La couche réseau - PLAN

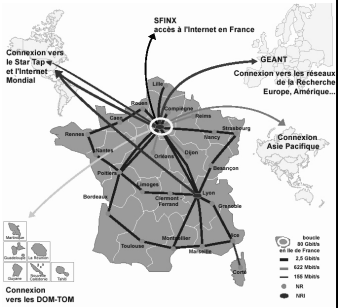
- ✓ La routage dans l'Internet
  - ✓ Routage dans l'Internet (aujourd'hui) …
  - ✓ Routage Intra-domaine (IGP)
  - ✓ Routage Inter-domaines (EGP)
  - ✓ Les futurs directions du routage IP

Routage dans l'Internet (aujourd'hui)

- L'Internet est composé aujourd'hui d'un très grand nombre de « systèmes autonomes »
  - Un AS est un ensemble de réseaux qui sont sous un même contrôle "administratif " → Internet Service Provider (ISP) ou Fournisseur de Service Internet (FAI)
    - Exemples : Orange, Renater, Sprint, UUNET, ...
- Les AS sont interconnectés entre eux: ex. RENATER

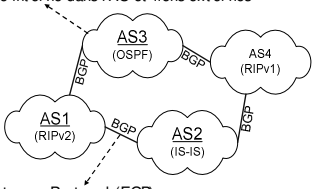
Routage dans l'Internet (aujourd'hui)

- Renater
  - Connexion directe avec le réseau de la Recherche européen : Géant
  - Connexion directe avec le réseau de la recherche en Corée : KREONET. Réseaux qui assure a son tour la connectivité avec APAN le réseau de la recherche de l'Asie-Pacifique
  - Connexion directe avec le réseau OpenTransit de FT : service offrant une connectivité avec l'Asie et l'Amérique notamment
  - Connexion à Internet en France via SFINX (Service for French INternet eXchange) : une infrastructure commune qui permet l'interconnexion de tous les opérateurs opérant en France (Orange, Free, ..)



Routage dans l'Internet (aujourd'hui)

- Taille de l'Internet 10.000 à 20.000 AS incluant 100.000 à 150.000 réseaux
- D'où la nécessité d'un routage hiérarchique
  - Interior Gateway Protocol (IGP)
    - Topologie interne dans l'AS et liens externes



  - Exterior Gateway Protocol (EGP)
    - Distribue les routes globales en considérant chaque AS comme une boite noire

La couche réseau - PLAN

- ✓ La routage dans l'Internet
  - ✓ Routage dans l'Internet (aujourd'hui) …
  - ✓ Routage Intra-domaine (IGP)
    - ✓ RIPv1 et RIPv2
    - ✓ OSPF
  - ✓ Routage Inter-domaines (EGP)
    - ✓ BGP
  - ✓ Les futurs directions du routage IP

Routage Intra-domaine

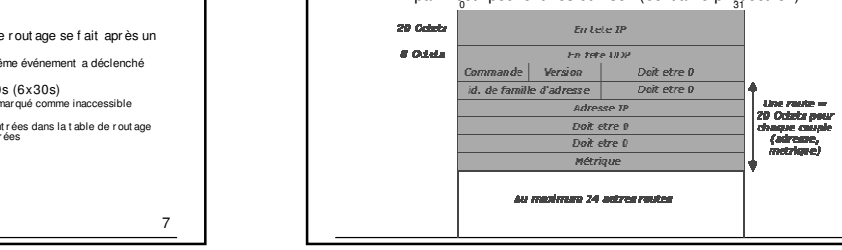
- Routage statique
- Routage avec vecteurs de distance
  - RIP (Routing Information Protocol)
    - Fréquemment utilisé dans de petits réseaux simples
- Routage avec états de liaison
  - Open Shortest Path First (OSPF)
    - Fréquemment utilisé dans des réseaux d'entreprises
  - IS-IS (Intermediate system to intermediate system)
    - Fréquemment utilisé par des ISPs
    - Initialement créé pour le réseau CLNP (OSI) mais utilisé dans IP

RIPv1 : Routing Information Protocol version 1

- Routage à vecteurs de distances
- Décrit par le document IETF
  - RFC 1058
- Caractéristiques
  - Deux types de nœuds
    - Actifs : les routeurs
    - Passifs : les hôtes
  - Temporisateur de broadcast → 30s
  - Les envois de VD après modification des tables de routage se fait après un temps aléatoire variant entre 1 à 5s
    - Limiter le phénomène de tempête de MAJ que le même événement a déclenché chez plusieurs voisins
  - Temporisateur de rafraîchissement → 3min = 180s (6x30s)
    - Si on ne reçoit aucun message, la route silencieuse est marqué comme inaccessible (distance infini)
    - Si aucune mise à jour après 240 secondes, toutes les entrées dans la table de routage correspondant au routeur qui ne répond pas seront retirées
  - Métrique unique : nombre de sauts
    - Entre 1 et 15 (16 = infini)

RIPv1 : format de paquet

- RIP est un protocole associé à IP mais qui n'utilise pas directement IP
  - Il est implémenté au dessus d'UDP
  - Utilise le port UDP 520 pour l'émission et la réception des messages
    - N° port <1024 => seul les processus privilégiés exécutés par 'root' peuvent les utiliser (certaine protection)



RIPv1 : format de paquet (2)

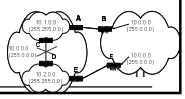
- Commande
  - RESPONSE : celle utilisée pour diffuser les infos de routage (vecteurs de distances avec split horizon) toutes soit toutes les 30 secondes soit pour transmettre des "mises à jour déclenchées"
  - REQUEST : utiliser pour deux raisons :
    - Un nouveau routeur qui demande à ses voisins leurs VD pour former ses tables de routage
      - Il annonce une seule route 0.0.0.0 avec distance ∞
      - Le récepteur du msg répond normalement (en utilisant split horizon)
    - Demande spécifique par rapport à une entrée spécifique de la table de routage
      - Utilisée pour un but de déboguage
      - La réponse ne tient donc pas compte du split horizon

RIPv1 : format de paquet (3)

- Version
  - 1 pour RIPv1
  - 2 pour RIPv2
- Identifiant de famille d'adresse
  - 2 pour IP (en réalité c'est la seule valeur utilisée !)
- Maximum de 25 entrées par message
  - Chaque entrée est codée sur 20 bytes
  - La taille maximale d'un message RIP est de 512 bytes
  - Un VD peut nécessiter plusieurs messages

RIPv1 : fonctionnement de RIPv1

- A la réception d'un message RESPONSE, le routeur examine toutes les entrées
- L'adresse IP dans le msg RIP peut être celle :
  - D'un réseau
    - Identifié grâce à la classe d'adresse (A, B ou C)
    - Partie hôte et sous-réseau = 0
  - D'un sous réseau ou d'un hôte
    - Les entrées correspondant aux autres sous-réseaux du même réseau que le routeur sont eux reconnus puisque chaque routeur connais son propre masque de sous réseau
    - Toutes les entrées correspondant à un autre réseau sont censées être regroupées dans une même entrée (même route)
      - Le support de routage par hôte est optionnel mais non utilisé car risque d'avoir des tables de routage énormes
    - Remarques:
      - Pas possible de séparer entre sous-réseau et hôte car il n'y a pas un champs « type d'adresse IP » et le masque de sous réseau non plus n'est pas disponible dans le message RIPv1 -> Sol RIPv2
      - Le masque de sous réseau n'est pas censé être propagé en dehors du réseau concerné
    - Supposition forte : on route vers le réseau car on suppose que tous les sous-réseaux d'un même réseaux sont interconnectés entre eux : chacun des routeurs du réseau est supposé joindre n'importe quel sous-réseau
      - Problème de panne d'un lien reliant deux sous-réseaux
      - Problème d'optimalité du routage



RIPv1 : fonctionnement de RIPv1 (2)

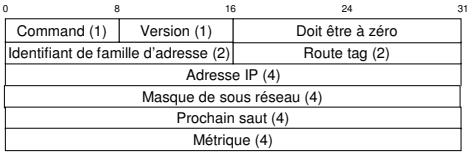
- Chaque entrée de la table de routage contient plusieurs entrées dont :
  - L'adresse destination (réseau, sous réseau ou hôte)
  - La métrique associée à la destination (nb. sauts, etc..)
  - L'adresse du prochain routeur (next-hop)
  - Un fanion 'MAJ récente'
  - Plusieurs timers
- Un routeur traite un message de réponse RIPv1 comme suit :
  - Vérifie si l'entrée est valide
    - Adresse de classe A, B ou C ne commençant pas par 127 ou 0 (sauf pour 0.0.0.0)
    - La partie hôte de l'adresse n'est pas une adresse de broadcast
    - La métrique < ∞
  - Pour chaque entrée valide, incrémenter la métrique de 1 et faire les vérifications et mise à jours d'usage
    - Si une entrée est rajoutée (mise-à-jour ou juste confirmée), alors réinitialiser son timer (180s)
    - Si une entrée est rajoutée ou mise-à-jour (meilleur chemin) elle doit être "marquée" en tant que tel
    - Dans le cas ou son timer va arriver à échéance, alors le champs « next hop » d'une entrée est mis-à-jour même si la métrique reçue est inférieure ou égale à la valeur stockée (non pas seulement inférieure)

RIPv1 : fonctionnement de RIPv1 (3)

- Dans le cas ou il y'a eu un changement dans la table de routage
  - Préparer une réponse par interface (suivant split horizon) et l'envoyer après un temps aléatoire (entre 1 et 5s)
- Dans les autres cas
  - Envoyer une réponse par interface (suivant split horizon) toutes les 30s
- Configurations des interfaces d'un routeur
  - RIP en sa version plus simple doit connaître pour chaque interface
    - L'adresse IP et le masque de sous réseau correspondants
    - Sauf mention contraire, RIP initialisera dans sa table de routage une entrée pour chacun des sous-réseaux qui lui sont connectés avec une distance de 1
    - Ensuite, le routeur envoi un msg de "requête" à tous ses voisins pour pouvoir construire ses tables
      - L'administrateur peut interdire le broadcast sur une des interfaces
        - Parce qu'il connaît la topologie et il est seul routeur d'un réseau local et envoyer des messages périodiquement serait un gaspillage
        - Maintenir des routages fixes sur certains interfaces (entrée permanente sans timers)
        - Utiliser un autre protocole que RIP
        - Interdire certaines detinations avec lesquels on ne veut pas communiquer
      - L'administrateur peut aussi jouer sur les métriques des interfaces
      - L'absence de messages RIP implique que la route est inaccessible après 180s
    - RIP passera ensuite le résultat de ces calculs à IP

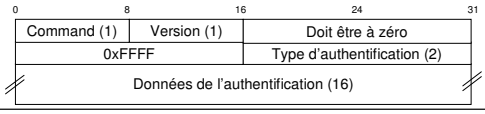
RIPv2 : quoi de neuf ?

- Routage à vecteurs de distances (toujours)
  - Améliorer les manques constatés de RIP (routage par sous-réseau, support CIDR, authentification des messages,)
- Décrit par le document IETF
  - RFC2453
- Idée : garder le même format de paquets mais l'améliorer en utilisant les champs "à zéro"



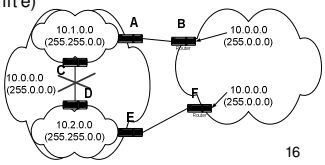
RIPv2 : format de paquet

- Le champs « Identifiant de famille d'adresse »
  - 2 pour IP et …0xFFFF pour l'authentification
  - Constat : Aujourd'hui n'importe qui peut être root sur une machine (pas le cas avant) et se faire passer pour un routeur RIP…quoi faire ?
    - Authentifier les infos que l'on reçoit en remplaçant la 1ere entrée du paquet RIP par un segment d'authentification
  - Permet aux routeurs de vérifier l'origine du paquet (ceci est optionnel et configurable)
- NB : compatibilité avec RIPv1 (message ignoré vu que le type ≠2)
- Les champs :
  - Type d'authentification : algorithme d'authentification utilisé
  - Données d'authentification : données utilisées par l'algorithme pour permettre l'authentification



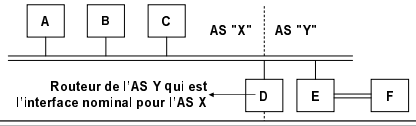
RIPv2 : format de paquet (2)

- Le champs « Masque de sous-réseaux »
  - Permettre un routage par sous-réseaux en dehors du réseau d'origine en adjoignant le masque de sous réseau à chaque adresse IP
    - Ceci permet également :
      - Le support de masques de tailles variables au niveau du "même" réseau et donc la possibilité de décomposition de sous réseaux en sous-réseaux
      - Le support de l'agrégation de plusieurs adresses de classe C en une entrée groupée dans la table de routage → CIDR
  - NB : Les entrées de type sous-réseau seront ignorées par RIPv1 (compatibilité)



RIPv2 : format de paquet

- Le champs « Prochain saut »
  - On a remarqué que souvent plusieurs AS partagent un même lien, par exemple un câble Ethernet ou un anneau FDDI pouvant jouer le rôle de backbone permettant l'interconnexion de ces AS -> RIPv2 supporte ce type d'utilisation
    - Si A veux envoyer un paquet à F, il l'enverra à D qui le remettra sur le lien partagé pour l'envoyer à E (sous-optimal)
    - Comme D sait que les routes vers F peuvent passer par E directement, il annonce la destination F dans le domaine X en envoyant adresse + métrique + « Prochain saut = E»
    - Grâce à ce champs D peut dire à A « La distance vers F est f, mais le meilleurs relais vers F n'est pas moi, D, mais plutôt E »
    - Un routeur utilisera la valeur 0.0.0.0 pour le champs « Prochain saut » pour dire que c'est par lui que passe le meilleur chemin (compatibilité avec RIPv1)
- Le champs « Route Tag »
  - Utilisé pour la déclaration des routes externe (en conjonction avec un EGP) → Cas général et non le cas ou plusieurs AS partagent un même lien



RIPv2 : les problèmes et améliorations à faire

- Problème de la synchronisation des envois de VD
  - RIP, de part son fonctionnement, mène à des synchronisations des envois des VD par tous les routeurs du réseau
  - La charge et les congestions dans le réseau augmentent toutes les 30s
  - Plusieurs améliorations sont proposées dans la littérature (et même dans le RFC 1058!) pour résoudre ce problème
- Amélioration : Mise à jour acquittées
  - Etait parfois nécessaire dans le passé car une partie des x paquets formant le VD pouvait être systématiquement perdus
- Support multi-métrique
  - Il n'est pas recommandé d'utiliser RIP (avec la métrique nombre de sauts) dans un réseau composé de liens ayant des caractéristiques très variables
    - Un lien X25 (lent et cher) face à un anneau FDDI (rapide et gratuit)
  - Une nouvelle valeur pouvant servir de métrique pour choisir la liaison la plus rapide = taille nominal d'un paquet / débit du lien → à utiliser avec prudence (difficile de choisir la valeur de l'infini)
- Problème des boucles (cf. routage avec vecteurs de distances)
  - Une solution existe mais elle est complexe (contrairement à la philosophie de RIP)
- Convergence lente (cf. routage avec vecteurs de distances)

RIPv2 : le prix de la simplicité

- RIP est un protocole contenant deux messages et une table
  - Peut être implémenté en quelques heures (simple algo)
  - Le risque d'avoir un problème grave avec RIP est réduit
  - RIP donne des résultats très satisfaisant avec des topologies de réseaux simple et peu de pannes de liens (liaisons fiables)
  - ⇨ C'est pour cela qu'il est simple et qu'on l'utilise tellement
- Mais, pour des réseaux plus large et plus complexe (en terme de topologie), il est carrément inadéquat
  - Ne sais pas faire si on a plus de 15 sauts
    - A cause des problème de convergence lente de protocoles à VD
    - Du fait que l'on préfère choisir une valeur petite pour l'infini (=16)
    - …
  - C'est pour cela que certains préfèrent utiliser des protocoles de la famille LS

La couche réseau - PLAN

- ✓ La routage dans l'Internet
  - ✓ Routage dans l'Internet (aujourd'hui) …
  - ✓ Routage Intra-domaine (IGP)
    - ✓ RIPv1 et RIPv2
    - ✓ OSPF
  - ✓ Routage Inter-domaines (EGP)
    - ✓ BGP
  - ✓ Les futurs directions du routage IP

OSPF : Open Shortest Path First (2)

- Protocole de routage avec états de liaisons (LS)
- Décrit par les documents IETF
  - RFC 1131 pour la version 1 et le RFC 2328 pour la version 2 (plusieurs RFC de OSPFv2 ont été rendus obsolète par celui-ci)
- C'est un protocole qui reprend tous ce qui a été défini pour le routage avec LS :
  - Définition du voisinage
    - Envoie de messages HELLO pour découvrir ses voisins
  - Base de données (LSPs) distribuée et procédure de diffusion
    - Envoi de LSPs
      - Acquittements, numéro de séquence et age
        - Age = 0 au début, il est incrémenté de 1 à chaque saut puis de 1 à chaque seconde passée dans la base des LSP. MaxAge = 1 heure.
        - MaxAgeDiff > 15 minutes → prendre le plus récent
    - Envoi périodique de LSP (30min) et lorsqu'une ligne change fortement
    - Un routeur peut demander à un routeur adjacent un LSP particulier
  - Paquets spéciaux pour les routes externes

OSPF : Open Shortest Path First (3)

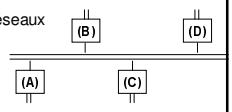
- Christian Huitema écrit dans son livre :
  - RIP est simple et limité, OSPF est très puissant et quelque peu complexe …
  - Pourquoi est il si puissant ?
    - Converge plus rapidement (Une seule itération via inondation)
    - Pas de boucles
    - Support de métriques à valeurs plus fines (car pas de notion d'infini)
    - Support plus aisé du multi-métriques (car pas de notion d'infini)
    - Support naturel du routage multi-chemins
      - Si deux chemins de même coût existes entre un point A et un point B, il est prouvé que le partage de la charge entre ces deux chemins est mieux que d'en utiliser qu'un
        - Faut quand même faire attention …peut avoir un impact négatif sur des connexions TCP par exemple
    - Meilleurs choix des routes externes par rapport aux algorithmes à VD
      - Car pas de borne max et une complexité au niveau du calcul qui est moindre
  - Pourquoi est il si complexe ?
    - Car il a été spécialement conçu pour supporter l'hétérogénéité de l'Internet
      - Séparation entre hôtes et routeurs
      - Prise en compte des réseaux à lien partagé (diffusions)
      - Prise en compte des réseaux à circuits virtuels (non-diffusions)
      - Division des grand réseaux en régions

OSPF : séparation entre routeurs et hôtes

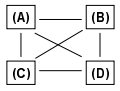
- Cette séparation est faite par le fait
  - Qu'un lien vers un routeur voisin est identifié par l'adresse IP de ce routeur
    - On parle de lien réseau (*network link*)
  - Qu'un lien vers un réseau terminal (stub network) est identifié par son adresse de sous réseau
    - Un stub network est un sous réseau
    - On parle de lien vers un stub network (*link to stub network*)

OSPF : les réseaux à lien partagé

- OSPF offre un support générique aux réseaux à lien partagé (dit aussi broadcast)
  - FDDI, Ethernet, Token ring, …
  - Connectivité complète



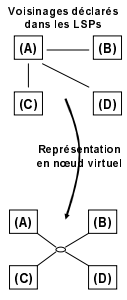
- Représentation en graphe totalement connecté
  - Non fidèle à la réalité (panne du lien A-B -> panne de tout le réseau)
  - Le problème du "carré de N" pour informer tous les routeurs du réseau :
    - Si N routeurs sont connectés à un même réseaux à lien partagé Alors chaque routeur déclare N voisins (liés au lien partagé) dans son LSP
      - N-1 pour les autres routeurs du lien partagé
      - +1 pour le stub network (les hôtes joignables par ce lien partagé)
    - ⇨ N^2 est la taille total de ce qui est déclaré par les N nœuds



- Idée d'OSPF, réduire la taille du graphe à N

OSPF : les réseaux à lien partagé (2)

- Comment réduire la taille du graphe ?
  - Un des routeurs est désigné comme représentant du réseau local
  - Les autres routeurs sont dits adjacents
    - Ils n'établissent leurs voisinages que par rapport au routeur désigné et synchronisent leurs bases des LSPs avec lui
    - Ils n'envoient leur LSP que vers l'adresse multicast des routeurs désignés (224.0.0.6)
      - Reçu par le routeur désigné et le routeur désigné backup (voir transparent suivant)
    - Les routeurs désignés utilisent l'adresse multicast (224.0.0.5) pour envoyer leurs LSP à tous les routeurs OSPF
  - L'élection du routeur désigné se fait grâce au protocole HELLO
  - Représenter le réseau local comme un noeud virtuel auquel tous les routeurs sont connectés
    - C'est le routeur désigné qui joue le rôle de ce nœud virtuel



OSPF : les réseaux à lien partagé (3)

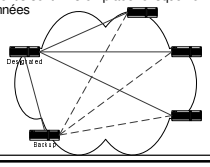
- Comment OSPF résout les problèmes qui peuvent survenir de cela ?
  - Cas ou le routeur désigné tombe en panne (détecté par le protocole Hello)
    - Election d'un "*backup designated router*" en même temps que le "*designated router*"
    - Tous les routeurs adjacents doivent maintenir leurs voisinages vers le *backup designated router*
      - Il est membre du groupe multicast des routeurs désignés (224.0.0.6)
    - Le *backup designated router* reste muet et ne retransmet aucun LSP
    - En cas de panne du routeur désigné, il le remplace et un autre backup est élu
  - Certains liens auront une métrique nulle vu que le nœud du milieu est virtuel (cela est permis par OSPF)
    - Le calcul de distance entre 2 nœuds du même réseau = distance du nœud source vers réseau + distance du réseau vers le nœud destination (artificielle et donc nulle pour la calcul de la distance à un hôte du réseau)

OSPF : les réseaux à circuits virtuels

- La plupart des réseaux WAN mettent en place des circuits virtuels (au niveau liaison de données)
  - Dans les années 80 : X25
  - Dans les années 90 : Frame Relay et ATM (débits + élevés)
- Les routeurs IP était alors connectés via ces technologies : IP over X25, IP over ATM, …
  - Pas de diffusion ni de transmission multipoint
- Deux routeurs IP qui communiquent doivent donc mettre en place des circuits virtuels entre eux …
  - Problème du "carré de N"
    - Pour avoir une connectivité complète entre toutes les paires de routeurs voisins on aurait besoins de N.(N-1)/2 circuits virtuels (VC) à mettre en place
    - Les infos de routage sont diffusées sur tous les VC
    - Ces réseaux sont des réseaux ou on paye au volume
      - ⇨ Le trafic de routage peut coûter très cher
        - Ou lorsque le critère économique se mêle à la technique …

OSPF : les réseaux à circuits virtuels (2)

- Comment résoudre ce problème ?
  - Même manière que pour les réseaux à diffusion
    - Election d'un routeur désigné et d'un backup …
      - Chacun envoi ces LSP directement au routeur désigné et au backup (point-à-point)
      - Le routeur désigné répond séparément à chaque routeur (en point-à-point puisqu'il n'y pas de diffusion) pour synchroniser avec lui sa table des LSP
    - La découverte du voisinage est toujours réalisée grâce au protocole Hello
      - Chacun envoi régulièrement un paquet Hello au routeur désigné qui envoi dans un même paquet Hello la liste de tous les routeurs connecté au réseau
  - Ceci permet de limiter le trafic des LSP dans les VC
    - Les VC entre routeurs adjacents sont mis en place lorsque l'on en a besoins pour le trafic de données



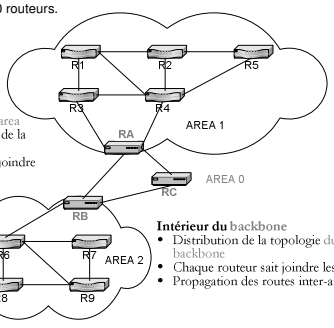
OSPF : support de grands réseaux

- Objectif :
  - Eviter d'avoir de grosses tables de routage sinon …
    - Base des LSPs énormes
    - Volume important pour les messages de routage
    - Durée de recalcul des routes élevée
    - …
- Solution : hiérarchisation en régions (areas)
  - Région backbone : cœur du réseau
    - Tous les routeurs connectés à deux areas ou plus font partie du backbone
  - Toutes les autres areas doivent se connecter au backbone
    - Au moins un routeur de chaque areas doit être connecté au backbone

  -> Structure en étoile dont le centre est le backbone
  - Routage doit permettre d'aller de n'importe quelle area vers une autre à travers le backbone

OSPF : support de grands réseaux (2)

Un area va contenir 10-200 routeurs.

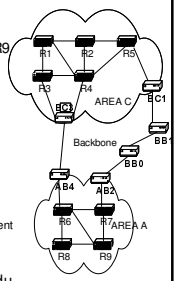


Intérieur d'un AREA
- Distribution de la topologie de l'area
- Les routeurs ne connaissent rien de la topologie des autres areas
- Chaque routeur sait comment rejoindre le backbone

Intérieur du backbone
- Distribution de la topologie du backbone
- Chaque routeur sait joindre les areas
- Propagation des routes inter-areas

OSPF : support de grands réseaux (3)

- Un routeur de la zone A contient
  - Les LSPs de tous les routeurs dans la zone (R6-R9 AB2, AB4)
  - Un compte rendu résumé émis par ses routeurs backbone (AB2, AB4) pour tous les réseaux et sous-réseaux IP qui font partie du backbone et des autres zones (zone C)
  - Les LSPs dit externes émis par les routeurs du backbone et qui n'appartiennent qu'au backbone (BB0, BB1) et relayés par AB2 et AB4
    - Permet de savoir comment joindre un routeur du backbone à partir de ses propres routeurs de backbone
      - c-à-d, savoir comment joindre les réseaux directement connectés à un routeur du backbone
- Remarque : OSPF ne fait pas un routage hiérachique stricte entre les zones (compte rendu résumé des autres zones):
  - Si un lien tombe et divise une région en deux, les comptes rendus résumés font en sorte de ne pas se tromper dans la mise en place du routage vers chacune des parties de la région



OSPF : Améliorations

- Gestion des stub areas ou les très petites zones → ne contenant 'en général' qu'un seul routeur qui fait partie du backbone
  - Dans ce routeur, toutes les routes externes à la zone sont remplacées par une seule route par défaut
    - Permet d'éviter la surcharge des comptes rendus résumés et externes
      - Car on n'a plus besoin dans ce cas de décrire ces routes externes qui peuvent être très nombreuse (# 10.000) face aux routes d'une zone OSPF (<200)
      - Contenu dans OSPFv2
- Gestion des "Not So Stubby Areas – NSSA (RFC 1587)" ou zones contenant qu'un seul routeur qui fait partie du backbone et un voir plusieurs routeurs inter-domaines
  - Toutes les routes appartenant à une autre zone du même AS sont remplacées par une route par défaut
  - Les routes inter-domaine, apprises grâce à un autre protocole EGP, ne le sont pas
    - On dit alors qu'une NSSA est une zone ou toutes les routes externes (à la zone) sont remplacées par une route par défaut ...sauf certaines.
- D'autres améliorations au stade de la recherche
  - Amélioration de la sécurité
  - Améliorer l'ingénierie de trafic avec OSPF (routage multi-chemins)
  - Débordement de la base des LSPs

OSPF : complexité et services

- Par rapport à RIP :
  - La documentation d'OSPF est 5 x plus grosse
  - Plusieurs informations à gérer (Notion de MIB)
  - L'implémentation d'OSPF prend beaucoup plus de temps et nécessite beaucoup plus de lignes de code :-)
  - RIP contient 2 messages ...OSPF, 5 messages différents et 3 procédures
  - Les LSP sont acquittés contrairement aux VD
  - Pas besoin d'élire de routeurs particuliers dans RIP
  - Besoins d'une table de routage et d'une base de LSPs dans OSPF (appelé LSA – A pour *Advertisements*)
- Pourquoi toute cette complexité
  - Le routage est important et OSPF est plus efficace que RIP
  - Essayez de router avec RIP les 60.000 routes externes qu'un routeur OSPF apprend à partir des routeurs de bordures de la zone backbone (60.000 c'était en Juin 1999 ..)

La couche réseau - PLAN

- ✓ La routage dans l'Internet
  - ✓ Routage dans l'Internet (aujourd'hui) ...
  - ✓ Routage Intra-domaine (IGP)
    - ✓ RIPv1 et RIPv2
    - ✓ OSPF
  - ✓ Routage Inter-domaines (EGP)
    - ✓ BGP
  - ✓ Les futurs directions du routage IP

Routage inter-domaines

- Comment faire communiquer des AS entre eux ?
  - RIP, OSPF, IS-IS, ...
    - Protocole de routage intra-domaine (intra-AS)
    - Caractéristiques liées aux contraintes des AS, non adaptés au routage inter-domaines
  - EGP : Exterior Gateway Protocol
    - Premier protocole de routage inter-domaines
      - En réalité il permet tout juste la connectivité entre un ensemble réduits d'AS ...aujourd'hui nous avons près de 9000 AS
      - EGP construit un réseau backbone sous forme d'arbre contenant tous les routeurs inter-domaines de l'Internet (pas réaliste avec 9000 AS)
  - BGP : Border Gateway Protocol
    - Protocole à Vecteur de ...Chemins (en remplacement des VD)
    - Remplaçant d'EGP
    - La version 4 (dénotée BGP4) est celle déployée aujourd'hui
      - Son déploiement a commencé en 1995
      - Décrite par le document IETF : RFC 1771

Routage inter-domaines (2)

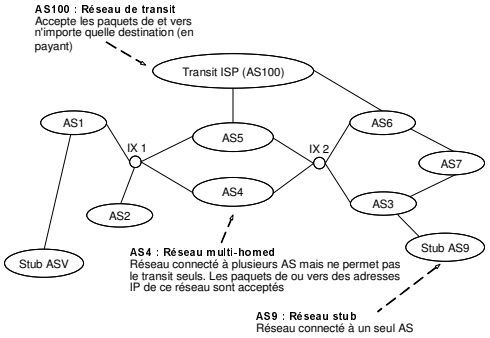
- Routage hiérarchique
  - L'Internet se résume aux différents AS et aux routeurs se trouvant au bord de ces AS
    - BGP ne connaît pas le détail interne de la topologie de chaque AS
- Deux routeurs BGP directement connectés s'échangent les informations de routage de façon fiable (en utilisant le protocole TCP)
  - TCP assure
    - La vérification de l'état des connexions (perte de connexion => routeur down)
      - TCP keep-alive
    - La fiabilisation de l'échange des informations de routage

BGP : Border Gateway Protocol

- Objectif de BGP
  - Déterminer les routes entre AS, tout en prenant compte des contraintes politiques ou commerciales
    - Belgacom refuse de faire transiter du trafic provenant de France Télécom et destiné à BT mais s'il vient de WIN (filiale Belgacom) et est à destination de BT, il sera accepté
      - FT n'a qu'a se débrouiller par elle-même pour joindre BT
    - RENATER accepte le trafic à destination des universités françaises mais refuse de servir de relais pour du trafic commercial
      - Car, par exemple, la liaison France-USA n'est pas payé par Renater pour servir du trafic autres que celui des universités françaises
    - Le trafic entre deux bâtiments de Microsoft ne doit jamais passer par un bâtiment d'IBM ou SUN ou Apple, ...
      - Car c'est des concurrents commerciaux
  - Les politiques de routage sont aujourd'hui spécifiées et configurées manuellement dans chaque routeur

37

Types de réseaux



38

BGP

- Principe de fonctionnement de BGP
  - Routage avec vecteurs de chemins (vecteurs de distance modifié) pour éviter les boucles
    - Chaque vecteur de chemins contient le chemin complet (liste d'AS) et pas seulement l'adresse de la destination
      - Si un AS reçoit un vecteur de chemins qui contient déjà son identification => boucle => ne pas prendre ce vecteur en compte
  - Chaque AS choisit lui même la route qu'il utilise sur base des vecteurs reçus
    - La route choisie n'est pas nécessairement la plus courte ...elle dépend des contraintes (politiques) configurées
- Avantage
  - Permet d'éviter les problèmes du vecteur de distance
    - Il est possible de détecter les boucles lorsque l'on connaît le chemin complet
  - Permet de connaître les AS par lesquels un paquet passera avant d'arriver à la destination
    - Nécessaire pour supporter du routage politique
    - Exemple : chaque fois que Microsoft recevra un vecteur qui lui signal une route passant par IBM, elle pourra ignorer cette route et en choisir une autre
      - NB: ceci ne serait pas vraiment convergent avec un algorithme LS
        - Contraintes difficilement réalisables avec l'algo de Djikstra

39

BGP : Exemple

Table de Routage [AS1]
AS3 : Sud
AS2 : Est
AS5 : Est [AS2 - AS5]
AS4 : Sud [AS2 - AS4]

Table de Routage [AS2]
AS3 : Sud [AS4 - AS3]
AS1 : Ouest
AS5 : Est
AS4 : Sud

Table de Routage [AS5]
AS1 : Ouest [AS1 - AS2]
AS2 : Ouest
AS3 : Est [AS2 - AS4 - AS3]
AS4 : Sud-Ouest

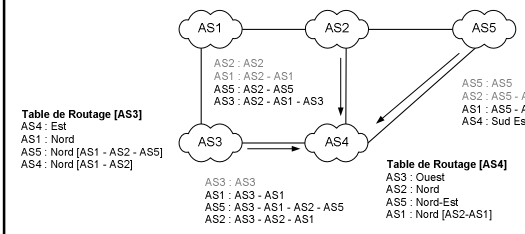
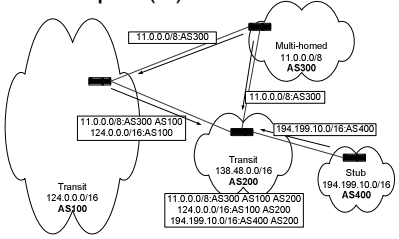


Table de Routage [AS3]
AS4 : Est
AS1 : Nord
AS5 : Nord [AS1 - AS2 - AS5]
AS4 : Nord [AS1 - AS2]

Table de Routage [AS4]
AS3 : Ouest
AS2 : Nord
AS5 : Nord-Est
AS1 : Nord [AS2-AS1]

40

BGP : Exemple (2)



- Le choix des routes se fait en trois étapes :
  1. Eliminer les routes interdites par la politique de routage
     - Microsoft éliminera tous les vecteurs de chemins passant par le réseau d'IBM, par exemple
  2. Choisir la route la plus courte pour une destination (nombre d'AS intermédiaires)
  3. Si plusieurs routes de même longueur, choisir l'une des route suivant ses propres critères (souvent aléatoires)

41

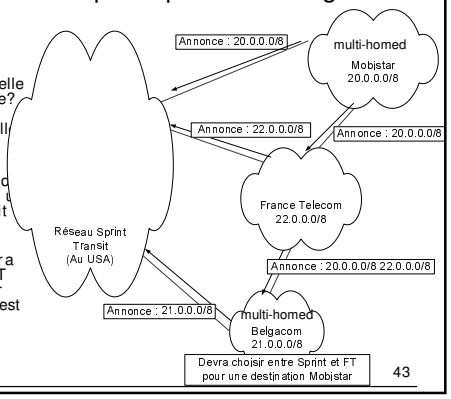
BGP: support de politiques de routage

- 2 façons pour le support de politiques de routage:
  1. Chaque routeur BGP choisit quelles routes il annonce aux AS voisins
     - Tous les voisins ne recevront pas nécessairement la même route
       - France Télécom n'annoncera pas à Belgacom les même routes qu'elle annonce à Mobistar (sa filiale)
     - Permet de contrôler le trafic qu'un AS accepte
       - Un AS accepte d'être le transit pour router le trafic de et/ou vers certains AS et pas d'autres
  2. Chaque routeur choisit parmi les routes qu'il reçoit celle qui est la meilleur
     - Permet de contrôler quelle direction prend le trafic qui sort de l'AS
       - Permet à Microsoft de ne pas faire passer son trafic par les réseaux de ses concurrents

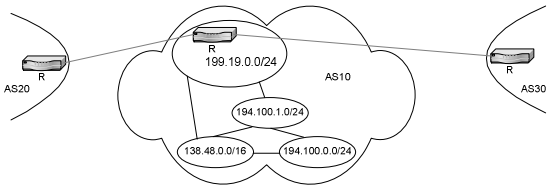
42

BGP: support de politiques de routage - exemple

- Quelle route on annonce vers l'extérieur? Quelle route on propage? Quelle route on accepte quand elle arrive?
- Mobistar filiale de FT. Donc FT est un réseau de transit mais uniquement vers Mobistar
- Belgacom choisira de passer par FT au lieu de passer vers l'US car c'est moins cher



BGP : agrégation d'adresses



- Rappel : un AS est un ensemble de réseaux, sous-réseaux et routeurs appartenant à la même autorité administrative
- Un routeur BGP annonce aux routeurs extérieurs les sous-réseaux qui peuvent être atteints via lui
  - Pour l'AS10, on annonce trois routes même si on en a quatre en fait :
    - 199.19.0.0/24
    - 138.48.0.0/16
    - 194.100.0.0/23 → agrégés avec CIDR

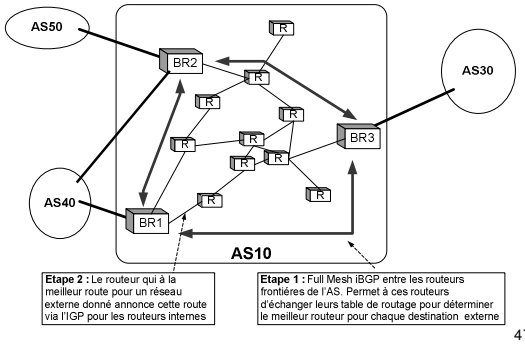
CIDR : Classless InterDomain Routing

- Vue la croissance des demandes d'adresses de Classe B (16 bits pour le réseau), l'IANA a décidé de donner plusieurs adresses de classe C (24 bits pour le réseau) **contiguës** à toutes les organisations ayant moins de 4096 hôtes à connecter et moins de 16 sous-réseaux
  - Idée :
    - Restructuration de l'assignation des réseaux IP
      - Préfixes d'adresses de taille variable (typiquement entre 13 et 27 bits)
    - Dans le cas ou les sous réseaux sont géographiquement bien distribuées, l'idée est de faire un routage dit CIDR en regroupant les adresses contiguës grâce à la technique de masquage de sous réseau → on parle de *Mulinetworks* ou d'*agrégation hiérarchique*
      - Exemple : 197.8.0/23 regroupera 197.8.0/24 et 197.8.1.0/24
- Support de CIDR dans BGP se fait comme suit :
  - Soit T un routeur ayant comme vecteurs de chemins
    - Path1 : Through {T} reaches 197.8.0/23
    - Path2 : Through {T,X} reaches 197.8.2/24
    - Path3 : Through {T,Y} reaches 197.8.3/24
  - On peut donc annoncer l'agrégation suivante au routeur Z :
    - Path1 : Through (Sequence (T), Set (X,Y)) reaches 197.8.0.0/22

Interactions entre routage intra- et inter-domaines

- Le routage intra-domaine construit les meilleurs routes à l'intérieur du réseau d'un AS
- Le routage inter-domaines construits les meilleurs routes externes
- La questions découlant de cela est :
  - Que faire si un AS possède plusieurs liens externes ?
    - Comment savoir lequel utiliser pour aller vers une destination externe donnée ?
      1. Un mécanisme qui va permettre aux routeurs se trouvant aux frontières de l'AS de décider entre eux quel est le meilleur routeur pour atteindre une destination externe
         - Solution : **liaison iBGP interne (à l'intérieur de l'AS) entre les routeurs connectés à l'extérieur**
      2. Ensuite, un mécanisme pour informer les routeurs internes de l'AS en redistribuant l'information apprise
         - Solution : interaction entre BGP et le protocole de routage intra-domaine IGP

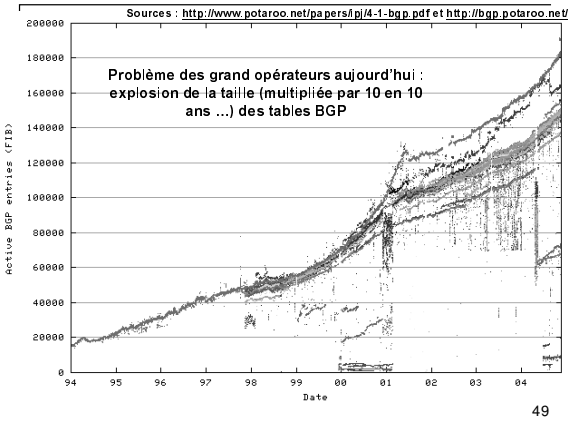
Interactions entre routage intra- et inter-domaines (2)



BGP et CIDR : sauveurs de l'Internet

- CIDR a été conçu en 1992 pour prévenir la mort de l'Internet en 1994
  - Cause du décès : manque d'adresses de classe B
- BGP a été conçu en 1994 pour prévenir la mort de l'Internet
  - Cause du décès : explosion des taille des tables de routage
- BGP à permit de quitter le modèle réseau dorsal unique organisé en arbre pour relier les AS (EGP) en un modèle complètement maillé
- BGP4 a cependant atteint ses limites
  - Important nombre de configurations manuelles
    - Configuration des politiques de routage
  - Une erreur de configuration peut se propager à tous l'Internet
  - Aujourd'hui les tables BGP sont devenue énormes

Sources : http://www.potaroo.net/papers/ipj/4-1-bgp.pdf et http://bgp.potaroo.net/



La couche réseau - PLAN

- ✓ La routage dans l'Internet
  - ✓ Routage dans l'Internet (aujourd'hui) …
  - ✓ Routage Intra-domaine (IGP)
    - ✓ RIPv1 et RIPv2
    - ✓ OSPF
  - ✓ Routage Inter-domaines (EGP)
    - ✓ BGP
  - ✓ Les directions actuelles et futurs du routage IP

Protocoles existants

Famille IGP

- RIPv1 (Routing Information Protocol)
- RIPv2 (Routing Information Protocol)
- IGRP(2) (Interior Gateway Protocol)
- EIGRP(2) (Enhanced IGRP)
- OSPFv1(1) (Open Shortest Path First)
- OSPFv2 (Open Shortest Path First)
- IS-IS(3) (Intermediate System Intermediate System)

Famille EGP

- BGP4 (Border Gateway Protocol)
- IDPR(4) (Inter-Domain Policy Routing)
- IDRP(3) (Inter-Domain Routing Protocol)
- EGP(1) (Exterior Gateway Protocol)

(1) Obsolète
(2) Propriétaire CISCO
(3) OSI. Indiqué par soucis de complétude
(4) Peu/pas déployé

Protocoles existants (2)

	Basé sur	Norme/ propriétaire	Commentaire
RIPv1 & v2	Dist Vector	IETF RFC1058 RFC2453	· Adapté aux petits AS (15 hops max) · Simple mais pb temps de convergence
OSPFv1 & v2	Link State	IETF RFC1131 RFC2328	· Concurrent IGRP · Résoud les limitations de RIP · Complexe
IS-IS	Link State	OSI	· Support adressage hiérarchique · support CLNP et IP · Complexe
IGRP/EIGRP	Dist Vector	CISCO	· Résout les pbs de boucles · Complexe

Protocoles existants (3)

	Basé sur	Norme/ propriétaire	Commentaire
BGP4	Dist Vector	IETF RFC1771	· Remplace EGP, BGP et BGP3 · Le plus répandu actuellement
IDPR	Link State	IETF RFC1478	· Permet de spécifier des restrictions d'accès · N'a jamais été supporté par des produits
IDRP	Dist Vector	OSI	· Perçu comme "BGP5"
EGP	Dist Vector	IETF	· Remplacé par BGP · C'est plus un protocole de découverte de connectivité qu'un protocole de routage

Le futur du routage

- Des sujets n'ont pas été traité dans ce cours et ils génèrent encore une activité substantielle de recherche
- Le routage multicast : jusqu'à présent, on a vu comment calculer des routes entre une source et une destination. Le routage multicast calcule une route entre une source et plusieurs destinations. Il est principalement utilisé pour des services multimédia (vidéoconférence)

Le futur du routage (2)

- Le routage avec qualité de service (QoS) : jusqu'à présent on calcule la meilleure route selon une certaine métrique « unique ». Le routage QoS calcule une route adéquate en fonction de plusieurs métriques (ex : route la plus rapide avec une fiabilité comprise dans un certain intervalle)
- Le routage dans des environnement mobiles : si les routeurs sont mobiles (réseau dynamique), la problématique de routage change
  - RIP et OSPF ne sont pas adaptés

Biblio.

- Christian Huitema, "Routing in the Internet", Second Edition, Prentice Hall PTR. ISBN 0-13-022647-5