RFC 7837 : IPv6 Destination Option for Congestion Exposure (ConEx)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 15 mai 2016

Date de publication du RFC : Mai 2016

http://www.bortzmeyer.org/7837.html

Le mécanisme ConEx, normalisé dans le RFC 7713 [1], permet d'informer les routeurs situés en aval qu'un flot de données se heurte à la congestion. Il y a plusieurs façons d'indiquer cette congestion, et ce RFC le fait par une option dans l'en-tête "*Destination Options*" d'IPv6.

En effet, le mécanisme décrit dans le RFC 7713 est abstrait : il spécifie des concepts, pas comment ils sont représentés sur le câble. Notre RFC 7837, au contraire, est concret : il indique comment mettre l'information ConEx dans des champs du paquet que les équipements réseau, notamment les routeurs, pourront lire. ConEx est actuellement un projet expérimental (et même très expérimental) et il n'est pas sûr qu'il soit déployé avec enthousiasme. En attendant, puisque c'est expérimental, le but est de récolter de l'information et, pour cela, il faut du code qui tourne, avec des paquets concrets (section 1 de notre RFC). Mettre l'information ConEx dans le champ "*Options*" d'IPv4 est délicat : ce champ est de taille limitée, et pose souvent des problèmes dans le réseau <http://www.bortzmeyer.org/options-interdites.html>. L'idée est donc d'utiliser le protocole du futur, IPv6, et ses en-têtes d'extension (RFC 2460, section 4).

La section 3 détaille les choix techniques effectués et leurs raisons. Les sections 3.3 et 4 du RFC 7713 expliquent les contraintes d'encodage concret de ConEx dans les paquets. Ici, les exigences considérées ont été :

— Les marques ConEx doivent être visibles par tous les nœuds ConEx du chemin (donc, pas question de les mettre tout au bout du paquet),
— Pour que les paquets marqués arrivent à bon port, même en traversant les équipements actuels qui ne connaissent pas ConEx, il faut un mécanisme standard et déjà largement accepté (pas question de changer le format des paquets IP, cela empêcherait le déploiement incrémental),

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc7713.txt

— La présence des marques ConEx ne doit pas ralentir le traitement des paquets (cf. section 5),
— Les marques ConEx doivent pouvoir être protégées contre les manipulations ultérieures (exigence pas réellement satisfaite en pratique, sauf à utiliser IPsec).

Quatre solutions IPv6 avaient été envisagées par le groupe de travail à l'IETF :

— Une option "*Hop-by-hop*" (RFC 2460, section 4.3).
— Réutiliser le champ "*Flow label*" qui ne sert quasiment pas (cf. RFC 6437),
— Créer un nouvel en-tête d'extension,
— Une option "*Destination*" (RFC 2460, section 4.6), le choix qui a finalement été fait.

L'en-tête d'extension "*Hop-by-hop*" aurait été l'option logique puisqu'elle est examinée par chaque routeur, ce qui est bien ce qu'on veut pour ConEx. Elle aurait été conforme aux principes d'IPv6. Mais, dans les routeurs actuels, le traitement de cette option se fait de manière extrêmement lente (elle n'emprunte pas le chemin rapide dans le routeur, mis en œuvre en ASIC ou FPGA), ce qui viole la troisième exigence. Le choix de l'en-tête "*Destination*", qui est normalement de bout en bout et que les intermédiaires ne sont pas censés regarder, est donc surprenant, mais justifié. Il viole un peu la première exigence (si le paquet est encapsulé, le routeur aura du mal à voir cet en-tête). Et, surtout, analyser la chaîne des en-têtes d'extension IPv6 est anormalement compliqué <http://www.bortzmeyer.org/analyse-pcap-ipv6.html>. Mais il n'y avait guère d'autre choix réaliste. En pratique, certains routeurs auront donc besoin d'un changement de leurs règles de traitement des en-têtes d'extension s'ils veulent voir les marques ConEx.

(Sur la survivabilité des en-têtes d'extension IPv6 dans l'Internet, voir l'étude de Mehdi Kouhen <http://pirl.tech/pdf/Mehdi_Kouhen_SymposiumV2.pdf> en février 2016 et le RFC 7872.)

La section 4 présente le format concret. La nouvelle option "*Destination*" se nomme CDO, pour "*ConEx Destination Option*". Elle est mise dans un en-tête d'extension "*Destination Options*" (RFC 2460, section 4.6). Comme les autres options "*Destination*", elle est encodée en TLV. Le type de l'option est 0x1E (30, valeur réservée aux expérimentations, non définitive), sa longueur est 1 (un seul octet, et encore, tous les bits ne sont pas utilisés) et sa valeur est composée de quatre bits (RFC 7713, notamment la section 2.1) : X (« je sais faire du ConEx »), L (« des paquets ont été perdus »), E (« de la congestion a été signalée par ECN ») et C (« pas (encore) de congestion, j'accumule des crédits »). Le dernier bit, C, est à utiliser avant qu'on détecte la congestion (RFC 7713, notamment la section 5.5).

Au passage, si vous écrivez des programmes en C qui veulent ajouter des options dans l'en-tête "*Destination*", vous pouvez consulter mon article <http://www.bortzmeyer.org/destination-options-ipv6.html>.

On a dit plus haut que la principale raison pour utiliser l'en-tête "*Destination*" et pas le "*Hop-by-Hop*" (qui aurait été plus logique), est le souci que les paquets restent sur le "*fast path*" des routeurs (le traitement fait par le matériel, par opposition au "*slow path*", confié au processeur généraliste, bien plus lent). Mais le problème est que l'en-tête "*Destination*", n'étant pas prévu pour être lu par les équipements réseau sur le chemin, peut se trouver n'importe où dans la chaîne des en-têtes (alors que l'en-tête "*Hop-by-hop*" est forcément au début, cf. RFC 2460, section 4.1). Et l'option CDO pourrait, en théorie, se trouver n'importe où dans l'en-tête "*Destination*". Notre RFC est donc obligé de recommander (section 5) que l'option CDO soit la première dans l'en-tête "*Destination*".

Reste à voir s'il sera effectivement possible de déployer cette option. L'ossification de l'Internet rend tout déploiement de ce type difficile (les en-têtes d'extension sont rares dans les paquets IPv6 actuels).