Groupe de travail Réseau
Request for Comments : 3032
Catégorie : En cours de normalisation
janvier 2001
Traduction Claude Brière de L'Isle

E. Rosen, D. Tappan, G. Fedorkow
Cisco Systems, Inc.
Y. Rekhter, Juniper Networks
D. Farinacci, T. Li, Procket Networks, Inc.
A. Conta, TranSwitch Corporation

Codage de la pile d'étiquettes MPLS

Résumé
La commutation d'étiquettes multi protocoles (MPLS, *Multi-Protocol Label Switching*) [RFC3031] requiert un ensemble de procédures pour ajouter des "piles d'étiquettes" aux paquets de couche réseau, en faisant par là des "paquets étiquetés". Les routeurs qui prennent en charge MPLS sont appelés des "routeurs de commutation d'étiquette" (LSR, *Label Switching Router*). Pour transmettre un paquet étiqueté sur une certaine liaison de données, un LSR doit prendre en charge une technique de codage qui, avec une pile d'étiquettes et un paquet de couche réseau produit un paquet étiqueté. Le présent document spécifie le codage à utiliser par un LSR afin de transmettre des paquets étiquetés sur des liaisons de données au protocole point à point (PPP), sur des liaisons de données de LAN, et éventuellement aussi bien sur d'autres liaisons de données. Sur certaines liaisons de données, l'étiquette au sommet de la pile peut être codée d'une manière différente, mais les techniques décrites ici DOIVENT être utilisées pour coder le reste de la pile d'étiquette. Le présent document spécifie aussi des règles et procédures de traitement des divers champs du codage de pile d'étiquettes.

Table des matières

1. Introduction

La commutation d'étiquettes multi protocole (MPLS, *Multi-Protocol Label Switching*) [RFC3031] exige un ensemble de procédures pour ajouter des "piles d'étiquettes" aux paquets de couche réseau, les transformant ainsi en "paquets étiquetés". Les routeurs qui prennent en charge MPLS sont appelés des "routeurs de commutation d'étiquettes" (LSR, *Label Switching Router*). Pour transmettre un paquet étiqueté sur une certaine liaison de données, un LSR doit prendre en charge une technique de codage qui, avec une pile d'étiquettes et un paquet de couche réseau, produit un paquet étiqueté.

Le présent document spécifie le codage qu'utilise un LSR afin de transmettre des paquets étiquetés sur des liaisons de données PPP et sur des liaisons de données en LAN. Le codage spécifié peut aussi bien être utile pour d'autres liaisons de données.

Le présent document spécifie aussi les règles et procédures de traitement des divers champs du codage de la pile d'étiquettes. Comme MPLS est indépendant de tout protocole de couche réseau particulier, la majorité de ces procédures est aussi indépendante du protocole. Quelques unes diffèrent cependant pour les différents protocoles. Dans le présent document, on spécifie les procédures indépendantes du protocole, et on spécifie les procédures dépendantes du protocole pour IPv4 et IPv6.

Les LSR qui sont mis en œuvre sur certains appareils de commutation (comme les commutateurs ATM) peuvent utiliser différentes techniques de codage pour coder les une ou deux entrées supérieures de la pile d'étiquettes. Lorsque la pile d'étiquettes a d'autres entrées, la technique de codage décrite dans le présent document DOIT cependant être utilisée pour les entrées supplémentaires de la pile d'étiquettes.

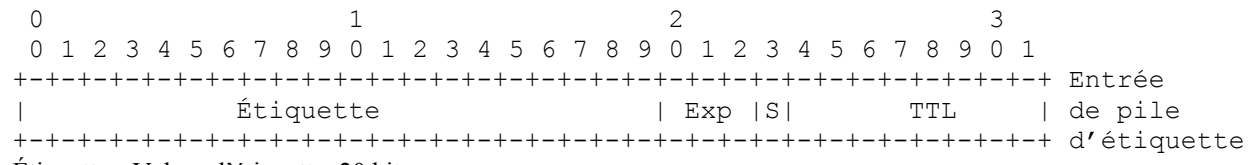
1.1 Spécification des exigences

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" dans le présent document sont à interpréter comme décrit dans la [RFC2119].

2. Pile d'étiquettes

2.1 Codage de la pile d'étiquettes

La pile d'étiquettes est représentée comme une séquence d'entrées de pile d'étiquettes. Chaque entrée de pile d'étiquettes est représentée par 4 octets. C'est ce qui est montré à la Figure 1.



Étiquette : Valeur d'étiquette, 20 bits
Exp : Utilisation expérimentale, 3 bits
S : Bas de la pile, 1 bit
TTL : Durée de vie, 8 bits

Figure 1

Les entrées de pile d'étiquette apparaissent APRÈS les en-têtes de couche de liaison des donnes, mais AVANT tout en-tête de couche réseau. Le sommet de la pile d'étiquettes apparaît le plus tôt dans le paquet, et le bas apparaît en dernier. Le paquet de couche réseau suit immédiatement l'entrée de pile d'étiquette qui a le bit S établi.

Chaque entrée de pile d'étiquette se compose des champs suivants :

1. Bas de pile (S)
   Ce bit est établi à un pour la dernière entrée dans la pile d'étiquette (c'est-à-dire, pour le bas de la pile) et zéro pour toutes les autres entrées de la pile d'étiquettes.

2. Durée de vie (TTL, *Time To Live*)
   Ce champ de huit bits est utilisé pour coder une valeur de durée de vie. Le traitement de ce champ est décrit au paragraphe 2.4.

3. Utilisation expérimentale
   Ce champ de trois bits est réservé pour une utilisation expérimentale.

4. Valeur d'étiquette
   Ce champ de 20 bits porte la valeur réelle de l'étiquette. Lorsque un paquet étiqueté est reçu, la valeur de l'étiquette au sommet de la pile est examinée. À la suite d'un examen réussi, on apprend :
   a) le prochain bond auquel le paquet est à transmettre;
   b) l'opération à effectuer sur la pile d'étiquettes avant la transmission ; cette opération peut être de remplacer l'entrée du sommet de la pile par une autre, ou de supprimer une entrée de la pile, ou de remplacer l'entrée du sommet de la pile puis de pousser une ou plusieurs entrées supplémentaires dans la pile d'étiquettes.

En plus d'apprendre le prochain bond et l'opération sur la pile d'étiquettes, on peut aussi apprendre l'encapsulation de liaison de données sortante, et éventuellement d'autres informations qui sont nécessaires afin de transmettre correctement le paquet.

Il y a plusieurs valeurs d'étiquette réservées :

i. Une valeur de 0 représente "l'étiquette NUL explicite IPv4". Cette valeur d'étiquette n'est légale qu'au bas de la pile d'étiquettes. Elle indique que la pile d'étiquettes doit être sautée, et que la transmission du paquet doit alors se fonder sur l'en-tête IPv4.
ii. Une valeur de 1 représente "l'étiquette d'alerte de routeur". Cette valeur d'étiquette est légale partout dans la pile d'étiquettes sauf en fin. Lorsque un paquet reçu contient cette valeur d'étiquette au sommet de la pile d'étiquettes, il est livré à un module logiciel local pour traitement. La transmission réelle du paquet est déterminée par l'étiquette qui se trouve en dessous de celle-ci dans la pile. Cependant, si le paquet est transmis plus loin, l'étiquette d'alerte de routeur devrait être repoussée en arrière dans la pile d'étiquettes avant la transmission. L'utilisation de cette étiquette est analogue à l'utilisation de "l'option d'alerte de routeur" dans les paquets IP [RFC2113]. Comme cette étiquette ne peut pas survenir au bas de la pile, elle n'est pas associée à un protocole de couche réseau particulier.
iii. Une valeur de 2 représente "l'étiquette NUL explicite IPv6". Cette valeur d'étiquette n'est légale qu'au bas de la pile d'étiquettes. Elle indique que la pile d'étiquettes doit être sautée, et que la transmission du paquet doit alors se fonder sur l'en-tête IPv6.
iv. Une valeur de 3 représente "l'étiquette NUL implicite". C'est une étiquette qu'un LSR peut allouer et distribuer, mais qui n'apparaît en fait jamais dans l'encapsulation. Lorsque un LSR remplacerait autrement l'étiquette au sommet de la pile par une nouvelle étiquette, mais que la nouvelle étiquette est "NUL implicite", le LSR va sauter la pile au lieu de faire le remplacement. Bien que cette valeur puisse ne jamais apparaître dans l'encapsulation, elle doit être spécifiée dans le protocole de distribution d'étiquettes, de sorte qu'une valeur lui est réservée.
v. Les valeurs 4 à 15 sont réservées.

2.2 Détermination du protocole de couche réseau

Lorsque la dernière étiquette est sautée dans la pile d'étiquettes d'un paquet (d'où il résulte que la pile est vidée) le traitement ultérieur du paquet se fonde sur l'en-tête de couche réseau du paquet. Le LSR qui saute la dernière étiquette de la pile doit donc être capable d'identifier le protocole de couche réseau du paquet. Cependant la pile d'étiquettes ne contient aucun champ qui identifie explicitement le protocole de couche réseau. Cela signifie que l'identité du protocole de couche réseau doit être déductible de la valeur de l'étiquette qui est sautée à partir du bas de la pile, éventuellement avec le contenu de l'en-tête de couche réseau lui-même.

Donc, lorsque la première étiquette est poussée sur un paquet de couche réseau, l'étiquette doit être de celles qui ne sont utilisées QUE pour les paquets d'une certaine couche réseau, ou bien de celles qui ne sont utilisées QUE pour un ensemble spécifique de protocoles de couche réseau, où les paquets des couches réseau spécifiées peuvent être distingués par inspection de l'en-tête de couche réseau. De plus, chaque fois que cette étiquette est remplacée par une autre valeur d'étiquette durant le transit d'un paquet, la nouvelle valeur doit aussi être de celles qui satisfont aux mêmes critères. Si ces conditions ne sont pas satisfaites, le LSR qui saute la dernière étiquette d'un paquet ne sera pas capable d'identifier le protocole de couche réseau du paquet.

L'adhésion à ces conditions ne donne pas nécessairement aux nœuds intermédiaires la capacité d'identifier le protocole de couche réseau d'un paquet. Dans des conditions ordinaires, cela n'est pas nécessaire, mais il y a des conditions d'erreur dans lesquelles cela est souhaitable. Par exemple, si un LSR intermédiaire détermine qu'un paquet étiqueté est indélivrable, il peut être souhaitable pour ce LSR de générer des messages d'erreur spécifiques de la couche réseau du paquet. Le seul

moyen qu'a le LSR intermédiaire pour identifier la couche réseau est l'inspection de l'étiquette du sommet et l'en-tête de couche réseau. Donc, si les nœuds intermédiaires sont capables de générer des messages d'erreur spécifiques du protocole pour les paquets étiquetés, toutes les étiquettes de la pile doivent satisfaire aux critères spécifiés ci-dessus pour les étiquettes qui apparaissent au bas de la pile.

Si un paquet ne peut pas être transmis pour une raison quelconque (par exemple, il excède la MTU de la liaison de données) et si le protocole de couche réseau ne peut pas être identifié, ou si il n'y a pas de règle spécifique dépendant du protocole pour le traitement de la condition d'erreur, le paquet DOIT alors être éliminé en silence.

2.3 Génération de messages ICMP pour les paquets IP étiquetés

Le paragraphe 2.4 et la section 3 exposent des situations dans lesquelles il est souhaitable de générer des messages ICMP pour les paquets IP étiquetés. Afin qu'un certain LSR soit capable de générer un paquet ICMP et que ce paquet soit envoyé à la source du paquet IP, deux conditions doivent être satisfaites :

1. il doit être possible à ce LSR de déterminer qu'un certain paquet étiqueté est un paquet IP ;
2. il doit être possible à ce LSR d'acheminer à l'adresse de source IP du paquet.

La condition 1 est exposée au paragraphe 2.2. Les deux paragraphes suivants exposent la condition 2. Cependant, il y aura des cas dans lesquels la condition 2 n'est pas satisfaite du tout, et dans ces cas, il ne sera pas possible de générer le message ICMP.

2.3.1 Tunnelage à travers un domaine d'acheminement de transit

Supposons qu'on utilise MPLS pour "tunneler" à travers un domaine d'acheminement de transit, où les chemins externes ne s'écoulent pas dans les routeurs intérieurs du domaine. Par exemple, les routeurs intérieurs peuvent fonctionner avec OSPF, et savent seulement comment atteindre des destinations au sein de ce domaine OSPF. Le domaine pourrait contenir plusieurs routeurs frontières de système autonome (ASBR, *Autonomous System Border Router*) qui parlent BGP entre eux. Cependant, dans cet exemple, les chemins à partir de BGP ne sont pas distribués dans OSPF, et les LSR qui ne sont pas ASBR ne fonctionnent pas avec BGP.

Dans cet exemple, seul un ASBR saura comment acheminer à la source d'un certain paquet. Si un routeur intérieur a besoin d'envoyer un message ICMP à la source d'un paquet IP, il ne saura pas comment acheminer le message ICMP.

Une solution est d'avoir un ou plusieurs des ASBR qui injectent "par défaut" dans l'IGP. (Note : cela N'EXIGE PAS qu'il y ait un portage "par défaut" par BGP.) Cela assurerait alors que tout paquet non étiqueté qui doit quitter le domaine (comme un paquet ICMP) soit envoyé à un routeur qui a toutes les informations d'acheminement. Les routeurs qui ont toutes les informations d'acheminement vont étiqueter les paquets avant de les renvoyer à travers le domaine de transit, de sorte que l'utilisation de l'acheminement par défaut au sein du domaine de transit ne cause pas de boucle.

Cette solution ne fonctionne que pour les paquets qui ont des adresses uniques au monde, et pour les réseaux dans lesquels tous les ASBR ont les informations d'acheminement complètes. Le paragraphe suivant décrit une solution qui fonctionne lorsque ces conditions ne sont pas satisfaites.

2.3.2 Tunnelage d'adresses privées à travers un cœur de réseau public

Dans certains cas où MPLS est utilisé pour tunneler à travers un domaine d'acheminement, il peut n'être pas possible du tout d'acheminer à l'adresse de source d'un paquet fragmenté. Cela serait le cas, par exemple, si les adresses IP portées dans le paquet étaient des adresses privées (c'est-à-dire, non uniques au monde) et si MPLS était utilisé pour tunneler ces paquets à travers un cœur de réseau public. L'acheminement par défaut à un ASBR ne fonctionnera pas dans cet environnement.

Dans cet environnement, afin d'envoyer un message ICMP à la source d'un paquet, on peut copier la pile d'étiquettes à partir du paquet d'origine sur le message ICMP, puis envoyer le message ICMP en commutation d'étiquettes. Cela sera cause que le message va voyager en direction de la destination du paquet d'origine, plutôt que vers sa source. Sauf si le message est en commutation d'étiquette tout le long du chemin vers l'hôte de destination, il va finir, non étiqueté, dans un routeur qui ne sait pas comme acheminer à la source du paquet d'origine, point auquel le message sera envoyé dans la bonne direction.

Cette technique peut être très utile si le message ICMP est le message "Durée dépassée" ou un message "Destination injoignable à cause de la fragmentation nécessaire et le bit DF est établi".

Lorsque on copie la pile d'étiquettes à partir du paquet d'origine dans le message ICMP, les valeurs d'étiquettes doivent

être copiées exactement, mais les valeurs de TTL dans la pile d'étiquettes devraient être réglées à la valeur de TTL qui est placée dans l'en-tête IP du message ICMP. Cette valeur de TTL devrait être assez longue pour permettre d'effectuer le circuit que le message ICMP aura besoin de suivre.

Noter que si l'expiration du TTL d'un paquet est due à la présence d'une boucle d'acheminement, si cette technique est utilisée, le message ICMP peut aussi être en boucle. Comme un message ICMP n'est jamais envoyé à la suite de la réception d'un message ICMP, et comme de nombreuses mises en œuvre réduisent le taux de génération des messages ICMP , on ne pense pas que cela puisse poser de problème.

2.4 Traitement du champ Durée de vie

2.4.1 Définitions

Le "TTL entrant" d'un paquet étiqueté est défini comme étant la valeur du champ TTL de l'entrée du sommet de la pile d'étiquettes lorsque le paquet est reçu.

Le "TTL sortant" d'un paquet étiqueté est défini comme étant le plus grand de :
a) un moins le TTL entrant,
b) zéro.

2.4.2 Règles indépendantes du protocole

Si le TTL sortant d'un paquet étiqueté est 0, celui-ci NE DOIT alors PAS être transmis plus loin ; la pile d'étiquettes ne doit pas être retirée et le paquet être transmis comme un paquet non étiqueté. La durée de vie du paquet dans le réseau est considérée comme étant terminée.

Selon la valeur de l'étiquette dans l'entrée de pile d'étiquettes, le paquet PEUT être simplement éliminé, ou il peut être passé à la couche réseau "ordinaire" appropriée pour un traitement d'erreur (par exemple, pour générer un message d'erreur ICMP, voir au paragraphe 2.3).

Lorsque un paquet étiqueté est transmis, le champ TTL de l'entrée de pile d'étiquettes au sommet de la pile d'étiquettes DOIT être réglé à la valeur du TTL sortant.

Noter que la valeur de TTL sortant est seulement une fonction de la valeur du TTL entrant, et est indépendante de la poussée ou du tirage d'étiquettes avant la transmission. La valeur du champ TTL n'a pas de signification pour l'entrée de la pile d'étiquettes qui n'est pas au sommet de la pile.

2.4.3 Règles dépendantes de IP

On définit le champ "TTL IP" comme étant la valeur du champ TTL IPv4, ou la valeur du champ Limite de bonds IPv6, selon celui qui est applicable.

Lorsque un paquet IP est étiqueté pour la première fois, le champ TTL de l'entrée de pile d'étiquettes DOIT être réglé à la valeur du champ TTL IP. (Si le champ TTL IP doit être décrémenté, au titre du traitement IP, on suppose que cela a déjà été fait.)

Lorsque une étiquette est sautée, et que la pile d'étiquettes résultante est vide, la valeur du champ TTL IP DEVRAIT alors être remplacée par la valeur du TTL sortant, comme défini ci-dessus. Dans IPv4, cela exige aussi la modification de la somme de contrôle de l'en-tête IP.

On reconnaît qu'il peut y avoir des situations dans lesquelles un administrateur de réseau préfèrera diminuer le TTL IPv4 de un lorsque il traverse un domaine MPLS, plutôt que de décrémenter le TTL IPv4 du nombre de bonds du LSP compris au sein du domaine.

2.4.4 Traduction entre différentes encapsulations

Parfois, un LSR peut recevoir un paquet étiqueté sur, par exemple, une interface ATM contrôlée par commutation d'étiquette (LC-ATM, *Label switching Controlled ATM*) [RFC3035], et peut avoir besoin de l'envoyer sur une liaison PPP ou de LAN. Le paquet entrant ne sera alors pas reçu en utilisant l'encapsulation spécifiée dans le présent document, mais le paquet sortant sera envoyé en utilisant l'encapsulation spécifiée dans ce document.

Dans ce cas, la valeur du "TTL entrant" est déterminée par les procédures utilisées pour porter les paquets étiquetés sur, par exemple, des interfaces LC-ATM. Le traitement de TTL se passe alors comme décrit ci-dessus.

Un LSR peut parfois recevoir un paquet étiqueté sur une liaison PPP ou de LAN, et peut avoir besoin de l'envoyer, disons, sur une interface LC-ATM. Le paquet entrant sera alors reçu en utilisant l'encapsulation spécifiée dans ce document, mais le paquet sortant ne sera pas envoyé en utilisant l'encapsulation spécifiée dans ce document. Dans ce cas, la procédure pour porter la valeur du "TTL sortant" est déterminée par les procédures utilisées pour porter les paquets étiquetés sur, par exemple, des interfaces LC-ATM.

3. Fragmentation et découverte de la MTU de chemin

Tout comme il est possible de recevoir un datagramme IP non étiqueté qui est trop gros pour être transmis sur sa liaison de sortie, il est possible de recevoir un paquet étiqueté qui est trop gros pour être transmis sur sa liaison de sortie.

Il est aussi possible qu'un paquet reçu (étiqueté ou non) qui était à l'origine assez petit pour être transmis sur cette liaison devienne trop gros du fait qu'une ou plusieurs étiquettes supplémentaires ont été poussées sur sa pile d'étiquettes. Dans la commutation d'étiquettes, un paquet peut grossir si des étiquettes supplémentaires sont poussées dessus. Donc, si on reçoit un paquet étiqueté avec une charge utile de trame de 1500 octets, et qu'on pousse une étiquette supplémentaire, on a besoin de le transmettre comme une trame avec une charge utile de 1504 octets.

Cette section spécifie les règles pour traiter les paquets étiquetés qui sont "trop gros". En particulier, elle donne les règles qui assurent que les hôtes qui mettent en œuvre la découverte de la MTU du chemin [RFC1191], et les hôtes qui utilisent IPv6 [RFC1885], [RFC1981], seront capables de générer des datagrammes IP qui n'ont pas besoin de la fragmentation, même si ces datagrammes sont étiquetés lorsque ils traversent le réseau.

En général, les hôtes IPv4 qui ne mettent pas en œuvre la découverte de la MTU du chemin [RFC1191] envoient des datagrammes IP qui ne contiennent pas plus de 576 octets. Comme les MTU en usage aujourd'hui sur la plupart des liaisons de données font 1500 octets ou plus, la probabilité que de tels datagrammes aient besoin d'être fragmentés, même si ils deviennent étiquetés, est très faible.

Certains hôtes qui ne mettent pas en œuvre la découverte de la MTU de chemin [RFC1191] vont générer des datagrammes IP qui contiennent 1500 octets, dans la mesure où les adresses IP de source et de destination sont sur le même sous-réseau. Ces datagrammes ne vont pas passer à travers des routeurs, et donc ne seront pas fragmentés.

Malheureusement, certains hôtes vont générer des datagrammes IP qui contiennent 1500 octets, et dont l'adresse IP de source et de destination ont le même numéro de classe de réseau. C'est le seul cas dans lequel il y a un risque de fragmentation quand de tels datagrammes se trouvent étiquetés. (Même comme cela, la fragmentation n'est probable que si le paquet doit traverser un ethernet entre le moment où il est étiqueté pour la première fois et le moment où il est désétiqueté.)

Le présent document spécifie les procédures qui permettrent de configurer le réseau de telle sorte que de grands datagrammes provenant d'hôtes qui ne mettent pas en œuvre la découverte de la MTU du chemin ne soient fragmentés qu'une seule fois, quand ils sont étiquetés pour la première fois. Ces procédures rendent possible (en supposant une configuration convenable) d'éviter d'avoir besoin de fragmenter des paquets qui ont déjà été étiquetés.

3.1 Terminologie

Par rapport à une certaine liaison de données, on peut utiliser les termes suivants :

- Charge utile de trame
  Le contenu d'une trame de liaison de données, excluant tout en-tête ou en-queue de couche de liaison de données (par exemple, les en-têtes MAC, les en-têtes de LLC, les en-têtes 802.1Q, l'en-tête PPP, les séquences de vérification de trame, etc.). Lorsque une trame porte un datagramme IP non étiqueté, la charge utile de trame est juste le datagramme IP lui-même. Lorsque une trame porte un datagramme IP étiqueté, la charge utile de trame consiste en les entrées de pile d'étiquettes et le datagramme IP.

- Taille conventionnelle maximum de charge utile de trame :
  C'est la taille maximum de charge utile de trame permise par les normes de liaison des données. Par exemple, la taille conventionnelle maximum de charge utile de trame pour ethernet est de 1500 octets.

- Vraie taille maximum de charge utile de trame :
  C'est la taille maximum de charge utile de trame qui peut être envoyée et reçue correctement par le matériel d'interface rattaché à la liaison de données. Sur les réseaux ethernet et 802.3, on estime que la vraie taille maximum de charge utile de trame est de 4 à 8 octets plus grande que la taille conventionnelle maximum de charge utile de trame (pour autant que ni un en-tête 802.1Q ni un en-tête 802.1p ne soit présent, et que ni l'un ni l'autre ne puisse être ajouté par un commutateur ou un pont lorsque un paquet est en transit vers son prochain bond). Par exemple, on estime que la plupart des équipements d'ethernet pourraient correctement envoyer et recevoir des paquets portant une charge utile de 1504 ou peut-être même 1508 octets, au moins, pour autant que l'en-tête ethernet n'ait pas un champ 802.1Q ou 802.1p. Sur les liaisons PPP, la vraie taille maximum de charge utile de trame peut être virtuellement non limitée.

- Taille effective maximum de charge utile de trame pour les paquets étiquetés :
  C'est soit la taille conventionnelle maximum de charge utile de trame, soit la vraie taille maximum de charge utile de trame, selon les capacités de l'équipement sur la liaison de données et la taille de l'en-tête de liaison de données utilisé.

- Datagramme IP initialement étiqueté :
  On suppose qu'un datagramme IP non étiqueté est reçu à un certain LSR, et que le LSR pousse une étiquette avant de transmettre le datagramme. Un tel datagramme sera appelé un datagramme initialement étiqueté à ce LSR.

- Datagramme IP précédemment étiqueté :
  C'est un datagramme IP qui a déjà été étiqueté avant d'être reçu par un certain LSR.

3.2 Taille maximum de datagramme IP initialement étiqueté

Chaque LSR qui est capable de
a) recevoir un datagramme IP non étiqueté,
b) d'ajouter une pile d'étiquette au datagramme, et
c) de transmettre le paquet étiqueté résultant,

DEVRAIT accepter un paramètre de configuration appelé "taille maximum de datagramme IP initialement étiquetée", qui peut être réglé à une valeur non négative.

Si ce paramètre de configuration est réglé à zéro, il n'a pas d'effet.

Si il est réglé à une valeur positive, il est utilisé de la façon suivante. Si :
a) un datagramme IP non étiqueté est reçu et si
b) ce datagramme n'a pas le bit DF établi dans son en-tête IP, et si
c) ce datagramme a besoin d'être étiqueté avant d'être transmis, et si
d) la taille du datagramme (avant étiquetage) excède la valeur de ce paramètre,

alors
a) le datagramme doit être cassé en fragments, dont la taille de chacun ne soit pas supérieure à la valeur du paramètre,
b) chaque fragment doit être étiqueté puis transmis.

Par exemple, si ce paramètre de configuration est réglé à une valeur de 1488, alors tout datagramme IP non étiqueté qui contient plus de 1488 octets sera fragmenté avant d'être étiqueté. Chaque fragment sera capable d'être porté sur une liaison de données à 1500 octets, sans autre fragmentation, même si jusqu'à trois étiquettes sont poussées sur sa pile d'étiquettes.

En d'autres termes, régler ce paramètre à une valeur différente de zéro permet d'éliminer toute fragmentation des datagrammes précédemment étiquetés, mais cela peut causer une fragmentation inutile de datagrammes initialement étiquetés.

Noter que le réglage de ce paramètre n'affecte pas le traitement des datagrammes IP qui ont le bit DF établi ; donc, le résultat de la découverte de la MTU du chemin n'est pas affectée par le réglage de ce paramètre.

3.3 Quand un datagramme IP étiqueté est-il trop grand ?

Un datagramme IP étiqueté dont la taille excède la taille conventionnelle maximum de charge utile de trame de la liaison de données sur laquelle il doit être transmis PEUT être considéré comme étant "trop gros".

Un datagramme IP étiqueté dont la taille excède la vraie taille maximum de charge utile de trame de la liaison de données sur laquelle il doit être transmis DOIT être considéré comme étant "trop gros".

Un datagramme IP étiqueté qui n'est pas "trop gros" DOIT être transmis sans fragmentation.

3.4 Traitement des datagrammes IPv4 étiquetés qui sont trop grands

Si un datagramme IPv4 étiqueté est "trop gros", et si le bit DF n'est pas établi dans son en-tête IP, le LSR PEUT alors éliminer le datagramme en silence.

Noter qu'éliminer un tel datagramme n'est une procédure délicate que si la "taille maximum de datagramme IP initialement étiqueté" est réglée à une valeur différente de zéro dans tous les LSR du réseau qui sont capables d'ajouter une pile d'étiquettes à un datagramme IP non étiqueté.

Si le LSR choisit de ne pas éliminer le datagramme IPv4 étiqueté qui est trop gros, ou si le bit DF est établi dans ce datagramme, il DOIT alors exécuter l'algorithme suivant :

1. Effacer les entrées de pile d'étiquettes pour obtenir le datagramme IP.
2. Soit N le nombre d'octets dans la pile d'étiquettes (c'est-à-dire,, 4 fois le nombre d'entrées de pile d'étiquettes).
3. Si le datagramme IP N'A PAS le bit "Ne pas fragmenter" établi dans son en-tête IP :
   a. le convertir en fragments, dont chacun DOIT avoir au moins N octets de moins que la taille effective maximum de charge utile de trame,
   b. ajouter devant chaque fragment le même en-tête d'étiquette qui aurait été sur le datagramme original si la fragmentation n'avait pas été nécessaire,
   c. transmettre les fragments
4. Si le datagramme IP a le bit "Ne pas fragmenter" établi dans son en-tête IP :
   a. le datagramme NE DOIT PAS être transmis.
   b. créer un message ICMP Destination injoignable :
      i. régler son champ Code [RFC0792] à "Fragmentation exigée et DF établi",
      ii. régler son champ MTU du prochain bond [RFC1191] à la différence entre la taille effective maximum de charge utile de trame et la valeur de N,
   c. Si possible, transmettre le message ICMP Destination injoignable à la source du datagramme éliminé.

3.5 Traitement des datagrammes IPv6 étiquetés qui sont trop grands

Pour traiter un datagramme IPv6 étiqueté qui est trop gros, un LSR DOIT exécuter l'algorithme suivant :

1. Effacer les entrées de la pile d'étiquettes pour obtenir le datagramme IP.
2. Soit N le nombre d'octets dans la pile d'étiquettes (c'est-à-dire, 4 fois le nombre d'entrées de pile d'étiquettes).
3. Si le datagramme IP contient plus de 1280 octets (non comprises les entrées de pile d'étiquettes) ou si il ne contient pas d'en-tête de fragment, alors :
   a. Créer un message ICMP Paquet trop gros, et régler son champ MTU de prochain bond à la différence entre la taille effective maximum de charge utile de trame et la valeur de N.
   b. Si possible, transmettre le message ICMP Paquet trop gros à la source du datagramme.
   c. Éliminer le datagramme IPv6 étiqueté.
4. Si le datagramme IP fait moins de 1280 octets, et si il contient un en-tête de fragment, alors,
   a. le convertir en fragments, dont chacun DOIT faire au moins N octets de moins que la taille effective maximum de charge utile de trame,
   b. ajouter en tête de chaque fragment le même en-tête d'étiquette qu'il y aurait eu si le datagramme original n'avait pas eu besoin d'être fragmenté,
   c. transmettre les fragments.

Le réassemblage des fragments sera fait chez l'hôte de destination.

3.6 Implications par rapport à la découverte de la MTU du chemin

Les procédures décrites ci-dessus pour le traitement des datagrammes qui ont le bit DF établi, mais qui sont "trop gros", ont un impact sur les procédures de découverte de la MTU du chemin de la [RFC1191]. Les hôtes qui mettent en œuvre ces procédures vont découvrir une MTU qui est assez petite pour permettre de pousser n étiquettes sur les datagrammes, sans qu'il soit besoin de les fragmenter, où n est le nombre d'étiquettes qui sont en fait poussées le long du chemin actuellement utilisé.

En d'autres termes, les datagrammes provenant d'hôtes qui utilisent la découverte de la MTU du chemin n'auront jamais besoin d'être fragmentés à cause de la nécessité d'y mettre un en-tête d'étiquette, ou d'ajouter de nouvelles étiquette à un en-tête d'étiquette existant. (De plus, les datagrammes provenant d'hôtes qui utilisent la découverte de la MTU du chemin

ont généralement le bit DF établi, et ne seront donc de toute façon jamais fragmentés.)

Noter que la découverte de la MTU du chemin ne va fonctionner correctement que si, au point où la fragmentation d'un datagramme IP étiqueté doit survenir, il est possible de provoquer l'acheminement d'un message ICMP Destination injoignable à l'adresse de source du paquet. Voir au paragraphe 2.3.

Si il n'est pas possible de transmettre un message ICMP de l'intérieur d'un "tunnel" MPLS à l'adresse de source d'un paquet, mais si la configuration du réseau rend possible au LSR du côté émetteur du tunnel de recevoir les paquets qui doivent passer à travers le tunnel, mais qui sont trop gros pour passer non fragmentés à travers le tunnel, alors :
- Le LSR à l'extrémité émettrice du tunnel DOIT être capable de déterminer globalement la MTU du tunnel. Il PEUT le faire en envoyant des paquets à travers le tunnel à l'extrémité de réception du tunnel, et en effectuant la découverte de la MTU du chemin avec ces paquets.
- Chaque fois que le point d'extrémité émetteur du tunnel a besoin d'envoyer un paquet dans le tunnel, et que ce paquet a le bit DF établi, et qu'il excède la MTU du tunnel, le point d'extrémité émetteur du tunnel DOIT envoyer le message ICMP Destination injoignable à la source, avec le code "Fragmentation exigée et DF établi", et le champ MTU du prochain bond réglé comme décrit ci-dessus.

4. Transport des paquets étiquetés sur PPP

Le protocole point à point (PPP) [RFC1661] fournit une méthode standard pour le transport de datagrammes multi-protocoles sur des liaisons point à point. PPP définit un protocole extensible de contrôle de liaison et propose une famille de protocoles de contrôle de réseau pour établir et configurer différents protocoles de couche réseau.

La présente section définit le protocole de contrôle de réseau pour établir et configurer la commutation d'étiquettes sur PPP.

4.1 Introduction

PPP a trois composants principaux :
1. une méthode pour encapsuler les datagrammes multi-protocoles,
2. un protocole de contrôle de liaison (LCP, *Link Control Protocol*) pour établir, configurer, et vérifier la connexion de liaison des données,
3. une famille de protocoles de contrôle du réseau pour établir et configurer les différents protocoles de couche réseau.

Afin d'établir les communications sur une liaison en point à point, chaque extrémité de la liaison PPP doit d'abord envoyer des paquets de LCP pour configurer et vérifier la liaison de données. Après l'établissement de la liaison et la négociation des facilités optionnelles nécessaires pour le LCP, PPP doit envoyer des paquets de "protocole de contrôle MPLS" pour permettre la transmission de paquets étiquetés. Une fois que le "protocole de contrôle MPLS" a atteint l'état Ouvert, les paquets étiquetés peuvent être envoyés sur la liaison.

La liaison va rester configurée pour la communication jusqu'à ce que des paquets explicites de LCP ou de protocole de contrôle MPLS closent la liaison, ou jusqu'à ce que survienne quelque événement externe (l'expiration d'un temporisateur d'inactivité, ou une intervention de l'administrateur du réseau).

4.2 Protocole de contrôle de réseau PPP pour MPLS

Le protocole de contrôle de MPLS (MPLSCP, *MPLS Control Protocol*) est chargé d'activer et de désactiver l'utilisation de la commutation d'étiquettes sur une liaison PPP. Il utilise le même mécanisme d'échange de paquets que le protocole de contrôle de liaison (LCP, *Link Control Protocol*). Les paquets MPLSCP ne peuvent pas être échangés tant que PPP n'a pas atteint la phase protocole de couche réseau. Les paquets MPLSCP reçus avant cette phase devraient être éliminés en silence.

Le protocole de contrôle MPLS est exactement le même que le protocole de contrôle de liaison [RFC1661] avec les exceptions suivantes :

1. Modifications de trame
   Le paquet peut utiliser toutes les modifications au format de base de trame qui ont été négociées durant la phase d'établissement de liaison.

2. Champ Protocole de couche de liaison des données
   Exactement un paquet MPLSCP est encapsulé dans le champ Information PPP, où le champ Protocole PPP indique le type hexadécimal 8281 (MPLS).

3. Champ Code
   Seuls les codes 1 à 7 (Demande-de-configuration, Accusé-de-réception-de-configuration, Non-accusé-de-réception-de-configuration, Rejet-de-Configuration, Demande-de-fin, Accusé-de-réception-de-fin et Rejet-de-code) sont utilisés. Les autres codes devraient être traités comme non reconnus et devraient résulter en Rejet-de-code.

4. Fins de temporisation
   Les paquets MPLSCP ne peuvent pas être échangés jusqu'à ce que PPP ait atteint la phase de protocole de couche réseau. Une mise en œuvre devrait être prête à attendre que se terminent l'authentification et la détermination de la qualité de la liaison avant de terminer l'attente d'un Accusé-de-réception-de-configuration ou d'une autre réponse. Il est suggéré qu'une mise en œuvre n'abandonne qu'après une intervention de l'utilisateur ou une durée configurable.

5. Types d'option de configuration
   Aucune.

4.3 Envoi des paquets étiquetés

Avant qu'aucun paquet étiqueté puisse être communiqué, PPP doit atteindre la phase de protocole de couche réseau, et le protocole de contrôle MPLS doit atteindre l'état Ouvert.

Exactement un paquet étiqueté est encapsulé dans le champ Information de PPP, où le champ PPP Protocole indique soit le type hexadécimal 0281 (envoi individuel MPLS) soit le type hexadécimal 0283 (diffusion groupée MPLS). La longueur maximum d'un paquet étiqueté transmis sur une liaison PPP est la même que la longueur maximum du champ Information d'un paquet encapsulé PPP.

Le format du champ Information lui-même est défini à la section 2.

Noter que deux codets sont définis pour les paquets étiquetés ; un pour la diffusion groupée et un pour l'envoi individuel. Une fois que MPLSCP a atteint l'état Ouvert, les étiquettes commutées en diffusion groupée et en envoi individuel peuvent être envoyées sur la liaison PPP.

4.4 Options de configuration du protocole de contrôle de commutation d'étiquettes

Il n'y a pas d'option de configuration.

5. Transport des paquets étiqueté sur un support de LAN

Exactement un paquet étiqueté est porté dans chaque trame.

Les entrées de pile d'étiquettes précèdent immédiatement l'en-tête de couche réseau, et suivent tout en-tête de couche de liaison des données, incluant, par exemple, tout en-tête 802.1Q qui pourrait exister.

La valeur d'ethertype hexadécimale 8847 est utilisée pour indiquer qu'une trame porte un paquet MPLS en envoi individuel.

La valeur d'ethertype hexadécimale 8848 est utilisée pour indiquer qu'une trame porte un paquet MPLS en diffusion groupée.

Ces valeurs d'éthertype peuvent être utilisées avec l'encapsulation ethernet ou l'encapsulation 802.3 LLC/SNAP pour porter les paquets étiquetés. La procédure pour choisir laquelle de ces deux encapsulations utiliser sort du domaine d'application du présent document.

6. Considérations relatives à l'IANA

Les valeurs d'étiquette de 0 à 15 inclus ont une signification spéciale, comme spécifié dans le présent document, ou selon les affectations qui seront faites par l'IANA.

Dans le présent document, les valeurs d'étiquettes de 0 à 3 sont spécifiées au paragraphe 2.1.

Les valeurs d'étiquettes de 4 à 15 peuvent être affectées par l'IANA, sur la base du consensus de l'IETF.

7. Considérations pour la sécurité

L'encapsulation MPLS qui est spécifiée ici ne soulève aucun problème de sécurité qui ne soit déjà présent dans l'architecture MPLS [RFC3031] ou dans l'architecture du protocole de couche réseau contenu au sein de l'encapsulation.

Deux considérations de sécurité sont héritées de l'architecture MPLS et peuvent être mentionnées ici :

- Certains routeurs peuvent mettre en œuvre des procédures de sécurité qui dépendent de ce que l'en-tête de couche réseau est à une place fixée par rapport à l'en-tête de couche de liaison des données. Ces procédures ne vont pas fonctionner lorsque l'encapsulation MPLS est utilisée, parce que cette encapsulation est de taille variable.

- Une étiquette MPLS a sa signification par suite d'un accord entre le LSR qui met l'étiquette dans la pile d'étiquettes (le "rédacteur d'étiquette") et le LSR qui interprète cette étiquette (le "lecteur d'étiquette"). Cependant, la pile d'étiquettes ne fournit aucun moyen de déterminer qui était le rédacteur d'étiquette pour quelque étiquette que ce soit. Si les paquets étiquetés sont acceptés de sources qui ne sont pas de confiance, le résultat peut être que des paquets seront acheminés d'une manière illégitime.

8. Propriété intellectuelle

L'IETF a reçu des notifications de revendications de droits de propriété intellectuelle à l'égard de tout ou partie de la spécification contenue dans le présent document. Pour plus d'informations, consulter la liste en ligne des revendications de droits.

9. Adresse des auteurs

Eric C. Rosen
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824
mél : erosen@cisco.com

Dan Tappan
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824
mél : tappan@cisco.com

Yakov Rekhter
Juniper Networks
1194 N. Mathilda Avenue
Sunnyvale, CA 94089
mél : yakov@juniper.net

Guy Fedorkow
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824
mél : fedorkow@cisco.com

Dino Farinacci
Procket Networks, Inc.
3910 Freedom Circle, Ste. 102A
Santa Clara, CA 95054
mél: dino@procket.com

Tony Li
Procket Networks, Inc.
3910 Freedom Circle, Ste. 102A
Santa Clara, CA 95054
mél : tli@procket.com

Alex Conta
TranSwitch Corporation
3 Enterprise Drive
Shelton, CT, 06484
mél : aconta@txc.com

10. Références

[RFC0792] J. Postel, "Protocole du message de contrôle Internet – Spécification du protocole du programme Internet DARPA", STD 5, septembre 1981.

[RFC1191] J. Mogul et S. Deering, "Découverte de la MTU de chemin", novembre 1990.

[RFC1661] W. Simpson, éditeur, "Protocole point à point (PPP)", STD 51, juillet 1994. *(MàJ par la RFC 2153)*

[RFC1885] A. Conta, S. Deering, "Protocole de contrôle de message Internet (ICMPv6) pour le protocole Internet

version 6 (IPv6)", décembre 1995. (*Obsolète, voir* RFC2463) *(P.S.)*

[RFC1981] J. McCann, S. Deering, J. Mogul, "Découverte de la MTU de chemin pour IP version 6", août 1996. *(D.S.)*

[RFC2113] D. Katz, "Option d'alerte de routeur IP", février 1997.

[RFC2119] S. Bradner, "Mots clés à utiliser dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997.

[RFC3031] E. Rosen, A. Viswanathan, R. Callon, "Architecture de commutation d'étiquettes multi protocoles", janvier 2001. *(P.S.)*

[RFC3035] B. Davie et autres, "Utilisation de MPLS dans la commutation de circuit virtuel LDP et ATM", janvier 2001. *(P.S.)*

11. Déclaration complète de droits de reproduction

Remerciement

Le financement de la fonction d'édition des RFC est actuellement fourni par l'Internet Society.