

La commutation de labels

par Bernard Cousin

Laboratoire IRISA

Université de Rennes-1

bcousin@irisa.fr

<http://www.irisa.fr/prive/bcousin>

- Plan

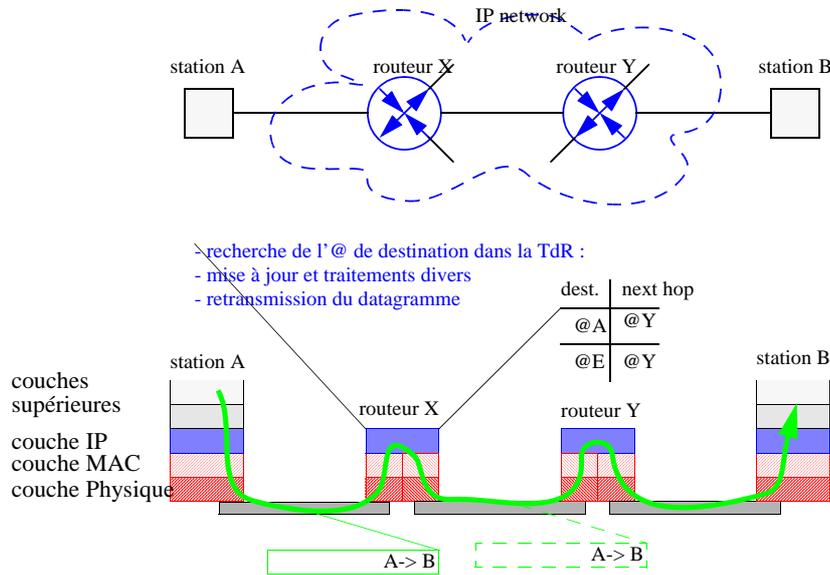
- Présentation de la commutation de labels
- Les autres techniques
- Gestion des labels
- Protocole de distribution : LPD
- MPLS, ATM et les autres réseaux

Présentation

- Acheminement sous Internet
- La commutation
- Comparaison

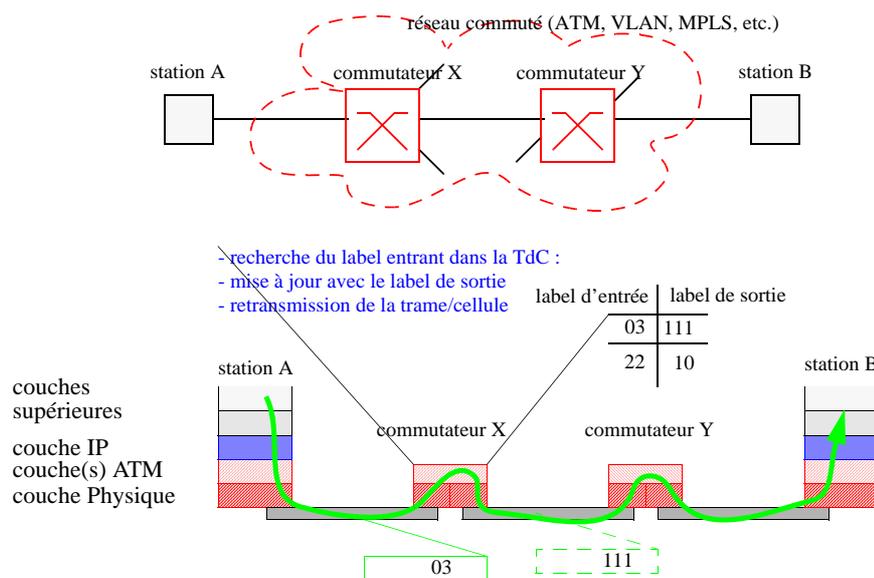
L'acheminement sous Internet

- "IP forwarding"



Switching

- La commutation



Comparaison

- Niveau Liaison de données/niveau Réseau
 - le "switch" est multi-protocole (idem LANE) !
- Adresse/label
 - longue/courte
 - sémantique :
 - . les adresses ont une sémantique globale au réseau
 - . les labels ont une sémantique locale au lien
- Longueur des unités de données :
 - longueur variable des paquets
 - longueur fixe et petite des cellules (ATM seulement)
 - => traitement plus rapide !
- Flexibilité
 - l'adresse de destination (+ToS) détermine un seul chemin
 - . tous les paquets suivront la même route
 - à chaque LSP (label) est associé un flux (FEC). Chaque LSP peut suivre un chemin indépendamment des autres LSP, même ceux ayant même source et même destination

- Traitement :
 - l'échange ("swapping") des labels est systématique
 - le traitement des adresses varie :
 - unicast/multicast;
 - netid/subnetid/"longest prefix match";
 - avec/sans options;
 - => traitement plus complexe !
- "IP forwarding" versus "switching"
 - plan de commande => IP !
 - plan de données => "datagram forwarding" ou "label switching"

combiner le meilleur des 2 techniques :
=> flexibilité + performance

Les applications de MPLS

- Une infrastructure d'interconnexion indépendante
- IP over MPLS
- Ingénierie de trafic
 - équilibrage de charge
 - sélection des chemins en fonction des besoins exprimés (QoS)
- "Fast ReRoute"
 - protection et rétablissement d'une route alternative
- "Layer 3 virtual private network"
 - séparation des trafics des différents utilisateurs
 - sécurisation des flux
- "Layer 2 virtual private network" (ATOM : "Any Transport over MPLS")
 - idem L3VPN
 - transport de diverses trames de niveau 2
 - . par ex. : FR, Ethernet, ATM, PPP, etc.

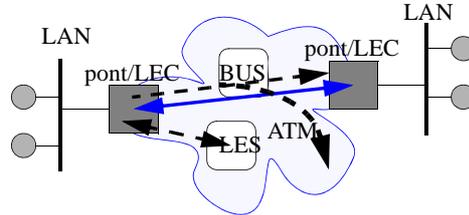
Les autres techniques

- IP sur ATM
 - LANE
 - CLIP
 - NHRP
- Apparition de MPLS

IP sur ATM

• LANE :

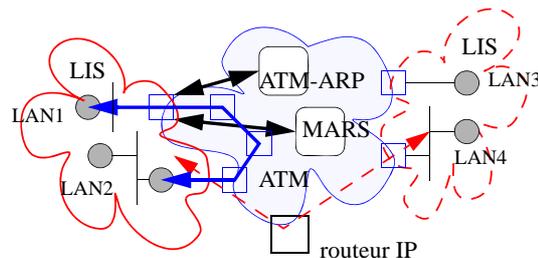
- interconnexion de LAN à l'aide d'un réseau de transit
- "tunnelling" à travers ATM : encapsulation frame/AAL5



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

• CLIP :

- "Classical IP" : IP directement sur ATM (AAL5/IP)



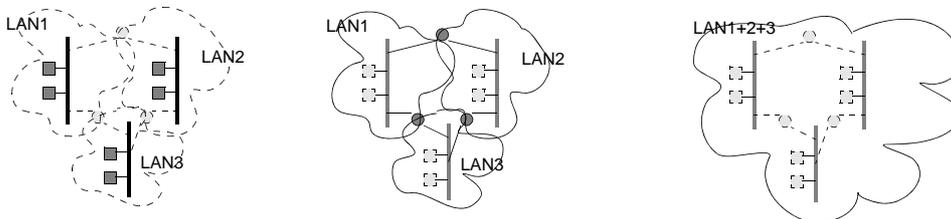
- LIS : "Logical IP subnet"
- MARS : "Multicast address resolution server"
- ATMARP : résolution d'adresse IP/ATM

• NHRP :

- "Next hop resolution protocol"
- interconnexion de stations ATM situées sur des LIS différents interconnectables directement sans utiliser de routeurs externes

VLAN/Interconnexion de LAN

- Interconnexion de LAN
 - augmentation de l'étendue
 - augmentation du débit si le trafic est localisé
 - flexibilité de la topologie
- Utilise des équipements d'interconnexion
 - ponts/ "bridges"/"switches"
- Méthode d'interconnexion
 - "Transparent Bridging" + "Spanning Tree"
 - . pas de configuration
 - pas de différence lors d'une communication si toutes les stations sont connectées à un seul LAN ou à un ensemble de LAN interconnectés
 - . les trames sont inchangées, les stations sont inchangées



VLAN/Interconnexion de LAN

par Bernard Cousin

VLAN

- "Virtual LAN"
 - on définit des réseaux virtuels, auxquels appartiennent un sous-ensemble des stations
 - une station ne communique qu'avec les stations du même réseau virtuel
 - . chaque réseau virtuel est identifié par un VID
 - . chaque trame est munie du VID
 - par interface /par station/par application
 - . une station connectée à une certaine interface d'un commutateur appartient à un VLAN donné
 - . une station appartient à un seul VLAN
 - . une station peut appartenir à plusieurs VLAN, en fonction de l'application concernée

Historique

- 95 : Cell switching - Toshiba
- 96 : IP switching - Ipsilon et 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 (ATM, FR, ...)
 - . n'importe quel protocole supérieur (IPv4/v6, IPX, Appletalk)

Label switching

- Définitions
 - FEC, label, LSR, LSP
- La gestion des labels
 - l'allocation, association, distribution
 - label et port
 - associations ordonnées/indépendantes
 - réservation
 - désallocation des labels
 - prise de décision
 - mode de distribution

Définitions

• FEC

- "Forward Equivalence Class"
- L'ensemble des paquets qui subissent
 - . la même décision de routage ("next hop")
 - . le même traitement (par ex., même file d'attente)
- Par exemple, sous IP les paquets ayant
 - . le même préfixe d'adresse de destination
 - . la même adresse de destination
 - . idem + même numéro de port (source ou/et dest.)

• La création d'une FEC peut être basée sur :

- le type de flux (QoS : RSVP, etc)
- informations de routage
 - . par ex., tous les paquets convergeant vers le même point

• Label

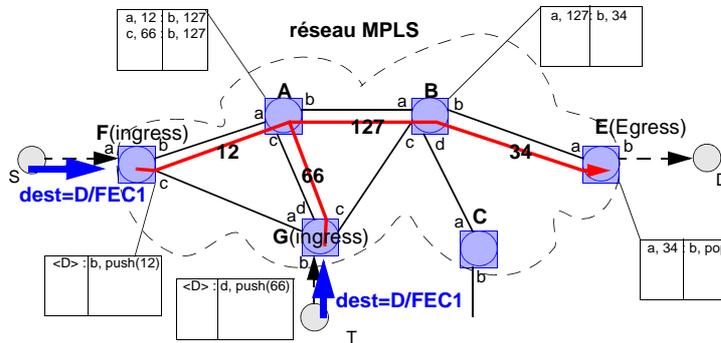
- MPLS permet d'associer une FEC à un paquet pour toute sa traversée du réseau.
- Cette association est déterminée à l'ingress LSR en fonction :
 - . des informations contenues par le paquet
 - . de la politique de gestion du domaine
- **Pour chaque lien** cette association est identifiée par une valeur courte et fixe :
=> **le label**
 - . accélère la prise de décision à chaque noeud lors du routage/de la commutation

• LSR

- "Label switching router"
- un routeur qui offre les fonctions MPLS
- routeur "ingress"/"egress" pour un LSR donné

• LSP

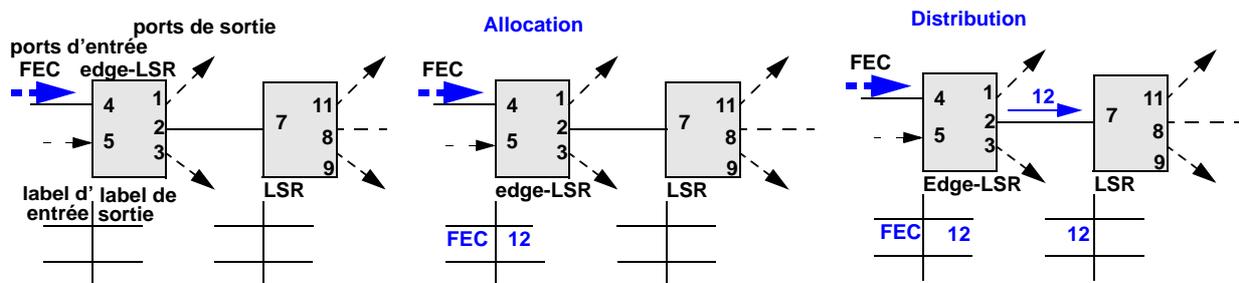
- "Label switch path" :
 - . une connexion MPLS



- défini pour une FEC
 - . un chemin dans le réseau MPLS
 - . une suite de labels
 - . un arbre (racine = 1 egress, feuilles = des ingress)
- optimisation :
 - . l'avant-dernier routeur fait le "pop" : le dernier étiquetage est inutile

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

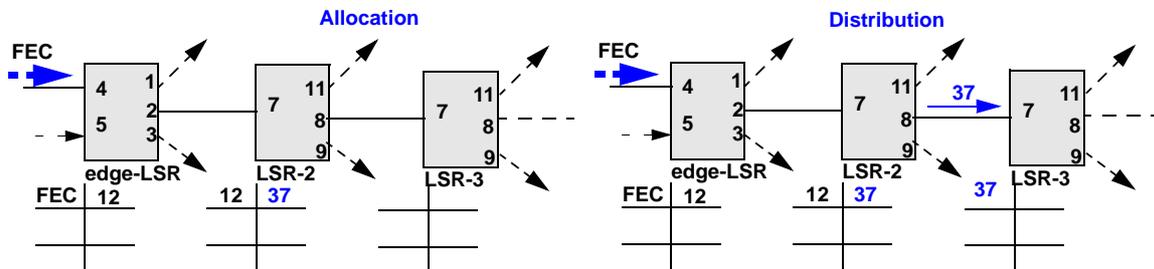


- Distribution

- 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

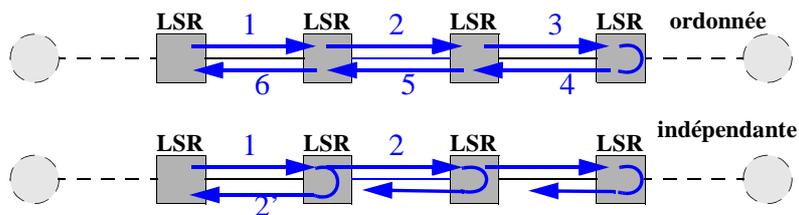
- un commutateur possède plusieurs ports de sortie
- label d'entrée -> label de sortie + port de sortie

- 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.

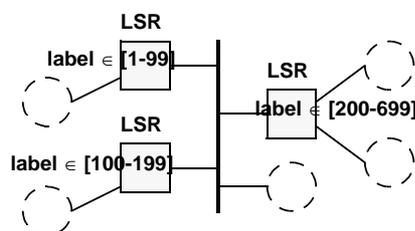
Enchaînement des associations

- La suite d'associations constituant un chemin (LSP) peut être construite de manière **ordonnée** (en série) ou **indépendamment** (en parallèle)
- L'enchaînement ordonné :
 - 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



Libération des LSP

- 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
 - => **"soft state" !**
 - notification d'évènements :
 - . protocole de routage, protocole de distribution

Etablissement des LSP

- La décision d'allouer des labels peut être :
 - déterminée par la présence des données contrôlée de manière externe au moyen d'un protocole
- Déterminée par les données ("**data-driven**")
 - lors de la réception d'un premier paquet de données
 - traitement "normal" par défaut : nécessite la présence obligatoire du composant traditionnel de routage
 - traitements plus fréquents
- Par contrôle explicite ("**control driven**")
 - lors de la 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 ($N^2/2$ LSP à établir au maximum),
 - 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
 - . chg^t 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

- Les fonctions de LDP
- Le format des messages
- Agrégation
- MTU
- Optimisation
- Cohérence
- Multicast

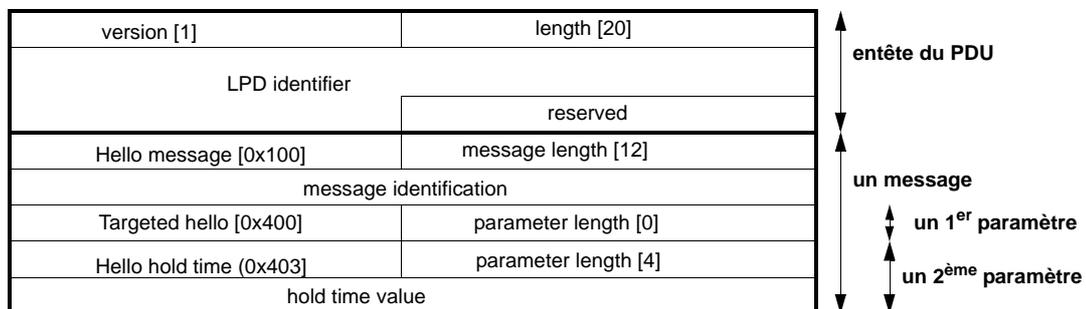
Fonctions de LDP

- 4 fonctions :
 - découverte, session, information, notification
- Fonction de découverte
 - annonce et surveille la présence de LSR dans le réseau
 - utilisation d'UDP ("LDP discovery port")
 - 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 ("LDP port")
 - 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

Formats

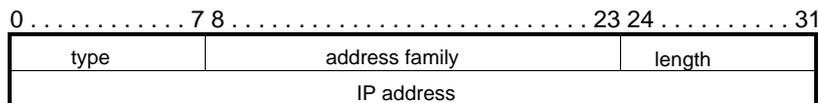
- Format des LDP PDU
 - plusieurs messages dans un PDU
 - encodage par TLV
 - . par exemple :



- Les messages de distribution de labels :

- Label_Request (FEC z)
 - . le LSR aval demande un label pour la FEC z
- Label_Mapping (FEC z, Label l)
 - . le LSR amont associe le label l à la FEC z
- Label_Withdraw(FEC z, Label l or *)
 - . le LSR aval informe que ce ou tout ses labels ne sont plus valides
- Label_Release(FEC z, Label l)
 - . le LSR amont n'utilise plus ce label

- Format de description des FEC

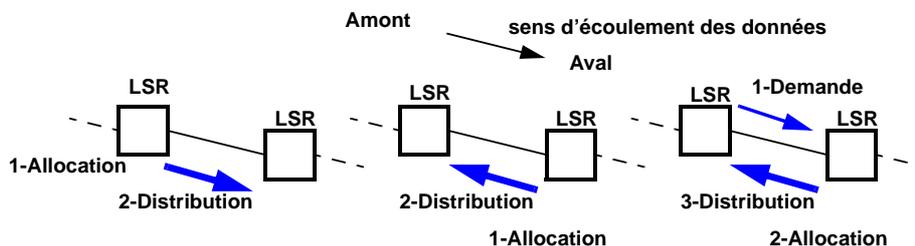


type code : 0x02, 0x03

Mode de distribution 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

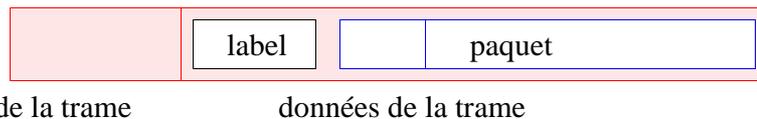
- niveau Liaison de données

- . par exemple : VPCI d'ATM ou DLCI de Frame relay



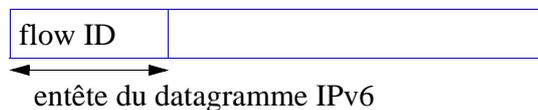
- niveau intermédiaire

- . par exemple : "Shim label"



- niveau Réseau

- . par exemple : identificateur de flot d'IPv6



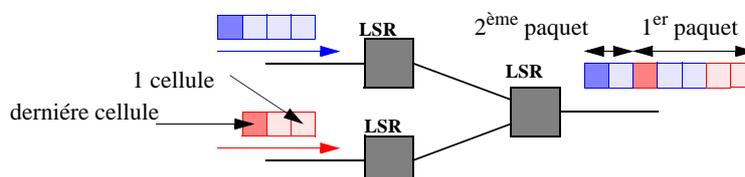
Agrégation

- Regroupement de plusieurs FEC ("Forwarding equivalence class")

- éviter 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



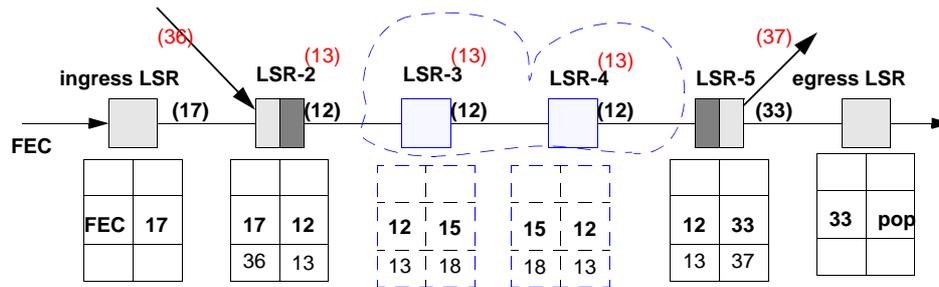
- stockage et transmission groupée des cellules du même paquet => complexe et coûteux

- autre solution :

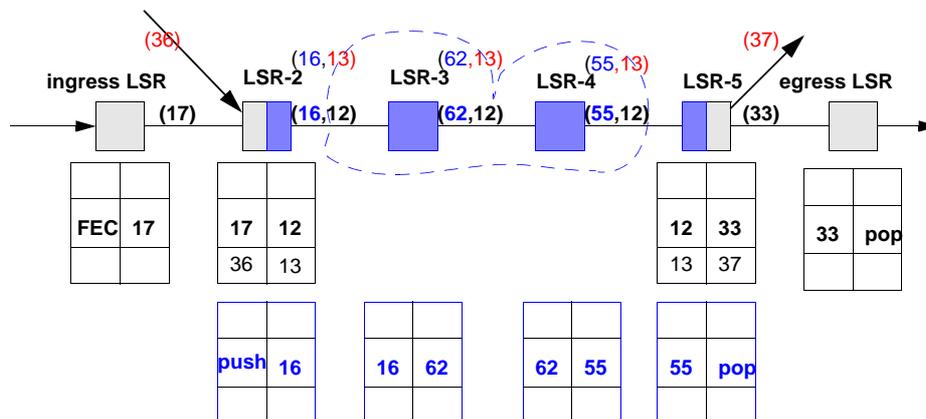
=> l'empilement de labels

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
- Exemple sans empilement :



- Exemple avec empilement



Cohérence des entêtes IP

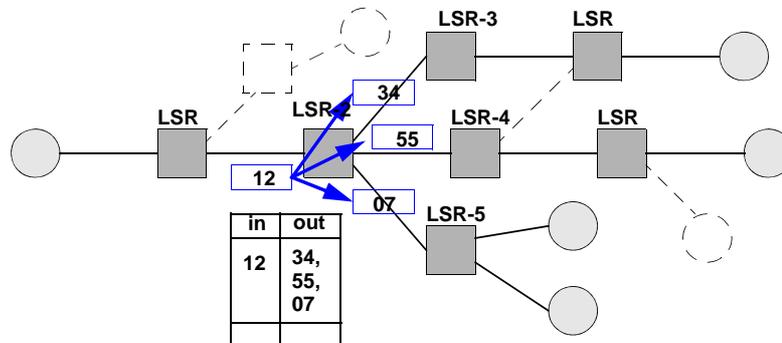
- Les datagrammes IP possèdent certains champs dont il convient d'assurer la cohérence :
 - TTL : décrémenté à chaque "hop"
 - conséquence sur le "checksum"
- Selon la disponibilité des informations :
 - soit les LSR assurent cette cohérence
par ex : champ TTL du "Shim header"
 - soit seuls, les Edge-LSR 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 informations peuvent être retirées 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 :
$$\text{TTL_de_sortie} = \text{TTL_d_entree} - \text{hop_count}$$
 - 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
 - 0283_{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
 - 8848_{16} = paquet MPLS multicast

Références

- "IETF working group" :
 - routing area - MPLS working group
 - "On line" :
<http://www.ietf.org/html.charters/mpls-charter.html>
 - conférence IETF :
43^{ème} 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