RFC 7663 : IAB Workshop on Stack Evolution in a Middlebox Internet (SEMI) Report

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 22 octobre 2015

Date de publication du RFC : Octobre 2015

https://www.bortzmeyer.org/7663.html

Une caractéristique importante de l'Internet d'aujourdhui, et qui le distingue considérablement de l'architecture initialement prévue, est la prévalence de "*middleboxes*" un peu partout sur le trajet. Alice et Bob ne peuvent plus, dans la plupart des cas, se parler directement, ils doivent passer par des "*middleboxes*", dont les intérêts ne sont pas ceux des utilisateurs, et qui sont en outre fréquemment boguées et/ou limitées dans leurs fonctions. Un des effets est l'**ossification** de l'Internet : innover devient de plus en plus difficile, car il y aura toujours une "*middlebox*" sur le trajet qui ne sera pas d'accord. Quelles sont les conséquences de cette "*middleboxation*" de l'Internet ? Comment peut-on résoudre les problèmes qu'elle posent ? C'est le but du programme "*IP Stack Evolution*" <https://www.iab.org/activities/programs/ip-stack-evolution-program/> de l'IAB que de trouver des réponses à ces questions. C'est dans ce cadre qu'un atelier a été organisé en janvier 2015 à Zurich pour explorer la question. Ce RFC est le compte-rendu de l'atelier SEMI ("*Stack Evolution in a Middlebox Internet*").

On comprend mieux le problème en revenant à l'architecture prévue pour l'Internet (très bien décrite dans le RFC 1958 [1]), et qui a assuré son succès, le **principe de bout en bout**, formalisé dans l'article de J.H. Saltzer, D.P. Reed et David Clark, « "*End-To-End Arguments in System Design*" <http://web.mit.edu/saltzer/www/publications/endtoend/endtoend.pdf> » ("*ACM TOCS, Vol 2, Number 4, November 1984*"). Ce principe dit en gros que deux machines connectées à l'Internet doivent décider elles-mêmes de ce qui est bon pour elles, les équipements intermédiaires qui forment le réseau ne sont pas bien placés pour prendre des décisions. Ce principe a pour conséquence un réseau « bête », où les routeurs transmettent les paquets sans chercher à comprendre. Des décisions comme le contrôle de congestion sont faites aux extrémités du réseau, dans les machines terminales <https://www.bortzmeyer.org/terminal-host.html>. Par exemple, un protocole de transport, dont l'une des fonctions est de

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc1958.txt

réagir en cas de congestion, sera mis en œuvre uniquement dans ces machines. Conséquence de ce principe, l'innovation ne nécessite pas de changement dans l'infrastructure. Si on trouve que le protocole de transport TCP n'est pas terrible, on en invente un autre (comme SCTP) et du jour au lendemain, tout couple de machines Internet peut l'utiliser, sans qu'on ait eu à modifier les routeurs de l'Internet.

Mais l'Internet a bien dégénéré depuis le RFC 1958. D'innombrables équipements supplémentaires, les "*middleboxes*" ont été installés et leur but n'est pas de fournir des services aux utilisateurs mais au contraire de les contrôler et de limiter ce qu'ils peuvent faire. C'est ainsi que boitiers NAT, pare-feux, et autres "*middleboxes*" se permettent désormais de regarder le trafic et de dire, par exemple « ah, non, SCTP, je ne connais pas, je ne laisse pas passer ». C'est ce qu'on nomme l'ossification : déployer un nouveau protocole de transport, voire un nouveau protocole d'application (« un seul autorisé, HTTP », est un principe fréquent sur les "*middleboxes*"), devient de plus en plus difficile. Peut-on encore désossifier l'Internet ? Le problème a été identifié il y a longtemps (cf. l'exposé de S. Deering, « "*Watching the Waist of the Protocol Hourglass*" <https://www.ietf.org/proceedings/51/slides/plenary-1> »), il reste à travailler sur les solutions.

L'atelier SEMI avait donc invité les experts à soumettre des articles autour de questions comme « quels chemins dans l'Internet acceptent encore tous les transports ? », « quels chemins sont ouverts à toutes les applications ? », « comment les applications peuvent-elle faire pour contourner les blocages ? », « comment revenir à une situation plus saine ? ». Des tas de pistes étaient possibles :

— Migrer certaines fonctions de la couche transport vers la couche application (d'une certaine façon, c'est un peu ce que fait HTTP/2, cf. RFC 7540),
— Développer de meilleurs outils pour que l'application puisse découvrir si un chemin est propre ou pas (et, dans ce cas, quels contournements auront le plus de chances de marcher),
— Développer des méthodes pour découvrir les "*middleboxes*" et leurs caractéristiques,
— Réfléchir aux questions "*business*" : ce n'est pas tout d'avoir des solutions, il faut aussi les déployer et pour cela, avoir des arguments sonnants et trébuchants, pas juste « on va rétablir les principes originels ».

Les articles acceptés sont tous disponibles sur le site Web de l'atelier SEMI <https://www.iab.org/activities/workshops/semi/>. Ce RFC est très court et n'en fait qu'un résumé (et moi, dans cet article, un résumé du résumé).

À l'époque de l'exposé de Deering cité plus haut, le principal type de "*middlebox*" repéré était le routeur NAT. Aujourd'hui, les problèmes causés par le NAT, sans avoir été supprimés, sont bien compris et il existe des solutions de contournement (imparfaites, contrairement à ce que laisse entendre l'optimisme du RFC, qui parle d'un problème « largement résolu »). Plusieurs groupes de travail de l'IETF ont travaillé pour cela, notamment BEHAVE <https://www.bortzmeyer.org/behave-wg.html>. Mais les "*middleboxes*" sont devenues plus variées et plus intrusives, au fur et à mesure que leurs possibilités techniques augmentaient (section 2 de notre RFC). Aujourd'hui, le problème qu'elles posent devient d'autant plus intolérable que les préoccupations de sécurité, et notamment de protection de la vie privée, poussent au développement de la cryptographie. Celle-ci permet de préserver l'intégrité des communications de bout en bout et s'oppose donc directement aux "*middleboxes*". (C'était le coeur de l'exposé de Huitema à SEMI <https://www.iab.org/wp-content/IAB-uploads/2014/12/semi2015_huitema.pdf>.)

Qu'est-ce qui fait que des "*middleboxes*" sont déployées ? La section 3 du RFC se penche sur les motivations ; après tout, il y a peu de chances de pouvoir revenir sur ce déploiement si on ne comprend pas pourquoi il a lieu. Parmi les causes, il y a la loi de Moore, qui fait qu'il est de plus en plus réaliste de mettre des traitements importants sur le chemin des communications, malgré la concentration de trafic. Il y a le désir des opérateurs réseau de ne pas être « uniquement un tuyau » (ce que voudraient les utilisateurs) mais d'imposer des traitements (qu'on baptise « valeur ajoutée ») qui justifieront des augmentations de tarif, ou des violations de la neutralité <https://www.bortzmeyer.org/neutralite.html>. Il y a des services informatiques qui trouvent qu'il est plus facile de déployer une politique dans

le réseau que dans les machines terminales, d'autant plus qu'ils ne contrôlent pas forcément celles-ci (cas du BYOD). Et il y a bien sûr le marketing des vendeurs de "*middleboxes*" qui insiste lourdement pour qu'on achète ces boitiers, dotés de vertus miraculeuses (en sécurité ou en performances). C'est vrai qu'acheter est un acte simple, et il est tentant de se procurer un service juste en faisant un chèque. Toutes ces motivations contribuent à l'ossification du réseau.

Les organisateurs de SEMI se placent dans le cadre d'un système capitaliste pur (rebaptisé, car c'est plus joli, « libre marché »), vu comme incontournable. Dans ce cadre, les désaccords ne sont adressables que par le jeu du marché et il faut donc fournir des motivations "*business*" pour tout changement. Cela mène à la question « comment vendre les alternatives aux "*middleboxes*" ? »

La section 4 explore ce que font exactement les "*middleboxes*". Car il y en a de plein de sortes différentes et, pour résoudre les problèmes qu'elles posent, il faut les étudier et les classer (c'était le papier de Edeline et Donnet <https://www.iab.org/wp-content/IAB-uploads/2014/12/semi2015_edeline.pdf>).

Cette étude est notamment nécessaire pour savoir si les applications pourront coopérer avec les "*middleboxes*" ou bien s'il faudra envisager des mesures plus « ninja » comme de tout tunneler (la contribution de Raiciu et Olteanu <https://www.iab.org/wp-content/IAB-uploads/2015/01/ninja.pdf>).

Et sur les protocoles de transport ? La section 5 discute de comment faire évoluer la couche 4 dans un Internet gelé par les "*middleboxes*". S'inspirant d'un épisode du film « La vie de Brian », le RFC classe les propositions en deux camps :

— Le « "*TCP Liberation Front*" », ceux qui veulent avant tout permettre de reprendre l'évolution de TCP, déployer de nouvelles extensions (qui entrainent actuellement souvent une réaction négative des "*middleboxes*").
— Le « "*People's Front of UDP*" », qui considère que c'est fichu et que tous les services « couche 4 ou à peu près » futurs devront être encapsulés dans l'UDP (c'est par exemple ce que propose le RFC 6951 pour arriver à faire passer SCTP et c'est également pour cette raison que LISP est forcément transporté dans UDP alors qu'il aurait très bien pu être directement sur IP).

Ajoutez la crypto à cela et la solution UDP se traduira peut-être dans le futur par « tout sur DTLS »... Des tas de solutions ont été imaginées par les « transporteurs », les gens de la couche 4. Par exemple, la contribution de Briscoe <https://www.iab.org/wp-content/IAB-uploads/2014/12/semi2015_briscoe.pdf> propose de mettre les options et extensions nouvelles de TCP dans la charge utile et non pas dans l'en-tête TCP, les mettant ainsi à l'abri des "*middleboxes*" (jusqu'à ce qu'elles se mettent à faire du DPI lourd...)

À noter que tous les chemins sur l'Internet ne sont pas forcément affligés de "*middleboxes*". Quelles que soient les solutions de contournement adoptées, il serait dommagee qu'elles soient imposées à tous, avec leurs coûts en octets gaspillés et en complexité, alors qu'on peut s'en passer dans beaucoup de cas. Welzl, Fairhurst et Ros avaient ainsi proposé <https://www.iab.org/wp-content/IAB-uploads/2014/12/semi2015_welzl.pdf> de consacrer des efforts à la détection des "*middleboxes*". On teste le chemin et, s'il est propre, on se dispense des horreurs nécessitées par le contournement des "*middleboxes*". Comme ce test va prendre du temps, il faudra sans doute le faire en parallèle des tentatives de connexion, comme dans le cas du RFC 6555.

Bon, maintenant, place à l'action. Qu'est-ce qui doit être fait pour avancer ? La section 6 de notre RFC liste les points à traiter. D'abord, est-ce qu'on ne pourrait pas envisager de signaler au réseau certaines informations sémantiques (du genre « ce paquet est le début d'un nouveau flot »), dispensant les "*middleboxes*" d'essayer de (mal) trouver cette information elles-mêmes ? Cela serait d'autant plus utile

que cette information est masquée en cas de chiffrement. Cette idée est à la base du projet SPUD ("*Substrate Protocol for User Datagrams*", « un traité de paix entre machines et réseau »). Ce projet a suscité des controverses lors de sa présentation à l'IETF 92 <http://www.ietf.org/meeting/92/> à Dallas (si on chiffre, ce n'est pas ensuite pour donner des informations à un réseau en qui on n'a pas confiance) et aucun groupe de travail n'a encore été créé. Les objections à SPUD portent notamment sur le fait que la publication des informations devrait être volontaire, et le résultat d'un choix délibéré.

Autre sujet sur lequel il faudrait travailler, la mesure. Cela ne vaudrait pas la peine de développer des hacks d'enfer dans les couches 4 ou 7 pour contourner une perturbation qui est rare. Or, à l'heure actuelle, on manque de données sur les différents types de problèmes causés par les "*middleboxes*". Par exemple, combien de chemins dans l'Internet ne laissent pas passer un protocole de transport inconnu ? (Avant de crier « aucun, tout le monde est derrière un routeur NAT », pensez aux réseaux d'entreprise ou d'organisation, aux réseaux IPv6, etc.) Ces mesures sont le travail de la liste de diffusion hops <https://www.ietf.org/mailman/listinfo/hops> ("*How Ossified is the Protocol Stack ?*") qui discute des meilleurs moyens de collecter cette information, par exemple à partir des journaux des applications qui tentent d'utiliser tel ou tel protocole mais se replient ensuite sur une autre solution. Le pourcentage de repli serait une information utile.

Le RFC n'en parle pas mais un des problèmes posés par les "*middleboxes*" est typiquement que leurs programmeurs sont inconnus, et ne participent pas aux forums comme le RIPE ou l'IETF, où se discutent les problèmes de l'Internet. Même si on rédige des bonnes pratiques à suivre pour les auteurs de logiciel de "*middlebox*", il y a peu de chances qu'ils les lisent, ou même apprennent leur existence. (Les "*middleboxes*" actuelles ont manifestement été programmées sans lire les RFC ou en tout cas sans lire tous les RFC pertinents, cf. RFC 5625 pour des exemples).

Mais l'IETF est optimiste et croit que l'humanité peut s'améliorer : d'où des documents comme le RFC 3234. Un travail est en cours pour le compléter avec des conseils pratiques.

Bref, la lutte <https://www.bortzmeyer.org/tussle-cyberspace.html> entre les machines terminales, qui veulent un tuyau bête et neutre, et le réseau qui cherche à imposer ses propres vues, n'est pas terminée, loin de là.