Les Tunnels GMPLS

Jean-Claude Bermond,
David Coudert,
Joanna Moulierac,
Stephane Pérennes,
Hervé Rivano,
Ignasi Sau
Fernando Solano Donado

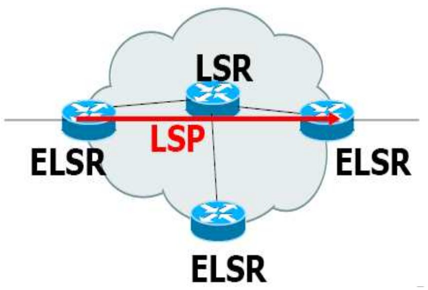
Références : Présentation MPLS Emilie Camisard (*www.renater.fr/IMG/pdf/MPLS_V4.pdf*)

MPLS

- Milieu des années 90 : apparition de solutions propriétaires multicouches, afin de simplifier le modèle IP over ATM et de réduire les congestions au niveau des routeurs.
- Puis, standardisation de ces solutions par l'IETF: naissance de MPLS.
- Les goulots d'étranglement disparaissent ensuite avec l'arrivée des routeurs basés sur du hardware: focalisation sur les services offerts par MPLS: Traffic Engineering, VPN de niveaux 2 et 3…

Terminologie MPLS

- **LSR**: Label Switching Router
- **ELSR**: Edge Label Switching Router
- **LSP**: Label Switched Path
- **FEC**: Forwarding Equivalence Class



Principes de MPLS

Next hops des routeurs déterminés par un protocole de routage

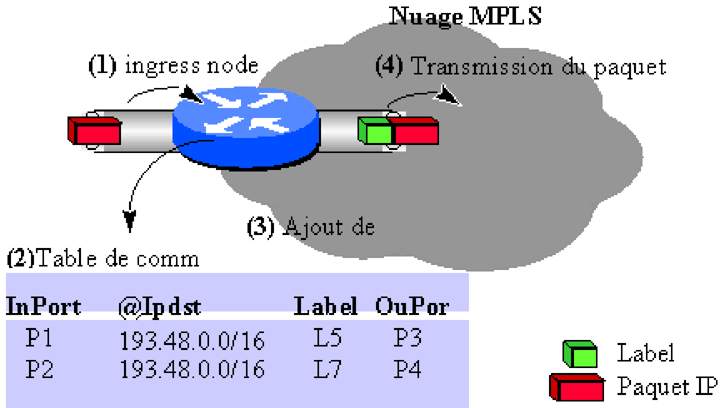
- Classification en FEC à l'entrée du réseau
- Les noeuds du réseau MPLS associent un label à chaque FEC
- Ajout d'un label à chaque paquet entrant
- Commutation au coeur du réseau grâce au label
- Une FEC peut être basée sur l'@ destination

et/ou d'autres critères (QoS, TE…)

Forwarding MPLS

- **Forwarding**: basé sur l'analyse du label, qui détermine la route vers une destination.
- **Contrôle**: création et maintien de la base des labels.
- Routes apprises par un IGP (*Interior Gateway Protocol*)
- Lien fait entre l'IGP et le label à apposer
- Distribution des labels avec un protocole spécifique (LDP, RSVP-TE)

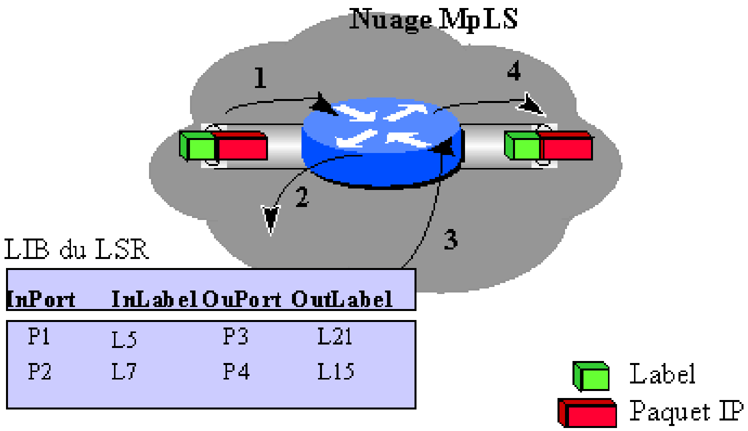
A l'entrée du domaine MPLS



InPort	@Ipdst	Label	OuPor
P1	193.48.0.0/16	L5	P3
P2	193.48.0.0/16	L7	P4

1 - Le paquet IP arrive sur l' ingress node
2 - Le protocole de routage IP détermine, à partir de l'adresse IP de l'eggress node, la FEC, le label et le port de sortie.
3 - Ajout de l'en-tête
4 - Paquet IP + Label envoyé vers le noeud suivant

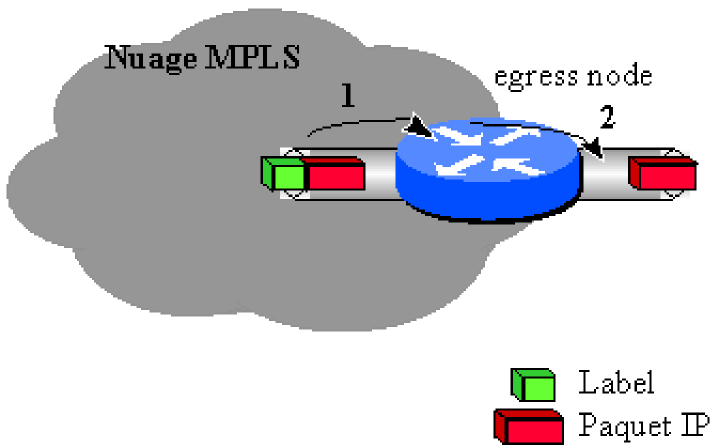
Dans le cœur du domaine MPLS



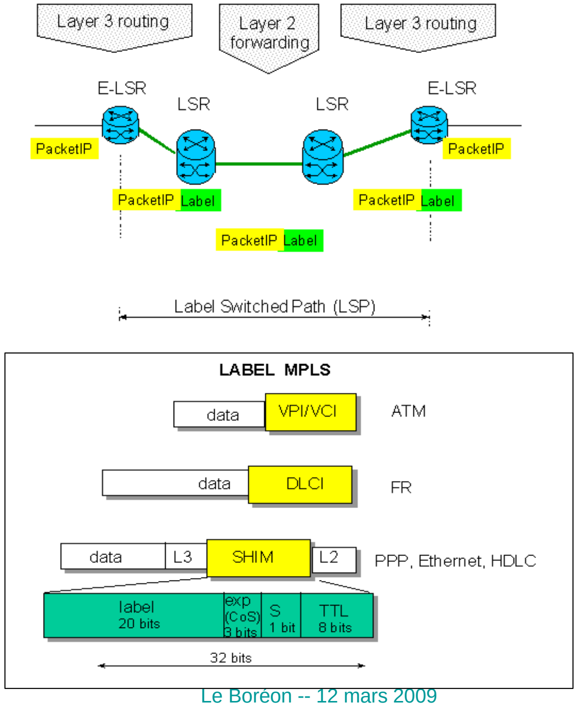
InPort	InLabel	OuPort	OutLabel
P1	L5	P3	L21
P2	L7	P4	L15

1 - Paquet IP + Label arrive sur le LSR
2 - Protocole de routage détermine, à partir de la LIB le next hop LSR ou E-LSR
3 - Mise à jour du label MPLS
4 - Paquet IP + Label envoyé vers le noeud suivant

A la sortie du domaine MPLS



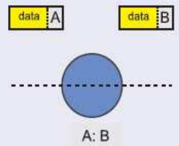
1 - Paquet IP + label arrive sur l'egress node
2 - Retrait du label et transmission du paquet IP à la couche réseau



Label stacking

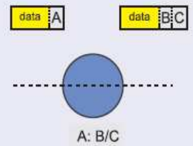
Packets contain a stack of labels...
Nodes may alter the stack of labels performing one of the following operations:

Swap the Top



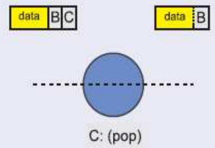
... with the neighbor's locally 'understandable' label.

Swap & Push



... as many as you want.

Pop the Stack

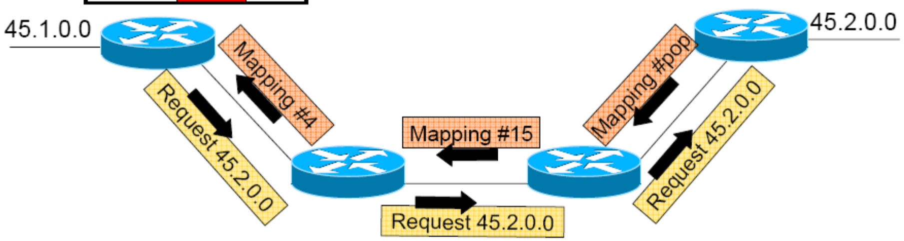


... only one.

Distribution de Labels

@ dest	Label out	If Out
45.2.0.0	#4	2

@ dest	If In	Label in	If out
45.2.0.0	2	# pop	1



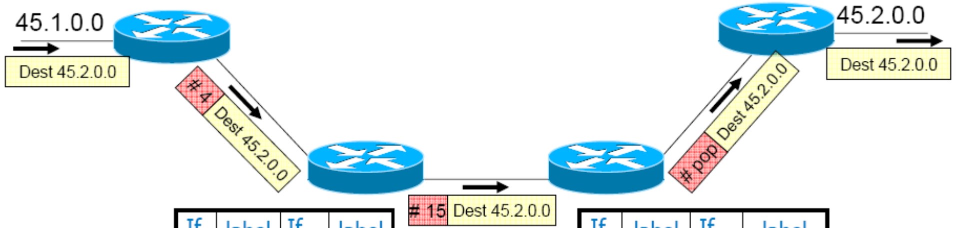
If In	label in	@ dest	If Out	label out
2	# 4	45.2.0.0	1	# 15

If In	label in	@ dest	If Out	label out
2	# 15	45.2.0.0	1	# pop

Commutation de labels

@ dest	Label out	If Out
45.2.0.0	#4	2

@ dest	If In	Label in	If out
45.2.0.0	2	# pop	1



If In	label in	If Out	label out
2	# 4	1	# 15

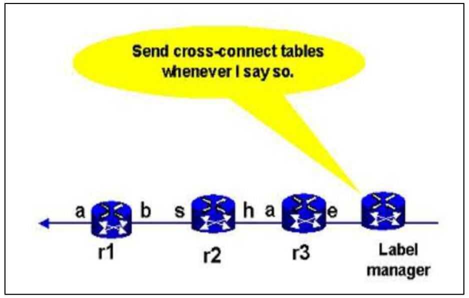
If In	label in	If Out	label out
2	# 15	3	# pop

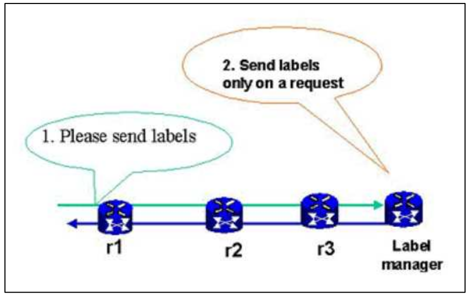
Distribution de Labels

- Permet de s'assurer que le mapping FEC label est bien cohérent dans tous les routeurs.
- 2 protocoles de distribution, compatibles IPv4 et IPv6:
  - CR-LDP: Constraint-based Routed Label Distribution Protocol
  - RSVP-TE: ReSerVation Protocol – Traffic Engineering

Attribution de Labels : LDP (Label Distribution Protocol)

- Downstream Unsolicited (DOU)
- Downstream On Demand (DOD)





RSVP-TE

RSVP-TE permet de :

- Créer un LSP le long d'une route explicite
- Etablir un LSP en distribuant des labels
- Définir les besoins en bande passante des liens qui forment un LSP

RSVP-TE

Path message

(Session, label request, ERO, RRO, Session attribute)

RESV message

(Session, label, style, RRO)

Session: LSP transportant de l'IPv4 ou IPv6 ?

ERO: EXPLICIT_ROUTE (par quels LSR veut-on passer?)

RRO: ROUTE_RECORD (quels LSR ont été traversés?)

Style: quelle sorte de réservation? Avec QoS sur le lien? Etc.

Session attribute:LSP_TUNNEL_RA ou LSP tunnel

GMPLS

Extension de MPLS-TE qui permet aux LSR de supporter plusieurs types de commutations:

– Paquets

– TDM (SDH/SONET)

– Lambdas

– Fibres…

• Généralisation de la définition d'un label, qui peut être un slot, une fibre, un lambda, une bande de longueurs d'onde ou un nombre ajouté au paquet.

Conséquences : extensions de CR-LDP et RSVP-TE, ajout de LMP (Link Management Protocol) pour la gestion des liens et des erreurs.

GMPLS

1 seul plan de contrôle pour plusieurs sortes de commutation

• Réutilisation :

– du paradigme de la commutation de labels

– des protocoles et mécanismes de MPLS-TE

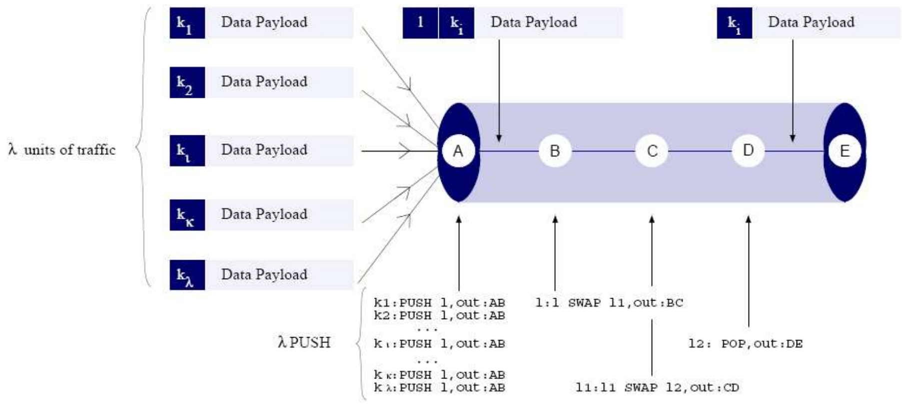
– du modèle de routage et d'adressage IP

Les applications de MPLS

- **Ingénierie de trafic** : diriger le trafic via un chemin spécifique. Pour Distribution optimale du trafic et améliorration de l'utilisation globale du réseau.
- **Bande-passante garantie :** allouer des largeurs de bande passante et des canaux garantis. Organiser le trafic 'prioritaire' et 'au mieux' (voix, données…)
- **Reroutage rapide** : reprise rapide après la défaillance d'une liaison ou d'un noeud. Empêcher l'interruption des applications utilisateur et la perte de données.
- **VPN MPLS : l**orsque le nombre de routes et de clients augmente, les VPN MPLS peuvent facilement monter en charge, tout en offrant le même niveau de confidentialité que les technologies de niveau 2.
- **La fonction Classe de service** (CoS) MPLS assure que le trafic important est traité avec la priorité adéquate sur le réseau et que les exigences de latence sont respectées.

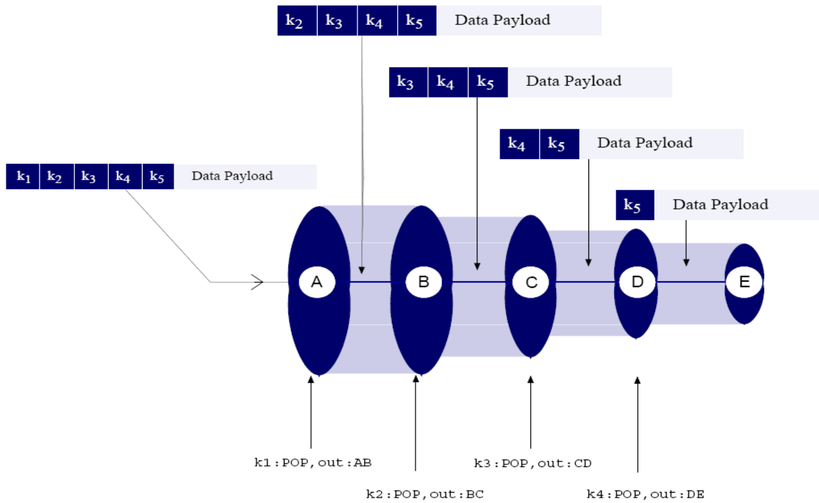
Notre problème : Label Space Reduction (LSPR)

Cost of a tunnel



Cost of the tunnel = #Units of traffic + length (tunnel)
= λ + 3

Label Stripping



Hypothesis

- Traffic can enter in any node of the tunnel, but can exit only at the end of the tunnel
- Label stacks of size 2
- Shortest paths used for the routing

LSPR Problem

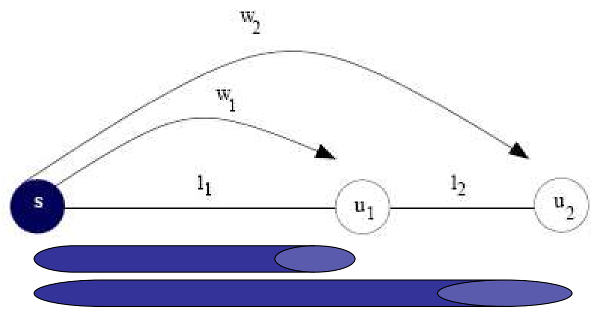
Label Space Reduction in a GMPLS Network: LSPR

INPUT: a network (digraph) $G = (V, E)$ and a set of requests $\mathcal{R}$, where in the request $r \in \mathcal{R}$, $r = (s_i, u_j)$, $s_i \in V$ sends w_r units of traffic to $u_j \in V$.

OUTPUT: A set $\mathcal{T}$ of tunnels enabling to route the traffic and a dipath composed of tunnels in $\mathcal{T}$ for each request (s_i, u_j).

OBJECTIVE: minimize the total cost of $\mathcal{T}$, that is $c(\mathcal{T}) = \sum_{T_k \in \mathcal{T}} c(T_k)$ where the cost $c(T_k)$ of a tunnel T_k which contains λ_k units of traffic and is of length $l(T_k)$ (number of arcs in G associated to the path joining the end-vertices of T_k) is $c(T_k) = \lambda_k + l(T_k) - 1$.

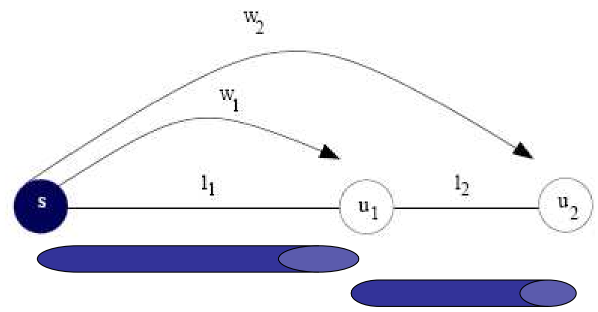
A solution on a simple example



If $l_1 \leq w_2$, the solution is composed of tunnels (s,u_1) and (s, u_2)

Cost = (w1 +l1 −1)+(w2 +l1 + l2 − 1) = w1 + w2 + 2l1 + l2 − 2

A solution on a simple example



If $l_1 \leq w_2$, the solution is composed of tunnels (s,u_1) and (s, u_2)

Else, the solution is composed of tunnels (s,u_1) and (u_1,u_2)

Cost = (w1+w2+l1−1)+(w2+l2−1) = w1 + 2w2 + l1 + l2 − 2

Cost of a solution

$$c(\mathcal{T}) = \sum_{k=1}^{|\mathcal{T}|} (l(T_k) - 1) + \sum_{r=1}^{|\mathcal{R}|} h_r w_r.$$

- hr is the number of consecutive tunnels for the request r in T,
- wr is the number of units of traffic of the request r and
- l(Tk) is the length of the tunnel Tk in terms of number of hops.

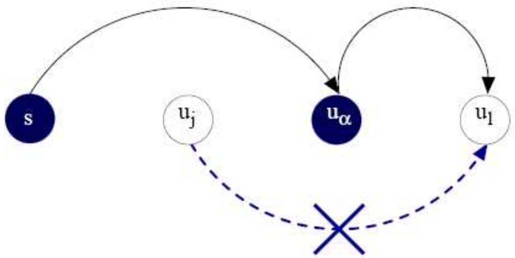
The cost c(T) is the sum of (1) the cost for the configuration of the tunnels and (2) the cost for the requests to enter the tunnels

General network

- In any network, there exists an optimal solution for LSPR problem s.t. all the units of traffic of requests (si,uj) are routed via a unique dipath (set of consecutive tunnels) from si to uj

The line network, one source

- When the network is a directed line, with source s, an optimal solution T for LSPR-L1 problem is such that, if (s, u_α) is the longest tunnel from s, then there is no tunnel (u_j, u_l) in T with $j < \alpha < l$.



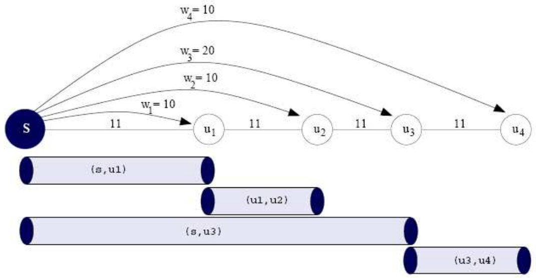
C is a table of size n^2 indicating the costs all the sub-solutions;
S is a table of size n^2 indicating the splitting points u_α associated to the optimal sub-solutions;
W is a table of size n storing partial sums of weigths, $W[0] = 0$, $W[j] = \sum_{i=1}^{j} w_i = W[j-1] + w_j$, and so $\sum_{i=\alpha}^{\beta} w_i = W[\beta] - W[\alpha - 1]$;

for $i \in [0, n]$ **do**
- $C[u_i, u_i] = 0$;
- $C[u_i, u_{i+1}] = w_{i+1} + l(u_i, u_{i+1}) - 1$;
- $S[u_i, u_{i+1}] = u_{i+1}$;

for $i \in [2, n]$ **do**
- **for** $\forall k \in [0, n-i]$ **do**
  - $min = +\infty$;
  - **for** $\forall \alpha \in [k+1, k+i]$ **do**
    - $value = (W[k+i] - W[\alpha - 1]) + l(u_k, u_\alpha) - 1 + C[u_k, u_{\alpha-1}] + C[u_\alpha, u_{k+i}]$;
    - **if** $value < min$ **then**
      - $min = value$;
      - $C[u_k, u_{k+i}] = c(P_{u_k \to u_{k+i}}) = value$;
      - $S[u_k, u_{k+i}] = u_\alpha$;

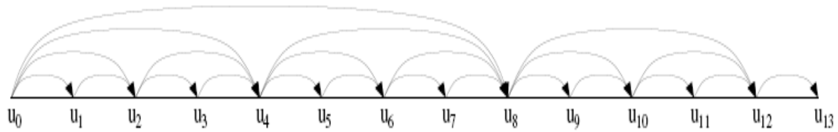
Compute the optimal set of tunnels from the table S;

Solution computed with dynamic programming algorithm



	$s = u_0$	u_1	u_2	u_3	u_4
$s = u_0$	0	20	50 (u_1)	101 (u_2)	132 (u_3)
u_1	-	0	20	61 (u_3)	91 (u_3)
u_2	-	-	0	30	60 (u_3)
u_3	-	-	-	0	20
u_4	-	-	-	-	0

log(n)-approximation



Proof. The cost of a solution computed by the algorithm is (1) the cost of the configuration of the tunnels plus (2) the cost for entering the tunnels.

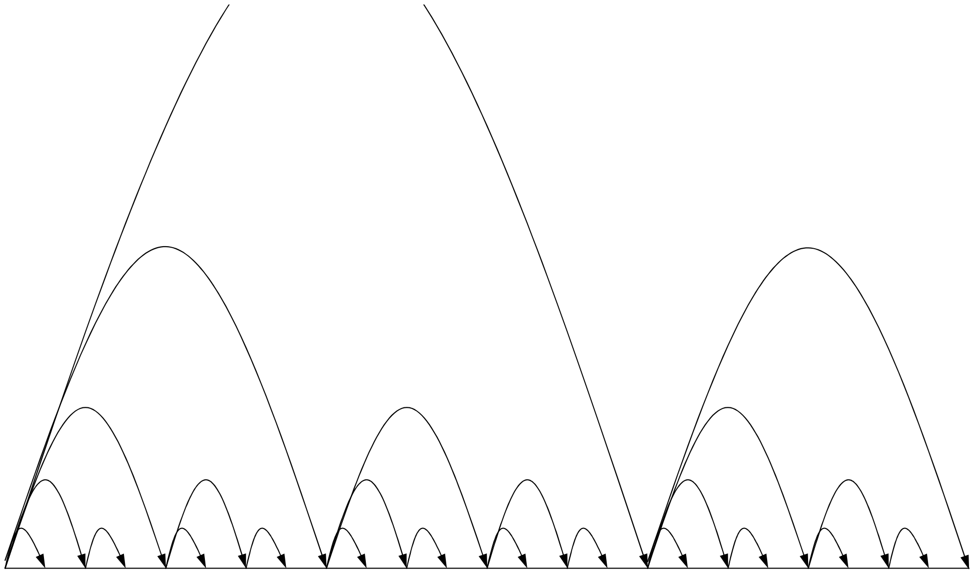
To configure each level of consecutive tunnels, at most $L-1$ labels are needed. There are at most $\log(n)$ different levels of tunnels. So, the overall number of labels needed for the configuration of tunnels is at most $(1) \leq (L-1)\log(n)$.

When that set of tunnels has been configured, any source can join any destination in at most $\log(n)$ hops. Therefore, the total cost needed to enter the tunnels is at most $(2) \leq \sum_{r=1}^{|\mathcal{R}|} w_r \log(n)$.

Then, the cost of this solution is at most: $(1) + (2) \leq \sum_{r=1}^{|\mathcal{R}|} w_r \log(n) + (L-1)\log(n) = \log(n)(\sum_{r=1}^{|\mathcal{R}|} w_r + L - 1)$.

In the best case, an optimal solution will be of cost $\sum_{r=1}^{|\mathcal{R}|} w_r + (L-1)$, giving a $\log(n)-$approximation. □

Less tunnels are needed for simulations

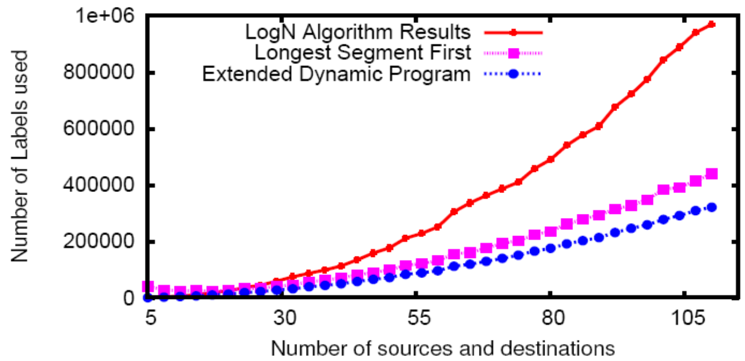


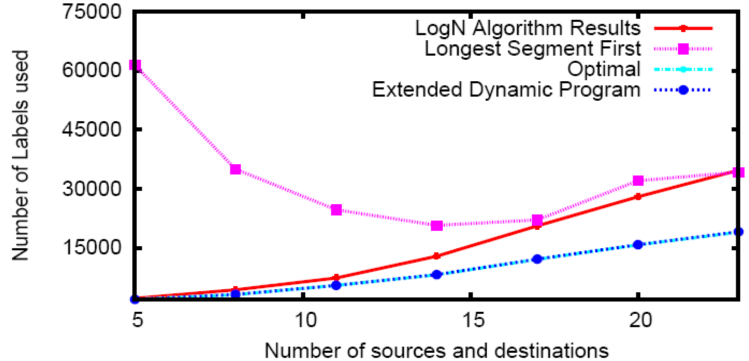
log(n)-approximation

- log(n)-approximation for the tree and for the general network.

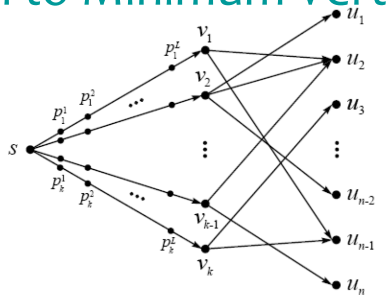
Parameters of the simulations

- Line network consisting of 500 nodes.
- Each experiment consists of a different number of sources and destinations.
- The number of sources equals the number of destinations in each experiment.
- Between a pair of source and destination nodes, a demand is generated (with a probability of 80%) with a random capacity between one and 500 (uniform).





Reduction to Minimum Vertex Set-cover



The LSPR problem cannot be approximated within a factor Clogn for C>0 even if the instance is a partial boradcast unless P=NP.

To a set cover instance with sets S1, S2,…,Sk, with Si ⊆{a1,a2,…an}

- We start with a distinguished node s
- For each set Si, we introduce a node vi and a directed path of length L+1 from s to vi through L new vertices
- For each element aj, we introduce a new vertex uj and for each vertex vi we add the arcs (vi,uj) if aj ∈Si
- The requests are from s to uj, for i=1,…,n.

Reduction to Minimum Vertex Set-cover

- If OPT is the optimal cost of LSPR and OPT-VC is the optimal cost of MIN-Vertex Set Cover
- Any cover induces a solution of our LSPR

$$L \cdot OPT_{VC} + n \leq OPT \leq L \cdot OPT_{VC} + 2n.$$

For symmetric graphs

- The LSPR problem for symmetric graphs is APX-Hard even if the instance is a partial broadcast. Therefore, it does not accept a PTAS unless P=NP.
- Minimum Steiner Tree

In AOLS Network

- All-optical forwarding without OEO conversion
- One hardware device for each label processed in the node
- What if the network is already designed and the number of devices available is given and limited per node?

Conclusion

- Proof of NP-Completeness when LSPs for requests are given
- Algorithms for trees, meshs, general graphs…
- If paths are not directed? (i.e., we can go backwards)
- What if label stacks are larger than 2…
- Alcatel – routage + groupage