MPLS

par Bernard Cousin
Laboratoire IRISA
Université de Rennes-1
bcousin@irisa.fr
http://www.irisa.fr/adp

- Plan
  - Introduction
  - Le "label switching"
  - Gestion des labels
  - Protocole de distribution
  - MPLS, ATM et les autres réseaux

Historique

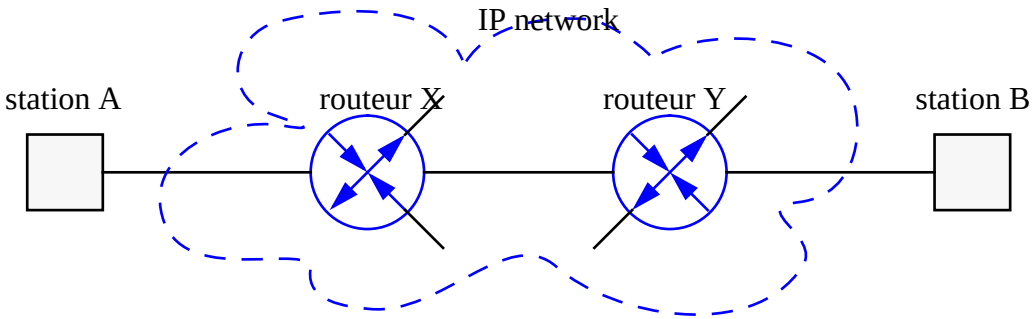
- 95 : Cell switching - Toshiba
- 96 : IP switching - Ipsilon ➟ Nokia
- 96 : Tag switching - Cisco
- 96 : ARIS - IBM

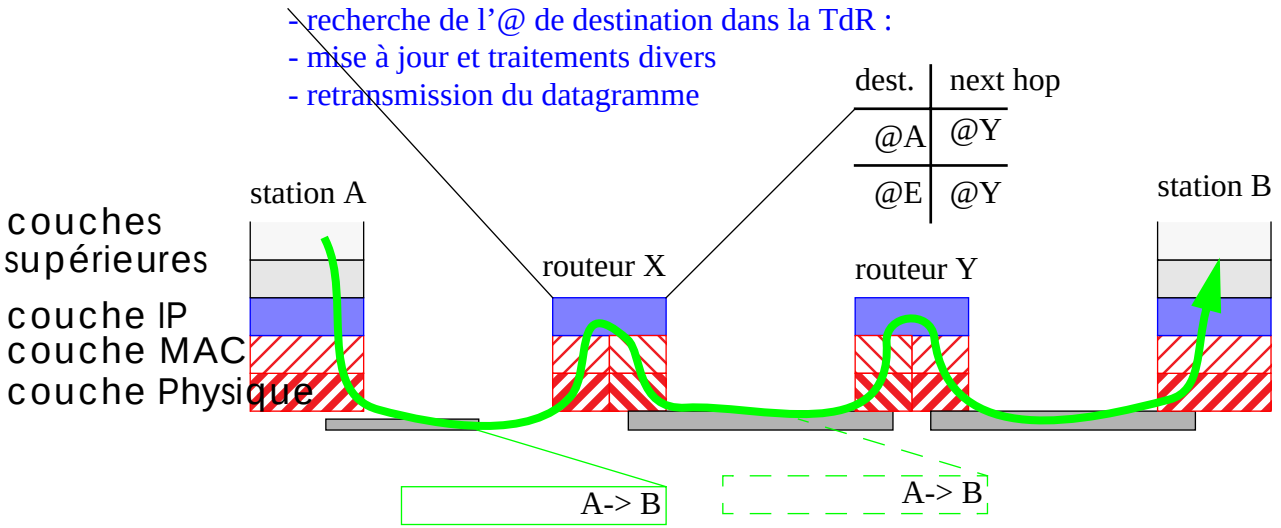
☞ Label switching

- MPLS :
  - 97 : IETF working group
  - "MultiProtocol Label Switching"
  - fusion des techniques précédentes
- Multi-protocole :
  - n'importe quelle infrastructure sous-jacente
  - n'importe quel protocole supérieur

L'acheminement sous Internet

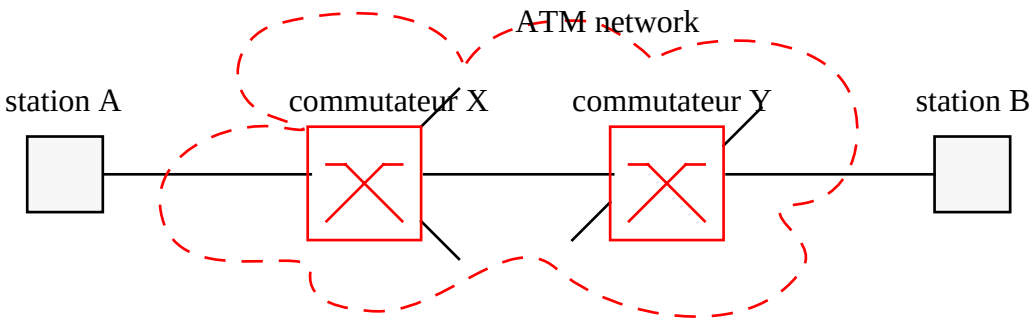
"IP forwarding"

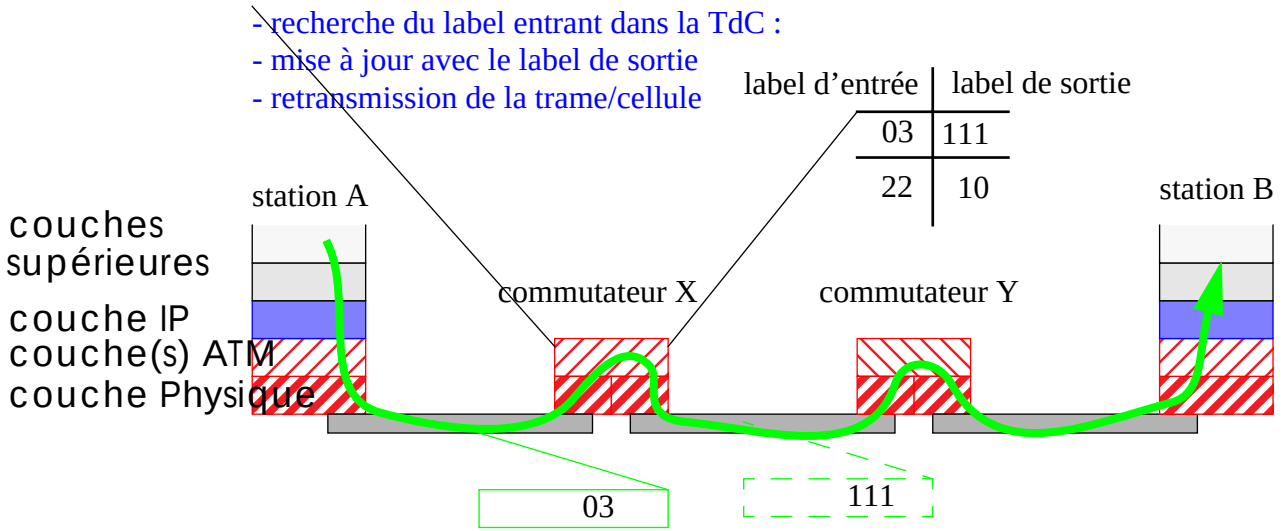




Switching

La commutation





Comparaison

- **Niveau Liaison de données/niveau Réseau**
  - **Le switch est multi-protocole !**
- **Adresse/label**
- **Sémantique :**
  - **les adresses ont une sémantique globale au réseau/**
  - **les labels ont une sémantique locale au commutateur**
- **Longueur des unités de données :**
  - **longueur variable des paquets/**
  - **longueur fixe et petite des cellules**

☞ traitement plus complexe !

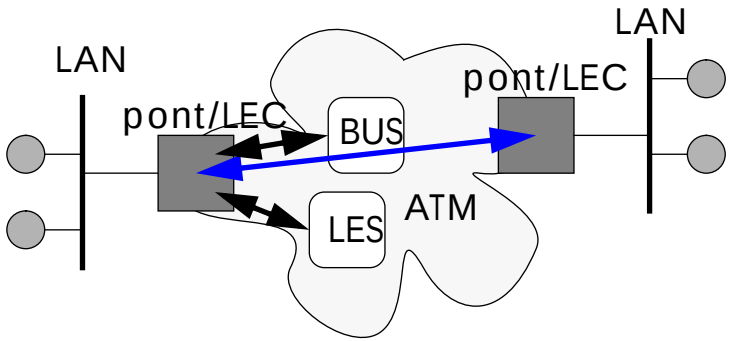
- **Traitement :**
  - **- l'échange ("swapping") des labels est systématique/**
  - **- le traitement des adresses varie :**
    unicast/multicast;
    "netid"/"subnetid"/"longest prefix match";
    ☞ traitement plus complexe !

☞ **combiner le meilleur des 2 techniques**

flexibilité + performance

IP sur ATM

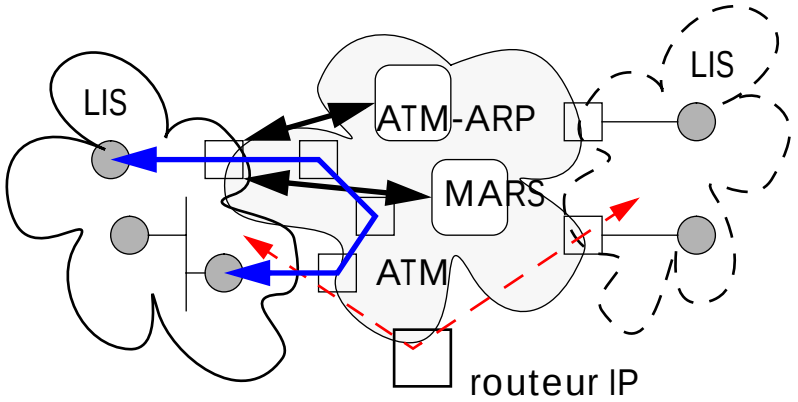
- **LANE :**
  - **interconnexion de LAN : pont**
  - **"tunnelling" à travers ATM : AAL5/trame**



  - **"LAN emulation client/server"**
  - **BUS : "Broadcast/unknown server"**
  - **indépendance vis-à-vis des protocoles de niveau supérieur donc d'IP**

☞ ATM/LAN/IP

- **CLIP :**
  - **Classical IP : IP directement sur ATM (AAL5/IP)**



  - **LIS : "Logical IP subnet"**
  - **MARS : "Multicast address resolution server"**
  - **ATMARP : résolution d'adresse IP/ATM**
- **NHRP :**
  - **"Next hop resolution protocol"**
  - **interconnexion de stations ATM situées sur des LIS différents interconnectables directement sans utiliser de routeurs externes**

Label switching

- **FEC**
  - **- "Forward Equivalence Class"**
  - **- l'ensemble des paquets pour lesquels on prend la même décision de routage ("next hop")**
  - **- Par exemple sous IP :**

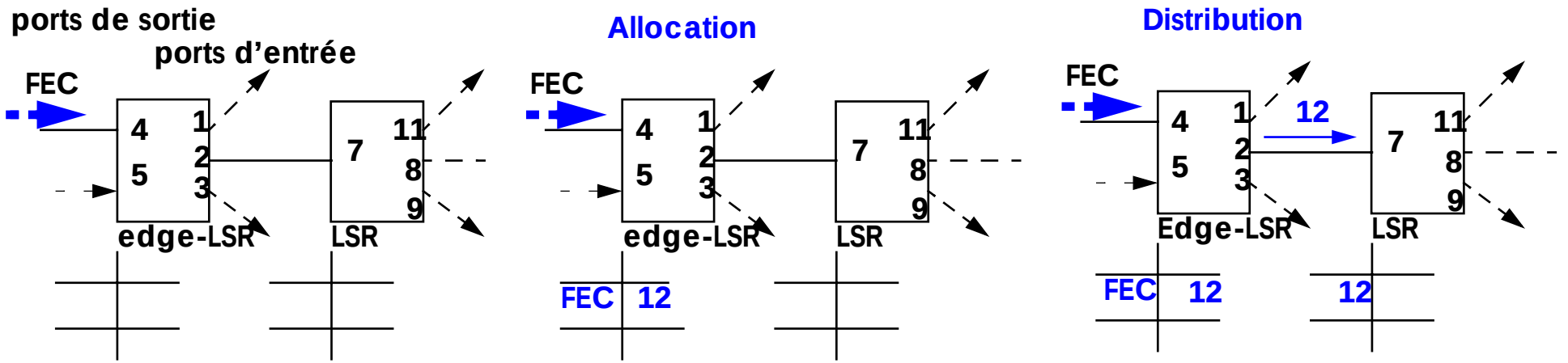
    les paquets ayant le même préfixe d'adresse de destination
  - **- MPLS permet d'associer une FEC à un paquet pour toute sa traversée du réseau en lui associant une valeur courte et fixe : ☞ le label**

    accélère la prise de décision à chaque noeud lors du routage

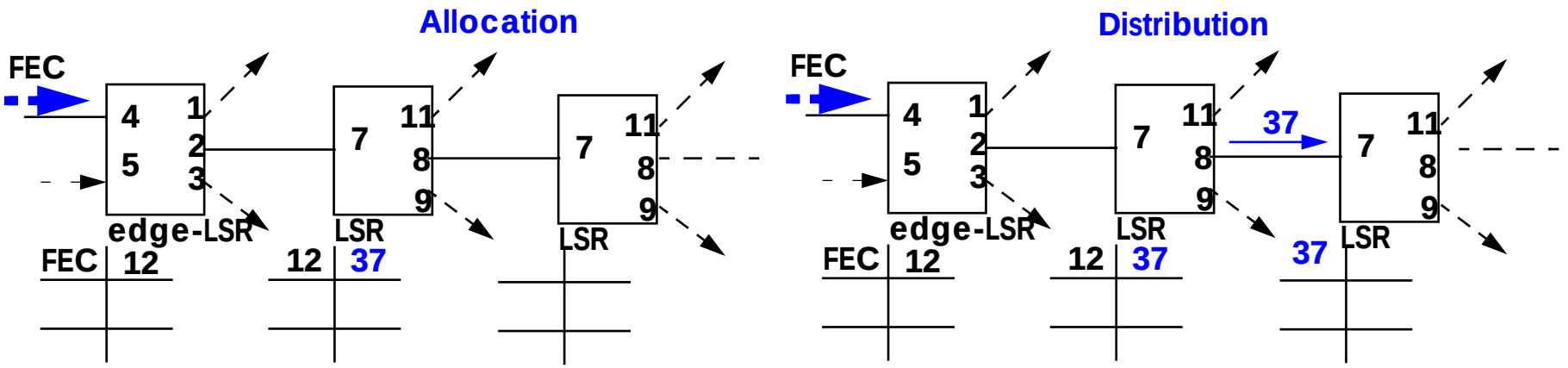
- l'association d'une FEC à un paquet peut être basée sur des informations multiples
  - informations externes au paquet (QoS : RSVP, LDP, etc)
  - informations profondes (type de protocole, numéro de port, etc)
- LSR
  - "Label switching router"
  - un routeur qui offre les fonctions MPLS

Gestion des labels

- **Allocation**
  - Allocation d'un label à un FEC
  - L'association est locale à une liaison, et temporaire
- **Association ("binding")**
  - Association entre un label d'entrée et un label de sortie
  - autour d'une liaison
  - Une association nécessite une allocation et une distribution



- **Distribution**
  - Un des LSR doit communiquer à l'autre l'allocation qu'il a réalisée
  - c'est le rôle du protocole de distribution des labels
- **Propagation**
  - Itération sur les LSR suivants constituant le LSP



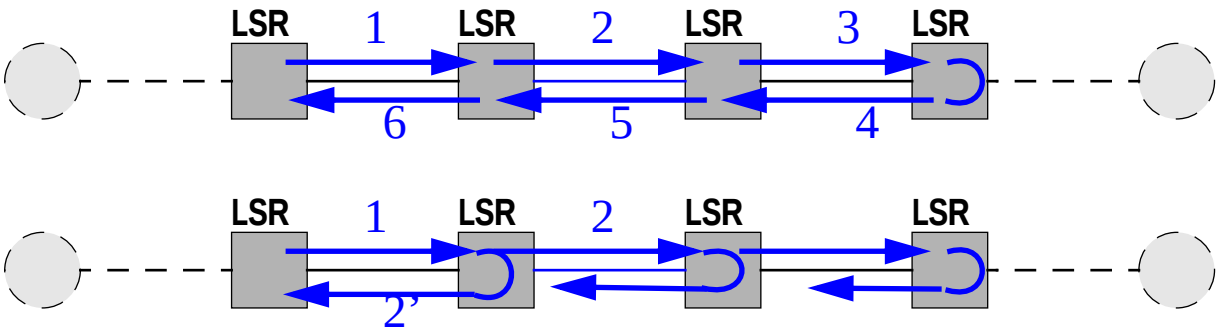
Labels et ports

- **Labels et ports**
  - **un commutateur possède plusieurs ports de sortie**
  - **label d'entrée -> label de sortie + port de sortie**
- **Allocation des labels par port/par commutateur**
  - **un commutateur possède plusieurs ports d'entrée**
  - **(label + port) d'entrée -> (label + port) de sortie**
  - **augmente l'espace des labels disponibles**
  - **exemple : ensemble complet de VPI+VCI par liaison ATM**

    un commutateur peut allouer plusieurs fois le même VPCI sur des liaisons différentes.

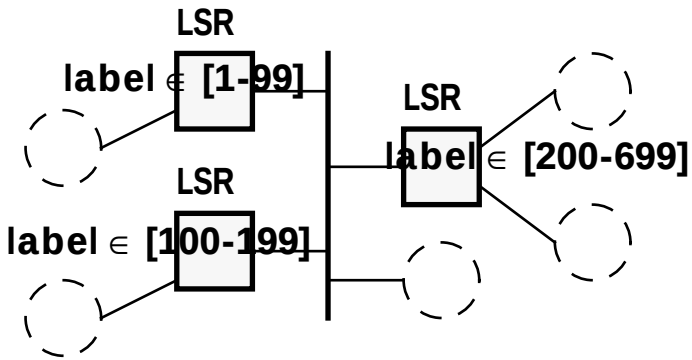
Enchaînement des associations

- La suite d'associations constituant un chemin (LSP) peut être construite de manière ordonnée (en série) ou indépendamment (en parallèle)
  - augmentation du délai de mise en oeuvre
  - minimise les interdépendances
  - simplifie la gestion



Réservation

- **Attribution d'un ensemble (un intervalle) de labels à un LSR**
  - **partition des labels entre les différents LSR**
  - **évite les collisions d'allocation d'un même label sur les liaisons multipoints (LAN)**
    --- inutile sur les liaisons bipoints
  - **Les labels réservés à un LSR vont être alloués par celui-ci**
  - **le partitionnement est mis en oeuvre par un protocole**



Désallocation

- **Problème de détection de l'inactivité de la FEC**
- **Cause de l'inactivité :**
  - **modification de routage**
  - **inactivité des émetteurs**
- **Explicitement /implicitement**
  - **surveillance et temporisateur :**
    test de l'activité, test de la connexité
  - **notification d'évènements :**
    protocole de routage, protocole de distribution

Prise de décision

- La décision d'allouer des labels peut être :
  - pilotée par les données ("data-driven")
  - contrôlée de manière externe au moyen d'un protocole
- Pilotée par les données,
  - ⇨ réception d'un paquet de données
  - traitement "normal" par défaut : nécessite la présence obligatoire du composant traditionnel de routage
  - traitements plus fréquents
- Par contrôle explicite
  - ⇨ réception d'un message de routage (OSPF, PIM) ou de gestion des ressources (RSVP)
  - simplicité : le composant de contrôle est géré uniquement à partir d'informations de contrôle
  - contrôle explicite : moins d'approximation,

- contrôle + souple : adaptation, pérennité
- la prise de décision peut dépendre d'événements autres que l'arrivée et le contenu des données
- l'association peut être anticipée

mais

- c'est plus lourd à gérer,
- les labels peuvent être alloués sans que le flux de données soit actif (sur-allocation des labels).

❑ MPLS utilise un procédé de contrôle externe :

☞ protocole de distribution

Distribution des labels

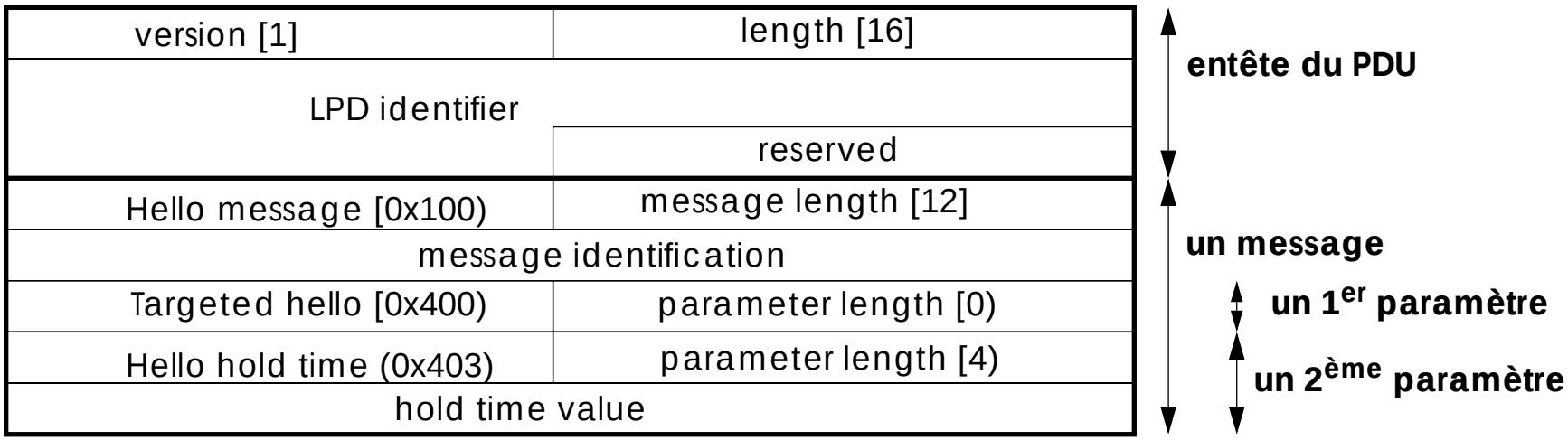
- **La distribution peut se faire en utilisant :**
  - **un protocole spécifique**
  - **un protocole pré-existant (par "piggy-backing")**
- **Utilisation d'un protocole de routage**
  - **synchronisation naturelle**
    chgt du routage/modif des associations de labels
  - **économie**
  - **modification de protocole impossible (pas d'option)**
  - **besoin de transmettre des informations rapidement**
- **Les propositions**
  - **PIMv2 :**
    nouveau format d'adresse "tagée", nouvelle option "tag parameter"
  - **LDP: "Label Distribution Protocol"**

LDP

- **4 fonctions :**
  - **découverte, session, information, notification**
- **Fonction de découverte**
  - **annonce et surveille la présence de LSR dans le réseau**
  - **utilisation d'UDP**
  - **entre LSR adjacents, entre LSR distants**
  - **messages "Hello" or "Targeted hello"**
  - **envoi périodique et contrôle par temporisateur**
- **Fonction de session LDP**
  - **établissement, maintien et libération de la connexion**
  - **utilisation de TCP**
  - **message d'initialisation**

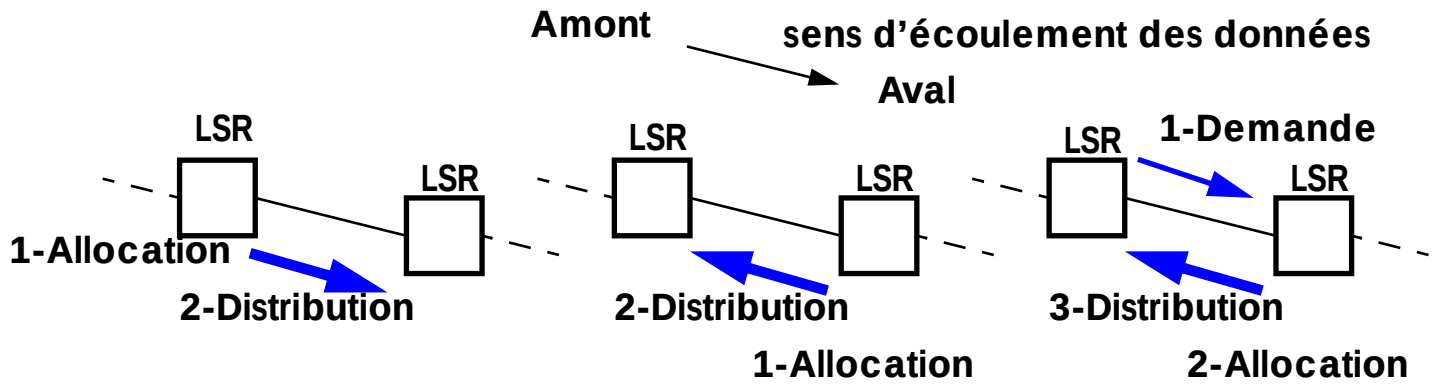
    version, temporisateur, mode de distribution, intervalle de labels réservés, etc.

- **Fonction d'échange d'information sur les labels**
  - **Création, changement ou destruction d'une association**
- **Fonction de notification**
  - **Erreur ou information**
- **Format des LDP PDU**
  - **Plusieurs messages dans un PDU**
  - **Encodage par TLV**
  - **Par exemple :**



Association des labels

- **Origine de l'association :**
  - **Amont ("upstream")**
  - **Aval ("downstream")**
  - **Aval à la demande ("downstream on demand")**

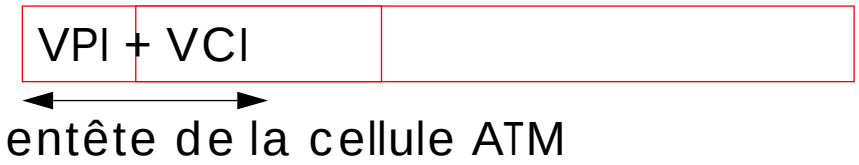


Définit le LSR à l'origine de l'association.
L'amont/l'aval sont définis par rapport au flux de données

Les différents niveaux de labels

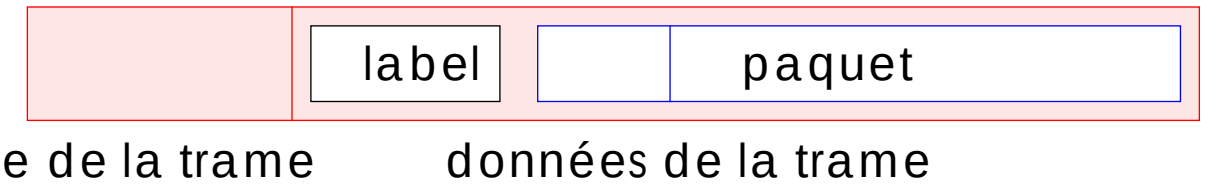
- **On définit 3 niveaux d'utilisation des labels**
  - **Niveau Liaison de données**

    Par exemple : VPCI d'ATM ou DLCI de Frame relay



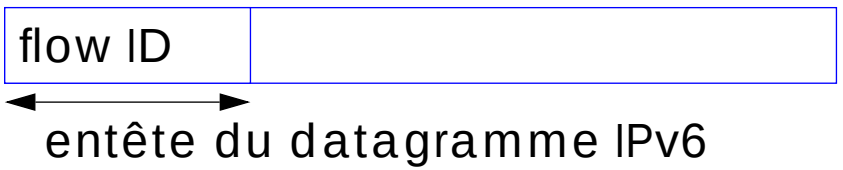
  - **Niveau intermédiaire**

    Par exemple : "Shim label"



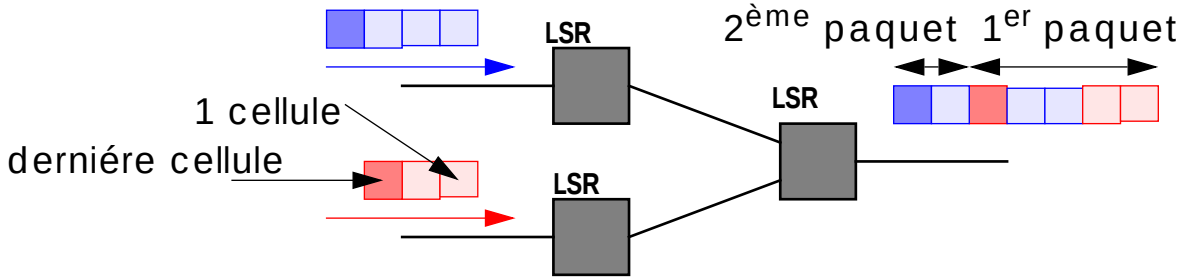
  - **Niveau Réseau**

    Par exemple : identificateur de flot d'IPv6



Agrégation

- Regroupement de plusieurs FEC ("Forwarding equivalence class")
  - Eviter la multiplication des labels au sein des LSR
- Problème d'entrelacement des cellules de paquets différents
  - des cellules successives partageant le même label peuvent appartenir à des paquets différents

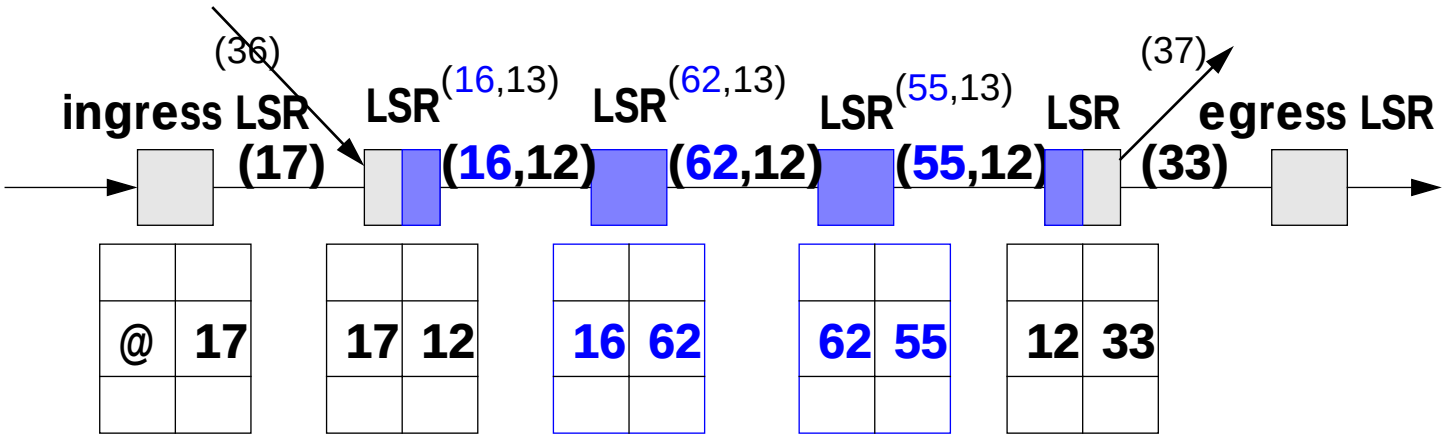


  - stockage et transmission groupée des celulles du même paquet ☞ complexe et couteux
  - autre solution :

☞ l'empilement de labels

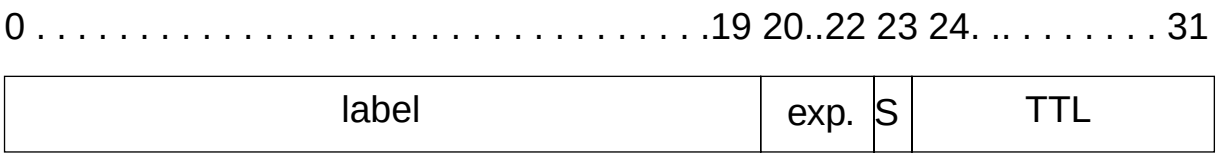
Empilement de labels

- Un paquet peut être muni de plusieurs labels
  - Seul le premier label est traité (haut de pile) à chaque LSR
  - procédé similaire aux commutateurs de VP d'ATM
  - c'est du "Tunnelling"
  - regroupement de plusieurs flux : accélère la commutation



"Shim label"

- **Format d'un "shim label" :**



  - S : indique le bas de pile

- **Le label permet de connaître :**
  - **le prochain routeur**
  - **les opérations à effectuer :**

    remplacement du label, empilement d'un nouveau label ou dépilement.

- **"Time To Live" : durée de résidence résiduelle**

MTU et label

- **Les labels augmentent la longueur totale du paquet**
  - **cela peut provoquer leur fragmentation :**
    traitement complexe
  - **certains paquets à cause de l'interdiction de fragmenter (DF bit) sont détruits alors que sans la labélisation ils auraient pu être transmis**
    message ICMP : "Dest. unreachable/Fragt. required"
- **"MTU discovery"**
  - **ne pas perturber le procédé de découverte du "MTU path"**
- **Franchissement d'un "tunnel"**
  - **les commutateurs intermédiaires ne peuvent pas connaître l'émetteur, donc la notification est impossible**
  - **le site en entrée du tunnel doit connaître le MTU effectif interne au tunnel, pour interdire l'accès aux longs paquets**

Cohérence des entêtes IP

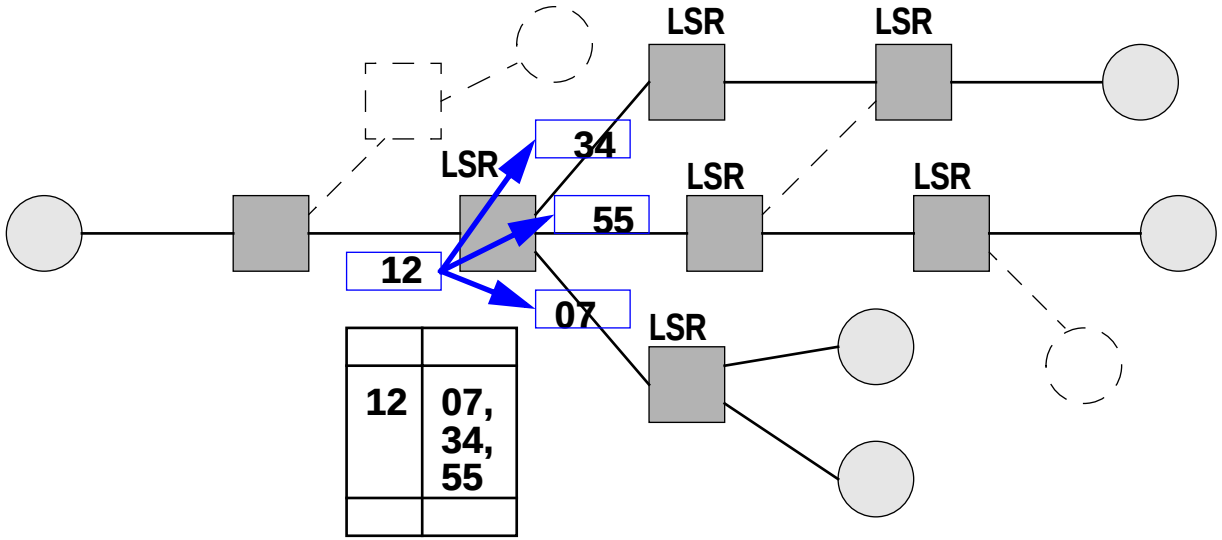
- **Les datagrammes IP possèdent certains champs dont il convient d'assurer la cohérence :**
  - **TTL : décrémenté à chaque "hop"**
  - **conséquence sur le "ckecksum"**
- **Selon la disponibilité des informations :**
  - **Soit les LSR assurent cette cohérence**
    par ex : champ TTL du "Shim header"
  - **Soit seuls, les Edge-LSR sont chargés de cette tâche**
    par ex : pour ATM
- **Les cycles**
  - **les routes cycliques sont néfastes au trafic et au réseau**
  - **a posteriori : le TTL sert à détruire les paquets errants**
  - **a priori : le protocole de distribution des labels peut, en détectant les cycles, refuser d'allouer des labels.**

Optimisation

- Il peut être possible d'omettre certaines informations du paquet si celles-ci sont implicitement représentées par le label.
  - par exemple : l'adresse de l'émetteur et de destination, le type de protocole, les numéros de port, le TTL, etc...
- Dans ce cas, ces information peuvent être retirés des données transmises au sein du réseau de labels.
- On obtient les avantages suivants :
  - la sécurité puisque les informations de l'entête sont tenues secrètes
  - respect du contrat puisque le détournement du LSP, pour transmettre des données non contractuelles, est rendu impossible.
  - l'optimisation du volume de données transmises

Multicast

- L'acheminement des paquets multicasts peut être réalisé par le "label switching"
  - À un label d'entrée on associe plusieurs branches de sortie. Sur chacune de ces branches des labels quelconques sont utilisés



MPLS et ATM

- **ATM-LSR**
  - **Label=VPI+VCI ou au sein d'un "virtual path" label=VCI**
  - **Problème d'entrelacement (n VCI -> 1 VCI)**
  - **au-dessus d'AAL5**
  - **"downstream on demand"**
  - **hétérogène : ATM-LSR/ATM-LSR ou ATM-LSR/frame-based LSR :**

    le chemin suivi par les paquets labellés peut traverser successivement un nombre quelconque de portions de réseaux ATM ou frame-based

    ...de réseaux utilisant le "label switching" ou non
  - **utilisation de LDP**
  - **utilisation des protocoles de routage : OSPF ou IS-IS**
  - **Possibilité d'avoir un commutateur ATM hybride :**

    compatibilité entre les règles de gestion de l'ATM forum/

et celles du label switching : partition de l'espace VPI/VCI.

- **Une connexion ATM entre 2 ATM-LSR :**
  - **VPI=0, VCI=32**
  - **permet d'échanger les paquets LDP**
  - **permet d'échanger les paquets d'autres protocoles (par ex. OSPF)**
  - **utilise l'encapsulation LLC/SNAP définie par le RFC 1483**
- **Empilement des labels**
  - **Afin de permettre l'empilement de labels les paquets transmis au sein d'un domaine d'ATM-LSR peuvent être munis d'un "shim label"**
  - **Le label en haut de la pile est inutilisé, car**
  - **au sein du domaine d'ATM-LSR seul est utilisé le VPI/VCI**

- **Traversée d'un nuage VP-ATM :**
  - **à travers un "virtual path"**
  - **le label est encodé dans le seul VCI**
- **LSR de bordure**
  - **les frame-based LSR connectés à un ATM-LSR.**
  - **lorsqu'ils reçoivent un paquet ils mettent à jour le TTL à partir du "hop count" qui a été obtenu lors de l'établissement de l'association :**

    TTL_de_sortie = TTL_d_entree - hop_count
  - **Si le TTL devient négatif le paquet n'est pas transmis, et un ICMP message est retourné.**

MPLS et PPP

- Un seul paquet labélisé par trame PPP
  - au format "shim header"
- Le code du champ "PPP protocol" :
  - 0281_{16} = paquet MPLS unicast
  - 0213_{16} = paquet MPLS multicast
  - 8281_{16} = paquet du protocole de contrôle de MPLS

MPLS et LAN

- Exactement un paquet labélisé par trame
  - après l'entête du niveau Liaison de données (tous : 802.1Q), et avant l'entête de niveau Réseau
  - au format standard "shim header"
- 2 formats possibles :
  - soit encapsulation directe (ex. : Ethernet)
  - soit encapsulation par LLC/SNAP
- Le code du champ "protocol type" de la trame:
  - 8847_{16} = paquet MPLS unicast
  - 8846_{16} = paquet MPLS multicast

Références

- **"IETF working group" :**
  - **routing area - MPLS working group**
  - **"On line" :**
    http://www.ietf.org/html.charters/mpls-charter.html
  - **prochaines conférences IETF :**
    43ème conférence à Orlando, Floride, USA, 7-11 Décembre 1998.
  - **proceedings :**
    http://www.ietf.org/proceedings/directory.html
- **Livres :**
  - **B.Davie, P.Doolan, Y Rekhter. Switching in IP Networks. Morgan Kaufmann. 1998.**

Documents techniques sur le "label switching":

E.Rosen & al.. Multiprotocol Label Switching Architecture. IETF MPLS working group. July 1998.

L.Anderson & al.. LDP Specification. IETF MPLS working group. August 1998.

E.Rosen & al.. MPLS Label Stack Encoding. IETF MPLS working group. September 1998.

B.Davie & al.. Use of Label Switching with ATM. IETF MPLS working group. September 1998.

A.Conta & al.. Use of Label Switching on Frame Relay Networks. IETF MPLS working group. August 1997.

B.Davie & al.. Use of Label Switching with RSVP. IETF MPLS working group. March 1998.

- **Documents historiques :**

  R. Woundy & al.. ARIS : Agregate Route-based IP Switching. IETF MPLS working group. November 1996.

  P.Newman & al.. Ipsilon's General Switch Management Protocol Specification, version 1.1. RFC 1987. August 1996.

- **Liste de discussion ("mailing list"):**

  subscribe to mpls-request@external.cisco.com