

BGP, le protocole de routage de l'Internet

Stéphane Bortzmeyer stephane+cnam@bortzmeyer.org

CNAM, 9 janvier 2018

Plan du cours

- 1 Le problème
- 2 Le protocole
- 3 Opérationnel
- 4 Sécurité
- 5 Conclusion

L'Internet

- Réseau mondial,
- Pas de Chef Suprême (BGP va être pair-à-pair),
- Pas de maillage total. Comment je vais de SFR à l'Université de Bangkok, sachant que les deux opérateurs n'ont aucun lien ?

Rappel IP

- Adresses comme `2001:db8:42:cafe::1:85f`,
- **Préfixes** comme `2001:db8:42::/48` (BGP route des préfixes),

Table de routage

(Devrait s'appeler table de transmission, en fait.) Sur un Linux :

```
% ip -6 route show
local ::1 dev lo proto kernel metric 256
2001:4b98:dc2:43::/64 dev eth0 proto kernel metric 256 expires 2591706sec
2001:4b98:dc2:45::/64 dev eth1 proto kernel metric 256 expires 2591545sec
fe80::/64 dev eth0 proto kernel metric 256
fe80::/64 dev eth1 proto kernel metric 256
default via fe80::643 dev eth0 proto ra metric 1024 expires 1506sec hoplimit 64
```

Table de routage, suite

Sur un Cisco :

```
route-views>show ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       ...
       1.0.0.0/8 is variably subnetted, 2638 subnets, 20 masks
B       1.0.4.0/22 [20/0] via 114.31.199.1, 2d01h
B       1.0.4.0/24 [20/0] via 114.31.199.1, 2d01h
B       1.0.5.0/24 [20/0] via 114.31.199.1, 2d01h
B       1.0.6.0/24 [20/0] via 114.31.199.1, 2d01h
B       1.0.7.0/24 [20/0] via 114.31.199.1, 2d01h
B       1.0.16.0/24 [20/0] via 202.232.0.2, 4d20h
B       1.0.64.0/18 [20/2523] via 208.51.134.254, 6d21h
B       1.0.128.0/17 [20/0] via 5.101.110.2, 2d00h
B       1.0.128.0/18 [20/0] via 5.101.110.2, 2d00h
B       1.0.128.0/19 [20/0] via 5.101.110.2, 2d00h
B       1.0.128.0/24 [20/0] via 95.85.0.2, 6d21h
       ...
```

Table de routage, attention

- Routage (*routing*) : calculer les tables de routage. C'est le travail de BGP.
- Transmission (*forwarding*) : envoyer le paquet. C'est le travail d'IP.

The data plane (IP) does not always follow the control plane (BGP).

La transmission est ultra-rapide, le routage prend plus de temps et est plus complexe.

traceroute

```
traceroute to yeti-ns.lab.nic.cl (2001:1398:1:21::8001), 30 hops max, 80 byte pa
...
 6  2001:67c:2218:255::8:10 (2001:67c:2218:255::8:10)  13.816 ms  13.223 ms  12.
 7  xe-7-0-3.bar1.Marseille1.Level3.net (2001:1900:5:2:2::421)  16.124 ms  15.92
 8  lo-22-v6.edge3.Paris1.Level3.net (2001:1900:2::8)  12.697 ms  12.871 ms  12.
 9  gblx-level3-xe.Paris1.Level3.net (2001:1900:5:3::126)  19.284 ms  19.286 ms
10  * * *
11  2001:450:2002:af::2 (2001:450:2002:af::2)  235.567 ms  235.570 ms  235.191 m
12  2800:1f0:8000::4 (2800:1f0:8000::4)  235.150 ms  235.463 ms  235.481 ms
13  * * *
14  yeti-ns.lab.nic.cl (2001:1398:1:21::8001)  263.425 ms  260.123 ms  260.081 m
```

(2800:1f0:8000::4 est le FAI chilien Adexus)

Les opérateurs

Les organisations qui font du BGP

- Pas le petit FAI ou petit hébergeur Web
- Critère possible : au moins deux connexions Internet
- En France : Gandi, Renater, Free/Proxad, SFR, Orange, OVH, Gitoyen, AFNIC...

Notion de système autonome (AS)

Un AS est un ensemble de routeurs sous une direction unique. Dans un AS, on peut faire des choix communs (comme « on déploie IPv6 »).

Identifié par un ASN (numéro d'AS). 12322 est Free, 16276 OVH, 2484 l'AFNIC, 42 identifie PCH...

Appairage et transit

- Appairage (*Peering*) : connexion entre pairs, en général gratuite et informelle ; le pair n'envoie que ses routes
- Transit : connexion à un opérateur plus gros, payante : le transitaire envoie toutes les routes
- *Tier 1* : opérateur qui n'achète pas de transit du tout (Tata, Level 3, OpenTransit...)
- Et c'est ainsi que l'Internet émerge, par ces appairages et transits.

Les routeurs

- Fondamentalement, un routeur BGP est un ordinateur,
- Il a beaucoup de logiciel donc beaucoup de bogues (exemple de l'attribut 99 sur Cisco),
- Première différence : il a beaucoup d'interfaces,
- Deuxième différence : le processeur qui fait BGP ne fait pas la transmission des paquets (routage, mais pas transmission).

IGP et EGP

- IGP (*Interior Gateway Protocol*, à l'intérieur d'un AS) : une seule administration, des choix cohérents. RIP, OSPF, IS-IS... Chaque opérateur choisit le sien.
- EGP (*Exterior Gateway Protocol*, entre AS) : multi-organisations, pas de chef, pas d'uniformité. Forcément un seul, BGP.

Les préfixes IP

- Les RIR (*Regional Internet Registry*) attribuent les préfixes IP. En Europe, le RIR est le RIPE-NCC.
- Un opérateur est un LIR (*Local Internet Registry*). Il est membre du RIR.
- Le LIR demande des préfixes au RIR, qui garde la base de données des préfixes alloués.
- On peut récupérer cette information avec whois ou RDAP.

whois

```
% whois 2001:678:c::1
...
inet6num:      2001:678:c::/48
netname:       NIC-FR-DNS-ANYCAST-AFNIC-V6
country:       FR
org:           ORG-AFp11-RIPE
created:       2009-06-03T15:13:50Z
last-modified: 2016-04-14T09:47:44Z

organisation:  ORG-AFp11-RIPE
org-name:      AFNIC (Association Francaise pour le Nommage Internet en Coopera
org-type:      LIR
address:       Immeuble Le Stephenson 1 rue Stephenson
address:       78180
address:       Montigny-le-Bretonneux
address:       FRANCE
...
```

Et avec RDAP

```
% curl -s https://rdap.lacnic.net/rdap/ip/2800:1f0:8000::4 | \
jq .events
[
  {
    "eventAction": "registration",
    "eventDate": "2007-12-27T14:00:00Z"
  },
  {
    "eventAction": "last changed",
    "eventDate": "2011-11-30T21:45:00Z"
  }
]
```

Avantage : format de sortie analysable par un programme.

Principes de base

Border Gateway Protocol, RFC 4271

- Deux routeurs décident de s'appairer,
- Ils établissent une connexion TCP, port 179 (ces connexions peuvent durer des semaines),
- Ils annoncent les nouveautés, nouvelles routes (ANNOUNCE) et routes supprimées (WITHDRAW),
- BGP transmet uniquement les nouveautés.

Exemple d'une annonce BGP

(Format de bgpdump, à partir de MRT, RFC 6396)

```
TIME: 01/01/18 00:00:57
TYPE: BGP4MP/MESSAGE/Update
FROM: 66.185.128.1 AS1668
TO: 128.223.51.102 AS6447
ASPATH: 1668 10310 3356 16637
NEXT_HOP: 66.185.128.1
ANNOUNCE
  41.181.164.0/24
  41.181.174.0/24
```

```
TIME: 01/01/18 00:00:57
TYPE: BGP4MP/MESSAGE/Update
FROM: 196.7.106.245 AS2905
TO: 128.223.51.102 AS6447
WITHDRAW
  41.180.0.0/16
```

Les chemins d'AS

- Le chemin d'AS se lit de droite à gauche,
- Chaque routeur ajoute son AS avant de propager aux pairs,
- Le premier AS (tout à droite) est l'origine.

Décisions

- Le routeur refuse les annonces qui incluent son AS. Ensuite,
- S'il y a plusieurs possibilités pour un même préfixe,
- On prend celle ayant la meilleure préférence locale,
- Si égalité, on prend la route ayant le chemin d'AS le plus court,
- Si égalité, on prend celle émise par le routeur ayant le plus petit identificateur BGP.

(L'algorithme est ici simplifié, il y a d'autres critères.)

Envoi au voisin/pair

- Le routeur BGP propage ensuite l'annonce à ses voisins,
- Au bout de quelques minutes, tous les routeurs l'auront reçu (le test de la présence de son AS évite les boucles),
- Acceptation et propagation sont contrôlés par une politique locale. BGP est du routage politique, plus que technique.

Pas de vision globale

- Les routeurs ne connaissent donc pas tout le réseau (\neq OSPF), puisque chacun ne transmet que les routes sélectionnées,
- Les routeurs n'ont pas tous exactement la même information,
- Quand on observe, il faut donc toujours dire depuis quel routeur.

Configuration d'un routeur BGP

BIRD sur Unix

```
router id 192.0.2.1;
protocol static static_bgp {
    import all;
    route 192.0.2.0/24 reject;
}
protocol bgp_64510 {
    import all;
    export where proto = "static_bgp";
    local as 64496;
    neighbor 203.0.113.101 as 64510;
}
protocol bgp_64511 {
    import all;
    export where proto = "static_bgp";
    local as 64496;
    neighbor 203.0.113.128 as 64511;
}
```

Configuration d'un routeur BGP, suite

Cisco IOS

```
router bgp 64497
  bgp router-id 198.51.100.1
  neighbor 2001:db8:36a::1:1 remote-as 64508
  neighbor 2001:db8:ff99:f53d::126 remote-as 64509

  address-family ipv6
    neighbor 2001:db8:36a::1:1 activate
    neighbor 2001:db8:ff99:f53d::126 activate
    network 2001:db8:42::/48
  exit-address-family
```

Attention avec les exemples simples

Ils ne présentent aucune sécurité (pas de filtrage).

Une configuration BGP est souvent simple mais doit être étudiée soigneusement.

Communautés

- RFC 1997,
- Des étiquettes qu'on attache aux annonces,
- Servent à tout : indiquer où la route a été apprise, quel est le traitement souhaité. . .
- Traditionnellement sous la forme AS:XXX,
- La sémantique dépend de l'AS, sauf pour les communautés bien connues.

Annonce avec communautés

```
TIME: 02/17/17 15:00:00
TYPE: BGP4MP/MESSAGE/Update
FROM: 208.51.134.246 AS3549
TO: 128.223.51.102 AS6447
ASPATH: 3549 3356 2914 30259
NEXT_HOP: 208.51.134.246
ATOMIC_AGGREGATE
AGGREGATOR: AS30259 10.11.1.1
COMMUNITY: 2914:410 2914:1001 2914:2000 2914:3000 3356:3 3356:86
           3356:575 3356:666 3356:2011 3356:11940 3549:2011 3549:2017
           3549:2521 3549:2582 3549:2950 3549:2991 3549:30840 3549:31826
           3549:32344 3549:33036 3549:34076
WITHDRAW
 93.181.192.0/19
ANNOUNCE
199.193.160.0/22
```

Communautés bien connues

- 0xFFFFF01 (alias NO_EXPORT : ne pas transmettre cette annonce en dehors de son AS),
- 0xFFFFF02 (NO_ADVERTISE, ne transmettre cette annonce à aucun autre routeur),
- 0xFFFF029A (BLACKHOLE, jetez-moi le trafic vers ce préfixe, RFC 7999).

Trouver la documentation sur les communautés

- Site Web de l'opérateur. Exemple
<http://www.us.ntt.net/support/policy/routing.cfm>. Notez le *European country origins* (là où la route a été apprise), qui peut permettre de faire du routage Schengen.
- Une liste assez complète en <https://onestep.net/communities/>
- whois.

Exemple communautés avec whois

```
% whois AS51706
...
aut-num:        AS51706
as-name:        FRANCE-IX-AS
...
remarks:        The following communities can be used by members:
...
remarks:        0:peer-as = Don't send route to this peer as
remarks:        51706:peer-as = Send route to this peer as
remarks:        0:51706 = Don't send route to any peer
remarks:        51706:51706 = Send route to all peers

remarks:        51706:64601 = Prefix received from a peer on RS1 Paris
remarks:        51706:64602 = Prefix received from a peer on RS2 Paris
remarks:        51706:64611 = Prefix received from a peer on RS1 Marseille
remarks:        51706:64612 = Prefix received from a peer on RS2 Marseille
...
```

Filtrage des annonces

Il est important de filtrer les annonces entrantes (ne pas se laisser envoyer n'importe quoi) et sortantes (ne pas envoyer n'importe quoi). RFC 7454
Juniper :

```
policy-statement no-small-and-big-prefixes {
    from {
        route-filter 0.0.0.0/0 prefix-length-range /25-/32 reject;
        route-filter 0.0.0.0/0 prefix-length-range /0-/7 reject;
    }
}
protocols {
    bgp {
        ...
        import no-small-and-big-prefixes;
    }
}
```

Un filtre utile, le nombre maximal de préfixes

```
group My-Nice-Peer {
    family inet {
        unicast {
            prefix-limit {
                maximum 100;
            }
        }
    }
}
```


BGP est politique

Applications de politiques locales

Cisco :

```
!--- Sets community 100:300 for routes matching access-list 101.
route-map Peer-R1 permit 10
  match ip address 101
  set community 100:300

!--- Sets local preference 130 for all routes
!--- matching community list 1.
route-map Peer-R3 permit 10
  match community 1
  set local-preference 130
```

Tout est transparent, tout (?) est public

- Du moment qu'on a un routeur BGP connecté à la DFZ (*Default-Free Zone*), on a l'information.
- Pas de routeur BGP dans la DFZ ? Pas grave, plusieurs services vous y permettent un accès.
- C'est ainsi qu'on a pu voir en octobre 2017 que la Corée du Nord avait un nouveau transitaire, un russe.
- Mais attention : pas de vue unique en BGP. Un routeur peut ne pas tout voir.

Looking glasses

- Un *looking glass* est un service donnant accès à un routeur en temps réel pour voir ce qu'il voit.
- Si ce routeur est sur la DFZ, vous voyez la DFZ.

Looking glass avec telnet

```
% telnet route-views.oregon-ix.net
route-views>show ip bgp 175.45.176.15
BGP routing table entry for 175.45.176.0/24, version 395583
 4826 174 701 4837 131279
 114.31.199.1 from 114.31.199.1 (114.31.199.1)
   Origin IGP, localpref 100, valid, external
   Community: 174:21000 174:22013 4826:5901 4826:6150 4826:59011
   rx pathid: 0, tx pathid: 0
...
```

Looking glass sur le Web

Chez Hurricane Electric <https://lg.he.net/>

Looking Glass

Welcome to Hurricane Electric's Network Looking Glass. The information provided by and the support of this service are on a best effort basis. These are snapshots of the network. We also operate a public route server accessible via telnet at route-server.he.net.

Show options

core1.bru1.he.net> show ip bgp routes detail 175.45.176.0/24							
Matching Routes	2						
Status Codes	A - Aggregate B - Best b - Not Install Best C - Confederation eBGP D - Damped E - eBGP H - History I - iBGP L - Local M - Multipath m - Not Suppressed F - Filtered s - Stale x - Best-External						
Status	Network	Next Hop	Metric	LocPrf	Weight	Path	
BI	175.45.176.0/24	216.66.64.178	1260	100	0	701, 4837, 131279	
I	175.45.176.0/24	216.66.64.178	1260	100	0	701, 4837, 131279	
Last Update		30d17h33m19s ago (1 path installed)					
Entry cached for another 55 seconds.							

core1.fmt1.he.net> show ip bgp routes detail 175.45.176.0/24							
Matching Routes	0						
Status Codes	A - Aggregate B - Best b - Not Install Best C - Confederation eBGP D - Damped E - eBGP H - History I - iBGP L - Local M - Multipath m - Not Suppressed F - Filtered s - Stale x - Best-External						
Status	Network	Next Hop	Metric	LocPrf	Weight	Path	
Last Update		Unknown					
Entry cached for another 21 seconds.							

RIPE stat

- <https://stat.ripe.net/> vous donne accès aux routeurs du RIS (réseau de sondes BGP),
- Jolie interface Web, très chargée en JavaScript compliqué.

>> You are here: [Home](#) > [Analyse](#) > [Statistics](#) > RIPEstat

2605:4500:2:245b::bad:dcaf

Search

permalink

At a Glance (5)

Routing (6/8)

DNS (3)

Anti Abuse (1)

Database (8)

Geographic (2)

Activity (2)

Suggestions (1)

+ MyView ?

Prefix Overview (2605:4500:2:245b::bad:dcaf)

✓ Announced


This prefix is part of 2605:4500::/32 announced by

AS46636
NATCOWEB - NatCoWeb Corp.

Resource	RIR	Country
2605:4500::/32	ARIN	US

Show IANA Registry Information


Showing results for 2605:4500::/32 as of 2018-01-07 00:00:00 UTC


 Given resource is not announced but result has been aligned to first-level less-specific (2605:4500::/32).

Geoloc (2605:4500:2:245b::bad:dcaf)




Geoloc details


 Data is based on MaxMind's GeoLite City data set and valid for the stated query time (see below)

Showing results for 2605:4500:2:245b::bad:dcaf as of 2018-01-06 00:00:00 UTC

[source data](#) [embed code](#) [permalink](#) [info](#)

Routing Status (2605:4500:2:245b::bad:dcaf)


 At 2018-01-07 00:00:00 UTC, 2605:4500::/32 was 100% visible (by 159 of 159 RIS full peers).



 First ever seen announced by AS46636, on 2011-07-02 08:00:00 UTC.

Originated by: AS46636

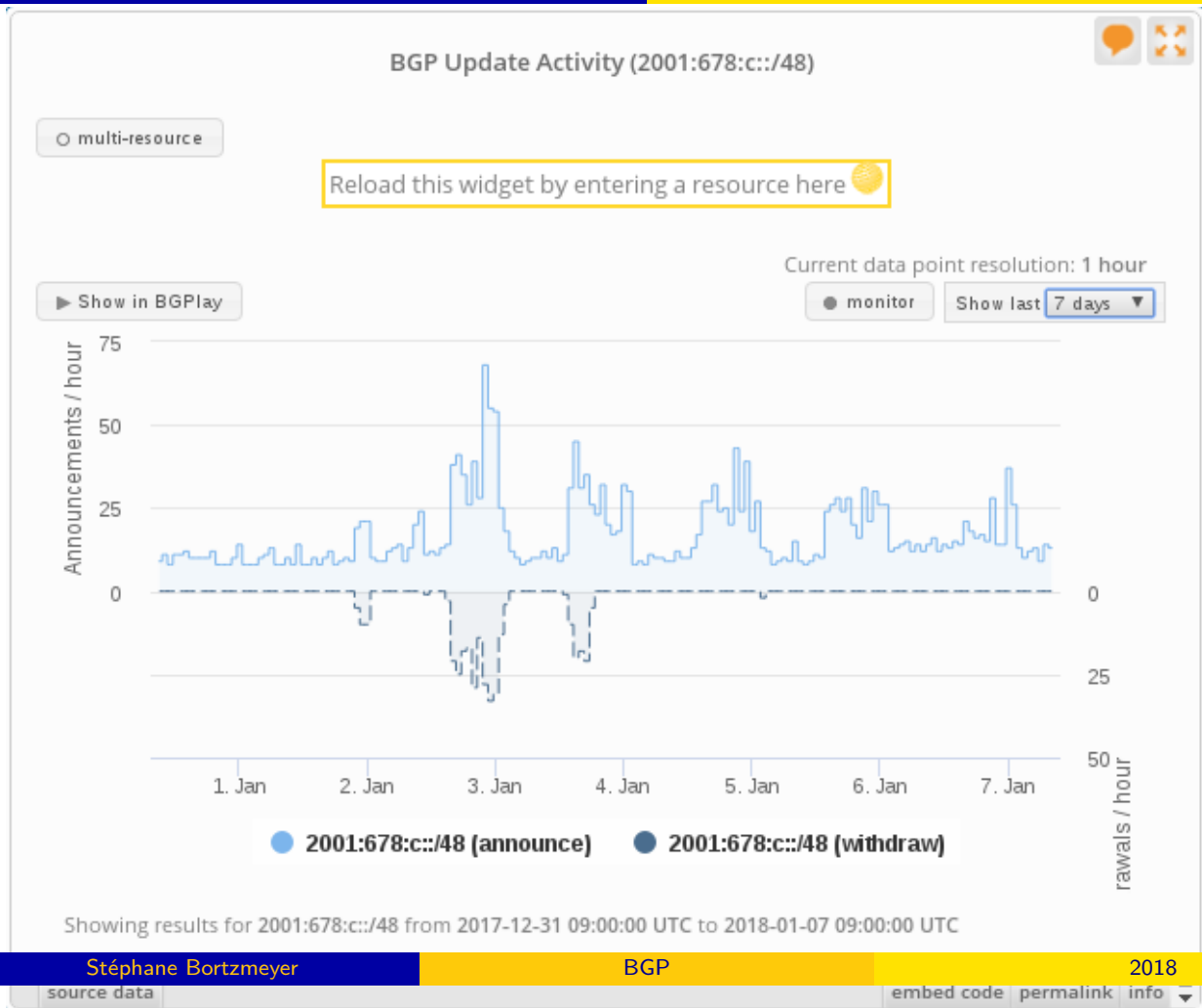
No less-specific covering prefixes.


[Advanced Settings](#)

Showing results for 2605:4500::/32 as of 2018-01-07 00:00:00 UTC


 IP address (2605:4500:2:245b::bad:dcaf) has been converted to its encompassing routed prefix (2605:4500::/32)


 Results exclude routes with very low visibility (less than 3 RIS full-feed peers seeing).
[source data](#)
[embed code](#) [permalink](#) [info](#)



RouteViews

<http://www.routeviews.org/>

- *Looking glass* telnet,
- Annonces BGP au format MRT, et tables BGP complètes, archivées (génial pour les chercheurs et les étudiants),
- Divers services, comme une passerelle DNS.

La passerelle DNS de RouteViews

Utilisation de “aspath.routeviews.org”

```
% bgproute 185.26.126.156
AS path: 2905 6939 29169
Route: 185.26.124.0/22
```

bgproute est juste un petit script qui appelle dig.

Utilisation du serveur whois de Cymru

Trouver le numéro d'AS à partir d'une adresse IP :

```
% whois -h whois.cymru.com 2001:4b98:dc2:45:216:3eff:fe4b:8c5b
AS      | IP                               | AS Name
29169   | 2001:4b98:dc2:45:216:3eff:fe4b:8c5b | GANDI-AS Registrar
```

Statistiques sur la DFZ avec RouteViews

On importe la table dans PostgreSQL et (au 9 janvier 2018) :

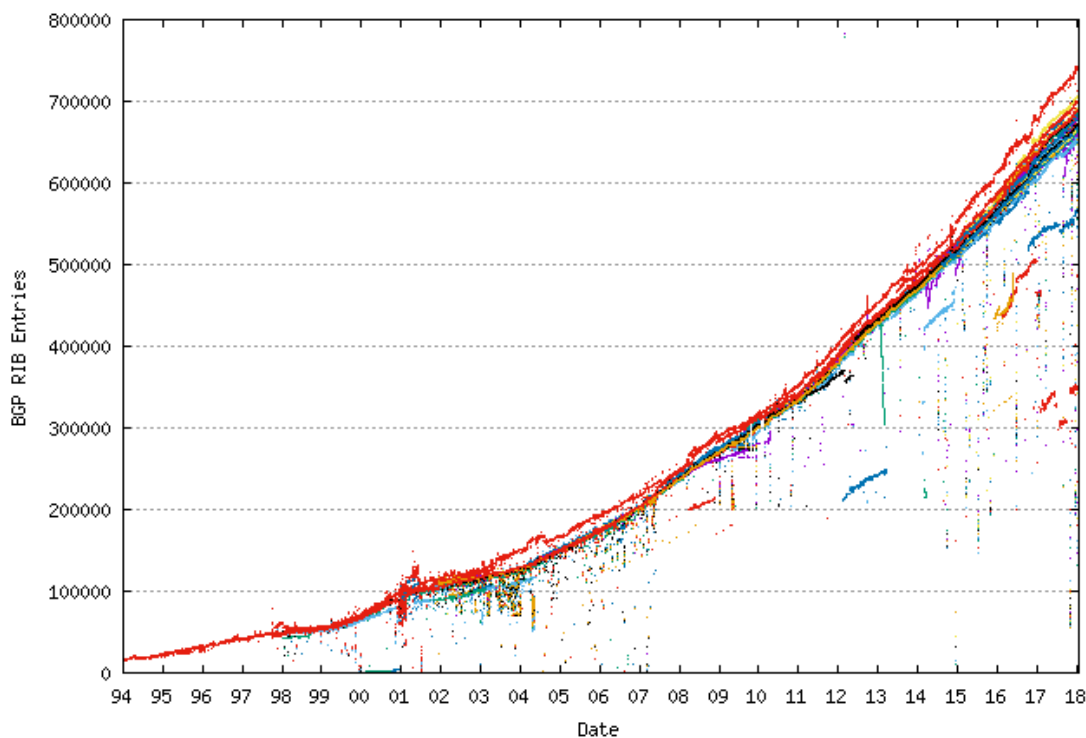
```
bgp=> SELECT count(*) FROM Prefixes;
794880
```

```
bgp=> SELECT count(*) FROM Prefixes WHERE family(pfx) = 4;
743916
```

```
bgp=> SELECT count(*) FROM Prefixes WHERE family(pfx) = 6;
50964
```

Plein de recherches possibles

Par exemples celles de Geoff Huston <https://bgp.potaroo.net/> :



Sécurité

- Par défaut, BGP croit tout ce qu'on lui raconte. « Pour parler à 8.8.8.8, passe par moi. »
- Ce n'est **pas** le résultat d'un oubli, c'est une conséquence de l'absence de hiérarchie, et de la complexité des relations.
- Mais cela permet des erreurs, et des attaques.

Les fuites

- (*Leaks*)
- Un routeur (ré-)annonce des routes qu'il n'aurait pas dû annoncer.
- Exemple : un client de deux transitaires annonce à un des transitaires les routes reçues de l'autre, attirant ainsi tout le trafic.
- Un grand classique de l'Internet. Une fuite massive arrive tous les deux ou trois ans.
- Dernier exemple : Google en août 2017, annonçant les préfixes de points d'échange.

Autre fuite : Telekom Malaysia (AS 4788)

- Juin 2015
- Annonce de 200 000 routes (le tiers de la DFZ)

```
TIME: 06/12/15 08:43:29
TYPE: BGP4MP/MESSAGE/Update
FROM: 208.51.134.246 AS3549
TO: 128.223.51.102 AS6447
ORIGIN: IGP
ASPATH: 3549 4788 3491 4651 9737 23969
NEXT_HOP: 208.51.134.246
COMMUNITY: 3549:4992 3549:7000 3549:7003 3549:7004 354
9:32344 4788:400 4788:410 4788:415
ANNOUNCE
  1.0.208.0/22
  1.0.212.0/23
  1.1.176.0/22
  ...
```

Détournement BGP

- Une annonce incorrecte est en général une erreur. « Il y a davantage d'incompétents que de malhonnêtes. »
- Mais cela peut aussi être une attaque.
- En général, on ne sait pas (sauf si on est Fox News ou BFM TV). 99 % des articles et reportages sont de la pure spéculation.

Détournement qui peut être une attaque

- Décembre 2017, 80 préfixes mais que des boîtes sensibles, annoncées par l'AS 39523 (DV-LINK, en Russie).
- Mars 2014, détournement des préfixes de mineurs Bitcoin, peut-être pour voler des bitcoins.

Le préfixe OVH 142.4.195.0/24 est utilisé par le *pool* de mineurs Bitcoin Hashfaster.

```

TIME: 03/23/14 18:32:38
TYPE: BGP4MP/MESSAGE/Update
FROM: 195.66.224.21 AS6939
TO: 195.66.225.222 AS6447
ASPATH: 6939 21548 34272 2093 2871 3721
NEXT_HOP: 195.66.224.21
ANNOUNCE
...
  142.4.195.0/24
  107.170.47.0/24
  54.194.173.0/24
...

```

Détecter les problèmes

Tout est public, donc tout peut être supervisé

Ici, avec le système d'alarme BGPmon :

Alerts Details



On Wednesday April 2nd 2014 at 18:51 UTC we detected a Origin AS Change event for your prefix (192.93.0.0/24 NIC-FR-DN). The detected prefix: 192.93.0.0/24, was announced by AS4761 (INDOSAT Internet Network Provider)

Alert description: Origin AS Change
 Detected Prefix: 192.93.0.0/24
 Detected Origin AS: 4761
 Expected Origin AS: 2483

This alert was detected by 1 unique probes in 1 unique countries

Thailand: 1 Peers



BGPstream et ses alarmes sur Twitter

Possible BGP hijack

Beginning at 2018-01-07 10:00:59 UTC, we detected a possible BGP hijack.

Prefix 103.230.226.0/24, is normally announced by AS132566 Skynet Broadband Plus Solution.

But beginning at 2018-01-07 10:00:59, the same prefix (103.230.226.0/24) was also announced by ASN 135830.

This was detected by 36 BGPMon peers.

Expected

Start time: 2018-01-07 10:00:59 UTC

Expected prefix: 103.230.226.0/24

Expected ASN: 132566 (Skynet Broadband Plus Solution)

Event Details

Detected advertisement: 103.230.226.0/24

Detected Origin ASN 135830 ()

Detected AS Path 61102 8551 9583 55644 59165 135830

Detected by number of BGPMon peers: 36

Réagir aux problèmes

- Les incidents sont relativement fréquents,
- Mais la plupart ne durent pas : les techniciens agissent,
- La vraie sécurité de BGP, ce sont eux.

 **bgpstream** @bgpstream · 7 janv. ▼
 BGP,HJ,hijacked prefix AS16276 87.98.182.0/24, OVH SAS,-,By AS44901
 BelCloud Hosting Corporation, bgpstream.com/event/122779
 🌐 À l'origine en anglais
 2 4 6

 **Andree Toonk** @atoonk · 7 janv. ▼
 CC'ing @olesovhcom @as16276 #bgp
 1 1

 **OVH Network** Suivre ▼
 @as16276

En réponse à @atoonk @bgpstream @olesovhcom

Contact done, thanks for notifying us !

🌐 À l'origine en anglais

08:44 - 7 janv. 2018

1 Retweet 3 J'aime

1 3

Empêcher les problèmes, les IRR

- Filtrer les routes selon le contenu des IRR (*Internet Routing Registry*),
- Cela suppose que les opérateurs maintiennent le contenu des IRR à jour, ce qui n'est pas le cas,
- Les IRR peuvent utiliser un langage plus perfectionné, RPSL (RFC 2622),
- Il existe de nombreux IRR.

Exemple d'une entrée IRR simple

```
route6:      2001:4b98::/32
descr:      GANDI is an ICANN accredited registrar
descr:      GANDI is a virtual server provider
descr:      for more information:
descr:      Web:  http://www.gandi.net
origin:     AS29169
mnt-by:     GANDI-NOC
created:    2009-06-15T16:04:25Z
last-modified: 2009-06-15T16:04:25Z
source:     RIPE
```

Les deux lignes importantes sont route6 et origin.

Empêcher les problèmes, avec la RPKI

L'idée de base est de partir d'assertions, arborescentes, et cryptographiquement signées.

- ① IANA : « RIPE-NCC gère 192.0.0.0/8 »,
- ② RIPE-NCC : « le FAI Example est titulaire de 192.0.2.0/24 »,
- ③ Example : « l'AS 64641 est autorisé à être l'origine d'une annonce de 192.0.2.0/24 » (cette dernière assertion étant un ROA, *Route Origin Authorizations*).

On met ces assertions dans la RPKI

- RPKI = *Resource Public Key Infrastructure*, RFC 6481,
- Parmi elles, les ROA (RFC 6482) indiquent l'AS d'origine utilisé,
- La RPKI est ensuite distribuée chez chaque opérateur,
- On peut donc prouver cryptographiquement les autorisations.

RPKI et ROA en pratique

- En dehors de l'Europe, peu de gens signent,
- Peu de gens valident.

L'infrastructure de l'Internet

- BGP fonctionne depuis des années,
- Il s'est montré très robuste (l'Internet n'est pas un environnement facile!),
- Il n'est pas parfait mais attention si vous croyez faire mieux : relisez bien le cahier des charges d'abord !

Bonnes lectures

- La conférence BGP de Sarah Nataf <http://www.iletaitunefoisinternet.fr/bgp-sarah-nataf/>,
- Le rapport sur la résilience de l'Internet en France, avec notamment plein de chiffres BGP <https://urlz.fr/6mYx>,
- Le guide des bonnes pratiques de l'ANSSI <https://urlz.fr/6mYu>,
- La conférence *BGP and the rule of custom* au CCC https://media.ccc.de/v/34c3-9072-bgp_and_the_rule_of_custom
- <https://bgpmon.net/>, <https://dyn.com/blog/> et <https://labs.ripe.net/> pour les nouvelles et les analyses. <https://twitter.com/bgpstream> pour suivre en temps réel. Listes de diffusion FRnog et Nanog.