MPLS et GMPLS

L'environnement IP est devenu le standard de raccordement à un réseau pour tous les systèmes distribués provenant de l'informatique. De son côté, la technique de transfert ATM a incarné la solution préférée des opérateurs pour relier deux routeurs entre eux avec une qualité de service. Il était donc plus que tentant d'empiler les deux environnements pour permettre l'utilisation à la fois de l'interface standard IP et de la puissance de l'ATM. Cette opération a donné naissance aux architectures dites IP sur ATM.

La difficulté de cette solution se situe au niveau de l'interface entre IP et ATM, avec le découpage des paquets IP en cellules, et lors de l'indication dans la cellule d'une référence correspondant à l'adresse IP du destinataire. En effet, le client que l'on souhaite atteindre est connu par son adresse IP, alors que les données doivent transiter par un réseau ATM. Pour ouvrir le chemin, ou circuit virtuel, il faut nécessairement connaître l'adresse ATM du client récepteur. La problématique vient de la correspondance d'adresses : en connaissant l'adresse IP du destinataire, comment trouver son adresse ATM ?

On peut regrouper les solutions à ce problème en trois grandes classes :

- Les techniques d'émulation, lorsque la correspondance d'adresses utilise un intermédiaire, l'adresse MAC.
- Le protocole CIOA (Classical IP over ATM), lorsqu'il n'y a qu'un seul sous-réseau ATM.
- Les techniques de serveur de routes MPOA (MultiProtocol Over ATM), PNNI (Private Network Node Interface) et NHRP (Next Hop Resolution Protocol), lorsqu'il y a plusieurs sous-réseaux ATM potentiels à traverser.

Ces trois techniques sont de plus en plus remplacées par un protocole beaucoup plus homogène, normalisé par l'IETF sous le nom de MPLS (MultiProtocol Label-Switching). Comme Ethernet et ATM, MPLS utilise des techniques de commutation de références, ou label-switching, mais avec d'autres types de trames, comme LAP-F ou PPP. MPLS fait appel à un chemin LSP (Label Switched Path), qui n'est autre qu'un circuit virtuel. Les paquets qui suivent ce chemin sont commutés dans les nœuds.

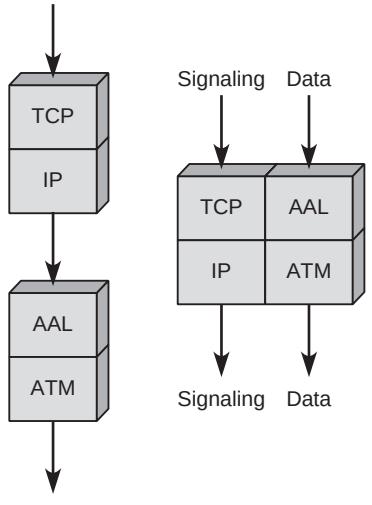
Pour le monde des opérateurs de télécommunications et, de façon plus fragmentaire, pour les très grandes sociétés internationales dotées de leur propre réseau, MPLS est devenu la technique des années 2000.

Des extensions à MPLS ont été apportées avec GMPLS (Generalized MPLS), qui introduit de nouveaux paradigmes de commutation. Ce chapitre commence par décrire les techniques IP sur ATM avant de détailler MPLS puis GMPLS.

IP sur ATM

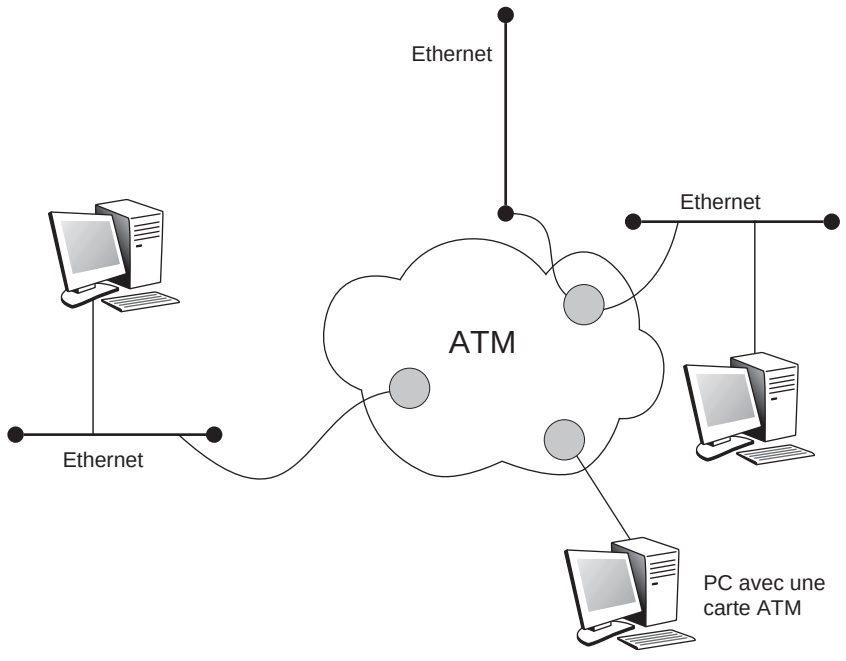
La figure 19.1 illustre deux architectures potentielles pour IP sur ATM. L'architecture de gauche (IP over ATM) est celle qui a été retenue par la quasi-totalité des constructeurs et des opérateurs. L'architecture de droite est une solution non implémentée, qui consiste à mettre en parallèle une infrastructure ATM et une pile TCP/IP. L'idée est de faire passer la signalisation par le plan TCP/IP et les données par le plan ATM. L'intérêt de cette solution est d'utiliser l'universalité de l'adressage IP et la puissance de transfert de l'ATM. Son inconvénient est de devoir mettre sur pied un double réseau et de ne pas avoir d'interface native ATM. C'est cette architecture qui va servir de base à MPLS.

Figure 19.1
Deux architectures IP sur ATM



Une troisième solution, encore utilisée aujourd'hui, est ce que l'on appelle l'émulation de réseau, ou LANE (LAN Emulation). Elle est illustrée à la figure 19.2. Dans cette solution, on se sert d'une adresse Ethernet comme intermédiaire entre l'adresse IP et l'adresse ATM. Cela permet d'ajouter une infrastructure ATM sans que les équipements terminaux aient à s'en soucier. C'est une façon d'introduire de l'ATM dans l'entreprise de manière transparente pour l'utilisateur. La section suivante en détaille les caractéristiques.

Figure 19.2
Architecture LANE



LANE (LAN Emulation)

Le protocole LANE poursuit trois objectifs :

- remplacer un sous-réseau par un réseau ATM ;
- conserver les interfaces utilisateur ;
- faire communiquer des équipements terminaux ATM avec des équipements terminaux LAN.

L'un des inconvénients majeurs de cette solution est qu'elle nécessite une double correspondance IP-MAC et MAC-ATM.

Il existe de nombreuses façons de définir une émulation, dont l'une des meilleures est proposée par l'ATM Forum sous le sigle L-UNI (LAN emulation User-to-Network Interface). Comme elle est de niveau MAC, cette émulation supporte toutes les applications existantes.

L'émulation L-UNI comporte quatre parties :

- L'émulation client, ou LEC (LAN Emulation Client), qui travaille comme un délégué pour le terminal ATM.
- L'émulation serveur, ou LES (LAN Emulation Server), qui résout la correspondance des adresses MAC et ATM.
- L'émulation serveur pour les applications multipoint, ou BUS (Broadcast and Unknown Server), qui résout la correspondance des adresses multipoint.
- L'émulation serveur de configuration, ou LECS (LAN Emulation Configuration Server), qui permet de mettre à jour une station qui se connecte.

Le logiciel LEC, que doit posséder toute station ou tout routeur qui veut être émulé, détient une adresse ATM d'accès. Le LES mémorise toutes les adresses MAC des stations des réseaux locaux qui sont logiquement attachés et leur adresse ATM associée.

Le BUS est un serveur du même type que le LES mais pour les adresses de diffusion et multipoint. Enfin, le LECS possède les informations de configuration, comme l'adresse du LES du réseau émulé auquel appartient une station qui s'active.

Quand un client désire envoyer une trame vers une autre station, il fait parvenir au serveur LES une requête sur l'adresse ATM correspondant à l'adresse MAC de la station destinataire. Le serveur répond avec l'adresse ATM du LEC auquel la station destination est connectée. Ensuite, le LEC ouvre un circuit virtuel avec son correspondant, déterminé par l'adresse ATM que lui a procurée le LES, et convertit la trame MAC en plusieurs trames ATM et envoie les cellules. Au LEC d'arrivée, les cellules sont converties en trames MAC, qui sont alors envoyées vers le terminal approprié.

Le cheminement des flots s'effectue de la façon illustrée à la figure 19.3. Le parcours 1 correspond à l'envoi par le client d'une requête, portant une demande de conversion d'une adresse IP en une adresse ATM, envoyée au serveur LES. Le parcours 2 illustre la réponse à cette requête. Le client connaissant maintenant l'adresse ATM de son correspondant, il peut lui envoyer un flot de paquets IP encapsulés dans des trames ATM et circulant sur le circuit virtuel ouvert vers l'adresse ATM du destinataire. Le parcours 3 correspond à l'ouverture du circuit virtuel avec la machine distante dont l'adresse ATM a été obtenue grâce à la conversion effectuée.

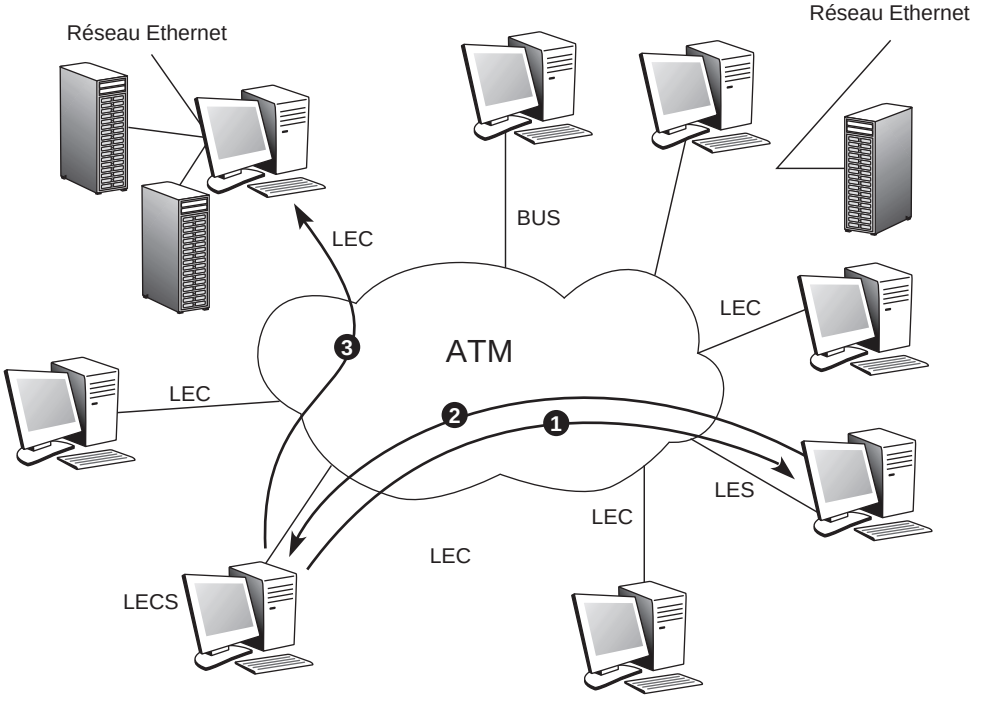


Figure 19.3

Émulation L-UNI

Si le serveur LES n'est pas capable d'effectuer la traduction d'adresse, il faut envoyer une demande de traduction au BUS. Celui-ci émet en diffusion cette demande vers l'ensemble des récepteurs du réseau ATM. La station de réception qui se reconnaît comme étant le correspondant, grâce à l'adresse IP incluse dans la demande, renvoie son adresse ATM à l'émetteur, qui peut enfin ouvrir un circuit virtuel, où transitera le flot des paquets IP.

Le serveur BUS est également utilisé lorsque l'adresse IP du récepteur est multicast. Le serveur BUS possède pour cela des circuits virtuels ouverts avec l'ensemble des machines participant au réseau ATM.

La figure 19.4 illustre l'architecture du LANE. La pile protocolaire de droite représente une machine terminale connectée à un réseau local. Celui-ci mène à un équipement de connexion au réseau ATM. La pile protocolaire de gauche représente une station ATM attachée directement au réseau ATM mais travaillant en émulation LAN. Les piles protocolaires du milieu représentent, à droite, un commutateur ATM et, à gauche, la passerelle de passage entre le réseau local et le réseau ATM.

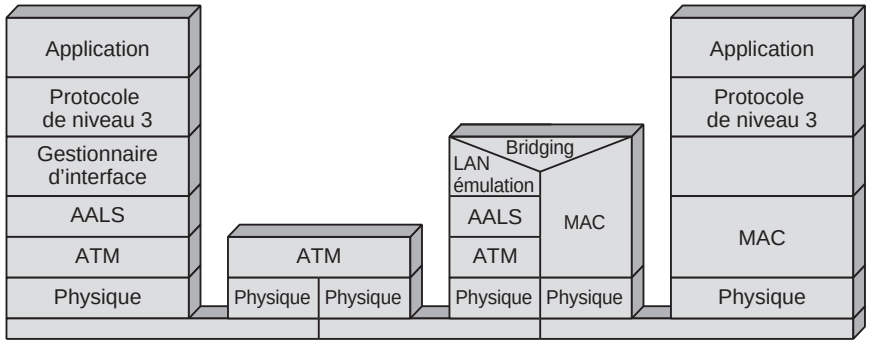


Figure 19.4

Architecture de l'émulation de réseaux locaux LANE

LANE 2.0 introduit une évolution notable par rapport à cette première génération en ajoutant le respect de la qualité de service et le support d'applications multipoint. Cette génération n'a cependant pas eu le temps de s'étendre, du fait de l'arrivée de MPLS, que nous détaillons dans la suite de ce chapitre.

CIOA (Classical IP over ATM)

La solution CIOA permet de transporter les paquets IP par l'intermédiaire d'un réseau ATM sans émulation de réseau local. Pour ce faire, l'adresse IP est traduite directement dans une adresse ATM. Pour réaliser le transport de l'information, il suffit d'encapsuler les paquets IP dans des cellules ATM. À la différence de la solution précédente, on ne passe pas par une première encapsulation dans une trame Ethernet, elle-même encapsulée dans des cellules ATM.

Issue du groupe de travail ION (Internetworking Over NBMA), chargé par l'IETF en 1996 de redéfinir les environnements IP sur ATM, CIOA est la solution la plus répandue aujourd'hui. Le sigle NBMA (Non Broadcast Multiple Access) a été attribué à tous les réseaux qui n'offrent pas une diffusion au niveau physique, comme celle obtenue dans un réseau Ethernet partagé. Un réseau ATM est un NBMA au même titre qu'un réseau relais de trames.

Pour réaliser la correspondance d'adresses, comme dans le couple IP-Ethernet, il faut un protocole de type ARP (Address Resolution Protocol), ici ATMARP (ATM's Address Resolution Protocol). Ce protocole est défini dans la RFC 1577, qui précise la notion de sous-réseau IP, ou LIS (Logical IP Subnetwork). Tous les utilisateurs connectés sur un LIS ont un préfixe d'adresse en commun. Un LIS regroupe l'ensemble des machines et des routeurs IP appartenant au même sous-réseau au sens IP. Un LIS comporte un serveur ATMARP, connu de toutes les machines connectées sur le LIS et contenant les correspondances d'adresses IP et ATM des stations du LIS.

CIOA (Classical IP over ATM) *(suite)*

Une station qui veut communiquer avec une autre station sur le LIS envoie une requête au serveur (phase 1), lequel, dans le cas standard, lui communique l'adresse ATM correspondante (phase 2), permettant à la station source d'ouvrir un circuit virtuel avec la station destination (phase 3). Ces trois phases sont illustrées à la figure 19.5.

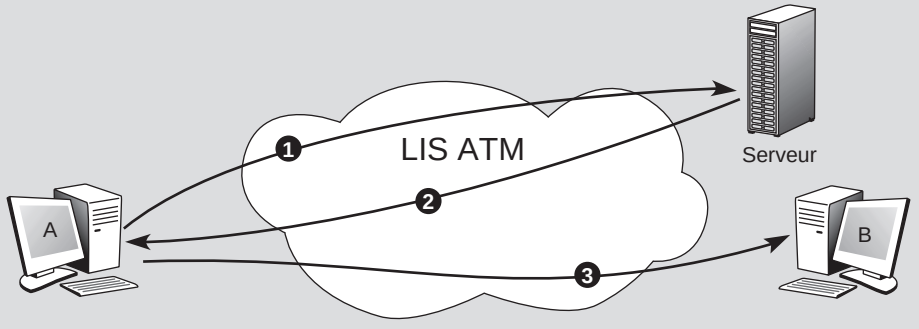


Figure 19.5
Connexion CIOA

Les paquets IP sont encapsulés dans des cellules ATM au moyen d'une fragmentation effectuée par l'AAL-5. Les spécifications de la fragmentation et du réassemblage sont indiquées dans les RFC 1577 et 1483. L'architecture protocolaire de l'encapsulation CIAO est illustrée à la figure 19.6.

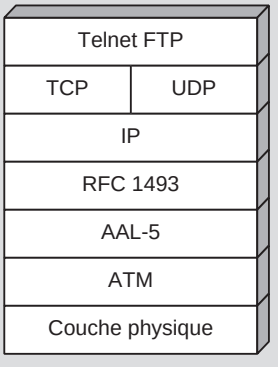


Figure 19.6
Architecture protocolaire de l'encapsulation CIAO

CIOA se place au niveau 3 de l'architecture OSI et utilise une résolution d'adresse IP directement en ATM. La résolution d'adresse de l'émulation LAN opère pour sa part au niveau MAC. Ce sont bien sûr deux solutions incompatibles. CIOA est beaucoup plus simple que l'émulation LAN, mais elle ne permet pas de gérer la diffusion.

De nombreuses autres possibilités d'encapsulation ont été proposées, dont les plus connues sont les suivantes :

- TULIP (TCP and UDP Lightweight IP), RFC 1932 ;
- TUNIC (TCP and UDP over a Non-existing IP Connection), également décrite dans la RFC 1932.

Le rôle de ces deux encapsulations concernant deux stations appartenant au même LIS est de simplifier les traitements dans le niveau IP en supprimant en grande partie l'en-tête.

Comme nous venons de le voir, la corrélation d'adresses dans CIOA se fait au niveau IP-ATM, ce qui simplifie la recherche de la correspondance d'adresses mais limite son utilisation aux réseaux IP. Avec la deuxième génération du protocole CIOA, la résolution d'adresse s'effectue par un mécanisme InAT-MARP (Inverse ATM's ARP), qui est une extension du mécanisme RARP (Reverse ARP) d'Internet *(voir le chapitre 17)*. Dans CIOA 2, le protocole NHRP, que nous détaillons à la section suivante, peut être utilisé.

Le groupe ION de l'IETF a mis au point le système MARS (Multicast Address Resolution Server) pour émuler le multicast au-dessus d'ATM. Ce service est étendu à tous les protocoles de la couche réseau au-dessus des réseaux NBMA. Le système MARS comporte un serveur et des clients. Dans le cadre d'IPv4, MARS ne travaille que sur un LIS. Il est possible d'ajouter des serveurs spécialisés MCS (Multicast Cluster Server) pour remplacer le serveur MARS dans la distribution des paquets et de la gestion des applications multicast. Il n'est toutefois pas évident de trouver l'architecture optimale entre une centralisation dans un serveur MARS unique et une distribution totale dans un réseau de serveurs MCS.

Avec l'arrivée d'IPv6, le protocole ATMARP ne peut plus être exploité. IPv6 au-dessus d'ATM remplace le processus ATMARP par ND (Neighbor Discovery), ce qui empêche le fonctionnement du protocole CIOA tel que nous l'avons décrit. De ce fait, l'IETF a normalisé dans les RFC 2491 et 2492 deux nouveaux protocoles pour le remplacer. Ces protocoles spécifient la mise en place d'IPv6 au-dessus de réseaux NBMA. Le protocole MARS est repris mais étendu pour transporter du trafic IPv6 unicast. Dans ce nouvel environnement, les LIS sont remplacés par des LL (Logical Link). Le serveur MARS réalise les fonctions auparavant réalisées par le serveur ATMARP.

NHRP et MPOA

Les deux solutions décrites précédemment, LANE et CIOA, s'appliquent facilement à un LIS (Logical IP Subnetwork) unique. Les protocoles utilisés par le BUS ou la procédure ATMARP requièrent une diffusion. Si les requêtes peuvent traverser des passerelles, la diffusion devient difficile à maîtriser. Il faut donc un protocole pour rechercher l'adresse du destinataire sans diffusion afin que l'environnement IP puisse se mettre au-dessus d'un ensemble de sous-réseaux ATM.

Considérons un ensemble de LIS ATM formant un NBMA. Chacun des réseaux ATM interconnectés a donc des utilisateurs possédant un préfixe d'adresse en commun et formant un LIS. Connaissant l'adresse IP du destinataire, il est possible de déterminer l'adresse ATM correspondante. La figure 19.7 illustre le processus consistant à trouver l'adresse ATM du destinataire en connaissant son adresse IP alors qu'il ne se trouve pas sur le même réseau.

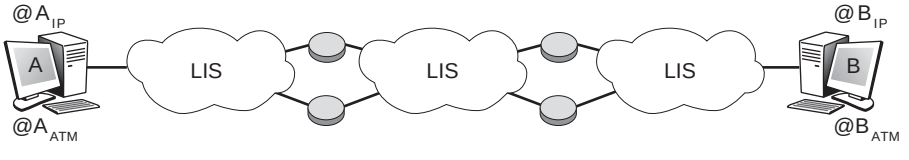


Figure 19.7
IP sur plusieurs LIS interconnectés

NHRP (Next Hop Resolution Protocol)

Le protocole NHRP provient du monde Internet et est décrit dans la RFC 1932. Il permet de rechercher l'adresse ATM correspondant à une adresse IP dans un réseau NBMA composé de plusieurs LIS. Plus précisément, NHRP permet la résolution d'une adresse IP d'une station de travail se trouvant sur un LIS distant en une adresse du réseau NBMA (adresse ATM, relais de trames, etc.).

Chaque LIS possède un serveur de route, appelé NHS (Next Hop Server), souvent situé dans un routeur. Lorsqu'un client demande une connexion, il s'adresse au NHS du LIS auquel il appartient pour obtenir les informations de routage sur son paquet. Si le NHS local ne peut résoudre le problème de la localisation, il adresse une requête vers les NHS connexes, et ainsi de suite jusqu'à arriver au LIS auquel le destinataire appartient.

Cette solution permet de trouver une route beaucoup plus directe que le passage par les différents NHS, comme l'illustre la figure 19.8. La phase 1 correspond à la demande de conversion d'adresse au NHR Routeur du premier LIS, lequel s'adresse avec la phase 2 au NHR Routeur du LIS dont dépend l'utilisateur distant. Les phases 3 puis 4 correspondent au retour de la conversion d'adresse. Avec l'adresse ATM le client ouvre un circuit virtuel avec le distant : c'est la phase 5. On peut ainsi obtenir une connexion directe en mode ATM de deux stations appartenant à des LIS distants, sans qu'il soit nécessaire de remonter au niveau IP du routeur.

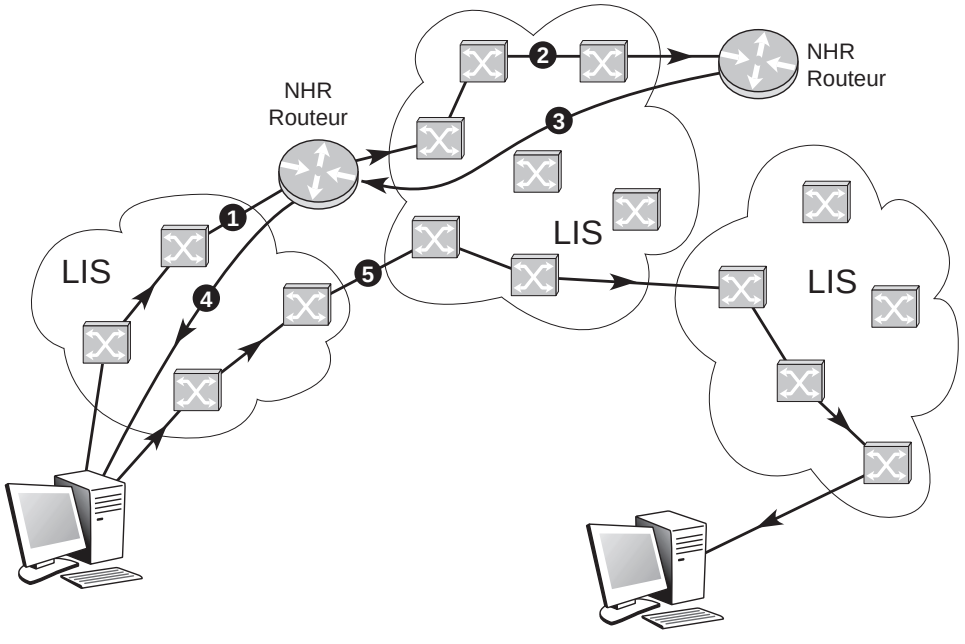


Figure 19.8
Mise en place d'une route par NHRP

MPOA (MultiProtocol Over ATM)

MPOA est un protocole mis au point par l'ATM Forum. Plus complexe que NHRP, il se sert des techniques décrites aux sections précédentes en les unissant et en les complétant

pour réaliser le transport de paquets IP ou de paquets d'autres protocoles, comme IPX, sur une interconnexion de réseaux ATM. La route peut être déterminée soit par une solution centralisée de type serveur de route, soit par une solution distribuée utilisant les protocoles PNNI ou NHRP.

Le rôle de MPOA est toujours de trouver l'adresse ATM du correspondant pour ouvrir une connexion directe, ou shortcut, entre deux stations ATM qui ne se connaissent au départ que par leur adresse IP.

Les deux composantes de MPOA sont les suivantes :

- MPC (MPOA Client), qui, à la demande d'un client, recherche la meilleure route pour ouvrir un circuit virtuel avec un client dont il connaît l'adresse IP.
- MPS (MPOA Server), situé dans un routeur, qui, à l'aide d'un routage classique, tel que RIP (Routing Information Protocol), OSPF (Open Shortest Path First), etc., achemine les requêtes NHRP de demandes de correspondance.

La figure 19.9 illustre le fonctionnement de MPOA. La phase 1 correspond à la demande de conversion d'adresse qui remonte jusqu'au serveur MPS connaissant la réponse. La phase 2 transporte la réponse à la demande de conversion qui permet l'ouverture du circuit virtuel vers le distant.

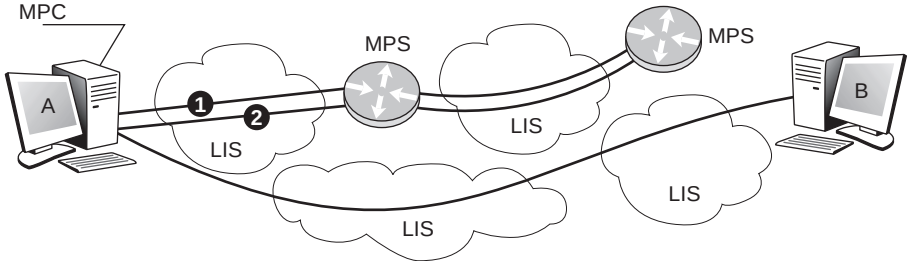


Figure 19.9
Fonctionnement de MPOA

PAR et I-PNNI

Issu de l'ATM Forum, le protocole PNNI a pour fonction de mettre en place une connexion entre deux utilisateurs en subdivisant le réseau en sous-réseaux, chaque sous-réseau possédant un nœud leader capable de connaître l'état des autres sous-réseaux et de renvoyer ces informations à ses propres nœuds dans son sous-réseau.

Lorsque des routeurs IP sont interconnectés par un ensemble de réseaux ATM, il est difficile de déterminer le chemin à suivre. Une solution pour trouver un chemin consiste à utiliser le protocole PNNI. Les mécanismes PAR et I-PNNI ont pour objet d'établir cette jonction entre les routeurs et le protocole PNNI.

- PAR (PNNI Augmented Routing) permet d'élire un serveur de route sur une machine de chaque sous-réseau ATM. Ce routeur est appelé DR (Designated Router). C'est lui qui est capable de faire la résolution d'adresse entre la partie IP et le réseau ATM et qui déclenche le protocole PNNI pour mettre en place une route sur l'interconnexion de réseaux ATM.
- I-PNNI (Integrated PNNI) étend le protocole PNNI de sorte qu'il puisse être utilisé sur les sous-réseaux IP. Dans chaque sous-réseau LIS, on indique en ce cas un leader.

MPLS (MultiProtocol Label-Switching)

MPLS est une norme proposée par l'IETF, l'organisme de normalisation d'Internet, pour l'ensemble des architectures et des protocoles de haut niveau (IP, IPX, AppleTalk, etc.). Cependant, son implémentation la plus classique concerne uniquement le protocole IP.

Les nœuds de transfert spécifiques utilisés dans MPLS sont appelés LSR (Label Switched Router). Les LSR se comportent comme des commutateurs pour les flots de données utilisateur et comme des routeurs pour la signalisation. Pour acheminer les trames utilisateur, on utilise des références, ou *labels*. À une référence d'entrée correspond une référence de sortie. La succession des références définit la route suivie par l'ensemble des trames contenant les paquets du flot IP.

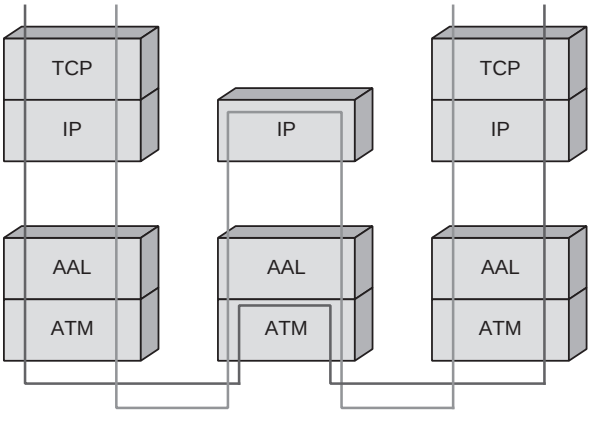
Toute trame utilisée en commutation, ou label-switching, peut être utilisée dans un réseau MPLS. La référence est placée dans un champ spécifique de la trame ou dans un champ ajouté dans ce but.

IP-switching

Introduite par la société Ipsilon, IP-switching a été la première version du label-switching. Dans cette architecture, la route est déterminée par le flot IP. Les nœuds IP-switch remplacent les routeurs en travaillant soit en mode routeur, pour tracer le chemin avec le premier paquet IP du flot, soit en mode commutation de cellules ATM, pour toutes les cellules qui suivent le chemin tracé. Le premier paquet IP est routé normalement, comme dans un réseau Internet. La route est déterminée par un algorithme de routage d'Internet.

Une fois déterminé le premier routeur à traverser, le paquet IP est subdivisé en cellules ATM pour traverser le premier LIS. Le paquet IP est recomposé au premier routeur IP-switch, lequel décide de la route à suivre, toujours à l'aide d'un algorithme de routage classique d'Internet. En même temps, une table de commutation est déterminée pour commuter les cellules du même flot. Après avoir franchi le premier nœud, le paquet IP est de nouveau fragmenté en cellules ATM, lesquelles sont émises vers le nœud suivant puis regroupées, et ainsi de suite. Tous les paquets IP suivants appartenant au même flot sont subdivisés en cellules ATM par l'émetteur et commutés sur le chemin tracé. Ce dernier devient un circuit virtuel ATM de bout en bout. La solution de routage-commutation de cette technique est illustrée à la figure 19.10.

Figure 19.10
TCP/IP sur ATM en IP-switching



L'ouverture de la route s'effectue à chaque nouveau flot se présentant dans le réseau. Si cette solution est assez fastidieuse à gérer, elle permet d'affecter une qualité de service à

chaque flot. Le principal reproche qui lui est adressé est de ne pas passer l'échelle, c'est-à-dire de ne pas pouvoir atteindre un très grand nombre de flots à gérer simultanément. Toshiba a proposé une solution comparable appelée, CSR (Cell Switching Router).

Les autres solutions pré-MPLS

Les autres solutions, principalement le tag-switching de Cisco Systems et ARIS (Aggregate Route-based IP Switching) d'IBM, utilisent des routes déterminées par la topologie et non plus par le flot. La norme MPLS a également choisi cette solution du choix de la route déterminée par la topologie pour des raisons de passage à l'échelle. Le tag-switching a été proposé quelques mois après l'annonce de l'IP-switching. Cisco a essayé de promouvoir sa solution *via* l'IETF, et de nombreuses RFC sont disponibles sur le type de sous-réseaux — ATM, PPP, Ethernet — à utiliser, ainsi que sur la possibilité de faire du multicast et d'utiliser le protocole RSVP.

La proposition ARIS d'IBM a été soumise à l'IETF sous forme de RFC. Comme dans le tag-switching, la mise en place des routes s'effectue par un algorithme dépendant de la topologie du réseau et non par une signalisation utilisant le premier paquet du flot de données. Les routes sont donc déterminées à l'avance.

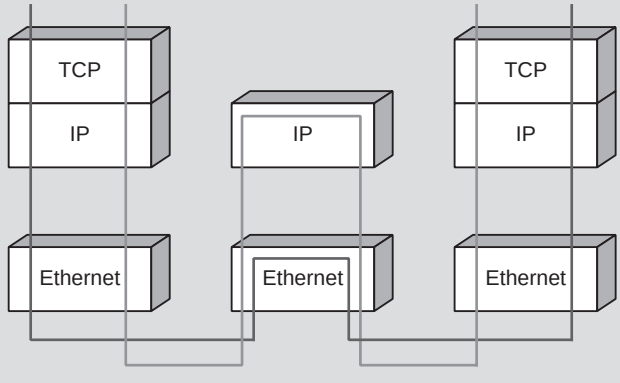
Ces solutions reposent sur le principe de la détermination d'une route entre l'émetteur et le récepteur, cette route étant établie par des serveurs de route. Les fragments de paquets IP sont étiquetés à l'entrée du réseau pour suivre la route déterminée. La route peut traverser des réseaux divers, aussi bien ATM que relais de trames ou Ethernet. La référence se trouve dans la zone VPI/VCI de la cellule ATM, dans la zone DLCI de la trame LAP-F d'un réseau relais de trames ou dans une zone supplémentaire de la trame Ethernet. On retrouve là la solution de mise en place d'une route à l'intérieur du réseau et de commutation de trames le long de cette route.

Les différences d'implémentation proviennent des antécédents des constructeurs. Si le constructeur propose des routeurs à son catalogue, il doit ajouter la partie ATM pour commuter les cellules ATM. Si le constructeur provient de l'environnement ATM, c'est un serveur de route IP qui est ajouté.

Le même type d'architecture peut se déployer directement au-dessus d'un environnement Ethernet, comme illustré à la figure 19.11.

Figure 19.11

Architecture d'un environnement IP sur Ethernet



Cette architecture est dictée par l'environnement Ethernet, qui brille par sa simplicité de mise en œuvre. Elle a l'avantage de s'appuyer sur l'existant, les coupleurs et les divers réseaux Ethernet, que de nombreuses sociétés ont mis en place pour créer leurs réseaux locaux. Puisque les données produites au format IP, IPX ou autre sont placées dans des trames Ethernet afin d'être transportées dans l'environnement local, il est tentant de commuter directement les trames Ethernet d'un réseau local vers un autre. Comme tous les réseaux de l'environnement Ethernet sont compatibles et parlent le même langage, les machines émettant des trames Ethernet peuvent s'interconnecter facilement. On peut ainsi réaliser des réseaux extrêmement complexes avec des segments partagés sur les parties locales, des liaisons commutées sur les longues distances ou entre les commutateurs Ethernet et des passages par des routeurs lorsqu'une remontée jusqu'au niveau IP est exigée.

Caractéristiques de MPLS

MPLS est l'aboutissement logique de toutes les propositions qui ont été faites dans les années 90. L'idée de l'IETF a été de proposer une norme commune pour transporter des paquets IP sur des sous-réseaux travaillant en mode commuté. Les nœuds sont des routeurs-commutateurs capables de remonter soit au niveau IP pour effectuer un routage, soit au niveau trame pour effectuer une commutation.

Les caractéristiques les plus importantes de la norme MPLS sont les suivantes :

- Spécification des mécanismes pour transporter des flots de paquets IP avec diverses granularités des flots entre deux points, deux machines ou deux applications. La granularité désigne la grosseur du flot, qui peut intégrer plus ou moins de flots utilisateur.
- Indépendance du niveau trame et du niveau paquet, bien que seul le transport de paquets IP soit réellement pris en compte.
- Mise en relation de l'adresse IP du destinataire avec une référence d'entrée dans le réseau.
- Reconnaissance par les routeurs de bord des protocoles de routage de type OSPF et de signalisation comme RSVP.
- Utilisation de différents types de trames.

Quelques propriétés supplémentaires méritent d'être soulignées :

- Ouverture du chemin fondée sur la topologie, bien que d'autres possibilités soient également définies dans la norme.
- Assignation des références faite par l'aval, c'est-à-dire à la demande d'un nœud qui émet un message dans la direction de l'émetteur.
- Granularité variable des références.
- Stock de références géré selon la méthode « dernier arrivé premier servi ».
- Possibilité de hiérarchiser les demandes.
- Utilisation d'un temporisateur TTL.
- Encapsulation d'une référence dans la trame incluant un TTL et une qualité de service.

Le principal avantage apporté par le protocole MPLS est la possibilité, illustrée à la figure 19.12, de transporter les paquets IP sur plusieurs types de réseaux commutés. Il est ainsi possible de passer d'un réseau ATM à un réseau Ethernet ou à un réseau relais de trames. En d'autres termes, il peut s'agir de n'importe quel type de trame, à partir du moment où une référence peut y être incluse. Nous verrons plus loin comment ajouter une référence lorsque la trame ne le prévoit pas.

Fonctionnement de MPLS

La transmission des données s'effectue sur des chemins nommés LSP (Label Switched Path). Un LSP est une suite de références partant de la source et allant jusqu'à la destination. Les LSP sont établis avant la transmission des données (control-driven) ou à la détection d'un flot qui souhaite traverser le réseau (data-driven).

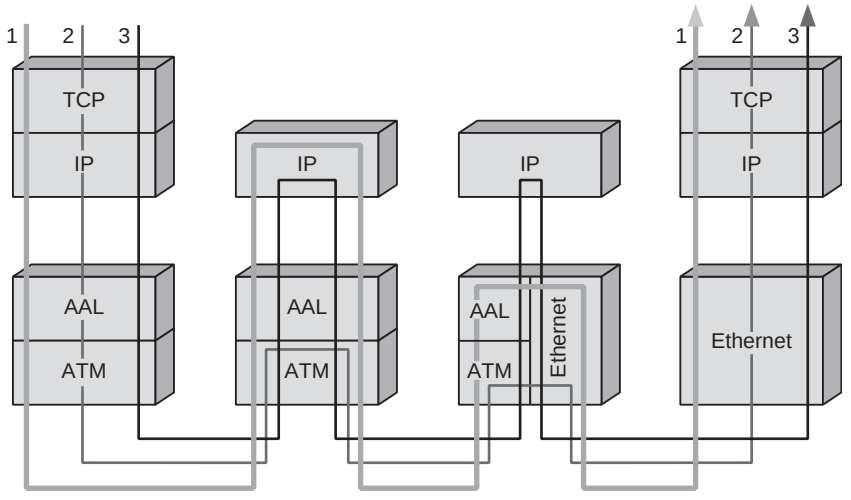


Figure 19.12

Réseau MPLS avec des sous-réseaux distincts

Les références incluses dans les trames sont distribuées en utilisant un protocole de signalisation. Le plus important de ces protocoles est LDP (Label Distribution Protocol), mais on utilise aussi RSVP, éventuellement associé à un protocole de routage, comme BGP (Border Gateway Protocol) ou OSPF. Les trames acheminant les paquets IP transportent les références de nœud en nœud.

LSR et LER (Label Edge Router)

Les nœuds qui participent à MPLS sont classifiés en LER et LSR. Un LSR est un routeur dans le cœur du réseau qui participe à la mise en place du circuit virtuel par lequel les trames sont acheminées. Un LER est un nœud d'accès au réseau MPLS. Un LER peut avoir des ports multiples permettant d'accéder à plusieurs réseaux distincts, chacun pouvant avoir sa propre technique de commutation. Les LER jouent un rôle important dans la mise en place des références.

LSR (Label Switched Router)

Un équipement qui effectue une commutation sur une référence s'appelle un LSR. Les tables de commutation LSFT (Label Switching Forwarding Table) consistent en un ensemble de références d'entrée auxquelles correspondent des ports de sortie. À une référence d'entrée peuvent correspondre plusieurs files de sortie pour tenir compte des adresses multipoint.

La table de commutation peut être plus complexe. À une référence d'entrée peut correspondre le port de sortie du nœud dans une première sous-entrée mais aussi, dans une deuxième sous-entrée, un deuxième port de sortie correspondant à la file de sortie du prochain nœud qui sera traversé, et ainsi de suite. De la sorte, à une référence peut correspondre l'ensemble des ports de sortie qui seront empruntés lors de l'acheminement du paquet.

Les tables de commutation peuvent être spécifiques de chaque port d'entrée d'un LSR et regrouper des informations supplémentaires, comme une qualité de service ou une demande spécifique de ressources.

FEC (Forwarding Equivalency Classes)

Dans MPLS, le routage s'effectue par l'intermédiaire de classes d'équivalence, appelées FEC. Une classe représente un flot ou un ensemble de flots ayant les mêmes propriétés, notamment le même préfixe dans l'adresse IP. Toutes les trames d'une FEC sont traitées de la même manière dans les nœuds du réseau MPLS. Les trames sont introduites dans une FEC au nœud d'entrée et ne peuvent plus être distinguées à l'intérieur de la classe des autres flots.

Une FEC peut être bâtie de différentes façons. Elle peut avoir une adresse de destination bien déterminée, un même préfixe d'adresse, une même classe de service, etc. Chaque LSR possède une table de commutation qui indique les références associées aux FEC. Toutes les trames d'une même FEC sont transmises sur la même interface de sortie. Cette table de commutation est appelée LIB (Label Information Base).

Les références utilisées par les FEC peuvent être regroupées de deux façons :

- Par plate-forme : les valeurs des références sont uniques sur l'ensemble des LSR d'un domaine, et les références sont distribuées sur un ensemble commun géré par un nœud particulier.
- Par interface : les références sont gérées par interface, et une même valeur de référence peut se retrouver sur deux interfaces différentes.

MPLS et les références

Une référence en entrée permet donc de déterminer la FEC par laquelle transite le flot. Cette solution ressemble à la notion de conduit virtuel dans le monde ATM, où les circuits virtuels sont multiplexés. Ici, nous avons un multiplexage de tous les circuits virtuels à l'intérieur d'une FEC, de telle sorte que, dans ce conduit, nous ne puissions plus distinguer les circuits virtuels.

Le LSR examine la référence et envoie la trame dans la direction indiquée. On voit bien ainsi le rôle capital joué par les LER, qui assignent aux flots de paquets des références qui permettent de commuter les trames sur le bon circuit virtuel. La référence n'a de signification que localement, puisqu'il y a modification de sa valeur sur la liaison suivante.

Une fois le paquet classifié dans une FEC, une référence est assignée à la trame qui va le transporter. Cette référence détermine le point de sortie par le chaînage des références. Dans le cas des trames classiques, comme LAP-F du relais de trames ou ATM, la référence est positionnée dans le DLCI ou dans le VPI/VCI.

La signalisation nécessaire pour déposer la valeur des références le long du chemin déterminé pour une FEC peut être gérée soit à chaque flot (data driven), soit par un environnement de contrôle indépendant des flots utilisateur. Cette dernière solution est préférable dans le cas de grands réseaux du fait de ses capacités de passage à l'échelle.

Les références peuvent être distribuées pour :

- un routage unicast vers une destination particulière ;
- une gestion du trafic, ou TE (Traffic Engineering) ;
- un multicast ;
- un réseau privé virtuel ;
- une qualité de service.

Le format de la référence MPLS est illustré à la figure 19.13. La référence est encapsulée dans l'en-tête de niveau trame du champ normalisé pour transporter la référence ou juste entre l'en-tête de niveau trame et l'en-tête de niveau paquet.

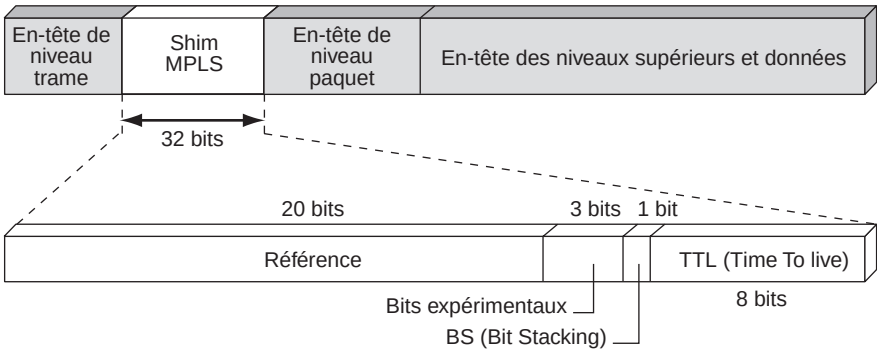
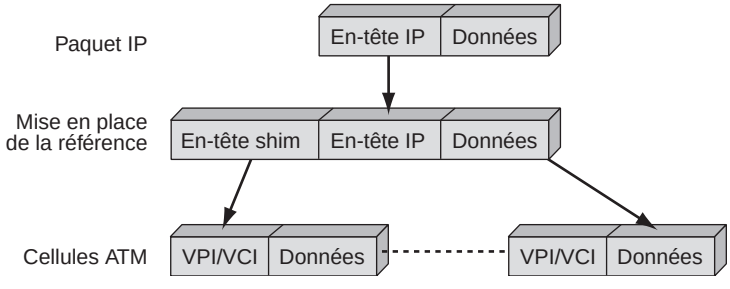


Figure 19.13
Format générique d'une référence dans MPLS

Les figures 19.14 et 19.15 illustrent la mise en place de la référence respectivement dans le cas d'ATM et du relais de trames. La figure 19.16 concerne le cas où la trame n'est pas conçue au départ pour un label-switching, comme la trame PPP.

Figure 19.14
Mise en place des références dans l'ATM



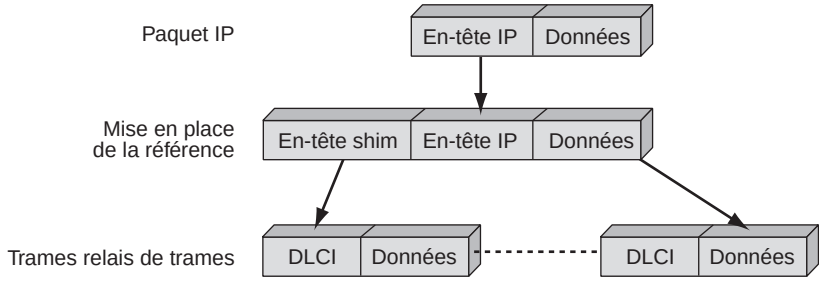
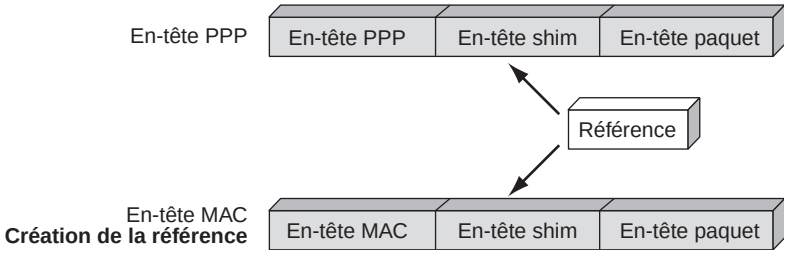


Figure 19.15
Mise en place des références dans le relais de trames

Figure 19.16
Mise en place des références dans la trame PPP



Distribution des références

MPLS normalise plusieurs méthodes pour réaliser la distribution des références. La distribution indique que chaque nœud possède ses propres références et qu'il doit les mettre en correspondance avec les références de ses voisins.

Les méthodes de distribution des références sont les suivantes :

- Topology-based, ou fondée sur la topologie, qui utilise les messages destinés à la gestion du routage, comme OSPF et BGP.
- Request-based, ou fondée sur le flot, qui utilise une requête de demande d'ouverture d'un chemin pour un flot IP. C'est le cas de RSVP.
- Traffic-based, ou fondée sur le trafic : à la réception d'un paquet, une référence est assignée à la trame qui le transporte.

Les méthodes fondées sur la topologie et sur le flot correspondent à un contrôle (control-based), tandis que celle fondée sur le trafic correspond à des données.

Les protocoles de routage, dont IGP (Interior Gateway Protocol), ont été améliorés pour transporter une référence supplémentaire. De même le protocole RSVP comporte une version associée à MPLS qui lui permet de transporter une référence. La version la plus aboutie est RSVP-TE (Traffic Engineering), qui permet l'ouverture de chemins en tenant compte des ressources du réseau.

L'IETF a également normalisé un nouveau protocole de signalisation, LDP (Label Distribution Protocol), pour gérer la distribution des références. Des extensions de ce protocole, comme CR-LDP (Constraint-based Routing-LDP), permettent de choisir les routes suivies par les clients des FEC avec une qualité de service prédéfinie.

Les principaux protocoles de signalisation sont les suivants :

- LDP, qui fait correspondre des adresses IP unicast et des références.
- RSVP-TE et CR-LDP, qui ouvrent des routes avec une qualité de service.
- PIM (Protocol Independent Multicast), qui fait correspondre des adresses IP multicast et des références associées.
- BGP, qui est utilisé pour déterminer des références dans le cadre de réseaux privés virtuels.

LSP (Label Switched Path)

Un domaine MPLS est déterminé par un ensemble de nœuds MPLS sur lesquels sont déterminés des FEC. Les LSP sont les chemins déterminés par les références positionnées par la signalisation. Les LSP sont déterminés sur un domaine avant l'arrivée des données dans le cas le plus classique. Deux options sont utilisées à cette fin :

- Le routage saut par saut (hop-by-hop). Dans ce cas, les LSR sélectionnent les prochains sauts indépendamment les uns des autres. Le LSR utilise pour cela un protocole de routage comme OSPF ou, pour des sous-réseaux de type ATM, PNNI *(voir les chapitres 15 et 17).*
- Le routage explicite, identique au routage par la source. Le LER d'entrée du domaine MPLS spécifie la liste des nœuds par lesquels la signalisation a été routée, le choix de cette route pouvant avoir été contraint par des demandes de qualité de service.

Le chemin suivi par les trames dans un sens de la communication peut être différent dans l'autre sens.

Agrégation de flots

Les flots provenant de différentes interfaces peuvent être rassemblés et commutés sur une même référence s'ils vont vers la même direction de sortie. Cela correspond à une agrégation de flots. Cette technique est déjà exploitée sur les réseaux ATM, dans lesquels un conduit peut agréger plusieurs flots venant de différents nœuds d'entrée vers un point commun, où les flots sont désagrégés.

L'agrégation de flots a pour objectif d'éviter l'explosion du nombre de références à utiliser ou, ce qui est équivalent, d'empêcher les tables de commutation de devenir trop importantes.

Signalisation

Comme expliqué précédemment, plusieurs mécanismes de distribution des références, appelée signalisation, peuvent être implémentés dans les nœuds d'un réseau MPLS, notamment les suivants :

- Demande de référence : un LSR émet une demande de référence à ses voisins vers l'aval (downstream), qu'il peut lier à la valeur d'une FEC. Ce mécanisme peut être utilisé de nœud en nœud jusqu'au nœud de sortie du réseau MPLS.
- Correspondance de référence : en réponse à une demande de référence d'un nœud amont, un LSR envoie une référence provenant d'un mécanisme de correspondance connu déjà mise en place pour aller jusqu'au nœud de sortie.

La figure 19.17 donne une illustration de ces deux mécanismes.

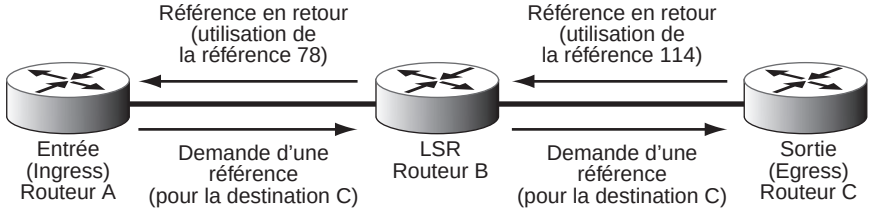


Figure 19.17

Mécanismes de signalisation de MPLS

LDP (Label Distribution Protocol)

LDP est le protocole de distribution des références qui tend à devenir le standard le plus utilisé dans MPLS. Ce protocole tient compte des adresses unicast et multicast. Le routage est explicite et est géré par les nœuds de sortie. Les échanges s'effectuent sous le protocole TCP pour assurer une qualité acceptable.

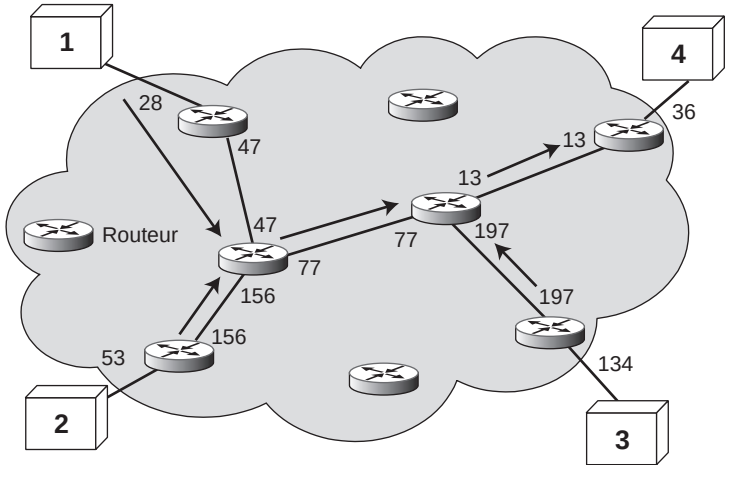
Deux classes de messages sont acceptées, celle des messages adjacents et celle des messages indiquant les références. La première permet d'interroger les nœuds qui peuvent être atteints directement à partir du nœud origine. La seconde classe de messages transmet les valeurs de la référence lorsqu'il y a accord entre les nœuds adjacents. Ces messages sont encodés sous la forme classique, qui permet de décrire un objet : on indique dans un premier champ le type d'objet, dans un deuxième la longueur totale du

message décrivant l'objet et dans un troisième la valeur de l'objet. Cet encodage s'appelle TLV (Type Length Value).

Le routage s'effectue, comme nous l'avons vu, par des classes d'équivalence, ou FEC (Forward Equivalent Class). Une classe représente une destination ou un ensemble de destinations ayant le même préfixe dans l'adresse IP. De ce fait, un paquet qui a une destination donnée appartient à une classe et suit une route commune avec les autres paquets de cette classe. Cela définit un arbre, dont la racine est le destinataire et dont les feuilles sont les émetteurs. Les paquets n'ont plus qu'à suivre l'arbre jusqu'à la racine, les flots se superposant petit à petit en allant vers la racine. Cette solution permet de ne pas utiliser trop de références différentes.

La granularité des références, c'est-à-dire la taille des flots qui utilisent une même référence, résulte de la taille des classes d'équivalence : s'il y a peu de classes d'équivalence, les flots sont importants, et la granularité est forte ; s'il y a beaucoup de classes d'équivalence, les flots sont faibles, et la granularité est fine. Par exemple, une destination peut correspondre à un réseau important, dans lequel toutes les adresses ont un préfixe commun. La destination peut aussi correspondre à une application particulière sur une machine donnée, ce qui donne une forte granularité. Ce dernier cas est illustré à la figure 19.18, dans laquelle le récepteur est la machine 1 et la FEC est déterminée par l'arbre dont les feuilles sont les machines terminales 1, 2 et 3. La classe d'équivalence, en descendant l'arbre à partir de 1, commence par les références 28 puis 47 et se continue par les branches 77 puis 13 puis 36. À partir de 2, les références 53 puis 156 sont utilisées pour aller vers la racine. De même, à partir de 3, les références 134 et 197 sont utilisées. Toutes les références que nous venons de citer appartiennent à la même classe d'équivalence.

Figure 19.18
Classes d'équivalence (FEC) dans un réseau MPLS



Dans cet exemple, les terminaux 1, 2 et 3 souhaitent émettre un flux de paquets IP vers la station terminale 4. Pour cela la station 1 émet ses trames (encapsulant les paquets IP) avec la référence 28, qui est commutée vers la référence 47 puis commutée vers les références 77 puis 13 puis 36. Le flot partant de la station 2 est commuté de 53 en 156 puis en 77, 13 et 36. Enfin, le troisième flot, partant de la station 3, est commuté à partir des valeurs 134 puis 197, 13 et 36. On voit que l'agrégation s'effectue sur les deux premiers flots avec la seule valeur 77 et que les trois flux sont agrégés sur les valeurs 13 et 36. La station 4 aurait pu être remplacée par un sous-réseau, ce qui aurait certainement permis d'agréger beaucoup plus de flux et d'avoir une granularité moins fine.

Un problème posé par les tables de routage impliquant les FEC est celui des boucles potentielles, c'est-à-dire d'un possible retour à une station qui a déjà vu passer la trame. Si le routage utilise un protocole comme OSPF, on évite les boucles en utilisant un message d'information.

Le protocole LDP comprend les messages suivants :

- Message de découverte (DISCOVERY MESSAGE), qui annonce et maintient la présence d'un LSR dans le réseau.
- Message de session (SESSION MESSAGE), qui établit, maintient et termine des sessions entre des ports LDP.
- Message d'avertissement (ADVERTISEMENT MESSAGE), qui crée, maintient et détruit la correspondance entre les références et les FEC.
- Message de notification (NOTIFICATION MESSAGE), qui donne des informations d'erreur ou de problème.

Les tables de commutation peuvent être construites et contrôlées de différentes façons. Les protocoles de routage d'Internet, tels que OSPF, BGP, PIM, etc., sont généralement utilisés à cet effet. Il faut leur ajouter des procédures pour faire correspondre les références et les classes d'équivalence FEC.

Nous avons indiqué que la distribution des références s'effectuait par l'aval en remontant vers la station d'émission. En réalité, il est indiqué dans la norme MPLS que la distribution des références peut s'effectuer par l'aval (downstream) ou par l'amont (upstream). Dans le premier cas, le destinataire indique aux nœuds amont la valeur de la référence à mettre dans la table de commutation. Dans le second cas, le paquet arrive avec une référence, et le nœud met à jour sa table de commutation.

Dans la distribution amont (upstream), un nœud aval envoie la valeur de la référence qu'il souhaite recevoir pour commuter un paquet sur une FEC. Ce sont les nœuds situés le plus en aval qui déclenchent le processus et indiquent les destinataires et leur granularité. Les modifications s'effectuent lors de la réception d'une trame ou par l'intermédiaire d'informations de supervision.

La distribution des identificateurs peut s'effectuer par l'intermédiaire des protocoles RSVP-TE ou PIM.

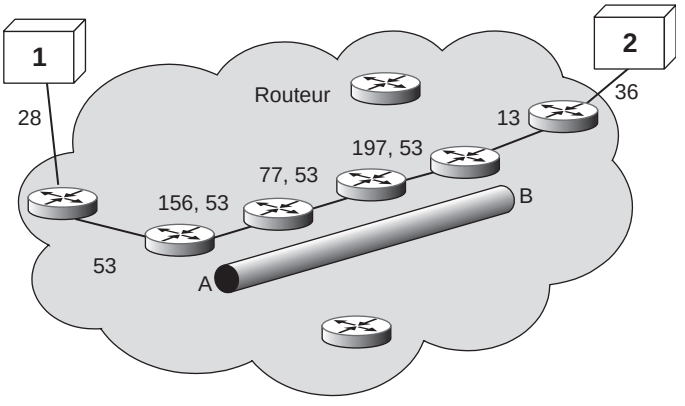
Les piles de références

Le mécanisme de piles de références de MPLS permet à un LSP de transiter par des nœuds non-MPLS ou par des domaines hiérarchiques. Pour cela, la zone portant la référence peut stocker non plus une valeur mais une pile de valeurs, c'est-à-dire une pile de références. Suivant le niveau de la hiérarchie de références on utilise la référence de la hiérarchie correspondante dans la pile.

Les piles de références permettent de réaliser des tunnels, dans lesquels sont regroupées les références d'un même niveau de la hiérarchie. À la sortie du tunnel, on revient à la hiérarchie juste en dessous, comme illustré à la figure 19.19. Sur cette figure, le flot partant de la station 1 est commuté sur les valeurs 28 puis 53. Au nœud A, une pile de références est créée avec l'ajout de la référence 156, qui est utilisée dans le nœud suivant pour commuter sur les valeurs 77 puis 197. Le nœud B permet la sortie du tunnel en utilisant de nouveau la référence 53 après avoir dépilé les références. On voit qu'entre le

nœud A et le nœud B un tunnel est constitué, qui, à une référence d'entrée 53, fait correspondre une référence de sortie 13.

Figure 19.19
Tunnel MPLS réalisé grâce à une pile de références



MPLS et l'ingénierie de trafic

Il est difficile de réaliser une ingénierie du trafic dans Internet du fait que le protocole BGP n'utilise que des informations de topologie du réseau. L'IETF a introduit dans l'architecture MPLS un routage à base de contrainte et un protocole de routage interne à état des liens étendu afin de réaliser une ingénierie de trafic efficace.

Comme nous l'avons vu, chaque trame encapsulant un paquet IP qui entre dans le réseau MPLS se voit ajouter par le LSR d'entrée, ou Ingress LSR, une référence au niveau de l'en-tête permettant d'acheminer la trame dans le réseau. Les chemins sont préalablement ouverts par un protocole de réservation de ressources, RSVP ou LDP. À la sortie du réseau, la référence ajoutée à l'en-tête de la trame est supprimée par le LSR de sortie, ou Egress LSR.

Des attributs permettant de contrôler les ressources attribuées à ces chemins sont associés au LSP, qui est le chemin construit entre le LSR d'entrée et le LSR de sortie. Ces attributs sont récapitulés au tableau 19.1. Ils concernent essentiellement la bande passante nécessaire au chemin, son niveau de priorité, son aspect dynamique, par l'intermédiaire du protocole utilisé pour son ouverture, et sa flexibilité en cas de panne.

Attribut	Description
Bande passante	Besoins minimaux de bande passante à réserver sur le chemin du LSP
Attribut de chemin	Indique si le chemin du LSP doit être spécifié manuellement ou dynamiquement par l'algorithme CBR (Constraint-Based Routing).
Priorité de démarrage	Le LSP le plus prioritaire se voit allouer une ressource demandée par plusieurs LSP.
Priorité de préemption	Indique si une ressource d'un LSP peut lui être retirée pour être attribuée à un autre LSP plus prioritaire.
Affinité ou couleur	Exprime des spécifications administratives.
Adaptabilité	Indique si le chemin d'un LSP doit être modifié pour avoir un chemin optimal.
Flexibilité	Indique si le LSP doit être rerouté en cas de panne sur le chemin du LSP.

TABLEAU 19.1 • Attributs des chemins LSP dans un réseau MPLS

L'algorithme CR (Constraint-based Routing)

L'algorithme CR est appliqué lors de l'ouverture du chemin ou de sa réouverture si le chemin est dynamique.

En plus des contraintes de topologie utilisées par les algorithmes de routage classiques, l'algorithme CR calcule les routes en fonction de contraintes de bande passante ou administratives. Les chemins calculés par le protocole CR ne sont pas forcément les plus courts. En effet, le chemin le plus court peut ne pas satisfaire la capacité de bande passante demandée par le LSP. Le LSP peut donc emprunter un autre chemin, plus lent mais disposant de la capacité de bande passante demandée. De la sorte, le trafic est distribué de manière plus uniforme sur le réseau.

L'algorithme CR peut s'effectuer en temps réel ou non. Dans le premier cas, le nombre de LSP à traverser est calculé à des instants quelconques par les routeurs sur la base d'informations locales. Dans le second cas, un serveur se charge, à partir d'informations recueillies sur tout le réseau, de calculer les chemins périodiquement et de reconfigurer automatiquement les routeurs avec les nouveaux chemins calculés.

Le protocole de routage est nécessaire pour le transport des informations de routage. Dans le cas de l'algorithme CR, le protocole de routage doit transporter, en plus des informations de topologie, des contraintes telles que les besoins en bande passante. La propagation de ces informations se fait plus fréquemment que dans le cas d'un IGP standard, puisqu'il y a plus de facteurs susceptibles de changer. Pour ne pas surcharger le réseau, il faut toutefois veiller que la fréquence de propagation des informations ne soit pas trop importante. Un compromis doit être trouvé entre le besoin d'actualiser les informations et celui d'éviter les propagations excessives.

La conception d'un système MPLS pour l'ingénierie de trafic nécessite de parcourir les étapes suivantes :

1. **Définition de l'étendue géographique du système MPLS.** Dépend de la politique administrative et de l'architecture du réseau.
2. **Définition des routeurs membres du système MPLS.** Il s'agit de définir les LSR d'entrée, de transit et de sortie du système MPLS. Pour diverses raisons, ce dernier ne contient pas nécessairement tous les routeurs du réseau, notamment si un routeur n'est pas assez puissant ou s'il n'est pas sécurisé.
3. **Définition de la hiérarchie du système MPLS.** Deux cas sont possibles : connecter tous les LSR du système MPLS et créer un seul niveau de hiérarchie formant un grand système MPLS ou diviser le réseau en plusieurs niveaux de hiérarchie. Dans ce dernier cas, les LSR de premier et deuxième niveau de la hiérarchie, qui forment le cœur du réseau, sont fortement maillés.
4. **Définition des besoins en bande passante des LSP.** Les besoins en bande passante peuvent être définis par la matrice de trafic de bout en bout, qui n'est pas toujours disponible, ou par un calcul statistique fondé sur l'exploitation des LSP et la mise à jour régulière de cette information en observant constamment leur trafic.
5. **Définition des chemins des LSP.** Les chemins sont généralement calculés de manière dynamique par un CR temps réel. Lorsqu'il se révèle difficile de réaliser ce calcul en temps réel, on peut utiliser un algorithme CR non-temps réel.
6. **Définition des priorités des LSP.** On peut attribuer la plus haute priorité à des LSP devant écouler un trafic volumineux. Cela permet d'emprunter les chemins les plus courts et d'éviter de surcharger un grand nombre de liens dans le réseau, tout en offrant une stabilité du routage et une meilleure utilisation des ressources.

7. **Définition du nombre de chemins parallèles entre deux extrémités quelconques.** On peut configurer plusieurs chemins en parallèle ayant des routes physiquement différentes. Cela garantit une distribution de la charge du trafic plus uniforme. L'idée sous-jacente est de définir des LSP de petite taille en vue d'une meilleure flexibilité du routage. Cette flexibilité est la première motivation des LSP parallèles.
8. **Définition de l'affinité des LSP et des liens.** Des couleurs peuvent être attribuées aux LSP et aux liens en fonction de contraintes administratives. Ces couleurs servent à déterminer les chemins à choisir pour les LSP.
9. **Définition des attributs d'adaptation et de flexibilité.** Selon l'évolution du comportement du réseau, il est possible de trouver des chemins optimaux pour les LSP déjà calculés. L'administrateur réseau peut accepter ou refuser une nouvelle optimisation des LSP. Il ne faut pas que cette dernière soit trop fréquente, car elle pourrait introduire une instabilité du routage. Il faut aussi prévoir des mécanismes de reroutage des LSP en cas de panne d'un LSR.

L'exploitation d'un réseau MPLS suit les étapes énumérées ci-dessous :

1. **Recueil des données statistiques en utilisant les LSP au démarrage du système.** L'objectif de cette étape est de calculer le taux de trafic entre chaque paire de routeurs. Les méthodes statistiques existantes permettent de calculer le taux de trafic à l'entrée et à la sortie d'une interface mais pas celui allant vers une destination particulière. La construction de la matrice de trafic de bout en bout est effectuée par estimation, ce qui rend l'ingénierie de trafic difficile et peu efficace. L'utilisation des LSP au démarrage d'un système MPLS donne précisément le taux de trafic entre deux extrémités quelconques en fonction des destinations.
2. **Exploitation des LSP avec les contraintes de bande passante définies à l'étape précédente.** L'étape 1 ci-dessus ayant permis de connaître les besoins en bande passante de chaque LSP, cette information est utilisée par l'algorithme CR pour recalculer les LSP avec leur besoin réel en bande passante.
3. **Mise à jour périodique des bandes passantes des LSP.** Une mise à jour périodique des bandes passantes des LSP est nécessaire pour assurer l'évolution et l'adaptation du réseau au changement du trafic dans le réseau.
4. **Exécution de l'algorithme CR en temps réel.** Pour une utilisation efficace des liens, l'algorithme CR doit être exécuté sur un serveur spécialisé. Calculé sur un serveur disposant de toutes les informations de topologie et d'attributs de tous les LSP, cet algorithme peut permettre d'atteindre le temps réel. L'algorithme propose des LSP ayant de meilleures performances comparées à celles des LSP déjà ouverts. L'algorithme CR doit pouvoir s'exécuter en temps réel pour tenir compte d'une panne d'un LSP. L'algorithme peut alors déterminer rapidement un nouvel LSP capable d'écouler le trafic en attente.

La qualité de service dans MPLS

Nous venons de voir que MPLS permettait de faire de l'ingénierie et d'effectuer des calculs pour déterminer les ressources à affecter à un chemin lorsque le système est relativement statique. Si le système est dynamique, des chemins doivent s'ouvrir et se fermer pour satisfaire à des contraintes qui s'expriment sur des laps de temps plus courts. L'idée de base est d'ouvrir les chemins grâce à un algorithme tenant compte des ressources. Nous avons déjà

examiné la proposition CR-LDP. Cet algorithme ayant été partiellement abandonné, un autre algorithme, RSVP-TE, a pris une place de choix parmi les équipementiers.

Dans CR-LDP, les deux ports qui doivent communiquer s'échangent leur ensemble de références pour établir la connexion. Dans RSVP-TE, il n'y a pas de négociation de références. C'est le plan de gestion qui prend à sa charge cette négociation. Pour de très grands réseaux, la mise en place du chemin avec LDP peut nécessiter des ressources considérables, ce qui explique son échec pour le moment.

CR-LDP peut spécifier la route à partir de la source par un champ de type TLV et RSVP-TE par le biais de l'objet « explicit route ». Les deux protocoles envoient une réponse au nœud d'entrée pour indiquer le succès ou l'échec de l'ouverture du chemin.

Les tableaux 19.2 et 19.3 récapitulent respectivement les similitudes et différences entre les deux techniques.

Caracteristique	CR-LDP	RSVP-TE	Commentaire
Initialisation de l'ouverture	Message LABEL_REQUEST	RSVP-TE Message PATH contenant l'objet LABEL_REQUEST	
Ouverture	DIFF-SERV_PSC TLV	Objet DIFFSERV_PSC	Les deux contiennent l'information correspondant au DSCP (DiffServ Code Point) inclus dans le message de demande d'ouverture.
Accepte les LSP point-à-multipoint	Non	Non	En attente d'une RFC
Possibilité d'un routage par la source	Transporté pa la liste TLV de EXPLICIT_ROUTE	Transporté par l'objet EXPLICIT_ROUTE	Spécifie le chemin à suivre.

TABLEAU 19.2 • Similitudes entre RSVP-TE et CR–LDP

Caracteristique	CR-LDP	RSVP-TE	Commentaire
Étape de développement	Le plus jeune mais non utilisé aujourd'hui	Le plus ancien, avec des ajouts pour tenir compte des divers réseaux disponibles dans MPLS	Certains objets de RSVP ont été modifiés pour être utilisés dans MPLS.
Signalisation	UDP pour la découverte et TCP pour la session	Paquets IP ou encapsulation dans UDP pour l'échange de messages	Pas de détection de panne déterministe avec RSVP-TE. Un problème sur TCP peut avoir un impact catastrophique sur les chemins dans CR-LDP.
État de la connexion	Hard State	Soft State	Le Soft State ne passe généralement pas l'échelle. RSVP prend en charge l'agrégation des messages de rafraîchissement.
Fiabilité	Défini pour prendre en charge la plupart des techniques trame, comme ATM, le relais de trames ou Ethernet.	Tunneling à travers le réseau ATM qui doit être configuré manuellement.	

TABLEAU 19.3 • Différences entre RSVP-TE et CR–LDP

GMPLS (Generalized MPLS)

Comme son nom l'indique, GMPLS est une généralisation du protocole MPLS. Cette généralisation est assez simple à expliquer, puisque tout ce qui peut jouer le rôle d'une référence — numéro d'une longueur d'onde, numéro d'un slot, etc. — peut entrer dans GMPLS. La structure de GMPLS est toutefois plus complexe qu'il n'y paraît, et une gestion globale est nécessaire pour arriver à bien contrôler cet environnement.

Les extensions de MPLS

MPLS ne travaille que sur des structures de trame de niveau 2, le L2S (Level 2 Switching). Des extensions permettent toutefois d'introduire des références sur d'autres supports, comme le numéro d'une tranche de temps dans un partage temporel ou un numéro de longueur d'onde sur une fibre optique.

Les principales possibilités d'extension de MPLS sont les suivantes :

- PSC (Packet Switching Capable), pour les paquets capables de recevoir une référence. On pourrait imaginer un paquet IPv6 avec le flow-label comme référence, mais cette solution n'est pas acceptable directement car un paquet ne peut être transmis directement sur un support physique. Pour cela, il faut encapsuler le paquet dans une trame. C'est généralement la trame PPP qui sert de transporteur.
- L2SC (Level 2 Switching Capable), qui correspond au label-switching utilisé dans la norme MPLS.
- TDMC (Time Division Multiplexing Capable), qui introduit la référence en tant que slot dans un multiplexage temporel. Toutes les techniques qui comportent une structure sous forme de trame avec des slots à l'intérieur font partie de cette classe. En particulier, toutes les techniques hertziennes avec division temporelle s'intègrent dans GMPLS.
- LSC (Lambda Switching Capable), qui prend le numéro de la longueur d'onde à l'intérieur d'une fibre optique comme référence de commutation. Cette technique a été la première extension de MPLS sous le nom de MPλS.
- FSC (Fiber Switching Capable), qui prend le numéro d'une fibre optique parmi un faisceau de fibres optiques comme référence de commutation. Dans un faisceau, les fibres sont numérotées de 1 à *n*, *n* correspondant au nombre de fibres optiques.

Le tableau 19.4 récapitule les techniques de transfert offertes par un réseau GMPLS.

Domaine de transfert	Type de trafic	Type de transfert	Exemple de station	Nomenclature
Trame	ATM, Ethernet	Utilisation de références	Commutateur ATM ou Ethernet	L2SC (Layer 2 Switching Capable)
Paquet	IP	Routage	Routage IP	PSC (Packet Switching Capable)
Temps	TDM/SONET	Slot de temps se répétant par cycle	Brasseur et commutateur	TDMC (Time Division Multiplexing Capable)
Longueur d'onde	Transparent	Lambda	DWDM	LSC (Lambda Switching Capable)
Espace physique	Transparent	Fibre optique	OXC (Optical Cross Connect)	FSC (Fiber Switching Capable)

TABLEAU 19.4 • Techniques de transfert de GMPLS

D'autres extensions sont imaginables, comme l'association d'un code dans une communication, que ce soit dans un CDMA ou dans une transmission quelconque. Par ces extensions, il est possible de faire correspondre en entrée et en sortie des références qui ne proviennent pas de la même technologie. En revanche, les différentes solutions ne donnent pas forcément des débits identiques. Par exemple, si l'on choisit comme référence une tranche avec un numéro bien déterminé d'un multiplex temporel hertzien, qui risque de donner au mieux quelques mégabits par seconde, il est difficile de lui faire correspondre en sortie une longueur d'onde d'une fibre optique qui peut avoir une capacité de 10 Gbit/s. Une hiérarchisation des supports est donc nécessaire.

Hiérarchie des supports et réseaux overlay

La figure 19.20 illustre une hiérarchie possible entre les supports qui peuvent être utilisés dans GMPLS. Dans cette figure, un flot de paquets IP donne naissance à un PSC, lui-même intégré dans un L2SC de type FEC, c'est-à-dire rassemblant plusieurs flots IP ayant une propriété commune, comme un même LSR de sortie.

Les flots de niveau L2CS peuvent eux-mêmes être encapsulés dans un slot d'une technique de type SONET-SDH. En continuant dans la hiérarchie, les flots TDMC peuvent être à leur tour multiplexés dans une même longueur d'onde, c'est-à-dire dans un LSC. En continuant la hiérarchie pour arriver au plus haut niveau, les longueurs d'onde peuvent elles-mêmes être intégrées dans une fibre particulière d'un faisceau de fibre optique.

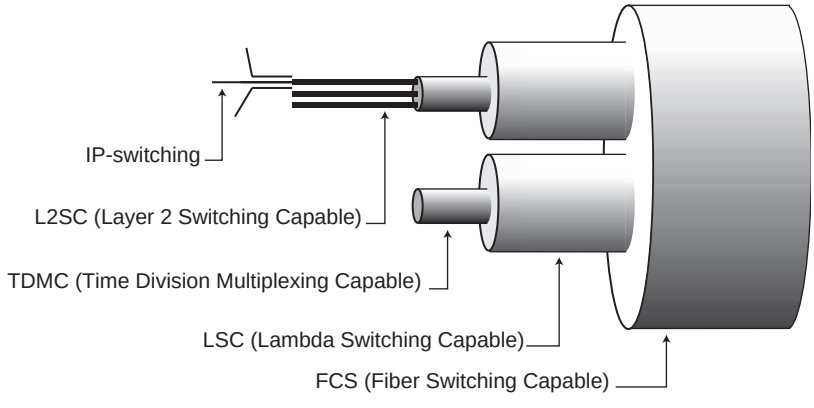


Figure 19.20

Hiérarchie des techniques de transfert dans GMPLS

Une autre façon de voir cette hiérarchie consiste à raisonner en réseau overlay, c'est-à-dire en une hiérarchie de réseaux, comme illustré à la figure 19.21, où trois niveaux sont représentés.

Si l'on suppose, pour simplifier, que le réseau global ne comprend que deux niveaux de hiérarchie, comme illustré à la figure 19.22, chaque nœud du réseau overlay dessert un réseau du niveau sous-jacent. Pour aller d'un point à un autre, de A à D par exemple, le paquet doit être envoyé par le réseau local au nœud d'entrée du réseau overlay, c'est-à-dire de A à B sur la figure, puis transmis dans le réseau overlay de B à C et enfin dans le réseau local d'arrivée de C à D.

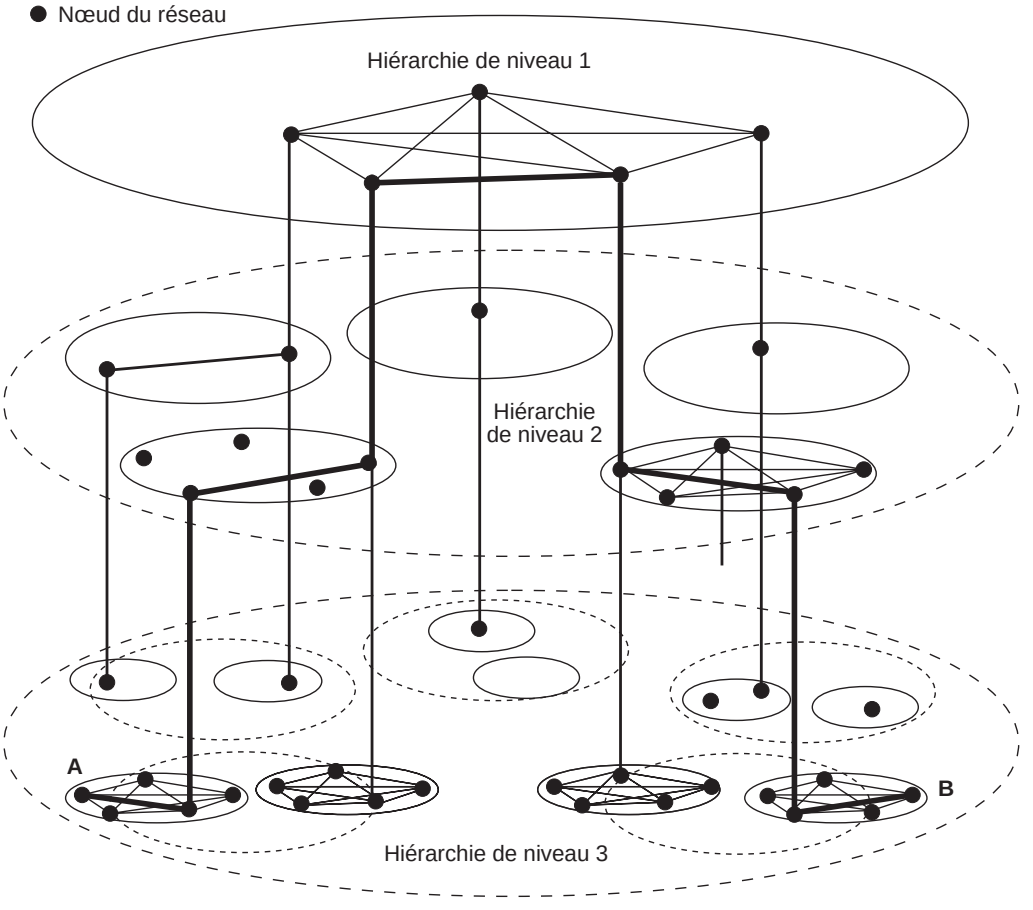
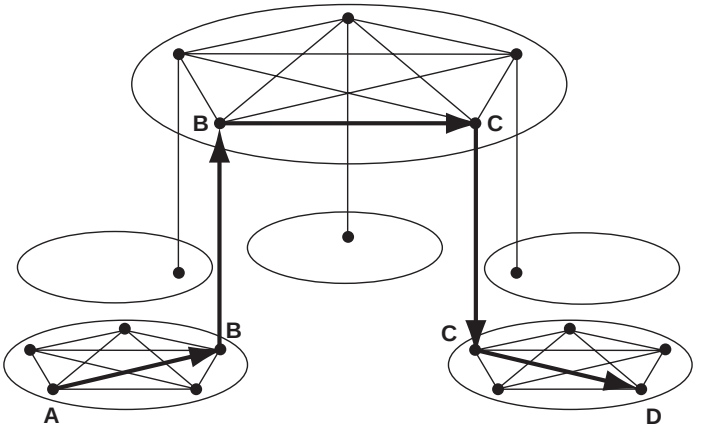


Figure 19.21

Hiérarchie de réseau à trois niveaux

Figure 19.22

Fonctionnement d'un réseau overlay



Si les différents niveaux de la hiérarchie comportent des réseaux maillés, qui permettent d'aller directement d'un point à un autre dans le réseau, on voit que cette solution de réseau permet de limiter le nombre de nœuds à traverser. Dans le cas de la figure 19.22, pour aller de A à D, l'on ne passe que par deux nœuds intermédiaires, alors que si tous les nœuds du réseau avaient été au même niveau, il aurait fallu peut-être une dizaine de sauts.

La structure hiérarchique des supports de transmission de GMPLS permet de mettre en place ce type de réseau. On peut, par exemple, dans un cas simple, avoir des domaines MPLS de niveau 2 interconnectés par un réseau overlay utilisant une longueur d'onde sur une fibre optique. Ce réseau overlay relie les points des domaines de base choisis pour faire partie du réseau overlay.

Pour ouvrir des chemins sur des réseaux différents les uns des autres, un ensemble de protocoles de contrôle et de surveillance est nécessaire. Un premier problème posé par le routage dans les réseaux overlay concerne le contrôle de la connectivité, qui est pris en charge par des messages de type HELLO, envoyés régulièrement sur toutes les interfaces. Chaque HELLO doit être acquitté explicitement. Lorsque aucun ACK n'est reçu, la ligne est considérée comme étant en panne. Dans le cas de GMPLS sur fibre optique, il n'est pas possible d'envoyer des messages HELLO. Le contrôle de la connectivité doit donc se faire par un nouveau protocole.

Un second problème posé par les réseaux overlay provient de l'impossibilité pour des nœuds de même niveau mais n'appartenant pas au même domaine de se transmettre directement des messages de contrôle. Il faut passer par un réseau de niveau supérieur, lequel peut ne pas être capable d'interpréter les messages des niveaux inférieurs. Il n'y a donc pas de vision globale du réseau.

Pour améliorer le contrôle et la gestion, il est nécessaire de bien séparer les plans utilisateur, gestion et contrôle, surtout si le réseau est complexe. Cela vaut encore davantage dans les réseaux utilisant de la fibre optique.

Comme pour l'ATM, on distingue trois plans dans GMPLS :

- Le plan utilisateur, qui est chargé de transporter les données utilisateur d'une extrémité à l'autre.
- Le plan de contrôle, destiné à mettre en place les circuits virtuels puis à les détruire à la fin de la transmission ou à les maintenir si nécessaire.
- Le plan de gestion, qui transporte les messages nécessaires à la gestion du réseau.

Les groupes de travail de GMPLS ont développé une telle architecture pour permettre de contrôler par un plan spécifique l'ensemble des composants du réseau.

Pour s'adapter au protocole GMPLS, les protocoles de signalisation (RSVP-TE, CR-LDP) et les protocoles de routage (OSPF-TE, IS-IS-TE) ont été étendus. Un nouveau

protocole de gestion, appelé LMP (Link Management Protocol), a été introduit pour gérer les plans utilisateur et de contrôle. LMP est un protocole IP qui contient des extensions pour RSVP-TE et CR-LDP.

Le tableau 19.5 récapitule les propriétés de ces protocoles et leurs extensions dans le cadre de GMPLS.

Protocole	Description
Routage (OSPF-TE, IS-IS-TE)	Destiné à la découverte automatique de la topologie du réseau et à la mesure de la disponibilité des ressources (bande passante, type de protection). Les principales améliorations sont les suivantes : - Indication du type de protection (1+1, 1:1, non protégé, trafic en plus). 1+1 indique qu'un chemin de secours est ouvert en permanence, 1:1 qu'en cas de panne un chemin de secours est prévu mais sans réservation de ressource. - Implémentation de lignes de dérivation pour améliorer le passage à l'échelle. - Acceptation et indication de liaisons qui n'ont pas d'adresse IP; utilisation d'une identification Link ID. - Identité des interfaces d'entrée et de sortie (interface ID). - Découverte d'un chemin pour un back-up utilisant un chemin différent du chemin primaire (shared-risk link group).
Signalisation (RSVP-TE, CR-LDP)	Destiné à la mise en place des chemins par une ingénierie de trafic. Les principales améliorations sont les suivantes : - Échange des références avec des réseaux non paquet (référence généralisée). - Établissement de chemin LSP bidirectionnel. - Signalisation pour l'ouverture d'un chemin de back-up. - Proposition de références suggérées. - Accepte la commutation de longueur d'onde.
LMP (Link Management Protocol)	Inclut les extensions suivantes : - Control Channel Management : établit, lors de la négociation, les paramètres de la liaison, tels la fréquence d'émission des messages KEEP_ALIVE et HELLO. - Link Connectivity Verification : permet de s'assurer de la connectivité physique entre les nœuds voisins grâce à des messages de type PING. - Link Property Correlation : détermine les mécanismes de protection. - Fault Isolation : isole les fautes simples ou multiples du domaine optique.

TABLEAU 19.5 • Propriétés et extensions des protocoles de GMPLS

Les différentes couches que nous avons examinées forment l'architecture de GMPLS illustrée à la figure 19.23.

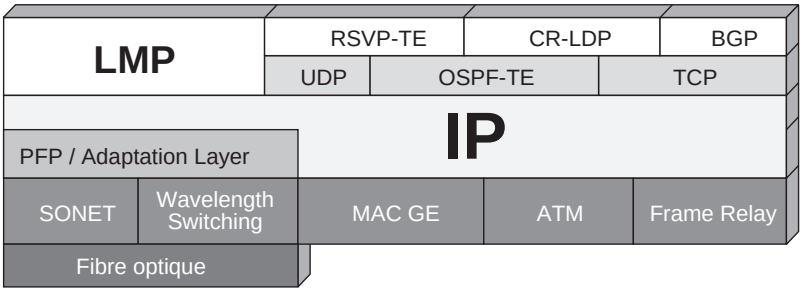


Figure 19.23

Architecture de GMPLS

Conclusion

Les réseaux MPLS et GMPS sont promis à un bel avenir. Presque tous les grands opérateurs ont investi dans cette direction, non sans une certaine appréhension quant à la complexité globale de cette nouvelle architecture, qui peut être vue comme un compromis entre un grand nombre d'architectures différentes.

Le plan utilisateur semble bien conçu et permet d'optimiser assez facilement la mise en place du réseau et son ingénierie, notamment pour ce qui concerne la qualité de service, la maintenance et la gestion. Cependant, de nombreux problèmes de compatibilité entre équipementiers se posent encore.

MPLS a aussi été retenu pour réaliser des réseaux privés virtuels grâce à ses chemins qu'il est relativement facile de protéger. Nous examinons ces solutions de VPN au chapitre 32.

Références

Ce livre de la collection Cisco Press donne des détails intéressants sur la mise en place de MPLS :

V. ALWAYN – *Advanced MPLS Design and Implementation*, Cisco Press, 2001

Un article qui donne une bonne présentation de MPLS, avec ses avantages et ses inconvénients :

G. ARMITAGE – *"MPLS: The Magic Behind the Myths", IEEE Communications Magazine*, janvier 2000

Autre article proposant une synthèse sur l'ingénierie du trafic dans MPLS :

D. O. AWDUCHE – *"MPLS and Traffic Engineering in IP Networks", IEEE Communications Magazine*, décembre 1999

La technologie MPLS peut également être utilisée pour les réseaux métropolitains :

M. J. BAGAJEWICZ – *MPLS for Metropolitan Area Networks*, Auerbach Publications, 2004

Excellent livre sur toutes les techniques modernes et notamment de commutation :

D. P. BLACK – *Building Switched Networks: Multilayer Switching, Qos, IP Multicast, Network Policy, and Service-Level Agreements*, Addison Wesley, 1999

Ce bon livre sur MPLS est un des très nombreux livres de U. Black :

U. BLACK – *MPLS and Label Switching Networks*, Prentice Hall, 2002

Le livre le plus orienté vers les techniques de commutation-routage de type label-switching et TCP/IP sur ATM :

B. DAVIE, P. DOOLAN, Y. REKHTER – *Switching in IP Networks*, Morgan Kaufmann Publishers, 1998

Livre complet sur MPLS :

B. S. DAVIE, Y. REKHTER – *MPLS: Technology and Applications*, Morgan Kaufmann Publishers, 2000

Excellent livre sur la qualité de service dans les différentes technologies (IP, ATM et Ethernet) :

P. FERGUSON, G. HUSTON – *Quality of Service*, Wiley, 1998

Un très bon livre pour comprendre jusque dans les détails d'implémentation les points sensibles de la technologie MPLS :

R. GALLAHER – *MPLS Training Guide: Building Multi Protocol Label Switching Networks*, Syngress Publishing, 2003

Juniper et Cisco sont les deux grands équipementiers pour les routeurs et les LSR. Ce livre beaucoup plus large que MPLS introduit bien leurs deux visions, qui ne sont pas toujours compatibles :

W. J. GORALSKI – *Juniper and Cisco Routing: Policy and Protocols for Multivendor Networks*, Wiley, 2002

Les réseaux qui résistent aux pannes sont particulièrement importants pour les opérateurs de télécommunications. Le livre suivant introduit bien les différentes solutions et en particulier la solution utilisant MPLS :

W. D. GROVER – *Mesh-based Survivable Transport Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*, Prentice Hall, 2003

Le livre suivant est fortement orienté VPN (Virtual Private Networks) :

J. GUICHARD, I. PEPELNJAK – *MPLS and VPN Architectures: A Practical Guide to Understanding, Designing and Deploying MPLS and MPLS-Enabled VPNs*, Cisco Press, 2000

Excellent livre de départ sur MPLS :

S. HARBEDY – *The MPLS Primer: An Introduction to Multiprotocol Label Switching*, Prentice Hall, 2001

Un bon livre sur les technologies qui permettent de faire du multiservice. MPLS en fait partie :

D. PAW – *ATM & MPLS Theory & Application: of Multi-Service Networking*, McGraw-Hill, 2002

Le livre suivant décrit en détail les différentes technologies de commutation et d'émulation :

D. MINOLI, A. ALLES – *LAN, ATM and LAN Emulation Technologies*, Artech House, 1997

Un livre orienté gestion de réseau et donc MPLS :

S. B. MORRIS – *Network Management, MIBs and MPLS: Principles, Design and Implementation*, Prentice Hall, 2003

Un excellent livre sur l'ingénierie du trafic dans les réseaux MPLS, qui est la raison d'être première de ce type de réseau :

E. OSBORNE, A. SIMHA – *Traffic Engineering with MPLS*, Pearson Education, 2002

Les deux livres suivants donnent un panorama complet et d'excellente qualité des architectures MPLS et de leur application pour la réalisation de réseaux privés virtuels :

I. PEPELNJAK, J. GUICHARD – *MPLS and VPN Architectures*, vol. 1, Cisco Press, 2000

I. PEPELNJAK, J. GUICHARD, J. APCAR – *MPLS and VPN Architectures*, vol. 2, Cisco Press, 2003

L'architecture du protocole MPLS provient de cette proposition de l'IETF :

> E. ROSEN, A. VISWANATHAN, R. CALLON – *"A Proposed Architecture for MPLS"*, IETF, *Internet draft*, juillet 1997

Livre proposant une bonne description de l'ATM et de ses possibilités en tant que technique de transfert supportant différents types de protocoles, en particulier IP :

> G. C. SACKETT, C. Y. METZ – *ATM and Multiprotocol Networking*, McGraw-Hill, 1997

Les protocoles de routage jouent un rôle important dans MPLS. Le livre suivant part de cette constatation pour introduire MPLS et les techniques IP sur ATM :

> S. A. THOMAS – *IP Switching and Routing Essentials: Understanding RIP, OSPF, BGP, MPLS, CR-LDP, and RSVP-TE*, Wiley, 2001