

# Le routage dans Internet

**Thomas Vantroys**

`thomas.vantroys@univ-lille1.fr`

Polytech'Lille  
Université Lille 1

**IMA 4**  
2018 - 2019



- 1 Introductions
  - Principes
  
- 2 Les algorithmes de routages
  - Présentation
  - Les protocoles à vecteur de distance
  - Les protocoles à états de liaison
  
- 3 Algorithmes intra domaines
  - RIP
  - OSPF
  
- 4 Protocoles inter-domaines
  - EGP
  - BGP

# Principes

- Deux modes de remise des paquets
  - remise direct
    - les deux machines sont sur le même réseau
  - remise indirect
    - il es nécessaire de passer par au moins un routeur pour délivrer le paquet à destination
- Comment apprendre de nouvelles routes ?

# Comment choisir le routeur suivant ?

Les algorithmes de routage IP utilisent une *table de routage*

- elle contient des informations relatives aux destinations et à la façon de les atteindre
- le schéma d'adressage IP permet de ne conserver que des adresses réseaux
- la table contient des paires
  - numéro de réseau, routeur de saut suivant
  - seuls les routeurs accessibles directement sont dans la table de routage

# Les types de routages

- Routage statique
  - réalisé par un administrateur
  - pour définir des routes par défaut
  - pour forcer des routes vers des destinations particulières (test/debug)
- Routage dynamique
  - réalisé par un protocole qui permet aux routeurs d'échanger de l'information pour prendre des décisions de routage

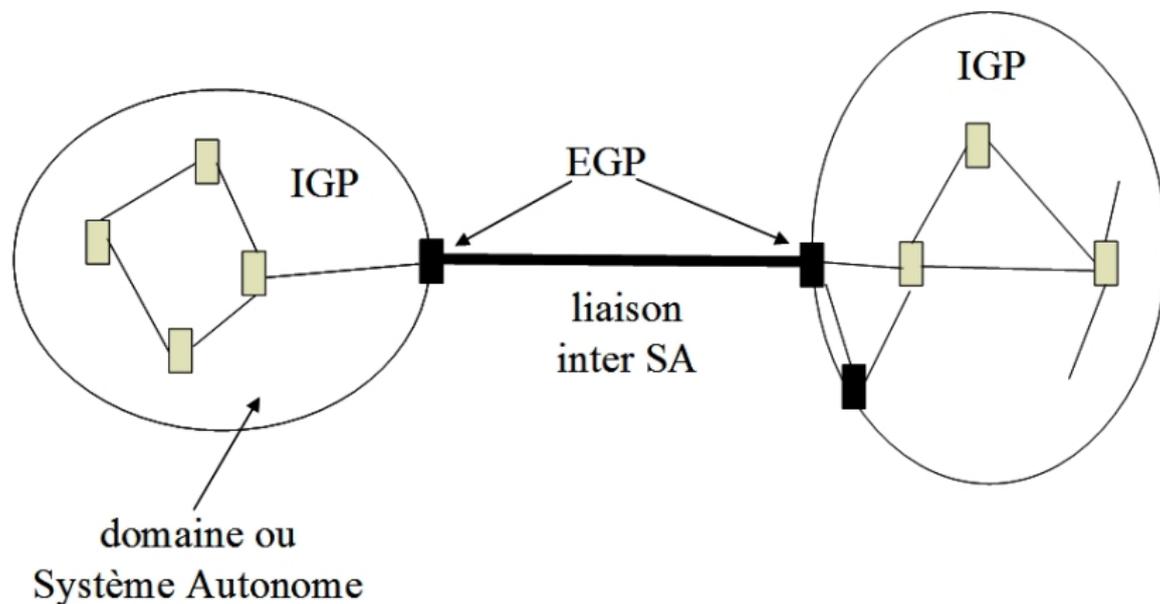
# Les protocoles de routages

On distingue :

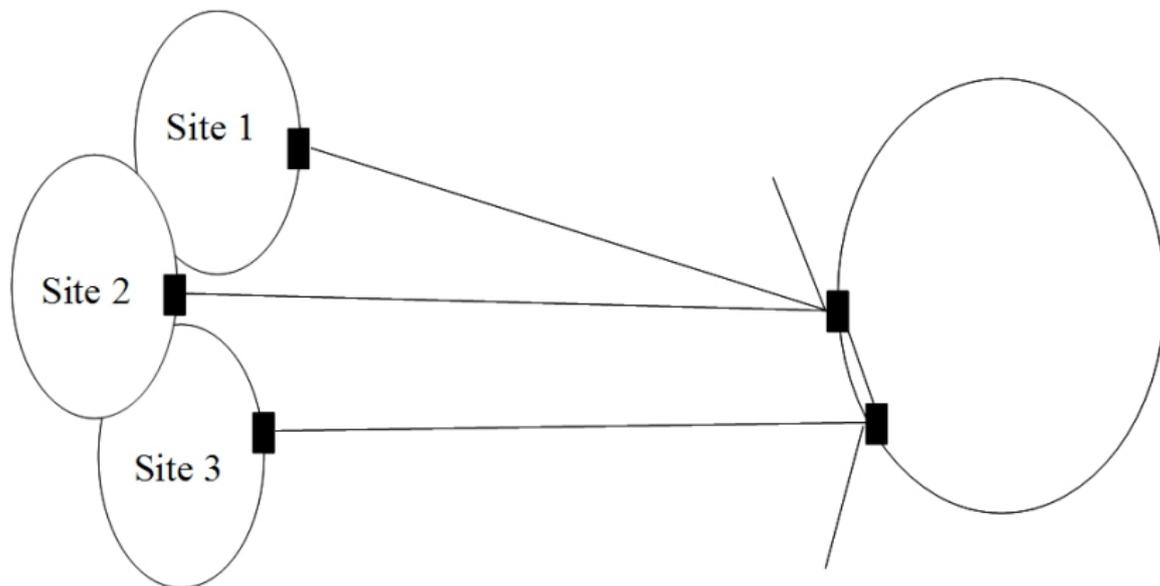
- les protocoles intra-domaine (IGP : Interior Gateway Protocol)
- les protocoles inter-domaines (EGP : Exterior Gateway Protocol)

Un domaine dépend d'une autorité d'administration unique

# Système autonome



# Système autonome



# Les algorithmes de routages

On distingue deux types d'algorithmes de routage :

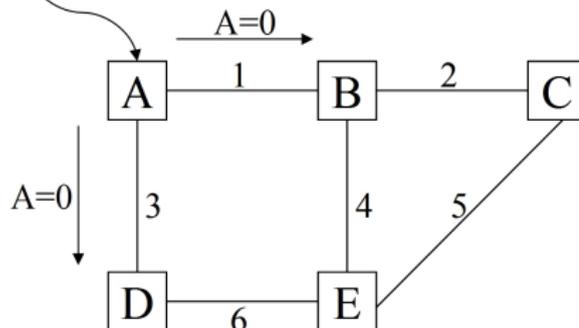
- les algorithmes dit à **vecteurs de distance**
  - basé sur les algorithmes de calcul du plus court chemin dans un graphe de Bellman (1957) et Ford/Fulkerson (1962) pour la version distribuée
- les algorithmes dit à **état de liaison**
  - basé sur l'algorithme du chemin le plus court d'abord (*shortest path first*) de Djisktra (1959)

# Principes des protocoles à vecteur de distance

- chaque nœud maintient une table de routage
- au départ chaque table ne contient que l'entrée qui correspond au nœud local
- les tables sont mises à jour par des diffusions régulières d'informations entre les différents nœuds

# Initialisation

De A à	Liaison	Coût
A	locale	0

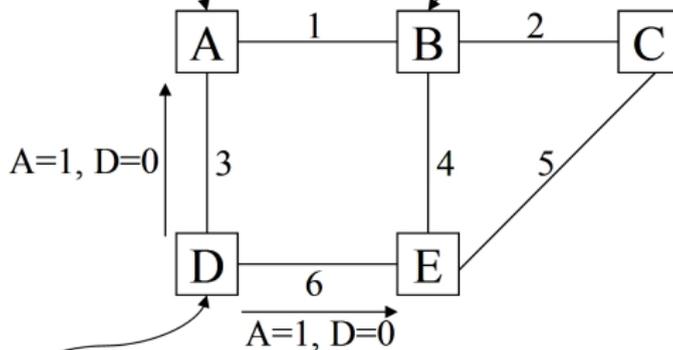


De E à	Liaison	Coût
E	locale	0

# Évolution

De A à	Liaison	Coût
A	locale	0

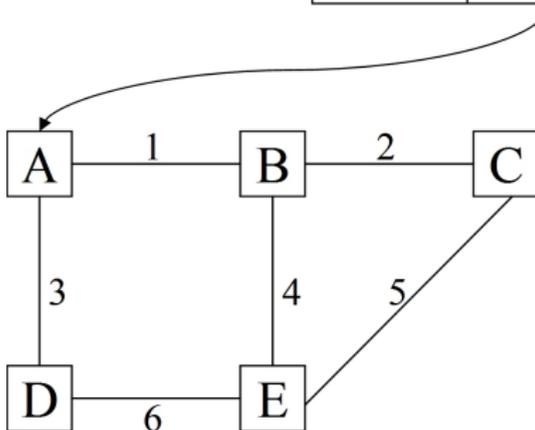
De B à	Liaison	Coût
B	locale	0
A	1	1



De D à	Liaison	Coût
D	locale	0
A	3	1

# Convergence

De A à	Liaison	Coût
A	locale	0
B	1	1
D	3	1
C	1	2
E	1	2



# L'effet de rebond

De	Liaison	Coût
A à C	1	2
B à C	2	1
C à C	locale	0
D à C	6	2
E à C	5	1

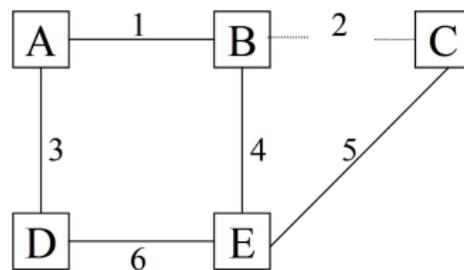
Rupture de la liaison 2

De	Liaison	Coût
A à C	1	2
<b>B à C</b>	<b>2</b>	<b><math>\infty</math></b>
C à C	locale	0
D à C	6	2
E à C	5	1

A diffuse son vecteur

A=0, B=1, D=1,  
C=2, E=2

De	Liaison	Coût
A à C	1	2
<b>B à C</b>	<b>1</b>	<b>3</b>
C à C	locale	0
D à C	6	2
E à C	5	1



# Le comptage à l'infini

La liaison 1 est coupée et les tables ont convergé. La liaison 6 tombe.

De D à	Liaison	Coût
D	locale	0
A	3	1
B	6	$\infty$
E	6	$\infty$
C	6	$\infty$

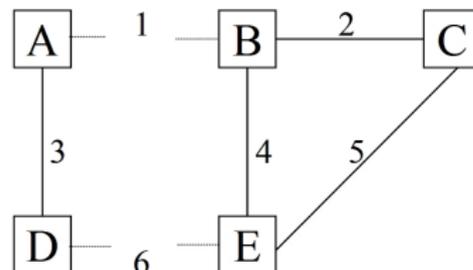
A diffuse son vecteur

A=0, B=3,  
D=1, C=3, E=2

De D à	Liaison	Coût
D	locale	0
A	3	1
B	3	4
E	3	3
C	3	4

D diffuse son vecteur

A=1, B=4,  
D=0, C=4, E=3



# L'horizon partagé

- On ne donne pas d'informations à un nœud X sur les destinations qu'on atteint via X
- Deux variantes :
  - on ne mentionne pas ces destinations
  - on les mentionne avec des distances infinies

# Mises à jour déclenchées

- en fonctionnement normal
  - les vecteurs de distances sont réemis périodiquement
  - chaque ligne de la table à un temporisateur. Si celui-ci tombe à zéro, la distance est marquée infinie
- en cas de défaillance d'un lien
  - on propage l'information le plus rapidement possible

# Bilan

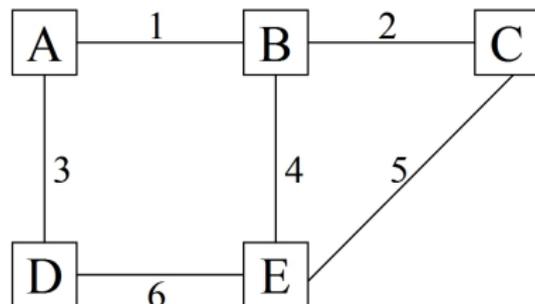
- Avantages
  - simple à mettre en œuvre
  - messages simples à construire
  - peu de gestion pour les tables de routage
- Inconvénients
  - risque de boucles dans le réseau
  - la convergence peut être lente
  - inadapté aux ruptures de liaisons

# Principes des protocoles à état de liaison

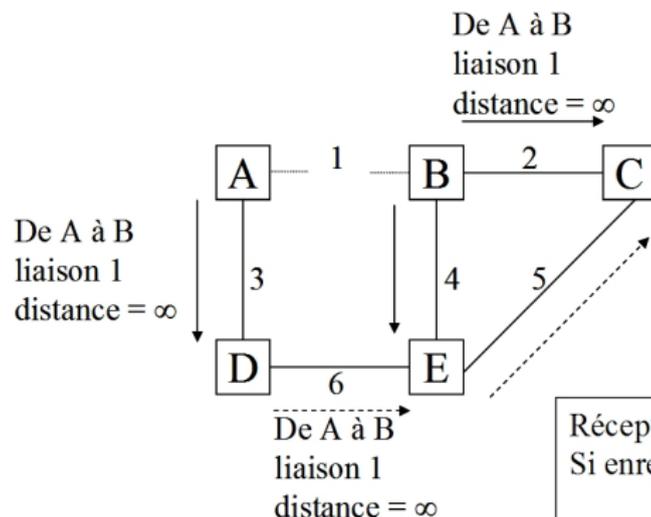
- les nœuds possèdent une carte identique du réseau
- les mises à jours sont diffusées rapidement
- chacun effectue le même calcul avec les mêmes données :  
donc les décisions sont cohérentes
- utilisation de l'algorithme *shortest path first* (SPF)

# La base de données

De	à	Liaison	Coût
A	B	1	1
A	D	3	1
B	A	1	1
B	C	2	1
B	E	4	1
C	B	2	1
C	E	5	1
D	A	3	1
D	E	6	1
E	B	4	1
E	C	5	1
E	D	6	1



# Le protocole d'inondation



Réception d'un message d'inondation  
 Si enregistrement non présent  
     ajouter enregistrement  
     rediffuser le message  
 Sinon si date < date enregistrement  
     remplacer enregistrement  
     rediffuser le message

# Algorithme Shortest Path First

- L'algorithme SPF calcule le chemin le plus court entre un nœud source et tous les autres nœuds du réseau
- Il est nécessaire que tous les nœuds utilisent exactement les mêmes métriques

# Support de métriques multiples

- Il est possible de définir des métriques multiples
  - IP : plus haut débit, plus bas délai, plus bas coût, meilleure fiabilité
- Il est alors nécessaire de :
  - fixer et transmettre plusieurs métriques par liaison
  - calculer autant de tables que de métriques
  - préciser la métrique souhaitée dans chaque paquet

# Établissement de voisinage

- Si le réseau est scindé, les bases de données des éléments connexes évoluent de manière autonome
- Lors de la reconnexion, il faut remettre les bases de données en phase :
  - les nœuds s'échangent une description de leurs tables
  - chacun demande les entrées "intéressantes"
  - les valeurs mises à jour sont ensuite transmises par inondation

# Bilan

Avantages :

- convergence rapide sans boucles
- métriques précises et au besoin multiples
- possibilité de chemins multiples
- traitement séparé des routes "externes"

# Algorithmes intra domaines

# RIP : Routing Information Protocol

- RIP est un protocole à vecteur de distance
- La table ne contient que les meilleures routes exprimées en nombre de sauts
- Le nombre de sauts est fixé à 15 (infini = 16)
- RIP v2
  - routage par sous-réseaux à l'extérieur du réseau
  - possibilité de définir plusieurs domaines sur le même médium
  - "sécurisation" par authentification

# Messages

- Utilise le protocole UDP (port 520)
- Deux types de messages :
  - requête : demande de route à un voisin
    - au démarrage requête pour la route par défaut (0.0.0.0) et une métrique infinie
    - demande de routes particulières seulement pour les tests
  - réponse : annonce de route
    - toutes les 30 secondes sauf mises à jour déclenchées
    - les mises à jour déclenchées sont séparées par un délai aléatoire de 1 à 5 secondes

# Table de routage

La table de routage contient :

- la destination finale
- la distance vers la destination
- le routeur de saut suivant
- un indicateur de "mise à jour récente"
- des temporisateurs

# Algorithme

```
réception message
pour (chaque entrée)
  si (entrée inconnue)
    alors
      ajouter une entrée (destination, métrique+1, émetteur)
      initialiser un temporisateur
  sinon
    si (métrique > métrique reçue +1)
      alors
        mettre à jour (métrique reçue+1, émetteur)
        réinitialiser temporisateur
    sinon
      si (routeur suivant = émetteur)
        alors
          mettre à jour métrique
          réinitialiser temporisateur
```

# Timers RIP

- Timer pour la mise à jour (*route update timer*) : intervalle entre les mises à jour (généralement 30 secondes)
- Timer pour les routes invalides (*route invalid timer*) : Temps avant qu'une route ne soit déclarée invalide (180 secondes), i.e., le routeur n'a pas reçu d'information sur cette route depuis une durée supérieure à ce timer
- Timer de sauvegarde (*holddown timer*) : Une route rentre dans l'état *holddown* lorsqu'un paquet est reçu indiquant qu'une route est inaccessible. Cet état persiste jusqu'à la mise à jour de la route ou le déclenchement du timer (180 secondes)
- Timer d'effacement (*route flush timer*) : durée entre le moment une route devient invalide et son effacement de la table de routage (240 secondes).

# Configuration des interfaces

- Pour chaque interface valide
  - donner l'adresse et le masque de sous-réseau
  - éventuellement la métrique (fixée à 1 par défaut)
- Choix des interfaces valides (sur lesquelles on va diffuser les informations de routage)
- Pour les médiums sans diffusion :
  - préciser la liste des adresses avec lesquelles on échange des informations
- Indiquer les routes fixes, interdire des routes

# Format des messages RIPv1

## RFC 1058

Commande	Version	Laissé à zéro
Id. de famille d'adresse		Laissé à zéro
Adresse IP		
Laissé à zéro		
Laissé à zéro		
Métrique		

- Commande : REQUEST (1) ou RESPONSE (2)
- Id. de famille : IP (2)

Filter:  Expression... Clear Apply

No.	Time	Source	Destination	Protocol	Info
3051	13.715031	134.206.3.1	255.255.255.255	RIPv1	Response
3052	13.715034	ff:ff:ff:ff:ff:ff	ff:ff:ff:ff:ff:ff	Ethernet II	Broadcast

- Frame 3052 (346 bytes on wire, 346 bytes captured)
  - Ethernet II, Src: Cisco\_a8:30:00 (00:0c:cf:a8:30:00), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
    - Destination: Broadcast (ff:ff:ff:ff:ff:ff)
    - source: Cisco\_a8:30:00 (00:0c:cf:a8:30:00)
    - Type: IP (0x0800)
  - Internet Protocol, Src: 134.206.3.1 (134.206.3.1), Dst: 255.255.255.255 (255.255.255.255)
    - Version: 4
    - Header Length: 20 bytes
    - Differentiated Services Field: 0x00 (DSCP 0x00: default; ECN: 0x00)
    - Total Length: 532
    - Identification: 0x0000 (0)
    - Flags: 0x00
    - Fragment Offset: 0
    - Time to live: 2
    - Protocol: UDP (0x11)
    - Header checksum: 0x2d0b [correct]
    - Source: 134.206.3.1 (134.206.3.1)
    - Destination: 255.255.255.255 (255.255.255.255)
  - User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
    - source port: router (520)
    - destination port: router (520)
    - Length: 512
    - Checksum: 0x3f1c [correct]
  - Routing Information Protocol (RIPv1)
    - Command: Response (2)
    - Version: RIPv1 (1)
      - IP Address: 0.0.0.0, Metric: 2
        - Address Family: IP (2)
        - IP Address: 0.0.0.0 (0.0.0.0)
        - Metric: 2
      - IP Address: 10.0.0.0, Metric: 1
        - Address Family: IP (2)
        - IP Address: 10.0.0.0 (10.0.0.0)
        - Metric: 1
      - IP Address: 172.16.0.0, Metric: 1
        - Address Family: IP (2)
        - IP Address: 172.16.0.0 (172.16.0.0)
        - Metric: 1
      - IP Address: 172.17.0.0, Metric: 1
        - Address Family: IP (2)
        - IP Address: 172.17.0.0 (172.17.0.0)
        - Metric: 1

```

0000 ff ff 02 08 00 00 00 00 00 00 00 00 00 00 00 00
0030 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0060 00 02 00 00 00 00 0a 00 00 00 00 00 00 00 00
0090 00 00 00 00 01 00 02 00 00 ac 15 00 00 00 00
00c0 00 00 00 00 00 00 00 00 00 01 00 02 00 00 ac 11
00f0 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0120 00 00 ac 12 00 00 00 00 00 00 00 00 00 00 00 00
0150 00 01 02 00 00 ac 13 00 00 00 00 00 00 00 00 00
0180 00 00 00 00 01 00 02 00 00 ac 15 00 00 00 00
01b0 00 00 00 00 00 00 00 00 00 01 00 02 00 00 ac 1a
01e0 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0210 00 00 ac 1b 00 00 00 00 00 00 00 00 00 00 00 00
0240 00 02 00 02 00 00 ac 1f 00 00 00 00 00 00 00 00
0270 00 00 00 00 02 00 02 00 00 c9 a8 02 00 00 00
0300 00 00 00 00 00 00 00 00 00 01 00 02 00 00 c0 a8
0330 14 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0360 00 00 c0 a8 05 00 00 00 00 00 00 00 00 00 00 00
0390 00 00 01 00 00 00 c0 a8 00 00 00 00 00 00 00 00
03c0 00 00 00 00 01 00 02 00 00 c0 a8 15 00 00 00
  
```

Routing Information Protocol (rip), 504 bytes | Packets: 5297 Deployed: 5297 Marked: 0

# Format des messages RIPv2

RFC 1388 et RFC 1387

Commande	Version	Domaine de routage
Id. de famille d'adresse		Identifiant de route
Adresse IP		
Masque de sous-réseau		
Prochain relais		
Métrique		

- adresse multicast 224.0.0.9

# OSPF : Open Shortest Path First

- OSPF est un protocole à état de liaison
- Créé par l'IETF pour pallier les défauts de RIP
- Cherche à limiter le nombre de voisins pris en compte :
  - distingue hôtes et routeurs
    - liaison entre routeurs (adresse du routeur)
    - liaison à un réseau terminal (adresse du sous-réseau)
  - traite les réseaux à diffusion et sans diffusion
  - découpe les grands réseaux en zones

# Réseaux à diffusion

Tous les routeurs sont voisins :

- on réduit les voisinages en désignant un routeur
- les autres routeurs établissent le voisinage avec ce routeur désigné :
  - on limite la quantité de messages de synchronisation
  - on limite le nombre d'enregistrements d'états de liaison
- un routeur désigné de backup est choisi pour prendre le relais en cas de défaillance
- les liaisons sur le même réseau ont une métrique nulle

# Réseaux sans diffusion

Les communications sont en point à point :

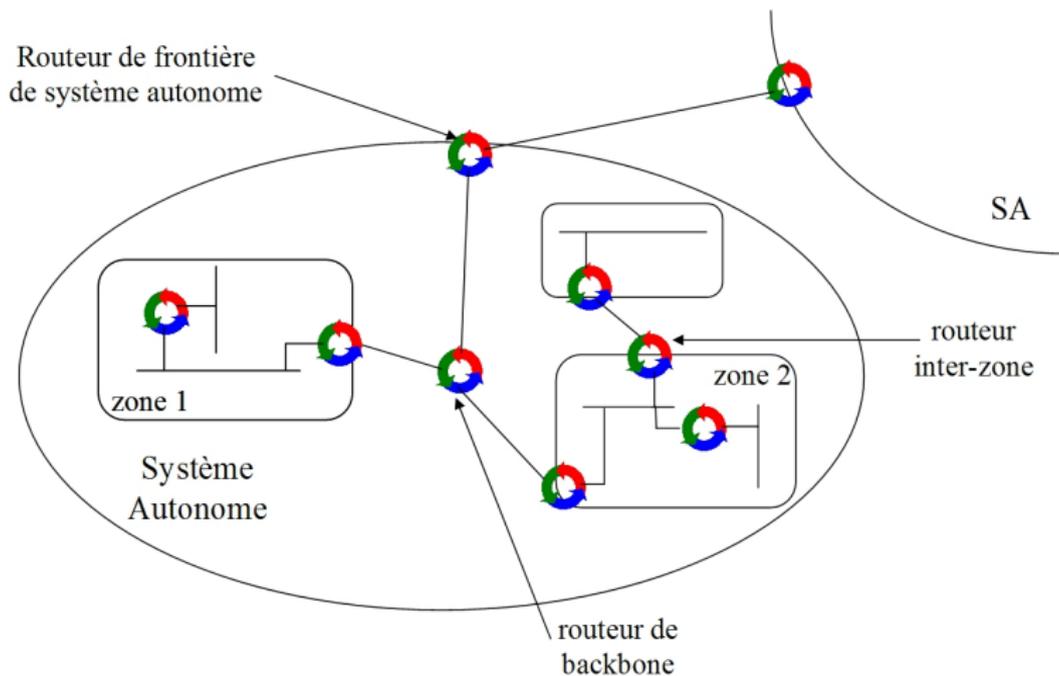
- on choisit là aussi un routeur désigné et un backup
- seuls les circuits virtuels entre les routeurs et le routeur désigné sont utilisés de manière permanente pour le routage

# Le protocole *Hello*

## Objectifs :

- vérifier la connectivité
  - intervalle de hello (réémission)
  - intervalle de mort (perte de connectivité)
- élire un routeur désigné et un backup
  - choix sur la priorité assignée à chaque routeur

# Découpage en zones



# Types d'enregistrements

- Routeur
  - liaisons issues d'un routeurs
- Réseau
  - liaisons d'un routeur désigné avec les autres routeurs du réseau
- Récapitulation de réseaux IP
  - liaisons publiées par un routeur inter-zones concernant les routes internes à la zone
- Récapitulation de routeur externe
  - liaisons publiées par un routeur inter-zone concernant les routes vers les routeurs externes
- Externe
  - liaisons publiées par un routeur externe concernant les routes vers l'extérieur du système autonome

# Format des trames OSPF

Version	Type	Longueur du message
Adresse IP du routeur source		
Identification de la zone		
Total contrôle	Type d'authentification	
Authentification		
Authentification		
Champ données		

# Format du message Hello

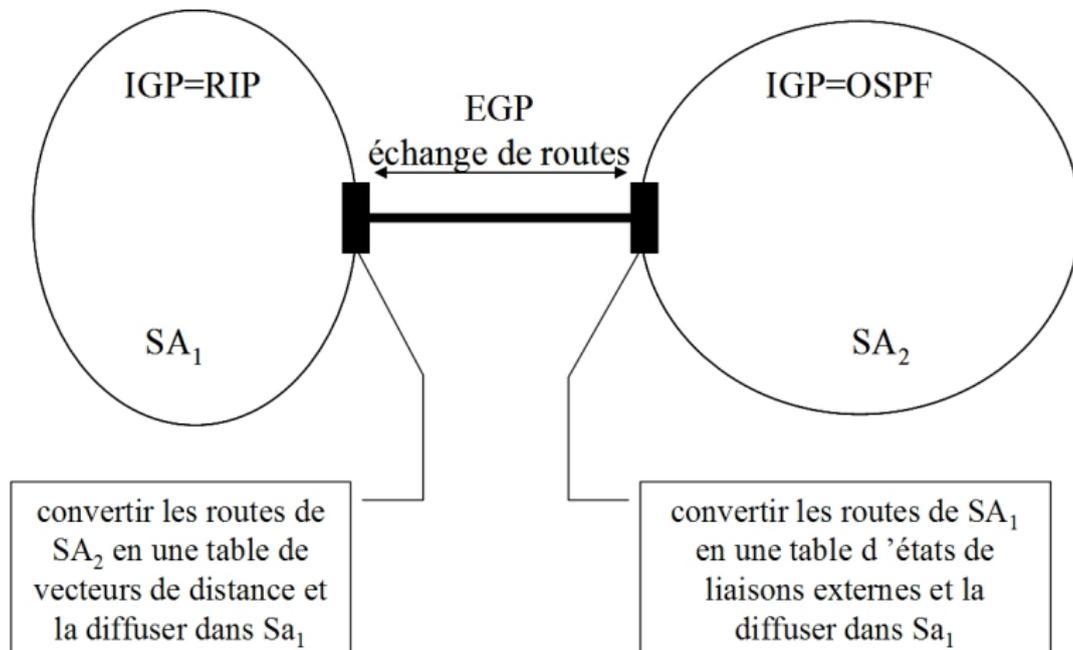
En-tête OSPF Type=1 (Hello)		
Masque de sous-réseau		
Total contrôle	xxxxxxET	Priorité
Intervalle de mort		
Routeur désigné		
Routeur backup		
Adresses IP des routeurs voisins ...		

# Protocoles inter-domaines

# Principes

- L'objectif des EGP est d'échanger des informations de routage entre systèmes autonomes voisins
- Un système autonome :
  - est sous administration unique
  - toutes les parties doivent être connexes
  - il est numéroté sur 16 bits (ex : RENATER : 1717)

# Principes



# EGP : External Gateway Protocol

- EGP permet l'échange d'informations de connectivité entre deux routeurs externes voisins
- Il est composé de trois procédures
  - une procédure d'acquisition de voisin détermine si deux routeurs acceptent de devenir voisins
  - une procédure d'accessibilité de voisin surveille la connectivité
  - une procédure d'accessibilité de réseaux permet l'échange des informations de connectivité

# Acquisition de voisin

- Le protocole d'acquisition est simple :
  - le routeur demandeur envoie une requête d'acquisition de voisin
  - le partenaire envoie une réponse d'acquisition de voisin ou un refus
  - les messages permettent de préciser les intervalles de hello et d'interrogation
- Devenir voisin signifie accepter de relayer une partie du trafic du SA
- Il s'agit d'une décision administrative

# Accessibilité d'un voisin

Permette vérifier que la liaison est opérationnelle :

- Envois de messages "hello" et de réponses "je vous entends"
  - les messages sont numérotés
  - un voisin devient inaccessible si on reçoit moins de  $i$  "JVE" pour les  $n$  derniers "hello"
  - un voisin redevient inaccessible si on reçoit au moins  $j$  "JVE" pour les  $m$  derniers "hello"
  - le protocole peut être asymétrique (un voisin actif et un passif)

# Accessibilité des réseaux

Permet d'échanger les listes des réseaux accessibles via chacun des voisins

- chaque voisin demande une liste à intervalles réguliers
- l'absence de réponse est assimilée à une liste vide

# Interaction EGP/IGP

- Les métriques d'EGP : 0 à 255 (inaccessible)
  - utilisées pour exprimer des routes préférées
    - une liaison de secours aura une métrique plus grande
    - une liaison dédiée entre deux AS doit avoir une métrique plus petite qu'une route plus générale
- Les tables de routage
  - diffusée dans l'IGP si
    - la destination est accessible
    - la route est la "meilleure" route. Pour pouvoir comparer, il faut convertir les distances EGP en métrique IGP
- Convertir les distances EGP en métrique IGP
  - OSPF :
    - utilisation d'états de liaison externes, on peut donc comparer directement les distances annoncées par EGP
  - RIP :
    - le nombre de relais est compris entre 0 et 15
    - métrique = nombre de relais dans le SA + distance convertie en métrique
    - difficile de définir des routes préférées dans ces conditions !

# Les limites d'EGP

- Nécessite une topologie arborescente
- Pas de moyen de contrôler et contenir des informations incorrectes
  - surcroît d'administration
- Pas de routage politique

# BGP: Border Gateway Protocol

- Amélioration d'EGP pour gérer des topologies arbitraires
- Basé sur des "vecteurs de chemins"
  - Chaque mise à jour de routage comprend la liste complète des SA traversés par le chemin
    - permet de vérifier l'absence de boucles facilement
    - permet de transmettre des métriques hétérogènes
    - la taille des messages croît avec celle de l'Internet
- Implémenté sur TCP

# Attributs de chemin

BGP-3 gère les chemins entre SA :

- un chemin est décrit par une liste d'attributs
  - la liste des SA traversés
  - la liste des réseaux accessibles
  - des attributs politiques ou métriques permettant de choisir un "meilleur" chemin
- tous les routeurs externes d'un SA doivent être connexes
  - pour propager les informations BGP dans le SA
  - pour propager uniquement la meilleure route dans l'IGP

# Procédures BGP

- Échange initial
  - La connexion est faite sur le port 179
  - un refus de connexion est identifié par un message de notification
  - une acceptation est identifiée par un message de sonde
- Mises à jour
  - un message de mise à jour se rapporte à un seul chemin
  - le chemin est conservé s'il est plus court que l'ancien
  - dans ce cas, on propage la mise à jour aux voisins
- Sondage
  - permet de vérifier la connectivité
  - le partenaire doit aussi envoyer des sondes dans le délai de garde fixé lors de l'échange initial

# Interaction EGP/IGP

- OSPF :
  - un routeur externe importe les chemins auxquels il est directement relié
  - exporte les routes indirectes avec un chemin de SA conforme aux décisions de routage d'OSPF
- RIP :
  - BGP indique qu'il ne connaît pas le chemin
  - RIPv2 conserve le numéro de SA pour les routes externes. Celui-ci est recopié par BGP

# CIDR : Classless Inter Domain Routing

- Éviter la mort de l'Internet
  - par épuisement des classes B
  - par explosion des tables de routage
- On alloue plusieurs réseaux de classe C
  - les numéros alloués sont contigus. Ils ont donc les mêmes préfixes (bits de poids fort)
  - un routage sur les préfixes permettrait de limiter la taille des tables
  - un préfixe par "région" permettrait de faciliter le routage

# Répartition géographique

Multirégionales	192.0.0.0	193.255.255.255
Europe	194.0.0.0	195.255.255.255
Autres	196.0.0.0	197.255.255.255
Amérique du Nord	198.0.0.0	199.255.255.255
Amérique centrale, Amérique du sud	200.0.0.0	201.255.255.255
Zone Pacifique	202.0.0.0	203.255.255.255
Autres	204.0.0.0	205.255.255.255
Autres	206.0.0.0	207.255.255.255

# Super-réseaux et BGP4

## BGP4 implémente CIDR

- l'attribut chemin de SA est décomposé en :
  - *une séquence de SA* : le nombre de SA dans la séquence permet de déterminer la longueur du chemin et de détecter les boucles
  - *un ensemble de SA* : qui agrège les réseaux accessibles via le chemin
- lors d'interactions avec des protocoles antérieurs, il est nécessaire de désagréger les chemins