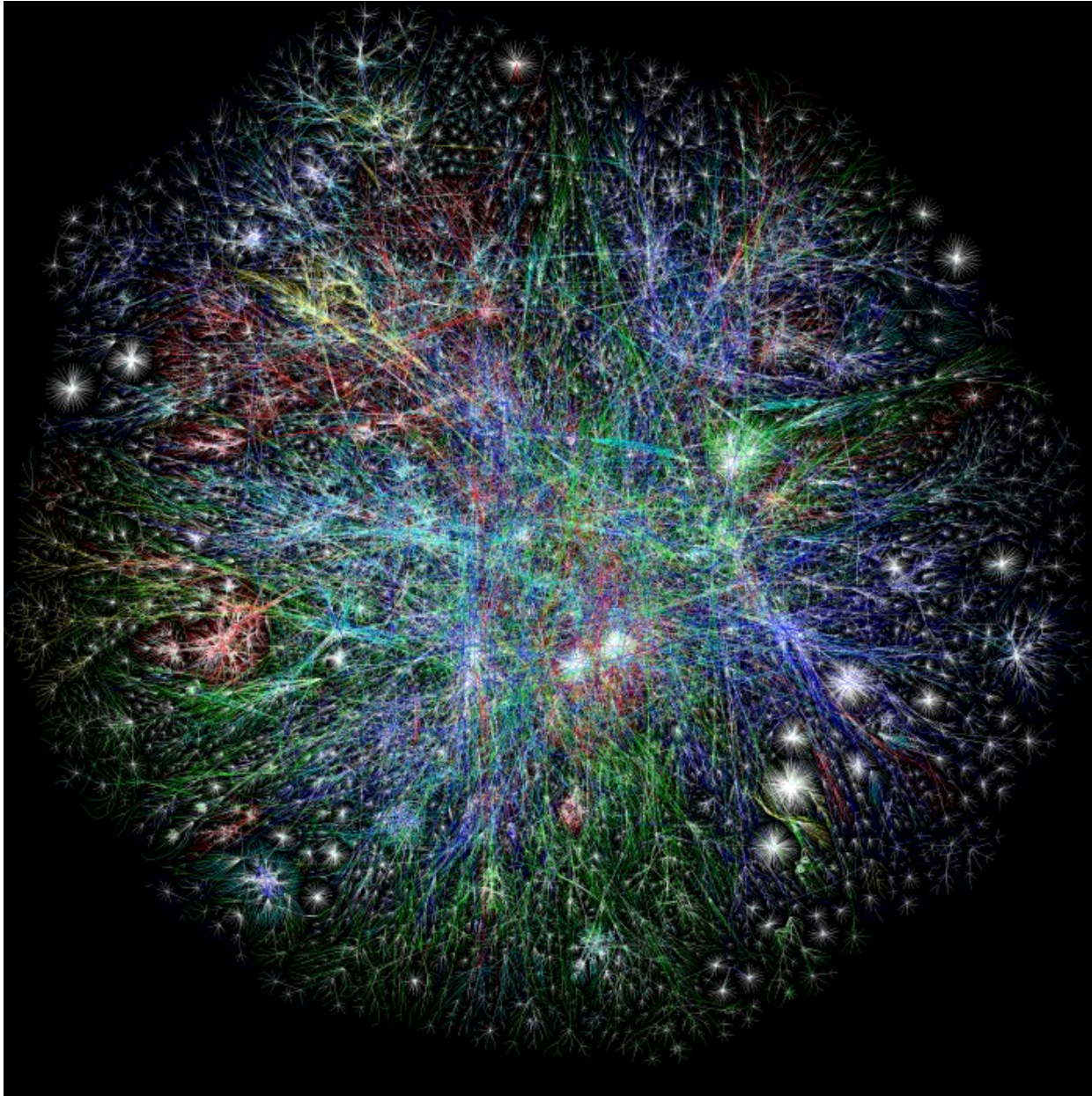


Le routage dans Internet



Jean-Patrick Gelas
Université de Lyon

Crédit : OPTE project

Sources

- Analyse structurée des réseaux, *J.Kurose, K.Ross*, Pearson.
- Réseaux (3ème éd.), *A.Tanenbaum*, Dunod.
- Le routage, *Robert Jean-Sébastien, Somerlinck Pierre*,
<http://www.tele.ucl.ac.be/EDU/ELEC2920/2000/Routing/routage.htm>

LO (*crash*)

Leonard Kleinrock inventeur du principe de la **commutation de paquets** devant le premier IMP (Interface Message Processor), 1969.



A quoi ressemble un routeur ?

Un routeur est une machine possédant **plusieurs interfaces**. Elle est connectée à plusieurs réseaux en même temps et peut faire passer un paquet d'un réseau à un autre.

*Alcatel-Lucent, Cisco, Foundry,
Ericsson, HP, Huawei, Juniper,...*



Table de routage

Exemple

Destination	Gateway	Genmask	Flags	MSS	Window	irrtt	Iface
10.135.249.1	0.0.0.0	255.255.255.255	UH	0	0	0	tun0
10.135.249.3	0.0.0.0	255.255.255.255	UH	0	0	0	tun1
129.88.70.248	0.0.0.0	255.255.255.248	U	0	0	0	eth0
129.88.70.0	0.0.0.0	255.255.255.192	U	0	0	0	br0
10.135.250.0	10.135.249.1	255.255.255.0	UG	0	0	0	tun0
192.168.22.0	129.88.70.62	255.255.255.0	UG	0	0	0	br0
10.135.248.0	10.135.249.3	255.255.255.0	UG	0	0	0	tun1
192.168.133.0	129.88.70.62	255.255.255.0	UG	0	0	0	br0
192.168.11.0	129.88.70.58	255.255.255.0	UG	0	0	0	br0
129.88.99.0	129.88.70.62	255.255.255.0	UG	0	0	0	br0
192.168.159.0	129.88.70.62	255.255.255.0	UG	0	0	0	br0
129.88.98.0	129.88.70.62	255.255.255.0	UG	0	0	0	br0
131.254.202.0	129.88.70.62	255.255.254.0	UG	0	0	0	br0
172.28.52.0	129.88.70.62	255.255.252.0	UG	0	0	0	br0
138.96.20.0	129.88.70.62	255.255.252.0	UG	0	0	0	br0
10.69.0.0	129.88.70.62	255.255.128.0	UG	0	0	0	br0
172.24.0.0	129.88.70.62	255.255.0.0	UG	0	0	0	br0
172.16.0.0	129.88.70.62	255.240.0.0	UG	0	0	0	br0
10.0.0.0	129.88.70.62	255.0.0.0	UG	0	0	0	br0
0.0.0.0	129.88.70.254	0.0.0.0	UG	0	0	0	eth0

Table de routage

Comment la tables de routage d'un routeur est-elle remplie ?

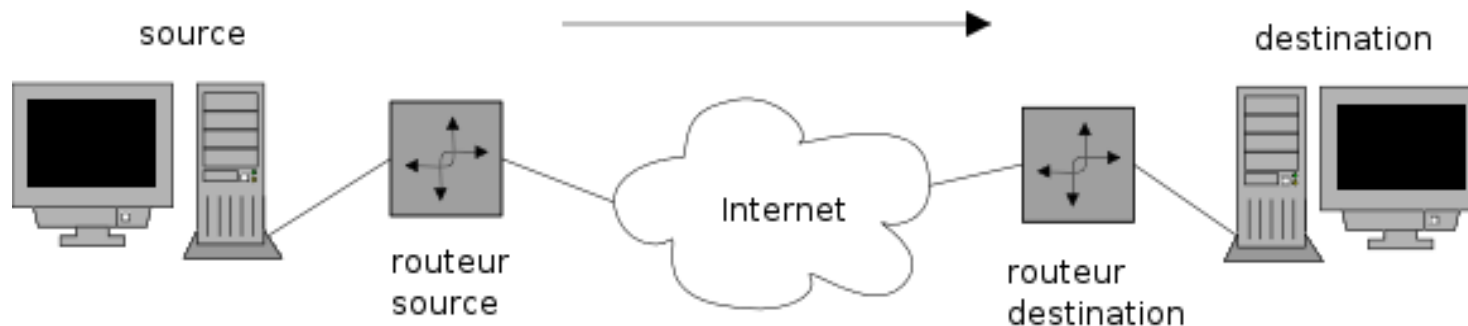
- Manuellement
 - Quid des changements de topologie (panne de liens, liens engorgés ?)
- Automatiquement
 - Nécessité d'un **Protocole de routage** pour échanger les informations avec les routeurs voisins.

Principe du routage

- Pour transférer des paquets, la couche **réseau** doit déterminer le parcours (ou la route) à emprunter, que ce soit
 - un service à datagramme (parcours différents) ou
 - un service à circuit virtuel (même parcours).
- Cette fonction incombe au **protocole de routage** de la couche **réseau**.

Principe du routage

- Un ordinateur est généralement connecté à un routeur spécifique = le routeur par défaut (ou routeur de premier bond ou passerelle (gw)).
- Acheminer un paquet d'une *source* à une *destination* revient à l'acheminer entre un routeur source et un routeur destination.



Introduction au routage

- Le routage est accompli par un **protocole de routage** qui établit des tables de routage pour chaque routeur.
- Une table de routage contient principalement trois colonnes :
 - # 1 : les adresses des destinations (préfixes),
 - # 2 : les masques associés,
 - # 3 : les adresses des nœuds (ou l'identifiant de l'interface de sortie) vers lesquels il faut envoyer les messages pour accéder aux destinations .
- Un protocole de routage repose sur un **algorithme de routage**.
 - Sa mission : trouver le « bon » parcours entre le routeur source et le routeur de destination.
 - Le bon parcours est le parcours le « moins onéreux ».

Introduction au routage (suite)

- La décision du routeur est **locale** -> pourtant elle dépend de la topologie globale du réseau !
- Chaque protocole de routage doit donc communiquer une information de topologie à chaque routeur pour prendre des décisions optimales.
- Cette information est difficile à collecter.
 - volumineuse
 - variable dans le temps

Spécification des protocoles de routages

Un protocole de routage doit répondre à plusieurs critères :

- Minimiser la table de routage
- Minimiser les messages de contrôle
- Être robuste
- Permettre l'utilisation de chemins optimaux

-> besoin de compromis...

Spécification des protocoles de routages : Minimiser la table de routage

Plus la table de routage est grande :

- Plus les échanges de messages entre routeurs sont volumineux (ex: RIP).
- Plus l'opération de *lookup* est longue.

Spécification des protocoles de routages : Minimiser les messages de contrôle

Les messages échangés entre routeurs constituent une surcharge sur le réseau et doivent donc être minimisés.

Spécification des protocoles de routages :

Robustesse : Trous noirs, boucles et oscillations

- **Trous noirs** : la pire chose que peut faire un routeur est d'envoyer des paquets dans une mauvaise direction, de sorte qu'ils n'atteignent jamais leur destination.
- **Boucles** : Si les tables de routages sont inconsistantes, des boucles peuvent se former.
- **Oscillations** : Des oscillations peuvent apparaître si les protocoles de routage choisissent un chemin en tenant compte d'un paramètre qui évolue au cours du temps (ex: la quantité de trafic dans le réseau).

Spécification des protocoles de routages : Robustesse : Trous noirs, boucles et oscillation

- Trous noirs, boucles et oscillations sont rares dans des conditions normales sauf si :
 - des tables de routages sont corrompues,
 - si l'admin spécifie des informations incorrectes,
 - si des liaisons disparaissent ou réapparaissent régulièrement,
 - si des paquets de contrôle sont corrompus, ...
- Le protocole de routage doit se protéger par des tests périodiques, l'usage de *checksums*, la numérotation de séquences, ...

Spécification des protocoles de routages : Utilisation de chemins optimaux

- Tout paquet devrait suivre le chemin optimal vers sa destination.
- Le chemin optimal :
 - le plus court ? (!\ Pas nécessairement !)
 - le plus petit délai, le plus grand débit,
 - les liaisons les plus sécurisées
 - le coût financier le plus faible, ...
- La détection de chemin optimaux nécessite une collaboration entre tous les routeurs.

Différents mécanismes de routage

Différents mécanismes de routage

Routage centralisé

- Dans le routage centralisé, un processeur central :
 - collecte l'information sur l'état de chaque liaison,
 - établit une table de routage pour chaque noeud et l'envoie à ceux-ci.
- Le routage centralisé n'est pas envisageable sur des réseaux de grande taille :
 - Et si les liaisons au voisinage du processeur central tombe en panne ou que celui-ci tombe en panne ?
 - Quels seraient les temps de calcul ?
 - Les routeurs situés près du processeur central reçoivent les tables de routage bien avant ceux qui sont plus éloignés (inconsistance dans le réseau pendant un certain temps).
 - Les liaisons aboutissant au processeurs central peuvent être surchargées si trop de routeurs envoient des informations.

Différents mécanismes de routage

Routage distribué

- Dans le routage distribué, les routeurs s'envoient périodiquement des informations pour créer des tables de routage dynamique.
- **Le routage sur Internet est résolument distribué.**

Différents mécanismes de routage

Routage « à partir de la source »

- L'en-tête d'un paquet contient les adresses de tous les noeuds par lesquels il va devoir passer pour arriver à destination.
- La source doit connaître la **topologie** de tous le réseau.
- Inconvénients :
 - Si une liaison ou un routeur sur le chemin disparaît, le paquet n'atteindra pas la destination.
 - Si le chemin est long, l'en-tête du paquet peut-être très grande.

Différents mécanismes de routage

Routage « noeud après noeud »

- L'en-tête d'un paquet contient juste l'adresse de destination.
- C'est aux routeurs de déterminer le noeud suivant.
- Sur Internet on utilise le routage « noeud après noeud ».

Différents mécanismes de routage

Routage « à chemin unique » ou routage « multi-chemins »

- Un routeur à **chemin unique** maintient un seul chemin pour chaque destination.
- Un routeur **multi-chemin** maintient un chemin primaire pour chaque destination, et des chemins alternatifs au cas où le chemin primaire serait indisponible (Remarque: dans le routage stochastique, un routeur peut envoyer des paquets sur un chemin alternatif alors que le chemin primaire est disponible).
- Sur Internet on utilise le routage à **chemin unique**.

Différents mécanismes de routage

Routage adaptatif ou routage statique

- Routage adaptatif :
 - Le choix du chemin peut dépendre de l'état du réseau (trafic, files d'attente, ...).
 - Meilleur choix des chemins, mais peut mener à des oscillations.
 - Requiert une charge supplémentaire sur le réseau pour l'établissement de l'état des liaisons.
- Routage statique : on ignore l'état du réseau.
- Internet utilise les deux types de routage.

Différents mécanismes de routage

Routage aléatoire

- Acheminement par inondation (Flooding)
 - Grande robustesse, simple à implanter.
 - Atteint toujours le destinataire par le chemin le plus court (délai minimum) même si la topologie change.
 - Pas de table de routage.
 - Mauvaise utilisation des ressources.
 - Risque de congestions élevé.
- Le *Selective Flooding* est une alternative.

Algorithmes de routage fondamentaux

- L'algorithme du **vecteur de distance** (*DV ou Distance-Vector routing*) : un noeud transmet à ses voisins le coût pour atteindre chaque noeud du réseau.
- L'algorithme à **état des liens** (*state-link routing*) : un noeud transmet à chaque noeud du réseau le coût pour atteindre ses voisins.

Ces deux algorithmes supposent que chaque routeur connaît l'adresse de ses voisins et le coût pour atteindre ceux-ci.

Ce sont des **algorithmes distribués** qui conviennent particulièrement bien à Internet.

Algorithmes de routage fondamentaux

Vecteur de distance

- Chaque routeur maintient un vecteur de distance (DV), liste de couples (*destination, coût*), qu'il recalcule à chaque fois qu'il reçoit une copie du vecteur de distance (DV) d'un de ses voisins.
- L'information se propage noeud après noeud à chaque échange de vecteurs, jusqu'à parcourir tout le réseau.

Routage par information d'état des liens

Link State Routing

- Le problème principal du routage par *vecteur de distance* est qu'il converge trop lentement.
- Remplacé par un algorithme de routage entièrement nouveau.
- Diverses variantes de cet algorithme sont très en vogue aujourd'hui.
- L'idée sous-jacente du routage par informations d'état des liens peut s'énoncer en 5 points. Tout routeur doit :
 1. Découvrir ses voisins et apprendre leur adresse réseau respective.
 2. Mesurer le temps d'acheminement vers chacun de ses voisins.
 3. Construire un paquet spécial exprimant tout ce qu'il vient d'apprendre.
 4. Envoyer ce paquet à tous les autres routeurs.
 5. Calculer le plus court chemin vers tous les autres routeurs.

Routage par information d'état des liens

1) Découvrir ses voisins

- La première tâche d'un routeur en cours d'initialisation est de savoir qui sont ses voisins.
- Il envoie sur chacune de ses lignes un paquet spécial HELLO et attend la réponse.
- Les routeurs aux extrémités de ces lignes répondent en donnant des informations de routage (nom unique, adresse IP,...)

Routage par information d'état des liens

2) Mesurer le temps d'acheminement

- Cet algorithme exige que chaque routeur connaisse le temps d'acheminement avec chacun de ses voisins.
- Utilisation d'un paquet spécial : ECHO.
- Prendre en compte ou non la charge de la liaison (*i.e* déclencher la mesure de temps dès la mise en file d'attente ou dès que le paquet atteint la tête de la file et qu'il est transmis).
 - Pour : utile lorsqu'un routeur dispose de deux lignes identiques pour atteindre une destination donnée.
 - Contre : Engendre des oscillations incessante rythmant les mises à jour des tables.

Routage par information d'état des liens

3) Construire un paquet d'informations d'état de lien

- Une fois les informations nécessaires au routage obtenues, chaque routeur construit un paquet spécial contenant les données qu'il a collecté.
- Ce paquet contient : *id.* du routeur src. ; seq# ; âge du paquet et la liste des routeurs voisins (+ le temps d'acheminement associé).
- Ces paquets peuvent être construit :
 - Périodiquement
 - lorsqu'un événement significatif apparaît (ex: coupure de ligne)

Routage par information d'état des liens

4) Envoyer les paquets d'informations d'état de lien

- La faiblesse de cet algorithme est la diffusion fiable des paquets. Il n'y a pas de concertation entre les routeurs ce qui peut conduire à :
 - différente topologie du réseau ;
 - donner lieu à des boucles ;
 - rendre certaine machines inaccessibles.
- Distribution des paquets par inondation (contrôlé par l'id. et le numéro de séquence).
- Trois problèmes avec le *seq#* ...

Routage par information d'état des liens

5) Calculer les nouvelles routes

- Dès que le routeur a accumulé un jeu de paquets d'informations d'état, il peut construire le graphe complet du sous-réseau.
- Une ligne peut avoir deux représentations (une dans chaque direction). Le routeur peut prendre la valeur moyenne des deux valeurs ou les utiliser séparément.
- L'algorithme de *Dijkstra* peut-être utilisé localement sur chaque routeur. Les résultats sont inscrits dans les tables de routage et le fonctionnement du sous-réseau peut reprendre.

Algorithmes de routage fondamentaux

État des liens : Bilan

- La philosophie du routage par état des liens est de distribuer la topologie du réseau et le coût de chaque liaison à travers tout le réseau.
- Ensuite chaque routeur calcule indépendamment les chemins optimaux vers chaque destination.
- Les routeurs communiquent entre eux par des LSP (link-state packets) qui décrivent leurs liens.
 - Un LSP contient l'adresse du routeur, l'adresse de ses voisins et le coût de chaque liaison avec ses voisins.
 - Un LSP reçu est stocké dans une « LSP database », et renvoyé sur chaque interface sauf celle par laquelle il est arrivé

Algorithme de routage fondamentaux

État des liens : Algorithme de Dijkstra

- L'algorithme de Dijkstra calcule le chemin le plus court entre deux noeuds d'un réseau.
- L'idée de base est de maintenir pour chaque ensemble de noeuds P le chemin le plus court ayant été trouvé.
- Chaque noeud extérieur à P doit être atteint à partir d'un noeud déjà dans P .

Algorithme de routage fondamentaux

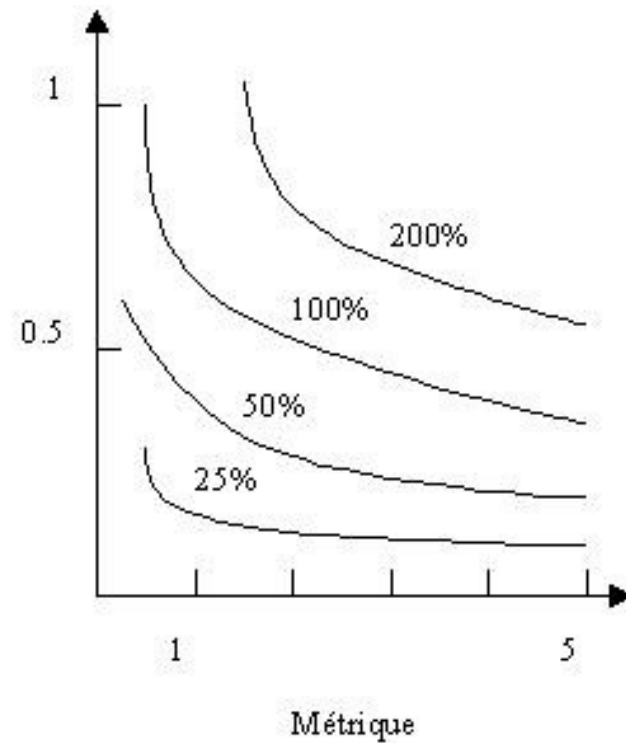
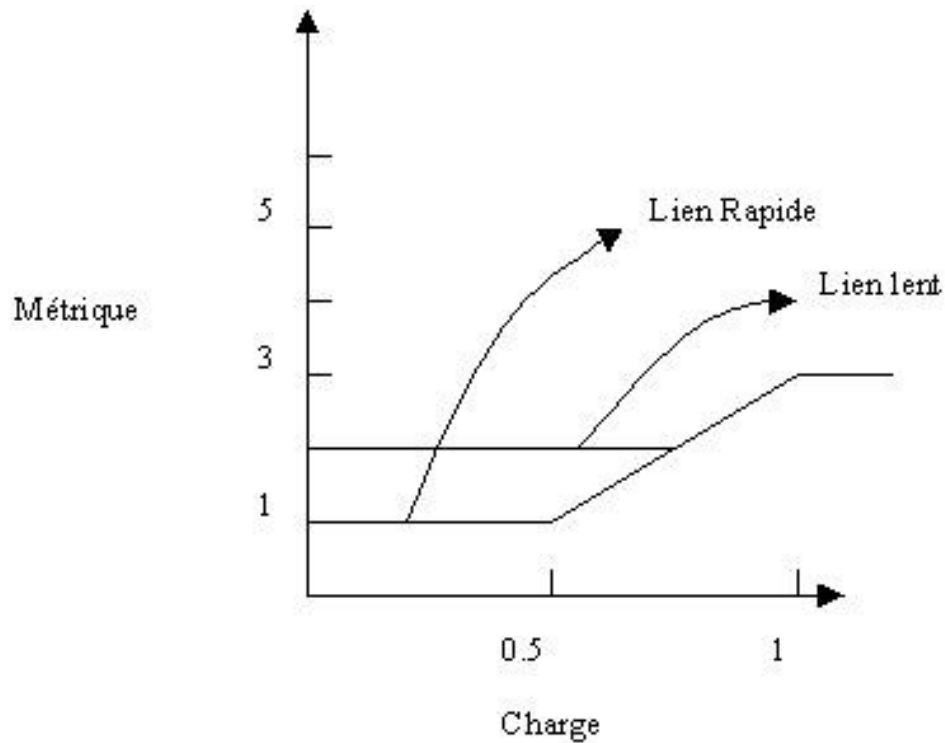
Vecteur de distance vs. État des liens

- Les algorithmes à état des liens sont :
 - **Plus stables** car chaque routeur connaît la topologie de l'entièreté du réseau.
 - Ils supportent des métriques multiples (*i.e.* plusieurs fonctions de coût des liaisons).
 - Ils **convergent généralement plus rapidement** que les algorithmes à vecteur de distance.
 - Les *LSP database* requièrent généralement moins d'espace mémoire et semblent mieux adaptés aux réseaux gérés par plusieurs entités administratives différentes.
- Les algorithmes à vecteurs de distance :
 - Moins stable. On y ajoute des optimisations (*poison reverse, path vector, trigger updates, ...*).

Choix du coût des liaisons

- Le coût d'une liaison peut dépendre de plusieurs paramètres :
 - bande passante ; coût financier ; valeur fixé par l'administrateur ; délai ; nombre de sauts ; MTU ; fiabilité ; ...
- Dans le cas du routage dynamique, le coût d'une liaison peut-être fonction de son encombrement (charge).
- On parle de **métrique dynamique** par opposition à une **métrique statique**.

Choix du coût des liaisons



- Coût d'un lien en fonction de sa charge.
- Charge d'un lien en fonction de son coût (réponse du réseau).

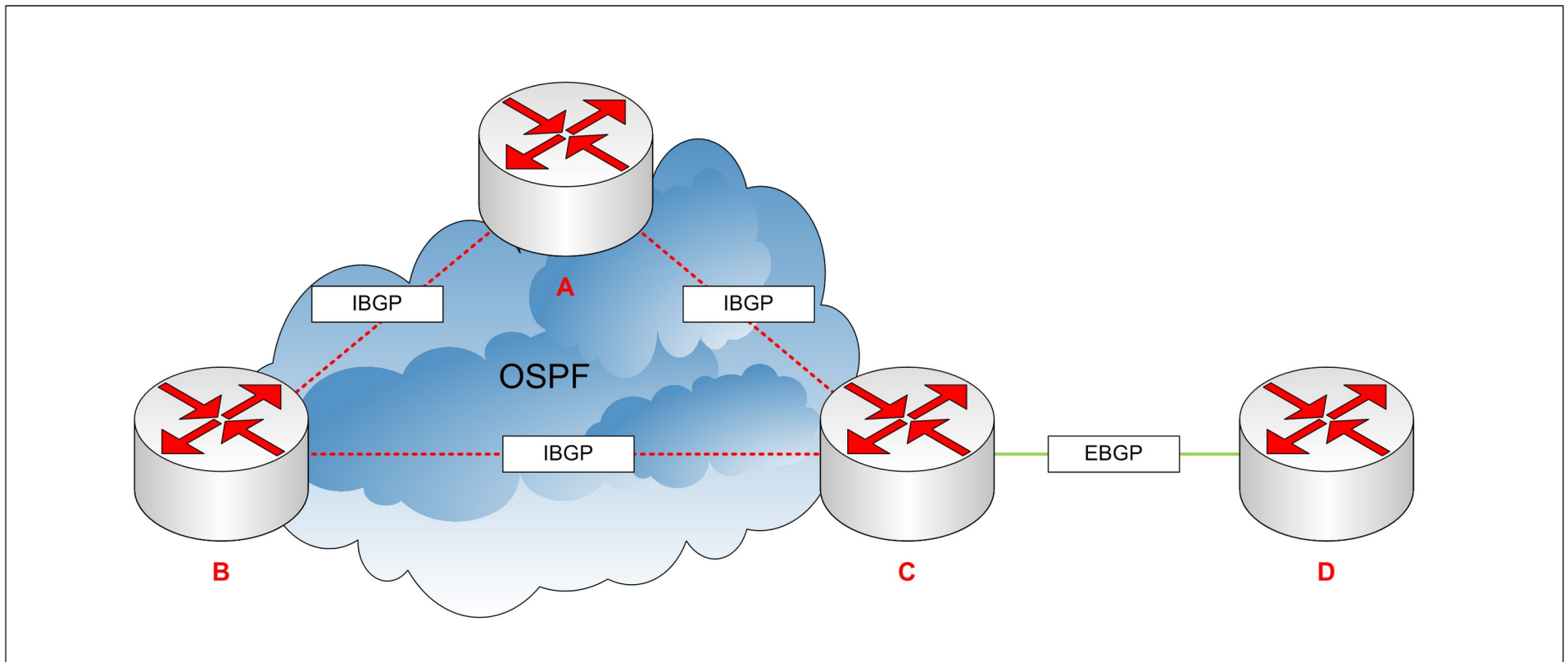
Dans le cas d'une **métrique dynamique**, un coût élevé va diminuer le trafic dans une liaison, ce qui va diminuer le coût, ce qui va augmenter à nouveau le trafic, etc... (besoin de trouver une fonction de coût appropriée afin que trafic et coût convergent au point désiré, tout en évitant les oscillations).

Routage hiérarchique

- Dans un réseau de la taille d'Internet, le routage hiérarchique est une nécessité (ex: taille des tables).
On distingue deux niveaux d'adressage : réseau et machine.
- Au plus haut niveau se trouve la colonne vertébrale (*backbone*) d'Internet qui interconnecte les AS (*Autonomous System*) ou **default-free zone** (constitué de routeurs et AS sans route par défaut).
- Le routage entre AS utilise des protocoles dits **extérieurs** (EGP *Exterior Gateway Protocol*)
 - Des routeurs interconnectant deux AS ne se font pas nécessairement confiance donc les protocoles extérieurs doivent être sécurisés.
- Au sein d'un AS on utilise des protocoles dits **intérieurs** (IGP *Interior Gateway Protocol*)
 - Ils sont libres de problèmes administratifs et divisent le système en zone (*areas*).
- Au niveau le plus bas, on retrouve les LANs (utilisant Ethernet).

Routage hiérarchique

- Il faut ensuite interconnecter les protocoles intérieurs (IGP) et extérieurs (EGP) qui n'utilisent pas les mêmes techniques de routage.



Implémentations des protocoles de routage usuels

Protocoles de routage usuels

Protocoles intérieurs

- RIP (*Routing Information Protocol*) : protocole à **vecteur de distance** utilisant une métrique statique.
 - les routeurs s'échangent leurs vecteurs de distance toutes les 30 secondes (déclaré HS au bout de 180 s)
 - Utilise le *split horizon* pour éviter le problème du comptage jusqu'à l'infini.
 - Utilisé pour les petits réseaux où sa simplicité compense ses lacunes (en cas de rupture d'une liaison).

Protocoles de routage usuels

Protocoles intérieurs

- OSPF (*Open Shortest Path Protocol*) : protocole à état des liens. Il utilise le routage hiérarchique via la notion de zone (area) pour diriger un paquet dans un AS.
- Peut être plus complexe à mettre en oeuvre que RIP.

OSPF : Protocole de routage intra-système

- Protocole de routeur interne : IGP (*Interior Gateway Protocol*)
- Devenu un standard en 1990 (RFC 1247)
- Protocole supporté par de nombreux routeurs.
- Devient le protocole de routage interne principal.

OSPF : Caractéristiques

Lors de sa conception les exigences suivantes ont été incluses :

1. Utilisation d'un algorithme ouvert (***Open Shortest ...***)
2. Support de divers métriques (distance géographique, délai, ...)
3. Algorithme dynamique capable de s'adapter automatiquement et rapidement aux changements de topologies.
4. Supporter le routage par type de service (ex: trafic temps réel) en exploitant le champ ToS de IP.
5. Faire de l'équilibrage de charge (*i.e* répartition sur plusieurs liens).
6. Support de systèmes hiérarchiques.
7. Offre un minimum de sécurité.

OSPF : Connexions et réseaux

OSPF supporte 3 types de connexions et de réseaux :

- les liaisons point-à-point entre deux routeurs,
- les réseaux multi-accès à diffusion (la plupart des réseaux LAN),
- les réseaux multi-accès sans diffusion (la plupart des réseaux WAN).

Un réseau *multi-accès* est un réseau qui contient plusieurs routeurs, chacun d'eux pouvant directement communiquer avec tous les autres.

OSPF : Division d'un AS en zone

- Un système autonome (AS) est souvent vaste et complexe.
- Le protocole OSPF permet de les diviser en zones numérotées
- Une zone est un ensemble qui comprend un ou plusieurs réseaux contigus et des routeurs.
- Les zones ne se chevauchent pas (extension du concept de sous-réseau).
- Tous les AS ont une zone épine dorsale (*backbone area*) appelée « Zone 0 » (noté 0.0.0.0).
- Toutes les zones d'un AS sont connectées à la zone épine dorsale (éventuellement par des tunnels) => de n'importe quel zone on peut accéder à une autre zone via l'épine dorsale.

OSPF : Zones

- A l'intérieur d'une zone, chaque routeur possède une base de données topologiques (état des liens) et exécute le même algorithme du plus court chemin (entre tous les routeurs de la zone + le routeur connecté à la zone épine dorsale).
- Un routeur connecté à deux zones a besoin des bases de données de ces deux zones et doit exécuter l'algorithme du plus court chemin pour chacune des zones séparément.

OSPF : Type de service

- Le protocole OSPF peut gérer le routage par type de service au moyen de trois graphes (métriques différentes) :
 - le délai d'acheminement ;
 - le débit ;
 - la fiabilité.

Bien que cela triple les calculs nécessaires, OSPF dispose de différentes routes possibles pour optimiser le *délai*, le *débit* et la *fiabilité*.

OSPF : Types de chemin

- En fonctionnement normal, trois types de chemins peuvent-être demandés :

- **intra-zone** : le plus simple, puisque le routeur source connaît déjà le chemin le plus court vers le routeur destinataire.

- **inter-zone** : Le routage inter-zone s'effectue en trois étapes.

- 1) Aller de la source vers l'épine dorsale,

- 2) transiter au travers de l'épine dorsale jusqu'à la zone de destination,

- 3) aller jusqu'au destinataire.

(Configuration en étoile, la zone épine dorsale étant le foyer, les zones étant les rayons)

Les datagrammes sont routés de la source vers la destination tels quels (ni encapsulés, ni transmis par un tunnel sauf s'ils vont vers une zone dont la seule connexion avec la zone épine dorsale est un tunnel).

- **inter-système autonome**

OSPF : Classes de routeur

- Le protocole distingue quatre classes de routeurs :
 - les routeurs intra-zones (internes) entièrement à l'intérieur d'une zone,
 - les routeurs inter-zones (*border routers*) connectés à deux zones (ou plus),
 - les routeurs fédérateurs (*backbone routers*) connectés à l'épine dorsale,
 - les routeurs inter-systèmes autonomes (*boundary routers*) connectés aux routeurs d'autres systèmes autonomes.
- Ces classes de routeurs peuvent se recouvrir (ex: tous les routeurs inter-zones font partis du *backbone*).

OSPF : Démarrage

- Quand un routeur démarre, il envoie un message *HELLO* sur toutes ses lignes de sorties.
 - sur LAN les autres routeurs sont accessibles en diffusion multi-destinataire,
 - sur WAN le routeur a besoin d'informations complémentaires (de configuration) pour savoir comment les contacter.
- Les réponses qu'il reçoit lui permettent de savoir qui sont ses voisins.

OSPF : Fonctionnement

- Fonctionnement par échange de messages d'information entre **routeurs adjacents**.
- En fonctionnement normal, chaque routeur transmet périodiquement un message de *Mise à jour d'état de lien* à chacun de ses routeurs adjacents.
 - Ce message indique l'état des liens et leurs *poids* à la base de données topologique.
 - Demande un accusé de réception pour améliorer la fiabilité.
 - Chaque message a un numéro de séquence.
- Le même type de message est émis lorsqu'une liaison se crée, se coupe ou change de *poids*.

OSPF : Messages

Les cinq types de messages OSPF sont envoyés sous forme de datagrammes IP :

- **Hello** : Permet de découvrir qui sont les routeurs voisins.
- **Mise à jour d'état de lien** : Information d'état fournie à la base de données topologique.
- **Accusé de réception de mise à jour** : Acquiescement d'une mise à jour d'état de lien.
- **Demande d'état de lien** : Demande d'information à la base de données topologique sur un partenaire.
- **Description de lien** : La base de données topologique donne les informations d'état de lien à qui en a besoin.

Routeur désigné et routeur de secours

- DR (*Designated Router*) et BDR (*Backup Designated Router*).
- Le DR sert de référent pour la base de données topologique. Objectifs :
 - réduire le trafic lié à l'échange d'informations sur l'état des liens ;
 - améliorer l'intégrité de la base de données topologique ;
 - accélérer la convergence.

Élection du DR

- Le routeur élu est celui qui a la plus grande priorité (Router ID ou RID).
- La priorité est un nombre fixé à 1 par défaut sur tous les routeurs.
- Pour départager les routeurs qui ont la même priorité, celui qui est élu a la plus grande adresse IP.
- Le BDR sera celui avec la seconde plus grande adresse IP.
- Une fois élu le DR n'est jamais remis en cause même si un routeur avec une priorité plus grande apparaît.

OSPF : Bilan

- Par inondation, chaque routeur informe tous les autres routeurs de sa zone à propos de ses voisins et de ses *poids*.
- Ces informations permettent à chaque routeur de construire le graphe pour sa (ses) zone(s) et de calculer le plus court chemin.
- De même pour la zone épine dorsale.
- Les routeurs fédérateurs acceptent des informations venant des routeurs inter-zones afin de calculer le meilleur chemin, de chaque routeur fédérateur vers chaque autre routeur.
- Cette information est ensuite propagé à tous les routeurs inter-zones, qui les communiquent à l'intérieur de leur zones.
- Un routeur est alors capable d'envoyer un datagramme inter-zone par la meilleur route de sortie vers la zone épine dorsale.

Intermediate system to intermediate system (IS-IS)

- Autre protocole à état des liens
- Similaire à OSPF, mais moins connu
- Moins verbeux qu'OSPF sur le réseau
=> souvent préféré par les ISP, car permet de gérer un plus grand nombre de routeurs avec les mêmes ressources

Autonomous System (AS)

- Un AS est une collection de réseaux et routeurs IP sous le contrôle d'une ou plusieurs entités, qui présente une politique de routage commune vers l'internet.
- Un ISP doit avoir un numéro d'AS (ASN) publique enregistré.
- Un ASN identifie un réseau unique (utilisé par BGP).
- Début 2013 on répertoriait plus de 45 000 AS différents.

Autonomous System

- L'ASN est assigné par l'IANA (*Internet Assigned Numbers Authority*)
- Puis par le RIR (*Regional Internet Registry*)
 - Ex: RIPE (Europe), parmi 5 (APNIC, AfriNIC, ARIN).
- Un ASN est codé sur 16 bits (32 bits depuis 2007)
 - 2-64511 : publique
 - 64512-65535 : privé

BGP : Border Gateway Protocol

- Entre AS on utilise un protocole de routage BGP (protocole de type EGP), RFC 1654, RFC 1628.
- Les objectifs d'un protocole IGP sont différents d'un protocole de type EGP :
 - IGP transfère de la source vers la destination le plus efficacement possible.
 - EGP se préoccupe de stratégie.

Exemples :

- Le trafic sortant et entrant de Xinc. ne doit pas transiter par chez Ycorp.
- Aucun trafic ne doit transiter à travers certains AS ou pays.
- Transiter par un pays que s'il n'y a pas d'alternatives.

BGP

- La stratégie de routage est paramétrée manuellement dans chaque routeur BGP – Elle ne fait pas partie du protocole lui-même.
- Du point de vue d'un routeur BGP : Le « monde » est constitué de routeurs BGP interconnectés.
 - Deux routeurs BGP sont « connectés » s'ils partagent un réseau commun.

BGP pour le trafic de transit

Trois catégories de réseaux :

- **réseaux souches** : une seule connexion au graphe BGP.
- **réseaux multiconnectés** : peuvent être utilisés pour le trafic de transit (à l'exception de celui qu'il refuse)
- **réseaux de transit** : épines dorsales disposées à acheminer les datagrammes d'un tiers avec d'éventuelles restrictions

BGP : Communications

- Deux routeurs BGP communiquent via des connexions TCP.
- BGP : Protocole à DV (un peu différent de RIP) -
> **Path Vector**
 - poids des liaisons vers chaque destination
 - garde la trace du chemin exact utilisé
 - envoie à ses voisins le chemin exact qu'il utilise

BGP : Principe de fonctionnement

- Le routeur ignore tous les chemins passant par lui même.
- Les routes restantes sont évalués en fonction de
 - la distance ;
 - la stratégie de routage (si violé -> distance infinie).
- La fonction d'évaluation ne fait pas partie du protocole BGP (fonction choisie par l'admin. du système).

Routage inter-AS

Réseau d'une organisation (Université, grande entreprise, ISP
(*Internet Service Provider*))

Exemples :

AS2200 : Renater

AS1945 : LyRes Lyon Recherche et Enseignement Supérieur

AS2060 : ROCAD : Réseau Optique Campus de la Doua

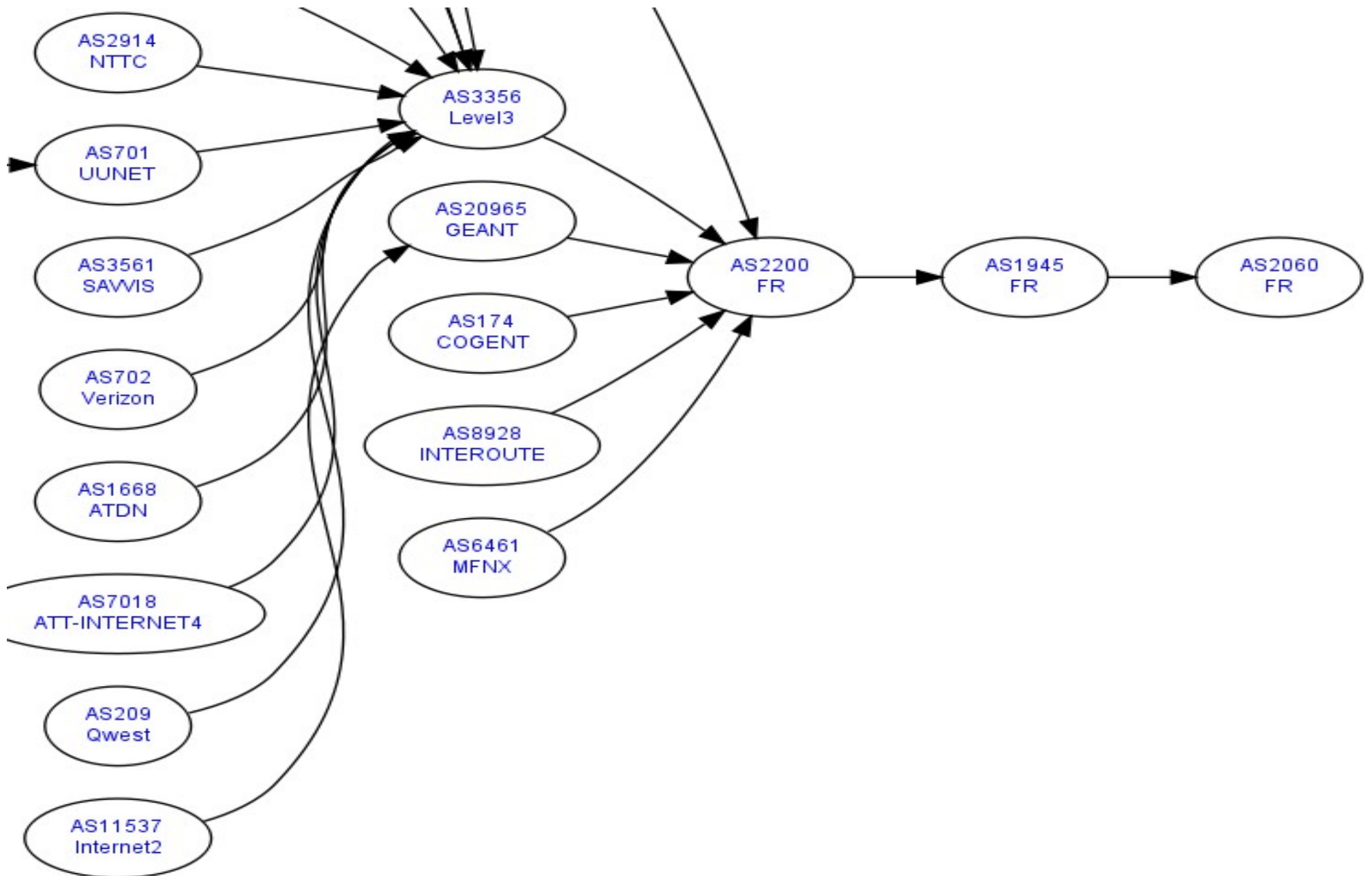
AS12322 : Proxad

AS3 : MIT

AS6 : Bull

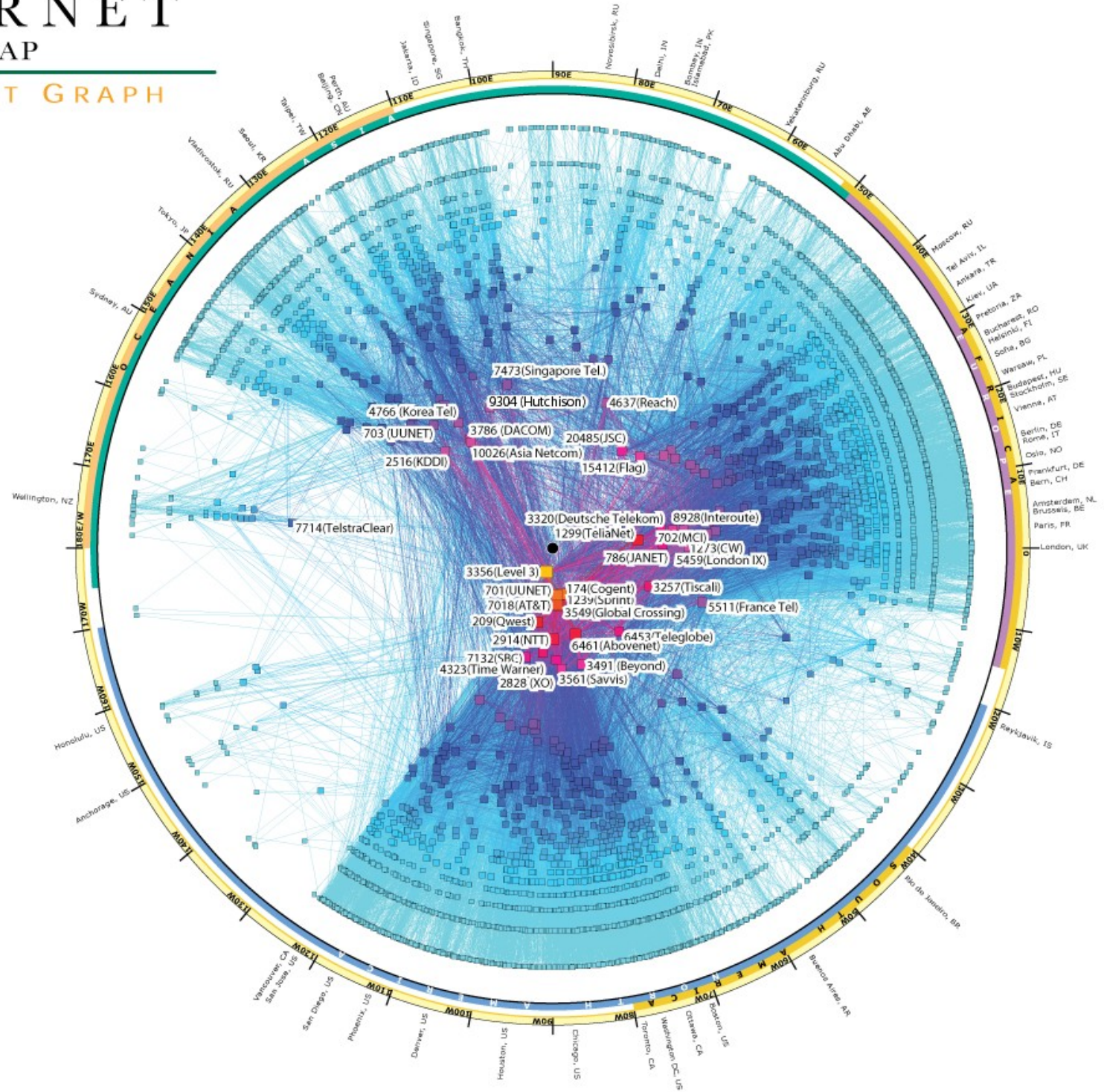
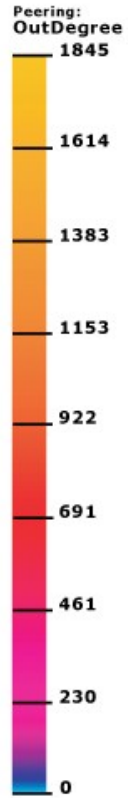
voir <http://www.robtex.com/as/>

Exemple : AS2060 – ROCAD



IPv4 INTERNET TOPOLOGY MAP

AS-level INTERNET GRAPH

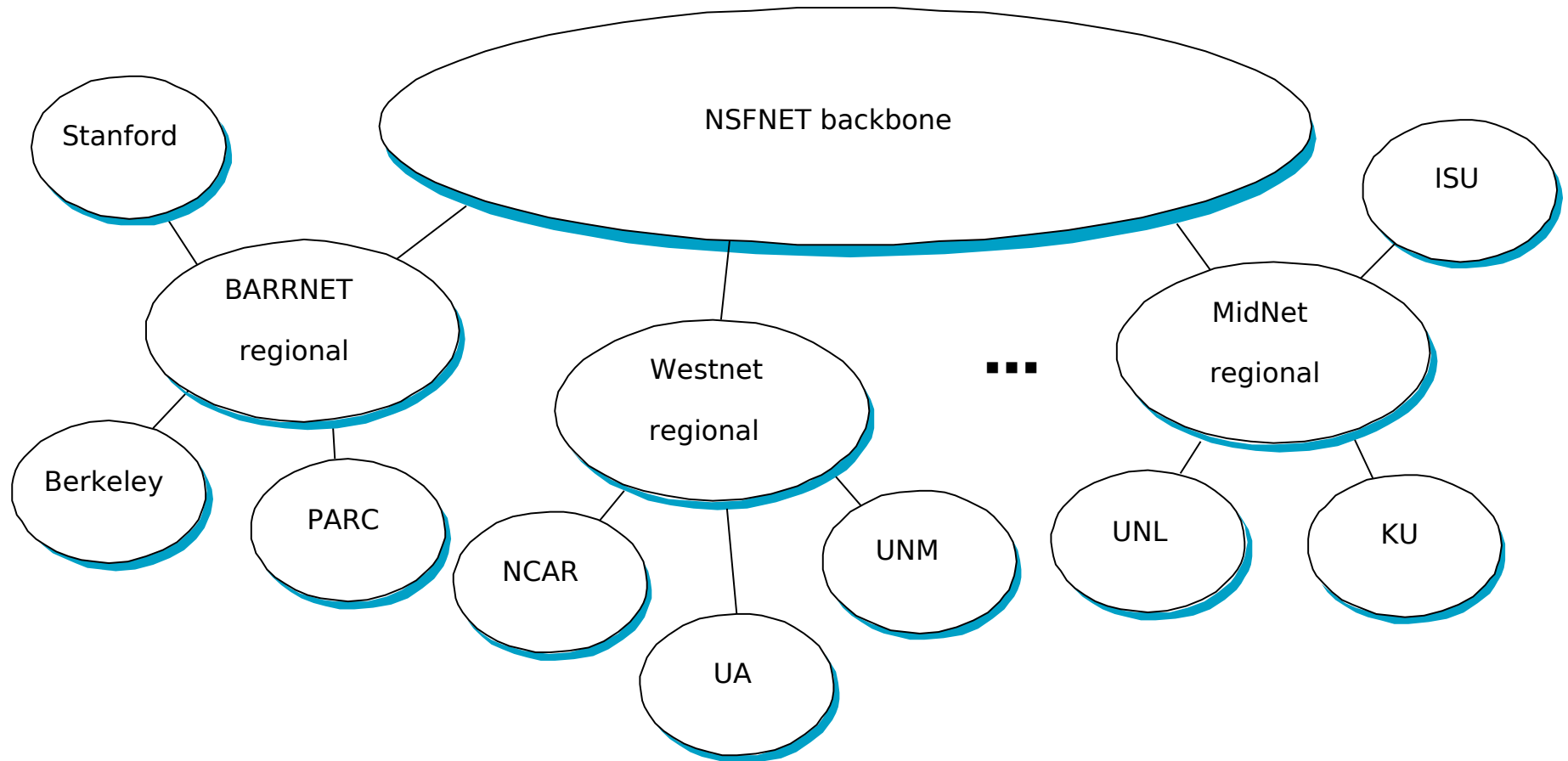


Routage inter-AS

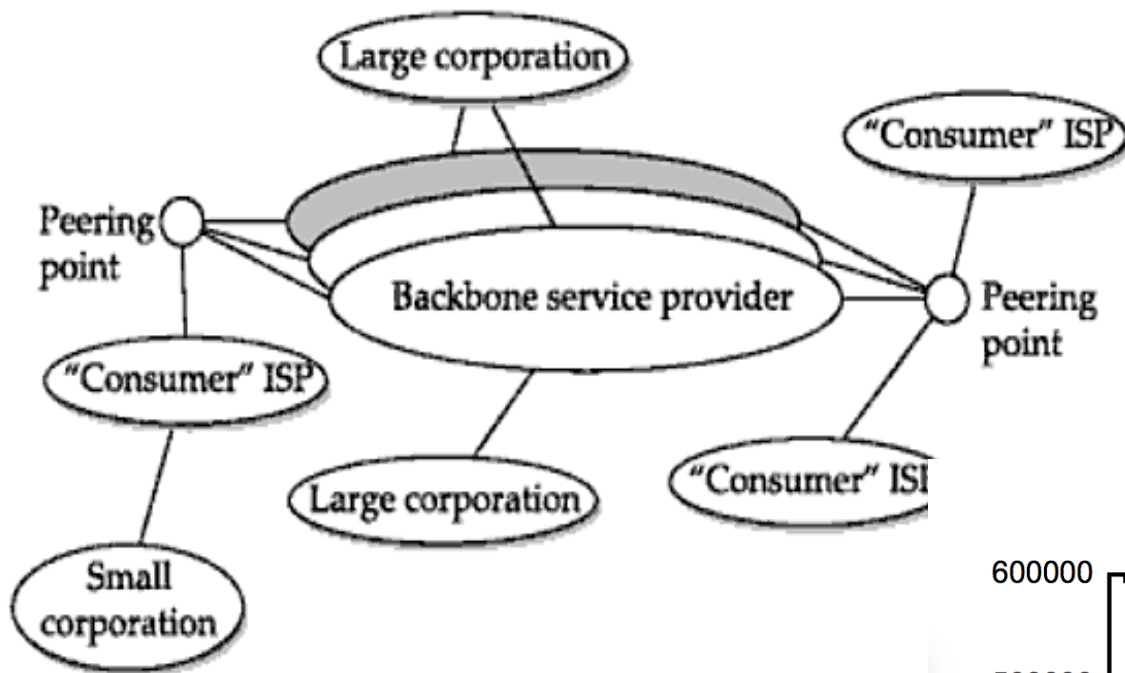
Problématique

- Nécessité de prendre en compte des contraintes économiques / politiques :
 - Accords commerciaux entre opérateurs : transit possible seulement via certains AS.
 - Choix entre plusieurs routes avec des coûts (financiers) différents.

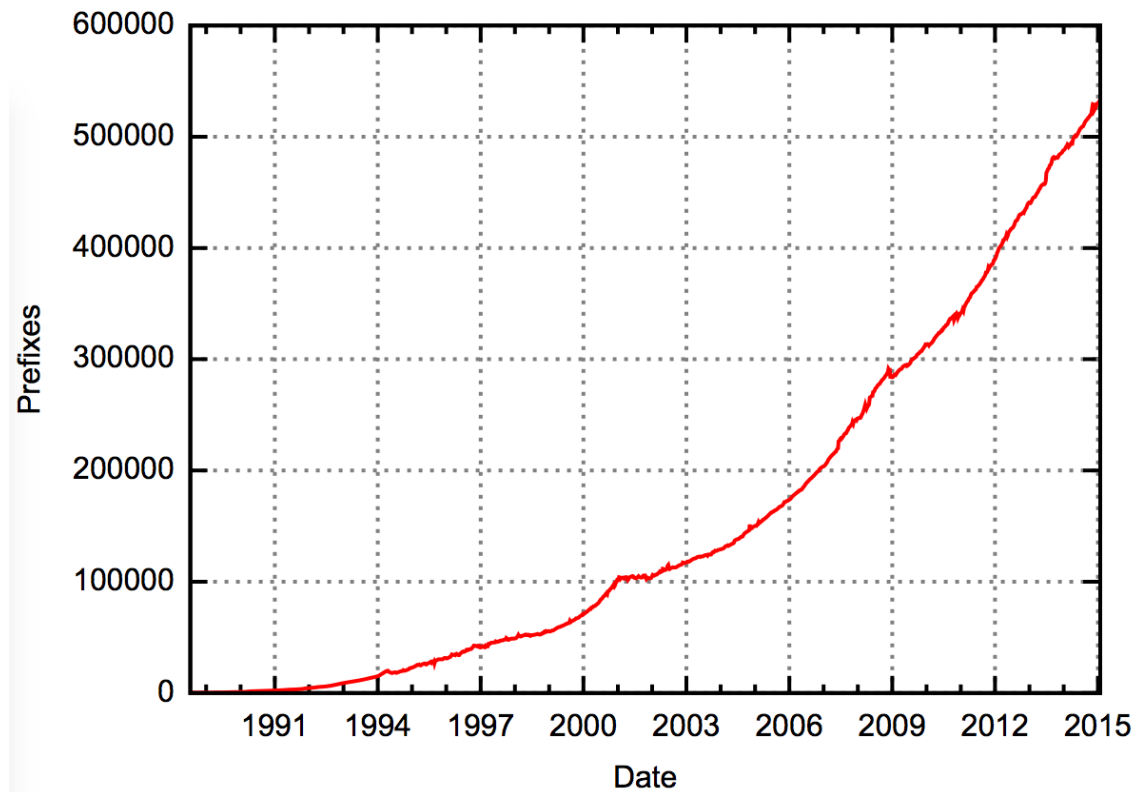
NSFNet : Internet en 1990



Internet aujourd'hui



Prefixes announced on the Internet



Peering

- Quand on est opérateur, 2 solutions pour pouvoir atteindre d'autres réseaux :
 - Payer un fournisseur de **transit** IP
 - Passer des accords réciproques avec d'autres opérateurs (**peering**), souvent dans un *Internet eXchange Point (IXP, GIX)*.
En France: SFINX (Renater), FreeIX (Proxad), PARIX (FT), ...
Lyonix.

Possible source de conflits

(2003: Free vs France Telecom ; 2008: Sprint vs Cogent)

Internet “Tiers”

- *Tier 1 network* :
 - Un réseau qui peut atteindre tous les autres réseaux sur Internet sans acheter du transit IP à un fournisseur ou payer un autre opérateur.
 - AT&T, Global Crossing, Level 3, NTT (Verio), Qwest, Sprint, Verizon (ex-UUNET), SAVVIS, AOL, AboveNet, Cogent, TeliaSonera, Teleglobe, XO Communications, ...
- *Tier 2 network* : réseau qui a passé des accords de peering avec d'autres opérateurs, mais qui achète aussi du transit IP.
- *Tier 3 network* : réseau qui assure sa connectivité uniquement en achetant du transit.

ISP de niveau 1 (tier 1)

- Les routeurs doivent être capable de réacheminer des paquets à des débits extrêmement élevés.
- Autres caractéristiques des ISP de niveau 1:
 - connexion **directe** à tous les autres ISP de niveau 1 (*peering*);
 - connexion à un grand nombre d'ISP de niveau 2 et à d'autres réseaux d'abonnés (*transit*);
 - couverture internationale.
- Également appelés réseaux **fédérateur d'internet**.
- Ex: France Telecom, UUNet (filiale de Worldcom), Sprint, AT&T, Genuity, MCI, AOL, NTT, Qwest,...

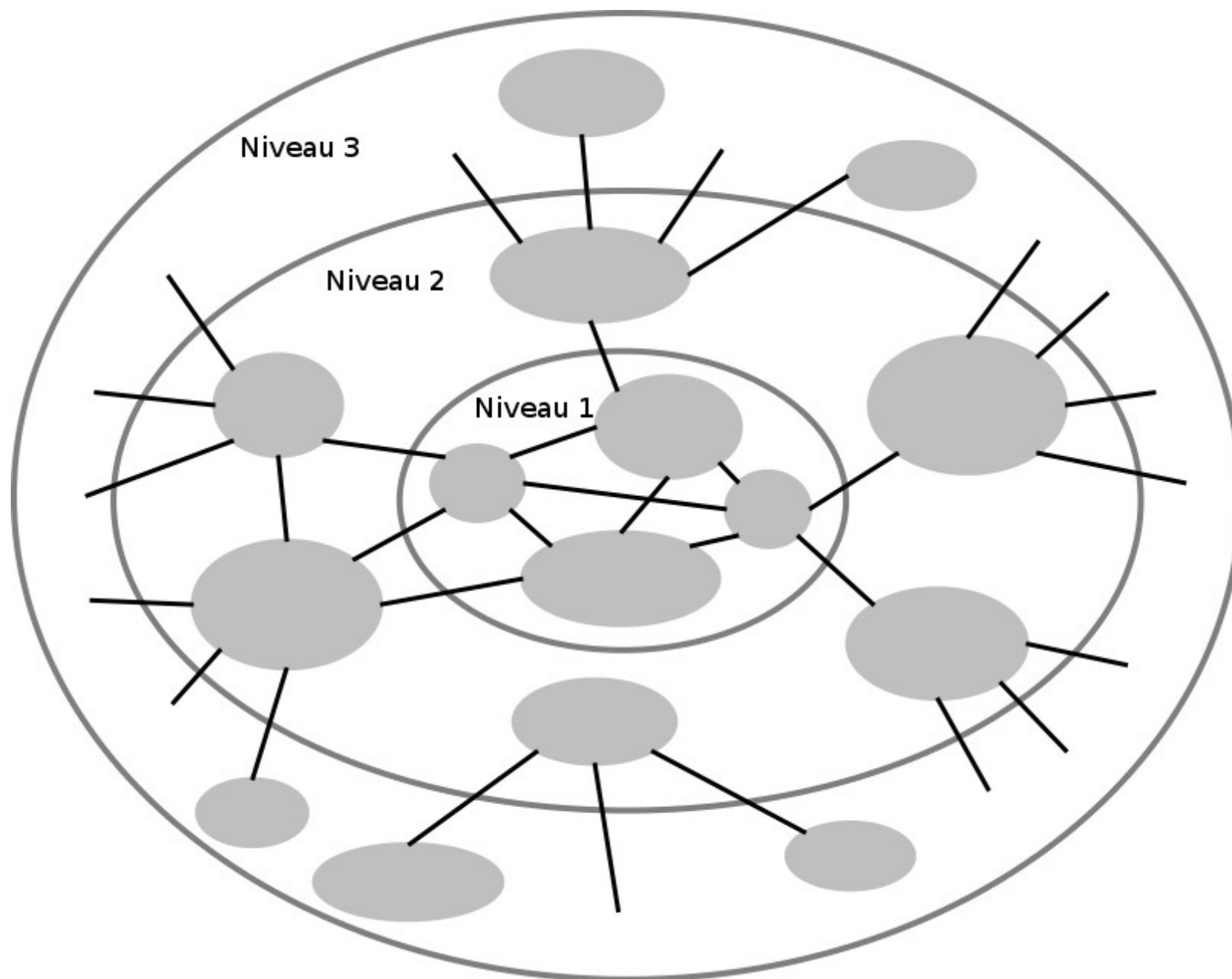
ISP de niveau 2

- Couverture de portée régionale ou nationale
- Il n'assure une connexion qu'à un nombre limité d'ISP de niveau 1 (*transit*).
- Pour accéder à *tout* l'internet, un ISP de niveau 2 doit acheminer du trafic via l'un des ISP de niveau 1 auquel il est relié. L'ISP de niveau 2 est alors **client** de l'ISP de niveau 1, ce dernier étant considéré comme **fournisseur**.
- Une grande entreprise (ou institution) relie son réseau directement à des ISP de niveau 1 ou 2 devenant ainsi leur client.
- Les tarifs établis par un ISP fournisseur à ses ISP clients dépend généralement du débit de la liaison.
- Un réseau de niveau 2 peut également se connecter à d'autres réseaux de même niveau (échange de flux horizontaux (i.e sans passer par un ISP 1)).

ISP de niveau inférieur

- Sous les ISP de niveau 2 se trouvent des ISP de niveau inférieur, qui proposent une connexion à l'internet via un ou plusieurs ISP de niveau 2.
- Tout en bas de la hiérarchie se trouvent les fournisseurs d'accès.
- Enfin, pour compliquer, certains ISP de niveau 1 sont également ISP de niveau 2 (intégration verticale).

Interconnexion entre les différents type d'ISP



3 catégories d'AS

- *Stub AS* : AS qui n'a qu'une seule connexion vers un autre AS.
- *Multihomed AS* : AS qui a des connexions vers plusieurs autres AS, mais qui ne transporte pas de trafic de transit.
- *Transit AS* : AS qui a des connexions vers plusieurs autres AS, et qui transporte du trafic de transit.

Préfixes BGP

- Les routeurs s'échangent des **préfixes**
= promesses de routage
= “je peux atteindre 192.100.0.0/16 via AS7,
AS100 et AS200”
- Ces préfixes sont filtrés (éventuellement) sur les routeurs.

Exemple : <http://www.robtex.com/as/as2200.html>

BGP : Confiance transitive

- Un fournisseur exporte aux autres fournisseurs les préfixes exportés par ses clients.
- En théorie, un fournisseur filtre les préfixes fournis par ses clients.
- Sinon, un client peut exporter un préfixe qui ne lui appartient pas.

BGP hijacking

- Annonce par un routeur d'un préfixe qui ne lui appartient pas.
- Février 2008 : *Pakistan Telecom* annonce une route pour 208.65.153.0/24 à son provider PCCW. Celui-ci ne la filtre pas, et annonce à son tour ce préfixe.
Résultat : En moins de 3 minutes (délai de propagation de la nouvelle route pour le préfixe), le trafic destiné à *Youtube* est envoyé au Pakistan pour 2/3 d'Internet, pendant plus de 2 heures.
- Solutions :
 - Systèmes d'alerte : PHAS, RIPE MyASN, Renesys...
 - Internet Routing Registry, pour détecter des incohérences

BGP : Bilan

- Les algorithmes à DV font souvent un mauvais choix car ils sont incapables de déterminer lesquels de leurs voisins ont des **routes indépendantes**.
- BGP (*Border Gateway Protocol*) utilise un protocole à vecteur de chemin (*path vector*).
 - Pas de boucles mais des tables de routage beaucoup plus grandes.
 - Les routeurs utilisent TCP (contrairement à tous les autres protocoles).
 - BGPv4 est complexe et difficile à maintenir.

Fin