Groupe de travail Réseau
Request for Comments : 4781
Catégorie : Sur la voie de la normalisation
Traduction Claude Brière de L'Isle

Y. Rekhter, Juniper Networks
R. Aggarwal, Juniper Networks
janvier 2007

Mécanisme de redémarrage en douceur pour BGP avec MPLS

Résumé
Un mécanisme pour BGP qui aide à minimiser les effets négatifs sur l'acheminement causés par le redémarrage de BGP a déjà été développé et est décrit dans un document séparé ("Mécanisme de redémarrage en douceur pour BGP"). Le présent document étend ce mécanisme pour minimiser les effets négatifs sur la transmission MPLS causés par le redémarrage du plan de contrôle du routeur de commutation d'étiquettes (LSR, *Label Switching Route*) et spécifiquement par le redémarrage de son composant BGP quand BGP est utilisé pour porter les étiquettes MPLS et que le LSR est capable de préserver l'état de transmission MPLS à travers le redémarrage.

Le mécanisme décrit dans ce document est neutre par rapport aux types d'adresses portées dans le champ Informations d'accessibilité de couche réseau (NLRI, *Network Layer Reachability Information*) de BGP. À ce titre, il fonctionne en conjonction avec toutes les familles d'adresses qui peuvent être portées dans BGP (par exemple, IPv4, IPv6, etc.).

Table des matières

1. Introduction

Dans le cas où un routeur de commutation d'étiquettes (LSR, *Label Switching Router*) pourrait préserver son état de transmission MPLS à travers les redémarrages de son plan de contrôle, et spécifiquement de son composant BGP, et si BGP est utilisé pour porter des étiquettes MPLS (par exemple, comme spécifié dans la [RFC3107]), il peut être désirable de ne pas perturber les LSP qui passent à travers ce LSR (et spécifiquement, les LSP établis par BGP) après une défaillance ou un redémarrage du composant BGP au plan de contrôle. Dans le présent document, on décrit un mécanisme qui permet d'atteindre cet objectif. Le mécanisme décrit dans le présent document fonctionne en conjonction avec le mécanisme

spécifié dans la [RFC4724]. Le mécanisme décrit dans le présent document ne fait pas de restrictions sur les types d'adresses (familles d'adresses) qu'il peut prendre en charge.

Le mécanisme décrit dans le présent document est applicable à tous les LSR, ceux qui ont la capacité de préserver l'état de transmission durant le redémarrage de BGP et ceux qui ne l'ont pas (bien que ces derniers n'aient besoin de mettre en œuvre qu'un sous ensemble de ce mécanisme). Prendre en charge un sous ensemble du mécanisme décrit ici, par les LSR qui ne peuvent pas préserver leur état de transmission MPLS à travers le redémarrage ne réduirait pas l'impact négatif sur le trafic MPLS causé par le redémarrage de leur plan de contrôle. Cependant, l'impact va être minimisé si leur ou leurs voisins sont capables de préserver l'état de transmission à travers le redémarrage de leur plan de contrôle, et si ils mettent en œuvre le mécanisme décrit ici. Ce sous ensemble inclut toutes les procédures décrites dans le présent document, excepté les procédures des paragraphes 4.1, 4.2, 4.3, et de la Section 5.

Pour faire court, par "état de transmission MPLS" on entend une des transpositions suivantes :
<étiquette entrante en (étiquette sortante, prochain bond)>
<classe d'équivalence de transmission en (étiquette sortante, prochain bond)>
<étiquette entrante en saut d'étiquette, prochain bond>
<étiquette entrante en saut d'étiquette>

Dans le contexte du présent document, l'état de transmission auquel on se réfère dans la [RFC4724] signifie l'état de transmission MPLS, comme défini ci-dessus. Le terme "prochain bond" se réfère au prochain bond annoncé dans BGP.

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" en majuscules dans ce document sont à interpréter comme décrit dans le BCP 14, [RFC2119].

2. Exigences générales

Avant tout, un LSR DOIT mettre en œuvre le mécanisme de redémarrage en douceur pour BGP, comme spécifié dans la [RFC4724]. Ensuite, le LSR DEVRAIT être capable de préserver son état de transmission MPLS à travers le redémarrage de son plan de contrôle (incluant le redémarrage de BGP). Troisièmement, pour les liens entre <classe d'équivalence de transmission (FEC) -> étiquette> distribués via BGP, le LSR DEVRAIT être capable soit (a) de reconstruire les mêmes liens qu'avait le LSR avant le redémarrage (Section 4), soit (b) de créer de nouveaux liens <FEC -> étiquette> après le redémarrage, tout en maintenant temporairement l'état de transmission MPLS correspondant à la fois aux liens d'avant le redémarrage, et aux nouveaux liens créés (voir à la Section 5). Quatrièmement, tant que le LSR conserve l'état de transmission MPLS que le LSR a préservé à travers le redémarrage, les étiquettes provenant de cet état ne peuvent pas être utilisées pour créer de nouveaux liens d'étiquettes locales (mais pourraient être utilisées pour reconstruire les liens existants, selon les procédures de la Section 4). Finalement, pour chaque prochain bond, si le prochain bond est accessible via un chemin d'étiquettes commuté (LSP, *Label Switched Path*) alors le LSR qui redémarre DOIT être capable de préserver l'état de transmission MPLS associé à ce LSP à travers le redémarrage.

Dans le scénario où le lien d'étiquette sur un LSR est créé/maintenu non seulement par le composant BGP du plan de contrôle, mais aussi par d'autres composants de protocole (par exemple, LDP, RSVP-TE) et où le LSR prend en charge le redémarrage des composants individuels du plan de contrôle qui créent/maintiennent le lien d'étiquette (par exemple, le redémarrage de BGP, mais pas le redémarrage de LDP) le LSR DOIT être capable de préserver à travers le redémarrage les informations sur quel protocole a alloué quelles étiquettes.

Après le redémarrage du LSR, il DOIT suivre les procédures spécifiées dans la [RFC4724]. De plus, si le LSR est capable de préserver son état de transmission MPLS à travers le redémarrage, le LSR DEVRAIT l'annoncer à ses voisins en réglant de façon appropriée le fanion pour le champ Famille d'adresses dans la capacité Redémarrage en douceur pour toutes les paires applicables de AFI/SAFI.

3. Annonces de capacités

Un LSR qui prend en charge le mécanisme décrit dans le présent document l'annonce à son homologue en utilisant la capacité de redémarrage en douceur, comme spécifié dans la [RFC4724]. L'identifiant de famille d'adresse suivante (SAFI,

Subsequent Address Family Identifier) dans la capacité annoncée DOIT indiquer que le champ Informations d'accessibilité de la couche réseau (NLRI, *Network Layer Reachability Information*) ne porte pas seulement les informations d'adressage, mais aussi des étiquettes (voir la [RFC3107] pour un exemple où les NLRI portent des étiquettes).

4. Procédures pour le LSR qui redémarre

Les procédures de cette Section s'appliquent quand un LSR qui redémarre est capable de reconstruire les mêmes liens <FEC -> étiquette> que le LSR avait avant le redémarrage.

Les procédures décrites dans cette Section sont conceptuelle et n'ont pas à être mises en œuvre précisément comme décrit, pour autant que les mises en œuvre prennent en charge la fonction décrite et que leur comportement visible de l'extérieur soit le même.

Une fois que LSR a achevé son choix de chemin (comme spécifié au paragraphe 4.1, "Procédures pour le locuteur qui redémarre", de la [RFC4724]), alors en plus de ces procédures, le LSR effectue un de ce qui suit :

4.1 Cas 1

Ce qui suit s'applique quand (a) le meilleur chemin choisi par le LSR a été reçu avec une étiquette, (b) cette étiquette n'est pas un NUL implicite, et (c) le LSR annonce ce chemin avec lui-même comme prochain bond.

Dans ce cas, le LSR cherche son état de transmission MPLS (celui qui a été préservé à travers le redémarrage) pour une entrée avec <étiquette sortante, prochain bond> égale à celle du chemin reçu. Si il trouve une telle entrée, le LSR ne marque plus l'entrée comme périmée. De plus, si l'entrée est du type <étiquette entrante, (étiquette sortante, prochain bond)> plutôt que <Classe d'équivalence de transmission (FEC), (étiquette sortante, prochain bond)>, le LSR utilise l'étiquette entrante provenant de l'entrée quand il annonce le chemin à ses voisins. Si l'entrée trouvée n'a pas d'étiquette entrante, ou si aucune entrée n'est trouvée, le LSR alloue une nouvelle étiquette quand il annonce le chemin à ses voisins (en supposant qu'il y a des voisins auxquels le LSR doit annoncer le chemin avec une étiquette).

4.2 Cas 2

Ce qui suit s'applique quand (a) le meilleur chemin choisi par le LSR a été reçu sans étiquette, avec une étiquette Nulle implicite, ou que le chemin a été généré par le LSR ; (b) le LSR annonce ce chemin avec lui-même comme prochain bond; et (c) le LSR doit générer une étiquette (non Nulle implicite) pour le chemin.

Dans ce cas, le LSR cherche son état de transmission MPLS pour une entrée qui indique que le LSR doit effectuer un saut d'étiquette, et que le prochain bond est égal au prochain bond du chemin considéré. Si une telle entrée est trouvée, alors le LSR utilise l'étiquette entrante provenant de l'entrée quand il annonce le chemin à ses voisins. Si une telle entrée n'est pas trouvée, le LSR alloue une nouvelle étiquette quand il annonce le chemin à ses voisins.

La description du paragraphe ci-dessus suppose que le LSR génère la même étiquette pour tous les chemins avec le même prochain bond. Si ce n'est pas le cas et si le LSR génère une étiquette unique pour chacun de ces chemins, alors le LSR a besoin de préserver à travers le redémarrage non seulement la transposition <étiquette entrante, (étiquette sortante, prochain bond)>, mais aussi la classe d'équivalence de transmission (FEC) associée à cette transposition. Dans ce cas le LSR va chercher son état de transmission MPLS pour une entrée qui (a) indique un saut d'étiquette (ce qui signifie pas d'étiquette sortante) (b) indique que le prochain bond est égal au prochain bond du chemin, et (c) a la même FEC que le chemin. Si une telle entrée est trouvée, alors le LSR utilise l' étiquette entrante provenant de l'entrée quand il annonce le chemin à ses voisins. Si une telle entrée n'est pas trouvée, le LSR alloue une nouvelle étiquette quand il annonce le chemin à ses voisins.

4.3 Cas 3

Ce qui suit s'applique quand le LSR ne règle pas le prochain bond BGP à lui-même.

Dans ce cas, le LSR, quand il annonce son meilleur chemin pour une NLRI particulière, utilise juste l'étiquette qui a été reçue avec ce chemin. Et si le chemin a été reçu sans étiquette, le LSR annonce aussi le chemin sans étiquette. De toutes façons, le LSR n'alloue pas d'étiquette pour ce chemin.

5. Procédures de remplacement pour le LSR qui redémarre

Dans cette Section, on décrit une solution de remplacement des procédures décrites à la Section "Procédures pour le LSR qui redémarre" (Section 4).

Les procédures de cette Section s'appliquent quand un LSR qui redémarre ne reconstruit pas les mêmes liens <FEC -> étiquette> que ceux qu'il avait avant le redémarrage, mais crée plutôt de nouveaux liens <FEC -> étiquette> après le redémarrage, tout en maintenant temporairement l'état de transmission MPLS correspondant à la fois aux liens d'avant le redémarrage ainsi qu'aux liens nouvellement créés.

Les procédures décrites dans cette section exigent que pour l'utilisation par le redémarrage BGP en douceur, le LSR DEVRAIT avoir (au moins) autant d'étiquettes non allouées que d'étiquettes allouées pour les liens <FEC -> étiquette> distribués par BGP. Ces dernières forment l'état de transmission MPLS que le LSR a réussi à préserver à travers le redémarrage. Les premières sont utilisées pour allouer les étiquettes après le redémarrage.

Pour créer des (nouveaux) liens d'étiquette locaux après le redémarrage, le LSR utilise les étiquettes non allouées (c'est tout à fait la procédure normale).

Le LSR DEVRAIT conserver l'état de transmission MPLS que le LSR a préservé à travers le redémarrage au moins jusqu'à ce que le LSR envoie un marqueur Fin-de-RIB à tous ses voisins (à ce moment le LSR a déjà terminé son processus de choix de chemin, et aussi annoncé son Adj-RIB-Out à ses voisins). Le LSR PEUT conserver l'état de transmission même un peu plus longtemps (la quantité de temps supplémentaire PEUT être contrôlée par configuration sur le LSR) afin de permettre aux voisins de recevoir et traiter les chemins qui ont été annoncés par le LSR. Après cela, le LSR DEVRAIT supprimer l'état de transmission MPLS qu'il avait préservé à travers le redémarrage.

Noter que alors qu'un LSR est dans le processus de redémarrage, le LSR peut avoir non un, mais deux liens d'étiquette locaux pour un chemin BGP donné – un qui a été conservé d'avant le redémarrage, et un autre qui a été créé après le redémarrage. Une fois que le LSR a achevé son redémarrage, l'ancien sera supprimé. Cependant, ces deux liens vont avoir la même étiquette sortante (et le même prochain bond).

6. Procédures pour un voisin d'un LSR qui redémarre

Le voisin d'un LSR qui redémarre (la terminologie de routeur receveur utilisée dans la [RFC4724]) suit les procédures spécifiées dans la [RFC4724]. De plus, le voisin traite les étiquettes MPLS reçues du LSR qui redémarre de la même façon qu'il traite les chemins reçus du LSR qui redémarre (à la fois avant et après le redémarrage).

Remplacer les chemins périmés par les mises à jour d'acheminement reçues du LSR qui redémarre implique de remplacer/mettre à jour les étiquettes MPLS appropriées.

De plus, si les fanions dans la capacité de redémarrage en douceur reçus du LSR qui redémarre indiquent que le LSR n'était pas capable de conserver son état MPLS à travers le redémarrage, le voisin DEVRAIT immédiatement supprimer toutes les NLRI et les étiquettes MPLS associées qu'il avait précédemment acquises via BGP du LSR qui redémarre.

Un LSR, une fois qu'il a créé un lien entre une étiquette et une classe d'équivalence de transmission (FEC) DEVRAIT conserver la valeur de l'étiquette dans ce lien pendant tout le temps que le LSR a un chemin pour la FEC dans le lien. Si le chemin pour la FEC disparaît et ensuite réapparaît à nouveau plus tard, il peut alors en résulter l'utilisation d'une valeur d'étiquette différente, car quand le chemin réapparaît, le LSR va créer un nouveau lien <étiquette, FEC>.

Pour minimiser les mauvais acheminements potentiels causés par le changement d'étiquette, quand il crée un nouveau lien <étiquette, FEC>, le LSR DEVRAIT prendre l'étiquette la plus récemment utilisée. Une fois qu'un LSR a libéré une étiquette, il NE DEVRA PAS réutiliser cette étiquette pour annoncer un lien <étiquette, FEC> à un voisin qui prend en charge le redémarrage en douceur pendant au moins la durée de redémarrage, comme annoncé par le voisin au LSR. Cette règle DEVRA s'appliquer à toute libération d'étiquette à tout moment.

7. Comparaison entre les procédures de remplacement pour le LSR qui redémarre

Les procédures décrites à la Section 4 impliquent plus de frais généraux de calcul sur le routeur qui redémarre que les procédures de la Section 5.

Les procédures décrites à la Section 5 exigent deux fois plus d'étiquettes que celles de la Section 4.

Les procédures décrites à la Section 4 causent moins de changements à l'état de transmission MPLS chez les voisins du routeur qui redémarre que les procédures décrites à la Section 5.

En principe, il est possible à un LSR d'utiliser les procédures décrites à la Section 4 pour certaines AFI/SAFI et les procédures décrites à la Section 5 pour d'autres AFI/SAFI.

8. Considérations sur la sécurité

Les considérations relatives au protocole BGP [RFC4271] restent pertinentes.

De plus, le mécanisme décrit ici rend les LSR qui le mettent en œuvre vulnérables à des attaques de déni de service supplémentaires.

Un intrus peut se faire passer pour un homologue BGP afin de forcer une défaillance et reconnexion de la connexion TCP, où l'intrus règle le bit État de transmission (F) (comme défini dans la [RFC4724]) à 0 à la reconnexion. Cela force la libération de toutes les étiquettes reçues de l'homologue.

Un intrus pourrait intercepter le trafic entre homologues BGP et changer le réglage du bit État de transmission (F) pour être à 0. Cela force la libération de toutes les étiquettes reçues de l'homologue.

Ces attaques peuvent être contrées en utilisant un schéma d'authentification entre les homologues BGP, comme le schéma mentionné dans la [RFC2385].

Comme avec BGP portant des étiquettes, un problème de sécurité peut exister si une mise en œuvre de BGP continue d'utiliser des étiquettes après l'expiration de la session BGP qui a causé leur première utilisation. Cela peut arriver si le LSR en amont détecte la défaillance de la session après que le LSR en aval a libéré et réutilisé l'étiquette. Le problème est évident avec l'espace d'étiquettes au niveau de la plate-forme et pourrait résulter en un déroutement des données sur des destinations autres que celles prévues ; et il est concevable que ces comportements puissent être exploités délibérément, soit pour obtenir des services sans autorisation, soit pour dénier les services aux autres.

Dans le présent document, la validité de la session BGP peut être étendue du temps de redémarrage, et la session peut être rétablie dans cette période. Après l'expiration du temps de redémarrage, la session doit être considérée comme ayant échoué, et le même problème de sécurité s'applique comme décrit ci-dessus.

Cependant, le LSR en aval peut déclarer l'échec de la session avant l'expiration de son temps de redémarrage. Cela augmente la période durant laquelle le LSR aval pourrait réallouer l'étiquette alors que le LSR amont continue de transmettre des données en utilisant le vieil usage de l'étiquette. Pour réduire la portée de ce problème, le présent document exige que les étiquettes ne soient pas réutilisées pendant au moins le temps de redémarrage.

9. Remerciements

Nous tenons à remercier Chaitanya Kodeboyina et Loa Andersson de leur relecture et commentaires. L'approche décrite dans la Section 5 se fonde sur l'idée suggérée par Manoj Leelanivas.

10. Références

10.1 Références normatives

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

(MàJ par RFC8174)

[RFC2385] A. Heffernan, "Protection des sessions de BGP via l'option de signature MD5 de TCP", août 1998. *(P.S. ; MàJ par la* RFC6691) *; remplacée par* RFC5925)

[RFC4271] Y. Rekhter, T. Li et S. Hares, "Protocole de routeur frontière version 4 (BGP-4)", janvier 2006. *(D.S.) (MàJ par* RFC6608, RFC8212)

[RFC4724] S. Sangli et autres, "Mécanisme de redémarrage en douceur pour BGP", janvier 2007. *(P.S.)*

10.2 Références pour information

[RFC3107] Y. Rekhter et E. Rosen, "Portage des informations d'étiquette dans BGP-4", mai 2001. (*MàJ par* RFC6790, RFC8277)

Adresse des auteurs

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

Rahul Aggarwal
Juniper Networks
1194 N.Mathilda Ave
Sunnyvale, CA 94089
mél : rahul@juniper.net

Déclaration complète de droits de reproduction

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