Prototyping the GMPLS UNI Implementation for End-to-end LSP Re-routing

by D.Verchere (dominique.verchere@alcatel.fr)
and D.Papadimitriou (dimitri.papadimitrou@alcatel.be)

MPLS 2004

ALCATEL

Table of Content

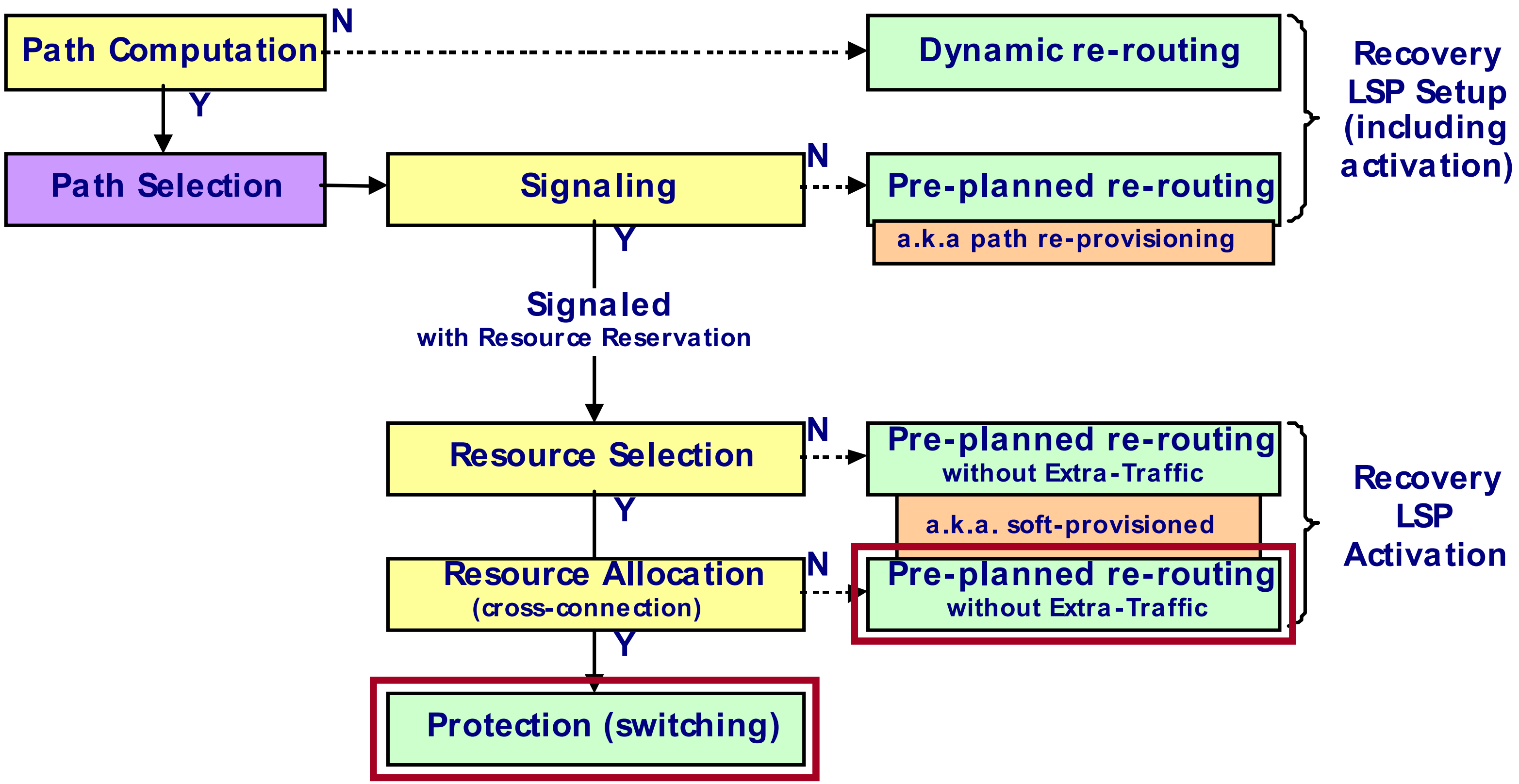
MPLS 2004

ALCATEL

Introduction

- Better collaboration between IP/MPLS and SONET/SDH layers implies overcoming drawbacks of independent recovery mechanisms
  - without coordination of the recovery actions at both layers, outcome of recovery procedure is unpredictable
  - a fallback recovery mechanism at a different layer can be triggered quickly in case of failure of the recovery attempt
  - since each layer governs independently its recovery resources, the overall resource utilization is not optimal
- Efficient coordination between actions taken by the different layers must be delivered during both provisioning and recovery phases

Introduction: Recovery Mechanisms



MPLS 2004

ALCATEL

Introduction

- Realization in the overlay model context
  - Interactions between networks through a User-to-Network Interface (UNI)
  - Note: recovery mechanisms equally applicable to augmented and unified control plane models but the tighter the integration b/w control planes, the easier the implementation of multi-layer recovery (since avoiding specialization allows easier coordination & faster processing)
- For the overlay model
  - OIF UNI v1.0: opaque and operationally complex, very restricted set of capabilities, and limited extensibility $\Rightarrow$ imposes severe restrictions on multi-layer recovery mechanisms
  - GMPLS UNI: enhanced interface developed by IETF
    - Addresses shortcomings of the OIF UNI
    - Provides capabilities required in support of multi-layer recovery

MPLS 2004

ALCATEL

GMPLS UNI Positioning

OIF Model Limitations (1)

- Separated routing instances between technology domains with
  - No physical end-point reachability information exchange
  - No (a priori) routing adjacencies between source and destination client
  - No address resolution between logical/physical end-points

  ⇒ Requires an out-of-band mechanism to bootstrap the system
- Per layer Traffic Engineering (per layer TEDB)
  - Only allows for manual triggering of connections

  ⇒ Default operational model is provisioned
- Network address allocation (network-to-client) does not allow for dynamic client learning of reachable end-points

  ⇒ Requires address resolution (logical to physical) for switched connections or (physical to logical) for soft-permanent connections, in turn, this precludes best exit point selection for multi-homed clients

OIF Model Limitation (2)

- Network unaware of client-initiated connection semantic
  - Usually part of client layer control plane topology (client control plane performance is strongly correlated to the data plane performance)

  ⇒ No distinction between a protecting versus protected connection
  - Impossible to provide soft-reservation of network capacity

  ⇒ This precludes any client-driven end-to-end re-routing mechanism
- Strict separation of the signalling domains, in turn, requires split of a single end-to-end RSVP session in (at least) 2 sessions for a single end-to-end connection

  ⇒ Operational limitations in using OIF UNI model (limits the gain of GMPLS-based control plane usage)

MPLS 2004

ALCATEL

Breaking the End-to-end Principle

- Split into multiple RSVP sessions per connections:
  - Increases the number of RSVP sessions to be managed per end-to-end connection: N+1 instead of 1 (for N sub-networks)
  - Decreases <u>interoperability level</u> and therefore one of the KEY objectives for carriers to adopt an IP-based distributed control plane
  - Does not increase flexibility or provide more features while creating <u>divergence</u> wrt to RFC 2205, RFC 2210, RFC 2961, RFC 3209, RFC 3473, and RFC 3477
  - Implies additional processing at sub-networks edges (as Tunnel endpoint address ≠ connection destination ⇒ additional look-up to a "non-associated" object) that substantially impacts performance
  - Error control/troubleshooting impossible since (IF_ID_)ERROR_SPEC objects carried as part of RSVP Sessions to which they are not associated
  - Vendors to implement Inter-working Functions (IWF) that in turn <u>increases carriers' OPEX</u>

MPLS 2004

ALCATEL

GMPLS UNI Signaling – Main Features (1)

- End-to-end RSVP sessions

  ⇒ Simplifies implementation (compliance with GMPLS RSVP-TE, no Generalized_UNI object) and software maintenance
- Error reporting
  - Since GMPLS UNI does not break the end-to-end principle, failures and other errors occurring at the destination UNI can be reported without loosing any capability to correlate this information with affected end-to-end connection(s)
- Reachable end-points are numbered (IPv4/IPv6) or un/numbered TE link identifiers

  ⇒ No NSAP support (<u>note</u>: ATM edges devices do not require NSAPs for control plane end-point identification)

GMPLS UNI Signaling – Main Features (2)

- Allows for Fast Notification mechanisms and keeps exact semantic of recovery connections through the network
  - Better collaboration between "domains" in terms of resource consumption and recovery speed using bulk recovery and soft-provisioning (no resource allocation for recovery connections throughout the optical network)
  - Capability to request and receive Notify messages (aggregating multiple LSP failures) that timely trigger any recovery action

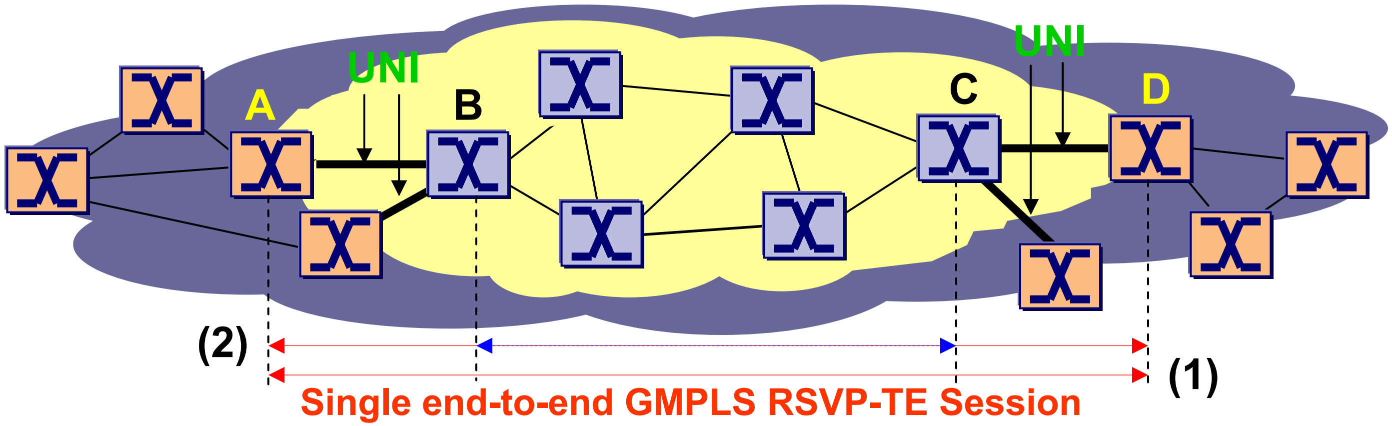
  ⇒ GMPLS UNI delivers crucial mechanisms that OIF UNI is incapable of
- Facilitates the delivery of "diversely routed" connections
  - Using explicit exclusion mechanism (eXclusion Route Object – XRO)

  ⇒ Mechanisms defined for multi-domain networks can be re-used in the overlay context

GMPLS UNI Signaling – Main Features (3)

- Explicit routing (client node-driven)
  - (typically loose routing in this context) that in turn provides built-in explicit label control capabilities
  - GMPLS UNI efficiently uses the signaling capabilities already delivered by [RFC 3473] without requiring additional extensions that specialize the signaling interface
- Route recording (client node-driven)
  - Possibility for client LSRs to diagnose connections they have initiated
  - Flexibility in using "feedback" information through RECORD_ROUTE object ⇒ edge client nodes learning process

GMPLS UNI: Contiguous and Stitching



Contiguous (1)

Path message w/ EXPLICIT_ ROUTE:

- Sender address: IP source address A
- Tunnel address: IP dest. address D

LSR ERO/RRO Processing:

- ERO: ingress core node (B strict) and egress edge node (D loose)
- B computes path to reach node D, append to the ERO included in outgoing Path message

LSP Stitching (2)

Path message w/o EXPLICIT_ROUTE:

- Sender address: IP source address A
- Tunnel address: IP dest. address D

OXC ERO/RRO Processing

- ERO: ingress core node (B strict), egress core node (C loose) and egress edge node (D loose)
- B computes path to reach C and include ERO in outgoing Path message

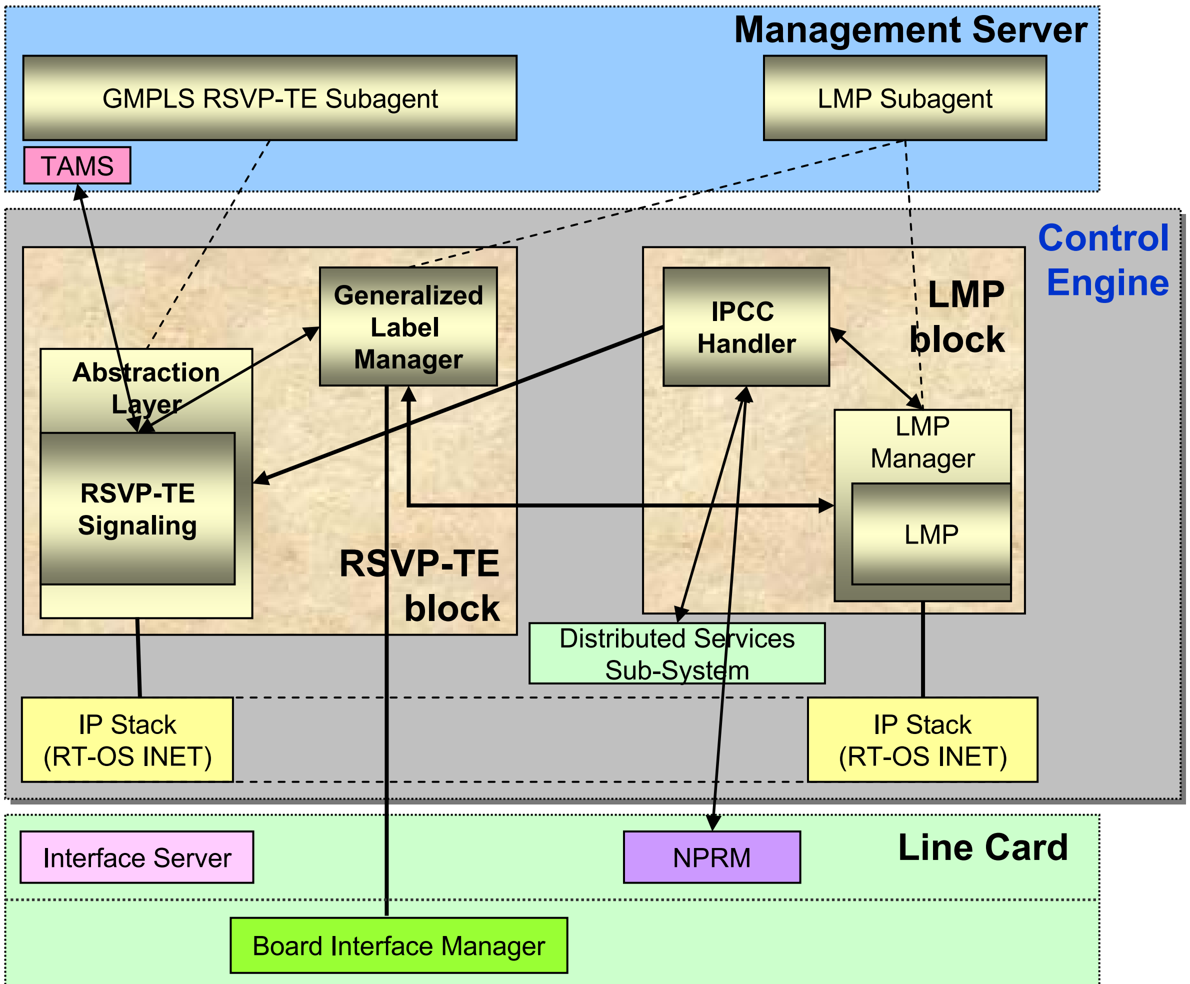
Comparison between Signaling Interfaces

Network Model	OIF UNI	GMPLS UNI	GMPLS Unified (Integrated)
Signaling	Direct and Indirect	Direct	Direct
Symmetry Scope	Asymmetrical Local	Asymmetrical End-to-end	Symmetrical End-to-end
Routing protocol	None	None / Optional	Link state preferred
Routing information	None	Network exchanges (based on policy) of end-point reachability information with client nodes is allowed	Reachability and traffic engineering information
Address space (client/network)	Must be distinct	May be common	Common
Discovery	Optional and only local	Optional and may be global	Through routing and global
Security Cooperation	No trust None	Limited trust Limited (*)	High trust Full
	Signaling must be domain specific (a separate signaling protocol instance must be running in the network)	End-to-end signaling that may be domain specific (a separate signaling protocol instance may be running in the network)	Inherently multi-layer capable (so also referred to as end-to-end integrated signaling)

(*) but suitable for end-to-end LSP re-routing

GMPLS UNI Implementation

GMPLS UNI Software Architecture



TAMS: Traps & Alarms Management Server

IPCC: IP Control Channel

LMP: Link Management Protocol

NPRM: Network Processor Resource Manager

Description RSVP-TE Block

- Generalized RSVP-TE
  - View upon all the data links (component links or ports) and channels
  - Maintains tables of originating and terminating LSPs
- Generalized Label Manager
  - Allocates and verifies labels (TDM data bearing links and channels) per LSP tunnel based on
    - Originating: constraints of the incoming request (e.g. bandwidth)
    - Terminating: traffic parameters of the incoming request (SENDER_TSPEC in Path msg) vs previous hop label selection
- IPCC Handler
  - Selects primary IPCC to be used by RSVP-TE and LMP
  - Handles selection of a different IPCC in case of CC failure
  - Keeps relationship between IPCCs and LMP neighbors
  - Filtering to forward incoming RSVP and LMP messages to the Control Engine and stop routed traffic from entering IPCCs

MPLS 2004

ALCATEL

Description LMP Block

- LMP is the owner of operational data of data (bearing) links, TE links, IPCCs, LMP Neighbors
- LMP is responsible for maintaining the following Finite State Machines (FSMs)
  - IP Control Channel (IPCC)
    - keeps the state of each IPCC
    - owner of IPCC related operational data
    - IPCC related procedures (CC configuration and maintenance) will be performed autonomously by LMP application.
  - TE and data links
    - Link Property Correlation: performed at initialization and on change of configuration of TE Links (by periodical re-initiation of Link Property Correlation), reply on neighbor requests
    - Link Verification

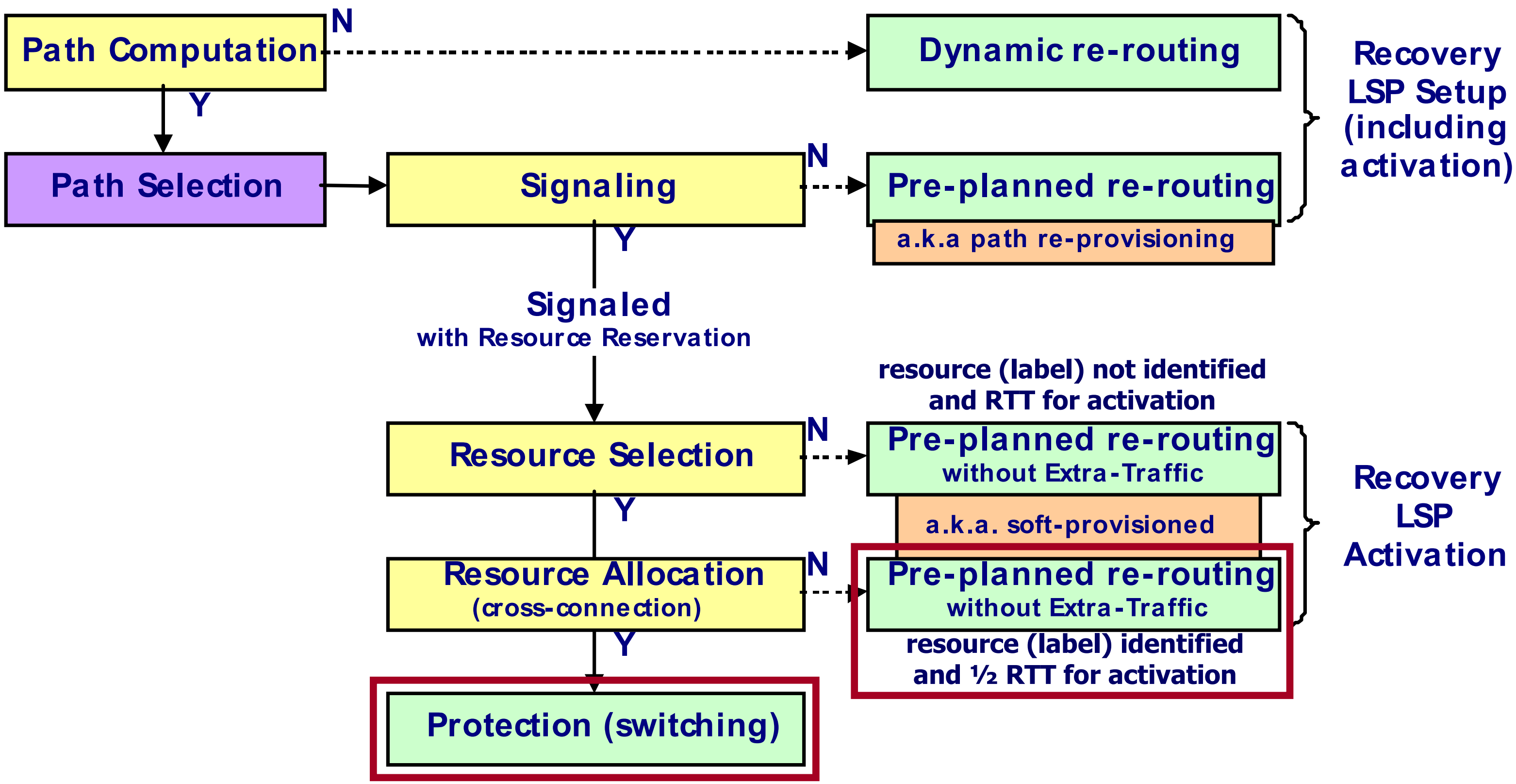
MPLS 2004

ALCATEL

Recovery Mechanisms Implemented (1)

- Pre-planned end-to-end re-routing without extra-traffic
  - Hard-provisioned working LSP and soft-provisioned protecting LSP (with label resource selection)
  - Failure detection/notification handled by the adjacent OXC
    - Fast notification (Notify message) towards ingress/egress LSR
  - Upon Notify message reception, the ingress client LSR activates the soft-provisioned protecting LSP
- End-to-end protection with extra-traffic
  - Hard-provisioned working LSP and hard-provisioned protecting LSP
  - Failure detection/notification handled by the adjacent OXC towards the Ingress LSR
    - Fast notification based on Notify message
  - Upon Notify message reception, the ingress client LSR redirects normal traffic into the protecting LSP (data plane switchover)

Recovery Mechanisms Implemented (2)



MPLS 2004

ALCATEL

End-to-end Recovery Objects

- Protection object (Path message)

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Length            | Class-Num(37) |   C-Type (2)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|P|N|O| Reserved  | LSP Flags |      Reserved     | Link Flags|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Reserved                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

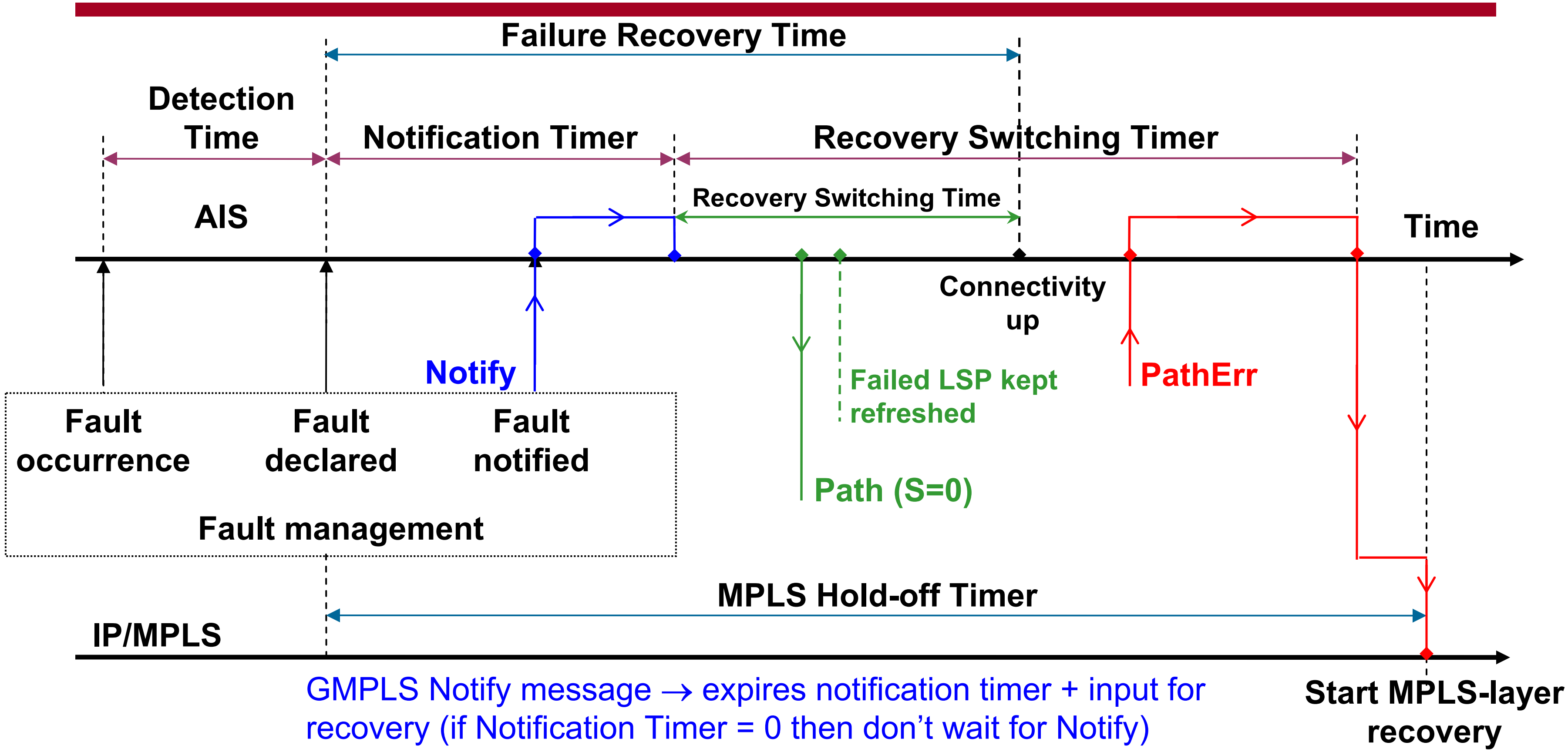
- Association object (Path message)

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Length            | Class-Num(198)|   C-Type (1)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Association Type (= 1)     |         Association ID        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    IPv4 Association Source                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

In addition to existing [RFC 3473] objects and messages:

- Notify_Request object
- Admin_Status object
- IF_ID_Error_Spec object
- Notify message

LSP Re-routing : Timers (Ingress)



GMPLS Notify message → expires notification timer + input for recovery (if Notification Timer = 0 then don't wait for Notify)

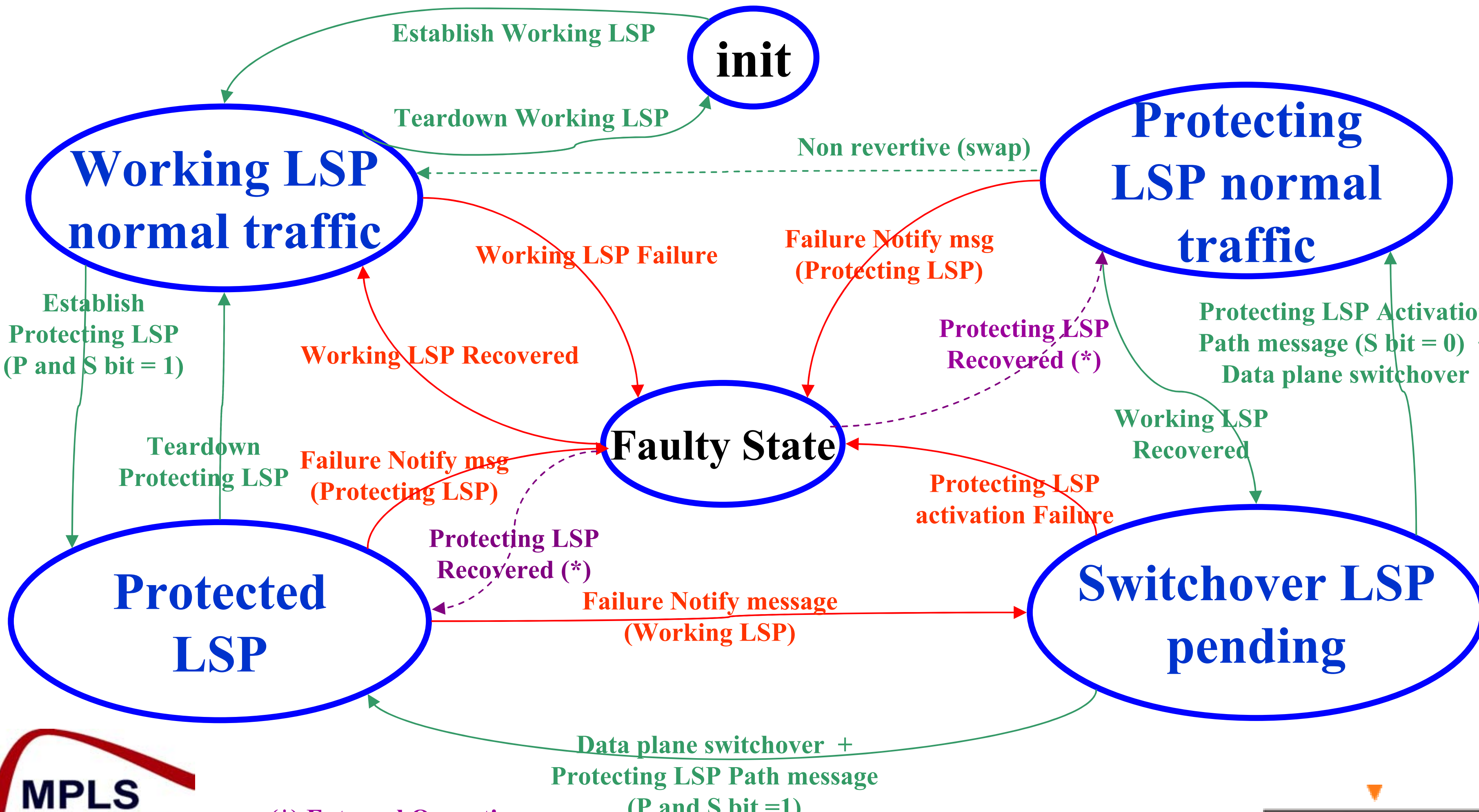
Protecting LSP Path message (S = 0) – connectivity up expires MPLS Hold-off timer

Protecting LSP PathErr message → expires recovery switching timer + trigger MPLS layer recovery

MPLS 2004

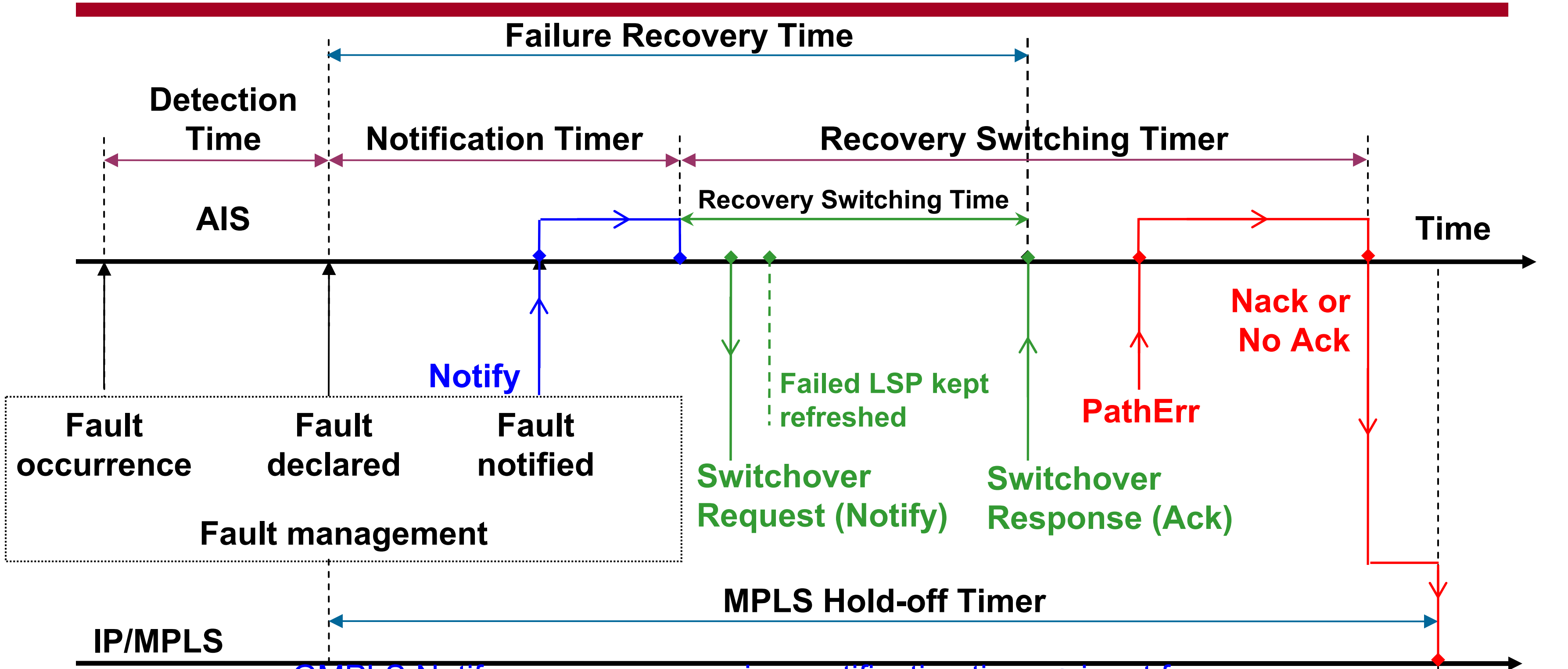
ALCATEL

LSP Re-routing: State Machine (Ingress)



(*) External Operation

Protection with Extra-Traffic : Timers

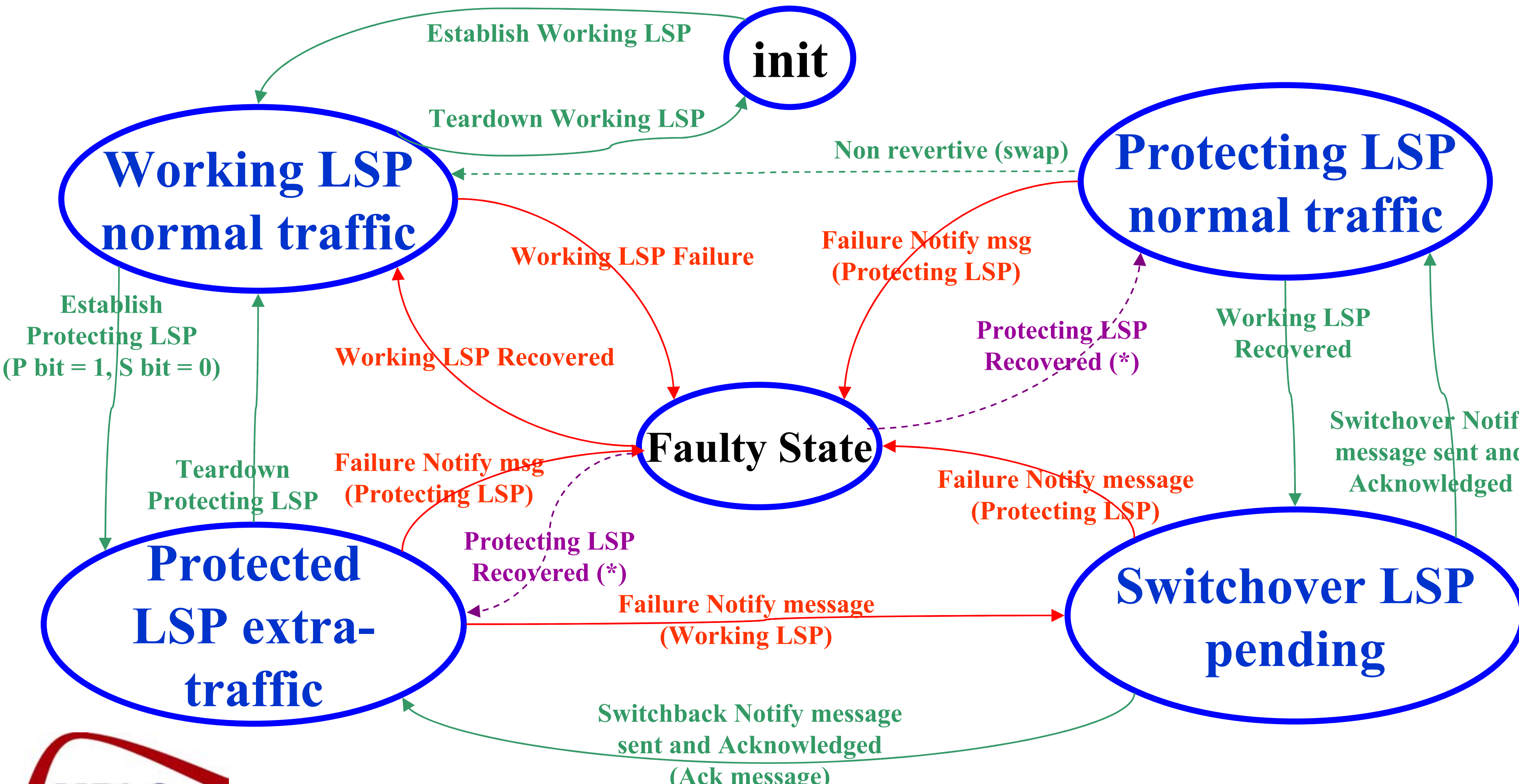


GMPLS Notify message → expires notification timer + input for recovery (if Notification Timer = 0 then don't wait for Notify)

Switchover Request + Response (Ack) → expires MPLS Hold-off timer

Protecting LSP PathErr message (or Nack or No Ack b/f expiration) → expires recovery switching timer + trigger MPLS layer recovery

Protection with Extra-Traffic : State Machine (Ingress)

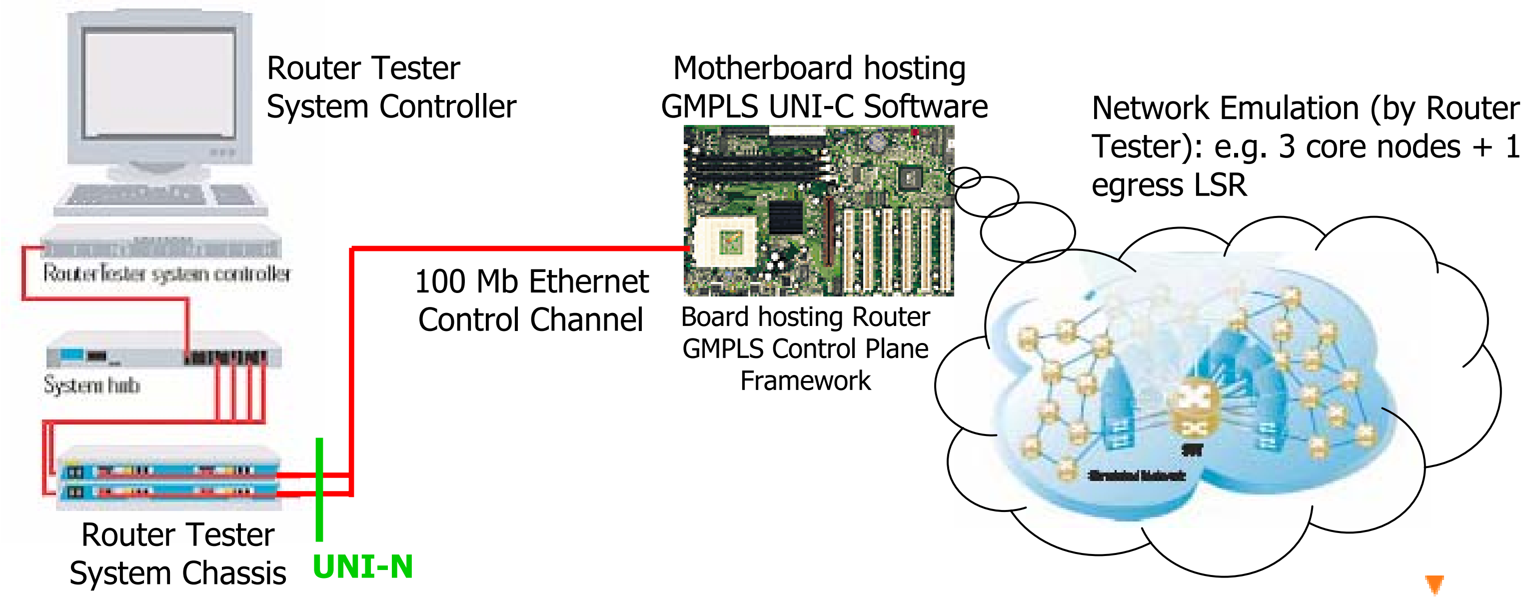


(*) External Operation

Testbed and Performance Gain

Testbed Configuration

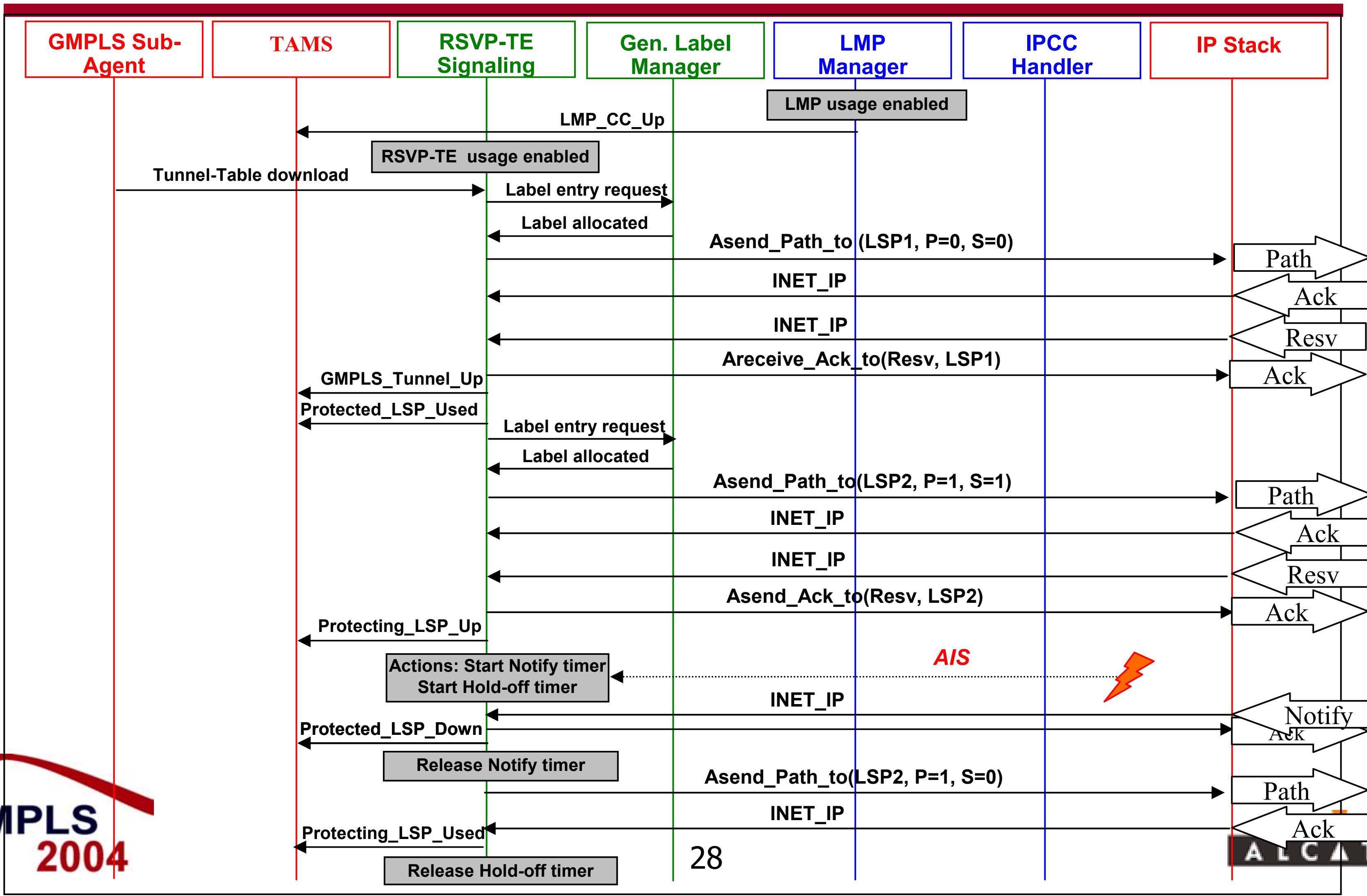
- GMPLS UNI (RSVP + LMP) hosted on dedicated board
- Router tester emulates: edge core node, network and egress client LSR
- Ethernet 100Mb Control Channel



MPLS 2004

ALCATEL

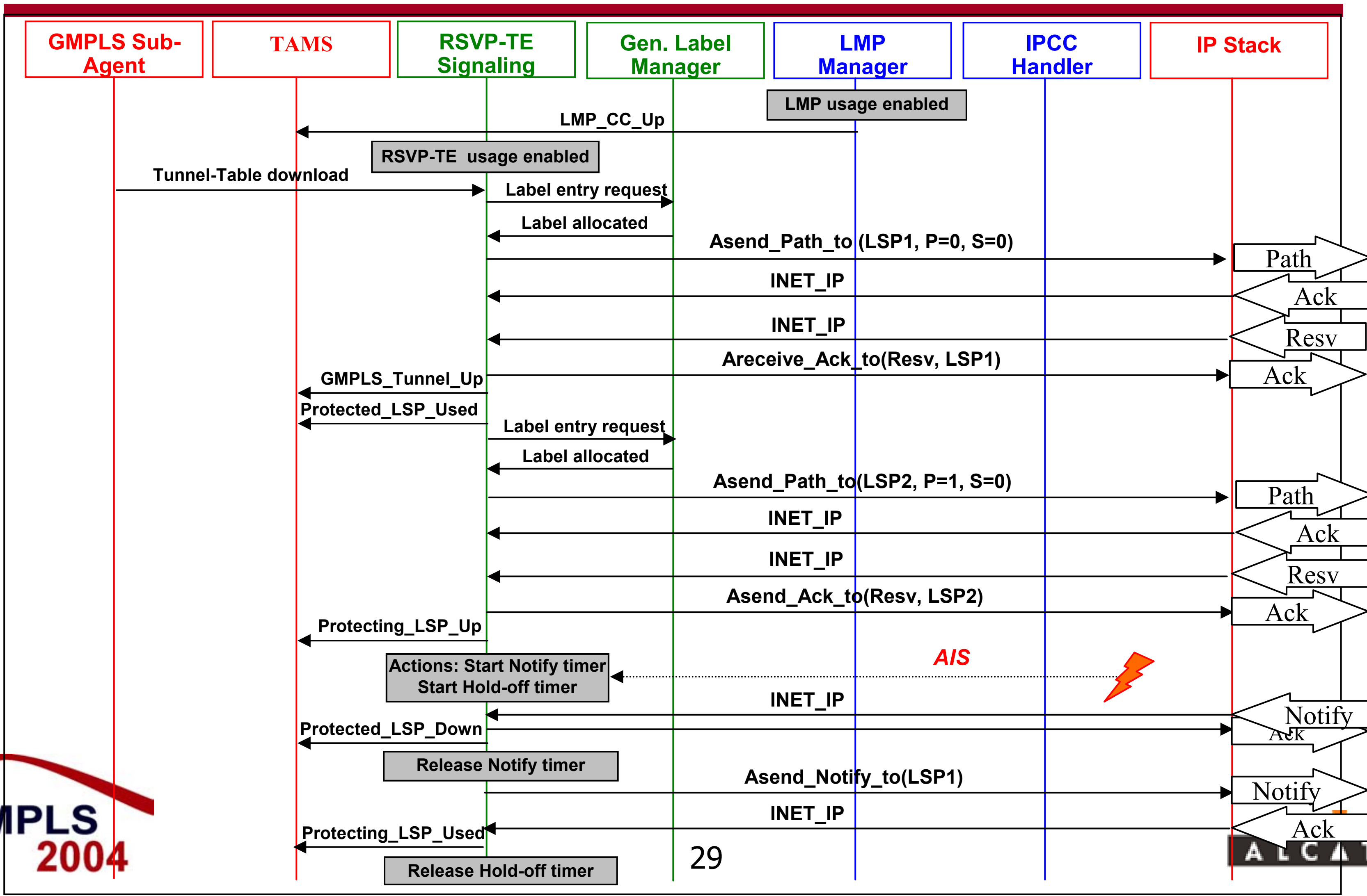
LSP Re-routing : Sequence Diagrams



MPLS 2004

ALCATEL

LSP Protection with Extra-Traffic: Sequence Diagrams



MPLS 2004

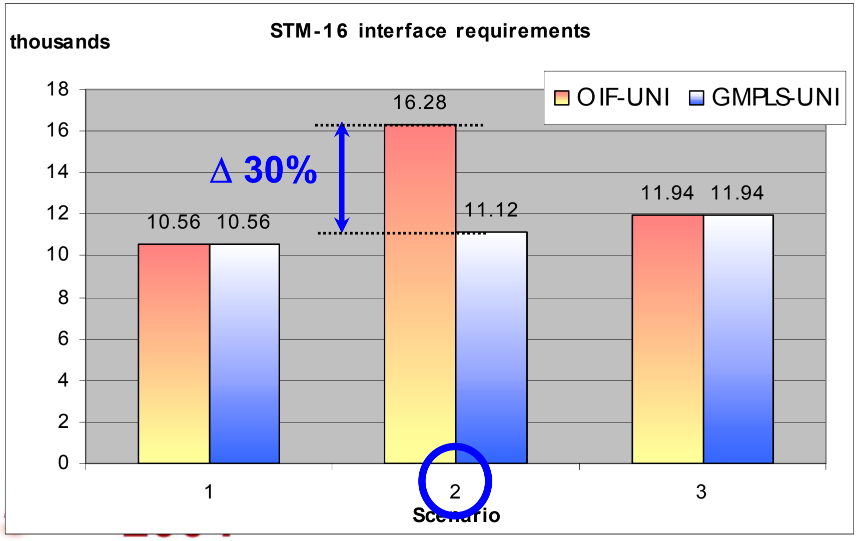
ALCATEL

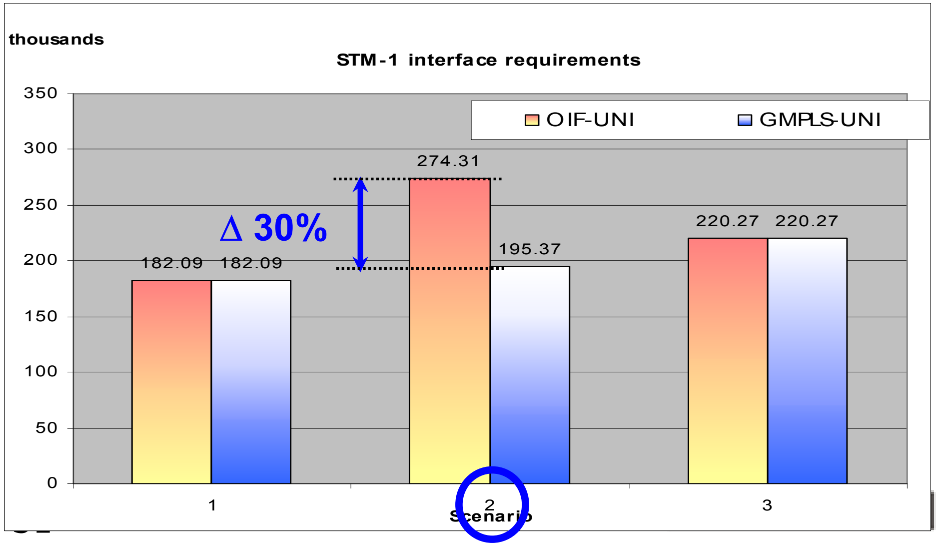
Recovery Mechanisms Comparison

	1. SONET/SDH LSP Segment Recovery	2. SONET/SDH LSP End-to-end Recovery	3. Packet LSP Local Recovery: N-HOP FRR
Responsible layer for recovery	Circuit (SONET/SDH)	Circuit (SONET/SDH)	Packet (IP/MPLS)
Recovery resource / router driven extra traffic or sharing of recovery resources	Soft-provisioned SONET/ SDH LSP segments (with resource sharing) / No possibility for client LSR driven extra-traffic	Soft-provisioned SONET/ SDH end-to-end LSP (resource sharing) / Possibility for LSR driven extra traffic (low priority	Bandwidth protection using shared detour LSP nested into pre-established SONET/SDH end-to-end LSP
Recovery switching initiating entity	Edge OXC	Client LSR	(Client) LSR
Link protection between client LSR & edge OXC	Dedicated mechanism	Inherent	Inherent
Protection of edge devices	No	No (edge OXC protection in case of dual homing)	Yes (except for source/ destination LSRs)
Complexity of implementation	Low / Intermediate (if resource sharing capable	Intermediate	High
Resource efficiency	Low (if no OXC driven extra-traffic) Intermediate (if resource sharing between recovery LSP segments)	Intermediate High (if good level of protection resources sharing)	Low (if no LSR driven extra-traffic) Intermediate (if resource sharing between detour LSPs)
Recovery granularity (assuming packet LSP granularity lower than SONET/SDH LSP)	Coarse (per SONET/SDH LSP)	Coarse (per SONET/SDH LSP)	Fine (per packet LSP)

GMPLS UNI – Resource Performance

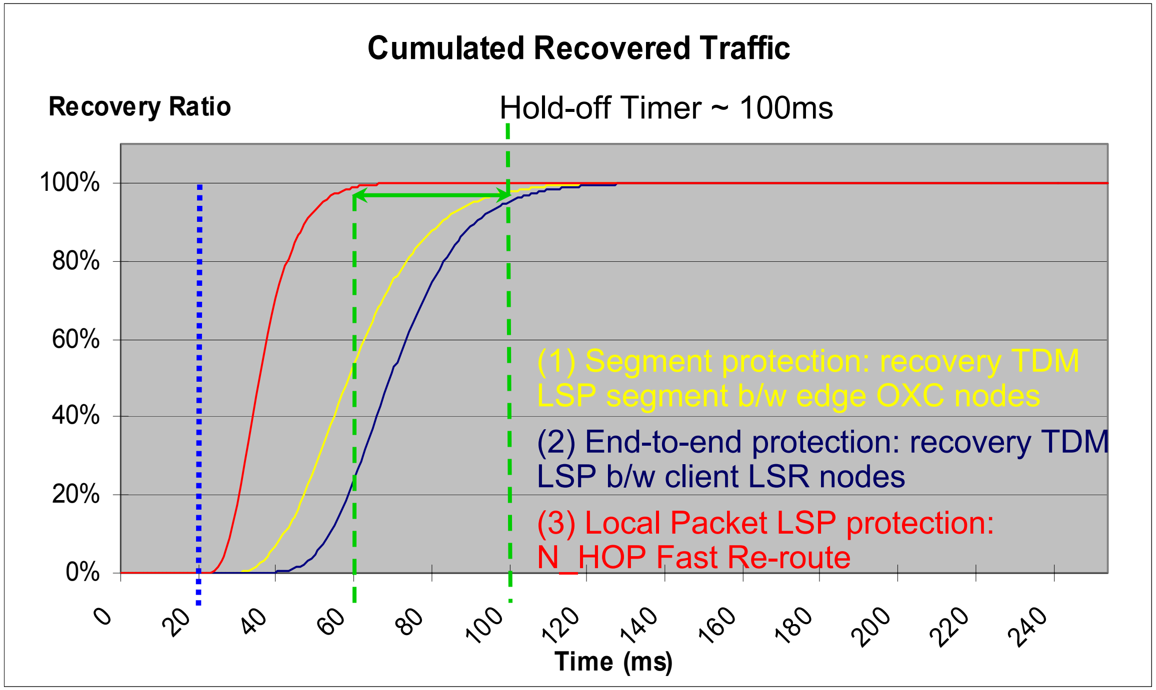
- Improvement of resource usage efficiency: difference of 30% wrt number LSR-OXC Interfaces between the OIF UNI and the GMPLS UNI
  - OIF UNI does not allow maintaining semantic of client-initiated LSPs in the network ⇒ impossible for the network to discriminate b/w soft-provisioned protecting LSPs and hard-provisioned working LSPs
  - GMPLS UNI allows performing such distinction ⇒ protecting LSPs can be soft-provisioned. Moreover, soft-provisioned resources for protecting LSPs can be used by lower priority LSPs (preemption during activation phase)





GMPLS UNI – Time Performance

- Recovery speed for (1) and (2) slower (~ 50%) than (3) because of the additional ½ RTT required to perform resource activation (consequence of soft-provisioning technique applied in (1) and (2))
- For (1) and (2), when MPLS-based recovery used as fallback mechanism in case of GMPLS recovery failure, an MPLS hold-off time of about 100 ms can be applied (before which the IP/MPLS layer should not initiate any recovery attempt)



MPLS 2004

ALCATEL

Conclusion

Conclusion

- GMPLS UNI capabilities are superior to those of its OIF counterpart since more flexible, more extensible and more feature-rich than the OIF UNI (and fully GMPLS compliant)
- GMPLS UNI design meets the increasing need for more efficient, more robust and deterministic recovery sequences for multi-layer networks as well as more resource efficient provisioning
- Simulation results confirms quantitatively the benefits of the proposed multi-layer recovery solution
- Additionally, the GMPLS UNI provides a smooth evolutionary path towards integrated routing

References

References

[1] G. Swallow et al., *GMPLS-UNI RSVP Support for the Overlay Model*, Work in progress, draft-ietf-ccamp-gmpls-overlay-04.txt, April 2004.

[2] J.P. Lang et al., *RSVP-TE Extensions in support of End-to-End GMPLS-based Recovery*, Work in progress, draft-ietf-ccamp-gmpls-recovery-e2e-signaling-01.txt, May 2004.

[3] J.Lang (Editor), *Link Management Protocol (LMP) v1.0*, Work in progress, draft-ietf-ccamp-lmp-10.txt, October 2003.

[4] L.Berger et al., *RSVP Refresh Overhead Reduction Extensions*, RFC 2961, April 2001.

[5] D.Awduche et al., *RSVP-TE: Extensions to RSVP for LSP Tunnels*, RFC 3209, December 2001.

[6] L.Berger (Editor) et al., *Generalized Multi-Protocol Label Switching (GMPLS) Signaling – Resource Reservation Protocol - Traffic Engineering (RSVP-TE) Extensions*, RFC 3473, January 2003.

[7] K.Kompella, and Y.Rekhter, *Signaling Unnumbered Links in Resource Reservation Protocol - Traffic Engineering (RSVP-TE)*, RFC 3477, January 2003.

Backup Slides

Fundamental OIF UNI Problems

1. UNI refers to horizontal partitioning of the control plane (orthogonal to inter-layer exchanges)
2. Reachability information can be exchanged between clients edge nodes (orthogonal to routing issues)
3. "Canonical" overlay model implies address resolution between logical end-points and their network (physical) counter-part for enabling connection service
4. Tunnel_ID + LSP_ID values are globally significant and Label values are locally significant
5. When using non-associated control plane the "control channels" are pre-configured (I.e. some auto-discovery mechanisms are useless)

GMPLS UNI vs OIF Identification

GMPLS: TE links and Bundles

- Bundled TE link := set of component links (numbered or unnumbered)
- Bundled TE link can themselves be numbered or unnumbered
- Component or Bundled Link_id := <Local Link_Id, Remote Link_Id> with Link_Id being either Interface_Id or IP address
- Local Label (Resource) identification := <Bundled Link_Id, Component Link_Id, Label> with Label as per [RFC 3471]

OIF UNI: Transport Network Address

- Values assigned by the network that identifies (bundled or unbundled) TE links using IPv4/IPv6/NSAP
- Correspond “names” assigned to TE links (I.e. <Local Link_Id, Remote Link_Id>)
- Local Resource identification := <TNA, label> with label := <Link_Id, Label> with Label as per [RFC 3471]