

# RFC 5773 : Analysis of Inter-Domain Routing Requirements and History

Stéphane Bortzmeyer  
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 19 février 2010

Date de publication du RFC : Février 2010

<https://www.bortzmeyer.org/5773.html>

---

Ce RFC est essentiellement un document historique. Écrit il y a pas mal d'années, il est publié aujourd'hui (avec quelques mises à jour et annotations) dans le cadre des efforts de définition d'une future architecture de routage inter-domaine dans l'Internet.

Sur Internet, il existe en effet deux sortes de routage, l'**intra-domaine** qui concerne les opérations se déroulant à l'intérieur d'un unique système autonome, donc sous une direction commune, et l'**inter-domaine**, qui traite du routage **entre** systèmes autonomes (les « domaines »), lorsqu'il faut franchir des frontières administratives, ce qui est bien plus complexe que de simplement pousser des paquets. Aujourd'hui, l'essentiel du routage inter-domaine est fait avec le protocole BGP (RFC 4271<sup>1</sup>, et BGP avait été originellement conçu, en 1989, en réponse aux exigences du RFC 1126) et l'architecture sur laquelle il repose montre des sérieuses limites, notamment en matière de passage à l'échelle : pourrions-nous faire encore croître l'Internet, et à un coût raisonnable ? (La section 2 du RFC résume l'état actuel du routage inter-domaine et de ses limites. Elle est complétée sur ce dernier point par la section 5. Le routage intra-domaine est, lui, considéré comme globalement satisfaisant.)

Il existe en ce moment beaucoup d'activité à l'IETF et à l'IRTF autour de ce thème d'une future architecture de routage. Elle porte, par exemple, sur des idées comme la séparation de l'identificateur et du localisateur <<https://www.bortzmeyer.org/separation-identificateur-localisateur.html>>. Le groupe de l'IRTF concerné est le "*Routing Research Group*" <<http://www.irtf.org/charter?gtype=rg&group=rrg>> et ce groupe travaille à la définition d'un cahier des charges pour la future architecture. Parmi les travaux étudiés à cette occasion, figurent les discussions qui ont eu lieu il y a plusieurs années, à partir du RFC 1126, et qui sont désormais synthétisées dans ce RFC. (Attention, donc, en le lisant, une bonne partie a été écrite au tout début du siècle.)

---

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc4271.txt>

La section 1 résume la longue genèse de ce document, qui prend sa source dans les travaux d'un groupe informel nommé Babylon <<http://www.cdt.luth.se/babylon/>> en 2001. Le texte original a été préservé, des ajouts indiquent ce qui a pu changer dans l'Internet depuis.

La section 3 du RFC détaille chacune des exigences du RFC 1126, dans l'ordre du RFC original, en expliquant dans quelles mesures elles sont satisfaites aujourd'hui, avec un Internet radicalement différent de ce qu'il était en 1989, où sa taille était minuscule, selon les critères d'aujourd'hui, et où sa connectivité était encore très hiérarchique, avec NSFnet au sommet.

Ainsi, l'exigence indiquée en section 3.1 e) du RFC 1126, "*Provide paths sensitive to user policies*" est décrite dans notre RFC 5773 en section 3.1.2.1.5 comme toujours valide (selon le texte écrit en 2001 par Babylon) mais très imparfaitement faite ("*loose source routing*", QoS) et les notes de 2006 à nos jours ajoutent qu'il faut plutôt parler d'échec complet que de déploiement insuffisant.

Mais il y a aussi des succès, le principal étant évidemment que l'Internet marche. Des points qui semblaient primordiaux en 1989 et même encore en 2001 ont sombré dans une relative indifférence (comme la QoS, justement).

Les plus gros problèmes sont peut-être quantitatifs. Le RFC 1126, section 3.4 c) demandait innocemment que le futur (à l'époque) protocole permette 10 000 systèmes autonomes et 100 000 réseaux. Ces nombres ont été largement dépassés mais il n'est pas garanti, loin de là, que cette croissance pourra durer éternellement. En 2001, Babylon s'inquiétait pour l'épuisement de l'espace de numérotation des systèmes autonomes (16 bits à l'époque) et cette inquiétude se retrouve dans notre RFC en section 3.1.2.4.3. Le RFC a finalement une note disant que le déploiement du RFC 6793 est en train de résoudre ce problème mais que ce déploiement prend plus de temps que prévu.

Dans tout exercice d'ingénierie, le plus dur n'est pas en général de définir les buts (« Que ça marche bien! Que ça ne soit pas cher! Que n'importe qui puisse s'en servir! Que cela fasse le café! ») mais les « non-buts », ce qu'on renonce à obtenir car cela nous entrainerait trop loin et condamnerait le projet. Il est rare que les non-buts soient explicites, car cela focaliserait la critique. Mais le RFC 1126 avait une telle section, la numéro 4, analysée en section 3.1.3 de notre RFC 5773. Par exemple, le RFC 1126 expliquait que la connectivité de tous n'était pas un but. En effet, elle nécessite la coopération de systèmes autonomes intermédiaires, coopération qui ne peut pas être obtenue par des moyens techniques. Ce simple non-but déclenche une grande discussion dans le RFC 5773 en section 3.1.3.1. N'est-il pas contraire à la mission de connectivité totale de l'Internet? Le RFC 1126 n'était-il pas excessivement prudent car il avait été écrit dans un monde où IP n'était pas encore le protocole universel qu'il était devenu en 2001? (Et qu'il n'est plus depuis que IPv4 et IPv6 coexistent.) Faut-il chercher la connectivité universelle (qu'on n'a pas, même avec IPv4 seul, notamment à cause du NAT) ou le « routage universel »?

De même, la répartition de charges était considérée par le RFC 1126 comme un nom but, même si la section 3.1.3.3 du RFC 5773 fait observer que le désir de faire passer le trafic d'un domaine par plusieurs fournisseurs est une des causes de la désagrégation des préfixes annoncés, et donc de la croissance de la table de routage.

La section 3 remet les choses dans le contexte de l'époque. En 1989, lorsque le RFC 1126 a été écrit, la famille de protocoles OSI était encore considérée sérieusement par certains (elle sera abandonnée au début des années 1990, sans jamais avoir connu de déploiement significatif). Le développement de BGP s'est donc fait dans un contexte marqué par la présence d'un concurrent, IDRP (alias ISO 10747, section 3.2 de notre RFC). La section revient donc sur l'histoire tourmentée (et parfois contestée) de cette époque, marquée par l'émergence du concept de système autonome et par celle de l'idée de routage non-hiérarchique. Parmi les documents importants cités par le RFC, il y a, par exemple, "*Internet Architecture*

*Workshop : Future of the Internet System Architecture and TCP/IP Protocols* <<http://www.eecis.udel.edu/~mills/database/papers/inarc.pdf>> ou bien le chapitre 14 <[http://www.interisle.net/OSN/Chapter\\_14.pdf](http://www.interisle.net/OSN/Chapter_14.pdf)> du livre *"Open Systems Networking"* <<http://www.interisle.net/OSN/OSN.html>>. Le RFC considère que, si IDRP n'a jamais été réellement déployé, du moins certaines des idées qu'il contenait ont inspiré les développements dans le monde Internet. (Beaucoup d'autres ont été abandonnées : pensez au chapitre sur les non-buts. Comme tous les protocoles OSI, IDRP ne pouvait pas résister à la conception en comité, où toute fonction demandée était forcément incluse, de peur de fâcher quelqu'un.) D'autres idées d'IDRP, comme l'utilisation de certificats X.509 pour signer les annonces, n'ont pas encore percé, bien qu'elles soient régulièrement évoquées pour BGP.

BGP a donc suivi son bonhomme de chemin, première version dans le RFC 1105 en juin 1989, deuxième dans le RFC 1163, en juin 1990, troisième dans le RFC 1267 publié en octobre 1991 et enfin quatrième dans le RFC 1771 en mars 1995 (BGP-4 est désormais normalisé dans le RFC 4271). IDRP est, lui, bien oublié, il n'a même pas d'article dans Wikipédia .

Parmi les autres efforts pour développer un mécanisme de routage inter-domaine, une place particulière doit être faite à Nimrod <<http://ana-3.lcs.mit.edu/~jnc/nimrod/>> (RFC 1753 et RFC 1992, section 3.3 de notre RFC). Le projet Nimrod, de 1994 à 1996, visait à créer une architecture de routage complètement nouvelle. S'il n'a pas débouché, les idées explorées à ce moment ont néanmoins beaucoup influencé les recherches ultérieures. Par exemple, Nimrod, contrairement à pas mal de projets « table rase » qui croient naïvement qu'on les laissera tout détruire et repartir de zéro, mettait explicitement au premier plan la nécessité d'être **déployable progressivement**, c'est-à-dire de ne pas rester bloqué dans le dilemme de l'œuf et de la poule, où personne n'adopte le nouveau système car il n'y a aucun intérêt à le faire, tant qu'on reste tout seul. Cette exigence de déploiement progressif reste essentielle aujourd'hui : l'Internet n'est plus un jouet de chercheurs, on ne peut plus l'arrêter pour voir, ni imposer un changement d'un seul coup, comme l'avait été l'abandon de NCP.

À noter que l'architecture de Nimrod faisait partie des projets concurrents pour le système « IPng », qui devait prendre la suite d'IPv4. Trop ambitieux, Nimrod avait été rejeté au profit du futur IPv6 qui se limitait au format des paquets IP et ne tentait pas de réformer l'architecture de routage inter-domaine (qui reste donc la même qu'avec IPv4).

Si Nimrod est relativement connu des gens de l'IETF, PNNI, résumé en section 3.4, l'est beaucoup moins. Il venait des travaux de l'"ATM forum" et n'avait guère eu de succès, peut-être parce que trop lié à une architecture vite dépassée, ATM.

Le travail des chercheurs sur le routage interdomaine ne s'est jamais arrêté. La section 4 est donc consacrée aux travaux récents en ce domaine. Ces recherches sur un objet très mouvant, le gigantesque Internet d'aujourd'hui, doivent s'adapter sans cesse. Ainsi, rappelle notre RFC, la connexion d'un site à plusieurs FAI devient de plus en plus fréquente et il est urgent de trouver un mécanisme qui permette de la faire dans de bonnes conditions.

Parmi les travaux de recherche des dernières années, le RFC cite NewArch <<http://www.isi.edu/newarch/>> (section 4.2). Datant de 2000-2001, financé par une agence gouvernementale états-unienne, NewArch avait, dans son cahier des charges, des préoccupations inquiétantes comme la protection des rapaces détenteurs de « propriété intellectuelle » ou comme la nécessité de développer des systèmes de surveillance plus perfectionnés.

Sans doute trop récents, puisque l'essentiel du RFC 5773 avait été fait avant 2006, les projets comme Clean Slate et GENI qui, pour l'instant, ont surtout produit du vent, ne sont pas mentionnés.

Quel est l'état de la réflexion sur les limites et défauts du modèle actuel de routage inter-domaine ? La section 5 répond à cette question, dans la suite du RFC 3221. Parmi les nombreux problèmes identifiés dans cette section, la propagation mondiale des erreurs locales (section 5.3), magnifiquement illustrée par l'attaque involontaire contre YouTube <<https://www.bortzmeyer.org/pakistan-pirate-youtube.html>> (depuis, une solution a été normalisée <<https://www.bortzmeyer.org/securite-routage-bgp-rpk.html>>, mais pas encore massivement déployée), l'absence de solution satisfaisante à la forte demande des utilisateurs pour les connexions à plusieurs FAI (section 5.4 et RFC 4116), la question de la sécurité, puisque n'importe quel routeur BGP peut annoncer n'importe quelle route (section 5.11, RFC 4593 pour le cahier des charges des efforts actuels visant à normaliser une solution et, pour un exemple récent d'alerte de sécurité, la faille Kapela & Pilosov <<https://www.bortzmeyer.org/faille-bgp-2008.html>>).

Comme toutes les listes de problèmes, celle-ci peut donner l'impression que BGP est fichu et on peut se demander, en la lisant, pourquoi l'Internet marche encore. C'est parce que les faiblesses de BGP, notamment la propagation mondiale de n'importe quelle annonce, même fausse, même mensongère, sont aussi ses forces. Elles permettent à tous de détecter le problème et de réagir rapidement <<https://www.bortzmeyer.org/securite-bgp-et-reaction-rapide.html>>. C'est bien à tort que le RFC prétend, dans sa section 5.14, qu'il n'existe pas d'outils de détection en temps réel des changements BGP, il y en a au contraire une pléthore <<https://www.bortzmeyer.org/alarmes-as.html>>!