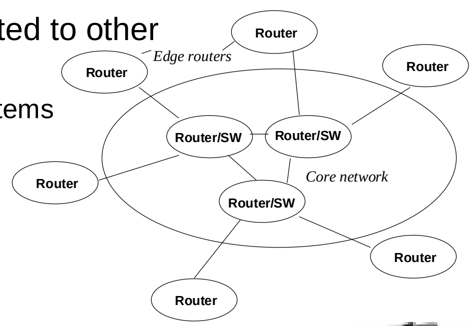
MPLS

Lecture for S-38.180 QoS in the Internet

10.10.2002 Mika Ilvesmäki

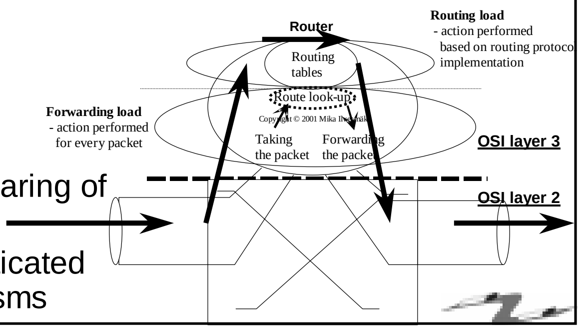
Internet structure

- Internet basically consists of routers connected to each other
  - Hierarchical structure
    - Edge routers, Core routers
    - Networks connected to other networks
      - Autonomous systems



IP router architecture

- The basic network block
  - Routing 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

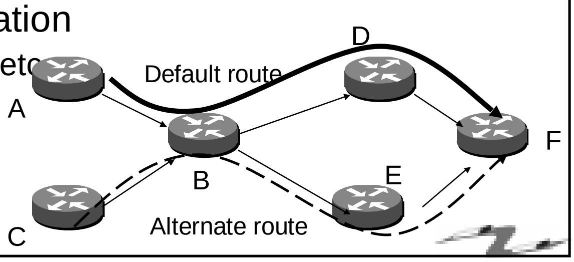


Routing, protocol, algorithm

- 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

- Current 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, BGP, etc



Multiprotocol Label Switching I

- a.k.a multi-layer routing or IP switching
  - Reduce the workload of the standard router by “caching” the forwarding decision to the link layer (OSI 2), i.e. before it reaches the forwarding processor
  - Distribute the cached forwarding info -> separation of route lookup and forwarding decision
- Map flows to (ATM) connections
  - Switching is quicker than routing
- Assumption: Routers are too slow to forward enough traffic.
  - What about Gbit-routers?

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

Multiprotocol Label Switching II

- Standardization work began 1997 in IETF
- Combines features of several IP switching solutions
  - Mainly Cisco Tag switching
- Control/topology driven with data driven capabilities
- 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.)

24.6.2002: http://www.ietf.org/html.charters/mpls-charter.html

- MPLS workgroup drafts
  - Definitions of Managed Objects for the Multiprotocol Label Switching, Label Distribution Protocol (LDP)
  - Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base
  - Framework for IP Multicast in MPLS
  - Multiprotocol Label Switching (MPLS) Label Switch Router (LSR)Management Information Base
  - Improving Topology Data Base Accuracy with LSP Feedback in CR-LDP
  - LSP Hierarchy with Generalized MPLS TE
  - Framework for MPLS-based Recovery
  - Multiprotocol Label Switching (MPLS) FEC-To-NHLFE (FTN) Management Information Base
  - Fault Tolerance for LDP and CR-LDP
  - Multi Protocol Label Switching Label Distribution Protocol Query Message Description
  - Signalling Unnumbered Links in RSVP-TE
  - Signalling Unnumbered Links in CR-LDP
  - Definitions of Textual Conventions and OBJECT-IDENTITIES for Multi-Protocol Label Switching (MPLS)Management
  - Link Bundling in MPLS Traffic Engineering
  - Link Bundling Management Information Base
  - Multiprotocol Label Switching (MPLS) Management Overview
  - Time to Live (TTL) Processing in MPLS Networks (Updates RFC 3032)
  - Graceful Restart Mechanism for BGP with MPLS
  - Graceful Restart Mechanism for LDP
  - Fast Reroute Extensions to RSVP-TE for LSP Tunnels
  - Detecting Data Plane Liveliness in MPLS

MPLS RFCs

- MPLS workgroup RFCs as of June 24th, 2002
  - Requirements for Traffic Engineering Over MPLS (RFC 2702)
  - Use of Label Switching on Frame Relay Networks Specification (RFC 3034)
  - MPLS using LDP and ATM VC Switching (RFC 3035)
  - Multiprotocol Label Switching Architecture (RFC 3031)
  - LDP Specification (RFC 3036)
  - MPLS Label Stack Encoding (RFC 3032)
  - 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)
  - Applicability Statement for Extensions to RSVP for LSP-Tunnels (RFC 3210)
  - RSVP-TE: Extensions to RSVP for LSP Tunnels (RFC 3209)
  - 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)

Forwarding Equivalence Class

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

Features of MPLS

- Datalink independent, not just ATM
  - 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
- 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 a FEC at the ingress may be based on variety of information
  - Forwarding decisions may be based on ingress router

MPLS primary objectives

- Primary objectives
  - Improve routing performance
    - Routing is one way to manage resources in the Internet
    - Traffic engineering
  - Improve scalability
  - Obtain flexibility to introduce new services
    - VPNs

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

MPLS implementation issues

- What initiates the connection set up?
  - Incoming traffic?
  - Knowledge on network topology?
- How are the label bindings distributed
- For whom are the connections meant for?
  - Users and/or application flows
    - # of flows?
  - Traffic aggregates
    - ability to provide for user needs?

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

MPLS core technologies

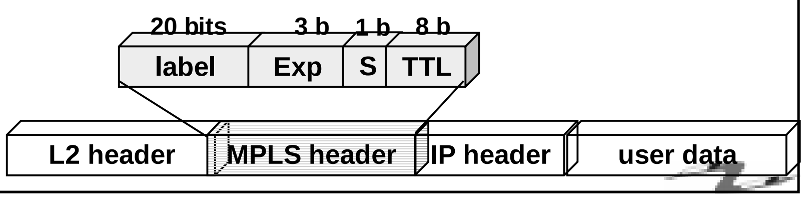
- The LSR, Label switch router
- Label swapping (forwarding mechanism)
- The LDP, Label distribution (protocol)
  - The former technologies act as mechanisms that form paths, Label Switched Paths (LSPs) in the network.
    - Paths may be traffic, topology or reservation (RSVP) initiated

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



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

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
- LDP is built over TCP (keepalive), uses TLV messages
  – (Almost) Infinite extendability
- Message types
  1. Discovery
  2. Adjacency
  3. Label advertisements
  4. Notification

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

Using RSVP-TE for label distribution

- New functions:
  – Label distribution
  – Explicit routing, rerouting, route tracking
  – Bandwidth/Resource reservation
- New objects (cf to slide 7 in RSVP-lecture)
  – PATH-message
    - LABEL_REQUEST
    - EXPLICIT_ROUTE
    - RECORD_ROUTE
    - SESSION_ATTRIBUTE
  – RESV-message
    - LABEL
    - RECORD_ROUTE

RSVP-TE in action

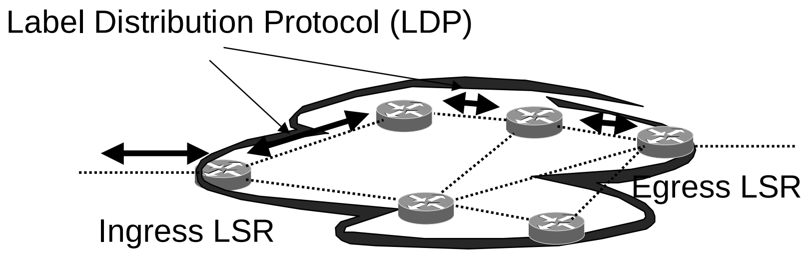
- Addition of Label_request –message in RSVP *PATH*-message
  - Downstream label allocation
- Addition of Label –object to be carried in RSVP *RESV*-message
  - Labels propagate upstream in the RESV-message
- LSPs are set up with FF-reservation

Comparing RSVP_TE and CR-LDP

Property	CR-LDP	RSVP_TE
Transport mechanism	Transport on TCP (reliable)	Raw IP packets (unreliable)
State management	Hard state	Soft state; needs per-flow refresh management
Msgs required for LSP set-up and maintenance	Request, mapping	Path, Resv, Resv_Conf
Base architecture	Based on LDP for MPLS	Based on RSVP, may require major changes
Signalling of QoS and traffic parameters	Can signal DiffServ and ATM traffic classes	Extendable, currently based on IntServ
Types of LSPs	Strict, loose, and loose pinned	Strict and loose, no pinning
Models of label distribution and LSP set-up	All modes	Only downstream on demand
Failure notification	Reliable procedure	Unreliable procedure
Loop detection/prevention	Employs path vector TLV to prevent Label Request –loops. Hop Count TLV used to find looping LSPs	May be done using Record_Route -object

Creating the Label Switched Path

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



- We still need routing protocols to find the paths
  - QoS routing in the future
- LSP is like ATM VC
  - ATM Forum signaling vs. LDP
- What initiates the LSP creation?
  - Traffic (Reactive), prediction of future traffic (Proactive, control)

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

Creating and using the label space

- Control of label distribution
  - Independent
    - Advertise the label assignments to neighbors
  - Ordered
    - Label assignment proceeds in an end-to-end fashion
      - Ingress or egress initiated
- Binding the label to a FEC
  - Local and remote
  - Remote options: **Downstream (always in MPLS)** or Upstream
    - Downstream on demand (request) and unsolicited downstream (distribute)
- Saving the label information
  - Liberal or conservative
- Save the label space!
  - Use label merging (and lose information on the packet arrival data)

Design of optimal MPLS networks

- The LSP design problem
  - Constrained non-linear optimization problem
    - Find such LSP configuration $\mathbf{x}_{opt}$ that maximizes the revenue earning rate F(x) subject to constraints such that each LSP has a strictly positive bandwidth, and that the bandwidths of the LSPs passing through link use the entire bandwidth of the link.
    - The necessary condition for the configuration to be locally optimal says that the change in revenue obtained by moving an inifitesimal amount of bandwidth to a route (of an aggregate) is equal to the revenue lost in acquiring this bandwidth from aggregates whose LSP sets include direct LSPs over the links of the route, and vice versa.

$$\max_x \sum_{a=1}^{A} F(a, x_a)$$

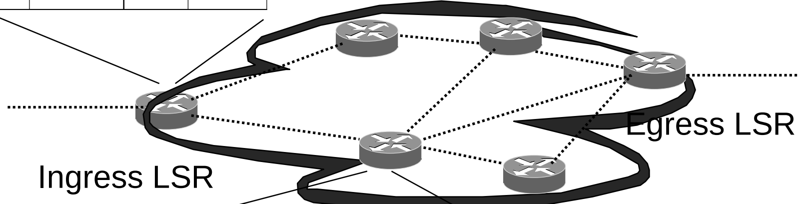
$$x_r > 0 \qquad \sum_{r \in A_l} x_r = B_l$$

$$\frac{\partial}{\partial x_r} F(a, x_r) = \sum_{l \in r} \frac{\partial}{\partial x_l} F(\cdot, x_l)$$

Using labels/tags in forwarding

- Different FECs (Forwarding Equivalence Classes) for different traffic

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

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.

MPLS objectives fulfilled

- Improve routing performance
  - Layer 3 performance results from pushing layer 3 processing to the edges
  - Separation of the route lookup and the packet forwarding processes
- Improved scalability
  - Aggregate flows
- Flexibility to bring new services to the network
  - use routing and LDP to map various FECs to alternative routes

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
  - Map trunks to LSP
  - Map LSP to physical network

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)

General problems with MPLS approach

- The original conclusions that lead to MPLS are no longer valid
  - Routers are too slow, routing tables are too big
    - How come, then, there are Gigabit-routers available off-the-shelf?
- Complex management of the MPLS network
  - Traffic or topology based path creation
    - or RSVP
- 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

More problems with MPLS

- 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.
- How much of the functionality existing in the lower layer(s) is taken into the concept
  - Signaling, QoS features, traffic management
  - What about different layer 2 technologies and their QoS support

Summary of the MPLS

- Partly a router workload reduction method
  - Acts as an enabler for Quality of Service networks
- Mostly a new way to use routing information in a flexible way
- Makes use of the connection oriented layer 2 technologies
  - ATM, IP over SDH, ISDN
- Standardization is well on the way
- Competing solutions on the edge of release

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...