Agrégation dans les réseaux MPLS

Miled Tezeghdanti
{miled@rennes.enst-bretagne.fr}
Tél : (+33) 2 99 12 70 28
Fax : (+33) 2 99 12 70 30
ENST Bretagne, Rennes

Jean-Marie Bonnin
{bonnin@rennes.enst-bretagne.fr}
Tél : (+33) 2 99 12 70 07
Fax : (+33) 2 99 12 70 30
ENST Bretagne, Rennes

Bernard Cousin
{bcousin@irisa.fr}
Tél : (+33) 2 99 84 73 33
Fax : (+33) 2 99 84 71 71
IRISA, Rennes

Laurent Toutain
{Laurent.Toutain@enst-bretagne.fr}
Tél : (+33) 2 99 12 70 26
Fax : (+33) 2 99 12 70 30
ENST Bretagne, Rennes

Résumé

Avec l'augmentation du nombre des utilisateurs et l'apparition des nouvelles applications multimédia, Internet rencontre des sérieux problèmes pour garantir une qualité de service acceptable par ce genre d'applications. Un mécanisme d'acheminement rapide est nécessaire pour résoudre ce problème. La commutation de labels a été présentée comme la solution miracle, elle consiste à commuter le trafic IP au niveau de la couche Liaison de données tout en conservant la flexibilité des mécanismes d'acheminement traditionnels. L'IETF propose l'architecture MPLS comme standard de cette technologie. Dans les réseaux de grande taille (réseaux de transit) où plusieurs millions de flux de données sont acheminés à travers les routeurs, l'agrégation est une autre technique qui permet d'augmenter les performances des routeurs via la diminution de la taille des tables de routage. L'objet de cet article est l'étude de la combinaison de MPLS et de l'agrégation pour obtenir un gain considérable de performances.

Mots clés

Internet, réseau à haut débit, commutation de labels, MPLS, agrégation, LDP.

Abstract

With the increasing number of users and the emergence of new multimedia applications, Internet has serious problems to deliver acceptable quality of service to these applications. A high speed forwarding

mechanism is needed to resolve this problem. Label Switching has been presented as the miracle solution, it replaces IP traditional forwarding paradigms by link layer switching. The IETF proposes MPLS architecture as the standard of this technology. In huge networks (transit networks) where millions of data flows are processed by routers, aggregation is another technique which increases routers' performances by reducing the size of routing tables. The goal of this article is to study the combination of MPLS and aggregation to obtain good performances.

Keywords

Internet, High Speed Network, Label Switching, MPLS, Aggregation, LDP.

1 Introduction

Internet est sans doute le moyen de communication qui a marqué cette fin de siècle. Initialement conçu pour interconnecter quelques sites de recherche et académiques aux États-Unis, il couvre actuellement la majorité du globe terrestre.

De nouvelles applications telles que la téléphonie sur IP, la vidéo et la visio-conférence vont voir le jour grâce à l'apparition des réseaux à haut débit (Fast Ethernet, ATM, Giga Ethernet, etc.).

Afin de répondre aux besoins de ces applications, Internet doit mettre en oeuvre des nouveaux mécanismes et protocoles qui assurent un service d'acheminement à haut débit. Les mécanismes conventionnels de routage ne sont plus suffisants. En effet, un routeur IP manipule les paquets un par un. Il extrait l'adresse de destination de chaque paquet et détermine, en scrutant la table de routage, l'entrée qui possède le plus long préfixe correspondant à cette adresse. À partir de cette entrée, il détermine le prochain routeur vers qui envoyer le paquet. Enfin, il décremente le champ TTL *(Time To Live)* et met à jour le champ de contrôle d'erreur de l'en-tête. Pour chaque paquet, cette procédure est répétée autant de fois qu'il y a de routeurs traversés. Il est évident que c'est un processus complexe qui consomme beaucoup de temps.

Au printemps 1996, Ipsilon une «*start-up*» américaine, de nos jours possédée par Nokia, a présenté le concept de commutation IP ("IP Switching") [Newman 96], où les modes de fonctionnement des routeurs Internet et des commutateurs ATM ont été combinés. Plusieurs autres parties ont depuis proposé leurs propres versions : «*Tag Switching*» de Cisco[Katz 97], *CSR* de Toshiba [Katsube 97] et *ARIS* d'IBM [Viswanathan 97]. Le dernier proposé étant le futur standard MPLS «*MultiProtocol Label Switching*» de l'IETF [Callon 99] [Rosen 1 99].

MPLS est une architecture de réseau qui permet de combiner l'efficacité d'ATM et la flexibilité d'IP. Il commute les paquets au niveau Liaison de données et utilise les protocoles de routage classiques (RIP, OSPF, BGP, etc.) pour établir les différents chemins. Il permet d'augmenter les performances tout en assurant une scalabilité du réseau. Cependant, un chemin au niveau MPLS est établi pour chaque entrée de la table de routage et sert à acheminer le trafic relatif à cette entrée. Comme les tables de routage d'un routeur d'interconnexion d'Internet contiennent plus que 60 000 entrées, les tables de commutation auront une taille très grande. Ainsi, bien que plus rapide que l'algorithme de best match appliqué sur les tables de routage, le temps de recherche

reste important. L'agrégation consiste à regrouper plusieurs entrées de la table de commutation en une seule, ce qui permet de réduire la taille de la table de commutation et par conséquent d'améliorer les performances.

Dans cet article, nous présentons l'architecture MPLS, qui est en cours de spécification, et ses mécanismes associés. Ensuite, nous proposons des méthodes qui permettent d'agréger le trafic dans un domaine MPLS.

2 MPLS

MPLS [Rosen 1 99] est un protocole en cours de standardisation par un groupe de travail au sein de l'IETF. Il sera le futur standard de la technologie "*label switching*". Il s'appuie sur les propositions existantes : IP Switching d'Ipsilon [Newman 96], Tag Switching de Cisco [Katz 97], ARIS (Aggregate Route-based IP Switching) d'IBM [Viswanathan 97] et CSR (Cell Switched Router) de Toshiba [Katsube 97].

MPLS se propose de fournir un bon rapport prix/performance, tout en assurant une indépendance par rapport au protocole de la couche Réseau (IPv4, IPv6, IPX, AppleTalk, etc.) et de la technologie de la couche Liaison de données (ATM, FR, Ethernet, etc.).

2.1 Commutation de labels

L'idée fondamentale de la commutation de labels consiste à remplacer les mécanismes conventionnels de routage par des mécanismes plus rapides. L'acheminement des paquets n'est plus basé sur l'analyse de l'adresse de destination. Un label de taille fixe, ajouté au paquet sert comme index à une base d'informations pour déterminer le port de sortie du paquet et la nouvelle valeur du label. L'analyse de l'en-tête réseau du paquet n'est faite qu'une seule fois à l'entrée du réseau. Cette analyse permet de choisir le label adéquat à appliquer au paquet. Quand ce dernier quitte le domaine MPLS, le label est enlevé et l'acheminement est assuré en analysant l'en-tête IP.

Un routeur qui supporte la commutation de labels est appelé LSR ("Label Switching Router"). Les paquets qui possèdent le même label au niveau d'un LSR donné appartiennent à une même classe d'équivalence appelée FEC ("Forwarding Equivalence Class") et reçoivent le même traitement au sein du domaine. Le chemin pris par tous les paquets d'une même FEC est appelé chemin commuté ou LSP ("Label Switched Path"). Il est constitué par l'ensemble des LSR traversés par un paquet appartenant à cette FEC (voir figure 1).

2.2 Distribution de labels

Les informations collectées par les protocoles de routage sont utilisées pour attribuer puis distribuer les labels aux voisins MPLS. Un LSR reçoit une association d'un label sortant de la part du prochain routeur d'une FEC à laquelle il associe un label entrant et distribue le résultat aux pairs MPLS en amont.

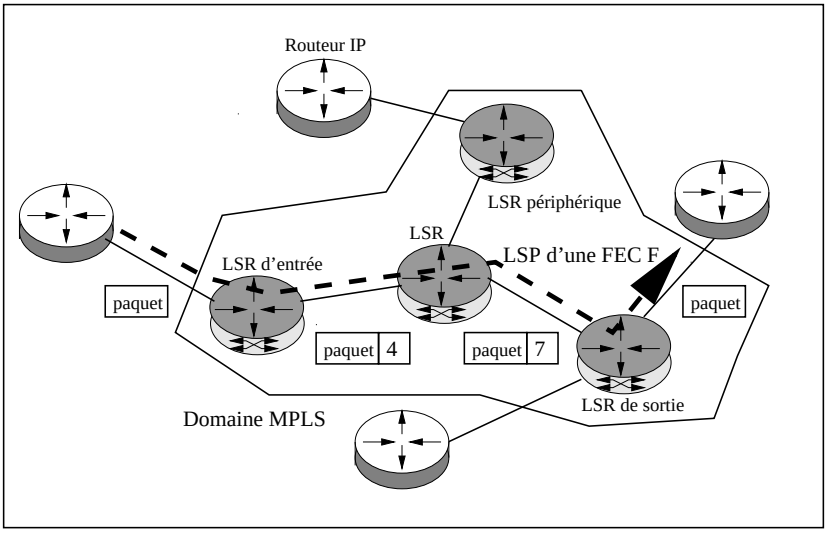

FIG. 1 – *Terminologie MPLS*

La distribution de labels est généralement faite d'une manière dynamique. Un protocole de distribution de labels LDP ("Label Distribution Protocol") a été défini par l'IETF [Andersson 99] [Jamoussi 99]. L'allocation de labels est faite par le LSR en aval par rapport au sens de l'écoulement de données. LDP prévoit deux formes d'allocation : allocation en aval et allocation en aval à la demande. Dans la dernière forme, les labels sont attribués aux FEC et distribués aux pairs qui ont formulé une demande explicite. Cette forme est utilisée dans les réseaux ATM où les commutateurs ATM ne sont pas capables d'agréger différents flux. Dans la première forme, les labels sont affectés aux FEC et distribués aux voisins MPLS sans aucune demande explicite de ceux-ci.

LDP définit aussi, deux mode de contrôle d'allocation de labels. Dans le premier, le contrôle indépendant, chaque LSR attribue un label à une FEC indépendamment des autres voisins. Dans le mode ordonné, seul le LSR de sortie (LSR par lequel le paquet quitte le domaine MPLS) peut attribuer un label à une FEC, les autres LSR ne peuvent attribuer un label à une FEC que s'ils ont reçu déjà une association pour cette FEC.

La distribution de labels peut être assuré par une extension des protocoles de routage (BGP, PIM, etc...) [Rehkter 99] ou des protocoles de contrôle (RSVP) [Awduche 99]. La distribution de labels en utilisant les protocoles de routage permet d'avoir une cohérence instantannée entre la table de routage et la table de commutation. Cependant, les protocoles de routage à état de liens ne peuvent pas être utilisés pour la distribution de labels car ils ne distribuent pas les routes mais la topologie du réseau. Dans ce cas, l'utilisation du protocole LDP est nécessaire.

3 Agrégation

3.1 Introduction

L'agrégation dans les réseaux IP classiques permet de réduire la taille des tables de routage en regroupant les adresses ayant un préfixe en commun et le même prochain routeur. MPLS permet de généraliser cette approche par l'utilisation judicieuse des labels.

Généralement dans un domaine MPLS, une FEC est associée à chaque préfixe d'adresse qui apparaît dans la table de routage. Cependant, il peut y avoir plusieurs préfixes qui empruntent le même chemin. L'agrégation consiste à appliquer un label unique à une union de préfixes. Elle permet de réduire le nombre de labels nécessaires pour traiter l'ensemble des paquets et ainsi de réduire la taille des tables de commutation. En effet, les avantages de l'agrégation dans le cadre de MPLS sont nombreux :

- Accélération du processus de commutation via la réduction de la taille des tables de commutation.
- Économie sur l'espace des labels ce qui est important surtout dans le cas des ATM-LSR (commutateurs ATM qui supportent le protocole MPLS) où les labels constituent une ressource rare (champ VCI : 12 bits) [Rosen 2 99] [Davie 99]. Il est aussi intéressant de faire des économies sur les labels lorsque le réseau est très grand.
- Distribution de labels plus facile. En effet, il est plus simple de distribuer une association entre un label et une FEC que de distribuer plusieurs associations.
- Agrégation d'un ensemble d'adresses IP non forcément contiguës. Un exemple d'agrégation possible dans MPLS (ce qui est impossible avec le CIDR d'IP) consiste à regrouper deux adresses réseaux appartenant à deux classes différentes en une seule FEC. Ce type d'agrégation n'est pas possible avec les protocoles de routage d'où l'avantage de l'agrégation au niveau MPLS.

Dans les prochains paragraphes, nous présentons trois algorithmes d'agrégation en fonction du LSR de sortie. Il s'agit en premier lieu de regrouper les paquets en classes d'équivalence. Ensuite, construire les différents LSP pour acheminer les paquets du LSR d'entrée vers le LSR de sortie du réseau. Le nombre de FEC pourra être égal au nombre des LSR périphériques du domaine MPLS.

Pour que le LSR d'entrée puisse attribuer le label adéquat à un paquet, il est nécessaire qu'il sache déterminer le LSR de sortie du réseau pour ce paquet. Trois cas de figure sont possibles :

1. Tous les LSR du domaine MPLS possèdent cette information.
2. Seulement les LSR périphériques ont cette information.
3. Aucun LSR n'a cette information, sauf le LSR par lequel les paquets de cette FEC quittent le réseau.

Les algorithmes proposés dans la suite correspondent respectivement aux cas de figure définis ci-dessus.

3.2 Agrégation en présence d'un protocole de routage à état de liens

La méthode d'agrégation que nous décrivons dans cette section a été proposée par IBM dans ARIS ("Aggregate Route-based IP Switching") [Viswanathan 97] [Davie 98]. Elle suppose l'existence d'un protocole de routage à état de liens comme OSPF [Moy 98] ou IS-IS.

ARIS permet de préétablir des chemins commutés vers des LSR de sortie connus à l'avance. Ces LSR de sortie sont connus via les protocoles de routage à état de liens comme OSPF et IS-IS. Aucune modification de ces protocoles n'est nécessaire puisque ils échangent naturellement cette connaissance.

Le LSR de sortie initialise l'établissement des chemins commutés en envoyant à tous ses voisins du même domaine, un message d'association entre son identifiant et une valeur bien déterminée d'un label. Ces voisins relayent le message à leurs voisins (qui appartiennent au même domaine) après s'être assurés que le chemin ne contenait pas de cycle. Ainsi chaque LSR de sortie est la racine d'un arbre où les LSR d'entrée sont les feuilles.

ARIS établit les chemins commutés en utilisant l'identifiant du LSR de sortie. Un tel identifiant qui doit être unique dans le domaine, peut être l'identifiant OSPF d'un routeur dans un domaine qui utilise OSPF comme protocole de routage. Lorsqu'un protocole de routage à état de liens est utilisé, chaque routeur dispose de la base de données topologique du domaine entier. Il en résulte que chaque routeur peut déterminer le LSR de sortie de n'importe quelle destination qui figure dans sa table de routage. En fait, le protocole de routage comme OSPF n'offre pas directement cette information, mais les données nécessaires sont disponibles et c'est au LSR de les utiliser pour déterminer le LSR de sortie moyennant un calcul supplémentaire. Chaque LSR détermine le LSR de sortie pour chaque destination qui figure dans sa table de routage et le sauvegarde dans un champ supplémentaire "Egress ID". Il ajoute ce champ à chaque entrée de la table de routage.

La méthode de distribution de labels utilisée est l'aval avec le mode de contrôle ordonné. Chaque LSR de sortie associe un label à son identifiant et envoie cette association à tous ses voisins. Un LSR ayant reçu une association entre un label et un identifiant d'un LSR de sortie, vérifie dans sa table de routage que le LSR qui est à l'origine de ce message est bien le routeur prochain ("next hop") vers le LSR de sortie. Si c'est le cas, il associe un nouveau label à cet "Egress ID" et propage l'association à tous ses autres voisins. Chaque LSR construit une table de commutation dans laquelle il maintient les correspondances entre les labels reçus et les labels attribués. À la fin du processus d'établissement des chemins commutés, un arbre est établi pour chaque LSR de sortie du réseau dont la racine est le LSR de sortie et les feuilles sont les LSR d'entrée.

3.3 Agrégation dans les réseaux de transit

Un réseau de transit est un réseau qui achemine des flux de données générés par un réseau externe et destinés à un autre réseau externe. Généralement, tous les réseaux qui forment le backbone Internet sont des réseaux de transit. Les routeurs périphériques d'un réseau de transit utilisent le protocole BGP ("Border Gateway protocol") [Rehkter 95] pour échanger les informations de routage relatives aux réseaux externes.

Ces informations permettent à chaque LSR périphérique qui exécute le protocole BGP de connaître le LSR de sortie pour tous les flux de transit. En effet, le LSR de sortie pour un paquet donné est le LSR qui a annoncé la route BGP correspondante. Chaque LSR de sortie établit un LSP qui a la forme d'un arbre ayant pour racine le LSR de sortie et dont les feuilles sont les LSR d'entrée. Le même label est utilisé sur toutes les branches de l'arbre qui ont le même père. Pour pouvoir le faire, le mode de contrôle ordonné (cf. paragraphe 2.2) sera utilisé pour la distribution de labels. Notons que l'établissement des LSP se fait de la même manière que celui de l'approche précédente.

Une fois tous les arbres construits, un LSR qui reçoit un paquet de transit détermine son LSR de sortie à l'aide des informations de routage distribuées par BGP. Puis il ajoute au paquet le label qui correspond à la branche de l'arbre dont la racine est le LSR de sortie.

Cette approche ne permet d'agréger que le trafic de transit. Sa mise en oeuvre est intéressante dans les réseaux formant le backbone de l'Internet et plus particulièrement dans les réseaux d'opérateurs transportant le trafic de la téléphonie sur IP.

3.4 Agrégation dans le cas général

Lorsque le protocole de routage sous-jacent est un protocole de type vecteur de distance, les LSR ne peuvent pas connaître le LSR de sortie pour une FEC donnée. Une solution, consistant à effectuer l'agrégation par LSR de sortie, est proposée ci-dessous. L'idée consiste à nous placer dans un état où tous les LSR connaissent les LSR de sortie pour chaque destination dans leurs tables de routage.

Pour cela, chaque LSR de sortie associe le même label à tous les préfixes qui figurent dans sa table de routage et dont le prochain routeur ("next hop") correspondant est un routeur externe au domaine. Ensuite, il envoie cette association à ses voisins. Un LSR ayant reçu le message précédent, scrute sa table de routage et détermine toutes les destinations qui ont pour prochain routeur le LSR de sortie ayant envoyé le message. Ce LSR associe un seul label en entrée pour toutes ces destinations sélectionnées et distribue le résultat à tous ses voisins. Ainsi, chaque LSR qui reçoit un tel message le traite de la même manière. En effet, il sélectionne parmi les destinations, qui ont pour prochain routeur le LSR qui a envoyé le message, celles qui correspondent aux destinations contenues dans le message. Ensuite, il associe un label à tous ces préfixes sélectionnés et envoie le message à tous ses voisins (sauf le LSR qui est la source de ces informations).

À la fin de ce processus, nous obtenons un arbre dont la racine est le LSR de sortie. Notons aussi, que cette méthode ne permet pas d'établir plusieurs chemins de même coût pour une destination donnée car les protocoles de routage de vecteur de distance ne conservent pas dans les tables de routage cette information. En effet, les routeurs choisissent une seule route parmi celles disponibles.

La figure 2 illustre l'établissement des LSP relatifs aux différents LSR de sortie. Le LSR B attribue le label 15 aux préfixes X, Y et Z et envoie l'association aux LSR E et H. Ces derniers effectuent les tests nécessaires et propagent l'association à leurs voisins afin d'établir l'arbre pour les préfixes dont le LSR de sortie est B. La table de routage et la table de commutation du LSR F sont données à titre indicatif. Le LSR F reçoit les associations suivantes : (S, T, W, 61) de la part de G, (X, Y, Z, 34) de la part de E et (X, U, V, 37) de la part de C. Nous remarquons que F

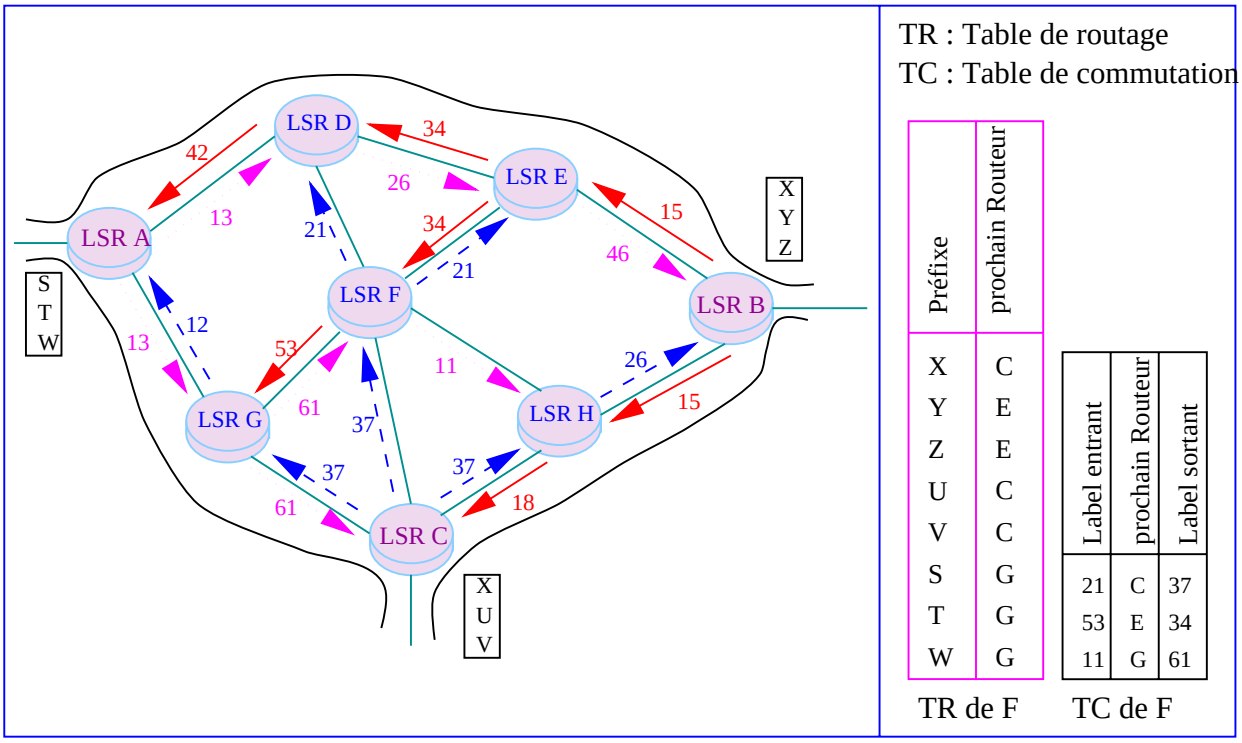

FIG. 2 – *Exemple d'agrégation dans le cas général*

reçoit deux associations qui contiennent le préfixe X. Étant donné que le prochain routeur relatif au préfixe X dans la table de routage de F est le LSR C, les paquets dont la FEC correspond au préfixe X seront étiquetés avec le label 37 et non pas avec le label 34. Les associations envoyées par F aux autres voisins sont : (S, T, W, 11), (Y, Z, 53) et (X, U, V, 21).

3.5 Comparaion de trois méthodes

Contrairement aux deux autres méthodes, l'agrégation dans le cas général est indépendante du protocole de routage utilisé. Elle fonctionne avec n'importe quel protocole de routage interne (RIP, EIGRP, OSPF). Cependant en présence d'un protocole de routage à état de liens, l'approche proposée par ARIS est plus avantageuse. Cette dernière exploite les informations spécifiques au protocole de routage pour minimiser les données de contrôle échangées afin d'établir les chemins commutés. En effet, tous les messages de contrôle contiennent uniquement une association entre l'identifiant du LSR de sortie et un label. Dans l'approche générique, les messages de contrôle contiennent une association entre un ensemble des préfixes (dont le nombre peut atteindre 60 000) et un label. L'agrégation dans les réseaux de transit ne permet d'agréger que le trafic de transit.

4 Conclusion

Dans cet article, nous avons présenté le concept de la commutation de labels comme une nouvelle technique permettant l'acheminement à haut débit des données dans le réseau Internet. Le futur standard MPLS a été introduit. Ce standard en cours d'étude et de préparation par un groupe de l'IETF qui porte le même nom verra le jour dans les mois à venir avec une première version quelque peu simplifiée. Les prochaines versions devront tenir compte de plusieurs aspects non traités par la première version du protocole notamment le multicast et la qualité de service et le routage contraint.

Nous avons présenté le concept d'agrégation qui permet d'accélérer le processus d'acheminement des paquets IP et d'économiser les ressources du réseau. Ensuite, nous avons étudié la mise en place de l'agrégation dans le cadre de MPLS. Elle permet de réduire la taille des tables de commutation et économise les labels qui sont une ressource précieuse. Enfin, elle permet l'agrégation de flux ayant des préfixes non contigus appartenant au même domaine de routage. Ce qui n'est pas faisable avec les protocoles de routage habituels.

Cet article a détaillé trois approches permettant l'agrégation du trafic. Celles-ci définissent une FEC par LSR de sortie. La dernière, plus générique, permet d'agréger le trafic indépendamment du protocole de routage utilisé.

Cette technologie s'adressant à un réseau backbone, il est difficile de déployer à court terme une maquette réaliste. C'est pourquoi, nous projetons de simuler les différentes approches. Cela permettra en premier lieu de démontrer la capacité d'établissement des différents chemins agrégés à partir des tables de routage, d'étudier la faisabilité et d'évaluer les performances, en particulier la reconstruction des LSP suite à un changement dans les tables de routage. Cette simulation mettra en évidence les points forts et les limites de chaque approche proposée.

Références

[Andersson 99] L. Andersson, P. Doolan, N. Feldman, A. Fredette and B. Thomas, "LDP specification", Internet Draft, June 1999.

[Awduche 99] D. Awduche, L. Berger, D. Gan, T. Li, G. Swallow and V. Srinvasan, "Extensions to RSVP for LSP Tunnels", Internet Draft, March 1999.

[Callon 99] R. Callon, N. Feldman, A. Fredette, G. Swallow and A. Viswanathan, "A Framework for Multiprotocol Label Switching", Internet Draft, June 1999.

[Davie 98] B. Davie, P. Doolan and Y. Rekhter, "Switching in IP Networks", Morgan Kaufmann Publishers, Inc. 1998.

[Davie 99] B. Davie, J. Lawrence, K. McGloghrie, Y. Rekhter, E. C. Rosen, G. Swallow and P. Doolan, "MPLS using LDP and ATM VC Switching", Internet Draft, April 1999.

[Jamoussi 99] B. Jamoussi, “Constraint-Based LSP Setup using LDP”, Internet Draft, February 1999.

[Katsube 97] Y. Katsube, K. Nagami and H. Esaki, “Toshiba’s Router Architecture Extensions for ATM: Overview”, RFC2098, February 1997.

[Katz 97] D. Katz, D. Farinacci, B. Davie, G. Swallow, Y. Rekhter and E. C. Rosen, G. Swallow and Farinacci, “Tag Switching Architecture - Overview”, Internet Draft, July 1997.

[Moy 98] J. Moy, “OSPF version 2”, RFC 2328, April 1998.

[Newman 96] P. Newman, W. L. Edwards, R. Hinden, E. Hoffman, F. Ching Liaw, T. Lyon and G. Minshall, “Ipsilon Flow Management Protocol Specification for IPv4 Version 1.0”, RFC 1953, May 1996.

[Rehkter 95] Y. Rekhter and T. Li, “A Border Gateway Protocol 4 (BGP-4)”, RFC 1771, March 1995.

[Rehkter 99] Y. Rehkter and E. Rosen, “Carrying Label Information in BGP-4”, Internet draft, February 1999.

[Rosen 1 99] E. C. Rosen, A. Viswanathan and R. Callon, “Multiprotocol Label Switching Architecture”, Internet Draft, April 1999.

[Rosen 2 99] E. C. Rosen, Y. Rekhter, D. Tappan, D. Farinacci, G. Fedorkow, T. Li and A. Conta, “MPLS Label Stack Encodings”, Internet Draft, April 1999.

[Viswanathan 97] A. Viswanathan, N. Feldman, R. Boivie and R. Woundy, “ARIS: Aggregate Route-Based IP Switching”, Internet Draft, March 1997.