# MPLS
# Multi-Protocol Label Switching

Mario Baldi

Politecnico di Torino
(Technical University of Torino)

mario.baldi [at] polito.it

staff.polito.it/mario.baldi

# Nota di Copyright

# MPLS Generic Label

- Label: 20 bits
  - 0-16 reserved
- Exp: 3 bits
  - For experimental use
  - QoS/CoS (Quality of Service/Class of Service) related functionality
- S: 1 bit
  - Identifies the label at the bottom of the stack
- TTL (Time to Live): 8 bits

# Encapsulation of Labeled Packets

- Multiprotocol support
  - No explicit identification of upper layer protocol
  - Network layer protocol inferred from label value
- "Too big" packets
  - Processing depends on higher layer protocol
  - Non fragmentable packets are discarded
    - e.g., IPv4 packet with DF (Don't Fragment) bit set
  - Fragmentable packets can be discarded
  - IPv4 is fragmented
    - If allowed by DF bit
  - IPv6 fragmented only if there is a fragment header

# Generic Label Encoding/MPLS Shim Header: Point-to-Point Protocol (PPP)

- MPLS Control Protocol
- One labeled packet per PPP frame
- PPP Protocol Field = 0281h → MPLS Unicast
- PPP Protocol Field = 0283h → MPLS Multicast

# Generic Label Encoding/MPLS Shim Header: Ethernet/IEEE 802

- One labeled packet per Ethernet Frame
- Label stack immediately precedes network layer
  - After other data link layer headers (e.g., IEEE 802.1Q)
- Ethernet encapsulation or IEEE 802 LLC/SNAP encapsulation
  - Ethertype = 8847 $\rightarrow$ MPLS Unicast
  - Ethertype = 8848 $\rightarrow$ MPLS Multicast

# MPLS Label Encoding with ATM

- Labeled packet segmented over multiple cells
- SVC Encoding (Switched Virtual Circuit)
  - VPI/VCI (Virtual Path Identifier/Virtual Circuit Identifier) encode label at top of stack
  - Label distribution protocol eliminates ATM signaling
  - No push and pop capability within ATM network
- SVP Encoding (Switched Virtual Path)
  - VPI encodes top label - VCI encodes second label
  - VP switching
  - The egress ATM-LSR is capable of a pop operation
- SVP multipoint encoding
  - VPI encodes top label - Part of VCI encodes second label - reminder of VCI encodes LSP ingress
  - Multipoint-to-point VPs can be used

# MPLS Label Encoding with Frame Relay

- Null encapsulation
  - DLCI implicitly identifies higher layer protocol
- Top label in DLCI (Data-Link Circuit Identifier)
- Virtual Circuits can be treated as half-duplex
  - Each direction has its own label

# TTL (Time To Live) Handling

- At exit from MPLS domain TTL should have been decremented by the number of hops traversed
- The way it is handled depends on how labels are carried
  - MPLS shim header
    - The shim header has TTL
    - Initialized at the value in the packet
    - Decremented by each hop
  - Layer 2 header
    - e.g., ATM, Frame relay
    - Non-TTL LSP segment
    - Adjsted upon entrance or exit

# Label Switched Path (LSP)

A level *m* LSP is a sequence of LSRs such that

- The first LSR (LSP Ingress) pushes on the stack a level *m* label
- All the intermediate LSRs base their routing decision on a level *m* label
- The last hop (LSP Egress) takes a routing decision based on either
  - a level *n* label, where $n < m$
  - non MPLS information
    - e.g., normal IP routing

# Penultimate Hop Popping

- The purpose of an *m* level label is to take its packet to the end of a level *m* LSP
- Once the penultimate hop has routed the packet it will reach the last hop of the LSP
- Lookup of the *m* level label by last hop is useless
  - The label is anyway to be popped in order to then
    - Lookup the next label
    - Forward an unlabeled packet
  - Single lookup simplifies the realization of a data "fastpath" in switching equipment

# Penultimate Hop Popping

- Not mandatory
  - There might be switching engines unable to pop
- It is requested by last hop
  - Request tells upstream node to be penultimate
  - Initial label distribution protocol negotiations must have ways of knowing whether an upstream node is pop capable
- It is necessary if last hop does not support MPLS

# FEC (Forwarding Equivalence Class)

- Packets belonging to same FEC receive same treatment
  - E.g., packets with same prefix
- Binding between a FEC and a label
- Packets belonging to multiple FECs might follow same route
  - Multiple *aggregatable* FECs can be
    - assigned different labels
    - Aggregated in a set of FECs
    - Aggregated in one FEC
  - Different granularity of aggregation
  - With independent control adjacent LSRs might aggregate FECs with different granularity

# Next Hop Label Forwarding Entry (NHLFE)

- Next hop for a packet
- Operation to be performed on label
  - Replace label
  - Pop label
  - Replace top label and push one or more new labels
- Data link encapsulation to be used on output link
- Way to encode label stack on output link

# Mapping packets to NHLFE

- ILM (Incoming Label Map)
  - Maps a labeled packet to a set of NHLFE
  - Enables forwarding of labeled packet
- FTN (FEC-to-NHLFE)
  - Maps an FEC to a set of NHLFE
  - Enables forwarding of unlabeled packets that are to be labeled before being forwarded
- Why mapping to a set of NHLFE?
  - It might be useful for specific applications
    - E.g., load balancing among alternate links/paths
  - Criteria for choice within set is not specified

# Label Assignment

- Upstream-assigned
  - An upstream node assigns one or more labels to a FEC
  - Communicates the assignment to the downstream node

# Label Binding

A LSR determines the label it intends to use for packets belonging to a given FEC

- Downstream:
  - Downstream node decides the label that shall be prepended to incoming packets
- Downstream-on-demand
  - Upstream node asks the next hop to bind a label to a FEC
- Unsolicited downstream
  - A node distributes bindings upstream without having been asked for them

# Label Distribution

- No single label distribution protocol
- Existing protocols have been modified
  - RSVP (Resource reSerVation Protocol)
  - BGP (Border Gateway Protocol)
- New protocols have been defined
  - LDP (Label Distribution Protocol)
  - CR-LDP (Constraint Routing-Label Distribution Protocol)

# Label Information Base (LIB)

- Bindings between FEC and labels
- For each FEC
  - One local binding
    - Label assigned by the LSR
  - Multiple remote bindings
    - One label assigned (and distributed) by a neighbor

# Label Retention Mode

When an LSR receives a label binding for a FEC *F* from a neighbor *N*

- Conservative label retention mode
  - Discards the binding unless *N* is his next hop for reaching *F*
- Liberal label retention mode
  - Keeps the binding anyway
  - The binding is used (to create a NHLFE) if *N* becomes next hop for *F*

Tradeoff between higher reactivity (liberal) and smaller memory requirements (conservative)

# Independent LSP Control

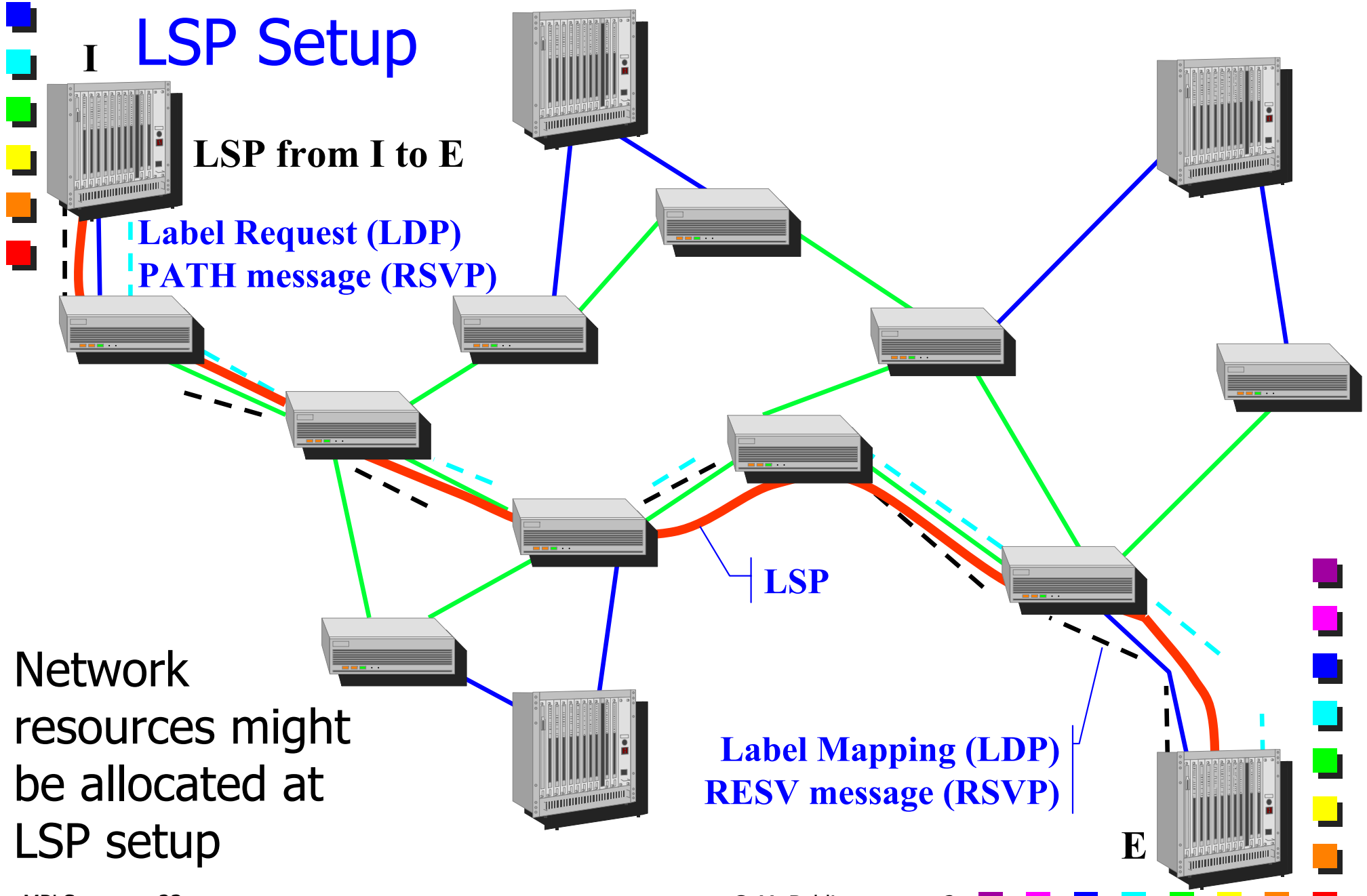
- Upon recognizing a FEC, an LSR makes independent binding and communicates it
- Similar to IP routing:
  - Every node decides independently
  - Relies on fast convergence of routing protocols
- Packets forwarding independent of LSP setup
  - Packets forwarded before completion might follow different paths

# Ordered LSP Control

- An assignment is performed only if
  - Egress LSR for FEC
  - Received binding request from next hop
- Can be initiated by either ingress or egress LSR
- Suitable to ensure properties
  - E.g., same node not traversed twice
- Suitable when FEC recognition is a consequence of LSP setup
  - E.g., an LSP is created for a given packet flow

The two LSP control techniques are interoperable, but the properties of ordered control hold true only if all nodes support ordered control

# LSP Setup




Network resources might be allocated at LSP setup

# LDP vs RSVP

- RSVP provides only downstream-on-demand label distribution
  - LDP provides every mode
- LDP uses reliable transport (TCP)
  - RVSP cannot guarantee fast failure notification
- LDP is hard-state, RSVP is soft-state
  - RSVP has poor scalability due to periodic refreshes

# LSP Tunnel

- Hop-by-hop tunnel
  - IP encapsulation
- Explicitly routed tunnel
  - Source routed packet
- LSP tunnel
  - Create a FEC and label binding for tunneled packets

# Routing

- Hop-by-hop routing
  - Each LSR chooses independently the next hop for each FEC
- Explicit routing
  - A single LSR specifies the LSR on the LSP
  - Normally ingress LSR or Egress LSR
  - Strictly explicitly routed: all LSRs are specified
  - Loosely explicitly routed: some LSRs are specified
  - Explicit route can be configured or calculated dynamically
  - Policy routing and traffic engineering

# Constraint Based Routing

- Route choice is based on multiple criteria (constraints)
  - Not necessarily shortest path
    - E.g., load balancing
- Routing protocol carries constraint data
  - Information on link and nodes
    - Status, characteristics
  - Constraint data change fast compared to topological information
    - Explicit routing
    - Loop risk with distributed control
    - Crackback capabilities

# MPLS and DiffServ

- Differentiate service on an LSP basis
- How to map behavior aggregates (BAs) to LSPs
  - BA: set of packets requiring same treatment
  - Marked with same DiffServ Code Point (DSCP)

# E-LSP (EXP-Inferred-PSC LSP)

- LSP carries multiple OAs (Ordered Aggregates)
  - OA: set of BA that share an ordering constraint
- PSC (PHB Scheduling Class): set of per-hop behaviors (PHBs) applied to the BAs of an OA
- EXP bits identify PHB to apply
  - Scheduling treatment
  - Drop precedence
- Mapping from EXP to PHB can be
  - Signaled at LSP setup
  - Pre-configured

# Label-only-Inferred-PSC LSP (L-LSP)

- LSP carries only one OA
- Label provides scheduling treatment (PSC)
- EXP or layer 2 information provide drop precedence
  - E.g., CLP (Cell Loss Priority) in ATM

# Traditional IP Fault Recovery

- In the Internet it is left to routing protocols
- Yesterday, strength of the Internet
  - Robustness
  - No single point of failure
- Today, shortcoming
  - Recovery might take tens of seconds to minutes
  - Certain applications require 50 ms
    - E.g., voice, circuit emulation

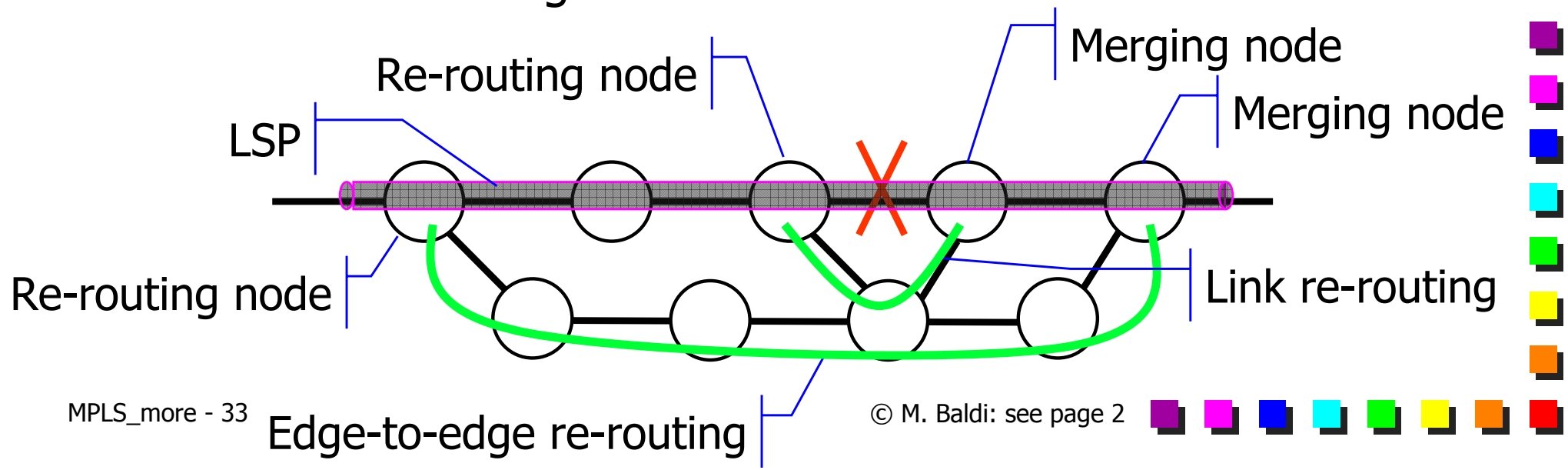
# MPLS Enabled Fault Recovery

- Protection
  - Pre-determined action
    - E.g., switch to other LSP
  - Pre-allocated resources
    - E.g., reserved backup LSP
  - Non-optimal resource utilization
  - Extremely fast
  - Change in the FTN (FEC-to-NHLFE)
- Restoration
  - Dynamically determined action
  - Optimization of resource utilization
  - Slower but done while network is operating properly

# Protection

## Link re-routing  Edge-to-edge LSP re-routing

- Pre-allocated resources on alternate path
  - Assigned labels
  - Buffering and switching resources
- Reaction time dependent on fault notification
  - LDP message could be used



# VoMPLS: Voice over MPLS

- Implementation agreement by MPLS Forum (now MPLS/Frame Relay Alliance)
- IETF dismissed it in favor of VoIPoMPLS
- VoIP strenght is IP ubiquity
  - VoMPLS is limited to the edge of an MPLS cloud
  - A LAN cannot be used to reach the desktop
  - More complicated wiring from desktop to MPLS cloud
- Advantage of VoMPLS is increased efficiency
  - Does not hold if header compression/suppression is used

# VoIPoMPLS

- VoMPLS cannot replace VoIP and
- VoIPoMPLS with header suppression obviates the need for VoMPLS
- VoIPoMPLS can benefit from MPLS advantages
  - IP header compression/suppression efficiency
  - QoS enabling features
  - Scalability
  - Fast restoration

# References


- IETF MPLS Working Group, http://www.ietf.org/html.charters/mpls-charter.html
- E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, Standards Track, Jan. 2001
- E. Rosen, D. Tappan, G. Fedorkow, Y. Rekhter, D. Farinacci, T. Li, A. Conta, "MPLS Label Stack Encoding," RFC 3032, Standards Track, Jan. 2001
- F. Le Faucheur, et al., "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services," RFC 3170, Standards Track, May 2002

# References


- Davie, B., Lawrence, J., McCloghrie, K., Rekhter, Y., Rosen, E., Swallow, G. and P. Doolan, "MPLS using LDP and ATM VC Switching", RFC 3035, January 2001.
- Andersson, L., Doolan, P., Feldman, N., Fredette, A. and B. Thomas, "LDP Specification", RFC 3036, January 2001.

# Loop Control

- TTL does not protect in non-TTL LSP segments

# Future issues

- Label merging (RFC 3031)
- Label distribution peering and hierarchy (RFC 3031)
- MPLS and IP Multicast (RFC 3353)
- MPLS and DiffServ (RFC 3270)
  - Tunnel mode
  - Pipe mode
- Aggiungere BGP per trasporto etichette
- Modello di routing BGP

# More on

- LDP and CR-LDP
- RSVP and RSVP-TE
- OSPF-TE
- G-MPLS
- Improve protection