Multi-Protocol Label Switching
Agenda

• Introduction to MPLS
• MPLS forwarding
• Label Distribution Protocol
• Traffic Engineering
• MPLS VPN
• MPLS QoS
MPLS Concept

At Edge:
- classify packets
- label them

In Core:
- forward using labels
- as opposed to IP addr

- Enable ATM switches to act as routers
- Create new IP capabilities via flexible classification
MPLS Overview

Label Distribution Protocol (LDP)

label Switches
(ATM Switch or Router)

Label Edge Routers
MPLS Operation

1a. Existing routing protocols (e.g. OSPF, IS-IS) establish reachability to destination networks
1b. Label Distribution Protocol (LDP) establishes label to destination network mappings.

2. Ingress label Edge Router receives packet, performs Layer 3 value-added services, and “MPLS” packets

3. Label Switches switch labeled packets using label swapping

4. Label Edge Router at egress removes tag and delivers packet
Control Planes in MPLS

- Internet
- intranet
- extranet
- voice
- data
- multimedia

Connectivity & Svc Class
- BGP
- ISIS
- EIGRP
- OSPF
- traffic engineering

Switching Paths
- LDP
- BPX 8600
- MGX 8800
- 7500
- 12000
- LS1010
- other Frame or ATM switches

SERVICES
ROUTE DETERMINATION
FORWARDING
Advanced MPLS

- Basic label switching: destination-based unicast
- Many additional options for assigning tags
- The Key: separation of routing and forwarding

<table>
<thead>
<tr>
<th>Destination-based Unicast Routing</th>
<th>IP Class of Service</th>
<th>Resource Reservation (e.g., RSVP)</th>
<th>Multicast Routing (PIM v2)</th>
<th>Explicit &amp; Static Routes</th>
<th>Virtual Private Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label Forwarding Information Base (TFIB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per-Label Forwarding, Queuing, and Multicast Mechanisms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Encapsulations

**ATM Cell Header**

```
+----+----+----+----+----+----+----+
| GFC | VPI | VCI | PTI | CLP | HEC | DATA |
+----+----+----+----+----+----+----+
```

**PPP Header**

(Packet over SONET/SDH)

```
+----+----+----+----+----+----+----+
| PPP Header | Label Info Header | Layer 3 Header |
+----+----+----+----+----+----+----+
```

**LAN MAC Label Header**

```
+----+----+----+----+----+----+----+
| MAC Header | Label Info Header | Layer 3 Header |
+----+----+----+----+----+----+----+
```
Generic Label Header Format

- **Generic**: can be used over Ethernet, 802.3, PPP links, Frame Relay, ATM PVCs, etc.
- **Uses 2 new Ethertypes/PPP PID/PPP SNAP values/etc.**
  - one for unicast, one for multicast
- **4 octets (per tag level)**

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>01234567890123456789012345678901</td>
<td>Label</td>
<td>EXP</td>
<td>S</td>
</tr>
</tbody>
</table>

Label = 20 bits
EXP = Experimental, 3 bits
S = Bottom of stack, 1bit
TTL = Time to live, 8 bits
ATM MPLS

- VPI/VCI field is used as a ‘tag’
- Label is applied to each cell, not whole packet
- Label swapping = ATM switching
Carrying Labels on Ethernet Links

- Extra four bytes might lead to fragmentation of 1492-byte packets
- Path MTU discovery will detect need to fragment (MTU discover packets will be sent tagged)
- But: many Ethernet links actually support 1500 or 1508-byte packets
- And: most packets will normally be carried over ATM, or PPP/SDH links, not Ethernet
MPLS Basics: Summary

- MPLS puts IP routing functions on ATM switches. This provides better IP and ATM integration and better scaling.

- On non-ATM equipment, MPLS simplifies the forwarding operation and introduces ‘lightweight virtual circuits’. This allows advanced features like MPLS Traffic Engineering.
Agenda

• Introduction to MPLS
• **MPLS forwarding**
• Label Distribution Protocol
• Traffic Engineering
• **MPLS VPN**
• **MPLS QoS**
MPLS: Forwarding

- A pair of routers handle a class of packets with similar parameters

The first router classifies the packets...

.. summarizes its decision with a label on the packet.

The next router just looks at the label

- MPLS simplifies forwarding, pushes packet classification back towards the edge
Label Distribution Protocol

Standard Routing Protocols

Label Distribution Protocol (LDP)

Routing Table

Label Information Base (TIB)

Label Edge Router

Label Switch

Label Edge Router
**Router Example: Distributing Routing Information**

You can reach 128.89 and 171.69 through me.

Routing updates (OSPF, IS-IS, ...)

You can reach 128.89 through me.

You can reach 171.69 through me.

<table>
<thead>
<tr>
<th>address prefix</th>
<th>I/F</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>171.69</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>address prefix</th>
<th>I/F</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>171.69</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>address prefix</th>
<th>I/F</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Router Example: Forwarding Packets

Packets forwarded based on IP address

<table>
<thead>
<tr>
<th>address prefix</th>
<th>I/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>1</td>
</tr>
<tr>
<td>171.69</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>address prefix</th>
<th>I/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>0</td>
</tr>
<tr>
<td>171.69</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>address prefix</th>
<th>I/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>0</td>
</tr>
</tbody>
</table>

0 128.89.25.4 Data
128.89 128.89.25.4 Data
171.69
You can reach 128.89 and 171.69 through me

Routing updates (OSPF, IS-IS, ...)

You can reach 171.69 through me
### ATM MPLS Example: Assigning Labels

<table>
<thead>
<tr>
<th>In Tag</th>
<th>Address Prefix</th>
<th>Out I/F</th>
<th>Out Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>128.89</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>-</td>
<td>171.69</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In Tag</th>
<th>In I/F</th>
<th>Address Prefix</th>
<th>Out I/F</th>
<th>Out Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2</td>
<td>128.89</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>128.89</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>171.69</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In Tag</th>
<th>In I/F</th>
<th>Address Prefix</th>
<th>Out I/F</th>
<th>Out Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1</td>
<td>128.89</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>128.89</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Use tag 4 for 128.89
Use tag 7 for 171.69
Use tag 8 for 128.89
Use tag 9 for 128.89
Use tag 10 for 128.89
ATM MPLS Example: Requesting Labels

<table>
<thead>
<tr>
<th>In Tag</th>
<th>Address Prefix</th>
<th>Out I'face</th>
<th>Out Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.89</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>171.69</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Label Distribution Protocol (LDP) (downstream allocation on demand)
MPLS Example: Forwarding Packets

Label Switch forwards based on tag

128.89.25.4 Data

128.89.25.4 Data

128.89.25.4 Data
MPLS Example: More Details

Prefixes that share a path can share tag

Remove tag one hop prior to de-aggregation point

De-aggregation point does L3 lookup

Prefixes:
- 128.89.
- 171.69
- 117.59

Data:
- 128.89.25.4
- 128.89.26

Tags:
- 7
- 4
- X

De-aggregation point does L3 lookup.
Internet IGP Labelling

- Apply labels to IGP routes
  - Conserves labels
- Shields core from BGP routes
  - No BGP route flaps in core
  - Smaller tables
  - Faster convergence

At Edge:
- Look up IP address, find BGP next hop
- Look up BGP next hop, find IGP route & label
- Apply IGP label, forward

In Core:
- Forward using labels
- Labels assigned to IGP routes only
MPLS Across Non-MPLS ATM Networks

- ATM SVCs created as needed; VCIs mapped to tags
- Labelled cells transported in Virtual Path

MPLS Network

ATM Network
Label Forwarding: Summary

- Helps routing scale: analyze packets only at edge
- Makes full-featured routing feasible
  - Labelling on destination, source, ToS, (RSVP)
  - Multicast labelling, other modes
- Will run on any MAC layer
- Basic mechanism is extensible to traffic engineering, multicast
Introduction to MPLS
MPLS forwarding
Label Distribution Protocol
Traffic Engineering
MPLS VPN
MPLS QoS
• **FIB**: for unlabelled packets
  - New function: outgoing *labelled* packet
• **TFIB**: for incoming labelled packets
TIB and TFIB

- TIB is populated by LDP/TDP
- TFIB is derived from TIB and used for packet forwarding
Label distribution

- **Upstream tag distribution**
  - when tag is assigned (based on destination) by upstream router

- **Downstream tag distribution**
  - current LDP/TDP implementation
Downstream label distribution

- Downstream LSR (Rd) distributes all tags to upstream neighbors (Ru)
- Used for frame interfaces
- When downstream LSR is ready to forward labelled packets for destination D, it assigns a label and distribute it to all upstream neighbors
**Label Distribution**

- **Downstream on demand label distribution**
  - Downstream LSR distribute part of its label space
  - Based on upstream neighbors requests
  - Used for ATM interfaces
  - When upstream LSR is ready to forward packets for destination D, it requests a tag for D from the next-hop (Rd)
Protocol enhancements in order to carry labels

- BGP
  Used to distribute labels for external destinations (MPLS-VPN)
- RSVP
  Used for LSP tunnels (Traffic Engineering)
- PIMv2
  Used to distribute labels for (S,G) or (*,G) entries in multicast state table
LDP transport

- LDP uses TCP as transport layer
- Well-known TCP port 711
- One TCP session per LDP session
  - No multiplexing at this stage
  - when label is assigned (based on destination) by upstream router
LDP Identifier

- Identifies label space for
  - The router
  - The interface
- Exchanged during LDP session set up
- 6 bytes
LDP neighbor discovery

• Discovery is done through Hello packets
  • Hello are periodically sent via UDP
  • Hello are sent on all label-enabled interfaces
  • Source address is the outgoing interface
  • Hellos packets contain
    - LDP Identifier
    - Label space
Once discovery is done the LDP session is established over TCP.

LSRs send periodically keepalive LDP packets to monitor the session.
LDP Identifiers and Next-Hop addresses

Tag Information Base (TIB):
- Stores tags with peer LDP Identifier

Routing Information Base (RIB):
- Maintains next-hop IP addresses

<table>
<thead>
<tr>
<th>Dest</th>
<th>In tag</th>
<th>(Peer, Out tag)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>tR1</td>
<td>(R2:0,tR2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dest</th>
<th>Next-Hop</th>
<th>Int</th>
<th>Pctl</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>a.b.c.d</td>
<td>e0</td>
<td>OSPF</td>
<td>10</td>
</tr>
</tbody>
</table>
LDP Identifiers and Next-Hop Addresses

- TFIB requests labels assigned by next-hop to destination
- LDP maps next-hop address into peer LDP Identifier in order to retrieve a label
- LSRs advertise interface addresses via LDP
- LSRs map peer LDP ID to addresses
  Using learned addresses
LDP Sessions

Session for L1, L2 and L3

Session for L1, L3

Session for L2

Session for L1, L3

Session for L1

Session for L2

Session for L3

Session for L4

Session for L2 (ATM)
LDP Sessions between non directly connected LSRs

LDP session is established between R1 and R2
End of tunnel is BGP next-hop for destination
Hello mechanism is different
Direct Hello packets
Label Distribution Protocol (LDP)

- Run in parallel with routing protocols
- Distributes \(<\text{tag, prefix}>\) bindings
- Incremental updates over TCP
- Other tag distribution mechanisms can run in parallel with it
Agenda

• Introduction to MPLS
• MPLS forwarding
• Label Distribution Protocol
• Traffic Engineering
• MPLS VPN
• MPLS QoS
Traffic Engineering Motivation

• “For a given network topology and traffic load, where should my traffic go and how do I make it go there?”
Traffic Engineering Motivation

- Link not available
- Economics
- Size of pipes
- Failure scenarios
- Unanticipated growth
- Class of service routing
Traffic Engineering

IP (Mostly) Uses Destination-Based Least-Cost Routing
Flows from R8 and R1 Merge at R2 and Become Indistinguishable
From R2, Traffic to R3, R4, R5, R9 Use Upper Route

Alternate Path Under-Utilised
LSP tunnels

- Labelled packets are forwarded based on tag, not IP destination
- In conjunction with signaling mechanism. Label forwarding can be used to create a multi-hop LSP tunnel: TE tunnel
- LSP tunnel is used to reach BGP next-hop
LSP tunnel setup via RSVP

- RSVP extensions
- Initiated at source router
- Complete path in forward messages
- Label established by reply messages
- Rapid tear down on link failure
LSP tunnel setup via RSVP

- Possible future resource capabilities
- Unidirectional data flow
- May traverse ATM LSR, but not begin or end there
**Setup:** Carries Path (R1->R2->R6->R7->R4) and Tunnel ID

**Reply:** Communicates Labels and Establishes Label Operations
LSP tunnel configuration

- IOS tunnel interface with tag-switching encapsulation (not GRE)
- Source route
  - Specified as the sequence of IP addresses
- Configured only at the head of the tunnel
LDP tunnels

Distribution

R8
R2
R6
R3
R4
R7
R1 R5
R9

S1

D2

IP routed

LDP Adjacency
R4 -> R1, R5 = Label 25

LDP Adjacency
R5 -> R4, R5 = Pop
Tunnel routed by filter
BGP NH = R5
Label 25 pushed
Label 49 pushed

Label-Switched swap label 49 ->17
Label-Switched swap label 17 ->22
Label-Switched Pop TE label
LSP Tunnels forwarding

• Build around CEF

• At head
  uses CEF (IP-->tag)
  TFIB (tag-->tag)

• At midpoint uses TFIB (tag-->tag)

• MPLS performance
Agenda

• Introduction to MPLS
• MPLS forwarding
• Label Distribution Protocol
• Traffic Engineering
• MPLS VPN
• MPLS QoS
Benefits of Internet-Scale VPNs

Connection-Oriented VPN Topology

Connectionless VPN Topology

VPN Aware Network: VPNs are “built-in” rather than “overlaid”
VPN Models - The Overlay model

- Private trunks over a TELCO/SP shared infrastructure
  - Leased/Dialup lines
  - FR/ATM circuits
  - IP (GRE) tunnelling
- Transparency between provider and customer networks
- Optimal routing requires full mesh over the backbone
VPN Models - The Peer model

- Both provider and customer network use same network protocol
- CE and PE routers have a routing adjacency at each site
- All provider routers hold the full routing information about all customer networks
- Private addresses are not allowed
- May use the virtual router capability

Multiple routing and forwarding tables based on Customer Networks
VPN Models - MPLS-VPN: The True Peer model

- Same as Peer model BUT !!!
- Provider Edge routers receive and hold routing information only about VPNs directly connected
- Reduces the amount of routing information a PE router will store
- Routing information is proportional to the number of VPNs a router is attached to
- MPLS is used within the backbone to switch packets (no need of full routing)
MPLS Operation

1a. Existing routing protocols (e.g. OSPF, ISIS) establish reachability to destination networks

1b. Label Distribution Protocol (LDP) establishes tag to destination network mappings.

2. Ingress Label Switch Router receives packet, performs Layer 3 value-added services, and “tags” packets

3. Core LSR switch packets using label swapping

4. Egress LSR removes label and delivers packet
MPLS VPN Routing Architecture

- **P** router = Provider Router (Core LSR)
- **PE** router = Provider Edge router (Edge LSR) knows which VPN each CE belongs to (by sub-interface)
- **CE** router = Customer Edge router
- **RD** (Route Distinguisher) = uniquely identify a VPN (AS#,VPN_ID)
- IPv4 Addresses are unique within VPN
- IPv4 Addresses might overlap across VPN’s
• Each P routers, including PE has to maintain Internal Routes reachability and associated internal Labels.

• The FIB is populated by an IGP (I-ISIS, OSPF, EIGRP)

• TFIB populated by LDP
Ingress PE routers, learns routes from CE
- Static routing, eBGP or RIPv2
- In order to guarantee the uniqueness of the customer address, the ingress PE router converts IPv4 address into a globally unique "VPN-IPv4" address
- A 64 bits "Route Distinguisher" is prepended to the customer IPv4 address and propagated via BGP to the egress PE’s (BGP Multiprotocol Extension)
Per VPN FIB (Forwarding Information Base)

- VPN-IPv4 address are propagated together with the associated Label in “BGP multi-protocol extension” (NLRI field)
- Additional community fields (64 bits Extended Community attribute) are associated to VPN-IPv4 address, to build a per VPN FIB:
  - “Target VPN” (list of), “VPN of Origin”, Site of Origin
- Filters (route-maps) are applied to tightly control intra-VPN and inter-VPN connectivity
- Creation of a per VPN RIB and FIB
Label Binding to VPN-IPv4 addresses

- iBGP (Multiprotocol Extension) has distributed the Label associated with `<VPN-IPv4>`. Filters are applied on extended community attributes.
- LDP has distributed the Label associated with Interior routes (`BGP next hop` add).
- **Recursive lookup**
  For each customer address the PE does a recursive lookup to find the path to the “`BGP next hop`”, and build its TFIB.
- Each `<VPN-IPv4 address>` is assigned, an **Interior Label** AND an **Exterior Label**
Scaling: BGP Hierarchical Architecture

- Full mesh of BGP peers => scalability issues for Very Large VPN’s
- Use of BGP Route Reflector to scale the VPN BGP peering
- For resiliency peers “multiple VPN PE” to multiple VPN RR
- PE needs to have the routing information only for the VPN’s it is connected to.
- Peer RR together to allow inter VPN communications
Forwarding and Isolation: Stacks of Label

- Ingress PE receives normal IP Packets from CE router
- PE router does “IP Longest Match” from VPN_B FIB, find iBGP next hop PE2 and impose a stack of Labels’s: exterior Label T2 + Interior Label T8
Forwarding and Isolation: Stack of Label

- All Subsequent P routers do switch the packet **Solely on Interior Label**
- Egress PE router, removes **Interior Label**
- Egress PE uses **Exterior Label** to select which VPN/CE to forward the packet to.
- **Exterior Label** is removed and packet routed to CE router
Closed User Group Servers

- Green VPN customers access to Green Server only
- There may be “public” servers in a common public “VPN”
- Server IPv4 address is advertised only in the VPN it belongs to.
- VLAN are used to isolate per VPN servers, in the “server farm”
Inter VPN’s communications

- Inter VPN’s communication is controlled by mean of “Community filtering” (VPN of Origin, Target VPN)
- VPN Leakage point control the inter-VPN point (may be multiple)
- *intra-VPN* can be *any to any* while *inter-VPN* can be *hub and spoke*
  - Central Firewall control
- Internet Connectivity can be provided in the same manner
VPN Spanning multiple domains

- VPN Membership can be extended across SP boundaries
- Private BGP peering
- *Multi-Protocol extension* and *community* attributes are carried through the external BGP private peer.
- *RD’s* are affected independently by both SP
- Reachability is controlled by both BGP peers (VPN of Origin, Target VPN)
Agenda

• Introduction to MPLS
• Label forwarding
• Label Distribution Protocol
• Traffic Engineering
• MPLS VPN
• MPLS QoS
What is Label/MPLS QoS?

Support of Consistent IP Diff-Serv Classes of Service end-to-end when part of the network is running MPLS
MPLS QoS: 3 Steps

1) in non-MPLS part:
   - existing IP mechanism (CAR) to mark IP DS-byte
   - existing IP Mechanisms (WRED/WFQ) for service differentiation

2) Mapping IP DS-byte into EXP field on MPLS Edge

3) Supporting Differentiation based on EXP field in MPLS Backbone
Mapping IP QoS into EXP

At MPLS Imposition: DS-Byte (initially Precedence) mapped into EXP (3 bits)
On MPLS Frame Interface (ie non-ATM), it’s simple:

- Every MPLS packet has explicit indication of QoS in MPLS Header
- Use EXP field to trigger Selective Scheduling (WFQ) and Selective Discard (WRED); exactly like use of IP DS-byte in non-MPLS

Net result is end-to-end QoS indistinguishable from non-MPLS network
Main challenges:
- No QoS field in ATM cell header
- No WRED in switches

Two modes:
- Single `VC’ ABR
- Multi-`VC’ TBR
  (closer to Frame QoS)
- Each has advantages and drawbacks

TBR = Tag Bit Rate
ATM Service Category better suited to IP
Single-ABR and Multi-TBR

• **Multi-VC TBR Mode:**
  
  • Congestion managed directly at every hop (IP and ATM hops)
  • Possible Discard at every hop
  • Resource Allocation per QoS per link; does not have to concern itself with topology and geography

• **Single-VC ABR:**
  
  • No Loss in the ATM fabric
  • Discard possible only on the Edge performed by Routers
  • Resource Allocation optionally per Pair of Edge Routers. Sharing of bandwidth across QoS indirect via WRED profiles