Internet Technology

The “inner network“ view, part 2 (D): MPLS

Michael Welzl http://www.welzl.at

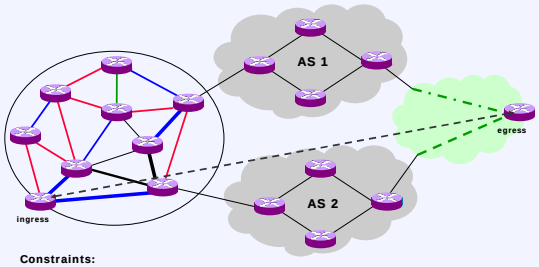
DPS NSG Team http://dps.uibk.ac.at/nsg
Institute of Computer Science
University of Innsbruck, Austria

Interdomain Traffic Engineering

Introduction

- Remember: path computation requires information about links
  - Done by adding TE-specific extensions to IGP (OSPF and IS-IS)
  - Hence limited to AS (IGP scope)
- Interdomain LSPs necessary for
  - TE in the large
  - large-scale deployment of services, e.g. connecting voice gateways, pseudowires, BGP/MPLS Layer 3 VPNs
- Inter-AS LSPs can exist across different ISP but also within one ISP
  - E.g. when ISP 1 just bought ISP 2 in a different geographical location
- Limitation is not signaling path setup (RSVP TE can do that), but computing the path
  - Has to be done offline: online calculation requires IGP information
    - Multiprovider environment: calculation based on information about all the links, but ISPs may not want to share such details
    - Also no FRR

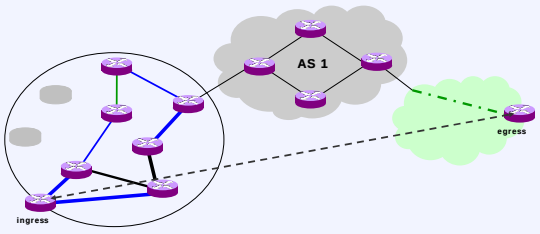
Interdomain Constraint Based Routing



Constraints:
- Do not cross AS 2
- Do not use red links
- Use TE links with unreserved bandwidth > 2 Mbps

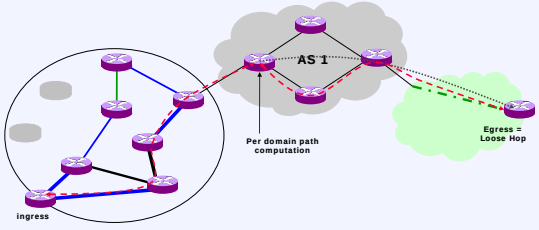
Source: EuroNGI Summer School'05 (D.Kofman & D.Papadimitriou)

Interdomain Constraint Based Routing /2



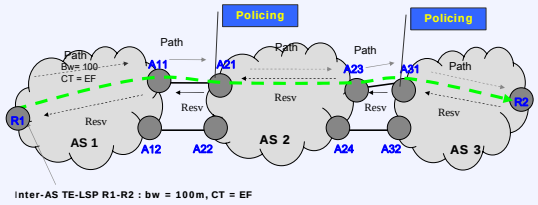
Constraints:
- Do not cross AS 2
- Do not use red links
- Use TE links with unreserved bandwidth > 2 Mbps

Interdomain Constraint Based Routing /3



Constraints:
- Shortest path
- Loose routing

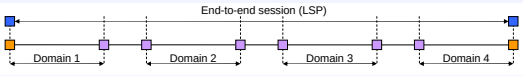
Interdomain LSP signaling



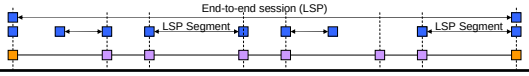
- Inter-AS TE-LSP signaling based on RSVP-TE
  - Explicit routing, local admission control, resources reservation & label distribution
- Path/REsv include QoS parameters
- Three signaling modes: Contiguous LSPs, LSP stiching, LSP nesting
- Need for proper policing and filtering of RSVP-TE messages at SP boundaries
  - Filter/modify QoS parameters

Interdomain path setup

- Contiguous LSP
  - End-to-end LSP is built across domain boundaries, using hop-by-hop signaling between adjacent neighbors; similar to setup within one domain

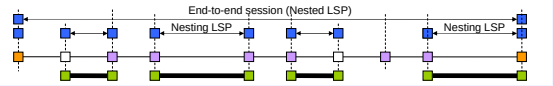


- LSP stitching
  - Smaller LSPs (TE LSP segments) set up in different domains
  - "Stitched" together (connected) at stitching points (domain boundaries)
  - 1:1 mapping of forwarding state at stitching points ⇒ creating a new TE LSP requires establishing new LSPs within all domains
  - LSP functions such as reoptimization and FRR limited to domains
  - End points are usually a domain's ingress and another domain's egress (Provider Edges, PEs)
  - Setup: preconfigured or triggered by LSP setup message from neighboring domain



Interdomain path setup /2

- LSP nesting
  - Tunnel TE LSP within LSP with per-domain scope (Forwarding Adjacency (FA) LSP) as it crosses a domain
  - Stack labels: FA LSP head end pushes FA LSP label on top of nested LSP's label stack
  - Possible to map multiple TE LSPs into one FA LSP (1:N mapping)
    - Stitching requires 1:1 mapping ⇒ forwarding state increases linearly with no. of LSPs
  - Several benefits: e.g. admission control possible for TE LSPs at FA LSP head end, FA LSP's reoptimization and FRR can be used

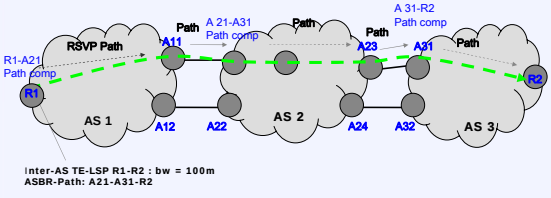


- LSP nesting more efficient – so why stitch?
  - Easier configuration for TE because of 1:1 mapping: ensure that all per-domain LSPs match requirements of TE LSP (requires admission control when nesting)
- Decision for method based on administrative policies at border router

Interdomain path computation

- Limited by per-domain knowledge available to the computing node
  - Constrained by location / ownership of that node: head end, offline tool, domain boundary node?
- Straightforward possibility: establish contiguous LSP where all hops are precomputed and signaled with Explicit Route Object (ERO)
  - Path computation must have interdomain scope (and visibility)
  - Possible with offline tool
- More sophisticated: separate per-domain path computation
  - ERO expansion: calculate path segment between border routers, add to ER
  - But TE constraints can vary between domains, and be incompatible (e.g. DiffServ based vs. link colors vs. max. capacity) ⇒ mapping needed
    - Implies that administrators cooperate to agree on a mapping
    - Topology information and TE characteristics (or mapping) needed

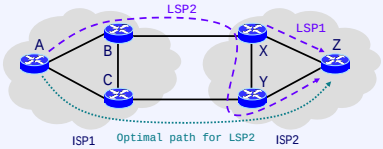
Per-domain path computation



- Done when visibility not given across all domains
  - from one border router to the next; assumption: address of egress border router known
  - Configured as loose hop or discovered dynamically based on IP reachability for LSP destination address ⇒ path to border router
- Path to border router can be used…
  - Contiguous LSP setup: during ERO expansion
  - Stitching: for setting up relevant TE LSP segment
  - Nesting: for setting up FA LSP

Per-domain path computation /2

- Problem: limited knowledge
  ⇒ per-domain optimal path may not yield optimal end-to-end path

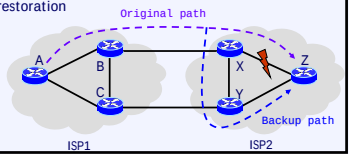


- Example above: assume similar link bandwidths
- From A's point of view, B is as good as C
  - but path via C would have been better – unknown to A
  - can yield bad performance, but also admission control failure (e.g. if link X-Y does not exist or does not have enough resources)

Intradomain failures

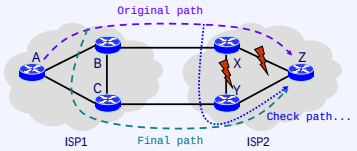
- Example on previous slide: what if reason to choose C instead of B (LSP1) is on the interdomain link (B-X)?
  - Admission control fails
  - Reason undetected because interdomain link is not in TED
  - Possibility: include it (via IGP advertisements)
    - But LSP setup can still fail – e.g. if resources become unavailable between time of computing and signaling
    - One ISP can prevent this from happening for its own links only
- Recall: intradomain link failure restoration
  - Node where admission control fails sends path error message to head end
  - Head end calculates backup path; topology knowledge assumed

⇒ different solution needed



Crankback

- What if link X-Y does not exist or does not have enough resources?
  - X cannot find a backup path
  - But a backup path would exist: A-C-Y-Z
  - Solution: treat X as the blocked resource, move computation back one step at a time, away from point of failure (signaled with RSVP extensions)



- As above, result may not be ideal; crankback has several other issues
  - e.g. need to avoid pointless probing for paths which are unavailable

Path Computation Element (PCE)

- Idea: problems before do not occur when the *whole* path is specified at the head end
  - Path Computation Element (PCE) was designed for that purpose: know everything, enabling offline computation of ideal interdomain path
- Reasons for offline computation:
  - Constraint communication:
    - intradomain: some constraints (e.g. link colors) must only be known at the head end
    - interdomain: all nodes participating in path computation must learn constraints
  - Need for extra visibility: for interdomain LSPs which have head and tail ends in the same domain but traverse other domains
  - Constraint translation: static mapping or have a different entity translate
  - Optimality of the solution: CSPF may not suffice

Path Computation Element /2

- PCE can be a module on a router or a separate entity that a router communicates with
- Issues that must be solved for towards a PCE based solution (current work in progress in IETF PCE WG)
  - Router-PCE communication protocol
  - PCE-PCE communication protocol
  - PCE discovery: how does a router find a PCE? Autodiscovery desirable
    - extensions for IS-IS and OSPF have been specified
  - Acquiring TED
  - Develop suitable computation algorithm (not standardized, left flexible)

Reoptimization

- Reminder about reoptimization:
  - dynamic calculation of better path, triggered by e.g. operator request, expiration of a timer, .. at head end
  - usually make-before-break
  - two steps: path computation and signaling
- Interdomain environment: path computation method and signaling method influence how reoptimization works
  - per-domain computation is good because...
    - reoptimization can be done within one domain
    - reasons for reoptimization are usually local ones
    - interdomain LSP's head end is unaware ⇒ better scalability
  - ... but can violating tight constraints (consider previous examples of nonoptimal solutions from local decisions)
  - Contiguous LSP: head end must be in control
    - RSVP extension for signaling reoptimization requests to downstream nodes
    - Better control of interdomain LSP but increased complexity

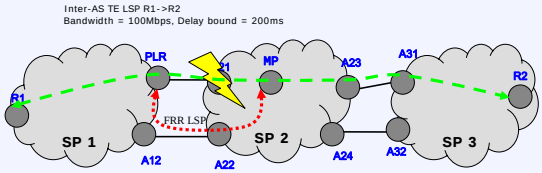
Multi-domain Route Recording

- Signaling mechanism providing for diagnostic information about the path of an established LSP
- Route record processing at inter-domain boundaries
  - boundary node may remove, filter or aggregate some of the recorded information for trust/confidentiality/policy reasons
    - route record may not be available on a Path message
  - in case of per domain boundaries path computation
    - computed path may loop back into a domain that has already been crossed by the LSP
    - loop avoidance: information pruning during path computation using the route record in Path message

Protection and FRR

- End-to-end protection path diversity desirable for primary / secondary path
  - attaining this in a multi-provider environment is more difficult
- Local protection
  - within each domain: no difference to interdomain LSPs
    - Stitching: protection path is applied to TE LSP segment
    - Nesting: protection path is applied to FA LSP
    - no other special mechanisms needed
  - between domains: PLR and MP in different domains
    - how to identify the MP and compute path to it?
    - independent of LSP setup method

Fast Reroute



- How does PLR learn about MP?
  - typically done via RRO using interface addresses
  - Not available in interdomain setup ⇒ FRR specification extended to advertise node ids (loopback addresses)
  - Contiguous LSP: MP can be any downstream node
  - Stitching or nesting: MP can only be the tail of TE LSP segment or FA LSP

Summary

- Interdomain TE somewhat more complex than intradomain
  - path calculation based on incomplete information, mapping / translation of constraints, ...
- Three methods of signaling LSP setup (contiguous, stitching, nesting) influence operations on intradomain LSP
- Management challenges: policies and contracts needed between provider edges
  - interprovider LSPs require a certain level of trust
  - requests to head end may be rate limited at domain edge to avoid DoS
  - LSP setup requests must be authenticated, types of requests typically negotiated ahead of time at domain boundaries

References

- Ina Minei, Julian Lucek: "MPLS-Enabled Applications", John Wiley & Sons, 2005, ISBN: 0-470-01453-9
- Slides from Dimitri Papadimitriou
  - Thanks!!!