January 2015

Dig into MPLS: Transit Tunnel Diversity

Yves VANAUBEL
Pascal MÉRINDOL
Jean-Jacques PANSIOT
Benoit DONNET
Summary

- Motivations
- MPLS Background
- Measurement Campaign
- Label Pattern Recognition Algorithm
- Traffic Engineering Analysis
- Conclusions
Motivations

- Multi Protocol Label Switching (MPLS) [RFC3031] widely used in the Internet

- Actual studies essentially about MPLS discovery [Sommers2011, Donnet2012] and its impact on the topology inference [Flach2012]

- Actual usage of MPLS by operators not yet studied

- **Objective**: Distinguish 3 types of MPLS usage (mono path and transit tunnel with or without Traffic Engineering (TE))

- Measurement campaign (weekly during 10 months with traceroute) and offline analysis of the data
FEC

- **Forwarding Equivalence Class (FEC)**
- Set of packets forwarded in the same way
- Choosing the next hop can be seen as:
  1. Associating the packet to a FEC
  2. Get the next hop based on the FEC
- Can be used for traffic engineering
Different fields of the packet’s header can be used to create FECs (TOS, source and/or destination IP prefix, ...)

Packet associated to a FEC only once, at the entrance of the MPLS network

FEC encoded as a short fixed length value called label

Label sent with the packet
MPLS Label Stack Entries

- **Label Stack Entries (LSE):**
  - 32 bits
  - Inserted between the MAC and the IP layer

<table>
<thead>
<tr>
<th>Label</th>
<th>TC</th>
<th>S</th>
<th>TTL</th>
</tr>
</thead>
</table>

- **Label**: Label value, 20 bits
- **TC**: Traffic Class field, 3 bits
- **S**: Bottom of stack, 1 bit
- **TTL**: Time To Live, 8 bits
MPLS Network

Ingress LSR (LER) → ISP X → Egress LSR (LER)

FH LSR → LSR → LH LSR

IP/1.1.1.0/24 → 1.1.1.1 → 2.2.2.0/24 → 2.2.2.2

LER : Label Edge Router
LSR : Label Switching Router
LSP : Label Switching Path
PHP : Penultimate Hop Popping

Source: 1.1.1.1
Destination: 2.2.2.2
Label Distribution Protocol

- **Label Distribution Protocol (LDP) [RFC5036]**
- Allows distribution FEC-to-label bindings among LSRs
- In this case, FECs are prefixes in IGP routing tables, or loopback addresses
- Downstream operation
RSVP

- **Resource ReSerVation Protocol (RSVP) [RFC2205]**
- Used to allocate resources on the path
- Messages follow IP route
RSVP - TE

- Resource ReSerVation Protocol - Traffic Engineering (TE) [RFC3209]

- RSVP-RESV can piggyback MPLS labels

- Explicit Route Object (ERO) extension:
  - Allows the source to pre-calculate the LSP (not necessary the IP route)
Several LSPs may exist for a given <Ingress, Egress> pair

- **LSP 1** and **LSP 2**: same IPs, different labels
- **LSP 1** and **LSP 3**: different IPs, different labels
Measurement-based classification of MPLS tunnels (traceroute)

Depend on two options:

- **ICMP extension** ([RFC4950]):
  - If an MPLS router must forge an ICMP *time exceeded* message, it should quote the MPLS label stack into it.

- **TTL propagation** ([RFC3443]):
  - The ingress router of an MPLS tunnel should initialize the LSE-TTL with the value inside the IP-TTL field (iTTL).
  - The opposite operation is done by the egress LER (oTTL).
- TTL-propagate activated
Measurement Technique

- No TTL-propagate

```
Ingress LER
IP/to:2.2.2.2 TTL:10

IP/to:2.2.2.2 TTL:9 MPLS/L:10 TTL:255

MPLS/L:10 TTL:254

LSP

Egress LER
IP/to:2.2.2.2 TTL:9
```
Measurement Campaign

- PlanetLab platform
- 1M of destinations from CAIDA data
- 200 Vantage Points (VP), i.e. 5000 destinations / VP
- Scamper with paris-traceroute (ICMP-echo)
- 43 cycles during 10 months (August 2013 to June 2014)
- Each week on Friday noon
- Same destinations and VPs for each cycle
Measurement Campaign

Proportion of paths containing at least one MPLS tunnel
LPR

- Label Pattern Recognition algorithm
- Allow to distinguish multi-FEC from IP load balancing
- Passive classification method (off-line)
- Recognize behaviors of LDP versus RSVP-TE based on labels distribution
- Inter-domain MPLS tunnels seem negligible => focus on transit tunnels
LPR - Static Filters

- Three static filters:
  1. **Intra AS**: IP addresses of an LSP must be in the same AS (removes ~0.5% of LSPs).
  2. **Target AS**: Destination of traceroute must be in a different AS than the LSP (removes ~10% of LSPs).
  3. **Transit Diversity**: For a given <Ingress, Egress> pair (i.e. MPLS tunnel), destinations of traceroute must be in different ASes (removes ~15% of LSPs).
LPR - AS Selection

- Per AS study
- Select Top 15 ASes containing the largest set of IP addresses flagged as MPLS
- Contain almost 75% of the LSPs in the filtered dataset
A temporal filter:

- **Persistence**: Keep an LSP if it was also seen in measurement cycle $X + 1$ or $X + 2$ (removes \( \sim 20\% \) of LSPs)

- Removes noise due to routing changes
LPR - Classification

- Algorithm classifies each MPLS tunnel (i.e. \(<\text{ingress, egress}>\) pair) based on the IP addresses and the labels of the LSP

- Four classes:
  - Mono LSP
  - Multi FEC
  - ECMP
  - Unclassified
LPR - Mono LSP

Trace LSP1:
1. A
2. B - Label L1
3. C - Label L2
4. D

Trace LSP2:
1. A
2. B - Label L1
3. C - Label L2
4. D

Same IPs and same Labels
**LPR - Multi FEC**

**Trace LSP1:**
1. A
2. …
3. B - Label
4. D\(_1\) - Label
5. E - Label L\(_1\)
6. …
7. F\(_1\) - Label
8. G

**Trace LSP2:**
1. A
2. …
3. C - Label
4. D\(_2\) - Label
5. E - Label L\(_2\)
6. …
7. F\(_2\) - Label
8. G

Different labels for at least 1 common IP
LPR - ECMP Mono FEC (1)

❖ Disjoint Routers

Ingress LER
A

…

B

1

D

2

E

1

F

2

Egress LER
G

…

C

Trace LSP1:
1. A
2. …
3. B - Label
4. D₁ - Label L₁
5. E - Label L₂
6. …
7. F₁ - Label
8. G

Trace LSP2:
1. A
2. …
3. C - Label
4. D₂ - Label L₁
5. E - Label L₂
6. …
7. F₂ - Label
8. G

Common IP

❖ Disjoint Routers

Ingress LER
A

…

B

1

D

2

E

1

F

2

Egress LER
G

…

C

Trace LSP1:
1. A
2. …
3. B - Label
4. D₁ - Label L₁
5. E - Label L₂
6. …
7. F₁ - Label
8. G

Trace LSP2:
1. A
2. …
3. C - Label
4. D₂ - Label L₁
5. E - Label L₂
6. …
7. F₂ - Label
8. G

Same label ∀ common IPs
LPR - ECMP Mono FEC (2)

❖ Parallel Links

- Ingress LER
- L1
- A

- L1
- B

- L2
- C

- L2
- D

- L3
- E

- L3
- F

- L4
- PHP

- L4
- EGRESS LER

Same labels along all the LSPs

Different IPs are aliases!

**Trace LSP1:**
1. A
2. ...
3. B - Label L1
4. C1 - Label L2
5. D1 - Label L3
6. ...
7. E1 - Label L4
8. F

**Trace LSP2:**
1. A
2. ...
3. B - Label L1
4. C2 - Label L2
5. D2 - Label L3
6. ...
7. E2 - Label L4
8. F
If PHP is used, the Egress LER does not exhibit labels

May happen that LSPs do not intersect on a common IP

In this case, tunnel arbitrarily tagged as unclassified
LPR - Global View

Filtering and formatting

Classification

Raw Dataset

Explicit MPLStunnels

IntraAS per LSP

TargetAS per LSP

Transit Diversity per <I,E>

Persistence per AS per <I,E> per LSP

Top20AS

Cleaned Data

Mono-LSP

Mono-FEC

Multi-FEC

Parallel Links

Routers Disjoint

≠ Labels on common IP

No common IP

Unclassif.
Gray-tagged ASes exhibit a highly dynamic LSP profile

Multi FEC not widely deployed, and limited to some operators

ECMP highly used and mainly made of parallel links (more than half of load balanced paths)

Results quite similar in time
Traffic Engineering Analysis

- Limitations due to **unclassified** tunnels
- **On traceroute paths**, all previous IPs of a given common IP should belong to the same router (assuming no layer-2 devices and routers answer to probes using the incoming interface)
- Common IP set may be extended with such previous IPs

![Diagram showing the relationships between A, B, and C with A and B (previous to C) should be the same router]
Traffic Engineering Analysis

- Impact of common IP set and temporal filtering:

  - M1: Only common IPs
  - M2: Common IPs + previous IPs
  - M1F: With temporal filtering
  - M2F: Without temporal filtering

![Diagram showing classes distribution across ASes for M1, M1F, M2, M2F]

- M1: Only common IPs
- M2: Common IPs + previous IPs
- F: With temporal filtering
- NF: Without temporal filtering
Conclusion

- New algorithm to reveal TE usage in ASes:
  - Label distribution (mono or multi FEC)
  - ECMP load balancing (parallel links or disjoint routers)
  - Dynamics (temporal evolution of MPLS tunnels)
Conclusion

- Most operators seems to deploy MPLS for LDP purpose
- Some prefer multi FEC and highly dynamic tunnels
- In case of ECMP, multi link present as often as when IP interfaces are not router aliases
References

