

Algorithmes de routage IP

3 Le routage inter-domaine...

Filière F5 - Isima 2007-2008

Mickael Meulle

mickaelmeulle@gmail.com

michael.meulle@orange-ftgroup.com

Plan

1. Généralités sur la granularité inter-domaine

- › 1.1 les domaines de routage autonomes
- › 1.2 Interconnexions de réseaux
- › 1.3 Routage et « Internet Connectivity »

2. Introduction au protocole BGP

- › 2.1 Généralités
- › 2.2 Granularité
- › 2.3 Sessions et échanges de messages
- › 2.4 iBGP et eBGP
- › 2.5 Les messages et les attributs
- › 2.6 Sélection de routes, propagation

Plan

3. Le protocole BGP en détail

- › 3.1 le processus de décision en détail
- › 3.2 L'attribut Local Preference
- › 3.3 L'attribut AS-PATH
- › 3.4 L'attribut MED
- › 3.5 L'attribut Origin

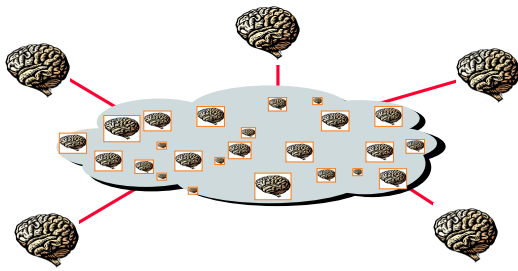
4. BGP dans la pratique: « Tuning BGP »

- › 4.1 Route reflectors
- › 4.2 convergence et stabilité
- › 4.3 Traffic Engineering sortant
- › 4.4 Traffic Engineering entrant
- › 4.5 Communautés
- › 4.6 Confédérations

1. Généralités sur la granularité inter-domaine

- 1.1 les domaines de routage autonomes
- 1.2 Interconnexions de réseaux
- 1.3 Routage et « Internet Connectivity »

Les domaines de routage autonomes



Les domaines de routage autonomes

- ▶ **Chaque administration dispose d'un réseau**
 - ▶ hétérogénéité
 - Beaucoup de technologies
 - Beaucoup de réseaux différents
 - Beaucoup de politiques de routage
 - Fournisseurs d'accès, opérateurs, services de transit spécialisés ...
 - ▶ Le routage IP évolue
 - Nouvelles extensions et protocoles
 - MPLS, QoS, Sécurité, ...
 - Les équipements évoluent
 - Puissance, fonctionnalités
 - Les pratiques évoluent
 - Avec ou sans les mêmes équipements
 - Mettant en jeu la même architecture ou non

L'adressage IP: découpage du réseau

- ▶ **L'Internet est avant tout découpé suivant l'adressage**
 - › Un réseau est représenté par une plage d'adresse
 - › Un réseau de réseaux est représenté par un ensemble de plages d'adresses
- ▶ **A un domaine de routage autonome correspond un ensemble de plages d'adresses**
 - › il permet pour toutes les plages d'adresses l'accessibilité et l'accès au reste de l'Internet via ses interconnexions
- ▶ **Les routeurs du domaine sont responsables du « Forwarding »: le routage**

L'Internet est un système distribué

- ▶ **Chaque Domaine autonome de routage remplit deux tâches**
 - › Assure le routage pour les réseaux qu'il héberge
 - › Assure le routage pour d'autres ARD connectés
- ▶ **Entre les domaines l'intelligence est distribuée**
 - › Chaque ARD contrôle ses décisions
 - › Chaque ARD n'a pas possibilité de contrôler les décisions des autres ARD
- ▶ **Le routage dans sa globalité nécessite une concertation mais minimale**
 - › Chaque machine ne peut pas connaître tous les routeurs de l'Internet!

1. Généralités sur la granularité inter-domaine

- 1.1 les domaines de routage autonomes
- 1.2 Interconnexions de réseaux**
- 1.3 Routage et « Internet Connectivity »

Le réseau des réseaux

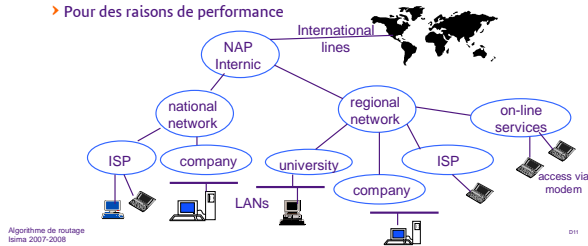
- ▶ A la couche IP, toutes les machines sont censées être joignables entre elles
- ▶ Les réseaux sont tous indirectement connectés pour pouvoir communiquer
- ▶ les réseaux sont connectés:
 - ▶ Localement en une ou plusieurs villes suivant leur étendue
- ▶ Des réseaux permettent le transit pour d'autres réseaux
- ▶ Les réseaux aux extrémités ne participent qu'aux communications qui les concernent

Algorithme de routage
Isma 2007-2008

010

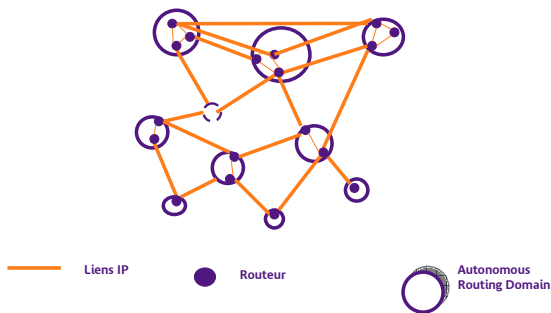
Les réseaux s'interconnectent pour communiquer

- ▶ Les réseaux s'interconnectent mais veulent contrôler les échanges d'informations
 - ▶ Pour des raisons économiques
 - ▶ Pour des raisons de performance



011

Schématiquement

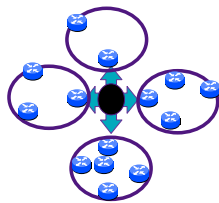


012

Les interconnexions

Interconnexions couche 2 ou couche 3

- Points d'échanges
- Fermes d'interconnexion



Interconnexion par circuit privé



Il existe même des interconnexions via Tunnel

- Très récent
- rare

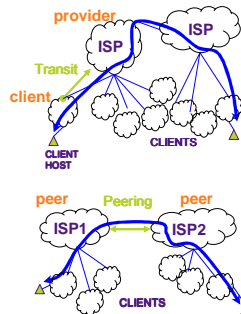
Algorithme de routage
Isma 2007-2008

D13

Les contrats d'interconnexion

Il existe deux grands types de contrats d'interconnexion de réseaux

- La vente de capacités et d'accessibilité
 - Client/Fournisseur
- Le partage gratuit des communications entre clients
 - Peering
 - Ne pas confondre avec Peer-To-Peer !!!!



Algorithme de routage
Isma 2007-2008

D14

Les accords d'interconnexion

Les clients payent pour recevoir et envoyer du trafic

- Un client a souvent plusieurs fournisseurs
 - Peut choisir par exemple un fournisseur pour accéder à l'Internet et d'autres fournisseurs pour être accessible sur l'Internet
- L'Internet est non régulé et les négociations d'accords tournent à l'avantage des plus gros
 - Les fournisseurs cherchent toujours à vendre leurs capacités
- « Interconnexion War »

Les relations de Peering sont mises en place pour des raisons de performances

- Les délais sont réduits
- Permet d'avoir une partie de connectivité gratuite

Algorithme de routage
Isma 2007-2008

D15

Les politiques de routage

- ▶ **Chaque domaine de routage autonome veut que son trafic IP suive un cheminement particulier**
 - Dépend des accords d'interconnexion négociés
 - Dépend des performances voulues
 - Un paquet devrait-il faire:
 - New York – San Francisco – Tokyo – Shanghai – Moscou – Amsterdam – Paris – Londres*
 - Lorsque tout simplement la route suivante est possible:
 - New York – Londres*
 - Les besoins sont au niveau « Data plane »
- ▶ **Chaque domaine de routage autonome met en place une politique d'échange des informations de routage**
 - La configuration des équipements de routage (« control Plane ») doit être telle qu'au niveau « Data plane », les routes des tables de routage sont telles que désirées
 - Les routes respectent au moins les accords économiques négociés

Algorithme de routage
Ismaïl 2007-2008

D16

1. Généralités sur la granularité inter-domaine

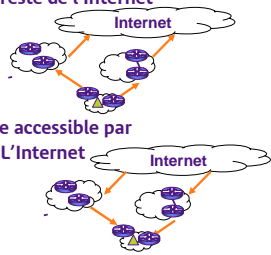
- 1.1 les domaines de routage autonomes
- 1.2 Interconnexions de réseaux
- 1.3 Routage et « Internet Connectivity »**

Algorithme de routage
Ismaïl 2007-2008

D17

La connectivité Internet

- ▶ **Chaque réseau veut accéder au reste de l'Internet**
 - Connectivité
- ▶ **Chaque réseau veut pouvoir être accessible par n'importe quelle machine dans l'Internet**
 - L'accessibilité
- ▶ **Accessibilité + Connectivité**
 - = « Internet Connectivity »



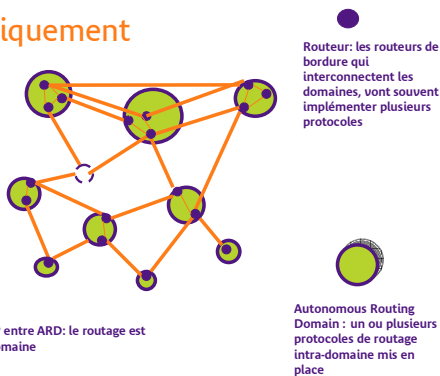
Algorithme de routage
Ismaïl 2007-2008

D18

Le routage est cloisonné

- ▶ **Typiquement, à l'intérieur d'un domaine**
 - La main mise administrative est totale
 - La taille en nombre de plages d'adresses et nombre de routeurs est modeste
 - Les besoins sont axés sur la performance, la flexibilité, la sécurité
 - Pas de considérations économiques avec les partenaires
- ▶ **Typiquement entre domaines**
 - Les considérations économiques jouent un rôle moteur
 - Les performances de bout en bout dépendent aussi des autres domaines
 - Les décisions sont locales, l'information est incomplète
 - « ROUTAGE AVEC POLITIQUES »

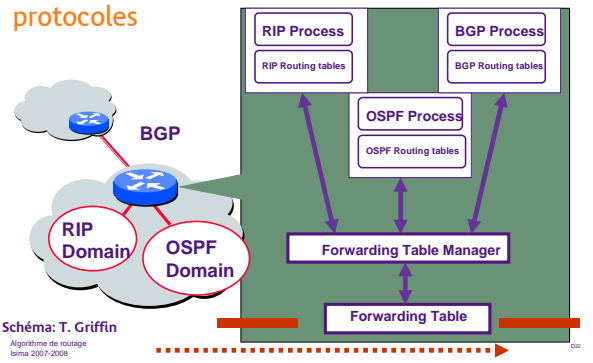
Schématiquement



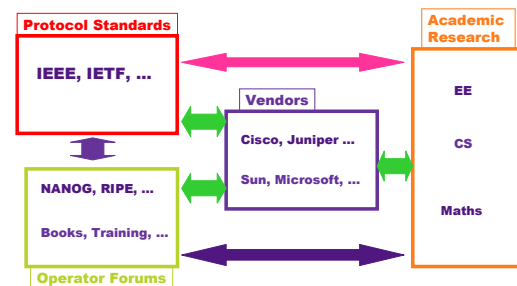
Le routage

- ▶ **Beaucoup d'ARD configurent un routage statique**
 - Aux bords de l'Internet
- ▶ **Tous les ARD non triviaux**
En particulier les opérateurs (ISP)
 - Le routage est dynamique
 - Les protocoles intra-domaine et inter-domaine interagissent

Les routeurs et les interactions entre protocoles



Anthropologie du routage



L'Internet

- Le routage a convergé vers l'utilisation d'un protocole inter-domaine unique:
 - Le protocole BGP
- Le routage dans l'Internet:
 - « Solving The puzzle »
- Au niveau intra-domaine
 - Ce sont les meilleurs chemins de bout en bout (calculés) qui déterminent les décisions locales
- Au niveau inter-domaine
 - Pour chaque réseau destination
 - Ces sont les décisions locales de proche en proches qui déterminent les chemins de bout en bout

Le contrôle du routage

- ▶ La configuration des routeurs conditionne la formation des tables de routage
- ▶ Les tables de routage impliquent la distribution des trafics IP
 - Un protocole inter-domaine permet la concertation entre les ARD
 - Les routes de bout en bout empruntées par les paquets dépendent des décisions relatives à chaque politique de routage
- ▶ Contrôle d'une politique de routage
 - « Traffic Engineering »
 - « Traffic engineering » entrant
 - « Traffic engineering » sortant
- ▶ « Control Plane » ↔ « Data Plane »

2. Introduction au protocole BGP

- 2.1 Généralités
- 2.2 Granularité
- 2.3 Sessions et Echange de messages
- 2.4 iBGP et eBGP
- 2.5 Les messages et les attributs
- 2.6 Sélection de routes, propagation

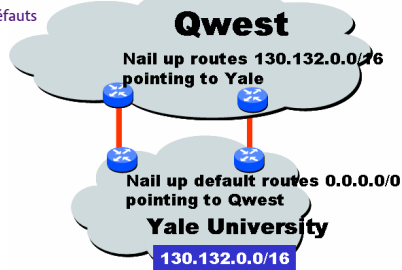
Les systèmes autonomes

- ▶ Un système autonome est une entité de routage BGP
 - La politique de routage d'un AS doit être unique
- ▶ Les Systèmes autonomes ne sont qu'un détail protocolaire!
 - Un opérateur peut disposer de plusieurs AS
 - Un AS peut être en réalité beaucoup plus d'entités
- ▶ Les As sont une réalité provoquée par le standard auquel on ne peut échapper
 - Le protocole BGP

Les Systèmes autonomes

- ▶ Un système autonome n'a pas lieu d'exister si sa configuration est triviale

▶ Routes par défauts



Algorithme de routage
Isma 2007-2008

008

Alors Pourquoi BGP?

- ▶ Pour beaucoup d'opérateurs, BGP reste encore un mystère...
- ▶ Meilleur contrôle des destinations
- ▶ Contrôle de la décision des routes
 - ▶ Pour « tuner » les tables de routage
 - ▶ en cas de pannes
- ▶ On peut contrôler beaucoup mieux les annonces de réseaux
 - ▶ Annonces sélectives
 - ▶ Découpage des plages réseaux
- ▶ Mieux choisir le fournisseur d'accessibilité
- ▶ Mieux choisir le fournisseur d'accès

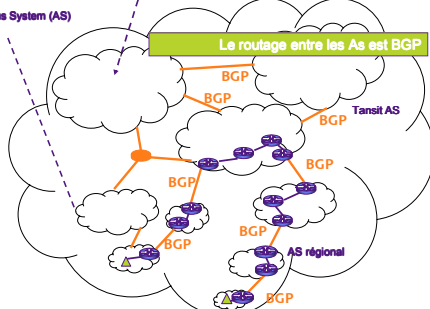
Algorithme de routage
Isma 2007-2008

009

architecture

Le routage dans un As est régi par des protocoles comme IS-IS, OSPF,...

Autonomous System (AS)

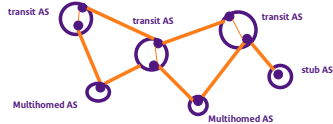


Algorithme de routage
Isma 2007-2008

010

Les AS de transit

- ▶ **stub AS:**
 - Un AS qui n'a qu'une connexion logique vers un unique autre AS
 - Naturellement, cet AS ne supporte que du trafic local
- ▶ **multihomed AS:**
 - Un AS qui a des connexions avec plusieurs AS
 - Cet AS ne supporte que du trafic local aussi
- ▶ **transit AS:**
 - Un AS qui a des connexions vers de multiples AS
 - Sous certaines conditions relatives à la politique de routage (e d'interconnexion), l'AS supporte du trafic local et du trafic de transit



Algorithme de routage
Isma 2007-2008

001

Le protocole BGP

- ▶ **C'est un protocole à vecteur de chemin**
 - Les informations de routage échangées concernent les réseaux destinations
 - Fonctionne à base d'annonces et de révocations de routes
 - Les routes sont des chemins d'AS muni d'attributs globaux ou locaux
 - Il y a volontairement perte d'information concernant la topologie, les détails concernant une granularité plus fine sont sous forme d'attributs (voir plus loin)
- ▶ **BGP est simple, mais la configuration est complexe**

Algorithme de routage
Isma 2007-2008

002

BGP est inévitable

- ▶ **BGP est une réponse à la problématique du routage inter-domaine**
 - Désormais, beaucoup trop d'équipements sont en place pour migrer vers autre chose
 - La réalité économique joue un rôle central
 - Beaucoup de problèmes sont expérimentés
 - Les tables de routage explosent
 - Les erreurs sont fréquentes et peuvent impacter sur l'Internet tout entier
- ▶ **Le protocole BGP est la glue de l'Internet**
 - Les réseaux physiques sont interconnectables grâce au protocole IP
 - La concertation du routage est propagée entre les AS uniquement via BGP ou de façon manuelle
 - BGP permet l'expression des politiques de routage

Algorithme de routage
Isma 2007-2008

003

Historique du routage BGP

Le routage avec politiques

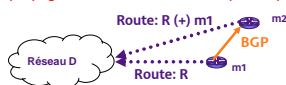
- ▶ 1989 : BGP-1 [RFC 1105]
 - Remplace EGP (1984, RFC 904)
- ▶ 1990 : BGP-2 [RFC 1163]
- ▶ 1991 : BGP-3 [RFC 1267]
- ▶ 1995 : BGP-4 [RFC 1771]
 - Support de CIDR

2. Introduction au protocole BGP

- 2.1 Généralités
- 2.2 **Granularité**
- 2.3 Sessions et Échange de messages
- 2.4 iBGP et eBGP
- 2.5 Les messages et les attributs
- 2.6 Sélection de routes, propagation

Les destinations sont des réseaux

- ▶ Les routeurs annoncent les réseaux qu'ils routent
 - "network-layer reachability information"
 - NLRI
- ▶ Les annonces sont des promesses
 - Un routeur qui annonce un réseau devient responsable pour l'acheminement vis-à-vis du récepteur de la route
 - Ils révoquent aussi des routes préalablement annoncées
- ▶ Les routeurs annoncent des **routes**
 - La propagation de l'annonce forme petit à petit le chemin

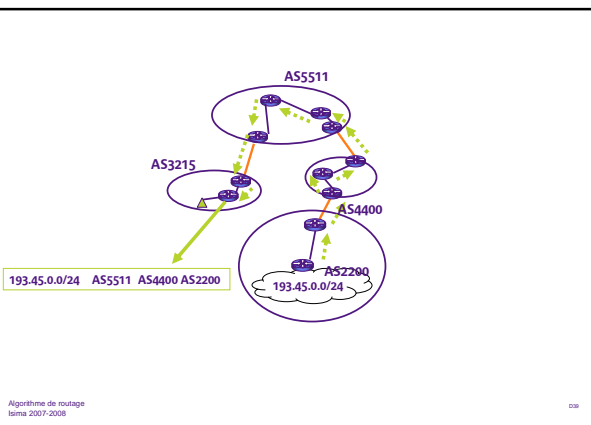


Injection des destinations

- Pour joindre une machine parmi une plage d'adresse précise
La plage d'adresse doit avoir été annoncée
 - Par BGP, les annonces de la plage d'adresse vont se propager entre les AS
- Deux façons de configurer les annonces
 - Directives BGP "network x.y.z.q [mask a.b.c.d]"
 - La route IGP doit exister
 - Redistribution statique ou par IGP (comme OSPF)
 - redistribute <igp-protocol> <protocol-id>
 - Faire attention à bien filtrer les routes
Comme les routes locales par exemple
 - redistribute static

Les routes

- Les routes sont des routes d'AS munies d'attributs
 - Chaque AS est considéré identique au sens des routes
 - Les attributs concernent la route de façon
 - Locale
 - Seulement propagés entre routeurs de l'AS
 - Ou seulement à un bond
 - Globale
 - Les attributs sont propagés entre les AS
- Les routeurs s'échangent donc des routes qui sont des chemins composés d'AS, à destination de réseaux, et muni d'attributs



La notion de NEXT HOP BGP

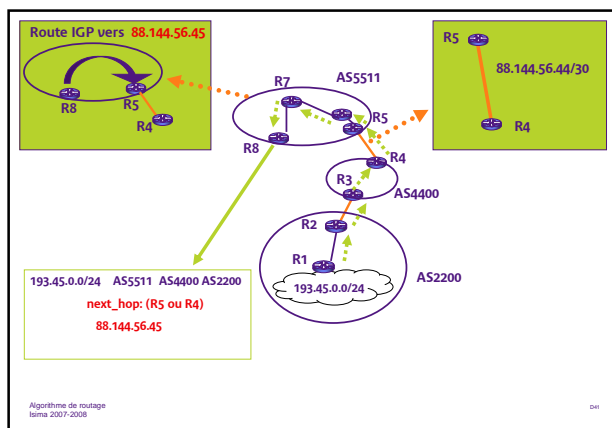
► Pour constituer la FIB

- un routeur doit connaître l'adresse du prochain bond pour envoyer un paquet pour qu'il puisse de bond en bond rejoindre la destination désirée

► Le next hop BGP d'une route est un attribut

► Le next hop BGP ne correspond pas forcément à la machine suivante

- Il correspond au point de sortie de l'AS
 - L'IGP d'un AS maintient les routes à destination des machines de l'AS
 - Au niveau BGP on peut vérifier la disponibilité d'une route IGP
- L'attribut n'est pas modifié dans les messages propagés dans l'AS



2. Introduction au protocole BGP

- 2.1 Généralités
- 2.2 Granularité
- 2.3 Sessions et Échange de messages
- 2.4 iBGP et eBGP
- 2.5 Les messages et les attributs
- 2.6 Sélection de routes, propagation

Les sessions BGP

► Les routeurs établissent des sessions entre eux

- Le protocole BGP est un protocole applicatif
 - du point de vue du modèle OSI, BGP au dessus de la couche Sessions et des autres couches inférieures
 - BGP requiert l'établissement de sessions
- Les sessions s'effectuent via TCP
 - Les sessions peuvent s'établir entre routeurs ne partageant pas de liens
 - *Attention très dangereux*

► Un routeur établit des sessions BGP avec tous les routeurs voisins configurés

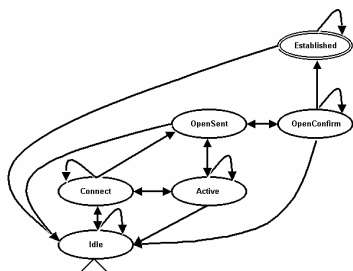


Etablissement de sessions

► 6 états ont été défini dans la RFC 1771

- IDLE : le routeur vient d'être initialisé (après erreur par exemple), ou attende d'une session
 - Des messages de NOTIFICATION sont envoyés après des erreurs
- CONNECT: après l'initialisation, le routeur commence une session par l'envoi d'un message OPEN
- ACTIVE: le routeur est actif et a commencé la phase d'ouverture TCP
 - La connexion TCP n'est pas établie
 - Si une erreur survient, le routeur retourne dans l'état IDLE
- OPEN_SENT: après une initialisation avec succès, le routeur envoie un paquet TCP SYN
 - Port 179
- OPEN_CONFIRM: la session est SYNchronisée (le routeur a reçu un TCP SYN de retour
- ESTABLISHED
 - Le seul état qui permet l'échange d'informations BGP
 - Dans un autre état, la session BGP est inutilisable (le lien aussi d'ailleurs!)

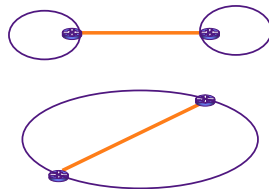
Machine à états (RFC 1771)



Fonctionnement des sessions BGP

▶ Tant que la connexion est active

- ▶ Les routeurs s'envoient leurs mises à jour



Algorithme de routage
Isma 2007-2008

D46

Remarques sur les sessions

▶ Si une session meurt, toutes les routes correspondantes sont révoquées

- ▶ « Withdraw »

▶ Les routeurs avec lesquels les sessions sont établies sont spécifiés au préalable

- ▶ Exemple avec Cisco:

```
Router A in AS 1
router bgp 1
neighbor 129.213.1.1 remote-as 2
Router B in AS 2
router bgp 1
neighbor 129.213.1.2 remote-as 1
```

Algorithme de routage
Isma 2007-2008

D47

2. Introduction au protocole BGP

- 2.1 Généralités
- 2.2 Granularité
- 2.3 Sessions et Échange de messages
- 2.4 iBGP et eBGP**
- 2.5 Les messages et les attributs
- 2.6 Sélection de routes, propagation

Algorithme de routage
Isma 2007-2008

D48

BGP est deux protocoles

▶ BGP

- « Internal BGP »: iBGP
- « External BGP »: eBGP

▶ Quelques différences

- Les attributs manipulés ne sont pas toujours les mêmes
- Avec iBGP, NEXT HOP n'est pas modifié
- Les routeurs connectés par iBGP ont normalement des configurations cohérentes et similaires
- iBGP a pour but la bonne dissémination des routes
- eBGP a pour but la propagation efficace de messages BGP qui vérifie les règles constituant la politique de routage de l'AS.

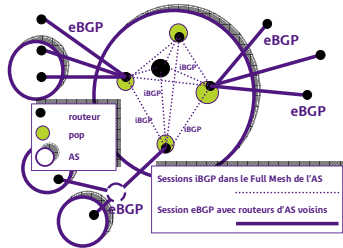
Sessions iBGP et eBGP

▶ Les sessions entre les AS

- sessions eBGP

▶ Les sessions entre routeurs du même AS

- sessions iBGP



Le full Mesh iBGP

▶ Les routeurs d'un AS ne propagent pas les messages iBGP entre voisins iBGP

- Les messages eBGP sont propagés à tous les voisins iBGP
- Les messages iBGP ne sont jamais propagés

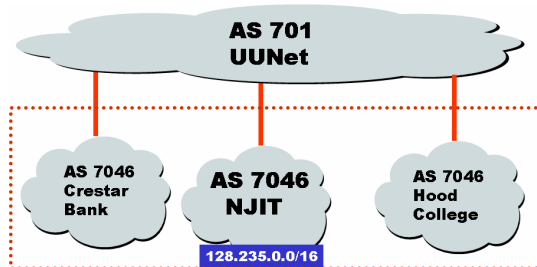
▶ Les routeurs doivent tous établir des sessions entre eux 2 à 2

- Cela assure une dissémination efficace de l'information
- Full Mesh iBGP

▶ Deux alternatives aux Full-Mesh iBGP (voir plus loin dans le cours)

- Les « route reflector »
 - propagent les messages iBGP à des routeurs dits 'esclaves'
- Les confédérations
 - Plusieurs AS privés à l'intérieur d'un même AS (invisible de l'extérieur)

Les AS peuvent être partagés (RFC 2722)

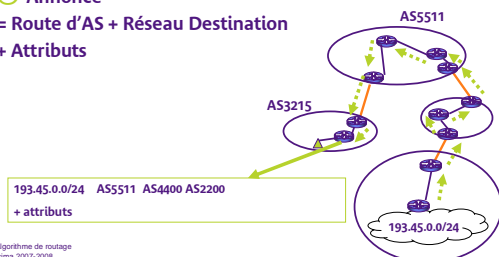


2. Introduction au protocole BGP

- 2.1 Généralités
- 2.2 Granularité
- 2.3 Sessions et Échange de messages
- 2.4 IBGP et eBGP
- 2.5 Les messages et les attributs**
- 2.6 Sélection de routes, propagation

Les messages BGP

- ▶ Les seuls intéressants : « Update BGP »
 - ▶ Message = une ou plusieurs annonces ou révocations
 - ▶ Annonce
- = Route d'AS + Réseau Destination
+ Attributs



Messages

► Header:

Marker (16 bytes)		
Length (2 bytes)	Type (1 byte)	Data (variable)

- **Marker** (16 bytes): contient (0xFF) utilisé pour la synchronisation quand de multiples segments TCP sont envoyés
- **Length** (2 bytes): taille totale du message
- **Type** (1 bytes): type du message
- **Data** (variable): dépend du type de message (éventuellement absent)

Messages

► Messages OPEN

- Type 1
- RFC 1771

Version (1 byte)	
My Autonomous System (2 bytes)	
Hold Time (2 bytes)	
BGP Identifier (4 bytes)	
Opt Param Len (1 byte)	
Optional Parameters (variable)	

- **Version** (1 byte): BGP version, BGP-4 par défaut (inchangeable)
- **My Autonomous System** (2 bytes): numéro d'AS su speaker BGP
- **Hold Time** (2 bytes): temps en secondes en attente d'un message UPDATE ou KEEPALIVE avant que le routeur de soit considéré « down *BGP Identifier* (4 bytes): routeur ID = BGP ID
- **Optional Parameter Length** (1 byte): longueur de la partie variable (peut être 0)
- **Optional Parameters** (variable):
 - Header password authentication (RFC 1771), jamais implémenté
 - **Capacités advertisement** (RFC 2842), fournit un mécanisme de négociation des capacités d'utilisation des particularités de BGP

Messages

► Message Update

- Type 2
- RFC 1771
- Les 2-tuples sont des Réseaux + longueur
- Appelé NLRI pour annonces

Withdrawn Routes Length (2 bytes)
Withdrawn Routes (variable)
Total Path Attribute Length (2 bytes)
Path Attributes (variable)
Network Layer Reachability Information (variable)

- **Withdrawn Routes Length** (2 bytes): longueur des champs pour les withdrawn routes (0 si pas de révocation)
- **Withdrawn Routes** (variable): 2-tuple IP prefix(es) de routes à révoquer
- **Total Path Attribute Length** (2 bytes): Longueur de la partie d'annonce (0 si pas d'annonces)
- **Path Attributes** (variable): attributs du chemin (TLVs)
- **NLRI** (variable): 2-tuple IP prefix(es) of reachability information

Messages



KeepAlive

- Type 3
 - RFC 1771
- Envoyés tous les $\frac{1}{3}$ *Hold_Time pour conserver les sessions en bon fonctionnement entre les routeurs



Notification

- Type 4
 - RFC 1771
- Indique une erreur
 - Error Code (1 byte): type of error
 - Error Subcode (1 byte): indicates more details about the error
 - Data (variable): optional error data

Error Code (1 byte)	Error Subcode (1 byte)	Data (variable)
------------------------	---------------------------	-----------------

Algorithme de routage
Isma 2007-2008

008

Messages



Route Refresh

- Type 5
 - RFC 2918
- Demande retransmission d'une route
 - Save RAM
- Normalement indépendant du protocole
 - IPv4 unicast routes

AFI (2 bytes)	Reserved (1 byte)	SAFI (1 byte)
---------------	----------------------	------------------

- AFI (2 bytes): Address Family Identifier, par exemple IPv4 ou IPv6
- Reserved (1 byte): toujours à 0
- SAFI (1 byte): Subsequent Address Family Identifier, par exemple unicast ou multicast

Algorithme de routage
Isma 2007-2008

009

Les attributs

AS_PATH	NEXTHOP	MED
---------	---------	-----	------



Encodage du type: Type Longueur Valeur. (TLV)



4 bytes long

0	1	2	3	4	5	6	7	
W/O	H/T	C/P	EL	Unused (4 bits)				Attribute Type Code (1 byte)
Attribute Length Code (variable)								Attribute Value (variable)

- Bit 0: Well-known or Optional
- Bit 1: Non-transitive or Transitive
- Bit 2: Complete or Partial
 - Complete: l'attribut a été transporté sur le chemin complet
 - Partial: a router in the path did not implement an attribute, routing information may be lost (very rare in today's Internet as most routers support all path attributes)
- Bit 3: Extended Length, définit si la longueur de l'attribut est 1 ou 2 bytes
- Bits 4 – 7: inutilisé et mis à zéro
- Attribute Type Code: Code de l'attribut
- Attribute Length : longueur de l'attribut
- Attribute Value: Donnée ou valeur de l'attribut

Algorithme de routage
Isma 2007-2008

010

Les attributs

► 4 types:

- **Well-known mandatory:** doit être présent dans les messages UPDATE et doit être implémenté par tous les BGP speakers pour assurer un jeu minimal de fonctionnalités
- **Well-known discretionary:** pas toujours présent dans les messages UPDATE et doit être implémenté par tous les BGP speakers, quelquefois un attribut n'a pas lieu d'être
- **Optional transitive:** pas toujours présent dans les messages UPDATE et pas toujours implémenté par tous les BGP speakers; ces attributs sont propagés sans être sûr que les autres routeurs implémenteront, quelquefois même si le routeur local ne comprend pas l'attribut
- **Optional non-transitive:** pas toujours présent dans les messages UPDATE et pas toujours implémenté par tous les BGP speakers; ces attributs sont propagés seulement si le routeur local comprend l'attribut

Les principaux attributs

► Principaux

- **ORIGIN** (Type Code 1, Well-known Mandatory – RFC 1771)
- **AS_PATH** (Type Code 2, Well-known Mandatory – RFC 1771)
- **NEXT_HOP** (Type Code 3, Well-known Mandatory – RFC 1771)
- **MULTI_EXIT_DISC** or **MED** (Type Code 4, Optional Non-transitive – RFC 1771)
- **LOCAL_PREF** (Type Code 5, Well-known Discretionary – RFC 1771)
- **ATOMIC_AGGREGATE** (Type Code 6, Well-known Discretionary – RFC 1771)
- **AGGREGATOR** (Type Code 7, Optional Transitive – RFC 1771)
- **COMMUNITIES** (Type Code 8, Optional Transitive – RFC 1991)
- **EXTENDED COMMUNITIES** (Type Code 16, Optional Transitive – draft-ietf-idr-bgp-ext-communities-02.txt)

► + Attributs iBGP: iBGP salability

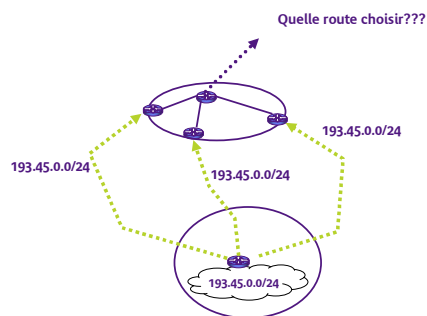
- confédérations et Route Reflection

Attribute	EBGP	IBGP
ORIGIN	Mandatory	Mandatory
AS_PATH	Mandatory	Mandatory
NEXT_HOP	Mandatory	Mandatory
MULTI_EXIT_DISC	Discretionary	Discretionary
LOCAL_PREF	Not Allowed	Required
ATOMIC_AGGREGATE	Discretionary	Discretionary
AGGREGATOR	Discretionary	Discretionary
COMMUNITY	Discretionary	Discretionary

2. Introduction au protocole BGP

- 2.1 Généralités
- 2.2 Granularité
- 2.3 Sessions et Échange de messages
- 2.4 iBGP et eBGP
- 2.5 Les messages et les attributs
- 2.6 Sélection de routes, propagation**

L'utilisation des attributs



La sélection et la propagation de routes

► BGP Decision Process

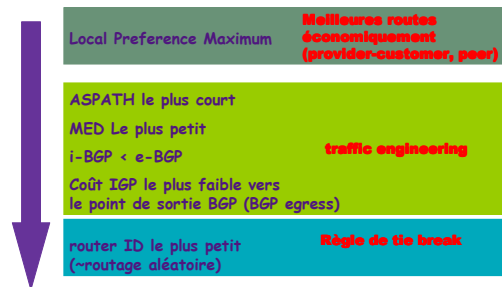
- Processus de décision **pas à pas** qui permet de choisir la meilleure route pour chaque réseau destination
- Le processus est en presque 10 points
 - Disponibilité de la route
 - Attributs de la route (globaux puis locaux)
 - Tie Break

► Les routeurs ne propagent qu'une seule « best route » par destination

► La propagation des messages dépend

- De la sélection des routes
- Des politiques de routage (filtre d'import et d'export de route)

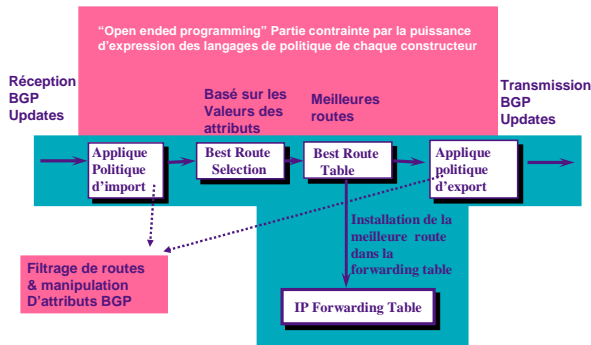
BGP Decision Process



Algorithme de routage
Isma 2007-2008

007

Le traitement des routes BGP



Algorithme de routage
Isma 2007-2008

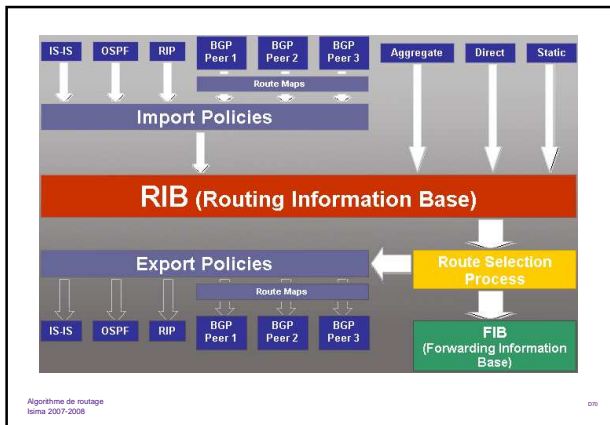
008

Le traitement d'une route BGP

- **Réception d'une route en provenance d'un voisin BGP**
 - Accepte ou rejette la route (d'après la politique d'import)
 - Stocke la route dans l'Adj-RIB-in (une Adj-RIB-in par voisin)
 - Traite les attributs et modifie éventuellement la route
- **Applique le BGP decision process pour le NLRI concerné par la route et installe la meilleure route dans la Loc-RIB (globale au routeur)**
 - La décision de la meilleure route s'effectue entre les routes de toutes les Adj-RIB-in pour le même NLRI
- **Pour chaque voisin:**
 - décide si la route peut être envoyée ou non (d'après la politique d'export)
 - Aggrège plusieurs routes en une seule, si applicable
 - Stocke la route dans l'Adj-RIB-out (une Adj-RIB-out par voisin)
 - Si l'Adj-RIB-out a changé, envoi un message update BGP
- **Ecrit une entrée dans la table de forwarding**
 - Écriture directement dans la table ou re-distribution de la route vers un protocole IGP

Algorithme de routage
Isma 2007-2008

009



010
