Workshop BGP

Routage, IPv4, IPv6, BGP

Arnaud FENIOUX – arnaud@afenioux.fr

Présentation + liens utiles : note.rezopole.net/p/bgp

Planning

- **Journée 1**
  - **MATIN**
    - Rappel IP/AS/Routage
    - Détails du protocol BGP
    - Différence IGP/EGP
    - Rappel CLI Cisco
  - **APRES MIDI**
    - Configuration des premières sessions
    - Filtrage : Prefix-list et Route-map
    - Local-pref : Comment influencer le trafic sortant
    - Session BGP en FULL Table

- **Journée 2:**
  - **MATIN**
    - Filtrage : as-path et route-map
    - Les communautés BGP
    - Exemples d'utilisation pour LyonIX
    - Trafic Shaping : AS-PATH prepend et Désagrégation
  - **APRES MIDI**
    - Configuration Ipv6
    - Fine BGP tunning : Convergence Rapide
      et Détection de PATHs sub-optimal / perte de paquets
    - Best Current Practices de configuration switch/routeur

Les réseaux

- LAN Local Area Network
- MAN Metropolitan Area Network
- WAN Wide Area Network
- Inter - network

- **Internet aujourd'hui**
  - Ethernet, TCP/IP norme de facto
  - Le modèle ISO n'a pas su s'imposer
  - Adresses IPv4
    - 2013 l'année de la pénurie
  - Adresses IPv6
  - Systèmes Autonomes (AS)

- **Routage**
  - En mode circuit (connecté)
    - Téléphonie
    - Allocation des ressources
  - En mode datagrammes (non connecté)
    - IP
    - Chaque paquet est traité indépendamment
    - Les paquets peuvent arriver dans le désordre
    - Pas d'allocation de bande passante

- **Routage**
  - **Algorithme**
    - **du plus court chemin (saut, distance, ms)**
    - **Inondation (toutes les lignes sauf celle entrante), utile dans le cas des réseaux sans fils**
    - **Routage par vecteur de distance RIP**
      - **Chaque routeur maintient sa propre table de routage**
      - **Convergence lente**
      - **Pb de la valeur infinie**
    - **Routage par informations d'états des liens OSPF**
      - **Envoie des infos à tous les routeurs**
      - **Convergence rapide**
      - **Routeur désigné**

- **Adresses MAC**
  - Codé en hexadécimal
    - 52:54:00:03:79:1b
  - 48 bits
  - Organizationnaly Unique Identifier / Network Interface Card
  - Limitées à la couche Ethernet (Layer 2), aux machines d'un même sous réseau

- IPv4
  - Notation CIDR
    - Codée du 32 bits
    - 77.95.64.0/24
    - Masque de sous réseau 255.255.255.0
    - Broadcast

- **NAT (Network Address Translation)**
  - IPv4 denrées rares
  - Masquer plusieurs adresses IP privées derrière une seule adresse IP publique (ADSL)
  - Inconvénients :
    - Chaque IP n'est pas unique dans le monde
    - Rupture avec le bout en bout
    - NAT altère internet en faisant un réseau avec connexion
    - Problème dans l'organisation des couches ISO
    - Que se passe-t-il si on décide d'utiliser autre chose que tcp/udp ?
    - FTP, H.323 => obliger de réécrire le code NAT

- IPv6
  - Notation CIDR
    - Codée du 128 bits
    - 8 blocs de 2 octets séparés par des « : »
    - Notation hexadécimale (RFC 5952)
    - 2001:07f8:0047:0047:0000:0000:0000:0001/64
    - 2001:7f8:47:47::1/64

- **Table de routage**
  - 0.0.0.0/0
  - ::/0
    - Passerelle par défaut
    - Ligne de sortie/routeur vers les autres réseaux (WAN, Internet)

RIPE, numéro d'AS, allocation des plages

- **Ou obtenir ses adresses IP ?**
  - **IANA et les 5 RIRs**
    - **RIPE NCC, ARIN, LACNIC, AfriNIC, APNIC**

  - **Directement auprès du RIPE NCC (devenir LIR)**
  - **Auprès d'un LIR (mais IP en PA)**

- **Provider Independant**
  - Les IP appartiennent au client mais ne peuvent pas être sous allouées
- **Provider Aggregate**
  - Les IPs appartiennent au fournisseur d'accès
- **LIR (Local Internet Registry)**
  - Permet de sous allouer son allocation

- **Combien ca coûte ?**
  - **LIR/DA**
    - **2000€ de droits d'entrée**
    - **1600€ annuel**
    - **50€ supplémentaire par assigement PI**
  - **Auprès d'un LIR ou fournisseur d'accès**
    - **Frais de mise en place possible**
    - **Récurrence annuel obligatoire libre**

Que fait un routeur BGP ?

- Apprend les chemins par l'intermédiaire de ses voisins BGP
- Choisit le meilleur chemin et l'installe dans sa table de routage (RIB)
- Le meilleur chemin est envoyé aux voisins BGP
- Les stratégies de filtrage sont appliquées pour influer sur le choix du meilleur chemin et sa diffusion

Définitions

- **Transit**
  - Acheminement de trafic sur tout internet
  - Moyennant un payement
    - Le client annonce les routes de son réseau
    - Le fournisseur annonce toutes les routes de l'internet
- **Peering**
  - Acheminement de trafic uniquement sur son réseau
  - Généralement gratuit
    - Chaque peer n'annonce que les routes de son réseau
    - Se réalise majoritairement sur des points d'échanges Internet (IXP)

Définitions

- **Default Route**
  - **0.0.0.0/0**
    - **où envoyer le trafic quand il n'y a pas de correspondance explicite dans la table de routage**
- **Default Free Zone**
  - **Lorsque l'on connait chacune des destinations**
  - **Pas de route pas défaut**
  - **Optimisations**

Border Gateway Protocol

- **Exterior gateway protocol**
  - Protocole de routage utilisé pour échanger des informations de routage entre différents réseaux
  - Initialement décrit dans la RFC1105 de Juin 89
  - TCP port 179

- **Le système autonome est la pierre angulaire de BGP**
  - Un AS est utilisé pour identifier de façon unique les réseaux ayant une politique de routage commune

Autonomous System

- Regroupe les réseaux avec la même politique de routage
- Habituellement en propriété et contrôle administratif unique
- Identifié par un entier 32 bits unique (ASN)
  - 1-64495 Internet public
  - 64496-65551 documentation et usage privé
  - 23456 Réservé AS16/32bits
  - 65552-4294967295 Internet public
- Actuellement 45 000 visibles sur Internet

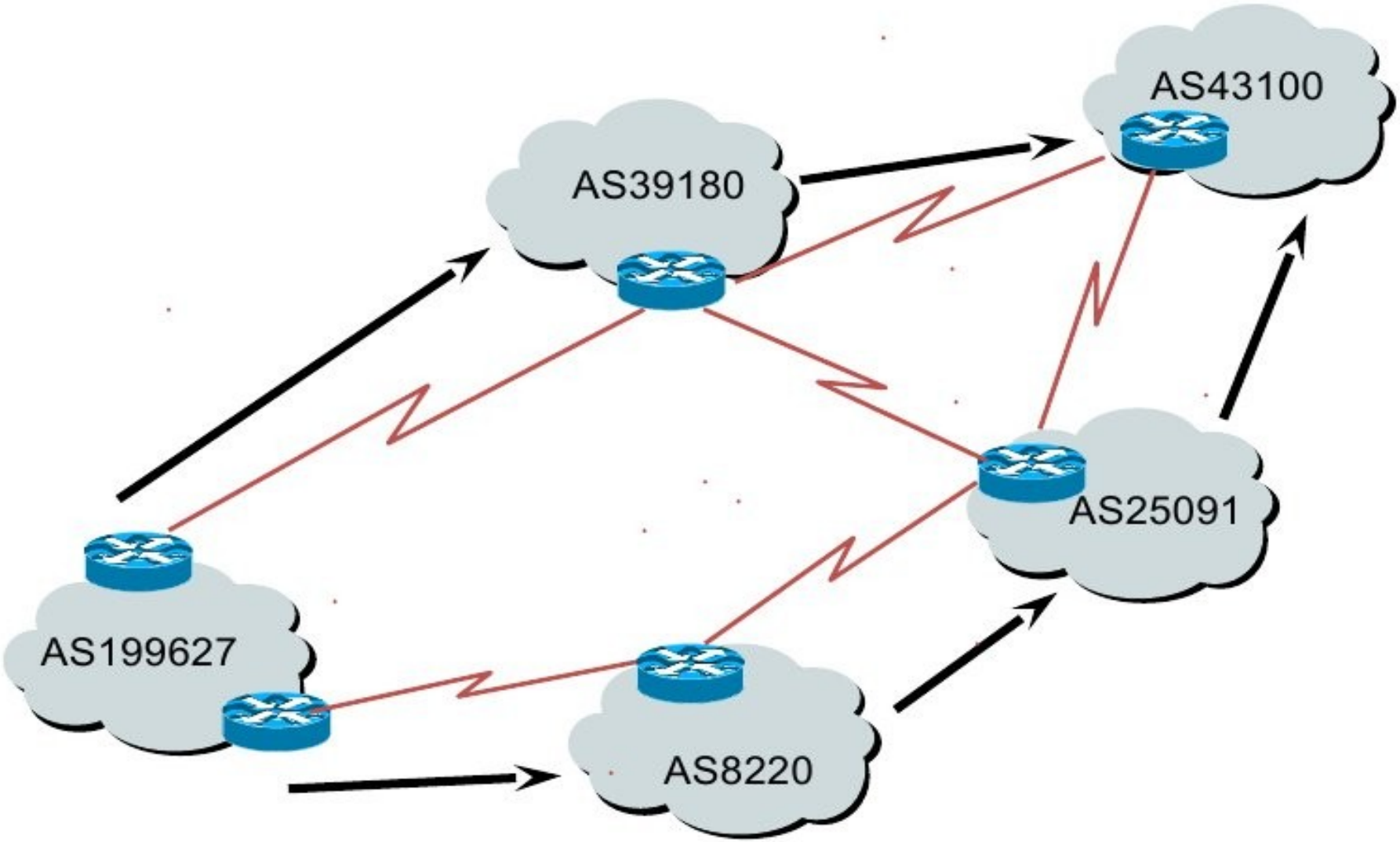
Border Gateway Protocol

- Largement utilisé pour le backbone Internet
- Sélection du meilleur chemin (next hop)
- Mises à jour incrémentales
- Path Vector Protocol
- Beaucoup d'options pour l'application des stratégies (filtrage)

- **MAJ incrémentale**
  - **Seules les modifications sont propagés**

```
cisco#show ip route 185.9.20.0/24
Routing entry for 185.9.20.0/24
  Known via "bgp 39180", distance 20, metric 0
  Tag 43100, type external
  Last update from 81.18.177.37 7w0d ago
  Routing Descriptor Blocks:
  * 81.18.177.37, from 81.18.177.37, 7w0d ago
      Route metric is 0, traffic share count is 1
      AS Hops 2
```

Border Gateway Protocol

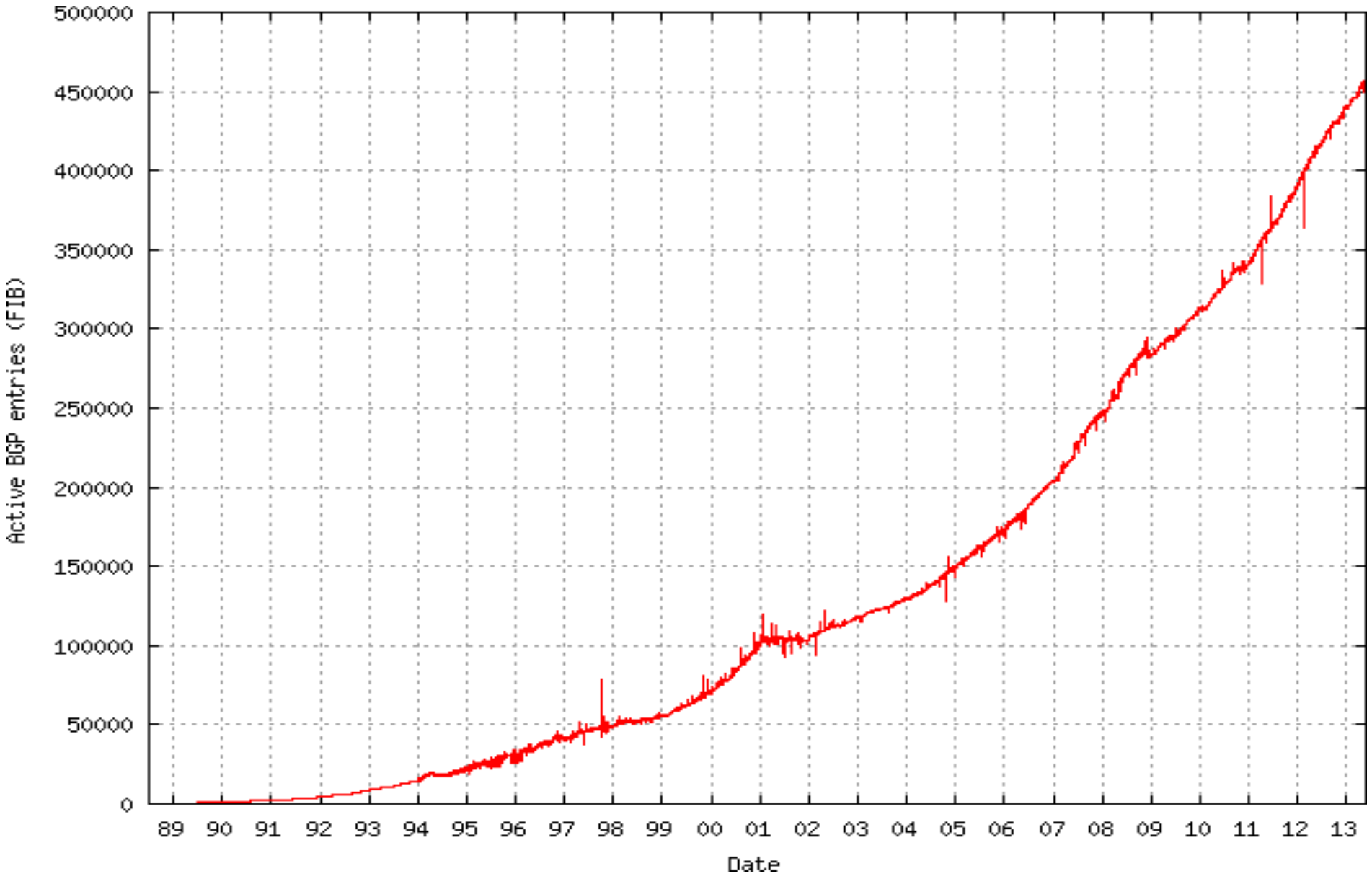


- **Protocole "path vector"**
  - **Liste les AS traversés jusqu'à la destination.**

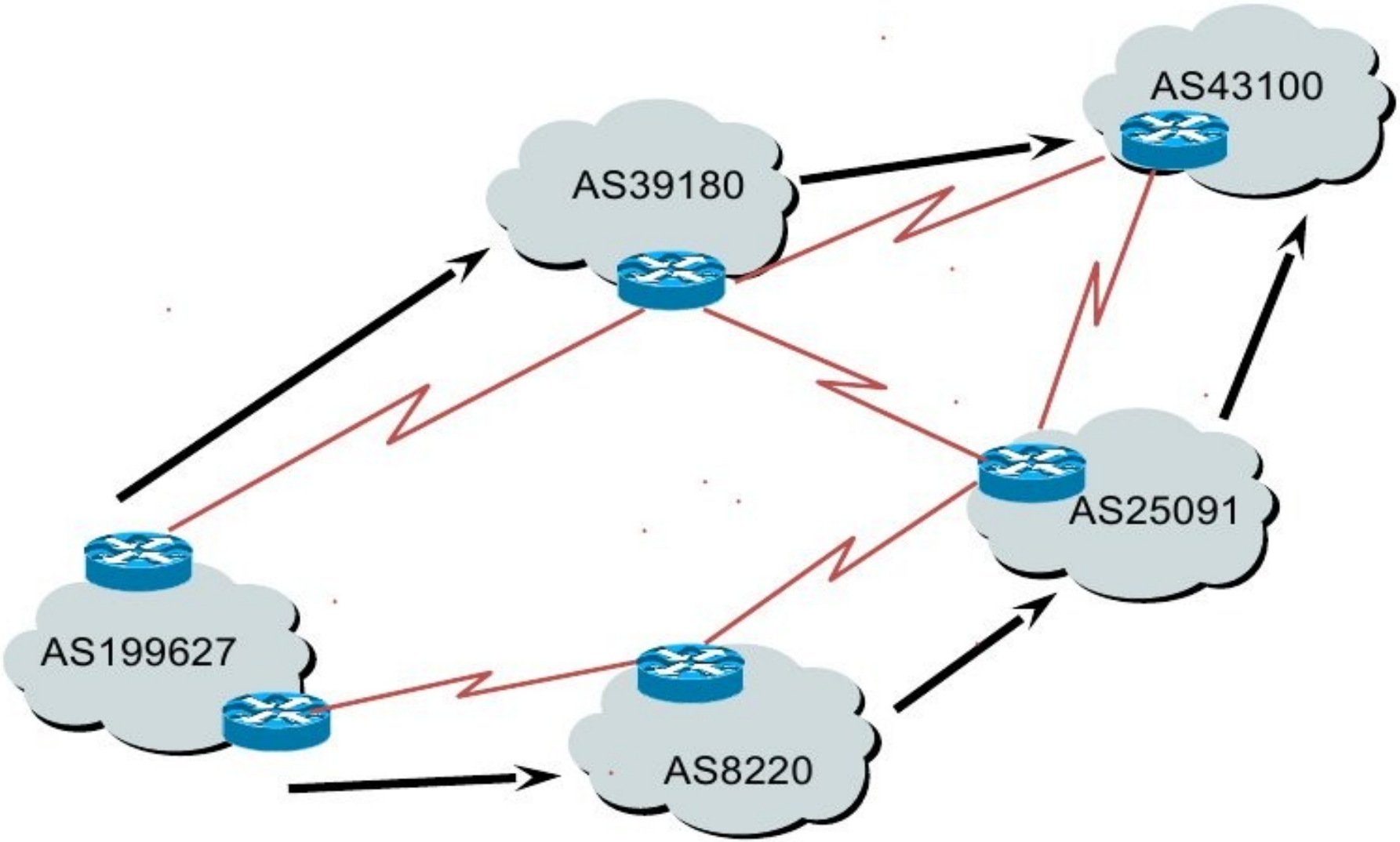
```
cisco#show ip bgp 185.9.20.0/24
BGP routing table entry for 185.9.20.0/24, version 2069615
Paths: (2 available, best #2, table Default-IP-Routing-Table)
  Not advertised to any peer
  25091 8220 199627
    46.20.247.42 from 46.20.247.42 (46.20.247.249)
      Origin IGP, metric 0, localpref 100, valid, external
  39180 199627
    81.18.177.37 from 81.18.177.37 (188.93.40.1)
      Origin IGP, localpref 100, valid, external, best
```

AS Path

Prefixes dans la FIB



Critères de sélection



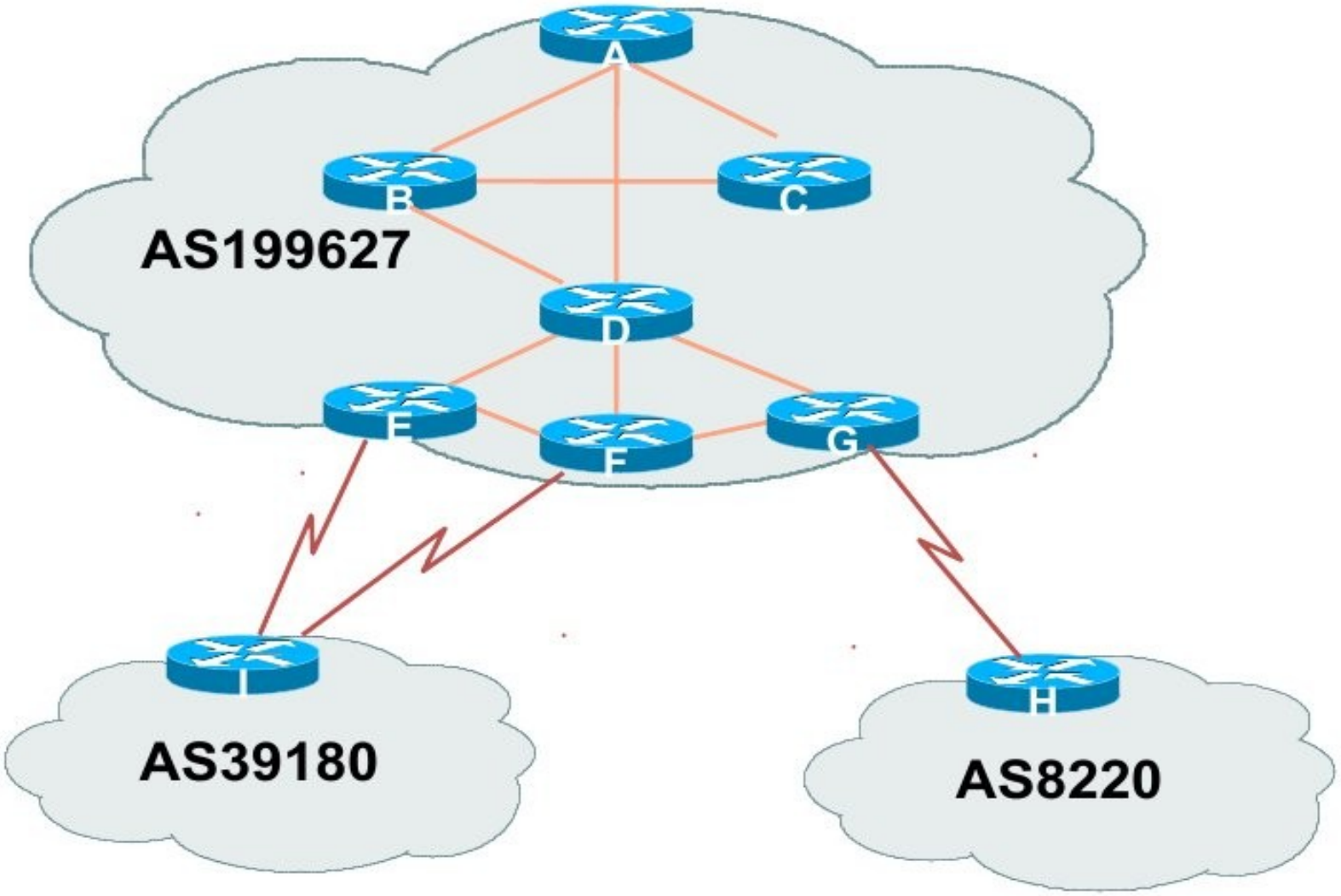
BGP Best Path Selection Algorithm

- Weight le plus grand (cisco only)
- Local Pref la plus grande
- AS Path le plus court
- Origine (IGP, EGP, INCOMPLETE)
- MED (metric) la plus petite
- Route apprise en eBGP puis iBGP
- Chemin le plus ancien

OSPF : Open Shortest Path First

- Interrior Gateway Protocol
  - Protocole de routage utilisé pour échanger des informations de routage dans un même AS
- Découverte des Voisins
- Link state protocol
  - Connaissance de tous les chemins
    - Forte utilisation CPU
- Convergence rapide
  - Si nombre de préfixes +/- égal au nombre de routeurs

EGP vs IGP



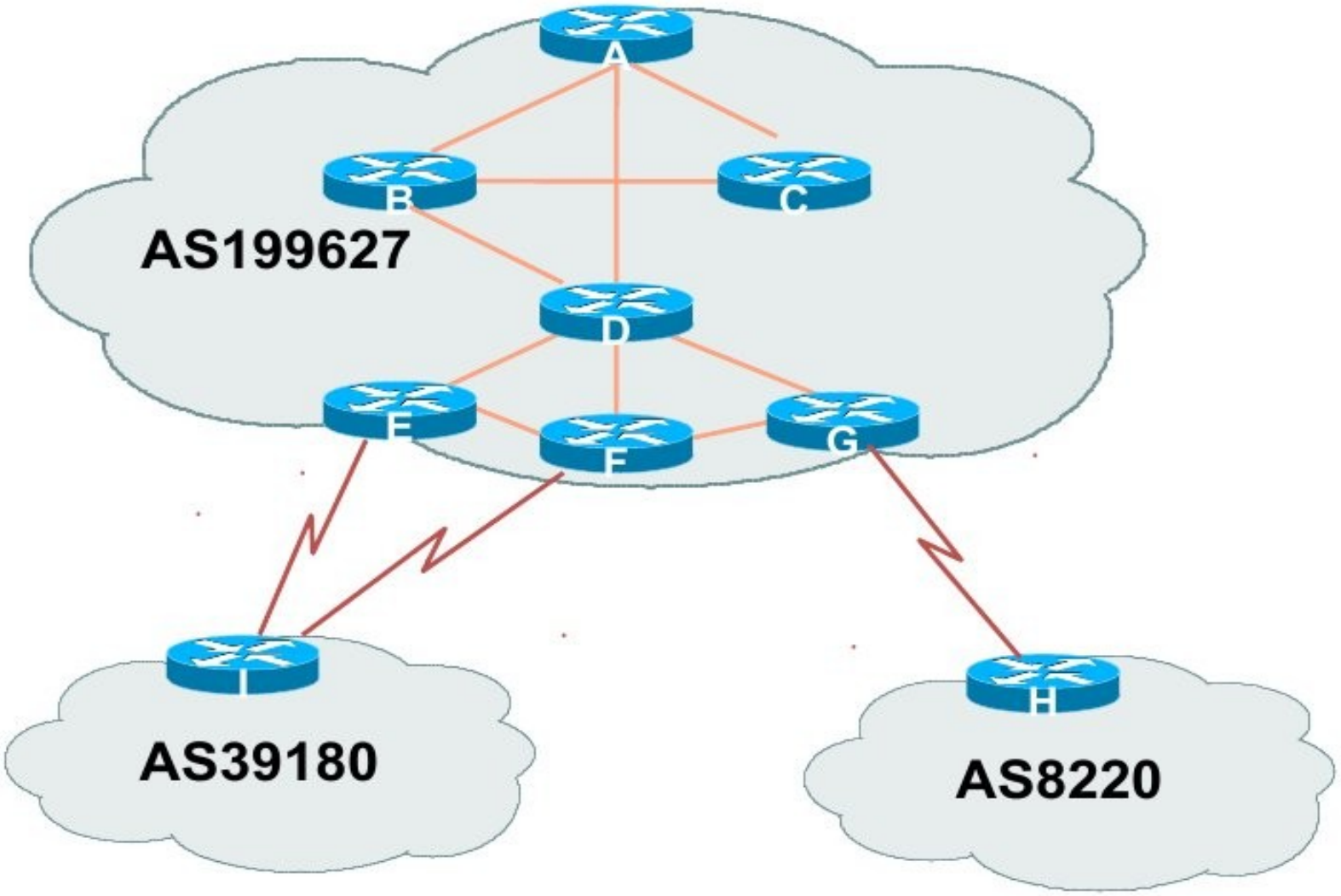
BGP Best Path Selection Algorithm

- Weight le plus grand (cisco only)
- Local Pref la plus grande
- AS Path le plus court
- Origine (IGP, EGP, INCOMPLETE)
- MED (metric) la plus petite
- Préférer les routes apprises en eBGP puis iBGP
- Chemin le plus ancien

eBGP & iBGP

- BGP utilisé en interne (iBGP) et externe (eBGP)
- iBGP utilisé pour transporter
  - Les préfixes Internet à travers le backbone d'un ISP
  - les préfixes des clients d'un ISP
  - Sessions entre les IP loopback
- eBGP utilisé pour
  - Echanger des préfixes avec d'autres AS
  - Mettre en œuvre la politique de routage
  - Sessions entre les IP sur les interfaces

EGP vs IGP



Description du protocole BGP

- Protocole de routage utilisé par tous les acteurs
- Assure la visibilité mondiale
- Ne protège pas les données échangées
- Pas de mécanisme de sécurité fort

- **Longueur de l'AS PATH**
  - Si un préfixe est annoncé par plusieurs AS, les données vont converger vers l'AS le plus proche
- **Préfixes plus spécifiques**
  - Lorsque des annonces se recouvrent
    - par exemple 192.0.2.0/23 et 192.0.2.0/24
  - Le préfixe le plus spécifique (ici 192.0.2.0/24) sera toujours privilégié quelle que soit la longueur de l'AS PATH

- **Contre mesures**
  - Utiliser une directive maximum-prefix (en cas de leak full table)
  - Filtrer grace aux objets Route des IRR
  - Filtrer grace a RPKI + ROA
  - Filtrer les bogons :

    http://www.team-cymru.org/Services/Bogons/http.html
- **Détections**
  - BGP Monitoring and analyzer : http://bgpmon.net
  - Routing Information Service : http://stat.ripe.net
  - BGPlay : https://stat.ripe.net/widget/bgplay
  - Cyclops : http://cyclops.cs.ucla.edu

Incidents médiatisés

- **2008 (hijack)**
  - Pakistan Telecom annonce a son transitaire PCCW des préfixe plus spécifique que ceux des Youtube.
- **2009 (bug d'implémentation)**
  - SuproNet, un operateur Tchèque annonces des chemins d'AS d'une longeur proche de 255 a l'aide d'un routeur MikroTik, déclanchant un bug jusqu'alors inconnu sur les routeurs Cisco.
- **2012 (leak full table)**
  - Dodo annonce a son transitaire Telstra l'intégralité des routes de l'Internet, et se retrouvent isolés de l'internet pendant 30 min.


The CIDR Report

Politique de filtrage

- Politiques basées sur l'AS-PATH, la communauté ou le préfixe
- Acceptation / refus de routes sélectionnées
- Définir les attributs pour influencer la sélection de chemin
- Outils
  - Prefix-list (filtre sur les préfixes)
  - Filter-list (filtre sur les ASs)
  - Route-maps et communautés
  - AS PATH prepend

Rappels CLI cisco 1/2

Passer enable

```
enable
```

passer en mode configuration

```
configure terminal
```

Configurer une interface réseau

```
interface gi1/0
  ip address 192.0.2.42 255.255.255.0
  no shutdown
```

Sortir du mode config de l'interface

```
exit
```

Rappels CLI cisco 2/2

Sortir du mode config

```
end (ou Ctrl + z)
```

Voir la config actuelle

```
show running-configuration
```

Sauvegarder la conf

```
copy running-configuration startup-configuration
```

ou

```
write memory
```

Première session BGP

```
router bgp 42
  network 10.42.0.0 mask 255.255.0.0
  neighbor 192.0.2.1 remote-as 1
exit

ip route 10.42.0.0 255.255.0.0  Null0
```

Deuxième session BGP

```
router bgp 42
  network 10.42.0.0 mask 255.255.0.0
  neighbor 192.0.2.1 remote-as 1
  neighbor 192.0.2.1 prefix-list PFL-AS42-OUT out
exit

ip route 10.42.0.0 255.255.0.0  Null0

ip prefix-list PFL-AS42-OUT seq 5 permit 10.42.0.0/16
```

Default routing

Solutions :

- ACL en INPUT
- Ne pas avoir de transit ni de route par défaut sur le router de l'IXP
- VRF

Spoofing IP source

Solutions :

- ACL en IN sur toutes les interfaces des clients
- ACL en IN sur nos interfaces de serveurs

Troisième session BGP

```
router bgp 42
  network 10.42.0.0 mask 255.255.0.0
  neighbor 192.0.2.1 remote-as 1
  neighbor 192.0.2.1 prefix-list PFL-AS42-OUT out
  neighbor 192.0.2.1 prefix-list PFL-TRANSIT-IN in
  neighbor 192.0.2.1 soft-reconfiguration inbound
exit

ip route 10.42.0.0 255.255.0.0  Null0

ip prefix-list PFL-AS42-OUT seq 5 permit 10.42.0.0/16

ip prefix-list PFL-TRANSIT-IN deny  10.42.0.0/16   le 32
ip prefix-list PFL-TRANSIT-IN deny  10.0.0.0/8     le 32
ip prefix-list PFL-TRANSIT-IN deny  172.16.0.0/12  le 32
ip prefix-list PFL-TRANSIT-IN deny  192.168.0.0/16 le 32
ip prefix-list PFL-TRANSIT-IN permit 0.0.0.0/0     le 24
```

Filtrage des annonces BGP

- **Ne pas coire sur parole**
  - **Je connais l'AS :**

```
whois as199422
whois 77.95.64.0/22
```

  - **Comment savoir ce qu'il peut annoncer**

```
whois 77.95.64.0/22 -T route
whois -i origin AS199422 -T route
whois -i origin AS199422 -T route6
```

  - **Scriptons !**

    http://irrtoolset.isc.org/

```
peval AS199422
peval 'afi ipv6 AS199422'
```

  - **Upstream ? As-set !**

```
whois AS43100:AS-MEMBERS
peval AS43100:AS-MEMBERS
```

Commandes utiles

```
show ip bgp summary
show ip bgp
show ip route
sh ip bgp neighbors X.X.X.X advertised-route
sh ip bgp neighbors X.X.X.X routes
sh ip bgp neighbors X.X.X.X received-routes
clear ip bgp X.X.X.X [in|out]
```

Template / peer-group

- **Utilisation de peer-group**
  - **Factorise la conf**
  - **N'allège pas le CPU (depuis 12.4)**

```
router bgp 42
  network 10.42.0.0 mask 255.255.0.0

  neighbor TRANSIT peer-group
  neighbor TRANSIT prefix-list PFL-AS42-OUT out
  neighbor TRANSIT prefix-list PFL-TRANSIT-IN in

  neighbor 192.0.2.1 remote-as 1
  neighbor 192.0.2.1 peer-group TRANSIT
  neighbor 192.0.2.1 route-map RTM-TRANSIT1-IN in

  neighbor 192.0.2.2 remote-as 2
  neighbor 192.0.2.2 peer-group TRANSIT
  neighbor 192.0.2.2 route-map RTM-TRANSIT2-IN in
exit
```

- **Limitation du nombre de routes apprises**

```
neighbor 192.0.2.1 maximum-prefix 25000
```

  - L'objectif est de couper la session BGP lorsque ce seuil est dépassé
  - Il peut s'agir d'une mauvaise configuration du routeur distant
  - La session ne pourra remonter qu'après l'exécution manuelle de la commande

```
neighbor 192.0.2.1 shutdown
no neighbor 192.0.2.1 shutdown
   ou
clear ip bgp 192.0.2.1
```

- **Local-preference ou weight**
  - On peut vouloir favoriser un transitaire ou un peering même si ce n'est pas le meilleur chemin d'après le protocole BGP
  - Ainsi du peering peut être préféré au transit car celui-ci est gratuit
  - Le poids sera local au routeur, alors que la local-pref pourra être propagé aux routeurs adjacents (même AS)

```
neighbor TRANSIT weight 100
neighbor LYONIX weight 200
```

- **Avec les route-maps, on peut créer des filtres plus complexes que des simples restrictions sur les addresses IP**

```
route-map RTM-LYONIX-OUT deny 5
  match ip address prefix-list PFL-NETWORK1
route-map RTM-LYONIX-OUT permit 10
  set community 43100:9999


route-map RTM-LYONIX-IN permit 5
  match ip address prefix-list PFL-NETWORK2
  set local-preference 1000
route-map RTM-LYONIX-IN permit 10
```

Session BGP

```
router bgp 42
  network 10.42.0.0 mask 255.255.0.0
  neighbor 192.0.2.1 remote-as 1
  neighbor 192.0.2.1 prefix-list PFL-AS42-OUT out
  neighbor 192.0.2.1 prefix-list PFL-TRANSIT-IN in
  neighbor 192.0.2.1 route-map RTM-AS1-IN in
exit

…

route-map RT-AS1-IN permit 5
  match ip address prefix-list PFL-NET1
  set local-preference 1000

route-map RT-AS1-IN permit 10
```

- **AS path**
  - **Ex : 1 23456 15557 43100**

```
router bgp 42
  network 10.42.0.0 mask 255.255.0.0
  neighbor 192.0.2.1 remote-as 1
  neighbor 192.0.2.1 route-map RTM-AS1-IN in
exit

ip as-path  access-list  1 permit  15557
ip as-path  access-list  1 permit  _23456_
ip as-path  access-list  1 permit  ^1
ip as-path  access-list  1 permit  43100$

route-map LYONIX-in permit 5
 match as-path 1
 set local-preference 1000
```

BGP Attributes

- **Attributs BGP**
  - **Well-knon Mandatory**

    doit etre supporté et propagé
    - **Origine (IGP, EGP, unknown)**
    - **AS Path**
    - **Next Hop**

  - **Well-know Discretionary**

    doit supporté, peut être propagé
    - **Local Preference**
    - **Atomic Aggregate**

      (AS supprimés en cas d'aggrégation de route)

BGP Attributes

- **Attributs BGP**
  - **Optional Transitive**

    peut etre supporté, doit être propagé (identifié comme partial)
    - **Aggregator**

      ID et AS du routeur agrégeant
    - **Community**
  - **Optional Nontransitive**

    supprimé si non supporté
    - **MED (Metric)**
    - **Originator ID**
    - **Cluster ID/List**
    - **Weight**

- **Communautés BGP**
  - Attribut utilité pour marquer les routes apprises ou annoncées
    - Utilisé pour filtrer
    - Utilisé pour indiquer a quel endroit a été apprise la route
    - Utilisé sur les Route Serveurs Rezopole
    - Permet de restreindre la diffusions a certaines zones géographiques

- **Une communauté par IXP**

  - 10 LyonIX
  - 20 TopIX
  - 70 TouIX
  - 80 EquinIX PA
  - 90 SfinX
  - 100 Fr-IX
  - 110 FranceIX

BGP Communities LyonIX

- A quel endroit a été apprise cette route ?
  - 43100:1xxx
  - 43100:1000 + IXP Number
    - 43100:1010 LyonIX
    - 43100:1020 TopIX
    - ...

BGP Communities LyonIX

- **Ne pas exporter mon préfixe sur un IXP spécifique**
  - 43100:9xxx
  - 43100:9000 + IXP Number
    - 43100:9090 pas d'export sur Sfinx
    - 43100:9110 pas d'export France-IX
  - Annoncer uniquement sur LyonIX
    - 43100:9999 ne pas annoncer sur les IXP en dehors de la Région Rhone Alpes

- Je ne veux pas annoncer mes préfixes a un ASN!
  - 0:ASN
  - Filtrage par ASN (ex AS1234)
    - 0:1234
  - Ne fonctionne qu'avec les ASN 16 bits

- **Communautés BGP**

```
set community community-number [additive]
set community [well-known-community]
```

  - **well known communities :**
    - **No-export (65535:65281)**

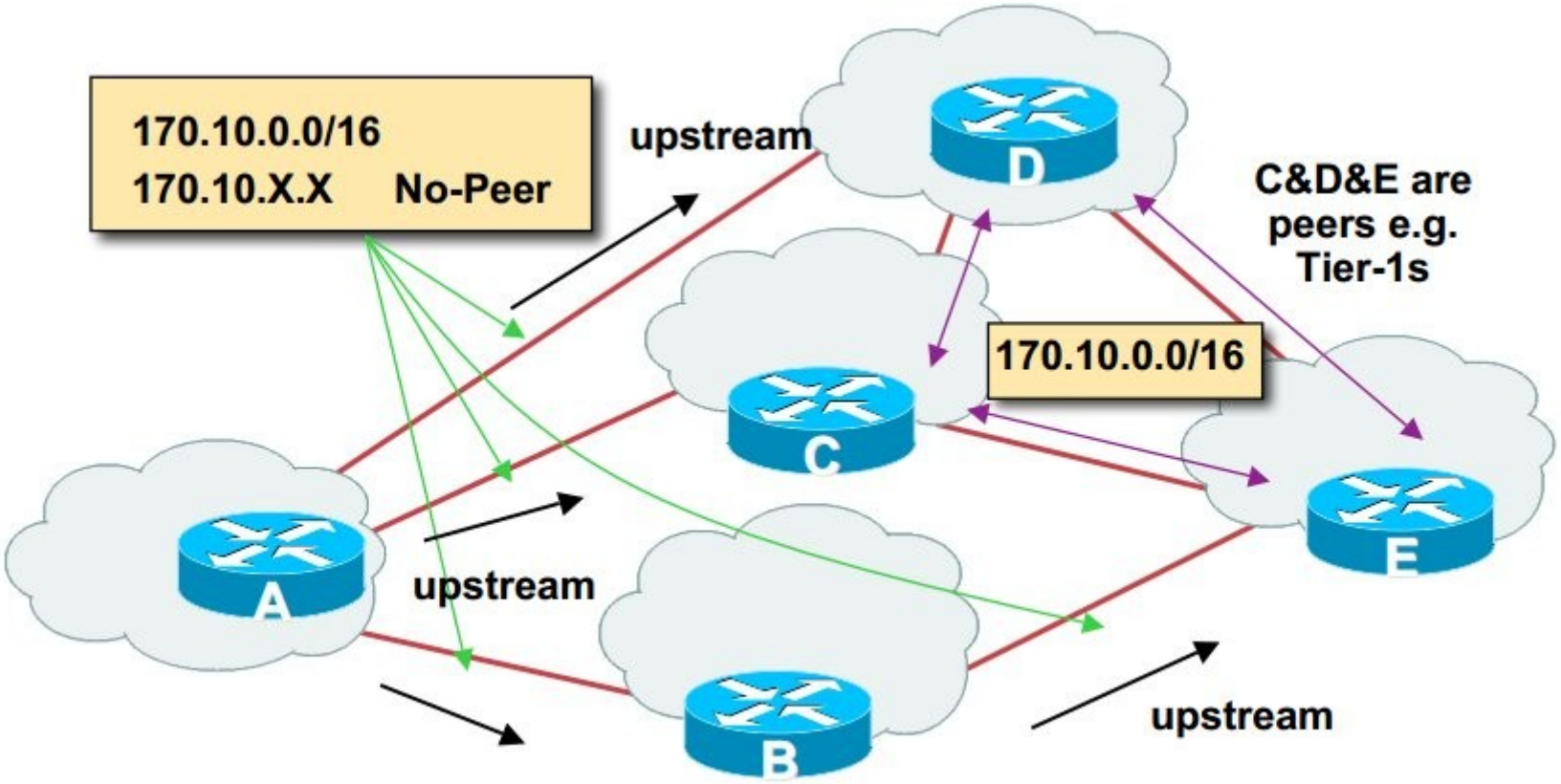
      Ne pas annoncer au voisins eBGP.

      Garder cette route dans l'AS.
    - **No-advertise (65535:65282)**

      Ne pas annoncer cette route a aucun voisin iBGP ou eBGP
    - **No-peer (65535:65284)**

      Ne pas annoncer en peering, uniquement aux upstreams

- **No-peer**


170.10.0.0/16
170.10.X.X No-Peer
upstream
C&D&E are peers e.g. Tier-1s
170.10.0.0/16
upstream
upstream
A
B
C
D
E

BGP Communities Examples

- **configuration**

```
ip bgp-community new-format

router bgp 42
  network 10.42.0.0 mask 255.255.0.0
  neighbor 192.0.2.1 remote-as 1
  neighbor 192.0.2.1 send-community
  neighbor 192.0.2.1 route-map RTM-LYONIX-IN in
  neighbor 192.0.2.1 route-map RTM-LYONIX-OUT out
exit

ip community-list standard CML-FROM-LYONIX permit 43100:1010

route-map RTM-LYONIX-IN permit 5
  match community CML-FROM-LYONIX

route-map RTM-LYONIX-OUT permit 5
  set community 43100:9999
```

Trafic shaping : AS Path Prepend

- Allonger la longueur de l'AS PATH en ajoutant plusieurs fois son numero d'AS
  - partage de charge sur trafic entrant

- Généralement utilisé sur le lien de backup
  - Si AS PATH plus long, trafic quasi nul

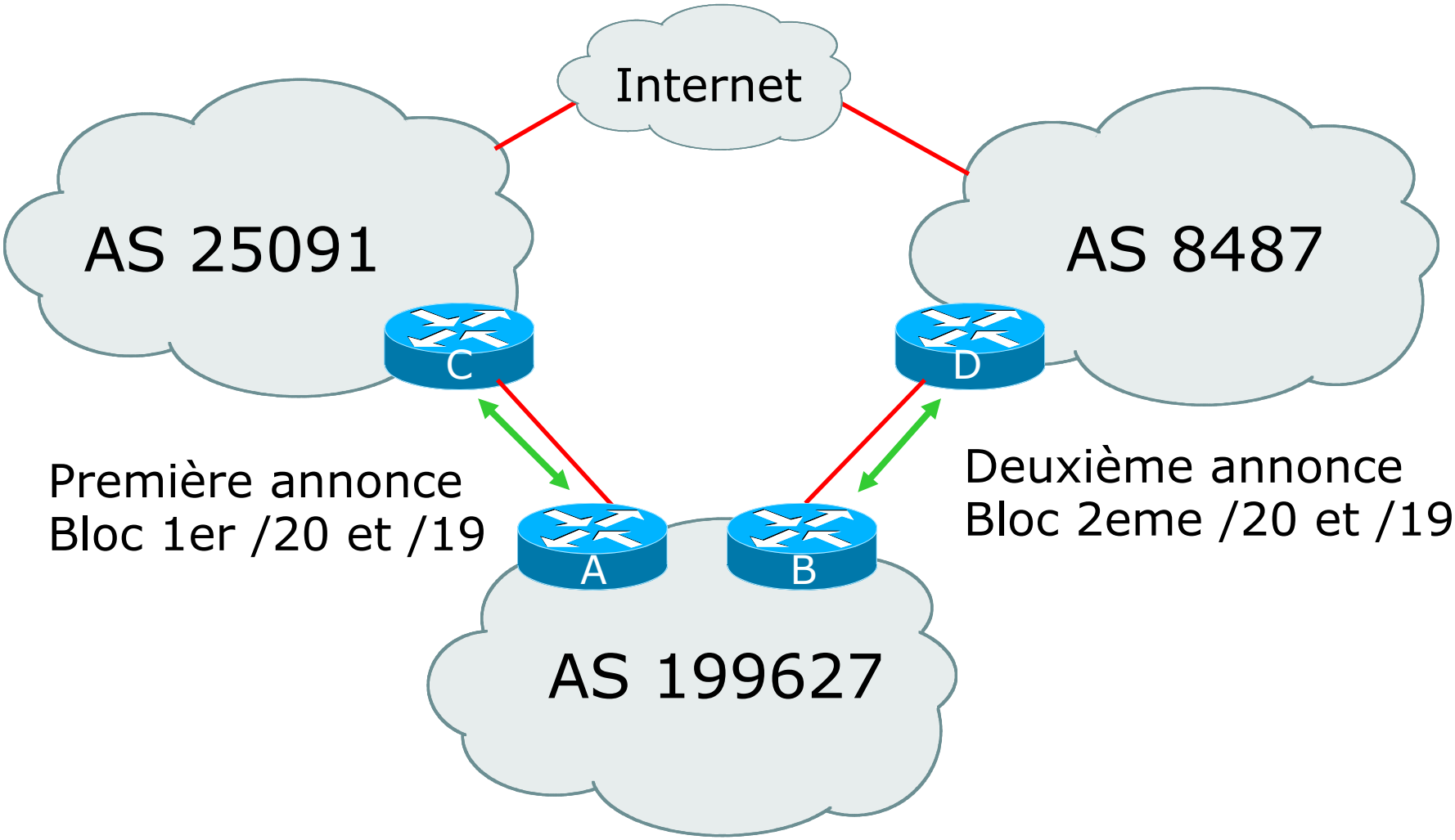
```
route-map RTM-LYONIX-OUT permit 10
 set as-path prepend 42 42 42
```

Trafic shaping : désagrégation

- Annoncer l'agrégat /19 sur chaque lien
- Annoncer deux / 20s, un sur chaque lien
  - partage de charge sur trafic entrant
- Ne partagera probablement pas la charge réseau en 50/50
  - Possibilité de désagréger par /24

    ! crée bcp de routes « inutiles » dans le DFZ !

Trafic shaping : désagrégation



Blackhole

- **Se protéger des DdoS**
  - **Ne pas recevoir le trafic pour une/plusieurs de nos IP**
    - **Annoncer une plage spécifique avec une communauté donnée**
    - **Le transitaire Null-route le trafic qui est a destination de ces IP**
    - **On perd une partie du service (serveur injoignable)**

      Mais le réseau survie

  - **Autres solutions :**
    - **Wanguard**
    - **Arbor networks / RedWare**
    - **CDN (CloudFlare, ...)**

- La pénurie d'adresses a été retardée par :
  - L'utilisation du NAT mais brise la philosophie de bout en bout (end to end)
  - L'utilisation de DHCP
  - routage CIDR
- IPv4 n'est pas un protocole adapté pour :
  - l'auto configuration (plug and play)
  - la mobilité → LISP ?
  - la gestion de la qualité de service (QoS)
  - la gestion de la sécurité en natif

- **Historique**
  - Début des travaux au milieu des années 1980 pour améliorer IP
  - IPv6 retenu comme nouveau standard (RFC 1752) et adopté vers la fin des années 1990
  - Type Ethernet 86DD (IPv4 = 0800)
- **Différence avec IPv4**
  - NDP vs ARP
  - Adresses Link-local fe80::/64
  - Addressage avec mac addr

IPv6 – les apports

- Adresses sur 128 bits (2^128 = 3,4x10^38)

- En tête plus simple (8 champs au lieu de 13 en IPv4)

  de taille fixe pour améliorer les performances et

  intégrer de nouvelles fonctionnalités par un mécanisme de liste chaînée d'extensions

- Couche IPSec intégrée au protocole
- extension du multicast et abandon du broadcast

- Chaque sous réseau devrait être un /64.

  On affecte en général un /48 à une entreprise

  (65 536 sous réseaux).

IPv6 – Espace d'adressage

- C’est la plus flagrante évolution mise en avant lorsque l’on parle d’IPv6.
- Ce plus grand nombre d’adresses disponibles permettra d’adresser et d’identifier tous les équipements communicants.
- Il permettra de s’affranchir des NATs, de déployer de nouvelles applications nécessitant des communications de bout en bout (téléphonie, vidéo-conférence, sécurité de bout en bout).
- NAT64 et Nat66 / NPT

- **NDP (Neighbor Discovery Protocol)**

  L'auto-configuration met en œuvre en un certain nombre de nouveaux protocoles associés à IPv6 : protocole de découverte des voisins, nouvelle version d'ICMP, etc.

- **Router Advertisment**

  - L'auto-configuration permet à un équipement de devenir complètement « plug-and-play ». Il suffit de connecter physiquement la machine pour qu'elle acquière automatiquement une adresse IPv6 et une route par défaut.

- **Pas de DHCP ?**

Session BGP

```
ipv6 unicast-routing
 router bgp 42
   neighbor 2001:7f8:47:47::1 remote-as 1

   address-family ipv4
     no neighbor 2001:7f8:47:47::1 activate
   exit

   address-family ipv6
     network 2001:db8:42::/48
     neighbor 2001:7f8:47:47::1 activate
     neighbor 2001:7f8:47:47::1 prefix-list PFL-AS42-OUT out
   exit
 exit

 ipv6 route 2001:db8:42::/48 Null0
```

Commandes utiles

```
show bgp ipv6 unicast summary
show bgp ipv6 neighbors X:X:X::X routes
show bgp ipv6 neighbors X:X:X::X advertised-routes

clear bgp ipv6 X:X:X::X
```

BGP Convergence Rapide

- **Timer BGP**
  - Durée pour laquelle la session est considée UP
    - Par défaut : keepalive 60s / holdtime 180s (KA 3s/HT 9s OK)

      `neighbor X.X.X.X timers` *keepalive holdtime*
- **Advertisement Interval**
  - Délais entre 2 updates
    - Par défaut : eBGP 30s / iBGP 0s

      `neighbor X.X.X.X advertisement-interval` *value*
- **Fast Peering Session Desactivation**
  - Retrait immédiat des routes apprises via l'interface si elle tombe
    - par defaut activé en eBGP uniquement

      `bgp fast-external-fallover`

BGP Convergence Rapide

- **Temps de covergence de l'IGP**
  - Valeurs à définir dans l'OSPF/ISIS
    - un évènement peut etre détecté en moins de 1s
- **BGP Scanner**
  - Vérification Next-Hop joignable
    - Par défaut : toutes les 60s

```
bgp scan-time value
```

- **Next Hop Trigger**
  - Temps pour prendre en compte un evenement IGP
    - Par défaut : 5s (0 seconde est OK)

```
bgp nexthop trigger delay value
```

BGP Convergence Rapide

- **BGP Multipath Load Sharing**
  - Avoir plusieurs chemins égaux dans la FIB
    - Par défaut : désactivé (seuls les next-hop sont différents)

```
maximum-paths value
```

- **BGP Multiple path propagation**
  - Propager une route backup vers la meme destination
    - Par défaut : désactivé (seulement la best path annoncée)

```
bgp additional-paths install
```

- **Birirectional Forwarding Detection**
  - Détécter une panne dans le Forwarding plane en qqlq ms

```
interface GiX/X
  bfd interval value min_rx value multiplier value
router bgp asn
  neighbor X.X.X.X fall-over bfd
```

BGP défauts

- Décisions de routage basées sur la longueur administrative des chemins

  (+Tie-Break aléatoire)
- Pas de contrôle de la performance
  - Longueur effective des chemins (Round Trip Delay)
  - Qualité du chemin (Perte de paquets, variation du délai)
- Mauvaise détection des pannes
  - e.g. Access-list
  - e.g. « Forwarding plane » rompu
- Pas de gestion de la capacité
  - Saturation d'un lien entrée ou sortie
  - Dissymétrie

Délais vers mon réseau

- Looking Glass

  http://lookinglass.org

- RIPE Atlas

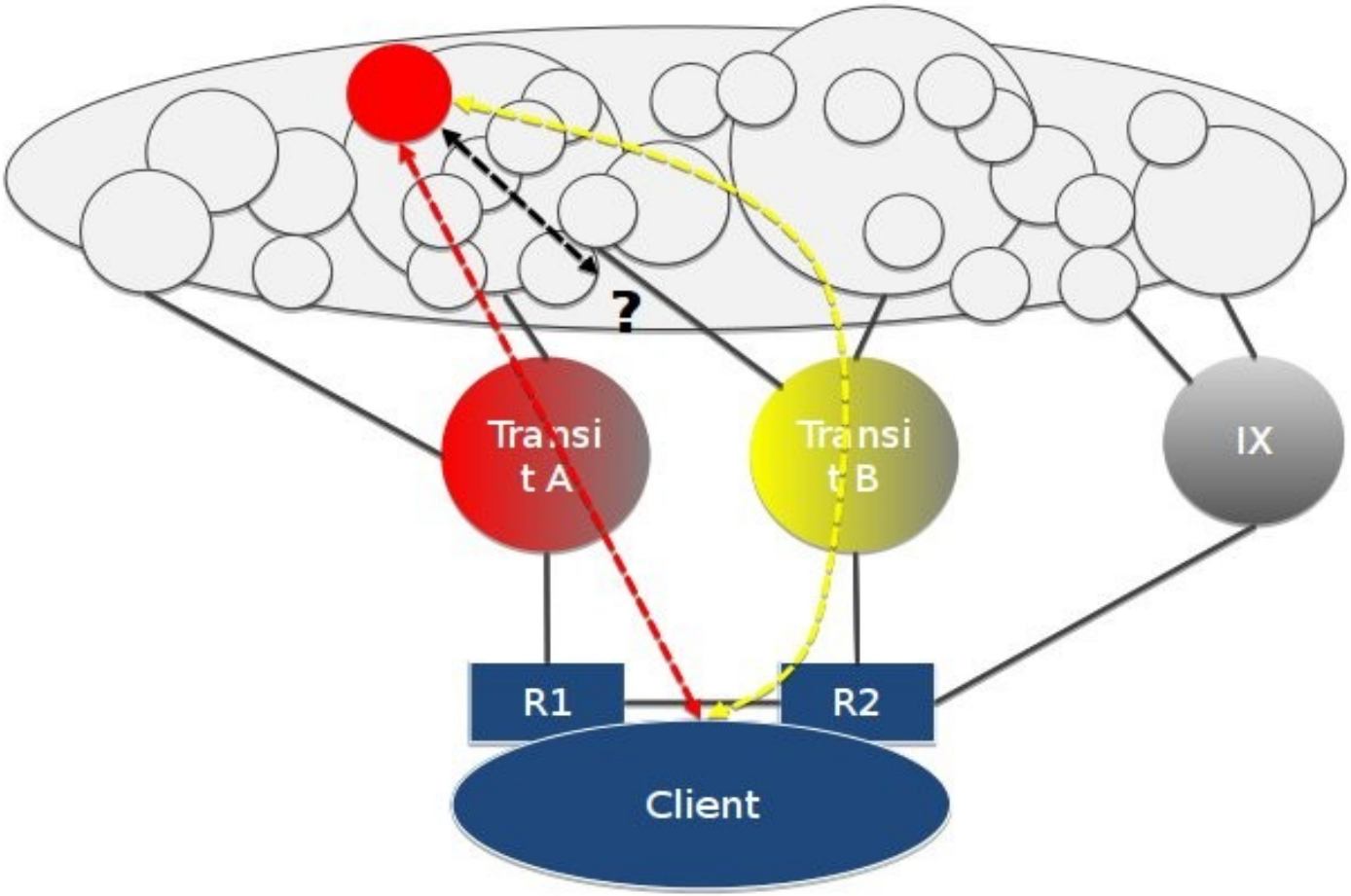
  https://atlas.ripe.net/

- NLNOG RING

  https://ring.nlnog.net

- www.just-ping.com

Delais depuis mon réseau ?



BGP Qualité/packetloss

	BGP	Transit A	Transit B	
Location	Average RTT	Average RTT	Average RTT	Difference
Paris, France:	6.1	101.7	6.5	**1464.6%**
Singapore, Singapore:	**334.7**	335.1	273.3	22.6%
Zurich, Switzerland:	27.5	194.5	28.3	**587.3%**
Groningen, Netherlands:	**82.7**	15.8	86.7	**448.7%**
Beijing, China:	**441.2**	355	441.3	24.3%

- **Délai**
  - **Transit**
    - **Vert = meilleur chemin**
  - **BGP**
    - **Blanc = BGP a choisi le meilleur chemin**
    - **Rouge = le chemin choisi est au moins 10% plus long que le chemin le plus court**

BGP Qualité/packetloss

ICMP uniquement ?

- Quid d'une QOS « mal » placé ?
- Mesurer TCP handshake
- Chargement des pages

Solutions

- Smokepings (avec curl) vers destinations importantes
- Weathermap/MRTG/Observium
- Alertes Nagios
- Cisco IP SLA
- Border6 / Noction / Internap FCP

Configurer son interface Switch sur un IXP

```
interface Gi0/1

  switchport mode access

  switchport access vlan 500

  spanning-tree bpdufilter enable

  no cdp enable

  no keepalive
```

Configurer son interface Routeur sur un IXP (1/2)

```
no ip arp gratuitous
no ip gratuitous-arps

interface Gi1/0
  ip address A.B.C.N 255.255.255.NNN
  ip access-group IXP-DMZ-IN4 in
  ip access-group IXP-DMZ-OUT4 out

  no ip redirects
  no ip proxy-arp
  no ip igmp version
  no cdp enable
```

Configurer son interface Routeur sur un IXP (2/2)

```
ipv6 address 2001:7F8:47:47::AAAA/64
ipv6 traffic-filter IXP-DMZ-IN6 in

ipv6 nd suppress-ra
ipv6 nd ra suppress
no ipv6 redirects

no ipv6 pim
no mld snooping
no ipv6 mld router
no ipv6 mfib forwarding
```

Configuration BGP sur un IXP

Limiter le TTL a 1

```
neighbor 192.0.2.1 remote-as 65000
neighbor 192.0.2.1 ttl-security hops 1
```

Utiliser des Mdp MD5

```
neighbor 192.0.2.1 password S3cr3tP4ssw0rd
```

Filtrer

```
neighbor 192.0.2.1 prefix-list PFL-ASN-IN in
neighbor 192.0.2.1 prefix-list PFL-MYPFX-OUT out
neighbor 192.0.2.1 maximum-prefix 50
```

Si Route-serveur

```
no bgp enforce-first-as
```

BGP en résumé ?

Filtrez !

Qqlq liens utiles

BGP Best Path Selection Algorithm
http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094431.shtm

BGP Convergence explained
http://blog.ine.com/2010/11/22/understanding-bgp-convergence/

Best Practice iBGP/eBGP
http://www.ripe.net/ripe/meetings/regional-meetings/manama-2006/BGPBCP.pdf

Cheat Sheets Cisco
http://packetlife.net/library/cheat-sheets/

Influence des bonnes pratiques sur les incidents BGP
https://www.sstic.org/2012/presentation/influence_des_bonnes_pratiques_sur_les_incidents_bgp

Best current practices BGP :
http://www.ssi.gouv.fr/IMG/pdf/guide_configuration_BGP.pdf
http://www.ssi.gouv.fr/IMG/pdf/rapport-observatoire-20130617.pdf
http://tools.ietf.org/html/draft-ietf-opsec-bgp-security-01

Liste Bogons :
http://www.team-cymru.org/Services/Bogons/http.html

BGPlay
https://stat.ripe.net/bgplay

Looking glass
https://stat.ripe.net/widget/looking-glass

Simultaeurs routeurs (cisco / juniper)
http://www.gns3.net/

Questions ?

Réponses !