RFC 8950 : Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 20 novembre 2020

Date de publication du RFC : Novembre 2020

https://www.bortzmeyer.org/8950.html

Le protocole de routage BGP annonce des préfixes qu'on sait joindre, avec l'adresse IP du premier routeur à qui envoyer les paquets pour ce préfixe. (Ce routeur est appelé le "*next hop*".) BGP a une extension, BGP multi-protocoles (RFC 4760 [1]) où les préfixes annoncés (NLRI pour "*Network Layer Reachability Information*") ne sont plus forcément de la même famille que les adresses utilisées dans la session BGP. On peut donc annoncer des préfixes IPv6 sur une session BGP établie en IPv4 et réciproquement. Notre RFC, qui succède au RFC 5549 avec quelques petits changements, étend encore cette possibilité en permettant que le "*next hop*" ait une adresse de version différente de celle du préfixe.

Normalement, BGP multi-protocoles (RFC 4760) impose la version (IPv4 ou IPv6) du "*next hop*" via l'AFI ("*Address Family Identifier*") et le SAFI ("*Subsequent Address Family Identifier*") indiqués dans l'annonce (cf. la liste actuelle des AFI possible <https://www.iana.org/assignments/address-family-numbers/address-family-numbers.xml#address-family-numbers-2> et celle des SAFI <https://www.iana.org/assignments/safi-namespace/safi-namespace.xml#safi-namespace-2>). Ainsi, un AFI de 1 (IPv4) couplé avec un SAFI valant 1 ("*unicast*"), lors de l'annonce d'un préfixe IPv4, impose que l'adresse du routeur suivant soit en IPv4. Désormais, cette règle est plus libérale, le routeur suivant peut avoir une adresse IPv6. Cela peut faciliter, par exemple, la vie des opérateurs qui, en interne, connectent des ilots IPv4 au-dessus d'un cœur de réseau IPv6 (cf. RFC 4925). Notez que cela ne règle que la question de l'annonce BGP. Il reste encore à router un préfixe IPv4 via une adresse IPv6 mais ce n'est plus l'affaire de BGP.

Il y avait déjà des exceptions à la règle comme quoi le préfixe et l'adresse du routeur suivant étaient de la même famille. Ainsi, le RFC 6074 permettait cela pour le couple AFI 25 (L2VPN) / SAFI 65 (VPLS).

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc4760.txt

Mais le couple AFI 2 / SAFI 1 (IPv6 / "*unicast*") ne permet pas de telles exceptions (RFC 2545). Une astuce (RFC 4798 et RFC 4659) permet de s'en tirer en encodant l'adresse IPv4 du "*next hop*" dans une adresse IPv6. (Oui, `::192.0.2.66` est une adresse IPv4 encodée dans les seize octets d'IPv6, cf. RFC 4291, section 2.5.5.2.) Quant au cas inverse (AFI IPv4, routeur suivant en IPv6), elle fait l'objet de notre RFC.

Lorsque l'adresse du "*next hop*" n'est pas de la famille du préfixe, il faut trouver la famille, ce qui peut se faire par la taille de l'adresse du "*next hop*" (quatre octets, c'est de l'IPv4, seize octets, de l'IPv6). C'est ce que propose le RFC 4684.

L'extension permettant la liberté d'avoir des "*next hop*" dans une famille différente du préfixe est spécifiée complètement en section 4. Elle liste les couples AFI / SAFI pour lesquels on est autorisé à avoir un "*next hop*" IPv6 alors que le préfixe est en IPv4. Le routeur BGP qui reçoit ces annonces doit utiliser la longueur de l'adresse pour trouver tout seul si le "*next hop*" est IPv4 ou IPv6 (la méthode des RFC 4684 et RFC 6074).

L'utilisation de cette liberté nécessite de l'annoncer à son pair BGP, pour ne pas surprendre des routeurs BGP anciens. Cela se fait avec les capacités du RFC 5492. La capacité se nomme "*Extended Next Hop Encoding*" et a le code 5 <`https://www.iana.org/assignments/capability-codes/capability-codes.xml#capability-codes-2`>. Cette capacité est restreinte à certains couples AFI / SAFI, listés dans l'annonce de la capacité. Par exemple, le routeur qui veut annoncer une adresse IPv6 comme "*next hop*" pour de l'"*unicast*" IPv4 va indiquer dans le champ Valeur de la capacité 1 / 1 (le couple AFI /SAFI) et le "*next hop AFI*" 2.

La section 2 du RFC résume les changements depuis le RFC 5549, que notre RFC remplace. L'encodage de l'adresse du "*next hop*" change dans le cas des VPN, et, pour les VPN MPLS, extension de l'exception au "*multicast*". Bref, rien de bien crucial.