UNIVERSITÄT
BERN

UNIVERSITÄT BERN

Advanced Networking and Future Internet

IV. Traffic Engineering

Prof. Dr. Torsten Braun, Institut für Informatik

Bern, 05.10.2020

Advanced Networking and Future Internet: Traffic Engineering

Table of Contents

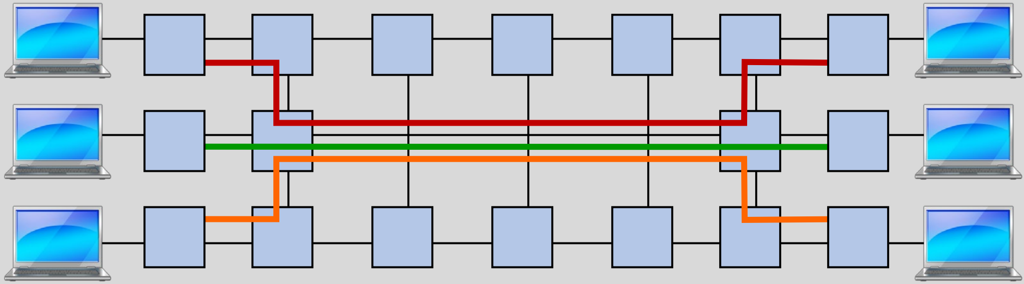
1. Traffic Engineering
    1. Motivation
    2. Optimization Objectives
    3. Steps
    4. Mechanisms
        1. Overlay Networks
        2. IP-Based Routing
        3. Constraint-based Routing
2. Multi-Protocol Label Switching
    1. ATM Virtual Circuit Switching
    2. IP Switching
    3. Multi-Protocol Label Switching
    4. Labels
    5. FEC and NHLFE
    6. Packet Processing
    7. Label Stack
    8. Label Distribution
    9. Label Switched Path Control
    10. Route Selection
    11. MPLS Applications
    12. Multiprotocol Lambda Switching
    13. Generalized MPLS
3. Overlay Networks
    1. Overcoming Routing Inefficiencies
    2. Transport Inefficiencies
    3. Example: Resilient Overlay Networks

1. Traffic Engineering

1. Motivation

Problem

– Shortest path routing leads to congestion at certain links while others remain unloaded.



Solutions

– Faster routers and links

– Traffic engineering =
process of controlling how traffic flows through a network in order to optimize resource utilization and network performance

1. Traffic Engineering

2. Optimization Objectives

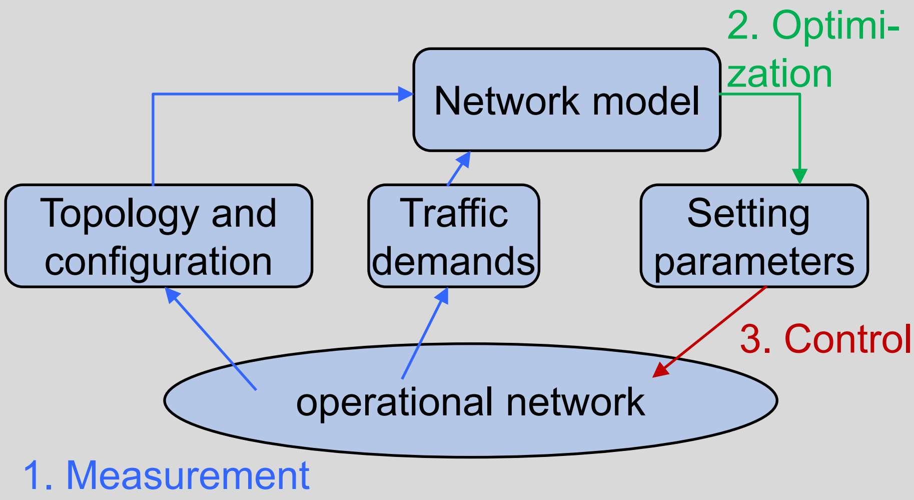
- Minimizing congestion and packet loss in the network
- Improving link utilization
- Minimizing total delay experienced by packets
- Increasing number of customers with current assets

UNIVERSITÄT BERN

1. Traffic Engineering

3. Steps

1. Acquisition of measurement data
2. Route optimization supported by modeling, analysis, simulation
   – Centralized
   – Distributed,
   e.g., by ingress routers
   → race conditions & oscillations
3. Assignment of traffic to routes



1. Traffic Engineering

4. Mechanisms

Mechanisms

- Overlay networks
- IP-based routing
  - e.g., based on interior / exterior gateway protocols for intra / inter-domain traffic engineering
- Constrained-based routing

Additional Issues

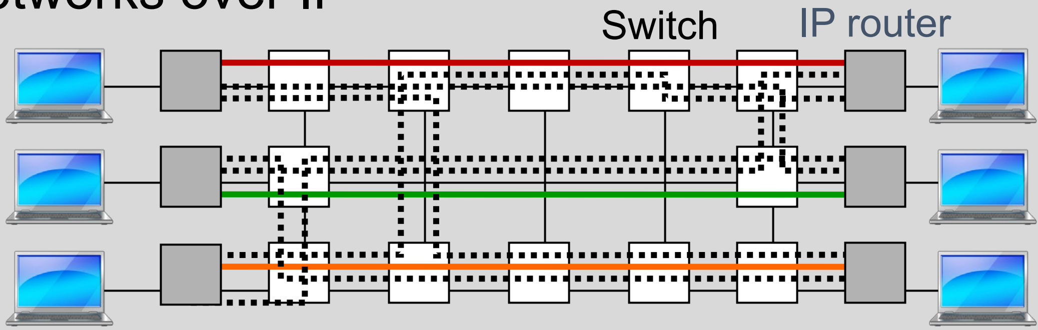
- Scope
  - Inter-domain
  - Intra-domain
- Timescale
  - Offline
  - Online

UNIVERSITÄT BERN

1. Traffic Engineering

4.1 Overlay Networks

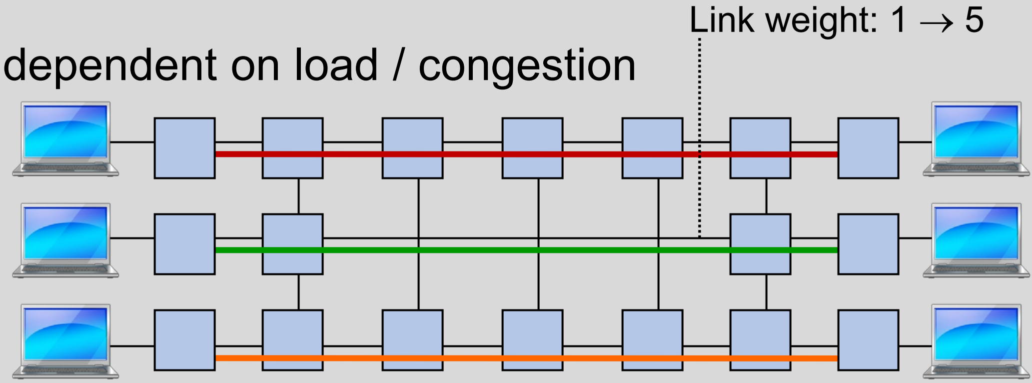
- IP over a connection-oriented technology, e.g., ATM, WDM, or overlay networks over IP
- Virtual topologies based on point-to-point links
- Drawbacks
  - Management of two networks
  - Complexity
  - Scalability (O(n^2) connections)



1. Traffic Engineering

4.2.1 Interior Gateway Protocol (IGP) Control

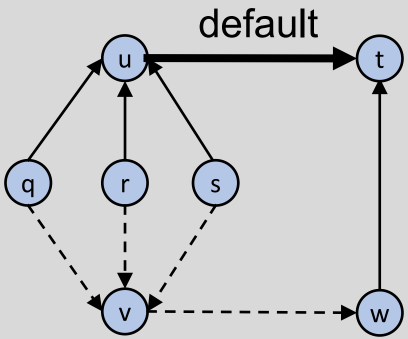
- Intra-domain traffic engineering
- OSPF link weights can be changed dependent on load / congestion
- Link state advertisement extensions describing maximum, maximum reservable and unreserved bandwidth
- QoS routing, e.g., using Shortest Widest Path algorithm
  - Bottleneck bandwidth = minimum of unused capacity over all links on the path
  - Selection of path with largest bottleneck bandwidth, hop number as 2nd criterion
- Risk of oscillations to be avoided by careful weight selection



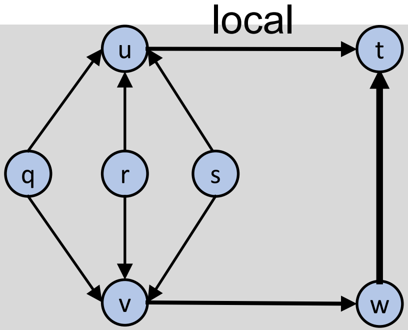
1. Traffic Engineering

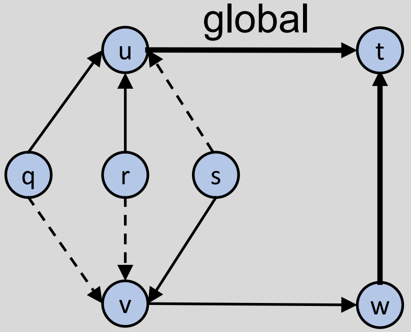
4.2.2 Local vs. Global Link Weight Changes

Example network with identical link capacity and identical load generated by nodes q, r, s, w



Link	Default unit weights		Increasing weight of overloaded link		Optimal global single change	
	Weight	Load	Weight	Load	Weight	Load
(q,u)	1	1	1	0.5	1	1
(r,u)	1	1	1	0.5	1	1
(s,u)	1	1	1	0.5	3	0
(u,t)	1	**3**	**2**	1.5	1	**2**
(q,v)	1	0	1	0.5	1	0
(r,v)	1	0	1	0.5	1	0
(s,v)	1	0	1	0.5	1	1
(v,w)	1	0	1	1.5	1	1
(w,t)	1	1	1	**2.5**	1	**2**





b UNIVERSITÄT BERN

1. Traffic Engineering

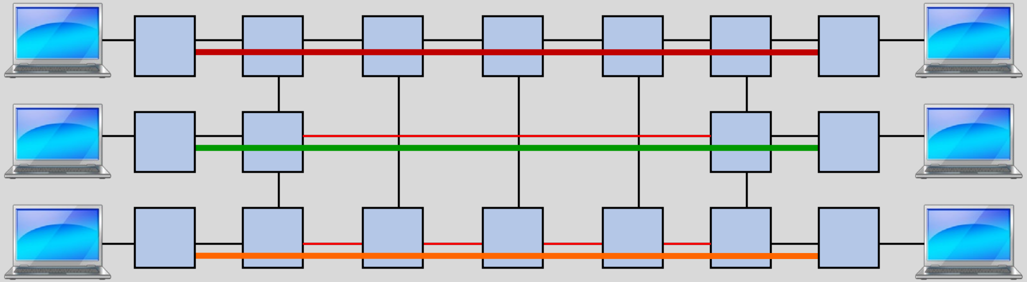
4.2.3 Inter-Domain Traffic Engineering with BGP

- BGP (Border Gateway Protocol) as de-facto exterior gateway protocol (EGP) in the Internet.
- Routers in different autonomous systems (ASs, domains) use BGP to exchange update messages about how to reach different destination prefixes.
- A router may receive routes for the same destination prefix from multiple neighbour ASs and
  - apply import policies to filter unwanted routes and to manipulate attributes of remaining routes.
  - invoke decision process to select exactly one best route for each destination prefix among all the routes it hears.
  - apply export policies to manipulate attributes and to decide whether to advertise the routes to neighbour ASs.
- Problems
  - BGP advertisements do not explicitly convey any information about resources available on a path.
  - BGP routing policies are complex and depend on many factors, e.g., commercial relationships.
  - Operators have only indirect influence on path selection.

1. Traffic Engineering

4.3.1 Constraint-based Routing with MPLS

- Building a network map with capacity information
  - Enhancement of routing protocols to advertise capacity information
  - Network management (measurement and monitoring)



- Constraint-based Routing (CR)
  - Prune links that do not satisfy constraints
  - Pick shortest path of remaining topology
- Set up constraint-based routed path between ingress and egress node using special signaling protocols such as CR-LDP

1. Traffic Engineering

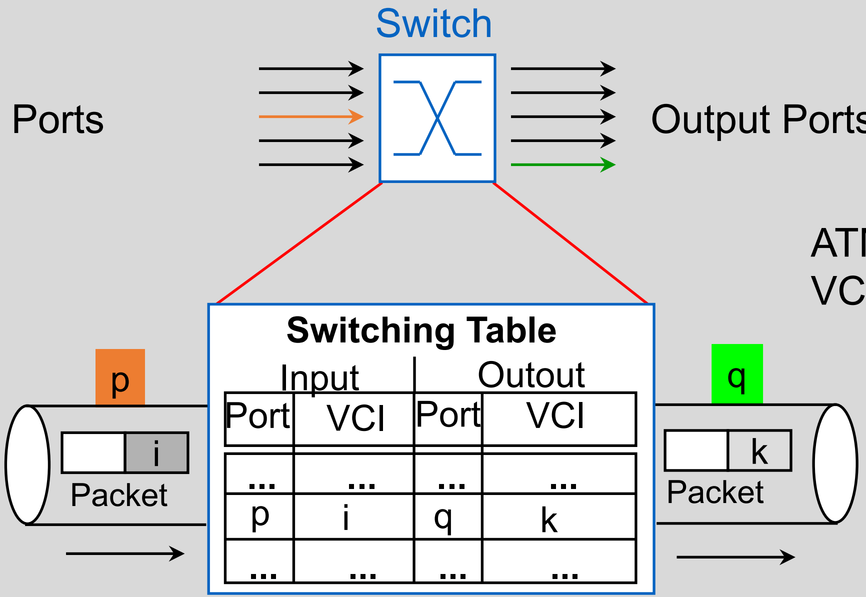
4.3.2 IP-Based Traffic Engineering vs. Constrained-Based Routing

- More fine-grained control and better flexibility by constrained-based routing

- Better scalability (no overhead for path setup) and robustness (automatic rerouting) by IP-based TE, e.g., IGP control u^b UNIVERSITÄT BERN

2. Multi-Protocol Label Switching

1. ATM Virtual Circuit Switching

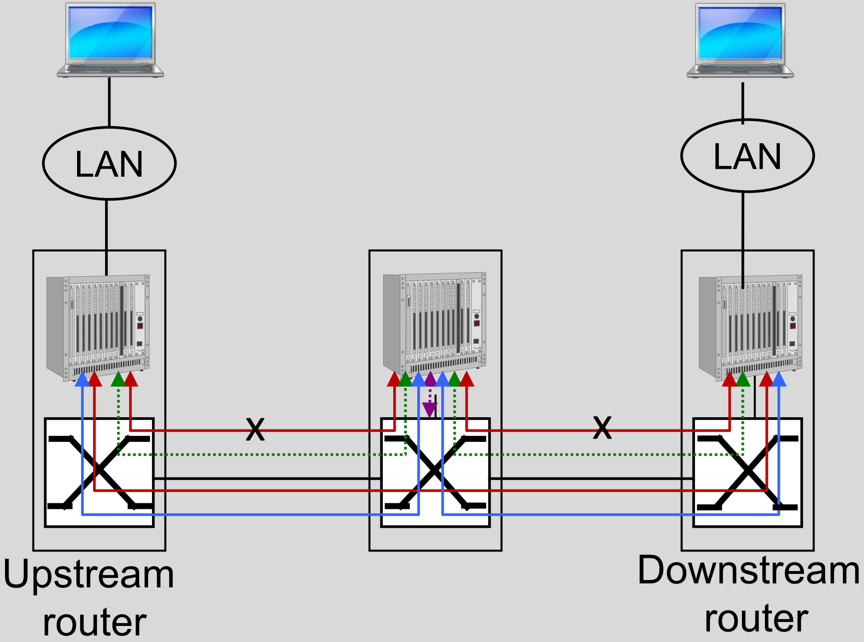


ATM: Asynchronous Transfer Mode
VCI: Virtual Circuit Identifier

2. Multi-Protocol Label Switching

2. IP Switching

- First: routing over default path
- Router identifies incoming flow, establishes a special ATM Virtual Circuit (VC) to upstream router, and signals this via IFMP (Ipsilon Flow Management Protocol, RFCs 1953/1954)
- Downstream router establishes also special ATM VC.
- Router splices both VCs using General Switch Management Protocol (GSMP, RFCs 1987/3292) → short-cut ATM-VC

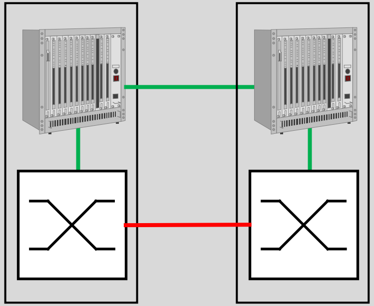


2. Multi-Protocol Label Switching

3. Multi-Protocol Label Switching

– Scalability problem of flow-based IP Switching due to fine granularity
→ Topology-based approaches establish short-cuts for aggregated flows, e.g., all flows between subnets.

– Harmonization of proprietary proposals:
Multi-Protocol Label Switching (MPLS)
  – Solution for any network technology (also: LANs)
  – Separation of IP router functionality into
    – Data packet forwarding (based on label swapping)
    – Control: routing protocols, signaling, management



2. Multi-Protocol Label Switching

4. Labels

Label

= short, fixed length,
locally significant,
IP-independent identifier

– Layer 2 information,
e.g. ATM VCI

– Shim header: header between
IP and layer 2 header

Label Swapping

– Table lookup to determine route and new label for outgoing packet (cf. ATM)

Label Switching Router

– forwards packets along unidirectional Label Switched Path.

UNIVERSITÄT BERN

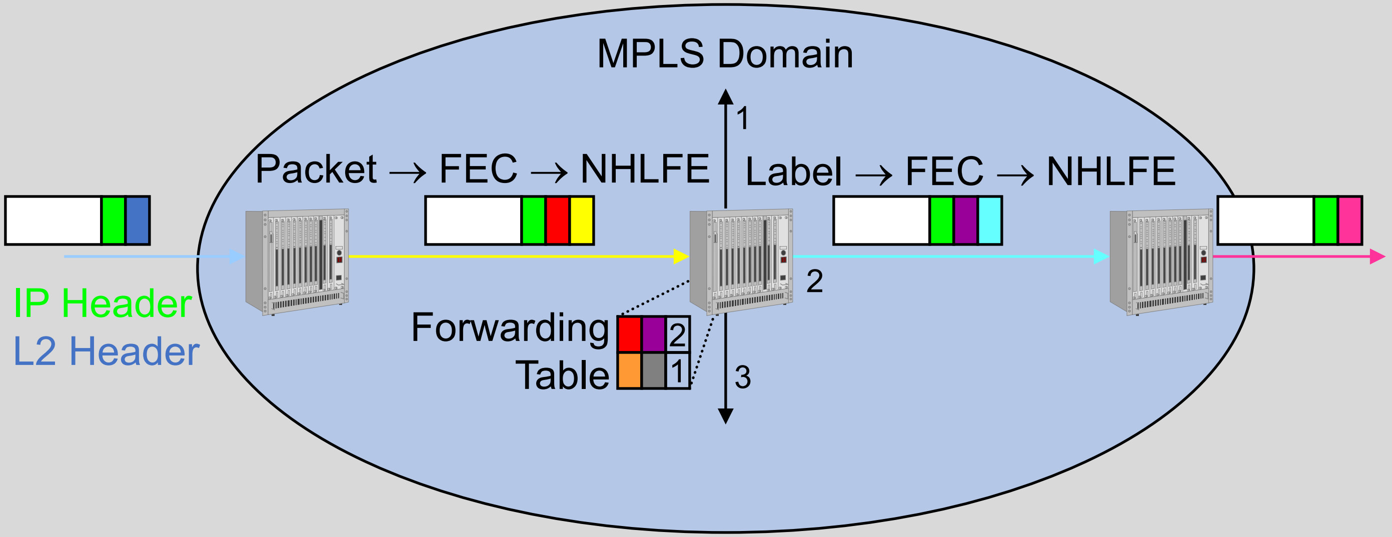
2. Multi-Protocol Label Switching

5. Forwarding Equivalence Classes and Next-Hop Label Forwarding Entries

- Forwarding Equivalence Class
  - a group of IP packets that are forwarded in the same manner over a path
  - Coarse-grained FEC, e.g. packets with the same destination address prefix
  - Fine-grained FEC, e.g. packets of the same application
- Next-Hop Label Forwarding Entry
  - Next hop
  - Label stack operation (push, pop, swap) and outgoing label
  - Optional information
    - Layer 2 Encapsulation
    - Encoding information for transmission
    - Further packet processing options, e.g., queue management etc.
- Mapping: FEC → NHLFE
  - Multiple NHLFEs per FEC possible
    - Load balancing: alternating usage of LSPs
    - Redundant NHLFEs: fast rerouting

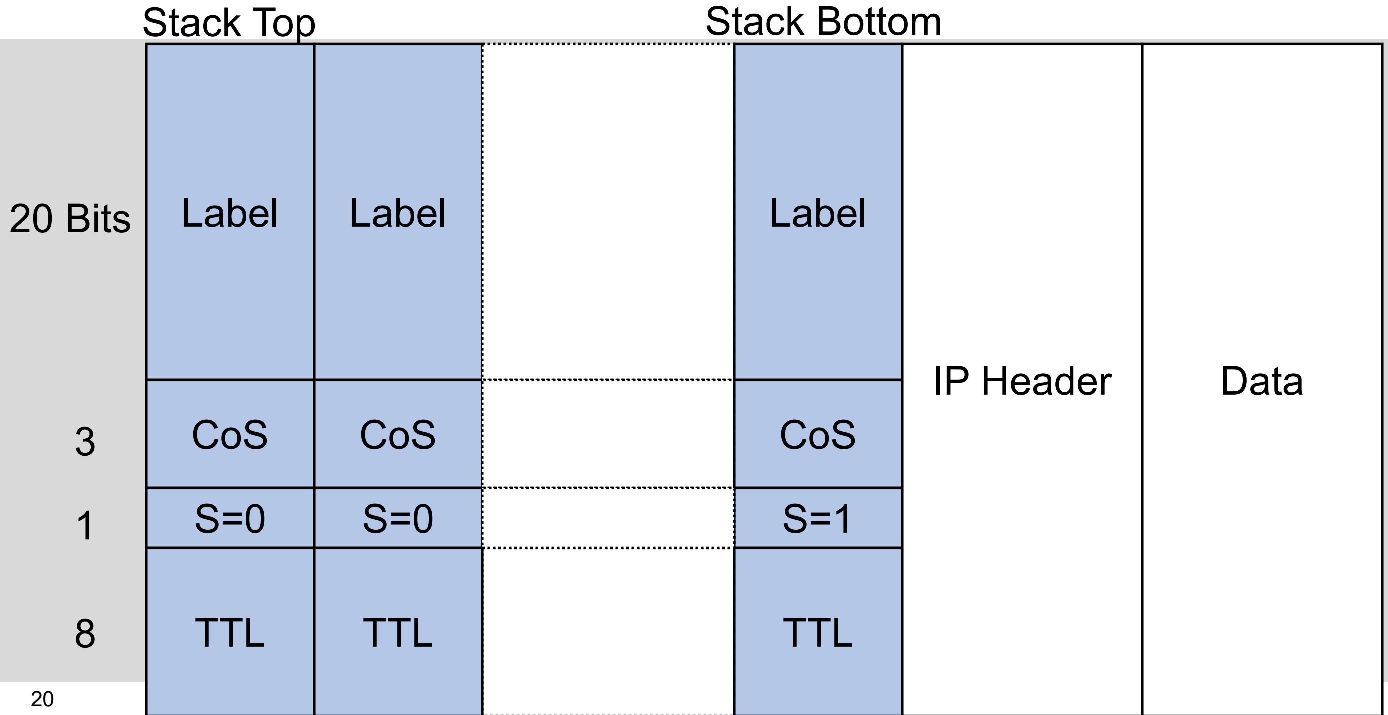
2. Multi-Protocol Label Switching

6. Packet Processing



2. Multi-Protocol Label Switching

7. Label Stack



CoS: Class of Service

u^b UNIVERSITÄT BERN

2. Multi-Protocol Label Switching

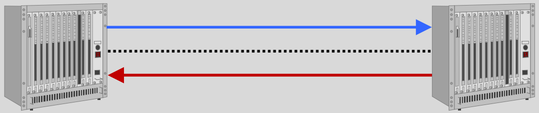
8. Label Distribution

Options

- Downstream Unsolicited: Downstream LSR distributes label bindings for FECs without solicitation to upstream LSR.
- Downstream On-Demand: Upstream LSR requests downstream LSR to distribute a label.

Protocols

- Piggybacked label distribution on existing protocols such as BGP
- New protocols such as Label Distribution Protocol



2. Multi-Protocol Label Switching

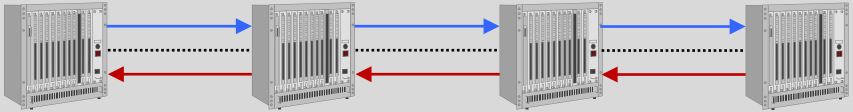
9. Label Switched Path Control

Independent Control

- Each LSR makes an independent decision to bind a label to a FEC and to distribute that binding to its peers.

Ordered Control

- An LSR only binds a label to a particular FEC, if it is the egress LSR for that FEC, or if it has already received a label binding for that FEC from its next hop.
- to ensure that traffic in a particular FEC follows a path with some specified properties, e.g.,
  - Traffic does not traverse any node twice.
  - Traffic follows an explicitly specified path.
  - Specified amount of resources are available to traffic.



UNIVERSITÄT BERN

2. Multi-Protocol Label Switching

10. Route Selection

Hop-by-hop Routing

- allows each node to independently choose the next hop for each FEC.
- Each LSR determines the next interface of an LSP based on the local IP forwarding table.

Explicit Routing

- A single LSR (e.g., LSP ingress or egress) specifies several or all LSRs in the LSP.
  - Strict explicit routing: LSR specifies entire LSP.
  - Loose explicit routing: LSR specifies some of LSP.
- Sequence of LSRs may be chosen by configuration or selected dynamically.
  - Example: Ingress/egress node may use topological information learned from a link state database in order to compute the entire path
- useful for policy routing or traffic engineering

UNIVERSITÄT BERN

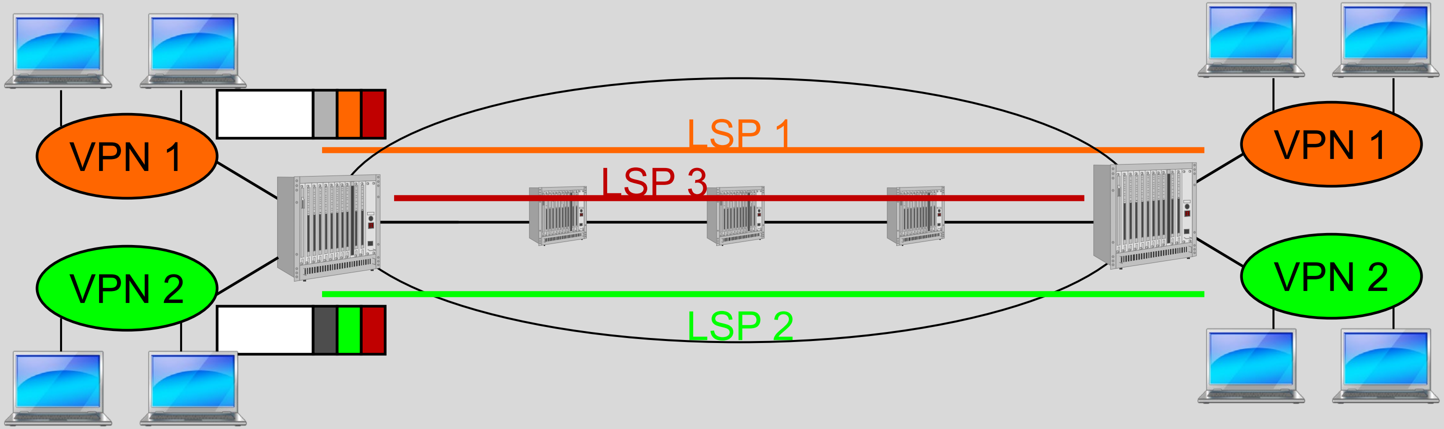
2. Multi-Protocol Label Switching

11. MPLS Applications

- Traffic Engineering
  - Problem: Shortest path routing may cause overload on certain links while others remain unloaded.
  - Establishment of LSPs for certain (aggregated) flows
- Virtual Private Networks (VPNs)
  - Forwarding between subnets based on MPLS labels
  - Replacement of IP-in-IP tunnels
- Load Balancing
  - Establishment of several LSPs for a single FEC
  - Switching between LSPs
- Quality-of-Service support
  - Resource reservation for certain LSPs
- Redirection in case of link failures
  - Establishment of several LSPs for a single FEC and switching between NHLFEs in case of detected failures
  - Establishment of bypass LSPs and label stacking
- Pseudo Wire Emulation Edge-to-Edge (PWE3, RFC 3985)
  - Emulation of services such as ATM, Ethernet, SONET/SDH over packet switched networks

2. Multi-Protocol Label Switching

11.1 MPLS VPNs



2. Multi-Protocol Label Switching

11.2.1 Rerouting

Optimized Rerouting

- Re-optimize traffic flows to a modified topology
- LSP head computes optimized LSP.
- Reasons for rerouting
  - Notification of LSP head about failure
  - Traffic monitoring by LSP head

Fast Rerouting

- Minimize service disruptions
- Techniques: splicing and stacking

2. Multi-Protocol Label Switching

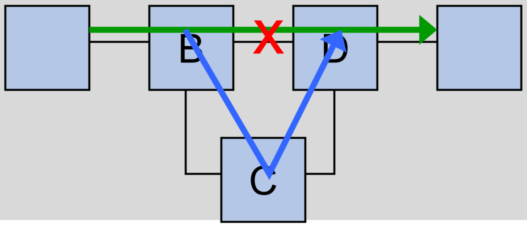
11.2.2 Fast Rerouting

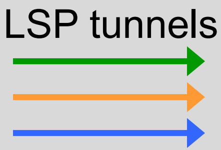
Stacking

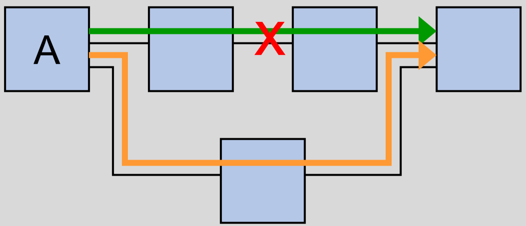
- LSP bypassing the failed link is created.
- B pushes label onto the stack.
- C pops label.
- D receives packet with expected label.

Splicing

- Pre-establishment of a bypass LSP to bypass a path
- A selects bypass LSP after failure detection.







12. Multiprotocol Lambda Switching

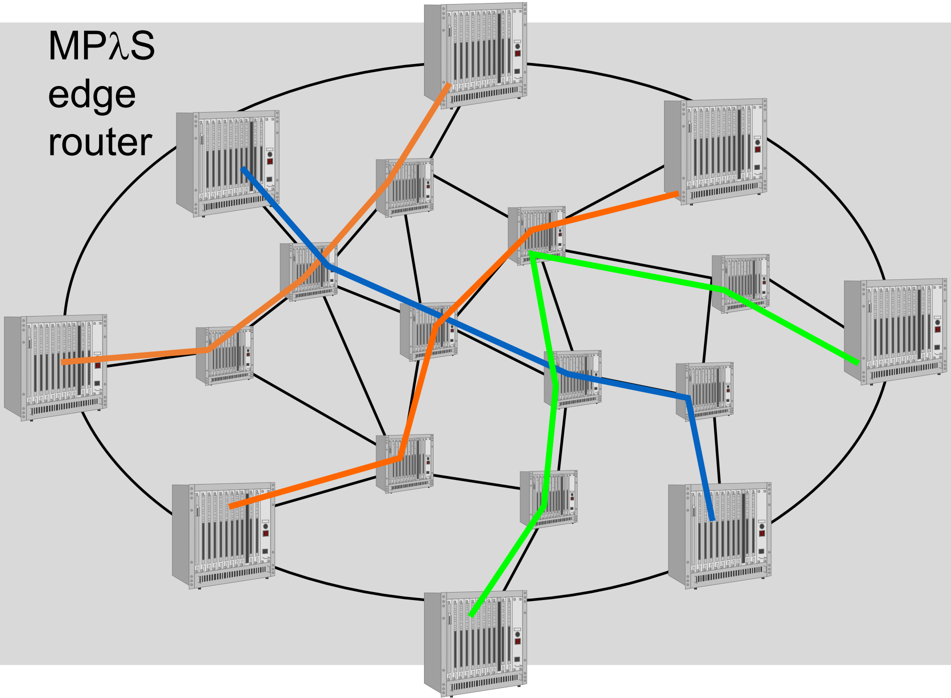
- Application of MPLS to optical networks → MPλS
- Wavelength conversion → Label = wavelength (λ)
- Optical switches need control plane for configuration and management. Usually: (proprietary) network management protocols

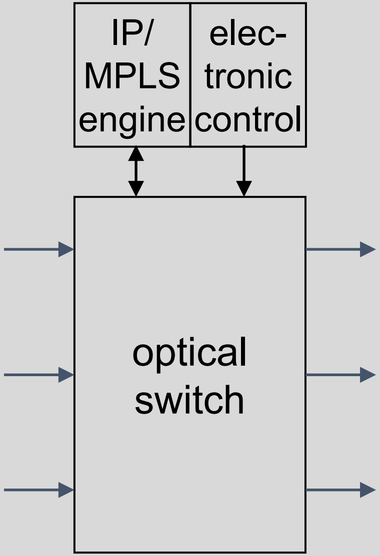
MPλS approach is similar to MPLS over ATM

- Extension of optical switches by MPLS engine
- IP/MPLS protocols as uniform control plane for optical equipment (together with network management protocols) with several control functions
  - Resource discovery
  - State information dissemination
  - Path selection
  - Path management
- Interconnection of MPλS router via dedicated λ's, e.g., for default routing and signaling

2. Multi-Protocol Label Switching

12.1 Multiprotocol Lambda Switching





2. Multi-Protocol Label Switching

12.2 MPλS Problems

- Number of λ's is relatively small compared to label space.
  → discrete set of bandwidth values
- Capacity of a λ is very large compared to a usual LSP.
- No push and pop operations
- No label merging
- Transparent optical switching
  → Optical switches are not able to recognize or modify packet headers.

2. Multi-Protocol Label Switching

13. Generalized MPLS

IGP / link state routing protocol, e.g., OSPF, extensions

– to advertise availability of optical resources

Generalized signaling, e.g., CR-LDP

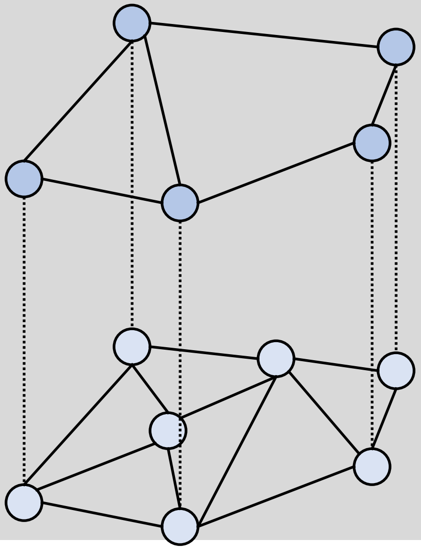
– support of TDM, λ, port switching
– suggesting and restricting of labels by upstream node
– bi-directional LSPs

Link Management Protocol

– Establishment and maintenance of control channels
– Link connectivity verification based on test message exchange
– Synchronization and exchange of link property summaries
– Fault detection and localization

3. Overlay Networks

- Overlay network = logical network on top of a physical network
- Examples:
  - IP network on top of a physical network
  - Overlay networks on top of IP networks
    - Virtual Private Networks
    - Application level forwarding

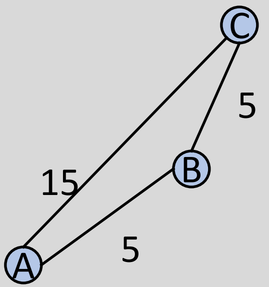


3. Overlay Networks

3.1 Overcoming Routing Inefficiencies by Overlay Networks

Overlay networks can be used to overcome routing inefficiencies in IP networks.

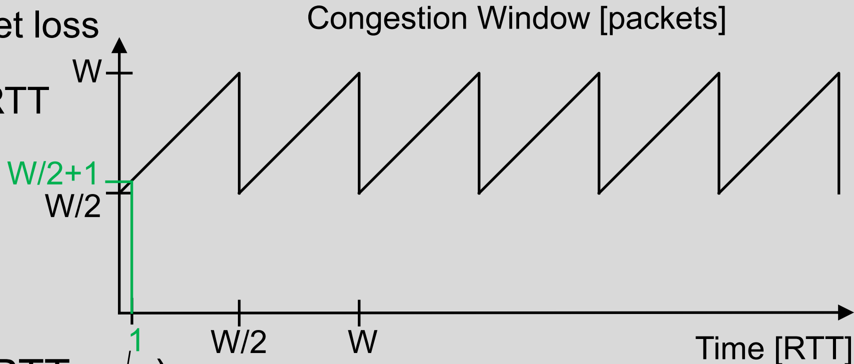
- Poor routing metrics
  - Routers typically exchange connectivity information but not performance information.
  - Routing decision by minimization of nodes / ASs along path to destination
  - Triangle inequality d(A,B) + d(B,C) > d(A,C) does not always hold in the Internet.
- Restrictive routing policies
  - Policy routing allows each AS to define its own rules
    (e.g., early exit, private peerings) for what traffic to carry and where to send.
- No automatic load balancing
  - Links may be underutilized.
- Single-path routing
  - Performance gains and robustness by multiple paths



3. Overlay Networks

3.2 Transport Inefficiencies

- TCP performance depends on round-trip time and packet error rate: BW < (MSS / RTT) · (1 / √p)
  - BW: bandwidth, MSS: maximum segment size, RTT: round trip time, p: packet error rate
  - Assumption: Delivery of 1/p packets followed by 1 packet loss
  - If receiver acknowledges each packet: Window opens by 1 per round trip, each cycle = W/2 · RTT
  - Data delivered in each cycle = $(W/2)^2 + \frac{1}{2}(W/2)^2 = \frac{3}{8}W^2 = 1/p$
  - BW = (data per cycle) / (time per cycle) = (MSS · ⅜ W^2) / (RTT · W/2) = (MSS/p) / (RTT √(2/(3p))) = (MSS · C) / (RTT · √p), C = √1.5 = 1.22

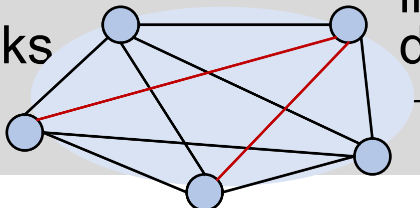


- Overlay links can be selected dependent on weight (RTT · √p)
- W: maximum window, W/2: minimum window in equilibrium

3. Overlay Networks

3.3.1 Example: Resilient Overlay Networks

- BGP takes several minutes to react on link failures.
- RON architecture allows distributed Internet applications to quickly detect and recover from path outages and performance degradation, e.g., caused by link breaks, overload, or denial-of-service attacks



- RON nodes monitor quality of Internet paths (RTT, packet error rate, throughput) and decide whether to route packets directly over the Internet or via other RON nodes. Typically, 1 RON node is sufficient.
- RON nodes exchange path quality information and establish link state database
  → limited scalability (~50 nodes)

UNIVERSITÄT BERN

3. Overlay Networks

3.3.2 RON: Routing and Path Selection

- Each RON node exchanges link information with N-1 RON nodes.
- Each RON node implements outage detection by active probing to determine whether another node is still working.
- Latency estimate on link l: $lat_l = \alpha \cdot lat_l + (1- \alpha)\ new\_sample_l$
- Path latency $lat_{path} = \sum_{l \in path} lat_l$
- $loss\_rate_{path} = \pi_{l \in path} (1- loss\_rate_l)$
- Throughput optimization: score = $C / (RTT \cdot \sqrt{p})$

Thanks

for Your Attention

UNIVERSITÄT BERN

Prof. Dr. Torsten Braun, Institut für Informatik

Bern, 05.10.2020