



UNIVERSITE DE RENNES 1

BGP

Le routage inter-domaine

Bernard Cousin

Plan

- Présentation de BGP
- Le protocole BGP
- Les attributs de BGP
- Quelques particularités de BGP
- iBGP

BGP version 4

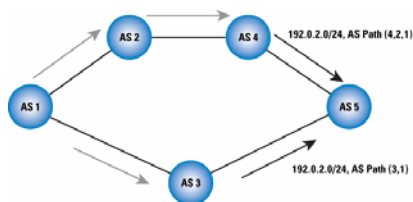
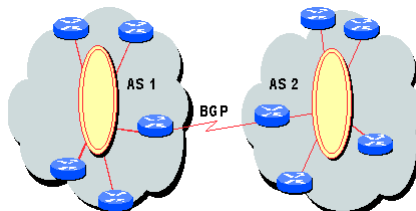
- Le protocole pour le routage inter-domaine :
 - "Exterior Gateway Protocol" : "[Border Gateway Protocol](#)"
 - Ne tient pas compte de la structure interne des AS
 - Gestion de plus de 90000 routes
- RFC 1771, mars 1995
- Sélection des routes basée sur
 - le préfixe le plus spécifique (cf. "best prefix match")
 - le meilleur chemin :
 - de plus court en nombre d' "Autonomous System" (AS) !
- Support essentiel au CIDR
 - "Classless InterDomain Routing" (CIDR)

Bibliographie

- Y. Rekhter, T.J. Watson, T. Li, "A Border Gateway Protocol 4 (BGP-4) ", Rfc 1771, March 1995
- C. Huitema, "Routing the Internet (partie III)", Prentice Hall, 2000
- Transparents, notamment les figures, sont inspirées de :
 - Dennis Ferguson (un des contributeurs à BGP)
 - Olivier Bonaventure (université de Louvain)

Introduction à BGP

- BGP est utilisé pour transporter des informations de routage entre AS
- Protocole à "path vector"
- Fonctionne au-dessus de TCP
 - Port 179
 - Fiabilité des transmissions



24 octobre 2007

Border Gateway Protocol

5

Les principes de BGP

- Apprentissage des chemins
 - Les préfixes des destinations + "next hop"
 - Grâce aux entités BGP (internes et externes)
- Sélectionne le meilleur chemin et configure la table de routage IP avec
 - Une variante du protocole de type "distance vector"
- La **politique d'administration du réseau**
 - Influe sur le processus de sélection du "meilleur chemin"
 - Utilise les attributs BGP

24 octobre 2007

Border Gateway Protocol

6

"Autonomous system"

- Un domaine de routage autonome (AS)
 - Un ensemble d'équipements (routeurs, stations) et de liens (LAN, switch, etc.)
 - Sous une même responsabilité administrative
 - Peut être très vaste ou non (mondial ou local), contenir un très grand nombre de routeurs et de stations ou un seul, etc.
- L'internet est composé (en janvier 2006) d'environ 40000 AS attribués
 - Numérotation $2^{16} \Rightarrow 2^{32}$, en 2007
- Les AS sont tous interconnectés
 - Cela permet d'envoyer un paquet à partir de partout à n'importe qui
 - En général un paquet traverse quelques AS avant de parvenir à son destinataire

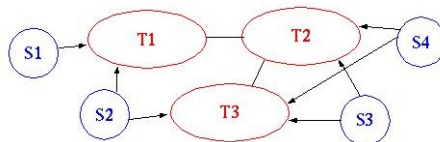
24 octobre 2007

Border Gateway Protocol

7

Les domaines de transit

- Un domaine de transit **autorise l'utilisation de son infrastructure** par certains domaines pour transmettre des paquets vers d'autres domaines
 - Exemple : UUNet, GEANT, Renater, etc.



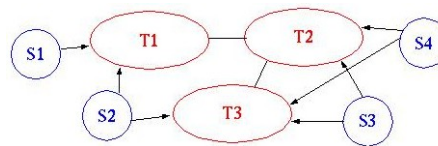
24 octobre 2007

Border Gateway Protocol

8

Domaine souche de routage

- Un domaine souche **n'autorise pas** d'autres domaine à transmettre des paquets **sur son infrastructure**
- Un "stub domain" est connecté à au moins un domaine de transit
 - "Single-homed stub domain"
 - Ex. : S1
 - "Dual-homed stub domain"
 - Ex. : S2
 - "Content-rich stub domain"
 - Ex : Google, Yahoo
 - "Access-rich stub domain"
 - Ex : n'importe quel ISP

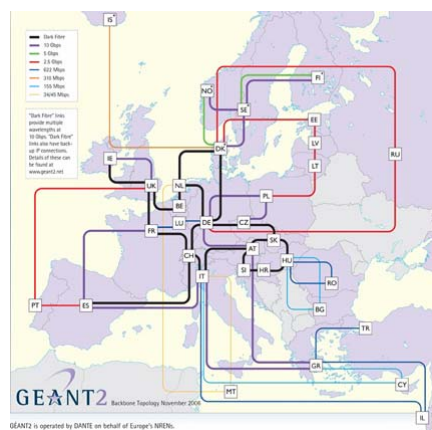


24 octobre 2007

Border Gateway Protocol

9

Un exemple de domaine de transit : GEANT



24 octobre 2007

Border Gateway Protocol

10

"Peering"

- Il y a, en pratique, 2 types de "peering" BGP :
 - "customer-provider peering"
 - Relation asymétrique dans laquelle un client (un domaine de routage) achète une connectivité à l'Internet auprès d'un fournisseur d'accès (un autre domaine de routage).
 - "shared-cost peering"
 - Relation symétrique où deux domaines de routage acceptent d'échanger gratuitement leurs paquets à travers un point d'interconnexion.

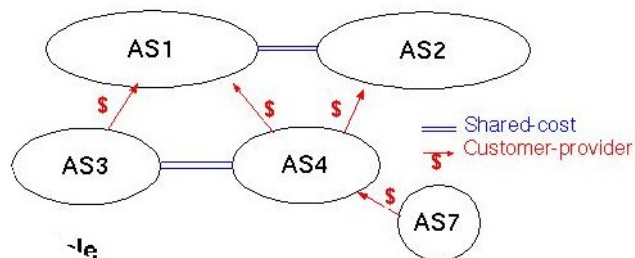
24 octobre 2007

Border Gateway Protocol

11

"Customer-provider peering"

- Le client **envoie ses routes internes et les routes apprises de ses propres clients** au fournisseur
 - Le fournisseur annoncera ces routes sur tout l'Internet
- Le fournisseur annonce à son client **toutes les routes** qu'il connaît
 - le client est capable d'atteindre n'importe qui sur l'Internet



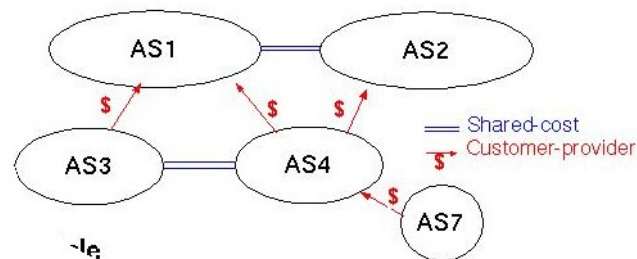
24 octobre 2007

Border Gateway Protocol

12

"Shared-cost peering"

- Chaque "peer" envoie à l'autre **ses propres routes et celles de ses clients**
- Le point d'interconnexion sera utilisé par l'un des pair BGP pour atteindre les destinations ou les destinations des clients de l'autre pair



24 octobre 2007

Border Gateway Protocol

13

Le protocole BGP

- Fonctionne au-dessus de TCP
 - Port 179
 - Fiabilité des transmissions
- Chaque routeur BGP échange avec ses voisins des messages pour ouvrir et négocier les paramètres la session BGP.
 - Message "open" [1]
- Les routeurs BGP échangent des informations concernant l'accessibilité de certains préfixes IP (destinations).
 - Ces informations sont formées principalement par le chemin (liste de numéro d'AS) qui doit être suivi pour atteindre une destination.
 - Ces informations permettent de construire un graphe formé d'AS (sans boucle) sur lequel une politique de routage peut être appliquée pour contraindre certains chemins.

24 octobre 2007

Border Gateway Protocol

14

Le protocole BGP

- Initialement les routeurs BGP échangent la totalité des informations de routage. Puis seules les modifications sont transmises.
 - Message "Update" [2]
- Un numéro est associé à chaque version des informations collectées par un routeur. Tous les voisins BGP doivent avoir le même numéro. Ce numéro est modifié à chaque mise à jour.
 - Un marqueur de 16 octets : Perte de synchronisation + authentification (par ex. MD5)
- Des messages sont transmis périodiquement pour vérifier le bon fonctionnement de la session BGP.
 - Message "Keepalive" [4]
- Des messages spéciaux sont utilisés pour informer les voisins BGP des erreurs et des cas spéciaux.
 - Message "Notification" [3]

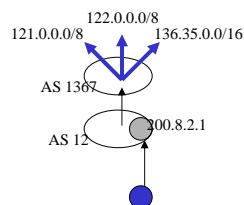
24 octobre 2007

Border Gateway Protocol

15

Exemple de message BGP

```
BGP : 16 byte Marker (all 1's)
BGP : Length = 64
BGP : BGP type = 2 (Update)
BGP : unfeasible Routes length = 0
BGP : No withdraw Routes in this Update
BGP : Path Attribute Length = 18 bytes
BGP : Attribute Flags = 0x4 (Well-know, Transitive)
BGP : Attribute type code = 1 (Origin)
BGP : Attribute Data Length = 1
BGP : Origin type = 0 (IGP)
BGP : Attribute Flags = 0x4 (Well-know, Transitive)
BGP : Attribute type code = 2 (AS Path)
BGP : Attribute Data Length = 4
BGP : AS Identifier = 1367
BGP : AS Identifier = 12
BGP : Attribute Flags = 0x4 (Well-know, Transitive)
BGP : Attribute type code = 3 (Next Hop)
BGP : Attribute Data Length = 4
BGP : Next Hop = [200.8.2.1]
BGP : Network Layer Reachability Information
BGP : 121.0.0.0/8
BGP : 122.0.0.0/8
BGP : 136.35.0.0/16
```



24 octobre 2007

Border Gateway Protocol

16

Différence avec un protocole DV

- BGP fonctionne différemment d'un protocole "distance vector"
 - Pas de transmission périodique des meilleures routes, mais uniquement des modifications
 - BGP utilise TCP
 - BGP mémorise toutes les routes vers toutes les destinations
 - Récupération rapide lorsque une destination devient inaccessible par la route initialement choisie
 - BGP construit des routes sans boucle
 - Le chemin suivi est décrit explicitement à l'aide de la liste des AS traversés
 - Les boucles sont facilement détectées

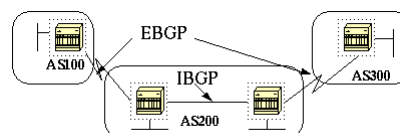
24 octobre 2007

Border Gateway Protocol

17

eBGP/iBGP

- Un AS peut servir de réseau de transit pour d'autres AS, s'il possède plusieurs routeurs BGP
 - par exemple AS 200
- Les routeurs internes à l'AS doivent être configurés avec les routes nécessaires avant d'être utilisé comme réseau de transit
 - obtenu par une combinaison du protocole de routage interne (IGP) et en redistribuant des informations BGP entre routeurs BGP du même AS (cf. iBGP)



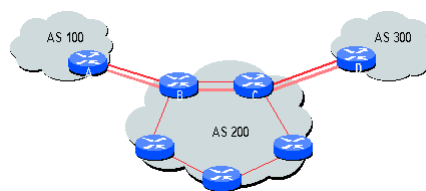
24 octobre 2007

Border Gateway Protocol

18

On doit utiliser BGP

- Lorsque le réseau est "dual" ou "multi homed"
- Pour fournir un routage complet ou partiel à un client en aval
- A chaque fois qu'une information sur le chemin vers un AS est nécessaire



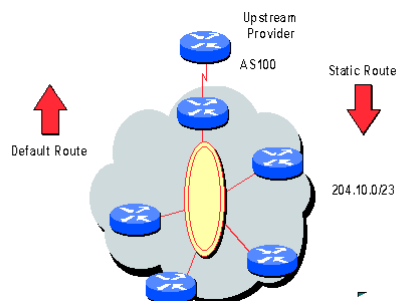
24 octobre 2007

Border Gateway Protocol

19

BGP n'est pas nécessaire

- Si votre AS est connecté par un seul point
- Et vous ne transmettez pas d'information de routage en aval
- Vous utilisez une route par défaut !



24 octobre 2007

Border Gateway Protocol

20

Les attributs de BGP

- A quoi servent-ils ?
- "AS path"
- "Next hop"
- "Local preference"
- "Multi-exit discriminator"
- Les autres attributs

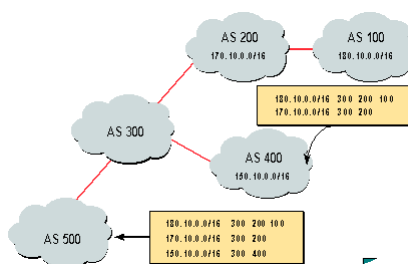
Les attributs de BGP

- Ils décrivent les caractéristiques associées à un préfixe particulier
 - Leur transmission est soit transitive soit non-transitive
 - Certains sont obligatoires
- 13 attributs (et d'autres...)
- Ils servent à sélectionner la meilleure route

- BGP appelle l'ensemble des préfixes partageant les mêmes attributs : "Network Layer Reachability Information" (NLRI)

L'attribut "AS path"

- La suite d'AS qu'un message BGP a traversé :
 - Détection des boucles
 - L'"AS path" le plus court est choisi, sauf si...
 - Permet d'appliquer certains politiques spécifiques au AS



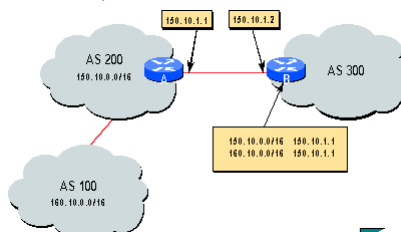
24 octobre 2007

Border Gateway Protocol

23

L'attribut "next hop"

- Le "next hop" pour atteindre un réseau
- Généralement :
 - Pour une session eBGP, c'est un routeur situé sur sous-réseau local (voisin)
 - Pour une session iBGP, c'est un routeur tiers



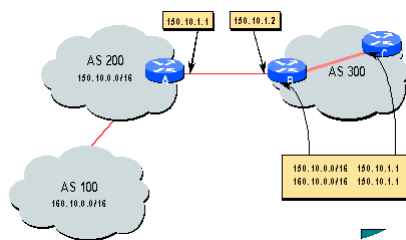
24 octobre 2007

Border Gateway Protocol

24

L'attribut "next hop"

- Le "next-hop" associé à une route externe n'est pas modifié lorsqu'elle est annoncée à un voisin iBGP
- Le protocole IGP doit permettre le routage vers le "next-hop"
 - Cela dissocie BGP de la topologie interne réelle



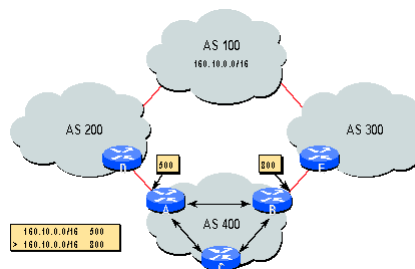
24 octobre 2007

Border Gateway Protocol

25

L'attribut "local preference"

- Interprétation locale à l'AS
- Utilisé pour influencer le processus de sélection du meilleur chemin
 - Le chemin avec la plus grande valeur de "local preference" est sélectionné



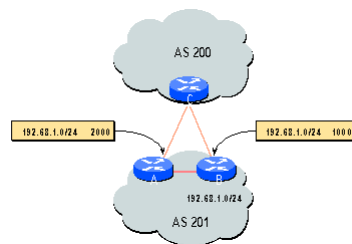
24 octobre 2007

Border Gateway Protocol

26

L'attribut "Multi-Exit Discriminator"

- N'est pas transitif
- Utilisé pour informer sur une préférence relative entre différents points d'entrée
- Comparable seulement si les chemins proviennent du même AS
- Les distances de l'IGP peuvent être utilisées



24 octobre 2007

Border Gateway Protocol

27

La politique de BGP

- La politique est basée sur :
 - l'"AS path",
 - la notion de "community"
 - le réseau
- Permet de rejeter ou d'accepter certaines routes
- Les attributs influencent la sélection des chemins

24 octobre 2007

Border Gateway Protocol

28

Algorithme de sélection du meilleur chemin

- Préférer le chemin ayant
 - Le "local preference" le plus grand
 - L'"AS-PATH" le plus court
 - Venant d'un EGP plutôt que d'un IGP ou qu'"incomplete"
 - Le MED le plus petit, s'ils viennent du même AS
 - Etc.

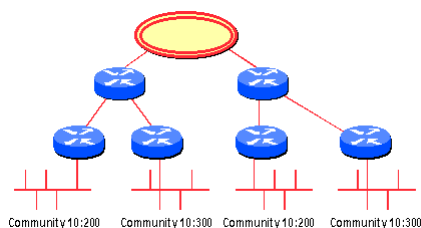
24 octobre 2007

Border Gateway Protocol

29

Une communauté sous BGP

- Regroupement de destinations
- Identifiée par un numéro
 - 0 à 4.10^9
 - $\langle n^{\circ}AS \rangle : \langle n^{\circ}communauté \rangle$



24 octobre 2007

Border Gateway Protocol

30

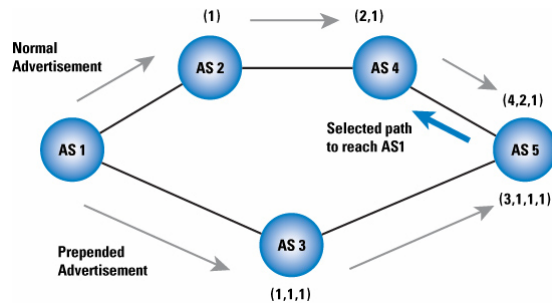
L'attribut "community"

- Une même destination peut être membre de plusieurs communautés
- L'attribut de communauté est conservé lors de la traversée des AS
- Permet une définition cohérente de la politique de BGP
- C'est un attribut optionnel
- Permet grâce à la commande "route-maps" d'appliquer des décisions de routage : "accept", "prefer", "redistribute", etc.

Les difficultés de BGP

- L'équilibrage de charge
- Le routage asymétrique
- L'inter relation entre IGP et BGP
- "Route flap" et "dampening"
- "AS path prepending"

AS path prepending



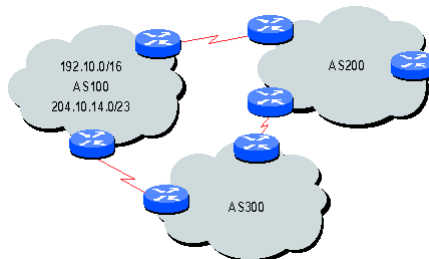
24 octobre 2007

Border Gateway Protocol

33

L'équilibrage de charge dans BGP

- L'équilibrage de charge
 - BGP n'est pas prévu pour faire de l'équilibrage de charge :
 - Il choisi et installe la meilleure route
 - Multi-lien
 - Multi-chemin



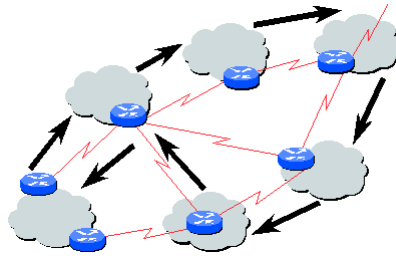
24 octobre 2007

Border Gateway Protocol

34

Le routage asymétrique

- Le routage asymétrique
 - Il est impossible de contrôler le chemin suivi
 - Les paquets peuvent ne pas emprunter le même chemin à l'aller et au retour



24 octobre 2007

Border Gateway Protocol

35

Relation entre BGP et IGP

- BGP gère un table de routage complète de l'Internet
- Les protocole IGP distribue les informations de routage **internes** au domaine de routage
- Les routes ne sont jamais redistribuées entre BGP et IGP (et vice versa)
- "resursive route lookup"

24 octobre 2007

Border Gateway Protocol

36

"Route flapping"

- Le "Route flap"
 - Une route apparaît ou disparaît
 - Change d'attribut
 - Influence tout l'Internet
 - Consomme des ressources : bande passante, processeur
- Le "Route flap damping"
 - Ne gêne pas les changements normaux de route
 - Supprime les routes qui oscillent
 - Les routes au-dessus de "suppress-limit" ne sont pas annoncées
 - Les routes au-dessous de "reuse-limit" sont annoncées

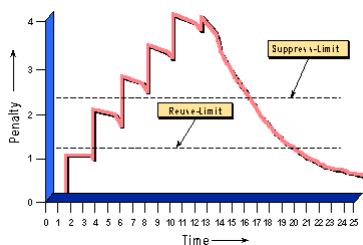
24 octobre 2007

Border Gateway Protocol

37

"Route flap damping"

- Sur les routes externes, seulement.
- Les routes alternatives sont toujours disponibles
- "additive penalty", "exponential decay"



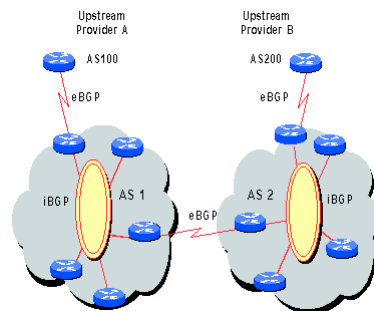
24 octobre 2007

Border Gateway Protocol

38

iBGP

- "Internal BGP"
 - Le même protocole dans un contexte différent
 - Doit être utilisé lorsqu'il existe plusieurs routeurs eBGP dans un même domaine
 - Une entité iBGP ne diffuse pas les routes trouvées par une autre entité
 - Prévention des boucles



24 octobre 2007

Border Gateway Protocol

39

Le maillage iBGP

- Chaque entité iBGP doit être connectée à toutes les autres entités :
 - Maillage complet = N^2 !
- Deux solutions :
 - Les confédérations de BGP
 - Les "Route Reflectors"
- Voisinage des entités iBGP
 - Il peut exister des routeurs intermédiaires entre voisins iBGP
 - Dans ces routeurs intermédiaires, toutes les routes vers les entités iBGP doivent avoir été correctement configurées :
 - Par exemple grâce à l'IGP

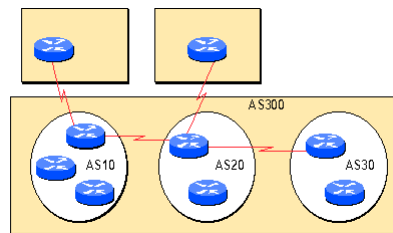
24 octobre 2007

Border Gateway Protocol

40

Les confédérations iBGP

- L'AS initial est partitionné en plusieurs sous-AS
- Seul l'AS global initial est annoncé par les entités BGP externes



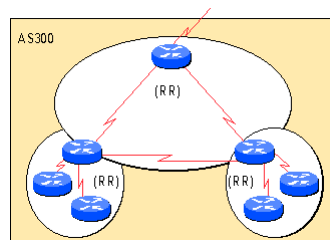
24 octobre 2007

Border Gateway Protocol

41

Les "route reflectors" iBGP

- Un "reflector" représente tous les entités iBGP d'un même groupe ("cluster")
 - une connexion est établie entre le "reflectors" de chaque membre de son "cluster"
 - Une hiérarchie de clusters peut être formées
 - Un "reflector" peut annoncer des routes dont il n'est pas à l'origine



24 octobre 2007

Border Gateway Protocol

42

Conclusion

- Utilisez BGP que lorsque c'est nécessaire
 - Pour le routage inter-AS
- BGP propose
 - Une gestion flexible de la politique et un contrôle des routes
- Quelques difficultés pour une parfaite gestion cohérente