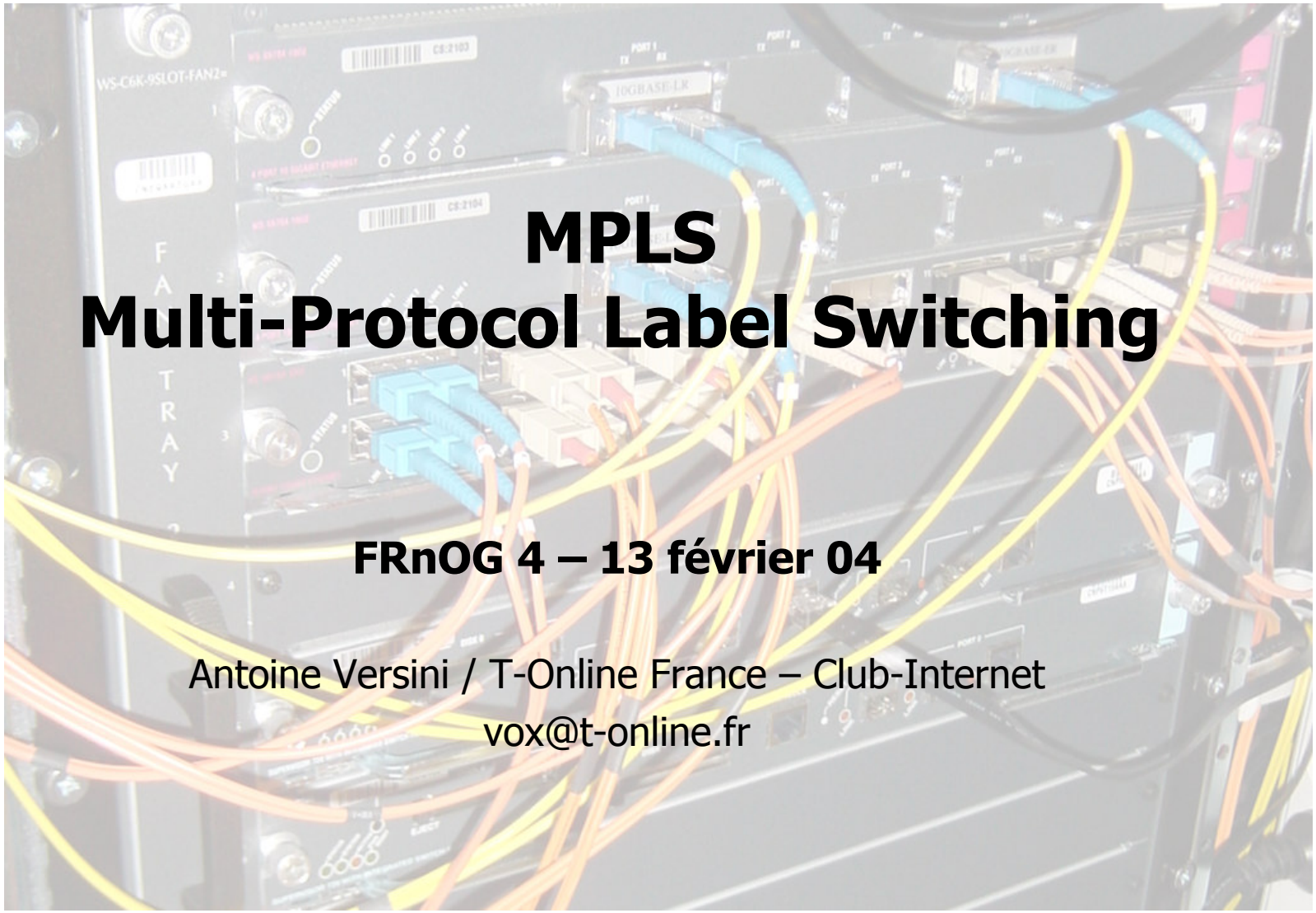


# MPLS
# Multi-Protocol Label Switching

**FRnOG 4 – 13 février 04**

Antoine Versini / T-Online France – Club-Internet
vox@t-online.fr

# Agenda

- MPLS : Concepts et principes de fonctionnement :
  - Genèse de MPLS : du "layered model" au réseau IP intelligent,
  - Principes de commutation de labels,
  - Protocoles mis en jeu.
- Fonctionnalités apportées par MPLS :
  - Décision de routage unique,
  - MPLS VPN : Les VRF,
  - AToM,
  - Traffic Engineering,
  - MPLS et IPv6 : 6PE.
- Feedback opérationnel :
  - Transition du réseau IP vers le réseau MPLS/IP,
  - Transition vers le Traffic Engineering.
- Questions / Réponses.

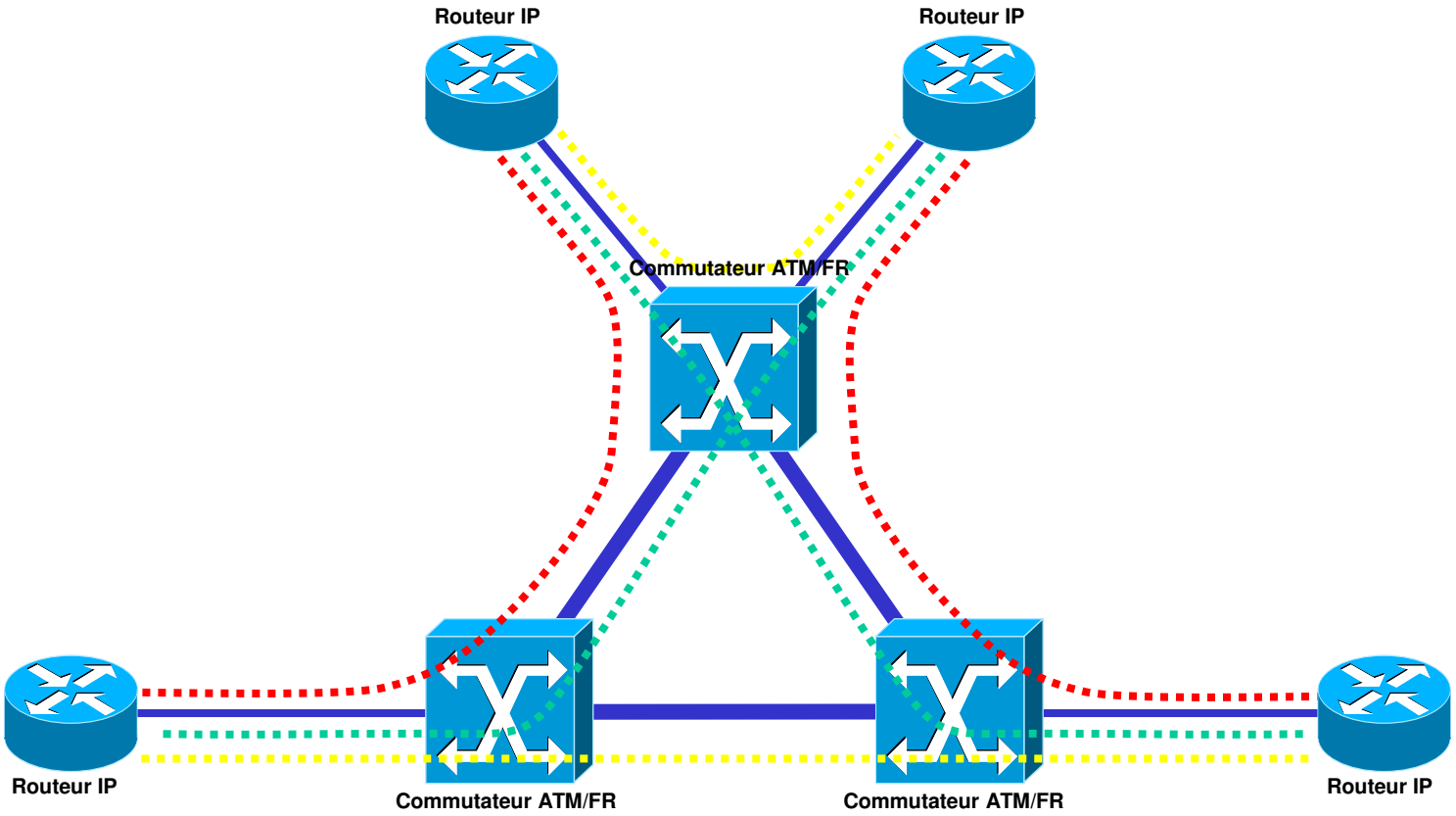
# MPLS : Concepts et principes de fonctionnement

- Genèse de MPLS - Historique des réseaux IP : le « layered model »

  - Le réseau IP repose sur une infrastructure de transport assurant la fonction « Couche de liaison » du modèle OSI (niveau 2), elle-même tributaire d'un réseau de télécommunication.
  - Tous les routeurs IP sont en bordure de réseau et assurent les interconnexions avec le monde extérieur et les réseaux locaux.
  - Les routeurs sont maillés entre eux par l'intermédiaire de circuits virtuels établis sur le réseau de commutation de niveau 2.
  - Le réseau de transport assure la redondance par re-routage des circuits virtuels établis entre les routeurs en cas de perte d'un lien physique sur le réseau de télécommunication.
  - Le réseau de transport assure la QoS (prioritisation des flux « temps réel », limitation de bande passante, allocation dynamique des besoins en bande passante avec gestion des débordements).
  - Les protocoles de routage internes et externes du réseau sont *de facto* maillés (iBGP full-mesh, adjacence IGP avec tous les autres routeurs).

# MPLS : Concepts et principes de fonctionnement

- Genèse de MPLS - Historique des réseaux IP : le « layered model »



# MPLS : Concepts et principes de fonctionnement

- Genèse de MPLS - Historique des réseaux IP : les réseaux purement IP

  - Glissement vers la technologie « Packet over Sonet/SDH » car les débits ATM et FR ne sont plus suffisants, le besoin de séparation des flux par circuit avec limitation de bande passante plus justifiable en cœur de backbone, le coût induit par le réseau de transport trop important et la maintenance & provisionning compliqués.
  - Le réseau IP n'est plus complètement maillé car repose directement sur le réseau de télécommunication en structure de boucles Sonet/SDH interconnectées géographiquement.
  - Nécessité d'introduire des routeurs purement cœur de réseau en lieu et place des commutateurs pour assurer la concentration et la rediffusion des flux vers les différents routeurs de bordure.
  - Les routeurs de cœur de réseau, même s'ils n'accueillent pas de lien de connectivité extérieure, doivent disposer de l'information de routage exhaustive afin que l'ensemble des routes dans le réseau soit cohérent.
  - Maillage complet BGP dans les grands réseau compliqué : utilisation de groupes de réflexion de routes, de confédérations de routage voire même de différents AS continentaux !

- Genèse de MPLS - Historique des réseaux IP : les réseaux purement IP (suite)

  - La décision de routage dépend du point d'entrée dans la table de routage (appelé FEC : Forwarding Equivalence Class) dont le caractère principal est le préfixe de destination du paquet mais peut également être l'adresse source dans le cas d'un partage de charge déterministe par hachage sur le couple adresse source et adresse destination, le tout pouvant être indexé sur le champs ToS du paquet à router.
  - La décision de routage est compliquée. En « layered model », elle n'est effectuée qu'une fois par le routeur de bordure sur lequel le paquet à router s'est présenté. Le routeur de sortie du réseau est alors directement joint par l'intermédiaire d'un circuit virtuel commuté par le réseau de transport. Dans un réseau purement IP, la FEC va être réévaluée à chaque saut : les routeurs de cœur de réseau concentrant tout le trafic vont donc effectuer un travail compliqué plus souvent.

# MPLS : Concepts et principes de fonctionnement

- Genèse de MPLS - Historique des réseaux IP : Création de MPLS

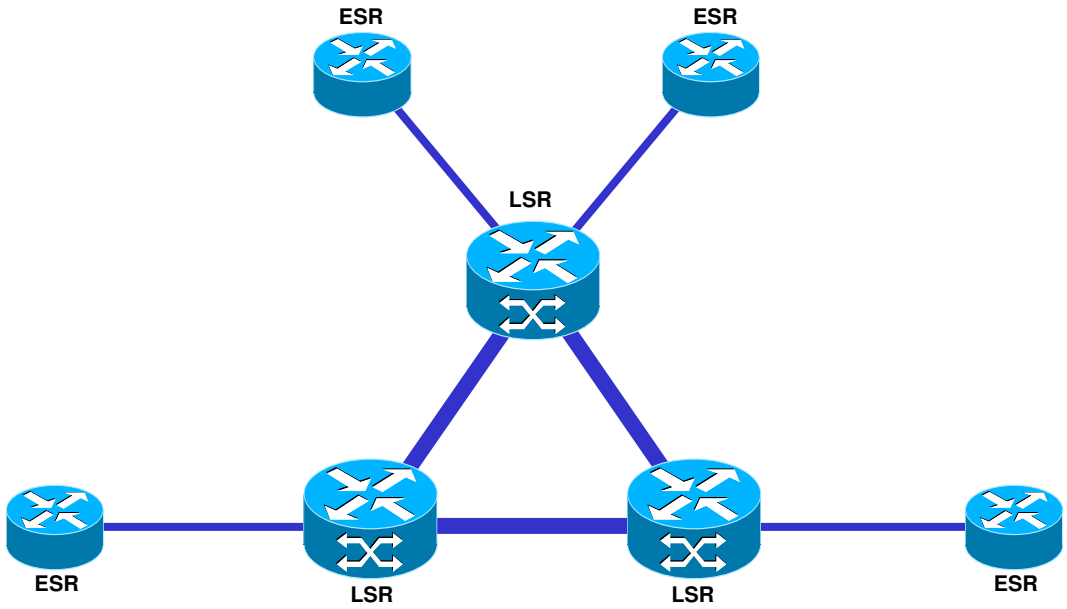
  - « Ce serait bien de pouvoir acheminer les paquets sur un chemin déterminé par le réseau sans devoir réévaluer la FEC à chaque saut » (A. Danthine, Professeur Emérite à l'Université de Liège, un des pères de MPLS).
  - Ceci peut être fait en établissant une connexion logique le long de laquelle les paquets seront routés en utilisant un identificateur unique par FEC.
  - L'identificateur de la FEC est appelé un label.
  - Les paquets qui se présentent en entrée du réseau sont préfixés d'une en-tête indiquant la FEC. Ce paquet n'est pas obligatoirement un paquet IP ! D'où la nature « Multi protocoles » de MPLS.

# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Bindings et LFIB
  - Les correspondances FEC / Label (appelés bindings en terminologie MPLS) sont distribuées de routeurs en routeurs le long du chemin déterminé par l'IGP de chacun d'eux vers le routeur de sortie.
  - Un protocole spécifique sur le routeur MPLS a pour charge d'échanger les bindings avec les routeurs voisins.
  - Chaque routeur crée une table des labels (la LFIB : « Label Forwarding Information Base ») à partir des bindings reçus par tout ses voisins MPLS et de la RIB générée par l'IGP.
  - L'IGP sert au routeur MPLS à déterminer la meilleure route vers le routeur de sortie et donc à sélectionner le meilleur binding pour chaque FEC : l'IGP reste ainsi seul maître du choix de la route ! Deux routeurs ne peuvent s'échanger des bindings s'ils ne se connaissent pas via l'IGP.
  - Une fois la LFIB générée, elle est transformée en FIB et uploadée dans les ASIC de commutation des cartes d'interface des routeurs. Auparavant, c'était la RIB qui était directement convertie en FIB.

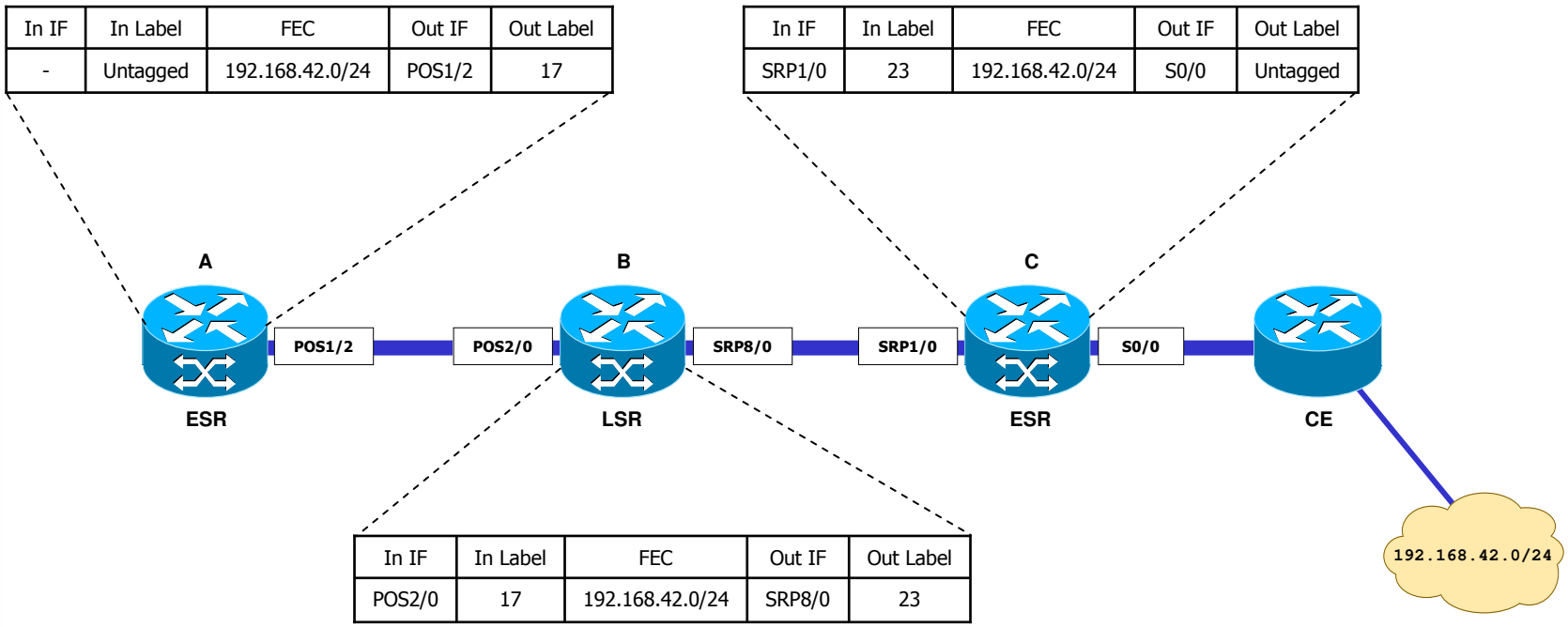
# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Terminologie
  - Un routeur assurant une fonction de commutation de labels en cœur de réseau est appelé LSR (« Label Switch Router ») ou « Routeur P » ( « Provider »).
  - Un routeur assurant une fonction de bordure entre les réseaux non MPLS et le cœur de réseau MPLS est appelé ESR (« Edge Switch Router »), ELSR (« Edge LSR ») ou routeur PE (« Provider Edge »).



# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Distribution des labels
  - Chaque routeur MPLS associe un label choisi localement avec une FEC et annonce ce binding à ses voisins : c'est la distribution des labels.

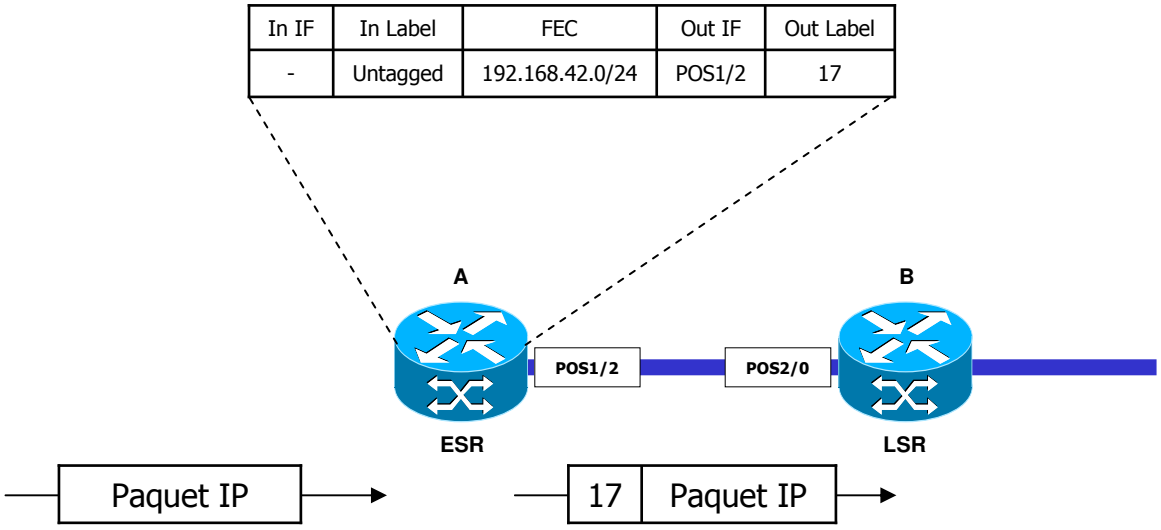


# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Distribution des labels (suite)
  - Le routeur C (Routeur PE de sortie) connaît une route via le routeur D (Routeur « Customer Edge ») pour le réseau **`192.168.42.0/24`** sur une interface non MPLS.
  - Cette route est distribuée normalement dans l'IGP.
  - Le routeur C associe un label avec cette FEC : 23.
  - C distribue à B sur l'interface SRP1/0 ce binding.
  - B installe ce label dans la table des bindings puis inspecte le SPF choisi par l'IGP pour déterminer quel binding utiliser pour insérer la FEC dans la LFIB. En l'occurrence, l'IGP désigne le routeur C comme nexthop préféré donc B choisi le label 23 sur l'interface SRP8/0 pour acheminer tout paquet dans cette FEC.
  - B associe son propre label avec cette FEC : 17.
  - B distribue à A sur l'interface POS2/0 ce binding.
  - A procède à l'installation du binding dans la LFIB après avoir déterminé avec l'IGP que B est la meilleure route.

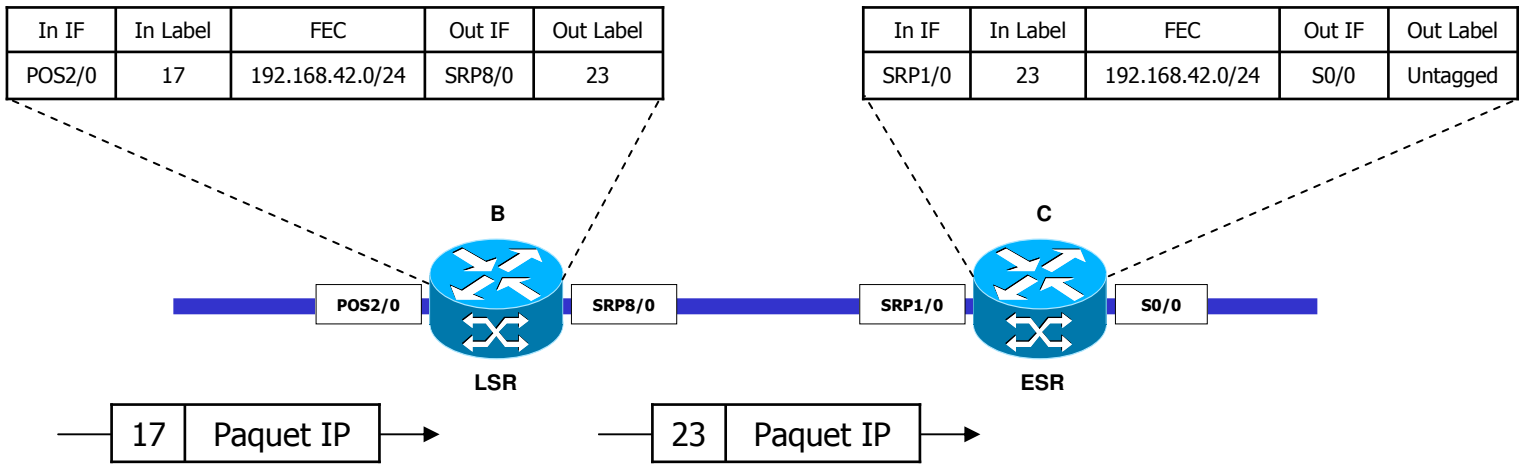
# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Acheminement du paquet
  - Un paquet IP avec comme destination la FEC **`192.168.42.0/24`** se présente en entrée du réseau sur le routeur PE « A ».
  - A procède à l'évaluation de la destination du paquet IP. La FIB générée à partir de la RIB et de la LFIB lui indique de procéder à l'imposition d'un label associé avec la FEC puis d'expédier le paquet IP labellisé sur l'interface POS1/2.



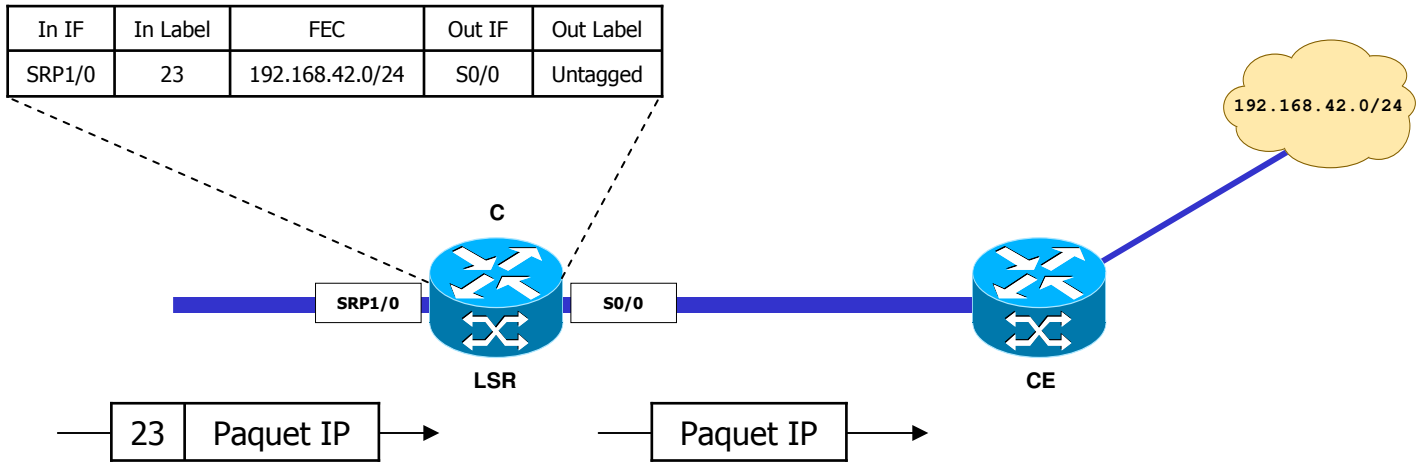
# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Acheminement du paquet (suite)
  - Le paquet ainsi labellisé se présente en entrée sur le LSR « B », interface POS2/0.
  - « B » n'évalue pas la FEC mais commute immédiatement le paquet vers l'interface de sortie avec le label correspondant au binding du routeur de downstream pour la FEC égale à celle désignée par le label 17 local.



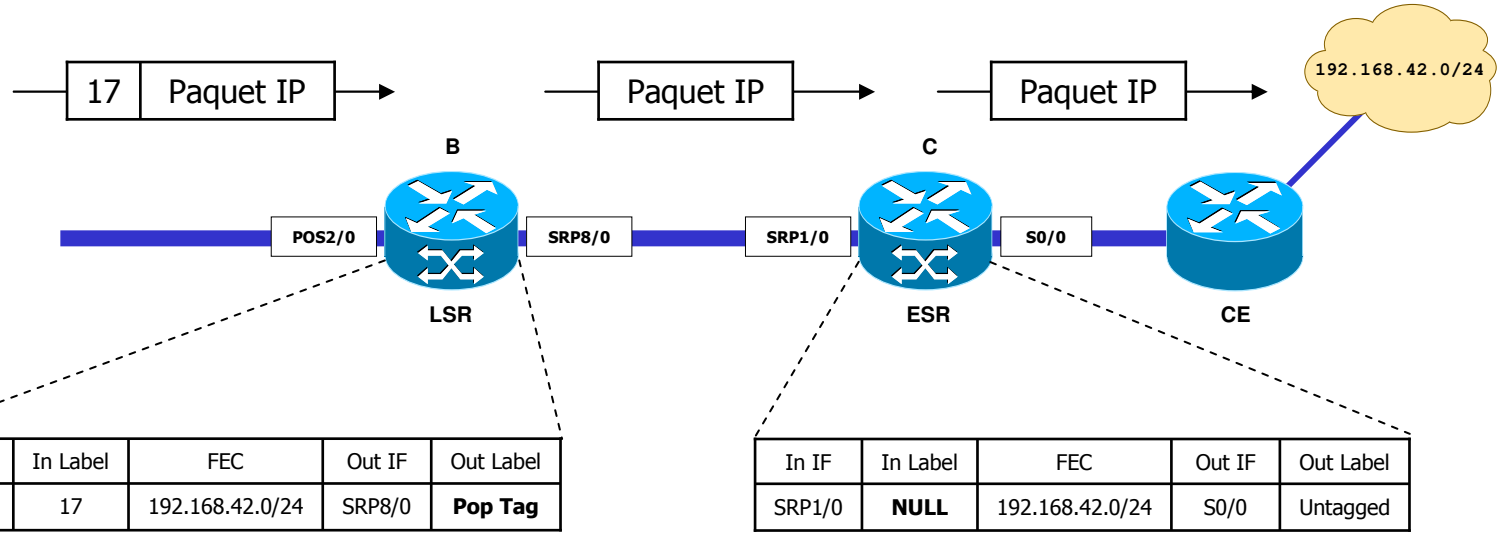
# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Acheminement du paquet (suite,fin)
  - Le routeur PE de sortie « C » dépose le label et envoie le paquet IP sur l'interface de sortie.



# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Penultimate Hop Popping
  - Le PE de sortie pour une FEC n'a pas besoin d'un label pour déterminer l'interface de sortie.
  - Ce routeur distribue donc un label nul pour cette FEC à ses voisins upstream pour leur demander de recevoir les paquets pour cette FEC sans label.
  - Les upstream déposent (« poppent ») le label avant de le forwarder au PE de sortie




| In IF | In Label | FEC | Out IF | Out Label |
|---|---|---|---|---|
| POS2/0 | 17 | 192.168.42.0/24 | SRP8/0 | **Pop Tag** |

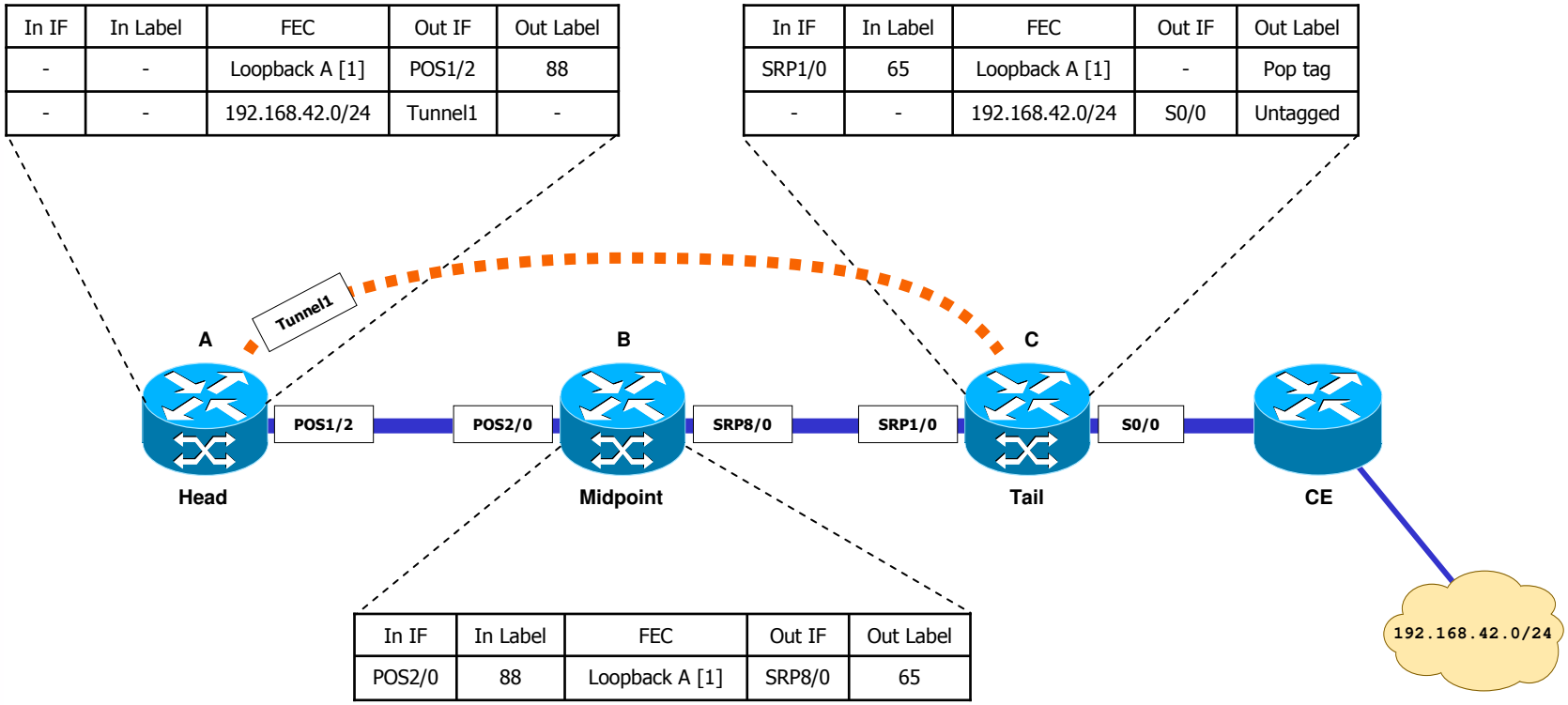
| In IF | In Label | FEC | Out IF | Out Label |
|---|---|---|---|---|
| SRP1/0 | **NULL** | 192.168.42.0/24 | S0/0 | Untagged |

# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Label Switched Path
  - Le trafic en entrée sur un routeur PE à destination d'un routeur PE de sortie commun pour n FEC est acheminé via un chemin constitué de routeurs P.
  - Cette succession de routeurs est un Label Switched Path dont la route peut être dérivé du SPF constitué par l'IGP ou forcée par l'administrateur.
  - Chaque LSP est associé avec un certain nombre d'attributs, notamment la bande passante désirée de bout en bout.
  - Un LSP est unidirectionnel : pour une communication duplex entre deux PE, il y a donc deux LSP d'établis dont les routes ne sont pas obligatoirement symétriques.
  - Les LSR peuvent ne pas disposer de binding pour la FEC de destination du paquet IP original, il leur suffit d'avoir un binding vers le routeur PE de sortie.

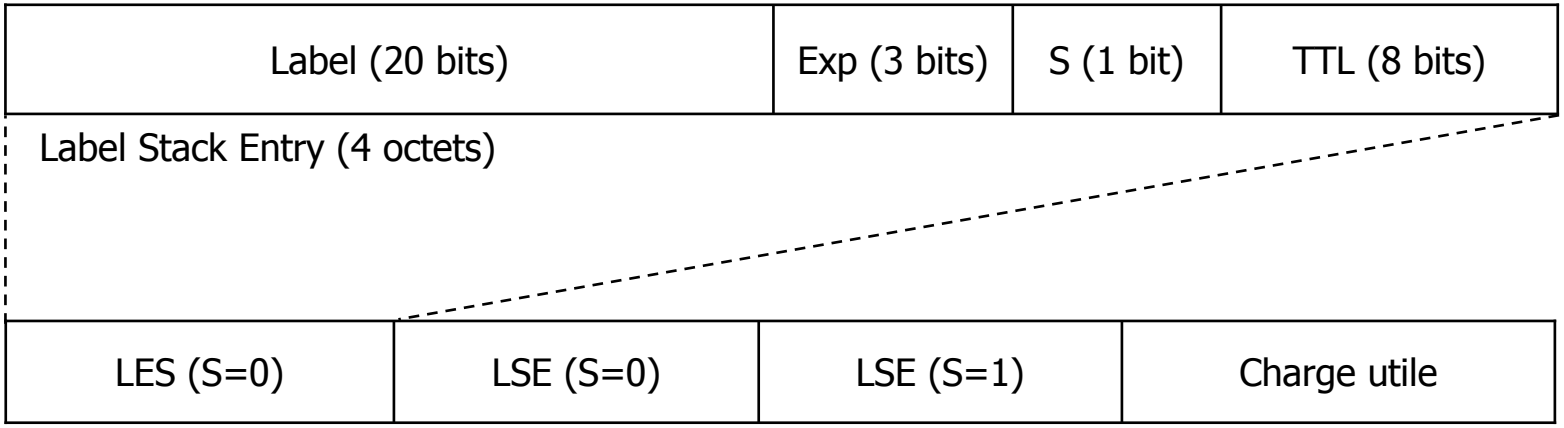
# MPLS : Concepts et principes de fonctionnement

- Principe de commutation de labels : Label Switched Path (suite)



# MPLS : Concepts et principes de fonctionnement

- Protocoles mis en jeu : MPLS
  - Le format de l'en-tête MPLS permet de cumuler différents labels de façon chaînée.



# MPLS : Concepts et principes de fonctionnement

- Protocoles mis en jeu : MPLS

  - L'en-tête MPLS est située entre la couche réseau et le header de la couche de liaison : ATTENTION AUX PROBLEMES DE MTU !

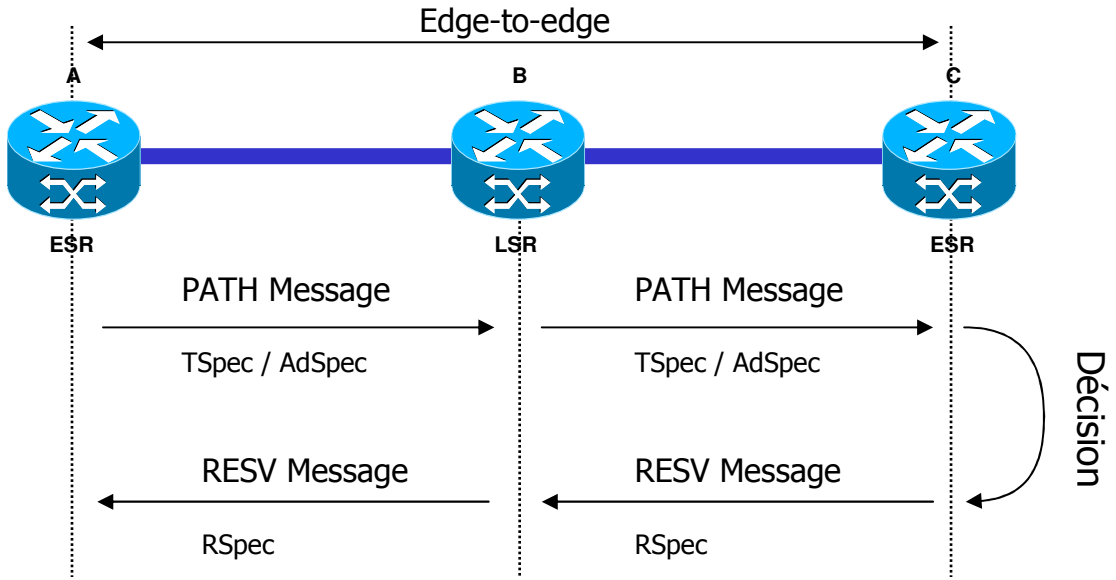
| Sonet/SDH | HDLC | MPLS | Charge utile |
|---|---|---|---|
| Layer 1 | Layer 2 | Layer 2 ½ | Layer 3 en MPLS/IP |

# MPLS : Concepts et principes de fonctionnement

- Protocoles mis en jeu : Distribution de labels
  - Plusieurs protocoles permettent à un routeur MPLS de découvrir ses voisins et d'échanger des bindings avec eux :
    - TDP : Tag Distribution Protocol, créé par Cisco à l'époque du « Tag Switching »,
    - LDP : Label Distribution Protocol, protocole normalisé iso fonctionnel avec TDP,
    - RSVP : Ressource Reservation Protocol, à l'origine utilisé comme protocole de signalisation dédié à la QoS a été étendu à la distribution de label et est particulièrement adapté pour le Traffic Engineering,
    - CR-LDP : Extension à LDP permettant le Traffic Engineering.
  - Deux routeurs MPLS s'échangeant des bindings sont appelés « Label Distribution Peers »,
  - Lorsque deux routeurs MPLS sont des « Label Distribution Peers », on parle de « Label Distribution Adjacency » entre eux.
  - TDP est tombé en désuétude et est propriétaire, il est recommandé de ne pas l'utiliser.
  - Le protocole le plus répandu est LDP, CR-LDP n'étant pas implémenté très largement, LDP et RSVP peuvent travailler de concert pour le Traffic Engineering.

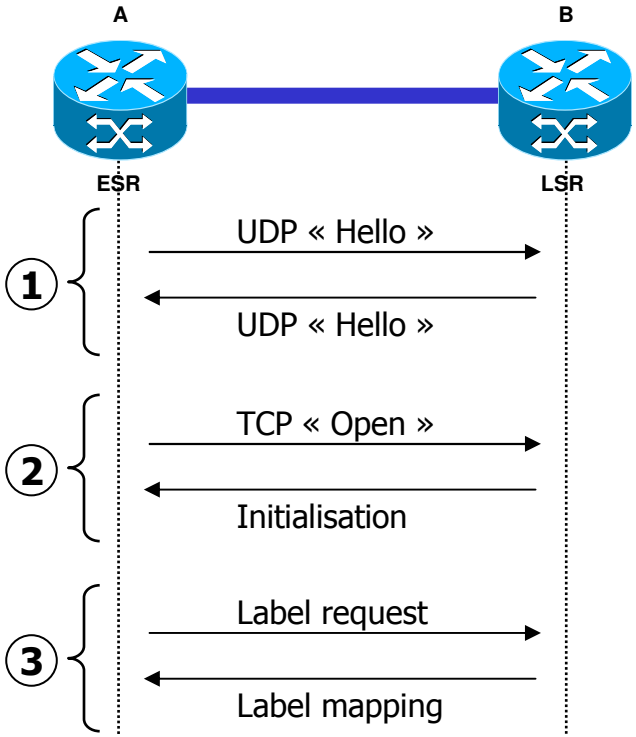
# MPLS : Concepts et principes de fonctionnement

- Protocoles mis en jeu : RSVP comme protocole de distribution de labels
  - A l'origine, RSVP permet aux applications de spécifier leurs besoins de QoS pour des flux de niveau 4 unidirectionnels.
  - Pour distribuer des labels, les paths messages RSVP de type PATH, un label pour le flux (correspondant à la FEC bindée) est transmis simultanément.
  - RSVP travaille en multicast.



# MPLS : Concepts et principes de fonctionnement

- Protocoles mis en jeu : LDP
  - Protocole spécialement conçu pour la distribution de bindings et le maintient d'adjacences.




1. Messages de découverte. Ils annoncent et maintiennent la présence d'un Label Distribution Peer dans le réseau ou sur le lien.
2. Messages de session. Ils permettent d'établir une Label Distribution Adjacency entre deux routeurs ou de la résilier.
3. Messages opérationnels dans la session. Ils permettent la création, la modification ou la destruction d'un binding.
4. Messages de notification. Ils permettent de fournir des notifications ou de signaler des erreurs.

# MPLS : Concepts et principes de fonctionnement

- Protocoles mis en jeu : Allocation des labels
  - Comme vu précédemment, les labels pour une FEC sont alloués localement et distribués aux routeurs d'upstream par les routeurs de downstream.
  - Il existe deux méthodes pour échanger les bindings :
    - Allocation « Downstream on demand » :

      Le label sera alloué par le routeur de downstream uniquement sur demande spécifique d'un binding pour une FEC par le routeur d'upstream et pour une interface spécifique.
    - Allocation « Unsolicited downstream » :

      Le label sera distribué à par le routeur de downstream à tout ses upstreams sur toutes les interfaces.

# Fonctionnalités apportées par MPLS

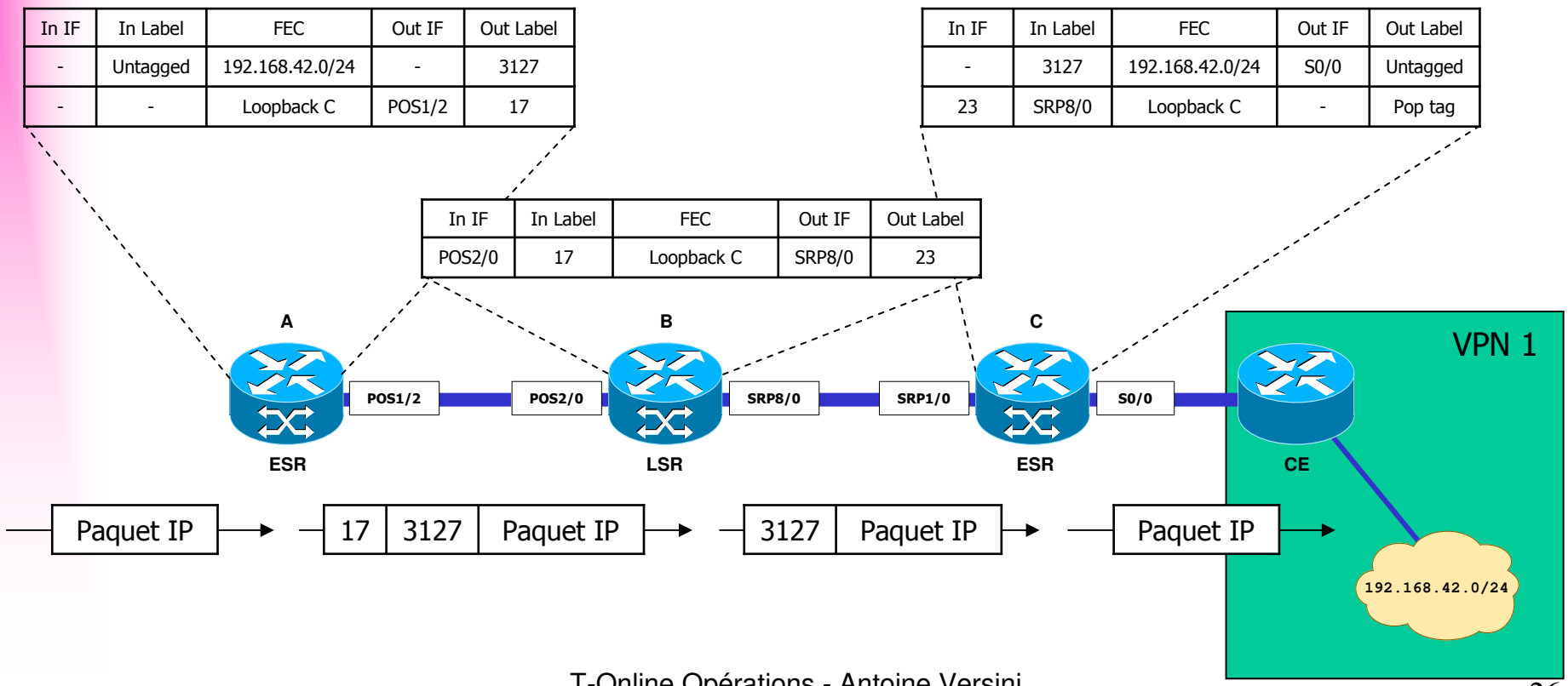
- Décision de routage unique
  - Lorsqu'un paquet se présente en entrée du réseau sur une interface externe d'un routeur PE, ce dernier évalue la FEC correspondant à l'adresse de destination du paquet et impose le label correspondant au binding distribué par le routeur de downstream choisi par l'IGP.
  - Dans le cas d'un paquet à destination d'un préfixe BGP appris par une session eBGP sur un autre PE, la FEC choisie est le next-hop BGP. Ainsi les routeurs P intermédiaires n'ont pas besoin de connaître un binding pour chaque préfixe de la table de routage complète, mais juste des bindings correspondant aux routeurs PE du réseau.
  - La conséquence est que les routeurs P peuvent ne pas disposer du full-routing et que leurs performances en sont largement améliorées, la FEC n'étant pas réévaluée à chaque saut.
  - Le meshing BGP des grands réseaux est simplifié puisque seuls les routeurs de bordure ont des sessions iBGP.
  - On retrouve le fonctionnement du "layered model", mais les commutateurs ATM/FR sont remplacés par des routeurs P et les circuits virtuels par des LSP.

# Fonctionnalités apportées par MPLS

- MPLS VPN : VPN Routing/Forwarding
  - Les VRF sont des tables de routage qui sont dissociées les unes des autres.
  - Une interface physique ou une interface logique (ex: une sous-interface dans un VLAN) non MPLS sur un PE peut être placée dans une VRF.
  - Les adresses de destination des paquets entrant sur cette interface seront évaluées uniquement avec les routes de la VRF associée à cette interface.
  - Les routes d'une VRF sont identifiées par des « route distinguisher » qui ont le format des communautés BGP.
  - Il est possible d'importer dans ou d'exporter vers une VRF les routes d'autres VRF, rendant possible à certains sites de voir d'autres sites qui peuvent par contre ne pas se voir entre eux.
  - Les routeurs PE distribuent les routes des VRF entre eux via des sessions MBGP avec des NRLI « vpnv4 ».
  - Pour chaque préfixe distribué en BGP dans une VRF entre deux PE, un label est associé par le routeur qui annonce ce préfixe.
  - Lorsque le PE reçoit sur une interface dans une VRF un paquet IP à destination d'un préfixe annoncé par un autre PE, il impose le label fourni par le PE de sortie en bas de la stack.

# Fonctionnalités apportées par MPLS

- MPLS VPN : VPN Routing/Forwarding (suite)
  - Ensuite le PE impose en haut de la stack le label distribué par le routeur P de downstream correspondant à la FEC du next-hop BGP du routeur PE de sortie.

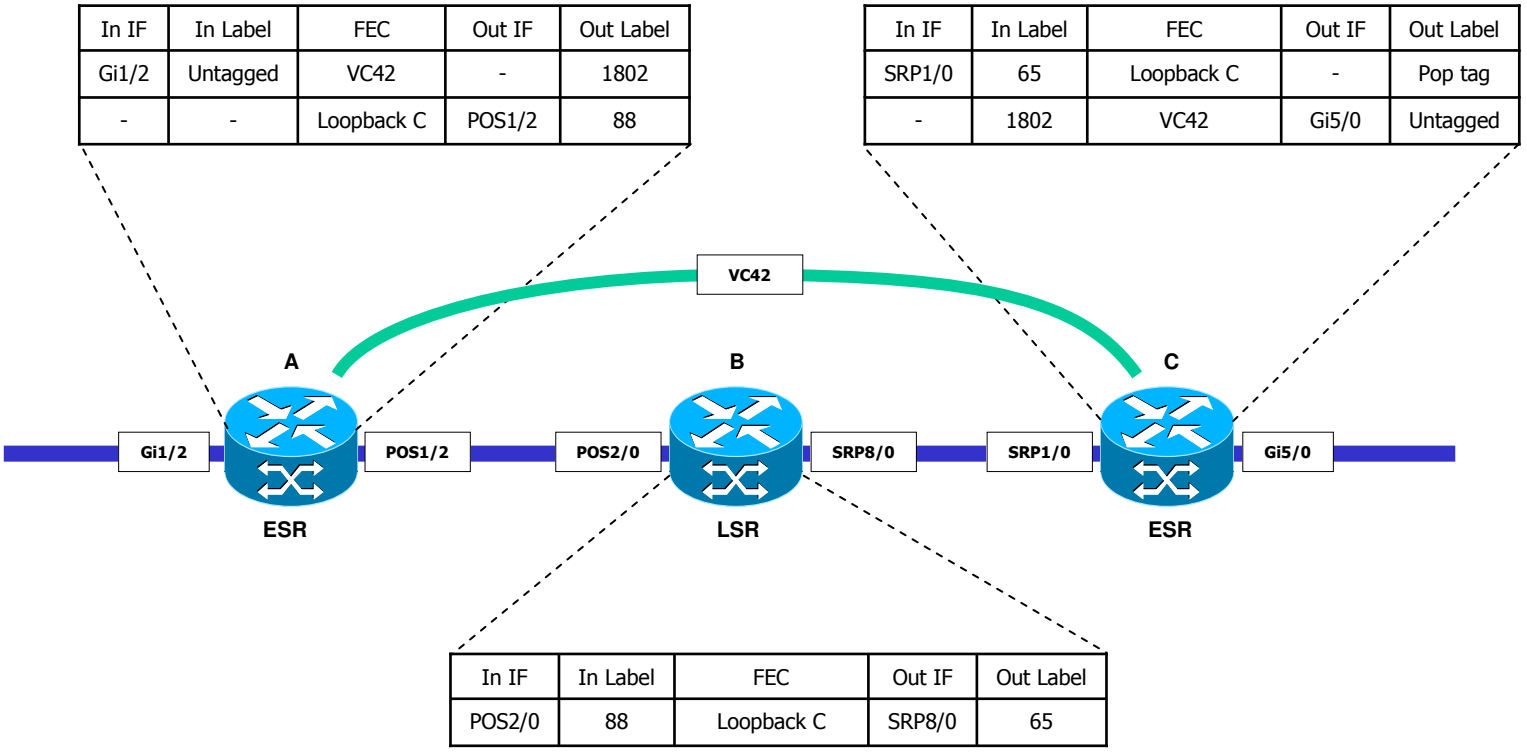


# Fonctionnalités apportées par MPLS

- AToM : Any Transport over MPLS
  - La nature multi protocolaire de MPLS permet de transporter d'autres charges utiles que des paquets IP.
  - Virtuellement, n'importe quel type de paquets, cellules, trames peuvent être directement labellisées par un PE en entrée avec un binding distribué par un PE de sortie qui va déposer le label et la retransmettre sans altération sur une interface de sortie.
  - La charge utile est transportée dans un circuit virtuel monté entre les deux PE via une Adjacence de Distribution de Label qui peut être multi-hop. LDP permet cela via les « targetted hello ».
  - Les deux PE négocient les bindings et se transmettent les charges utiles avec en bas de la stack le label identifiant le VC et en haut de la stack le label identifiant leurs next-hops respectifs.
  - AToM permet de créer des VPN de niveau 2.

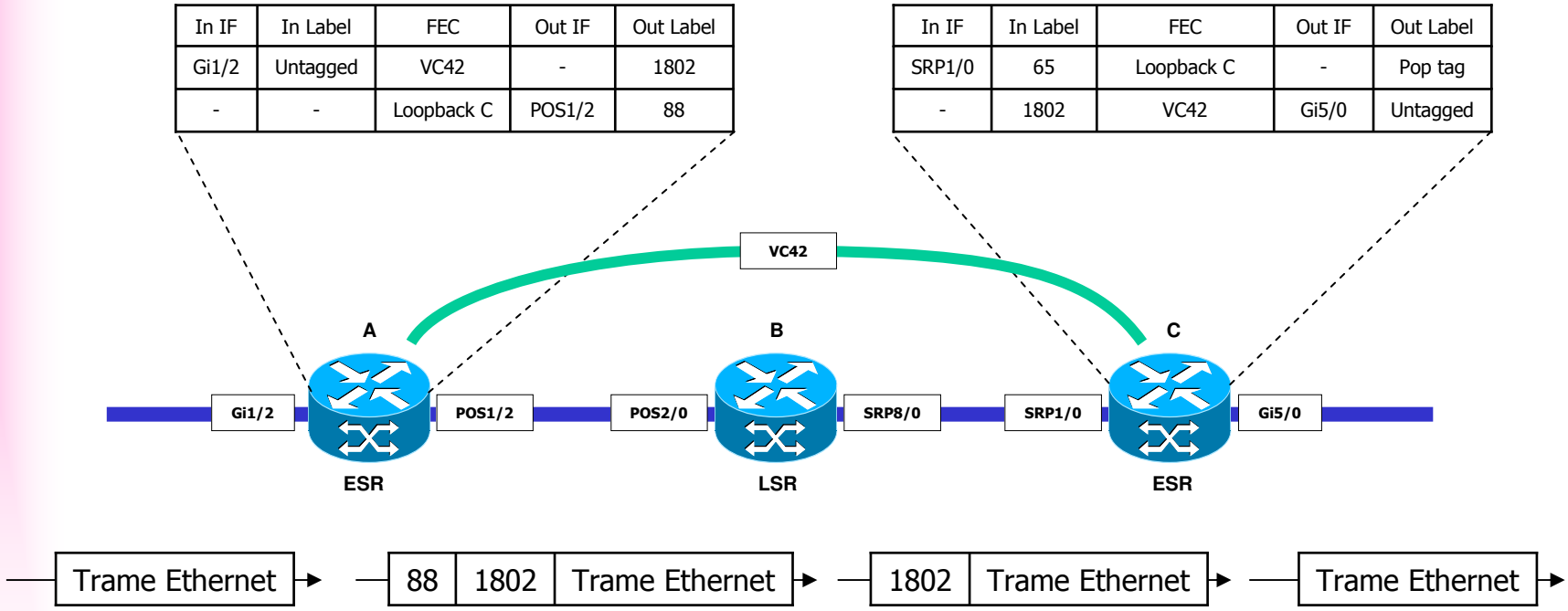
# MPLS : Concepts et principes de fonctionnement

- AToM : Any Transport over MPLS (suite)



# MPLS : Concepts et principes de fonctionnement

- AToM : Any Transport over MPLS (fin)



# Fonctionnalités apportées par MPLS

- Traffic Engineering
  - Un LSP est accompagné de plusieurs attributs, notamment son besoin en bande passante et sa priorité.
  - L'IGP transporte entre ses adjacences des informations telle que la capacité d'un lien.
  - Ainsi, lorsqu'un routeur PE de tête de LSP détermine le chemin que va prendre le LSP pour se terminer sur le routeur PE de queue, il sait quelle est la bande passante disponible hop par hop pour la priorité du LSP.
  - Si le chemin le plus court ne dispose pas d'assez de bande passante, un chemin moins court peut être sélectionné.
  - L'algorithme de l'IGP est modifié pour permettre à un PE de queue d'annoncer au PE de tête des préfixes à router via le LSP (résolution du problème récursif : adresse de destination d'un tunnel annoncée dans le tunnel conduisant à sa chute pour raison de boucle). C'est le « enhanced SPF ».

# Fonctionnalités apportées par MPLS

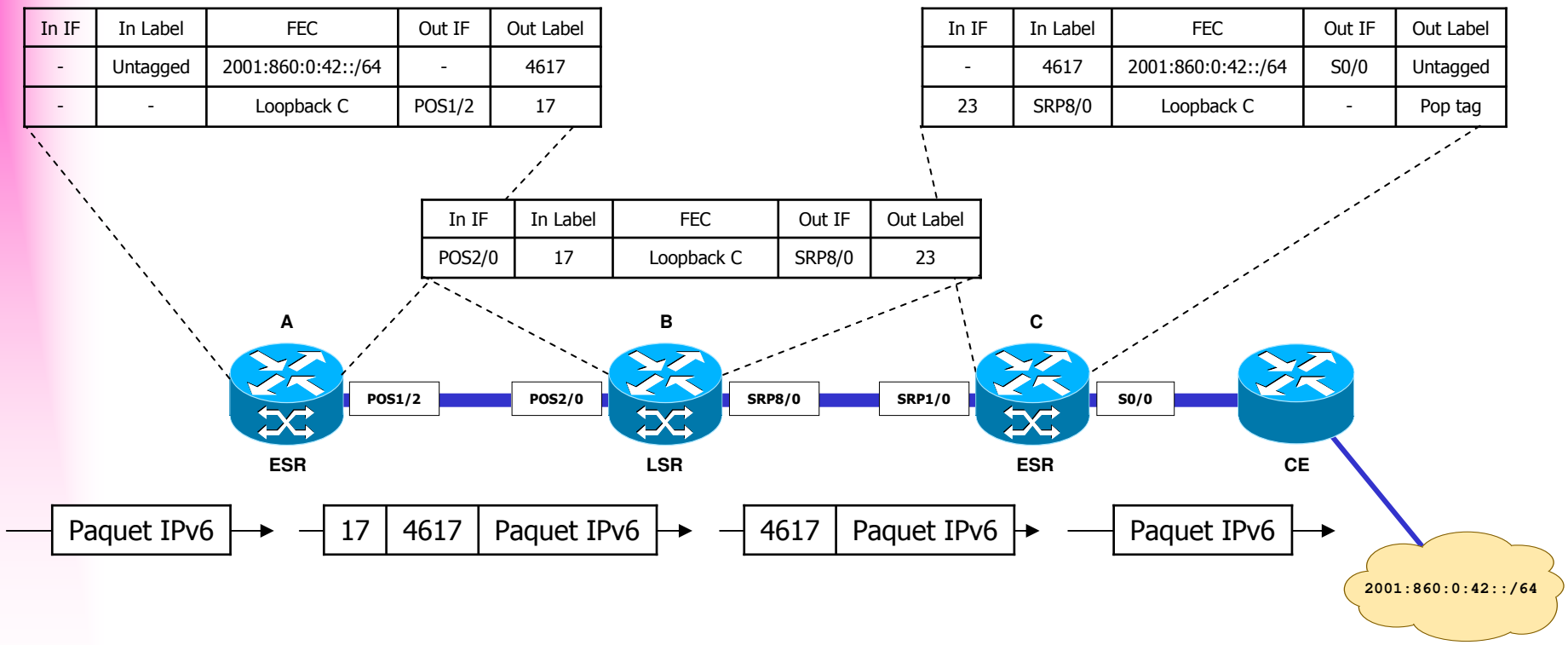
- Traffic Engineering (suite)
  - Si la communication entre tous les PE passe via des LSP signalisés avec des besoins en bande passante et des niveau de priorité, le backbone sait calculer la charge de chacun des liens.
  - Si dans le chemin le plus court il n'y a pas la bande passante nécessaire, un autre chemin sera choisi conduisant à une utilisation optimale des ressources du réseau.
  - Le traffic engineering permet de retarder les upgrades de capacité, de délester les administrateurs de la charge de positionner les métriques IGP pour déplacer du trafic vers des liens plus disponibles à la main.
  - Il existe sur certain routeurs des fonctionnalités d'adaptation de la demande de bande passante pour un LSP en fonction du pic de trafic mesuré sur une période de temps.
  - Le fait que les LSP soit unidirectionnels permet également de router sur des liens chargés dans un sens mais pas dans l'autre le LSP « aller » et de router par ailleurs le LSP « retour ».

# Fonctionnalités apportées par MPLS

- MPLS et IPv6 : 6PE
  - 6PE permet à un réseau ne sachant pas router de l'IPv6 d'acheminer les paquets IPv6 dans des labels via les routeurs P.
  - Les routeurs PE 6PE ont entre eux des sessions iBGP en IPv4 dans lesquels ils s'annoncent des NLRI « ipv6 unicast » avec des labels.
  - Le routeur 6PE d'entrée impose le label correspondant au binding distribué par le routeur 6PE de sortie pour la FEC correspondant à l'adresses IPv6 de destination du paquet en bas de la stack, puis impose en haut de la stack le label du binding pour la FEC du routeur P de downstream pour le next-hop IPv4 du 6PE de sortie.
  - 6PE utilise donc le même mécanisme que MPLS VPN, considérant que le réseau IPv6 est une VRF.

# MPLS : Concepts et principes de fonctionnement

- MPLS et IPv6 : 6PE (suite)



# Feedback opérationnel

- Transition du réseau IP vers le réseau MPLS/IP
  - MPLS peut être activé graduellement sur des équipements en production !
  - La plupart du temps les équipements se comportent bien, mais il y a une technique pour éviter tout dérapages :

    Il s'agit d'activer MPLS entre des routeurs « en pointillé ». Sur une chaîne de 4 routeurs, activer MPLS entre les deux premiers et entre les deux derniers, mais pas entre les deux centraux. L'objectif étant que le penultimate hop popping fasee que jamais un paquet ne voyage labellisé sur le réseau à ce stade.

    Une fois que les « îlots » MPLS sont en place, établir les adjacences entre les routeurs centraux (toujours les routeurs P) pour que les LSP entre PE se mettent en place sans que l'IGP ne soit trop perturbé.

    Le trafic sur les routeurs P est alors label-switché.

# Feedback opérationnel

- Transition vers le Traffic Engineering
  - Un réseau peut-être mixte : commutation de label et LSP traffic engineerés

    Si un routeur backbone dispose d'une route pour un préfixe à travers un LSP, il sait la redistribuer un binding pour la FEC correspondante à ses voisins MPLS, même s'il n'y a pas de LSP avec TE de signalisés.

    Cela donne des traceroutes étonnants, où deux hops intermédiaire ne font pas de commutation de label : le routeur voisin de celui qui a la route via un LSP reçoit un paquet sans label (penultimate hop popping) et le routeur suivant impose le label du LSP.

    Cette technique surprenante permet d'activer le TE de façon incrémentale sur le réseau sans perturber le trafic. Mais la condition requise est que tous les routeurs doivent faire fonctionner un IGP avec des extensions TE !

# Feedback opérationnel

- Transition vers le Traffic Engineering (suite)
  - Un réseau mixte n'a pas grand intérêt. Il faut que tout le trafic soit conduit dans des LSP afin que les routeurs puissent déterminer la matrice de LSP entre les PE et la bande passante requise pour chacun d'eux. Ils en déduisent la charge des liens qui doit être proche du débit instantané observé.
  - Si du trafic est label-switché de façon classique en parallèle de trafic conduit dans des LSP signalisés TE, l'IGP-TE ne tiendra pas compte de ce trafic et pourrait éventuellement router des LSP sur un lien plein, car les routeurs ne vérifient pas la charge réelle des liens mais du cumul des besoins en bande passante des LSP transitant déjà sur ces liens.
  - L'administrateur peut forcer le LSP à suivre un chemin prédéterminé et autoriser le réseau à le re-router de façon dynamique si des LSP plus prioritaires exploitent la bande passante disponibles sur un ou plusieurs liens du chemin forcé. Certaines implémentations permettent également de demande au PE de tête du LSP d'éviter certains routeurs pour router un LSP.
  - L'administrateur peut évidemment forcer la bande passante requise par un LSP. Par exemple un service de LAN-to-LAN Gigabit Ethernet peut emprunter un LSP avec une bande passante forcée à 1 000 000 Kbit/s avec une priorité en fonction du niveau de service acheté.

# Questions / Réponses

- Pas trop compliquées, hein !
- Pas de guerre Cisco / Juniper.

Merci de votre attention !

☺