Use of Maximally Redundant Trees (MRTs) in Bridged Networks

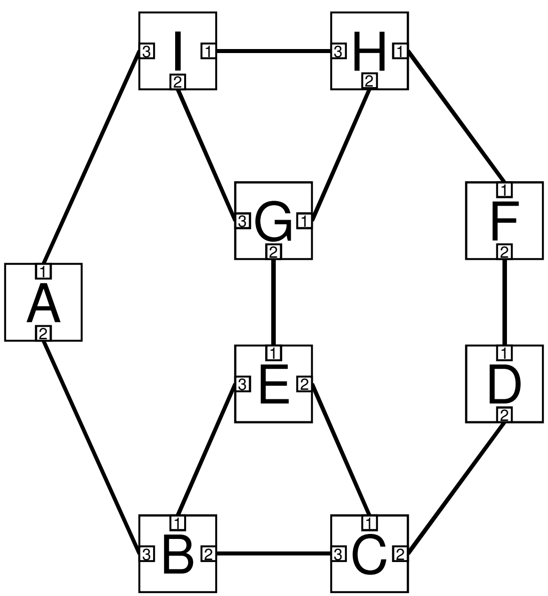
János Farkas

2015-05-19

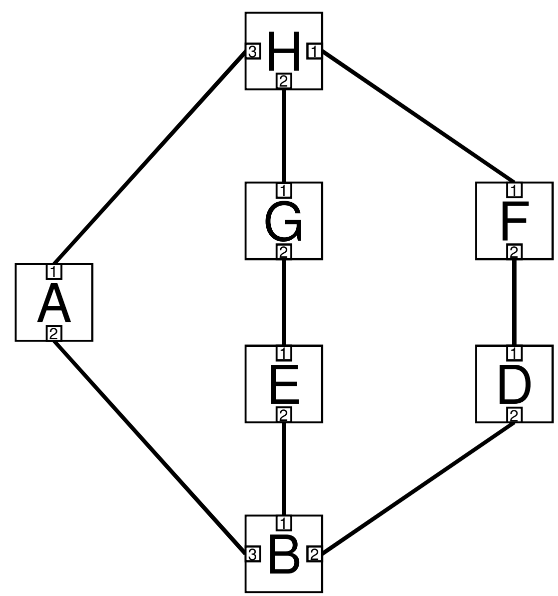
Background

- [1] https://datatracker.ietf.org/doc/draft-ietf-rtgwg-mrt-frr-architecture
- [2] https://datatracker.ietf.org/doc/draft-ietf-rtgwg-mrt-frr-algorithm/
- [3] http://www.ieee802.org/1/files/public/docs2014/ca-farkas-mrt-0114-v01.pdf

Example Topologies



and



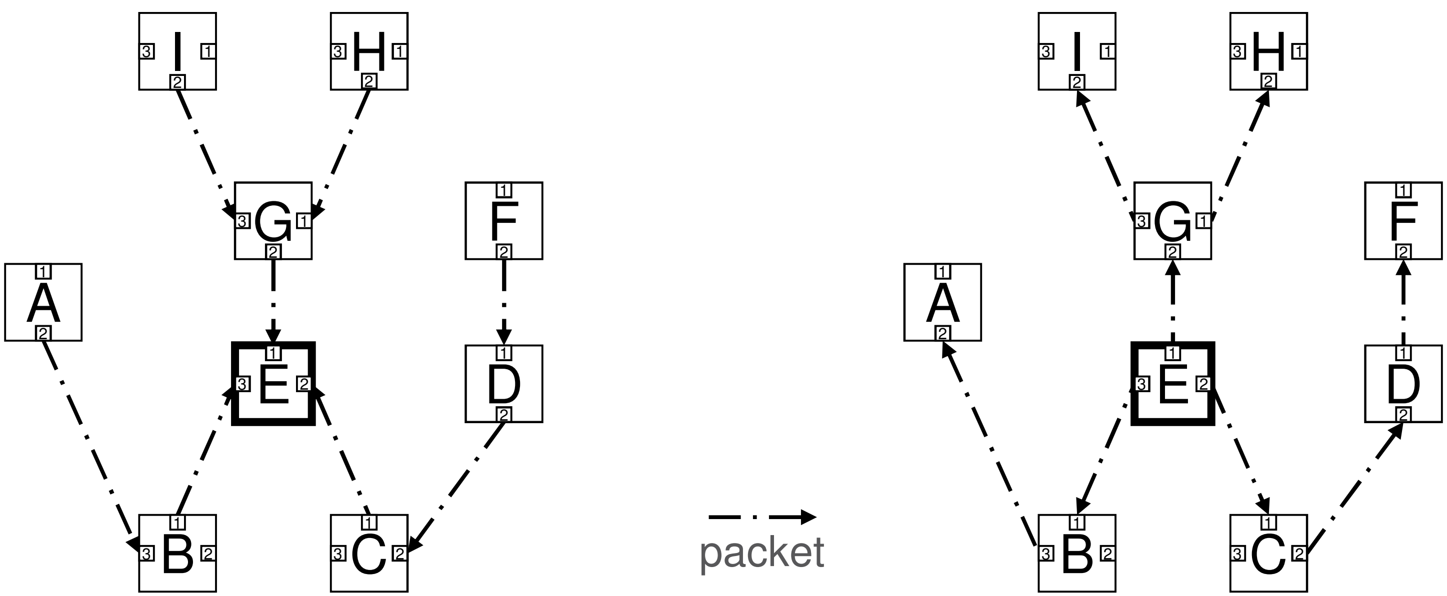
Shortest Path Trees

Shortest Path Tree
SPT Root = E

- **This is about non-learning for now; FDB entry: MAC+VID, e.g. SPBM**

- Each node determines its FDB entry to destination E
  - Next hop to reach E in IP terms
- Considering all FDB entries for E, an SPT can be drawn

- The source rooted tree can be retrieved by reversing direction
- Due to symmetric SPTs, it is equivalent with how each destination is reached from E



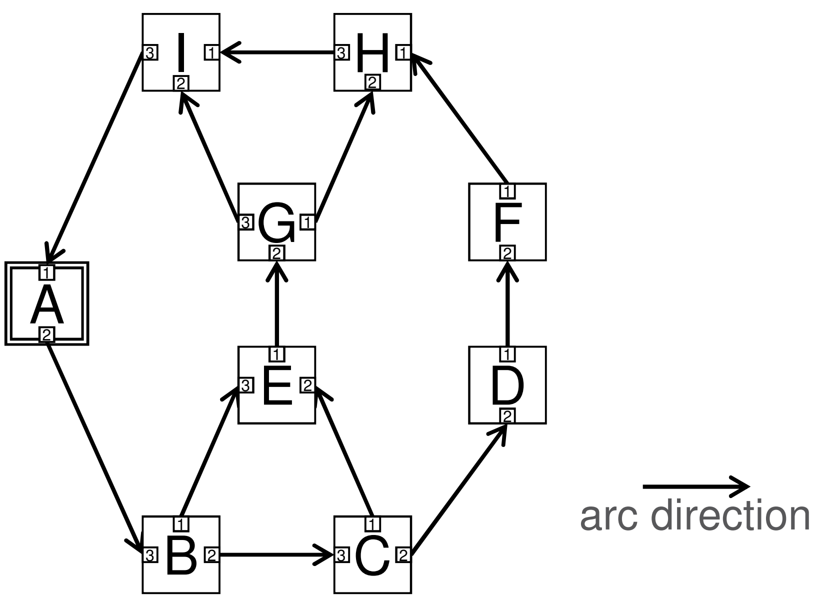
destination rooted source rooted

Maximally Redundant Trees

MRT Lowpoint Algorithm Basics

draft-ietf-rtgwg-mrt-frr-algorithm

1. GADAG is computed
2. Blue and Red next hops determined for each destination

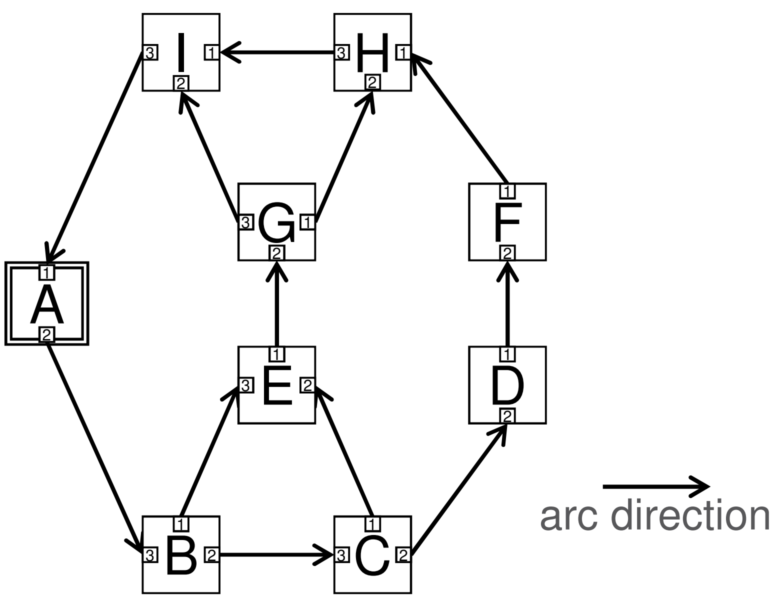


GADAG; GADAG Root = A

- GADAG provides ***partial order***
- Many nodes are ***ordered*** with each other
  - X > Y or X < Y
  - direction of arc(s) follows order
  - e.g. A < B < C < D < F < H < I
- Some nodes are ***unordered*** with each other
  - X ? Y
  - e.g. E ? D, E ? F, G ? D, G ? F

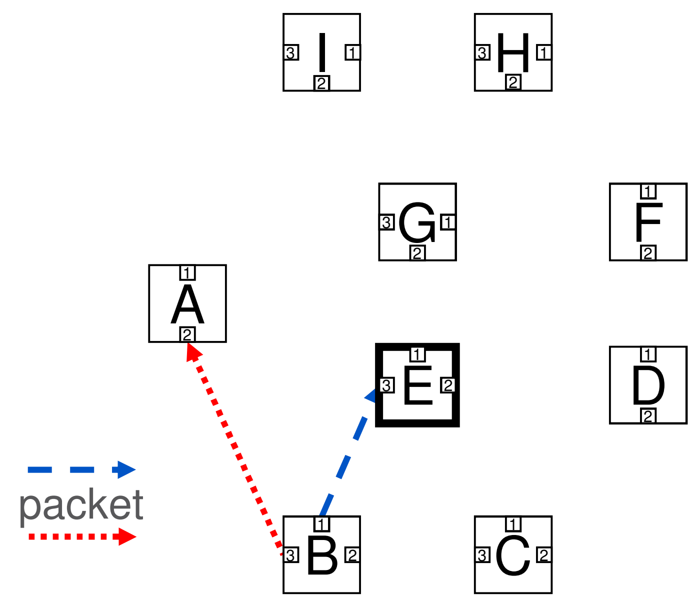
MRT Next Hops – X < Y

- X = computing node, Y = destination node
- Blue next-hop: the next-hop along a shortest directed path from X to Y (Blue follows arc direction)
- Red next-hop: the next-hop along a shortest directed path from X to the GADAG Root with reversed arcs (Red follows reversed arcs)



GADAG; GADAG Root = A

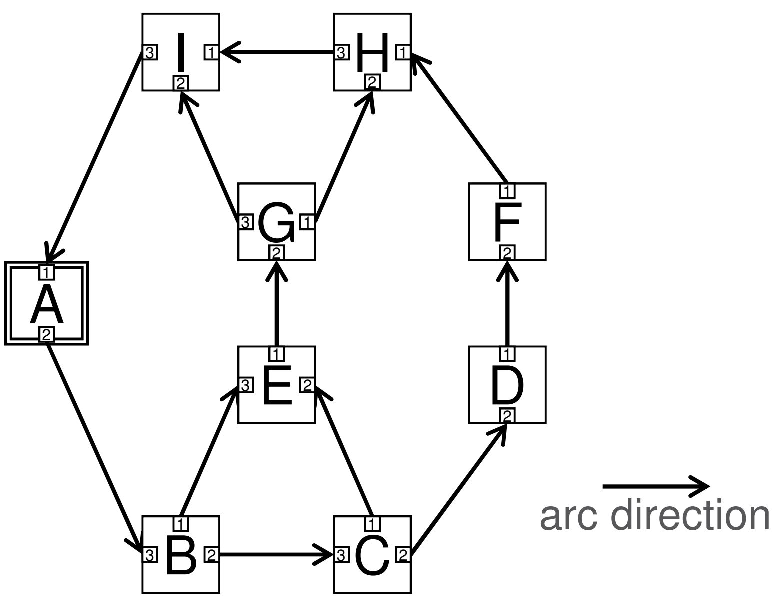
B < E



B's next hops to E

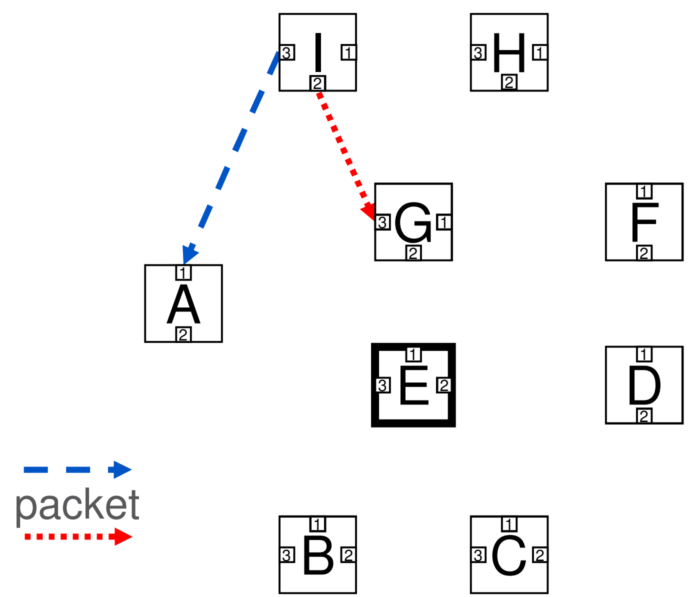
MRT Next Hops – X > Y

- X = computing node, Y = destination node
- Blue next-hop: the next-hop along a shortest directed path from X to the GADAG Root
- Red next-hop: the next-hop along a shortest directed path from X to Y with reversed arcs



GADAG; GADAG Root = A

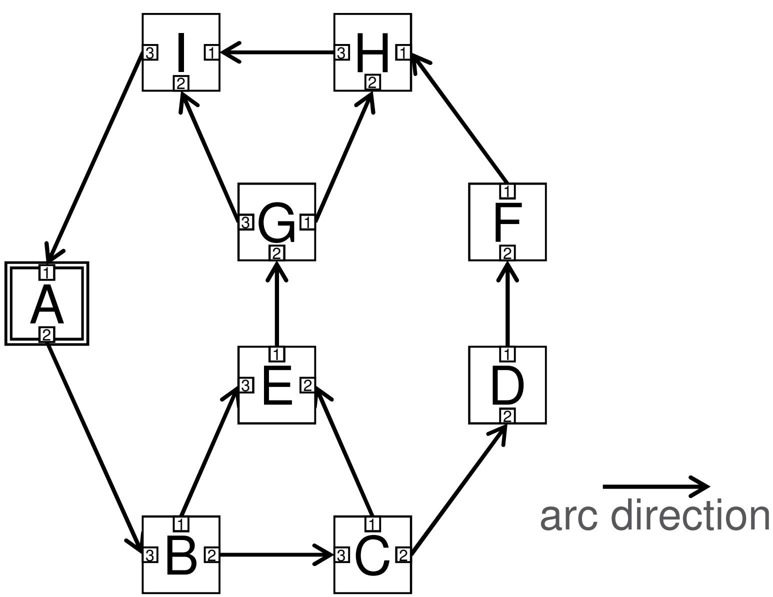
I > E



I's next hops to E

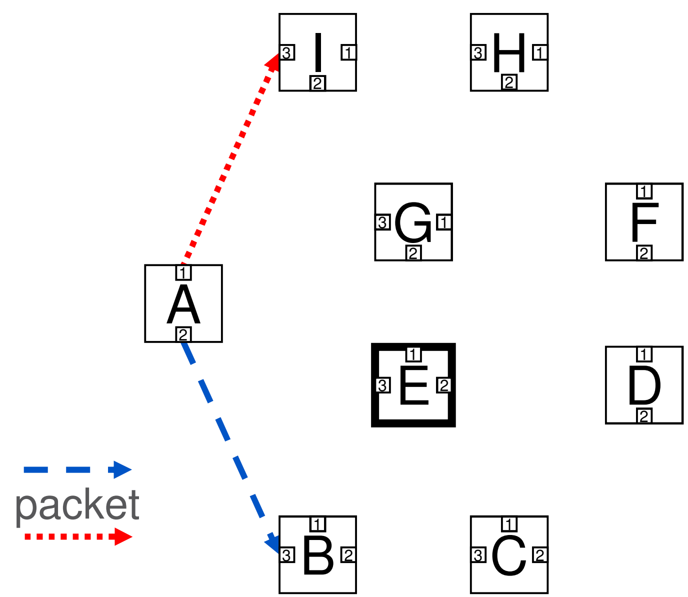
MRT Next Hops – GADAG Root

- X = computing node = GADAG Root, Y = destination node
- Blue next-hop: the next-hop along the shortest directed path from X to Y
- Red next-hop: the next-hop along the shortest directed path from X to Y with reversed arcs



GADAG; GADAG Root = A

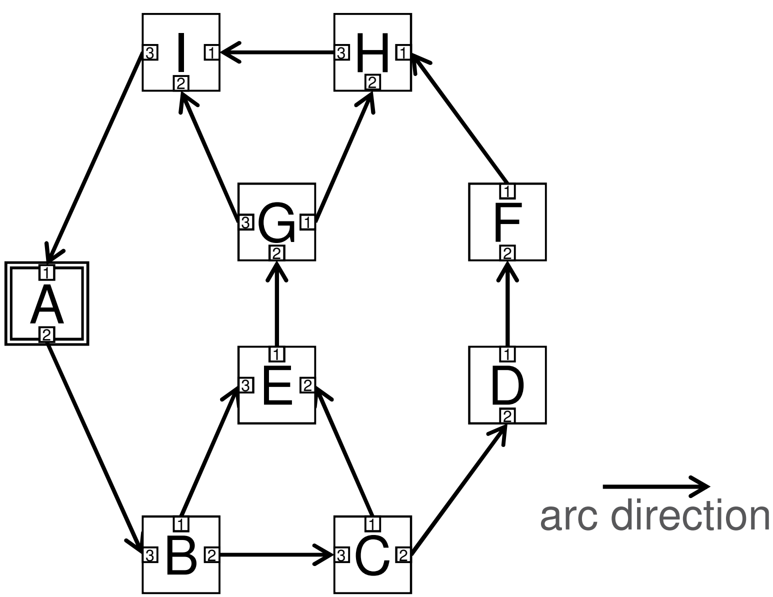
A = GADAG Root



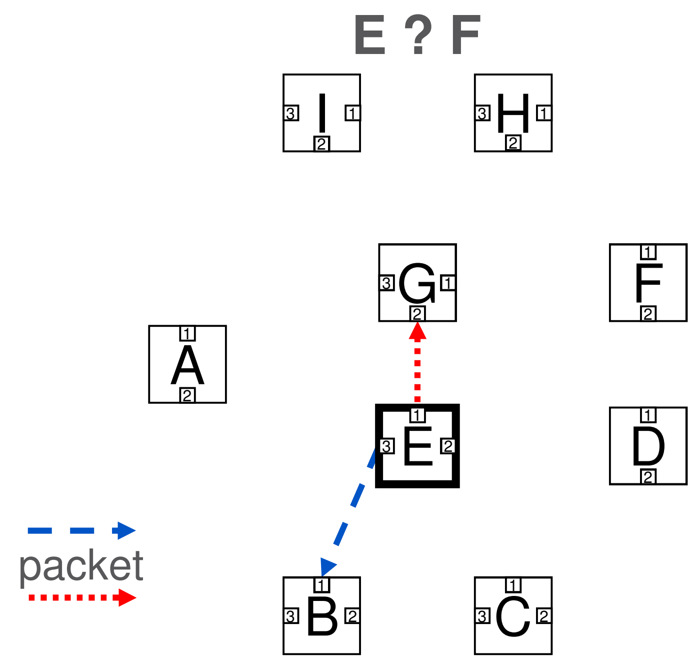
A's next hops to E

MRT Next Hops – Unordered Nodes: X ? Y

- X = computing node = GADAG Root, Y = destination node
- Blue next-hop: the next-hop along the shortest directed path from X to the GADAG Root with reversed arcs
- Red next-hop: the next-hop along the shortest directed path from X to the GADAG Root



GADAG; GADAG Root = A



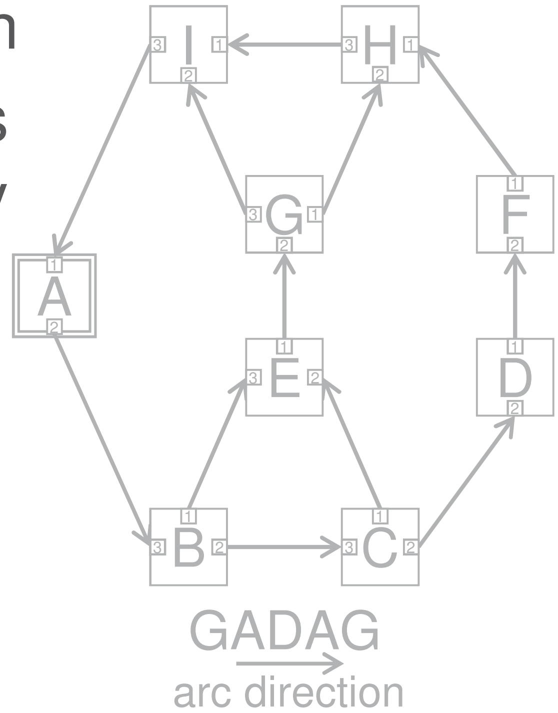
E's next hops to F

MRT Next Hops – Summary X = computing, Y = destination

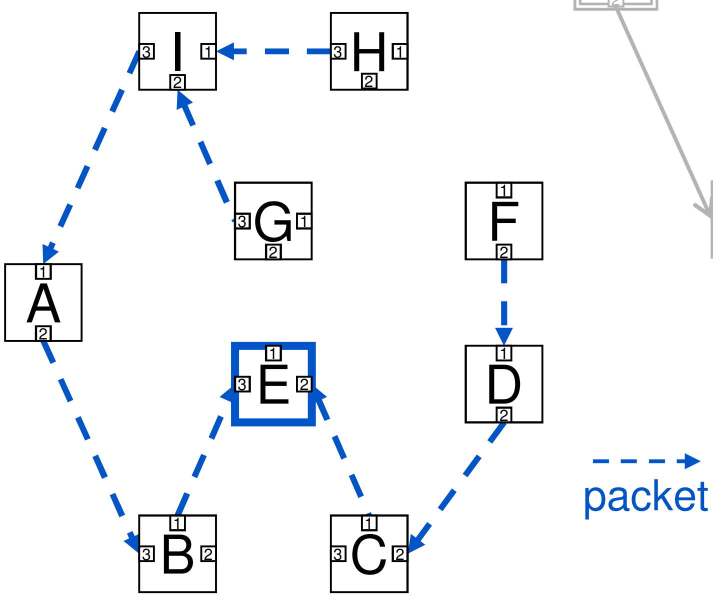
- If X < Y:
  - Blue: the next-hop along a shortest directed path from X to Y
  - Red: the next-hop along a shortest directed path from X to the GADAG Root with reversed arcs
- If X > Y:
  - Blue: the next-hop along a shortest directed path from X to the GADAG Root
  - Red: the next-hop along a shortest directed path from X to Y with reversed arcs
- If X is the GADAG Root:
  - Blue: the next-hop along the shortest directed path from X to Y
  - Red: the next-hop along the shortest directed path from X to Y with reversed arcs
- If X and Y are unordered (X ? Y):
  - Blue: the next-hop along the shortest directed path from X to the GADAG Root with reversed arcs
  - Red: the next-hop along the shortest directed path from X to the GADAG Root

MRT Paths to a Destination

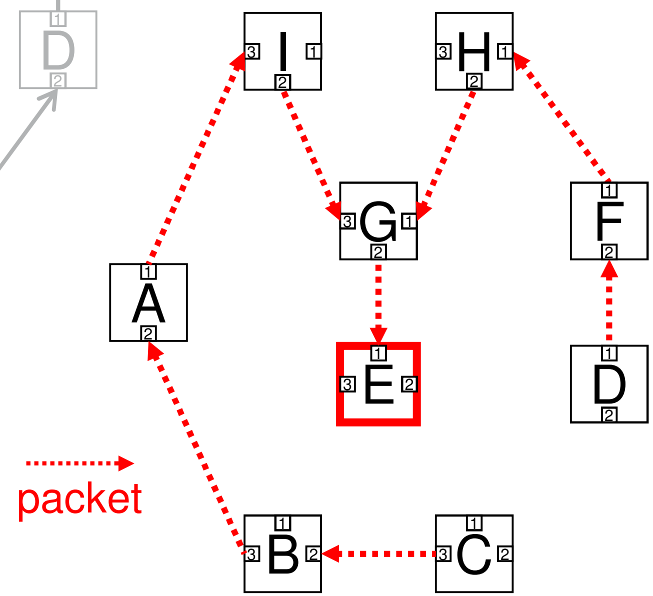
- Considering all FDB entries (next hops), destination rooted MRTs can be drawn
- Source rooted trees can be retrieved by reversing direction



It is still the non-learning case

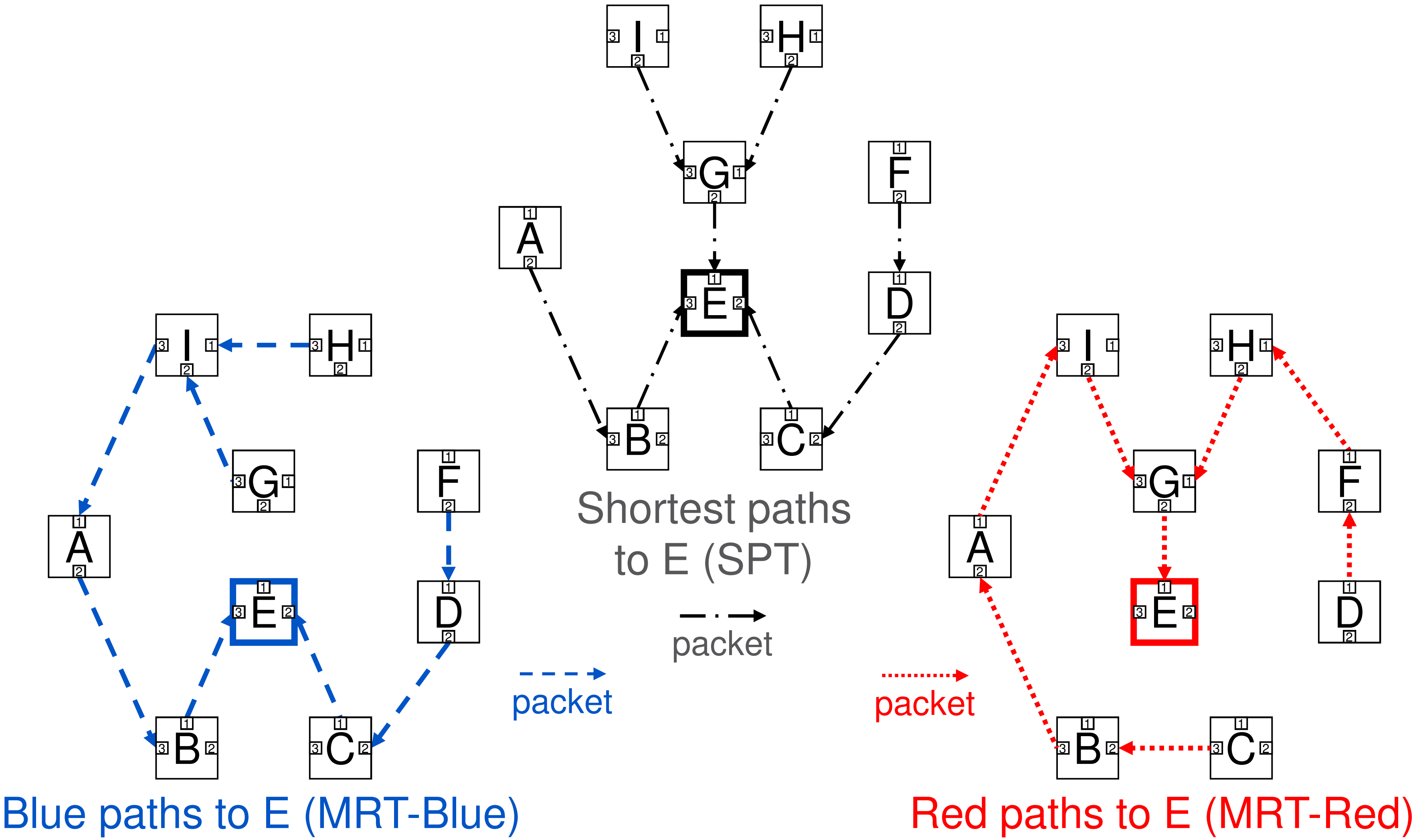


Blue paths to E (MRT-Blue)



Red paths to E (MRT-Red)

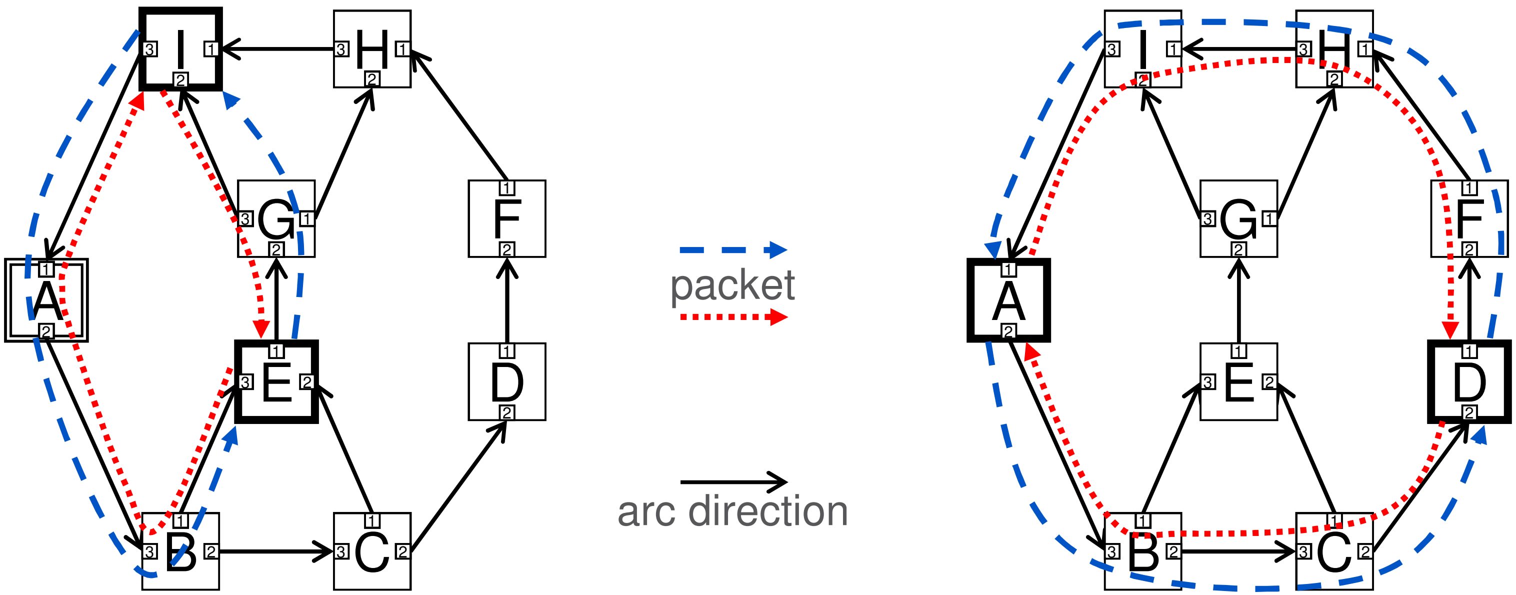
Paths to a Destination



MRT Paths Between a Pair of Nodes – Ordered Nodes

› Between E and I

› Between D and A (GADAG Root)

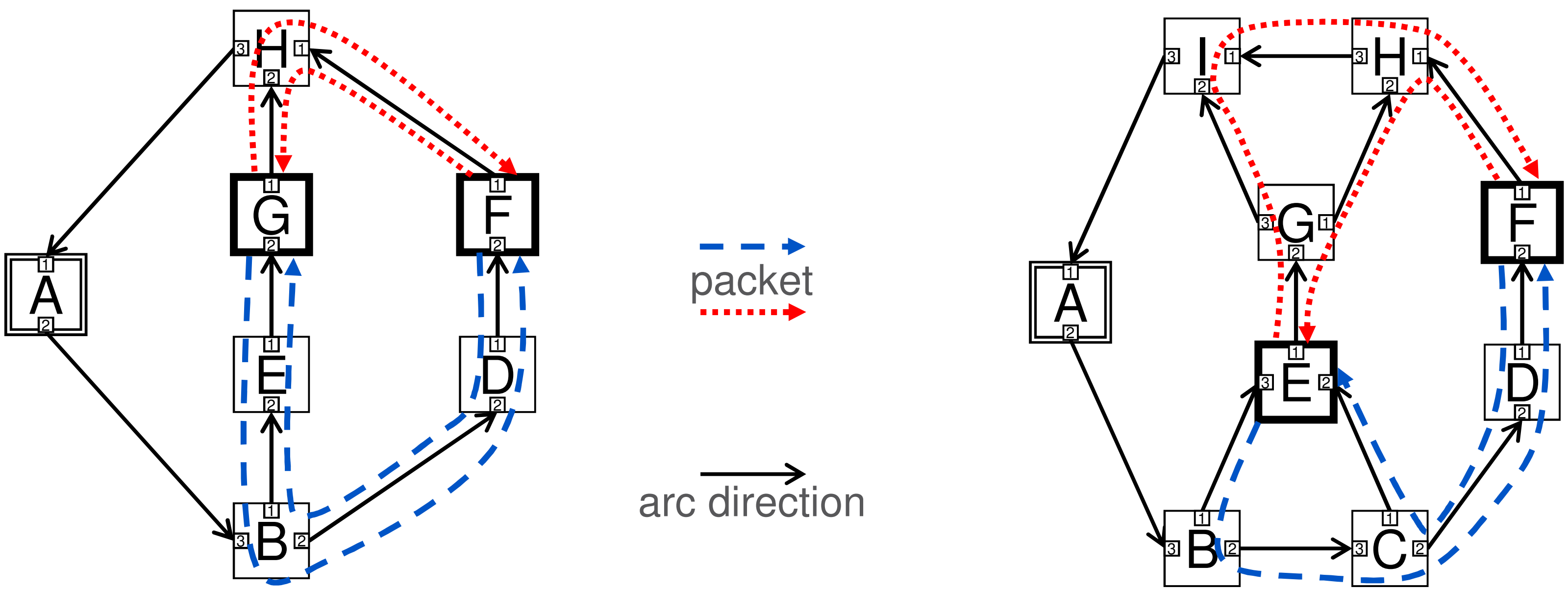


› Symmetry is provided between the nodes by paths having the other color

› Similar to rooted multipoint

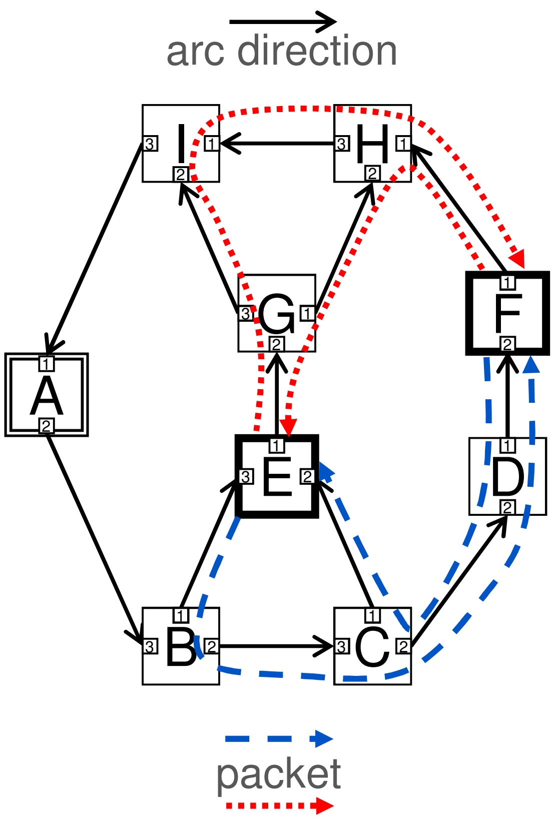
MRT Paths Between a Pair of Nodes – Unordered Nodes

- Note the difference between the two topologies below!
- Between G and F
- Between E and F



- Symmetric paths (same color)
- Not symmetric, see next slide
- Symmetry is not always provided, it depends on topology

Unordered Nodes – Asymmetry



Asymmetry appears if arcs change direction at different nodes in the different paths, see I vs H & B vs C

- Unordered
  - E ? F
    - E's next hops to F
      - Blue = B (the next-hop along the shortest directed path from E to the GADAG Root with reversed arcs)
      - Red = G (the next-hop along the shortest directed path from E to the GADAG Root)
    - F's next hops to E
      - Blue = D
      - Red = H
  - G ? F → G's Red next hop to F is I
  - D ? E → D's Blue next hop to E is C
- Ordered (Blue follows arc, Red is reverse arc)
  - H > E → H's Red next to E is hop G
  - G > E → G's Red next hop to E is E
  - I > F → I's Red next hop to F is H
  - H > F → H's Red next hop to F is F
  - B < F → B's Blue next hop to F is C
  - C < F → C's Blue next hop to F is D
  - D < F → D's Blue next hop to F is F
  - C < E → C's Blue next hop to E is E

Link Metric and Path Cost

- Link metric
  - SPB Link Metric sub-TLV (Type 29) must be used, see 28.12.7
  - Comment #54
- Equal cost paths
  - LowPATHID tie-breaking must be used
  - It is the default tie-breaking, see 28.7
  - Comment #77

Operation Details

IP FRR Operation
(Section 12.2 of [1])

- When a ***failure event*** happens, traffic is put by the Point of Local Repairs (PLRs) onto the MRT topologies.
- After that, each router recomputes its shortest path tree (SPT) and moves traffic over to that.
- Only after all the PLRs have switched to using their SPTs and traffic has drained from the MRT topologies should each router install the recomputed MRTs into the Forwarding Information Base (FIB).
- At each router, therefore, the ***restoration*** sequence is as follows:
  1. Receive failure notification
  2. Recompute SPT
  3. Install new SPT
  4. If the network was stable before the failure occurred, wait a configured (or advertised) period for all routers to be using their SPTs and traffic to drain from the MRTs.
  5. Recompute MRTs
  6. Install new MRTs.

IP FRR – Failure Handling

- PLRs redirect traffic from SPT to MRT
- It is known in advance whether MRT-Blue or MRT-Red to be used for handling a particular failure
- FIB entry belonging to SPT is replaced by FIB entry belonging to MRT-Blue or MRT-Red upon a failure event
  - An implementation possibility is to handle the three FIB entries of a given destination similar to ECMP such that SPT entry is chosen first, then the blue or the red entry depending on the failure

Establishment of MRTs by Qca

- MRT ECT Algorithm
  - Fully distributed computation, BLCE of each bridge computes GADAG and MRTs for each MRT Root
  - IS-IS restores the trees
- MRTG ECT Algorithm
  - GADAG is computed by PCE, flooded in a Topology sub-TLV
  - BLCE of each bridge computes MRTs for each MRT Root
  - Restoration is initiated by PCE flooding a new GADAG, IS-IS does the rest
- ST ECT Algorithm
  - PCE computes the GADAG and all the MRTs
  - Blue and Red MRTs encoded in Topology sub-TLV are individually flooded by PCE for each MRT Root
- The following can be applied to all

MRTs for Seamless Redundancy

- MRTs can be used with 802.1CB Seamless Redundancy
- A copy of a data frame is sent both on the Blue and the Red MRTs (or even on SPTs)
- No reaction is needed to handle the failure event as data frames are sent on multiple paths anyways
- There is no Point of Local Repair (PLR)
- Therefore, no concerns on how a failure event is exactly handled

- (Comment #58)

VLAN IDs

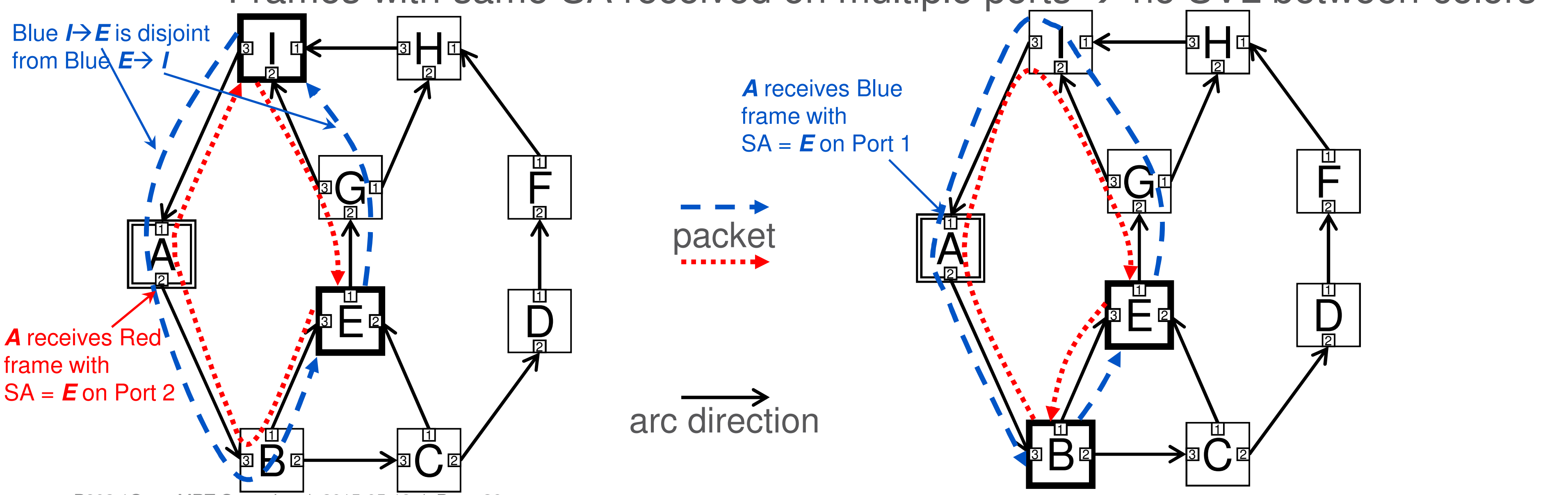
- MRTs are distinguished by VLAN IDs
- Non-learning VLAN (like SPBM)
  - One VID is enough for all the SPTs supporting a VLAN
  - Similarly, one VID is enough for all the MRT-Blue trees supporting a VLAN
  - And, one VID is enough for all the MRT-Red trees supporting a VLAN
- Learning VLAN (like SPBV)
  - Each bridge has its own VID for its own SPT in support of a VLAN
  - Similarly, each bridge has its own Blue VID for its own MRT-Blue in support of a VLAN
  - And, each bridge has its own Red VID for its own MRT-Red in support of a VLAN

Non-learning VLAN

- There is no MAC address learning from data frames
- All FDB entries are installed by the control plane both for data frames and OAM frames
- OAM frames are delivered as data frames
- MAs defined for SPBM can be used as specified in 27.18
  - SPBM VID MA
  - SPBM path MAs and
  - SPBM group Mas
  - The use of PBB-TE MIP TLV (21.7.5) is also specified

Learning VLAN – next slide too

- SPTs
  - Shared learning (SVL) among SPVIDs supporting a given VLAN
- MRTs
  - Independent learning (IVL) should be applied for MRT VIDs (Comment #56)
  - Otherwise MAC learning can be confused due to disjoint paths
    - The same color paths may differ in forward and reverse directions → no SVL within a color
    - Frames with same SA received on multiple ports → no SVL between colors



Learning VLAN – Cont'd

- Independent learning (IVL) for MRT VIDs → always broadcast unknown on MRTs within the VLAN
- If no traffic is going on MRTs during normal operation, then it is broadcast unknown after the failure event anyways
- OAM
  - VLAN MAs similarly to an E-TREE
  - CCM, LBM, LTM can be sent on MRTs
  - LBR and LTR have to be sent back with the VID of the other color MRT if the nodes are not unordered, which is known in advance for each node pair
  - PBB-TE MIP TLV (21.7.5) can also be used to specify reverse VID

MRT Protection Modes

- MRTs can protect SPT
  - Different protection modes can be applied, e.g. 2:1, 2+1 (e.g. in Seamless Redundancy fashion)

- MRTs can protect each other
  - Different protection modes can be applied, e.g. 1:1, 1+1 (e.g. Seamless Redundancy)

(Comment #82)

Failure Handling Modes

- The traffic is **NEVER** directed back from the protection tree to the normal tree**!!!**
- Seamless Redundancy, see above
- Point of Local Repairs (PLRs) handle the failure
  - Similar to IP FRR
  - The bridge detecting the failure redirects traffic by changing the VID
    - To one of the MRTs if they protect SPTs
    - To the other MRT if MRTs protect each other
- Edge Bridges handle the failure
  - CCMs monitor the trees (SPTs and MRTs)
  - Edge Bridges detecting the failure redirect traffic by changing the VID
    - To one of the MRTs if they protect SPTs
    - To the other MRT if MRTs protect each other
- No out-of-order frame delivery due to failure handling

(Comment #84)

Redirection by Point of Local Repair (PLR)

- It is known in advance whether MRT-Blue or MRT-Red is to be used for handling a particular failure
- Non-learning VLAN
  - Similar to IP FRR
  - FDB entry belonging to tree of normal operation (i.e. SPT or an MRT) is replaced with FDB entry belonging the protection tree (i.e. with MRT-Blue or MRT-Red) upon a failure event
- Learning VLAN
  - VID translation at the Ingress Port from the VID of normal operation (i.e. SPT or an MRT) to the VID of the protection tree (i.e. to MRT-Blue or MRT-Red) upon a failure event

OAM

- It is known in advance which paths are pairs
  - Ordered nodes: the paths with different color are pairs
    - Red CCM missing from an ordered node → Blue forward path is down or vice versa
  - Unordered nodes: the paths with the same color are pairs
    - Blue CCM missing from an unordered node → Blue forward path is down if the paths are symmetric
    - It is also know in advance whether or not the paths are symmetric
- RDI flag of CCMs is for handling asymmetry
- RDI flag has to be used if forward and reverse paths are asymmetric between unordered nodes

Restoration

- At each bridge the sequence specified by Section 12.2 of [1]) must be followed:
  1. Receive failure notification
  2. Recompute SPT
  3. Install new SPT
  4. If the network was stable before the failure occurred, wait a configured (or advertised) period for all routers to be using their SPTs and traffic to drain from the MRTs.
  5. Recompute MRTs
  6. Install new MRTs.
- The restoration should not be initiated if there is any reason to suspect that the bridges do not share a common view on the network topology. No mandatory method for that, several options:
  - The safest way is to use the Agreement Protocol for MRTs
  - Wait some time after the reception of the last LSP indicating a topology change
  - Using Controlled Convergence sub-TLV of draft-ietf-isis-mrt
  - Using OAM
  - Further options listed in http://www.ieee802.org/1/files/public/docs2013/ca-farkas-path-status-notification-0513-v01.pdf
- Probability of out-of-order delivery is similar to that of IS-IS restoration

(Comment #84)

Loop Avoidance

- SPTs
  - 6.5.4.1 Loop prevention
  - 6.5.4.2 Loop mitigation (ingress check, RPFC)
- MRTs
  - Seamless Redundancy and Edge Bridge based protection
    - Just like SPTs
  - PLR based operation
    - Loop mitigation cannot be applied
    - Loop prevention has to be used, unless it is ensured that MRTs are updated with no traffic on them when the network is stable
- Comment #83

Source Specific Multicast

- All Pairs Shortest Path computation is used for SPTs with multiple multicast sources
- Similarly, MRTs of other bridges have to be computed for VLANs having multiple multicast sources
- Comment #78