MPLS and More

08 May 2015 AN
Who am I

Marijke Kaat
SURFnet / UvA
4 days / 1 day

Work:
@SARA - network management & operations (NOC)

@SURFnet - network management & design

@HvA - lecturing

Education:
@UvA - Computer Science

marijke.kaat@surfnet.nl
m.e.kaat@uva.nl
Multi-Protocol Label Switching

...because its techniques are applicable to ANY network layer protocol.

• It was designed to provide a unified data-carrying service for both circuit-based clients and packet-switching clients which provide a datagram service model; (converged networks)

• It can be used to carry many different kinds of traffic, including IP packets, as well as native ATM, SONET, Frame Relay and Ethernet frames;
Label switching idea brilliant innovation?

- Back then ATM was popular, only encapsulations for flow over ATM were defined;
- ATM did not become the big hit and is mostly replaced now by IP/Ethernet devices;
Multi-Protocol Label Switching

- Next Cisco came with Tag Switching (1997);
- This was brought to the IETF for open standardisation;
- IETF working group involved other vendors and MPLS was defined;
- Tag renamed to Label;

**Really:**
Back then traffic was growing faster than router vendor and service providers could keep up with;

Existing routing equipment was very expensive; performance was not enough (no in hardware forwarding of packets);

Fixed length label lookup was faster;

RFC-3031: Multiprotocol Label Switching Architecture
• ATM switches offered higher-speed interfaces and faster forwarding;
• ISPs were building backbones with ATM switches and routers as edge devices;
• That required building full-mesh networks and that’s a lot of configuration and difficult management;
• Vendors were trying to implement tighter integration between router and ATM switch control planes;

Label idea comes from "label swapping" in Frame Relay, ATM.

**Basically:**

★ edge device applies tag
  ★ tag switch forwards according to label swapping table
  ★ edge device removes tag and forwards packet
• This works as long as you **don't try** to do it per flow;
  - (nice for a 20 minute phone call, not for thousands of sessions per second to a web server, it does not scale)

• It was invented for fast(er) "**routing**";
• Each flow might be special but numerous flows share the same forwarding behaviour;
• All packets with same label follow the same path;

• LSP is similar to a Virtual Circuit in ATM and Frame Relay;

**MPLS is just a tool** (every single network technology is a tool :-)

**NO** network technology has ever managed to magically create additional bandwidth...
Multi-Protocol Label Switching

- Protocol used in the core of networks
- Single domain (ISP)
- Used for Traffic Engineering (TE)

LER = Label Edge Router (or PE = Provider Edge router)
LSR = Label Switching Router (or P = Provider router)

LSP = Label Switched Path: unidirectional path between LERs

LER = Label Edge Router (or PE = Provider Edge router)
LSR = Label Switching Router (or P = Provider router)
Multi-Protocol Label Switching

- MPLS header applied to packet
  [This header is put between layer-2 and layer-3 header (shim header) in IP]
  
  Label field= 20 bits
  EXP field = 3 bits
  S field = 1 bit
  TTL field = 8 bits
  
  Label = number, picked by the router (local)
  EXP = experimental bits, for Class of Service (*)
  S = Bottom of Stack
  TTL = Time-to-Live (to detect loops)

FEC = Forwarding Equivalence Class

- The ingress router receives packet and determines to which FEC it belongs;
- Packets which should be forwarded in the same manner belong to same FEC;
- Forwarded with the same label (over the same LSP);

LIB = Label Information Base

- Mapping between previous hop (incoming port, label) and FEC;
- Mapping between FEC and next hop (outgoing port, label);
- Each router has its own LIB, generates LFIB (Label Forwarding Information Base);

(*) Renamed to Traffic Class field RFC-5462
How are bindings between labels and FECs distributed through network?
You need routing and signalling;

Manual configuration not an option, need protocol;
2 options: invent new protocol or extend existing protocol to carry labels;

Both were done:
New protocol: **LDP** Label Distribution Protocol
Two existing protocols: **RSVP** and **BGP**

⇒ Without all the details: what does what and how?
LDP - made by IETF [RFC-5036]

Fundamental concept in MPLS is that 2 LSRs must agree on the meaning of labels used to forward traffic. Protocol used where one LSR informs another of the LABEL BINDINGS it has made.

• Specifically designed for label distribution - does nothing else but that, no routing, in fact it relies on an IGP for all routing decisions;

• UDP discovery and TCP session with peers;
• Adjacent LSRs inform each other of the label bindings;
• An IGP protocol is configured on all LSRs;
• New IGP routes lead to new label bindings;
• Labels can be withdrawn when IGP routes are no longer valid;
• Hard-state;
  Expected to work until explicitly torn down
Label Distribution - Control Plane

• LDP works between directly connected neighbors or peers;
• Peers are automatically discovered (via multicast to well-known UDP port);

  • **Initialization**: exchange information regarding features and modes supported;
  • **Next**: information regarding binding Labels and FECs exchanged;
  • **After discovery** a TCP session is established and LDP session is set up;

  • [why chosen to use TCP? Reliable delivery and incremental updates, not periodic refreshes]
  • To keep session up keepalive messages are sent.

  **Label messages**: advertise new labels, withdraw labels
- LSR A receives mapping for Label N for FEC F from peer LSR B;
- LSR A will use Label N for forwarding if and only if B is on the IGP shortest path for destination F from A’s point of view;
  OR: LSPs set up via LDP follow the IGP shortest path and LDP uses IGP to avoid loops;
- LSPs shift with IGP path changes;
  - Danger of blackholing/looping during reconvergence;

But who assigns the labels? (so LDP can distribute the bindings between Label and FEC).
Goal is to build a forwarding table with mapping between incoming label and outgoing label;
➡ Routers pick the label values.

The MPLS architecture uses downstream label assignment: router expects to receive the traffic with label it picked locally.

Called downstream because label assigned to traffic at point X was picked by a router who is one hop further down in the direction of the traffic flow from X.
Prefixes distributed with OSPF/ISIS

PE1 pushes L4

PE1 assigns a Label (L1) to its own L0 address (FEC) and advertises that to its LDP peer X.

PE1 evaluates whether PE1 is on the IGP shortest path for that FEC. If successful X assigns L2 for FEC PE1, installs forwarding state swapping L2 and L1 and advertises a binding for L2 and FEC PE1 to Y.

Y will do similar processing. The LSP setup proceeds from egress to ingress.

Label actions: Push to the stack, Swap top label, PoP from the stack; S-bit is set to 1 in MPLS header if label is last label on the stack;
• Label 3 is announced by router $B$ to its neighbor, 3 is a special value, Implicit NULL label [RFC-3032];
• This triggers **Penultimate Hop Popping (PHP)**
  • the LSR ($E$) before the LER ($B$) pops the label and forwards normal IP packet to LER ($B$);
  • simplifies processing at LER (saves one lookup);
  • default behaviour of most implementations, not mandatory;
RSVP-TE  Resource ReserVation Protocol

• RSVP was invented before MPLS;
• To create bandwidth reservations for individual traffic flows in network as part of the int-serv model;
• Its mechanism is to reserve bandwidth along each hop of a network for an end-to-end session;

→ Doesn't scale (create, maintain, tear-down state for each traffic flow!), so it is not/hardly used.

• RSVP extensions for MPLS to create and maintain LSPs and to create associated bandwidth reservations [RFC-3209];
• Better scaling (single LSP can carry all traffic between ingress and egress router pair, not per flow);
RSVP-TE

- RSVP-signaled LSP does not necessarily follow IGP shortest path;
- Extensions allow for explicit routing (specify entire path or specific transit nodes)

Explicitly Routed LSP: An LSP whose path is established by a means other than normal IP routing. Bandwidth reservation is optional

- Creation of RSVP-signaled LSP initiated by ingress router by sending RSVP Path message;
- Destination is the egress router;
- Transit routers inspect the message and make modifications (define label, check and reserve bandwidth);
  - Path message:
    - label request object, Explicit Route Object (ERO), Record Route Object, Sender Tspec
  - ERO contains addresses of nodes through which the LSP must pass;
**RSVP-TE**

- in response the egress router sends an RSVP Resv message, this follows the reverse path back to ingress;
- establishes the LSP (send label in Resv message);

- Resv message:
  - label object, Record Route Object

  With Record Route Object Object
  routers can check if the path is loop-free

- Path and Resv messages travel hop-by-hop through network - establish state at each node;
- Periodic exchange of messages after establishment to refresh the state (if missed LSP is torn down);
- RSVP-signaled LSPs follow single path from ingress to egress (even in case of multiple available paths);
MP-BGP Multiprotocol Extensions to BGP

- It supports multiple address families, easy to define and carry new types of reachability information and associated attributes;
- Advertise prefix and label(s) associated with it;

[RFC-3107] Carrying Label Information in BGP-4:
- The label mapping information for a particular route is piggybacked in the same BGP Update message that is used to distribute the route itself.

- Can be used inter-domain (between AS-es BGP is used);
- Often BGP is already used so no need for another protocol;
- This is used for Layer3 VPN between sites interconnected by MPLS (provider) core network;
- Each VPN has its own VRF (Virtual Routing and Forwarding instance);
- MPLS forwarding uses stacked labels:
  - outer label for LSP forwarding
  - inner label to differentiate between different VPNs
Now back to the larger world...

Due to Moore’s Law lookup speed is no longer the biggest problem, but since 1997 a lot of new ways to use MPLS and Family have been found...
Traffic engineering

What is it?
• Process of manipulating traffic on an (IP) network to make better use of capacity;
• Is not network engineering, but linked;
• Reduce the overall cost of operations by more efficient use of bandwidth resources;

Just with IP and IP routing protocols difficult:
- tweaking link cost or weight to influence IGP behaviour.
- availability of resources (e.g. bandwidth) not taken into account.

• IGPs distribute network topology information through network;
• Can be used to calculate the routes of LSP automatically;
• When required to establish LSPs not following IGP routes, with guaranteed QoS characteristics and backup LSPs that avoid single points of failure you need more:

Traffic Engineering extensions “-TE”

RFC-3272: Overview and Principles of Internet Traffic Engineering
Traffic engineering

• Cost optimisation (better utilisation of network resources);
• congestion management;
• dynamic services & traffic profiles;
• Efficient routing (predictable, deterministic paths);
• Availability/ resilience / fast restoration;
• QoS / separate realtime latency-critical services from other traffic;

“MORE CONTROL”

• MPLS-TE: set of extensions to MPLS
  • explicit or constraint based routing;
  • use RSVP-TE to set up explicit paths;
  • bandwidth reservation;

More extensions: fast rerouting, restoration, QoS, Shared Risk Link Groups, link coloring, make-before-break, pre-emption, auto-bandwidth, etc.
MPLS-TP Transport Profile

• In 2006 the ITU-T started with its own MPLS-like technology: T-MPLS or Transport MPLS
• Continued as joint effort of IETF together with ITU;
• Now called: MPLS-TP;

• MPLS-TP is set of extensions to IP MPLS feature set that fulfils packet transport requirements;
  • MPLS is bi-directional LSPs (follow same path both ways)
  • No LSP merging, no PHP
  • No ECMP (Equal-cost multi-path routing)
  • Does not support connectionless mode
  • Adds carrier grade OAM

• MPLS-TP applies additional constraints, eliminates some complex functions that make networks unpredictable and non-deterministic
Tools to manage/monitor the network

- Continuity Check
- Connectivity Verification
- on-demand Route tracing
- Alarm suppression
- performance monitoring (delay, loss, jitter)

Lots of controversy on how to implement OAM - “work in progress”
IETF: BFD and LSP Ping
ITU-T: Y-1731

“packet transport requirements”

- More the ITU view than the IETF view;
- Requirements for things from SDH/ATM/Frame Relay etc. the virtual circuit instead of the packet switched networks;
- Clear separation of control plane and data plane
**VPLS** Virtual Private LAN Service

- A VPLS is a provider service that emulates the full functionality of traditional LAN.
- A Layer 2 Virtual Private Network (VPN)
- VPLS is "private" in that CE devices that belong to different VPLSs cannot interact.
- VPLS is "virtual" in that multiple VPLSs can be offered over a common packet switched network (over IP or MPLS network).

---

[RFC-4026: Provider Provisioned Virtual Private Network (VPN) Terminology]
VPLS Virtual Private LAN Service

• Ethernet service: frames sent to broadcast addresses and unknown dest MAC addresses are flooded to all ports
  • all unknown unicast, broadcast and multicast frames are flooded over the corresponding PWs to all PE nodes participating in the VPLS
• Responsibility of service provider to create loop-free topology
• Full-mesh of pseudo-wires connecting the edge sites
  • Using LDP for Signaling [RFC-5461]
  • Using BGP for Auto-Discovery and Signaling [RFC-5462]

• Number of limitations in redundancy, multicast, multihoming, provisioning simplicity
• New RFC on defining the requirements for a new solution: Ethernet VPN (EVPN) [RFC-7290 (May 2014)]
MPLS and more... a lot more...

• MPLS over 15 years old but...
• Still lot of activities on standardisation;

Just look at the number of drafts in the MPLS WG of the IETF:

https://datatracker.ietf.org/wg/mpls/
Next week’s (12 May) reading:

The IEEE 802.11 Universe
By: G. Hiertz et al
In” IEEE Communication Magazine vol.48 no.1 2010

Thanks for your attention!

Questions? marijke.kaat@surfnet.nl