

RFC 7476 : Information-centric Networking: Baseline Scenarios

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 27 mars 2015

Date de publication du RFC : Mars 2015

<https://www.bortzmeyer.org/7476.html>

Le concept d'ICN ("*Information-Centric Networking*") est assez à la mode en ce moment. Il désigne une conception du réseau où le composant essentiel n'est pas une machine mais une information, ayant un identificateur, et dont la récupération par l'utilisateur est le but principal de la connectivité. Ce modèle très minitélien colle bien à certains usages du réseau (comme le Web traditionnel, avant les applications Web, ou comme l'accès aux œuvres culturelles en pair à pair), ce qui explique son succès actuel. Ce RFC est le premier publié qui se penche sur ce paradigme : il décrit un certain nombre de scénarios où l'ICN a un sens.

N'attendez pas la description d'un protocole précis, c'est bien trop tôt. Ce RFC sort de l'IRTF, pas de l'IETF et est donc dans une démarche plus recherche qu'ingénierie. Il liste des cas, étudie une bibliographie mais ne donne pas encore de solutions.

Donc, d'abord, une introduction à l'ICN (section 1 du RFC). Ses promoteurs aiment expliquer que l'Internet actuel est "*host-centric*", c'est-à-dire construit autour de l'"*host*", l'ordinateur ou le routeur connecté à l'Internet et identifié par une ou plusieurs adresses IP. Ces machines, ordinateurs et routeurs, sont connectées plus ou moins en permanence et, dans le modèle original, ont une connectivité de bout en bout (tout le monde peut parler à tout le monde). Le modèle ICN repose sur un changement important de paradigme : ce qui compte, c'est l'information, c'est elle qui a un identificateur, les machines ne sont pas forcément connectées tout le temps, ce qui compte, c'est qu'on puisse récupérer l'information, et peu importe si elle vient d'un serveur précis, ou d'un cache ou d'autres équipements dans le réseau. Le réseau ne transporte donc pas seulement des bits, il peut aussi servir de mécanisme de distribution. Les partisans de l'ICN utilisent souvent le terme de "*location independance*" <<https://www.bortzmeyer.org/hostname-physical-location.html>>.

Naturellement, cela change la vision qu'on a de la sécurité. Aujourd'hui, si je télécharge le rapport ANSSI « Comprendre et anticiper les attaques DDoS » <<https://www.ssi.gouv.fr/uploads/>

2015/03/NP_Guide_DDoS.pdf> », ma confiance dans l'intégrité et l'authenticité de ce document vient du serveur où je l'ai récupéré (ici, un serveur de l'ANSSI) et des mécanismes techniques qui empêchent une modification en cours (comme HTTPS). En ICN, le serveur n'a plus d'importance, il faut donc bâtir la confiance sur autre chose, qui doit voyager avec l'information elle-même, comme par exemple des signatures. (Vous noterez que, contrairement à ce que prétendent les promoteurs de l'ICN, qui ne connaissent en général que HTTP, il existe plusieurs protocoles Internet qui fonctionnent ainsi depuis longtemps. Ainsi, pour le DNS, il est rare qu'on récupère l'information depuis le serveur original - qui peut être non-joignable, on passe par une série de caches, et la sécurité demande donc un mécanisme de signature, DNSSEC.)

Il existe plusieurs travaux de recherche autour de cette idée. C'est le cas de NetInf <<http://www.netinfo.org/>>, CCN <<https://www.parc.com/work/focus-area/content-centric-networking/>> (sur lequel j'ai déjà écrit <<https://www.bortzmeyer.org/van-jacobson-ccn.html>>), NDN <<http://named-data.net/>> et bien d'autres. Comme le note notre RFC, c'est l'équivalent du bouillonnement que connaissait le réseau de la fin des années 70 à la fin des années 80 quand DECnet, IPX et quelques autres concurrençaient sérieusement IP, avant que ce dernier ne s'impose.

Un travail est en cours, au sein du groupe de recherche ICN <<https://irtf.org/icnrg>>, pour évaluer les différentes façons de faire de l'ICN qui existent. (Voir le RFC 8569¹.)

Donc, en section 2, le cœur du RFC, les scénarios envisagés. Logiquement, on commence par les usages à la mode avec les réseaux sociaux. Un gros réseau social comme Facebook doit distribuer énormément d'informations à beaucoup de gens, situés dans des endroits très divers. Le service est donc une combinaison de "multicast" et de cache. Le contenu est disséminé sur plein de serveurs différents et le modèle ICN convient donc particulièrement bien puisque le client veut récupérer « les photos de Mère-Grand » et pas se connecter à telle ou telle machine (cf. Mathieu, B. et al., « *Information-centric networking : a natural design for social network applications* » <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6231278>> » en 2012). Dans le monde merveilleux de l'ICN, un réseau social comme Twitter pourrait être bâti très simplement : un utilisateur publie son tweet, certains utilisateurs le recevront tout de suite (ceux qui sont connectés), les autres le récupéreront à partir d'un cache dans le réseau dès qu'ils seront connectés. L'article de Kim, J. et al. « *Content Centric Network-based Virtual Private Community* » <<http://ieeexplore.ieee.org/xpl/articleDetails?arnumber=5722896>> » de 2011 décrit plus en détail cette vision, que le RFC qualifie, de manière franchement baratineuse, de "serverless" (il y a bien des serveurs, même s'ils ne sont pas explicitement nommés).

Autre scénario d'usage possible, la communication en temps-réel, avec multimédia, par exemple des services comme Skype. Cela inclut des communications entre deux personnes, des conférences d'un groupe de personnes, des ressources partagées comme un tableau virtuel où tout le monde peut écrire, etc. Le tout est soumis à des contraintes de performance assez strictes. Contrairement au scénario précédent, celui-ci semble aux antipodes de l'ICN, puisqu'il s'agit d'échanger et non plus d'« accéder à du contenu ». Et, pourtant, il y a déjà eu des recherches à ce sujet (j'en ai mentionné une dans mon article sur le CCN <<https://www.bortzmeyer.org/van-jacobson-ccn.html>>).

Le RFC cite également un exemple de MMORPG fait en ICN, par Chen, J. et al., « *G-COPSS : A Content Centric Communication Infrastructure for Gaming Applications* » <https://www.net.informatik.uni-goettingen.de/publications/1810/G-COPS_ICDCS2012.pdf> » en 2012.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc8569.txt>

Peut-on utiliser l'ICN pour le partage d'infrastructures, comme la capacité disque ou réseau? Le RFC dit que oui, en se basant sur le fait que l'ICN sécurise des objets et pas des canaux de communication et permet donc le partage de ressources. Par exemple, toute machine avec suffisamment de place disque peut devenir un cache. Nul besoin de lui faire confiance, puisque les objets seront sécurisés (par exemple par le biais d'une signature qui leur sera attachée). Cette possibilité est creusée dans l'article de Jacobson, V. et al., « *"Custodian-Based Information Sharing"* <<http://named-data.net/wp-content/uploads/06231277.pdf>> » en 2012 mais aussi dans le cas de trafic HTTP classique (Psaras, I. et al., « *"Modelling and Evaluation of CCN- Caching Trees"* <http://discovery.ucl.ac.uk/1341599/1/1341599_ccn-caching-networking11-ipsaras.pdf> », 2011) ou de trafic BitTorrent (Tyson, G. et al., « *"A Trace-Driven Analysis of Caching in Content-Centric Networks"* <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6289181>> », 2012). Les CDN actuels font déjà un travail analogue, mais au prix d'une configuration manuelle plus importante.

Si, arrivé à ce stade, vous pensez qu'il y a beaucoup de promesses et peu de réalisations concrètes, je dirais que vous avez tout à fait raison. L'ICN, c'est beaucoup de marketing, faisant miroiter d'extraordinaires changements positifs et se souciant peu des vulgaires problèmes pratiques. Il y a toutefois des applications plus raisonnables comme la distribution de contenu, certes le scénario le plus réaliste puisque l'ICN voit tout comme du contenu. Il est donc logique que ce soit le domaine où l'ICN ait été le plus étudié, voyez le RFC (section 2.5) pour une bonne bibliographie.

Autre exemple marketing et spectaculaire, les réseaux dans les véhicules en mouvement (les VANET). Aussi bien la communication entre véhicules, que celle entre le véhicule et des stations fixes, posent des problèmes très difficiles, et sont un vrai défi pour les concepteurs de réseaux. La topologie change très vite (à la vitesse du véhicule), et la connectivité n'est pas permanente. On est dans un cas adapté à l'ICN où on se soucie moins de qui a envoyé l'information, que de l'information elle-même. Les caches et la réplication de l'information seront certainement utiles ici.

Vous trouvez ce scénario trop terrestre? Plus en altitude, il y a les réseaux spatiaux, voire interplanétaires. Le concept de DTN (*"Delay-Tolerant Networking"*) était né autour des voyages dans l'espace mais peut s'appliquer à d'autres secteurs (comme les sous-marins ou comme des réseaux très perturbés, par exemple suite à une catastrophe naturelle). Comme son nom l'indique, le DTN est conçu pour des cas où le délai d'acheminement d'un bit d'une machine à une autre est tel que les protocoles de transport traditionnels comme TCP ne fonctionneront jamais. (On ne pourra jamais terminer une « triple poignée de mains » dans les temps, par exemple.) En plus du délai d'acheminement, le monde du DTN se caractérise par le fait que la connectivité est l'exception, plutôt que la règle. Pour travailler dans ces conditions, les systèmes DTN (RFC 4838) retrouvent l'architecture « *"store and forward"* » des anciens protocoles comme UUCP. Tout routeur est également un serveur de stockage et un cache, et garde l'information au moins jusqu'à ce qu'il ait pu la transmettre à quelqu'un d'autre. Le stockage se fait donc en partie dans le réseau. Ce paradigme peut même s'étendre au cas où certaines des communications se font par un mécanisme de « mule de données », de transport d'un mécanisme de stockage physique (une clé USB, par exemple). Le DTN et ICN ont en commun qu'il n'imposent pas une connectivité permanente et ils pourraient donc, en théorie, bien s'entendre ensemble (Tyson, G., Bigham, J. et E. Bodanese, « *"Towards an Information-Centric Delay-Tolerant Network"* <<http://s3.amazonaws.com/academia.edu.documents/30476154/1295052003-libre.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1427162015&Signature=vTu2iqqKHwX8J0%2FKXa7APhHMJhI%3D>> », 2013). De plus, les deux paradigmes reposent sur le traitement de l'information comme étant un objet en soi (que le DTN nomme un *"bundle"*), à longue durée de vie.

Une variante de ce scénario où le contenu est distribué entre des machines à la connectivité intermittente est celle du « partage de données selon l'occasion » où des informations sont disponibles et seront récupérées si on passe à côté. L'exemple typique (mais non cité par le RFC...) est celui de la PirateBox, qui attend tranquillement d'éventuels visiteurs. Le RFC préfère un cas où les deux parties sont mobiles, et

où on distribue des tweets (Hossmann, T., et al. « *Twitter in disaster mode : smart probing for opportunistic peers* » <<http://dl.acm.org/citation.cfm?id=2159576.2159601>> », 2012).

Seconde variante, celle où on distribue de l'information pendant une crise majeure, ayant mené à une importante perturbation des réseaux, perturbation d'autant plus sérieuse que l'énergie manque elle aussi, stoppant ainsi les équipements réseau qui n'avaient pas été mis hors de service pendant la crise. Les DTN peuvent se montrer très utiles dans ce cas, par exemple pour distribuer efficacement l'information utile (comme les alertes, ou l'emplacement des camps de réfugiés). Comme avec tous les réseaux pair-à-pair se posera le problème de la motivation à participer : si le tremblement de terre a frappé et que je me retrouve à la rue avec un "smartphone" qui n'a plus qu'une heure de batterie, sans espoir de recharge proche, pourquoi est-ce que je consacrerai ces précieux watt-heures à router des messages pour les autres ? Peut-être faudra-t-il compter uniquement sur les équipements apportés par les premiers secours ?

On l'a vu, l'ICN repose sur beaucoup de "buzzwords" et de promesses aussi floues que mirobolantes. Il était donc inévitable que le super-buzzword Internet des objets soit mentionné dans ce RFC. On sait que les progrès du matériel font que de nombreux objets qui n'étaient pas vus comme des ordinateurs sont désormais connectés à un réseau et même parfois à l'Internet. Prenons les capteurs, par exemple : on peut envisager plein de capteurs dispersés un peu partout, récoltant plein d'informations utiles, auxquelles il faut ensuite accéder. Les idées de l'ICN comme le cache réparti peuvent être utiles ici (cf. l'Internet-Draft de Kutscher, D. et S. Farrell, « *Towards an Information-Centric Internet with more Things* » <<https://tools.ietf.org/html/draft-kutscher-icn-iwmt>> », présenté à l'atelier "Interconnecting Smart Objects with the Internet" de l'IAB, en 2011, atelier décrit dans le RFC 6574, ou bien les très nombreuses références citées dans la section 2.8).

Un des points difficiles de tout réseau informatique est le nommage, et le RFC cite ici les possibilités de nommer directement le contenu, ce que permettent par exemple les URI ni : du RFC 6920 (un exemple concret et utile). Par contre, le RFC ne cite pas les "magnets" mais il est vrai qu'ils n'ont jamais fait l'objet d'une standardisation rigoureuse.

Autre "buzzword" à ne pas manquer, la "smart city". Le RFC tourne ici franchement au discours marketing, avec des envolées sur l'interaction des "smart people" avec la "smart city", le tout sous l'égide d'une bienveillante "smart governance".

Assez de scénarios d'usage, la section 3 du RFC passe ensuite aux questions transversales, communes à plusieurs scénarios. Cette fois, il y a davantage de concret. Premier problème, évidemment, le fric. La section 3.1 se penche sur l'économie de la connectivité. Par exemple, les engins connectés auront de plus en plus souvent le choix entre plusieurs connexions (pensez aux "smartphones" actuels, avec leur choix entre WiFi et 3G) et intégreront probablement des considérations financières dans leur choix. L'accès aux données indépendamment d'un serveur précis est ici utile, pour laisser davantage de liberté de choix aux engins connectés. À condition, comme toujours en économie, que les acteurs rationnels et égoïstes d'Adam Smith soient parfaitement informés des conséquences de leur choix. Le RFC note donc qu'il faudra un travail en ce sens, avec des offres simples et claires, et distribuées automatiquement aux engins qui auront à faire des choix de connectivité.

Ces questions de sous dans les réseaux ne concernent évidemment pas que la machine terminale mais également les opérateurs. L'ICN leur offre de nouvelles possibilités (le déploiement de caches dans leur réseau, automatiquement utilisés par les clients, permet de diminuer les coûts liés à la connectivité externe). Là encore, le RFC cite énormément de références vers des travaux existants.

Autre sujet transversal qui a une grande importance (tout à fait justifiée) aujourd'hui, la consommation énergétique. Le problème est bien réel (les factures d'électricité augmentent, la pollution et les

autres conséquences pour l'environnement aussi, les petits objets n'ayant qu'une batterie à capacité limitée souhaitent faire des économies d'électricité) mais son traitement dans le RFC est très décevant : ce ne sont que des vagues promesses comme quoi l'ICN permettrait de diminuer la consommation électrique. L'idée est que l'utilisation massive de caches, le traitement des données dans le réseau lui-même et pas dans les extrémités, et enfin la non-obligation d'être connecté en permanence seraient plus efficace énergétiquement. Le RFC cite plusieurs études dont j'ai l'impression qu'elles sont purement théoriques, sans mesures effectives.

Autre question qui s'applique à plusieurs scénarios d'usage de l'ICN, le cas où des réseaux de différents types, n'utilisant pas tous IP, sont disponibles. Dans le modèle original, dit du sablier <<http://commons.wikimedia.org/wiki/File:Internet-hourglass.svg>>, IP formait la ceinture, le point de passage obligé (aujourd'hui, dans la réalité, c'est plutôt HTTP qui joue ce rôle). Dans un monde qui utilise massivement l'ICN, ce pourrait être ICN qui serait la ceinture, l'unificateur de toutes les interactions (Trossen, D. et al., « *Arguments for an information centric internetworking architecture* » <<http://www.sigcomm.org/node/2840>>, 2010). Les auteurs du RFC sont évidemment conscients du caractère très spéculatif de ce projet, et notent que bien des points restent à résoudre pour avancer dans ce sens, comme la sécurité (traitée plus en détail, par la suite, dans le RFC 7927).