

RFC 3031 : Multiprotocol Label Switching Architecture

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 11 avril 2007. Dernière mise à jour le 18 mai 2007

Date de publication du RFC : Janvier 2001

<https://www.bortzmeyer.org/3031.html>

MPLS est un protocole réseau souvent décrit comme « de couche 2,5 » qui est aujourd'hui très utilisé dans l'Internet. Ce RFC est le premier décrivant MPLS.

L'IETF travaille surtout sur les protocoles de couche 3 et de couche 4 mais a aussi une activité importante dans la couche 7 et fait parfois des incursions dans les couches inférieures à 3, par exemple avec PPP ou bien avec MPLS, sujet de notre RFC.

MPLS est un mécanisme de commutation rapide de paquets, fondé sur des adresses de format fixe et sans structure hiérarchique, les **labels** (ou étiquettes).

En effet, avec IP, un routeur doit, pour décider de l'étape suivante ("*next hop*"), dérouler l'algorithme de l'adresse la plus spécifique ("*longest match*") puisque le routeur peut avoir une route pour 192.0.2.0/24 **et** une route plus spécifique (et qui doit donc être sélectionnée) pour 192.0.2.128/26. Bien qu'il existe de nombreuses implémentations de cet algorithme, et des structures de données spécialisées pour ranger la table de routage (comme les Patricia tries), il sera toujours plus lent qu'une table avec un index. C'est ce que propose MPLS. Les labels, contrairement aux adresses IP, n'ont pas de structure, ils ne sont qu'un index dans une table.

Les labels n'ont qu'une signification locale, à l'intérieur d'un même système autonome (ou groupe de systèmes autonomes). MPLS ne prétend pas pouvoir remplacer IP, simplement faciliter la commutation rapide des paquets à l'intérieur du réseau d'un opérateur.

MPLS fonctionne donc ainsi : à l'entrée du réseau MPLS, le premier routeur MPLS (on le nomme LSR pour "*Label Switching Router*"), décide d'une **classe**, la FEC ("*Forwarding Equivalence Class*", ensemble des paquets qui sont transmis au même routeur suivant et dans les mêmes conditions, par exemple parce

qu'ils font partie de la même classe de service), attribue un label au paquet et l'insère (le RFC 3032¹ donne un exemple d'une manière de réaliser cette insertion). Tant qu'on reste dans la partie MPLS du réseau de l'opérateur, la commutation se fait sur le label et l'adresse IP est ignorée. À la sortie de la partie MPLS du réseau, le label est retiré et le routage « normal » reprend.

La réalité est un peu plus compliquée, par exemple MPLS a le concept de « pile de labels » où des labels successifs sont ajoutés et retirés par les LSR, permettant ainsi de créer des sortes de tunnels, par exemple pour décider finement de l'endroit où doit passer le trafic ("*traffic engineering*").

Comment savoir si une liaison Internet passe par MPLS? Il existe un "*patch*" pour le programme `traceroute` qui affiche les labels MPLS en utilisant ICMP, si les LSR sont configurés pour signaler ces labels (tous les opérateurs ne le font pas). Ce patch est par exemple intégré dans Gentoo ou dans le `traceroute` de NANOG <<ftp://ftp.login.com/pub/software/traceroute/>> et normalisé dans le RFC 4950. Voici un exemple de ce qu'il affiche (étapes 6 et 7) :

```
% traceroute f.root-servers.net
traceroute to f.root-servers.net (192.5.5.241), 30 hops max, 46 byte packets
 1 208.75.84.1 (208.75.84.1) 0.256 ms 0.129 ms 0.201 ms
 2 host166.datotel.com (63.97.187.166) 0.478 ms 0.349 ms 0.496 ms
 3 66.236.121.49.ptr.us.xo.net (66.236.121.49) 0.983 ms 1.098 ms 0.993 ms
 4 p4-3-0.MAR1.MarylandHeights-MO.us.xo.net (207.88.84.73) 6.737 ms 1.346 ms 1.477 ms
 5 p5-2-0-2.rar1.chicago-il.us.xo.net (65.106.6.157) 7.069 ms 57.625 ms 10.694 ms
 6 p6-0-0.rar2.denver-co.us.xo.net (65.106.0.25) 30.537 ms 30.471 ms 30.487 ms
   MPLS Label=134469 CoS=0 TTL=0 S=1
 7 p0-0-0d0.rar1.denver-co.us.xo.net (65.106.1.73) 33.935 ms 31.420 ms 30.463 ms
   MPLS Label=441165 CoS=0 TTL=0 S=1
 8 p6-0-0.rar1.sanjose-ca.us.xo.net (65.106.0.21) 72.947 ms 186.167 ms 72.648 ms
```

À noter que la valeur du label change à chaque routeur. Ce qu'affiche `traceroute` n'a donc d'intérêt que si on a accès au routeur en question.

Comment sont distribués les labels? Par le biais d'un protocole spécialisé. Il en existe plusieurs comme le LDP (RFC 5036) ou comme BGP (RFC 3107). Certains opérateurs ont ainsi un réseau interne où toute la commutation est en MPLS. Parfois, ils se passent complètement de BGP pour le routage interne.

MPLS se nommait à l'origine "*tag switching*" et avait été popularisé par une société nommée Ipsilon, rachetée par Cisco depuis. C'est pourquoi ce terme de "*tag*" (officiellement remplacé par **label**) apparaît parfois dans les documentations.

MPLS est aujourd'hui largement déployé et on trouve des mises en œuvre de MPLS pour les routeurs Cisco (voir des exemples de configuration <http://www.cisco.com/en/US/tech/tk436/tk428/tsd_technology_support_protocol_home.html>), Juniper, pour Linux <<http://mpls-linux.sourceforge.net/>>, etc.

Merci à Sarah Tharan pour ses explications et clarifications.

MPLS n'a pas été apprécié par tout le monde, entre autre parce qu'il revient aux traditionnels circuits (l'Internet ayant été bâti autour de la commutation de paquets). Van Jacobson avait ainsi écrit une vigoureuse critique, « "*Circuits : the search for a cure*" <<http://www.sigcomm.org/sites/default/files/award-talks/vj-sigcomm01.pdf>> ».

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc3032.txt>