

L'Internet.

*Extrait de "Systèmes d'information, obstacles et succès - La pensée aux prises avec l'information".
Laurent Bloch (Vuibert, Paris, 2005), pages 155 à 161.*

Un réseau ouvert

S'il est un succès incontestable dans l'histoire des développements informatiques, c'est bien l'Internet, et il est donc intéressant de se pencher sur la façon dont ont été créés les logiciels d'ampleur considérable qui en assurent le fonctionnement.

Pour les moins informés de ses utilisateurs, l'Internet se confond avec le WWW et éventuellement avec le courrier électronique, qui n'en sont que deux usages parmi d'autres. Le WWW est un ensemble de méthodes destinées à permettre la publication sur l'Internet de documents indépendants de la nature du serveur et reliés entre eux par des liens, les URL (Universal Resource Locator). Les URL (ces formules un peu étranges de la forme <http://www.laurent-bloch.org/> qui apparaissent dans la barre d'adresses de votre navigateur) constituent un moyen très simple de désigner de façon précise et dépourvue d'ambiguïté un document quelconque sur le réseau.

L'Internet est un réseau de réseaux, un gigantesque système d'interconnexion que l'on peut comparer au système téléphonique international, même si ses principes de fonctionnement sont différents. C'est un système public: il suffit de s'abonner pour une somme modique, voir gratuitement, auprès d'un fournisseur d'accès (FAI) local pour avoir accès, le plus souvent sans paiement de redevances supplémentaires, à des serveurs dans le monde entier, ce qui n'est pas du tout le cas avec le système téléphonique international. De ce fait, l'Internet permet de publier toutes sortes de textes, d'images et d'autres informations qui seront accessibles à un public mondial, pour un coût minime. Toutes sortes d'informations dont la consultation demandait il y a dix ans de nombreuses heures de recherche dans des bibliothèques souvent accessibles aux seuls citoyens des pays riches sont désormais au bout des doigts de tout un chacun.

Si la grande presse a pu faire ses choux gras des abus de cette liberté, et si la recherche d'informations sérieuses aux dépens des moins sérieuses demande, comme par le passé, une bonne connaissance de son domaine de recherche, l'arbre ne doit pas cacher la forêt, et l'existence de ce moyen de diffusion universel et bon marché représente une véritable étape dans l'histoire de la pensée et de la civilisation. Son existence a permis l'apparition d'innombrables sources de créativité dans les domaines les plus variés, et qui n'auraient sans lui jamais trouvé la voie de la publication. On trouvera dans le livre de Lawrence Lessig [1] une analyse approfondie du rôle d'un système de publication libre d'accès, et des menaces qui pèsent aujourd'hui sur lui.

Modèle en couches et intelligence aux extrémités

La charpente de l'Internet (comme de toute architecture de réseau) est constituée de protocoles, c'est-à-dire de documents qui fixent les règles imposées à un type de communication donné. Les protocoles de réseau sont généralement représentés selon une architecture en couches imaginée par le chercheur néerlandais Edsger Wybe Dijkstra (1930-2002) dans un article fameux publié en mai 1968 par les CACM (Communications of the Association for Computer Machinery), «The structure of the THE multiprogramming system» [2].

Avant d'être le plan de construction du réseau concret, l'architecture en couches est un modèle destiné à se représenter intellectuellement les choses par des abstractions. Ce modèle est utile pour «penser un objet dans lequel plusieurs logiques

s'articulent» (Michel Volle, <http://www.volle.com/opinion/couches.htm>), lorsqu'il faut séparer différents niveaux d'abstraction. On nommera couches basses celles qui décrivent la transmission de signaux sur des supports physiques tels que câble téléphonique, faisceau hertzien ou fibre optique, tandis que les couches intermédiaires concernent l'acheminement de messages complexes à travers des réseaux à la topologie également complexe, et que les couches hautes traitent de la présentation de ces messages à travers une interface utilisateur, de l'identification de leur destinataire et de son authentification. Les ensembles de règles et de conventions qui régissent les communications entre les couches de même niveau de plusieurs systèmes communicants constituent des protocoles de communication. Les règles et les conventions qui régissent les échanges entre une couche donnée et la couche immédiatement inférieure d'un même système constituent une interface.

Une des choses qui a fait la force de l'Internet, c'est le principe «end to end», que l'on pourrait traduire par «l'intelligence est aux extrémités», c'est-à-dire dans les programmes installés sur les ordinateurs des utilisateurs finals; l'architecture du réseau reste la plus simple possible, on s'interdit d'introduire au sein des infrastructures des dispositifs ou des optimisations spécifiques pour tel ou tel usage particulier. Pour implanter un nouveau protocole qui devra permettre tel nouveau mode de communication entre deux ordinateurs connectés au réseau, par exemple la téléphonie par l'Internet, tout ce qu'il faudra, c'est que ces ordinateurs soient dotés chacun des protocoles de base (Internet Protocol, ou IP, et en général Transmission Control Protocol, ou TCP), et bien sûr y installer les logiciels propres à la nouvelle application. Toutes les fonctions nouvelles seront donc implantées dans les ordinateurs aux extrémités de la communication, sans nécessiter aucune adaptation du réseau et des noeuds intermédiaires. C'est-à-dire, par exemple, que lorsque Tim Berners-Lee a voulu créer HTTP, le protocole du WWW, il n'a eu à demander l'autorisation de personne, parce que ce nouveau protocole ne nécessitait aucune modification de l'infrastructure du réseau - bien sûr, pour que HTTP devienne un standard de l'Internet, il aura fallu ensuite quelques formalités, que nous évoquerons ci-dessous. Ce principe «end to end» garde l'infrastructure du réseau aussi simple, ouverte et favorable à l'évolution que possible.

Le fait que le contrôle de l'accès à l'Internet ne puisse être accaparé par personne garantit à ses usagers qu'ils pourront l'utiliser sans restriction tant pour y publier que pour y consulter des informations, et c'est ce qui a fait son succès, face à l'incompréhension totale des grandes entreprises de télécommunications comme de télédiffusion, qui continuent à n'y voir qu'un hybride entre TF1 et le catalogue de la Redoute. Tant pis pour elles.

Mise au point historique

Entamons cet examen par la réfutation d'une légende aussi fallacieuse que répandue et répétée ad nauseam: **l'Internet n'a pas été créé selon un cahier des charges rédigé par les militaires américains en vue de préserver une capacité de communication après une frappe nucléaire.** Il n'y a jamais rien eu qui ressemble de près ou de loin à un «cahier des charges de l'Internet». Cette thèse télescope plusieurs événements liés mais distincts. Paul Baran, du think tank RAND, contractant du DoD (Department of Defense), avait posé les principes d'un système de communications dépourvu de point de centralisation unique afin de maintenir certaines liaisons même si certains noeuds venaient à être détruits. Les travaux de Baran furent publiés entre 1960 et 1964. Le même DoD, plusieurs années plus tard, en 1969, a impulsé par son agence ARPA (Advanced Research Projects Agency) la création du réseau ARPANET, qui n'était pas destiné aux communications militaires mais à faciliter la collaboration entre centres de recherches universitaires et industriels sous contrat avec l'ARPA. À sa création ARPANET reliait quatre universités: Stanford à Palo Alto, les campus de

Los Angeles et de Santa Barbara de l'Université de Californie, et Utah à Salt Lake City. On pourra consulter à ce propos le livre passionnant de Katie Hafner et Matthew Lyon [3].

C'est en 1974 que Vinton Cerf (de l'université Stanford) et Robert Kahn de BBN (Bolt, Beranek & Newman) publièrent le premier article sur TCP /IP. En 1976, la clairvoyance de Vint Cerf et de Robert Kahn, le financement de BBN et les contrats de l'ARPA devenue entre temps la DARPA donnèrent naissance au protocole réseau de l'Internet, destiné à devenir TCP /IP. En 1977 ce protocole commença à fonctionner, et c'est de ce moment que l'on peut dater la naissance de l'Internet expérimental.

Nous avons déjà mentionné comme une innovation propre à l'Internet le principe de «l'intelligence aux extrémités» («end to end»), nous pouvons en ajouter une autre, liée à la précédente et qui a trait à la fiabilité. Voilà de quoi il s'agit: on dit qu'un protocole de transmission est fiable s'il comporte une vérification que les données émises ont été transmises sans erreur; la vérification a lieu au moyen d'informations redondantes ajoutées aux données émises, dont on vérifie à l'arrivée la cohérence avec le corps du message. Il est en effet dans la nature des communications à distance d'être sujettes aux erreurs de transmission. Si un message a été corrompu lors de sa transmission, il faut en général prévenir son émetteur pour qu'il le réémette. Si l'on dispose d'un protocole fiable pour faire circuler des données entre deux noeuds adjacents, et si l'extrémité réceptrice de la communication vérifie l'intégrité des données, conformément au principe de l'intelligence aux extrémités, il est inutile que le protocole intermédiaire chargé de faire circuler les données dans le réseau (la couche réseau) comporte lui-même un contrôle d'erreur. Ce principe d'une couche réseau non-fiable, formulé pour la première fois par Louis Pouzin pour le réseau français Cyclades, permet une grande simplicité du protocole de réseau.

C'est en 1978 pendant une pause d'une conférence à Marina del Rey que Vint Cerf et Jon Postel tirèrent les conséquences de ce principe de la couche réseau non-fiable en divisant leur protocole en deux parties: un protocole fiable de transport de bout en bout, TCP, et un protocole réseau non-fiable, IP. Cette innovation hardie allait constituer une des forces de l'Internet, avec quelques autres qu'il n'est pas possible d'exposer ici et pour lesquelles on se reportera avec profit au livre de Katie Hafner et Matthew Lyon [3].

Un peu plus tard, en 1979, le groupe de recherche en système (Computer Systems Research Group, CSRG) de l'Université de Californie à Berkeley allait incorporer TCP /IP à une nouvelle version d'Unix, dite BSD (Berkeley Software Distribution).

En 1982 la Défense américaine adopte officiellement les protocoles TCP et IP (Internet Protocol) pour le réseau ARPANET et accepte de les distribuer gratuitement sur le réseau. C'est ainsi que s'est nouée une situation qui allait unir les destins de l'Internet, d'Unix et des logiciels libres.

Tous ces éléments ont contribué à donner naissance en 1986 à l'Internet tel que nous le connaissons, quand les centres de recherche connectés à ARPANET ont voulu communiquer avec les universités de leur choix et que la NSF (National Science Foundation) a entrepris de leur en donner les moyens en créant son propre réseau, NSFNet.

Dès le début des années 1980, ARPANET acceptait des connexions à partir d'universités non américaines. Il s'agissait souvent (dans le cas de la France notamment) de liaisons indirectes, établies par des passerelles et des relais complexes, mais qui donnaient accès au réseau; la France eut accès au courrier électronique ainsi en 1984, mais ce n'est que le 28 juillet 1988 que notre pays fut vraiment raccordé à l'Internet par une liaison permanente en TCP /IP [4].

Organisation administrative de l'Internet

Il faut rappeler ici que l'Internet n'est la propriété de personne ni d'aucune institution, ce qui semble parfois difficile à comprendre. Nous invitons le lecteur à quelques secondes de réflexion admirative devant un dispositif technique conçu à l'origine pour un réseau d'une centaine de noeuds et qui a pu résister à une croissance de six ou sept ordres de grandeur, d'autant plus que cette conception n'était le fait ni d'un grand groupe industriel ou financier, ni d'un gouvernement, ni d'un conglomérat de telles puissances. Mais peut-être était-ce là le secret ? Le fonctionnement de l'Internet, à l'image de sa construction, repose sur la coopération volontaire. Nous donnons ici l'organigramme général de son organisation:

- L'Internet Architecture Board (IAB) est responsable des grandes orientations et de la coordination.
- L'Internet Engineering Task Force (IETF) se charge de la normalisation à court terme et émet les Requests for Comments (RFC), qui sont les documents de référence pour le fonctionnement du réseau*. Toutes les RFC sont accessibles par l'URL (Universal Resource Locator) <http://www.ietf.org/rfc/> ou sur de nombreux sites miroirs. Nous ne saurions trop en conseiller la lecture: même si leur qualité littéraire est inégale elles fournissent sur l'Internet une information de première main, souvent exposée très clairement.
- L'Internet Steering Group (IESG) coordonne l'IETF, dont l'effectif est devenu très important.
- L'Internet Assigned Numbers Authority (IANA) centralise et contrôle les conventions relatives à l'identification des objets du réseau, et notamment veille à l'unicité des adresses.
- L'Internet Corporation for Assigned Names and Numbers (ICANN) , créée en 1998 à l'initiative du gouvernement américain, supervise l'attribution des noms de domaines et des adresses.

Ainsi les décisions organisationnelles et techniques sont prises par des instances aux séances desquelles tout un chacun peut assister et participer, les séances des instances de pilotage (IAB, IESG, ICANN) étant toutefois réservées à leurs membres élus. Cette organisation coopérative ne signifie pas l'absence de rapports de force marchands ou politiques, mais elle exclut (au moins à court terme) la prise de contrôle par une entreprise unique. L'ICANN soulève bien des critiques du fait de l'influence déterminante qu'y exerce le gouvernement américain, mais il est significatif que la réaction à ce qui pourrait devenir une mainmise est rendue possible par la structure ouverte des autres instances, par contraste avec ce qui se passe à l'ISO ou à l'UIT (Union Internationale des Télécommunications).

* Citons ici le nom de Jon Postel, éditeur des RFC depuis la première en 1969 jusqu'à sa mort en 1998, et auteur ou coauteur de 204 d'entre elles, ce qui lui a conféré une influence considérable sur la physionomie du réseau.

Organisation topographique de l'Internet

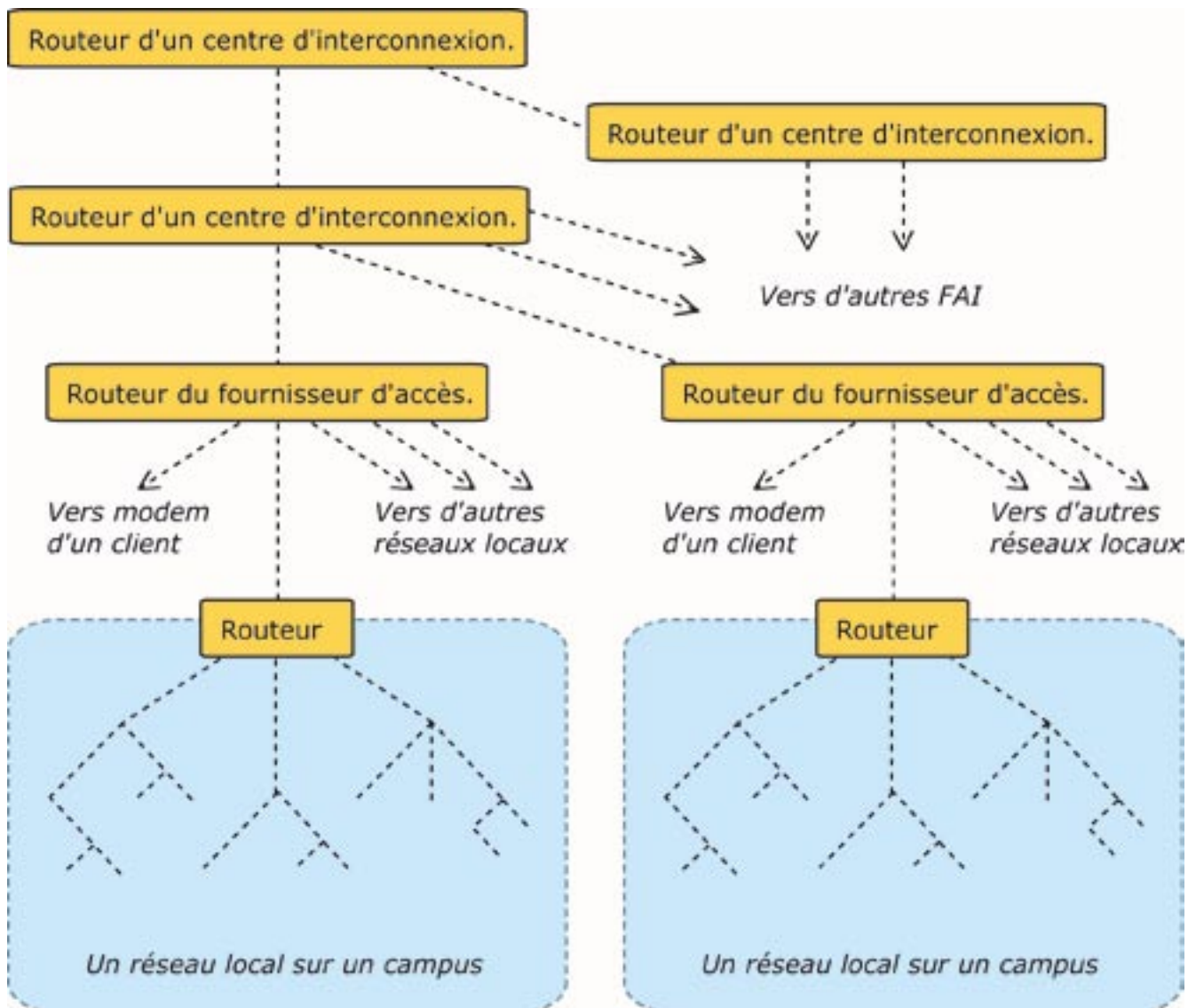


Figure 1 Topographie simplifiée de l'Internet.

La figure 1 donne une représentation schématisée de la topographie de l'Internet. Cette représentation est simplifiée, notamment elle est purement arborescente, alors que rien n'empêche une entreprise d'avoir plusieurs fournisseurs d'accès à l'Internet (FAI), ni un FAI d'être connecté à plusieurs centres d'interconnexion de niveau supérieur, ce qui complique le schéma et le transforme d'un arbre en un graphe connexe quelconque. Chaque noeud (ordinateur connecté, ou autre type d'équipement) est identifié par une adresse (dite adresse IP) unique à l'échelle mondiale, un peu comme un numéro de téléphone précédé de tous ses préfixes nationaux et régionaux. L'essentiel dans l'établissement d'une communication, c'est, à chaque noeud, de savoir quel est le prochain noeud sur l'itinéraire, et par quelle liaison l'atteindre. Les noeuds terminaux, origines ou destinations des communications, sont au sein des réseaux locaux de campus ou d'entreprise. Un routeur est un noeud particulier, doté de plusieurs interfaces réseau (et donc de plusieurs adresses), ce qui lui permet d'être connecté simultanément à plusieurs réseaux et de faire passer les paquets de données d'un réseau à un autre, en fonction de règles de routage qui auront été enregistrées dans sa mémoire.

Architecture de l'Internet

Les documents qui décrivent les protocoles - et par conséquent l'architecture - de l'Internet s'appellent des Requests for Comments (RFC) [5]. Les protocoles les plus connus sont IP (Internet Protocol) pour l'acheminement d'un paquet de données à travers un réseau complexe, TCP (Transmission Control Protocol) pour le transport d'un segment de données de bout en bout selon une connexion établie, puis au niveau applicatif SMTP (Simple Mail Transfer Protocol) pour le courrier électronique, HTTP (Hypertext Transfer Protocol) pour le WWW, FTP (File Transfer Protocol) pour le transfert de fichiers, etc.

Les protocoles sont publiés par l'IETF à partir de soumissions spontanées. Le RFC 2026 définit la procédure de validation des RFCs: une soumission doit suivre un processus nommé «standards track», qui comporte trois étapes: «Proposed Standard», «Draft Standard», et «Standard». L'introduction d'un document dans ce processus est effectuée à l'initiative de l'IESG, qui choisit les «Proposed Standards» parmi des documents nommés «InternetDrafts», soumis sans formalités à la discussion. Pour qu'une soumission soit examinée avec quelque chance de succès il est de coutume qu'elle soit accompagnée d'un prototype de logiciel pour l'illustrer.

Il est clair que la spontanéité des soumissions est variable, et, au fur et à mesure que l'Internet devenait un enjeu économique important, les industriels de l'informatique et des réseaux ont contribué de plus en plus intensément à la création de RFCs, mais sans que le caractère consultatif et concerté du processus soit fondamentalement altéré.

De fait, aujourd'hui, le fonctionnement de l'Internet repose sur des documents de normalisation librement accessibles à tous. Contrairement aux normes ISO qui ne sont accessibles que moyennant finances. Faire payer ou non l'accès aux normes, c'est une décision lourde de conséquences, qui définit le public auquel on s'adresse. Des logiciels libres sont disponibles pour tous les usages de l'Internet: serveurs de messagerie (Sendmail, Postfix, Qmail, Exim), serveurs WWW (Apache), serveurs de noms (Bind, Djbdns), navigateurs (Mozilla)... Ce processus de création par accréation d'une infrastructure mondiale, à la cohérence extrême, utilisée par des centaines de millions de personnes, en l'absence de tout pilotage centralisé, constitue un défi réellement fascinant pour tous les architectes de systèmes.

Bibliographie

[1] Lawrence LESSIG.

The future of ideas - The fate of the commons in a connected world.
Random House, New York, 2001.

[2] Edsger Wybe DIJKSTRA.

«The structure of the THE multiprogramming system ».

Communications of the ACM (CA CM), \01. 11 n° 5, May 1968.

[3] Katie HAFNER et Matthew LYON.

Where Wizards Stay Up Late - The Origins of the Internet.

Pocket Books, Londres, 1996.

[4] Christian HUITEMA.

Et Dieu créa l'Internet.

Eyrolles, Paris, 1995.

[5] IETF. « Requests for Comments ».

RFC, 2004. <http://www.ietf.org/rfc/>.