

Histoire et fonctionnement de l'Internet, le développement des infrastructures

Par **Stéphane BORTZMEYER**
AFNIC

Considérons un utilisateur, que nous appellerons M. Toutlemonde, et qui, depuis chez lui à Bobigny (Île-de-France), veut regarder le site Web du gouvernement burkinabé, <https://www.gouvernement.gov.bf/>. Supposons que M. Toutlemonde soit client du fournisseur d'accès Internet Free. Entre Bobigny et Ouagadougou, il y a 5 140 kilomètres et, surtout, beaucoup d'acteurs différents. Free ne connaît pas le gouvernement du Burkina Faso, et ce dernier ne connaît pas Free. Pourtant, la communication fonctionne et M. Toutlemonde voit la page Web sur son écran, moins d'une seconde plus tard. Comment est-ce que cela a été possible, alors que personne n'était en charge de coordonner la connexion de Free avec celle du gouvernement burkinabé ?

L'émergence de l'Internet

C'est la propriété la plus importante de l'Internet, l'**émergence**. Chaque acteur s'occupe de son réseau, achète ses machines, embauche des techniciens, configure le tout et se relie à quelques autres acteurs proches. La communication passera ensuite d'acteur en acteur, jusqu'à atteindre la destination. Personne n'avait de responsabilité globale et pourtant ça marche.

Car l'Internet est un **réseau de réseaux**. Si personne n'est responsable de la liaison entre Bobigny et Ouagadougou, en revanche chacun des réseaux qui composent l'Internet est géré par une association, une administration ou une entreprise, qui se chargent de tout faire marcher. Sur l'Internet, ces réseaux se nomment Systèmes autonomes⁽¹⁾. Free est le système autonome n°12322 et le site Web du gouvernement burkinabé est hébergé par l'ANPTIC, ou encore système autonome n°327871. Il y a environ 69 000 systèmes autonomes actifs et l'Internet, ce sont eux, ou plutôt leur interconnexion.

Si la notion de système autonome est un peu abstraite, il faut noter qu'un AS (acronyme anglais de Système autonome) n'est pas une entité désincarnée, planant loin au-dessus des réalités terrestres. Il a une nationalité, il obéit à des lois locales, il a un compte en banque, etc.

Les infrastructures matérielles

D'abord, l'Internet, c'est du matériel. Loin du discours *marketing* qui vend du rêve avec de jolis « *clouds* » éthérés, l'Internet repose sur des infrastructures matérielles. L'essentiel du trafic inter-continental passe par des câbles sous-marins, dont la pose et surtout l'entretien coûtent cher. Peu de sociétés ou d'États peuvent se permettre de financer seuls un de ces câbles, qui sont souvent gérés par des consortiums, ou par les plus grosses entreprises. Les câbles ne sont pas une infrastructure publique, contrairement aux routes. Le caractère critique de ces câbles est bien visible à chaque fois que l'un d'eux se rompt, comme le câble Sea-Me-We 3 en mars 2020, au moment où le confinement lié à la pandémie de Covid-19 rendait chaque élément de l'infrastructure d'Internet particulièrement crucial.

(1) En anglais, *Autonomous Systems*, ou AS.

Ensuite, il y a les routeurs. Ce sont des ordinateurs spécialisés dans le passage des messages d'un réseau à l'autre. Un routeur a plusieurs interfaces avec ces différents réseaux et doit ensuite transmettre les données très vite. Si la modeste « *box* » de votre domicile est bien un routeur, elle est nettement moins performante, moins chère et moins consommatrice de courant que les grands équipements des opérateurs Internet. On notera que le marché de ces routeurs « de cœur » est très concentré, trois entreprises, Cisco, Juniper et Huawei, s'en partagent l'essentiel. Entrer sur ce marché est une tâche difficile, du fait du concentré de haute technologie dans ces engins.

Et puis il faut parler des serveurs, les ordinateurs qui attendent, connectés à l'Internet, des requêtes des clients pour obtenir un message, télécharger un film, ou récupérer un document. Contrairement aux câbles et aux routeurs, les serveurs ne sont pas un élément indispensable de l'infrastructure de l'Internet. Il existe beaucoup d'applications, dites pair-à-pair, qui fonctionnent uniquement par communication directe entre les machines des utilisateurs finaux. Mais, en pratique, une grande partie des utilisations de l'Internet dépend de ces serveurs⁽²⁾.

Enfin, ces routeurs et ces serveurs sont hébergés dans des centres de données, des bâtiments spécialisés fournissant électricité, climatisation et sécurité physique. Là encore, il n'y a pas d'infrastructure publique : tous sont gérés par des sociétés privées.

Les logiciels

L'infrastructure, ce n'est pas seulement du matériel tangible. Elle comprend également beaucoup de logiciels. Prenons un routeur, par exemple, la machine qui a plusieurs interfaces, et va faire passer les données d'une interface à l'autre. C'est du matériel, un processeur, une mémoire, des cartes Ethernet... Et c'est aussi du logiciel. Le routeur doit faire circuler les paquets en suivant les règles générales de la transmission de paquets IP (*Internet Protocol*), et la politique décidée localement. Ce logiciel, sur les routeurs haut de gamme, se trouve dans des circuits spécialisés, ce n'est pas un processeur généraliste qui l'exécute, mais le principe reste le même. Et le routeur doit aussi pouvoir parler avec les autres routeurs, pour les algorithmes de routage, et avec l'administrateur du système qui le configure.

Au niveau de l'Internet global, le protocole de routage entre les AS est nommé BGP, pour *Border Gateway Protocol*. C'est le même pour tout l'Internet, puisqu'il doit pouvoir fonctionner entre AS différents, gérés par des organisations différentes. Au contraire, à l'intérieur d'un AS, les administrateurs des réseaux sont libres de se décider pour le protocole de leur choix, même non standard. Le principe de base du BGP est que chaque AS annonce à ses voisins les routes qu'il sait joindre et qu'il veut annoncer. Le BGP, c'est du routage politique : un AS n'annonce que ce qu'il veut. Ainsi, les gérants d'un AS peuvent, dans une certaine mesure, décider de mettre en œuvre des critères comme « je veux échanger avec tous les opérateurs de mon pays (s'ils sont d'accord...) » ou, au contraire, « pas question d'échanger avec Untel ». On parle, par exemple, de « routage Schengen » quand on configure son routage pour que le trafic entre deux pays de l'espace Schengen ou de l'Union européenne reste dans cet espace ou dans l'Union. Le BGP n'est pas toujours facile à maîtriser, et on est de toute façon dépendant des décisions des autres : si je veux échanger du trafic avec Orange mais qu'Orange ne le veut pas, je ne peux pas l'y forcer. C'est ce qui fait qu'en pratique, les souhaits politiques ne sont pas toujours parfaitement mis en œuvre.

Au-dessus du routage, il y a évidemment beaucoup d'autres logiciels. Les applications que vous utilisez quotidiennement sur votre *ordiphone* ou bien sur votre ordinateur en font partie. Pour beaucoup d'utilisateurs, elles sont l'Internet, et il est facile d'oublier qu'il y a beaucoup d'autres

(2) Ce qui est regrettable, car cela aggrave la concentration et permet aux gérants de ces serveurs davantage de contrôle sur les utilisations.

logiciels dans l'Internet, par exemple ceux des innombrables serveurs⁽³⁾ qui assurent des fonctions indispensables.

Les protocoles

J'ai utilisé le terme de protocole. Que désigne-t-il ? Un ensemble de règles que doivent suivre les machines qui communiquent. Par exemple, pour que deux routeurs puissent échanger des routes, et ainsi faire en sorte que l'Internet soit connecté, ils doivent se transmettre des messages formatés selon des directives précises, et dont la signification est connue des deux routeurs. Le protocole peut être spécifié informellement, mais il est préférable qu'il soit normalisé, c'est-à-dire décrit formellement dans un document écrit, validé et publié par une organisation de normalisation (en anglais, SDO, pour *Standard Development Organisation*).

Il existe de nombreuses organisations de normalisation, nationales ou internationales. Compte-tenu de l'enjeu stratégique que représente l'Internet, il n'est pas étonnant que, dans ces organisations, les débats soient parfois vifs. Pensons, par exemple, aux discussions sur la nouvelle version du protocole de sécurité TLS (*Transport Layer Security*), qui utilise la cryptographie pour assurer la confidentialité des communications. Cette version 1.3 corrigeait certaines faiblesses des versions précédentes, mais elle avait suscité la colère de certains qui exploitaient ces faiblesses pour analyser les données transmises, ce que normalement TLS devait empêcher.

Parmi les organisations de normalisation, on peut notamment citer l'IETF (*Internet Engineering Task Force*), qui normalise les protocoles BGP et TLS, déjà mentionnés, et de nombreux autres. L'IETF est très ouverte, il n'y a pas de barrière explicite à l'entrée, à part la compétence technique, et produit ses normes dans des documents nommés RFC⁽⁴⁾, qui forment les « textes sacrés » de l'Internet.

On l'a vu avec l'exemple de TLS, un protocole n'est pas neutre. Bien sûr, de nombreuses personnes, très différentes, utilisent ces protocoles pour des buts très variés, néfastes ou positifs. Mais les protocoles facilitent ou, au contraire, rendent plus difficiles, certains usages, et c'est en cela que la spécification des protocoles a des conséquences politiques. Ainsi, TLS 1.3 fait que la surveillance des communications est plus difficile, ce que certains verront comme un avantage et d'autres comme un inconvénient.

Les relations

Jusqu'à présent, j'ai surtout parlé de technique, mais l'Internet n'est pas géré selon des critères de pure technique. Les entreprises à but lucratif, par exemple, vont essayer de gagner de l'argent. Les États vont essayer de faire respecter leurs choix politiques. Les relations entre les nombreux acteurs de l'Internet sont complexes, et ne peuvent pas se comprendre uniquement en lisant les RFC. Prenons l'exemple de l'interconnexion entre opérateurs Internet. Qu'est-ce qui fait que deux opérateurs décident de se connecter, ou non ?

On distingue en général deux types de connexion entre opérateurs :

1) L'« appairage » (*peering*, en anglais) concerne en général des opérateurs de taille comparable (des « pairs »), et se fait souvent de manière informelle, sans contrepartie monétaire. Par contre, il ne donne accès, dans la plupart des cas, qu'au réseau du pair.

(3) Une légende courante aujourd'hui est celle du « serverless », selon laquelle on pourrait bâtir un service sans serveurs. C'est évidemment faux, à part pour le pair-à-pair pur, il y a toujours des serveurs, même s'ils sont « cachés », par exemple chez un sous-traitant comme AWS.

(4) Je ne donnerai pas la signification du sigle, désormais dépassée.

2) Le transit se fait entre deux opérateurs très différents, un client (en général payant) et un fournisseur. Il fait l'objet d'un contrat formel. Et, cette fois, il donne accès à tout l'Internet. Aussi bien M. Toutlemonde, quand il s'abonne à SFR, que la société Free, quand elle fait appel à Cogent, sont clients d'un transitaire, d'un fournisseur de transit (SFR dans le premier cas, Cogent dans le second).

C'est la combinaison de l'appairage et du transit qui forme l'Internet, permettant à M. Toutlemonde d'accéder à tout l'Internet, pas seulement au réseau de son FAI (fournisseur d'accès à l'Internet). Les petits achètent du transit aux gros, et font de l'appairage entre eux, et les gros, les « Tier 1 », ceux qui n'ont besoin d'aucun transitaire, s'appairent entre eux.

Aussi bien l'appairage que le transit nécessitent une connexion physique entre les deux acteurs. Elle peut se faire par un lien direct, ou bien par un point d'échange. L'idée du point d'échange est de simplifier l'interconnexion : au lieu d'avoir à faire un lien vers tous ses pairs, on fait seulement un lien avec le point d'échange, et l'établissement de l'appairage ne nécessite plus qu'une configuration, pouvant se faire sans intervention physique. Avoir des points d'échange locaux est crucial, pour que le trafic local reste, autant que possible, local.

Les services

Jusqu'à présent, j'ai parlé de l'Internet en tant que réseau. Le vécu de l'utilisateur est évidemment plus complexe ; quand il dit « je l'ai lu sur l'Internet », il ne pense certainement pas aux fibres optiques, aux routeurs, et au BGP. Il faut donc également dire un mot des services d'autant plus qu'ils sont souvent davantage centralisés que l'infrastructure, et plus souvent gérés par une entreprise étrangère.

Il y a deux sortes de services, les services d'infrastructure, et les services visibles. Les services d'infrastructure sont ceux qui, quoique ne faisant pas partie de l'Internet à proprement parler, sont indispensables ou très importants pour l'utilisation de l'Internet. Les plus connus sont :

- le DNS (*Domain Name System*) est le mécanisme qui est derrière les noms de domaine comme www.annales.org. Quasiment toute activité Internet va nécessiter des noms de domaine et le DNS est une infrastructure critique. Il est géré de manière décentralisée, par exemple le domaine de premier niveau national .fr est géré en France.
- les CDN (*Content Delivery Network*) sont des ensembles de machines servant du contenu (Web, par exemple, ou vidéo) et disposées un peu partout dans le monde pour réduire le temps d'accès. Ils sont gérés par les opérateurs de contenu eux-mêmes, ou bien par des entreprises spécialisées, toutes étatsunienne ou chinoises.

Et, il y a enfin les services visibles, ceux que M. Toutlemonde connaît, les sites Web, par exemple. Ils ne sont pas indispensables au fonctionnement de l'Internet⁽⁵⁾. Mais c'est pour eux que M. Toutlemonde accède au réseau. Ils sont à la fois très décentralisés (créer un site Web, ou un service d'hébergement, est assez facile et bon marché et beaucoup de gens le font) et, en pratique, très centralisés car un petit nombre d'acteurs, presque tous états-unien⁽⁶⁾, détiennent l'essentiel du marché.

(5) Il est à noter que, pendant le confinement imposé suite à l'épidémie de Covid-19, confinement qui a mené à un usage accru de l'Internet, ce sont toujours ces services qui, mal ou trop grossièrement conçus, ont connu des défaillances, et non pas l'Internet lui-même.

(6) Cette affirmation concerne la France. La situation est très différente en Chine, ou même en Russie.