Border Gateway Protocol
BGP4 et MP-BGP4
Section 2

AfNOG 2009

Le Caire, 11-15 Mai 2009

aalain@trstech.net

Attributs de chemin BGP

- Encodés sous la forme d'un triplet Type, Longueur & Valeur (TLV)
- Attributs Transitifs ou non transitif
- Certains attributs sont obligatoires
- Ils sont utilisés pour choisir le meilleur chemin
- Ils permettent d'appliquer des règles d'ingéniérie du trafic (routage politique)

Liste des attributs de chemins BGP

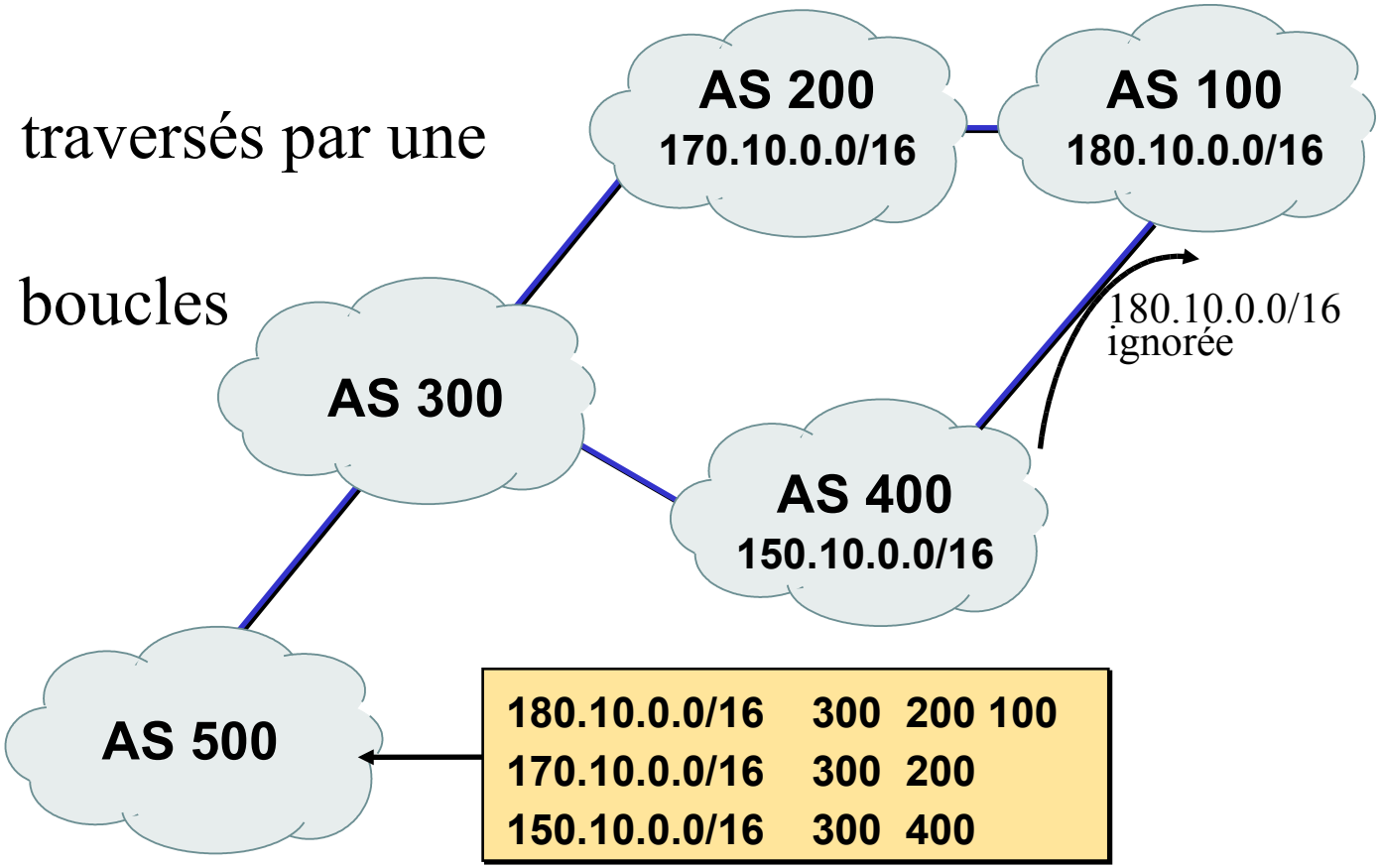
- Origine
- AS-path (chemin d'AS)
- Next-hop (prochain routeur)
- Multi-Exit Discriminator (MED)
- Local preference (préférence locale)
- BGP Community (communauté BGP)
- Autres...

AS-PATH (chemin d'AS)

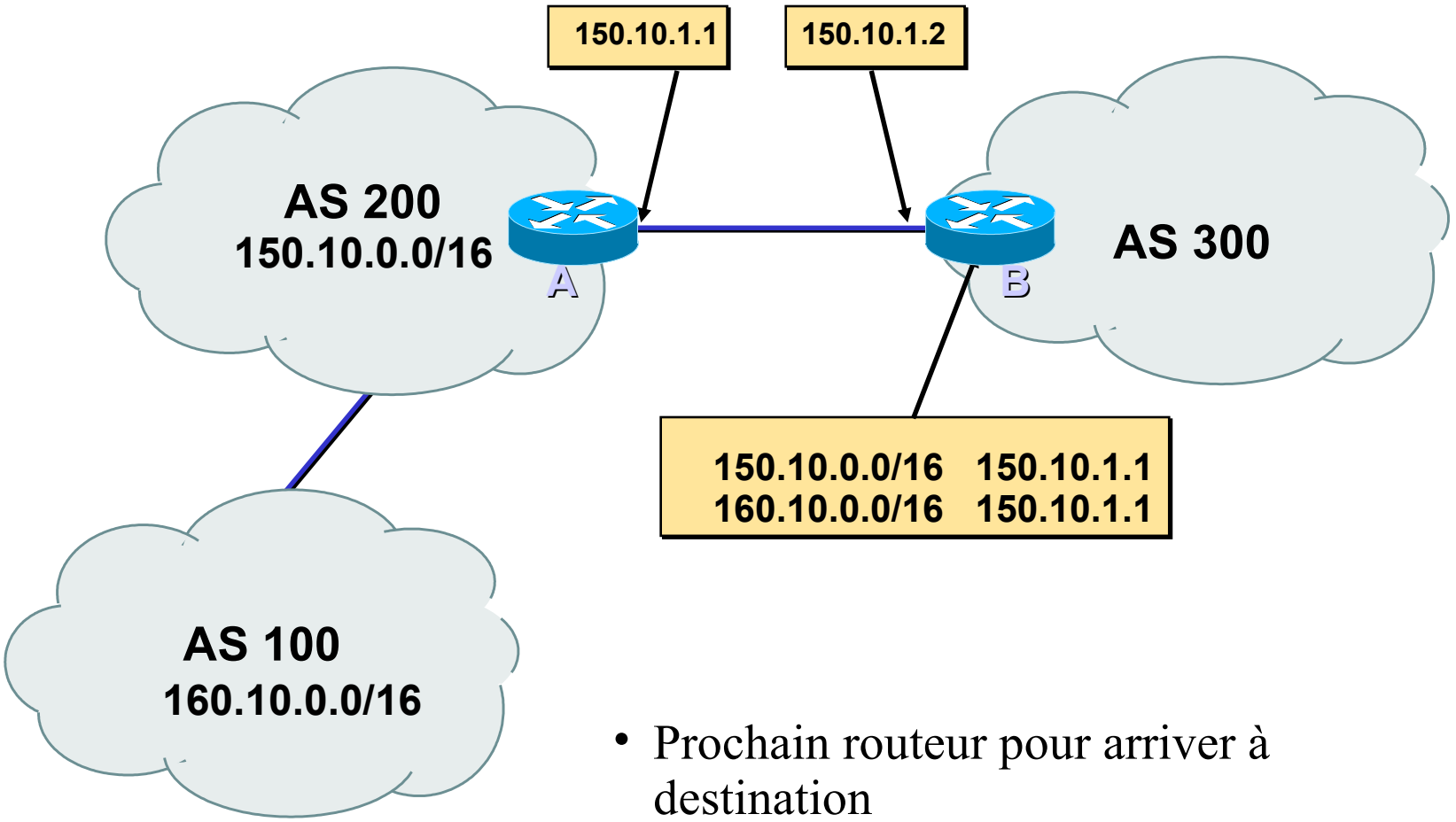
- Attribut mis à jour par le routeur envoyant un message BGP, en y ajoutant son propre numéro d'AS
- Contient la liste des AS traversés par le message
- Permet de détecter des boucles de routage
  - Une mise à jour reçue est ignorée si elle contient son propre numéro d'AS

AS-Path (chemin d'AS)

- Liste des AS traversés par une route
- Détection de boucles

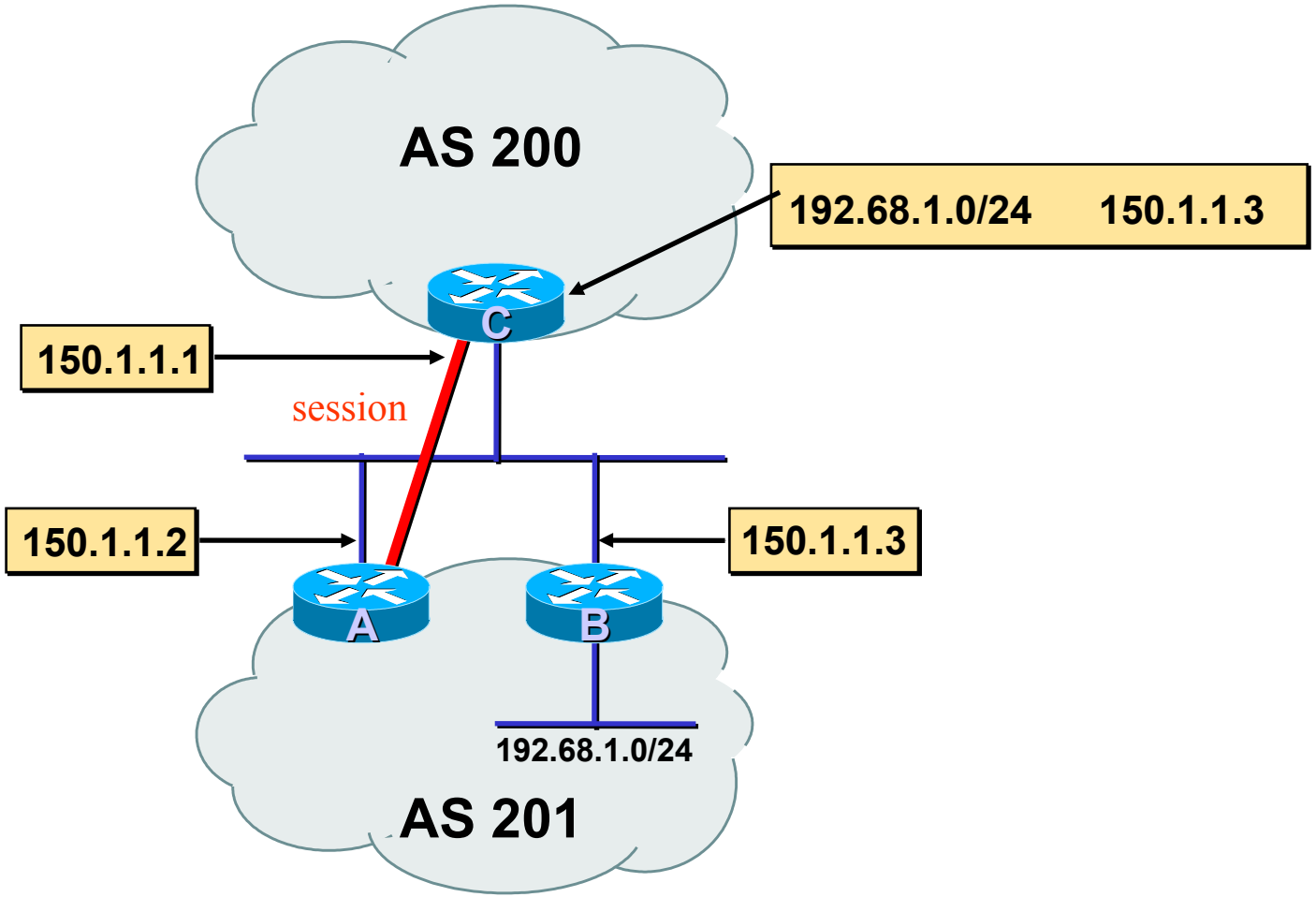


Next-Hop (prochain routeur)



- Prochain routeur pour arriver à destination
- Adresse de routeur ou de voisin en eBGP
- Non modifié en iBGP

Next-Hop sur un réseau tiers



- Serait plus efficace, mais c'est une mauvaise idée !

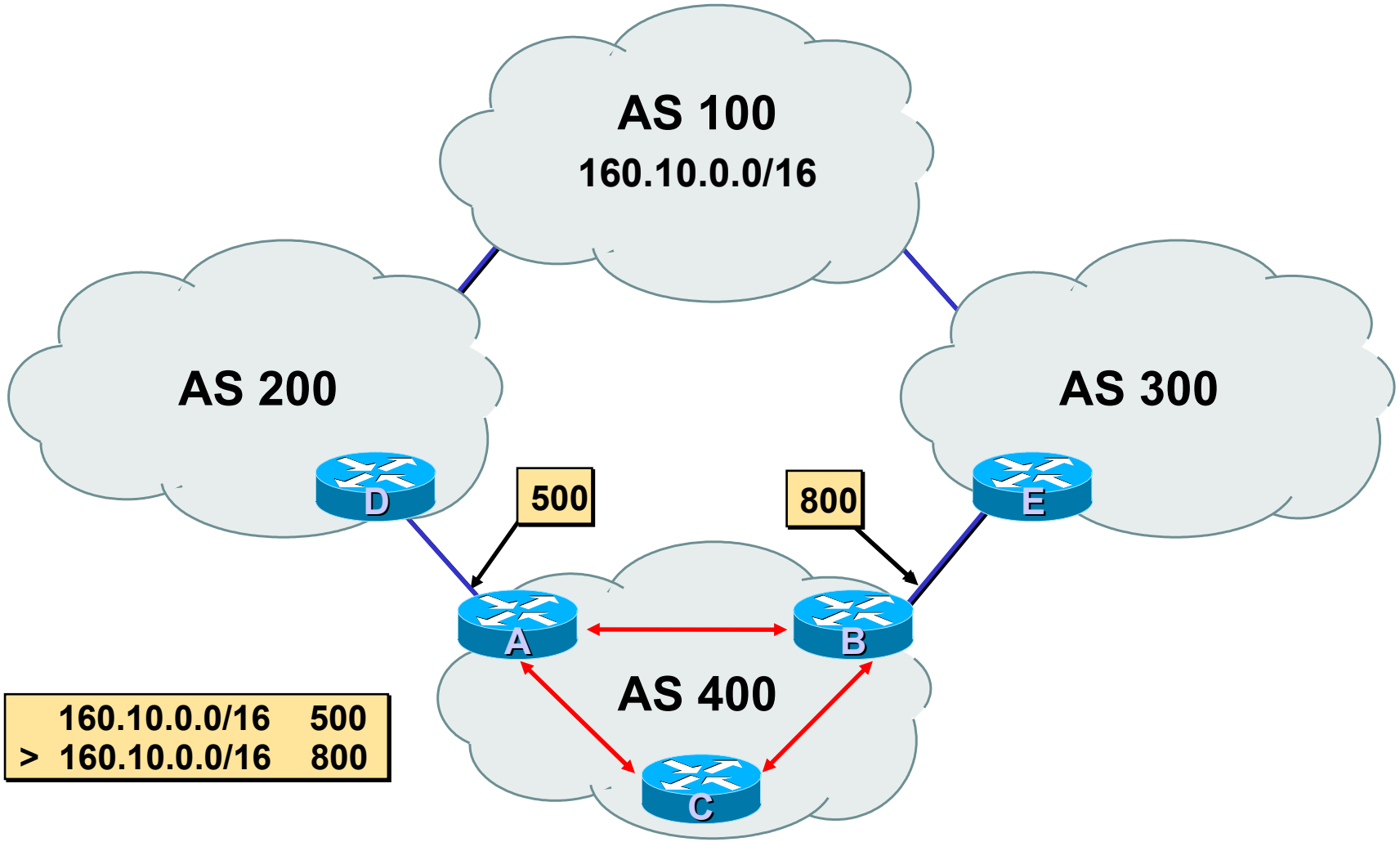
Next-Hop suite...

- Les routes vers l'ensemble des adresses de next-hop sont à transporter dans l'IGP
- Recherche de route récursive dans les tables
- BGP n'est plus lié à la topologie du réseau
- Les bonnes décisions de routage sont prises par le protocole IGP

Local Preference (préfence locale)

- Obligatoire pour iBGP, non utilisé dans eBGP
- Valeur par défaut chez Cisco : 100
- Paramètre local à un AS
- Permet de préférer une sortie à une autre
- Le chemin avec la préférence locale la plus élevée est sélectionné

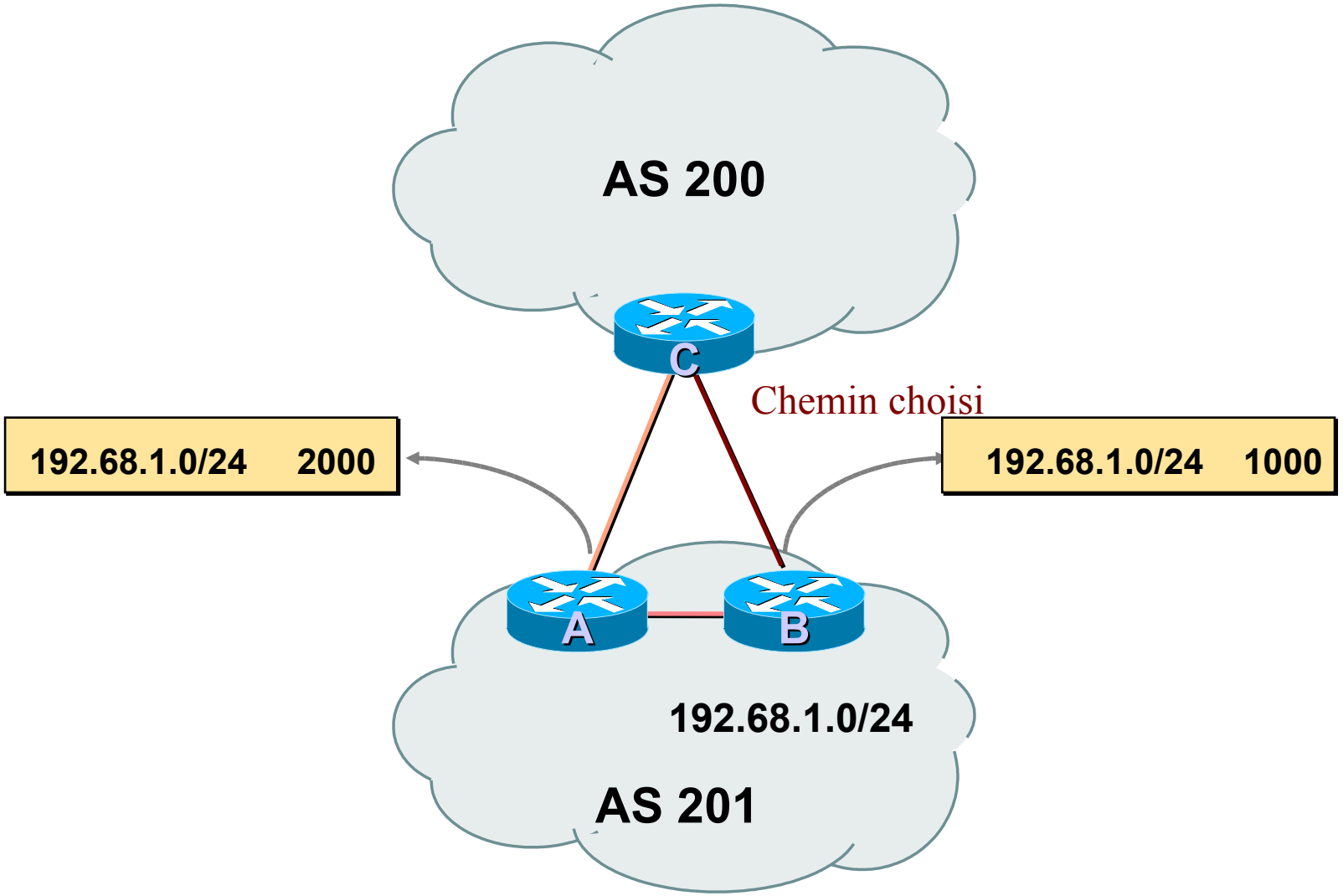
Local Preference (préférence locale)



Multi-Exit Discriminator

- Attribut non transitif
- Valeur numérique (0-0xffffffff)
- Permet de transporter des préférences relatives entre points de sortie
- Si les chemins viennent du même AS le MED peut être utilisé pour comparer les routes
- Le chemin avec le plus petit MED est sélectionné
- Le métrique IGP peut être choisi comme MED

Multi-Exit Discriminator (MED)



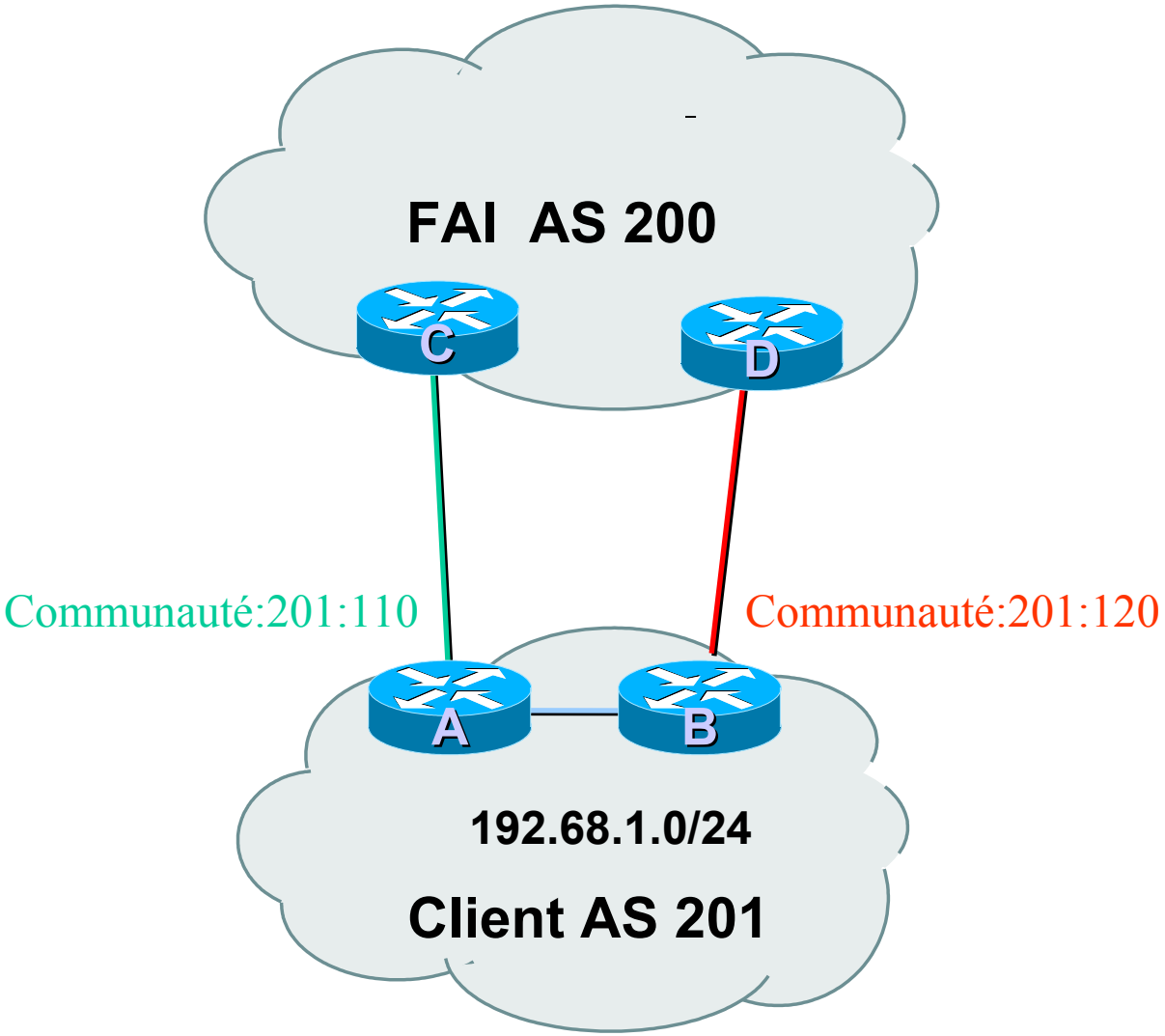
Origin (Origine de la route)

- Indique l'origine du préfixe
- Trois valeurs
  - IGP - préfixe obtenu avec une clause "network"
    - exemple : *network 35.0.0.0*
  - EGP - Redistribué par un EGP
  - Incomplete - Redistribué par un IGP
    - exemple : *redistribute ospf*
- IGP < EGP < INCOMPLETE

Communautés BGP

- Transitives, attribut facultatif
- Valeur numérique (0-0xffffffff)
- Permettent de créer des groupes de destinations
- Chaque destination peut appartenir à plusieurs communautés
- Attribut très flexible, car il permet de faire des choix avec des critères inter ou intra-AS

Communautés BGP



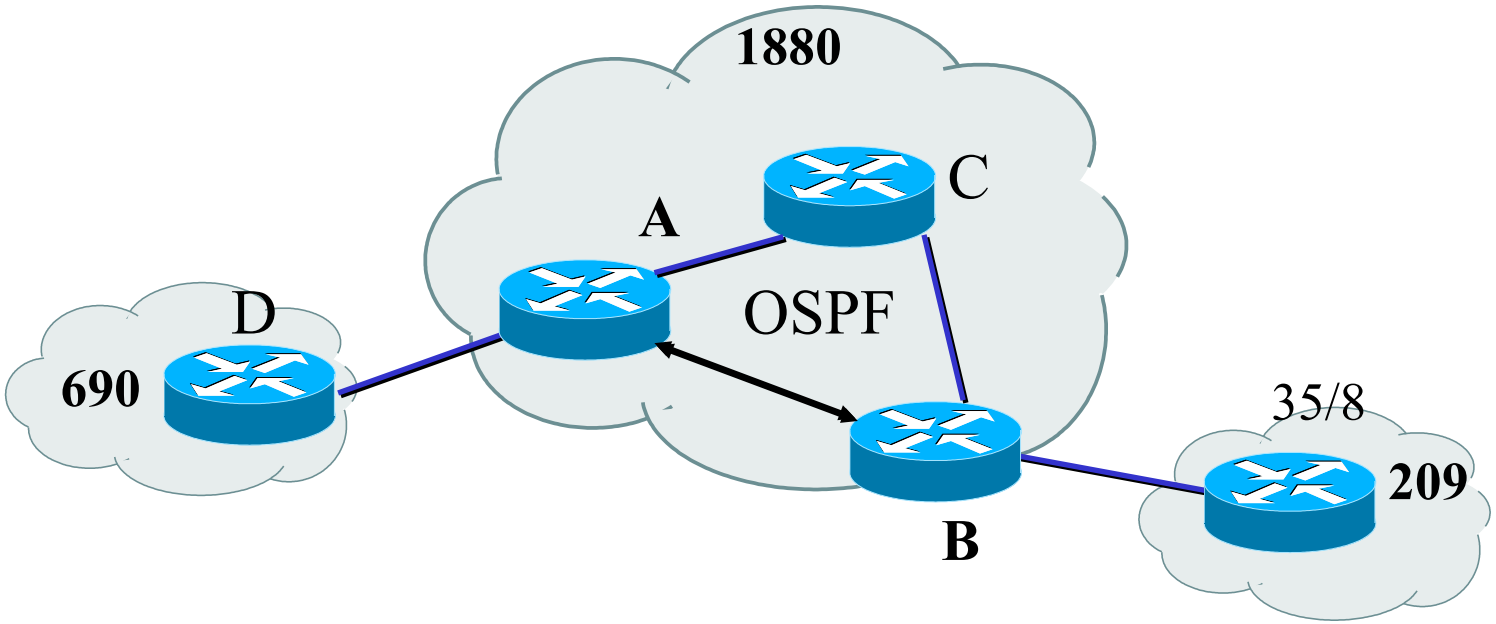
Poids (Weight)

- Attribut spécifique Cisco utilisé lorsqu'il y a plus d'une route vers la même destination
- Attribut local à un routeur (non propagé ailleurs)
- Valeur par défaut 32768 pour les chemins dont l'origine est le routeur et 0 pour les autres
- Lorsqu'il y a plusieurs choix, on préferra la route dont le poids est le plus élevé.

Distance administrative

- Les routes peuvent être apprises par plusieurs protocoles de routage
  - il faut les classifier pour faire un choix
- La route issue du protocole avec la plus faible distance est installée dans la table de routage
- Distances par défaut en BGP:
  - local (routes provenant du routeur) : 200
  - eBGP : 20, iBGP : 200
- Cela n'a pas d'impact dans l'algorithme de choix des chemins BGP, mais il y a un impact quand à installer ou pas une route BGP dans la table de routage IP

Synchronization (synchronisation)



Synchronization (synchronisation)

- Spécifique IOS Cisco : BGP n'annoncera pas une route avant que l'ensemble des routeurs de l'AS ne l'ait apprise par un IGP
- Désactiver la synchronisation si :
  - Votre AS ne sert pas d'AS de transit, ou
  - Tous les routeurs de transit tournent BGP, or
  - iBGP est utilisé sur le cœur de réseau (backbone)

Sélection d'une route BGP (bestpath)

Il ne peut y avoir qu'un seul meilleur chemin ! (sauf multipath)

- La route doit être synchronisée

C'est à dire être dans la table de routage

- Le "Next-hop" doit être joignable

Il se trouve dans la table de routage

- Prendre la valeur la plus élevée pour le poids (weight)

Critère spécifique Cisco et local au routeur

- Choisir la préférence locale la plus élevée

Appliqué pour l'ensemble des routeurs de l'AS

- La route est d'origine locale

Via une commande BGP "redistribute" ou "network"

Sélection d'une route BGP ...

- Choisir le plus court chemin d'AS

en comptant le nombre d'AS dans l'attribut AS-Path

- Prendre l'origine de valeur la plus faible

IGP < EGP < INCOMPLETE

- Choisir le plus petit MED

pour des chemins en provenance d'un même AS

- Préférer une route Externe sur une route Interne

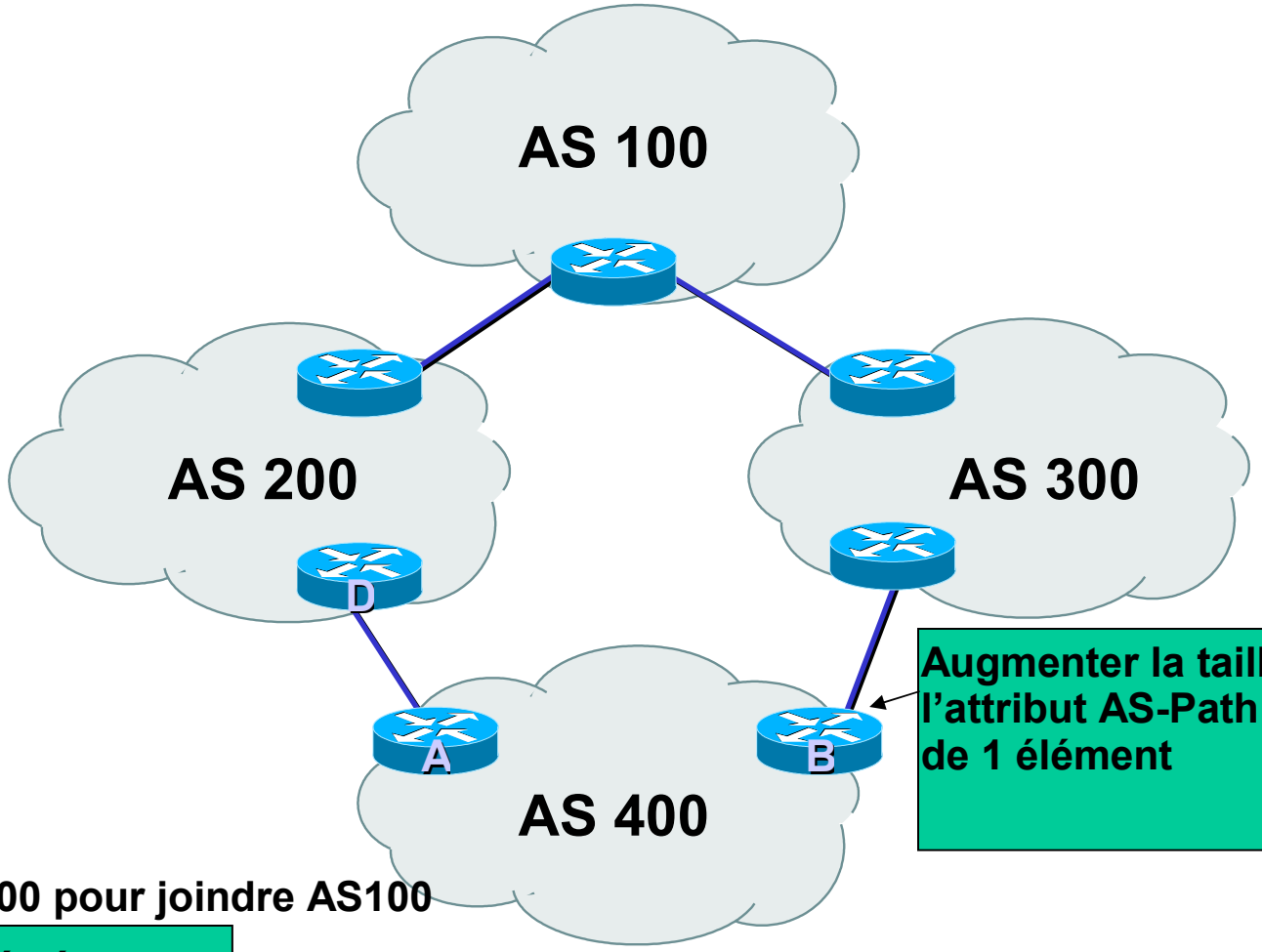
prendre la sortie la plus proche

- Choisir le "next-hop" le plus proche

Plus faible métrique IGP, donc plus proche de la sortie de l'AS

- Plus petit "Router-ID"
- Adresse IP du voisin la plus petite

Sélection d'une route BGP...



Augmenter la taille de l'attribut AS-Path au-moins de 1 élément

Politique AS 400 pour joindre AS100

AS 200 est préférré

AS 300 en secours

Politique de routage - Liste de préfixes, Route Maps et Listes de distribution (distribute lists)

Politique de routage

- Pourquoi ?
  - Pour envoyer le trafic vers des routes choisies
  - Filtrage de préfixes en entrée et sortie
  - Pour forcer le respect des accords Client-ISP
- Comment ?
  - Filtrage basé sur les AS - filter list
  - Filtrage basé sur les préfixes - distribute list
  - Modification d'attributs BGP - route maps

Politique - Liste de préfixes

- Filtrage par voisin
  - c'est une configuration incrémentielle
- Access-list utilisées très performantes
- Fonctionne en entrée comme en sortie
- Basé sur les numéros de réseaux (adressage IPv4 réseau/maque)
- Un "deny" est implicite à la fin de la liste

Liste de préfixes - Exemples

- Ne pas accepter la route par défaut
  - `ip prefix-list Exemple deny 0.0.0.0/0`
- Autoriser le préfixe 35.0.0.0/8
  - `ip prefix-list Exemple permit 35.0.0.0/8`
- Interdire le préfixe 172.16.0.0/12
  - `ip prefix-list Exemple deny 172.16.0.0/12`
- Dans 192/8 autoriser jusqu'au /24
  - `ip prefix-list Exemple permit 192.0.0.0/8 le 24`
  - Ceci autorisera toute route dans 192.0.0.0/8, sauf les /25, /26, /27, /28, /29, /30, /31 and /32

Listes de préfixes - Exemples 2

- Dans 192/8 interdire /25 et au-delà
  - `ip prefix-list Exemple deny 192.0.0.0/8 ge 25`
  - Ceci interdit les préfixes de taille /25, /26, /27, /28, /29, /30, /31 and /32 dans le bloc 192.0.0.0/8
  - Très ressemblant au précédent exemple
- Dans 192/8 autoriser les préfixes entre /12 et /20
  - `ip prefix-list Exemple permit 192.0.0.0/8 ge 12 le 20`
  - Ceci interdit les préfixes de taille /8, /9, /10, /11, /21, /22 et au-delà dans le bloc 192.0.0.0/8
- Autoriser tous les préfixes
  - `ip prefix-list Exemple 0.0.0.0/0 le 32`

Utilisation des listes de préfixes

- Exemple de configuration

```
router bgp 200
  network 215.7.0.0
  neighbor 220.200.1.1 remote-as 210
  neighbor 220.200.1.1 prefix-list PEER-IN in
  neighbor 220.200.1.1 prefix-list PEER-OUT out
!
ip prefix-list PEER-IN deny 218.10.0.0/16
ip prefix-list PEER-IN permit 0.0.0.0/0 le 32
ip prefix-list PEER-OUT permit 215.7.0.0/16
ip prefix-list PEER-OUT deny 0.0.0.0/0 le 32
```

Tout accepter du voisin, sauf nos réseaux

Envoyer uniquement nos réseaux au voisin

Distribute list - avec des ACL IP

```
access-list 1 deny 10.0.0.0
access-list 1 permit any
access-list 2 permit 20.0.0.0
```

… il faut créer des ACL avec l'ajout de nouveaux préfixes ...

```
router bgp 100
  neighbor 171.69.233.33 remote-as 33
  neighbor 171.69.233.33 distribute-list 1 in
  neighbor 171.69.233.33 distribute-list 2 out
```

Filtrage avec des expression régulières

- L'expression régulière décrit la forme que doit avoir l'argument
- Est utilisé pour comparer l'attribut AS-Path
- Exemple : ^3561.*100.*1$
- Grande fléxibilité qui permet de générer des expression complexes

Filtrage avec des expressions régulières

```
ip as-path access-list 1 permit 3561
ip as-path access-list 2 deny 35
ip as-path access-list 2 permit .*

router bgp 100
   neighbor 171.69.233.33 remote-as 33
   neighbor 171.69.233.33 filter-list 1 in
   neighbor 171.69.233.33 filter-list 2 out
```

Accepter les routes d'origine AS 3561. Tout le reste est rejeté en entrée ("deny" implicite).

Ne pas annoncer les routes de l'AS 35, mais tout le reste est envoyé (en sortie).

Route Maps

```
router bgp 300
neighbor 2.2.2.2 remote-as 100
neighbor 2.2.2.2 route-map SETCOMMUNITY out
!
route-map SETCOMMUNITY permit 10
match ip address 1
match community 1
set community 300:100
!
access-list 1 permit 35.0.0.0
ip community-list 1 permit 100:200
```

Route-map : clauses match & set

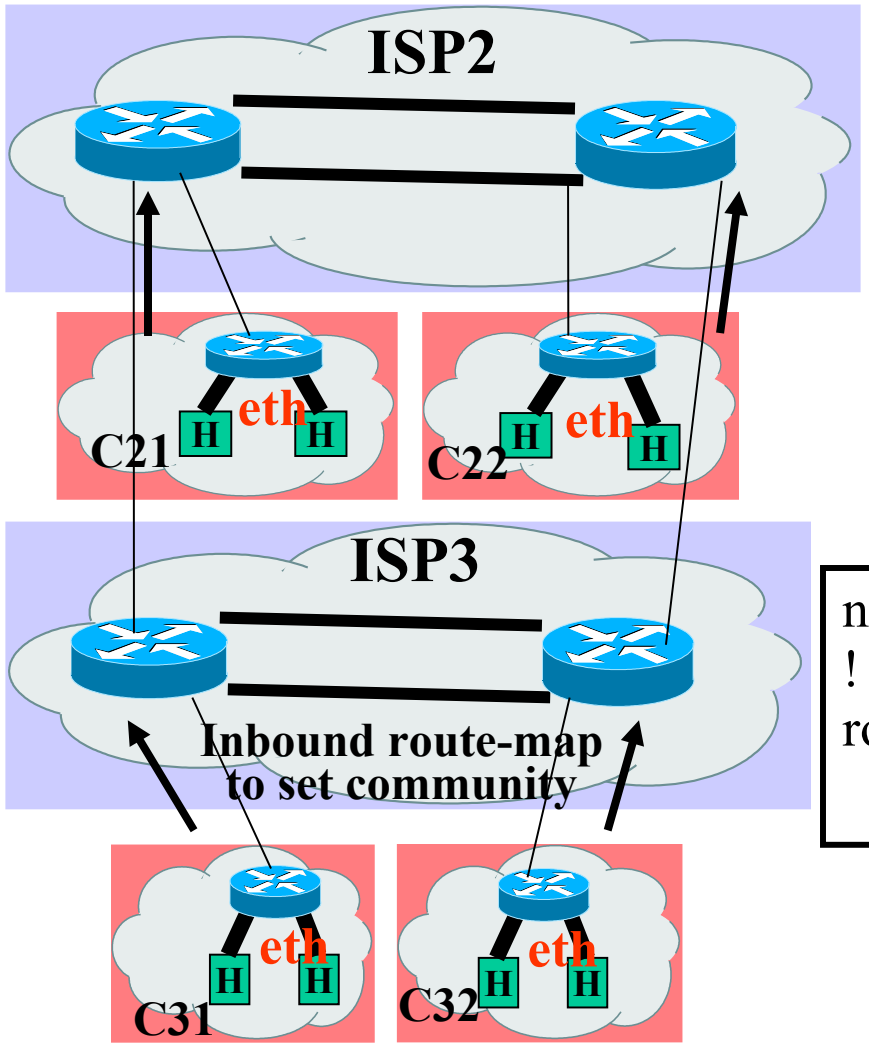
Match Clauses

- AS-path
- Community
- IP address

Set Clauses

- AS-path prepend
- Community
- Local-Preference
- MED
- Origin
- Weight
- Autres...

Exemple de configuration avecRoute-map




```
neighbor <y.y.y.y> route-map AS200_IN in
!
route-map AS200_IN permit 10
        match community 1
        set local-preference 200
!
ip community-list 1 permit 100:200
```

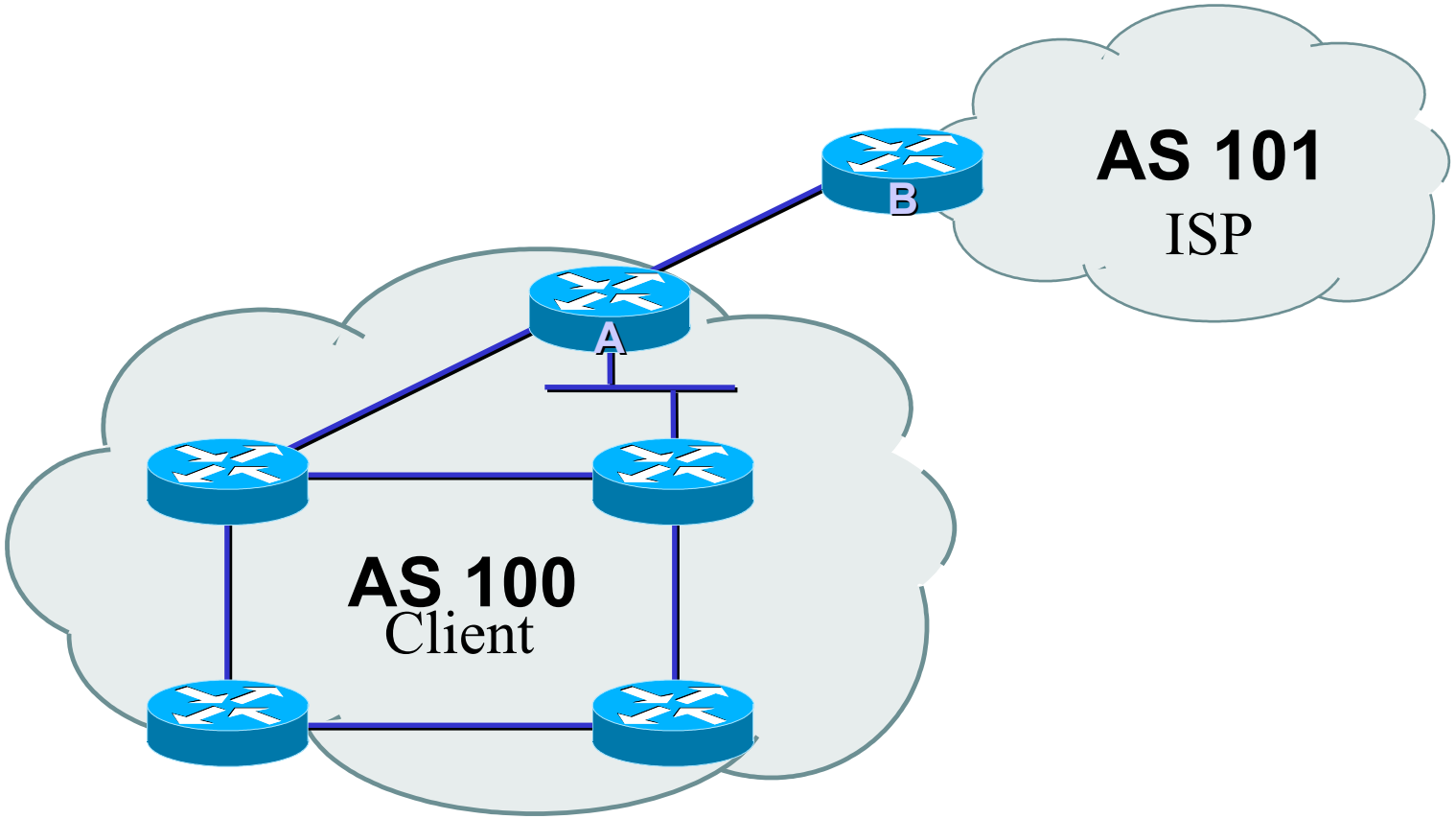
```
neighbor <x.x.x.x> route-map AS100_IN in
!
route-map AS100_IN permit 10
        set community 100:200
```

BGP et architecture de réseaux

AS “feuille” (stub AS)

- Situation ne nécessitant pas de BGP
- Route par défaut chez le FAI
- Le FAI annonce vos réseaux dans on AS
- La politique de routage de votre FAI est également la vôtre

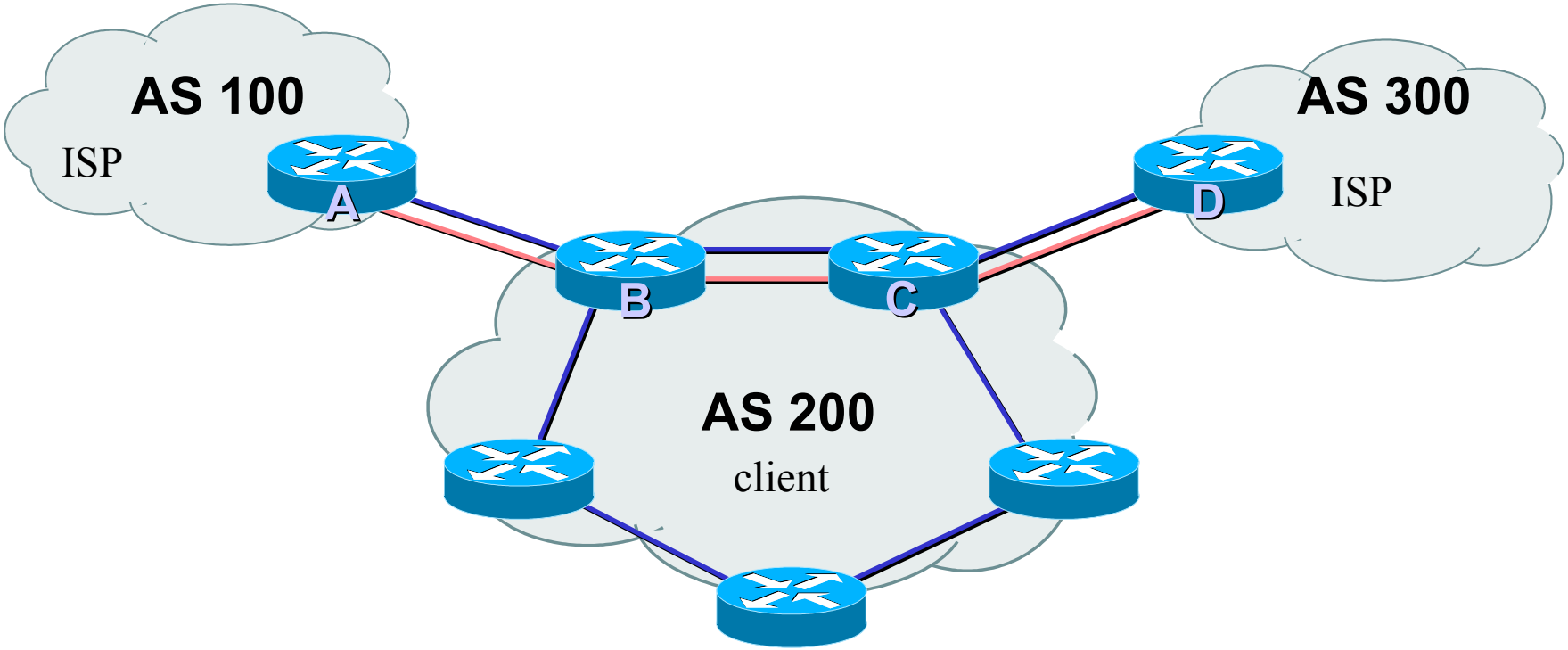
AS feuille



AS multi-raccordé (multi-homed)

- Les routeurs d'extrêmité font du BGP
- Sessions IBGP entre ces routeurs
- Il faut redistribuer les routes apprises avec prudence dans l'IGP, ou bien utiliser une route par défaut

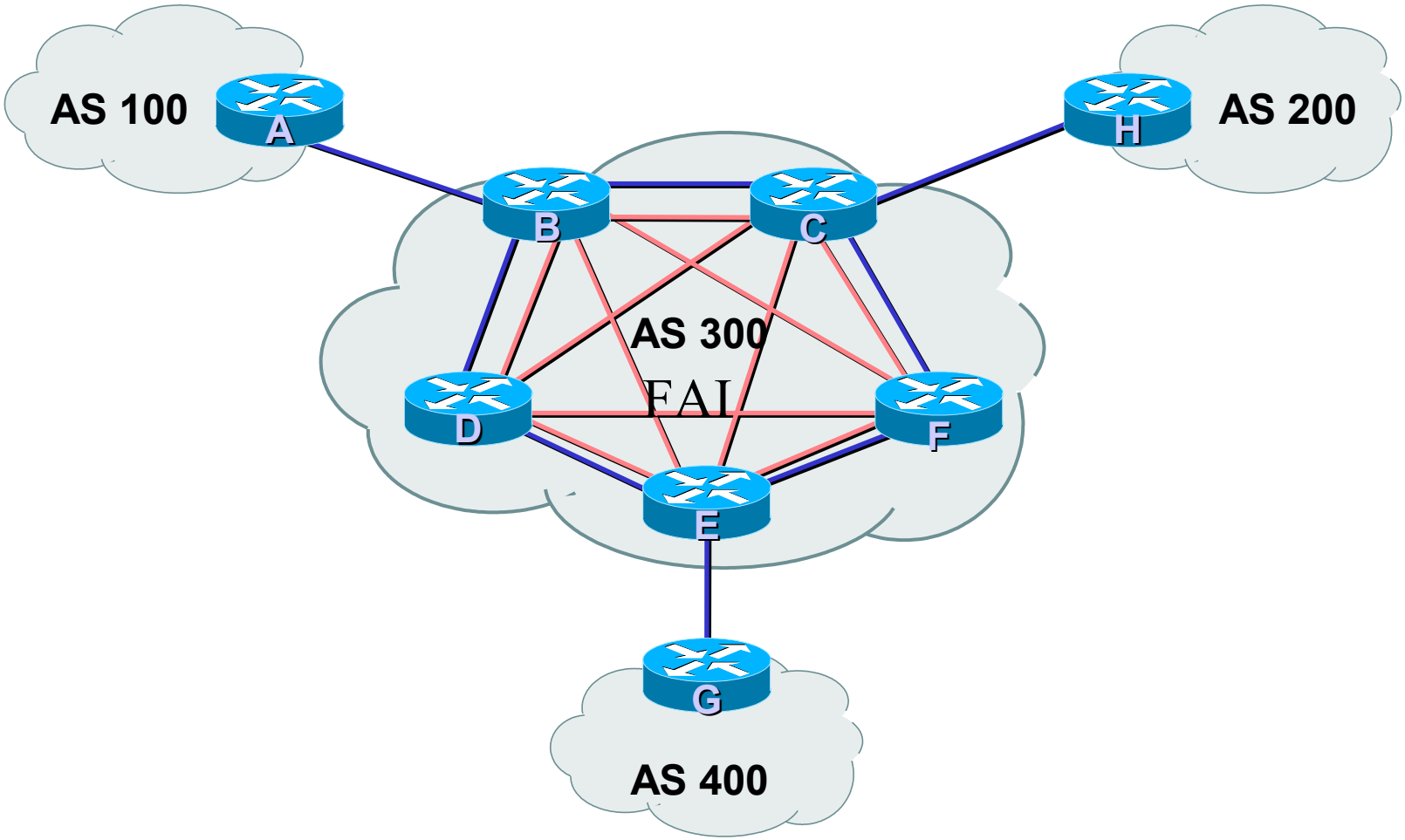
AS multi-homé



Réseau d'un FAI

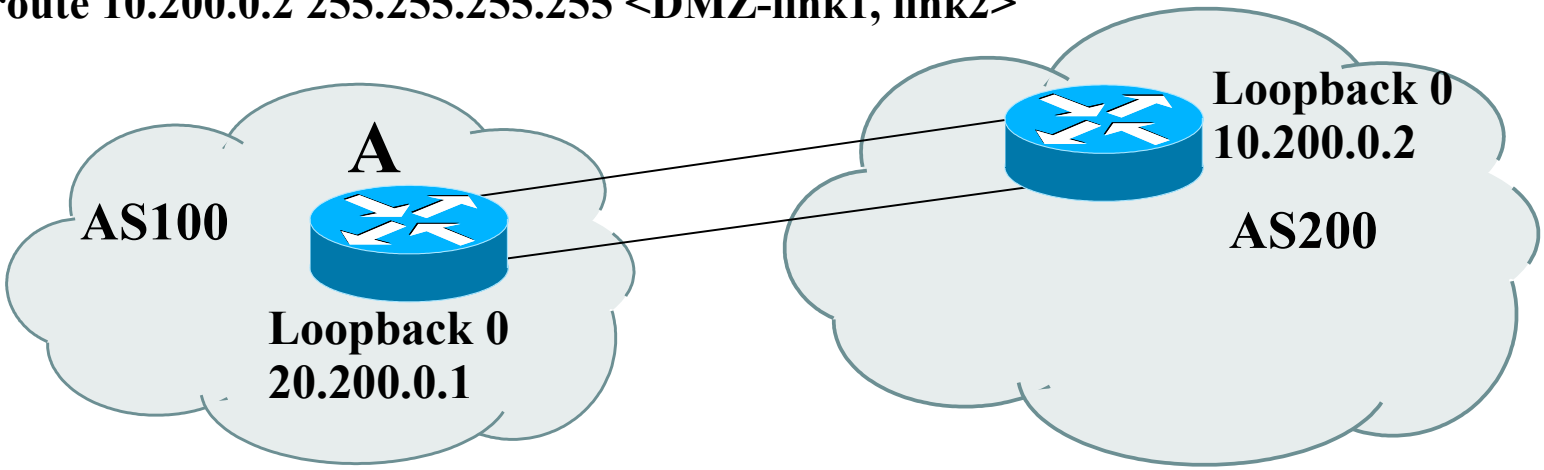
- IBGP permet de transporter les routes extérieures à l'AS
- Un IGP permet de gérer la topologie du réseau
- Un maillage complet iBGP est requis

Réseau typique d'un FAI



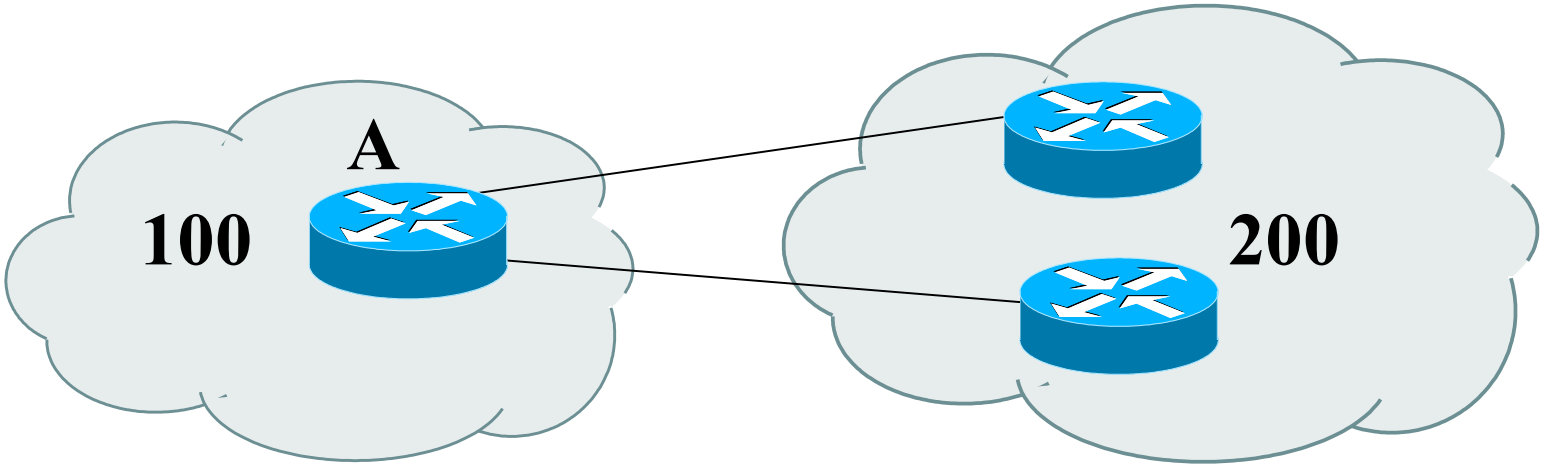
Partage de charge - 1 chemin

Routeur A:
interface loopback 0
ip address 20.200.0.1 255.255.255.255
!
router bgp 100
neighbor 10.200.0.2 remote-as 200
neighbor 10.200.0.2 update-source loopback0
neighbor 10.200.0.2 ebgp-multi-hop 2
!
ip route 10.200.0.2 255.255.255.255 <DMZ-link1, link2>



Partage de charge - Plusieurs chemins disponibles

Routeur A:
router bgp 100
neighbor 10.200.0.1 remote-as 200
neighbor 10.300.0.1 remote-as 200
maximum-paths 2



Note : A n'annoncera que 1 seul "bestpath" à ses voisins iBGP

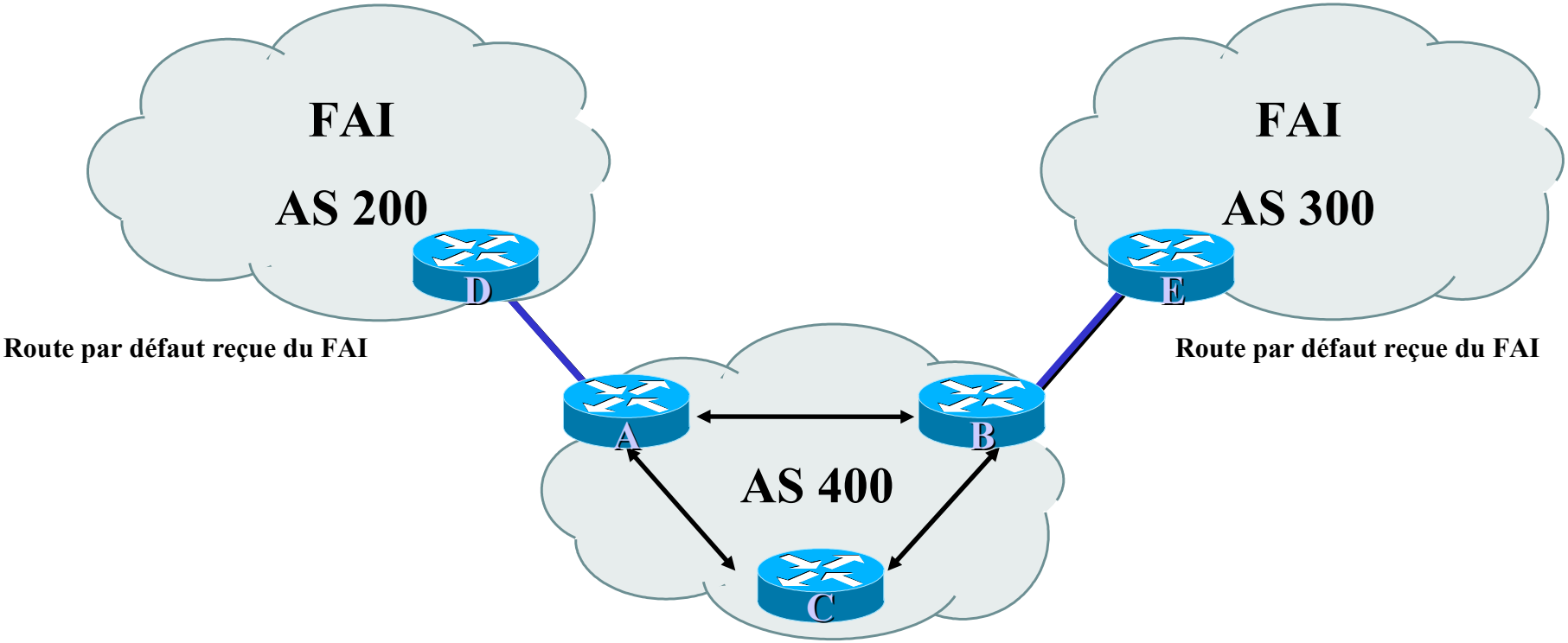
Redondance - Multi-homing

- Etre connecté de manière fiable à l'Internet
- 3 situations courantes en multi-homing
  - accepter la route par défaut des prestataires
  - clients + route par défaut chez les prestataires
  - recevoir toutes les routes de tous les voisins
- Adressage IP
  - fourni par les prestataires "upstream", ou
  - obtenu directement auprès d'un registre IP

Route par défaut des FAI

- Permet d'économiser la mémoire et la puissance de calcul
- Le FAI envoie une route par défaut BGP
  - le métrique IGP permet de choisir le FAI
- La politique des FAI détermine votre politique de trafic entrant
  - Il est cependant possible d'influencer cela en utilisant une politique de sortie, par exemple: AS-path prepend

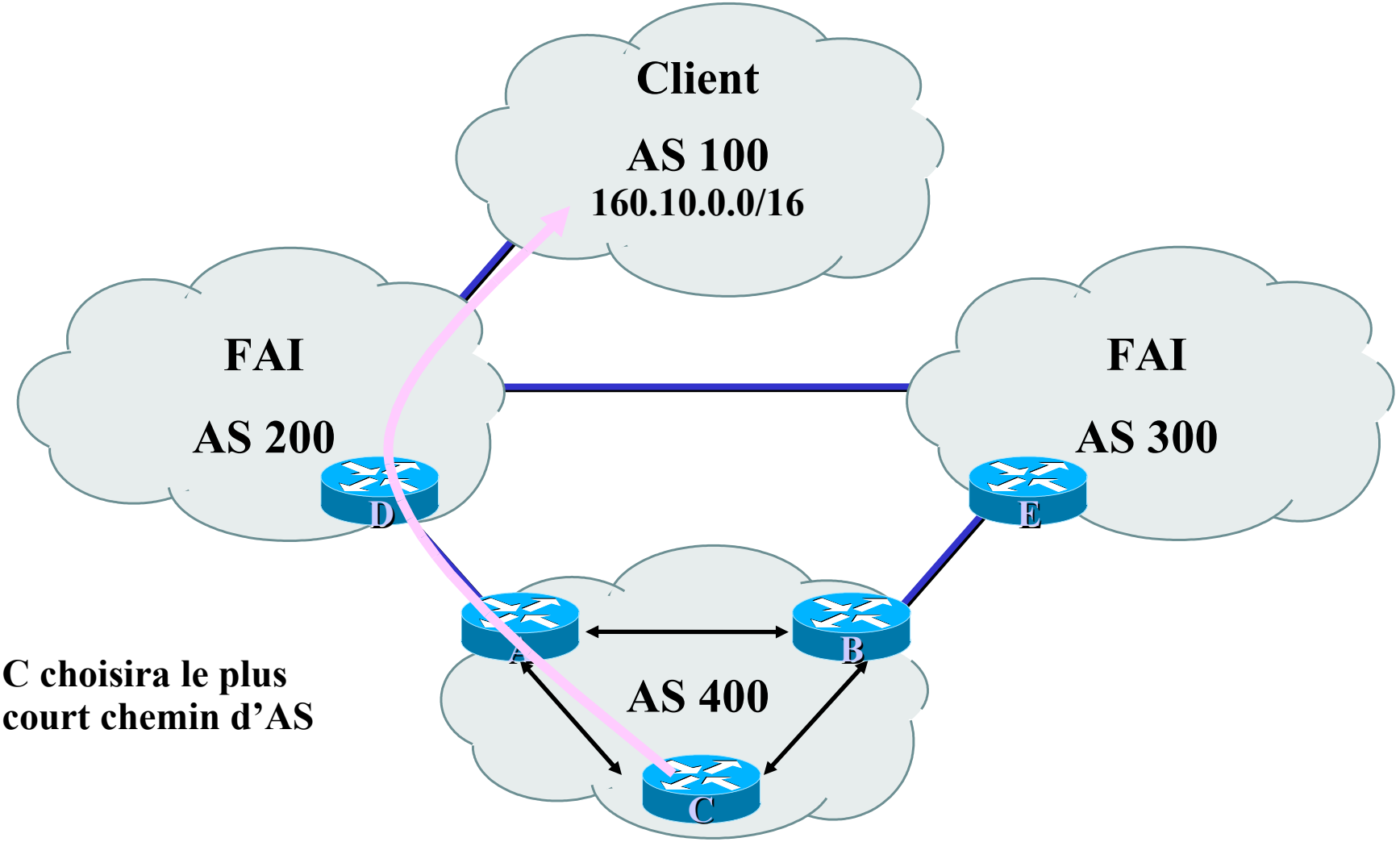
Route par défaut des FAI



Clients + route par défaut des FAI

- Consommation modérée de mémoire et CPU
- Gestion individuelle des routes des clients et route par défaut pour le reste
  - il est nécessaire de connaître les routes du client !
- Politique de routage entrant laissée aux FAI choisis
  - mais il est possible d'influencer ces choix (exemple : as-path prepend)

Les ISP annoncent les routes de leurs clients

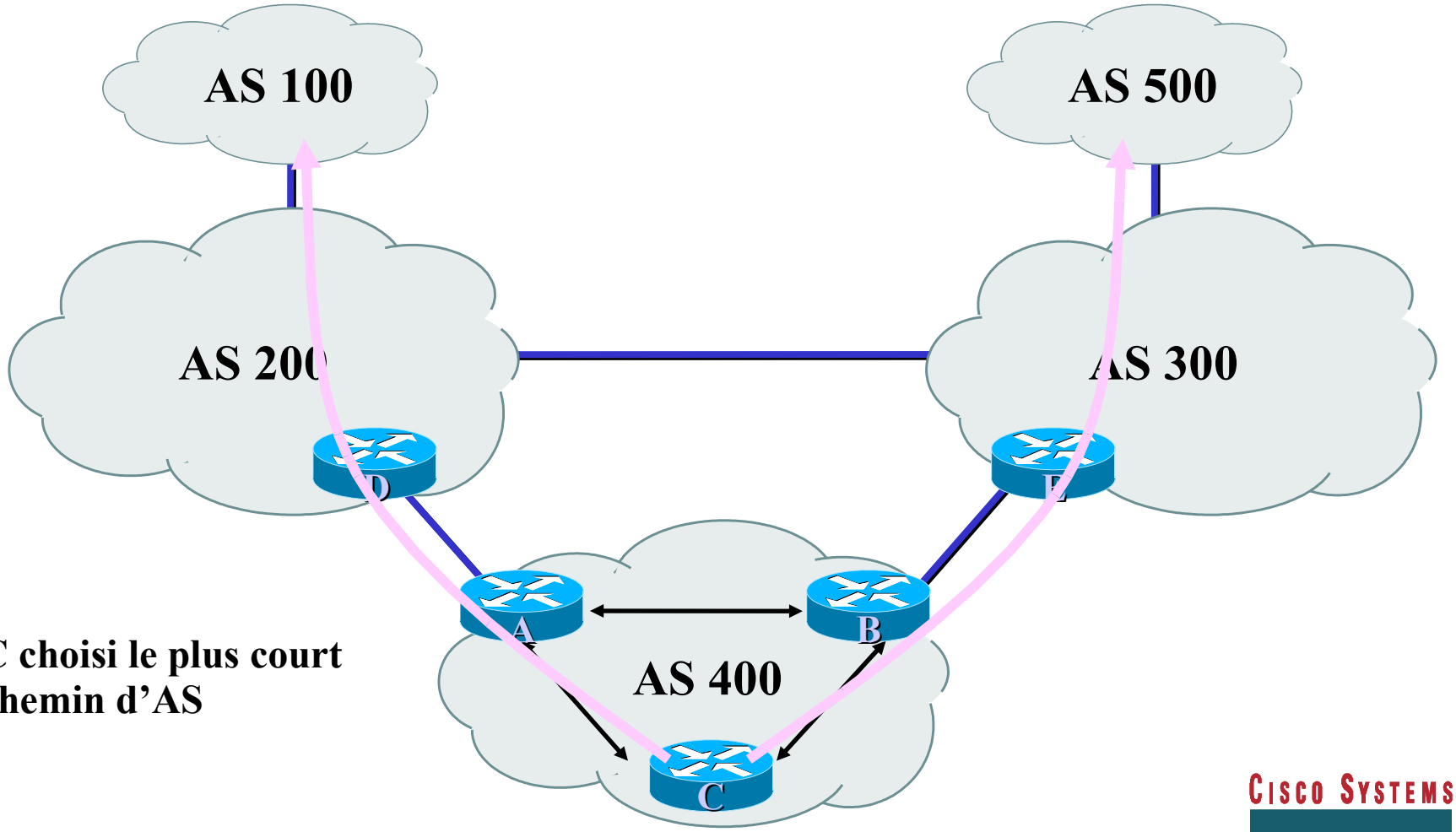


C choisira le plus court chemin d'AS

Gérer toutes les routes “full routing”

- Plus de consommation mémoire et CPU
- Contrôle plus poussé sur la politique de routage
- Les AS de transit gèrent généralement toutes les routes
- BGP est généralement le principal protocole de routage

Tous les prestataires envoient toutes les routes



C choisi le plus court chemin d'AS

Etat de l'art
Choix de l'IGP dans le Backbone

- L'IGP assure la gestion de la topologie de votre infrastructure - pas des réseaux de vos clients
- L'IGP doit converger rapidement
- L'IGP doit transporter les routes et masques - OSPF, IS-IS, EIGRP

Etat de l'art...
Raccorder un client

- Routes statiques
  - Vous les contrôlez directement
  - pas de "flaps"
- Protocole de routage dynamique
  - Vous devez filtrer ce que votre client annonce
  - Risque de "flaps"
- Utiliser BGP pour les clients "multi-homés"

Etat de l'art...
Se raccorder à d'autres FAI

- Annoncez uniquement vos réseaux
- Acceptez le minimum nécessaire
- Prendre le plus court chemin vers la sortie
- Agrégez les routes !!!
- FILTREZ ! FILTREZ! FILTEZ!

Etat de l'art...
Les points d'échange

- Les raccordements longue distance sont chers
- Ils permettent de profiter d'un point unique pour se raccorder à plusieurs partenaires