Le routage externe BGP4

Luc Saccavini

LE ROUTAGE BGP4(+)

septembre 2006

Luc.Saccavini@inria.fr

Classification des protocoles de routage

- Il existe 2 grandes familles de protocoles de routage
  - Les protocoles intérieurs (IGP)
    - Distance-vecteur : RIP, IGRP
    - État des liens : OSPF, IS-IS
    - Taille <100 routeurs, 1 autorité d'administration
    - Échange de routes, granularité = routeur
  - Les protocoles extérieurs (EGP)
    - EGP, BGP, IDRP
    - Taille = Internet, coopération d'entités indépendantes
    - Échange d'informations de routage, granularité = AS

Rappel sommaire sur les types de protocoles de routage :

- distance vecteur : la distance est le nombre de routeurs pour joindre une destination, chaque routeur ne connaît que son voisinage et propage les routes qu'il connaît à ses voisins (ex. RIP).

- états des liens : chaque routeur connaît la topologie et l'état de l'ensemble des liens du réseau, puis en déduit les chemins optimaux. À chaque interaction les routeurs s'envoient toute leur table de routage (ex. OSPF).

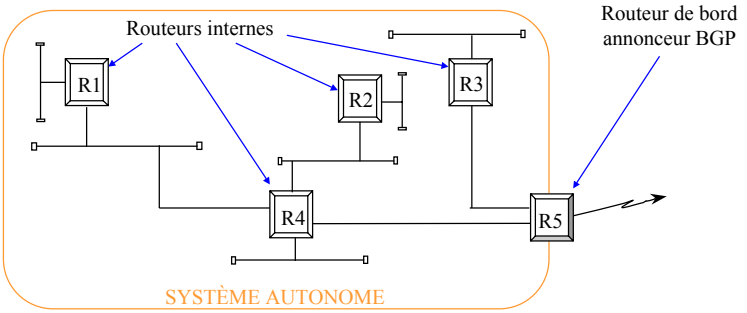
Le protocole BGP peut être considéré comme à mi-chemin entre les deux types de protocoles précédents. En effet, l'échange de chemins d'AS permet à chaque routeur de reconstruire une grande partie de la topologie du réseau, ce qui est caractéristique des protocoles de type «état des liens», mais deux routeurs voisins n'échangent que les routes qu'ils connaissent, ce qui est caractéristique d'un protocole de type «distance-vecteur».

Références sur les autres protocoles de routage :

	IPv4	IPv6
RIP (Routing Information Protocol)	RFC 2453, 11/98 «RIPv2»	RFC2080, 01/97 «RIPng»
IGRP	voir manuel IOS de Cisco	
EIGRP	voir manuel IOS de Cisco	
OSPF (Open Shortest Path First)	RFC 2328, 04/98 «OSPv2»	RFC 2740, 12/99 «OSPv3»
IS-IS (Intermediate System to Intermediate System ISO/IEC 10589, (ou RFC1142, 02/90)		
EGP (Exterior Gateway Protoco)	RFC 904 04/84	--------
IDRP (Inter Domain Routing Protocol)	ISO/IEC IS10747 10/93	
BGP (Border Gateway Protocol)	RFC 4271, 01/06 «BGP4»	RFC 2545, 03/99 «BGP4+»

Notion de système autonome (AS)

- Ensemble de routeurs sous une même entité administrative



Au sein d'un AS plusieurs IGP (et/ou un routage statique) peuvent être utilisés.

Fonctionnellement, on distingue 2 types de Systèmes Autonomes :

- les AS clients : ils sont les producteurs ou les consommateurs de paquets IP
- les AS de transit : ils ne font que transporter les paquets IP qui leurs sont confiés

Un AS n'est à priori pas lié à la localisation géographique des différents routeurs qui le constituent.

Objectifs généraux du protocole BGP

- Échanger des routes (du trafic) entre organismes indépendants
  - Opérateurs
  - Gros sites mono ou multi connectés
- Implémenter la politique de routage de chaque organisme
  - Respect des contrats passés entre organismes
  - Sûreté de fonctionnement
- Être indépendant des IGP utilisés en interne à un organisme
- Supporter un passage à l'échelle (de l'Internet)
- Minimiser le trafic induit sur les liens
- Donner une bonne stabilité au routage

BGP élimine les boucles de routage en examinant le chemin d'AS associé à une route.

Les RFC1265 et RFC1774 contiennent une étude des propriétés de mise à l'échelle du protocole BGP. Dans cette étude, si on appelle N le nombre total de préfixes annoncés dans l'Internet, M la distance moyenne entre les AS (exprimée en nombre d'AS), et A le nombre total d'AS de l'Internet, alors, le volume d'information échangé lors du premier échange entre deux voisins BGP est proportionnel à : O(N+M*A). Le volume de mémoire nécessaire dans chaque routeur étant proportionnel à : O((N+M*A)*K), avec K=nombre moyen de voisins BGP par routeur.

Nombre de préfixes (N)	Distance moy. inter-AS (M)	Nombre moy. d'AS (A)	Nombre moy. de voisins (K)	Volume initial échangé	Volume mém. utilisé
2100	5	59	3	9000	27000
4000	10	100	6	18000	108000
10000	15	300	10	49000	490000
20000	8	400		86000	
40000	15	400		172000	
100000	20	3000	20	520000	1040000

La première ligne de ce tableau correspond à la situation de début 1991, la quatrième à celle de fin 1994, la dernière au 1er semestre 2001.

Principes généraux du protocole BGP

- ❏ Protocole de type PATH-vecteur
- ❏ Chaque entité est identifiée par un numéro d'AS
- ❏ La granularité du routage est le Système Autonome (AS)
- ❏ Le support de la session BGP est TCP (port 179)
- ❏ Les sessions BGP sont établies entre les routeurs de bord d'AS
- ❏ Protocole point à point entre routeurs de bord d'AS
- ❏ Protocole symétrique
- ❏ (un annonceur BGP n'est pas forcément un routeur)

Le choix de TCP comme support du protocole est important car il le libère du problème de garantir une bonne transmission des informations.

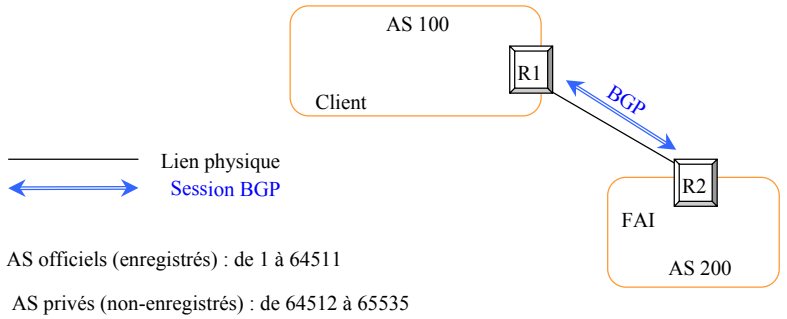
C'est ce choix qui a permis aux architectes du protocole de ne plus procéder que par mise à jour des informations modifiées après l'envoi initial de la table complète lors de l'ouverture de la session entre deux routeurs. Cela permet de minimiser le trafic induit.

La politique de routage se traduit par le filtrage des routes apprises et annoncées (ne jamais oublier qu'annoncer une route vers un réseau c'est accepter du trafic à destination de ce réseau).

Le filtrage (au sens BGP) peut agir en «tout ou rien» sur la route (annonce, prise en compte), mais aussi par modification des attributs de la route pour modifier la préférence accordée à la route comme on le verra plus loin.

Exemple de connexion BGP (1)

- Client connecté à un seul Fournisseur d'Accès Internet (FAI). Seuls les routeurs de bord de l'AS sont figurés.



Les routeurs qui échangent leurs informations en BGP doivent être directement connectés (liaison point à point ou LAN partagé).

C'est la conséquence logique de la frontière administrative qui les sépare et qui empêche que le routage à travers un réseau de routeurs puisse être assuré par un IGP.

Exceptionnellement, des routeurs de bord peuvent ne pas être en vis-à-vis (ex. le routeur où arrive le lien externe à l'AS ne connaît pas le protocole BGP).

L'utilisation de numéros d'AS privés est à éviter pour des AS terminaux (clients) car une connexion à un deuxième AS de transit (FAI) peut conduire à une configuration illégale.

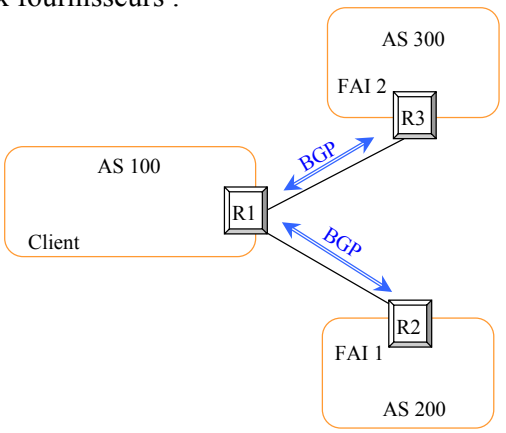
Les numéros d'AS officiels sont attribués par les mêmes organismes qui sont en charge de distribuer les réseaux IP :

- RIPE-NCC : zone Europe
- APNIC : zone Asie et Pacifique
- ARIN : zone Amérique du Nord
- AFRINIC : zone Afrique
- LACNIC : zone Amérique Latine et îles Caraïbes

C 'est le même numéro d'AS qui est utilisé pour les échanges de préfixes IPv4 et IPv6 (car BGP est multi-protocole)..

Exemple de connexion BGP (2)

- Client connecté à deux fournisseurs :



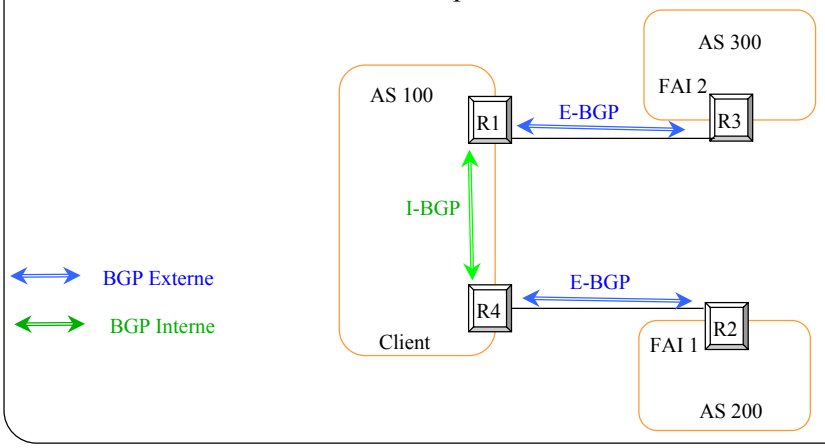
R1 à deux voisins : R2 et R3

L'AS client peut choisir de faire passer tout son trafic par FAI1 (AS 200) et garder sa liaison vers FAI2 (AS 300) en secours, ou équilibrer son trafic entre FAI1 et FAI2. C'est le cas typique qui amène à utiliser le protocole de routage BGP pour réagir dynamiquement en cas de défaillance d'un lien.

Dans le cas précédent, le seul intérêt d 'avoir un protocole de routage dynamique (par rapport à une simple route par défaut) est de pouvoir avoir une alerte (en provenance de la session BGP) en cas de défaillance du FAI.

Exemple de connexion BGP (3)

❑ Client connecté à 2 fournisseurs par 2 routeurs différents :



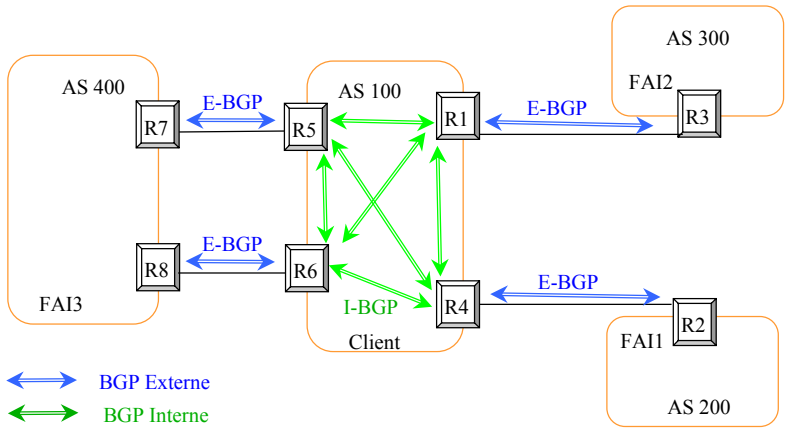
Ce schéma correspond au cas classique d'un client qui se connecte à deux fournisseurs pour s'assurer une protection contre la défaillance de l'un d'entre eux ou de l'un de ses routeurs de bord R1 ou R4.

On notera la présence d'une connexion BGP entre les routeurs de bord de l'AS 100. Cette connexion BGP «interne» (notée I-BGP) est nécessaire pour maintenir la cohérence entre ces 2 routeurs qui doivent posséder les mêmes informations de routage (se souvenir qu'en BGP la granularité du routage est l'AS).

L'un des principaux intérêts de l'I-BGP est de permettre la redondance des routeurs de bord d'un AS.

Exemple de connexion BGP (4)

❑ Client connecté à 3 fournisseurs avec redondance sur l'un :



Noter le maillage complet de sessions I-BGP entre R1, R4, R6, R5 dans l'AS 100. Pour les autres AS, les 4 routeurs de bord de l'AS 100 sont vus, du point de vue fonctionnel comme un seul routeur (avec 4 interfaces).

Cet exemple montre aussi une des limitations d'avoir à faire un maillage complet de sessions I-BGP entre les routeurs de bord d'un même AS (nombre de sessions = N*(N+1)/2). On verra à la fin de l'exposé qu'il existe des solutions (réflecteurs de routes) qui permettent de diminuer le nombre de sessions I-BGP.

Sauf mention explicite, tout ce qui est exposé dans la suite concerne les sessions BGP externes.

Dans le cas de deux AS multiplement connectés comme AS400 et AS100 et si l'ensemble des routeurs de bord des deux AS partagent un même LAN, les routeurs de bord ne sont pas forcément des annonceurs BGP, et vice-versa.

Règles pour les AS multi-connectés

- Les routeurs de bord d'un même AS échangent leurs informations de routage en I-BGP
- Les connexions en I-BGP forment un maillage complet sur les routeurs de bord d'un AS
- Ce sont les IGP internes à l'AS qui assurent et maintiennent la connectivité entre les routeurs de bord qui échangent des informations de routage en I-BGP
- Le numéro d'AS est un numéro officiel (si connexions vers 2 AS différents)

Attention, dans un même AS, c'est bien l'IGP (ou le routage statique) qui est responsable de la connectivité interne de l'AS. Si un routeur de bord ne peut pas atteindre une route de son AS (qui lui a été annoncée par un voisin interne par exemple), il ne la propagera pas à ses voisins BGP (externes ou internes).

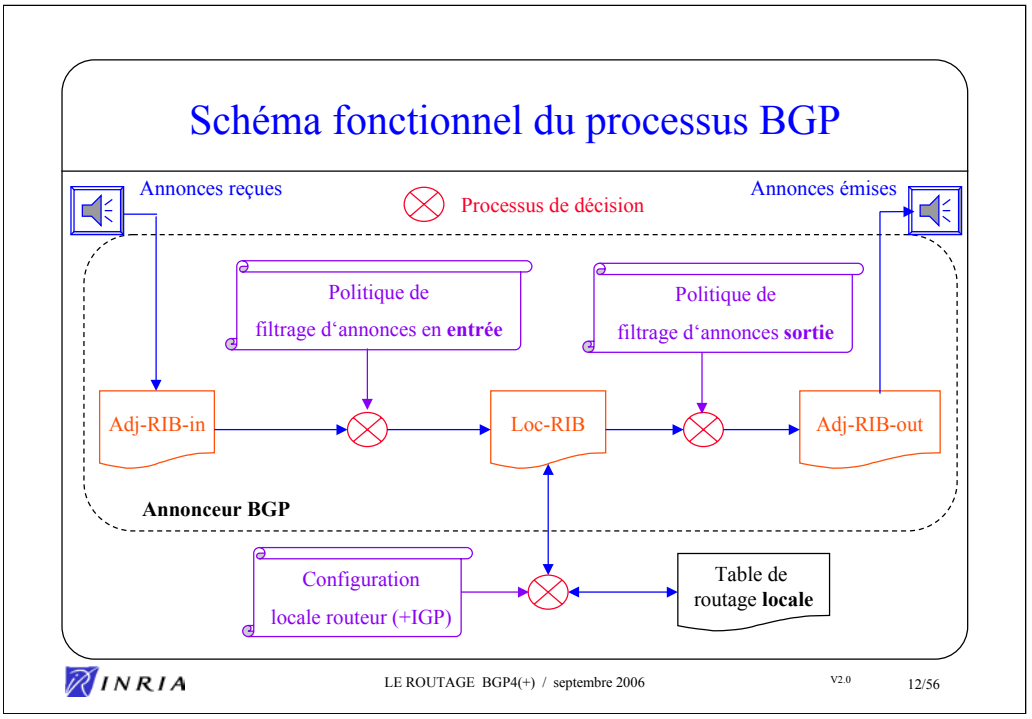
Les composants d'un annonceur BGP

- Une description des politiques de routage (entrée et sortie)
- Des tables où sont stockées les informations de routage
  - En entrée : Adj-RIB-in (informations reçues et non traitées)
  - En sortie : Adj-RIB-out (informations à annoncer)
  - En interne : Loc-RIB (informations locales au routeur)
- Un automate implémentant le processus de décision
- Des sessions avec ses voisins pour échanger les informations de routage

L'expression 'routeur BGP' est très souvent utilisée à la place de 'annonceur BGP' car il est peu fréquent qu'un annonceur ne soit pas aussi un routeur. Le cas pouvant cependant se produire (ex. serveurs de routes), le standard (RFC4271) utilise systématiquement l'expression 'annonceur BGP'.

Concernant les 3 tables où sont stockées les informations de routage, le standard ne spécifie pas qu'elles doivent être physiquement séparées, ce qui impliquerait un gaspillage de mémoire qui est une ressource critique sur les routeurs qui doivent connaître toutes les routes de l'Internet (environ 180 000 en septembre 2006).

La spécification de l'expression de la politique de routage dans BGP n'est pas standardisée, elle dépend donc des implémentations du protocole. Une telle standardisation n'est suggérée que dans RFC1786 (*status Informational*) pour les bases des organismes d'allocation (RIPE-NCC, ARIN, APNIC, LACNIC, AFRINIC).



Noter la flèche à double sens entre la table Loc-RIB et le processus de décision en bas du schéma. En effet, si c'est bien la table Loc-RIB qui permet au final de bâtir la table de routage, elle reçoit aussi des informations sur les routes locales de l'AS à travers des directives du fichier de configuration (annonces statiques ou redistribution des routes apprises par l'IGP dans BGP).

Ce schéma ne concerne que les annonces reçues et faites en E-BGP. En I-BGP, le schéma est plus simple (voir fin d'exposé).

Quand l'annonceur BGP est aussi un routeur, sa table de routage locale est construite à partir des informations de routage produites par le processus BGP, les autres protocoles de routage, et sa configuration. S'il existe plusieurs routes vers le même réseau, une métrique nouvelle est introduite (la 'distance administrative' dans l'implémentation de Cisco) pour régler le choix de la route à installer dans la table de routage.

La vie du processus BGP

- Automate à 6 états, qui réagit sur 13 événements
- Il interagit avec les autres processus BGP par échange de 4 types de messages :
  - OPEN
  - KEEPALIVE
  - NOTIFICATION
  - UPDATE
- Taille des messages de 19 à 4096 octets
- Éventuellement sécurisés par MD5

Les messages étant de longueur variable, ils sont marqués dans le flot d'octets du canal TCP par une séquence spéciale de trois octets qui repère leur début.

Le message OPEN

- 1er message envoyé après l'ouverture de la session TCP
- Informe son voisin de :
  - Sa version de BGP
  - Son numéro d'AS
  - D'un numéro identifiant le processus BGP
- Propose une valeur de temps de maintien de la session
  - Valeur suggérée : 90 secondes
  - Si 0 : maintien sans limite de durée
- Met le processus en attente d'un KEEPALIVE

En cas de démarrage simultané de deux sessions BGP par deux voisins, il faut choisir de ne conserver que l'une des deux connexions. Pour cela on ne conserve que celle ouverte par le processus de numéro identifiant le plus petit. Pour déterminer ce numéro identifiant, les implémentations de Cisco et Zebra choisissent par défaut le plus petit numéro IP de interfaces connues.

Le message KEEPALIVE

- Confirme un OPEN
- Réarme le minuteur contrôlant le temps de maintien de la session
- Si temps de maintien non égal à 0
  - Est ré-émis toutes les 30 secondes (suggéré)
- Message de taille minimum (19 octets)

En cas d'absence de modification de leur table de routage, les routeurs ne s'échangent plus que des messages KEEPALIVE toutes les 30 secondes, ce qui génère un trafic limité à environ 5bits/s au niveau BGP.

L'implémentation BGP de Cisco porte par défaut à 60 secondes l'intervalle entre 2 messages KEEPALIVE, celle de Zebra à 30 secondes.

Le message NOTIFICATION

- Ferme la session BGP
- Fournit un code et un sous code renseignants sur l'erreur
- Ferme aussi la session TCP
- **Annule toutes les routes apprises par BGP**
- Émis sur incidents :
  - Pas de KEEPALIVE pendant 90s (<*hold time*>)
  - Message incorrect
  - Problème dans le processus BGP
  - ….

Le message NOTIFICATION est envoyé au moindre incident lors du déroulement du processus BGP. Le fait de supprimer lors de son arrivée toutes les routes apprises par BGP peut provoquer des instabilités de routage injustifiées (un incident ne veut pas forcément dire que toutes les routes apprises précédemment sont devenues fausses).

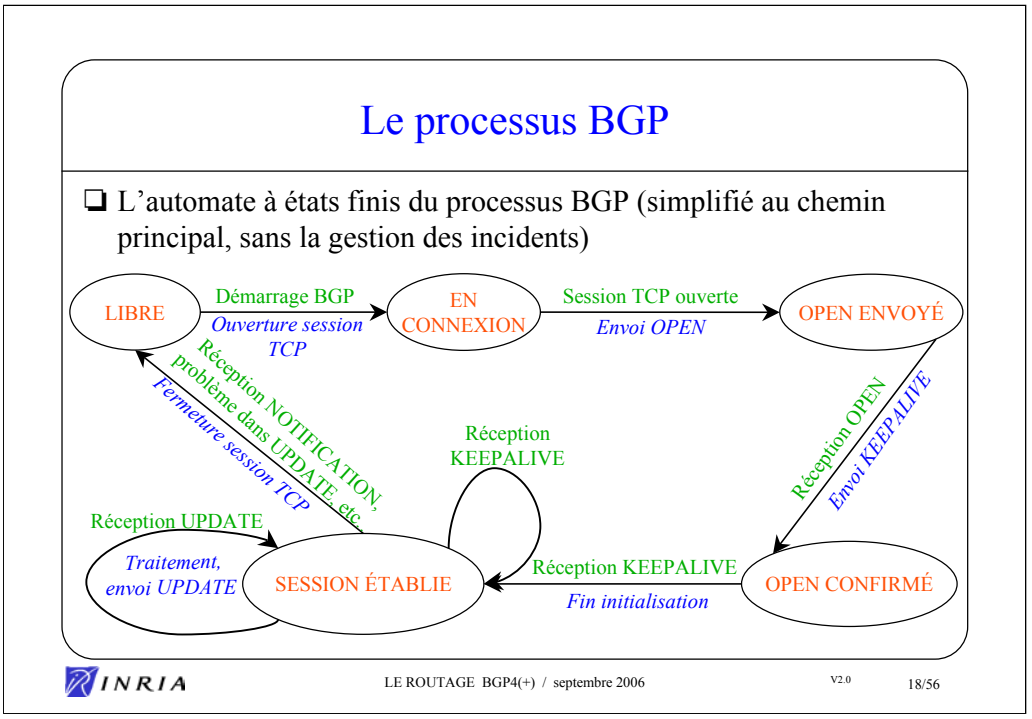
Dans son implémentation de BGP, Cisco donne la possibilité de supprimer cette fonctionnalité, en conservant telle quelle la table de routage en cas de réception d'un message NOTIFICATION.

Le message UPDATE

- Sert à échanger les informations de routage
  - Routes à éliminer (éventuellement)
  - Ensemble des attributs de la route
  - Ensemble des réseaux accessibles (NLRI)
    - Chaque réseau est défini par (préfixe, longueur)
- Envoyé uniquement si changement
- Active le processus BGP
  - Modification des RIB f(Update, politique de routage, conf.)
  - Émission d'un message UPDATE vers les autres voisins

C'est le message principal du protocole.

Lors du paramétrage d'un processus BGP il faut aussi faire un choix entre synchroniser ou pas les annonces de l'IGP et les annonces BGP.



L'état supplémentaire non figuré (ACTIF) sur le schéma se rapporte à la phase d'initialisation de la session BGP et concerne la gestion des incidents au niveau TCP pendant cette phase.

La liste complète des événements pouvant arriver est la suivante :

1 : Démarrage BGP

2 : Fin BGP

3 : Session TCP ouverte

4 : Session TCP fermée

5 : Ouverture session TCP échouée

6 : Erreur fatale dans session TCP

7 : Minuteur ConnectRetry expiré

8 : Minuteur Hold Time expiré

9 : Minuteur KeepAlive expiré

10 : Réception d'un message OPEN

11 : Réception d'un message KEEPALIVE

12 : Réception d'un message UPDATE

13 : Réception d'un message NOTIFICATION

Le message UPDATE : attributs de la route

- Classés en 4 catégories :
  - Reconnus, obligatoires
    - ORIGIN, AS_PATH, NEXT_HOP
  - Reconnus, non-obligatoires
    - LOCAL_PREF, ATOMIC_AGGREGATE
  - Optionnels, annonçables (transitifs ou non)
    - MULTI_EXIT_DISC (MED), AGGREGATOR
  - Optionnels, non-annonçables
    - WEIGHT (spécifique à Cisco)

Tout ces attributs de route concernent le cas principal qui est l'E-BGP. Un seul est spécifique de l'I-BGP, c'est le LOCAL_PREF qui n'est annoncé qu'à l'intérieur de l'AS dans les sessions I-BGP.

Pour un attribut de route, le fait d'appartenir à la catégorie «reconnu» impose au processus BGP de savoir le traiter s'il est présent dans une annonce.

Inversement, s'il appartient à la catégorie «optionnel» un processus BGP n'est pas dans l'obligation de savoir le prendre en compte pour le traiter.

Le caractère «transitif» d'un attribut lui donne une portée illimitée.

Le caractère «non-transitif» d'un attribut limite sa portée à l'AS (ex. LOCAL_PREF) ou à l'AS voisin (ex. MED).

Les attributs de route obligatoires (1)

- ORIGIN
  - Donne l'origine de la route, peut prendre 3 valeurs :
    - IGP : la route est intérieure à l'AS d'origine
    - EGP : la route a été apprise par **le protocole** EGP
    - Incomplète : l'origine de la route est inconnue ou apprise par un autre moyen (redistribution des routes statiques ou connectées dans BGP par exemple)

On ne voit dans la pratique que les valeurs "IGP" ou "Incomplete" qui sont positionnées. (même sur des routeurs de points d'échange qui connaissent environ 130 000 routes), le protocole EGP n'étant plus utilisé.

Dans les implémentations de Cisco ou de Zebra, les valeurs «IGP», «EGP» ou «incomplete», sont respectivement représentées par les lettres «i», «e» ou «?» dans les représentations des tables d'informations de routage.

Exemple (Cisco ou Zebra) :

cs7206>sh ip bgp

BGP table version is 28403, local router ID is 194.199.17.59

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal

Origin codes: **i** - IGP, **e** - EGP, **?** - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path	
*> 129.88.0.0	193.54.188.1	19		0	64515	**i**
*> 129.88.1.250/32	193.54.188.1	19		0	64515	**?**
*> 129.88.1.254/32	193.54.188.1	11		0	64515	**?**
*> 129.88.3.0/24	193.54.188.1	18		0	64515	**?**
*> 129.88.100.0/24	194.199.17.35	0		32768		**i**
*> 129.88.103.0/24	193.54.188.1	20		0	64515	**?**
*> 129.88.253.0/24	193.54.188.1	20		0	64515	**?**
*> 132.168.0.0	193.54.188.5	0		0	2063	**i**

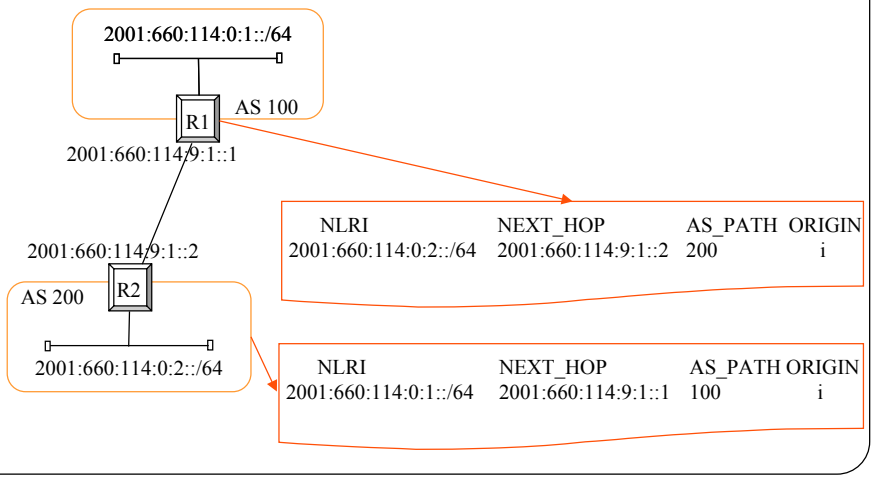
Les attributs de route obligatoires (2)

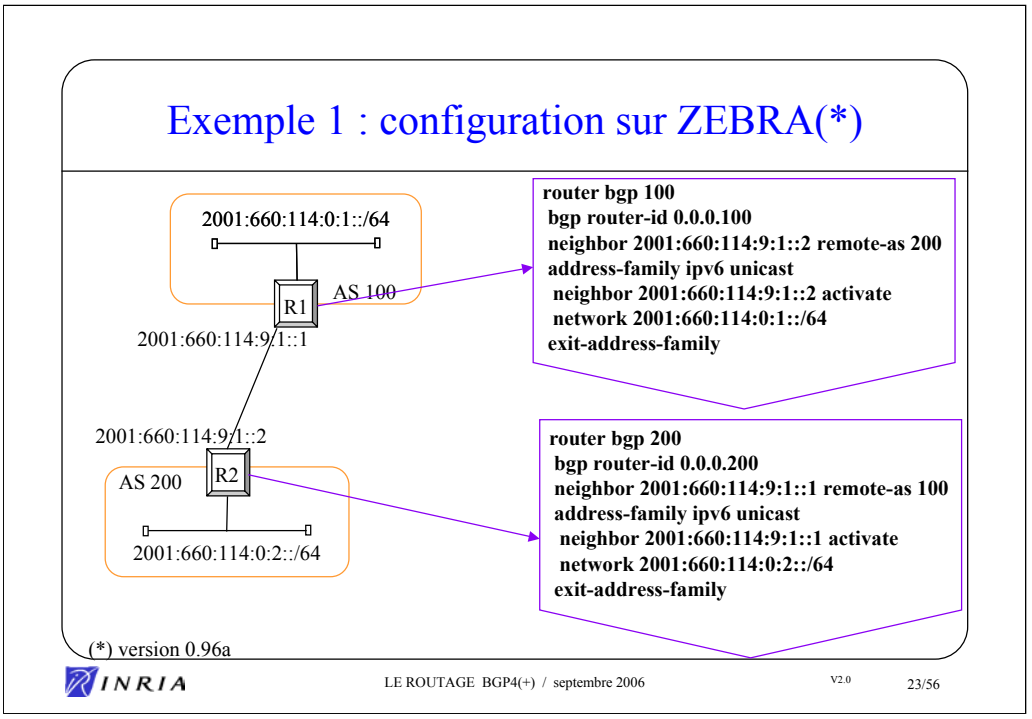
- AS_PATH
  - Donne la route sous forme d'une liste de segments d'AS
  - Les segments sont ordonnés ou non (AS_SET)
  - Chaque routeur rajoute son numéro d'AS aux AS_PATH des routes qu'il a apprises avant de les ré-annoncer
- NEXT_HOP
  - Donne l'adresse IP du prochain routeur qui devrait être utilisé (peut éviter un rebond si plusieurs routeurs BGP sont sur un même réseau local)

Les segments d'AS non ordonnés sont formés par un routeur qui a fait une opération d'agrégation. Ce dernier regroupe dans cet ensemble non ordonné tous les AS associés aux routes qu'il a agrégées. Cela permet aux autres routeurs de continuer à détecter d'éventuelles boucles concernant ces routes.

Dans l'implémentation de Cisco, les segments d'AS dans un AS_PATH sont encadrés par des accolades {}.

Exemple 1 : tables Adj-RIB-in





Noter que l'annonce des réseaux internes de l'AS se fait par une directive "network" qui positionne aussi l'attribut ORIGIN à la valeur "IGP" (cf. planche précédente).

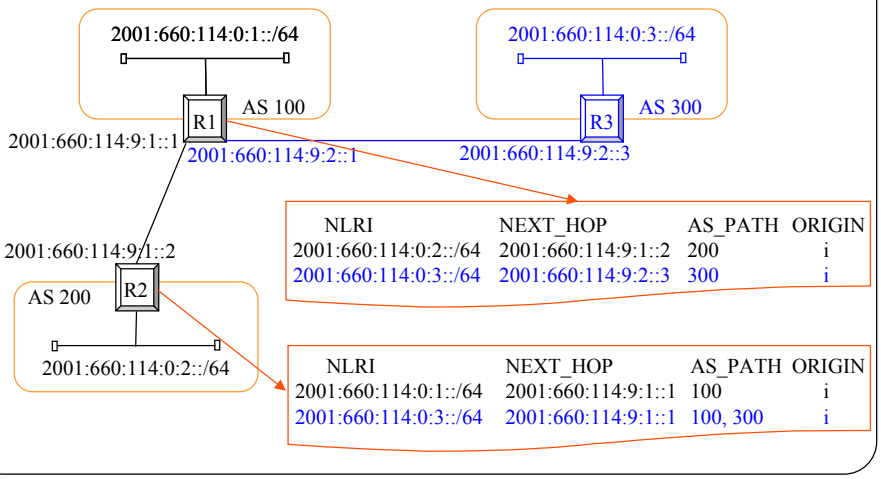
Attention, cette directive n'a pas du tout le même sens qu'avec certains IGP (ex. OSPF), de plus les implémentations de Cisco et Zebra diffèrent sensiblement quand à l'effet d'une directive "network" :

Pour Cisco, cette directive ne fait que positionner l'attribut ORIGIN à IGP, mais ne force pas l'annonce de la route concernant ce réseau en BGP. Cette annonce est conditionnée au fait que le routeur sache bien router ce réseau. Ce comportement est normal pour un routeur, mais une instabilité de l'IGP interne à l'AS se propage hors de l'AS et peut s'avérer pénalisant).

Pour Zebra, cette directive positionne l'attribut ORIGIN à IGP, et provoque l'annonce de la route concernant ce réseau en BGP. Ce comportement évite les instabilité d'annonces de route, mais peut provoquer un trafic inutile sur le lien inter-AS.

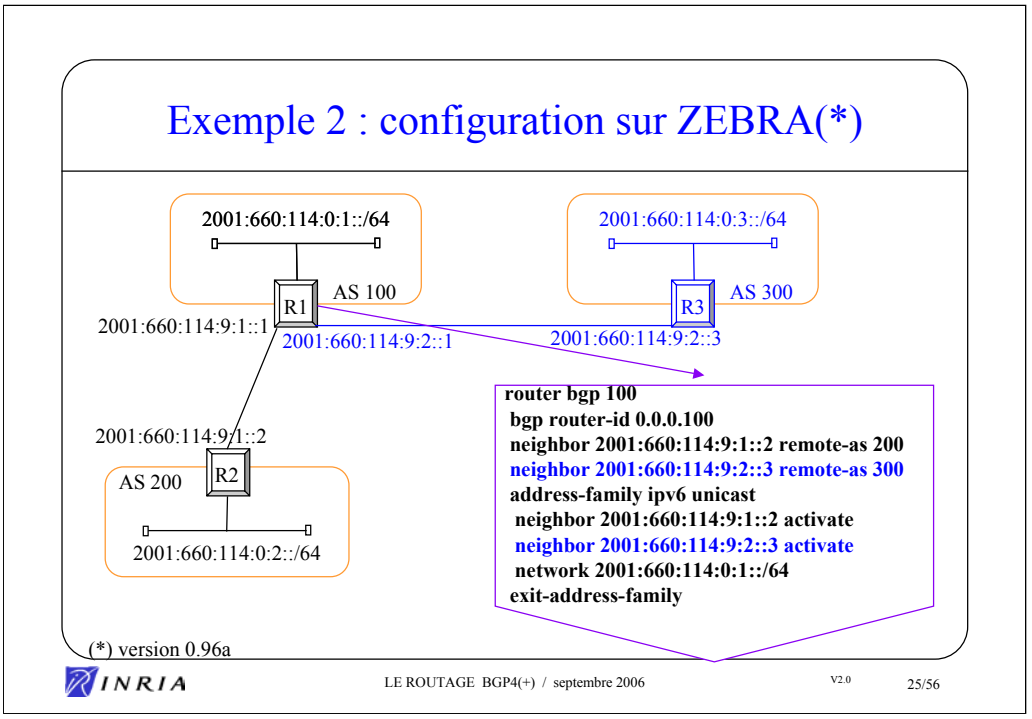
Noter aussi que la première directive neighbor (ex `neighbor 2001:660:114:9:1::2 remote-as 200`) identifie le voisin et le protocole IP de transport (IPv6 dans cet exemple). La deuxième directive neighbor (ex `neighbor 2001:660:114:9:1::2 activate`) qui est positionnée dans la séquence spécifique au protocole IPv6 (ex `address-family ipv6 unicast`) active spécifiquement des échanges d'informations de routage concernant le protocole IPv6.

Exemple 2 : tables Adj-RIB-in



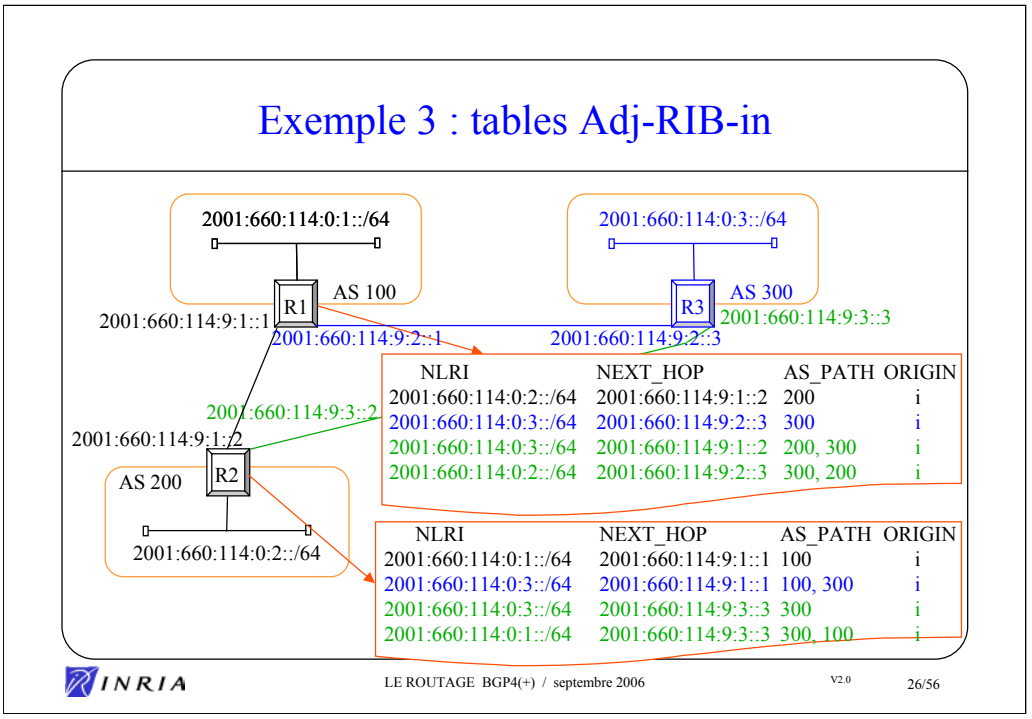
NLRI	NEXT_HOP	AS_PATH	ORIGIN
2001:660:114:0:2::/64	2001:660:114:9:1::2	200	i
2001:660:114:0:3::/64	2001:660:114:9:2::3	300	i

NLRI	NEXT_HOP	AS_PATH	ORIGIN
2001:660:114:0:1::/64	2001:660:114:9:1::1	100	i
2001:660:114:0:3::/64	2001:660:114:9:1::1	100, 300	i



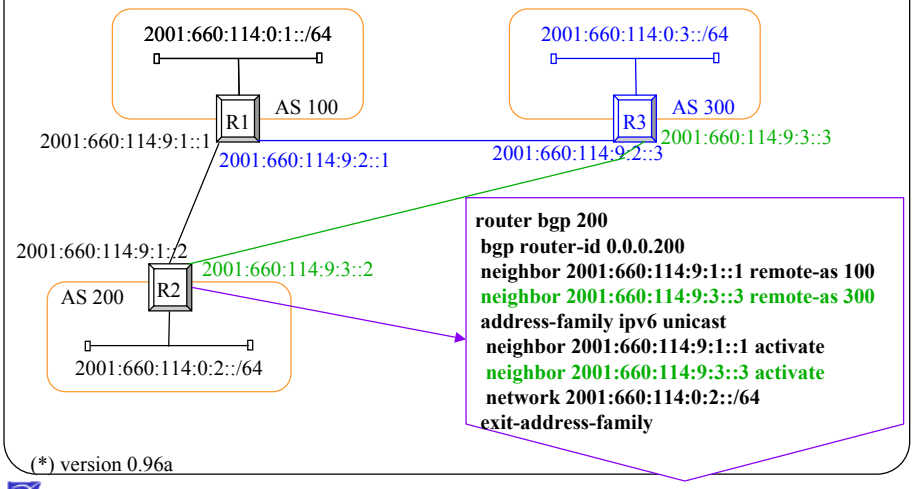
La configuration de R3 est symétrique de celle de R2.

L'AS 100 qui sert d'AS de transit pour les AS 200 et 300 doit synchroniser les annonces entre BGP et l'IGP interne à l'AS. Sinon un effet de «trou noir» peut se produire.



Noter la présence de plusieurs routes pour le même réseau dans les tables des routeurs R1, R2 (et R3 par symétrie).

Exemple 3 : configuration sur ZEBRA(*)



(*) version 0.96a

Les attributs de route optionnels (1)

- LOCAL_PREF (non transitif, discretionary)
  - Pondère la priorité donnée aux routes en interne à l'AS
  - Jamais annoncé en E-BGP
- ATOMIC_AGGREGATE (transitif, discretionary)
  - Indicateur d'agrégation
  - Quand des routes plus précises ne sont pas annoncées
- AGGREGATOR (transitif)
  - Donne l'AS qui a formé la route agrégée
  - L'adresse IP du routeur qui a fait l'agrégation

L'attribut LOCAL_PREF est un puissant outil d'expression de la politique de routage à l'intérieur d'un AS car il est pris en compte avant la longueur de l'AS_PATH dans le choix entres des routes concurrentes.

Noter le caractère non-transitif de l'attribut de route LOCAL_PREF qui n'est donc pas transmis hors de l'AS.

Les attributs de route optionnels (2)

- MULTI_EXT_DISC ou MED (non transitif)
  - Permet de discriminer les différents points de connexion d'un AS multi-connecté (plus faible valeur préférée)
- WEIGHT (non transitif, spécifique Cisco)
  - Pondère localement (au routeur) la priorité des routes BGP
- COMMUNITY (transitif)
  - Pour un ensemble de routeurs ayant une même propriété
  - Trois valeurs reconnues
    - no-export : pas annoncé aux voisins de la confédération
    - no-advertise : pas annoncé aux voisins BGP
    - no-export-subconfed : pas annoncé en E-BGP

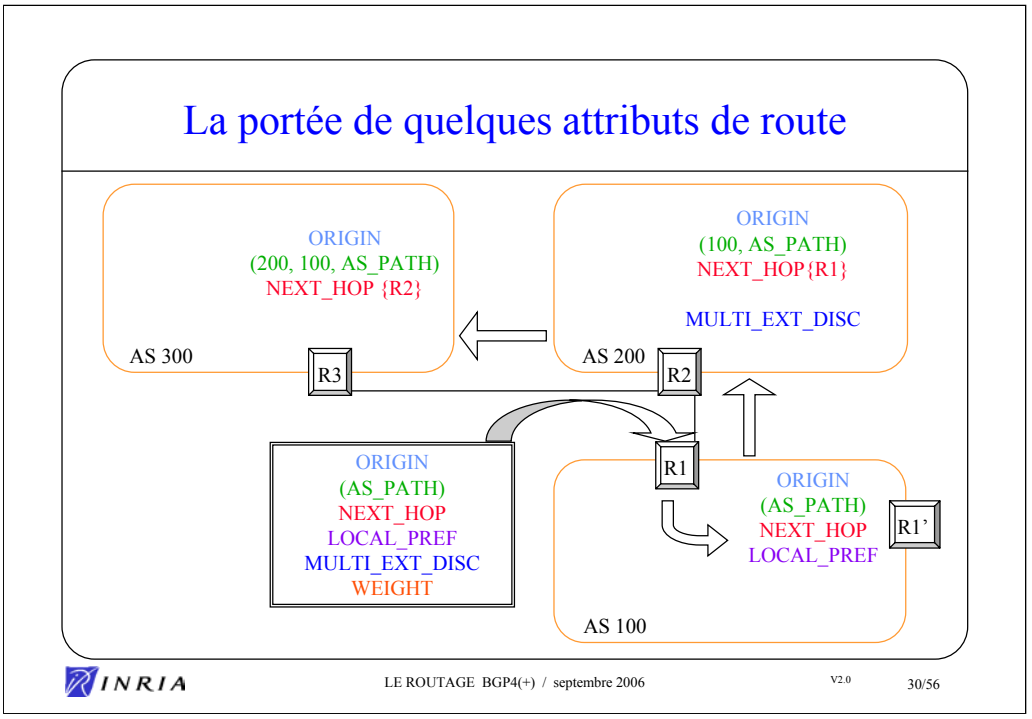
Dans la version 3 de BGP, l'attribut MED était appelé Inter-AS_Metric, l'implémentation Cisco de BGP-4 a gardé le terme de Metric pour certaines commandes manipulant le MED. Cette implémentation permet aussi de comparer des MED d'AS différents (*bgp always-compare-med* sur IOS Cisco).

On pourra aussi consulter le RFC4451 'BGP MULTI_EXT_DISC (MED) Considerations' sur l'usage et la pratique de l'attribut MED.

Pour l'attribut COMMUNITY, le standard (RFC1997) recommande de coder le numéro d'AS dans les deux premiers octets, les 2 suivants étant laissés à disposition des administrateurs de l'AS. Une communauté de routeurs peut s'étendre sur plusieurs AS. L'implémentation de Cisco connaît une valeur prédéfinie égale à «internet».

L'attribut WEIGHT peut varier de 0 à 32768, les plus grandes valeurs sont préférées. Par défaut, il prend les valeurs suivantes :

- 32768 si la route est initiée par le routeur
- 0 pour les autres routes



L'attribut MED n'est pas annoncé dans l'AS du routeur de bord R1 mais à son voisin extérieur (qui ne le propage pas).

L'attribut LOCAL_PREF n'est annoncé que dans l'AS du routeur de bord.

L'attribut NEXT_HOP est modifié par chaque traversée d'AS.

L'attribut AS_PATH est préfixé du numéro d'AS à chaque traversée d'AS.

L'attribut ORIGIN n'est jamais modifié.

Le processus de décision (1)

- Il est enclenché par une annonce de route
- Il se déroule en trois phases
  - Calcul du degré de préférence de chaque route apprise
  - Choix des meilleures routes à installer dans RIB-Loc
  - Choix des routes qui vont être annoncées
- Il applique aux informations de routage un traitement basé sur
  - Critères techniques : suppression boucles, optimisations, …
  - Critères administratifs : application de la politique de routage de l'AS.

Pour être prise en compte une annonce de route doit avoir son NEXT_HOP routable.

Une route interne n'est annoncée par un routeur que s'il sait la joindre.

Une route externe n'est annoncée par un routeur que s'il sait joindre le NEXT_HOP.

Une route dont l'attribut NEXT_HOP est l'adresse IP du voisin n'est pas annoncée à ce voisin (qui la connaît déjà !).

Le processus de décision (2)

- Critères de choix entre 2 routes (priorités décroissantes) :
  - WEIGHT (propriétaire Cisco, plus grand préféré)
  - **LOCAL_PREF le plus grand**
  - Route initiée par le processus BGP local
  - **AS_PATH minimum**
  - ORIGIN minimum (IGP -> EGP -> Incomplete)
  - **MULTI_EXT_DISC minimum**
  - Route externe préférée à une route interne (à l'AS)
  - Route vers le plus proche voisin local (au sens de l'IGP)
  - Route vers le routeur BGP de plus petite adresse IP

L'installation d'une route dans la table de routage doit prendre en compte le fait qu'une route peut être apprise par plusieurs protocoles de routage différents. L'implémentation de Cisco utilise la notion de distance administrative pour cela. Le choix entre 2 routes se fait en prenant celle qui a la distance administrative la plus faible. Les valeurs par défaut des distances administratives associées aux origines des routes sont :

Route directement connectée	0
Route statique	1
Route apprise en E-BGP	20
Route apprise en EIGRP (interne)	90
Route apprise en IGRP	100
Route apprise en OSPF	110
Route apprise en ISIS	115
Route apprise en RIP	120
Route apprise en EGP	140
Route apprise en EIGRP (externe)	170
Route apprise en I-BGP	200
Route apprise en BGP (local)	200
Route d'origine inconnue	255

Différences entre E-BGP et I-BGP

- ❑ Une annonce reçue en I-BGP n'est pas ré-annoncée en I-BGP
- ❑ L'attribut LOCAL_PREF n'est annoncé qu'en I-BGP
- ❑ Seuls les voisins E-BGP doivent être directement connectés
- ❑ Les annonces I-BGP ne modifient pas l'AS_PATH
- ❑ Les annonces I-BGP ne modifient pas le NEXT_HOP
- ❑ Le MED n'est pas annoncé en I-BGP

Le traitement différent appliqué aux attributs de route suivant que le voisin BGP est externe ou interne est résumé dans le tableau suivant :

ATTRIBUT	E-BGP	I-BGP
AS_PATH	=(local AS+AS_PATH)	non modifié si reçu en E-BGP
NEXT_HOP	=@IP annonceur	non modifié
MED	=métrique	non annoncé
LOCAL_PREF	pas annoncé	annoncé
ATOMIC_AGGREGATE		
AGGREGATOR		

Certains minuteurs (vus plus loin) sont aussi traités différemment :

MINUTEUR	E-BGP	I-BGP
MinRouteAdvertisement	pris en compte	pas pris en compte (pour accélérer la convergence dans l'AS)

L'annonce des routes internes d'un AS

- Statique
  - Pas d'instabilité de routage, mais trous noirs possibles
  - Exemples en IOS
    - ***redistribute [static|connected]*** -> ORIGIN: Incomplete
    - ***network <adresse réseau>*** -> ORIGIN: IGP
- Dynamique
  - Suit au mieux l'état du réseau, nécessite du filtrage
  - Exemples en IOS
    - ***redistribute <paramètres de l'IGP>*** -> ORIGIN: IGP

La redistribution de routes apprises dynamiquement est difficile à contrôler. Il est nécessaire de faire attention à ne pas faire boucler la redistribution de routes entre l'IGP et BGP (notamment la route par défaut !).

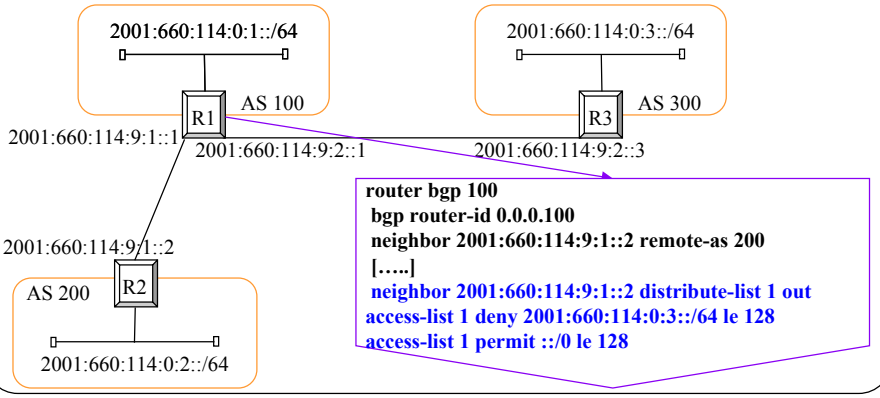
L'annonce statique est à préférer à priori pour annoncer les routes internes de l'AS par BGP.

La politique de routage

- Elle peut influencer :
  - Le traitement des routes reçues
  - Le traitement des routes annoncées
  - L'interaction avec les IGP de l'AS
- En pratique elle s'exprime par :
  - Du filtrage de **réseaux**
  - Du filtrage de **routes** (AS_PATH)
  - De la manipulation **d'attributs de routes**

Politique de routage : exemple de filtrage de réseaux sur ZEBRA

- Filtrage des réseaux annoncés : AS100 ne veut pas servir d'AS de transit pour le réseau 2001:660:114:0:3::/64 de l'AS300

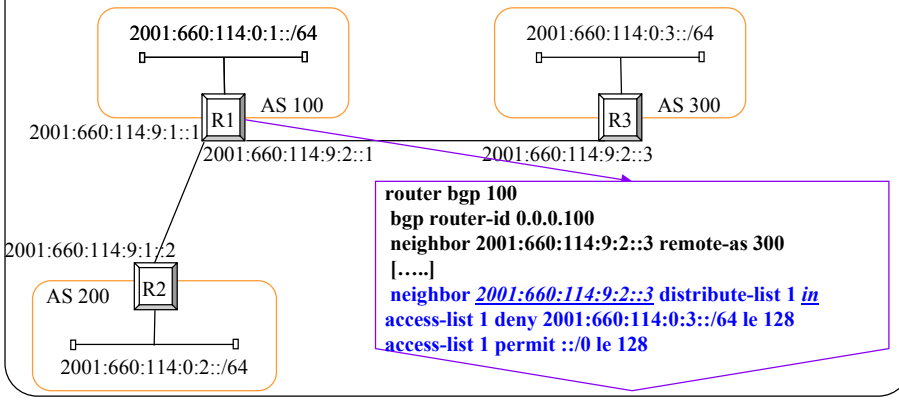


Le filtrage BGP s'appuie sur le même mécanisme des *access-list* qui est utilisé pour le filtrage des paquets IP. L'application de *l'access-list* à une session BGP (au lieu d'une interface dans le cas de filtrage de paquets IP) permet d'éliminer certains réseaux d'une annonce reçue (paramètre 'in') ou faite (paramètre 'out').

Dans le cas ci-dessus, *l'access-list* est à appliquer à toutes les autres sessions BGP que pourrait avoir le routeur R1.

Politique de routage : exemple de filtrage de réseaux sur ZEBRA

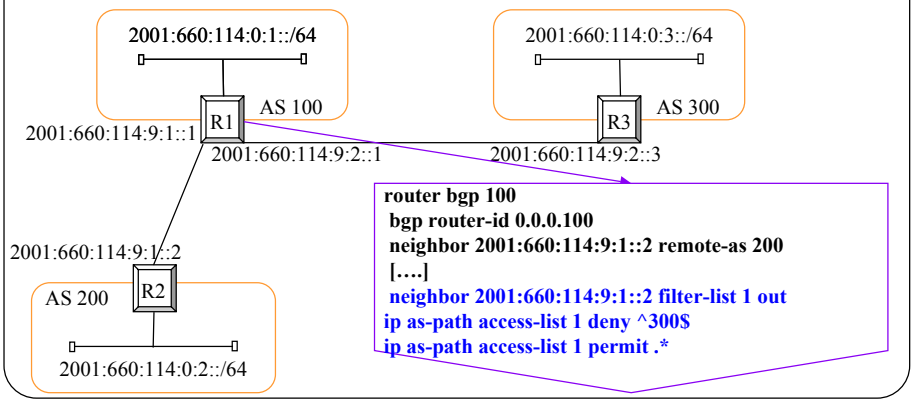
- Filtrage des réseaux annoncés : AS100 ne veut pas servir d'AS de transit pour le réseau 2001:660:114:0:3::/64 de l'AS300 *(variante)*



Dans cette variante, l'élimination de l'annonce du réseau 2001:660:114:0:3::/64/24 empêche bien le transit car ce réseau ne sera pas réannoncé, mais en plus, l'AS 100 ne sera pas capable de router ce réseau.

Politique de routage : exemple de filtrage de routes sur ZEBRA

- Filtrage des AS_PATH annoncés : AS100 ne veut pas servir d'AS de transit pour tous les réseaux internes d'AS300



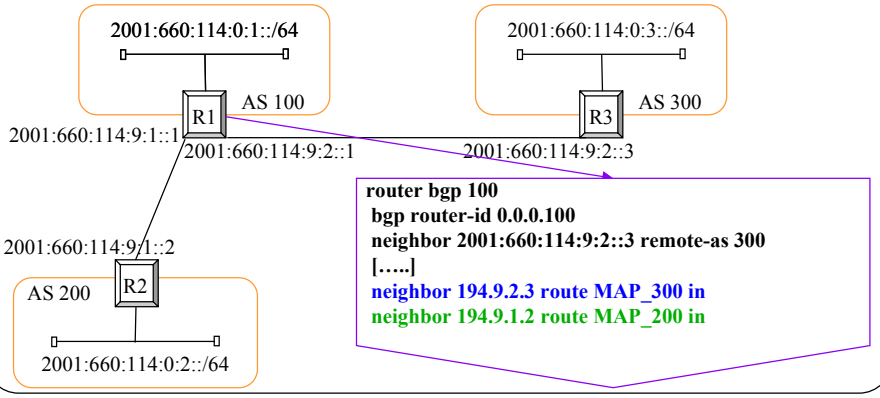
Les AS_PATH étant des chaînes de caractères, l'identification et la localisation d'un AS ou d'un ensemble d'AS se fait par des expressions régulières, en utilisant le sous ensemble suivant de caractères spéciaux :

Caractère	Symbole	Signification
Point	.	Représente n'importe quel caractère
Astérisque	*	Représente 0 ou N fois le caractère précédent
Plus	+	Représente 1 ou N fois le caractère précédent
Interrogation	?	Représente 0 ou 1 fois le caractère précédent
Circonflexe	^	Représente le début de la chaîne de caractères
Dollar	$	Représente la fin de la chaîne de caractères
Souligné	_	Représente l'un des 5 caractères servant à délimiter les N° d 'AS soit: ,{}() le début ou fin de chaîne
Crochet ouvrant	[	Début d'un intervalle
Crochet fermant	]	Fin d'un intervalle
Tiret	-	Sépare les 2 caractères définissant l'intervalle

Les 3 derniers caractères spéciaux s'utilisent conjointement, par exemple l'intervalle noté [1-6] représente un chiffre compris entre 1 et 6 inclus.

Politique de routage : exemple de manipulation sur ZEBRA

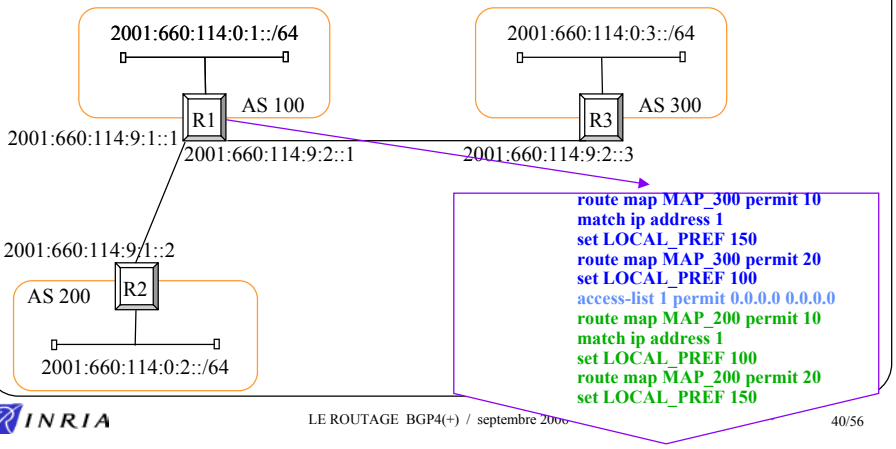
- Filtrage par route map : AS100 veut privilégier la route par défaut annoncée par AS300



Nota : confédération de routeurs dans AS100

Politique de routage : exemple de manipulation sur ZEBRA (suite)

- Filtrage par route map : AS100 veut savoir router uniquement 2001:660:114:0:3::/64, mais sans servir d'AS de transit pour AS300

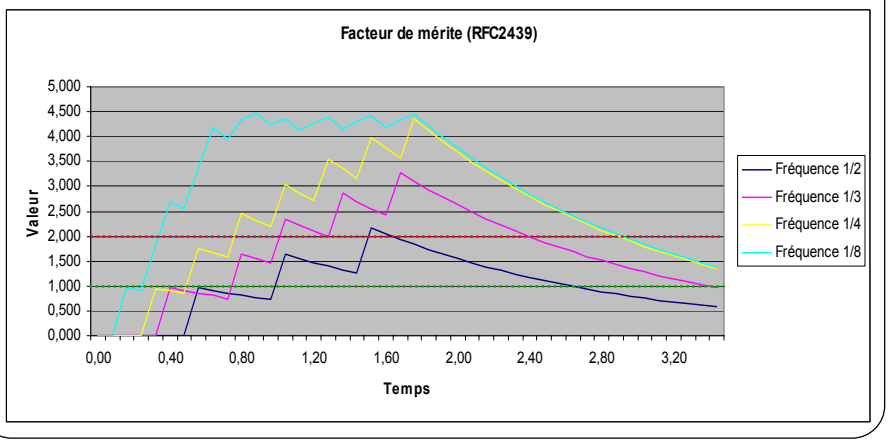


Optimisations : stabilité du routage (1)

- Les routes instables sont pénalisées
  - À chaque instabilité => X points de pénalité
  - Si Pénalité > limite L1 => route supprimée
  - Si Pénalité < limite L2 => route rétablie
  - Si : pas de nouvelle pénalité pendant T1 => Pénalité/2
  - Si Pénalité < limite L3 => on oublie tout
- Ne concerne que les annonces E-BGP

Optimisations : stabilité du routage (2)

❑ Allure du facteur de mérite associé à une route instable



Dans l'implémentation IOS de Cisco, on a :

- Pénalité pour une instabilité (X) = 1000 points
- Limite de suppression d'une route (L1) = 2000 points
- Limite de réutilisation d'une route (L2) = 750 points
- Valeur d'oubli des informations de pénalisation (L3) = 350 points
- Demie vie de la pénalisation (T1) = 120 secondes

Cette technique de pénalisation des routes instables est justifiée et standardisée dans le RFC2439.

Optimisations : contrôle du trafic BGP

- On peut agir sur différents minuteurs
  - MinRouteAdvertisementInterval
  - MinASOriginationInterval
  - La gigue dans la fréquence des annonces
- On peut réduire le volume des informations annoncées
  - NLRI agrégés
  - AS_PATH condensés

MinRouteAdvertisementInterval est le temps minimum entre 2 annonces de routes vers des voisins externes (uniquement).

MinASOriginationInterval est le temps minimum entre 2 annonces résultant d'une mise à jour des routes internes de l'AS (en provenance de l'IGP par exemple).

Le facteur de gigue est un paramètre global au routeur. C'est un nombre aléatoire à valeur dans l'intervalle [0,75-1] qui pondère l'ensemble des 5 minuteurs du processus BGP.

Rappel des valeurs (en secondes) des minuteurs d'un processus BGP :

Minuteur	Valeur suggérée par le RFC1771	implémentation Cisco	implémentation Zebra
ConnectRetry	120		
Hold Time	90	180	180
KeepAlive	30	60	60
MinRouteAdvertisementInterval	30	30	0
MinASOriginationInterval	15		

Optimisation : sécurisation des échanges BGP

- Mesures natives au protocole
  - Session BGP = {@IP1,numéro AS1},{@IP2,numéro AS2}
  - Signature MD5 de chaque message
- Compléments : mesures standard au niveau TCP ou IP
  - Filtrage du port 179
- MAIS : a toutes les vulnérabilités de TCP ou IP
  - Déni de service

Optimisations : les réflecteurs de routes

- Permet d'éviter une croissance en N^2 des sessions I-BGP
- Mais rajoute un point de panne singulier
- On met donc plusieurs réflecteurs de route par AS

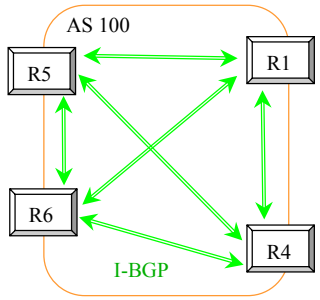


Schéma sans réflecteur de routes

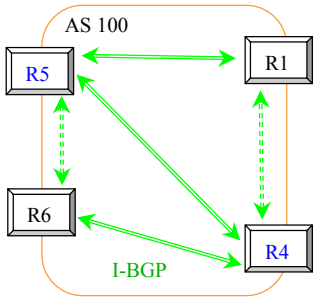


Schéma avec 2 réflecteurs de routes (R4 et R5)

Introduit à titre expérimental en 1996 par le RFC1966, modifié en 2000 par le RFC2756, actuellement défini comme standard par le RFC4456 (d'avril 2006).

Si l'on considère un AS avec N routeurs de bord, il aura un nombre de sessions I-BGP qui sera égal à : N(N-1)/2 sans réflecteur de route. Si certains de ces routeurs de bord sont aussi réflecteurs de route, le nombre de sessions I-BGP sera plus faible, mais pourra varier entre 2 valeurs suivant le niveau de redondance que l'on souhaite (noter que tous les réflecteurs de route doivent être complètement maillés en sessions I-BGP).

Ainsi le nombre de sessions I-BGP sera compris entre :

N-1	et	2N-3	avec 2 réflecteurs de route
N	et	3(N-2)	avec 3 réflecteurs de route
N-R + R(R-1)/2	et	NR - R(R+1)/2	avec R réflecteurs de route

La fonction Fmin(N,R)=N-R + R(R-1)/2 a une valeur minimale pour R=3/2 quel que soit N. Les valeurs entières de R qui la minimisent sont donc 1 et 2.

La fonction Fmax(N,R)=NR-R(R+1)/2 a une valeur maximale pour R=N-1/2 donc plus R est petit plus la valeur de Fmax sera faible.

Du point de vue de la minimisation du nombre de sessions I-BGP, la valeur optimale de R est donc égale à 2 quel que soit N, si l'on veut assurer une redondance des réflecteurs de route. Dans ce cas on a 2N-3 sessions I-BGP.

Extensions : les confédérations d'AS

- Permet de réduire le nombre de sessions I-BGP
- En divisant l'AS en mini-AS (ou sous AS)
- Les routeurs de bord d'un mini-AS établissent des sessions
  - I-BGP entre eux (maillage complet)
  - E-BGP avec leurs voisins d'autres AS
  - Pseudo E-BGP avec leurs voisins d'autres minis-AS
- Vu de l'extérieur, la confédération d'AS apparaît comme un seul et unique AS

Pour bien apparaître comme faisant partie d'un même AS vis-à-vis de l'extérieur, les routeurs de bord de deux mini-AS différents échangent des sessions E-BGP (car leurs numéros d'AS sont différents), mais ces sessions suivent les mêmes règles de modification des attributs de route que les sessions I-BGP. Lors de ces sessions, les attributs NEXT_HOP, MED, et LOCAL_PREF ne sont donc pas modifiés.

Les confédérations d'AS ont été introduites en 1996 à titre expérimental par le RFC1965, puis standardisées en 2001 par le RFC3065.

Extensions : les groupements de routeurs

- Les routeurs BGP d'un groupement partagent la même politique de routage (ex. routes maps, filtres d'annonces, …)
- Cette politique est définie sur l'un des routeurs du groupement
- Elle est propagée automatiquement sur les autres routeurs
- Un routeur du groupement peut modifier localement sa politique de routage (mais ne la propage pas aux autres)

Extensions : les serveurs de route

- Sur un grand point d'échange on peut avoir :
  - 100 fournisseurs d'accès Internet
  - Plus de 180 000 routes annoncées (en 2006)
- Ce qui pourrait impliquer :
  - Jusqu'à 10 000 sessions TCP !
- Solution : les serveurs de route
  - Réduit le nombre de sessions (quelques unes par fournisseur d'accès)

Introduit à titre expérimental en 1995 par le RFC1863, et classé en historique en 2005 par le RFC4223.

Extensions : le routage multi-protocole (IPv6)

- Dans BGP, seuls 3 attributs de route de dépendent d'IPv4
  - NLRI, NEXT_HOP, (AGGREGATOR)
- Pour rendre BGP multi-protocole, on introduit 2 attributs de route supplémentaires
  - MP_REACH_NLRI (optionnel, non-transitif)
  - MP_UNREACH_NLRI (optionnel, non-transitif)
- L'attribut de route MP_REACH_NLRI contient des triplets
  - *Adress_family* (ex. IPv4, IPv6, IPX), NEXT_HOP, NLRI
- Un message UPDATE contient MP_REACH_NLRI et les autres attributs de route déjà vus (ORIGIN, LOCAL_PREF...)

Introduit comme standard par le RFC2858. Les seules modifications de configurations correspondent aux format des adresses IPv6.

Exemple de configuration d'une session BGP en IPv6 sous Zebra :

```
router bgp 65400
 bgp router-id 192.108.119.167
 ipv6 bgp neighbor 2001:660:281:8::1 remote-as 1938
```

Exemple d'affichage des informations BGP en IPv6 sous Zebra :

```
bgpd# sh ipv6 bgp
BGP table version is 0, local router ID is 192.108.119.167
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network                Metric LocPrf Weight Path
*> ::194.182.135.0/120                       0 1938 2200 1103 766 278 6435 i
     2001:660:281:8::1(fe80::83fe:c80a)
*> 2001:200::/35                             0 1938 2200 3425 2500 i
     2001:660:281:8::1(fe80::83fe:c80a)
*> 2001:200:12a::/48                         0 1938 2200 5511 3549 ?
     2001:660:281:8::1(fe80::83fe:c80a)
[….]
```

Exemple de configuration BGP en IPv6 (Zebra)

```
router bgp 65400
 bgp router-id 192.108.119.167
 ipv6 bgp neighbor 2001:660:281:8::1 remote-as 1938
 ipv6 bgp neighbor 2001:660:281:8::1 prefix-list filtre_nlri in
 ipv6 bgp neighbor 2001:660:281:8::1 filter-list filtre_as in
!
ipv6 prefix-list filtre_nlri description    Refus des annonces de son préfixe et du 2002::/16
ipv6 prefix-list filtre_nlri seq 5 deny 3ffe:305:1014::/48 le 128
ipv6 prefix-list filtre_nlri seq 10 deny 2002::/16 le 128
ipv6 prefix-list filtre_nlri seq 15 permit any
!
ip as-path access-list filtre_as deny 1938 2200 5511 *
ip as-path access-list filtre_as permit .*
```


Résultat sous Zebra de la configuration ci-dessus (commande **'sh ipv6 bgp neighbors'**)

BGP neighbor is 2001:660:281:8::1, remote AS 1938, external link
BGP version 4, remote router ID 131.254.200.10
BGP state = Established, up for 00:04:16
Last read 00:00:16, hold time is 180, keepalive interval is 60 seconds
Neighbor capabilities:
Route refresh: advertised and received(old and new)
[…..]
For address family: IPv6 Unicast
Community attribute sent to this neighbor
Inbound path policy configured
Incoming update prefix filter list is *filtre_nlri
Incoming update AS path filter list is *filtre_as
225 accepted prefixes

Connections established 1; dropped 0
Local host: 2001:660:281:8::2, Local port: 1190
Foreign host: 2001:660:281:8::1, Foreign port: 179
Nexthop: 192.108.119.167
Nexthop global: 2001:660:281:8::2
Nexthop local: ::
BGP connection: non shared network
Read thread: on Write thread: off

Extensions : le routage multicast (MBGP)

- Vu comme un cas particulier du routage multi-protocole
- Utilisation de la notion de sous famille d'adresse
- Implémentations récentes (IOS, ….)

Introduit comme standard par le RFC2858 en juin 2000.

Extensions : annonce de capacité

- Standardisé initialement en mai 2000 par le RFC2842 (statut PS)
- Standardisé définitivement en novembre 2002 par le RFC3392 (DS)
- Introduit un paramètre optionnel : *capabilities*
- Annonce les capacités fonctionnelles d'un routeur lors de l'OPEN
- Permet une mise à niveau automatique des fonctionnalités utilisées dans cette session BGP
- Permettra des mises à niveau des implémentations de BGP non synchrones

Exemple sous Ios/Cisco du résultat de la commande **'sh ipv6 bgp neighbors'** :

BGP neighbor is 2001:660:281:1::1, remote AS 1938, external link
BGP version 4, remote router ID 131.254.200.10
BGP state = Established, up for 16:42:08
Last read 00:00:08, hold time is 180, keepalive interval is 60 seconds
Neighbor capabilities:
Route refresh: advertised and received
Address family IPv6 Unicast: advertised and received
Received 5601 messages, 0 notifications, 0 in queue
Sent 3785 messages, 0 notifications, 0 in queue
Route refresh request: received 0, sent 0
Minimum time between advertisement runs is 30 seconds

For address family: IPv6 Unicast
BGP table version 3272, neighbor version 3272
Index 1, Offset 0, Mask 0x2
264 accepted prefixes consume 16896 bytes
Prefix advertised 4, suppressed 0, withdrawn 0

Bibliographie : principaux RFC sur BGP

- RFC1772 Application of the Border Gateway Protocol in the Internet. Y Rekhter, P. Gross. 03/1995. (DS)
- RFC1773 Experience with the BGP-4 protocol. P. Traina. 03/1995. (INFO)
- RFC1774 BGP-4 Protocol Analysis. P. Traina, Editor. 03/1995. (INFO)
- RFC1997 BGP Communities Attribute. R. Chandra, P. Traina & T. Li. 06/1996. (PS)
- RFC1998 An Application of the BGP Community Attribute in Multi-home Routing. E. Chen & T. Bates. 06/1996. (INFO)
- RFC2042 Registering New BGP Attribute Types. B. Manning. 01/1997. (INFO)
- RFC2385 Protection of BGP Sessions via the TCP MD5 Signature Option. A. Heffernan. 08/1998. (PS)
- RFC2439 BGP Route Flap Damping. C.Villamizar, R.Chandra, R.Govindan. 11/1998. (PS)
- RFC2457 Definitions of Managed Objects for Extended Border Node. B. Clouston, B. Moore. 11/1998. (PS)
- RFC2545 Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing. P. Marques, F. Dupont. 03/1999. (PS)
- RFC2858 Multiprotocol Extensions for BGP-4. T. Bates, Y. Rekhter, R. Chandra, D. Katz. 06/2000. (PS)

Bref historique de l'évolution du protocole BGP (voir RFC1773)

BGP-1 : RFC1105, juin 1989

BGP-2 : RFC1163, juin 1990

La hiérarchisation des AS est supprimée (notion de liens inter-AS haut/bas/horizontaux), introduction des attributs de routes, beaucoup de changements dans les formats des messages.

BGP-3 : RFC1267, octobre 1991

Détection et gestion des collisions d'ouvertures de sessions BGP, introduction d 'un identifiant de routeur, le NEXT_HOP peut être situé dans un autre AS que celui du routeur qui fait l'annonce.

BGP-4 : RFC1771, mars 1995

Ajout des adresses CIDR, introduction des ensembles d'AS (non ordonnés) dans les AS_PATH, et ajout des attributs de route MED (remplace INTER-AS METRIC), LOCAL-PRFERENCE, AGGREGATOR.

BGP-4+ : RFC2283 en février 1998, RFC2545 en mars 1999, RFC2858 en juin 2000

Extensions multiprotocoles (RFC2283, remplacé par le RFC2858)

Support d'IPv6 (RFC2545)

Routage multicast

Réflecteurs de routes, RFC2796 en avril 2000

Annonces de capacités, RFC2842 en mai 2000, puis RFC3302 en novembre 2002

Confédérations d'AS RFC3065 en février 2001

Ré-écriture complète du RFC1771 par le RFC4271 en janvier 2006

BGP/MPLS, RFC4364 février 2006

Interaction entre OSPF et BGP/MPLS, RFC4577, juin 2006

Bibliographie : principaux RFC sur BGP

- RFC2918 Route Refresh Capability for BGP-4. E. Chen, 09/2000. (PS)
- RFC3065 Autonomous System Confederations for BGP. P. Traina, D. McPherson, J. Scudder. 02/2001. (PS)
- RFC3107 Carrying Label Information in BGP-4. Y.Rekhter, E.Rosen. 02/2001.(PS)
- RFC3345 Border Gateway Protocol (BGP) Persistent Route Oscillation Condition. D. McPherson, V. Gill, D. Walton, A. Retana, 08/2002. (INFO)
- RFC3392 Capabilities Advertisement with BGP-4. R. Chandra, J. Scudder. 11/2002. (DS)
- RFC4271 A Border Gateway Protocol 4 (BGP-4). Y. Rekhter, T. Li., S. Hares. 01/2006. (DS)
- RFC4272 BGP Security Vulnerabilities Analysis. S. Murphy. 01/2006 (INFO)
- RFC4273 Definitions of Managed Objects for BGP-4. J. Haas, Ed., S. Hares, Ed.. 01/2006. (PS)
- RFC4274 BGP-4 Protocol Analysis. D. Meyer, K. Patel. 01/2006. (INFO)
- RFC4276 BGP-4 Implementation Report. S. Hares, A. Retana. 01/2006. (INFO)
- RFC4364 BGP/MPLS IP Virtual Private Networks (VPNs). E. Rosen, Y. Rekhter. 02/2006. (PS)
- RFC4456 BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP). T. Bates, E. Chen, R. Chandra. 04/2006. (DS)

Bibliographie : livres

- Le routage dans l'Internet, C. Huitema, Eyrolles, 1994
- Interconnections with bridges and routers, R, Perlman, Addison-Wesley, 1996
- Internet Routing Architectures, B.Halabi, Cisco Press, 1997
- BGP4 Inter-Domain Routing in the Internet, J. W. Stewart III, Addison-Wesley, 1999

Bibliographie : Sites web

- www.rsng.net : Route Server Next generation Project
- www.merit.net : Nombreuses informations sur les points d'échange de trafic entre opérateurs des USA.
- www.gated.org : Site de distribution du logiciel gated (payant) qui implemente la plupart des logiciels de routage (dont BGP4)
- www.zebra.org : Site de distribution du logiciel zebra (licence GPL) qui implemente la plupart des logiciels de routage (dont BGP4)
- www.caida.org : Propose des outils de métrologie réseau, beaucoup de données sur le trafic.
- www.merit.edu/~ipma/ : outils de mesure de performances, beaucoup d'informations sur les tables BGP de certains routeurs des points d'échange
- www.ep.net : Liste des points déchange
- www.ra.net : Routing Arbiter Project
- telnet://route-server.cerf.net : Accès en ligne a un routeur BGP
- http://www.cisco.com/univercd/cc/td/doc/cisintwk/ics/icsbgp4.htm : Manuel de référence des commandes BGP sur IOS de Cisco.
- www.mcvax.org/~jhma/routing/ : nombreuses statistiques sur les tables de routage BGP