Van Jacobson et le réseau centré sur le contenu

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 13 juillet 2011

https://www.bortzmeyer.org/van-jacobson-ccn.html

J'ai déjà eu l'occasion de le dire : en matière de recherche sur les réseaux informatiques, j'apprécie les gens qui développent des idées originales, ne se limitent pas à penser dans le cadre existant, et déblaient le terrain pour les développements futurs. L'Internet tel qu'il est aujourd'hui n'est pas éternel, et il faut travailler à son remplaçant. Seulement, parmi les gens qui annoncent être des penseurs radicaux et des réformateurs audacieux, pour un vrai pionnier qui innove, il y a cent powerpointeurs, qui tiennent des grands discours, mais ne font rien de concret, et dont les idées ne dépassent jamais le stade « dos de l'enveloppe ». Le « réseau centré sur le contenu », prôné notamment par Van Jacobson, est de quelle catégorie ?

Pour mes lecteurs qui préfèrent les visions positives, je vais d'abord parler du projet de Van Jacobson et de ses collègues du PARC, le "*Content-Centric Networking*" (**CCN**), car c'est un projet très intéressant. Mais, comme rien n'est parfait en ce bas monde, je mentionnerai aussi d'autres efforts médiatiques qui utilisent ce même slogan de « réseau basé sur le contenu », et qui sont davantage dans la lignée des « raseurs de table » habituels (ceux qui prétendent faire « table rase » de l'Internet existant <https://www.bortzmeyer.org/science-et-vie-table-rase.html> et repartir de zéro <https://www.bortzmeyer.org/table-rase-et-john-day.html>).

L'article fondateur du projet est « "*Networking Named Content*" <http://conferences.sigcomm.org/co-next/2009/papers/Jacobson.pdf> ». Parmi les auteurs, Van Jacobson, légende de l'Internet, inventeur de traceroute, et auteur de pas mal de RFC comme le RFC 1144 [1], sur la technique de compression qui porte son nom. Il est surtout connu pour ses travaux sur le contrôle de congestion (cf. RFC 2001). Mais je vais essayer de ne pas utiliser d'argument d'autorité et de lire son article (après tout, il n'est qu'un des nombreux co-auteurs) objectivement.

Donc, que dit l'article en question ? Il part de l'affirmation comme quoi les communications sur l'Internet, dominées autrefois par des interactions entre machines (typiquement, telnet) le sont aujourd'hui

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc1144.txt

par de l'accès au contenu (je sais, c'est très Minitel, comme approche). Il faut donc réorganiser les protocoles Internet autour de cette idée. Cela change par exemple le système de nommage et d'adressage, où on ne nomme jamais une machine, mais toujours des données (le **quoi** plutôt que le **où**). La forme exacte des noms n'est pas complètement définie (c'est un projet de recherche, rappelez-vous) mais en gros, la machine qui a une copie de `/parc.com/media/art` annonce ce préfixe localement et une machine qui veut `/parc.com/media/art/carla-bruni.mp3` pourra alors le trouver (imaginez des URL annoncés en BGP...)

Le routage est donc modifié. Au niveau local, CCN fonctionne en diffusant à la cantonade (« qui a une copie de `carla-bruni.mp3` ? »). Au niveau global, les protocoles de routage peuvent eux-même être réécrits en terme de CCN (les routes étant le contenu auquel on cherche à accéder). À noter que cela ne marche qu'avec les protocoles de routage qui gardent toute l'information (comme OSPF). Ceux qui condensent (comme RIP) ne conviennent pas au CCN, où le même contenu peut se trouver à plusieurs endroits. Les FAI sont encouragés à déployer des caches locaux, pour accélérer le processus. Du point de vue pratique, CCN peut fonctionner directement sur la couche 2 ou bien sur IP, pour faciliter la transition.

L'idée de base me semble très contestable. Toutes les utilisations de l'Internet ne rentrent pas dans ce cadre. Même sur le Web, ce modèle de « récupération d'un contenu » ne colle pas du tout à l'utilisation de Facebook, de Twitter ou de Google Maps.

Un problème courant des nouveaux systèmes de nommage est la sécurité. Aujourd'hui, je fais confiance à <http://www.rue89.com/planete89/2011/07/12/baleines-algues-moules-un-ocean-radioacti
parce que les protocoles Internet garantissent que ce contenu vient bien des serveurs de Rue89. Dans un réseau « orienté contenu », on perd cette garantie. Comment la récupérer ? C'est l'objet de l'excellent article « "*Securing Network Content*" <http://conferences.sigcomm.org/co-next/2009/workshops/rearch/papers/Jacobson.pdf> ». Pour comprendre le problème, voyons le contenu accessible sur le bon vieil Internet en <http://www.bortzmeyer.org/files/exemple-de-contenu.txt>. Si on obtient ce fichier par un autre biais que le Web, par exemple par courrier, on ne peut pas le valider, on n'a pas de moyen de savoir qu'il est authentique. Il existe plusieurs solutions à ce problème dont la plus courante est d'utiliser des noms auto-validants. Par exemple, si je décide que les identificateurs du contenu sont des condensats cryptographiques dudit contenu (un certain nombre de protocoles pair-à-pair comme Freenet font effectivement cela), le nom `http://www.bortzmeyer.org/files/exemple-de-conte` devient (en SHA-256), `c4eb57876451fa515f483a4fa1ae7e2845c3cfcede5582f5133fb5e6a4245205`. Ce nom est auto-validant (si je récupère le contenu par un autre biais, je peux calculer le condensat et vérifier qu'il correspond) mais il a deux défauts (on ne peut pas avoir tous les avantages en même temps <https://www.bortzmeyer.org/no-free-lunch.html>) : il n'est pas très convivial et il ne permet pas de résolution facile (étant dans un espace plat, on est obligé d'utiliser un serveur centralisé, un protocole d'inondation, ou bien une DHT pour le résoudre). CCN utilise une autre approche : on signe le lien entre un localisateur (par exemple un URL) et le contenu. Prenons un exemple avec une technologie classique, PGP. Le condensat cryptographique du fichier cité plus haut a été calculé avec `sha256sum exemple-de-contenu.txt` et le texte du lien a été signé avec `gpg --sign --detach lien-nom-contenu.txt`. Vous pouvez donc, lorsque vous avez le lien ((en ligne sur `https://www.bortzmeyer.org/files/lien-nom-contenu.txt`)) et sa signature ((en ligne sur `https://www.bortzmeyer.org/files/lien-nom-contenu.txt.sig`)) vérifier la signature, puis récupérer le fichier <http://www.bortzmeyer.org/files/exemple-de-contenu.txt>, calculer son condensat, et vérifier que tout est normal. Naturellement, dans un vrai CCN, tout ceci sera fait automatiquement. Mais cet exemple permet de montrer le principe.

Bien sûr, la faiblesse évidente de ce concept de CCN est cette idée de tout centrer sur le contenu : ce modèle convient bien à la récupération de pages Web statiques ou à celle de fichiers (musique, film) en pair-à-pair. Pas étonnant que ce thème ait du succès, puisque ces deux utilisations sont les seules que connaissent les commerciaux, les journalistes et les hommes politiques. Mais l'Internet sert à bien

d'autres choses, notamment les usages **conversationnels** où on ne récupère pas du contenu mais on mène un dialogue. Ces usages sont par exemple la connexion SSH ou bien la messagerie instantanée. Ces conversations rentrent très mal dans un modèle « centré sur le contenu ». Les auteurs en sont conscients. Ils ont même consacré un très bon article, « "*VoCCN : Voice-over Content-Centric Networks*" <http://conferences.sigcomm.org/co-next/2009/workshops/rearch/papers/Jacobson.pdf> » à cette question. Dans cet article, ils décrivent VoCCN, un mécanisme permettant de faire passer de la voix (un bon exemple des applications conversationnelles) sur CCN. Le principe est de considérer le téléphone de l'appelé comme un contenu « en devenir », donc de lui donner un nom et de récupérer le contenu associé à ce nom. C'est astucieux mais cela me semble peu naturel et cela illustre bien le syndrôme « lorsque le seul outil qu'on a est un marteau, tous les problèmes ressemblent à des clous ». Les auteurs ont décidé arbitrairement que tout était contenu, et la réalité doit se plier à ce modèle.

Le projet du PARC, contrairement à l'écrasante majorité des projets « table rase », ne produit pas uniquement du PowerPoint. Sur le site Web officiel <http://www.ccnx.org/>, on peut trouver des descriptions plus détaillées et plus concrètes <http://www.ccnx.org/releases/latest/doc/technical/>, du protocole <http://www.ccnx.org/releases/latest/doc/technical/CCNxProtocol.html>, des noms <http://www.ccnx.org/releases/latest/doc/technical/Name.html> ou de leur représentation en URI de plan `ccnx:`. Le code source mettant en œuvre ces idées est disponible <https://github.com/ProjectCCNx/ccnx> (attention, c'est très expérimental, réservé aux barbus curieux) : il se compose d'une bibliothèque de bas niveau en C et d'une autre en Java pour les accès « normaux ».

À partir de là, je vais être nettement moins positif. Car un projet ne se limite pas aux articles scientifiques (que personne ne lit), il a aussi une face publique, des articles de vulgarisation, des discours à la télé et, là, les choses se gâtent sérieusement.

Prenons par exemple un article qui se veut scientifique sur la forme mais qui est bien faible sur le fond, « "*Towards a Content-Centric Internet*" <http://www.coast-fp7.eu/public/Zahariadis_FCCI_Camera_Ready.pdf> », financé par plusieurs projets de l'Union européenne. Que trouve-t-on dans cet article à l'apparence austère mais qui rassemble tous les clichés des raseurs de table ? Il commence évidemment par l'affirmation comme quoi la majorité des usages de l'Internet est de la récupération de contenu (j'ai expliqué pourquoi c'est faux). Contrairement aux articles des chercheurs du PARC, il se contredit plusieurs fois, comme lorsqu'il donne comme exemple d'un utilisateur voulant « récupérer un contenu indépendemment de sa localisation », l'exemple d'un utilisateur qui veut « lire les nouvelles sur le site de la BBC ». Il est vrai que l'article mélange allègrement localisation physique <https://www.bortzmeyer.org/hostname-physical-location.html> et emplacement dans le réseau. Comme beaucoup d'articles de raseurs de table, il peint un portrait erroné de la situation actuelle, par exemple en confondant nom d'un service et machine physique (lorsque j'écris à `page@gmail.com`, `gmail.com` ne désigne **pas** une machine). L'article part ensuite dans les habituelles promesses creuses qui font bien dans les colloques (« "*more efficient use of the available resource* », « *provide a business environment that encourages investment*" »). Parfois, le texte touche à la poésie (« "*A Content Object is an autonomous, polymorphic/holistic container*" »). Les dessins sont du même style (vides de contenu mais donnant l'impression d'une profonde sagesse, avec leurs boîtes qui s'interconnectent ; la figure 3 est particulièrement réussie de ce point de vue).

Comme beaucoup de réformateurs qui veulent changer radicalement l'Internet, ils lui reprochent surtout son côté trop libre. C'est ainsi qu'ils citent six fois le fait que leur approche « centrée sur le contenu » permettra de rendre les DRM plus efficaces (évidemment, si on voit le réseau comme un moyen de distribuer du contenu vers du temps de cerveau disponible, on arrive vite aux DRM).

Mais, au moins, les auteurs de cette vulgaire brochure commerciale ont le sens des réalités. Ils finissent par avouer « "*It is currently very difficult to imagine what a network architecture that support objects would look like*". ». Je ne peux pas dire mieux.

Encore meilleur (ou pire, selon le point de vue), l'article publicitaire qu'avait décroché les tenants du « réseau centré sur le contenu » dans Network World, « "*2020 Vision : Why you won't recognize the 'Net in 10 years*" <http://www.networkworld.com/news/2010/010410-outlook-vision.html> ». Cet article choisit la voie facile. D'abord, il compare l'Internet actuel (actuel, donc plein de défauts) avec un réseau idéal et très loin dans le futur (alors que, si on veut être honnête, il faut comparer le réseau des raseurs de table à l'Arpanet de Baran et Licklider, celui qui résistait aux attaques nucléaires). Le record étant lorsque Van Jacobson (dans une citation peut-être tronquée) dit qu'il va mettre fin au spam.

Naturellement, il y a zéro cahier des charges, ce qui permet de vagues promesses (« Meilleure sécurité ») sans indiquer ce qu'on abandonnera en échange (la sécurité est toujours un compromis...) Les seules fois où le cahier des charges est explicite, il va du côté sino-saoudo-hadopien : accord préalable avant d'envoyer des données, flicage obligatoire.

L'article de Network World ne fait pas de la publicité que pour CCN mais aussi pour d'autres projets, qui sont montrés comme s'ils allaient dans le même sens alors que les raseurs de table n'ont en commun que leur chasse aux subventions et présentent des projets incompatibles. L'exemple caricatural dans cet article est l'éloge de MPLS dans un article qui mentionne souvent Van Jacobson, inlassable pourfendeur de MPLS <http://www.sigcomm.org/sites/default/files/award-talks/vj-sigcomm01.pdf>.

Et enfin, l'article mélange des techniques qui nécessitent en effet une refonte complète du réseau et de techniques qui pourraient parfaitement être développées sur l'Internet existant (exemple : nouveaux mécanismes d'adressage et de localisation du contenu). Bref, une soupe catastrophique. Incompétence de journaliste, dont les chercheurs du PARC qui travaillent sur le CCN ne devraient pas etre tenus responsables ? Qui sait ? Van Jacobson est souvent cité dans l'article, avec des déclarations sensationnalistes, prétentieuses et inquiétantes comme « "*The security is so utterly broken that it's time to wake up now and do it a better way. The model we're using today is just wrong. It can't be made to work. We need a much more information-oriented view of security, where the context of information and the trust of information have to be much more central.*" »