

RFC 2923 : TCP Problems with Path MTU Discovery

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 16 Janvier 2008

Date de publication du RFC : Septembre 2000

<http://www.bortzmeyer.org/2923.html>

Au moment de sa normalisation, dans le RFC 1191¹, la découverte de la MTU du chemin semblait une excellente idée, en permettant d'adapter la taille des paquets IP de façon à minimiser la fragmentation. Mais plusieurs problèmes sont surgis à l'usage, notre RFC faisant la liste de ceux qui affectent plus particulièrement TCP.

Pour chaque problème, notre RFC donne un nom, décrit la question, fournit des traces tcpdump et suggère des pistes pour la résolution. Le plus fréquent des problèmes n'est pas spécifique à TCP, c'est le **trou noir** (section 2.1). Comme la découverte de la MTU du chemin repose sur des paquets ICMP indiquant que le paquet TCP était trop grand pour le lien, tout ce qui empêche de recevoir ces paquets ICMP perturbera la découverte de la MTU. Les paquets TCP seront abandonnés par les routeurs mais on ne verra pas pourquoi, d'où le nom de trou noir.

Qu'est-ce qui peut créer un trou noir ? Il existe plusieurs raisons (le RFC 1435 en donne une particulière) mais la plus fréquente, de loin, est l'ignorance abyssale de certains administrateurs de coupe-feux, qui bloquent complètement tout ICMP pour des « raisons de sécurité » en général fumeuses (un bon test pour détecter cette ignorance est de leur demander si le "*ping of death*" est spécifique à ICMP. S'ils disent oui, cela montre qu'ils n'ont rien compris.)

Le problème est suffisamment répandu pour qu'il ai fallu développer un nouveau protocole, où TCP fait varier la taille des paquets pour voir si les gros ont d'avantage de pertes, protocole décrit dans le RFC 4821. Ce protocole a été développé bien après notre RFC 2923, qui donne des conseils en ce sens mais fait preuve de prudence en craignant que, si les mises en œuvre de TCP réparent ce genre de problèmes, la cause (le filtrage d'ICMP) risque de demeurer (la section 3 revient sur cette question).

1. Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc1191.txt>

Ce débat est récurrent à l'IETF. Faut-il pratiquer la politique du pire, en refusant de contourner les erreurs de configuration (tenter de faire corriger les sites mal configurés est une entreprise courageuse, voire folle <<http://www.phildev.net/mss/>>), ou bien faut-il essayer de réparer ce qu'on peut, au risque de diminuer la pression sur ceux qui ont fait ces erreurs ?

D'autres problèmes sont, eux, réellement spécifiques à TCP comme les acquittements retardés (section 2.2) provoqués par un envoi des acquittements TCP en fonction de la MSS, qui peut être bien plus grande que la MTU du chemin (cf. RFC 2525). Ce problème n'a pas de solution simple et unique.

Un autre problème lié à la MSS est la détermination de celle-ci (section 2.3). Une implémentation naïve qui déduirait le MSS à partir de la MTU du chemin annoncerait une MSS trop faible et, surtout, ne pourrait pas l'augmenter si un nouveau chemin apparaît. La solution est de déduire la MSS de la MTU du réseau local (qui est stable), pas de la MTU du chemin (RFC 1122).