

Attributs BGP et sélection de route



SIF
AFNOG 2018, Dakar

Attributs BGP



Les “outils” disponibles pour faire le travail requis

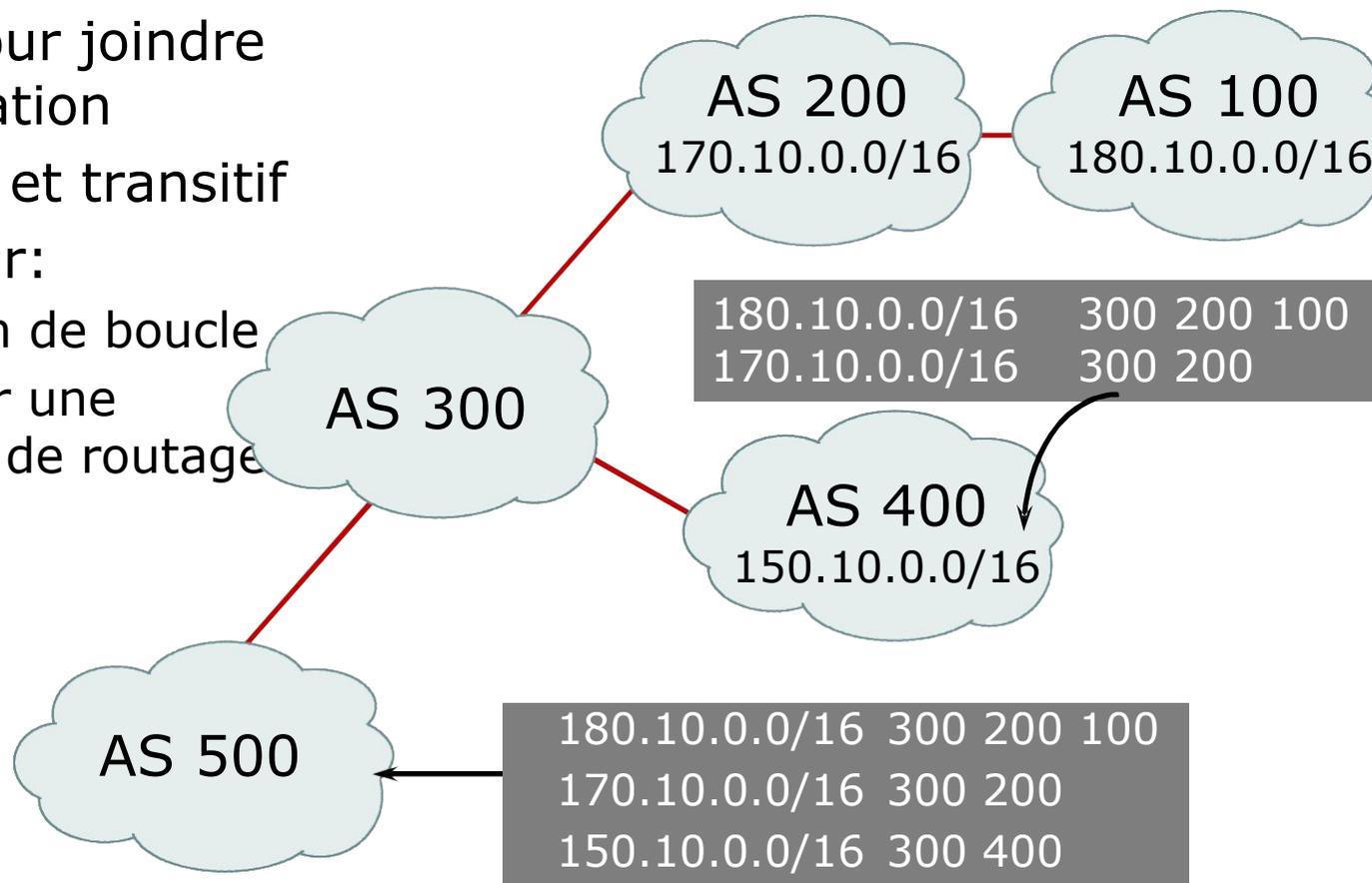
Qu'est ce qu'un Attribute ?

...	Next Hop	AS Path	MED
-----	----------	---------	-----	-----	-----

- ❑ Inclus dans les Update BGP
- ❑ Descrit les caractéristiques des préfixes
- ❑ Peut être soit transitif ou non transitif
- ❑ Certains sont obligatoires

AS-Path

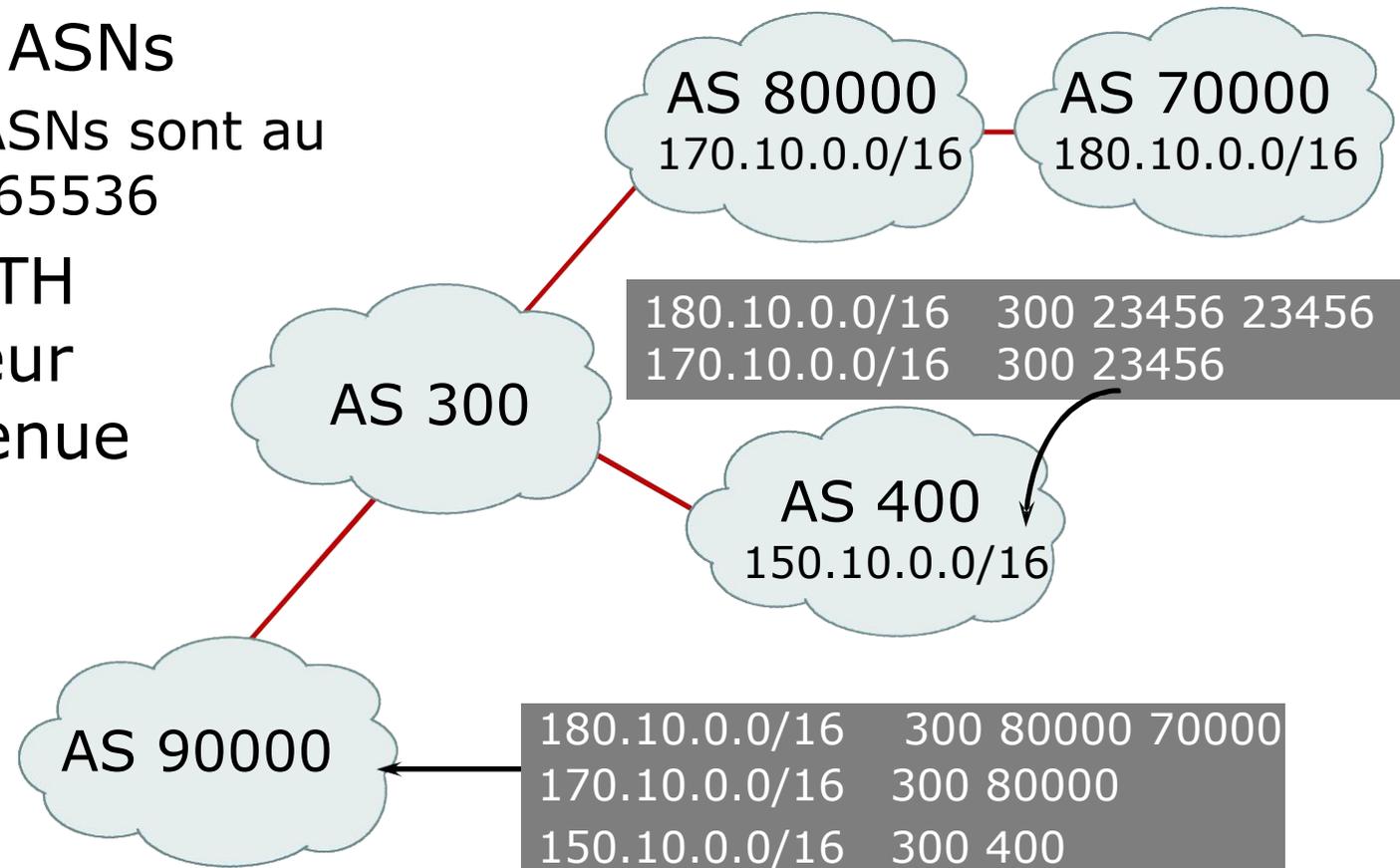
- Séquence d'AS traversé pour joindre une destination
- Obligatoire et transitif
- Utiliser pour:
 - Détection de boucle
 - Appliquer une politique de routage



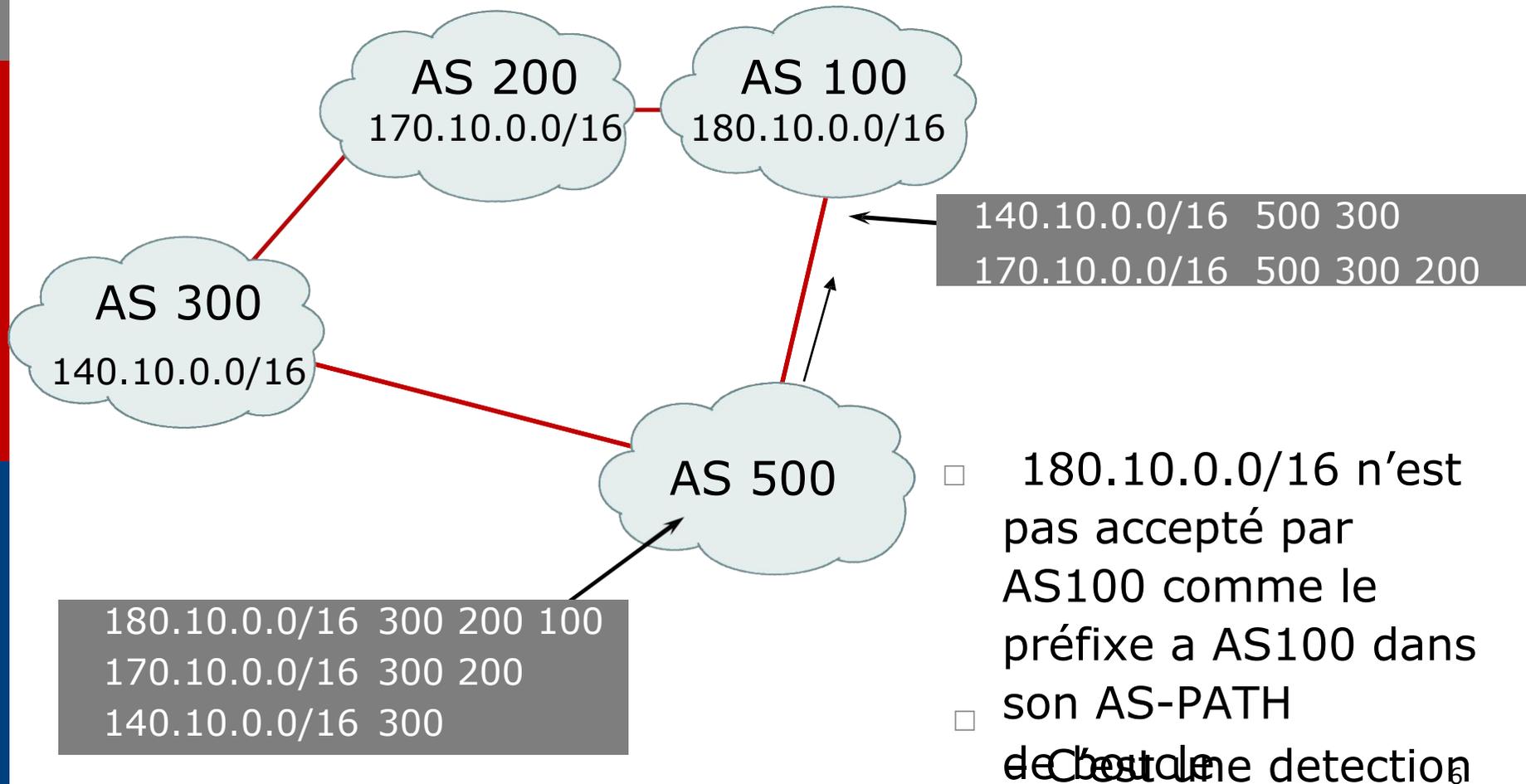
AS-Path (avec 16 et 32-bit ASNs)

- Internet avec 16-bit et 32-bit ASNs

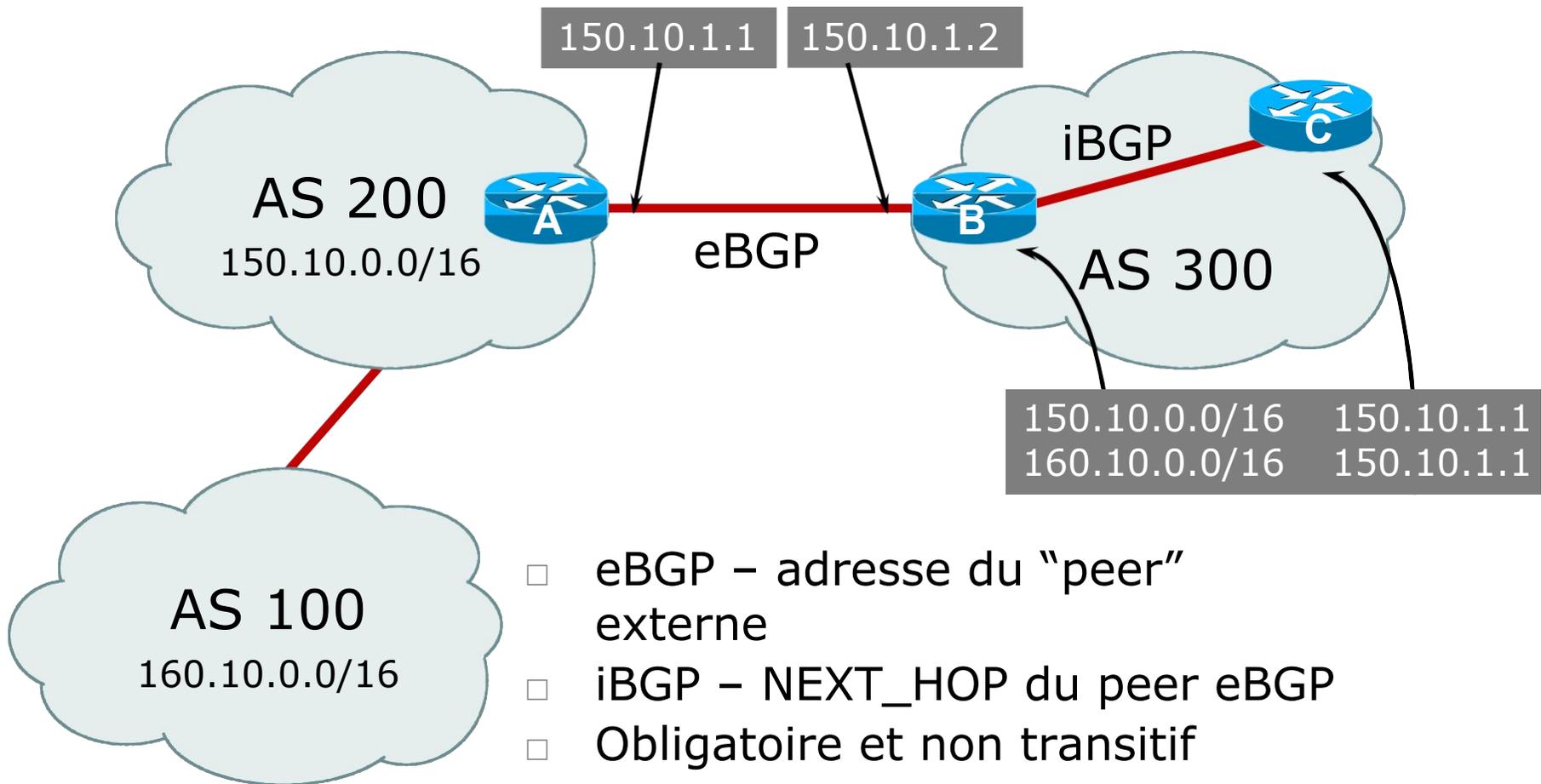
- 32-bit ASNs sont au dessus 65536
- AS-PATH longueur maintenue



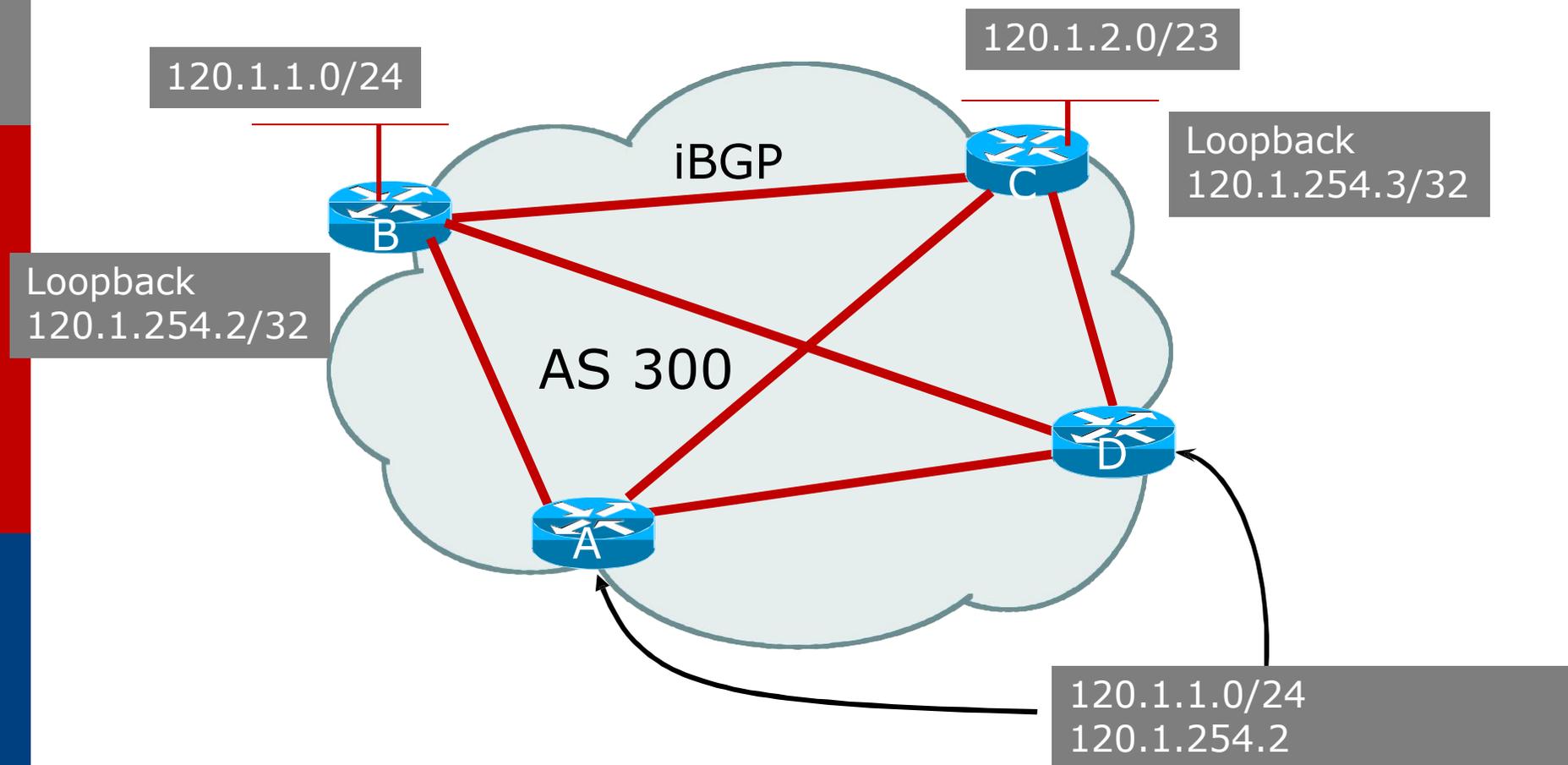
Detection de boucle AS-Path



Next Hop

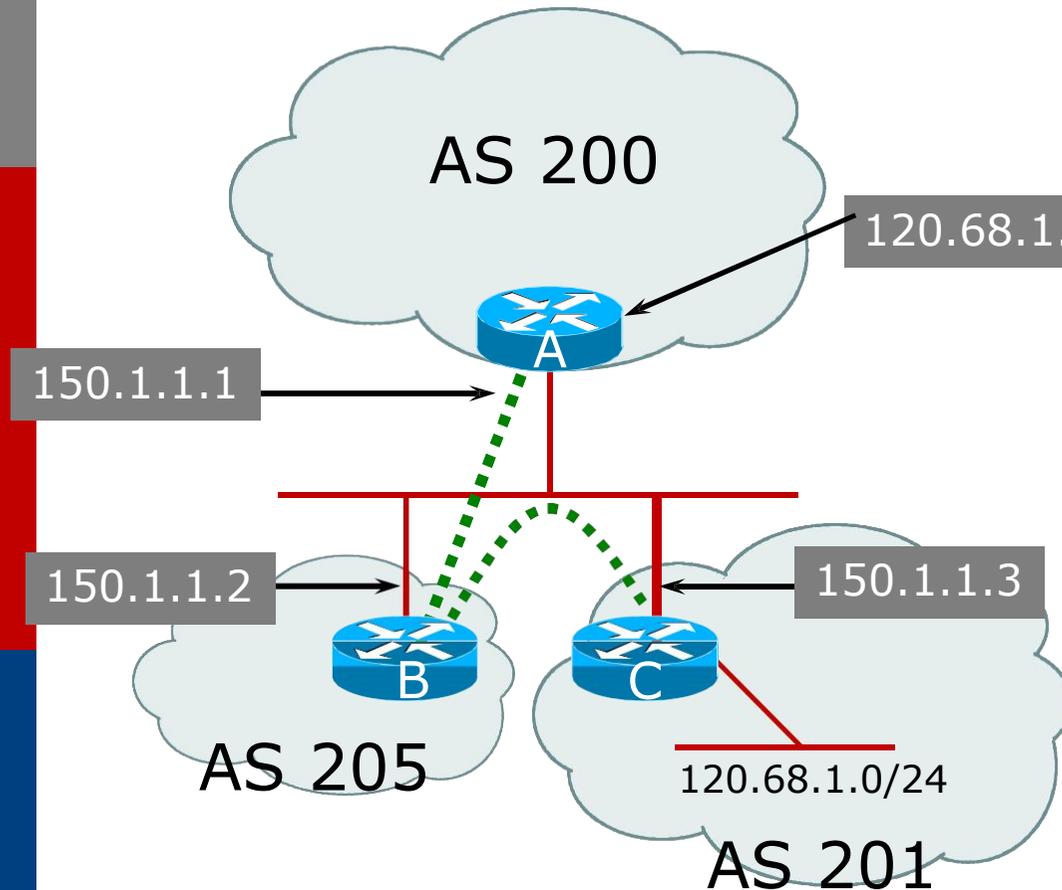


iBGP Next Hop



- Next-hop est la loopback adresse du routeur ibgp
- **Vérification de route recursive**

Next Hop



- eBGP entre Router A et Router B
- eBGP entre Router B et router C
- 120.68.1/24 a pour next Hop l'adresse 150.1.1.3 – Utilisé ici par Router A au lieu de 150.1.1.2 puisque les routeurs sont sur le même sous réseau
- Plus optimal configuration
- Requier aucune

Next Hop Best Practice

- Par défaut Cisco IOS propage le next-hop externe sans être modifié (eBGP) aux peers iBGP.
 - Ce qui signifie que l'IGP doit transporter ce Next-hop
 - En cas d'omission le réseau externe restera inaccessible
 - Avec plusieurs peers eBGP , C'est une charge supplémentaire pas forcément nécessaire sur l'IGP
- La bonne pratique est de changer le next-hop externe en le remplaçant par celui du routeur local

```
neighbor x.x.x.x next-hop-self
```

Next Hop (Summary)

- ❑ IGP doit transporter la route vers les next hops
- ❑ Verification de route Recursive
- ❑ Utiliser “next-hop-self” pour les next-hops externes
- ❑ Permet à l'IGP de prendre des decisions intelligentes de forwarding

Origin

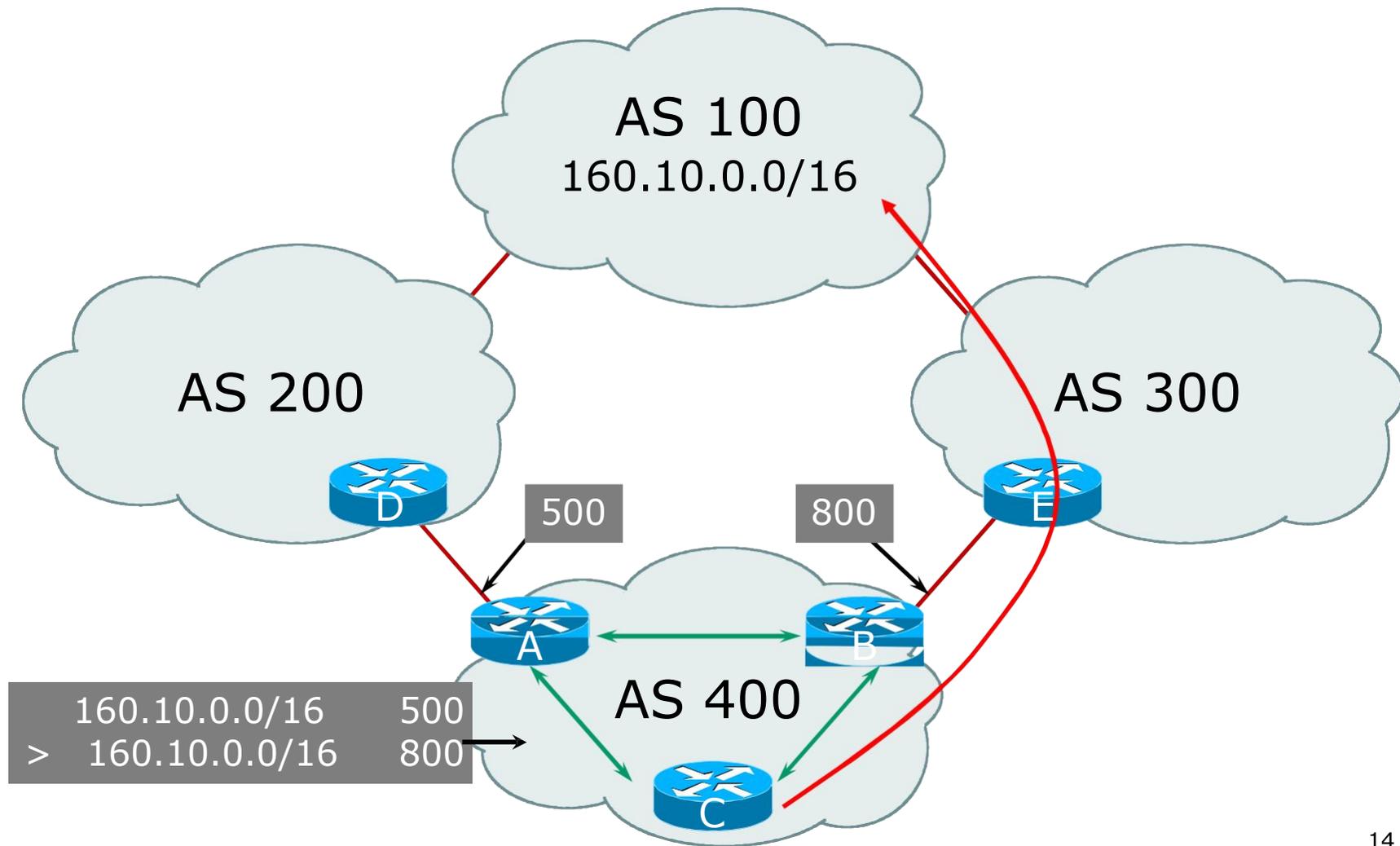
- Transporte l'origine d'un préfixe
- Attribut Historique
- Utilisé dans la transition d'EGP vers BGP
- Attribut Transitif et obligatoire
- Influence la sélection des meilleures routes
- Trois valeurs: IGP, EGP, incomplete
 - IGP – generé par la commande "Network" par BGP
 - EGP – generé par EGP
 - incomplete – redistribué d'un autre protocole de routage

Aggregator

- Transporte l'adresse IP du routeur ou du peer BGP qui a généré la route agrégée
- Attribut optionnel et transitif
- Utilé pour des déboggages
- N'a aucune influence sur la sélection des meilleurs chemins.
- Création de route agrégée en utilisant la commande "aggregate-address" active l'attribut aggregator

```
router bgp 100
  aggregate-address 100.1.0.0 255.255.0.0
```

Local Preference



Local Preference

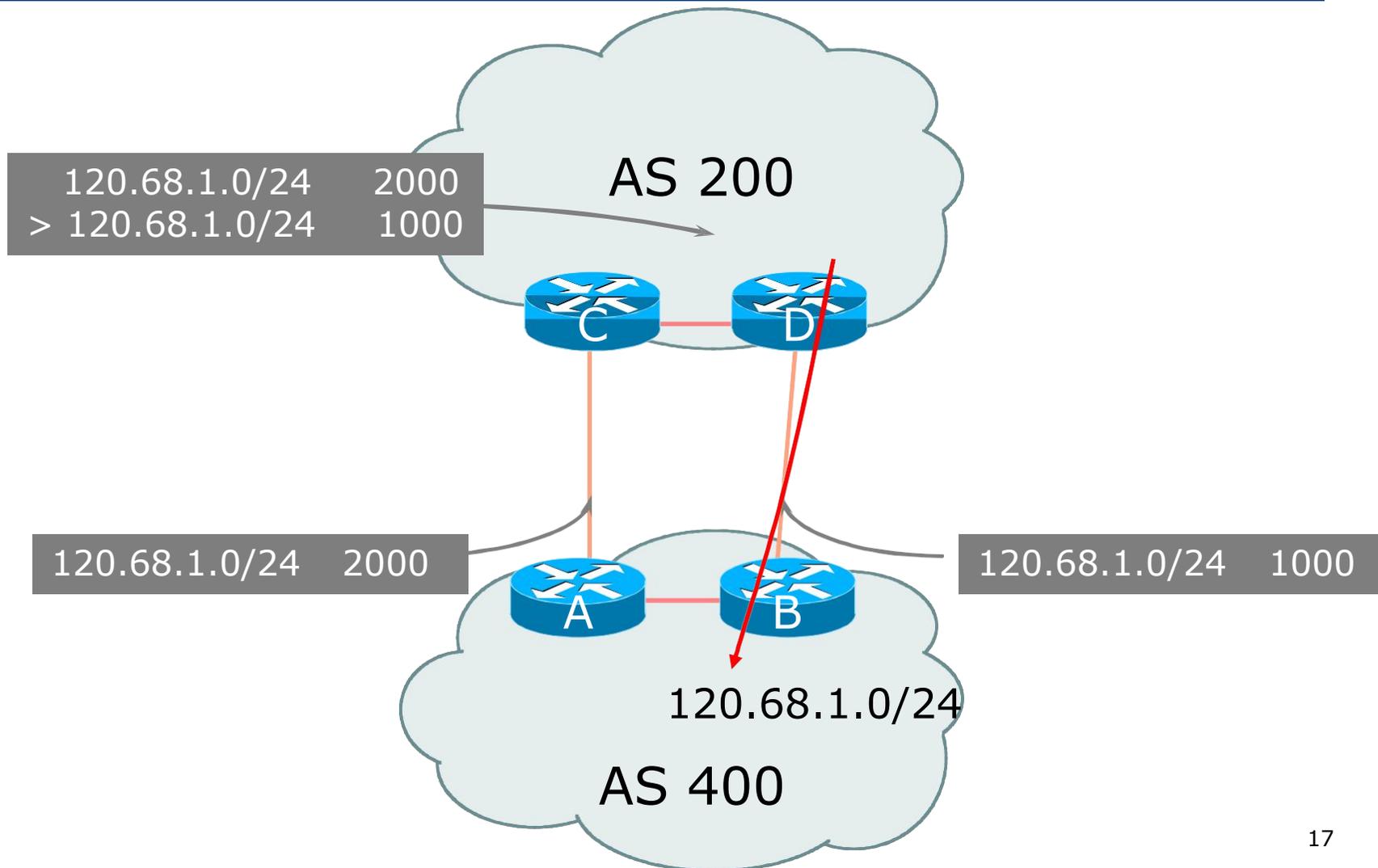
- Attribut non-transitif et optionel attribut
- Local à un AS uniquement
 - Valeur par Default "local preference" est 100 (IOS)
- Utiliser pour influencer la selection de route
 - Determine le meilleur chemin pour le trafic *outbound*
- Les chemins avec le plus grand local reference l'emporte

Local Preference

- Configuration du Router B:

```
router bgp 400
  neighbor 120.5.1.1 remote-as 300
  neighbor 120.5.1.1 route-map local-pref in
  !
  route-map local-pref permit 10 match ip
    address prefix-list MATCH set
    local-preference 800
  route-map local-pref permit 20
  !
  ip prefix-list MATCH permit 160.10.0.0/16
```

Multi-Exit Discriminator (MED)



Multi-Exit Discriminator

- Inter-AS – non-transitif et optionel attribut
- Utilisé pour choisir les préférences relatives aux points d'entrée
 - determine le meilleur chemin pour le trafic inbound
 - Comparable si les chemins proviennent du même AS
 - `bgp always-compare-med` permet de comparer des MED entre different AS
- Le chemin avec le plus petit MED l'emporte
- Absence d'attribut MED implique que le MED a une valeur **zero** (RFC4271)

MED déterministe

- IOS compare les chemins dans l'ordre de réception.
 - Conduit à des décisions inconsistantes lorsque les MED sont comparés
- MED Déterministe
 - Configurer sur tous les "peer" BGP au sein de l'AS
 - Ordonne les chemins selon l'ASN
 - Le meilleur chemin pour chaque group d'ASN est choisi
 - Tous les meilleurs chemins sont choisis parmi les bestpath selected from the winners of each group

```
router bgp 100
  bgp deterministic-med
```

MED & IGP Metric

- La metrique IGP peut être transporté comme un MED
 - **set metric-type internal** dans route-map
 - Activer BGP pour annoncer un MED qui correspond à la valeur du metrique de l'IGP
 - Changements sont monitorés (et annoncés à nouveau si necessaire) tous les 60s
 - **bgp dynamic-med-interval <secs>**

Multi-Exit Discriminator

□ Configuration of Router B:

```
router bgp 400
  neighbor 120.5.1.1 remote-as 200
  neighbor 120.5.1.1 route-map set-med out
!
route-map set-med permit 10
  match ip address prefix-list MATCH set
  metric 1000
route-map set-med permit 20
!
ip prefix-list MATCH permit 120.68.1.0/24
```

Weight

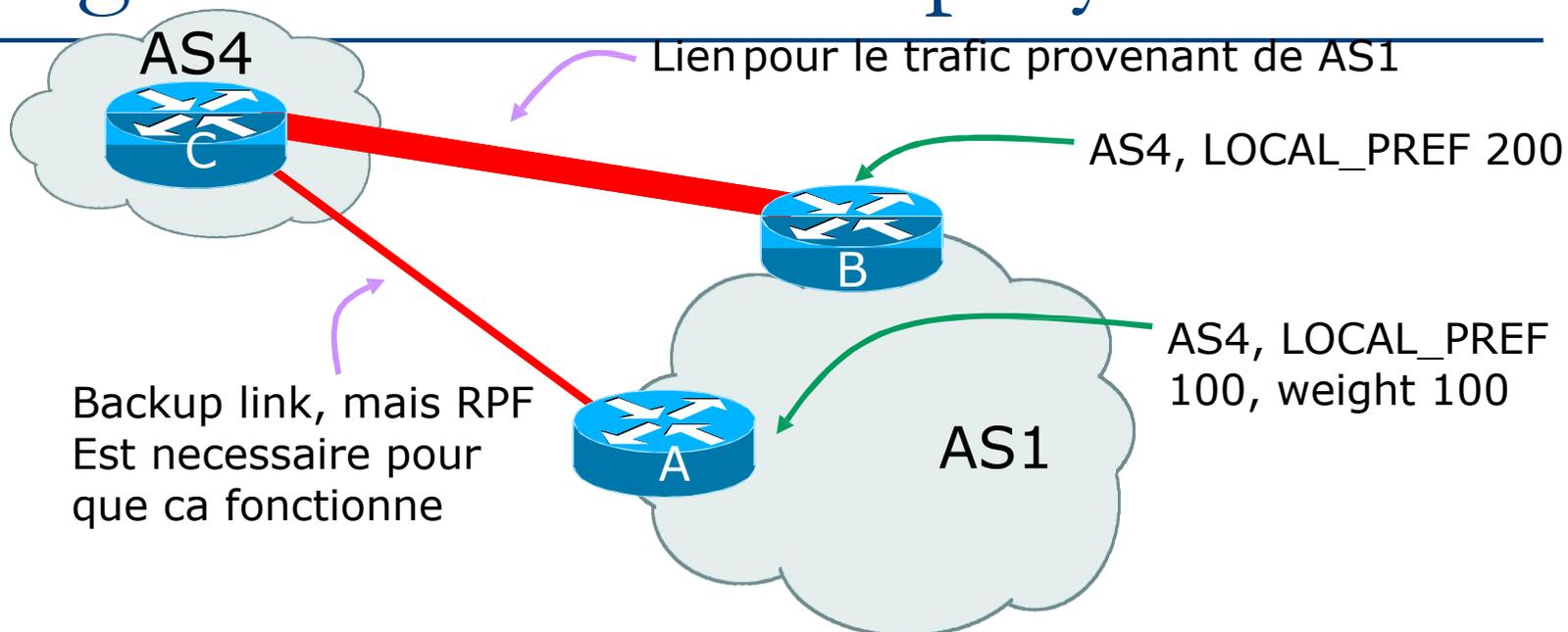
- Pas vraiment un attribut – local au routeur
- Le plus grand “weight” remporte
- Appliqué à toutes les routes reçues d’un peer BGP

- `neighbor 120.5.7.1 weight 100`

- Weight sont affectés aux routes en fonction des filtres

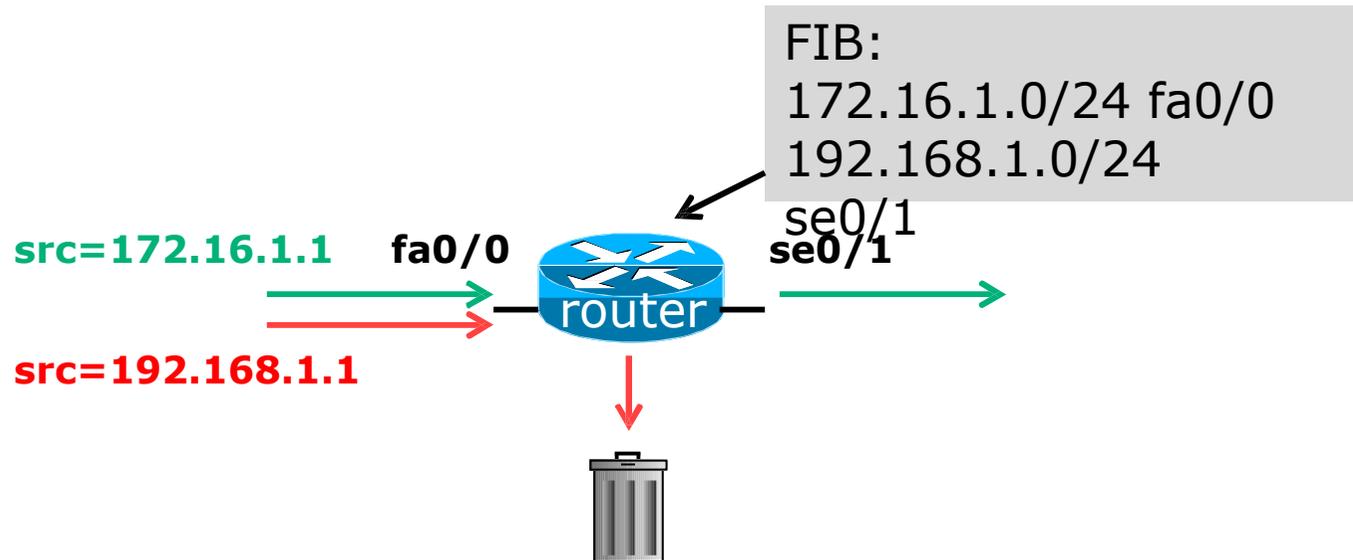
```
neighbor 120.5.7.3 filter-list 3 weight 50
```

Weight – Permet de deployer RPF



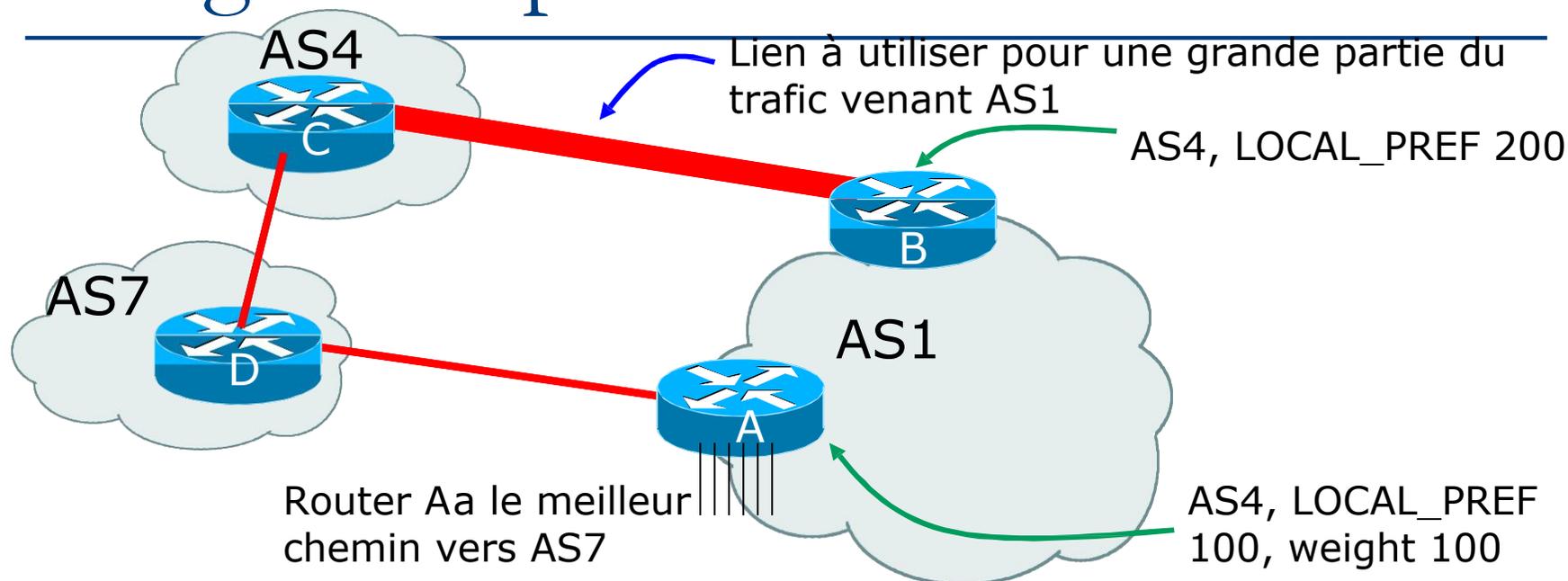
- Meilleur chemin vers AS4 en venant AS1 est toujours via B du fait de la "local-pref"
- Les paquets arrivant à A de AS4 via le lien direct C vers A passeront avec succès le test RPF du fait du Weight configuré
 - Si le weight n'était pas configuré le meilleur chemin aurait été via B et le test RPF aurait échoué

Qu'est ce que uRPF?



- Les routeurs comparent les sources d'adresse des paquets "inbound" avec l'entrée de la FIB
 - Si une entrée de la table FIB correspond à une interface inbound le paquet est transmis
 - Si l'entrée de la FIB ne correspond pas à une interface inbound le paquet est supprimé

Weight – Optimisation du trafic

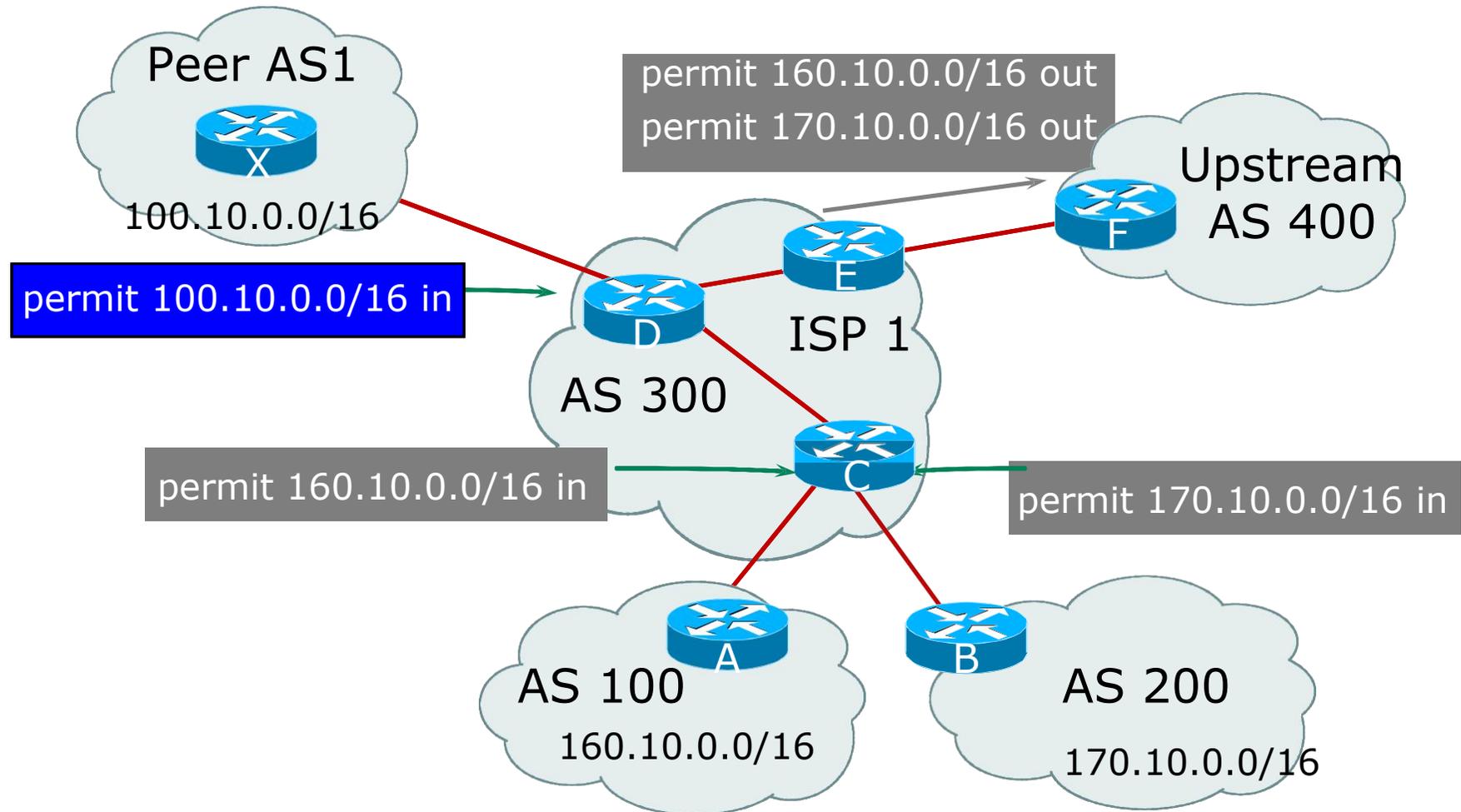


- Le meilleur chemin vers AS4 venant d'AS1 est toujours via B du fait de la "**local-pref**"
- Mais les clients connectés directement au Router A utilisent le lien vers AS7 comme meilleur "**outbound**" chemin du fait du plus grand "**weight**" appliqué aux routes apprises de AS7
 - Si le lien A vers D est défaillant, alors les clients de Router A auront

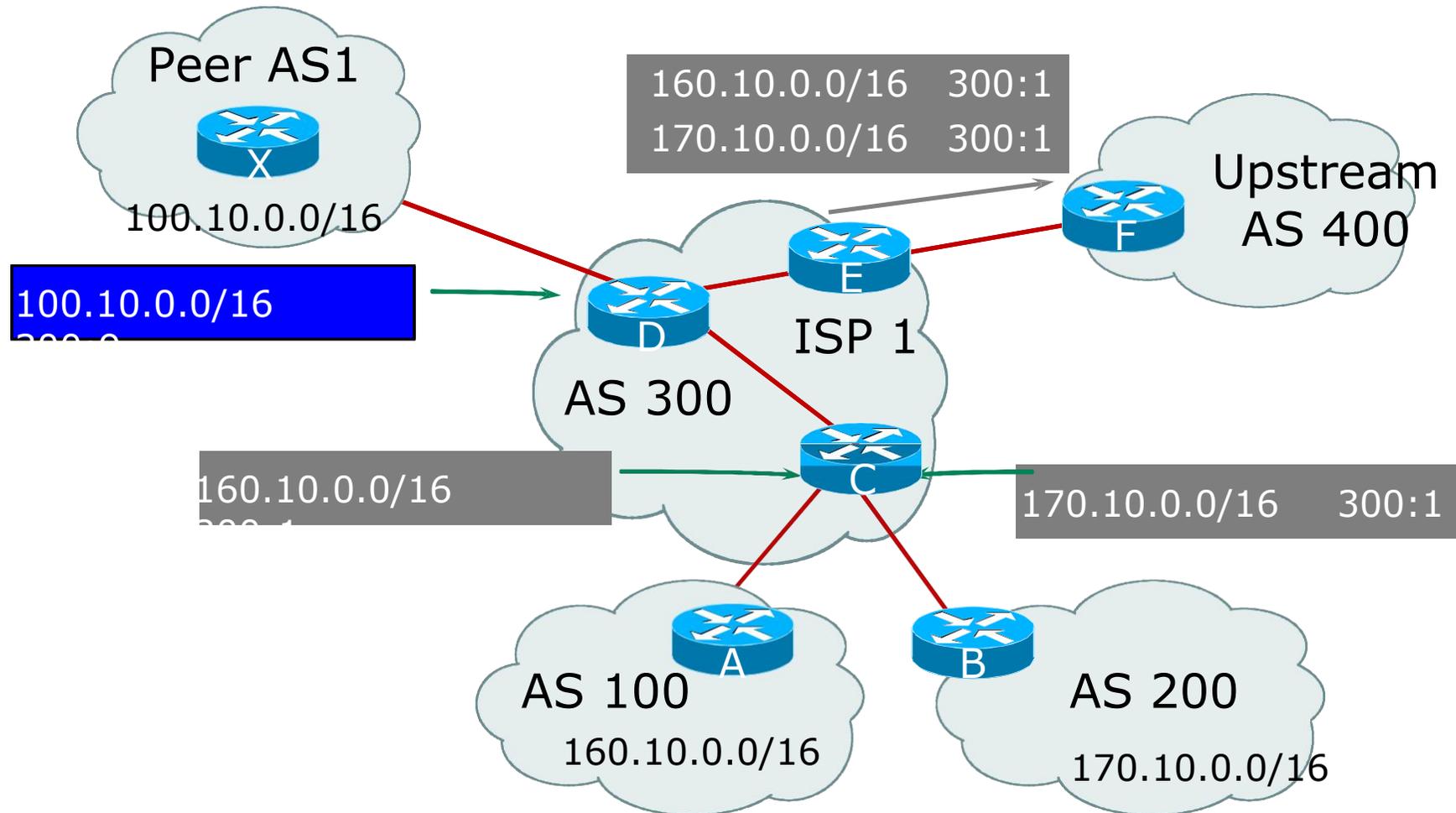
Community

- Les Communauté sont décrites dans le RFC1997
 - Attribut Transitif et Optional
- Entier de 32 bit
 - Representation en deux entiers de 16 bit (RFC1998)
 - Format standard <local-ASN>:xx
 - 0:0 à 0:65535 et 65535:0 à 65535:65535 est réservé
- Utilisé pour regrouper les destinations
 - Chaque destination peut être membre de plusieurs communauté
- Très utile pour appliquer des politiques de routage entre AS

Exemple de Community (avant)



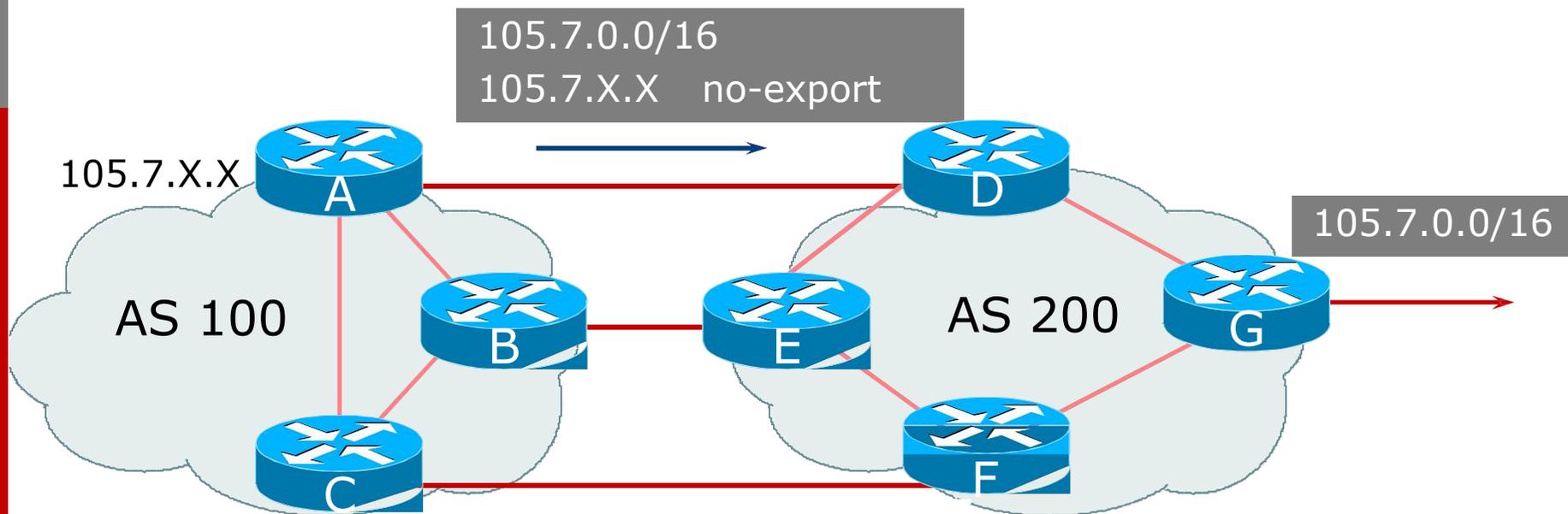
Exemple de Community (Après)



Communauté “Well-Known”

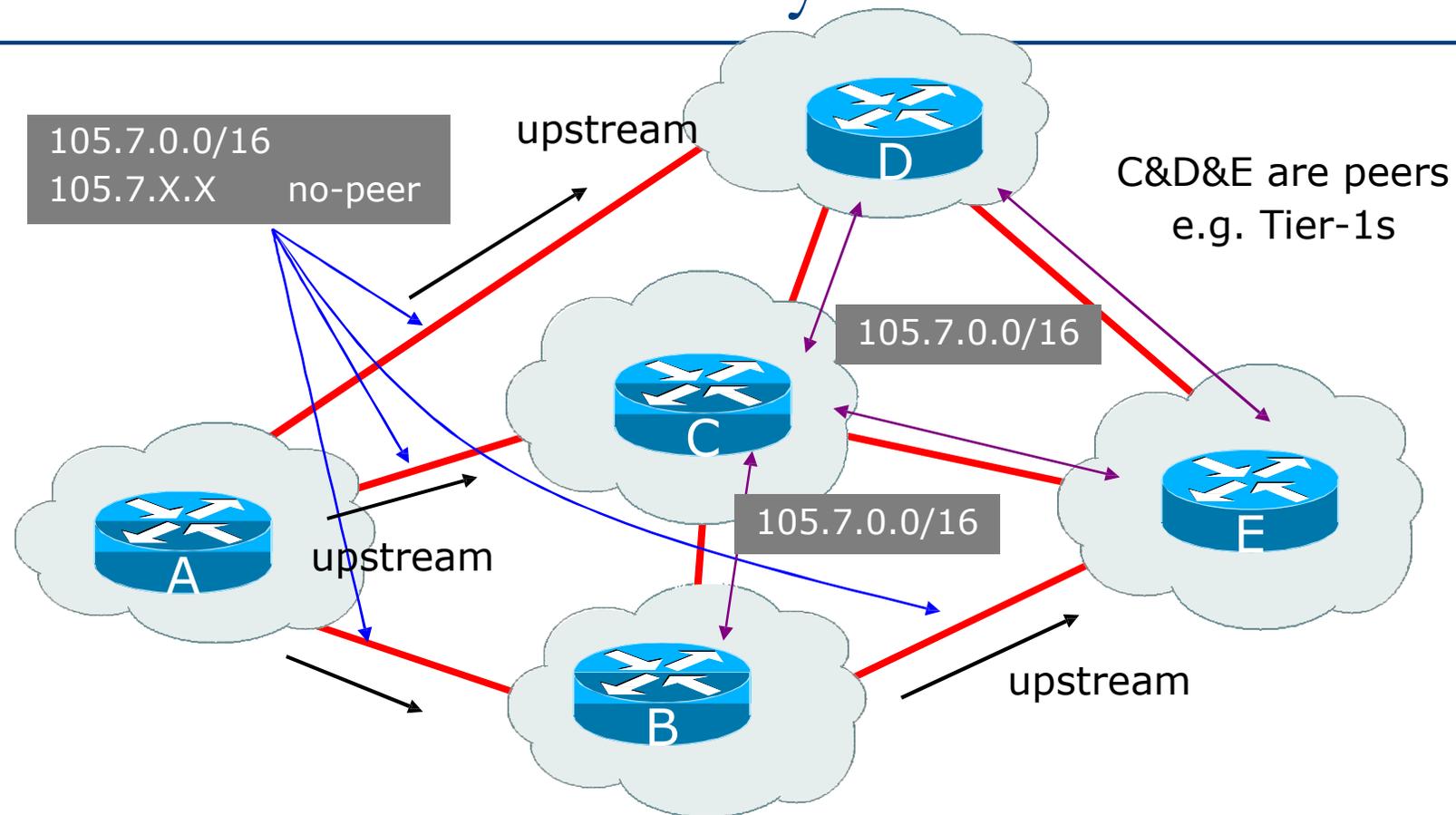
- Plusieurs communauté “well known”
 - www.iana.org/assignments/bgp-well-known-communities
- no-export 65535:65281
 - Ne pas annoncé aux autres peer eBGP
- no-advertise 65535:65282
 - Ne pas annoncé aux autres peer BGP
- no-export-subconfed 65535:65283
 - Ne pas annoncé hors du local AS (Utilisé dans les confederation)
- no-peer 65535:65284
 - Ne pas annoncé aux peer bi-lateral (RFC3765)

No-Export Community



- AS100 annonce la route agrégée et subprefixes
 - Intention est d'améliorer le partage de charge en bloquant le subprefix
 - Subprefixes marqué avec **no-export** community
- Router G dans AS200 n'annonce pas le préfixe avec la communauté marqué **no-export**

No-Peer Community



- Les Sub-prefixes marqués avec "no-peer" community ne sont pas envoyés aux peer bi-lateral
 - Ils sont seulement envoyés aux upstream provider

What about 4-byte ASNs?

- Communauté sont largement utilisée pour définir la politique de routage
 - 32 bit attribute
- Le format défini par le RFC1998 est maintenant un "standard"
 - ASN:number
- Adapté pour les ASN de 2-byte ASNs, mais ne peut être encode pour les ASN de 4-byte ASNs
- Solutions:
 - Utilisé "private ASN" for the first 16 bits
 - Attendre <http://datatracker.ietf.org/doc/draft-ietf-idr-as4octet-extcomm-generic-subtype/> soit implémenté

Summary

Attributes in Action

```
Router6>sh ip bgp
```

```
BGP table version is 16, local          ID is 10.0.15.246
```

```
router
```

```
Status codes: s suppressed, d          h history, * valid, > best, - internal,
damped,          x best-external, a additional-path, c
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete, m multipath, b backup-path, RT-Filter,
```

```
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*>i 10.0.0.0/26	10.0.15.241	0	100	0	i
*>i 10.0.0.64/26	10.0.15.242	0	100	0	i
*>i 10.0.0.128/26	10.0.15.243	0	100	0	i
*>i 10.0.0.192/26	10.0.15.244	0	100	0	i
*>i 10.0.1.0/26	10.0.15.245	0	100	0	i
*> 10.0.1.64/26	0.0.0.0	0		32768	i
*>i 10.0.1.128/26	10.0.15.247	0	100	0	i
*>i 10.0.1.192/26	10.0.15.248	0	100	0	i
*>i 10.0.2.0/26	10.0.15.249	0	100	0	i
*>i 10.0.2.64/26	10.0.15.250	0	100	0	i

```
...
```

BGP Path Selection Algorithm



Pourquoi est-il le meilleur chemin ?

Algorithm de selection de route pour Cisco IOS: Partie 1

1. Le chemin est ignoré si le next hop n'est pas joignable
2. Les chemins IBGP sont ignorés lorsqu'il ne sont pas synchronisés (Cisco IOS)
3. Le plus grand "weight" (local au routeur)
4. La plus grande "local preference" (au sein de l'AS)
5. Préférence sur routes "locally originated"
6. Plus petit "AS path"

Algorithm de selection de route pour Cisco IOS: Partie 2

7. La plus petite "origin code"
 - IGP < EGP < incomplete
8. La plus petite Multi-Exit Discriminator (MED)
 - si **bgp deterministic-med**, alors ordonner les chemins par AS number avant la comparaison
 - si **bgp always-compare-med**, alors comparer tous les chemins
 - Dans les autres cas MED seul considéré si les chemins proviennent du même AS (default)

Algorithm de selection de route pour Cisco IOS: Partie 3

9. Les routes eBGP sont préférées aux routes apprises par iBGP
10. Les routes avec le plus petit métrique IGP vers le next-hop
11. Pour les routes eBGP :
 - Si "multipath" est activé, installation de N routes parallèles dans la table de forwarding
 - Si le "router-id" est le même aller à l'étape suivante
 - Si le "router-id" ne sont pas identiques, sélectionner la route la plus agée

Algorithm de selection de route pour Cisco IOS: Partie 4

12. Plus petit "router-id" ("originator-id" pour les routes reflecteur)
13. Plus petit cluster-list
 - Les Clients doivent avoir connaissance des attributs du Route Reflector!
14. Plus petite adresse du "peer"

Attributs BGP et Sélection de Chemin



SIF
AFNOG 2018, Dakar