Derrière la scène : comment et grâce à qui l'Internet fonctionne t-il ?

Stéphane Bortzmeyer
<stephane+toulibre@bortzmeyer.org>

26 novembre 2011

Exposé libre

Dequoi ki parle ?

1. Du fonctionnement de l'Internet
2. Pas juste des protocoles TCP/IP. Je suppose que tout le monde ici sait faire un ping, configurer Apache et même utiliser dig.
3. Mais de l'Internet comme objet global et multi-acteurs. Un tel réseau de réseaux est radicalement différent d'un réseau local !

Que se passe-t-il derrière ma box ?

```
% traceroute6 www.ietf.org
```

1. freebox (2a01:e35:8bd9:8bb0::1) 6.543 ms
2. th2-crs16-1.intf.routers.proxad.net (2a01:e00:2:d::1) 25.218 ms
3. londres-6k-1-po101.intf.routers.proxad.net (2a01:e00:1:a::2) 34.033 ms
4. 20gigabitethernet1-3.core1.ams1.ipv6.he.net (2001:7f8:1::a500:6939:1) 46.654 ms
5. 10gigabitethernet7-4.core1.nyc4.he.net (2001:470:0:3e::1) 109.349 ms
6. cr1.n54ny.ip.att.net (::ffff:12.122.81.106) 188.216 ms
7. cr81.sj2ca.ip.att.net (::ffff:12.122.1.118) 190.344 ms
8. mail.ietf.org (2001:1890:123a::1:1e) 183.194 ms

Qu'est-ce qu'on a vu ?

- Un FAI (Free/Proxad), deux opérateurs (Hurricane Electric et AT&T),
- Un câble transatlantique
- Des codes géographiques (variés)
- Des adresses IP et des noms de domaine
- **Une multiplicité d'acteurs** La principale caractéristique de l'Internet, ce n'est pas d'utiliser TCP/IP ou d'être un réseau à commutation de paquets. C'est d'être **multi-acteurs**.

Système autonome

L'Internet est composé de dizaines de milliers de systèmes autonomes ou AS. L'AS est une entité administrative, un acteur de l'Internet. M. Michu n'a pas d'AS.

Un AS a un numéro (comme 42 ou 2486). À l'intérieur d'un AS, il y a unité de politique de routage. Entre AS, il n'y a rien de commun, et pas de chef au dessus.

Qu'est-ce qu'un Tier 1 ?

- La plupart des AS n'ont pas d'autres AS derrière (Proxad dans l'exemple ci-dessus). Ce sont des feuilles du graphe. Ils achètent du **transit** à des fournisseurs et se connectent (en général gratuitement) en **peering**.
- Une minorité d'AS fournissent du transit aux autres. Ce sont les AS de transit (HE et AT&T dans l'exemple ci-dessus).
- Une minorité de la minorité n'achète de transit à personne. Ce sont les **Tier 1** comme Level 3.
- Évidemment, en vrai, c'est encore plus compliqué.

Qui dirige l'Internet ?

Bon, il y a plein d'acteurs, mais qui les dirige ?

Cf. exposé de Mathieu Weill à JRES hier...

- Qui décide de déployer (ou pas) IPv6 ?
- Qui décide de déployer (ou pas) davantage de cryptographie ?
- Qui décide qu'on pourra voir du porno (ou pas) en .XXX ?
- Qui décide de <insérez votre demande personnelle> ?

Zoo

- J'ai parlé des opérateurs mais il y a aussi...
- Les fournisseurs de matériel et logiciel de routage (Juniper, Huawei, Cisco)
- Les éditeurs de logiciel (BitTorrent, Firefox, Ubuntu)
- Les fournisseurs de service (Google, Facebook)
- Les États (« L'Internet ne peut pas demeurer sans règles »)
- Les organismes de normalisation (IEEE, IETF)
- Les organismes de régulation (ICANN, RIR, demain CSA ?)
- Et même vous !

Le plus difficile sur l'Internet, pour l'administrateur réseaux débutant, c'est de s'y retrouver dans ce zoo. Le Minitel avec la DGT était plus simple !

Pourquoi Comcast et Level 3 s'accusent-ils réciproquement de « patate chaude » ?

Les acteurs cités plus haut sont souvent concurrents : ils ne travaillent pas pour le bien commun.

Par exemple, le routage sur l'Internet s'est toujours fait par la « patate chaude » : on fait sortir le paquet de son réseau le plus vite possible. Lors d'une communication symétrique, patate chaude ou froide donnent le même résultat.

Mais l'Internet d'aujourd'hui est souvent asymétrique : pousseurs de conneries vidéo d'un côté, globes oculaires passifs de l'autre. Les deux algorithmes ne sont plus équivalents. D'où des disputes comme lorsque Comcast demandait à Netflix/Level 3 d'utiliser la patate froide.

C'est le monde merveilleux de la **neutralité du réseau**.

À quoi sert un point d'échange ?

```
% traceroute6 a.dns.gandi.net
```

1. vl387-te2-6-paris1-rtr-021.noc.renater.fr (2001:660:300c:1002:0:131:0:2200) 2.175 ms
2. te0-0-0-3-paris1-rtr-001.noc.renater.fr (2001:660:7903:a:1::1) 4.924 ms
3. gandi.sfinx.tm.fr (2001:7f8:4e:2::118) 81.681 ms
4. po82-ip6-vd3.core3-d.paris.gandi.net (2001:4b98:82::42) 11.297 ms
5. a.dns.gandi.net (2604:3400:a::2) 10.556 ms

Entre `te0-0-0-3-paris1-rtr-001.noc.renater.fr` et `gandi.sfinx.tm.fr`, on traverse un point d'échange, le Sfinx.

Les points d'échange

- Simplifient l'interconnexion des opérateurs
- Permettent de garder en local le trafic local

- Fondamentalement, c'est juste un commutateur Ethernet, où les FAI et opérateurs viennent se brancher,
- Bon, j'exagère, c'est quand même quelque chose à faire soigneusement,
- Mais fondamentalement, ce n'est pas difficile à faire, il n'y a pas de raison que ça prenne des années,
- Le plus gros en France est désormais le France-IX.

BGP fait-il mal ?

1. J'envoie un paquet depuis ma connexion Free, à une université au Japon. Comment Free sait à qui la transmettre ? C'est le rôle de BGP (*Border Gateway Protocol*).
2. BGP est le protocole standard d'échanges de route sur l'Internet. Deux AS s'échangent leurs informations par BGP.
3. Chaque pair BGP annonce « Je sais joindre 2001:1890::/32 » « Moi, je sais joindre 2001:503:a83e::/48 ». Ensuite, il le propage à ses voisins. En quelques minutes, tout l'Internet est au courant. Presque 400 000 routes aujourd'hui.
4. Même si on n'a pas de routeur BGP (il faut un AS, et des pairs), on peut voir cette activité par les *looking glasses*. `http://www.traceroute.org/#Looking%20Glass`

Pourquoi un maladroit au Pakistan peut-il bloquer YouTube sur toute la planète ?

1. Comment on sait que celui qui annonce une route vers 2605:4500::/32 dit la vérité ? **On ne sait pas**.
2. N'importe qui peut le faire. C'est à l'origine de quelques bavures fameuses. Mais tout est vite corrigé, car les acteurs de l'Internet ne suivent pas aveuglément des procédures : ils ont un cerveau.
3. Sécuriser, d'accord, mais comment ? L'absence de Centre n'aide pas.
4. Déploiement en cours de la RPKI, usine à gaz de certification des adresses IP, et des ROA, annonces BGP dont l'origine (mais pas le chemin) est authentifiée.

Résilience

L'Internet est-il très fragile ou très robuste ?

Les deux. Très fragile localement et très robuste globalement.

La dernière grande panne : il y a trois semaines

Le 7 novembre, une annonce BGP plante un grand nombre des routeurs Juniper. Level 3, Tata et Neo Telecoms sont touchés. Trente minutes de perturbation. Fin du monde ou anecdote ?

L'important, c'est plutôt ce qu'on fait pour améliorer cette résilience : diversité, compétence, vigilance, coopération

Et puis pensons à ceux qui vont réparer nos fibres optiques dehors par tous les temps... `http://blog.level3.com/2011/08/04/the-10-most-bizarre-and-annoying-causes-of-fiber-cuts/`

Il n'y a pas que le routage

L'Internet dépend de bien d'autres choses.

Je ne vais pas parler de toutes ces choses, seulement de celles que je connais bien.

Les RIR

Les RIR (*Regional Internet Registries*) sont les cinq organisations qui attribuent les adresses IP dans le monde. C'est au sein des RIR que se décide si vous pouvez avoir un bloc d'adresses PI (*Provider Independent*, c'est-à-dire portables), et selon quels critères.

Les RIR sont un acteur essentiel de l'Internet, c'est pour cela qu'ils ne sont jamais mentionnés dans les médias.

Le RIR de l'Europe (Europe au sens d'Alexandre le Grand, cela inclut la Turquie, le Liban et l'Iran) est le RIPE-NCC, installé à Amsterdam.

Les rencontres des RIR rassemblent des centaines d'acteurs de la communauté du routage et sont une occasion privilégiée d'échanger. Prochaine réunion RIPE à Ljubljana en avril 2012.

Exemple de l'information stockée chez les RIR

```
inet6num:       2001:4b98::/32
netname:        FR-GANDI-20050224
descr:          GANDI
country:        FR
...
tech-c:         GNO4-RIPE
status:         ALLOCATED-BY-RIR
mnt-by:         RIPE-NCC-HM-MNT
mnt-lower:      GANDI-NOC

role:           Gandi Network Operations
address:        15 Place de la Nation
address:        75011 Paris
address:        France
phone:          +33 1 70 39 37 55
nic-hdl:        GNO4-RIPE
...
```

Le DNS

- DNS *Domain Name System*
- Une infrastructure devenue indispensable, car elle fournit de la stabilité aux identificateurs (les adresses IP changent).
- Contrairement au graphe de BGP, le DNS est un arbre, avec une racine, gérée par le gouvernement des États-Unis (qui vient de lancer un appel d'offres pour son sous-traitant)...
- ...et tout un écosystème de registres (comme l'AFNIC, DENIC ou ICM Registry), et de bureaux d'enregistrements (comme Gandi, OVH ou BookMyName). Sans compter les gérants des serveurs DNS. Le DNS est fortement décentralisé.
- L'OARC *DNS Operations, Analysis and Research Center* rassemble les acteurs techniques pour études et échanges d'expériences.
- Le transparent suivant se concentre sur les opérateurs techniques.

Les acteurs du DNS

- La racine du DNS est servie par les serveurs de onze organisations (question piège : qui les a choisis ?). Chaque organisation déploie parfois sur des dizaines de sites (grâce à l'*anycast*). Un des points les plus solides de l'Internet (il a survécu à deux grandes attaques, en 2002 et 2007), et les mieux connus.
- Les TLD (domaines de tête comme `.fr` ou `.org`) sont servis par les serveurs du registre (AFNIC pour `.fr`, Afilias pour `.org`, etc). Qualité très variable.
- Des tas d'organisations gèrent des serveurs faisant autorité (servent les données) ou des résolveurs (récupèrent les donnéees). Typiquement, le résolveur que vous utilisez est fourni par votre FAI ou par votre organisation (pour les gros réseaux locaux, par exemple d'une Université).

Exemple d'un TLD

```
% dig +nssearch tn.
... 2011111844 ... from server rip.psg.com in 581 ms.
... 2011111844 ... from server sunic.sunet.se in 347 ms.
... 2011111844 ... from server ns.ati.tn in 177 ms.
... 2011111844 ... from server ns2.nic.fr in 31 ms.
... 2011111844 ... from server ns2.ati.tn in 89 ms.
```


Normalisation

Le bon fonctionnement de l'Internet dépend de **normes** techniques communes. Typiquement, c'est l'IEEE qui fixe les normes pour les couches basses, l'IETF pour les couches 3 à 6, et une partie de la 7, d'autres organismes (W3C, OASIS) pour le reste de la couche 7.

L'IETF produit les fameux RFC, les textes sacrés de l'Internet. Contrairement à tant d'autres normes techniques, ils sont très lisibles. Lecture recommandées aux étudiants et ingénieurs.

Les rencontres de l'IETF rassemblent des centaines d'acteurs de l'Internet et sont une occasion privilégiée d'échanger. Prochaine réunion à Paris en mars 2012. Mais le gros du travail se fait en ligne et ne nécessite pas de venir.

L'IETF n'est pas la police de l'Internet, elle n'a aucun pouvoir (autrement, on utiliserait tous IPsec depuis longtemps).

Conclusion

L'un des gros avantages de l'Internet est que son fonctionnement est lui-même ouvert :

- Grâce à traceroute, on peut connaître la politique de routage de son opérateur (imaginez si France Telecom avait conçu l'Internet...).
- Les RFC sont disponibles facilement et gratuitement (contrairement aux normes ISO ou AFNOR) et sont très lisibles.
- Une bonne partie des discussions et des faits sont publics (lisez les compte-rendus des réunions NANOG, RIPE, OARC...).
- Bref, n'hésitez pas à aller voir « derrière la *box* ».