RFC 7679 : A One-Way Delay Metric for IPPM

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 1 février 2016

Date de publication du RFC : Janvier 2015

https://www.bortzmeyer.org/7679.html

Ce RFC définit une métrique, une grandeur à mesurer, en l'occurrence le délai d'acheminement d'un paquet d'un point à un autre du réseau. Cela semble trivial, mais la définition rigoureuse de la métrique, permettant des mesures scientifiques et leur comparaison, prend vingt pages... Elle remplace l'ancienne définition, qui était normalisée dans le RFC 2679 [1].

Comme tous les RFC du groupe de travail ippm <http://tools.ietf.org/wg/ippm>, celui-ci s'appuie sur les définitions et le vocabulaire du RFC 2330, notamment la notion de "*Type-P*" (paquets IP ayant la même « signature », c'est-à-dire même protocole de couche 4, même numéro de port, etc, car certains équipements réseaux traitent différemment les paquets selon ces variables).

Commençons par le commencement, c'est-à-dire l'introduction, en section 1. Le RFC définit une métrique pour une mesure unique (un singleton) nommé "*Type-P-One-way-Delay*", ainsi qu'un échantillonnage, pour le cas où la mesure est répétée, et enfin diverses statistiques agrégeant les résultats de plusieurs mesures. Pourquoi une telle métrique ? Comme le rappelle la section 1.1, celle-ci est utile dans des cas comme, par exemple :

— Estimer le débit maximal qu'on obtiendra (avec certains protocoles comme TCP, il est limité par le RTT, cf. le RFC 7323).
— La valeur minimale de ce délai d'acheminement nous donne une idée du délai dû uniquement à la propagation (qui est limitée par la vitesse de la lumière) et à la transmission. (Voir aussi la section 4.3.)
— Par contre, son augmentation nous permet de détecter la présence de congestion.

Mais pourquoi un délai d'acheminement « aller-simple » et pas « aller-retour », comme le fait ping, ou bien comme le normalise le RFC 2681 ? Après tout, le délai aller-retour est bien plus simple à mesurer. Mais il a aussi des défauts :

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc2679.txt

— Une bonne partie des routes sur l'Internet sont **asymétriques**. Le délai aller-retour dépend donc de deux délais aller-simple, qui peuvent être très différents.
— Même si le chemin est symétrique, les politiques de gestion de la file d'attente par les routeurs peuvent être très différentes dans les deux sens.
— Certains protocoles s'intéressent surtout aux performances dans une direction. C'est le cas du transfert d'un gros fichier, où la direction des données est plus importante que celle des accusés de réception.

Le RFC ne le mentionne apparemment pas, mais on peut aussi dire que la mesure d'un délai aller-retour dépend aussi du temps de réflexion par la machine distante (avec ping, si la machine visée est très chargée, la génération du paquet ICMP de réponse peut prendre un temps non négligeable ; c'est encore pire avec des protocoles applicatifs, où la réflexion n'est pas faite dans le noyau, et est donc sensible à par exemple, le "*swapping*").

Comme avec tout système de mesure, les horloges et leurs imperfections jouent un rôle crucial. La section 2 rappelle des concepts (voir le RFC 2330) comme la synchronisation (le fait que deux horloges indiquent la même valeur) ou le décalage (le fait qu'une horloge aille plus ou moins vite qu'une autre).

Enfin, après tous ces préliminaires, le RFC en arrive à la définition de la métrique, en section 3. "*Type-P-One-way-Delay*" est définie comme une fonction de divers paramètres comme les adresses IP des deux parties et l'instant de la mesure. La mesure dépend aussi du "*Type-P*" (par exemple parce que certains ports sont favorisés par la QoS ; notez que si la QoS dépend d'une DPI, la définition du Type-P ne suffira pas). Elle est en secondes. Et sa définition (section 3.4) est « Le temps entre l'envoi du premier bit du paquet sur le câble et la réception du dernier bit du paquet depuis le câble ». Si le paquet n'arrive pas, le délai est dit « indéfini » et, en pratique, pris comme étant infini. Les programmeurs noteront donc qu'il n'est pas possible de mesurer cette valeur avec l'API Posix, qui ne donne pas accès à des détails aussi précis.

Ce n'est pas tout de définir une métrique, il faut aussi la mesurer. Dans le monde réel, cela soulève quelques problèmes, couverts par la section 3.5. Le principal étant évidemment la synchronisation des horloges. Si le paquet part à 1247389451,578306110 (en secondes depuis le 1er janvier 1970) et arrive à 1247389453,018393817, le délai a-t-il réellement été de 1,44 secondes ou bien seulement de 1,32 secondes mais les horloges différaient de 0,12 secondes ? Comme des délais peuvent être aussi bas que 100 μs, une synchronisation précise est nécessaire. Le GPS est une bonne solution pour cela, NTP une moins bonne, car sa résolution est plus faible (de l'ordre de plusieurs ms) et il dépend du réseau qu'on veut mesurer.

Il faut aussi tenir compte de problèmes comme les paquets perdus (si le délai de garde est de cinq secondes, un paquet perdu ne doit pas compter pour un délai d'acheminement de cinq secondes, ou bien des agrégats comme la moyenne seront complètement faussés) ou comme la duplication de paquets (dans ce cas, le RFC précise que c'est la première occurrence qui compte). Le délai de garde (délai d'attente maximal avant de considérer un paquet comme perdu) ne doit pas être trop bas, sinon on risque de sur-estimer les performances du réseau, en éliminant les paquets lentement acheminés.

Enfin, pour celui qui voudrait concevoir un protocole de mesure de cette métrique, le RFC suggère une méthodologie, en section 3.6 :

— S'assurer de la synchronisation des horloges,
— Former le paquet et y mettre l'instant de départ sur le câble (sur une machine non temps réel, cela peut être délicat de mettre cet instant exact),
— À la réception, mesurer l'instant et calculer la différence avec ce qui est indiqué dans le paquet,
— Évaluer l'erreur (il y en aura forcément) et, éventuellement, corriger.

Cette question de l'analyse d'erreur fait l'objet de la section 3.7. Les deux principales sources d'erreur seront liées aux horloges (qui ne sont jamais parfaites) et à la différence entre le temps de départ ou d'arrivée mesuré et le temps réel. Sur un système d'exploitation multi-tâches et non temps réel comme Unix, le temps entre le moment où le paquet arrive dans la carte Ethernet et celui où l'application peut appeler `gettimeofday()` est souvent significatif (section 3.7.2) et, pire, variable et imprévisible (car on dépend de l'ordonnanceur). Parmi les mécanismes pour déterminer l'erreur, de façon à pouvoir effectuer les corrections, le RFC suggère la calibration (section 3.7.3), en répétant de nombreuses fois la mesure. Cette partie sur l'analyse des erreurs est détaillée et explique en partie la longueur du RFC : la métrologie n'est pas une science simple !

À noter que les deux horloges, à l'arrivée et au départ, ont besoin d'être synchronisées mais n'ont pas forcément besoin d'être correctes. La correction reste quand même utile, car les délais peuvent dépendre de l'heure de la journée et il faut donc indiquer celle-ci dans le rapport final.

Une fois qu'on a bien travaillé et soigneusement fait ses mesures, il est temps de les communiquer à l'utilisateur. C'est l'objet de la section 3.8. Elle insiste sur l'importance d'indiquer le "*Type-P*", les paramètres, la méthode de calibration, etc. Si possible, le chemin suivi par les paquets dans le réseau devrait également être indiqué.

Maintenant, la métrique pour une mesure isolée, un singleton, est définie. On peut donc bâtir sur elle. C'est ce que fait la section 4, qui définit une mesure répétée, effectuée selon une distribution de Poisson, "*Type-P-One-way-Delay-Poisson-Stream*".

Une fois cette métrique définie, on peut créer des fonctions d'agrégation des données, comme en section 5. Par exemple, la section 5.1 définit "*Type-P-One-way-Delay-Percentile*" qui définit le délai d'acheminement sous lequel se trouvent X % des mesures (les mesures indéfinies étant comptées comme de délai infini). Ainsi, le 95e percentile indique le délai pour lequel 95 % des délais seront plus courts (donc une sorte de « délai maximum en écartant les cas pathologiques »). La section 5.2 définit "*Type-P-One-way-Delay-Median*" qui est la médiane (équivalente au 50e percentile, si le nombre de mesures est impair). La moyenne, moins utile <https://www.bortzmeyer.org/mediane-et-moyenne.html>, n'apparait pas dans ce RFC.

Au moins un protocole a été défini sur la base de cette métrique, OWAMP, normalisé dans le RFC 4656 et mis en œuvre notamment dans le programme de même nom <http://e2epi.internet2.edu/owamp/>.

Le but principal du nouveau RFC (projet documenté dans le RFC 6576) était de faire avancer cette métrique sur le « chemin des normes » (cf. RFC 2026), de « proposition de norme » à « norme [tout court] ». Les tests du RFC 6808 (voir aussi le rapport sur le RFC 2679 <https://tools.ietf.org/html/draft-ietf-ippm-implement-02>) ont montré que les métriques Internet étaient sérieuses et utilisables, justifiant cet avancement sur le chemin des normes.

La section 7 décrit les (rares) changements depuis le RFC 2679. Le RFC 6808 notait plusieurs problèmes avec le RFC 2679, problèmes qui ont été traités par ce nouveau RFC 7679. Par exemple, la métrique `Type-P-One-way-Delay-Inverse-Percentile`, définie dans le RFC 2679 mais jamais mise en œuvre, est abandonnée. D'autre part, une bogue <http://www.rfc-editor.org/errata_search.php?eid=398> a été corrigée. Les autres changements sont l'introduction des références au RFC sur la publication de mesures, RFC 6703 et à d'autres RFC publiés depuis, ainsi que l'accent plus important sur la protection de la vie privée. Notre nouveau RFC note aussi que le groupe de travail a étudié les conséquences du RFC 6921 (dont vous noterez la date de parution) et conclut que, même avec des délais de propagation négatifs, les métriques décrites ici ne nécessitaient pas de mise à jour.