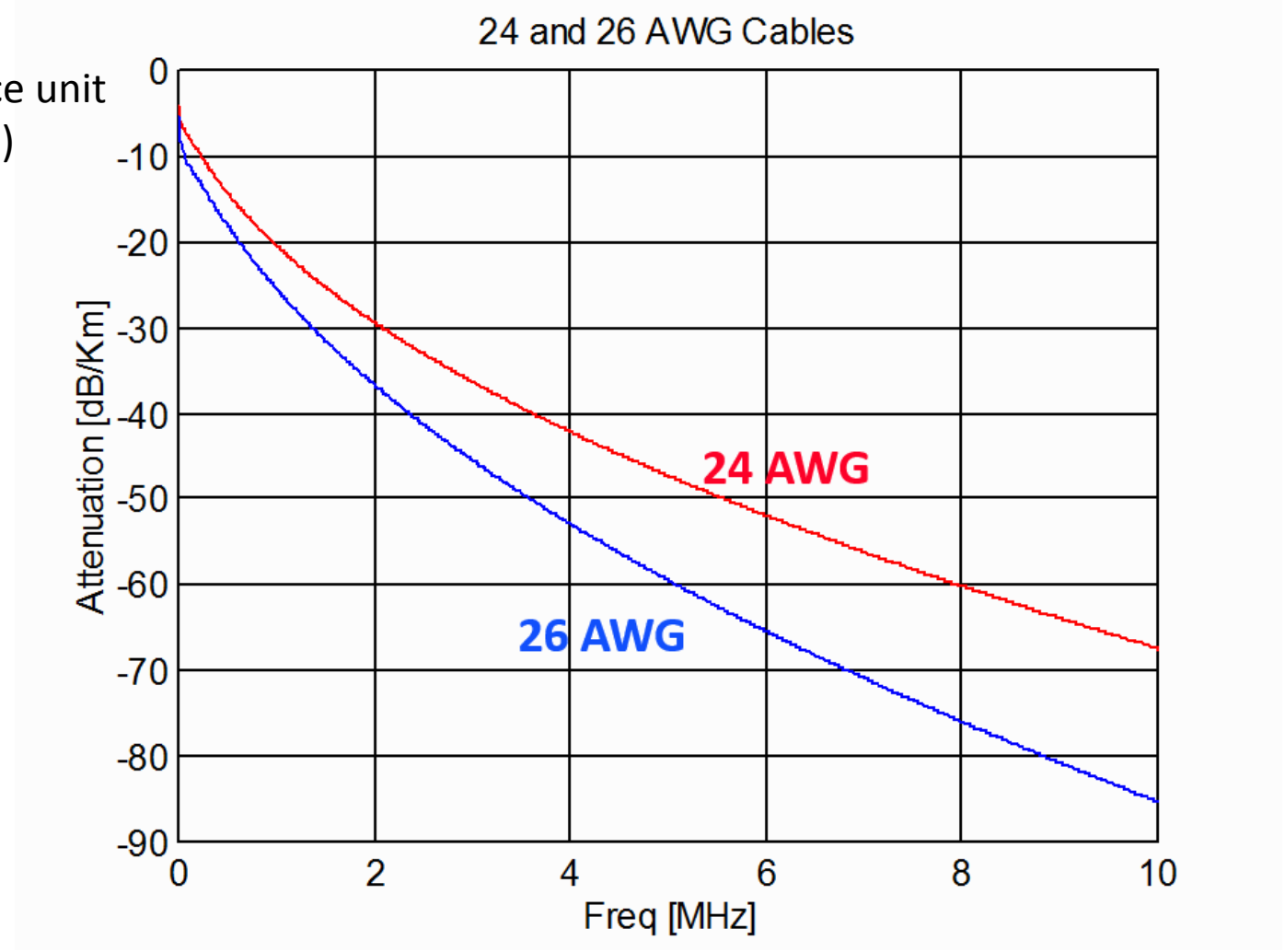
So by using more BW we can get higher transfer rates if SNR does not decrease

But: SNR decreases with distance **and frequency**

So what is the BW of UTP? And what how does SNR vary with BW ?

Attenuation vs. frequency

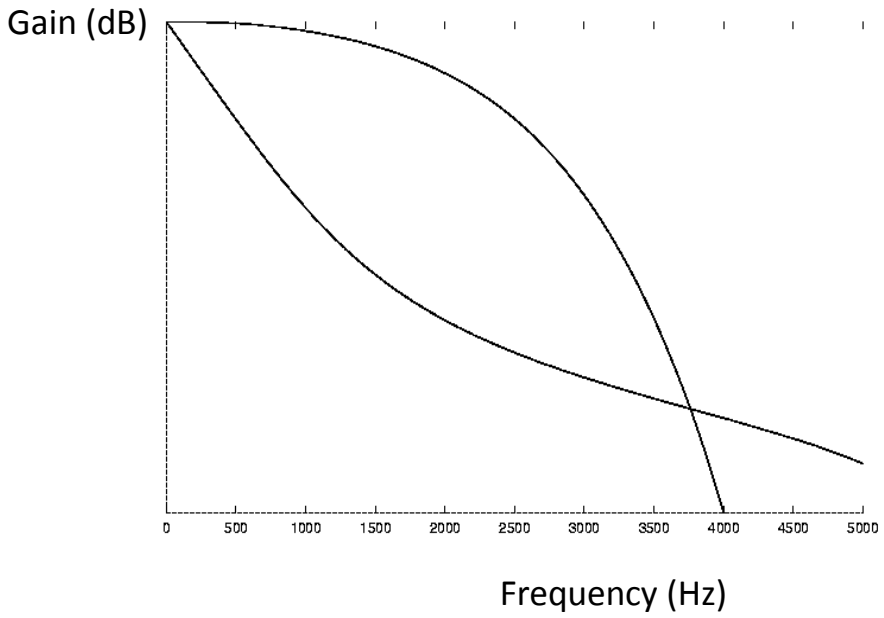
Per distance unit
SNR=f(freq)



Loading coils

Long loops have loading coils to prevent voice distortion

What does a loading coil do?



Flattens response in voice band

Attenuates strongly above voice frequencies

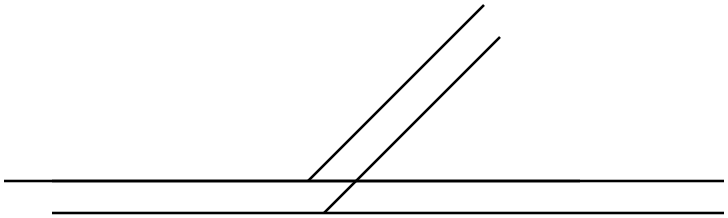
loops longer than 6 Km need loading coils: 88 mH every 2 Km starting 1 Km

Bridge taps

There may also be *bridged taps (from Party lines before 1940)*

Parallel run of unterminated UTP

- unused piece left over from old installation
- placed for subscriber flexibility



High frequency signals are reflected from the open end

A bridged tap can act like a notch filter!

Other problems

Splices

Subscriber lines are seldom single runs of cable

In the US, UTP usually comes in 170m lengths

So splices must be made every 170m

Average line has >20 splices

Splices are pressure connections that add to attenuation

Over time they corrode and may spark, become intermittent, etc.

Gauge changes

US binder groups typically start off at 26 AWG

Change to 24 AWG after 3 Km

In rural areas they may change to 19 AWG after that

Binder groups

UTP are not placed under/over ground individually

In central offices they are in cable *bundles*

with 100s of other UTP

In the outside plant they are in binder groups

with 25 or 50 pairs per group

These pairs interfere with each other

a phenomenon called cross-talk (XTALK)

CSA guidelines

1981 AT&T **C**arrier **S**ervice **A**rea guidelines advise as follows for new deployments

- No loading coils
- Maximum of 3 Km of 26 gauge (including bridged taps)
- Maximum of 4 Km of 24 gauge (including bridged taps)
- Maximum of 800m bridged taps
- Maximum single bridged tap 600m
- Suggested: no more than 2 gauges

In 1991 more than 60% of US lines met CSA requirements

Present US PSTN

We will see that for DSL - basically four cases

- Loaded line - no DSL possible
- Resistance design unloaded <6 Km <1300 Ω - ADSL
- CSA reach - HDSL
- DA (distribution area) 1-2 Km - VDSL

Higher rate - lower reach

(because of attenuation and noise!)

xDSL

What problems to overcome?

What goals to reach?

Alternatives for data services

Fiber, coax, Hybrid Fibre-Coax

COST: $10k-$20k / mile

TIME: months to install

T1/E1

COST: >$5k/mile for *conditioning*

TIME: weeks to install

DSL

COST: ≅ 0 (just equipment price)

TIME: ≅ 0 (just setup time)

xDSL

Need for higher speed digital connection to subscribers

Not feasible to replace UTP in the last mile

Older voice grade modems assume 4kHz analog line

Newer (V.90) modems assume 64kbps digital line

DSL modems don’t assume anything

Use whatever the physics of the UTP allows

Maximum reach

To use Shannon's capacity theorem we need to know how much noise there is

One type of noise that is always present (above absolute zero temperature) is thermal noise

Maximum reach is the length of cable for reliable communications at a given data rate, ASSUMING ONLY THERMAL NOISE

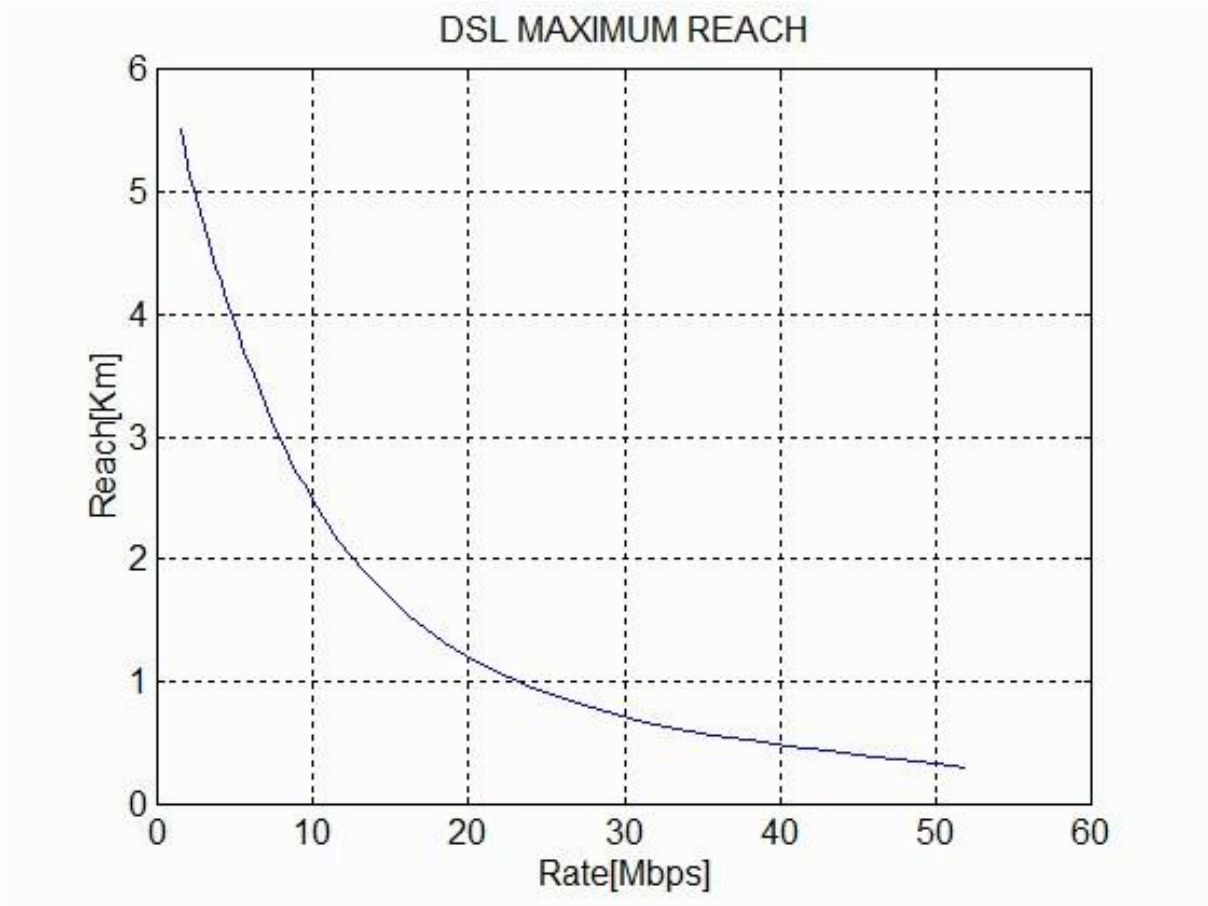
Bellcore study in residential areas (NJ) found

- -140 dBm / Hz
- white (i.e. independent of frequency)

is a good approximation

We can compute the maximum reach from known UTP attenuation

xDSL - Maximum Reach



Other sources of noise

But real systems have other sources of noise,

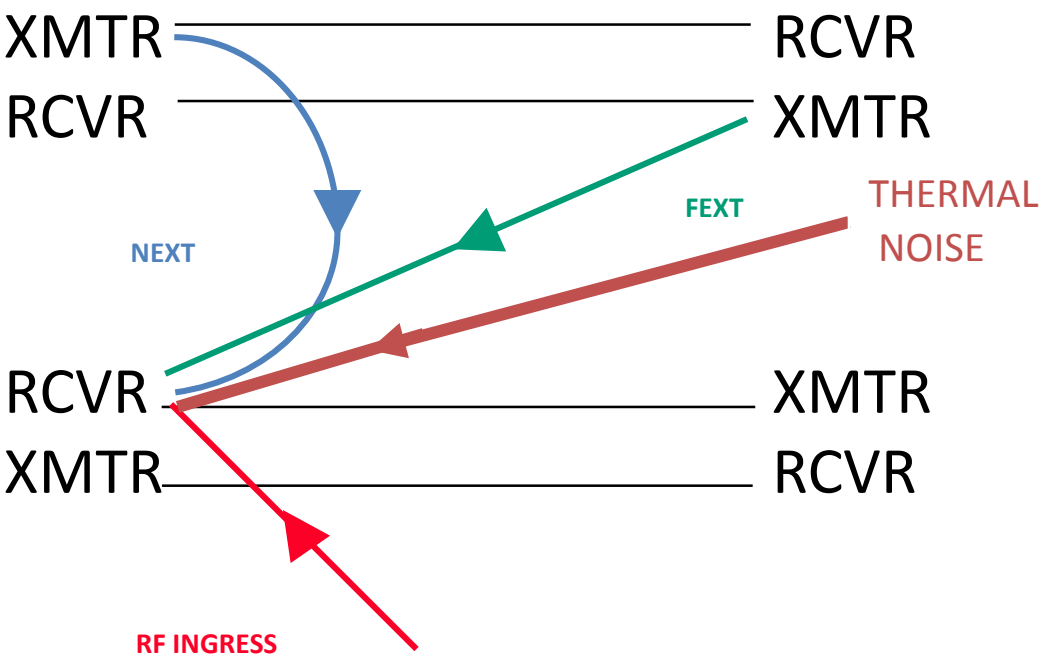
and thus the SNR will be lower

and thus will have lower reach

There are three other commonly encountered types of noise

- RF ingress
- Near End Cross Talk (NEXT)
- Far End Cross Talk (FEXT)

Sources of Interference



Duplexing

Up to now we assumed that only one side transmits

Bidirectional (full duplex) transmission
requires some form of ***duplexing***

For asymmetric applications we usually speak of
DS downstream and **US** upstream

Four methods are in common use:

- Half duplex mode
- Echo cancellation mode (ECH)
- **T**ime **D**omain **D**uplexing (requires syncing all binder contents)
- **F**requency **D**omain **D**uplexing

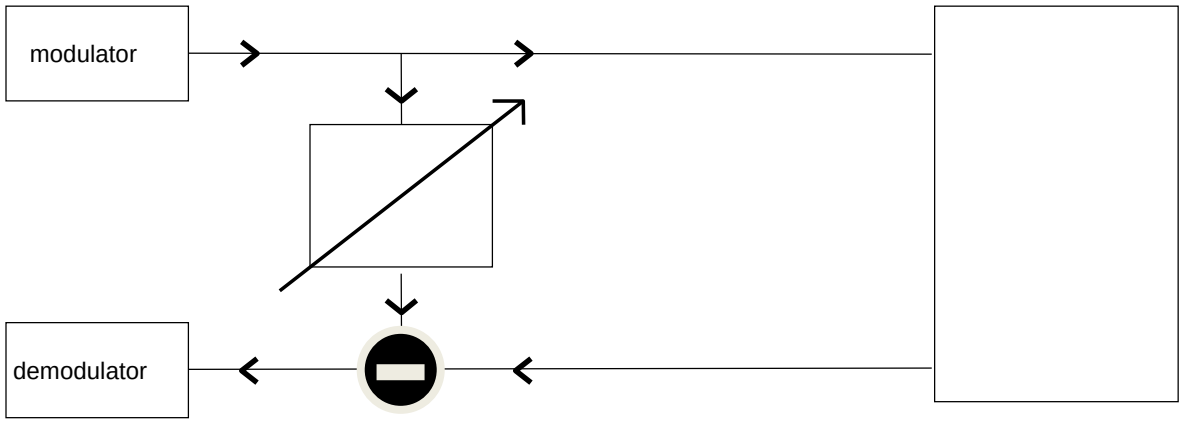
(Adaptive) echo cancellation

Signal transmitted is known to transmitter

It is delayed, attenuated and distorted in the round-trip

Using adaptive DSP algorithms we can

- find the delay/attenuation/distortion
- subtract



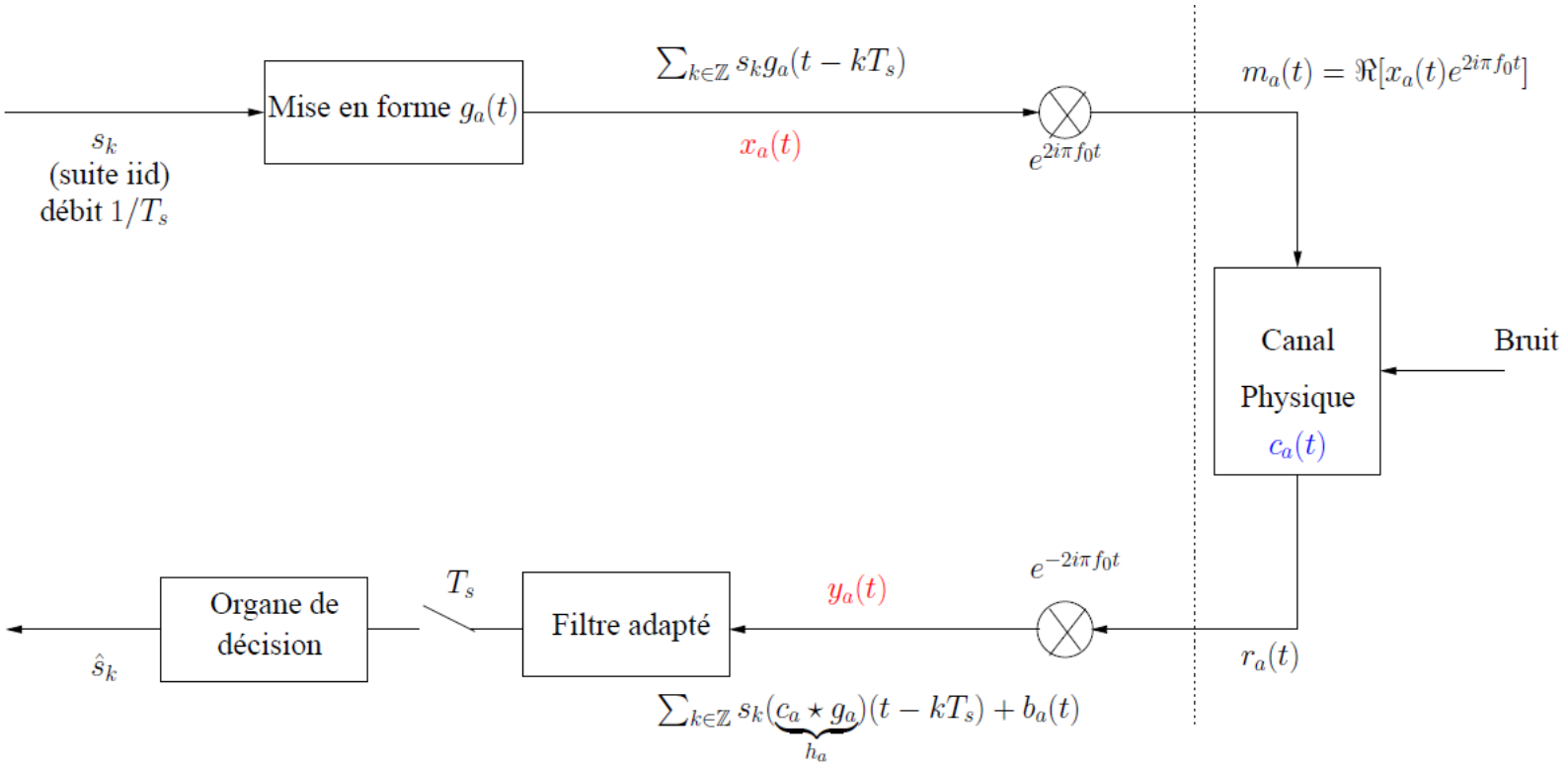
Plan général du cours

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
V. Accès : Réseaux optiques
VI. Accès : Technologies xDSL
    - VI.1. Principe : tirer le maximum de la paire de fils de cuivre
    - VI.2. Technique de TS pour xDSL : l'OFDM
    - VI.3. Fonctionnement protocolaire entre le FAI et l'abonné (cas dégroupé et non-dégroupé)

Principe de l'OFDM:
Orthogonal Frequency Division Multiplexing

Système mono-porteuse

- Modulation linéaire



Canal $c_a(t) \neq \delta(t) \Rightarrow \tilde{h}_a(t)$ différent de Nyquist.

=> Interférence entre symboles (ISI)

Canaux difficiles: Interférence entre symboles

$$y_a(t) = c_a \star x_a(t)$$

- Si $c_a(t) = \lambda\delta(t)$, alors $y_a(t) = \lambda x_a(t) + b_a(t)$

- Si $c_a(t) = \sum_{l=1}^{L_c} \lambda_l \delta(t - \tau_l)$ (cas multi-trajets),

  alors $y_a(t) = \sum_{l=1}^{L_c} \lambda_l x_a(t - \tau_l)$ (somme de versions retardées)

  $\longrightarrow$ Interference entre symboles possible !! (IES)

Interférence entre symboles

- L'IES n'arrive pas si

$$\tau_{max} << T_s \tag{1}$$

- Or on a (théorème de Shannon): $\frac{1}{T_s} > 2f_{max} \rightarrow T_s < \frac{1}{Bw_{signal}}$.

  Donc (1) est assuré si:

$$\tau_{max} << \frac{1}{Bw_{signal}}$$

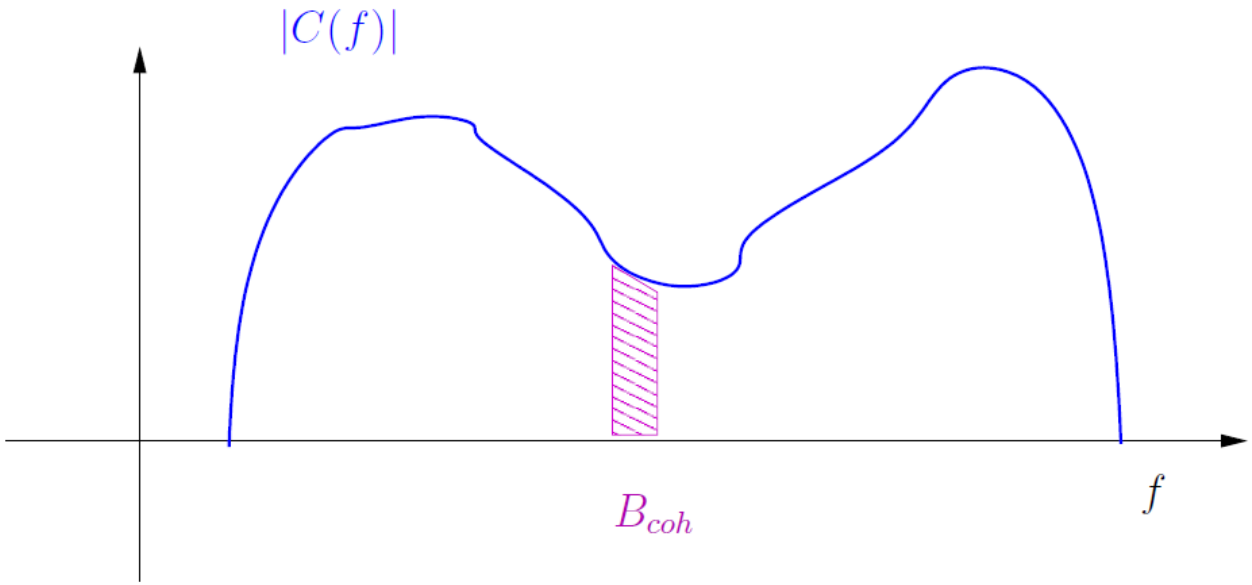
  Soit la bande de cohérence du canal $B_{coh} = \frac{1}{\tau_{max}}$

- Alors les IES sont évitées si

$$Bw_{signal} << B_{coh}$$

Le canal $c_a(t)$ (1/2)

Canal multi-trajets : $c_a(t) = \sum_{l=1}^{L_c} \lambda_l \delta(t - \tau_l)$.

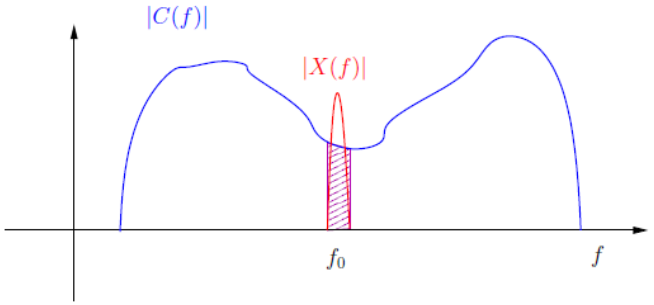


B_c : bande de cohérence du canal.

Le canal $c_a(t)$ (2/2)

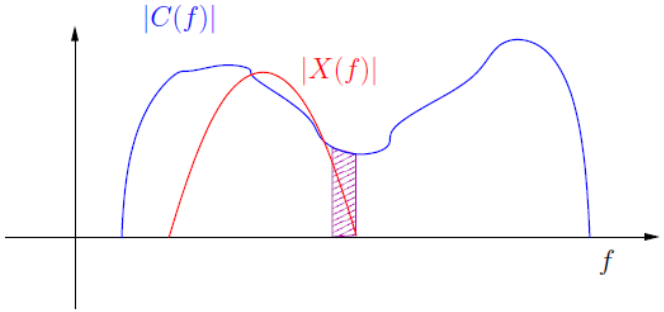
Soit B la bande du signal émis :

- $B < B_c$: seulement atténuation et déphasage



$$Y(f) \approx C(f_0)X(f) \Rightarrow y_a(t) \approx C(f_0)x_a(t)$$

- $B > B_c$: IES dû au canal physique $c_a(t)$.



$$Y(f) = C(f)X(f) \Rightarrow y_a(t) = c_a(t) \star x_a(t)$$

Comment se retrouver dans une configuration $B < B_c$?

Principe du multi-porteuses

Rappel : B requise $=$ Débit-symbole $(1/T_s)$

Idée naïve : Diviser la suite des symboles en N sous-suites de symboles (de période $T = NT_s$) telles que

$$\frac{1}{T} < B_c.$$

Chaque sous-suite n étant émise sur une sous-bande différente associée à une sous-porteuse f_n.

Intérêt : Pour chaque sous-bande, pas d'IES.

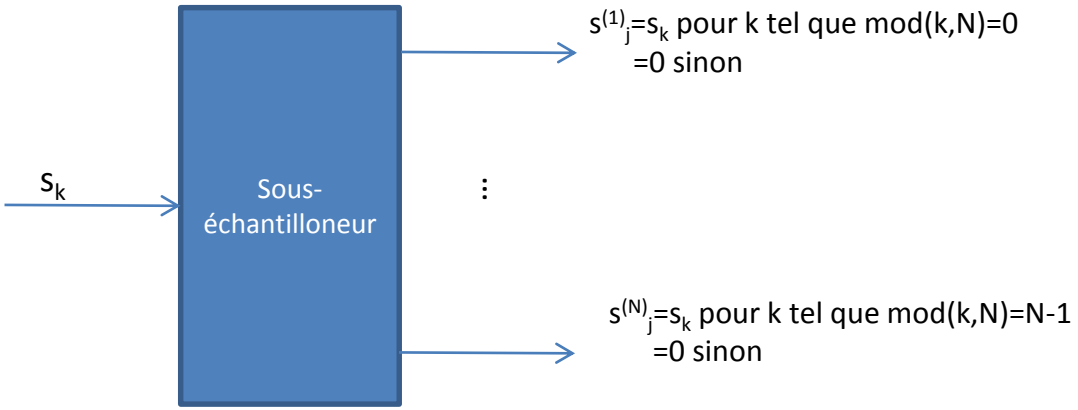
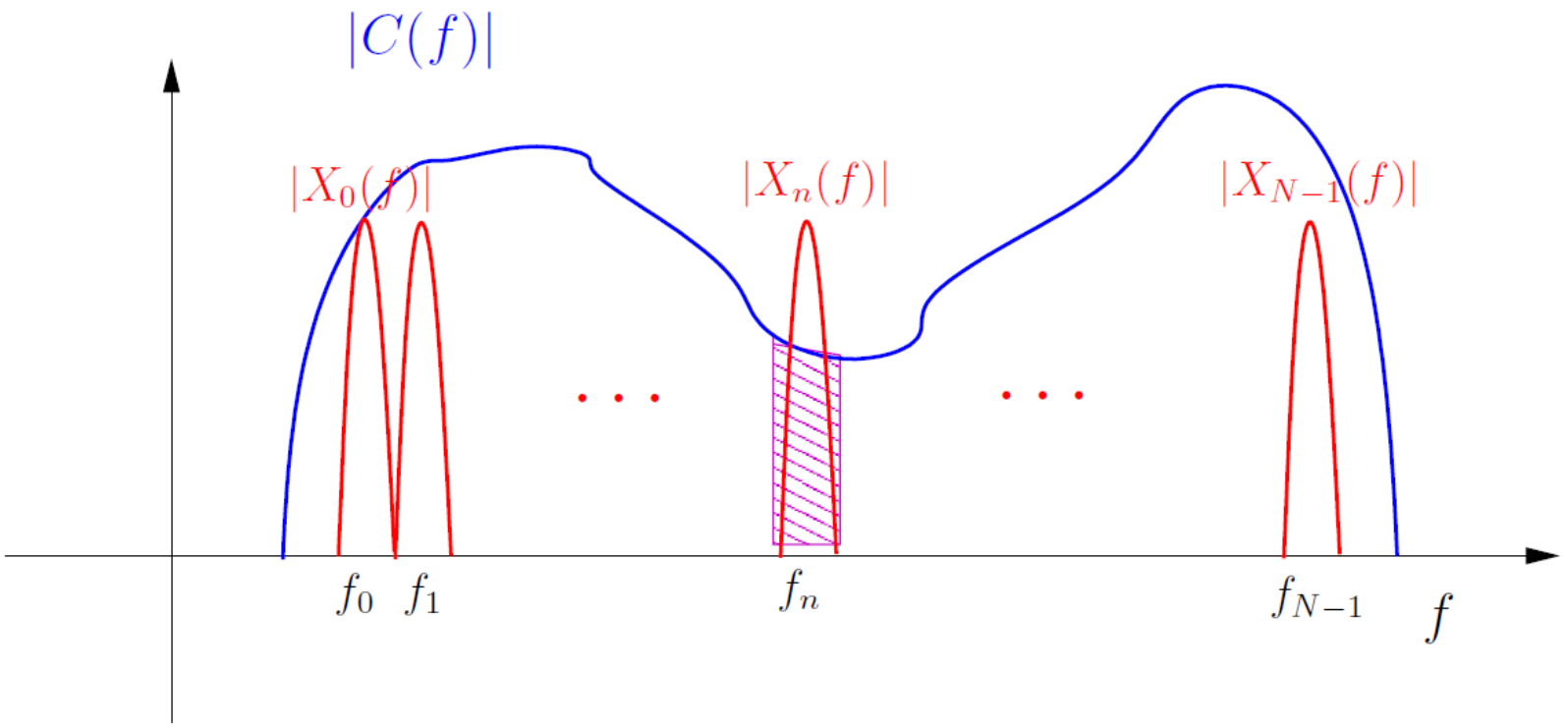
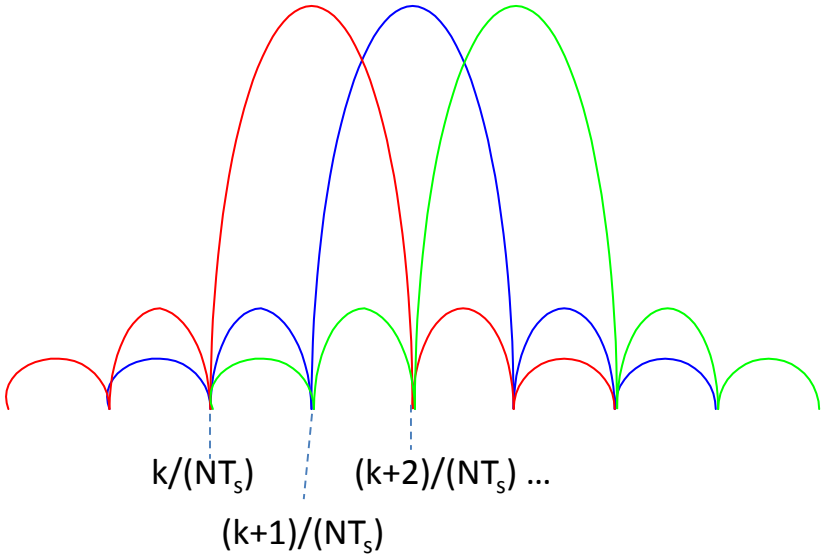


Schéma récapitulatif

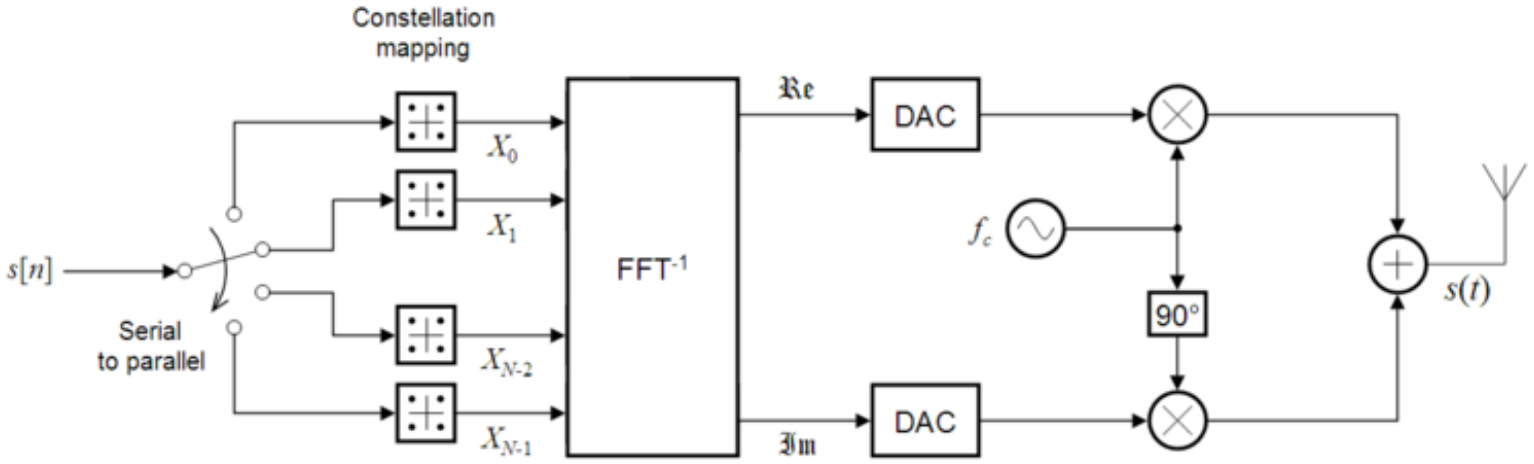


OFDM



- Sinc function has zero at center of nearby modem
- This implies that the signals are orthogonal - no ICI
- No guard frequencies are needed
- Don’t need N independent modems
  - efficient digital implementation by **FFT** algorithm

OFDM Transmitter

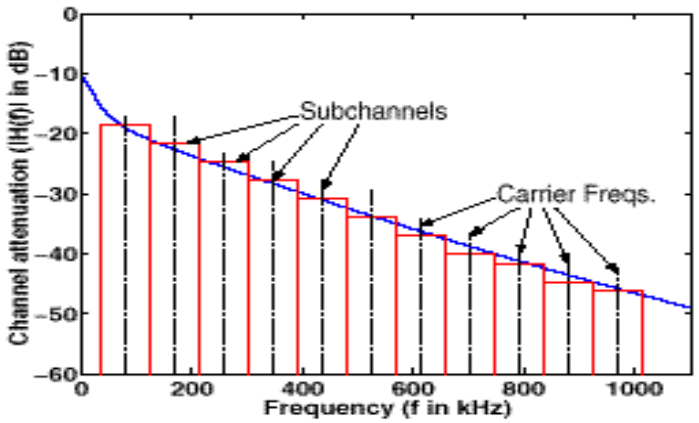


- Chaque séquence X_i résultant du sous-échantillonnage de s[n] est mise sur une porteuse différente

- Chaque porteuse est espacée de $1/(NT_s)$ si s[n] provient d'un échantillonnage à la période T_s

Discrete Multi-Tone (DMT)

- Measures SNR(f) during initialization
- Water fill QAM signals according to SNR(f)
- Symbol duration > channel impulse response time --- no ISI
- No equalizer required



xDSL

Modems et débits

Le modem ADSL

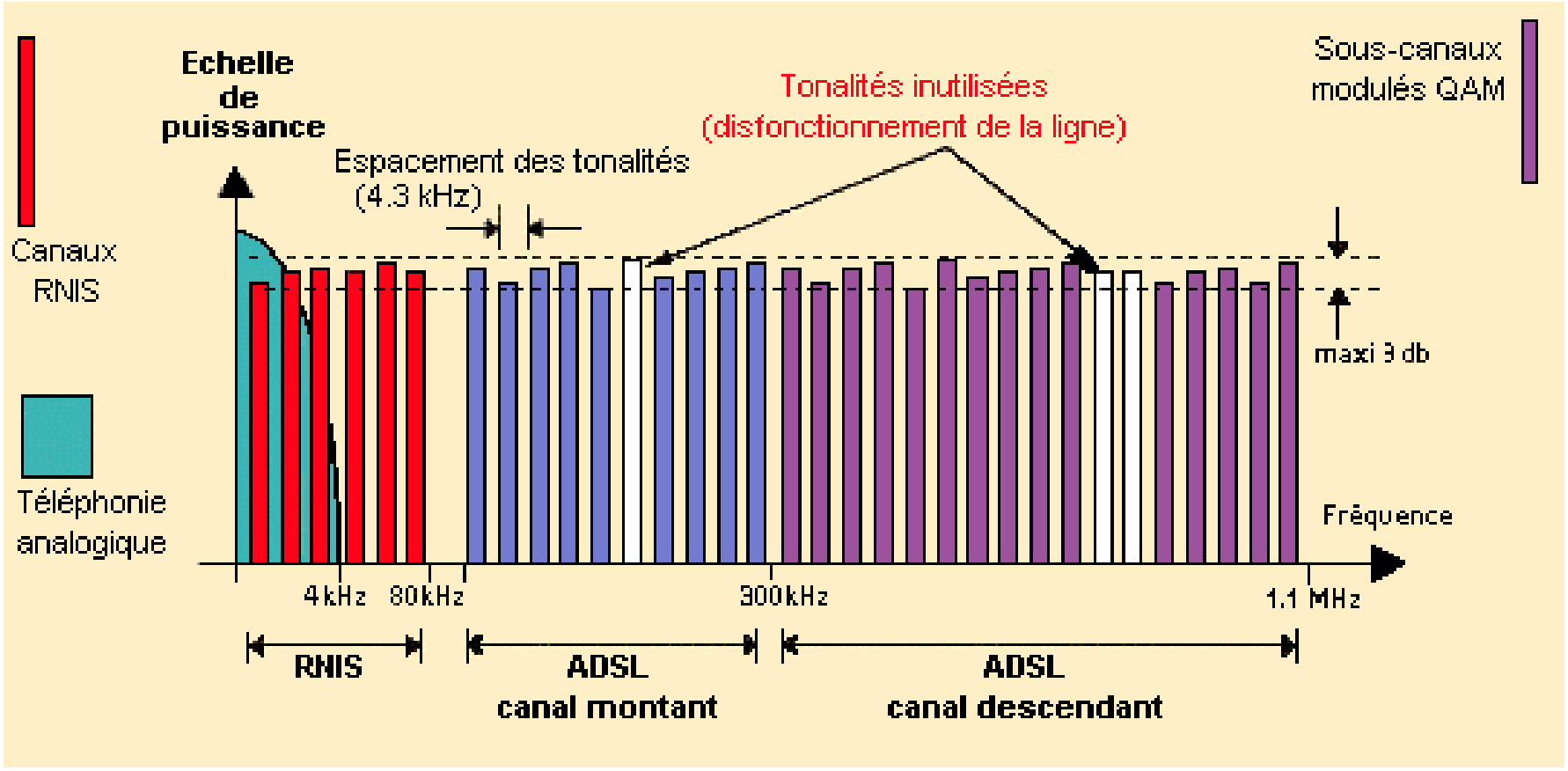
- La technologie ADSL a été normalisée par l'UIT-T sous la recommandation G.992.1.
- Une extension a été apportée dans la recommandation G. 992.3 déterminant l'ADSL2.

- Des extensions permettant l'amélioration de l' ADSL2 sont adoptées pour allonger la distance entre le client et le DSLAM. Ces améliorations ont permis l'adoption de l'ADSL2+ ou LDSL (Long-reach DSL) ou encore READSL (Range Extended ADSL).

- Les modems ADSL offrent une bande montante de 4 à 100 kHz, qui est utilisée pour des débits de 0,64 Mbit/s. La bande descendante utilise une bande comprise entre 100 kHz et 1,1 MHz, qui permet d'atteindre le débit de 8,2 Mbit/s.

- La parole analogique, entre 0 et 4 kHz, passe en même temps que les données utilisant le modem.

Spectre de signal sur ligne téléphonique



Modulation et débits

- La méthode DMT (Discrete MultiTone) :
  - partie US du spectre divisée en 20 sous-bandes de 4,3 kHz
  - chaque sous-bande est capable de transporter de 4 à 15 bits en parallèle
    - On choisit 8 bits par symb (i.e., par intervalle d'horloge)

- 4000 IT/s (4000 Bauds=nb symb/s) -> 4 000 x 8 bits = 32 Kbps par sous-bande
- 20 sous-bandes -> total de 32 x 20 = 640 Kbps
- 256 sous-bandes pour DS -> total de 4 000 x 8 bits x 256 = 8,2 Mbps
- Débit peut être augmenté en: augmentant le nb de bits par symb et/ou la BW

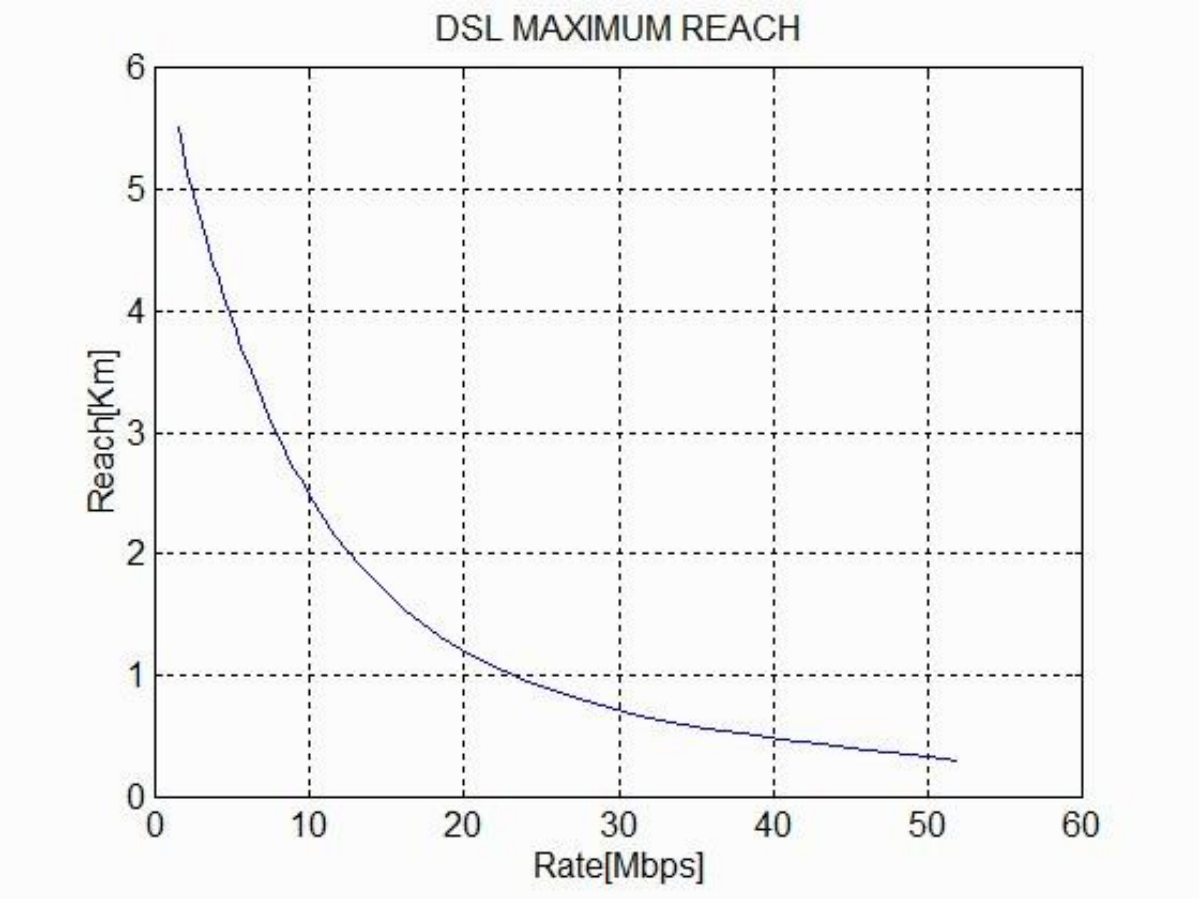
- Les modems ADSL2+ permettent ainsi de monter jusqu'à 28 Mbps en augmentant fortement la partie du spectre DS (jusqu'à 2,2 MHz).

Les différentes technologies xDSL

modem	speed	reach	main applications
ADSL	6 Mbps DS 640 Kbps US	3.5-5.5 km	residential Internet, video-on-demand
ADSL2	8 Mbps DS 800 Kbps US	> ADSL	Internet access, VoIP
ADSL2+	16 Mbps DS 800 Kbps US	< 2 km	"
VDSL	<= 52 Mbps	300m - 1 km	LAN interconnect, HDTV, combined services
VDSL2	200 Mbps (aggregate)	up to 1.8 km	"
cable modem	10-30Mbps DS shared	50 km	residential Internet
HPNA	1, 10 Mbps	home wiring	residential networking

Not DSL

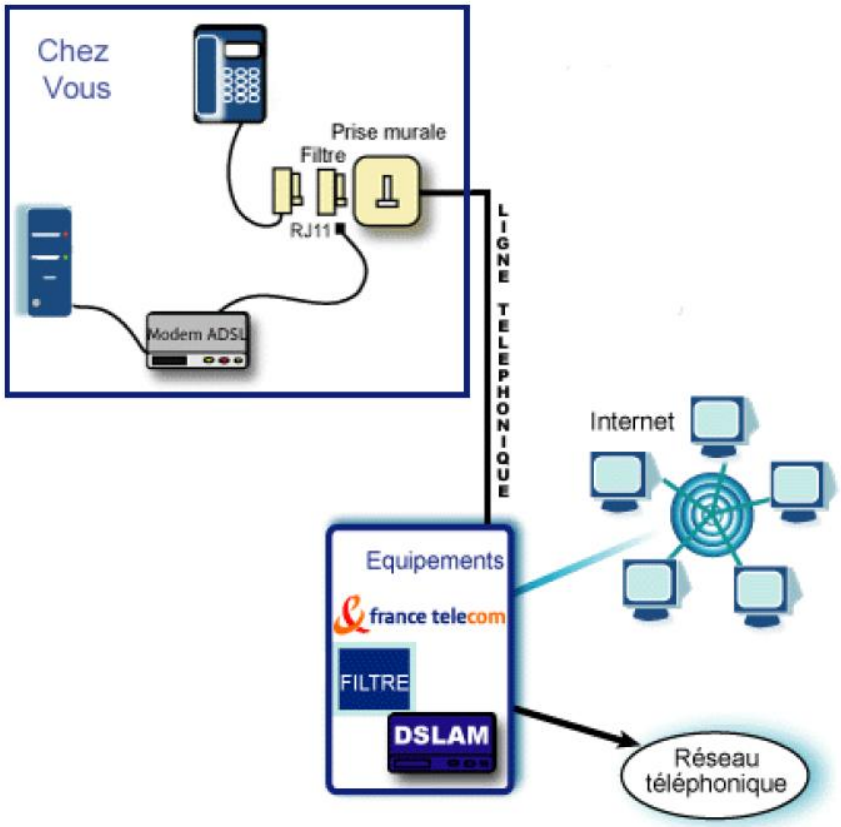
xDSL - Maximum Reach



Plan général du cours

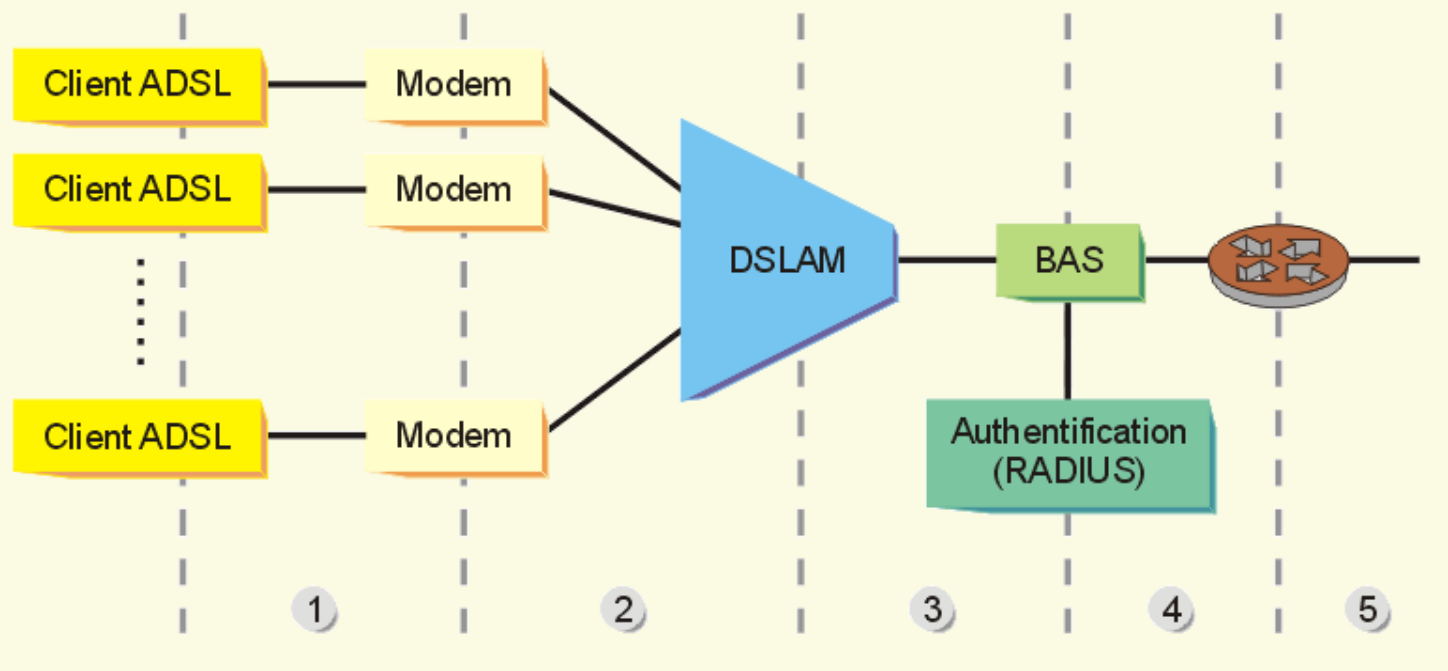
I. Organisation des opérateurs de l'Internet
II. TCP et Qualité de service
III. Commutation par circuits virtuels (VC) : le cas d'ATM
IV. Commutation par VC dans le mode IP : MPLS
V. Accès : Réseaux optiques
VI. Accès : Technologies xDSL
    - VI.1. Principe : tirer le maximum de la paire de fils de cuivre
    - VI.2. Technique de TS pour xDSL : l'OFDM
    - VI.3. Fonctionnement protocolaire entre le FAI et l'abonné (cas dégroupé et non-dégroupé)

xDSL System Reference Model

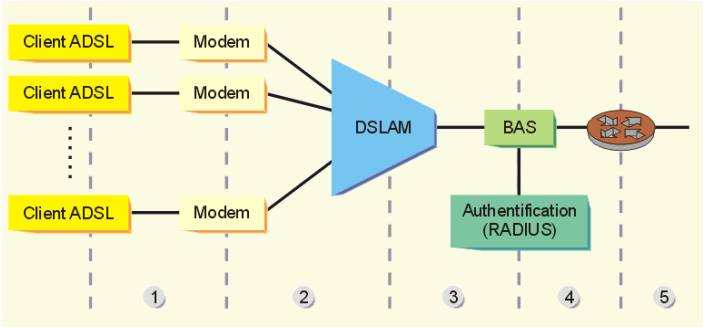


Non-dégroupé: Schéma simplifié

Le schéma qui suit, simplifié à l'extrême, n'est pas pour autant faux. Il peut convenir dans le cas du dégroupage

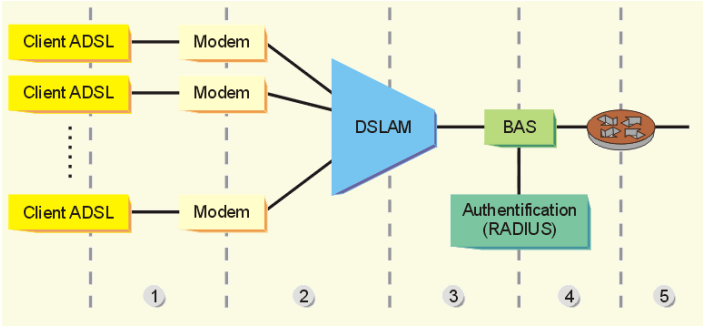


Les composants



- **Client ADSL**
  - La machine que l'abonné connecte à l'internet. Peu importent la plate forme matérielle et le système d'exploitation, pourvu que ce dernier supporte le réseau TCP/IP.
- **Modem**
  - Modulateur/Démodulateur. Une boîte dont la fonction est assez similaire à celle du modem RTC, à part qu'ici, elle est conçue pour la technologie DSL.
- **DSLAM**
  - Digital Subscriber Line Access Multiplexer. Multiplexage sur un seul lien à fort débit.
- **BAS**
  - Broadband Access Server
  - Lorsque l'on a réussi une connexion avec son FAI, on a établi un lien PPP (Point to Point Protocol) entre son ordinateur et le BAS. Ce lien PPP va transporter les protocoles supérieurs IP, TCP, UDP, ICMP...
    C'est au niveau du BAS que l'authentification du client va se faire et que les paramètres IP vont être transmis (serveur RADIUS, généralement).
- **Routeur**
  - C'est l'équipement qui va assurer la liaison entre le BAS et le réseau du fournisseur d'accès.
  - Le lien n°4 relie le BAS à ce routeur et les données circulent dans un tunnel de type L2TP (Layer 2 Tunnel Protocol). Il s'agit de construire un VPN (Virtual Private Network : réseau privé virtuel) entre le BAS et le réseau du fournisseur d'accès.

Les liaisons



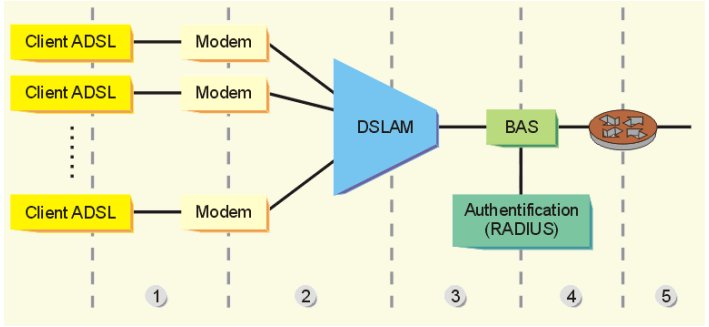
- 1. Représente le "bout de câble" qui relie votre machine au modem. Ce bout de câble peut être de type :
  - Ethernet, USB, Hertzien (pas de câble, mais une onde électromagnétique), dans le cas d'un modem/routeur WiFi.
- 2. Représente votre ligne téléphonique, c'est une paire de fils de cuivre qui va de votre prise murale jusqu'au concentrateur de l'opérateur de téléphonie. Cette ligne va véhiculer la téléphonie traditionnelle (POTS) et les porteuses ADSL. La séparation des services se fait au moyen d'un filtre séparateur (Splitter) qui peut être extérieur, mais aussi intégré au modem.
- 3. Représente le lien à fort débit (généralement de la fibre optique) qui véhicule les données numériques de l'ensemble des abonnés connectés au même DSLAM.
- 4. Est encore un lien à fort débit, qui relie le BAS au réseau de votre FAI.
- 5. Nous sommes maintenant sur le réseau du fournisseur d'accès.

Les protocoles du réseau

- **ATM**
  - Les opérateurs de télécom aiment ATM (Asynchronous Transfert Mode). Pourquoi ? Parce que ATM est très bien adapté au transport de paquets commutés, avec des circuits (virtuels) point à point, avec une gestion de QoS (Quality of Service) fine, autant de services utiles pour passer sur le même réseau de la téléphonie, de la vidéo et des données informatiques.

- **Ethernet**
  - Comparativement, les réseaux informatiques aiment Ethernet, qui n'a absolument rien à voir avec ATM, si ce n'est que, dans le modèle OSI à 7 couches, les deux protocoles pourraient à la rigueur être situés au même niveau (2). C'est en grande partie faux, ATM peut remplacer non seulement Ethernet mais aussi IP.
  - Mais pour ce qui nous intéresse ici, ATM transportera de l'IP ou du PPP et donc jouera un rôle similaire à Ethernet.

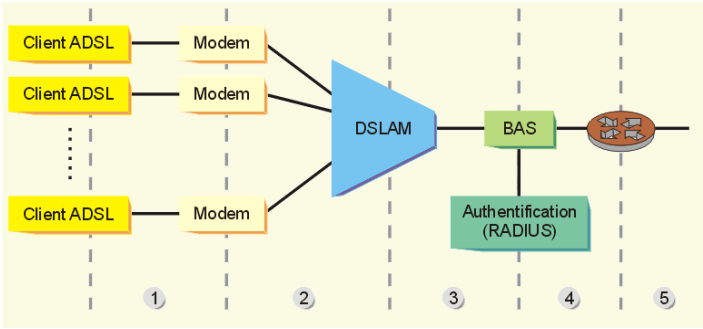
Les protocoles



- Le tronçon 3 de l'illustration a comme support ATM.
- Le tronçon 4 est spécifié comme du "réseau IP", mais par quoi est transportée la couche IP ? Seul l'opérateur du réseau pourra le dire. IP peut en effet être transporté par ATM, par Ethernet, par Ethernet sur ATM...
- Ce qui peut paraitre surprenant, c'est que le tronçon 2 (entre votre prise téléphonique et le DSLAM) utilise également ATM. Autrement dit, votre modem ADSL est un terminal ATM.
- Entre votre ordinateur et votre modem, ça pourra être :
  - de l'Ethernet si :
    - votre modem dispose d'une prise Ethernet et votre machine d'une interface réseau également Ethernet,
    - votre modem est un modem USB, qui n'est qu'une liaison série, sur laquelle on va simuler un lien Ethernet, grâce au driver à installer avec le modem,
  - de l'ATM si :
    - votre modem dispose d'une prise ATM et votre machine d'une interface réseau également ATM (cas fort peu probable, vu le prix),
    - votre modem est un modem USB, qui n'est qu'une liaison série, sur laquelle on va simuler un lien ATM, grâce au driver à installer avec le modem.

NB: La trame ATM est aujourd'hui de plus en plus remplacée par la trame Ethernet. Cette solution s'explique aisément pour réaliser une continuité Ethernet entre la connexion de l'utilisateur et l' utilisation de solution Gigabit Ethernet entre le DSLAM et le réseau de l' opérateur.

Non-dégroupé: Les protocoles



- **Et par dessus tout ça...**
  - On va construire un lien PPP, qui doit aller de votre ordinateur jusqu'au BAS, et qui sera porté :
    - par Ethernet (PPPoE : PPP over Ethernet) entre votre machine et le modem, si la liaison en dessous est de type Ethernet,
    - par ATM (PPPoA : PPP over ATM) entre votre machine et le modem, si la liaison en dessous est de type ATM.
  - Entre le modem et le BAS, souvent du PPPoA.
  - Ce lien PPP transportera IP entre votre ordinateur et le BAS. Comme c'est du PPP, il y a une IP à chaque bout :
    - la vôtre, généralement dynamique, qui vous est attribuée lors de l'établissement de la session PPP par le serveur d'authentification

Dans la pratique: dégroupage et non-dégroupage

- Dans la pratique, ça va être un peu plus compliqué, du moins dans le cas du non-dégroupage
- Qu'est-ce que le dégroupage ?
  - La boucle locale appartient à France Télécom. Les centres de raccordement (NRA) aussi.
  - Ces centres de raccordement sont souvent situés dans des locaux exigus, où il est difficile de laisser de la place à des opérateurs alternatifs. Pourtant, le dégroupage doit en passer par là.
  - Pour dégrouper une ligne, il faut qu'un ou plusieurs opérateurs alternatifs puissent installer leurs propres équipements dans le centre, et y connecter directement les lignes de leurs abonnés. France Télécom se contente de louer la ligne entre le centre et le client.
  - Lorsque ceci n'est pas fait, peut importe la raison, il faut trouver une autre solution pour que des opérateurs alternatifs puissent tout de même proposer des abonnements ADSL.

- **Les offres France Télécom**

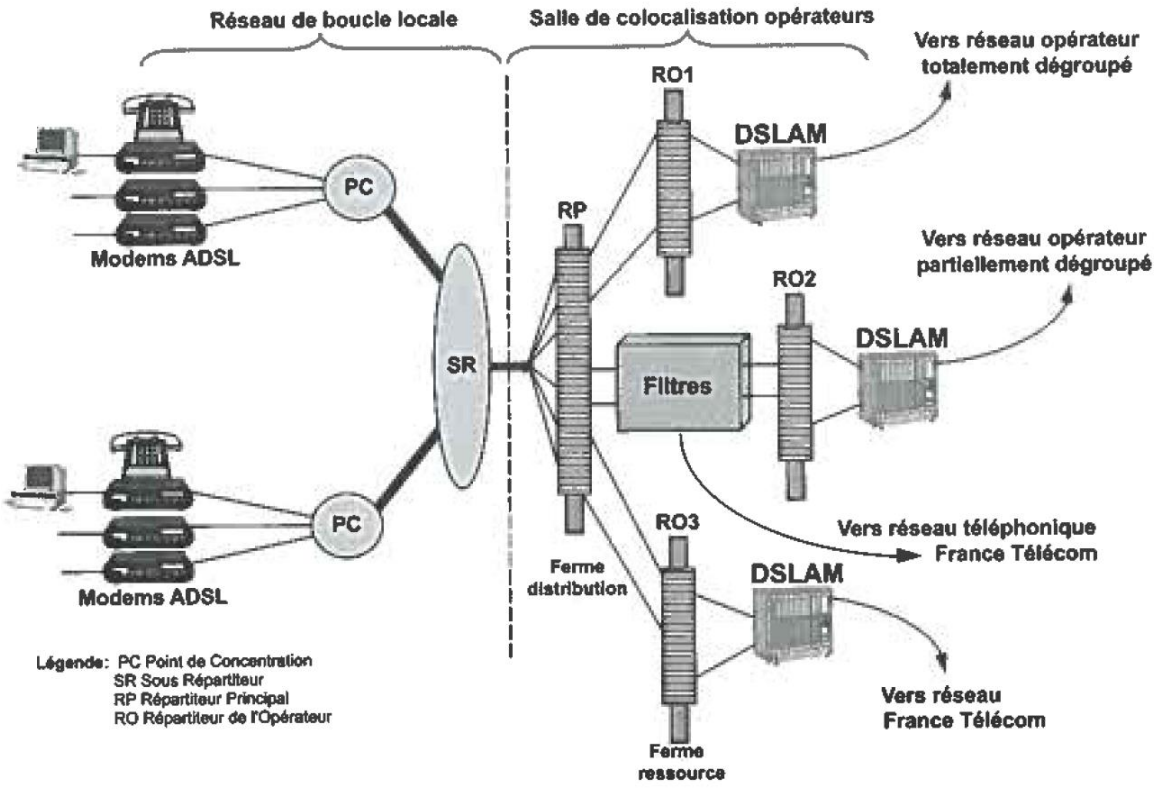
  - **L'option 1**
    - Non-dégroupé: France Télécom assure la totalité du transport entre l'abonné et le réseau du FAI.

  - **L'option 2**
    - Dégroupage: FT se contente de louer l'accès à la boucle locale aux FAIs. Les fournisseurs sont donc maîtres de la qualité de leur prestation. Ils ne sont dépendants que de la qualité physique de la boucle locale.

Dégroupage

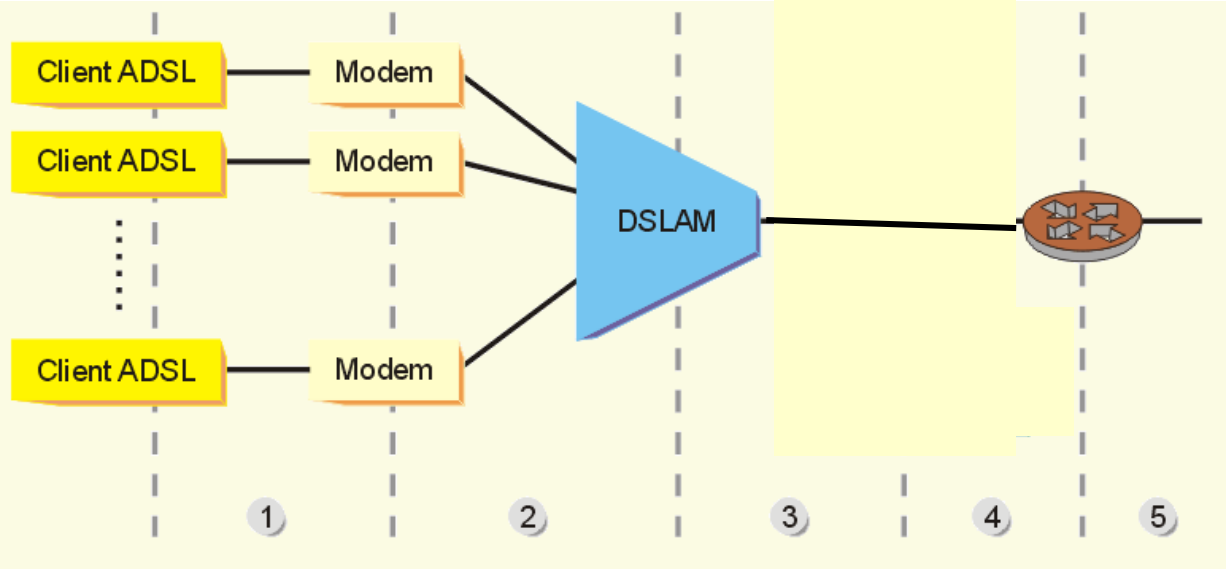
- Le **répartiteur téléphonique** est le lieu situé dans le CO (ou NRA) où se font toutes les connexions entre la LL et les infrastructures

- Aucun dégroupage: FAI absent du NRA
- Dégroupage partiel: FAI présent au NRA, mais abonnement FT encore possible
- Dégroupage total: FAI présent au NRA, abonnement FT impossible (téléphonie assurée sur spectre ADSL)



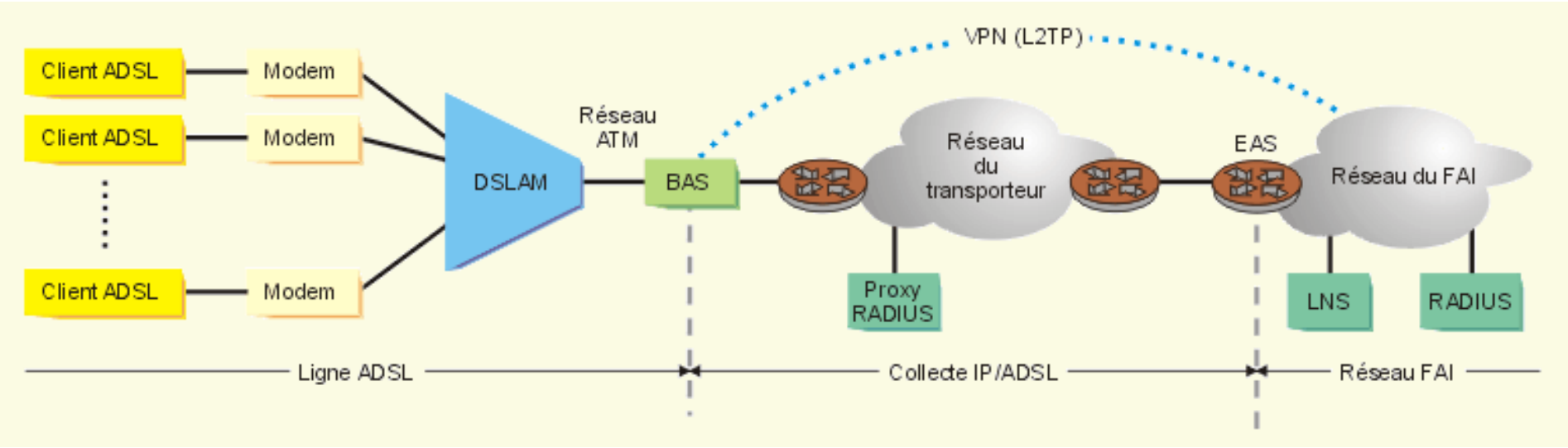
En dégroupé

- DSLAM du FAI présent au CO
- IP over ATM entre le modem et le FAI
- Adresse IP souvent fixe --> plus de PPP

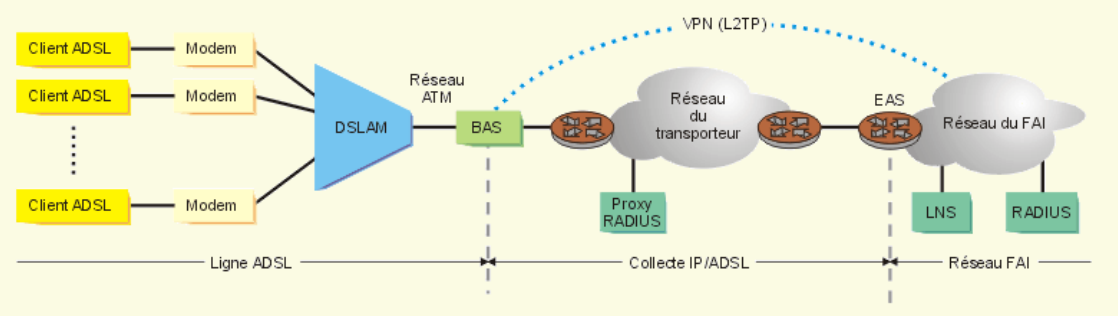


Le "vrai" schéma en non-dégroupé

- Nous allons donc détailler le cas de l'option 1.
- L'option 2 se rapproche beaucoup du schéma simplifié, vu plus haut.
- Dans le cas de la collecte IP/ADSL (option 1), c'est à dire le cas où FT prend totalement en charge l'acheminement des données depuis l'abonné jusqu'au FAI, tout ce qui est situé entre le modem de l'abonné et l'EAS (routeur assurant l'interconnexion avec le réseau du FAI) est géré par France Télécom.



Non-dégroupé: principe



- Permettre la mise en place d'un lien PPP entre l'utilisateur (abonné ADSL) et le serveur d'accès de son FAI à travers le réseau France Télécom.

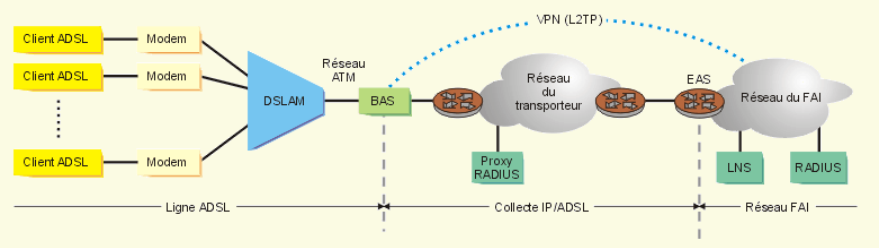
- Ce lien PPP va être supporté par le réseau ATM entre le modem de l'utilisateur et le BAS, puis par un tunnel L2TP (VPN) entre le BAS et le serveur du FAI.

- Cette méthode va permettre à plusieurs FAI de proposer leurs services aux abonnés à une ligne fixe FT.

- En effet, il est possible de créer dans le réseau IP de FT autant de VPN qu'il sera nécessaire pour connecter les utilisateurs à leur FAI respectif.

- La collecte ADSL a pour but de fournir la structure matérielle nécessaire à l'établissement du tunnel L2TP entre le BAS et les équipements du FAI concerné.

Non-dégroupé: quoi de plus



- Toute la partie "Collecte IP/ADSL". Cette partie a pour objet de créer un tunnel VPN entre le BAS et le réseau du FAI et de faire passer dans ce tunnel le lien PPP qui relie le client à son fournisseur. Pratiquement, ça permet d'utiliser le réseau IP France Télécom pour relier le BAS aux installations du FAI, qui peuvent être distantes, comme si elles étaient directement raccordées au BAS.
- Autrement dit :
  - Le client est authentifié via le proxy RADIUS fourni par FT. Un proxy va servir d'intermédiaire entre le client ADSL, quel que soit son fournisseur, et le serveur RADIUS de son fournisseur,
  - le tunnel VPN est construit sur L2TP (Layer 2 Tunneling Protocol) entre le BAS et le LNS (L2TP Network Server) du fournisseur,
  - ce tunnel va permettre le passage du lien PPP entre le client ADSL et le réseau du FAI, comme si ce lien était établi directement entre le BAS et le réseau du fournisseur, comme on l'a vu dans l'approche simplifiée.
- L'EAS est un routeur fourni par France Télécom, qui permet d'interconnecter le réseau IP FT à celui du fournisseur.
- --> C'est plus compliqué, mais ça ne change strictement rien au fonctionnement de la connexion, à partir de la couche IP.

Pour finir: Etablir PPP côté abonné

- Nous savons maintenant que le lien PPP va voyager "over ATM" entre le modem et le DSLAM. Reste à savoir comment le faire voyager entre votre ordinateur (ou routeur domestique) et votre modem.

- **Vous avez un modem Ethernet**
  - Pas le choix, ayant un lien Ethernet entre la machine et le modem, il faudra utiliser PPPoE.

- **Vous avez un modem USB**

  USB est une liaison série. Vous avez ici le choix de la couche de transport de PPP :
  - créer une couche Ethernet sur la liaison série et utiliser PPPoE,
  - créer une couche ATM sur la liaison série et utiliser PPPoA.
    Ceci est possible par l'installation sur l'ordinateur du driver approprié (suivant la disponibilité pour un type de modem donné).
- Q: Puisqu'on est sur une liaison série (USB), pourquoi ne pas faire passer PPP directement dessus, puisque c'est possible, et éliminer une couche inutile ? Elle n'est pas inutile. PPP directement sur une liaison série ne sait pas attribuer d'adresse MAC, adresse utilisable au niveau 2 du modèle OSI, et l'ajout de la couche Ethernet ou ATM contourne cette limite de PPP.

- **Vous avez un modem/routeur intégré**

  Vous aurez le choix que vous aura donné le constructeur (généralement PPPoE).

Etablissement d'un tunnel L2TP

- Manque de confidentialité de L2TP -> souvent implémenté avec Ipsec
  on parle alors de L2TP/Ipsec (standardisé dans IETF RFC 3193)

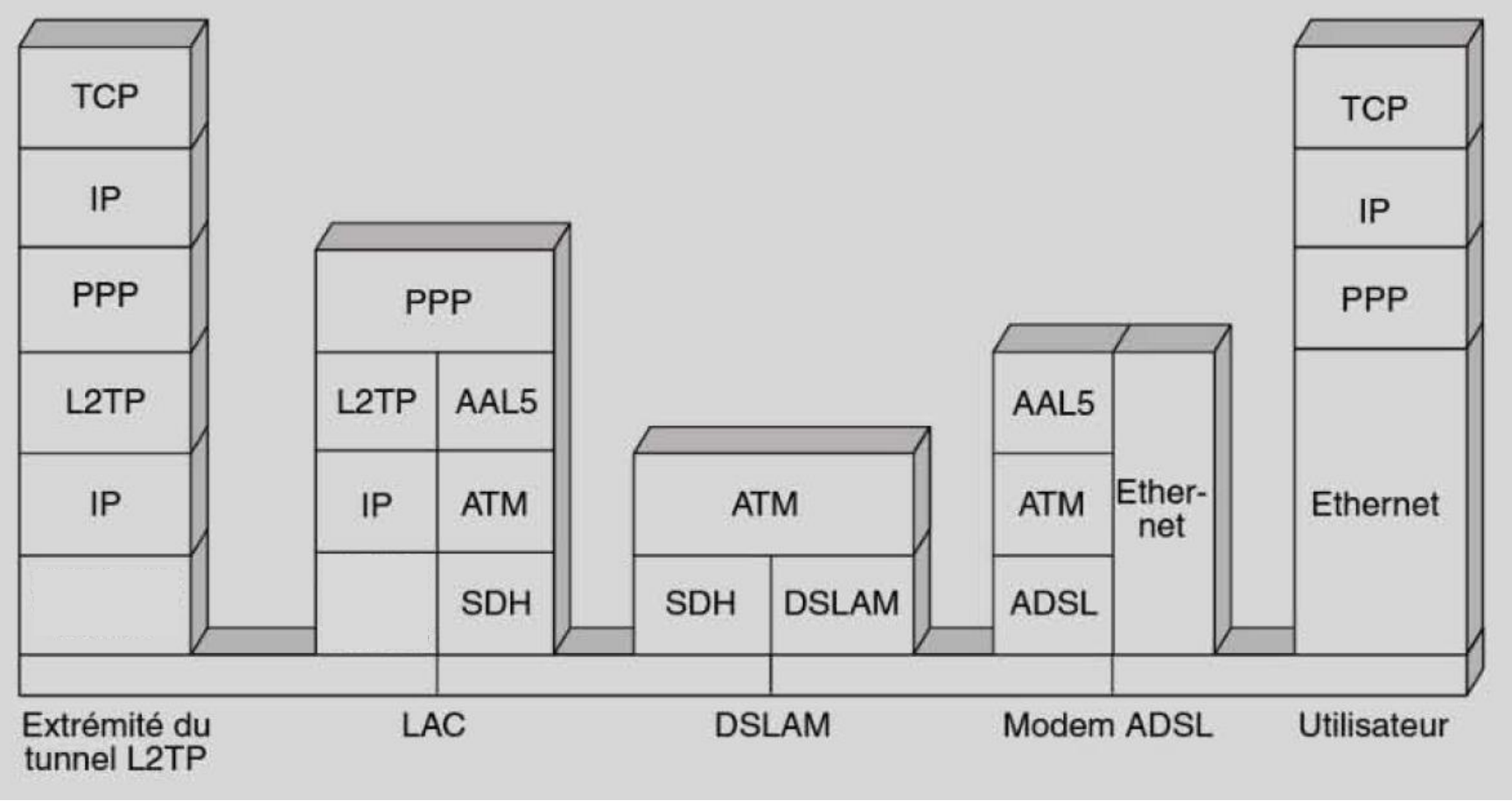
- On peut alors établir un VPN basé sur L2TP/Ipsec

- Confusion possible dans L2TP/IPsec entre les termes "tunnel" and "secure channel."

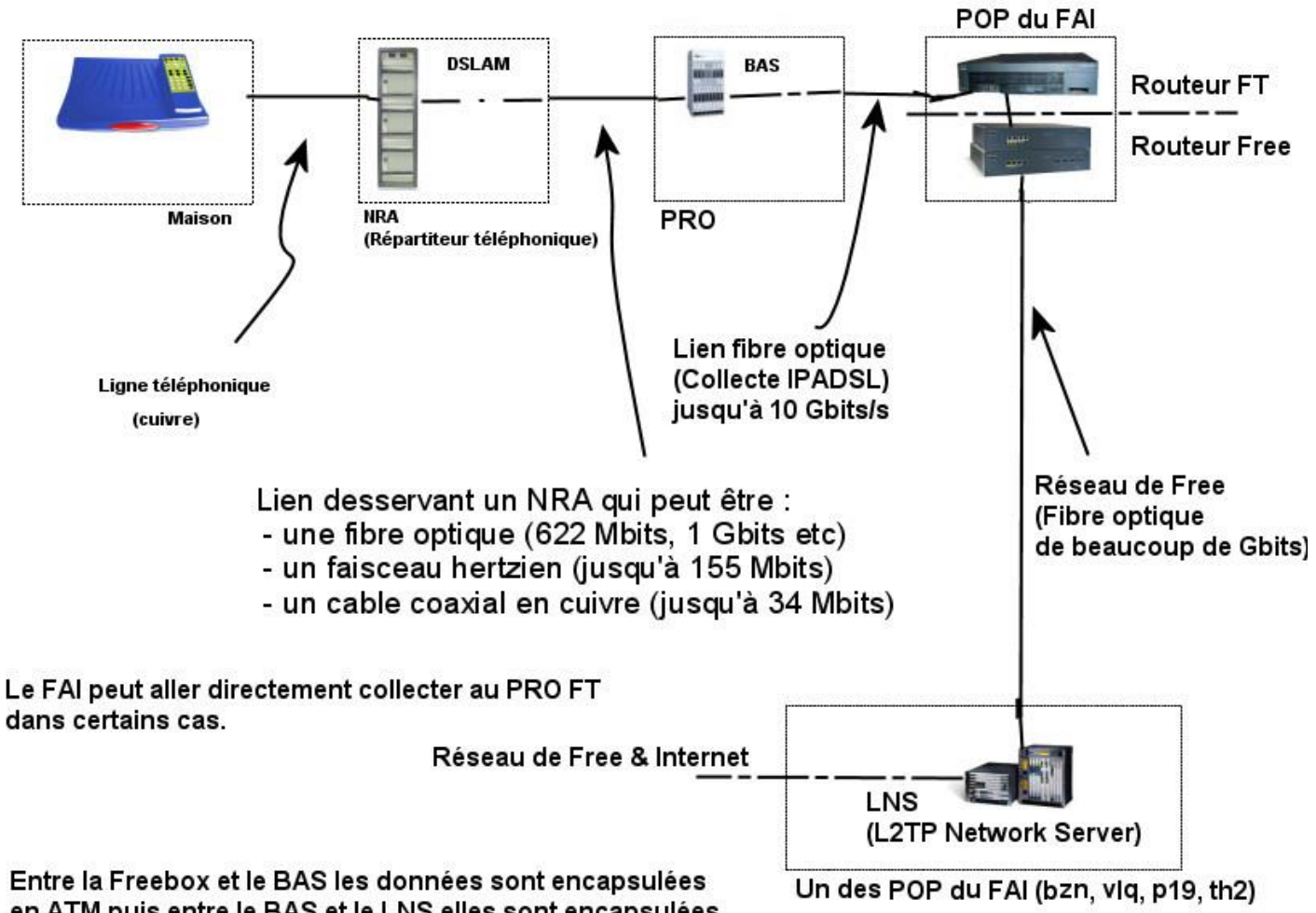
  – *Tunnel*: canal permettant de transporter des paquets d'un réseau sur un autre réseau sans les modifier ni nécessiter de connaitre leur contenu (pas de lecture, pas d'écriture)
    Dans le cas de L2TP/IP, les paquets L2TP sont transportés sur IP

  – *Secure channel*: connexion dans laquelle la confidentialité de toutes les données est garantie.
    Dans le cas de L2TP/IPsec, IPsec fournit un secure channel, puis L2TP fournit un tunnel.

Architecture protocolaire d'une communication ADSL en non-dégroupé



Free ADSL: schéma en zone non-dégroupée



MC Cob

Vocabulaire

- **ATM** : Asynchronous Transfert Mode
- **AAL5** : ATM Abstraction Layer 5
- **EAS** : Equipement d'Acces aux Services
- **L2TP** : Layer two Tunneling Protocol
- **LAC** : L2TP Access Concentrator
- **LNS** : L2TP Network Server
- **PAS** : Plateforme d'Acces aux Services (Proxy RADIUS)
- **PPP** : Point to Point Protocol
- **PPPoA** : PPP over ATM
- **PPPoE** : PPP over Ethernet
- **POTS** : Plain Old Telephone Service
- **RADIUS** : Remote Authentification Dial In User Service

xDSL types
and
history

DSL Flavors

DSL is often called **x**DSL

since there are many varieties (different **x**)

e.g. ADSL, HDSL, SHDSL, VDSL, IDSL, etc.

There were once many unconnected types

but now we divide them into three main families

The differentiation is by means of the application scenario

- HDSL (symmetric, mainly business, data + telephony)
- ADSL (asymmetric, mainly residential, Internet access)
- VDSL (very high rate, but short distance)

More xDSL flavors

modem	speed	reach	main applications
IDSL	160 (144) Kbps	5.5 km	POTS replacement, videoconferencing, Internet access
HDSL	2 Mbps (4-6W)	3.6-4.5 km	T1/E1 replacement PBX interconnect, FR
HDSL2	2 Mbps (2W)	3 km	same as HDSL
SHDSL	2.3 Mbps	3 km	same as HDSL
SHDSLbis	4.6 Mbps	3 km	same as HDSL

More xDSL flavors (cont.)

modem	speed	reach	main applications
ADSL	6 Mbps DS 640 Kbps US	3.5-5.5 km	residential Internet, video-on-demand
ADSL2	8 Mbps DS 800 Kbps US	> ADSL	Internet access, VoIP
ADSL2+	16 Mbps DS 800 Kbps US	< 2 km	"
VDSL	<= 52 Mbps	300m - 1 km	LAN interconnect, HDTV, combined services
VDSL2	200 Mbps (aggregate)	up to 1.8 km	"
cable modem	10-30Mbps DS shared	50 km	residential Internet
HPNA	1, 10 Mbps	home wiring	residential networking

Not DSL

T1 line conditioning

In order for a subscriber's line to carry T1

- Single gauge
- CSA range
- No loading coils
- No bridged taps
- Repeaters every 6 Kft (starting 3 Kft)
- One T1 per binder group
- Labor intensive (expensive) process
- Need something better … (DSL)

HDSL

1991: HDSL

- Replaced T1/E1 service, but
  - full CSA distance w/o line conditioning / repeaters
- AMI line code replaced with IDSL's 2B1Q line code
- Use 2 UTP pairs, but in ECH mode (DFE)
  - For T1 784 kbps on each pair
  - For E1, 1, 2, 3 and 4 pair modes (all ECH)
- Requires DSP for echo cancellation
- Mature DSL technology, now becoming obsolete

HDSL2

With the success of HDSL,

customers requested HDSL service that would :

- require only a single UTP HDSL
- attain at least full CSA reach
- be spectrally compatible w/ HDSL, T1, ADSL, etc.

The result, based on high order PAM, was called

- **HDSL2 (ANSI)**
- **SDSL Symmetric DSL (ETSI)**

and is now called

- **SHDSL Single pair HDSL (ITU)**

SHDSL

Uses Trellis Coded 16-PAM with various shaping options

Does not co-exist with POTS service on UTP

Can use regenerators for extended reach

single-pair operation

- 192 kbps to 2.312 Mbps in steps of 8 kbps
- 2.3 Mbps should be achieved for reaches up to 3.5 km

dual-pair operation (4-wire mode)

- 384 kbps to 4.608 Mbps in steps of 16 kbps
- line rate is the same on both pairs

Latest standard (G.shdsl.bis - G.991.2 2003 version)

- bonding up to 4 pairs
- rates up to 5696 kbps
- optional 32-PAM (instead of 16-PAM)
- dynamic rate repartitioning

ADSL

Asymmetric - high rate DS, lower rate US

Originally designed for video on demand

New modulation type - **D**iscrete **M**ulti**T**one

FDD and ECH modes

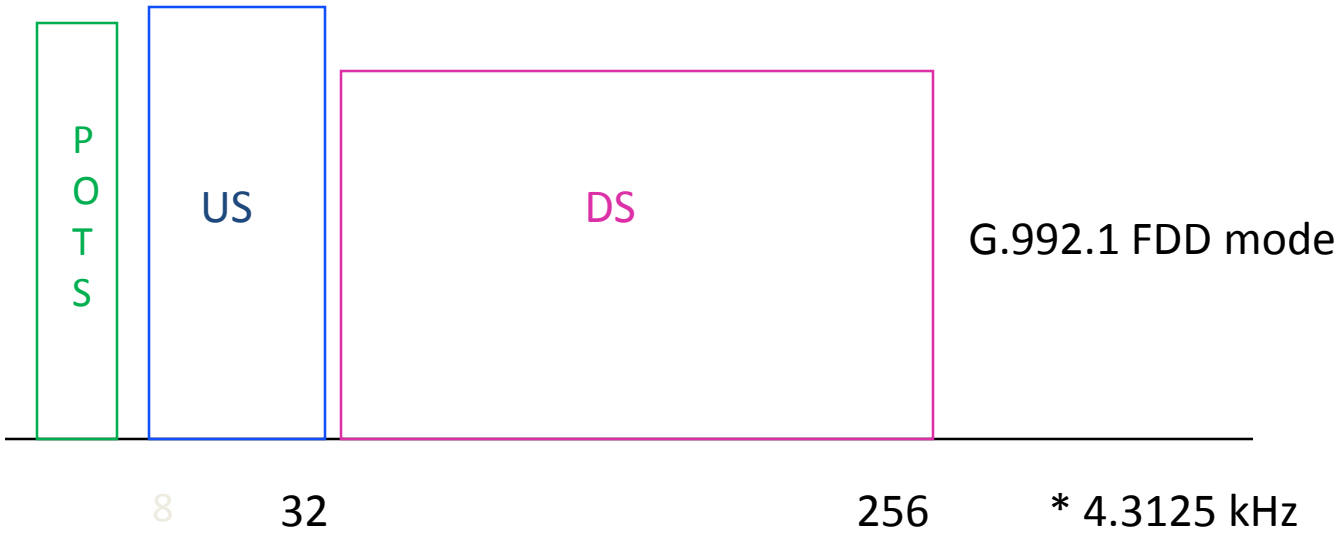
Almost retired due to lack of interest

…but then came the Internet

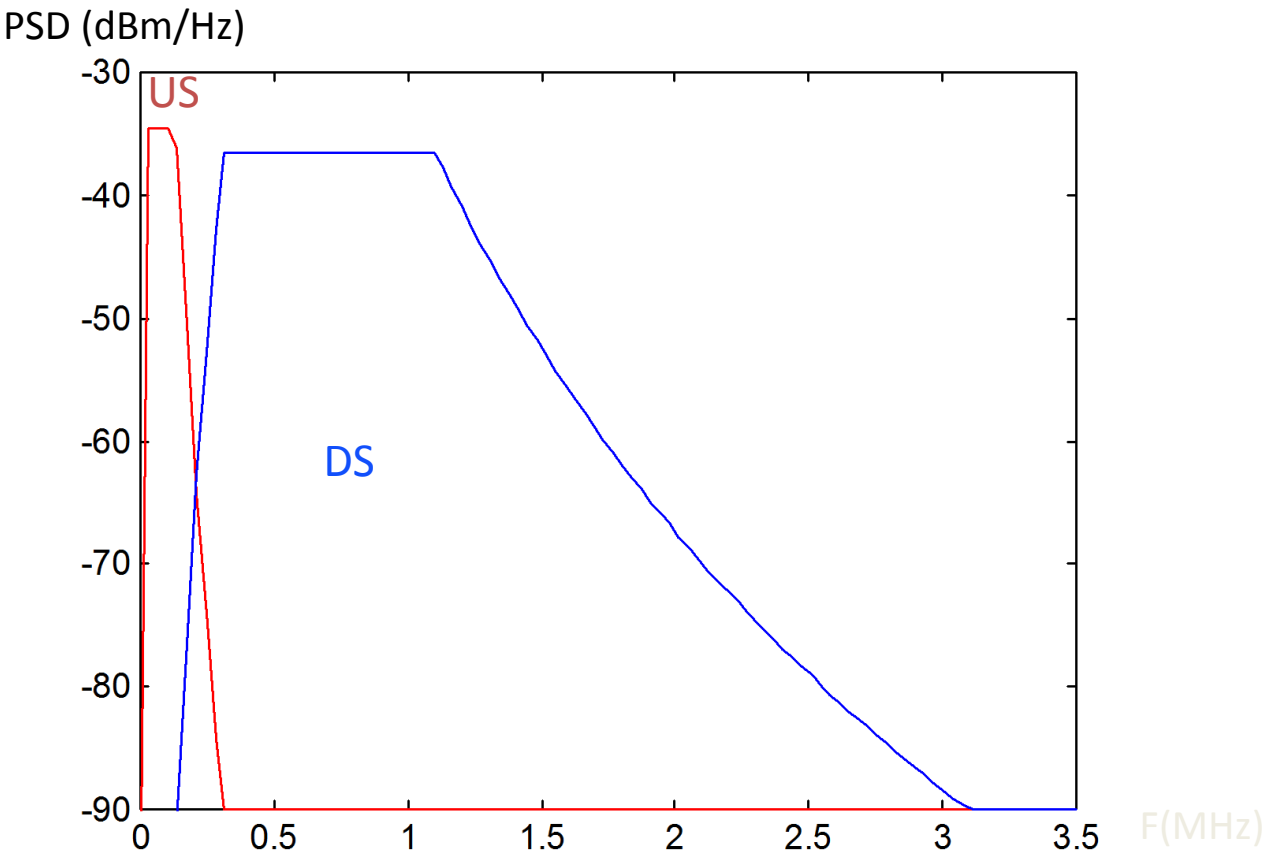
Studies - DS:US for both applications can be about 10:1

ADSL Duplexing

- US uses low DMT tones (e.g. 8 - 32)
- If over POTS / ISDN lowest frequencies reserved
- DS uses higher tones
  - If FDD no overlap
  - If ECH DS overlaps US



Why asymmetry?



ADSL standard

ITU (G.dmt) G.992.1, ANSI T1.413i2 standard

DS - 6.144 Mbps (minimum)

US- 640 kbps

First ADSL data implementations were CAP (QAM)

ITU/ANSI/ETSI standards are DMT with spacing of 4.3125 kHz

- DMT allows approaching water pouring capacity
- DMT is robust
- DMT requires more complex processing
- DMT may require more power

ADSL2

ADSL uses BW from 20 kHz to 1.1 MHz

ADSL2 Increases rate/reach of ADSL by using 20 kHz - 4.4 MHz

Also numerous efficiency improvements

- better modulation
- reduced framing overhead
- more flexible format
- stronger FEC
- reduced power mode
- misc. algorithmic improvements

for given rate, reach improved by 200 m

ADSL2+ dramatically increased rate at short distances

More ADSL2 features

Dynamic training features

- **B**it **S**wapping (dynamic change of DMT bin bit/power allocations)
- **S**eamless **R**ate **A**daptation (dynamic change of overall rate)

Frame bearers

- Multiple (up to 4) frame bearers (data flows)
- Multiple latencies for different frame bearers (FEC/interleave lengths)
- Dynamic rate repartitioning (between different latencies)

VDSL

Optical network expanding (getting closer to subscriber)

Optical Network Unit ONU at curb or basement cabinet

FTTC (curb), FTTB (building)

These scenarios usually dictates low power

Rates can be very high since required reach is minimal!

Proposed standard has multiple rates and reaches

VDSL - rate goals

Symmetric rates

```
 6.5 4.5Kft (1.4 Km)
13   3 Kft  (900 m)
26   1 Kft  (300 m)
```

Asymmetric rates (US/DS)

```
0.8/ 6.5 6 Kft    (1.8 Km)
1.6/13   4.5 Kft (1.4Km)
3.2/26   3 Kft    (900 m)
6.4/52   1 Kft    (300 m)
```

Summary of xDSL Line Codes

PAM

- IDSL (2B1Q)
- HDSL
- SHDSL/HDSL2 (with TCM and optionally OPTIS)
- SDSL

QAM/CAP

- proprietary HDSL/ADSL/VDSL

DMT

- ADSL
- ADSL2, ADSL2+
- G.lite
- VDSL2

DMT processing

bit handling ((de)framer, CRC, (de)scrambler, RS, (de)interleaver)

tone handling (bit load, gain scaling, tone ordering, bit swapping)

QAM modem (symbolizer, slicer)

signal handling (cyclic prefix insertion/deletion, (I)FFT, interpolation, PAR reduction)

synchronization (clock recovery)

channel handling (probing and training, echo cancelling, FEQ, TEQ)