APNIC eLearning: Introduction to MPLS

01 NOVEMBER 2017
01:00 PM AEST Brisbane (UTC+10)
Introduction

• Presenter/s

Jessica Bei Wei
Training Officer
jwei@apnic.net

Specialties:
Routing & Switching
MPLS, IPv6
QoS

• Reminder: please take time to fill-up the survey
Overview

• Definition of MPLS
• Advantages of MPLS
• MPLS Application
• MPLS Architecture
• MPLS Labels
• LSP Setup
• Forwarding Operations
Definition of MPLS

- **Multi Protocol Label Switching**
  - Multiprotocol, it supports ANY network layer protocol, i.e. IPv4, IPv6, IPX, CLNP, etc.
  - A short label of fixed length is used to encapsulate packets
  - Packets are forwarded by label switching instead of by IP switching
Initial Motivation of MPLS

• In mid 1990s, IP address lookup was considered more complex and take longer time.
  – Longest matching

A label-swapping protocol was the need for speed.
Decoupling Routing and Forwarding

- MPLS can allow core routers to switch packets based on some simplified header.

- But, hardware of routers became better and looking up longest best match was no longer an issue.

- More importantly, MPLS de-couples forwarding from routing, and support multiple service models.
### MPLS vs IP over ATM

#### MPLS
- Layer 2 devices run a Layer 3 routing protocol and establish virtual circuits dynamically based on Layer 3 information.
- MPLS provides a virtual full mesh topology.

#### IP over ATM
- Layer 2 topology may be different from Layer 3 topology, resulting in suboptimal paths.
- Layer 2 devices have no knowledge of Layer 3 routing – virtual circuits must be manually established.
MPLS VPN

- MPLS Layer 3/ Layer 2 VPN
Optimal Traffic Engineering

<table>
<thead>
<tr>
<th>IP TE</th>
<th>MPLS TE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortest path</td>
<td>Determines the path at the source based on additional parameters (available resources and constraints, etc.)</td>
</tr>
<tr>
<td>Equal cost load balancing</td>
<td>Load sharing across unequal paths can be achieved.</td>
</tr>
</tbody>
</table>
MPLS QoS

- MPLS does NOT define a new QoS architecture.
  - Similar parts with IP DiffServ: functional components and where they are used (such as marking and traffic policing at network edge, etc)
  - Difference: packets are differentiated by MPLS Traffic Class bits
## Technology Comparison

<table>
<thead>
<tr>
<th></th>
<th>IP</th>
<th>Native Ethernet</th>
<th>MPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forwarding</strong></td>
<td>• Destination address based</td>
<td>• Destination address based</td>
<td>• Label based</td>
</tr>
<tr>
<td></td>
<td>• Forwarding table learned from control plane</td>
<td>• Forwarding table learned from data plane</td>
<td>• Forwarding table learned from control plane</td>
</tr>
<tr>
<td></td>
<td>• TTL support</td>
<td>• No TTL support</td>
<td>• TTL support</td>
</tr>
<tr>
<td><strong>Control Plane</strong></td>
<td>Routing protocols</td>
<td>Ethernet loop avoidance and signaling protocols</td>
<td>Routing protocols</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Label distribution protocols</td>
</tr>
<tr>
<td><strong>Packet Encapsulation</strong></td>
<td>IP header</td>
<td>802.3 header</td>
<td>MPLS Header</td>
</tr>
<tr>
<td><strong>QoS</strong></td>
<td>8 bit TOS in IP header</td>
<td>3 bit 802.1p in VLAN tag</td>
<td>3 bit TC in label</td>
</tr>
<tr>
<td><strong>OAM</strong></td>
<td>IP Ping, traceroute</td>
<td>E-OAM</td>
<td>MPLS Ping, traceroute</td>
</tr>
</tbody>
</table>
Evolution of MPLS

• Technology Evolution and Main Growth Areas

1996, Ipsilon, Cisco and IBM announced label switching plans, till now, there are over 280 RFCs of MPLS tech.
MPLS Application Scenario

MPLS CORE

Enterprise

L3VPN

PE1

P

P

L2VPN

P

P

L2VPN

PE2

PE3

PE4

TE Main Path for PE1-PE3

TE Backup Path for PE1-PE3

QoS Operations: Traffic marking, police, shaping

QoS Operations: Congestion management, congestion avoidance

QoS Operations: Traffic marking, police, shaping

Enterprise

Enterprise

Enterprise
MPLS Architecture

Routing Information Exchange with other routers

Routing Information Base (RIB)

IP Routing Protocols

Label Distribution Protocols

Label Information Base (LIB)

Data Plane

Incoming IP Packet

Forwarding Information Base (FIB)

Label Forwarding Information Base (LFIB)

Control Plane

Incoming Labeled Packet

Label Binding and Exchange with other routers
MPLS Topology

- **LSR** (Label Switch Router) is a router that supports MPLS.
- **LER** (Label Edge Router), also called edge LSR, is an LSR that operates at the edge of an MPLS network.
- **LSP** (Label Switched Path) is the path through the MPLS network or a part of it that packets take.
MPLS Label

MPLS Label Encapsulation

Datalink Layer Header  MPLS Label  Layer 2/ Layer 3 Packet

Label - 20bits  TC  S  TTL-8bits

TC = Traffic Class: 3 Bits; S = Bottom of Stack: 1 Bit; TTL = Time to Live
MPLS Label Stacking

Multiple labels can be used for MPLS packet encapsulation in the network. This is done by packing the labels into a stack.

Some MPLS applications (VPN, etc.) actually need more than one label in the label stack to forward the labeled packets.

MPLS Label Stack

- Multiple labels can be used for MPLS packet encapsulation in the network. This is done by packing the labels into a stack.
- Some MPLS applications (VPN, etc.) actually need more than one label in the label stack to forward the labeled packets.
LSP Setup Overview

• Before forwarding packets, labels must be allocated to establish an LSP.
• Protocols for label distribution: LDP, RSVP-TE, MP-BGP.

Establishing an LSP
Labels are allocated from downstream LSRs to upstream LSRs.
Basic Concepts of MPLS Forwarding

• **FEC**
  – Forwarding Equivalence Class, is a group or flow of packets that are forwarded along the same path and are treated the same with regard to the forwarding treatment.
  – For example, packets with Layer 3 destination IP address matching a certain prefix.

• **Push**
  – A new label is added to the packet between the Layer 2 header and the IP header or to the top of the label stack.

• **Swap**
  – The top label is removed and replaced with a new label.

• **Pop**
  – The top label is removed. The packet is forwarded with the remaining label stack or as an unlabeled packet.
MPLS Forwarding Operations

- **Prefix**: 100.1.1.1/32
  - **Local Label**: Null
  - **Out Interface**: E1
  - **Out Label**: 100
  - **Operation**: Push

- **Prefix**: 100.1.1.1/32
  - **Local Label**: 100
  - **Out Interface**: E1
  - **Out Label**: 200
  - **Operation**: Swap

- **Prefix**: 100.1.1.1/32
  - **Local Label**: 200
  - **Out Interface**: E1
  - **Out Label**: 300
  - **Operation**: Swap

- **Prefix**: 100.1.1.1/32
  - **Local Label**: 300
  - **Out Interface**: --
  - **Out Label**: --
  - **Operation**: POP

The diagram illustrates the MPLS forwarding process for IP prefix 100.1.1.1/32 with different operations at each router (R1, R2, R3, R4): Push, Swap, and Pop. The loopback interface (100.1.1.1) is also shown.
Why PHP?

Review what R4 has done:
1. First, lookup the label in the LFIB; Remove the label
2. Then, IP lookup and forward IP packet.

Is the first lookup necessary? Can we simplify it?
Penultimate Hop Popping

The implicit NULL label is the label that has a value of 3, the label 3 will never be seen as a label in the label stack of an MPLS packet.
MPLS TTL Processing (1)

- MPLS processes the TTL to prevent loops and implement traceroute.
- By default, TTL propagation is enabled as above.
MPLS TTL Processing (2)

After disabled TTL propagation

- TTL propagation can be disabled to hide the MPLS network topology.
- Disabling TTL propagation makes routers set the value 255 into the TTL field of the label when an IP packet is labeled.
MPLS LSP Ping

R1#ping mpls ipv4 4.4.4.4/32
Sending 5, 100-byte MPLS Echos to 4.4.4.4/32,
timeout is 2 seconds, send interval is 0 msec:
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'L' - labeled output interface, 'B' - unlabeled output interface,
'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
'M' - malformed request, 'm' - unsupported tlvs, 'N' - no label entry,
'P' - no rx intf label prot, 'p' - premature termination of LSP,
'R' - transit router, 'I' - unknown upstream index,
'l' - Label switched with FEC change, 'd' - see DDMAP for return code,
'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 12/14/16 ms
Total Time Elapsed 128 ms
R1#traceroute mpls ipv4 4.4.4.4/32
Tracing MPLS Label Switched Path to 4.4.4.4/32, timeout is 2 seconds
Codes: '!': success, 'Q': request not sent, '.': timeout,
'L': labeled output interface, 'B': unlabeled output interface,
'D': DS Map mismatch, 'F': no FEC mapping, 'f': FEC mismatch,
'M': malformed request, 'm': unsupported tlvs, 'N': no label entry,
'P': no rx intf label prot, 'p': premature termination of LSP,
'R': transit router, 'I': unknown upstream index,
'1': Label switched with FEC change, 'd': see DDMAP for return code,
'X': unknown return code, 'x': return code 0

Type escape sequence to abort.
  0 12.1.1.1 MRU 1500 [Labels: 200 Exp: 0]
L 1 12.1.1.2 MRU 1500 [Labels: 19 Exp: 0] 16 ms
L 2 23.1.1.2 MRU 1504 [Labels: implicit-null Exp: 0] 12 ms
! 3 34.1.1.2 12 ms

Cisco IOS

MPLS LSP Trace
IP MTU

- MTU indicates the maximum size of the IP packet that can still be sent on a data link, without fragmenting the packet.
MPLS MTU Issue

- In MPLS L3VPN network, 2 labels are added into the packet, the labeled packets are slightly bigger than the IP packets. This would lead to the need to fragment the packet.
How to Optimize Fragmentation?

• Solution 1. Change MPLS MTU: Make sure that you configure this value on all the links in the path so that the packets are not dropped.

R1(config)#interface ethernet1/0
R1(config-if)#mpls mtu 1508
R1#show mpls interfaces Ethernet 1/0 detail
Interface Ethernet1/0:
   IP labeling enabled
   LSP Tunnel labeling not enabled
   BGP labeling not enabled
   MPLS not operational
   MTU = 1508

• Solution 2. Change the TCP MSS to be smaller:

R1(config)#interface ethernet 1/0
R1(config-if)#ip tcp adjust-mss 1452

For detailed, please refer to: https://blog.apnic.net/2014/12/ip-mtu-and-tcp-mss-mismatch-an-evil-for-network-performance/
Questions

- Please remember to fill out the feedback form
  - [https://www.surveymonkey.com/r/apnic-20171101-eL2](https://www.surveymonkey.com/r/apnic-20171101-eL2)

- Slides are available for download from APNIC FTP.

- Acknowledgement to Cisco System.
APNIC Helpdesk Chat

Helpdesk

APNIC Helpdesk provides assistance to all on matters related to APNIC Services, such as membership and IP address enquiries.

APNIC Helpdesk offers (through prior arrangement) multi-language phone support for the following: Bahasa Indonesia, Bahasa Malaysia, Bengali, Cantonese, English, Filipino (Tagalog), Hindi, Japanese, Malay, Mandarin, Sinhalese, Tamil and Telugu.

You may also find our FAQs helpful with your enquiries.

Contact details

Helpdesk hours
09:00 to 21:00 (UTC +10)
Monday - Friday
(closed for some public holidays)

Chat
APNIC Live Chat Online
Click here to chat

Skype
Call
Click here to Skype
ID: apnic-helpdesk

Email
helpdesk@apnic.net

Phone
+61 7 3858 3188

VoIP
helpdesk@voip.apnic.net

Fax
+ 61 7 3858 3199

Service Updates

Service announcement: 10 February 2016
Service disruption: APNIC services were disrupted on Wednesday, 10 February 2016
More announcements

Subscribe to APNIC Service Announcements
Learn more about system maintenance
Thank You!
END OF SESSION