HELSINKI UNIVERSITY OF TECHNOLOGY

MPLS

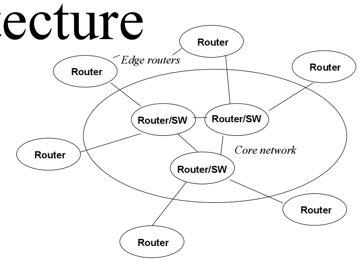
Lecture for S-38.192

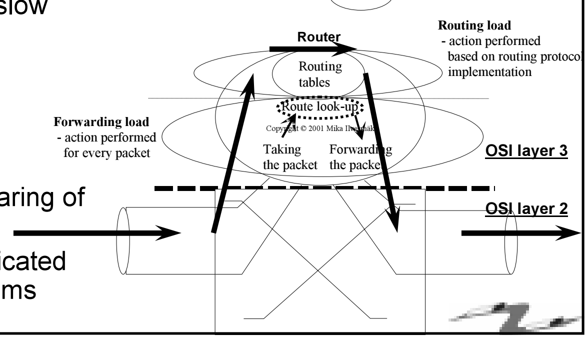
10.2.2005 Mika Ilvesmäki

Networking laboratory

IP router architecture

- The basic network block
  - Routing/control and forwarding
- The route is looked up for every packet
  - This is robust but slow and redundant
- No knowledge of previous or future packets
- FIFO-queues
  - Statistical (fair) sharing of resources
  - Also more sophisticated queuing mechanisms



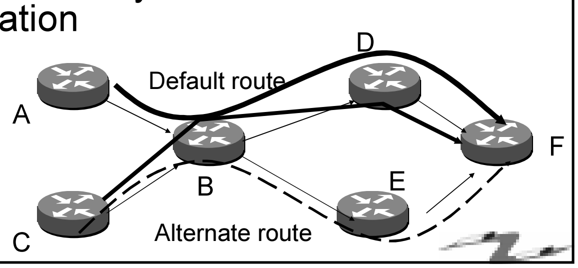


Routing, protocol, algorithm

- Routing is discovering the network structure and topology
- Routing is done with ROUTING PROTOCOLS in routers
  – Exchange of router positional information
    - distance to places, costs etc.
- Routing **protocols implement** routing **algorithms**
  – Dijkstra SPF, Bellman-Ford etc.

Routing in the Internet

- Traditional Internet routing is based on finding the shortest path to the destination regardless of the source
  – No possibility to optimize resource usage
  – Destination based routing offers the possibility to use only the default route
- shortest path refers usually to the number of hops to the destination
  – OSPF, RIP, etc.

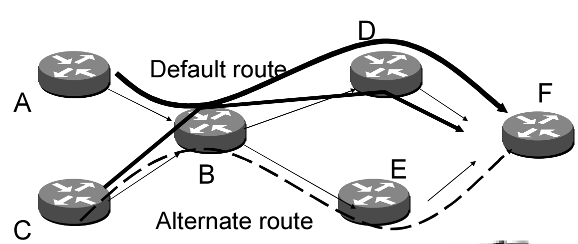


Constraint-Based routing

- Plain IP routing aims to find a path that optimizes a certain scalar metric (number of hops, typically)
- Constraint-Based routing aims to find a path that optimizes a certain scalar metric without violating a set of constraints.
- Constraints:
  - minimum bandwidth, delay or other performance constraint
  - link inclusion or exclusion or other administrative constraint
  - combination of performance and administrative constraints

What if…?

- Let’s label the route A-B-D-F with label X
- Let’s label the route C-B-E-F with label Y
  - And label the packets either with X or Y as we choose…
  - And we have (very simplified) MPLS



Multiprotocol Label Switching

- a.k.a multi-layer routing or IP switching
  - Distribute the cached forwarding info using IP address independent labels -> **separation of route lookup and forwarding decision**
    - Or concatenating several hops into one…
    - May save the forwarding resources but adds the need to distribute the label info -> increases the total work done
- Standardization work began 1997 in IETF
- Combines features of several IP switching solutions
  - Mainly Cisco Tag switching
- Separate signalling and label exchange protocol (LDP, CR-LDP, RSVP-TE, BGP)

Status of the standardization effort

HELSINKI UNIVERSITY OF TECHNOLOGY
Mika Ilvesmäki, Lic.Sc. (Tech.)

3.8.2004: http://www.ietf.org/html.charters/mpls-charter.html

- MPLS workgroup drafts
  - LSP Hierarchy with Generalized MPLS TE
  - Link Bundling in MPLS Traffic Engineering
  - Multiprotocol Label Switching (MPLS) Management Overview
  - Graceful Restart Mechanism for BGP with MPLS
  - Fast Reroute Extensions to RSVP-TE for LSP Tunnels
  - Detecting MPLS Data Plane Failures
  - Multiprotocol Label Switching (MPLS) Label-Controlled ATM and Frame-Relay Management Interface Definition
  - Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base for Fast Reroute
  - Maximum Transmission Unit Signalling Extensions for the Label Distribution Protocol
  - Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)
  - Requirements for Point to Multipoint Traffic Engineered MPLS LSPs
  - Definition of an RRO node-id subobject
  - MPLS Traffic Engineering Soft preemption
  - Traffic Engineering Link Management Information Base
  - Label Switching Router Self-Test
  - Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using RSVP-TE
  - Removing a Restriction on the use of MPLS Explicit NULL

MPLS RFCs

- MPLS workgroup RFCs as of August 3rd, 2004
  - Requirements for Traffic Engineering Over MPLS (RFC 2702)
  - Multiprotocol Label Switching Architecture (RFC 3031)
  - MPLS Label Stack Encoding (RFC 3032)
  - Use of Label Switching on Frame Relay Networks Specification (RFC 3034)
  - MPLS using LDP and ATM VC Switching (RFC 3035)
  - LDP Specification (RFC 3036)
  - LDP Applicability (RFC 3037)
  - VCID Notification over ATM link for LDP (RFC 3038)
  - The Assignment of the Information Field and Protocol Identifier in the Q.2941 Generic Identifier and Q.2957 User-to-user Signaling for the Internet Protocol (RFC 3033)
  - MPLS Loop Prevention Mechanism (RFC 3063)
  - Carrying Label Information in BGP-4 (RFC 3107)
  - RSVP-TE: Extensions to RSVP for LSP Tunnels (RFC 3209)
  - Applicability Statement for Extensions to RSVP for LSP-Tunnels (RFC 3210)
  - Constraint-Based LSP Setup using LDP (RFC 3212)
  - Applicability Statement for CR-LDP (RFC 3213)
  - LSP Modification Using CR-LDP (RFC 3214)
  - LDP State Machine (RFC 3215)
  - MPLS Support of Differentiated Services (RFC 3270)
  - Framework for IP Multicast in MPLS (RFC 3353)
  - Time to Live (TTL) Processing in MPLS Networks (Updates RFC 3032) (RFC 3443)
  - Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE) (RFC 3477)
  - Framework for MPLS-based Recovery (RFC 3469)
  - Graceful Restart Mechanism for Label Distribution Protocol (RFC 3478)
  - Fault Tolerance for the Label Distribution Protocol (LDP) (RFC 3479)
  - Signalling Unnumbered Links in CR-LDP (Constraint-Routing Label Distribution Protocol) (RFC 3480)
  - Applicability Statement for Restart Mechanisms for the Label Distribution Protocol (LDP) (RFC 3612)
  - Definitions of Textual Conventions for Multiprotocol Label Switching (MPLS) Management (RFC 3811)
  - Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base (RFC 3812)
  - Multiprotocol Label Switching (MPLS) Label Switching Router (LSR)Management Information Base (RFC 3813)
  - Multiprotocol Label Switching (MPLS) Forwarding Equivalence Class To Next Hop Label Forwarding Entry (FEC-To-NHLFE)Management Information Base (RFC 3814)
  - Definitions of Managed Objects for the Multiprotocol Label Switching, Label Distribution Protocol (LDP) (RFC 3815)

Features of MPLS

- Datalink independent
  - It seems that MPLS is capable of providing almost the same as ATM (flexibility in traffic management options). However, as with ATM, this comes with the high cost of extremely demanding network management.
  - Unicast and (multicast) capable
  - IntServ and DiffServ compatible (might be considered as QoS enabler, however, MPLS is not an QoS architecture in itself)
- MPLS is not
  - only a way to make switches to efficient routers
  - a replacement for traditional routing
- MPLS advantages (RFC3031):
  - Packet forwarding can be done by nodes not capable of analyzing IP packets (fast enough)
  - Assignment of packets to different *forwarding equivalence classes (FEC)* at the ingress may be based on variety of information
  - Forwarding decisions may be based on ingress router

Forwarding Equivalence Class

- Forwarding procedures for certain packets form a FEC
  - Procedures include
    - Next hop routers, queuing info
  - Based on network header information
- In MPLS the forwarding procedure is bound to a label
  - Mark different packets with different labels (-> FECs) to achieve different treatment of packets
    - QoS, optimal resource usage, customer wishes

MPLS core technologies

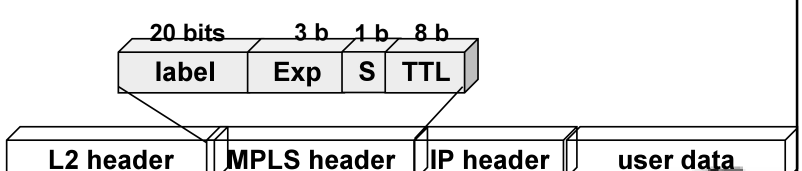
- The LSR, Label switch router
- Label swapping (forwarding mechanism)
- The LDP, Label distribution (protocol)
  - The former technologies act as mechanisms that form paths, Label Switched Paths (LSPs) in the network.
    - Paths may be traffic, topology or reservation (RSVP) initiated

MPLS Components - I

- LSR - Label switch router
  - ordinary IP router with the ability to switch on layer 2
  - has a specialized protocol (LDP) to co-operate with neighboring routers
  - LER - label edge router is able to communicate outside the MPLS domain

MPLS Components - II

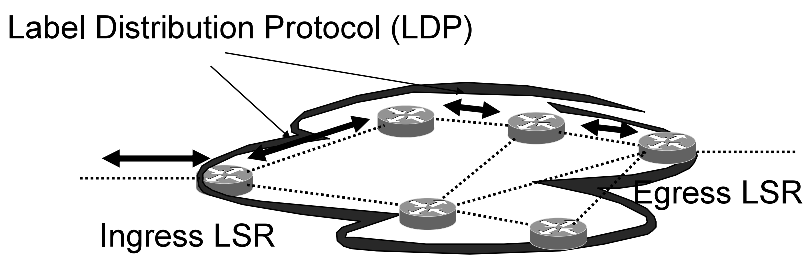
- Label
  - Use the existing connection identifiers (ATM VPI/VCI) or *update* with 32 bit L2/L3 shim
- LDP - Label distribution protocol
  - Distribute the knowledge on label use
- Traffic, topology or reservation (RSVP) initiated LSP creation



Label distribution protocol

- Labels may be distributed by piggybacking on existing protocol (BGP or RSVP) or with LDP (RFC 3036 stds track)
  - QoS reservations made possible with CR-LDP or RSVP-TE
- LDP is built over TCP (keepalive), uses TLV messages
  - (Almost) Infinite extendability
- Message types
  1. Discovery
  2. Adjacency
  3. Label advertisements
  4. Notification

Creating the Label Switched Path



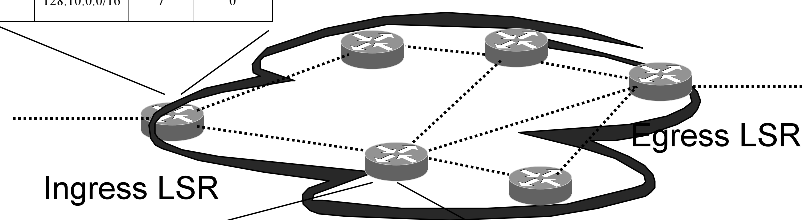
- We still need routing protocols to find the paths
  - QoS routing in the future
- What initiates the LSP creation?
  - Traffic (Reactive), prediction of future traffic (Proactive, control)

Using labels/tags in forwarding

- Different FECs (Forwarding Equivalence Classes) for different traffic

- LSR extracts the label from the incoming label, uses it to access the forwarding table, and places a new (outgoing) label into the packet

In Label	Prefix	Out Label	Out Interface
	199.1.1.0/24	6	0
	128.10.0.0/16	7	0



In Label	Prefix	Out Label	Out Interface
6	199.1.1.0/24	1	1
7	128.10.0.0/16	2	1

LIB - Label information base

Failure in the link

- Upon link failure the traffic has to be rerouted
  - Since paths are determined by the LSR at the end of the trunk, rerouting has to be performed at the same place
  - Rerouting uses IGP, RSVP or CR-LDP information to establish the new route.

MPLS and Fast rerouting

- Aim: Reduce packet loss during routing transients
- FRR:
  - Constraint-Based routing forms an alternate go-around for the damaged link
  - When failure occurs the original label is preserved and another label is pushed above the original
    - Label stacking

Binding labels to a FEC

- Both local and remote methods may used simultaneously
  - Local binding: LSR creates the binding with a label that is chosen and assigned locally
  - Remote binding: LSR receives label binding info from another LSR
    - Remote binding options: **Downstream (always in MPLS)** or Upstream
      - Downstream on demand (request) and unsolicited downstream (distribute)

Creating and using the label space

- Control of label information distribution
  - Independent
    - Advertise the label assignments to neighbors
  - Ordered
    - Label assignment proceeds in an end-to-end fashion
      - Ingress or egress initiated
- Saving the label information
  - Liberal or conservative
- Save the label space!
  - Use label merging (and lose information on the packet arrival data)

Stacking the labels

- It is possible to tunnel/stack MPLS-packets within/over MPLS-packets
  - To separate the core network from the edges
- Use the S-bit in the shim-header
  - When set you are at the bottom of the stack
- Ultimate or pen-ultimate LSRs strip the stacking away.

Label stacks

- Operations: Push, pop, swap
- Label stacks are used to
  - Merge and split traffic streams
    - Path sharing
    - Aggregate traffic trunks
  - Limit the spread of the routing information
- Enabler for MPLS VPNs.

GMPLS

- Generalized MPLS
  - MPLS supports packet and cell switching
  - GMPLS supports also TDM-, Wavelength- and physical port switching
  - uses generalized labels to include the new switching methods
  - opens routers to have a view into the transmission network topology

What can you do with MPLS?

- Integrate ATM with MPLS
  - MPLS acts as an VC aggregator
  - RFC 3035 (std)
- Traffic Engineering (TE)
  - Direct streams of traffic to non-default paths and balance the network load
    - Because of separated routing and forwarding
  - QoS/CoS with paths and FECs -> Service architectures (DiffServ)
    - CR-LDP
- VPN / Virtual Private Networks
  - Private traffic travels within public network
  - dedicated paths/FECs for VPN traffic
- Multicast (work in progress)
  - Labels to LSP trees

MPLS and TE

- Explicitly form LSPs (not using standard IP routing)
  - Map packets to FEC
  - Map FEC to a traffic trunk
    - Traffic trunk is an aggregation of traffic flows of the same class and path
  - Map trunks to an LSP formed with constraint-based routing
  - Map LSP to physical network

MPLS and IntServ

- Create and distribute binding info between flows (FlowSpec) and labels.
  - The packets that fall under the FlowSpec need their own FEC
- A new RSVP_ LABEL object is created and carried in the RSVP Resv-message
- As the Resv-message propagates upstream the reservation and the appropriate label bindings are created.
- Note, that only the edge router first receiving the flows needs to do packet/flow –mapping.
  - And these mapping rules could be extended to include a variety of packets to be associated with a reservation.
    - Guaranteed service –pipe.

MPLS and DiffServ

- How to map BAs onto LSPs?
  - LSPs carry several ordered aggregates
    - Exp-field separates classes from each other
    - Maximum of 8 (3 bits) BAs in single LSP
    - Exp->PHB mapping explicitly signalled or pre-configured
  - LSPs carry a single OA
    - packet treatment indicated in the label-field
  - Requires careful management of LSPs
  - Requires extending the signalling protocol (RSVP_TE or CR-LDP)

Problems with MPLS

- Complex management of the MPLS network
  - Traffic or topology based path creation
- Increase in overhead if the label is not present in layer 2
  - However, the overhead is not that large as it is with tunneling solutions
- MPLS may easily lead to unoptimal use of routes
  - The shortest path is not used as the primary route
- Where are the QoSR algorithms and protocols?
- No support for multicast, yet (Check RFC 3353/inf).

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ilvesmäki, Lic.Sc. (Tech.)

Summary of MPLS

- Mostly a new way to use routing information in a flexible way
  - Acts as an enabler for Quality of Service networks
- Standardization is well on the way
- Competing solutions on the edge of release or just released
  - Some commercial services available

Sources of information

- MPLS-workgroup in IETF
  - http://www.ietf.org/html.charters/mpls-charter.html
- MPLS resource center
  - http://www.mplsrc.com/
- MPLS tutorial (one of many)
  - http://www.nanog.org/mtg-9905/ppt/mpls/
- MPLS forum
  - http://www.mplsforum.org/
- www.google.com (type in MPLS and wait...)

To make the point the recent (edited) words from Fred Baker in an answer to anti-MPLS whining:

Date: Tue, 09 Jan 2001 15:12:32 +0800, **From: Fred Baker <fred@cisco.com>**

At 1/4/01, **someone wrote**:

>Despite the negative comments recently about MPLS from Fred and IESG members, **MPLS/TE solves real problems and is seen as easily deployable**, particularly relative to such things as Nimrod.

I'm sorry you see me as anti-mpls and anti traffic engineering. I'm not. What I am anti, if anything, is discarding IP routing in favor of MPLS. Yes, you see MPLS LSPs as extending IP routing, and bully for you. If you attended the CEOT BOF or the IPO BOF, you got a flavor of what I'm dealing with on other fronts. If a service provider wants to use MPLS to accomplish goals like traffic engineering or VPNs, I'm all for that.

But on the one hand I have a short list of folks who have deployed MPLS, and a long list of folks who don't want to - they want the same goals met in IP routing. On the telco and research side, I also have a long list of folks who are saying "well, if I can't make the world be ATM in the ITU, I'll call it MPLS and make the world be ATM in the IETF.“ The IETF may someday decide to go there, but I'm sufficiently narrow-minded that it won't do so on my watch. Of course, my watch ends in a couple of months :^) **Further, I also worry about people deciding that "MPLS is the answer, now what was your question?" To pick on one pet peeve, some bunch of jerks, probably from my company, are promulgating the belief that MPLS has something to do with QoS. You and I know it doesn't. Traffic engineering is a way to reduce the total cost of a network by maximizing the use of the individual links. What it ensures, if anything, is a slightly longer path for the average route (instead of taking the overloaded direct link from here to there, use the underutilized paths from here to over-thar, and then from over-thar to there). Neither increasing the mean traffic rate on a link nor increasing the total number of interfaces that a message must cross is a recipe for making delay more constant or reducing it. MPLS can certainly be used *with* bandwidth allocation to engineer peak rates (and therefore queue depths) so that delay is minimized and stabilized, and it can certainly be used *with* other QoS technologies to accomplish QoS goals. But it is not in and of itself a QoS solution: it is the antithesis.** For example, there is rather a largeish set of people who like IPSEC tunnels running over IP networks for certain classes of solutions. Is there a reason they should be forced into doing something with MPLS? Can the IETF be open-minded enough to keep that model in view rather than focusing all of its energies on MPLS? What I said rather a bunch of times at the IETF was that I was interested in the Internet Engineering Task Force being used as a venue to engineer solutions for the Internet. I said that I was willing to look at sub-ip technologies (mpls traffic engineering being an example) to the extent that they are useful for IP; I was not interested in going the extra twelve steps to taking on the general interworking problem that the ITU loves (make it be native voice on IP here, native ATM voice there, and native circuit switch voice somewhere else), or to try to put the ITU out of business. If we can make IP work on optics, perhaps using adapted MPLS technologies, fine. If voice-on-optics can use exactly the same technology to accomplish its goals, fine by me. But voice-on-optics is a non-goal; if it won't work, use voice-on-IP on Optics, or go somewhere else and engineer an appropriate solution. I was pretty frustrated to hear people instantly say "so I don't understand you, you're being ambiguous". I view that as intentional non-understanding - there is none so difficult to explain something to as someone who has decided that he doesn't like what you're saying is has therefore stopped listening. MPLS, by the way, is not a routing technology, nor is it normally interdomain. What on God's green earth does your view of MPLS and traffic engineering have to do with getting a better interdomain routing technology out there? I don't do well with whining...