

# Multi Protocol Label Switching

**Du routage IP à la commutation de labels  
Application à la mise en place de VPN**

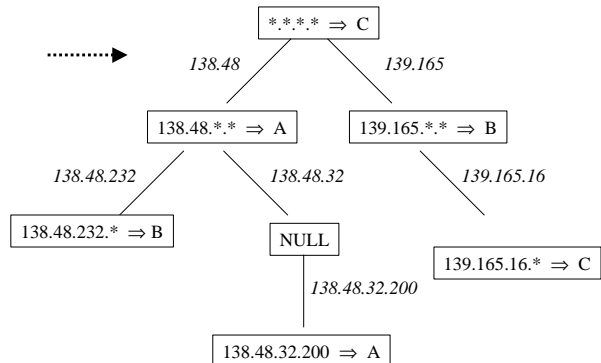
## Dans les réseaux IP traditionnels

- **Routage des paquets**
  - En fonction de **l'adresse destination** dans l'entête **de niveau 3** (Network Layer)
  - En recherchant le prochain saut (next-hop) à effectuer dans **les tables de routage** des routeurs
- **Problèmes posés**
  - Le mécanisme de recherche du prochain saut dans les tables de routage est fortement consommateur de CPU
  - La taille des tables de routage des routeurs a constamment augmenté
  - **Comment trouver une méthode plus efficace pour le routage des paquets de niveau 3 ?**

# Tables de routage IP

## • Représentation sous forme d'arbre

| SubNet        | Prefix | Next-Hop |
|---------------|--------|----------|
| 138.48.0.0    | 16     | A        |
| 139.165.0.0   | 16     | B        |
| 139.165.16.0  | 24     | C        |
| 138.48.232.0  | 24     | B        |
| 138.48.32.200 | 32     | A        |
| 0.0.0.0       | 0      | C        |

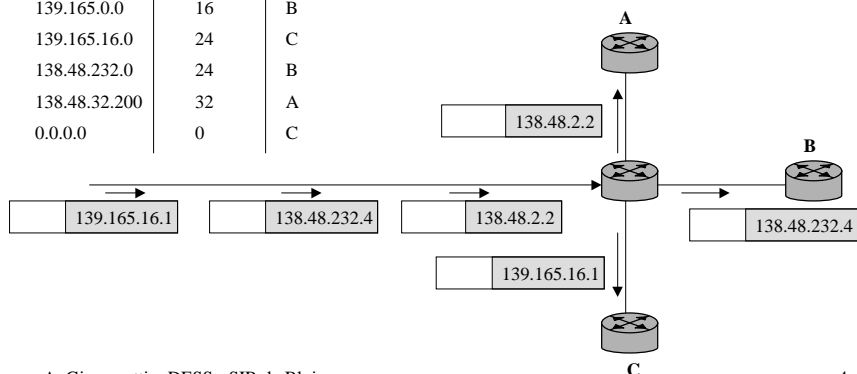


# Routage IP

## • Détermination du prochain saut

- En déterminant dans la table de routage le préfixe le plus long

| SubNet        | Prefix | Next-Hop |
|---------------|--------|----------|
| 138.48.0.0    | 16     | A        |
| 139.165.0.0   | 16     | B        |
| 139.165.16.0  | 24     | C        |
| 138.48.232.0  | 24     | B        |
| 138.48.32.200 | 32     | A        |
| 0.0.0.0       | 0      | C        |



## Dans les réseaux MPLS

- **Routage des paquets IP**

- En fonction d'une information de **label** (ou tag) insérée **entre le niveau 2** (Data-Link layer) **et le niveau 3** (Network Layer)
- En recherchant le prochain saut à effectuer dans **les tables de commutation de labels** des routeurs

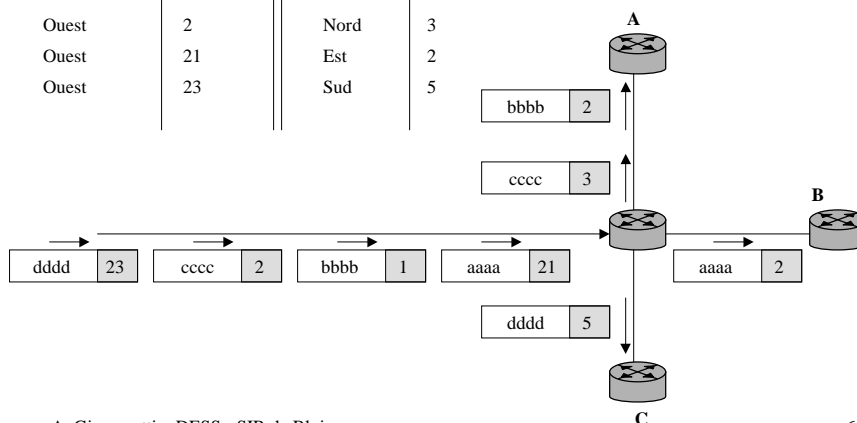
- **Problèmes posés**

- Comment attribuer un label à chaque paquet IP entrant dans un domaine MPLS ?
  - *Par classification des paquets IP dans des FEC (Forwarding Equivalence Class)*
- Comment construire les tables de commutation de labels des routeurs ?
  - *A partir des tables de routage des routeurs*
  - *Par distribution de labels entre routeurs*

## Tables de commutation

*Label Forwarding Table*

| InPort | InLabel | OutPort | OutLabel |
|--------|---------|---------|----------|
| Ouest  | 1       | Nord    | 2        |
| Ouest  | 2       | Nord    | 3        |
| Ouest  | 21      | Est     | 2        |
| Ouest  | 23      | Sud     | 5        |



## Historique et perspectives

- **Origines du protocole MPLS**

- Protocoles standards
  - Utilisation par « X.25 », « Frame Relay », « ATM » de labels identifiants des circuits, voies ou chemins virtuels
- Protocoles propriétaires
  - « Tag Switching » de Cisco, « IP Switching » de Ipsilon (nouvellement racheté par Nokia)
- Architecture MPLS définie par l'IETF dans la RFC 3031

- **Objectifs de MPLS**

- Donner aux routeurs une plus grande puissance de commutation
- Offrir de nouveaux services
  - Pour la définition de réseaux privés virtuels (en anglais, Virtual Private Network)
  - Pour l'ingénierie de trafic (en anglais, Traffic Engineering)

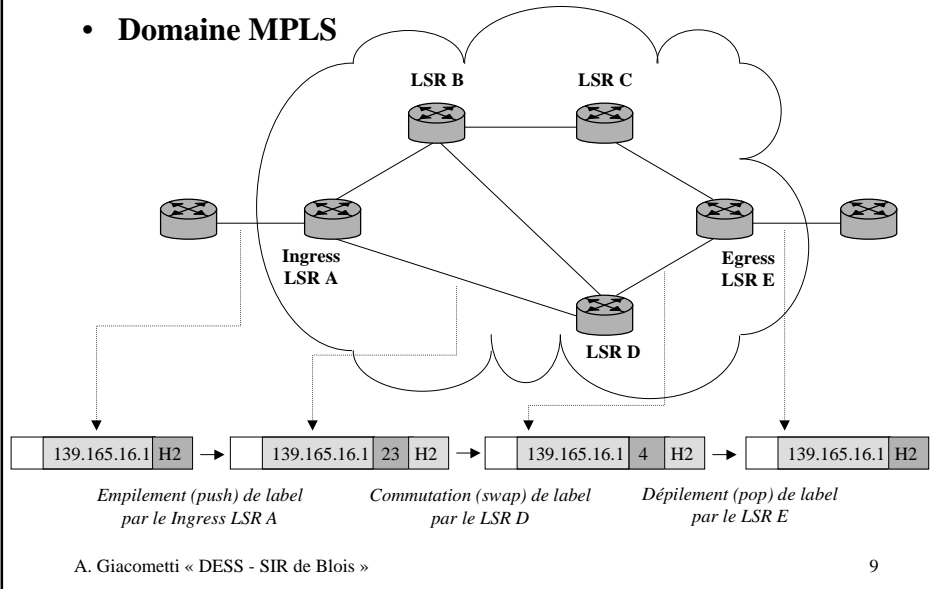
## Intégration de IP et MPLS

- **Dans quel type d'équipement ?**

- Label Switching Router (LSR)
  - Possède à la fois des tables de routage et des tables de commutation
- Trois types de LSR (relativement à un flux)
  - **Ingress Edge LSR** (routeur d'entrée)
    - » Reçoivent des paquets IP
    - » Classifient ces paquets dans des FEC (Forwarding Equivalence Class) à qui sont associées des étiquettes ou labels
    - » Encapsulent ces paquets dans des unités de données MPLS étiquetées
    - » Commutent en sortie les PDU MPLS construites
  - **Interior or Traffic LSR** (routeur interne)
    - » Reçoivent des unités de données MPLS étiquetées
    - » Commutent ces PDU MPLS en fonction de leur étiquette
  - **Egress Edge LSR** (routeur de sortie)
    - » Reçoivent des unités de données MPLS étiquetées
    - » Suppriment l'entête MPLS des PDU reçues
    - » Routent les paquets IP désencapsulés

# Architecture MPLS

## • Domaine MPLS



9

# Etiquetage des paquets IP

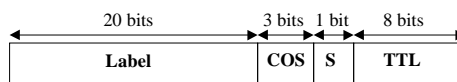
## • Deux types de solutions

- En utilisant les champs « labels » de protocoles standards
  - *DLCI (Data-Link Connection Identifier) des trames Frame Relay*
  - *VCI (Virtual Chanel Identifier) des paquets X.25*
  - *VCI/VPI (Virtual Path Identifier) des cellules ATM*
- En utilisant un entête MPLS spécifique placé entre
  - *L'entête IP de niveau 3 (Network Layer) et*
  - *Un entête de niveau 2 (Data-Link Layer)*
    - » Dans le futur, les PDU MPLS pourrait néanmoins être transmises directement au niveau physique (dans le cas de réseaux optiques)

## Entête MPLS

- **Codé sur au moins 32 bits**

- Avec un ou plusieurs labels codés sur 20 bits chacun
  - Une PDU MPLS peut contenir une pile de labels
  - Le label au sommet de la pile est transmis en premier
    - » Si S = 1, le label codé est au fond de la pile (dernier label)
    - » Si S = 0, le label codé est au-dessus d'un autre label
- Autres champs
  - TTL (Time To Live)
    - » Pour éviter que des PDU MPLS puissent boucler indéfiniment dans un domaine
  - COS (Class Of Service)
    - » Pour pouvoir appliquer plusieurs politiques de gestion des files d'attente



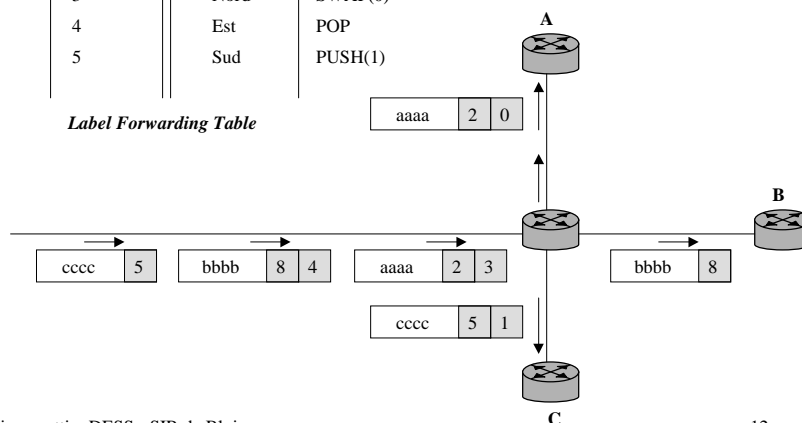
Format de l'entête MPLS

## Pile de labels

- **Modifie la structure des tables de commutation**

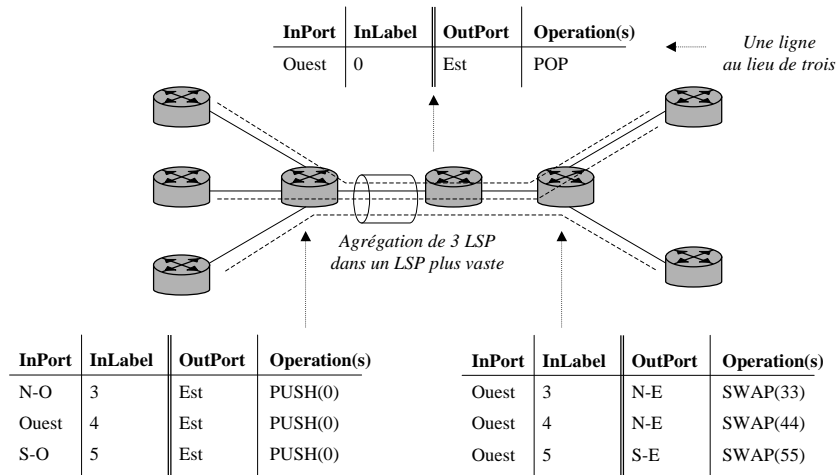
| InPort | InLabel | OutPort | Operation(s) |
|--------|---------|---------|--------------|
| Ouest  | 3       | Nord    | SWAP(0)      |
| Ouest  | 4       | Est     | POP          |
| Ouest  | 5       | Sud     | PUSH(1)      |

Label Forwarding Table



## Agrégation de LSP

- Réduction de la taille des tables de commutation



A. Giacometti « DESS - SIR de Blois »

13

## Définition de FEC (1)

- Identifie un ensemble de paquets IP
  - Seront étiquetés en sortie d'un Ingress LSR par le même label
  - Suivront ainsi le même chemin ou LSP (Label Switching Path)
- Plusieurs types de FEC
  - En fonction de leur mode de définition
    - Dans le cas le plus simple
      - » Deux paquets appartiendront à une même FEC si leurs adresses IP destination appartiennent à un même sous-réseau
    - Autres paramètres possibles
      - » Numéro de ports source et destination TCP/UDP
      - » Qualité de service demandée

A. Giacometti « DESS - SIR de Blois »

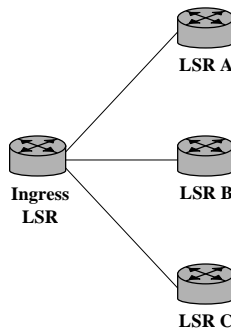
14

## Définition de FEC (2)

- Définition de FEC par sous-réseau

| SubNet       | Prefix | Next-Hop |
|--------------|--------|----------|
| 138.48.0.0   | 16     | A        |
| 139.165.0.0  | 16     | B        |
| 139.165.16.0 | 24     | C        |
| 138.48.232.0 | 24     | B        |

| FEC             | OutPort | OutLabel |
|-----------------|---------|----------|
| 138.48.0.0/16   | N-E     | 3        |
| 139.165.0.0/16  | Est     | 10       |
| 139.165.16.0/24 | S-E     | 200      |
| 138.48.232.0/24 | Est     | 4        |



Quelle est l'origine de ces labels ?

## Distribution de labels

- Problème posé

- Comment coordonner les tables de commutation des LSR ?
  - Par distribution d'associations entre FEC découvertes et labels (en anglais, FEC-Label mappings)

- Protocoles proposés

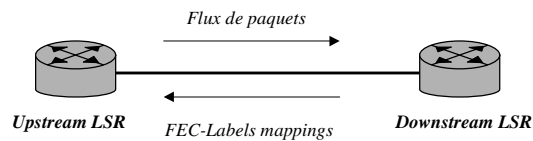
- TDP/LDP (Tag/Label Distribution Protocol)
  - TDP est un protocole CISCO propriétaire
  - LDP a été défini par l'IETF dans la RFC3036
  - Utilisés notamment dans le cas de FEC définies par sous-réseau
- Extensions de RSVP (Resource Reservation Protocol)
  - Permet d'établir des LSP en fonction de critères de ressources et de qualité de service
- Extensions de BGP (Border Gateway Protocol)
  - Permet de distribuer les labels en même temps que les routes découvertes, mais aussi l'échange de routes VPN



## Création d'associations

- **Par quels LSR ?**

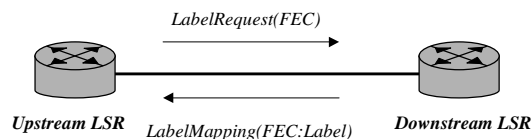
- Dans le cas d'un flux de paquets
  - Transmis d'un LSR amont (*Upstream LSR*) vers un LSR aval (*Downstream LSR*)
- Les associations FEC-Labels
  - Sont transmises du LSR aval vers le LSR amont



## Modes de distribution (1)

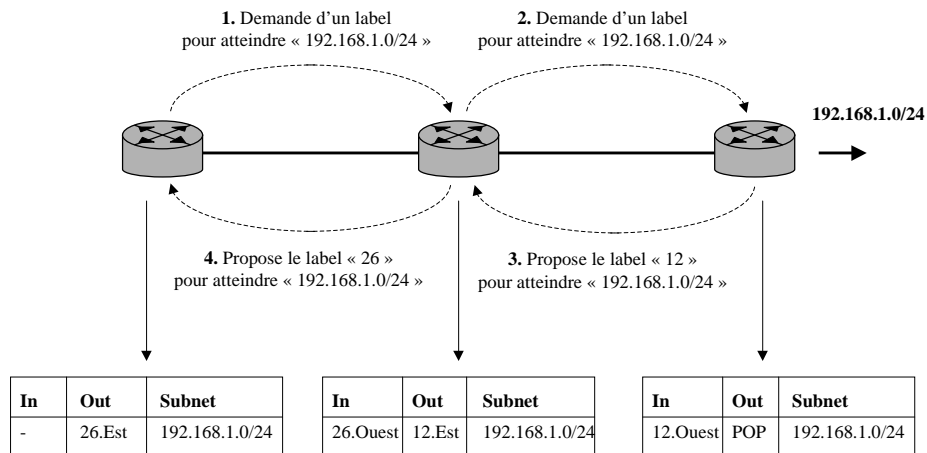
- **A quels moments sont distribués les associations ?**

- **Premier mode** : à la demande (downstream on demand label distribution)
  - Les LSR amont font des demandes d'association par FEC
  - Les LSR aval allouent des labels pour les FEC demandés



## A la demande

- **Downstream on demand**



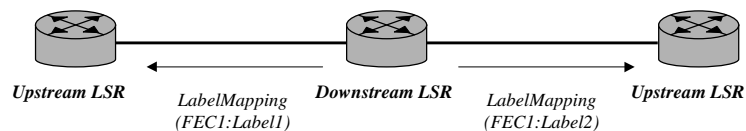
A. Giacometti « DESS - SIR de Blois »

19

## Modes de distribution (2)

- **A quels moments sont distribués les associations ?**

- **Deuxième mode** : non sollicité (unsolicited downstream label distribution)
  - Les LSR aval transmettent (indépendamment de toute demande) les associations FEC-label aux LSR amont

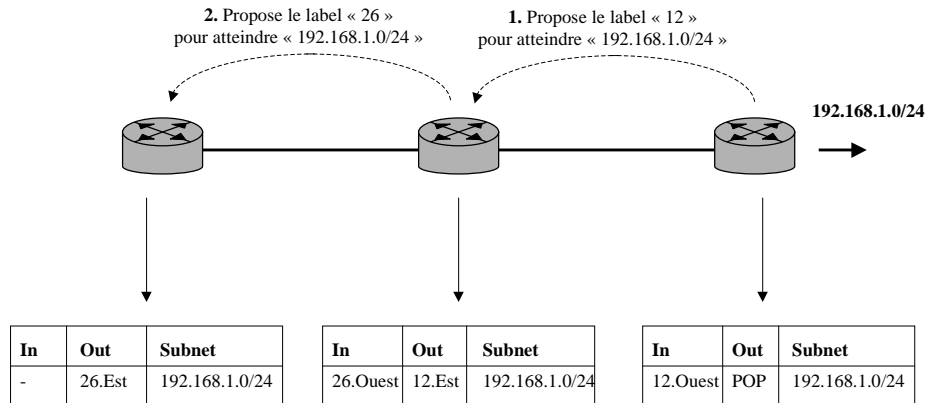


A. Giacometti « DESS - SIR de Blois »

20

# Non Sollicité

- **Unsolicited Downstream**



# Comparaison des modes

- **A la demande**

- **Avantage**
  - Seules les associations nécessaires sont transmises
  - Les LSR mémorisent uniquement les associations utilisées
- **Inconvénient**
  - Délai d'attente incompressible
    - » En cas de défaillance d'un « saut-suivant », une nouvelle requête d'association doit être faite au nouveau « saut-suivant » sélectionné (par routage dynamique)

- **Non sollicité**

- **Avantage**
  - Un même LSR (amont) peut obtenir plusieurs associations pour un même FEC (obtenues de différents LSR aval)
- **Inconvénient**
  - Des associations reçues peuvent ne pas être utilisées par la suite

## Modes de distribution (3)

- **Qui initie la distribution des labels pour un FEC donné ?**
  - Independent Control
    - *Chaque LSR choisit indépendamment des autres*
      - » A quel moment il distribue ses associations FEC-Label
    - *Chaque LSR assigne un label à tous les FEC qu'il découvre*
      - » Par exemple à tous les sous-réseaux qu'il découvre par routage dynamique
  - Ordered Control
    - *Les LSR frontaliers (Ingress et Egress LSR)*
      - » Initient la distribution d'associations FEC-Label pour un FEC donné
    - *Les autres LSR du domaine MPLS*
      - » Propagent seulement en amont les associations reçues

## Label Distribution Protocol

- **Historique**
  - Basé sur le protocole TDP (Tag Distribution Protocol) de Cisco
  - Défini par l'IETF dans la RFC 3036
    - *TDP et LDP diffèrent principalement par le format des paquets ou messages qu'ils transmettent*
- **Fonctions principales**
  - Découverte de voisins
    - *Pour un LSR donné, déterminer si ses voisins directs sont également des LSR ou seulement des routeurs classiques*
  - Distribution de labels
    - *Intègre l'ensemble des modes de distribution décrits précédemment*
      - » Unsolicited downstream or Downstream on demand
      - » Ordered or Independent control

## Découverte de voisins

- **Principe de base**

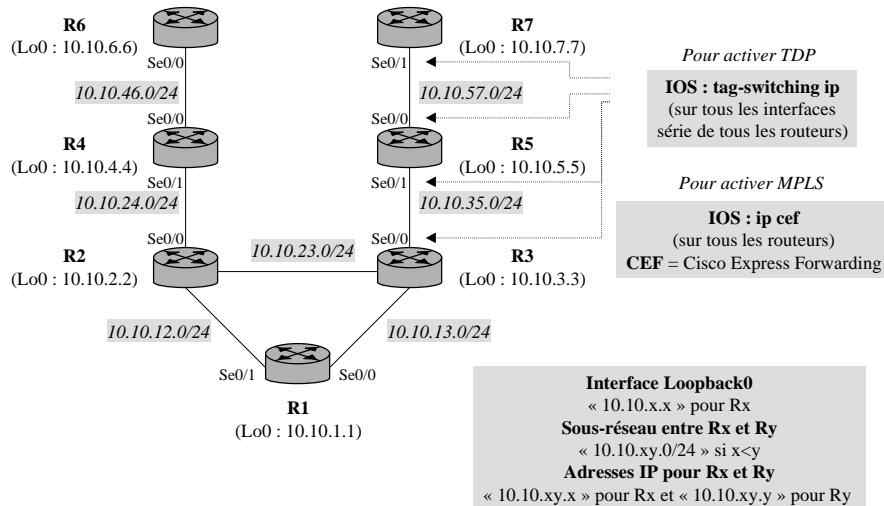
- Tout LSR transmet périodiquement à ses voisins
  - Un message *LDP Hello* avec comme adresse destination une adresse « multicast » réservée
- Tout LSR recevant un message LDP Hello
  - Répond par un autre message *LDP Hello* s'il utilise lui-même LDP
- Etant donné deux LSR voisins utilisant LDP
  - Le LSR avec la plus grande adresse IP
    - » Devient actif et établit une connexion TCP sur le port 646 (711 pour TDP) pour la transmission de messages LDP
  - Le LSR avec la plus petite adresse IP
    - » Devient passif et attend l'établissement de la connexion TCP avec son voisin actif
  - **Note** : des options peuvent être négociées pendant l'établissement de la connexion entre LSR voisins
    - » Modes de distributions des labels, valeur du temporisateur « keep-alive », etc.

## Messages LDP

- **Principaux types de messages**

- Message « Hello »
  - Utilisés pour la découverte de voisins
- Message « Keep-alive »
  - Transmis périodiquement en l'absence d'autres messages (panne)
- Message « Label Mapping »
  - Utilisé par un LSR pour annoncer une association FEC-Label
- Message « Label Withdrawal »
  - Utilisé par un LSR pour retirer une association annoncée auparavant
- Message « Label Release »
  - Utilisé par un LSR pour indiquer qu'il n'utilisera plus une association reçue précédemment
- Message « Label Request »
  - Utilisé par un LSR pour demander un label pour un FEC donné

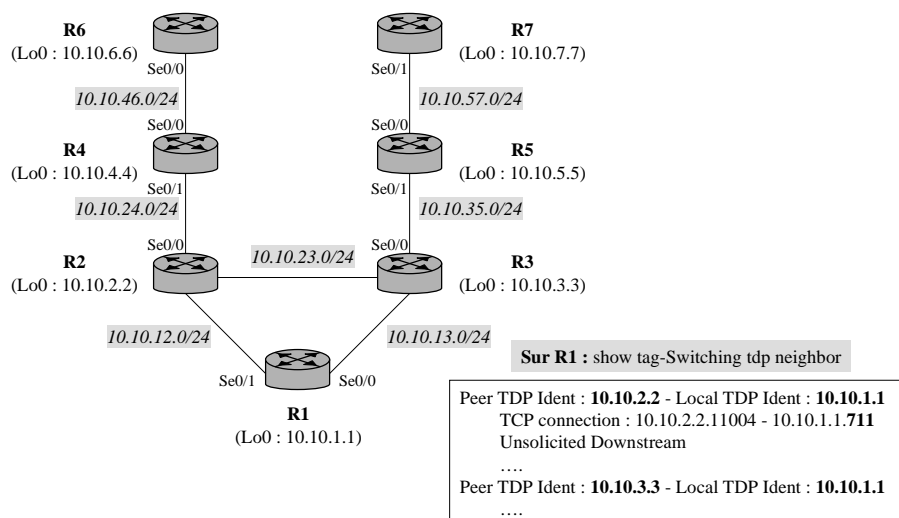
## Réseau exemple



A. Giacometti « DESS - SIR de Blois »

27

## Découverte de voisins



A. Giacometti « DESS - SIR de Blois »

28

# Tables de commutation

## • Rôle des TIB et TFIB

### ➤ Tag Information Base

- Cette table contient toutes les associations apprises par un TSR de ses TSR voisins (TSR = Tag Switching Router)
- Elle contient pour chaque FEC (sous-réseau IP) la liste des labels affectés par les TSR voisins

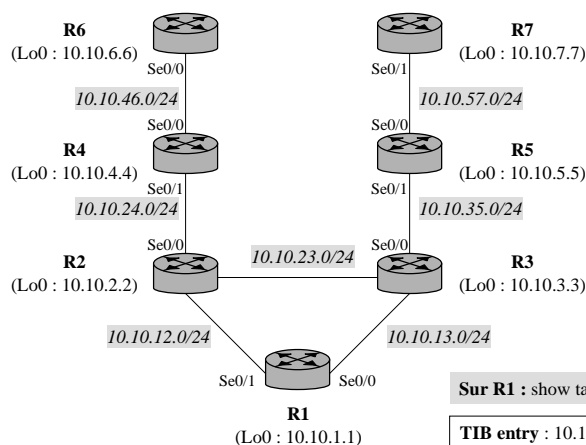
### ➤ Tag Forwarding Information Base

- Cette table est un sous-ensemble de la TIB
  - » Pour chaque sous-réseau, elle contient l'entrée de la TIB correspondant au plus court chemin
  - » Le plus court chemin (déterminé par routage dynamique) est retrouvé dans la table de routage du routeur
- Elle est utilisée pour la commutation des paquets MPLS

### ➤ Note : dans le cas de LDP

- Label Information Base « LIB »
- Label Forwarding Information Base « LFIB »

# Rôle de la TIB



Sur R1 : show tag tdp bindings 10.10.4.4 255.255.255.255

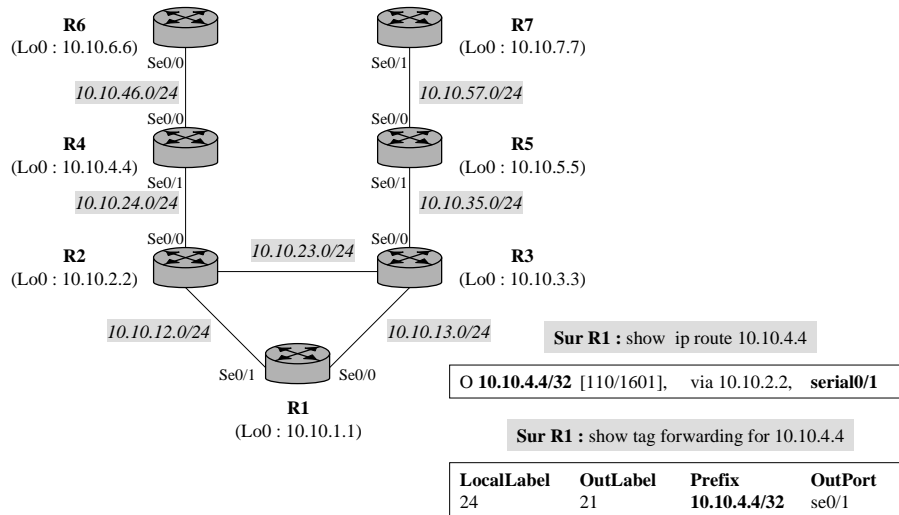
TIB entry : 10.10.4.4/32

Local binding: tag: **24**

Remote binding: tsr: 10.10.3.3, tag: **20**

Remote binding: tsr: 10.10.2.2, tag: **21**

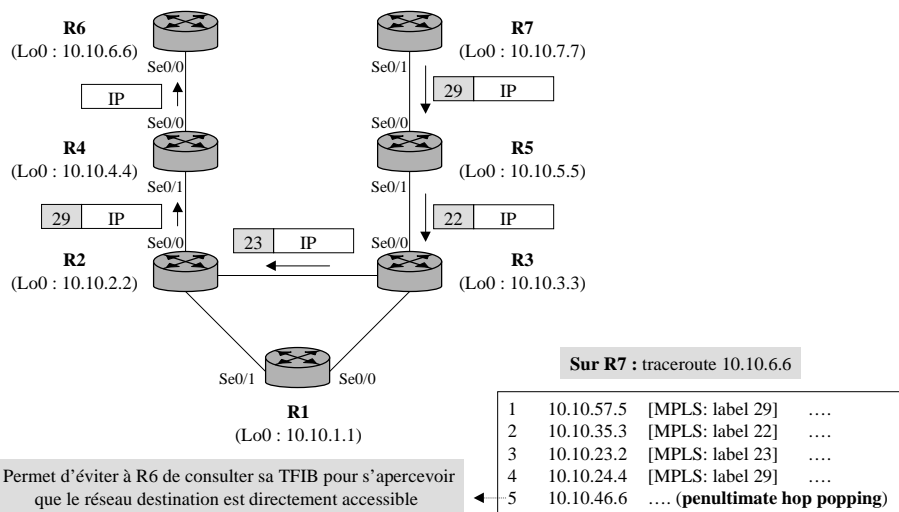
## Rôle de la TFIB



A. Giacometti « DESS - SIR de Blois »

31

## Label Switching Path

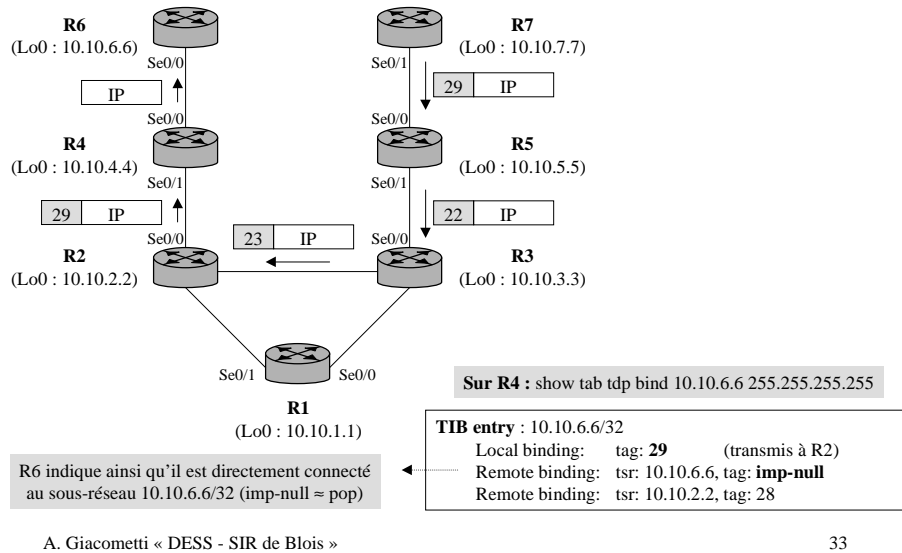


A. Giacometti « DESS - SIR de Blois »

32



## Penultimate Hop Popping



33

## Rétention des associations

### • Mode libéral

- Les LSR conservent dans leur TIB/LIB toutes les associations annoncées par leur voisins
  - Y compris celles provenant de voisins ne correspondant pas au « saut-suivant » pour le FEC en question (coûteux en mémoire)
  - En cas de perte de lien, permet de sélectionner rapidement un nouveau Label de sortie
    - » Sur notre exemple, si le lien entre R1 et R2 tombe en panne, R1 peut sélectionner dans sa TIB le label 20 annoncé par R3 pour 10.10.4.4/32

### • Mode conservatif

- Les LSR conservent dans leur TIB/LIB uniquement les associations sélectionnées dans leur TFIB/LFIB
  - En cas de perte de lien, nécessite d'attendre de nouvelles association pour remplacer les association défectueuses
    - » Sur notre exemple, R1 conserverait pour le sous-réseau 10.10.4.4/32 uniquement le label 21 provenant de 10.10.2.2

## Définition de VPN

- **Problème posé**

- Comment permettre à plusieurs réseaux indépendants de cohabiter sur une même infrastructure (FAI) ?
  - Ces réseaux doivent pouvoir utiliser le même adressage privé
  - Les trafics issus des différents réseaux doivent être isolés

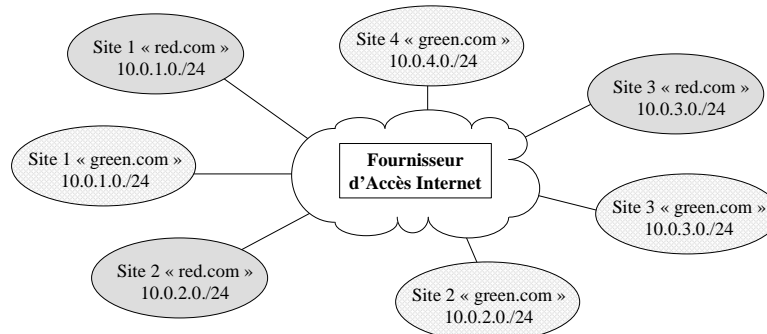
- **Solutions proposées**

- Solutions classiques
  - Utilisation de lignes spécialisées
  - Utilisation de tunnels IP
  - Utilisation de réseaux Frame Relay ou ATM
- Solutions alternatives
  - Utilisation simple de MPLS
  - Utilisation améliorée de MPLS

## Exemple d'architecture VPN

- **Comment interconnecter dans deux VPN distincts**

- Les réseaux des sites « red.com » et « green.com »
  - En supposant que les deux réseaux utilisent le même adressage privé « 10.0.X.0/24 » pour le site X



## Solutions classiques (1)

- **Utilisation de Lignes Spécialisées**

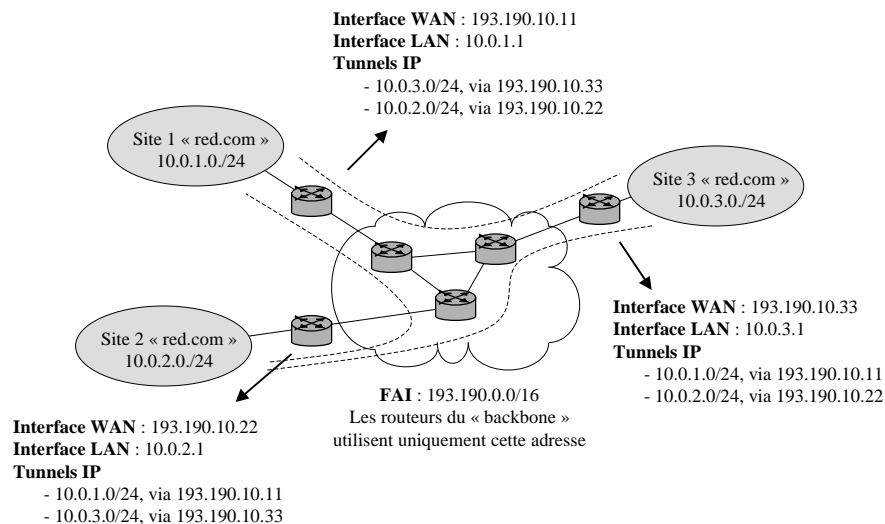
- **Avantages**
  - *Qualité de service a priori de très bonne qualité*
- **Inconvénients**
  - *Coût pouvant être très élevé*
    - » Même en cas de trafic inter-sites faible
  - *Nombre de lignes à louer potentiellement important*
    - »  $N.(N-1)$  pour une architecture à maillage complet
  - *Solution peu évolutive*
    - » Pour chaque nouveau site, au moins une nouvelle LS doit être mise en place

## Solutions classiques (2)

- **Utilisation de Tunnels IP**

- **Avantages**
  - *Coût moins important*
    - » Partage plus important des ressources du backbone
  - *Solution évolutive*
    - » Nécessaire de configurer une seule connexion physique pour chaque nouveau site à connecter (même en cas de tunnels à maillage complet)
- **Inconvénients**
  - *Qualité de Service non nécessairement garantie*
  - *Nombre de tunnels à configurer potentiellement important*
    - »  $N.(N-1)$  pour une architecture à maillage complet
  - *Sécurité dépendante du mécanisme de « tunneling » utilisé*
    - » Faible avec GRE (Generic Routing Encapsulation)
    - » Meilleure avec IPsec
  - *Configuration lourde*
    - » Principalement à la charge des clients

## Exemple simple de tunnels IP



A. Giacometti « DESS - SIR de Blois »

39

## Solution MPLS simple

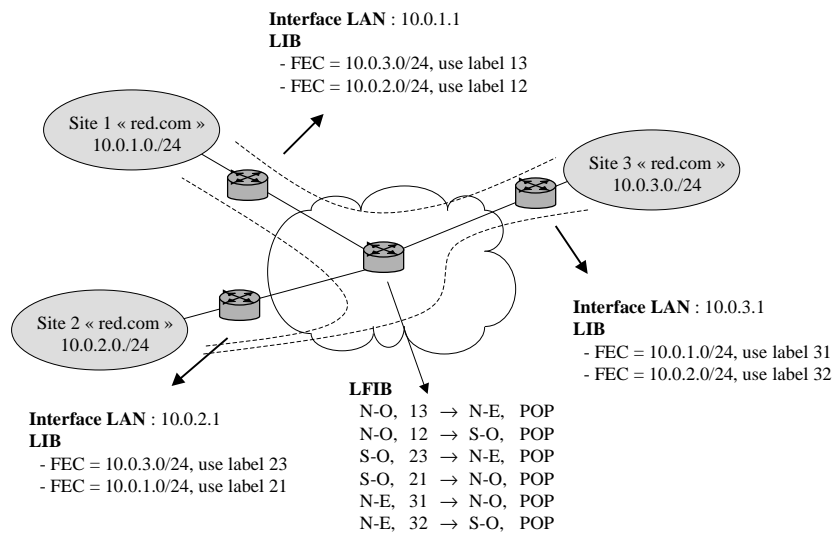
### • Principe de base

- Création **manuelle** de LSP entre routeurs « clients » des différents sites appartenant à un même VPN
  - *Même principe que les VPN basés sur Frame Relay ou ATM*
- Avantages
  - *Solution évolutive*
    - » Une seule connexion physique à configurer pour chaque nouveau site
    - » La bande passante offerte à chaque LSP peut être modifiée aisément
- Inconvénients
  - *Support de MPLS*
    - » Nécessaire à la fois au niveau des routeurs « clients » des sites et des routeurs du « backbone » MPLS
  - *Configuration lourde*
    - » Les routeurs « clients » doivent être reconfigurés pour chaque nouveau site
    - » Les routeurs du « backbone » doivent également être configurés pour chaque nouveau LSP

A. Giacometti « DESS - SIR de Blois »

40

## Architecture MPLS simple



A. Giacometti « DESS - SIR de Blois »

41

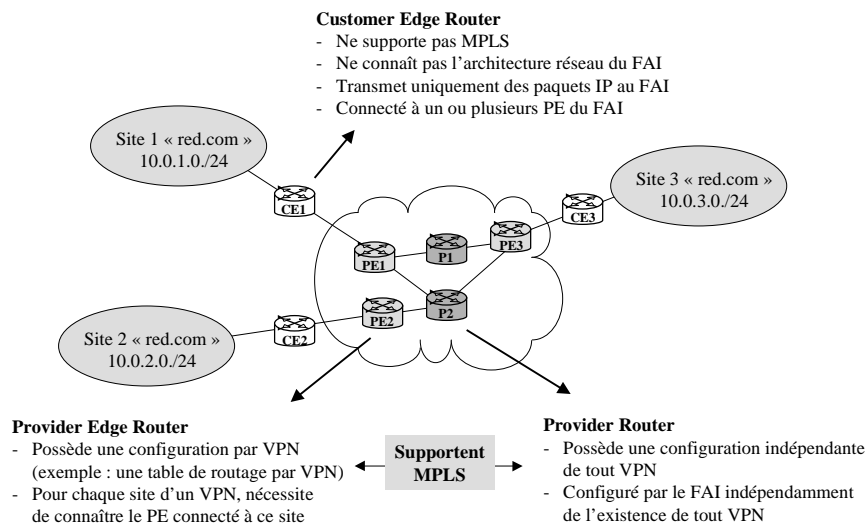
## Solution MPLS améliorée

- **Comment trouver une solution ?**
  - Qui soit la plus automatique possible
    - Pour les clients des différents VPN
    - Pour le fournisseur d'accès également
  - Qui facilite l'ajout de nouveaux sites à un VPN
    - En nécessitant seulement la configuration
      - » Des routeur « client » des nouveaux sites
      - » Des routeur du FAI connectés directement aux routeurs « client » des nouveaux sites

A. Giacometti « DESS - SIR de Blois »

42

## Architecture MPLS améliorée



A. Giacometti « DESS - SIR de Blois »

43

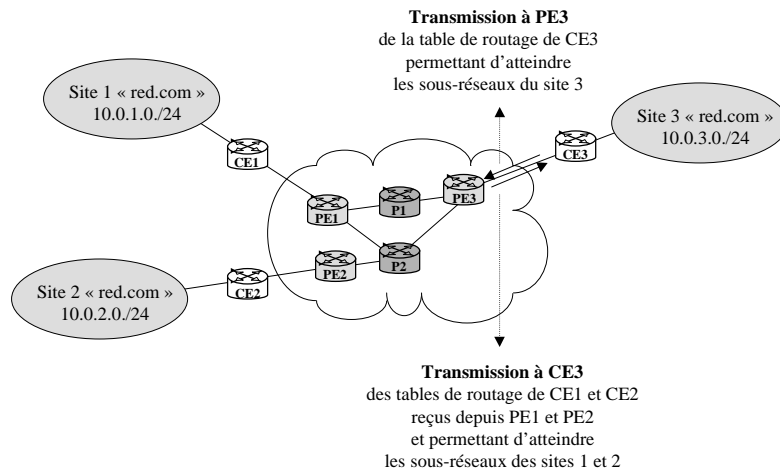
## Routage entre CE et PE (1)

- **Comment distribuer les routes apprises ?**
  - En utilisant un protocole de routage dynamique
    - Exemple : RIP, OSPF, etc.
  - Chaque CE d'un VPN( $\alpha$ ) doit transmettre à son ou ses PE paires
    - Les routes apprises pour atteindre les sous-réseaux localisées sur son site
  - En retour, le ou les PE paires doivent annoncer au CE en question
    - Les routes à suivre pour atteindre les sous-réseaux localisés sur d'autres sites, mais appartenant au même VPN( $\alpha$ )
    - Ces routes seront transmises depuis d'autres PE connectés à des CE du même VPN( $\alpha$ )

A. Giacometti « DESS - SIR de Blois »

44

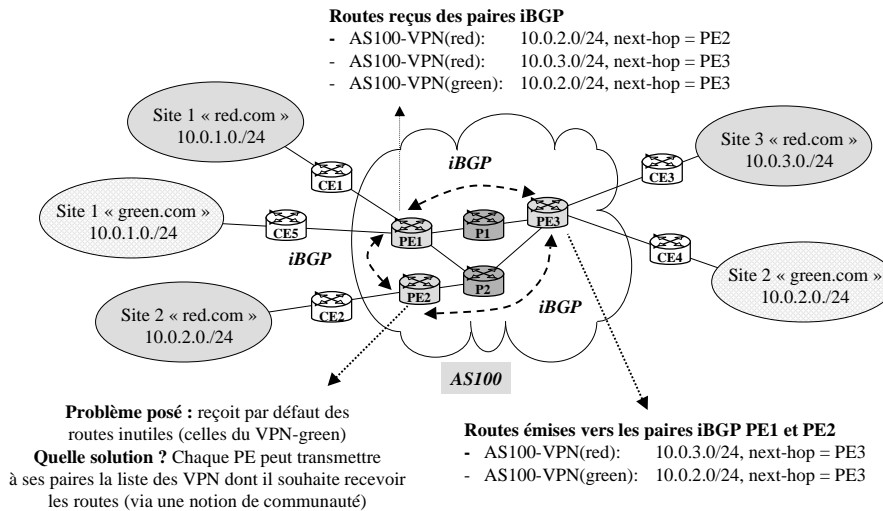
## Routage entre PE et CE (2)



## Routage entre PE (1)

- **Problème posé**
  - Impossible de distribuer directement les tables de routage des CE
    - Les sites des différents VPN peuvent en effet utiliser les mêmes systèmes d'adressage privés
- **Solution proposée**
  - Distribuer non pas des adresses IP, mais des adresses IP-VPN composées de deux éléments
    - Un identifiant appelé RD (Route Distinguisher) composé
      - » D'un identifiant de FAI (ex : numéro AS de système autonome)
      - » D'un identifiant de VPN
    - Une adresse de sous-réseau
      - » Située à l'intérieur du VPN identifié au sein du RD
  - Utiliser des extensions de BGP (Border Gateway Protocol)
    - MP-iBGP = entre PE d'un même AS (MP=MultiProtocol, i=interior)
    - MP-eBGP = entre PE de différents AS (e=exterior)

## Routage entre PE (2)



A. Giacometti « DESS - SIR de Blois »

47

## Commutation de labels

- **Utilisation de deux niveaux de labels**
  - Un premier niveau de label
    - Utilisé par les PE pour atteindre les PE apparaissant comme prochain saut dans leurs tables de routage
  - Un deuxième niveau de label
    - Utilisé par les PE pour atteindre le bon CE
      - » Chaque PE allouera un label à chaque CE/VPN auquel il est connecté
- **Distribution des labels**
  - Utilisation de LDP à l'intérieur du backbone
    - Entre routeurs P et PE, ou P et P
  - Utilisation de MP-BGP
    - Entre routeurs PE en transmettant simultanément routes et labels

A. Giacometti « DESS - SIR de Blois »

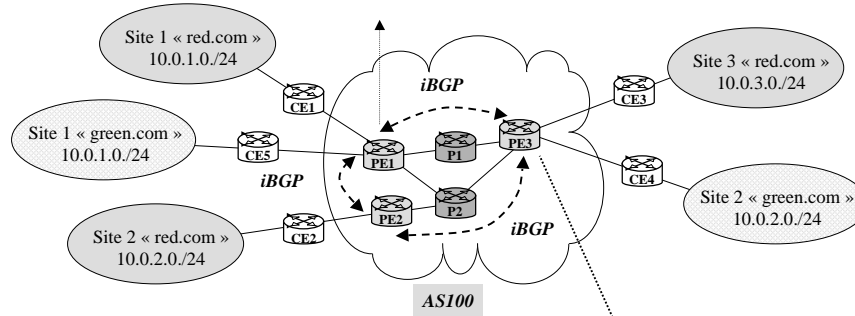
48



## Distribution des labels de niveau 2

### Routes et labels reçus du pair iBGP PE3

- AS100-VPN(red): 10.0.3.0/24, next-hop = PE3, **PUSH(22)**
- AS100-VPN(green): 10.0.2.0/24, next-hop = PE3, **PUSH(33)**



### Routes et labels émis vers les paires iBGP PE1 et PE2

- AS100-VPN(red): 10.0.3.0/24, next-hop = PE3, **PUSH(22)**
- AS100-VPN(green): 10.0.2.0/24, next-hop = PE3, **PUSH(33)**

#### Label Information Base

- Ouest, **Label = 22** → N-E, **POP**
- Ouest, **Label = 33** → S-E, **POP**

## Distribution des labels de niveau 1

### Label Information Base

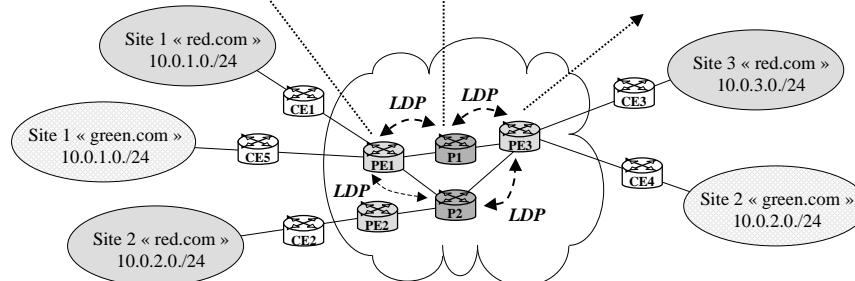
- PE3 → Est, **PUSH(5)**

### Annonces de P1 pour FEC=PE3

- A PE1, utilise **Label = 5**
- #### Label Information Base
- Ouest, **Label = 5** → Est, **POP**

### Annonces de PE3 pour FEC=PE3

- A P1, utilise **Label = NULL**
  - A P2, utilise **Label = NULL**
- #### Label Information Base
- Ouest, **Label = 22** → N-E, **POP**
  - Ouest, **Label = 33** → S-E, **POP**



**Label = NULL** pour indiquer  
un saut ultime (penultimatehop)

## Exemple final de commutation

### Routing Table

- AS100-VPN(red): 10.0.3.0/24, next-hop = PE3, PUSH(22)

### Label Information Base

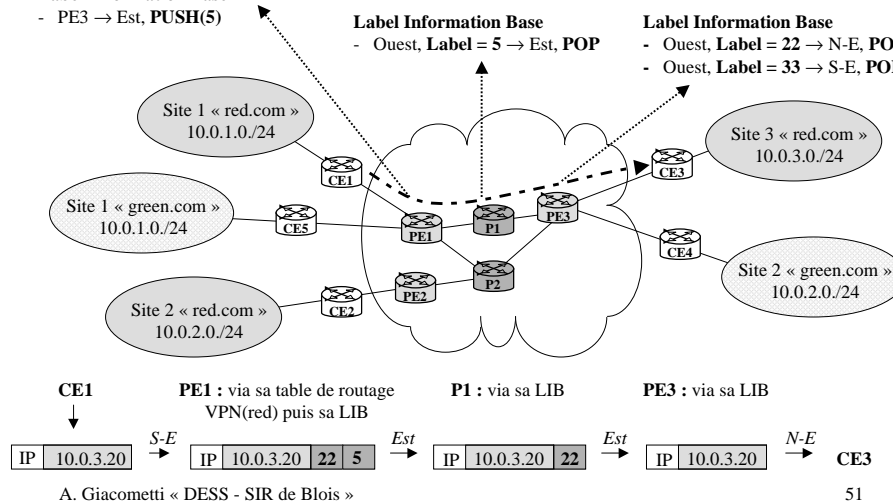
- PE3 → Est, PUSH(5)

### Label Information Base

- Ouest, Label = 5 → Est, POP

### Label Information Base

- Ouest, Label = 22 → N-E, POP
- Ouest, Label = 33 → S-E, POP



## Bibliographie

### • Quelques ouvrages

- I. Pepelnjack and J. Guichard, « **MPLS and VPN Architectures** », Cisco Press Editor (2001)
  - Exemples pratiques de configuration
- U. Black, « **MPLS and Label Switching Networks** », Prentice Hall Editor (2002)
  - Reprend les principales RFC liées à MLPS
- B. Davie and Y. Rekhter « **MPLS Technology and Applications** », Morgan Kaufman (2000)
  - Excellent rapports sur « [www.amazon.com](http://www.amazon.com) »