

RFC 8793 : Information-Centric Networking (ICN): Content-Centric Networking (CCNx) and Named Data Networking (NDN) Terminology

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 18 juin 2020

Date de publication du RFC : Juin 2020

<https://www.bortzmeyer.org/8793.html>

Le terme d'ICN ("*Information-Centric Networking*") désigne un paradigme de communication où tout est du contenu : plus d'adresses IP ou de noms de domaine, on désigne chaque ressource par un nom. Il en existe plusieurs déclinaisons comme le NDN ("*Named Data Networking*") ou le CCNx ("*Content-Centric Networking*"). Ce nouveau RFC décrit le vocabulaire de ces deux déclinaisons.

La section 1 du RFC résume les concepts essentiels de l'ICN. Il s'agit d'une architecture de réseaux où le composant de base est un contenu auquel on veut accéder. Pour récupérer, mettons, une image de chaton, on donnerait le nom de cette image et elle arriverait sur votre écran. Dans cette architecture, la notion de machine disparaît, il n'y a que du contenu. En théorie, cela permettra de faire du "*multicast*", et de la mémorisation ("*caching*") de contenu par les éléments du réseau, transformant chaque routeur en serveur miroir.

Le concept se décline ensuite en plusieurs architectures, utilisant des terminologies très variables, et peu normalisées. Toutes ne sont pas couvertes dans ce RFC, qui se focalise sur les sujets sur lesquels l'IRTF a le plus travaillé, NDN et CCNx (RFC 8569¹.)

La section 2 de notre RFC contient une description générale de l'ICN. Il y a un **demandeur** ("*requestor*"), également appelé consommateur, qui voudrait du contenu, mettons une photo de Scarlett Johansson. Il va manifester un **intérêt** ("*Interest*") qui est un message demandant du contenu. Un ou plusieurs nœuds du réseau (l'équivalent des routeurs, mais la principale différence est qu'ils peuvent

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

stocker du contenu) vont alors faire suivre cet intérêt jusqu'à une machine qui se trouve avoir le contenu en question, soit parce qu'elle en est la source (le producteur), soit parce qu'elle en a une copie. La machine répondra alors avec des **données** ("*Data*") qui seront relayées en sens inverse, jusqu'au demandeur. Il y a donc deux types de messages, qui pourraient s'appeler requête et réponse mais que, dans le monde ICN, on nomme Intérêt et Données. Le message d'Intérêt contient le nom du contenu convoité. Chaque protocole particulier d'ICN a une façon différente de construire ces noms. Comme le message de Données arrive via plusieurs relais, l'authentification de ce message ne peut pas être déduit de sa source. Il faut qu'il soit signé, ou bien que le nom soit auto-certificateur, par exemple parce que le nom est un condensat du contenu. (Attention, l'authentification n'est pas la même chose que la confiance. Si un message est signé de la porte-parole du gouvernement, cela prouve qu'il vient bien d'elle, pas qu'elle dit la vérité. Voir la section 6 du RFC.)

Pour faire son travail, le nœud ICN va avoir besoin de trois bases de données :

- La FIB ("*Forwarding Interest Base*"), les moyens de joindre les données,
- La PIT ("*Pending Interest Table*"), les requêtes en cours,
- Le stockage ("*Content Store*"), puisque tout nœud ICN peut stocker les données qu'il voit passer, pour les transmettre la fois d'après.

Parfois, les données sont de grande taille. En ICN, on dit qu'on va **segmenter** ces données en plusieurs messages. Le RFC utilise le terme de **paquet** pour une unité de données après la segmentation. Ces paquets sont potentiellement plus gros que la MTU et on parle de **trame** ("*frame*") pour l'unité qui passe sur le réseau, après le découpage du paquet.

Une fois qu'on a ce résumé en tête, on peut passer aux termes utilisés dans l'ICN. C'est la section 3 du RFC, le gros morceau. Je ne vais pas les lister tous ici, je fais un choix arbitraire. Attention, comme il s'agit d'un domaine de recherche, et peu stabilisé, il y a beaucoup de synonymes. On a :

- **ICN** ("*Information-Centric Networking*"); c'est un concept dont le CCNx du RFC 8569 est une des déclinaisons possibles.
- L'**immuabilité** est une notion importante en ICN; si le nom d'un contenu est un condensat du contenu, l'objet est immuable, tout changement serait un nouvel objet, avec un nouveau nom.
- Le contenu est produit par des **producteurs** et récupéré par des **consommateurs**. Il sera relayé par des **transmetteurs** ("*forwarders*", un peu l'équivalent des routeurs.) Si le transmetteur stocke les données, pour accélérer les accès ultérieurs, on l'appelle une **mule**. La transmission est dite **vers l'amont** ("*upstream*") quand il s'agit des Intérêts et **vers l'aval** ("*downstream*") pour les Données et les rejets (lorsqu'un Intérêt ne correspond à aucune donnée connue).
- Un transmetteur peut être à **état**, ce qui signifie qu'il se souvient des Intérêts qu'il a transmis (dans sa PIT), et qu'il utilisera ce souvenir pour transmettre les Données qui seront envoyées en réponse.
- Un Intérêt peut ne pas être **satisfait** si on ne trouve aucune Donnée qui lui correspond (zut, pas de photo de Scarlet Johansson). Un rejet ("*Interest NACK*") est alors renvoyé au demandeur. Notez que beaucoup de schémas d'ICN ont des noms hiérarchiques et que, dans ce cas, la correspondance dans la FIB entre un Intérêt et une interface se fait sur le préfixe le plus long.
- Un des points importants de l'ICN est la possibilité pour les nœuds intermédiaires de mémoriser les données, pour faciliter leur distribution (stockage dans le réseau, ou "*In-network storage*"). Cette mémorisation ne se fait en général pas systématiquement mais uniquement quand l'occasion se présente ("*opportunistic caching*"). Cela améliore la distribution de contenus très populaires mais cela soulève des gros problèmes de vie privée, puisque tous les nœuds intermédiaires, les transmetteurs, doivent pouvoir voir ce qui vous intéresse. (Une photo de Scarlet Johansson, ça va. mais certains contenus peuvent être plus sensibles.)
- Comme IP, ICN sépare le **routing** (construire les bases comme la FIB) et la **transmission** (envoyer des messages en utilisant ces bases).
- Un **lien** est un paquet de données qui contient le nom d'un autre contenu. Une indirection, donc.

- Et le **nom**, au fait, concept central de l'ICN? Il est typiquement hiérarchique (plusieurs composants, parfois également appelés segments, je vous avais prévenu qu'il y avait beaucoup de synonymes) pour faciliter la transmission. Le début d'un nom se nomme un **préfixe**. Le nom peut indiquer la localisation d'un contenu ou bien peut être sémantique (indiquant le contenu). Dans ICN, on utilise en général le terme d'**identificateur** ("*packet ID*") pour les identificateurs cryptographiques, calculés par condensation. Si le nom est l'identificateur, on parle de nommage à plat. (Et, évidemment, cela sera plus dur à router, il faudra sans doute une DHT.)
- L'ICN a une notion de **version**. Dans le cas de noms hiérarchiques, la version d'un contenu peut être mise dans le nom.
- Avec ICN, les propriétés de sécurité du contenu ne peuvent pas être garanties par le serveur où on l'a récupéré. Il faut donc une sécurité des données et pas seulement du canal. L'intégrité est vérifiée par la cryptographie, l'authenticité par une signature et la confidentialité par le chiffrement. Tout cela est classique. Plus délicate est la question de la confidentialité des noms, puisqu'elle empêcherait les nœuds intermédiaires de savoir ce qui est demandé, et donc de faire leur travail.