RFC 9687 : Border Gateway Protocol 4 (BGP-4) Send Hold Timer

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 8 novembre 2024

Date de publication du RFC : Novembre 2024

https://www.bortzmeyer.org/9687.html

Que doit faire un routeur BGP lorsque le pair en face ne traite manifestement plus ses messages ? Ce n'était pas précisé avant mais la réponse est évidente : raccrocher (mettre fin à la communication).

Un problème classique lors d'une connexion réseau, par exemple sur TCP, est de détecter si la machine en face est toujours là. Par défaut, TCP ne fournit pas ce service : s'il n'y a aucun trafic, vous ne pouvez pas savoir si votre partenaire est mort ou simplement s'il n'a rien à dire. Une coupure de réseau, par exemple, ne sera pas détectée tant que vous n'avez pas de trafic à transmettre (avec attente d'une réponse). Et BGP ne transmet que les changements donc l'absence de trafic ne signale pas forcément un problème. Il existe des solutions, comme d'envoyer périodiquement des messages même quand on n'a rien à dire (RFC 4271 [1], section 4.4), mais aucune n'est parfaite : un programme qui utilise TCP ne sait typiquement pas immédiatement si ses messages sont vraiment partis (et l'alarme actuelle ne couvre que la réception des messages, pas leur envoi). Et BGP n'a pas de fonction « ping », qui exigerait une réponse.

Quand la coupure est franche et détectée, aucun problème, la session BGP (RFC 4271) s'arrête et les routes correspondantes sont retirées de la table de routage. Mais ce RFC traite le cas de où le routeur BGP d'en face a un problème mais qu'on ne détecte pas. Un exemple : si ce routeur en face a complètement fermé sa fenêtre TCP de réception (RFC 9293, notamment la section 3.8.6), on ne pourra pas lui envoyer de messages, mais la session BGP ne sera pas coupée et les paquets continueront à être transmis selon des annonces de routage dépassées, alors qu'ils finiront peut-être dans un trou noir (le problème des « zombies BGP <https://labs.ripe.net/author/romain_fontugne/bgp-zombies/> »).

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 solution (section 3 de notre RFC) est de modifier l'automate de BGP (RFC 4271, section 8), en ajoutant une alarme (RFC 4271, section 10), `SendHoldTimer`. Quand elle expire, on coupe la connexion TCP et on retire les routes qu'avait annoncé le pair dont on n'a plus de nouvelles. Le RFC recommande une configuration par défaut de huit minutes de patience avant de déclencher l'alarme.

L'erreur « "*Send Hold Timer Expired*" » est désormais dans le registre IANA des erreurs BGP <https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xml#bgp-parameters-3> et tcpdump sait l'afficher <https://github.com/the-tcpdump-group/tcpdump/pull/1134>.
Il existe plusieurs mises en œuvre de ce RFC :

— Celle d'OpenBGPD (son source <https://github.com/openbsd/src/commit/ab9b1ccdd3ae2f6b9b7
— Celle de FRRouting (son source <https://github.com/FRRouting/frr/pull/11225>),
— Celle de BIRD (son source <https://gitlab.nic.cz/labs/bird/-/commit/bcf2327425d4dd96f381

Si les processus IETF vous passionnent, il y a une documentation des discussions <https://datatracker.ietf.org/doc/draft-ietf-idr-bgp-sendholdtimer/shepherdwriteup/> autour de ce RFC.