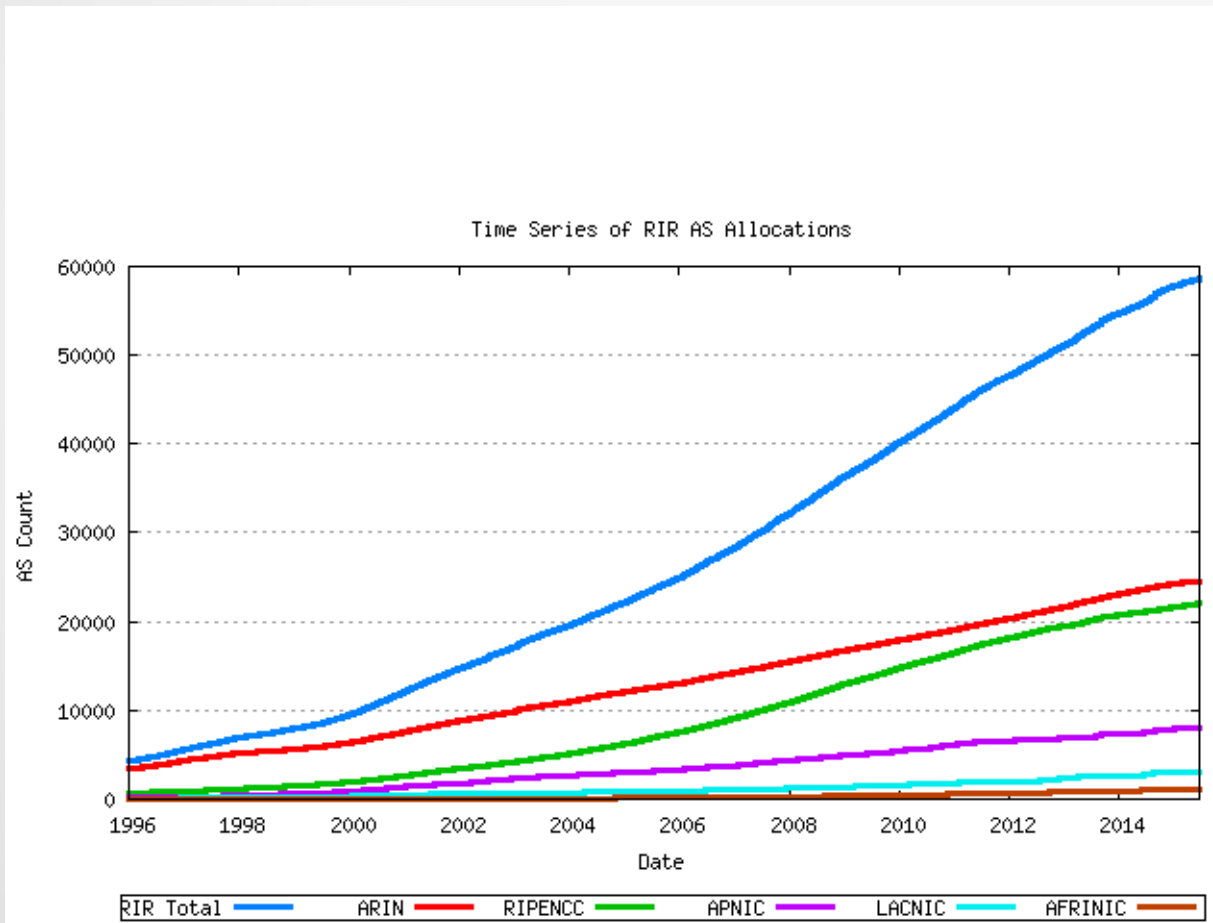


Introduction

- Routage inter AS
 - Utilisé pour router les informations entre les AS donc principalement routage « Internet » (ou inter domain routing), et nécessite de posséder un AS :
 - Opérateurs
 - Services spécifiques (multi homing) hébergeur de services.
 - 16bits puis 32bits

Introduction

- Routage inter AS



Source : iana.org

BGP : Historique

- Historique :
 - ARPANET : Gateway to Gateway protocol (RFC 823)
 - 1984 : Exterior Gateway Protocol (RFC 904), structure en arbre.
 - 1989 : BGP v1 (RFC 1105)
 - 1990 : BGP v2 (RFC 1163)
 - 1991 : BGP v3 (RFC1267)
 - 1994 : BGP v4 (RFC1771) (update RFC4271)
 - Others for MP-BGP BGP for IPv6....

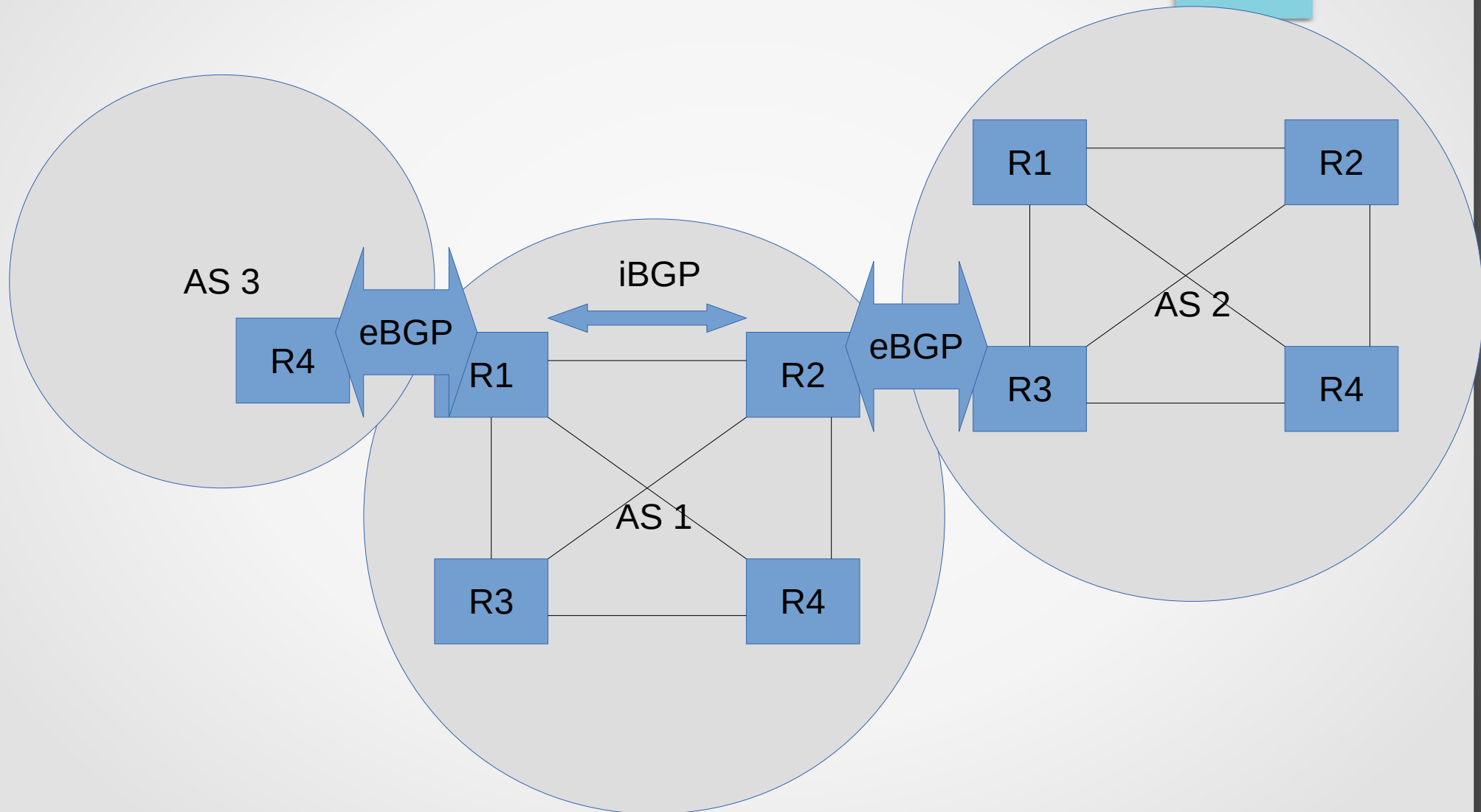
BGP : Fonctionnement

- Encapsulation et adressage :
 - BGP est un protocole utilisant la couche transport TCP (port 179)
 - Inhabituel, car la plupart des autres protocoles de routage utilisent IP ou UDP pour pouvoir utiliser une diffusion de groupe (broadcast ou multicast)
 - BGP n'en a pas besoin car il ne découvre pas ses voisins, ils doivent être renseignés dans la configuration du routeur.
 - Et profite des services offerts par TCP/IP (respect de l'ordre d'envoi, fragmentation, retransmission, etc..)
 - Attention : TCP permet le transport d'un flux de données, pas très adapté à une communication par messages.

BGP : Fonctionnement

- BGP est un protocole de routage de type PATH vector, le PATH est constitué des AS traversés pour rejoindre un subnet.
- Traffic engineering pour définir des politiques de routage :
 - Accord de type Peering/Traffic entre opérateurs
 - Eviter/favoriser certains AS
- BGP se compose de deux parties
 - eBGP (entre des routeurs d'AS différents)
 - iBGP (entre les routeurs d'un même AS)

BGP : Fonctionnement



BGP : Fonctionnement

- eBGP :
 - Les routeurs doivent être directement connectés (sur le même réseau niveau 2)
 - Uniquement entre des routeurs d'AS différents
 - Ne doit pas y avoir d'IGP entre ces routeurs

BGP : Fonctionnement

- iBGP :
 - Les routeurs NE doivent PAS FORCEMENT être directement connectés (l'IGP permet de les «relier»)
 - Uniquement entre des routeurs d'un même AS
 - Tous les routeurs iBGP doivent être reliés (full mesh)
 - Relayent les prefixs appris depuis l'exterieur de l'AS
 - Ne relayent pas les prefixs appris depuis un pair iBGP

BGP : Messages

- Format des messages (dans un flux de données TCP)

4 types de messages différents

- OPEN
- UPDATE
- NOTIFICATION
- KEEPALIVE

Un « entête » commun

- Marker
- Length
- Type

Wireshark capture showing BGP messages. The packet list displays the following data:

No.	Time	Source	Destination	Protocol	Length	Info
5	7.999977	192.168.0.15	192.168.0.33	TCP	74	elatelink > bgp [SYN]
6	8.003909	192.168.0.33	192.168.0.15	TCP	60	bgp > elatelink [SYN]
7	8.003954	192.168.0.15	192.168.0.33	TCP	60	elatelink > bgp [ACK]
8	8.004042	192.168.0.15	192.168.0.33	BGP	83	OPEN Message
9	8.208048	192.168.0.33	192.168.0.15	TCP	60	bgp > elatelink [ACK]
10	8.337997	192.168.0.33	192.168.0.15	BGP	83	OPEN Message
11	8.338027	192.168.0.15	192.168.0.33	TCP	60	elatelink > bgp [ACK]
12	8.338115	192.168.0.15	192.168.0.33	BGP	73	KEEPALIVE Message
13	8.342206	192.168.0.33	192.168.0.15	BGP	73	KEEPALIVE Message
14	8.349836	192.168.0.15	192.168.0.33	TCP	60	elatelink > bgp [ACK]
15	8.544101	192.168.0.33	192.168.0.15	TCP	60	bgp > elatelink [ACK]

Details of the selected packet (Frame 8):

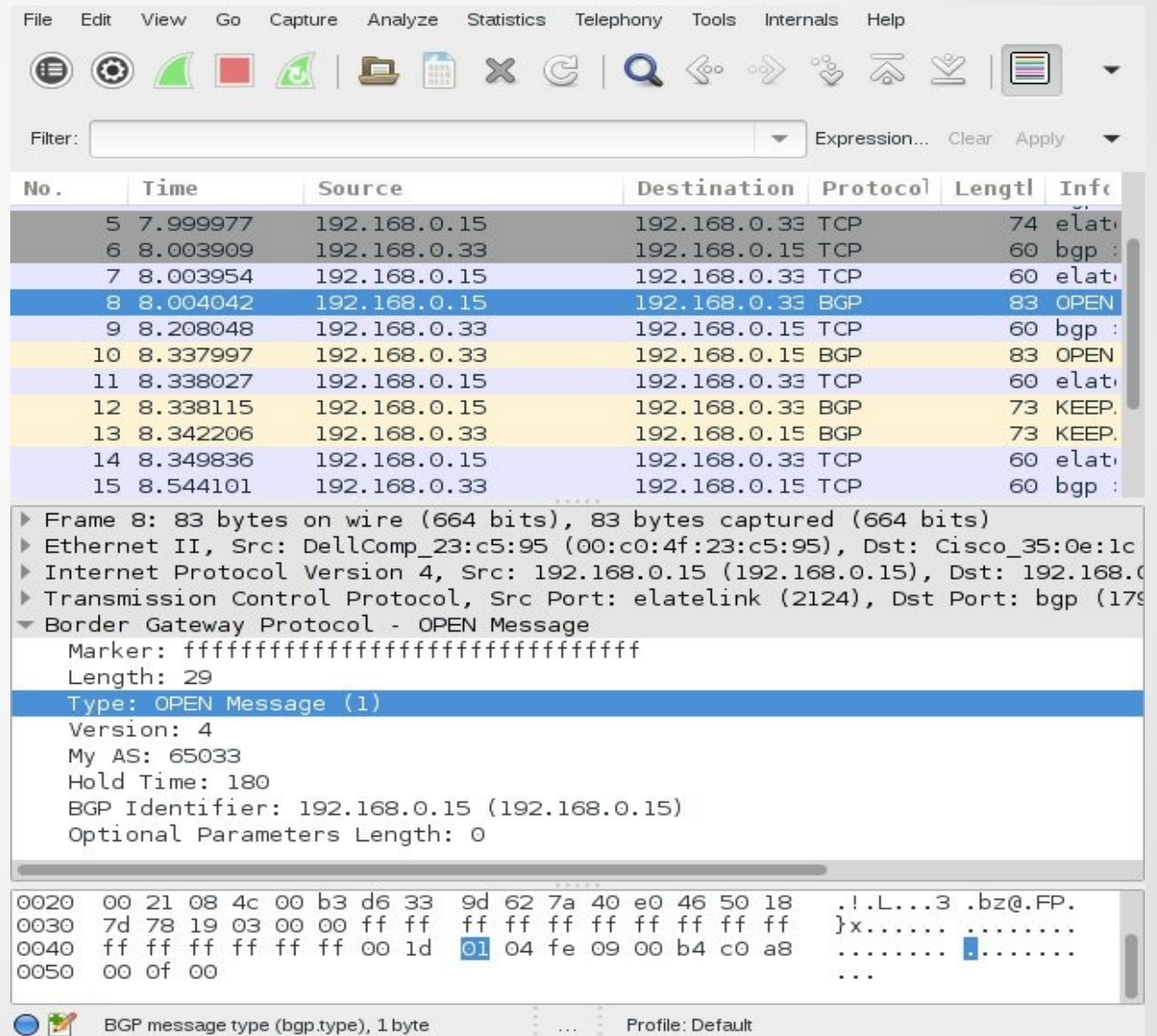
- Frame 8: 83 bytes on wire (664 bits), 83 bytes captured (664 bits)
- Ethernet II, Src: DellComp_23:c5:95 (00:c0:4f:23:c5:95), Dst: Cisco_35:0e:1c (00:00:0c:35:0e:1c)
- Internet Protocol Version 4, Src: 192.168.0.15 (192.168.0.15), Dst: 192.168.0.33 (192.168.0.33)
- Transmission Control Protocol, Src Port: elatelink (2124), Dst Port: bgp (179), Seq: 1, Ack: 1, Len: 83
- Border Gateway Protocol - OPEN Message
 - Marker: ffffffffffffffffffffffffffffffffff
 - Length: 29
 - Type: OPEN Message (1)
 - Version: 4
 - My AS: 65033
 - Hold Time: 180
 - BGP Identifier: 192.168.0.15 (192.168.0.15)
 - Optional Parameters Length: 0

Packet bytes (hex):

```
0020 00 21 08 4c 00 b3 d6 33 9d 62 7a 40 e0 46 50 18 .!.L...3 .bz@.FP.
0030 7d 78 19 03 00 00 ff ff ff ff ff ff ff ff ff }x.....
0040 ff ff ff ff ff ff 00 1d 01 04 fe 09 00 b4 c0 a8 .....
0050 00 0f 00
```

BGP : Message OPEN

- Message envoyé à l'ouverture de la connexion
- Version
- N° AS
- Hold time (temps max entre 2 keepalives)
- BGP Identifier (identifiant de routeur ex : @IP)



The image shows a Wireshark network packet capture. The top toolbar includes menus like File, Edit, View, Go, Capture, Analyze, Statistics, Telephony, Tools, Internals, and Help. Below the toolbar is a filter field and a list of captured packets. Packet 8 is selected, showing details for a BGP OPEN message. The details pane shows the message structure: Marker (29 bytes), Type (OPEN Message (1)), Version (4), My AS (65033), Hold Time (180), BGP Identifier (192.168.0.15), and Optional Parameters Length (0). The bottom pane shows the raw packet data in hexadecimal and ASCII.

No.	Time	Source	Destination	Protocol	Length	Info
5	7.999977	192.168.0.15	192.168.0.33	TCP	74	elate
6	8.003909	192.168.0.33	192.168.0.15	TCP	60	bgp :
7	8.003954	192.168.0.15	192.168.0.33	TCP	60	elate
8	8.004042	192.168.0.15	192.168.0.33	BGP	83	OPEN
9	8.208048	192.168.0.33	192.168.0.15	TCP	60	bgp :
10	8.337997	192.168.0.33	192.168.0.15	BGP	83	OPEN
11	8.338027	192.168.0.15	192.168.0.33	TCP	60	elate
12	8.338115	192.168.0.15	192.168.0.33	BGP	73	KEEP.
13	8.342206	192.168.0.33	192.168.0.15	BGP	73	KEEP.
14	8.349836	192.168.0.15	192.168.0.33	TCP	60	elate
15	8.544101	192.168.0.33	192.168.0.15	TCP	60	bgp :

Frame 8: 83 bytes on wire (664 bits), 83 bytes captured (664 bits)
Ethernet II, Src: DellComp_23:c5:95 (00:c0:4f:23:c5:95), Dst: Cisco_35:0e:1c
Internet Protocol Version 4, Src: 192.168.0.15 (192.168.0.15), Dst: 192.168.0.33
Transmission Control Protocol, Src Port: elatelink (2124), Dst Port: bgp (179)
Border Gateway Protocol - OPEN Message
Marker: ffffffffffffffffffffffffffffffffff
Length: 29
Type: OPEN Message (1)
Version: 4
My AS: 65033
Hold Time: 180
BGP Identifier: 192.168.0.15 (192.168.0.15)
Optional Parameters Length: 0

0020 00 21 08 4c 00 b3 d6 33 9d 62 7a 40 e0 46 50 18 .!.L...3 .bz@.FP.
0030 7d 78 19 03 00 00 ff ff ff ff ff ff ff ff }x.....
0040 ff ff ff ff ff ff 00 1d 01 04 fe 09 00 b4 c0 a8
0050 00 0f 00 ...

BGP message type (bgp.type), 1 byte Profile: Default

BGP : Message UPDATE

- Message envoyé pour échanger des informations de routage

The image shows a Wireshark packet capture of a BGP UPDATE message. The packet list at the top shows four packets: packet 16 is a BGP KEEPALIVE (270 bytes), packet 17 is the BGP UPDATE (118 bytes), packet 18 is a TCP segment (60 bytes), and packet 19 is another BGP KEEPALIVE (73 bytes). The details pane for packet 17 shows the following structure:

- Frame 17: 118 bytes on wire (944 bits), 118 bytes captured (944 bits)
- Ethernet II, Src: Cisco_35:0e:1c (00:00:0c:35:0e:1c), Dst: DellComp_23:c5:9
- Internet Protocol Version 4, Src: 192.168.0.33 (192.168.0.33), Dst: 192.168
- Transmission Control Protocol, Src Port: bgp (179), Dst Port: elatelink (21
- Border Gateway Protocol - UPDATE Message
 - Marker: ffffffffffffffffffffffffffffffffff
 - Length: 64
 - Type: UPDATE Message (2)
 - Unfeasible routes length: 0 bytes
 - Total path attribute length: 39 bytes
 - Path attributes
 - ORIGIN: EGP (4 bytes)
 - AS_PATH: empty (3 bytes)
 - NEXT_HOP: 192.168.0.33 (7 bytes)
 - MULTI_EXIT_DISC: 0 (7 bytes)
 - LOCAL_PREF: 100 (7 bytes)
 - COMMUNITIES: 65033:500 65033:600 (11 bytes)
 - Network layer reachability information: 2 bytes
 - 10.0.0.0/8

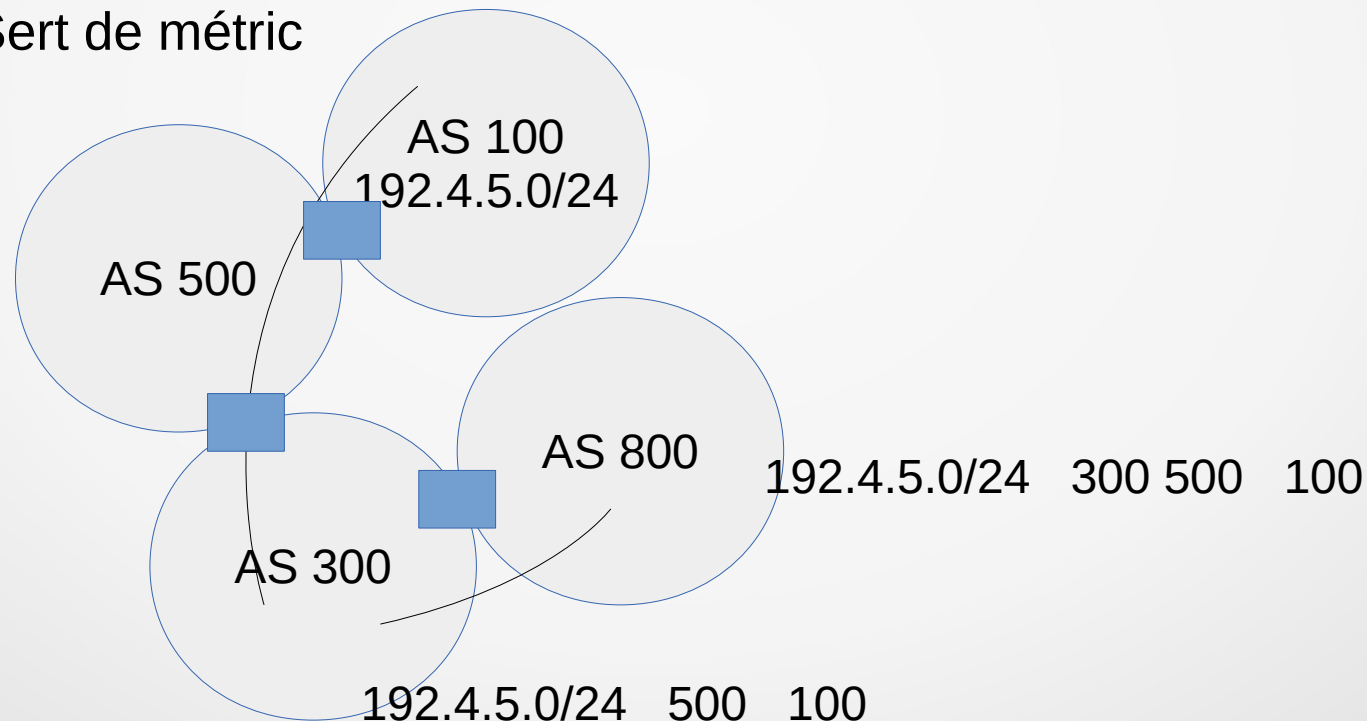
The packet bytes pane at the bottom shows the raw data of the packet, with the BGP UPDATE message structure visible in the hex and ASCII columns.

BGP : Message UPDATE

- Permettent de supprimer plusieurs routes
- Permettent d'ajouter une route v(un prefix)
 - Contient des attributs de plusieurs TYPES
 - Well-known mandatory.
 - Well-known discretionary.
 - Optional transitive.
 - Optional non-transitive.
 - Permet de savoir comment traiter des attributs mêmes s'ils sont inconnus

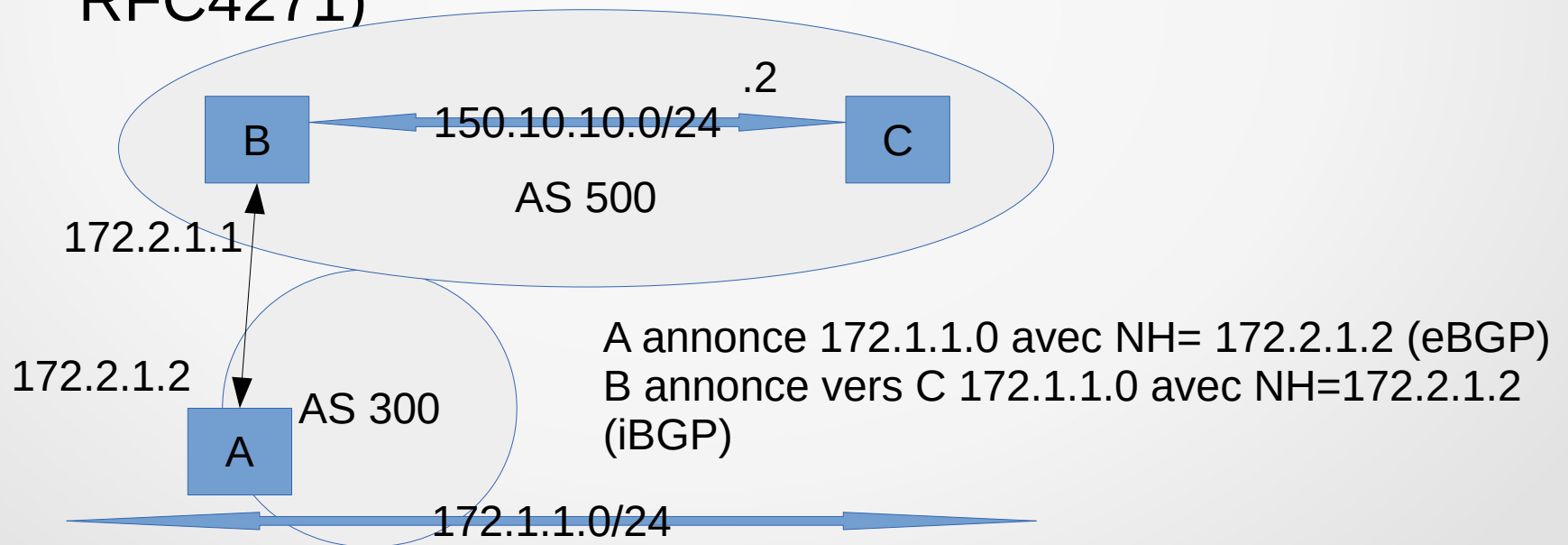
BGP : Attributs d'une route

- AS_PATH (Well-known, mandatory)
 - Séquence (ordonné) des AS qu'une route a traversés.
 - Evite les boucles (si mon AS dans liste des AS)
 - Sert de métrique



BGP : Attributs d'une route

- NEXT_HOP (Well-known mandatory)
 - L'adresse IP du routeur qui devrait être utilisée pour la route
 - Différents usages selon eBGP ou iBGP (Cf RFC4271)

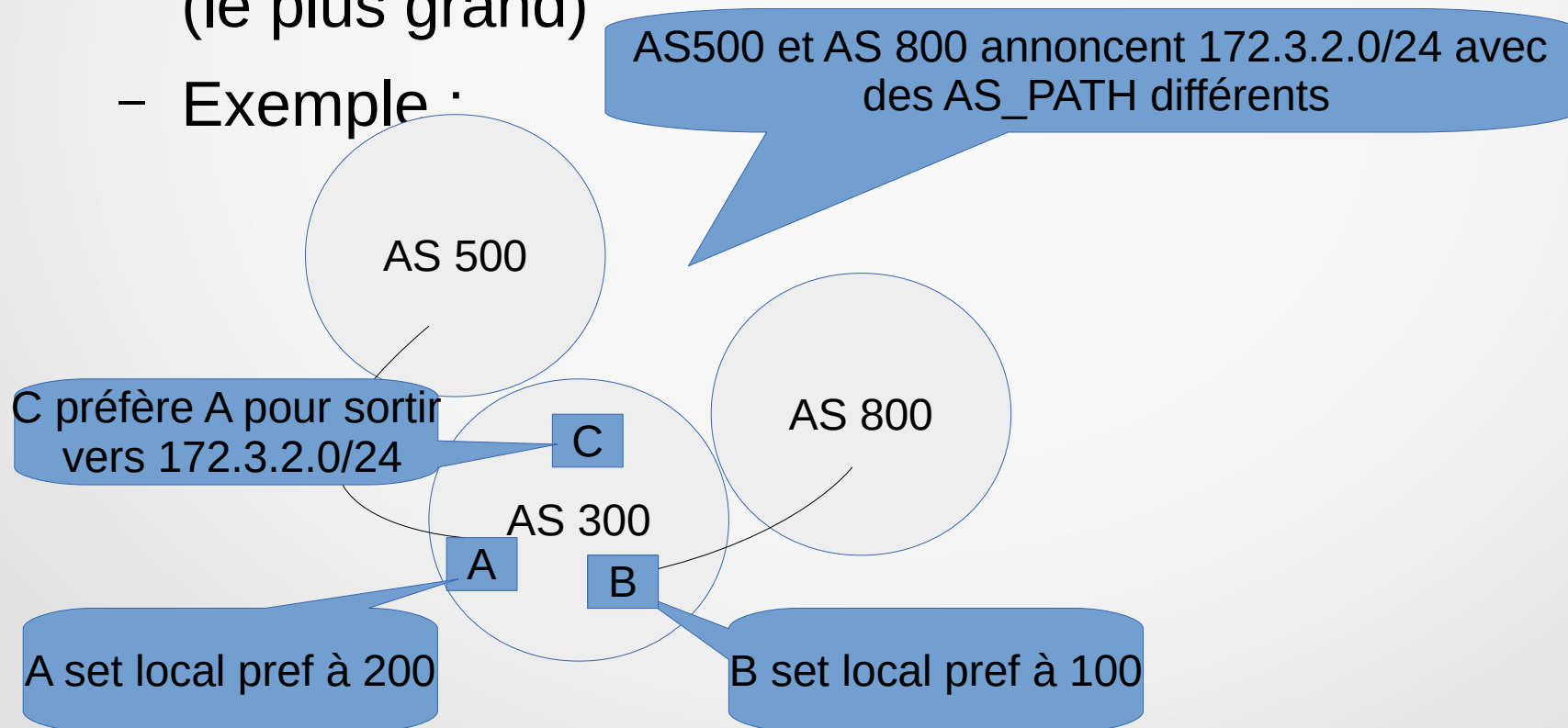


BGP : Attributs d'une route

- ORIGIN (Well-known mandatory)
 - Attribut historique (transition EGP/BGP)
 - Identifie l'origine de la route :
 - EGP
 - IGP (issu d'un AS)
 - Incomplete (redistribuée depuis un autre protocole de routage)
 - Peut influencer le mécanisme de sélection du meilleur chemin.

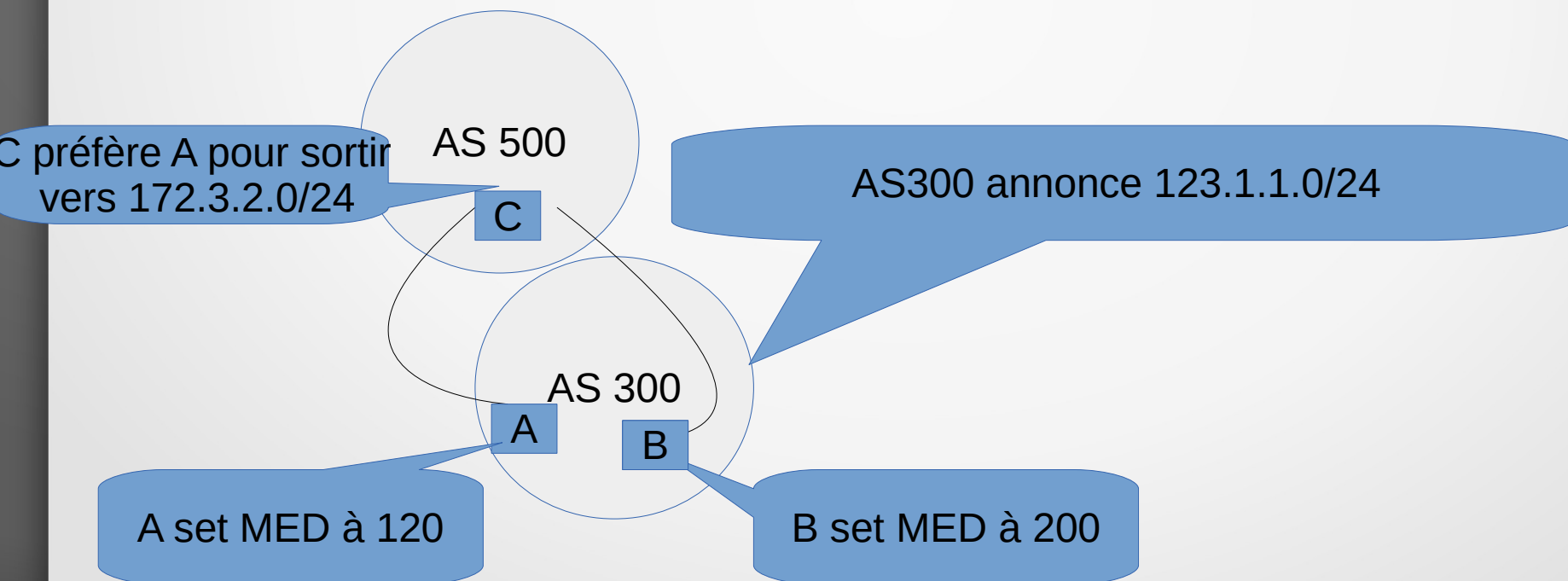
BGP : Attributs d'une route

- LOCAL_PREF (Well known)
 - Permet d'indiquer dans l'AS quel est le chemin privilégié pour sortir de cet AS vers le réseau considéré. (le plus grand)
 - Exemple :



BGP : Attributs d'une route

- MULTI_EXT_DESCRIPTOR (Optionnal, non transitive)
 - Attribut envoyé à un AS (mais pas propagé)
 - Permet d'influencer un AS voisin pour le choix (le plus petit)



BGP : Attributs d'une route

- COMMUNITY (Optional) – RFC 1997
 - Attribut (transitif) permettant de regrouper des réseaux destinations dans une même communauté et d'appliquer des décisions de routage en fonction de cette communauté.
 - Exemple :
 - Des communautés connues universellement :
 - no-export (Do not advertise to EBGP peers)
 - no-advertise (Do not advertise this route to any peer)

BGP : Traffic engineering

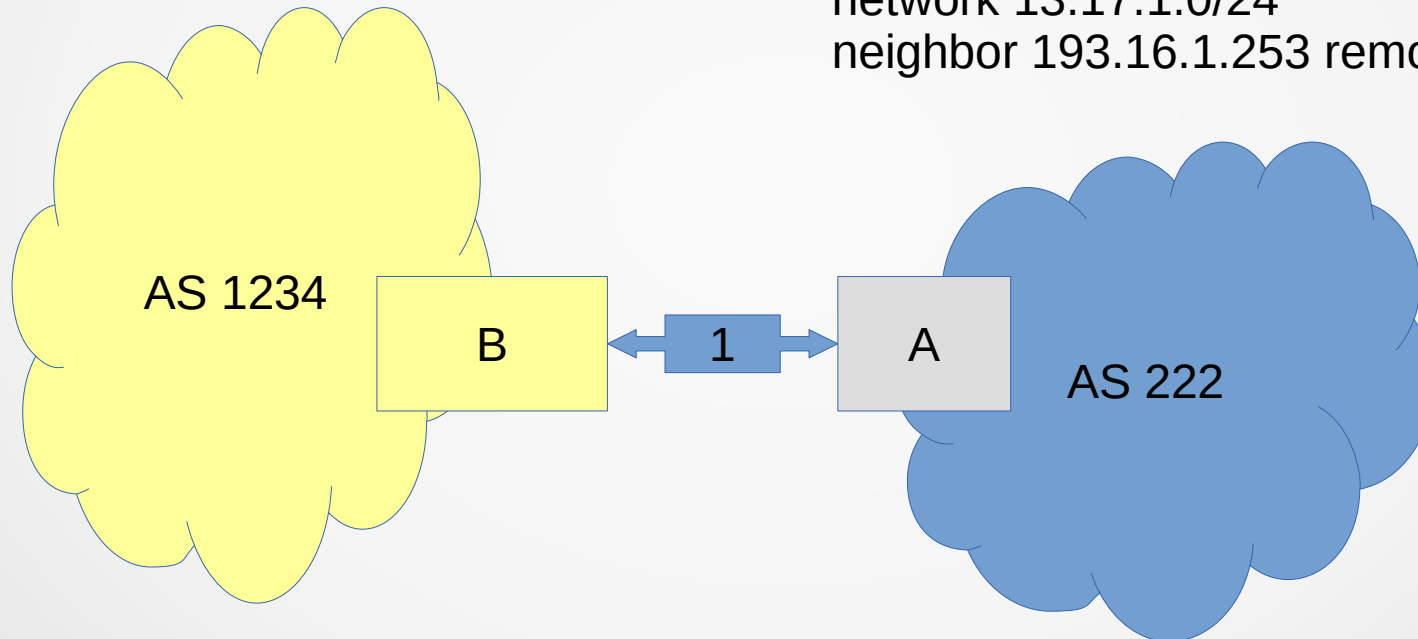
- Algorithme de sélection de la route :
 - Critères de choix (dans l'ordre)
 - LOCAL_PREFERENCE le plus élevé.
 - Le plus court AS_PATH
 - MED le plus petit
 - Autres critères de sélection :
 - Le coût le plus faible vers le NEXT_HOP (comme indiqué par l'IGP)
 - Route annoncée par le voisin eBGP ayant le plus petit BGPid
 - Route annoncée par le voisin iBGP ayant le plus petit BGPid

BGP : Traffic engineering

- En pratique, les constructeurs ajoutent de nouveaux critères, par exemple Cisco :
 - Ne pas considérer une route si le next_hop est injoignable.
 - Ne pas considérer les routes issues de l'iBGP si non joignable via IGP.
 - Avant de considérer LOCAL_PREFERENCE, utilisation du poids des routes
 - Avant de considérer MED, considérer l'origine des routes (privilégier les routes externes aux routes internes),
 - Etc...

BGP : Traffic engineering

- Exemple simpl(ist)e :

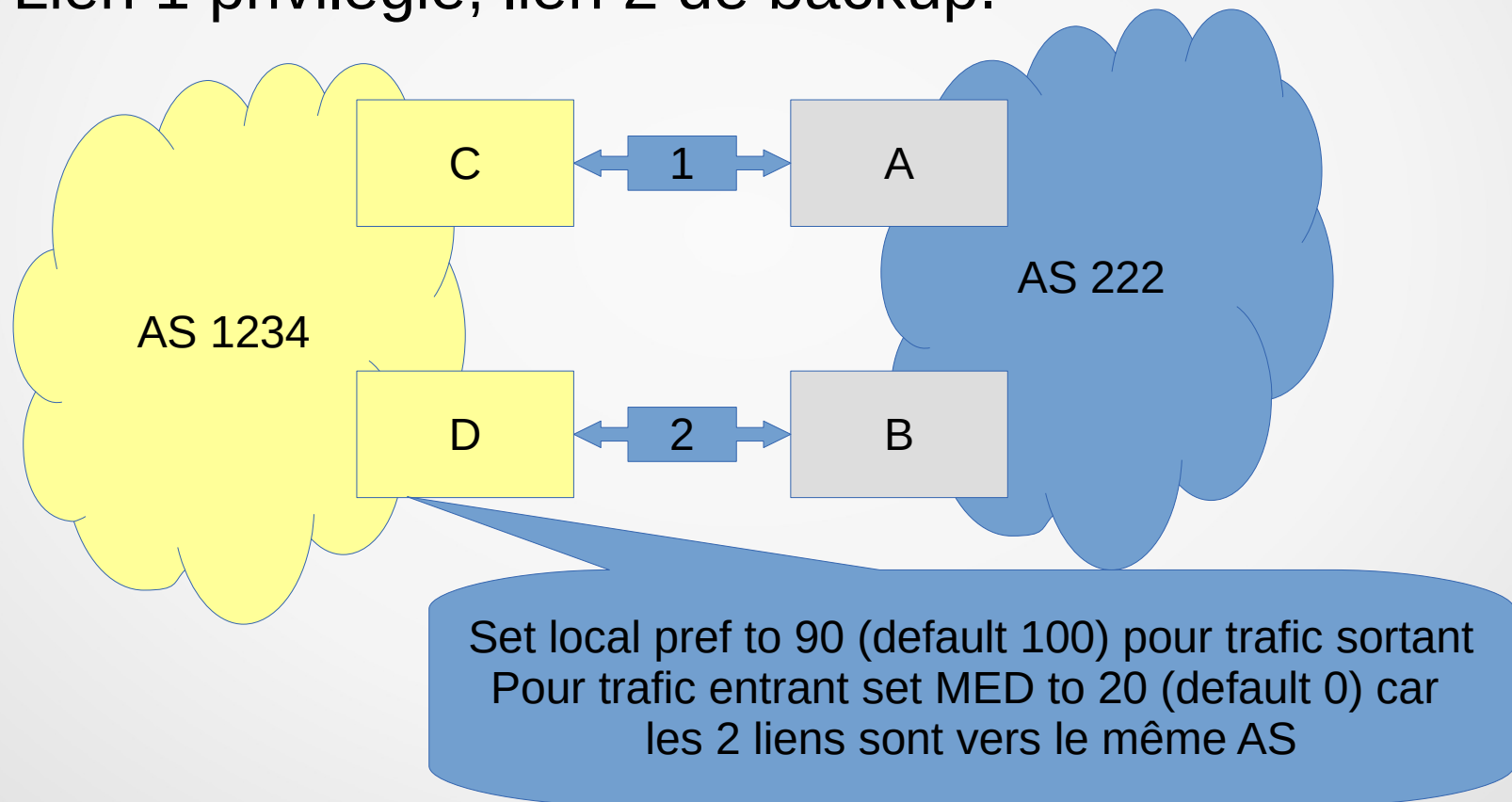


```
router bgp 222  
  bgp router-id 193.16.1.254  
  network 13.17.1.0/24  
  neighbor 193.16.1.253 remote-as 1234
```

```
router bgp 1234  
  bgp router-id 193.16.1.253  
  neighbor 193.16.1.254 remote-as 222
```

BGP : Traffic engineering

- Exemple partage de charge sur le même ISP :
 - Lien 1 privilégié, lien 2 de backup.



BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec :
 - neighbor peer distribute-list name/id [in|out]
 - access-list simple 1-99 --> ne filtre que sur la partie reseau de l'annonce ex:
 - access-list 1 permit 172.4.1.0 0.0.0.255
 - R1 recoit :
 - network 172.4.1.0/24
 - network 172.4.1.0/25

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec :
 - neighbor peer distribute-list name/id [in|out]
 - access-list extended 100-199
 - access-list 100 permit ip 172.4.1.0 0.0.0.255
255.255.255.128 0.0.0.127
 - R1 recoit :
 - network 172.4.1.0/25

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec :
 - neighbor peer distribute-list **name**/id [in|out]
 - access-list **toto** permit 172.4.1.0/24
 - R1 recoit :
 - Network 172.4.1.0/24
 - Network 172.4.1.0/25

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec :
 - neighbor peer distribute-list **name**/id [in|out]
 - access-list **toto** permit 172.4.1.0/24
 - R1 recoit :
 - Network 172.4.1.0/24
 - Network 172.4.1.0/25

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec :
 - neighbor peer **distribute-list** **name**/id [in|out]
 - access-list **toto** permit 172.4.1.0/24 exact-match
 - R1 recoit :
 - Network 172.4.1.0/24

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec : (forme plus simple)
 - neighbor peer prefix-list name [in|out]
 - ip prefix-list titi seq 10 permit 172.4.1.0/24 le 25
 - R1 recoit :
 - Network 172.4.1.0/24
 - Network 172.4.1.0/25

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec : (sur les attributs)
 - neighbor peer filter-list name [in|out]
 - filter-list s'applique pas sur les NLRI mais sur l'attribut AS_PATH
 - exemple:
 - ip as-path access-list toto permit ^20\$

BGP : Traffic engineering

- Le filtrage dans BGP (syntaxe quagga pour le TP)
 - R2 annonce les réseaux
 - network 172.4.1.0/24
 - Network 172.4.1.0/25
 - R1 filtre le voisin avec : (sur les attributs)
 - neighbor peer route-map name [in|out]
 - route-map permet de faire des autorisation et d'associer des actions en fonction des matchs
 -