MPLS Traffic Engineering

When you really need more than the best path…

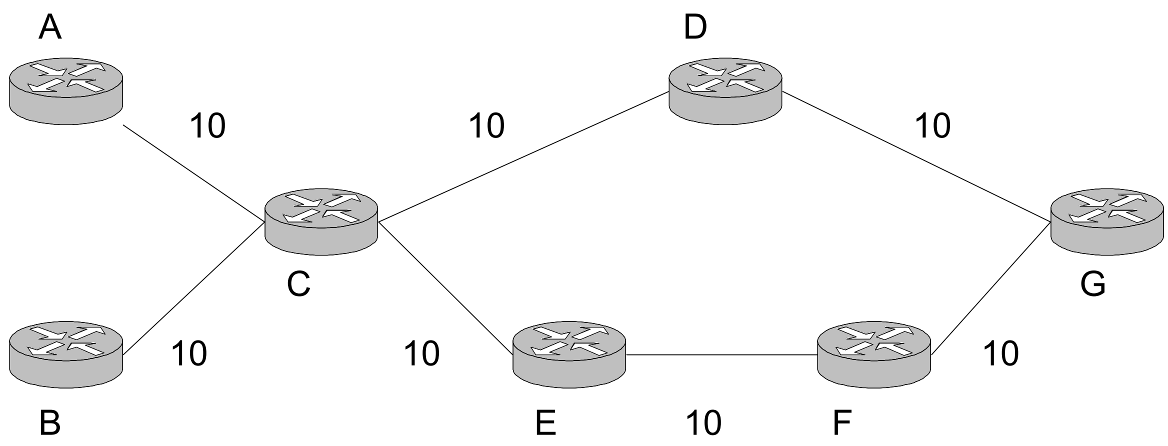
Agenda

- Introductions, disclaimers, etc.
- Traffic Engineering before MPLS
- Basics of TE tunnels
- Information distribution
- Path calculation and setup
- Forwarding traffic down tunnels

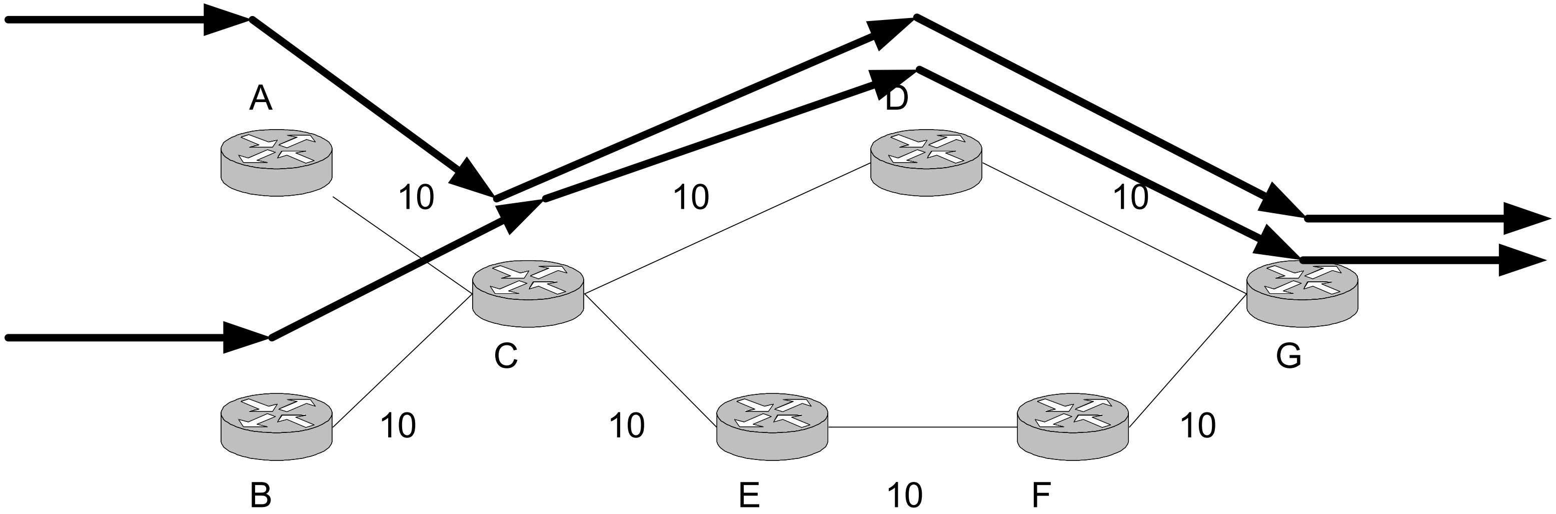
Introductions, disclaimers, etc.

- I’ve never been exposed to JunOS, so I don’t have any JunOS configuration examples.
  - Cisco is not necessarily the greatest, it just happens to be what I use.
- Ask questions!

Traffic Engineering before MPLS



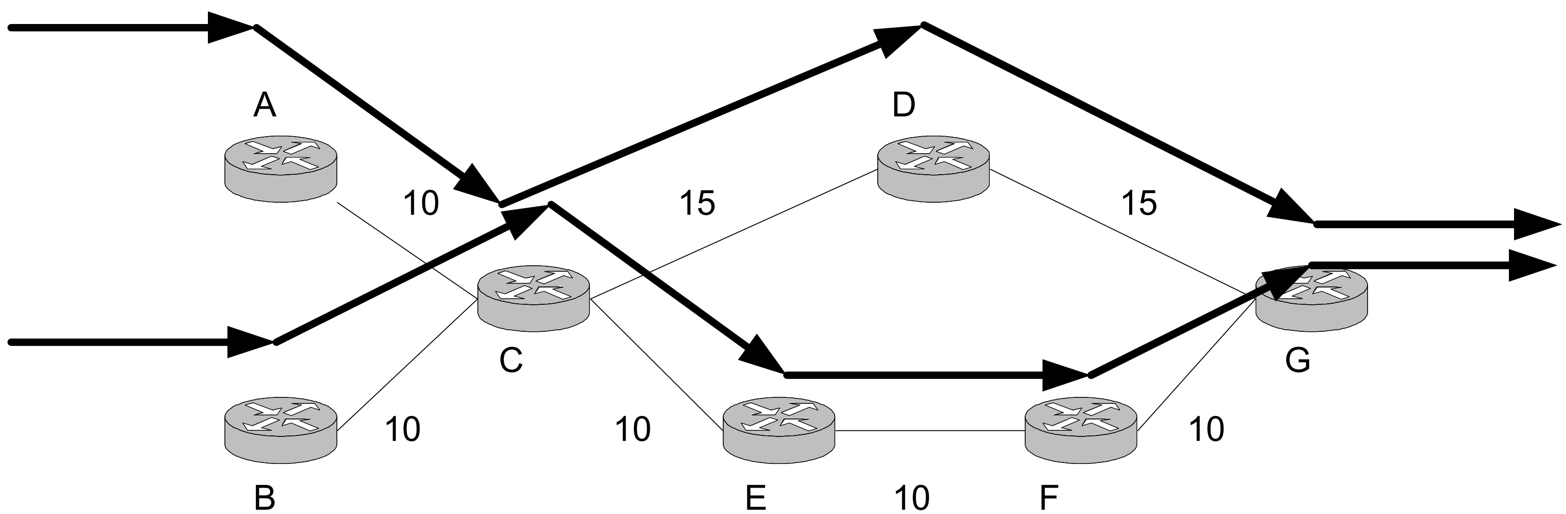
Traffic Engineering before MPLS



TE before MPLS

- IGP Metric adjustment
  - IGPs could be tweaked, but “getting it right” usually involved a network-wide plan.
  - Routing loops are possible if link metrics don’t match in each direction.

Traffic Engineering before MPLS



MPLS TE Fundamentals

- Packets are forwarded based on FIB or LFIB.
- FIB/LFIBs are built based on RIB.
- RIB contains next-hop data for prefixes.
- To route traffic over a deterministic path, we need an exit interface.
- Ergo: the TE tunnel.

TE Tunnel basics

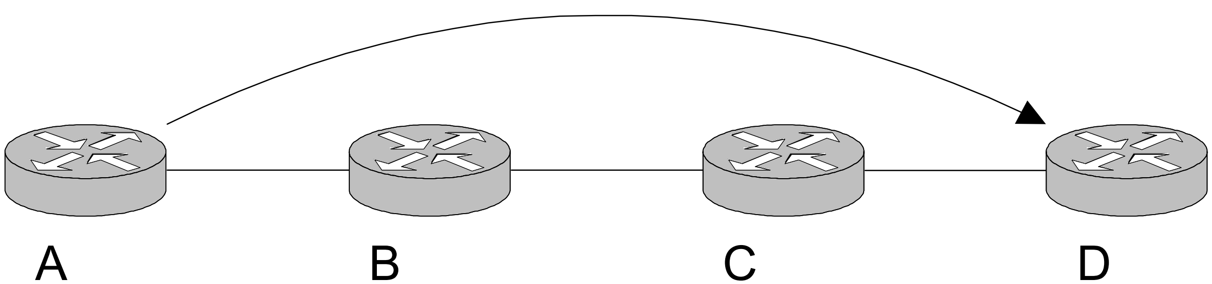
- TE tunnel interfaces are **unidirectional** logical links from one router to another router.
  - Once the tunnel is properly configured, a label is assigned for the tunnel that corresponds to the path through the MPLS network (LSP).
  - The path through the network will be deterministic, and not necessarily the IGP shortest path.

TE Tunnel basics

- Once traffic is routed onto the tunnel, the traffic flows through the tunnel based on the path.
- Return traffic could be placed onto a tunnel going the opposite direction, or simply routed by IGP.

Key TE terms

- Headend
  - Router on which the tunnel is configured.
- Tail
  - Destination address of tunnel
- Midpoint
  - Router(s) along the tunnel LSP.



Basic TE configuration

- Global:
  - “mpls traffic-eng tunnels”
- IGP: (must be OSPF or IS-IS)
  - “mpls traffic-eng router-id Loopback0”
  - “mpls traffic-eng [area X|level-2]”
- Physical Interface(s):
  - “mpls ip”
  - “mpls traffic-eng tunnels”
    - Tells IGP to share TE info with other TE nodes.

Basic TE tunnel configuration

- "interface TunnelX"
  - "ip unnumbered Loopback0"
    - Borrow the loopback's address, so we can forward traffic down this interface.
  - "tunnel mode mpls traffic-eng"
  - "tunnel destination <a.b.c.d>"
    - "Tunnel Tail"
  - "tunnel mpls traffic-eng path-option 10 dynamic"
    - Find a dynamic path through the network
    - We'll discuss path selection in a bit

Where are we at?

- Tunnels go from headend to tailend over a deterministic path.
- We know what commands go on a router:
  - Global commands
  - Physical interface commands
  - Tunnel interface (basic) commands
- Next: TE and bandwidth

TE and bandwidth

- Physical interfaces can be told how much bandwidth can be reserved (used).
  - “ip rsvp bandwidth X X”
- TE Tunnels can be configured with how much bandwidth they need.
  - “tun mpls traff bandw Y”
- Tunnels will reserve Y bandwidth on outbound interfaces, and find a path across the network over interfaces with X(unused)>Y bandwidth.

TE and bandwidth

- Operators can adjust the tunnel bandwidth values over time to match changes in traffic.
- If tunnels are dynamically placed, the tunnels will dynamically find a path through the network with sufficient bandwidth, or go down.

TE auto-bandwidth magic

- Tunnels can be configured to watch their **actual** traffic (as in “sh int <blah> | i rate ” every five minutes) and update their reservation to match, at periodic intervals.
  - Dynamic reservations to match the live network!
  - Bandwidth is “reserved” using RSVP.
    - But not “saved” for TE…

TE auto-bandwidth magic

- “tunnel mpls traffic-eng auto-bw frequency Y”
  - Each auto-bw tunnel does “sh int” to capture its rate every 300* seconds.
  - Each auto-bw tunnel updates “tunn mpls traff bandwidth X” every Y seconds.
  - The configuration ACTUALLY changes. This will impact your RANCID tracking, etc.

Where are we at?

- TE Tunnel basics
- Router config basics
- General concepts about TE and bandwidth
- Next: how do we find a path through the network?

Path Calculation and Setup

- Constrained SPF
  - Find the shortest path through the network that meets certain constraints.
  - In this case, the shortest path that has X bandwidth available for reservation.
    - Actually, bandwidth X at or below priority Y, but we'll get there.

SPF Calculations

- Step 0: Create a PATH list and a TENT list.
- Step 1: Put "self" on the PATH list with a distance of 0 and a next hop of self. Set the bandwidth to N/A.
- Step 2: For the node just placed on the PATH list, call it the PATH node.

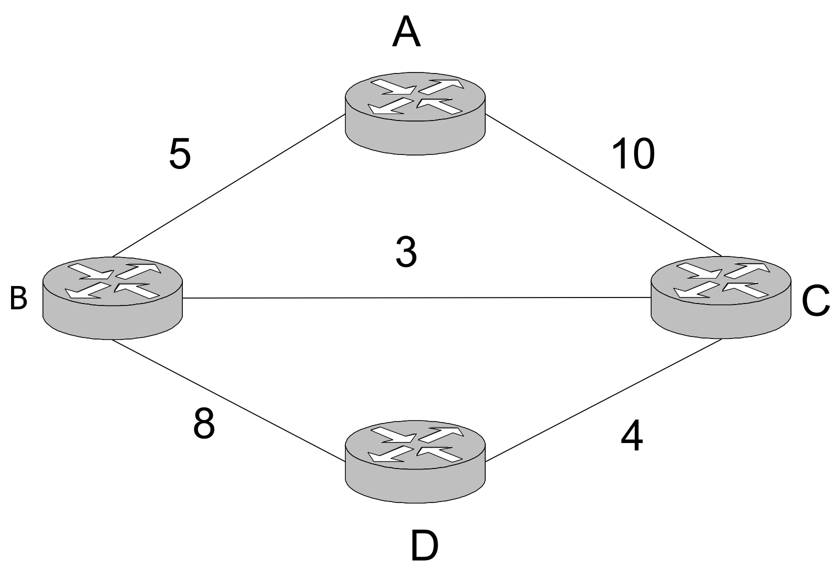
SPF

- Step 3: Put all of PATH node's neighbors on the TENT list:
  - Unless a neighbor is already on the TENT or PATH list with a lower cost.
- Step 4: If the TENT list is empty, stop.

SPF

- Step 5: Find the neighbor in the TENT list with the lowest cost.
  - Add that neighbor to the PATH list.
  - Return to Step 2.

Example Exercise

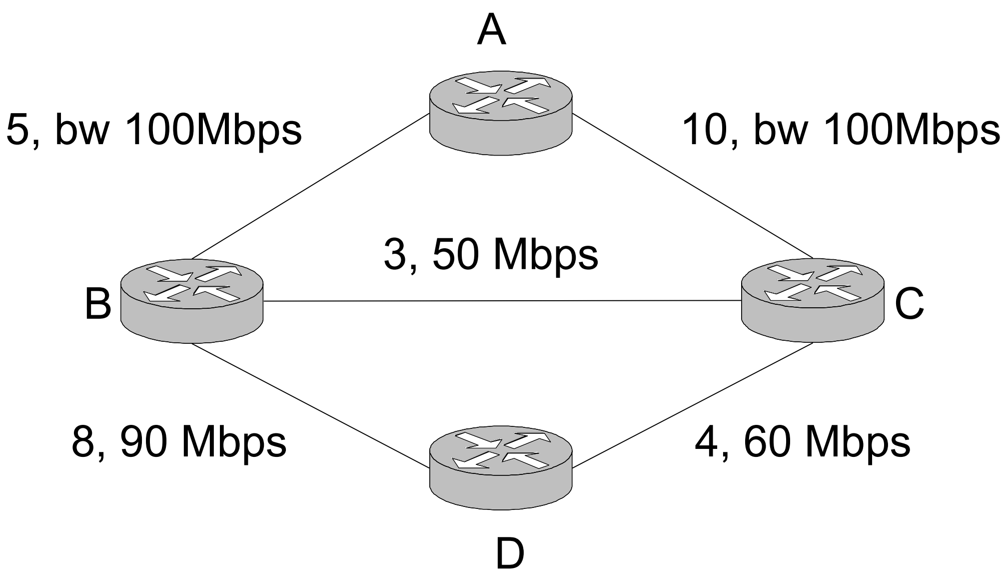


- We'll calculate router A's best path to router D.
- Use the handout supplied at the door, and we'll "think like a router" for a bit.

Constrained SPF

- Repeat the process, except modify step 3 such that you only put PATH node’s neighbors on the TENT list if the bandwidth available meets the requirements.

CSPF Exercise



- We'll calculate router A's best path to router D.
- Use the handout supplied at the door, and we'll "think like a router" for a bit.

CSPF notes

- No load sharing is performed within **a tunnel** – as soon as a path is found, it wins.
- CSPF tiebreakers:
  - Lowest IGP cost
  - Largest minimum available bandwidth.
  - Lowest hop count
  - Top node on the PATH list.

Where are we at?

- TE and config basics
- TE and bandwidth concepts
- Calculating a path through the network
- Next: configuring path selection

Creating paths

- Paths can be created dynamically, by allowing the TE-enabled IGP to perform CSPF.
    - “tunnel mpls traff path-option X dynamic”

Explicit paths

- Paths can be created manually, by explicitly creating a path.
  - "ip explicit-path name <name>"
    - "next-address a.b.c.d"
    - "next-address a.b.c.d"
  - "tunnel mpls traff path-option X explicit name <name>"
  - Manual paths can specify router loopbacks or physical interfaces.
    - Paths can be written to go from router to router, and/or from router to router over a specific interface.

Explicit paths

- Paths can be created manually, by explicitly configuring a path that **excludes** an address.
  - “ip explicit-path name <name>”
    - “exclude-address a.b.c.d”
      - Can be an interface address (to avoid a link) or a loopback address (to avoid a node).
    - Cannot combine exclude-address and next-address on the same explicit path!

Assembling paths

- A TE tunnel can have multiple path options.
  - Lowest-cost path option is attempted.
    - Higher-cost paths attempted sequentially.
    - Until a path can be successfully established.
  - Usually best to have a “dynamic” option as the highest-cost option (“find a way through the network with bandwidth X, wherever it may be”).

Using paths

- Upon setup or reoptimization, CSPF checks the tunnel’s path options in sequence for a path that has sufficient bandwidth.
  - A static path option will be checked hop-by-hop for available bandwidth.
  - A dynamic path option will do a complete CSPF calculation to find the best available path.
  - Lack of any available configured paths will result in the tunnel going (line protocol) “down”.

Where are we at?

- TE and config basics
- TE and bandwidth concepts
- Calculating and configuring paths through the network
- Next: distributing TE information around the network

Information distribution

- Once TE is enabled on routers and interfaces, new information about the network is shared among the routers.
  - Available bandwidth per interface
    - Eight priority levels.
    - High priority tunnels can push low priority tunnels out of the way.
    - Some dynamics as far as tunnel vs. interface sizing.
  - Administrative weight (TE-specific "IGP" metric)
  - Affinity (customizable)

Information distribution

- This information is distributed:
  - Immediately, for "significant" changes
  - Periodically, for "insignificant" changes
  - Immediately, if a change causes an error
- Significant changes occur when available bandwidth on an interface passes preset thresholds.
  - Customizable with 16 thresholds

Information distribution

- Insignificant changes are flooded every 3 minutes (by default).
- If a path setup fails because of insufficient bandwidth, the information is flooded immediately.

Tunnel reoptimization

- What if a new path (dynamic or static) appears that's "better" than the current one?
- Reoptimize!
  - Periodically, every hour by default.
  - Manually ("mpls traffic-eng reoptimize").
  - Event-driven
    - When a link comes up.
    - Optional; requires "mpls traff reo events link-up"

Where are we at?

- TE, config, bandwidth concepts
- Calculating and configuring paths through the network
- Distributing TE information around the network
- Next: putting IP traffic over the tunnels

Routing traffic over tunnels

- Static routes
  - “ip route x.x.x.x x.x.x.x tuX”
- Policy-based routing
  - “route-map PBR permit 20”
    - “match ip address <ACL>”
    - “set ip next-hop tuX”
- Autoroute

Autoroute

- "Treat this tunnel as though it's a directly connected link to the tunnel tail"
  - "Send any packets down the tunnel that are destined for either the tunnel tail or anything behind that tunnel tail."
  - Updates the RIB/FIB with "tunnelX" in place of the IGP next-hop; preserves the IGP cost to the tunnel tail.

Autoroute config basics

- “tunn mpls traff autoroute announce”
- Autoroute and load-sharing:
  - Parallel tunnels will load-share inversely proportional to their configured bandwidth.
    - Auto-bandwidth can really muck with these values!
  - Load-sharing can be tuned separately with “tunn mpls traff load-share X”.

Load sharing

- IGPs can load-share over equal-cost paths.
- A TE tunnel cannot load-share over multiple physical interfaces.
  - But multiple tunnels head->tail can be built, and traffic can be load-shared over the tunnels.

Where are we at?

- TE, config, bandwidth concepts
- Calculating and configuring paths through the network
- Distributing TE information around the network
- Putting IP traffic over the tunnels
- Next: TE diagnostics

TE diagnostics

- “sh ip route x.x.x.x”
  - If routed over a tunnel, “sh run int tuX”.
- “sh ip rsvp reservation”
  - Shows all RSVP reservations and their egress interface.
- “sh mpls traff tun suboptimal constr none”
  - Shows headend tunnels taking suboptimal paths to the tunnel tail.

TE diagnostics

- “sh mpls traff tun”
  - Detailed info for all tunnel headends.
    - Bandwidth information (auto-bw)
    - MPLS labels, hop-by-hop path
  - Moderate info for all tunnel midpoints, tails.
  - “sh mpls traff tun role head|middle|remote|tail”
    - Restricts to only that type of tunnels.

TE diagnostics

- "sh ip rsvp interfaces"
  - Shows allocated and maximum RSVP bandwidth on each active interface.

TE Caveats

- MPLS VPNs:
  - If a tunnel tail is not the egress PE, add “mpls ip” to the tunnel configuration.
    - (PE—P—P—PE—PE)
  - Add “mpls ldp discovery directed-hello accept” to config.

TE Caveats

- Multicast:
  - RPF calculations are normally based on unicast RIB.
  - Unidirectional TE tunnels cause RPF failures.
  - Add "mpls traffic-engineering multicast-intact" to IGP configuration.
    - Bases RPF checks on RIB before TE tunnels are substituted.

Questions?

templin@templin.org