RFC 7276 : An Overview of Operations, Administration, and Maintenance (OAM) Tools

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 25 juin 2014

Date de publication du RFC : Juin 2014

https://www.bortzmeyer.org/7276.html

Ce RFC est entièrement consacré à la gestion de réseaux. Quels sont les outils existants dans les réseaux TCP/IP, notamment l'Internet ? Ce RFC est destiné aux opérateurs réseau, aux fournisseurs d'équipements et de logiciels réseau, et aux gens de la normalisation, pour s'assurer qu'ils ont une base d'information commune sur les outils existants.

La gestion de réseaux est ici désignée par le sigle **OAM** qui signifie "*Operations, Administration, and Maintenance*". OAM est défini dans le RFC 6291 [1]. Les outils de l'OAM sont tous les dispositifs qui permettent de détecter un problème et de l'isoler, ou bien de mesurer les performances. (Les outils permettant de configurer les équipements réseaux ne sont pas traités ici.) Ce nouveau RFC couvre les outils utilisés dans des contextes très divers, d'IP à MPLS en passant par TRILL mais je ne vais parler que des outils IP, ceux que je connais.

Ce RFC décrit les outils, logiciels mettant en œuvre un protocole particulier. Une liste des protocoles normalisés par l'IETF pour l'OAM est dans le RFC 6632. Par exemple, OWAMP (RFC 4656) est un protocole, le programme distribué par Internet 2 <http://e2epi.internet2.edu/owamp/> est un outil mettant en œuvre ce protocole.

Tous les réseaux sont bien sûr gérés, et tous ont évidemment des dispositifs d'OAM, plus ou moins perfectionnés. On a donc :

— O (pour "*Operations*") : tout ce qui contribue à ce que le réseau reste « "*up & running*" » en détectant les problèmes (et en les résolvant) avant même que l'utilisateur ne s'en aperçoive.
— A (pour "*Administration*") : garder à jour la liste des ressources et surveiller leur utilisation.

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6291.txt

— M (pour "*Maintenance*") : permettre les réparations et les évolutions (mise à jour du logiciel du routeur, par exemple).

Parfois, ces dispositifs d'OAM ont été ajoutés après, en utilisant (souvent de manière créative) les fonctions disponibles. De nos jours, on tend à prévoir l'OAM dès le début, comme ce fut le cas pour l'ATM.

Notre RFC commence (section 2) par des rappels de concept et de terminologie (outre le RFC 6291, déjà cité, le RFC 7087 définit les termes spécifiques à MPLS). D'abord, données vs. contrôle : on distingue souvent dans un équipement réseau le sous-système des données ("*data plane*" ou "*user plane*") et le sous-système de contrôle ("*control plane*"). Le premier désigne toutes les fonctions qui contribuent à transmettre les paquets (et donc les données) à l'équipement suivant. Le second, le sous-système de contrôle, comprend les fonctions qui permettent d'établir dynamiquement les informations dont le sous-système des données va avoir besoin. Par exemple, dans un routeur, le sous-système des données est composé des fonctions qui font passer le paquet d'une interface d'entrée à une interface de sortie (rôle de "*forwarding*"), alors que le sous-système de contrôle (rôle de "*routing*") comprend notamment les fonctions de calcul de la table de routage (les mises en œuvre d'OSPF, de BGP, etc), table de routage qu'utilisera le sous-système des données lors de la transmission. Si le routeur est une machine Unix, le sous-système de données est mis en œuvre dans le noyau, alors que celui de contrôle l'est par un démon séparé, genre Quagga ou BIRD. La distinction entre sous-systèmes de contrôle et de données n'est pas seulement conceptuelle, ni limitée au logiciel. Dans un routeur Juniper ou Cisco moyen ou haut de gamme, c'est même du matériel différent (ASIC pour les données, CPU généraliste pour le contrôle). L'OAM, tel qu'utilisé ici, mesure et teste le sous-système de données, mais il peut utiliser des informations fournies par le sous-système de contrôle, et être piloté par lui. (On voit parfois, par exemple dans le RFC 6123, un troisième sous-système, le sous-système de gestion, le "*management plane*". De toute façon, la différence entre ces sous-systèmes n'est pas forcément claire et nette dans tous les cas.)

Encore un peu de terminologie : les outils peuvent être « à la demande » ou bien « proactifs ». Les premiers sont déclenchés lorsqu'il y a un problème (c'est l'administrateur réseaux à qui on dit « le serveur ne marche pas » et qui lance ping tout de suite), les seconds tournent en permanence, par exemple pour détecter et signaler les pannes.

Quelle sont les fonctions précises d'un logiciel d'OAM ? Notre RFC cite le test de la connectivité, la découverte du chemin, la localisation d'un problème, et des mesures liées aux performances comme la mesure du taux de perte de paquets, ou comme la mesure des délais d'acheminement.

On entend parfois parler de « tests de continuité » et de « tests de connectivité ». Ces termes viennent de MPLS (RFC 5860) et désignent, dans le premier cas, un test que les paquets peuvent passer et, dans le second, que deux systèmes sont effectivement connectés, et par le chemin attendu (IP, étant sans connexion, n'a guère besoin de ces tests de connectivité).

En parlant de connexion, le monde TCP/IP comprend des protocoles avec connexion et des protocoles sans. L'OAM concerne ces deux groupes et les outils présentés dans ce RFC couvrent les deux. Un test utilisant un protocole sans connexion (ping est l'exemple archétypal) peut emprunter un autre chemin, d'autres ressources (et donc donner un autre résultat), qu'une utilisation réelle du réseau. Au contraire, en testant un protocole avec connexion (TCP, par exemple), on utilise les mêmes ressources que les vraies données. Notez aussi que des protocoles sans connexion peuvent être testés par des protocoles avec connexion, le RFC donnant l'exemple d'un test d'IP fait avec BFD (RFC 5880).

Enfin, dernier point de terminologie, les mots utilisés pour les problèmes. Il n'existe pas une norme unique et acceptée pour ces mots. Suivant la norme ITU-T G.806, notre RFC fait une différence entre défaillance ("*fault*"), qui est l'incapacité à effectuer une tâche, comme de livrer un paquet, interruption

("*defect*"), qui est un arrêt temporaire d'un service et panne ("*failure*"). Cette dernière, au contraire de l'interruption, n'est pas un accident ponctuel et elle peut durer longtemps.

Bien, après tous ces préliminaires, les outils eux-mêmes, présentés en section 4 du RFC. À tout seigneur, tout honneur, on va évidemment commence par l'outil emblématique de l'OAM, ping. Il envoie un paquet ICMP de type « demande d'écho » et la réception d'un paquet ICMP de réponse indiquera que la connectivité est complète. Il donnera également quelques indicateurs de performance comme le RTT, affiché après "*time=*" :

```
% ping -c 3 www.ietf.org
PING www.ietf.org (4.31.198.44) 56(84) bytes of data.
64 bytes from mail.ietf.org (4.31.198.44): icmp_req=1 ttl=44 time=227 ms
64 bytes from mail.ietf.org (4.31.198.44): icmp_req=2 ttl=44 time=222 ms
64 bytes from mail.ietf.org (4.31.198.44): icmp_req=3 ttl=44 time=223 ms

--- www.ietf.org ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2004ms
rtt min/avg/max/mdev = 222.689/224.685/227.533/2.067 ms
```

Certaines versions de ping peuvent faire des tests avec UDP plutôt qu'ICMP. ping donne un point de vue unilatéral, celui du côté qui initie la connexion. Si on veut le point de vue du pair, il faut que celui-ci fasse un test de son côté (par exemple, s'il y a un pare-feu à état entre les deux machines, un des tests peut échouer même si l'autre réussit). Malheureusement, il arrive que ces paquets ICMP soient filtrés (par exemple, suite à l'incompétence de l'administrateur qui croit que l'attaque mal nommée "*ping of death*" a un rapport avec ping alors qu'elle n'en a pas).

Après ping, l'outil d'OAM le plus connu et le plus utilisé est sans doute traceroute. Comme ping, il n'y a pas de texte sacré de référence le décrivant. À l'époque, on codait d'abord, on documentait ensuite (parfois). Les textes les plus canoniques sur traceroute semblent être les RFC 1470 et RFC 2151.

traceroute sert à découvrir le chemin pris par les paquets entre l'initiateur de la requête et une cible. Il envoie un paquet UDP à destination du port 33434 en mettant dans l'en-tête IP un TTL inhabituel, de 1. Vu avec tcpdump, cela donne (notez le `ttl 1`) :

```
22:10:20.369307 IP (tos 0x0, ttl 1, id 19164, offset 0, flags [none], proto UDP (17), length 60)
    192.168.2.1.44024 > 213.154.224.1.33434: UDP, length 32
```

Le premier routeur rencontré va donc considérer ce paquet comme trop ancien, le jeter, et envoyer un paquet ICMP de type "*Time exceeded*" à la source, qui aura ainsi l'adresse IP du premier routeur :

```
22:17:46.359084 IP (tos 0xc0, ttl 64, id 64893, offset 0, flags [none], proto ICMP (1), length 88)
    192.168.2.254 > 192.168.2.1: ICMP time exceeded in-transit, length 68
IP (tos 0x0, ttl 1, id 19239, offset 0, flags [none], proto UDP (17), length 60)
    192.168.2.1.34315 > 213.154.224.1.33434: UDP, length 32
```

Ensuite, traceroute itère : il incrémente le numéro de port et le TTL et trouve ainsi le deuxième routeur. Le jeu s'arrête lorsque traceroute reçoit un paquet ICMP "*Port unreachable*" lorsque le paquet UDP atteint sa cible et que celle-ci, n'ayant pas de service sur ce port, répondra négativement :

```
22:19:20.648790 IP (tos 0x0, ttl 53, id 11453, offset 0, flags [none], proto ICMP (1), length 56)
    213.136.31.102 > 192.168.2.1: ICMP 213.154.224.1 udp port 33445 unreachable, length 36
IP (tos 0x0, ttl 1, id 19325, offset 0, flags [none], proto UDP (17), length 60)
    192.168.2.1.57471 > 213.154.224.1.33445: UDP, length 32
```

Utilisant toutes ces informations, traceroute peut alors afficher le chemin :

```
% traceroute -q 1 www.nlnetlabs.nl
traceroute to www.nlnetlabs.nl (213.154.224.1), 30 hops max, 60 byte packets
 1  freebox (192.168.2.254)  13.646 ms
 2  88.189.152.254 (88.189.152.254)  37.136 ms
 3  78.254.1.62 (78.254.1.62)  37.133 ms
 4  rke75-1-v900.intf.nra.proxad.net (78.254.255.42)  46.845 ms
 5  cev75-1-v902.intf.nra.proxad.net (78.254.255.46)  39.030 ms
 6  p16-6k-1-po12.intf.nra.proxad.net (78.254.255.50)  43.658 ms
 7  bzn-crs16-1-be1024.intf.routers.proxad.net (212.27.56.149)  43.662 ms
 8  bzn-crs16-1-be1106.intf.routers.proxad.net (212.27.59.101)  43.637 ms
 9  londres-6k-1-po101.intf.routers.proxad.net (212.27.51.186)  55.973 ms
10  *
11  amsix-501.xe-0-0-0.jun1.bit-1.network.bit.nl (195.69.144.200)  60.693 ms
12  nlnetlabs-bit-gw.nlnetlabs.nl (213.136.31.102)  62.287 ms
```

(On note que le routeur n° 10 a refusé de répondre, ce qui arrive, ou que sa réponse s'est perdue, ce qui arrive aussi.)

Sur une machine Unix d'aujourd'hui, il existe souvent plusieurs mises en œuvre différentes de traceroute. traceroute est une idée, pas un programme unique. Par exemple, Paris traceroute <http://www.paris-traceroute.net/> fait varier l'en-tête IP pour essayer de découvrir plusieurs chemins lorsqu'il y a de l'ECMP. Des traceroutes ont des options pour utiliser ICMP et pas UDP (c'est le cas des sondes Atlas <https://www.bortzmeyer.org/traceroute-atlas.html>). Il existe des extensions à ICMP pour rendre traceroute plus utile, comme celles du RFC 4950 ou RFC 5837.

Attention aux limites de traceroute : il ne montre que le chemin aller. Le routage sur l'Internet pouvant être asymétrique, rien ne dit que le chemin de retour sera le même.

Après, le RFC passe à des outils moins connus du grand public. D'abord, les outils BFD. Normalisé dans le RFC 5880, BFD permet de tester la connectivité dans le contexte d'IP (RFC 5881) mais aussi de MPLS (RFC 5884) et de bien d'autres. Il fonctionne en mode connecté : les deux machines établissent une session BFD. Ensuite, elles échangent des paquets BFD, soit périodiquement, soit à la demande. La bonne arrivée de ces paquets indique que la connexion avec le pair fonctionne. Il existe aussi un mode où BFD envoie des paquets « demande d'écho » et attend une réponse.

Il existe ensuite plusieurs outils d'OAM pour le monde MPLS, le plus important étant LSP-Ping, décrit dans le RFC 4379. S'il assure des fonctions analogues à ping, comme son nom l'indique, il en fait aussi qui sont équivalentes à traceroute. Répondre à une demande d'écho est moins trivial en MPLS qu'en IP. Par exemple, un chemin (un LSP, "*Label Switched Path*") peut n'être établi que dans une seule direction, rendant impossible de répondre en MPLS à un LSP-Ping.

Les autres outils MPLS sont décrits en détail dans les sections 4.4 et 4.5 de notre RFC. Sinon, la problématique générale de l'OAM en environnement MPLS est dans les RFC 4377 et RFC 4378. (L'OAM de MPLS a d'ailleurs suscité un affrontement entre l'IETF et l'UIT, le point de vue de l'IETF étant exposé dans le RFC 5704.)

ping sur un réseau IP a plusieurs limites : il mesure un aller-retour (alors que les chemins peuvent être différents, avec des caractéristiques très différentes), et il n'est pas défini avec une grande précision donc la sémantique du RTT, par exemple, n'est pas rigoureuse. L'IETF, via son groupe de travail IPPM </search?pattern=ippm>, a deux protocoles qui traitent ces limites, OWAMP et TWAMP. Outre des définitions rigoureuses de métriques comme le délai d'acheminement aller-simple (RFC 7679), ou comme la connectivité (RFC 2678), IPPM a produit ces deux protocoles qui établissent une session de contrôle entre deux points, sur TCP, avant de lancer une mesure précise avec UDP. OWAMP (RFC 4656) mesure un aller-simple alors que TWAMP (RFC 5357) mesure un aller-retour (comme ping).

À noter que, pour mesurer les performances, il existe bien d'autres outils comme les outils OAM d'Ethernet <https://www.bortzmeyer.org/ethernet-oam.html>.

Enfin, la section 4 se termine par une section sur TRILL (RFC 6325). Trop récent, ce système n'a pas encore beaucoup d'outils OAM. Il existe un cahier des charges pour ces outils, le RFC 6905, qui demande des outils pour tester la connectivité, découvrir le chemin suivant et mesurer des indicateurs comme le temps d'acheminement d'un paquet.

La section 5 du RFC résume tous les outils décrits, sous forme d'un tableau donnant le nom des outils, un résumé de leur fonctionnement, et le protocole auquel ils s'appliquent (IP, MPLS, TRILL, etc). Pratique pour les lecteurs pressés.

Il reste enfin la classique section de sécurité, la section 6. Un système d'OAM soulève évidemment des enjeux de sécurité. Par exemple, un attaquant qui contrôle l'OAM pourrait créer l'illusion de pannes non-existantes, forçant les équipes du réseau à perdre du temps. Ou il pourrait empêcher la détection de pannes réelles, ce qui aggraverait le déni de service.

Certains des outils d'OAM présentés dans ce RFC ont des fonctions de sécurité, empêchant par exemple la modification des données par un tiers. BFD a un mécanisme d'authentification. OWAMP et TWAMP peuvent authentifier (via un HMAC) et/ou chiffrer leurs communications (en AES). Ceci dit, la confidentialité n'est pas en général considérée comme un problème pour l'OAM, seules l'authentification et l'intégrité le sont. D'autres outils n'ont aucune sécurité (traceroute peut facilement être trompé, comme l'avait montré The Pirate Bay <http://thenextweb.com/insider/2013/03/04/the-pirate-bay-may-or-may-not-have-been-invited-to-north-korea-but-its-not-being-hosted-t>).

Les outils d'OAM peuvent aussi être utilisés pour la reconnaissance : par exemple, l'option `-g` de fping <http://fping.org/> ou la possibilité d'indiquer un préfixe IP (et pas juste une adresse) à nmap, permettent de découvrir les machines existant sur le réseau.

L'annexe A termine le RFC en énumérant une (longue) liste de documents de normalisation sur l'OAM, y compris des documents non-IETF.