



Le routage dynamique

-
- les routages internes aux AS
 - à vecteurs de distance (RIP, EIGRP,..)
 - à état de liens (OSPF, IS-IS,..)
 - le routage externe aux AS
 - à vecteur de chemins (BGP)



AS : Autonomous System

- Un **AS** est un ensemble de réseaux informatiques dont la politique de routage interne est cohérente. **Un AS est sous le contrôle d'une entité/organisation unique.**
- Au sein d'un AS, on a un protocole de routage interne : un **IGP (Interior Gateway Protocol)**
- Un AS est identifié par un numéro de 16 ou 32 bits, fourni par les RIR (Registres Internet Régionaux). Pour l'Europe, **c'est le RIPE-NCC** (Réseaux IP Européens - Network Coordination Centre).
- 5 RIR : ARIN, APNIC, LACNIC (*Latin America and Caribbean Network Information Center*), AfriNIC

RIP (1)

✧ Routing Information Protocol

- ✧ Protocole intra-domaine (IGP)
 - RFC 1058, 2453
 - Version 1.0 et 2.0
- ✧ Protocole à vecteur de distance
 - Utilisation algorithme de Bellman-Ford
 - Inondation de messages
 - Itératifs
 - Distribué

RIP (2)

- ✦ Une route est associée à une métrique
 - ◆ le nombre de bonds (Hop)
 - ◆ Pas de poids sur les liens par défaut
 - ◆ l'infini est représenté par la métrique 16.
- ✦ **Routeurs échangent leurs routes avec leurs voisins**
(calcul du plus court chemin)
- ✦ Les routeurs annoncent leurs routes
 - ◆ Périodiquement (30 s)
 - ◆ Lorsqu'une information a été mise à jour (déclenché)

RIP (3)

✧ Table de routage

- ✧ l'@IPv4 de destination
- ✧ le nombre de sauts pour atteindre la destination
- ✧ l'@IPv4 du prochain routeur le long du chemin vers la destination finale
- ✧ un drapeau pour indiquer que l'information sur la route a changé récemment (drapeau de changement de route)
- ✧ différents temporisateurs...

RIP (4)

✦ Messages

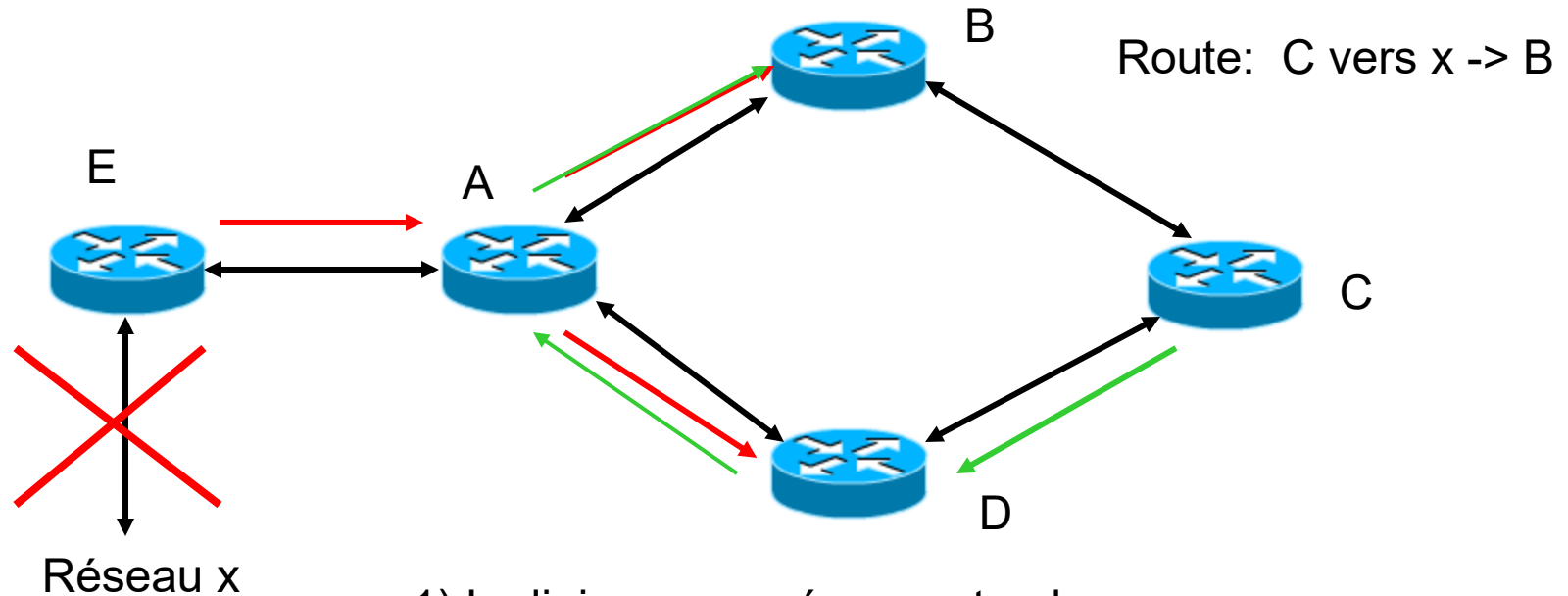
- ✦ les messages sont envoyés en UDP, port 520
- ✦ Itératif
 - Continue la propagation de message après mise à jour
- ✦ Auto terminant
 - Pas de signal pour stopper
- ✦ 2 sortes :
 - Request -> demande de routes (envoyé par broadcast ou multicast)
 - Response -> annonces de routes
(max 25 entrées par messages)

RIP (5)

✧ Maintenance table de routage

- ✧ un routeur qui ne reçoit plus rien d'un voisin
 - Premier temps : route inutilisable
 - Mise à la distance 16 de toutes les routes passant par ce voisin
 - Puis, suppression de la route
- ✧ A chaque réception,
 - Mise à jour pour les routes connues
 - Installation pour les routes non connues

Exemple de boucle RIP



- 1) La liaison vers réseau x tombe en panne
- 2) E avertit le routeur A → plus de route vers x
- 3) A avertit les routeurs B et D → plus de route vers x
- 4) C envoie une mise à jour de sa table vers D
route vers x en passant par B
- 5) D envoie à A route vers x en passant par C, puis
A envoie à B

PB

RIP (6)

✧ Eviter la formation de boucles

✧ Horizons partagés (split horizon)

- ✧ un routeur ne renvoie pas une route à un voisin si celui-ci vient de la signaler infinie

✧ Mise à jour déclenchée et « route poisoning »

- ✧ Dès qu'un changement est vu, envoi d'une MAJ.
- ✧ Si réseau indisponible, « route poisoning » = metric 16

RIPv2 (1)

✧ Différences v1, v2

- ✧ Utilisation du multicast
 - Adresse : 224.0.0.9
- ✧ Gestion des réseaux CIDR
- ✧ Permet l'authentification des paquets RIPv2
 - RFC 4822
 - authentifications possibles : Plain-text (à éviter), MD5 et SHA..
 - Actuellement, SHA non pris en charge par Cisco

RIPv2 (2)

✧ Configuration

- ✧ définition clé : key chain nom
puis son n° : key 1
et la clé partagée : key-string Maclé
- ✧ Mise en place sur l'interface
 - ip rip authentication (key-chain nom /mode md5)

Utilisation de RIPv2 pour prendre en charge IPv6

- Définition du nom : ipv6 router rip monroutage
- Sur chaque interface, déclaration : ipv6 rip monroutage enable
- Chiffrement des données par IPSec

OSPF (1)

✦ Open Shortest Path First (rfc 1247, 2328)

- ✦ Protocole intra-domaine (IGP)
- ✦ **Protocole à état de liens**

✦ Link-State Database (LSDB)

- ✦ Structure de données fondamentales
- ✦ Une copie est maintenue par tous les routeurs
- ✦ Contient une description du graphe des routeurs de l'area
- ✦ Chaque lien vers un réseau est associé à une entrée de la LSDB avec une métrique associée.

OSPF (2)

✧ 3 tables sont maintenues

◆ Base de données de contiguïté

- Table de voisinage
- Show ip ospf neighbor

◆ Base de données d'états de lien (LSBD)

- Base de données topologique
- Show ip ospf database

◆ Table de routage

- Show ip route

OSPF (3)

✦ Les routeurs utilisent des paquets LSA pour mettre à jour leurs informations

- ◆ Link State Advertisement
- ◆ Chaque routeur envoie ce qu'il sait du réseau

✦ **Tous les routeurs ont la même LSDB**
(mis à jour via les LSA)

✦ Chaque routeur utilise l'algorithme du plus court chemin pour construire sa table de routage (Algo de Dijkstra)

- ◆ A chaque changement, recalcule du plus court chemin
- ◆ Plusieurs métriques possibles, mais actuellement :
$$\text{cout} = 10^8 / \text{bande passante en bits} \quad \rightarrow 100\text{Mb/s} = 1$$

✦ Chaque routeur met à jour une table de voisinage

OSPF (3)

✧ Segmentation de la zone : **AREA**

- ✧ Obligation d'avoir une aire dorsale (backbone)
 - C'est l'area 0
- ✧ Chaque area est ensuite raccordée au backbone
 - Valeur de l'area ≥ 1
 - Pour communiquer entre 2 areas, obligation de passer par le backbone

✧ Différents routeurs :

- ✧ *Internal router* (raccordé à 1 seule area)
- ✧ *Area Border Router* (raccordé à plusieurs area)
- ✧ *Backbone Router* (Router dont 1 interface appartient à l'area 0)
→ les ABR sont des Backbones routers

OSPF (4)

✧ Messages

- ✧ paquets OSPF sont envoyés au-dessus de IP (type:89)
- ✧ TTL = 1, pour éviter la diffusion
- ✧ adresse multicast : 224.0.0.5, 224.0.0.6
- ✧ En_tête du paquet OSPF (24 octets) + payload suivant le type

Field length, n bytes	1	1	2	4	4	2	2	8	Variable
	Version number	Type	Packet length	Router ID	Area ID	Check- sum	Authent- ication type	Authentication	Data

✧ 5 types de messages

- ✧ Hello → pour l'adjacence
 - Type 1
 - 10 s ou 30 s si réseau NBMA (Non Broadcast Multi Access)
 - Lien cassé (dead interval) = 4 * hello-interval

OSPF (5)

✧ 5 types de messages (suite)

- ✧ Database Description (DBD) → description du LSDB
 - Utilisation d'un numéro de séquence unique (0x80000001)
 - A chaque nouvel envoi, le numéro augmente
- ✧ Link State Request (LSR) → pour avoir des infos sur une route
- ✧ Link State Update (LSU) → mise à jour
 - Contient une ou plusieurs LSA
 - 11 types de LSA
 - ♦ Type 1 : Router LSA : liste les liens vers les autres routeurs
 - ♦ Type 2 : Network LSA : le DR envoie la liste des routeurs
 - ♦ Type 3 : Summary LSA : Un ABR envoie des infos d'une autre zone
 - ♦ Type 4 : ...
- ✧ Link State Acknowledgment → accusé de réception

OSPF (6)

✧ Paquet Hello

- ✧ Boot du routeur
- ✧ périodiquement (permet de savoir si le routeur est vivant)
- ✧ Permet de découvrir les voisins
 - Si découverte voisin, envoie de Database Description
 - Puis LSR si besoin, puis LSU...

✧ Pour les réseaux à diffusion (ethernet)

- ✧ Election d'un DR et BDR (Designated Router)
- ✧ Utilisation de l'id du routeur : choix du plus grand
 - Router-id ou **loopback** ou autre

OSPF (7)

✧ Supporte l'authentification

- ✧ RFC 2328 et 5709

- ✧ 2 sortes d'authentification: plaintext ou MD5, et SHA

- Md5 sur l'interface : `ip ospf message-digest-key 1 md5 mac`
puis sur router ospf : `area 0 authentication message-digest`
- SHA : définition d'une clé avec `key chain`
puis sur interface : `ip ospf authentication key-chain name`

✧ Compatible CIDR

✧ Plus stable au niveau convergence et plus rapide

OSPF (8)

- ✦ Mais demande plus de mémoire que RIP
- ✦ Envoie plus de messages à la mise en place que RIP

Pour supporter IPv6, OSPFv3

```
conf global : ipv6 router ospf {nb}  
               router-id x.x.x.x  
  
interface ...  
               ipv6 ospf {nb} area {0,1..}
```

-> Chiffrement des données par IPsec

IS-IS (1)

✧ Intermediate System to Intermediate System

- ✧ protocole intra-domaine (IGP)
- ✧ développé à partir de 1985 par Digital Equipment, et nommé DECNET phase V Routing
- ✧ normalisé par l'ISO 10589 et nommé IS-IS.
- ✧ IS-IS fonctionne pour **la couche 3 OSI**, c.à.d pour le CLNP (Connectionless network protocol) qui est inclus dans CLNS (Connectionless Network Service)
 - IETF publie ce protocole dans RFC 1142
- ✧ En 1990, extension à l'IP : RFC 1195

IS-IS (2)

✧ IS-IS similaire à OSPF

- ✧ Protocole à état de liens
- ✧ Mise à jour via des LSP (Link State Packets)
- ✧ Table de voisinage
- ✧ Base de données topologiques
- ✧ Utilisation de Dijkstra
- ✧ Convergence rapide
- ✧ Supporte le CIDR

Rem : OSPF a été créé pour IP après IS-IS

IS-IS (3)

- ✧ Découpage en zone, mais toutes les zones sont équivalentes
- ✧ 4 niveaux de routage
 - ◆ Niveau L1 : routage intrazone
 - ◆ Niveau L2 : routage interzone
 - ◆ Niveau 0 : entre routeur et end user (?)
 - ◆ Niveau 3 : entre AS (non utilisé)
- ◆ 3 sortes de routeur ; routeur N1, routeur N2 , routeur N1-N2
- ◆ Pour changer de zone, obligation de passer par un routeur N2
- ◆ Routeur N1 → une BD topologique
- ◆ Routeur N1-N2 → 2 BDs topologiques

IS-IS (4)

✧ Utilisation du plus court chemin

- ✧ Métrique par défaut : coût
- ✧ Délais
- ✧ Taux d'erreur
- ✧ Coût monétaire

✧ IP, utilisation du coût seulement (10 pour chaque saut)

✧ CLNP, utilisation des 4 métriques

✧ Pour réseau à diffusion, élection d'un DIS (Designated Intermediate System)

IS-IS (5)

✧ Messages

- ✧ indépendance par rapport au protocole routé
- ✧ Directement sur la couche 2
- ✧ Utilisation du format **TLV** (Type, Longueur, Valeur)
(permet d'évoluer rapidement → IPv6)
- ✧ Authentification

- ✧ 3 sortes de messages
 - Hello
 - LSP Link State Packet
 - SNP Sequence Number Packets

IS-IS (6)

✧ Découverte des voisins

- ✧ Une fois l'interface activée, le protocole va chercher automatiquement s'il existe autre part une interface IS-IS
- ✧ Si interface trouvée, 3 messages Hello sont envoyés (poignée de main)
- ✧ Messages IIH (Intermediate System to Intermediate System Hello)
- ✧ Messages envoyés en multicast au niveau 2 :
@MAC : 0180C2000014 et 0180C2000015
- ✧ Message Hello pour niveau 1 ou pour niveau 2
- ✧ Comme OSPF, plus de message HELLO → fin du voisinage

IS-IS (7)

✧ LSP (Link State Packet)

- ✧ Echange de données topologiques
- ✧ Envoyé seulement aux voisins
- ✧ Chaque paquet LSP à
 - un LSP-id pour identifier l'expéditeur
 - Un n°séquence pour la fraîcheur de l'information (incrémenté de 1 à chaque nouvel envoi)

✧ SNP

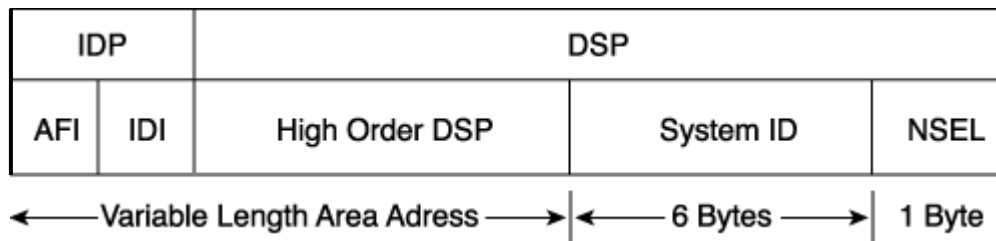
- ✧ contient plusieurs LSP

(mise à jour de la BD topologiques : **flooding**)

IS-IS (8)

✦ Adressage CLNS (adresse NSAP)

max 20 octets → 160 bits



IDP : Initial Domain Part → extérieur à l'AS

AFI : Authority Format Identifier : corps administratif attribuant l'adresse

IDI : Initial Domain Identifier : sous-organisation

Ex : 47.0005, civil du US. Government, 47.0006, US department of Defense,

DSP : Domain Specific Part

HODSP : sous-réseau

SID : adresse unique d'un équipement

NSEL : identifie un process sur équipement : 0 pour un routeur

Ex : 49.0001.0000.0C12.3456.00

IS-IS (9)

✧ IS-IS

- ✧ Moins de CPU utilisé que OSPF
- ✧ plus simple pour évoluer
- ✧ une seule LSP, alors que plusieurs LSA
- ✧ convergence plus rapide que OSPF

✧ OSPF

- ✧ Fait pour IP
- ✧ Plus de documentation

Routage inter domaine (1)

- ✦ Routage entre AS ou domaine
→ routage inter domaine
- ✦ AS gérés par des administrateurs différents
 - ◆ Pas la même culture
 - ◆ Pas la même organisation
 - ◆ Du matériel totalement différent
→ hétérogénéité entre AS
- ✦ Les routeurs du domaine sont responsables du
« forwarding » : le routage

Routage inter domaine (2)

✧ L'internet est avant tout découpé suivant l'adressage

- ✧ Un réseau est représenté par une plage d'adresse
- ✧ Un réseau de réseaux est représenté par un ensemble de plages d'adresses

✧ A un domaine de routage autonome correspond un ensemble de plages d'adresses

- ✧ *Un AS annonce les plages d'adresses qu'il contrôle*
- ✧ Il permet l'accessibilité et l'accès au reste de l'internet via ses interconnexions.

✧ Chaque AS remplit deux tâches:

- ✧ Assure le routage pour les réseaux qu'il héberge
- ✧ Assure le routage pour d'autres AS connectés

✧ Chaque AS à un numéro entre 1 et 65535 (privé entre 64212 et 65535)

Routage inter domaine (3)

✦ Routage entre AS est distribué

- ✦ Chaque administrateur contrôle ses décisions, mais ne peut pas contrôler le voisin

✦ But :

- ✦ A la couche IP, toutes les machines sont censées être joignables entre elles
- ✦ Les réseaux sont tous indirectement connectés les uns aux autres
- ✦ Des réseaux ne servent que de transit entre deux réseaux
- ✦ Les réseaux aux extrémités ne participent qu'aux communication qui les concernent.

Routage inter domaine (4)

✧ Interconnexion entre AS

- ✧ Lien privé.
- ✧ Ferme d'interconnexion (IXP : Internet Exchange Point)

Les routeurs des différents AS sont tous dans un bâtiment.

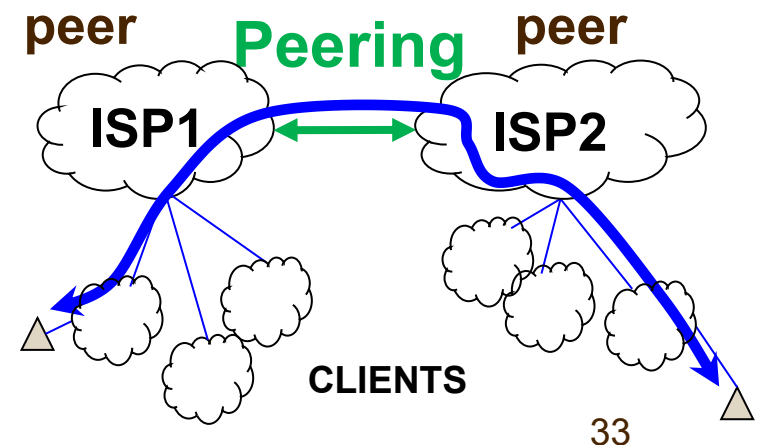
❑ Contrats d'interconnexion

- ❑ La vente de capacités et d'accessibilité

- ❑ Client/ fournisseur

- ❑ Le partage gratuit des communications entre clients

- ❑ **Peering**



Routage inter domaine (5)

✦ Les accords d'interconnexion

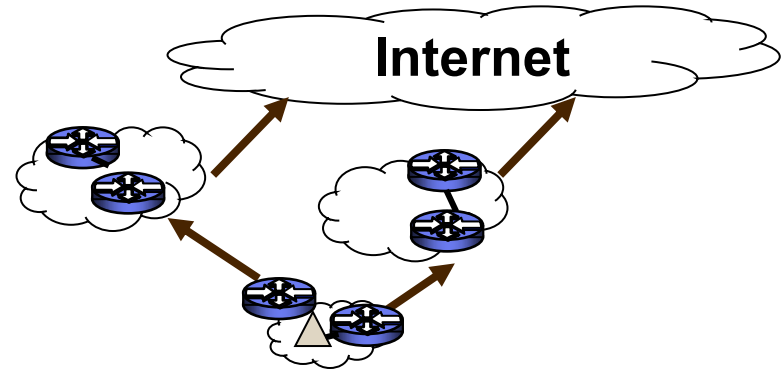
- ✦ **Les clients** payent pour recevoir et envoyer du trafic à **un fournisseur**
 - Un client a souvent plusieurs fournisseurs
(ex : un pour accéder à internet, et d'autres pour être accessible)
 - L'internet est non régulé et les négociations d'accords tournent à l'avantage des plus gros
- ✦ Les relations de Peering sont mises en place pour des raisons de performances
 - Les délais sont réduits
 - Permet d'avoir de la connectivité gratuite

Routage inter-domaine (6)

✧ Vocabulaire

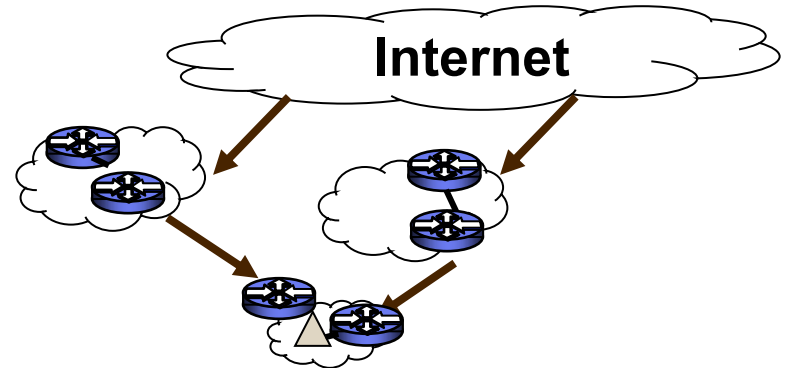
✧ Chaque réseau veut accéder au reste de l'Internet

- **Connectivité**



✧ Chaque réseau veut pouvoir être accessible par n'importe quelle machine dans l'internet

- **Accessibilité**



✧ Accessibilité + connectivité = « internet Connectivity »

Les Tiers (1)

Tier 1 : environs 14

- Compose le cœur d'internet
- Transmission des paquets gratuitement entre eux
- Raccordé à des plus petits qui payent, mais permet d'avoir une connectivité presque totale

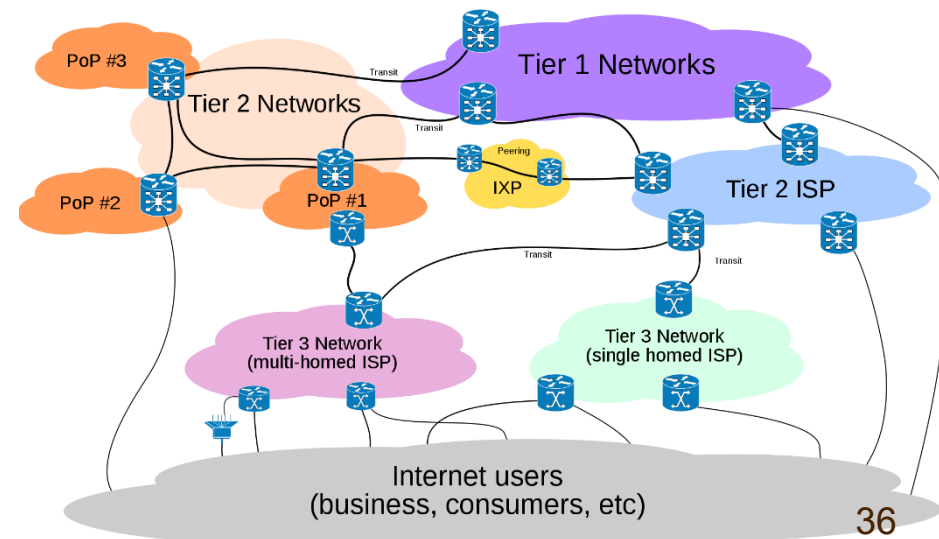
Tier 2 :

- Plus régional, ont des accords de peering entre eux
- Sont raccordés avec tiers 1

Tier 3 :

- Paye à des tiers 1 ou 2 le droit d'accéder à internet

Hyperscale network !!



Les Tiers (2)

Name	Headquarters	AS number	CAIDA AS rank ^[41]	Fiber route (km)	Peering policy
AT&T ^[42]	 United States	7018	29	660,000 ^[43]	AT&T Peering policy
Deutsche Telekom Global Carrier ^[44]	 Germany	3320	17	250,000 ^[45]	DTAG Peering Details
GTT Communications	 United States	3257	4	232,934 ^{[46][47]}	GTT Peering Policy
Liberty Global ^{[48][49]}	 Netherlands ^[50]	6830	26	800,000 ^[51]	Peering Principles
Lumen Technologies (formerly CenturyLink, formerly Level 3) ^{[52][53][54]}	 United States	3356	1	885,139 ^{[55][56]}	Lumen Peering Policy
NTT Communications (formerly Verio) ^[57]	 Japan	2914	5	?	Global Peering Policy
Orange ^[58]	 France	5511	11	495,000 ^[59]	OTI peering policy
PCCW Global	 Hong Kong	3491	10	?	Peering policy
Tata Communications (prev. VSNL prev. Teleglobe) ^[60]	 India	6453	6	700,000 ^[61]	Peering Policy
Telecom Italia Sparkle (Seabone) ^[62]	 Italy	6762	8	560,000	Peering Policy
Arelion (formerly Telia Carrier) ^[63]	 Sweden	1299	2	70,000 ^[64]	Arelion's IP Network Peering Policy 
Telxius (Subsidiary of Telefónica) ^[65]	 Spain	12956	15	65,000 ^[66]	Peering Policy 
Verizon Enterprise Solutions (formerly UUNET) ^[71]	 United States	701	25	805,000 ^[72]	Verizon UUNET Peering policy 701, 702, 703
Zayo Group (formerly AboveNet) ^[73]	 United States	6461	9	196,339 ^[74]	Zayo Peering Policy

Routage - rappel

✧ Routage IGP

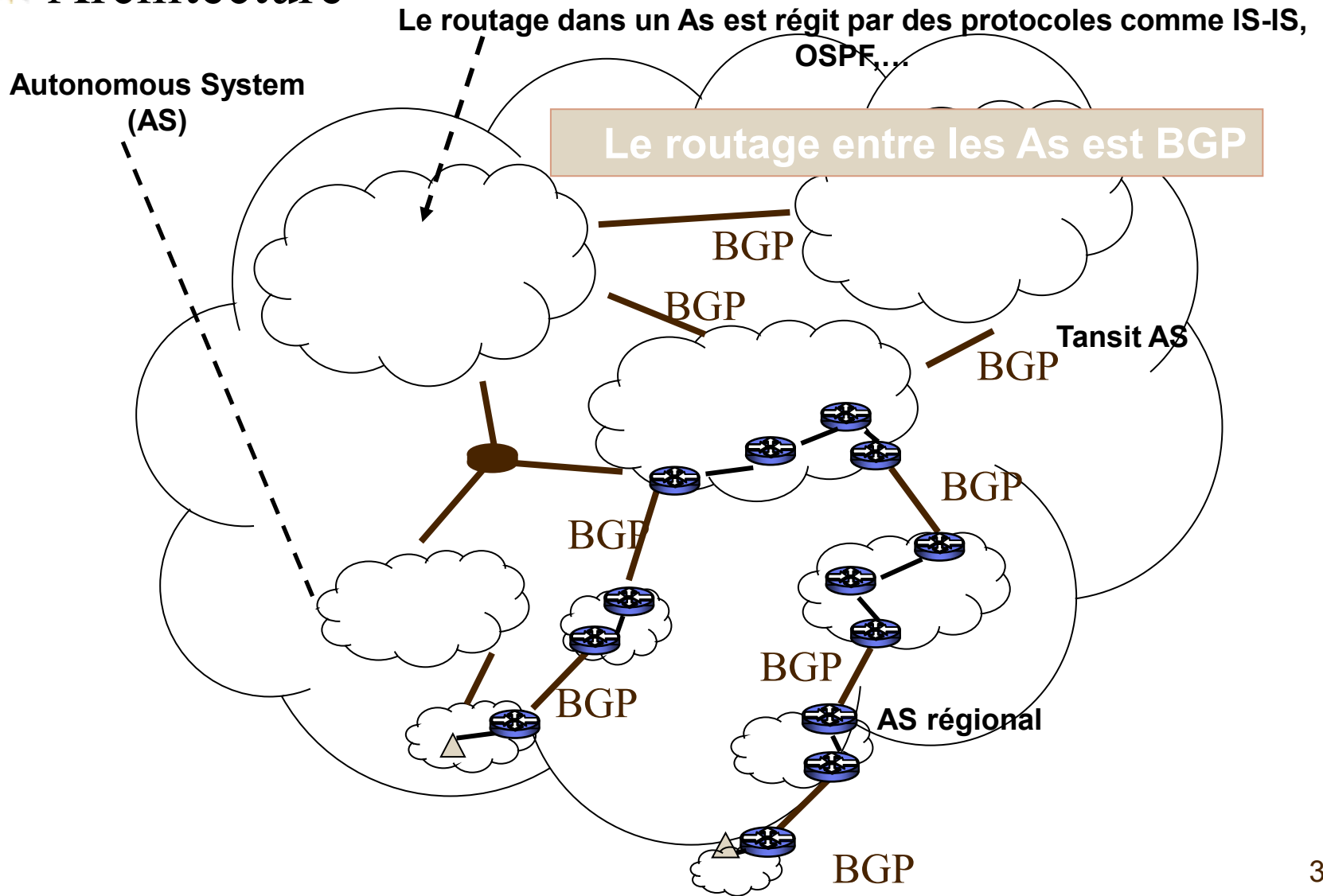
- ✧ la main mise administrative est totale
- ✧ la taille en nombre de plages d'adresses et nombre de routeurs est modeste
- ✧ les besoins sont axés sur la performance, la sécurité, la flexibilité

✧ Routage EGB

- ✧ **Les considérations économiques jouent un rôle moteur**
- ✧ Les performances dépendent des autres domaines
- ✧ Routage avec POLITIQUE (on choisit la route)

BGP (1)

✦ Architecture



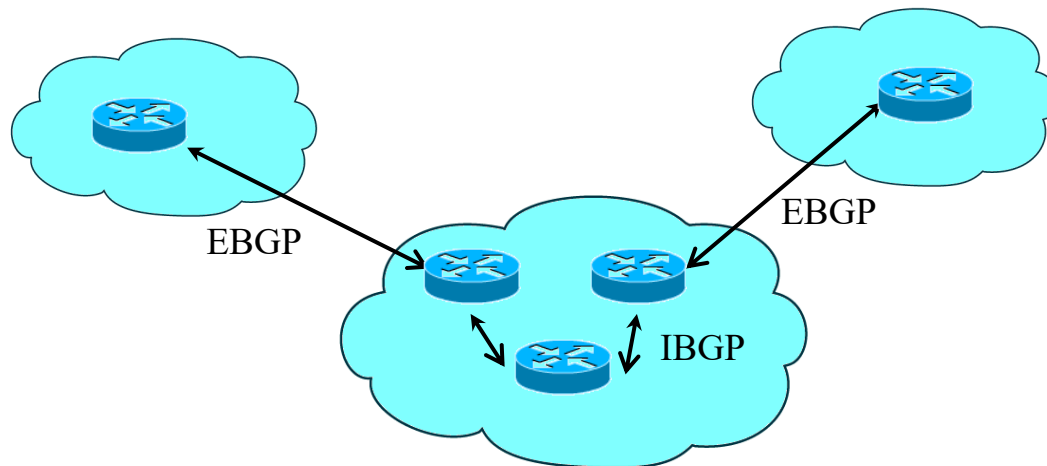
BGP (2)

★ Stub AS ou single homed:

- ◆ Un AS qui n'a qu'une connexion logique vers un unique autre AS
- ◆ Naturellement, cet AS ne supporte que du trafic local
- ◆ Utilisation d'une route par défaut pour sortir de l'AS
- ◆ BGP n'est pas obligatoire... ou alors utilisation d'un numéro privé

★ Multihomed AS:

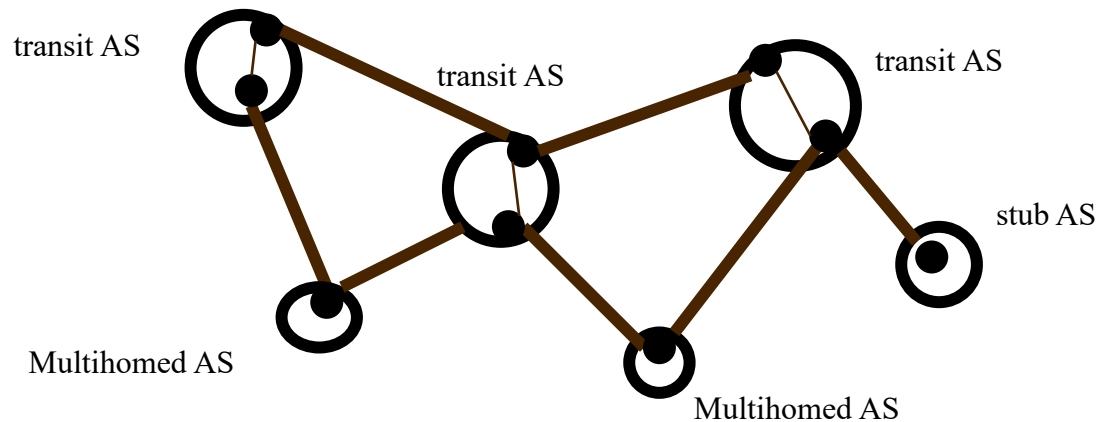
- ◆ Un AS qui a des connexions avec plusieurs AS, en général 2
- ◆ Cet AS supporte que du trafic local si Non-Transit, autrement Transit AS



BGP (3)

✧ Transit AS:

- ✧ Un AS qui a des connexions vers de multiples AS
- ✧ Sous certaines conditions relatives à la politique de routage (et d'interconnexion), l'AS supporte du trafic local et du trafic de transit



BGP (4)

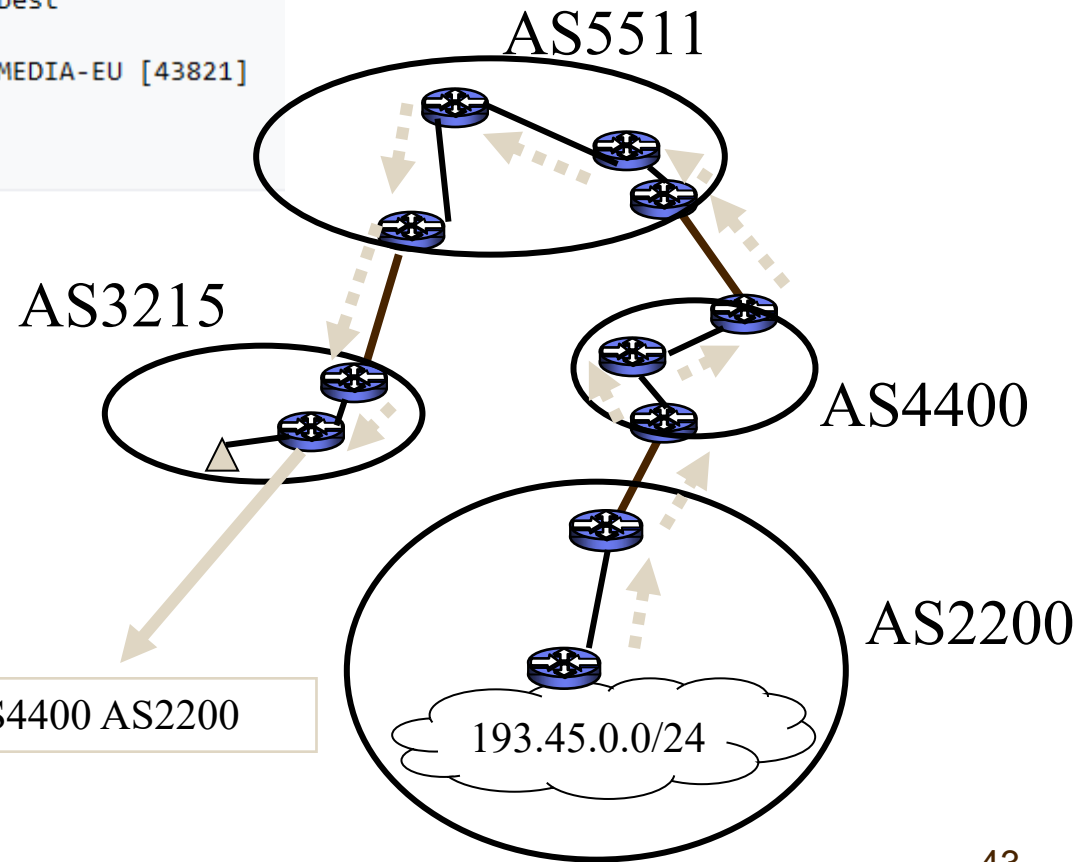
✧ Protocole à vecteur de chemin

- ◆ Informations de routage échangées concernent les réseaux destinations
- ◆ Fonctionne à partir d'annonces et de révocations de routes
- ◆ Les routes sont des chemins d'AS muni d'attributs globaux ou locaux
- ◆ Une route indique la liste d'ASs qu'il faut traverser pour arriver à destination (évite les boucles !!!)
- ◆ **BGP est la glue de l'internet**
 - La concertation du routage est propagée entre les AS uniquement via BGP ou manuellement
 - Mais les tables de routage explosent
 - Et une erreur peut impacter l'Internet tout entier (cf wikipedia)

BGP est inévitable actuellement

BGP (4 bis)

```
> show ip bgp 91.198.174.2
BGP routing table entry for 91.198.174.0/24, version 151383419
Paths: (2 available, best #1, table Default-IP-Routing-Table)
  Advertised to non peer-group peers:
    82.115.128.225
  TDC [3292] WIKIMEDIA-EU [43821]
    82.115.128.18 from 82.115.128.18 (62.95.30.10)
      Origin IGP, localpref 100, valid, external, best
      Community: 215745616 215746418
  IPO-EU [12552] PORT80-GLOBALTRANSIT [16150] WIKIMEDIA-EU [43821]
    82.115.128.26 from 82.115.128.26 (62.109.44.1)
      Origin IGP, localpref 100, valid, external
```



BGP (5)

✧ Objectif

- ✧ Échange d'informations de routage entre AS
- ✧ Garantir le choix d'une route sans boucle

✧ Caractéristiques

- ✧ Seul BGP4 supporte le CIDR et agrégation de routes
- ✧ BGP n'utilise pas de métrique
- ✧ Fait les décisions selon des politiques de routage
- ✧ N'indique pas les détails internes des ASs
- ✧ Ne représente qu'un arbre d'ASs

BGP (6)

-
- ✦ Les routeurs annoncent les réseaux qu'ils routent
 - ◆ Déclaration de la plage d'adresse
 - ◆ Un routeur qui annonce un réseau devient responsable pour l'acheminement vis-à-vis du récepteur de la route
 - ◆ La propagation de l'annonce via les routeurs BGP forme petit à petit un chemin
 - ◆ Les routes sont des routes d'AS munies d'attribut
 - Les attributs concernent la route de façon
 - ◆ Locale (seulement propagé entre routeurs de l'AS)
 - ◆ Globale (entre les AS)

Quelques problèmes (1)

✦ Attention :

- ✦ Par défaut, **les routes plus spécifiques sont privilégiées** par rapport aux autres routes
 - Ex : 192.10.1.0/24 est prioritaire par rapport à 192.10.0.0/16

✦ Piratage (volontaire ou non)

- Cas du pakistan et Youtube (en 2008-> IP Hijacking)
- AS55410 (Vodafone) en avril 2021
- Telekom malaysia en juin 2015
- Piratage janvier 2024 sur Orange espagne

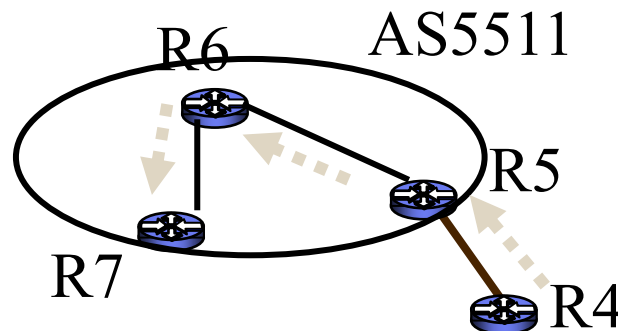
✦ Erreurs

- ✦ Tests sur des attributs en direct (2009, 2010)
- ✦ Annonce erronée pour la destruction des routes (Facebook en 2021)

BGP (7)

✧ Remarque:

- ✧ Un routeur doit connaître l'adresse du prochain bond pour envoyer un paquet pour qu'il puisse de bond en bond rejoindre la destination désirée
- ✧ **MAIS le next hop BGP ne correspond pas forcément à la machine suivante :**
 - Il correspond au point de sortie de l'AS
 - Routage en interne des AS vient d'un protocole IGP pour faire correspondre l'entrée et la sortie



BGP (8)

✧ 2 routeurs BGP « voisin » établissent une session entre eux

- ◆ Dialogue au-dessus de TCP, port 179
- ◆ Un routeur va avoir autant de sessions que de voisins.

✧ 6 états possibles (automate fini BGP):

- ◆ **IDLE** : routeur vient d'être initialisé
- ◆ **CONNECT** : volonté d'envoyer le message OPEN, début connexion TCP (Envoie SYN)
- ◆ **ACTIVE** : passe dans cet état si attente de TCP/SYN trop long après SYN, retour à IDLE si rien ne se passe
- ◆ **OPENSENT** : connexion TCP OK, envoie OPEN, et attend un OPEN, passe ensuite dans Openconfirm si OK, sinon IDLE
- ◆ **OPENCONFIRM** : envoie KEEPALIVE
- ◆ **ESTABLISHED** : réception KEEPALIVE → Session OK

BGP (9)

✦ BGP : 2 protocoles

◆ eBGP : external BGP (entre AS)

- Propagation des routes en dehors de l'AS en utilisant la bonne politique de routage

◆ iBGP : internal BGP (dans l'AS)

✦ Règle :

◆ Les routeurs d'un AS ne propagent pas les messages iBGP entre voisins iBGP

- Les messages eBGP sont propagés à tous les voisins iBGP
- Les messages iBGP ne sont jamais propagés

Utilisation possible d'un full-mesh iBGP ou de route reflector

BGP (10)

✦ Types de messages

✦ Après connexion TCP, 4 types:

- OPEN
 - ✦ Permet l'identification des voisins
 - ✦ N°AS, identifiant, etc...
- KEEPALIVE
 - ✦ Envoi entre 60 et 180 secondes.
- UPDATE
 - ✦ Mise à jour des tables de routage
- NOTIFICATION
 - ✦ Lorsqu'une erreur est détectée
 - ✦ Code d'erreur (8 b) , puis un sous-code (8 b) puis champ de données

BGP (11)

✧ Message UPDATE

- ✧ Une ou plusieurs annonces ou révocations dans **Update**
 - Annonce = route d'AS + réseau destination + attribut
 - Attribut codé en TLV (Type Longueur Valeur)
- ✧ 4 types d'attribut
 - Well-Know Mandatory (WM)
 - ✧ Présent dans les messages Update et implémenté par tous les routeurs
 - Well-Know Discretionary(WD)
 - ✧ Optionnel dans les messages Update et implémenté par tous les routeurs
 - Optional Transitive (OT)
 - ✧ Optionnel dans les messages Update et optionnel dans les routeurs
 - Optional Nontransitive (ON)
 - ✧ Rare dans les messages et optionnel dans les routeurs

BGP (12)

✦ Quelques attributs

Attribute	EBGP	IBGP	Description
Origin	WM	WM	Origine de la route
As_path	WM	WM	Liste des AS traversés
Next_hop	WM	WM	@IP du voisin eBGP
Multi_exit_disc (MED)	ON	ON	Indice de préférence de route externe
Local_pref	Non autorisé	WD	Indice de préférence de route interne
Aggregator	OT	OT	Id du routeur qui réalise l'agrégation
Community	OT	OT	Marquage de route

BGP (13)

✧ Choix de la route

- ✧ existe un processus de décision qui permet de choisir la meilleure route pour chaque réseau destination
- ✧ Choix est basé sur 10 critères (Attributs)
- ✧ les routeurs ne propagent qu'une seule « best route » par destination

Routage BGP (1)

✦ Chaque constructeur implémente le choix d'une façon un peu différente

1. **Weight** le plus fort

attribut local, n'existe que sur les routeurs Cisco

1. Local_Pref maximal

préférence de la route à l'intérieur de l'AS

2. Self-originated

préférence des réseaux dont l'origine est le routeur

3. AS_Path le plus court

4. Préférence en fonction de l'Origin de l'information

5. MED les plus petits

6. Préférence des routes eBGP, plutôt que iBGP

7. Préférer les routes avec métrique IGP minimal vers le Next_hop

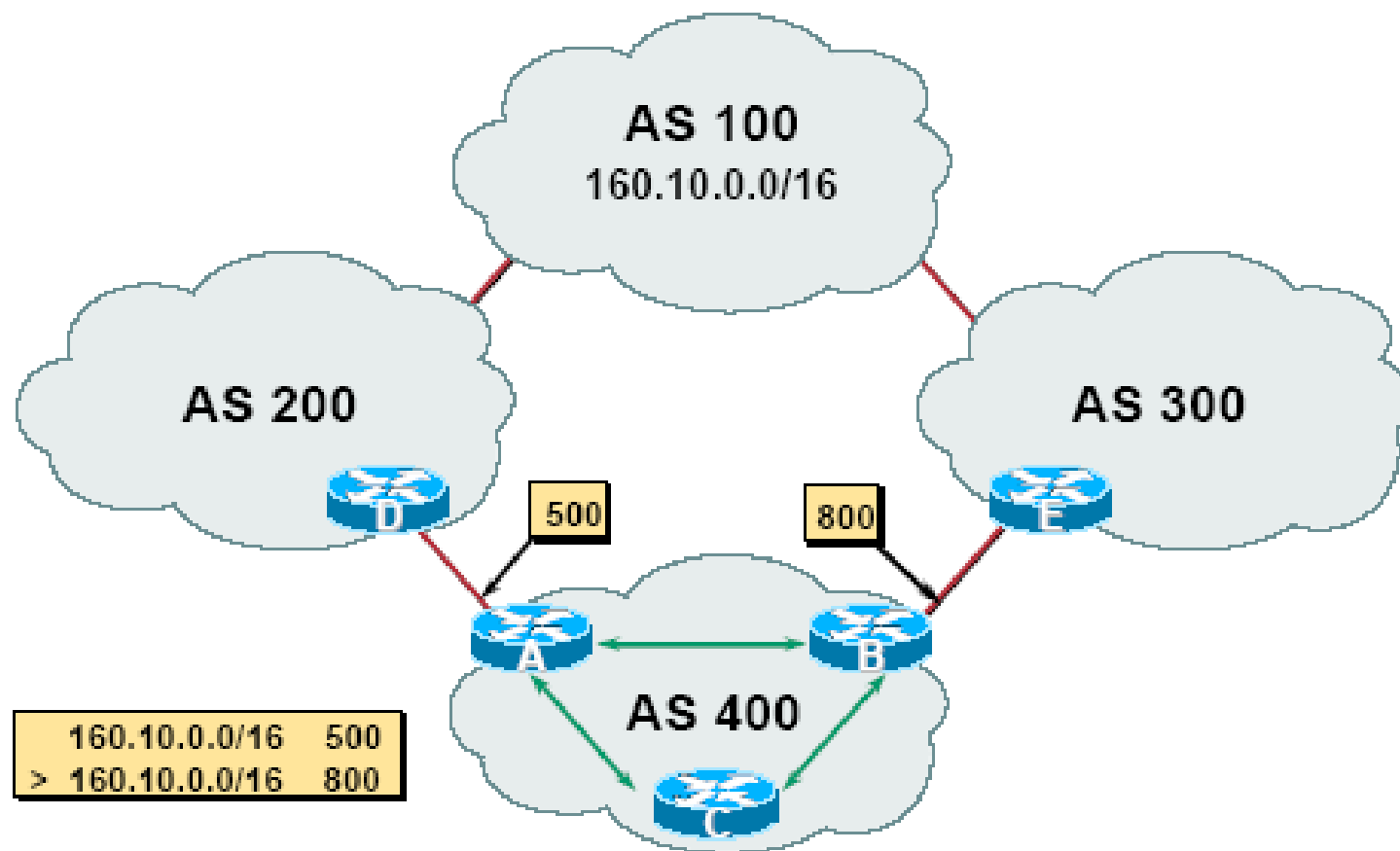
8. Préférer les routes les plus stables

9. Préférer la route avec la plus petite adresse IP du next_hop (router id)

Local-pref

- ✧ Attribut très utilisé pour préférer un AS de sortie
 - ✧ Indique un degré de préférence
 - ✧ Plus l'attribut est grand pour une route, plus cette route sera préférée
- ✧ Cet attribut est seulement propagé à l'intérieur de l'AS
 - ✧ Il indique donc une préférence globale pour tout l'AS
 - ✧ Si l'attribut est propagé entre AS, il doit être supprimé
- ✧ L'attribut est souvent utilisé pour prendre en compte des considérations économiques
 - ✧ Préférer les clients aux peers et les peers aux fournisseurs
- ✧ La valeur par défaut est 100

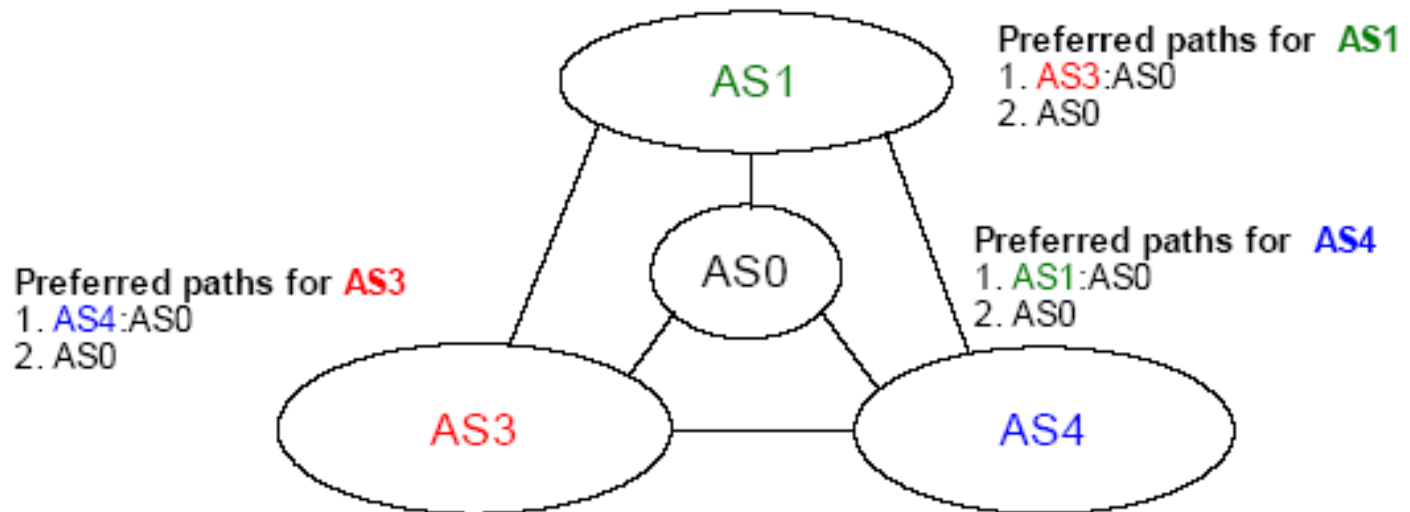
Local-pref (bis)



Local-pref (ter)

✦ Quelquefois, la convergence n'est pas assurée

- ◆ On voit même apparaître des boucles



Routage BGP (2)

✧ "Chemin d'AS"

- ✧ De droite à gauche
 - les AS successivement parcourus par l'annonce de la route
- ✧ De gauche à droite
 - Les AS entre le routeur local et le routeur qui a donnée origine à la route
- ✧ Ne contient pas l'AS du routeur local

✧ Utilisation/intérêt :

- ✧ Pas de boucles au niveau AS. Tout routeur recevant une route avec son AS dans l'attribut AS_PATH supprime cette route!
- ✧ Cela permet d'effectuer des décisions politiques en fonction de certains AS présents ou non dans l'AS_PATH
- ✧ Une route avec un AS_PATH de plus petite longueur sera préférée à une route avec un AS_PATH de taille plus grande

BGP configuration

Pour commencer à configurer le processus BGP, il faut utiliser la commande suivante :

Router (config)# router bgp AS-number

Remarque : Un routeur ne peut faire tourner qu'un processus BGP, donc un routeur ne peut appartenir qu'à un AS.

Router (config –router)# network network-number [mask network-mask]

- ☐ Pour déclarer un réseau
- ☐ Attention, les réseaux doivent être présents dans la table de routage pour être distribués (contraire à RIP ou OSPF)

Router (config-router)# neighbor ip-address remote-as AS-number

- ☐ Permet d'établir une relation de voisinage
- ☐ AS-number permet de mettre en place de E-bgp ou I-bgp

Conclusion

RIP	OSPF	BGP
Convergence lente	Convergence rapide	Annnonce des routes
Plus court chemin sur le nombre de saut	Plus court chemin sur différentes métriques (saut, bande passante, latence,..)	Chemin « économique »
Mémoire limitée	Mémoire normale	Très grosse mémoire
Débit assez faible	Débit assez faible	Très gros débit
Une centaine de lignes maximum	Une centaine de lignes maximum	> 150000 lignes (tous les préfixes Internet)