RFC 7196 : Making Route Flap Damping Usable

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 17 mai 2014

Date de publication du RFC : Mai 2014

https://www.bortzmeyer.org/7196.html

L'**amortissement** des annonces de routes lorsque ces routes sont instables (annoncées et retirées fréquemment) est une vieille technique pour limiter la charge sur les routeurs. Le principe est d'ignorer (pendant un certain temps) une route s'il y a eu trop de changements dans une période récente, de façon à éviter que le routeur BGP ne passe tout son temps à gérer une route qui part et revient en permanence. Le terme officiel est "*Route Flap Damping*" (RFD, on dit aussi "*dampening*"). L'idée est bonne mais l'expérience a montré que l'amortissement pénalisait excessivement les sites très richement connectés : plus on a de connexions, plus il y a des changements sur ses préfixes et plus on sera amorti. Résultat, ces dernières années, un certain nombre d'opérateurs ont préféré couper l'amortissement. Ce nouveau RFC propose de le rétablir, mais avec des nouveaux paramètres quantitatifs, qui devraient, cette fois, ne pénaliser réellement que les routes qui déconnent et pas les sites qui sont simplement très connectés.

L'amortissement a été formellement décrit dans ripe-178 <http://www.ripe.net/ripe/docs/ripe-178> puis dans le RFC 2439[1]. Le problème des faux positifs a été décrit en 2002, dans « "*Route Flap Damping Excacerbates Internet Routing Convergence*" <http://conferences.sigcomm.org/sigcomm/2002/papers/routedampening.pdf> ». Cela a mené à des révisions des politiques des opérateurs, allant dans le sens de **déconseiller** l'amortissement, comme raconté dans ripe-378 <http://www.ripe.net/ripe/docs/ripe-378>, qui dit que le remède (le RFD, l'amortissement) est pire que le mal et conclut « "*the application of flap damping in ISP networks is NOT recommended*" ». Des nouvelles études (comme « "*Route Flap Damping Made Usable*" <http://pam2011.gatech.edu/papers/pam2011--Pelsser.pdf> », par les auteurs du RFC) ont mené à une approche intermédiaire : garder l'amortissement, mais avec de nouveaux paramètres, comme déjà recommandé dans ripe-580 <http://www.ripe.net/ripe/docs/ripe-580> (document RIPE très proche de ce RFC), qui annule le ripe-378. Il existe aussi un "*Internet-Draft*" contenant le résultat d'une étude faite auprès des opérateurs sur leurs pratiques, le document draft-shishio-grow-isp-rfd-implement-survey.

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc2439.txt

En effet, un tout petit nombre de préfixes d'adresses IP est responsable de la majorité du travail (le "*churn*") des routeurs BGP, comme indiqué dans « "*BGP Extreme Routing Noise*" <http://meetings.ripe.net/ripe-52/presentations/ripe52-plenary-bgp-review.pdf> » ou bien dans l'article "*Route Flap Damping Made Usable*" cité plus haut. Cet article a testé des annonces BGP réelles pendant une semaine et note que 3 % des préfixes font 36 % des messages BGP. Ce sont ces préfixes qu'il faut pénaliser par l'amortissement, en épargnant les 97 % restants.

Quels sont les paramètres d'amortissement sur lesquels on peut jouer ? La section 3 les rappelle dans un tableau. Certains sont modifiables par l'administrateur du routeur, d'autres pas. Et le tableau, qui liste les valeurs par défaut existant chez Cisco et chez Juniper, montre qu'il n'y a pas de consensus sur ces valeurs par défaut. Attention, en lisant le tableau. Il n'existe malheureusement pas de vocabulaire unique pour ces paramètres (malgré la section 4.2 du RFC 2439) et le nom varie d'un document à l'autre. Ainsi, le RFC 2439 nomme "*cutoff threshold*" ce que ripe-580 et notre RFC nomment "*suppress threshold*". C'est dommage, c'est le paramètre le plus important : c'est le nombre de retraits ("*WITHDRAW*" BGP) de routes après lequel on supprime la route (multiplié par un facteur, la pénalité). Il vaut 2 000 par défaut sur IOS et 3 000 sur JunOS, ce qui est trop bas.

La section 4 de notre RFC cite en effet l'article "*Route Flap Damping Made Usable*" mentionné plus haut, qui estime qu'un seuil remonté à 6 000 permettrait de réduire le rythme de changement BGP de 19 %, contre 51 % avec un seuil de 2 000, mais en impactant dix fois moins de préfixes, donc en faisant beaucoup moins de victimes collatérales. Monter le seuil à 12 000 supprimerait presque complètement l'amortissement, très peu de préfixes étant à ce point instables.

Notre RFC recommande donc :

— De ne pas changer les valeurs par défaut dans les systèmes d'exploitation de routeurs, même si elles sont mauvaises, car cela changerait la sémantique de certaines configurations actuelles, qui comptent sur ces valeurs par défaut,
— Remonter le "*suppress threshold*" à 6 000 dans les configurations actives,
— Si possible, ajouter un mode de test ("*dry run*") aux routeurs où les calculs d'amortissement seraient faits mais pas appliqués aux routes. Cela permettrait aux opérateurs de tester les paramètres qu'ils envisagent.

Ces recommandations, notamment la valeur du seuil de déclenchement de l'amortissement ("*suppress threshold*") permettraient d'utiliser l'amortissement sans gros inconvénients.

Ah, un petit mot sur la sécurité (section 7) : un attaquent peut générer des faux retraits pour déclencher l'amortissement et mener à des dégâts supérieurs à ce qu'il aurait pu faire directement. Les paramètres plus conservateurs de ce nouveau RFC devraient limiter ce risque.

Pour configurer l'amortissement sur IOS, voir la documentation officielle <http://www.cisco.com/c/en/us/td/docs/ios/12_2/ip/configuration/guide/fipr_c/1cfbgp.html#wp1002395>, pour Quagga, c'est très semblable <http://www.nongnu.org/quagga/docs/docs-multi/BGP-route-flap-html>, et pour JunOS, voir leur documentation <http://www.juniper.net/techpubs/en_US/junos12.2/topics/topic-map/bgp-flap-damping.html>. Sur un routeur IOS, une route supprimée par l'amortissement sera affichée ainsi (voyez notamment la dernière ligne) :

```
R2# sh ip bgp
BGP table version is 12, local router ID is 192.168.0.1
Status codes: s suppressed, d damped, h history, * valid, > best, i { internal,
r RIB-failure, S Stale
Origin codes: i { IGP, e { EGP, ? { incomplete

Network               Next Hop  Metric  LocPrf  Weight  Path
d 12.12.12.12/32    192.168.0.2       0               0  2 i
*> 13.13.13.13/32   192.168.0.2       0               0  2 i
```

```
R2# sh ip bgp 12.12.12.12
BGP routing table entry for 12.12.12.12/32, version 12
Paths: (1 available, no best path)
Flag: 0x820
Not advertised to any peer
2, (suppressed due to dampening) (history entry)
192.168.0.2 from 192.168.0.2 (13.13.13.13)
Origin IGP, metric 0, localpref 100, external
Dampinfo: penalty 3018, flapped 4 times in 00:04:11, reuse in 00:03:20
```