

RFC 9107 : BGP Optimal Route Reflection (BGP ORR)

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 13 septembre 2021

Date de publication du RFC : Août 2021

<https://www.bortzmeyer.org/9107.html>

Les grands AS ont tellement de routeurs BGP qu'ils ne peuvent pas connecter chaque routeur à tous les autres. On utilise alors souvent des réflecteurs de route, un petit nombre de machines parlant BGP auxquelles tout le monde se connecte, et qui redistribuent les routes externes apprises. Mais une machine BGP ne redistribue que les routes qu'elle utiliserait elle-même. Or, le réflecteur risque de faire un choix en fonction de sa position dans le réseau, qui n'est pas la même que celle du routeur « client ». Les routeurs risquent donc d'apprendre du réflecteur des routes sous-optimales (la route optimale étant typiquement celle qui amène à la sortie le plus vite possible, en application de la méthode de la patate chaude). Ce RFC définit une extension de BGP qui va permettre de sélectionner des routes spécifiques à un client, ou à un groupe de clients.

Un petit rappel : un réflecteur de routes (*"route reflector"*) fonctionne sur iBGP (*"Internal BGP"*), à l'intérieur d'un AS, alors que les serveurs de routes (*"route server"*) font de l'eBGP (*"External BGP"*), par exemple sur un point d'échange. Ces réflecteurs sont décrits dans le RFC 4456¹. Ils ne sont pas la seule méthode pour distribuer l'information sur les routes externes à l'intérieur d'un grand AS, mais c'est quand même la solution la plus fréquente.

Le RFC 4456 notait déjà que, vu les coûts attribués aux liens internes à l'AS, le réflecteur ne choisirait pas forcément les mêmes routes que si on avait utilisé un maillage complet. Le routage « de la patate chaude » (qui consiste à essayer de faire sortir le paquet de son réseau le plus vite possible, pour que ce soit un autre qui l'achemine) risque donc de ne pas aussi bien marcher : le point de sortie lorsqu'on utilise le réflecteur sera peut-être plus éloigné que si on avait le maillage complet, surtout si le réflecteur est situé en dehors du chemin habituel des paquets et n'a donc pas la même vision que les routeurs « normaux ». Or, c'est un cas fréquent. Le réflecteur choisira alors des routes qui sont optimales pour lui, mais qui ne le sont pas pour ces routeurs « normaux ».

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc4456.txt>

La solution ? Permettre à l'administrateur réseaux de définir une localisation virtuelle pour le réflecteur, à partir de laquelle il fera ses calculs et choisira ses routes, au lieu d'utiliser sa localisation physique. Cette localisation virtuelle sera une adresse IP. Le réflecteur peut avoir plusieurs de ces localisations virtuelles, adaptées à des publics différents. Bref, le texte du RFC 4271 qui concerne la sélection de la meilleure route (section 9.1.2.2 du RFC 4271) est modifié pour remplacer « en fonction du saut suivant » par « en fonction de l'adresse IP configurée dans le réflecteur ».

Pour que cela marche, il faut que le réflecteur ait une vue complète du réseau, pour pouvoir calculer des coûts à partir de n'importe quel point du réseau. C'est possible avec les IGP à état des liens comme OSPF, ou bien avec BGP-LS (RFC 9552).

Et si le réflecteur a plusieurs clients ayant des desiderata différents, par exemple parce qu'ils sont situés à des endroits différents ? Dans ce cas, il doit faire tourner plusieurs processus de décision, chacun configuré avec une localisation virtuelle différente.

Les principales marques de routeurs mettent déjà en œuvre ce RFC, comme on peut le voir sur la liste des implémentations <<https://trac.ietf.org/trac/idr/wiki/draft-ietf-idr-bgp-optimal-route-20implementations>>. Du côté des logiciels qui ne sont pas forcément installés sur des routeurs, il semble que BIRD <<https://bird.network.cz/>> ne sache pas encore faire comme décrit dans ce RFC.