

# Routage IP

Roland Dirlewanger

CNRS - Délégation Aquitaine et Poitou Charentes

[rd@dr15.cnrs.fr](mailto:rd@dr15.cnrs.fr)

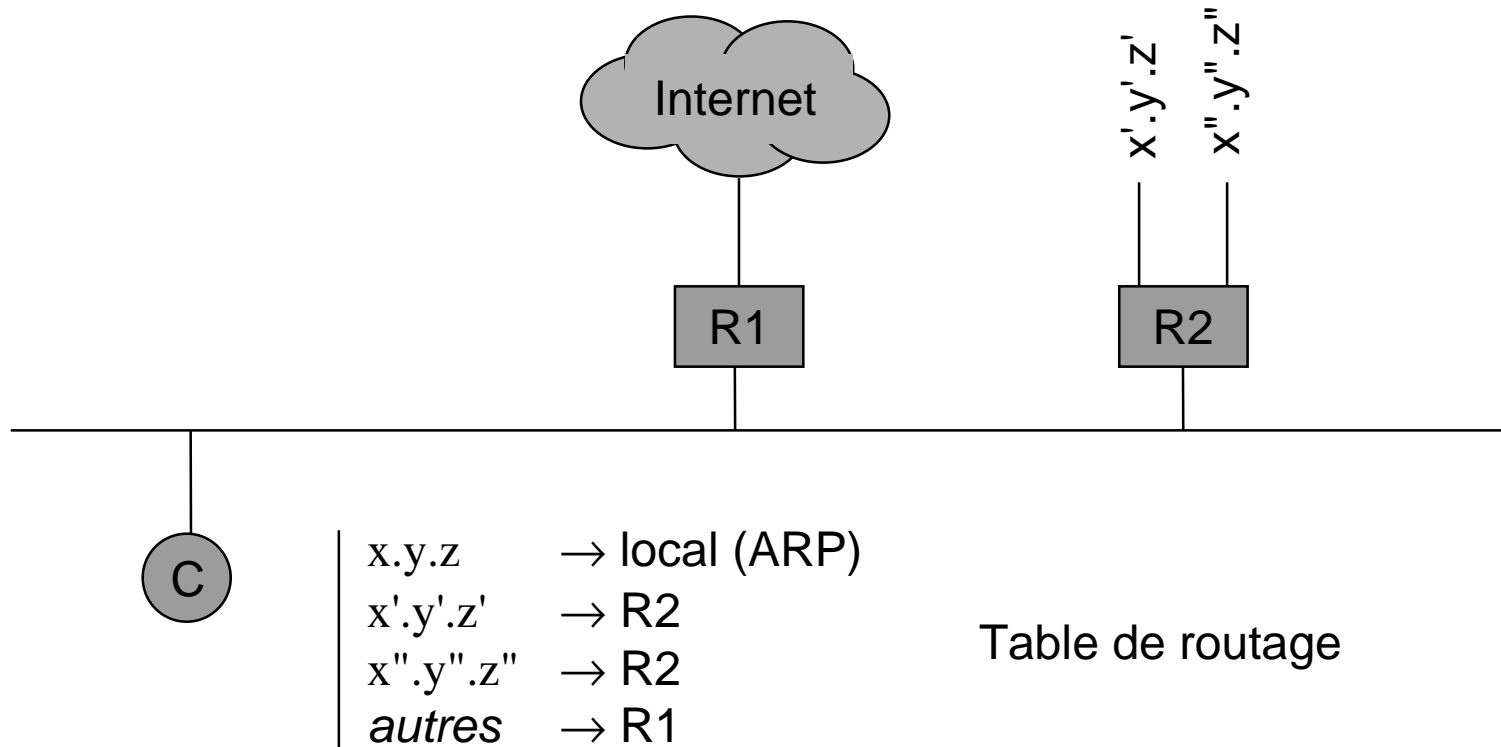
# Bibliographie (1)

---

- Christian Huitema, "Le routage dans l'Internet", Eyrolles, 1995.  
ISBN 2-212-08902-3
- [RFC 1654], "A Border Gateway Protocol 4 (BGP-4)", 1994
- [RFC 1655], "Application of the Border Gateway Protocol in the Internet", 1994

# Routage IP : le cas du site feuille

---



# Routage IP: l'acquisition des routes

---

- routage statique : par configuration de l'équipement
  - `route add destination relais distance`
- ICMP "redirect" : automatique
  - un routeur reçoit un paquet par une interface
  - il constate que pour l'acheminer, il doit le renvoyer par la même interface
  - il envoie un paquet "ICMP redirect" pour indiquer à l'émetteur qu'il peut atteindre le destinataire directement.

# Routage IP: protocoles de routage (1)

---

- Deux familles de protocoles
  - protocole de routage interne (*Internal Gateway Protocol*) : ex. RIP, OSPF, IGRP
  - protocole de routage externe (*External Gateway Protocol*) : ex. EGP, BGP
- Système autonome (AS, *Autonomous System*)
  - ensemble d'équipements sous une administration unique et obéissant à la même politique de routage

# Protocoles à vecteurs de distance

---

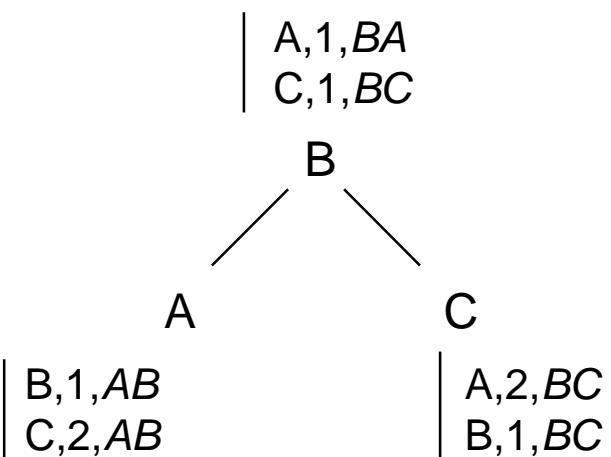
## Principe :

- distance ou métrique :
  - unité arbitraire liée soit au nombre de relais traversés, soit au coût de la liaison, soit à son débit, soit aux délais de transmission, pour atteindre la destination.
- émission :
  - diffusion du vecteur de distance (*destination, relais*)
- réception :
  - mise à jour de la table de routage ssi  $distance + 1$  pour *destination* est plus petite que la valeur actuelle
  - réarmement du temps d'expiration pour chaque destination telle que  $distance + 1 =$  valeur actuelle

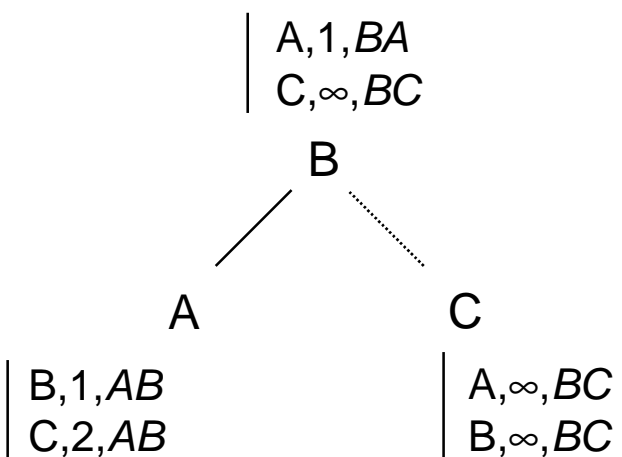
# Boucles et comptage jusqu'à

## l'infini (1)

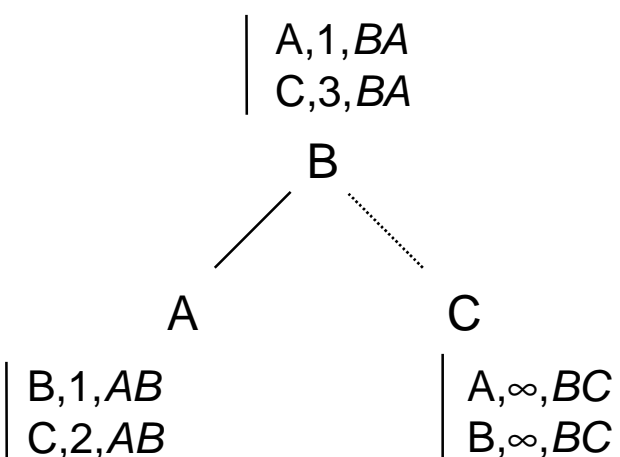
---



### Rupture de BC



### Emission de A



# Boucles et comptage jusqu'à l'infini (2)

---

- *étape 3* : A a émis avant B  $\Rightarrow$ 
  - il y a une boucle entre A et B pour tout paquet émis vers C
- *étapes suivantes* :
  - A expire la route vers C, mais en reçoit une nouvelle de B avec une distance + 1
  - B expire la route vers C, mais en reçoit une nouvelle de A avec une distance + 1
  - cycle qui converge lentement (distance  $\leftarrow$  distance + 2) vers la distance maximale  $\Rightarrow$  comptage jusqu'à l'infini.



# Boucles et comptage jusqu'à l'infini (3)

---

## Améliorer la convergence

- Horizon partagé
  - A n'envoie pas vers B les routes reçues de B
  - plus efficace : A envoie vers B les routes reçues de B en indiquant une distance infinie
- Mise à jour déclenchées
  - envoi immédiat des modifications notables (rupture de connectivité, etc.)
  - en fait, chaque routeur envoie la mise à jour avec un délai aléatoire pour éviter les tempêtes de mises à jour.

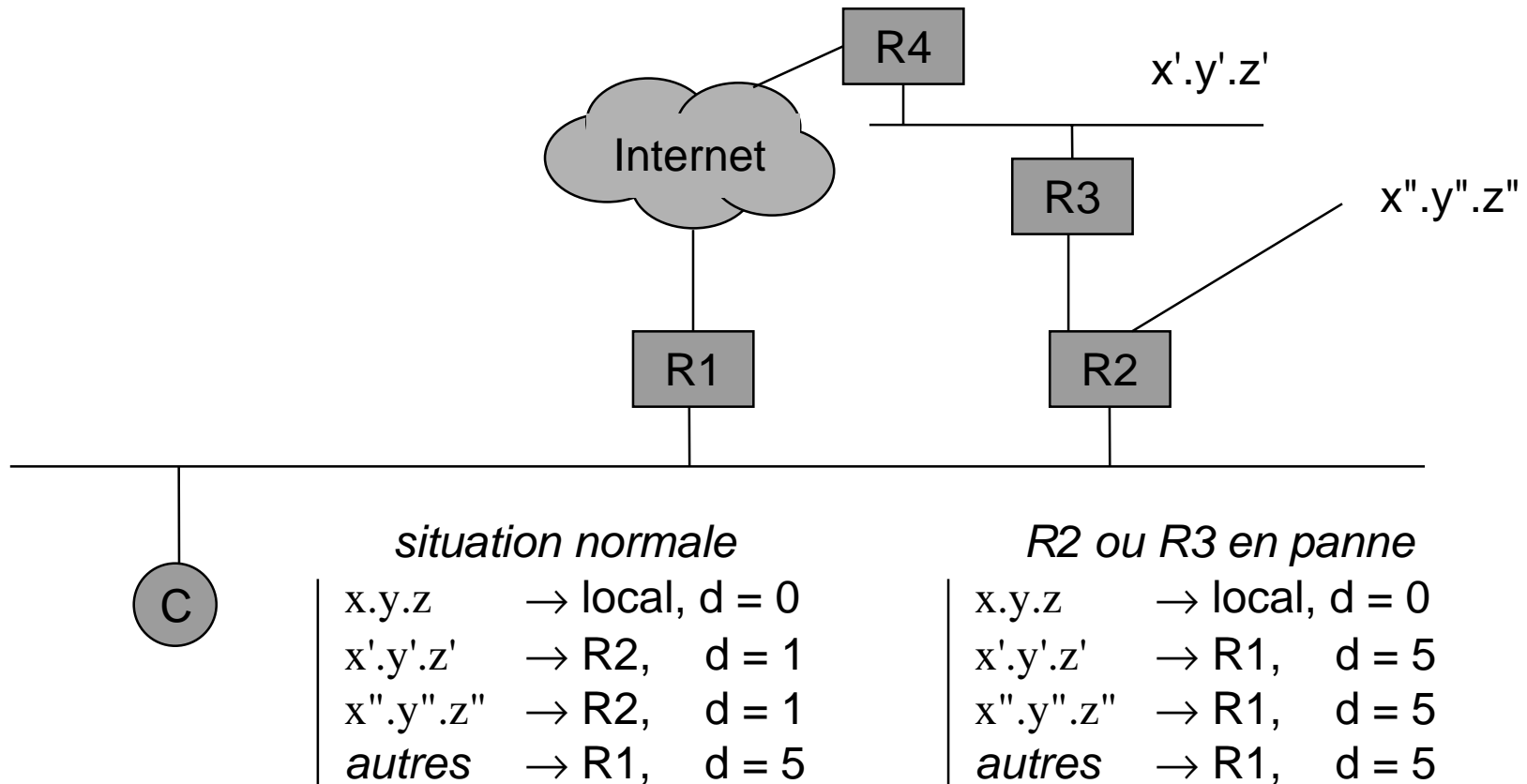
# Le protocole RIP

---

## RIP (*Routing Information Protocol*)

- Version 1
  - utilise la diffusion sur Ethernet et l'émission sur liaison PàP
  - chaque nœud diffuse sa table toutes les 30 secondes et expire les routes non rafraîchies toutes les 3 minutes
  - utilise l'horizon partagé et les màj déclenchées
- Version 2
  - gère les subnets
  - implémente le multicast
  - authentification
  - notion de relais suivant

# Exemple de liaison de secours



# Protocoles à vecteurs d'états de liaisons

---

## Principe :

- chaque nœud propage l'état de la liaison avec ses voisins par inondation
- chaque nœud maintient une carte complète du réseau :
  - il calcule le chemin le plus court pour atteindre une destination.

## Avantages :

- routeurs en phase => pas de boucles
- convergence très rapide vers un état stable en cas de modification

## Inconvénients :

- problème sur les gros réseaux (> 200 nœuds)

# La carte du réseau

---

- l'élément d'information est un enregistrement
  - identificateur de la liaison
  - destination atteinte par cette liaison
  - information de distance
  - numéro de séquence
- rôle de chaque nœud
  - propager l'état de ses liaisons
  - propager les enregistrements reçus de ses voisins sur toutes les interfaces, sauf celle par laquelle l'enregistrement est arrivé
  - ignorer ses propres enregistrements

# Plusieurs chemin vers la même destination

---

i.e distance (*dest* via *i*) = distance (*dest* via *j*)

Protocole à vecteur de distance :

- ne met pas à jour la table de routage => si on commence à router via *i*, on routera tous les paquets via *i*.
- si *i* tombe en panne, il faudra attendre la prochaine annonce de *dest* via *j*.

Protocole à vecteur d'état de liaisons :

- peut router via *i* ou via *j* ou les deux (équilibre de charge)
- si *i* tombe, le relais via *j* est immédiat

# Le protocole OSPF

---

- OSPF (*Open Short Path First*) recommandé par l'IETF (*Internet Engineering Task Force*) comme protocole de routage interne (RIP = trop d'inconvénients, IGRP = propriétaire cisco).
- protocole à vecteur d'états de liaison, avec :
  - découpage en zones des gros réseaux
  - gestion des métriques multiples (débit, délai, coût, fiabilité)
  - gestion des routes externes
  - temporisation des enregistrements, acquittements, checksums, authentification

# OSPF : démarrage

---

- démarrage d'un routeur :
  - numéro de séquence = valeur minimale, diffusion par inondation de l'état de liaison
  - s'il reçoit un état avec un numéro de séquence plus grand => numéro de séquence = ce numéro +1, diffusion par inondation de l'état de liaison
- rétablissement d'une liaison
  - échange avec ses voisins de l'état de ses liaisons
    - envoi de description (identifiant de la liaison, n° de séquence)
    - transmission uniquement des enregistrements de liaison différents



# OSPF : les numéros de séquence

---

- éviter que le numéro de séquence déborde ( $\geq 2^{32}$ )
  - numéros de séquence évoluent selon l'algorithme suivant ( $N = 2^{31} - 1$ )
    - initialisation :  $s = -N + 1$
    - progression : si  $s < N - 2$  alors  $s \leftarrow s + 1$ ; si  $s = N - 2$  alors  $s \leftarrow 0$
    - comparaison : si  $s_1 < 0$  ou  $s_2 < 0$ , alors comparaison classique  
si  $s_1 \geq 0$  et  $s_2 \geq 0$ . Supposons que  $s_1 < s_2$  alors  
si  $s_2 - s_1 > (N - 1)/2$  alors  $s_2$  plus récent  
sinon  $s_1$  plus récent

# OSPF : les temporisations

---

- retransmission par inondation de toutes les liaisons toutes les 30 minutes
- éviter de conserver des routes anciennes
  - les enregistrements d'états de liaison ont un champ "âge" :
    - vaut 0 lorsque l'enregistrement est diffusé par le routeur responsable de la liaison
    - augmente d'un à chaque routeur traversé puis d'un toutes les secondes
  - si l'âge atteint 1 heure sur un routeur => envoi par inondation de la suppression de cet enregistrement
  - pour chaque enregistrement, on compare l'âge reçu avec l'âge actuel.
    - Si la différence est de quelques unités, c'est le cas normal
    - Si la différence dépasse 15 minutes, on met à jour l'enregistrement correspondant

# CIDR : le routage sans classe

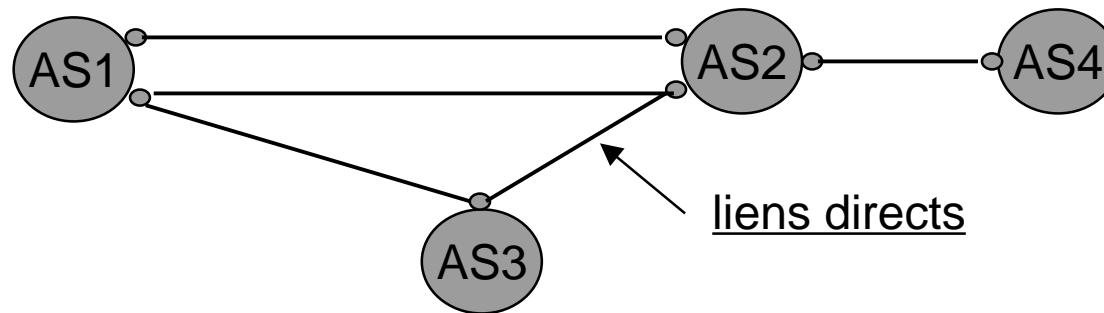
---

- CIDR (*Classless Internet Domain Routing*) définit la notion de *supernet* par opposition aux *subnets*: supernet = masque inférieur au masque naturel de la classes A, B ou C.
- CIDR privilégie la notation préfixe/longueur par rapport à la notation adresse/masque. On utilisera plutôt 193 . 51 / 8 que 193 . 51 . 0 . 0 / 255 . 255 . 0 . 0
- Objectif : agréger les réseaux IP pour diminuer la taille des tables de routage dans l'Internet.

# Les protocoles de routage externe

---

- Inconvénients des IGP :
  - vecteur de distance : problème des boucles, longs délais de convergence
  - vecteur d'état de liaison : calcul de la carte pour les gros réseaux
- Solution :
  - routage par Systèmes Autonomes (AS) plutôt que réseau par réseau



# BGP4 : un protocole à vecteur de chemin

---

## Principe

- Les routeurs à la frontière des AS échangent des informations de routage :
  - chemin d'AS et liste des routes accessibles par ce chemin
  - attributs des chemins
  - routes devenues inaccessibles
- Chaque routeur calcule pour chaque destination le chemin le plus court en terme d'AS traversés

# BGP4 : points forts

---

- Détection triviale des boucles :
  - si l'AS d'un routeur apparaît dans le chemin, le routeur ignore le chemin
- Utilise TCP/IP entre voisins => pas de gestion d'accusés de réception/retransmission (comme dans OSPF)
- Chaque routeur ne transmet que les modifications et que les informations concernant les routes qu'il utilise effectivement
- Chaque AS peut utiliser ses propres métriques et sa propre politique d'utilisation acceptable (*Acceptable Use Policy*, AUP)

# BGP4 : établissement d'une connexion

---

- Connexion TCP entre voisins avec détection de collision :  
A demande l'ouverture d'une connexion avec B pendant que B demande l'ouverture d'une connexion avec A.
- Echange de paramètres
  - n° d'AS
  - identifiant BGP, i.e. l'une des adresses IP : l'identifiant du routeur est le même pour toutes les interfaces
  - temps de garde (*Hold Time*) de la connexion
- Echange initial des tables de routage

# BGP4 : les échanges entre voisins

---

- TCP est un protocole en mode "flux" : l'entête de chaque message contient sa longueur et permet de déterminer le début de l'entête du message suivant
- Types de messages :
  - sondes (*keepalive*) toutes les (*temps de garde*) / 3 secondes
  - mises à jour
  - notifications (erreurs dans le protocole, fin de session)



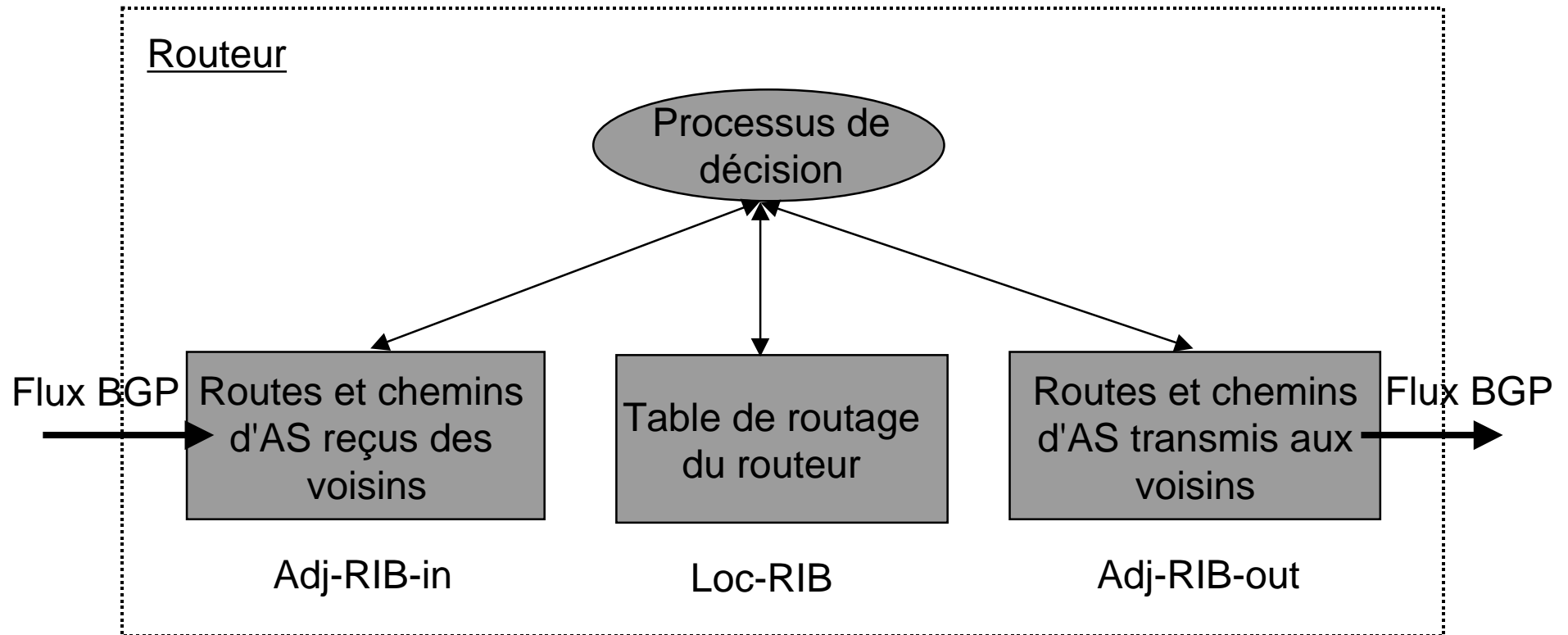
# BGP4 : les mises à jour

---

Mise à jour : deux parties pouvant être vides

- liste de routes devenues inaccessibles
- chemin d'AS
  - liste des AS formant le chemin
  - attributs de ce chemin
    - origine des routes (acquises de façon interne à l'AS, acquises par EGP, acquises autrement)
    - prochain relais
    - préférences sur le chemin
    - information d'agrégation
  - liste des préfixes des routes accessibles par ce chemin

# BGP4 : traitement par le routeur



RIB = Base d'Information de Routage (*Routing Information Base*)

# BGP4 : le processus de décision

---

- Phase 1 : réception des mises à jour
  - suppression de routes inaccessibles de Adj-RIB-in
  - calcule la préférence pour chaque chemin
- Phase 2 :
  - installe la meilleure route dans la table de routage (Loc-RIB)
- Phase 3 :
  - transmet les nouvelles informations aux voisins en rajoutant son propre n° d'AS en début de chemin.

# BGP4 : agrégation des adresses

