Border Gateway Protocol
BGP4 et MP-BGP4
Section 1

AfNOG 2009
Le caire, 11-15 Mai 2009
aalain@trstech.net

1

Border Gateway Protocol (BGP)

- Rappels : bases du routage
- Briques élémentaires
- Exercices
- Bases du protocole BGP
- Exercices
- Attributs de routes BGP
- Calcul du meilleur chemin
- Exercices

2

Border Gateway Protocol (BGP)...

- Topologies typiques avec BGP
- Politiques de routage
- Exercices
- Redondance / Partage de charge
- Etat de l'art (BCP, Best Current Practices)

3

Le routage : quelques bases

4

Routage IP

- Chaque routeur (ou machine) décide comment acheminer un paquet
- L'expéditeur n'a pas à connaître le chemin jusqu'à la destination
- L'expéditeur doit seulement déterminer le prochain saut (next-hop).
  - Ce processus est répété jusqu'à arriver à la destination
- La table de routage est consultée afin de déterminer le prochain saut

5

Routage IP

- Routage par préfixe (Classless routing)
  - une route est composée de
    - la destination
    - l'adresse du prochain routeur (next-hop)
    - le masque de réseau permet de déterminer la taille de l'espace d'adressage concerné (-> préfixe)
- Choix du préfixe le plus long
  - pour une destination donnée, il faut prendre la route la plus spécifique (le préfixe le plus grand)
  - exemple: adresse destination 35.35.66.42
    - la table de routage contient 35.0.0.0/8, 35.35.64.0/19 and 0.0.0.0/0

6

Routage IP

- Route par défaut (default route)
  - indique où expédier un paquet si la table de routage ne contient pas une route spécifique
  - c'est une configuration courant : la plupart des machines disposent d'une (et une seule) route par défaut
  - autre nom : passerelle par défaut (default gateway)

7

Les routes spécifiques sont utilisées en premier

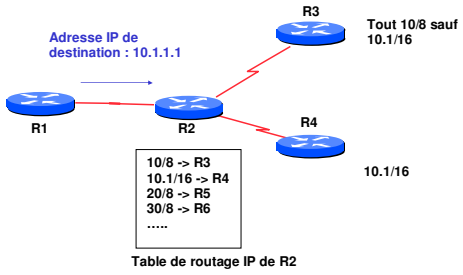


Table de routage IP de R2

8

Les routes spécifiques sont utilisées en premier

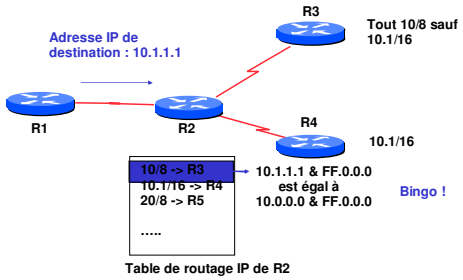


Table de routage IP de R2

9

Les routes spécifiques sont utilisées en premier

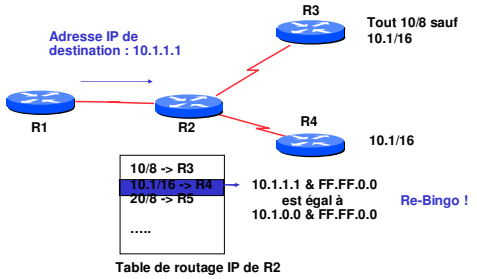


Table de routage IP de R2

10

Les routes spécifiques sont utilisées en premier

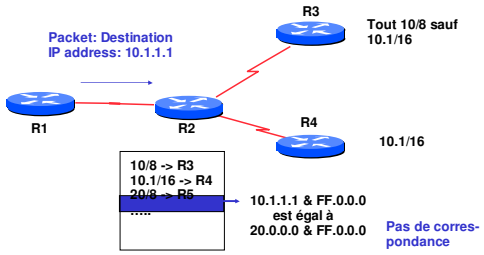


Table de routage IP de R2

11

Les routes spécifiques sont utilisées en premier

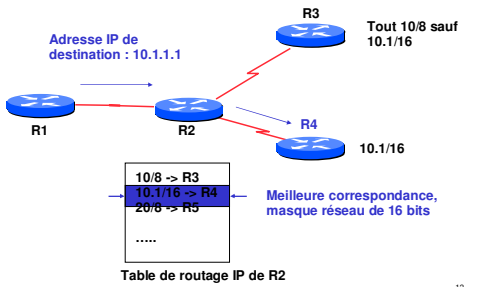


Table de routage IP de R2

12

Les routes spécifiques sont utilisées en premier

- On utilise toujours la route la plus spécifique (celle qui correspond au plus petit volume d'adresses IP)
- La route par défaut est notée 0.0.0.0/0
  - ce qui permet d'utiliser l'algorithme décrit ci-dessus
  - Il y a toujours correspondance. C'est la route la moins spécifique.

13

Routage dynamique

- Les routeurs déterminent leur table de routage automatiquement à partir des informations reçues des autres routeurs
- Les routeurs s'échangent les information de topologie en utilisant divers protocoles
- Les routeurs calculent ensuite un ou plusieurs "next-hops" pour chaque destination en essayant d'emprunter le meilleur chemin

14

Table d'acheminement

- En anglais : forwarding table
- Permet de déterminer comment acheminer un paquet dans le routeur
- Construite à partir de la table de routage
  - Les meilleurs routes sont choisies dans la table de routage
- Effectue une recherche pour déterminer le prochain saut et l'interface de sortie
- Commute le paquet sur l'interface de sortie avec l'encapsulation adéquate (ex : PPP, FR, POS)

15

Briques élémentaires

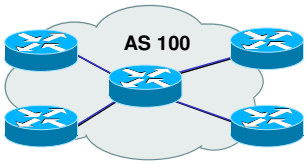
16

Briques élémentaires

- Système autonome - Autonomous System (AS)
- Type de routes
- IGP/EGP
- DMZ (zone démilitarisée)
- Politique
- Trafic sortant
- Trafic entrant

17

Système autonome (AS)



- Ensemble de réseaux partageant la même politique de routage
- Utilisation d'un même protocole de routage
- Généralement sous une gestion administration unique
- Utilisation d'un IGP au sein d'un même AS

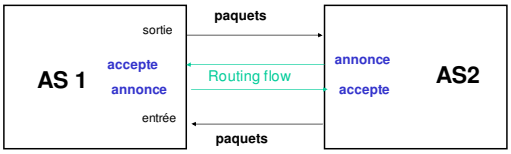
18

Système autonome (AS)...

- Caractérisé par un numéro d'AS
- Il existe des numéros d'AS privés et publics
- Exemples :
  - Prestataire de services Internet
  - Clients raccordés à plusieurs prestataires
  - Quiconque souhaite établir une politique de routage spécifique

19

Flux de routes et de paquets



Pour que AS1 et AS2 puissent communiquer :

AS1 annonce des routes à AS2
AS2 accepte des routes de AS1
AS2 annonce des routes à AS1
AS1 accepte des routes de AS2

20

Trafic en sortie

- Paquets qui quittent le réseau
  - Choix de la route (ce que les autres vous envoient)
  - Acceptation d'une route (ce que vous acceptez des autres)
  - Politique et configuration (ce que vous faites des annonces des autres)
  - Accords de transit et d'échange de trafic

21

Trafic entrant

- Paquets entrant dans votre réseau
- Ce trafic dépend de :
  - Ce que vous annoncez à vos voisins
  - Votre adressage et plan d'AS
  - La politique mise en place par les voisins (ce qu'ils acceptent comme annonces de votre réseau et ce qu'ils en font)

22

Types de routes

- Routes statiques
  - configurées manuellement
- Routes "connectées
  - crées automatiquement quand une interface réseau est "active"
- Routes dites "intérieures"
  - routes au sein d'un AS
  - routes apprises par un IGP
- Routes dites "extérieures"
  - routes n'appartenant pas à l'AS local
  - apprises par un EGP

23

Politique de routage

- Définition de ce que vous acceptez ou envoyez aux autes
  - connexion économique, partage de charge, etc...
- Accepter des routes de certains FAI et pas d'autres
- Envoyer des routes à certains FAI et pas à d'autres
- Préferrer les routes d'un FAI plutôt que d'un autre

24

Pourquoi a-t-on besoin d'un EGP ?

- S'adapter à un réseau de grande taille
  - hiérarchie
  - limiter la portée des pannes
- Définir des limites administratives
- Routage politique
  - contrôler l'accessibilité des préfixes (routes)

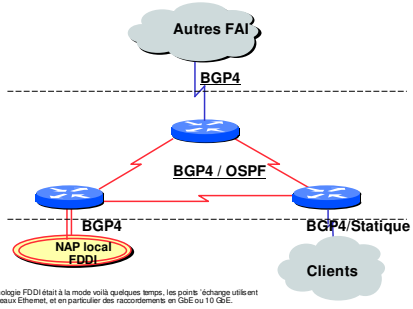
25

Protocoles intérieurs vs. extérieurs

- Intérieurs (IGP)
  - Découverte automatique
  - Confiance accordée aux routeurs de l'IGP
  - Les routes sont diffusées sur l'ensemble des routeurs de l'IGP
- Extérieurs (EGP)
  - Voisins explicitement déclarézs
  - Connexion avec des réseaux tiers
  - Mettre des limites administratives

26

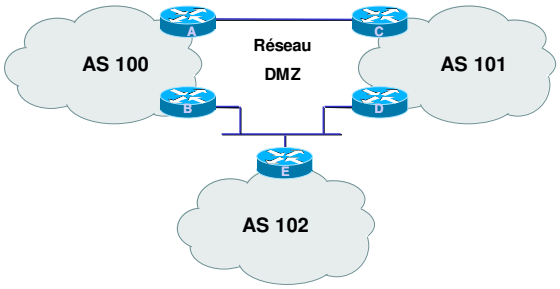
Hiérarchie dans les protocoles



Note: la technologie FDDI était à la mode voilà quelques temps, les points d'échange utilisent plutôt des réseaux Ethernet, et en particulier des raccordements en GbE ou 10 GbE.

27

Zone démilitarisée (DMZ)



- Le réseau démilitarisé est partagé entre plusieurs AS

28

Gestion de l'adressage (FAI)

- Il faut réserver des adresses IP pour son propre usage
- Des adresses IP sont également allouées aux clients
- Il faut prendre en considération la croissance de l'activité
- Le prestataire "upstream" attribuera les adresses d'interconnexion dans ses blocs

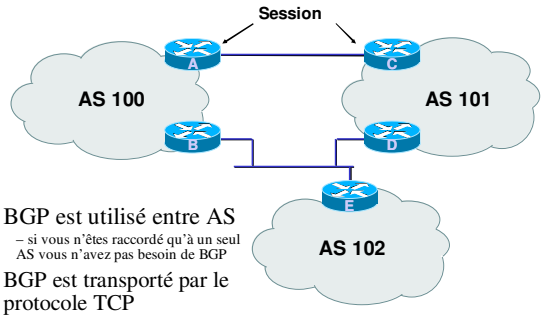
29

Bases de BGP

- Bases concernant le protocole
- Vocabulaire
- Messages
- Exploitation d'un routeur BGP
- Types de sessions BGP (eBGP/iBGP)
- Comment annoncer les routes

30

Principes de base du protocole



- BGP est utilisé entre AS
  - si vous n'êtes raccordé qu'à un seul AS vous n'avez pas besoin de BGP
- BGP est transporté par le protocole TCP

31

Principes de base (2)

- Les mises à jours sont incrémentielles
- BGP conserve le chemin d'AS pour atteindre un réseau cible
- De nombreuses options permettent d'appliquer une politique de routage

32

Vocabulaire

- Voisin (Neighbor)
  - Routeur avec qui on a une session BGP
- NLRI/Préfixe
  - NLRI - network layer reachability information
  - Informations concernant l'accessibilité (ou pas) d'une route (réseau + masque)
- Router-ID (identifiant de routeur)
  - Adresse IP la plus grande du routeur
- Route/Path (chemin)
  - Préfixe (NLRI) annoncé par un voisin

33

Vocabulaire (2)

- Transit - transport de vos données par un réseau tiers, en général moyennant paiement
- Peering - accord bi-latéral d'échange de trafic
  - chacun annonce uniquement ses propres réseaux et ceux de ses clients à son voisin
- Default - route par défaut, où envoyer un paquet si la table de routage de donne aucune information plus précise

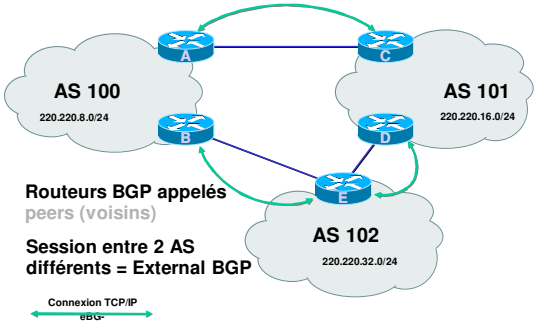
34

Bases de BGP ...

- Chaque AS est le point de départ d'un ensemble de préfixes (NLRI)
- Les préfixes sont échangés dans les sessions BGP
- Plusieurs chemins possibles pour un préfixe
- Choix du meilleur chemin pour le routage
- Les attributs et la configuration "politique" permettent d'influencer ce choix du meilleur chemin

35

Sessions BGP



Routeurs BGP appelés
peers (voisins)

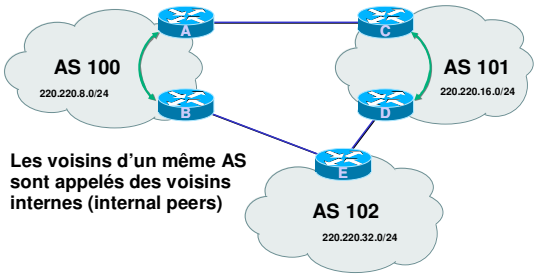
Session entre 2 AS différents = External BGP

Connexion TCP/IP eBGP

Note: les voisins eBGP doivent être directements raccordés.

36

Sessions BGP



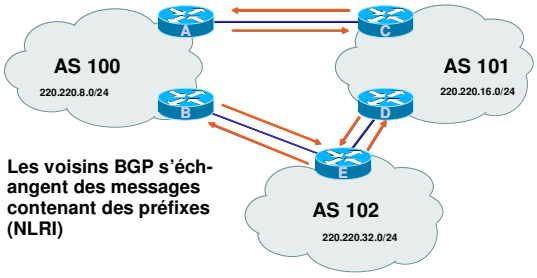
Les voisins d'un même AS sont appelés des voisins internes (internal peers)

Connexion TCP/IP iBGP

Note: les voisins iBGP peuvent ne pas être directement connectés.

37

Sessions BGP

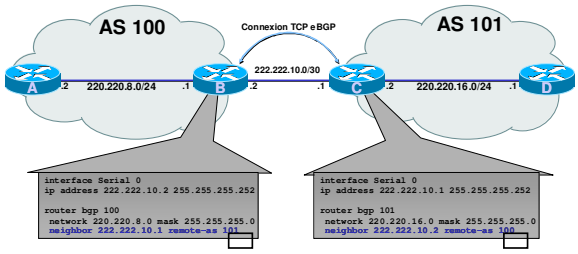


Les voisins BGP s'échangent des messages contenant des préfixes (NLRI)

Message de mise à jour BGP

38

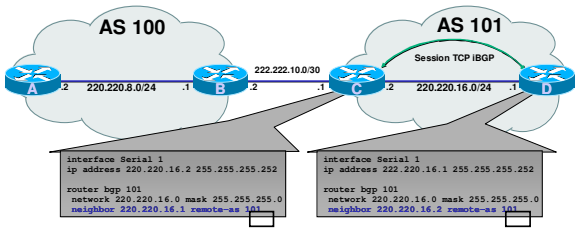
Configuration de sessions BGP



- **Les sessions BGP sont établies en utilisant la commande BGP "neighbor" du routeur**
  - Lorsque les numéros d'AS sont différents il s'agit d'une session BGP Externe (eBGP)

39

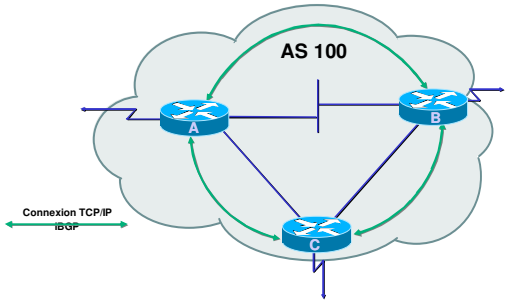
Configuration de sessions BGP



- **Les sessions BGP sont établies en utilisant la commande BGP "neighbor" du routeur**
  - Numéros d'AS différents -> BGP Externe (eBGP)
  - Numéros d'AS identiques -> BGP Interne (iBGP)

40

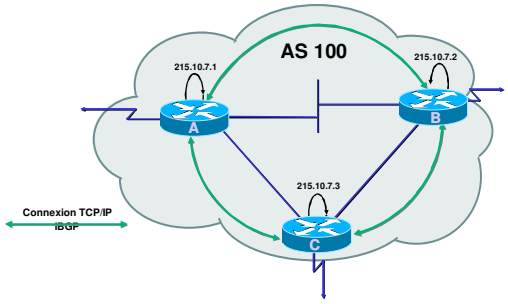
Configuration de sessions BGP



- Chaque routeur iBGP doit établir une session avec tous les autres routeurs iBGP du même AS

41

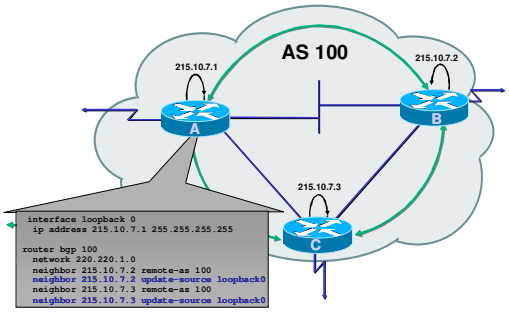
Configuration de sessions BGP



- Il est recommandé d'utiliser des interfaces Loopback sur les routeurs comme extrêmités des sessions iBGP

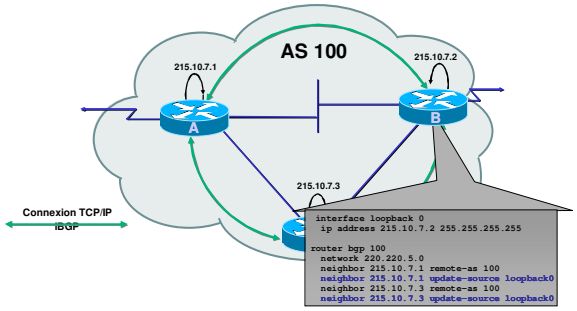
42

Configuration des sessions BGP



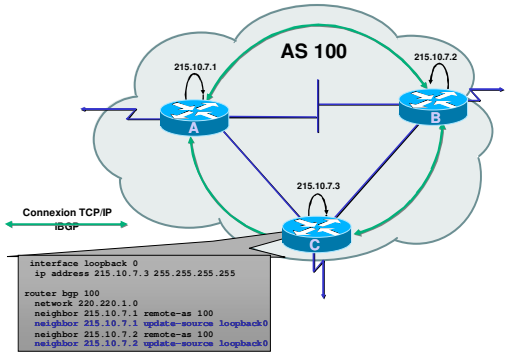
43

Configuration des sessions BGP



44

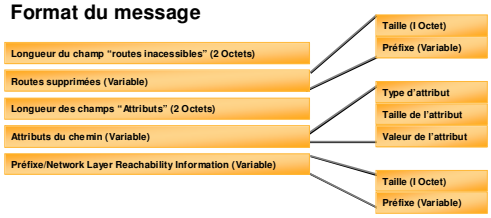
Configuration des sessions BGP



45

Messages de mise à jour BGP

Format du message



- Une mise à jour BGP permet d'annoncer plusieurs routes à un voisin, ou de supprimer plusieurs routes qui ne sont plus accessibles, ou de simutanément annoncer une route et supprimer plusieurs routes.
- Chaque message contient des attributs comme : origine, chemin d'AS, Next-Hop, ...

46

Mises à jour BGP — Préfixes/NLRI

- NLRI = Network Layer Reachability Information = Préfixes
- Permet d'annoncer l'accessibilité d'une route
- Composé des informations suivantes :
  - Préfixe réseau
  - Longueur du masque

47

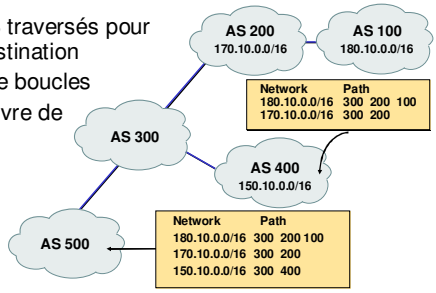
Mise à jour BGP — Attributs

- Permet de transporter des informations liées au préfixe
  - Chemin d'AS
  - Adresse IP du "next-hop"
  - Local preference (préférence locale)
  - Multi-Exit Discriminator (MED)
  - Community (communauté)
  - Origin (origine de la route)
  - Aggregator (IP d'origine si aggrégation)

48

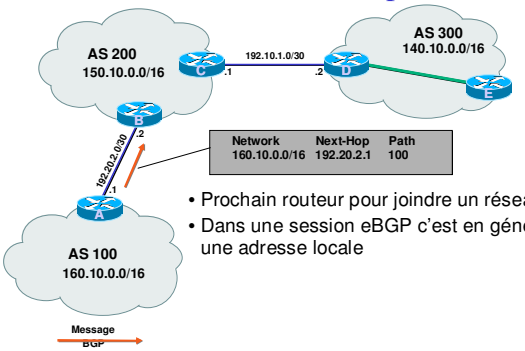
Attribut “chemin d’AS”

- Liste les AS traversés pour arriver à destination
- Détection de boucles
- Mise en œuvre de politiques



49

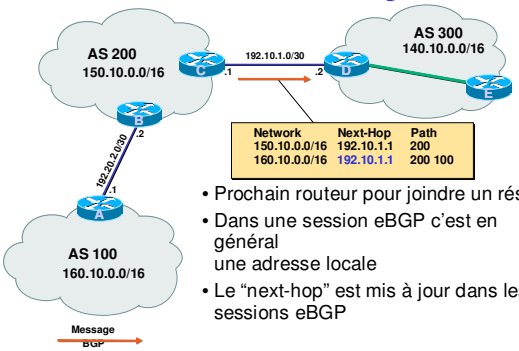
Attribut “Next-Hop”



- Prochain routeur pour joindre un réseau
- Dans une session eBGP c’est en général une adresse locale

50

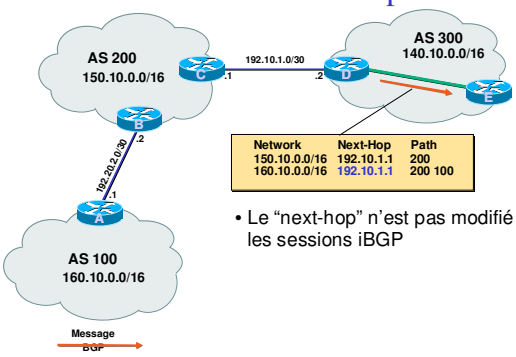
Attribut “Next-Hop”



- Prochain routeur pour joindre un réseau
- Dans une session eBGP c’est en général une adresse locale
- Le “next-hop” est mis à jour dans les sessions eBGP

51

Attribut “Next-Hop”



- Le “next-hop” n’est pas modifié dans les sessions iBGP

52

Attribut “Next-Hop” (suite)

- Les adresses des “next-hops” doivent circuler dans l’IGP
- Recherche récursive des routes
- Permet de concevoir la topologie BGP indépendemment de la topologie physique du réseau
- En interne les bonnes décisions de routage sont faites par l’IGP

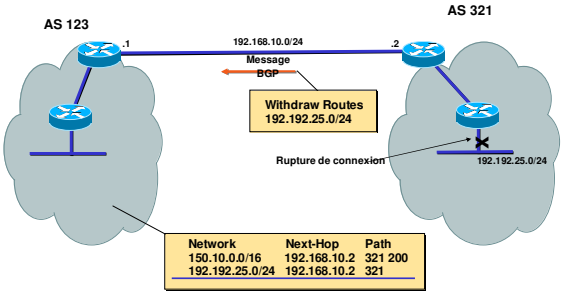
53

Mises à jour BGP — Suppression de routes

- Permet de retirer un réseau de la liste des réseaux accessibles
- Chaque route supprimée est composée de :
  - son Préfixe
  - la longueur du masque

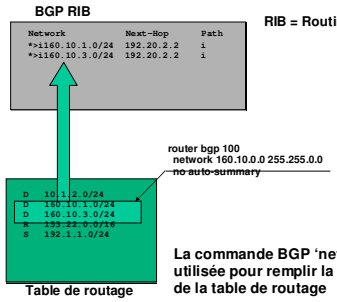
54

Mises à jour BGP - Suppression de routes



55

Table du routeur BGP

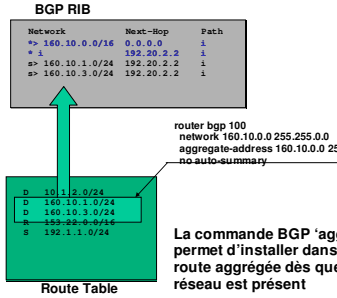


RIB = Routing Information Base

La commande BGP 'network' peut être utilisée pour remplir la table BGP à partir de la table de routage

56

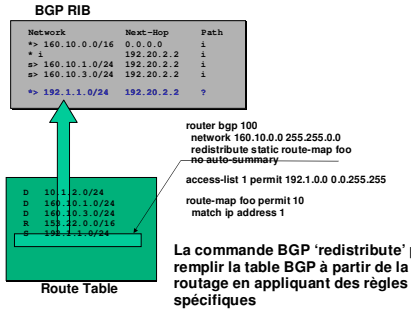
Table du routeur BGP



La commande BGP 'aggregate-address' permet d'installer dans la table BGP une route aggrégée dès que au-moins un sous-réseau est présent

57

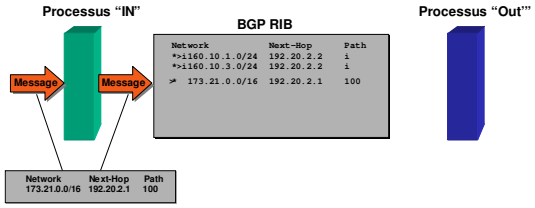
Table du routeur BGP



La commande BGP 'redistribute' permet de remplir la table BGP à partir de la table de routage en appliquant des règles spécifiques

58

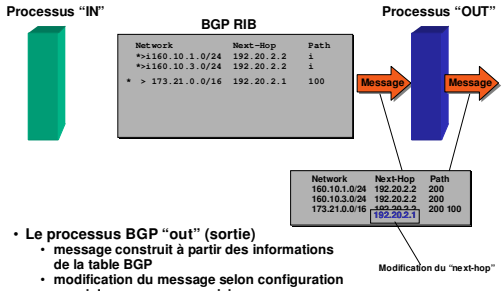
Table du routeur BGP



- **Le processus BGP "in" (entrée)**
  - **reçoit les messages des voisins**
  - **place le ou les chemins sélectionnés dans la table BGP**
  - **le meilleur chemin (best path) est indiqué avec le signe ">"**

59

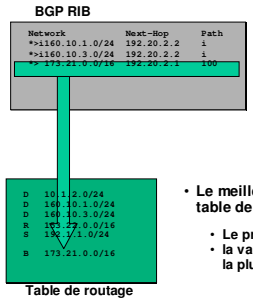
Table du routeur BGP



- **Le processus BGP "out" (sortie)**
  - **message construit à partir des informations de la table BGP**
  - **modification du message selon configuration**
  - **envoi du message aux voisins**

60

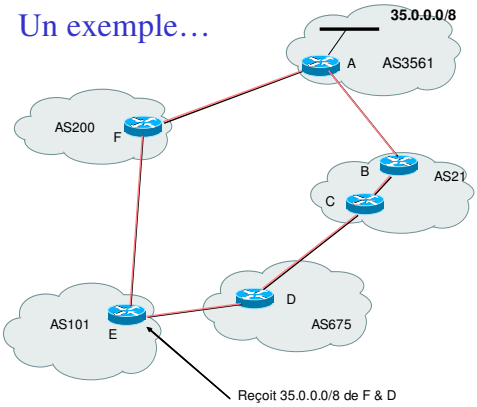
Table du routeur BGP



- **Le meilleur chemin est installé dans la table de routage du routeur si :**
  - **Le préfixe et sa taille sont uniques**
  - **la valeur "distance" du protocole est la plus faible**

61

Un exemple…



62

MP-BGP:MultiProtocol BGP

63

MP-BGP:MultiProtocol BGP(1)

- Pour rendre BGP-4 disponible pour d'autres protocoles réseaux, RFC 4760(Obsoletes: 2858) a défini les extensions multiprotocoles pour BGP-4
- Permet à BGP-4 de transporter des informations d'autres protocoles comme, MPLS, IPv6
- Utilise une combination de «Address Family Identifier (AFI)» et « Subsequent Address Family(SAFI) »
  - http://www.iana.org/assignments/address-family-numbers
    - IPv4, IPv6, etc.....
  - http://www.iana.org/assignments/safi-namespace
    - Unicast forwarding, multicast forwarding etc....

64

MP-BGP:MultiProtocol BGP(2)

- Les routeurs utilisent la négociation de capacités pour signaler leur support du MP-BGP
  - RFC3392
- Utilise deux nouveaux attributs BGP
  - MP_REACH_NLRI
  - MP_UNREACH_NLRI

65

Configuration de BGP

66

Commandes BGP de base(1)

Configuration

```
router bgp <AS-number>
no bgp default ipv4-unicast

address-family ipv6
neighbor <ipv6 address> remote-as  <as-number>
neighbor <ipv6 address> activate
exit-address-family

address-family ipv4
network 196.200.221.208 mask 255.255.255.248
neighbor 196.200.221.162 remote-as 1
neighbor 196.200.221.162 update-source loopback0
neighbor 196.200.221.162  activate
exit-address-family
```

67

Commandes BGP de base(2)

Consultation d'information

```
show bgp ipv4 unicast summary
show bgp ipv4 unicast neighbors

show bgp ipv6 unicast summary
show bgp ipv6  unicast neighbors

show bgp all  summary
show bgp all  neighbors
```

68

Ajout de préfixes dans la table BGP

- Cela peut se faire de deux grandes manières
  - "redistribute static" (redistribuer les routes statiques)
  - utiliser la commande BGP "network"

69

Pour insérer une route...

- Commande ***network*** ou redistribution

`network` <*ipaddress*> **`mask`** <*netmask*>

`redistribute` <*protocol name*>

- Il faut que la route soit présente dans la table de routage du routeur pour qu'elle soit insérée dans la table BGP

70

Utilisation de "redistribute static"

- Exemple de configuration

```
router bgp 109
   redistribute static
ip route 198.10.4.0 255.255.254.0 serial0
```

- La route statique doit exister avant que la redistribution ne fonctionne
- L'origine de la route sera "*incomplete*", mais il est possible de le changer avec une "route-map"
- A utiliser avec prudence !

71

Utilisation de "redistribute"

- Attention avec les redistributions
  - redistribute <protocole> signifie que toutes les routes du <protocole> seront transférées dans le protocole courant
  - cette solution doit être contrôlée (volumétrie)
  - à éviter dans la mesure du possible
  - préférer l'utilisation de "route-maps" et avec un contrôle administratif très strict

72

Utilisation de la commande “network”

- Exemple de configuration

```
network 198.10.4.0 mask 255.255.254.0
ip route 198.10.0.0 255.255.254.0 serial 0
```

- La route doit être présente dans la table de routage pour qu’il y ait une annonce BGP
- Origine de la route : IGP

73

Aggrégats et routes vers Null0

- Rappel : la route doit exister dans la table de routage pour être annoncée via BGP

```
router bgp 1
   network 198.10.0.0 mask 255.255.0.0
ip route 198.10.0.0 255.255.0.0 null0 250
```

- Une route vers “null0” est souvent utilisée pour faire de l’aggrégation
  - destination en dernier ressort pour le préfixe
  - distance de 250 pour être sûr d’être le dernier choix
- Très pratique pour la stabilité de la route
  - il ne peut y avoir de “flap” !

74

Choix pour les sessions iBGP

- Les sessions iBGP ne doivent pas être liées à la topologie du réseau
- L’IGP transporte les adresses de Loopback

router ospf *<ID>*
network *<loopback-address>* **0.0.0.0 area 0**

- Utiliser les adresses Loopback pour les sessions iBGP

router bgp *<AS1>*
neighbor *<x.x.x.x>* **remote-as** *<AS1>*
neighbor *<x.x.x.x>* **update-source** *loopback0*

75