

---

# Arbres multicast dans un réseau MPLS

**Ali Boudani — Bernard Cousin**

*IRISA/Université de Rennes I  
Campus de Beaulieu  
35 042 Rennes Cedex  
France  
{ali.boudani, bernard.cousin}@irisa.fr*

---

*RÉSUMÉ. Dans ce papier, nous étudions la construction d'arbres pour les communications multicast dans un réseau MPLS. Nous discutons d'abord de la difficulté de combiner multicast et MPLS dans le cadre d'un réseau. Nous décrivons quelques propositions MPLS pour le trafic multicast et nous justifions de la nécessité de définir un nouveau protocole. Par la suite nous proposons le protocole MMT (MPLS Multicast Tree), qui utilise des chemins MPLS entre les nœuds de branchement de l'arbre multicast afin de réduire le nombre d'états de routage dans les routeurs et ainsi d'augmenter la résistance au facteur d'échelle. Nous proposons une amélioration du protocole MMT et nous l'évaluons en termes de résistance au facteur d'échelle et d'efficacité. Nous présentons quelques résultats de simulation pour valider notre proposition et nous montrons en conclusion que le protocole MMT semble prometteur et bien adapté à une éventuelle implémentation de l'ingénierie de trafic multicast dans le réseau.*

*ABSTRACT. In this paper, we study multicast tree construction in MPLS network. We discuss the difficulty in combining multicast and MPLS in a network. We describe some MPLS proposals for the multicast traffic and we justify the need for defining a new protocol. Thereafter we propose MMT, the MPLS multicast Tree protocol, which uses MPLS LSP between multicast tree branching nodes in order to reduce the routing states and to increase scalability. We present improvements to MMT protocol and we evaluate it in term of scalability and efficiency. Finally, we present simulation results to validate our evaluation and we conclude that the MMT protocol seems promising and well adapted to a possible implementation of multicast traffic engineering in the network.*

*MOTS-CLÉS : Internet, Réseaux à haut débit, MPLS, Multicast, Ingénierie de trafic*

*KEYWORDS: Internet, High speed networks, MPLS, Multicast, Traffic Engineering*

---

## 1. Introduction

Augmenter l'efficacité de l'utilisation des ressources de l'Internet est primordial. De nombreux types d'applications multimédias utilisant les diffusions des données (comme les services d'audio, de vidéo à la demande et les téléconférences) consomment beaucoup de bande passante. En réduisant le nombre de paquets transmis à travers le réseau, le service *multicast* augmente essentiellement la QoS fournie aux utilisateurs. Cela augmente par conséquent les performances du réseau.

La commutation de labels avec MPLS, étudiée et standardisée au sein de l'IETF [ROS 01], est présentée comme une solution pour gérer efficacement la bande passante de l'Internet. Elle consiste à commuter le trafic en s'appuyant sur le label ajouté à chaque paquet. MPLS enrichit le routage classique en permettant aux paquets de suivre un chemin explicite (appelé aussi meilleur chemin) au lieu de suivre le plus court chemin traditionnel [AWD 99]. Les paquets qui possèdent le même label au niveau d'un LSR (*Label switching Router*) donné appartiennent à une même classe d'équivalence appelée FEC (*Forwarding Equivalence Class*) et reçoivent le même traitement au sein du domaine. Les chemins pris par tous les paquets d'une même classe d'équivalence est appelé chemin commuté ou LSP (*Label Switched Path*). Il est constitué de l'ensemble des LSR traversés par un paquet appartenant à une même classe d'équivalence. Les informations collectées par les protocoles de routage IP sont utilisées pour déterminer la topologie, la charge du réseau, attribuer les labels aux chemins puis distribuer les labels aux routeurs lors de la construction des chemins commutés.

La commutation de labels et le *multicast* sont deux techniques complémentaires. Nous allons montrer que la combinaison de ces deux techniques (où les arbres *multicast* sont construits dans des réseaux MPLS) augmente la performance du réseau et présente une solution résistante au facteur d'échelle. Le service *multicast* minimise l'utilisation de la bande passante du réseau, alors que MPLS utilisé comme outil d'ingénierie de trafic essaye d'approvisionner en bande passante les utilisateurs d'une façon appropriée.

Dans le deuxième chapitre de cet article, nous décrivons les différentes propositions d'utilisation de MPLS pour le trafic *multicast* et cela nous permet de justifier la nécessité de définir un nouveau protocole. Dans le chapitre 3, nous présentons le protocole MMT (*MPLS Multicast Tree*) et son extension le protocole MMT2. Dans le chapitre 4, nous évaluons nos propositions en termes de résistance au facteur d'échelle (la diminution en taille des tables de routage) et d'efficacité et nous présentons quelques résultats de simulation pour valider cette évaluation. Enfin, nous concluons notre travail dans le chapitre 5.

## 2. La combinaison de MPLS et du multicast

MPLS peut être employé dans un réseau pour acheminer le trafic *unicast* à l'aide de chemins explicites. De même, MPLS peut être employé pour le trafic *multicast* en

utilisant des arbres explicites<sup>1</sup>. Mais le trafic *multicast* possède des caractéristiques spécifiques dues à la nature du routage *multicast* [OOM 02]. En effet, le routage *multicast* d'Internet se base sur l'adresse IP *multicast* et c'est pourquoi il est très difficile d'agrèger<sup>2</sup> le trafic *multicast* puisque les destinataires appartenant au même groupe *multicast* peuvent être extrêmement dispersés. De plus, la structure arborescente du *multicast* pourrait nécessiter d'établir des LSP point-à-multipoint ou même des LSP multipoint-à-multipoint. Dans les implémentations actuelles de MPLS, les LSP point-à-point sont privilégiés même si MPLS n'exclut pas d'autres types de LSP.

Dans [OOM 02], un cadre pour le déploiement du service *multicast* dans un domaine MPLS a été proposé. Les avantages et les inconvénients des protocoles de routage IP *multicast* existants dans le contexte de MPLS sont décrits et la relation avec les différentes méthodes de distribution de labels sont discutées. Les caractéristiques des protocoles *multicast* suivantes sont considérées : l'agrégation de trafic, l'inondation et l'élagage, la coexistence d'arbres basés à une source et d'arbres partagés, les arbres uni-directionnels et bi-directionnels, les données *multicast* transmises en mode encapsulé (à travers un tunnel) ou en mode natif, le contrôle selon l'interface RPF, la résistance au facteur d'échelle, la complexité du calcul pour le routage de paquets, la latence d'adhésion à un groupe et le surcoût des messages de contrôle et d'entretien de l'arbre. L'étude n'a pas mené au choix d'un seul protocole de routage *multicast* mais l'auteur conclut que différents protocoles de routage *multicast* pourraient être déployés simultanément dans Internet.

Les labels MPLS favorisent l'agrégation d'arbres mais ne résolvent pas complètement le problème. En effet, on doit concevoir des techniques qui permettent d'agrèger des flux *unicast* avec des flux *multicast* ainsi que d'agrèger des flux *multicast* entre eux. Malheureusement, les études sur l'agrégation *multicast* sont limitées à l'agrégation des états de routage dans chaque routeur plutôt qu'à l'agrégation des LSP. Par la suite, nous présentons deux protocoles de routage MPLS *multicast* : PIM-MPLS [FAR 00] et *Aggregated Multicast* [CUI 03]. Ils sont représentatifs des propositions actuelles. Une présentation plus complète peut être trouvée dans [BOU 04].

## 2.1. PIM-MPLS

*Using PIM to Distribute MPLS Labels for Multicast Routes* [FAR 00] (appelée ci-après PIM-MPLS) est une proposition qui utilise les messages d'adhésion du protocole *multicast* PIM-SM [EST 98] pour distribuer des labels MPLS et construire ainsi l'arbre *multicast* MPLS (arbre *multicast* dont le routage est effectué au niveau de la couche "Liaison de données"). Les messages d'adhésion de PIM-SM sont modifiés pour porter un label MPLS alloué par un LSR en aval.

---

1. Un arbre explicite peut être construit à partir de plusieurs chemins explicites. C'est l'arbre des meilleurs chemins.  
2. Lors de l'agrégation plusieurs états de routage sont remplacés par un seul état dans la table de routage.

#### 4 CFIP 05.

Avec PIM-MPLS, MPLS n'est pas employé avec toute son efficacité comme un outil d'ingénierie de trafic. En effet, cette proposition ne tient pas compte de la charge du réseau. Aucune tentative d'équilibrage n'est proposée puisque l'arbre *multicast* est construit selon le protocole PIM-SM (c'est-à-dire en utilisant le point de rendez-vous). Ce routage n'est pas optimal. De plus, la table de commutation de chacun des routeurs d'un arbre contient un label *multicast* pour cet arbre. Ainsi le nombre d'états croît avec le nombre de groupes.

### 2.2. AGGREGATED-MULTICAST

*Aggregated Multicast* [CUI 03] suppose que plusieurs groupes *multicast* peuvent partager un même arbre *multicast* au lieu de construire un arbre *multicast* par groupe *multicast*. Ceci réduit le nombre d'états de routage dans les routeurs du domaine. Chaque groupe *multicast* est associé à un arbre agrégé. Pour la gestion de l'arbre agrégé et l'association entre les groupes *multicast* et les arbres agrégés, une entité de gestion appelé *tree manager* est introduite.

Lors de l'association entre groupes *multicast* et arbres agrégés, les complications surgissent quand il n'y a pas une association parfaite : aucun arbre agrégé existant ne couvre parfaitement l'ensemble des membres d'un groupe (association perméable). L'inconvénient de l'association perméable est qu'une certaine portion de la bande passante du réseau est gaspillée pour fournir des données à des nœuds qui ne sont pas impliqués dans le groupe considéré.

L'inconvénient d'*Aggregated multicast* est que le nombre de labels est lié au nombre d'arbres agrégés à construire dans le réseau. Si le nombre d'arbres continue à augmenter, le nombre de labels reste élevé malgré la réduction faite par rapport à PIM-MPLS. Ce phénomène est accentué en présence des contraintes de l'ingénierie de trafic puisque il faudra construire plus d'arbres agrégés pour équilibrer la charge du réseau.

Dans les paragraphes suivants, nous nous servons des idées précédentes pour justifier de la définition d'un nouveau protocole pour assurer le service *multicast* dans un réseau MPLS où MPLS est utilisé comme outil d'ingénierie de trafic.

### 3. Le protocole MMT

Le protocole MMT (*MPLS multicast tree*) construit un arbre *multicast* dans un réseau MPLS en considérant seulement les routeurs de branchement de cet arbre. En limitant la présence d'états de routage *multicast* aux routeurs de branchement, le protocole MMT convertit les flux *multicast* en multiple flux *unicast*. Dans MMT, au lieu de construire un arbre pour chaque canal<sup>3</sup> *multicast* individuel, on peut avoir plusieurs

---

3. Rappelons qu'un canal est un groupe identifié par le couple  $(S, G)$  où  $S$  est l'adresse de la source et  $G$  l'adresse du groupe.

canaux *multicast* qui partagent des branches de leurs arbres. Les LSP *unicast* sont utilisés entre les routeurs de branchement de l'arbre *multicast*. En utilisant cette méthode, nous réduisons la quantité d'informations à mémoriser dans les routeurs et nous assurons la résistance au facteur d'échelle.

### 3.1. MMT et les protocoles IP multicast

En comparaison avec l'IP *multicast*, le protocole MMT présente plusieurs avantages :

- Il simplifie l'installation des LSP : puisque les nœuds de branchement de l'arbre *multicast* sont situés aux routeurs de branchement, il n'y a aucun besoin de créer et de maintenir des LSP point-à-multipoint ou des LSP multipoint-à-multipoint. Un arbre est décomposé et chacune de ses branches est associée à un LSP point-à-point. Ces LSP point-à-point sont utilisés pour la transmission du trafic *multicast*.

- Il rend les flux *multicast* plus faciles à agréger : chaque branche d'un flux *multicast* peut être agrégée avec d'autres flux *unicast* qui partagent les mêmes LSR d'entrée et de sortie.

- Il utilise une entité de contrôle<sup>4</sup> pour assurer l'ingénierie de trafic *multicast* dans le réseau : cette entité de contrôle possède les informations nécessaires sur les LSP. Toutes les sources et tous les destinataires des différents groupes *multicast* ainsi que la bande passante associée sont connus. L'entité de contrôle est informée directement de tout changement de topologie du réseau (pannes de liens ou de routeurs) et de tout changement d'appartenance d'un destinataire à un groupe. Un arbre est calculé à l'aide de cette entité et transmis par la suite sur le réseau (cf. paragraphe 3.2).

- Il est inter-opérable avec d'autres protocoles de routage *multicast* : ce protocole peut être limité à un seul domaine (typiquement aux réseaux de cœur). Dans les autres domaines, les protocoles de routage *multicast* traditionnel peuvent être utilisés et une fois transmis dans le domaine MPLS, les paquets *multicast* seront traités facilement par les mécanismes du protocole MMT.

Dans le paragraphe suivant, nous présentons le rôle de l'entité de contrôle chargée de calculer les arbres et de faire une collecte des états des liens et des adhésions aux groupes ainsi que d'exécuter l'algorithme associant un groupe à un arbre. Par la suite nous présentons la construction de l'arbre MPLS ainsi que les nouveaux LSP.

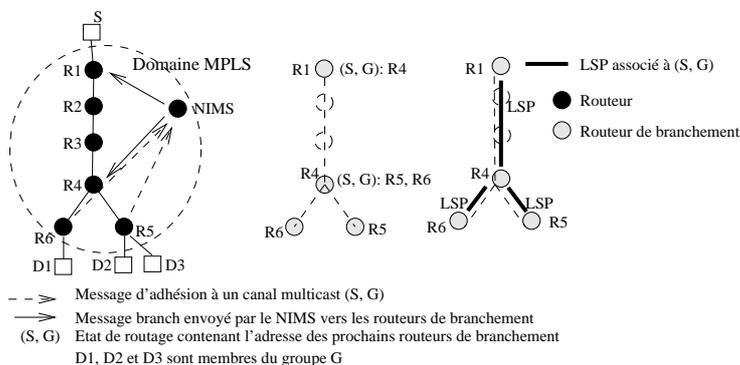
---

4. L'entité de contrôle est un point de défaillance critique. Une certaine redondance de l'entité de contrôle peut assurer la survivabilité du service. Une certaine distribution de l'entité centrale est envisageable. Nous ne le traitons pas ici car: (1) elle complexifierait inutilement le discours;(2) idéalement la distribution est indépendante de l'ingénierie de trafic *multicast*.

### 3.2. La construction de l'arbre multicast MPLS par le NIMS

Un routeur de branchement doit connaître ses prochains routeurs de branchement sur l'arbre *multicast*. Pour ceci, nous proposons d'utiliser une entité de contrôle dans chaque domaine, appelée ci-après NIMS (*Network Information Management System*), chargée de recevoir les messages d'adhésion et les messages de désabonnement de tous les membres des canaux dans ce domaine. Le NIMS est informé de la topologie du réseau et de l'existence des LSP *unicast*. Ceci est conforme avec les orientations de l'IETF au sujet du *multicast* MPLS. Le routeur supportant le NIMS peut par exemple être élu par un mécanisme semblable à celui du routeur rendez-vous dans PIM-SM [EST 98]<sup>5</sup>.

Le calcul de l'arbre *multicast* permet de connaître tous les routeurs de branchement pour un canal. Le NIMS envoie alors des messages *branch* à tous les routeurs de branchement pour les informer de leurs prochains routeurs de branchement. À la réception de ce message, un routeur de branchement crée un état de routage *multicast* pour le canal *multicast*. Une fois les routeurs de branchement et leurs prochains routeurs identifiés, les paquets de données seront envoyés d'un routeur de branchement aux autres jusqu'à atteindre leurs destinations. Les LSP *unicast* déjà établis sont employés entre les routeurs de branchement d'arbre *multicast* afin de réduire la taille des tables de routage et afin d'augmenter la résistance au facteur d'échelle (*cf.* figure 1).



**Figure 1.** La construction de l'arbre multicast MPLS.

*Edge Router Multicasting* (ERM) [YAN 02] est une proposition pour le *multicast* dans un réseau MPLS basée sur les mêmes principes que le protocole MMT. En conséquence, comme dans MMT, les installations de LSP *multicast*, l'association des flux *multicast* et l'agrégation du trafic *multicast* sont transformées en simples problèmes *unicast*. Cependant, ERM limite les points de branchement de l'arbre *multicast* aux routeurs de bordure des domaines MPLS. Les paquets sont envoyés sur les branches à

5. Le NIMS peut être différent au sein du même domaine pour chaque canal (S, G). On a ainsi naturellement un équilibrage de la charge, une répartition du service NIMS et une survivabilité du système accrue.

l'aide des tunnels MPLS établis entre les routeurs de bordure à travers les routeurs du cœur.

Dans ERM, contrairement à MMT, la réservation de la bande passante pour les flux *multicast* n'est pas traitée. De plus, le stress des liens autour des routeurs de bordure augmente puisque la duplication d'un paquet n'est permise que dans les routeurs qui n'appartiennent pas au réseau cœur. Une comparaison entre MMT et ERM est étudiée dans [BOU 04]. Finalement, d'après [ROS 03], les caractéristiques de ERM rendent la solution non recommandée.

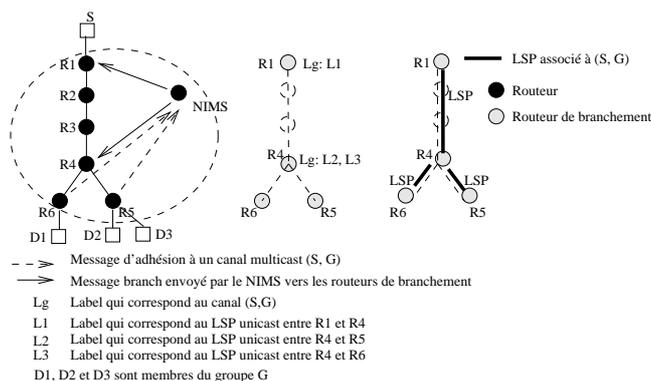
### 3.3. Une extension du protocole MMT : le protocole MMT2

Le protocole MMT nécessite que tous les LSR de branchement routent au niveau IP les paquets multicast. Premièrement, cette fonction n'est pas toujours disponible sur tous les LSR : par ex. sur les LSR de forte puissance et donc spécialisé en commutation. Deuxièmement, lorsque cette fonction est disponible, il est plus efficace de commuter le paquet que de le router. En conséquence nous proposons MMT2. MMT2 utilise un double niveau de labels, tout en préservant le principe de fonctionnement du protocole MMT. Le label du niveau inférieur est un label unique représentant un canal  $(S, G)$ . Un label (appartenant à un intervalle de labels réservé au protocole MMT2) est attribué à chaque canal  $(S, G)$  lors de la réception par le NIMS des messages d'adhésion pour ce canal. Ce label identifie le canal dans le domaine géré par le NIMS. Ce label peut être différent d'un domaine à un autre. Le NIMS informe tous les routeurs de branchement à propos de ce label ainsi que les labels qui correspondent aux prochains routeurs de branchement pour ce canal. Une extension du message *branch* est nécessaire pour porter les nouvelles informations. Le label qui correspond au canal  $(S, G)$  est ajouté au paquet *multicast* à l'entrée du domaine, le routeur *ingress* (routeur à l'entrée du domaine) ajoute aussi le label qui correspond au prochain routeur de branchement pour chaque branche. Dans les routeurs intermédiaires qui ne sont pas des routeurs de branchement, le paquet est analysé suivant le label entrant placé en haut de pile, label qui sera remplacé par un label sortant de MMT (comme dans le cas du MPLS *unicast*). Lorsque le paquet arrive dans le routeur intermédiaire de branchement, le label du niveau supérieur est enlevé, le label identificateur du canal est traité et les nouveaux labels qui correspondent aux prochains routeurs de branchement sont ajoutés (cf. figure 2). Cette opération est répétée jusqu'à l'arrivée aux routeurs *egress* (routeurs à la sortie du domaine). Tous les labels sont ainsi dépilés et de nouveau le paquet est envoyé vers les autres domaines ou bien directement vers les destinations appartenant aux sous réseaux IP des routeurs *egress*.

### 3.4. Le protocole MMT2 et les arbres agrégés

Pour résoudre le problème de pénurie de labels, MMT2 nécessite d'utiliser uniquement les arbres agrégés. On choisit (comme *Aggregated multicast*) que deux canaux soient associés à un même arbre agrégé dans un domaine si l'arbre calculé pour le

premier canal a exactement les mêmes branches que l'arbre calculé pour le deuxième canal dans le domaine. Ceci permet au NIMS d'associer plusieurs canaux à un même arbre agrégé, d'économiser l'utilisation de labels dans le domaine et de réduire les états de routage à conserver dans les routeurs de branchement.



**Figure 2.** La construction de l'arbre multicast MPLS avec le protocole MMT2.

Dans le chapitre suivant, nous évaluons l'approche proposée en termes de résistance au facteur d'échelle (la diminution de la taille des tables de routage *multicast*) et d'efficacité (le coût de l'arbre *multicast* et le temps de traitement des en-têtes *multicast* dans les routeurs).

#### 4. Évaluation du protocole MMT

Notre approche présente une certaine efficacité vis-à-vis des autres propositions *multicast* IP (DVMRP, PIM-SM, MOSPF) et *multicast* MPLS (PIM-MPLS, *Aggregated multicast*) puisque d'une part nous utilisons l'arbre des meilleurs chemins pour acheminer les paquets et d'autre part nous utilisons la technique de commutation rapide de MPLS dans les routeurs.

Nous allons comparer MMT et son extension le protocole MMT2<sup>6</sup> avec les principaux protocoles de routage MPLS *multicast*, notamment PIM-MPLS [FAR 00] et *Aggregated Multicast* [CUI 03]. Dans nos simulations PIM-MPLS se réfère au simulateur décrit dans [BOU 04]. Rappelons que dans [BOU 04] nous avons présenté un simulateur pour le routage *multicast* dans un réseau MPLS où nous avons choisi PIM-SM (avec un arbre spécifique enraciné à la source) comme protocole de routage *multicast*. Nous simulons le protocole MMT avec le simulateur NS<sup>7</sup> pour valider le comportement de base de l'approche et son efficacité à réduire le nombre d'états de routage, à diminuer le temps de traitement d'un paquet et à baisser le coût des arbres.

6. Nous ne considérons que les arbres agrégés.

7. [www.isi.edu/nsnam/ns](http://www.isi.edu/nsnam/ns)

#### 4.1. La diminution en taille des tables de routage multicast

Puisque notre approche ne stocke des états de routage que dans les routeurs de branchement, la réduction de taille des tables de routage est évidente.

Un domaine MPLS peut servir uniquement comme domaine de transit pour un canal. Ni la source, ni les destinataires ne sont obligatoirement présents dans le domaine pour ce canal. Nous nous proposons d'étudier deux cas limites que peuvent prendre les arbres dans un domaine MPLS. Un arbre ayant un ou plusieurs nœuds de branchement dans le domaine est appelé BT (*Branched Tree*). Un arbre avec un seul chemin dans le domaine où aucun nœud de branchement ne figure dans l'arbre est appelé OPT (*One Path Tree*). Le tableau 1 représente le nombre moyen d'états de routage dans les routeurs dans les deux cas : arbres BT de transit avec nœuds de branchement et arbres OPT de transit sans nœud de branchement.

Protocole/ Arbre	PIM-MPLS	<i>Aggregated multicast</i>	MMT	MMT2
BT	$\overline{n_T} * T$	$\overline{n_{T_{agg}}} * T_{agg}$	$\overline{n_{MMT}} * T$	$\overline{n_{MMT-agg}} * T_{agg}$
OPT	$\overline{n_T} * T$	$\overline{n_{T_{agg}}} * T_{agg}$	$2 * T$	$2 * T_{agg}$

**Tableau 1.** Le nombre moyen d'états de routage dans les routeurs.

$T$  représente le nombre d'arbres *multicast* présents dans le domaine,  $\overline{n_{MMT}}$  le nombre moyen de routeurs de branchement sur les arbres en utilisant le protocole MMT,  $\overline{n_{MMT-agg}}$  le nombre moyen de routeurs de branchement sur les arbres agrégés en utilisant le protocole MMT2,  $\overline{n_T}$  le nombre moyen de routeurs sur les arbres *multicast* en utilisant un protocole de routage *multicast* traditionnel,  $T_{agg}$  est le nombre d'arbres agrégés de *Aggregated Multicast*,  $\overline{n_{T_{agg}}}$  est le nombre moyen de routeurs sur l'arbre agrégé. Ces valeurs satisfont les relations suivantes :

$$T \geq T_{agg}; \overline{n_T} \geq \overline{n_{MMT}}; \overline{n_T} \geq \overline{n_{MMT-agg}};$$

et

$$\overline{n_{T_{agg}}} \geq \overline{n_{MMT-agg}}; \overline{n_T}, \overline{n_{T_{agg}}} \geq 2.$$

Il est donc évident d'après le tableau 1 que MMT offre de meilleures performances que PIM-MPLS. Dans le cas d'arbres OPT, le nombre d'états de routage par canal pour MMT dans les routeurs intermédiaires dans le réseau est égal à  $0^8$  et donc offre des meilleures performances par rapport à *aggregated multicast*. Par contre, dans le cas d'arbres BT, le nombre d'états de routage pour le protocole MMT n'est pas toujours inférieur à celui d'*Aggregated Multicast*. En effet, le protocole MMT présente de meilleures performances par rapport à *Aggregated Multicast* seulement dans le

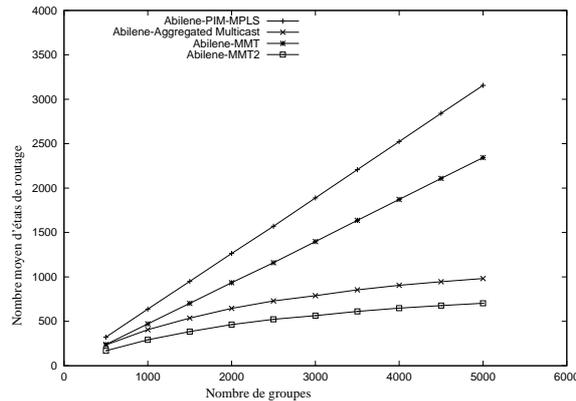
8. Dans le reste de cette évaluation, nous considérons que  $\overline{n_{MMT}}$  et  $\overline{n_{MMT-agg}}$  incluent aussi les états présents dans les *ingress* et *egress* LSR.

cas où  $\overline{n_{MMT}} * T < \overline{n_{T_{agg}}} * T_{agg}$  donc lorsque la formule suivante est respectée :  $\overline{n_{T_{agg}}} > \overline{n_{MMT}} * \frac{T}{T_{agg}}$ . Or, ceci est possible dans beaucoup de cas. Prenons l'exemple suivant :

D'après [CUI 03], le réseau vBNS est formé de 43 routeurs dont 16 sont des routeurs du cœur. Les 43 routeurs du réseau participent à la diffusion du trafic *multicast*. Un ensemble de 2500 canaux *multicast* sont présents dans le réseau. Ces 2500 canaux sont regroupés en 1150 arbres agrégés. Donc d'après la formule du paragraphe précédent,  $\overline{n_{T_{agg}}}$  doit être plus grand que  $\overline{n_{MMT}} * \frac{2500}{1150} \approx 2.2 * \overline{n_{MMT}}$  pour que MMT soit meilleur. Comme nous l'avons mesuré personnellement [BOU 04], dans l'Internet le nombre moyen de routeurs de branchement sur un arbre est généralement très petit (de l'ordre de 8% du nombre de routeurs de l'arbre *multicast*). Nous déduisons que  $\overline{n_{MMT}}$  est au maximum  $\approx 4$ . Si la valeur de  $\overline{n_{T_{agg}}}$  dépasse  $\overline{n_{MMT}} * \frac{T}{T_{agg}} \approx 9$ , MMT a de meilleures performances. Il est donc possible que MMT réduise mieux que *Aggregated Multicast* la taille des tables de routage *multicast* dans les routeurs. Dans tous les cas, d'après le tableau 1, MMT2 présente des meilleures performances par rapport à tous les autres protocoles.

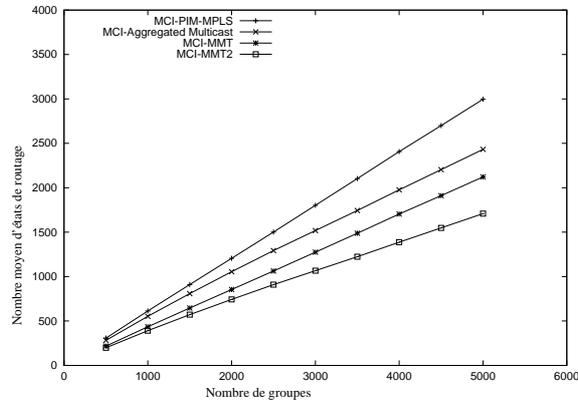
Pour valider notre évaluation, nous considérons les 3 réseaux : MCI<sup>9</sup> (18 nœuds dans le cœur du réseau), Abilene (11 nœuds dans le cœur du réseau) et NSFNET (14 nœuds dans le cœur du réseau), et nous calculons le nombre d'arbres agrégés pour 5000 arbres. Nous considérons qu'à chaque nœud du cœur du réseau est attaché un seul nœud qui joue le rôle d'une source ou d'un destinataire. Le nombre de membres destinataires pour chaque groupe est entre 2 et 10 pour le réseau Abilene, 2 et 12 pour le réseau NSFNET et 2 à 16 pour le réseau MCI.

Les figures 3, 4 et 5 représentent respectivement le nombre moyen d'états de routage dans un routeur pour les réseaux Abilene, MCI et NSFNET.

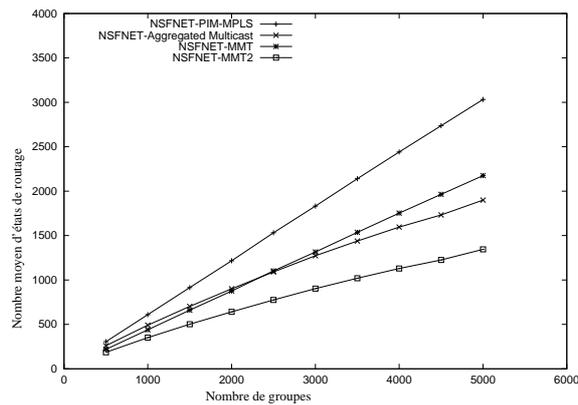


**Figure 3.** Le nombre moyen d'états de routage dans un routeur dans le réseau Abilene.

9. Rappelons que MCI a développé le réseau vBNS+.



**Figure 4.** Le nombre moyen d'états de routage dans un routeur dans le réseau MCI.



**Figure 5.** Le nombre moyen d'états de routage dans un routeur dans le réseau NSFNET.

Nous remarquons que le protocole MMT2 offre les meilleures performances. Nous remarquons aussi que PIM-MPLS présente les plus mauvais résultats. Pour MMT et *Aggregated multicast*, nous remarquons que MMT présente un nombre réduit d'états de routage par rapport à *Aggregated multicast* dans le réseau MCI mais c'est le contraire dans le réseau Abilene. À partir d'un certain nombre d'arbres, *Aggregated multicast* offre de meilleures performances par rapport à MMT dans le réseau NSFNET. En effet, le réseau Abilene contient seulement 11 nœuds : si le nombre de membres dans un groupe est grand, alors il est très probable que tous les routeurs du cœur soient considérés comme routeurs de branchement. Notons que nous considérons les états de routage dans la source et dans tous les destinataires. Si le nombre de membres dans les groupes est petit, le rapport  $\frac{T}{T_{agg}}$  devient grand et donc le protocole MMT ne convient

pas à ce genre de topologie. Dans tous les cas, le protocole MMT2 réduit la taille des tables de routage.

#### 4.2. Le temps de traitement des en-têtes multicast dans les routeurs

Le délai d'établissement de l'arbre et le temps de transmission d'un paquet sont deux paramètres importants à étudier.

Dans les trois protocoles MMT, MMT2 et *Aggregated multicast*, une entité de contrôle reçoit les messages d'adhésion aux canaux, calcule l'arbre pour chaque canal et déclenche l'association entre labels et canaux. Le délai d'établissement de l'arbre est donc le même pour les trois protocoles. Mais un avantage pour MMT et MMT2 est que des LSP *unicast* sont utilisés pour le trafic *multicast*.

En revanche, dans le cas d'un protocole de routage *multicast* traditionnel, chaque routeur entre l'*ingress* et l'*egress* doit contenir un état de routage pour chaque arbre *multicast*. Considérons  $t_{Trait}(Ri)$ , qu'on appelle temps de traitement, le temps nécessaire pour traiter un paquet dans un routeur  $Ri$ . Nous allons comparer le temps de traitement global  $t_{Traitg} = \sum t_{Trait}(Ri)$  d'un paquet pour les protocoles PIM-MPLS, *Aggregated multicast*, MMT et MMT2.

Nous faisons l'hypothèse que le réseau est composé des liens symétriques. Nous supposons que le temps de transmission d'un paquet sur les liens entre la source et le destinataire sur l'arbre *multicast* et le temps d'attente du paquet dans la file d'attente des routeurs sont constants pour les différents protocoles.

##### 4.2.1. PIM-MPLS

Nous pouvons constater que  $t_{Traitg} = \overline{n_T} * \overline{t_{MPLS\ multicast}}$ , où  $\overline{n_T}$  est le nombre moyen de routeurs sur un arbre *multicast* et  $\overline{t_{MPLS\ multicast}}$  est le temps moyen pour parcourir la table de commutation.

##### 4.2.2. MMT

Dans MMT, seuls les routeurs de branchement gardent les états de routage *multicast*. Dans ces routeurs, le protocole MMT cherche à trouver la *FEC* correspondante dans la table de commutation. Tous les autres routeurs sur l'arbre utilisent les tables de commutation MPLS *unicast* pour acheminer les paquets. D'où :  $t_{Traitg} = (\overline{n_T} - \overline{n_{MMT}}) * \overline{t_{MPLS\ unicast}} + \overline{n_{MMT}} * (\overline{t_{MPLS\ multicast}})$  avec  $\overline{t_{MPLS\ unicast}}$  est le temps moyen pour parcourir la table de commutation MPLS *unicast*.

$$\begin{aligned} \delta_{MMT, PIM-MPLS} &= \overline{n_T} * (\overline{t_{MPLS\ unicast}} - \overline{t_{MPLS\ multicast}}) \\ &+ \overline{n_{MMT}} * (\overline{t_{MPLS\ multicast}} - \overline{t_{MPLS\ unicast}}) \\ &= (\overline{n_T} - \overline{n_{MMT}}) * (\overline{t_{MPLS\ unicast}} - \overline{t_{MPLS\ multicast}}). \end{aligned}$$

La valeur de  $\overline{n_{MMT}}$  est plus petite ou égale à la valeur de  $\overline{n_T}$ . Comme on envisage de traiter un grand nombre d'arbres *multicast* supérieur au nombre de routeurs dans

le domaine MPLS, il est probable que la valeur de  $\overline{t_{MPLS\ multicast}}$  croît. En effet, si l'on suppose toutes choses égales, le temps de traitement associé à l'accès à la table de commutation *multicast* est inférieur ou égal au temps de traitement associé à l'accès à la table de commutation *unicast*. La valeur de  $\overline{t_{MPLS\ unicast}}$  devient plus petite que la valeur de  $\overline{t_{MPLS\ multicast}}$  et  $\delta_{MMT,PIM-MPLS}$  prend alors des valeurs négatives.

Ce qui nous permet de conclure que le protocole MMT offre des meilleures performances par rapport au protocole PIM-MPLS ainsi que tous les autres protocoles utilisant le même type de construction d'arbres *multicast* MPLS. Rappelons que d'après le site web de Juniper [MPL ], un routeur Juniper T640 peut traiter un paquet en  $10^{-9}s$  et le gain en temps de traitement d'un paquet se traduit en un débit plus élevé et une capacité à acheminer un nombre plus élevé de paquets.

#### 4.2.3. Aggregated multicast

Comme dans le cas de PIM-MPLS nous pouvons constater que :  $t_{Traitg} = \overline{n_{T_{agg}}} * \overline{t_{MPLS\ agg}}$ , où  $\overline{t_{MPLS\ agg}}$  est le temps moyen pour parcourir la table de labels présente dans les routeurs après utilisation du protocole *Aggregated multicast*. Nous obtenons ainsi :

$$\delta_{MMT,Aggregated\ multicast} = (\overline{n_T} - \overline{n_{MMT}}) * \overline{t_{MPLS\ unicast}} + \overline{n_{MMT}} * \overline{t_{MPLS\ multicast}} - \overline{n_{T_{agg}}} * \overline{t_{MPLS\ agg}}.$$

Il n'est pas facile de faire des approximations avec cette formule. En effet, le temps de recherche dans une table de routage n'est pas linéaire : elle peut être parfois logarithmique avec l'utilisation de différentes techniques de recherche [WAL 01]. La valeur  $\overline{t_{MPLS\ agg}}$  dépend aussi du taux de réduction des arbres *multicast*.

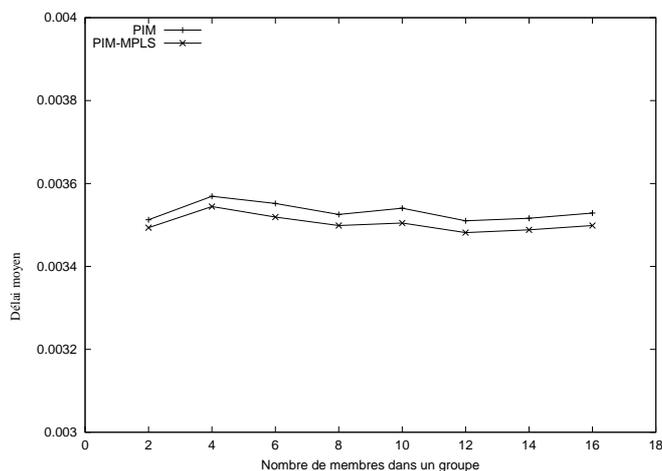
Par contre, en comparaison avec le protocole MMT2, il est facile de remarquer<sup>10</sup> que  $\delta_{MMT2,Aggregated\ multicast}$  prend souvent des valeurs négatives, ce qui nous permet de conclure que le protocole MMT2 offre des meilleures performances par rapport au protocole *Aggregated multicast* en terme de temps de traitement d'un paquet.

$$\delta_{MMT2,Aggregated\ multicast} = (\overline{n_{T_{agg}}} - \overline{n_{MMT-agg}}) * \overline{t_{MPLS\ unicast}} + \overline{n_{MMT-agg}} * \overline{t_{MPLS\ agg}} - \overline{n_{T_{agg}}} * \overline{t_{MPLS\ agg}} = (\overline{n_{T_{agg}}} - \overline{n_{MMT-agg}}) * (\overline{t_{MPLS\ unicast}} - \overline{t_{MPLS\ agg}}).$$

Remarquons qu'il est très complexe de simuler les valeurs exactes du temps de traitement global d'un paquet *multicast*. En effet, ce traitement dépend de la taille de la table de routage et de la technique de recherche utilisée pour trouver une entrée dans cette table. Nous nous contentons donc de faire la comparaison simple entre PIM-SM et PIM-MPLS. En effet, d'après [MPL ], MPLS peut réduire 25% le temps de traitement d'un paquet. Puisque les deux protocoles PIM-SM et PIM-MPLS ont exactement

10. Puisque  $(\overline{n_{T_{agg}}} - \overline{n_{MMT-agg}})$  est toujours  $\geq 0$  et  $(\overline{t_{MPLS\ unicast}} - \overline{t_{MPLS\ agg}})$  est souvent  $\leq 0$  (avec la croissance de nombre de canaux *multicast* dans le réseau).

le même nombre d'états de routage, la comparaison devient facile en supposant que le temps de traitement d'un paquet *unicast* et *multicast* est le même.



**Figure 6.** Le délai global moyen d'un paquet pour les protocoles PIM-SM et PIM-MPLS dans le réseau MCI.

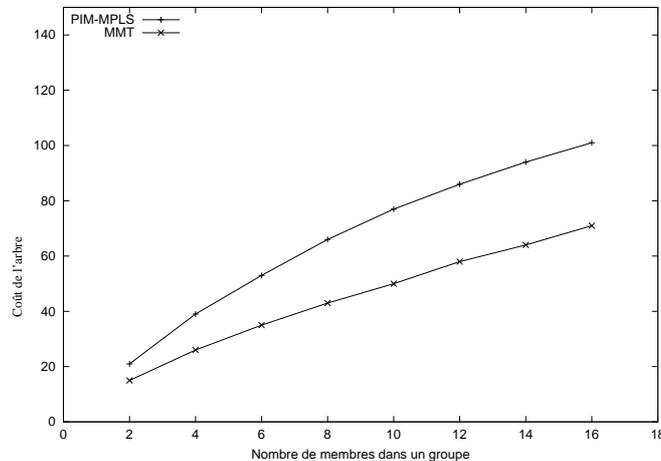
La figure 6 représente le délai de bout en bout moyen des arbres lorsque le nombre de destinataires varie de 2 à 16 pour les deux protocoles PIM-SM et PIM-MPLS dans le réseau MCI. Nous avons réalisé 100 simulations pour chaque valeur de 2 à 16. La valeur présentée dans les graphes est la valeur moyenne pour les 100 simulations. Nous en déduisons que MPLS permet de réduire le délai de bout en bout pour les différentes destinations.

### 4.3. Le coût de l'arbre

Les protocoles MMT, MMT2 utilisent l'arbre des meilleurs chemins. En l'absence de contraintes, l'arbre des meilleurs chemins est l'arbre des plus courts chemins. *Aggregated Multicast* utilise l'arbre des plus courts chemins tandis que le protocole PIM-MPLS utilise l'arbre des plus courts chemins inverses mais dans les deux cas une seule copie d'un paquet est transmise sur un lien. Remarquons que l'arbre des plus courts chemins est identique à l'arbre des plus courts chemins inverses si on considère que le réseau est symétrique (ce n'est pas toujours le cas comme nous l'avons déjà vu dans le chapitre 2 de [BOU 04]).

Nous considérons la topologie de MCI<sup>11</sup> [APO 98] et nous utilisons le scénario suivant : un nouveau nœud est attaché à chaque nœud du cœur du réseau MCI. Un

11. Au lien  $\langle n1, n2 \rangle$  qui connecte les nœuds  $n1$  et  $n2$  sont associés deux coûts,  $n1-n2$  et  $n2-n1$  choisis aléatoirement dans l'intervalle  $[1, 10]$ .



**Figure 7.** Le coût de l'arbre construit par les protocoles MMT et PIM-MPLS pour le réseau MCI.

nœud spécifique est fixé comme source et un nombre variable de destinataires est choisi aléatoirement parmi les nœuds qui ne sont pas du cœur du réseau. Pour chaque nombre de destinataires, nous avons réalisé 500 simulations par algorithme. Ayant identifié le coût de l'arbre comme étant le coût moyen des chemins de la source vers tous les destinataires, la figure 7 représente le coût moyen de l'arbre construit par les protocoles MMT<sup>12</sup> et PIM-MPLS pour le réseau MCI. Nous remarquons que le coût de l'arbre est moins élevé avec le protocole MMT qu'avec le protocole PIM-MPLS puisque ce dernier utilise l'arbre des plus courts chemins inverses.

## 5. Conclusion

Dans cet article, nous avons décrit le protocole MMT, qui utilise des LSP *unicast* entre les nœuds de branchement de l'arbre *multicast* afin d'enlever les états de routage dans les routeurs intermédiaires et ainsi d'augmenter la résistance au facteur d'échelle. Nous avons présenté le protocole MMT2 : une extension du protocole MMT. Alors que MMT impose de router au niveau IP les paquets dans les routeurs de branchement, MMT2 utilise un double niveau de labels qui permet de commuter les paquets multicast dans les routeurs de branchement.

Nous avons évalué MMT et MMT2 en termes de résistance au facteur d'échelle et d'efficacité. Nous avons remarqué une diminution attendue en taille des tables de routage *multicast* par rapport aux autres approches *multicast* MPLS. Nous avons re-

12. Les protocoles MMT2 et *Aggregated Multicast* offrent des résultats identiques au protocole MMT.

marqué aussi un temps de traitement rapide d'un paquet dû à l'utilisation de la technique de commutation de labels MPLS dans les routeurs. Nous avons validé le coût faible obtenu par MMT par rapport aux protocoles qui utilisent l'arbre des plus courts chemins inverses.

Actuellement, nous implémentons à l'IRISA les protocoles MMT et MMT2 sur une plate-forme de 8 machines Redhat Linux. Cette plate-forme permettra de valider notre évaluation réalisée jusqu'à maintenant avec le simulateur NS. Notre travail futur consistera à étudier le coût supplémentaire des messages de signalisation envoyés vers et à partir du NIMS, ainsi que le coût de la reconstruction de l'arbre dans le cas des groupes dynamiques.

## 6. Bibliographie

- [APO 98] APOSTOLOPOULOS G., GUERIN R., KAMAT S., TRIPATHI S., « Quality of Service Based Routing : A Performance Perspective », *ACM SIGCOMM*, July 1998, p. 913-919.
- [AWD 99] AWDUCHE D., MALCOM J., AGOGBUA J., O'DELL M., « Requirements for Traffic Engineering over MPLS », IETF RFC 2702, September 1999.
- [BOU 04] BOUDANI A., « Routage multicast : gestion des petits groupes et ingénierie de trafic », PhD thesis, Université de Rennes I, 2004.
- [CUI 03] CUI J., KIM J., MAGGIORINI D., BOUSSETTA K., GERLA M., « Aggregated Multicast: A Comparative Study », *Special Issue of Cluster Computing: The Journal of Networks, Software and Applications*, , 2003.
- [EST 98] ESTRIN D., FARINACCI D., HELMY A., THALER D., DEERING S., HANDLEY M., JACOBSON V., LIU C., SHARMA P., WEI L., « Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification », IETF RFC 2362, 1998.
- [FAR 00] FARINACCI D., REKHTER Y., ROSEN E., « Using PIM to distribute MPLS labels for multicast routes », IETF Internet draft, November 2000.
- [MPL ] ENABLING TECHNOLOGIES MPLS, [www.juniper.net/solutions](http://www.juniper.net/solutions).
- [OOM 02] OOMS D., SALES B., LIVENS W., ACHARYA A., GRIFFOUL F., ANSARI F., « Overview of IP Multicast in a Multi-Protocol Label Switching MPLS Environment », IETF RFC 3353, August 2002.
- [ROS 01] ROSEN E., VISWANATHAN A., CALLON R., « Multiprotocol label switching architecture », IETF RFC 3031, January 2001.
- [ROS 03] ROSEN E., CAI Y., TAPPAN D., REKHTER Y., FARINACCI D., « Multicast in MPLS/BGP VPNs », IETF Internet draft, October 2003.
- [WAL 01] WALDVOGEL M., VARGHESE G., TURNER J., PLATTNER B., « Scalable high-speed prefix matching », *ACM Transactions on Computer Systems*, vol. 19, n° 4, 2001, p. 440-482.
- [YAN 02] YANG B., MOHAPATRA P., « Edge Router Multicasting With MPLS Traffic Engineering », *IEEE International Conference on Networks (ICON'02)*, 2002.