QOSPF:
Quality of Service Extensions to OSFP
or
Quality of Service Path First Routing

Eric S. Crawley

Bay Networks, Inc.

esc@baynetworks.com

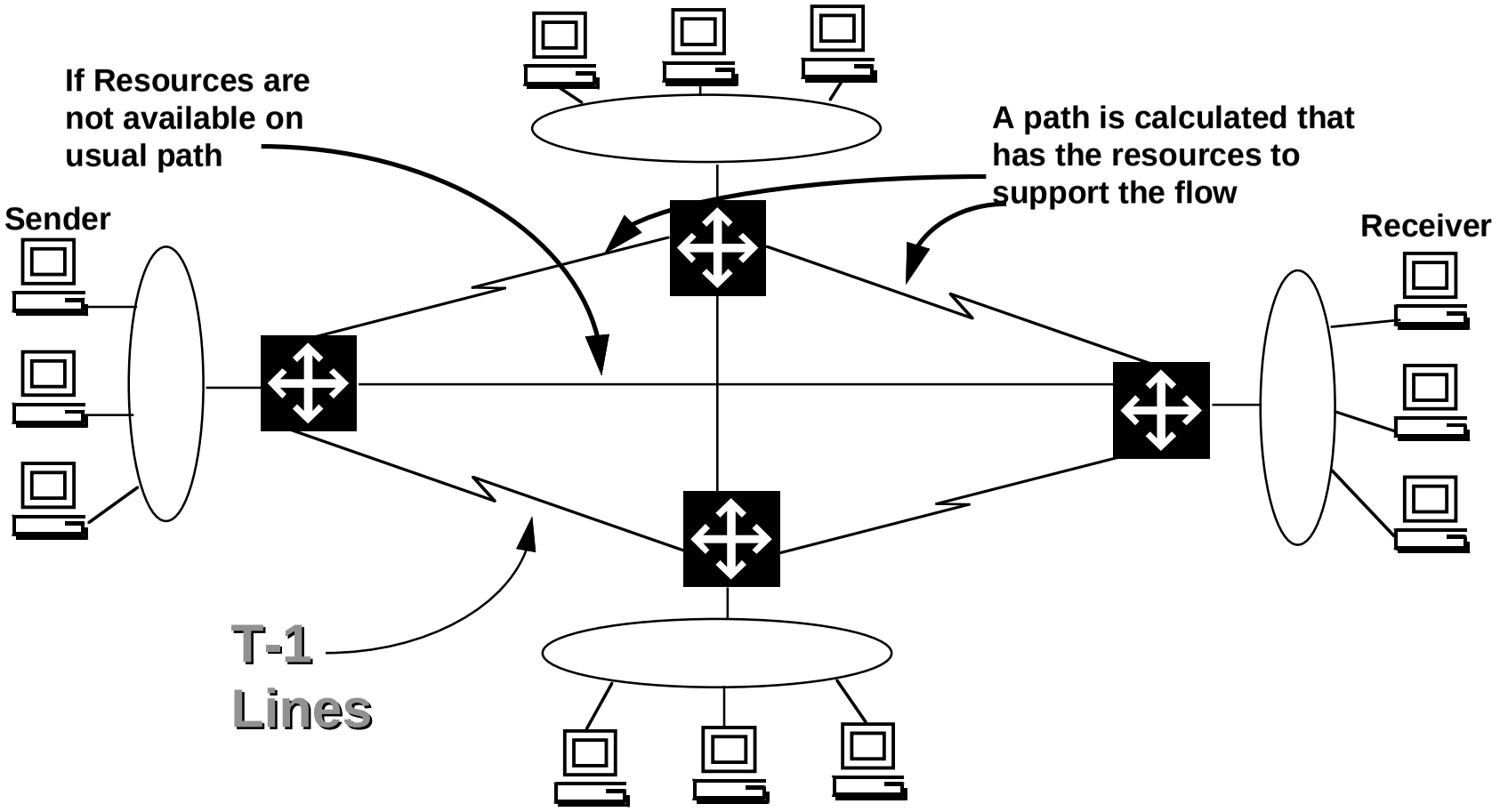
Overview

- The Challenge
- Protocol Overview
- Scaling Problems
- The Details
- The Remaining Problems

The Challenge

- A customer asked us to combine MOSPF and RSVP to provide Quality of Service routing in a WAN environment
- The WAN was mostly T-1 links but sufficiently redundant that there were multiple paths between multicast sources and destinations
- The application closely matched the “Cable TV” type of environment
  - Single sender
  - Multiple receivers coming and going
  - Fast receiver JOIN times needed (< 2 sec)

Example Topology



Basic Protocol Steps

- Network resources are advertised to the area
- Trigger event causes calculation of QoS route (demand driven)
  - RSVP messages containing resource requirements
- QoS route is computed as needed by routers on path
- Notification of actual reservation of resources to assure that path computation was correct and “store state”
- QoS route is removed when trigger events disappear (RSVP PATH state times out)

Protocol Overview: Resource Advertisements

- New OSPF LSAs for advertising router link resources
  - Controlled Load Model used
    - Advertise Largest chunk of reservable BW
    - Advertise Largest buffer/burst reservable
    - For each link
- Other service models/parameters can/should be added
- Provides information for QoS routing calculation

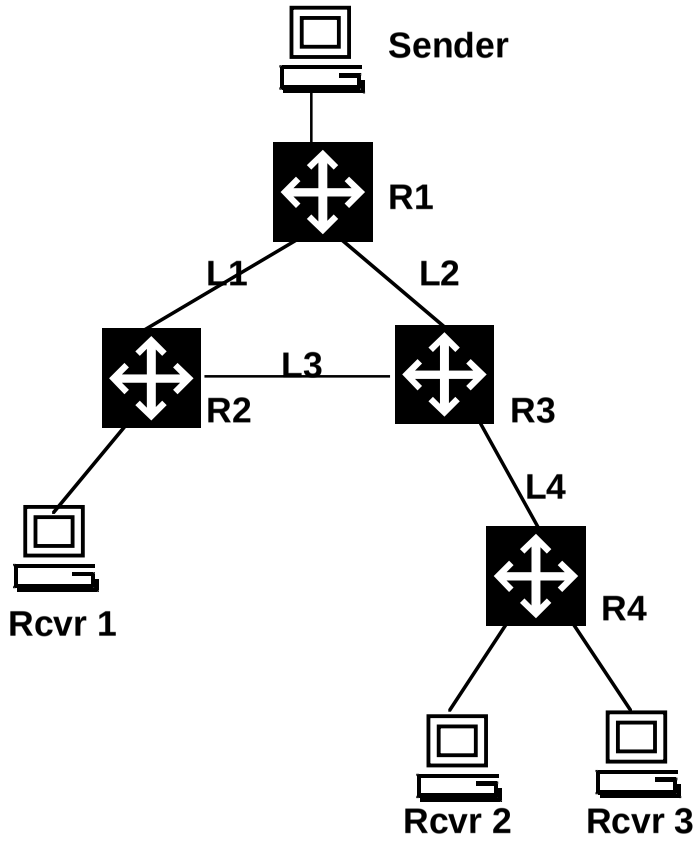
Protocol Overview: Route Computation

- QoS Dijkstra calculation triggered by receipt of new RSVP PATH message
  - Other triggers such as different signaling protocols and flow thresholds can be used
  - Multicast uses Group Membership LSAs for member list
  - QoS route computed for Source->Destination pair
  - This means ***all*** traffic for Source->Destination pair is routed on same path
- Only Links with adequate resources are used for Dijkstra
  - Recovery to best-effort path possible

Protocol Overview: Confirmation/ Recalculation

- Success on RESV triggers Opaque LSA indicating reservation on link for Source->Destination pair
- Recomputation could occur on RESV if resources different from PATH
- Routes may be pinned (e.g. not changed by new links or more resources becoming available)
- New resource LSAs (reserved and available) and topology changes cause recalculation favoring links currently reserved for flow

Example



1. RCVRs 1,2,&3 have joined group G

2. RSVP PATH Message sent from sender to G

3. R1 Computes QoS Path to RCVRs 1,2,&3
   - Link L1 does not have adequate resources for flow based on RSVP TSpec
   - Tree is computed using L2, L3, and L4

4. Process is repeated for R2, R3, and R4 as PATH message travels down tree

5. Rcvrs send RESV message back toward sender, reserving resources

6. Routers send Resource reserved advertisements indicating the resources used for the Source, Group pair

7. If significant resources are used, routers send new Resource Available advertisements

Why are Resource Reserved Advertisements Needed?

- Need to save tree state somehow
  - For adding new members and knowing what branches are currently in use
  - So new resource available advertisements don't affect paths currently in use (AKA "Stepping on your own Shadow")

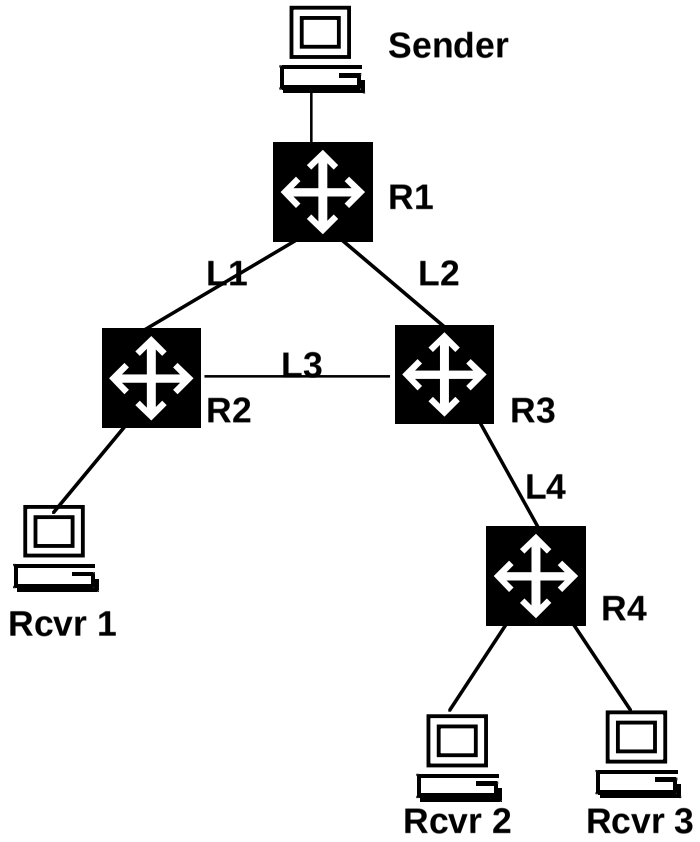
Scaling Problems

- Resource Available LSAs can be sent out too frequently
  - Solution: Use watermarks and/or resource reserved > resource available comparisons
- Resource Reserved Advertisements are flooded for every link on source->destination path generated on receipt of RESV
  - Solution: Explicit Routing

Explicit Routing (EROSPF)

- A router, usually the source router, computes the QoS route path from source->destination
- The route computation is passed down the tree using opaque LSAs (Explicit Route Advertisements, ERAs)
- ERAs encode the route in such a way that no recomputation is necessary for each router along the path
- Routes can be broken up into multiple ERAs to avoid MTU limits
- Resource Reserved Advertisements need only be sent back to the computing/source router
- Flushing ERAs are used to remove routes

Explicit Routing Example



1. RCVRs 1,2,&3 have joined group G

2. RSVP PATH Message sent from sender to G

3. R1 Computes QoS Path to RCVRs 1,2,&3
   - Link L1 does not have adequate resources for flow based on RSVP TSpec
   - Tree is computed using L2, L3, and L4

4. R1 sends ERA to R3 with subtree of R2 and R4

5. R3 installs ERA state and sends ERAs to R2 and R4

6. Process is repeated on R2 and R4 installing state

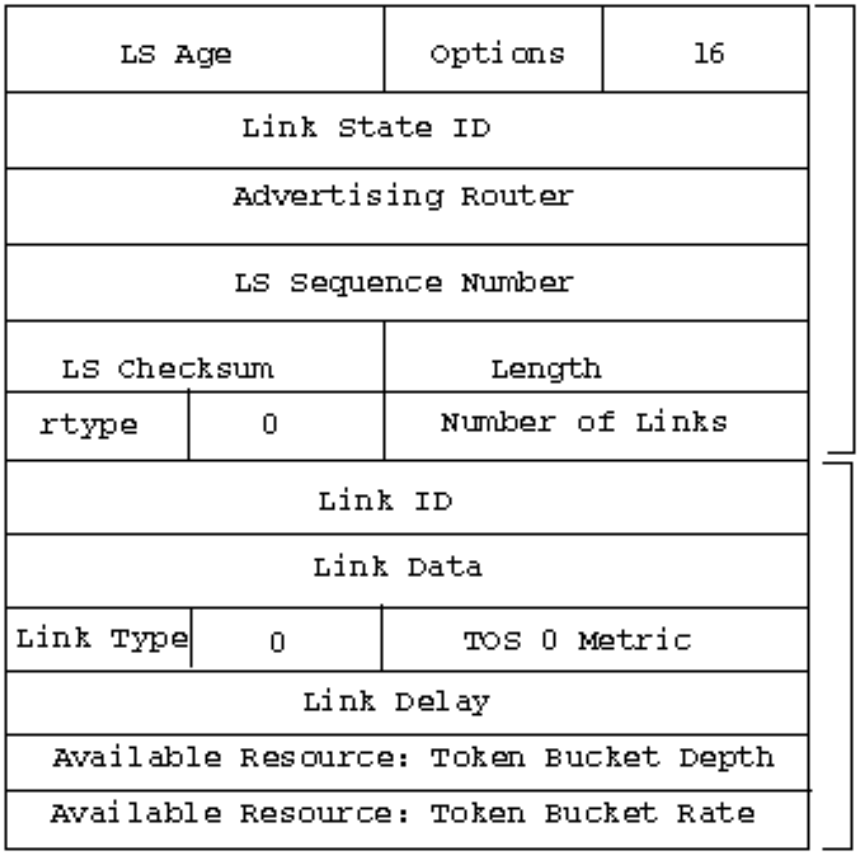
7. Rcvrs send RESV message back toward sender, reserving resources

8. Routers send Resource reserved advertisements *only to R1* indicating the resources used for the Source, Group pair

OSPF/MOSPF Extensions

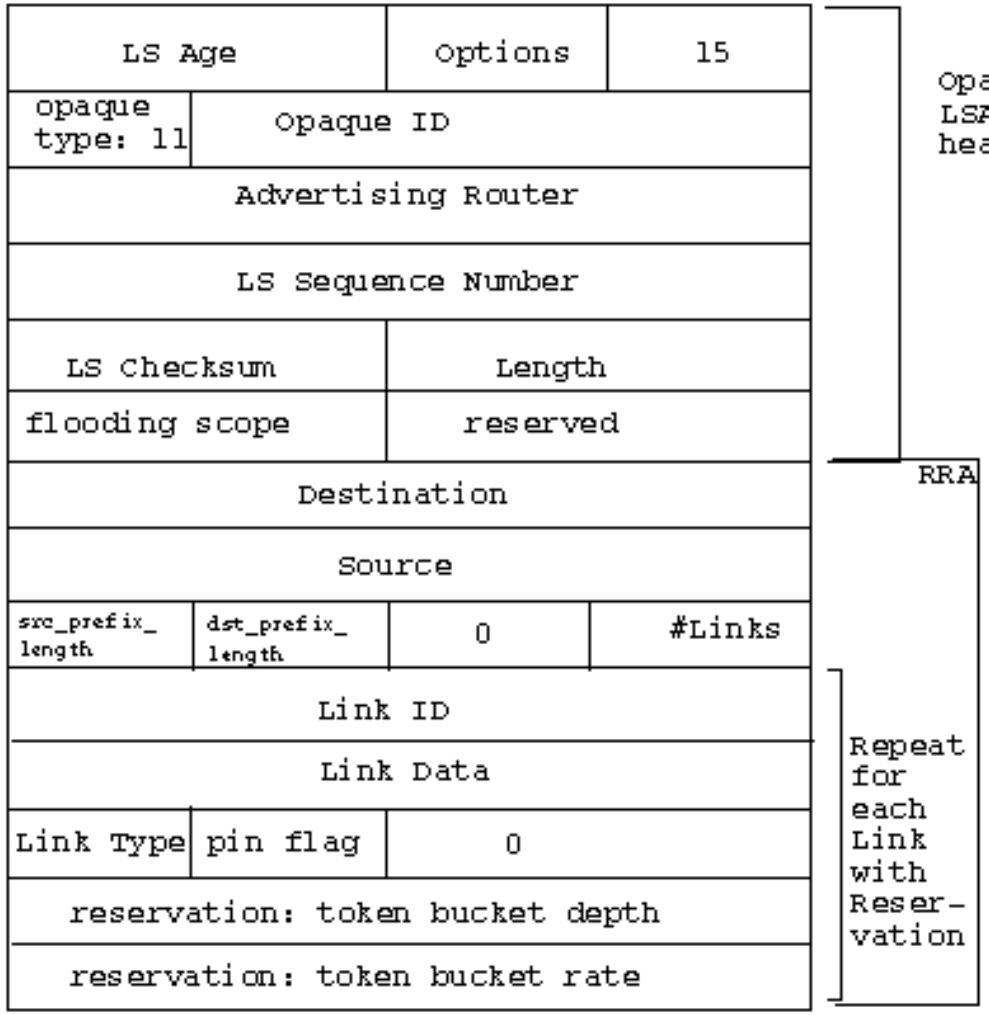
- Resource Available LSA
- Resource Reservation Advertisement
- Explicit Route Advertisements
- Border Router Advertisements
- Route Computation Changes

Resource Available LSA (RES-LSA)



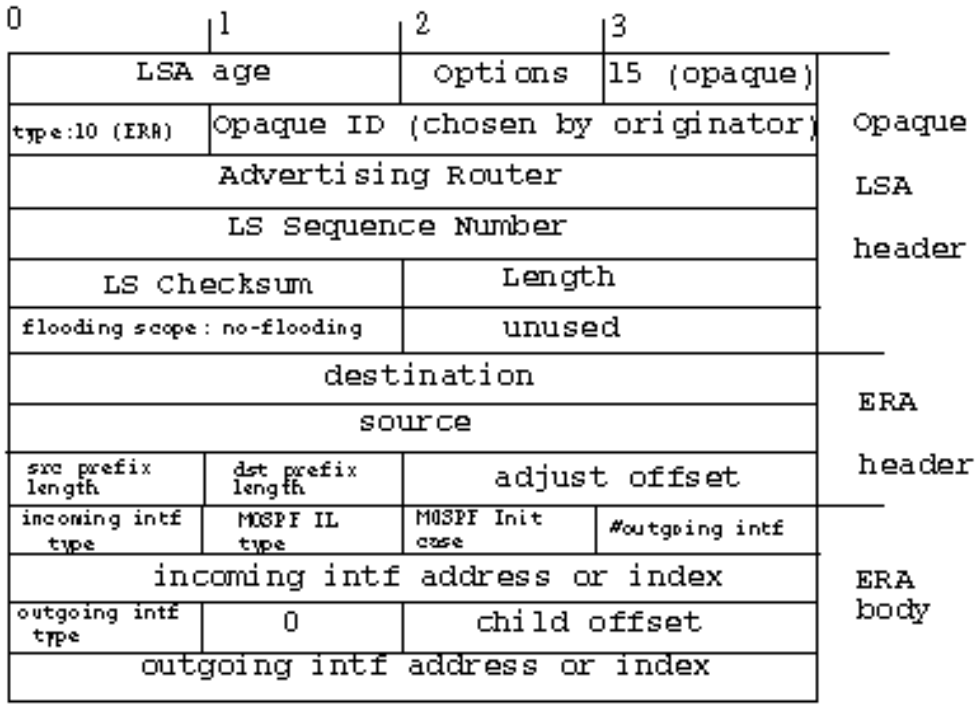
- Flooded through Area
- Token Bucket parameters are single precision floating point (IntServ std)
- Link Delay is used in place of TOS 0 metric for QoS route calculation
- Could be expanded for other service models

Resource Reservation Advertisement (RRA)



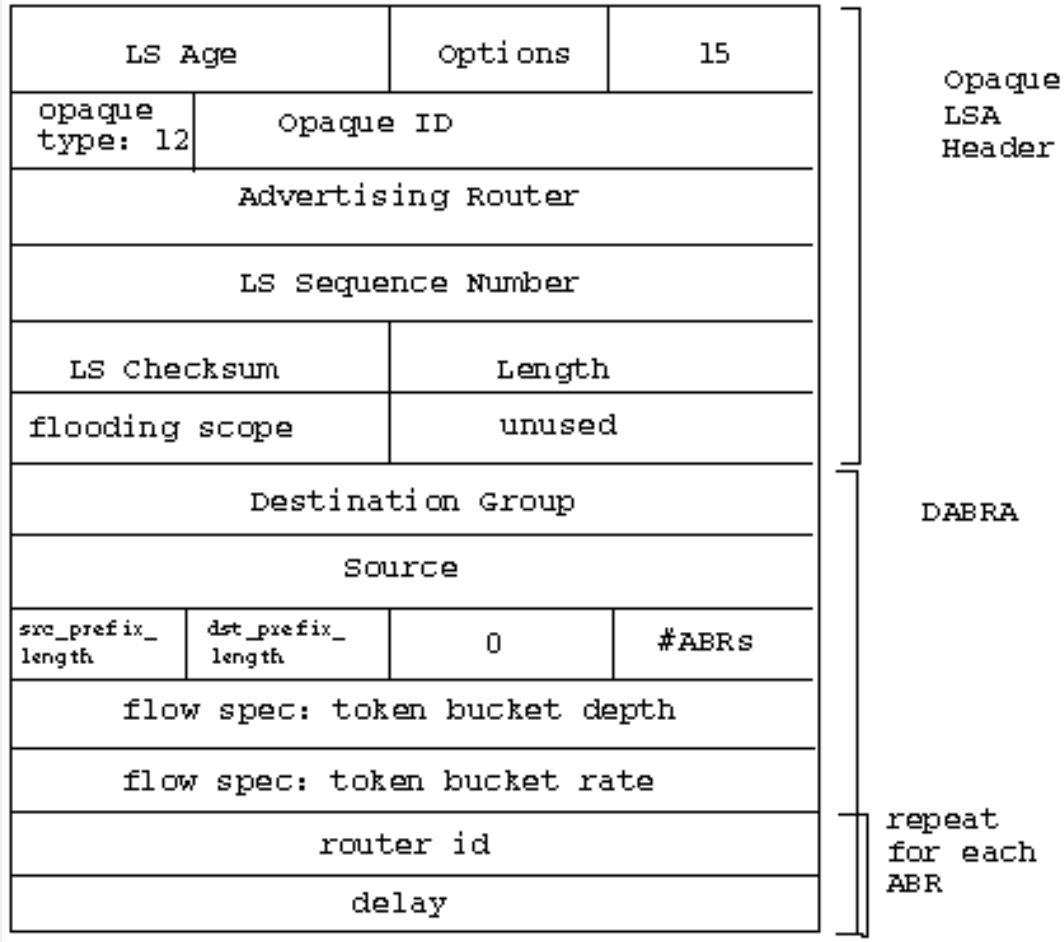
- Flooded “Area Local” without ER
- Flooded “No Flooding” and “Link Local” with ER
- Source and Destination are the IP source and destination addresses
- Pin Flag indicates whether route pinning has been requested

Explicit Route Advertisement (ERA)



- "No Flooding" scope
- ERA header indicates source-> destination pair
- ERA body encodes path
- Adjust Offset is used to avoid recomputing offsets for subsequent ERAs
- Child offset indicates next child in for interface
- LSA Age = MAXAGE indicates flushing LSA

Border Router Advertisement (DABRA)



- Indicates to “downstream” areas how to root a tree (which area border router to use) for source-> destination
- Flooded to all downstream areas from borders that receive flow without ER
- Sent only to border routers on path with ER

Route Computation Changes

- Topology Change - all routes recomputed (conventional or QoS)
  - Must use RRAs to determine resources used by current path and pinned routes
- New MOSPF Group Membership LSAs
- New RES-LSAs - all QoS routes recomputed
- New RRAs - all QoS routes related to the RRAs recomputed
- New DABRAs- all QoS routes related to the DABRAs recomputed

Multicast Issues

- Adding members/branches can be tricky so you don’t “avoid your own shadow”
  - Must know where current distribution tree passes
- Border router issues are interesting
  - Enough information to prevent loops
  - Not too much to keep summarization
- Solutions may be very different for shared tree protocols
- Interdomain solutions will be very interesting...

Possible RSVP Additions

- Route pinning flag so receivers can request a “pinned” route
  - Possible length of time to pin a route?
- Indication that a RESV differs from a PATH to help determine if recalculation is needed

Remaining Problems

- Unicast can possible suffer from route loops if packets “fall off” the QoS computed path
  - May need to tag or mark packets on QoS paths to note different treatment (TOS bits or Flow ID?)
  - If packets “fall off” QoS path mark is removed
- Need to think through other RSVP reservation styles (SE & WF)
- Need more experience, simulation, and real use