The Great Debate: TRILL Versus 802.1aq (SBP)

NANOG 50

October 4, 2010

What is this all about?

- Enlarging / extending Layer 2 Ethernet networks
- Getting multipath and redundancy in a way that is better than classic STP
- Two prevalent solutions have been emerging over the years, and are becoming viable
- TRILL: IETF
  - TRansparent Interconnection of Lots of Links
- 802.1aq (SBP): IEEE
  - Shortest Path Bridging

Introductions

- Donald Eastlake
  - Stellar Switches & IETF TRILL Working Group
- Peter Ashwood-Smith
  - Huawei
- Srikanth Keesara
  - Avaya
- Paul Unbehagen
  - Alcatel-Lucent

Intro to 802.1aq / SPB

Shortest Path Bridging IEEE 802.1aq Trill Debate

NANOG 50 Oct 4th 2010

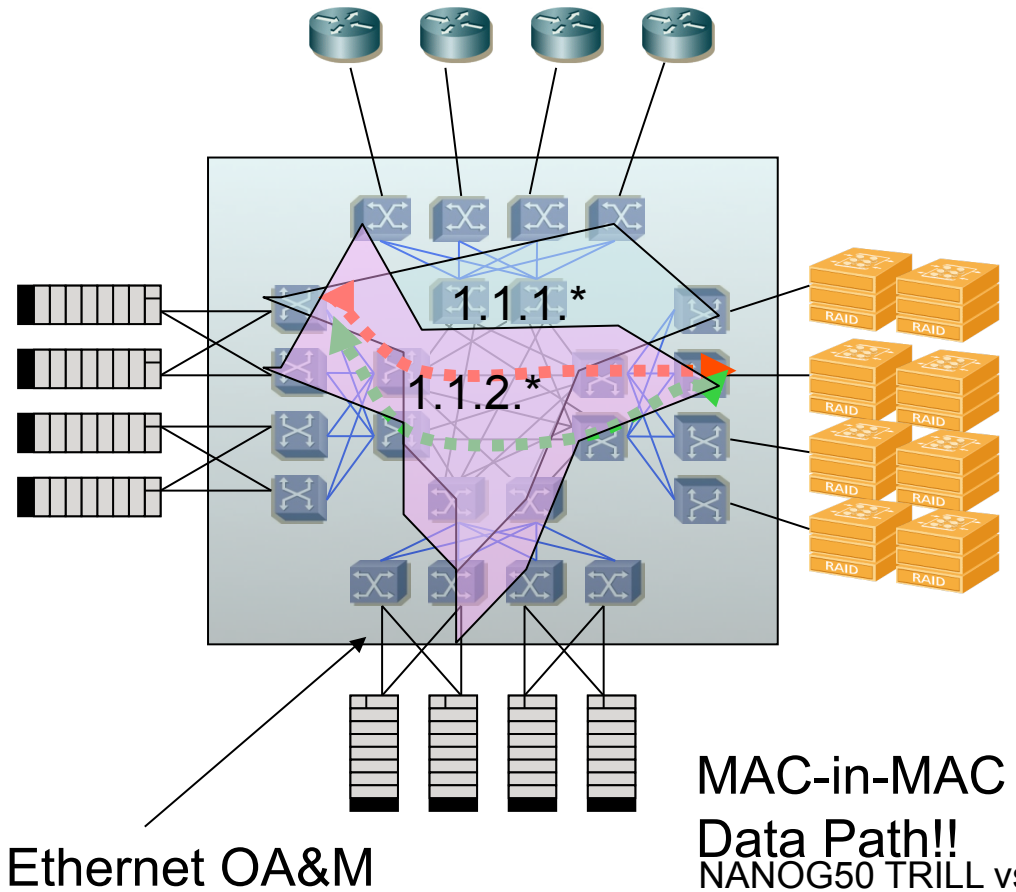
Peter Ashwood-Smith

peter.ashwoodsmith@huawei.com

802.1aq Shortest Path Bridging

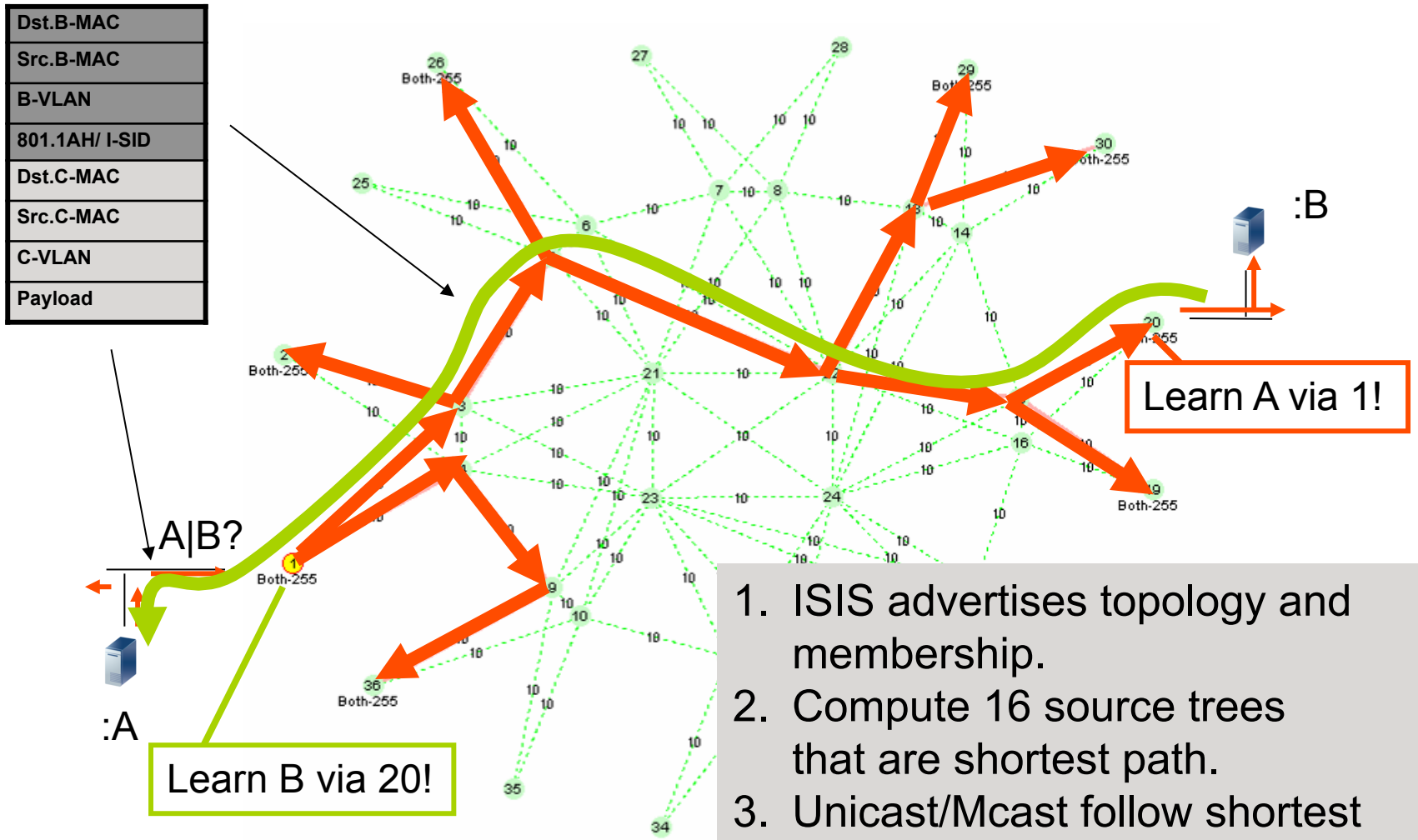
- Industry Standard, widely deployed, vendor supported, test tool supported, **Ethernet data planes** 802.1ah/ad.
- Industry Standard, widely deployed **Ethernet OA&M** 802.1ag.
- Industry Standard, widely deployed IS-IS **link state** protocol with only minor TLV extensions.
- New calculations that produce multiple **shortest** equal cost paths for both unicast **and multicast** traffic L2 VPNs.
- Supports **10's of thousands** of services with 802.1ah **I-SID** on the data path.
- Building on 10's of **thousands of man years** of engineering effort.
- Applications include large L2 in the Data Center in support of Virtualization, Internet L2 exchanges, Metro Ethernet, Wireless backhaul .. Anywhere L2VPN is important.

Supports Wider scope Virtualization and better routing in DC



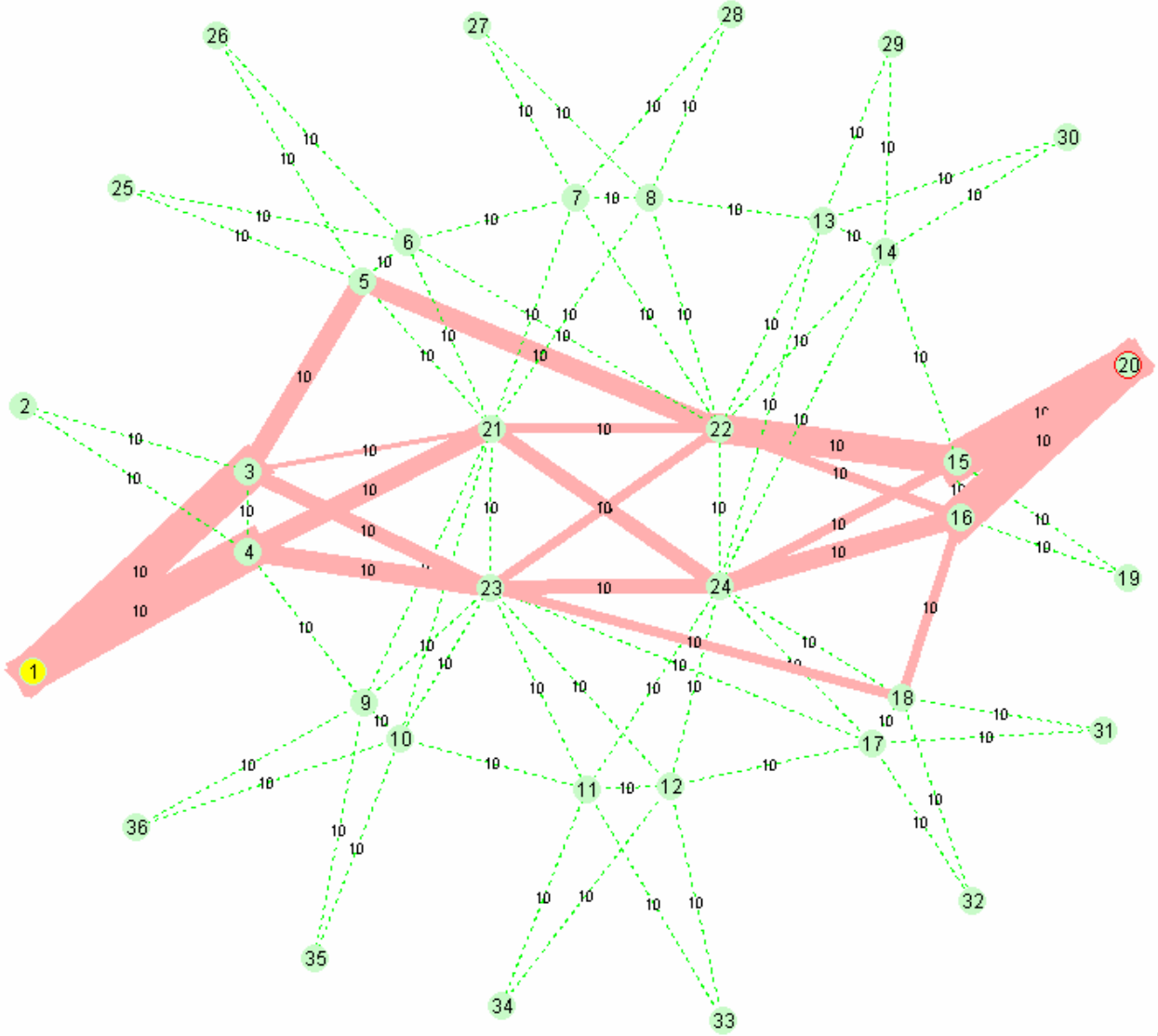
- Support 100s – 1000s of multi Terra bit switches.
- Supports non blocking "Fat Tree" connectivity for 100's of Tera fabrics.
- Compatible with all 802.1 Data Center Protocols.
- Subnet virtualization anywhere with Single point add/remove.

802.1aq - Visually



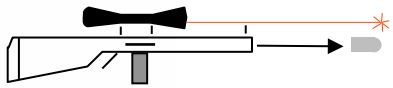
1. ISIS advertises topology and membership.
2. Compute 16 source trees that are shortest path.
3. Unicast/Mcast follow shortest paths. Edge learning only.
4. Mac in Mac data path.

802.1aq - Visually



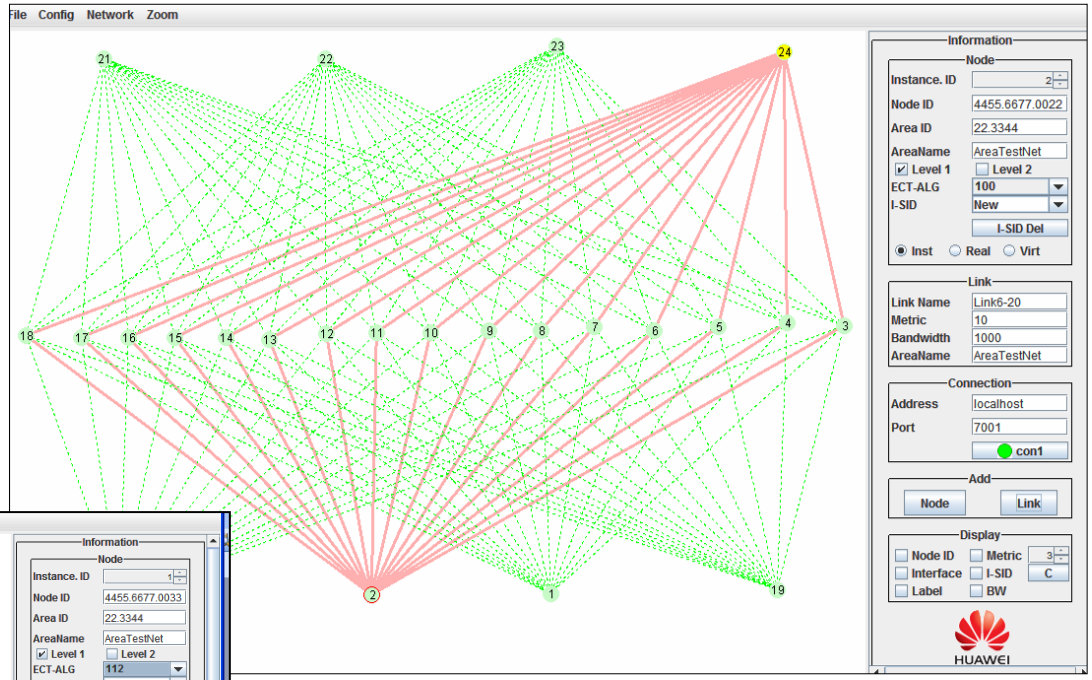
16 x ECMP available at head end

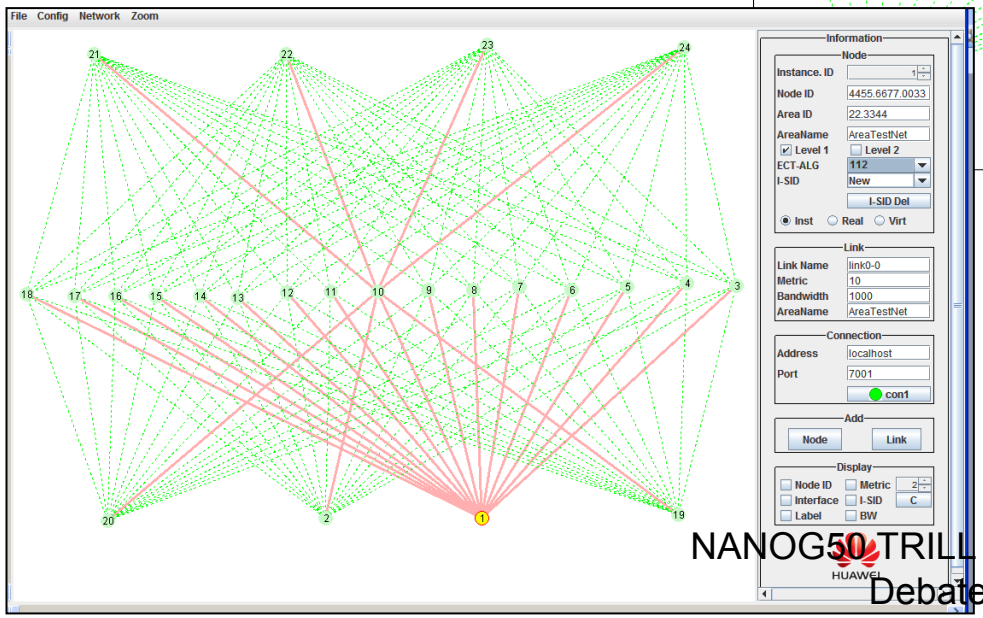
Rifle



802.1aq ECMP in DC Fabric

Can get perfect balance down spine of a two layer 16 ECT L2 Fabric. Shown Are all 16 SPF's from 2<->24

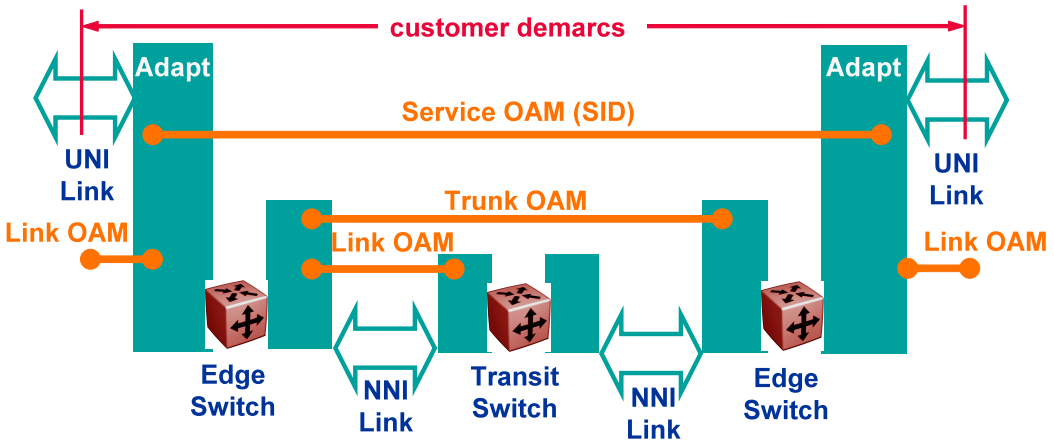




16 different SPF trees Each use different spine as replication point. Shown is one of the 16 SPF's from/to node 1.

802.1aq OAM capabilities = 802.1ag!!!

1. **Continuity Check (CC)**
   a) Multicast/unidirectional heartbeat
   b) Usage: Fault detection
2. **Loopback – Connectivity Check**
   a) Unicast bi-directional request/response
   b) Usage: Fault verification
3. **Traceroute (i.e., Link trace)**
   a) Trace nodes in path to a specified target node
   b) Usage: Fault Isolation



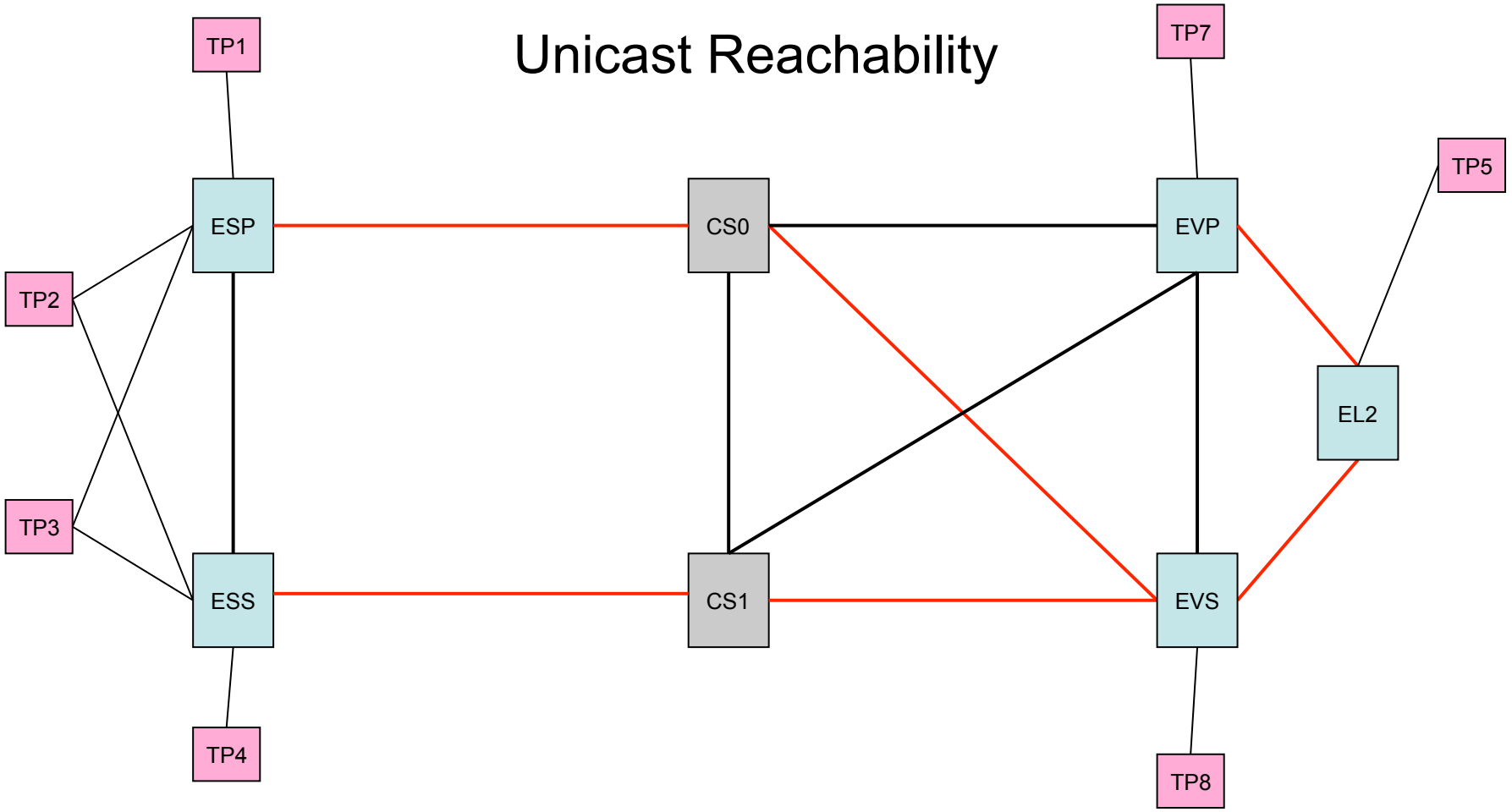
4. **Discovery** (not specifically supported by .1ag however Y.1731 and 802.1ab support it)
   a) Service (e.g. discover all nodes supporting common service instance)
   b) Network (e.g. discover all devices common to a domain)
5. **Performance Monitoring** (MEF10 and 12 - Y.1731 for pt-pt now extending to pt-mpt and mpt-mpt)
   a) Frame Delay, Frame Loss, Frame Delay Variation (derived)
   b) Usage: Capacity planning, SLA reporting

SPB Deployment Experience and OAM

Srikanth Keesara

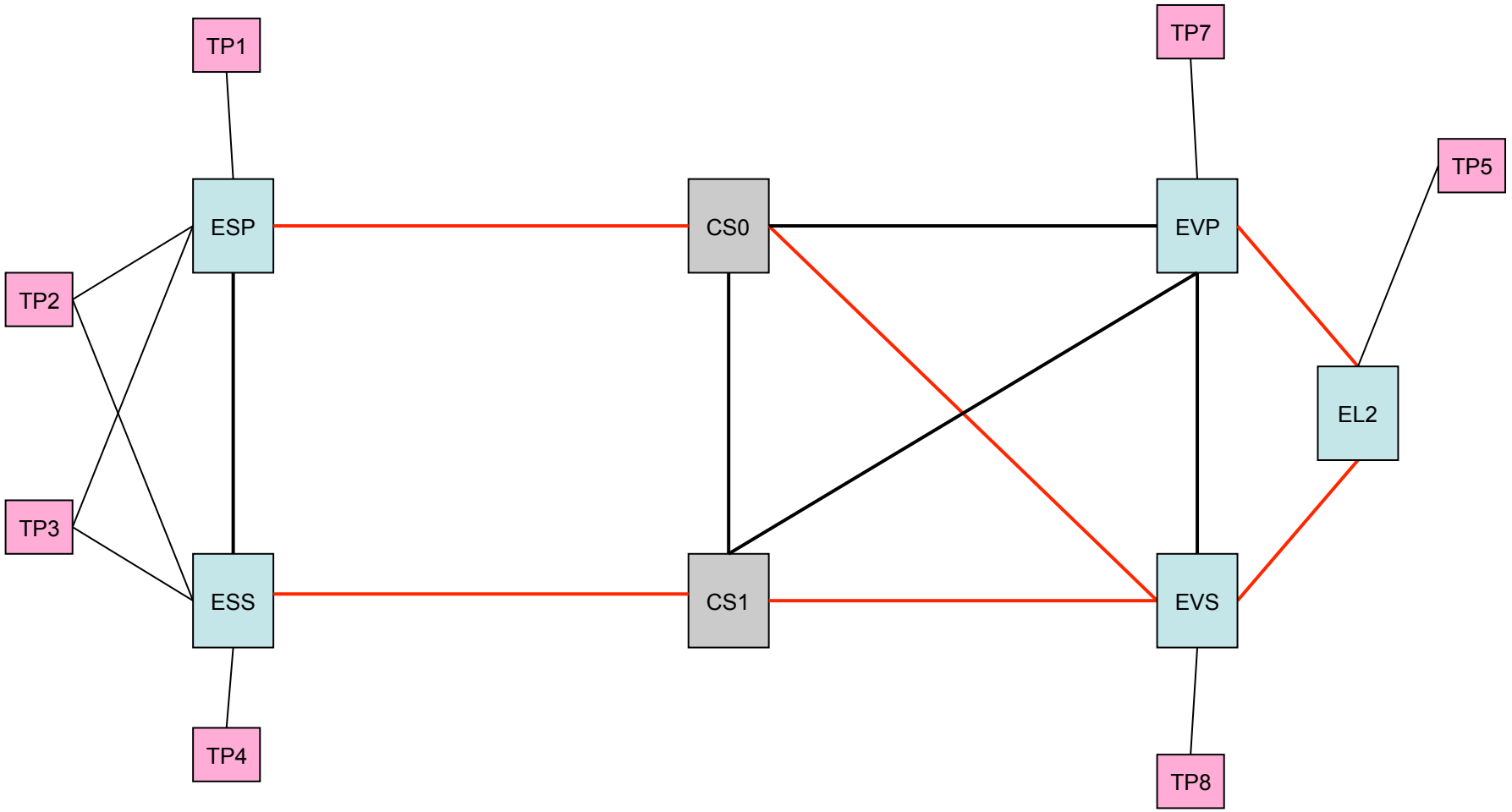
(skeesara@avaya.com)

Unicast Reachability



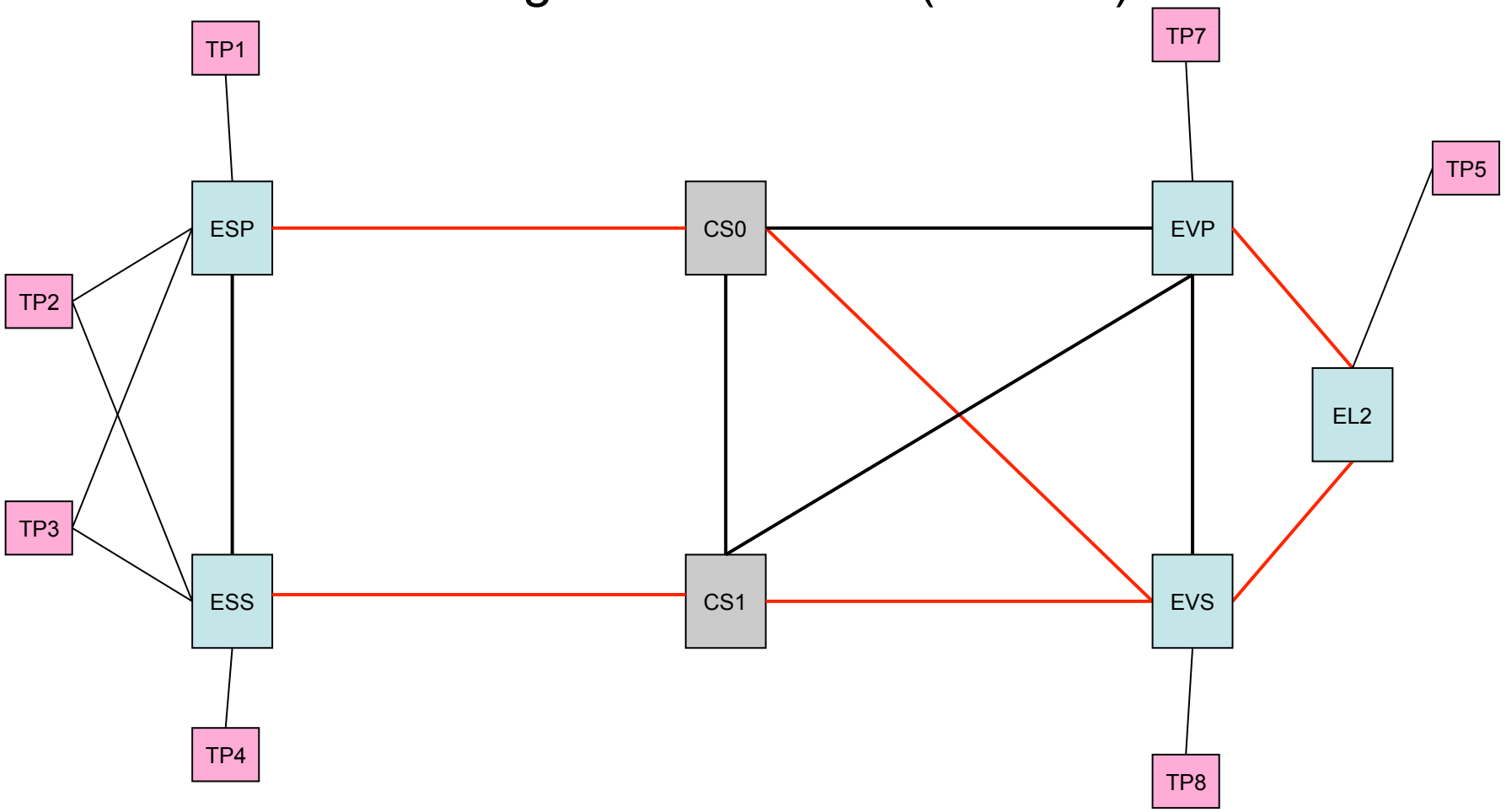

```
el2:6# l2ping 10.esp
----00:09:97:f7:9b:df      L2 PING Statistics----  0(68) bytes of data
1 packets transmitted, 1 packets received,   0.00% packet loss
  round-trip (us) min/max/ave/stdv =  772/772/772.00/ 0.00

el2:6# l2ping 10.esp burst-count 10
----00:09:97:f7:9b:df      L2 PING Statistics----  0(68) bytes of data
10 packets transmitted, 10 packets received,   0.00% packet loss
  round-trip (us) min/max/ave/stdv =  493/778/555.30/ 85.96
```

Tracing Unicast Path



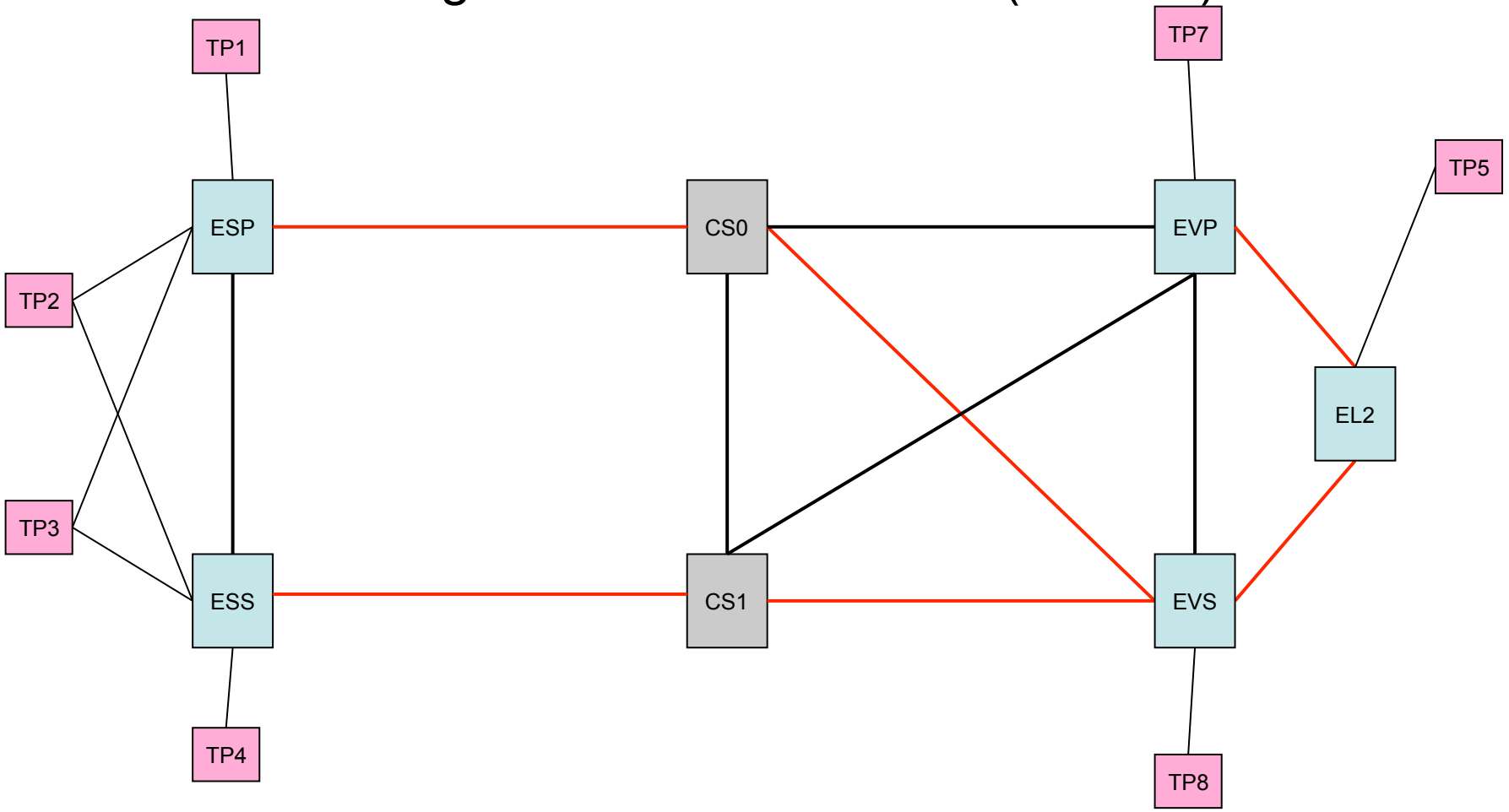

```
el2:6#l2traceroute 10.esp
 0    el2              (00:0c:f8:03:83:df)
 1    evs              (00:13:0a:e6:43:df)
 2    cs0              (00:04:dc:6c:03:df)
 3    esp              (00:09:97:f7:9b:df)
```

Tracing Multicast Tree (Service)



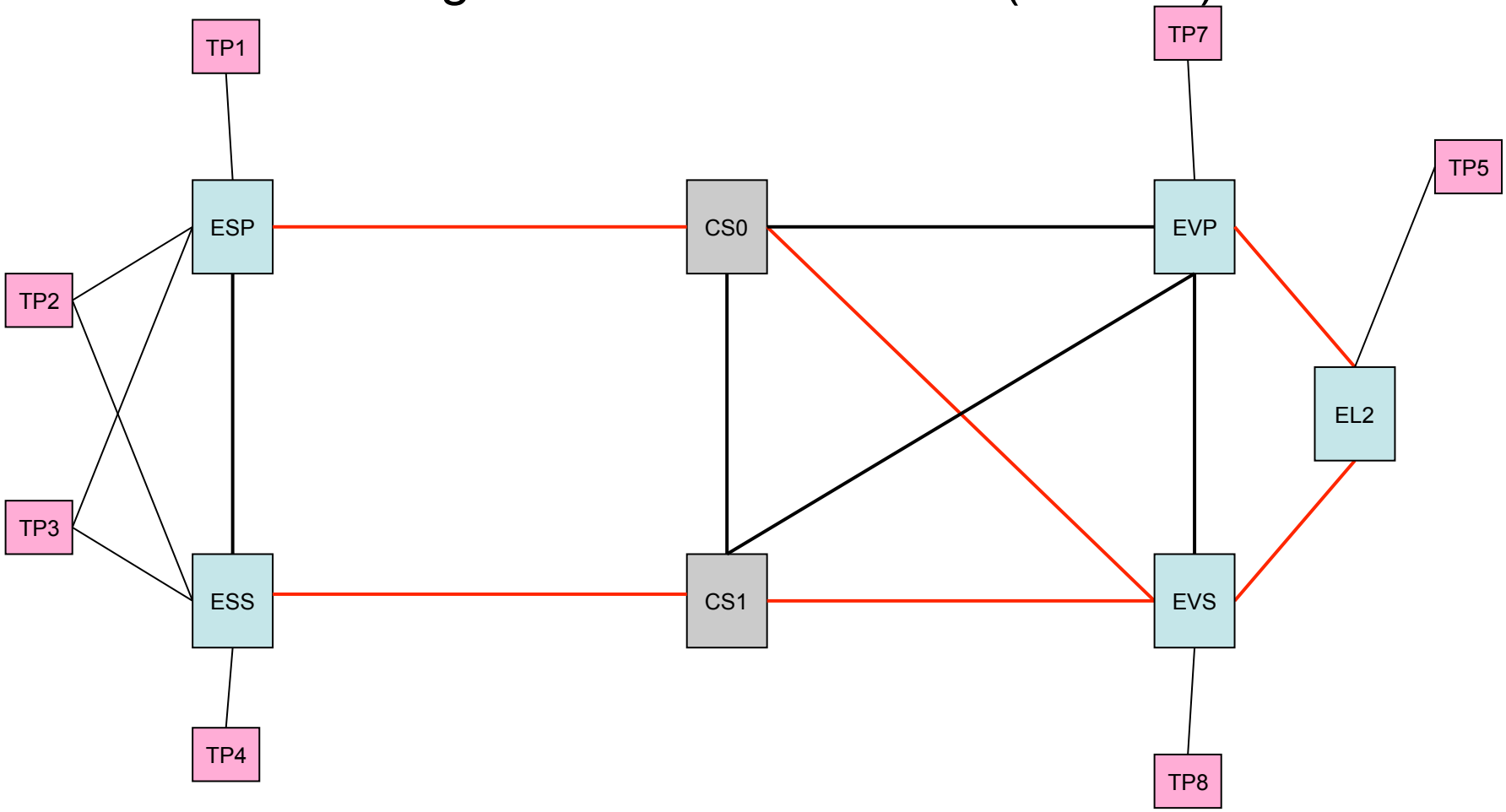

```
el2:6# l2tracetree 10.5010
l2tracetree to 13:00:01:00:13:92, vlan 10 i-sid 5010 nickname 1.00.01 hops 64
1   el2              00:0c:f8:03:83:df -> evp              00:13:0a:e6:73:df
1   el2              00:0c:f8:03:83:df -> evs              00:13:0a:e6:43:df
2   evs              00:13:0a:e6:43:df -> cs0              00:04:dc:6c:03:df
2   evs              00:13:0a:e6:43:df -> cs1              00:e0:7b:bd:a3:df
3   cs0              00:04:dc:6c:03:df -> esp              00:09:97:f7:9b:df
3   cs1              00:e0:7b:bd:a3:df -> ess              00:15:e8:f0:53:df
```

Tracing Partial Multicast Tree (Service)



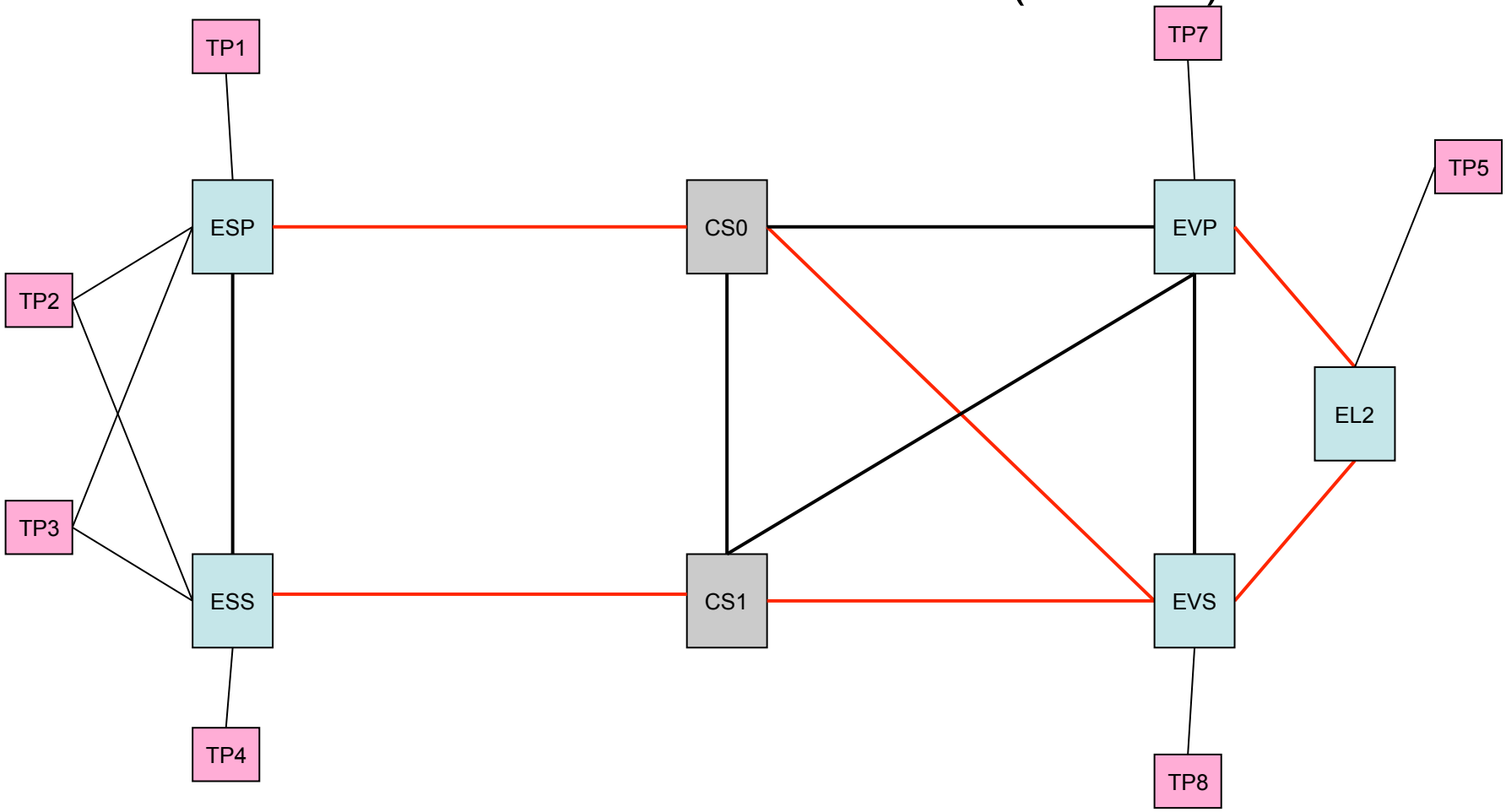

```
evs:6# l2tracetree 10.5010.el2
l2tracetree to 13:00:01:00:13:92, vlan 10 i-sid 5010 nickname 1.00.01 hops 64
1   evs                 00:13:0a:e6:43:df -> cs0                 00:04:dc:6c:03:df
1   evs                 00:13:0a:e6:43:df -> cs1                 00:e0:7b:bd:a3:df
2   cs0                 00:04:dc:6c:03:df -> esp                 00:09:97:f7:9b:df
2   cs1                 00:e0:7b:bd:a3:df -> ess                 00:15:e8:f0:53:df
```

Tracing Partial Multicast Tree (Service)




```
cs0:6# l2tracetree 10.5010.el2
l2tracetree to 13:00:01:00:13:92, vlan 10 i-sid 5010 nickname 1.00.01 hops 64
1   cs0              00:04:dc:6c:03:df -> esp              00:09:97:f7:9b:df
```

Multicast Tree Trace on a leaf (Service)




```
esp:6# l2tracetree 10.5010.el2
l2tracetree to 13:00:01:00:13:92, vlan 10 i-sid 5010 nickname 1.00.01 hops 64
  Leaf node for the tree
```

SPB - Current Status

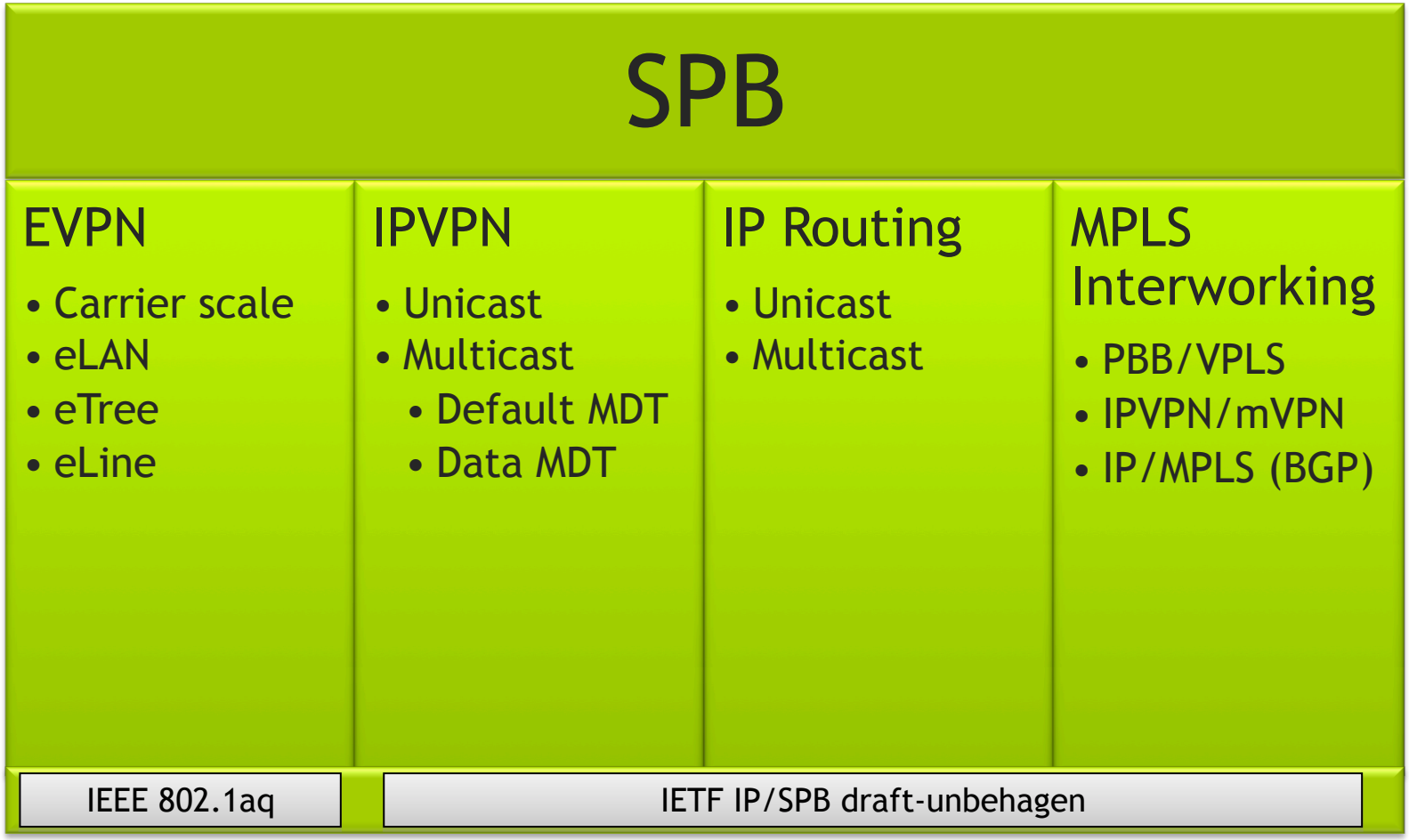
- Live in several networks
- Current Production Networks up to 80 nodes.
- Live Topologies –
  - Mesh
  - Ring
  - Hierarchical Ring
- Enterprise and Carrier
- Business services as well as residential aggregation
- Access networks include – STP based as well as vendor proprietary.

Shortest Path Bridging

Paul Unbehagen

NANOG | October 2010

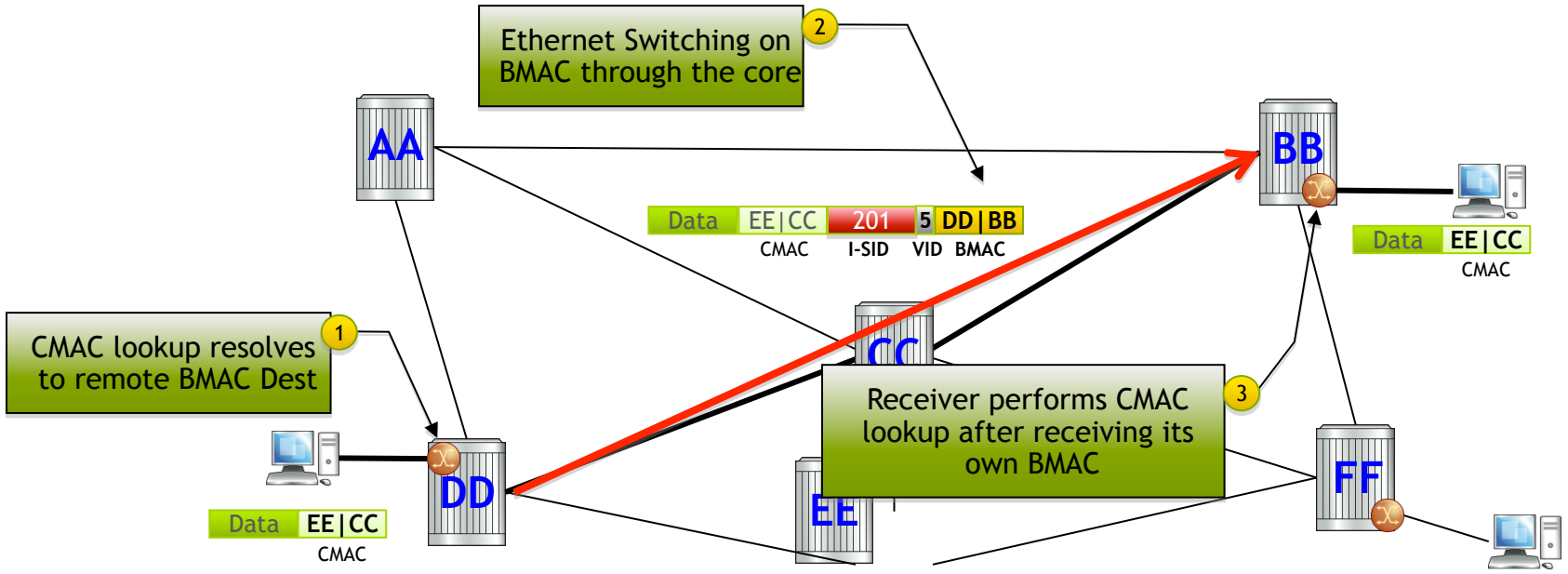
Applications of Shortest Path Bridging (802.1aq) and IP/SPB



Operational Simplicity, single end-point provisioning, very easy to trouble shoot.

Ethernet LANs with SPB-M

- ISID and/or VLAN create large to medium Virtual LANs (16mil ISID's or 4K[4K] VLANs)
- As soon as a Node is configured with SPB, IS-IS will automatically create a SPF unicast FIB from/to each node in the domain based on the nodal MAC, derived from the Sys-ID.
- ISIDs are defined on Access interfaces are then announced in the Link State Update.
- Multicast FIB state is only calculated when a new ISID is configured.
- All forwarding within the SPB-M domain is performed on the BMAC, enabling large scale deployments



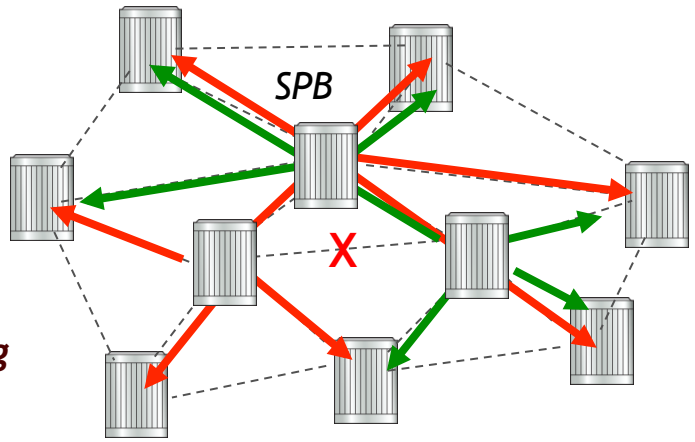
ELAN Service VSI

Default Tree == Network Wide Fast Convergence

Each node joins a control ISID that is used to notify every node know of a link failure at the same time

Upon any link failure IS-IS would use the default ISID to tell everyone reachable of the failure by putting their LSP (LSA) on the tree

- All other nodes receive the update on the multicast tree and can converge at nearly the same time.
- This adds to, but does not replace the standard hop by hop spread of LSP updates.



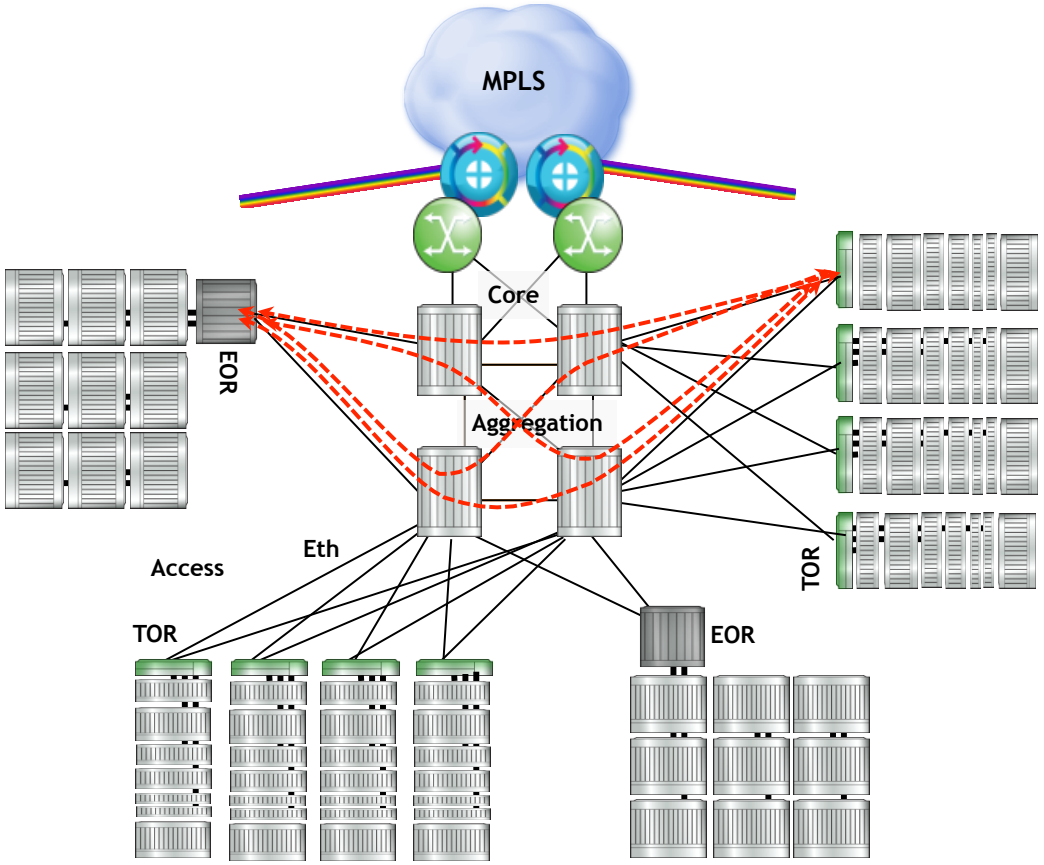
Acts like a bearer channel

Automatically created in the background

Upon any failure becomes a fast delivery of LSP flooding

Use in a Datacenter Architecture

Link State Bridging & L2MP



New Ethernet control plane needed

- Exponential jump from STP
- Link State control of topology

Aware of the full topology

Service awareness

MPLS like control of native Ethernet

Broadcast containment

- Protect core from VM MAC scaling
- Optimized Multicast Algorithm

Easy Subnet management

Equal Cost Path Forwarding

Operationally Simpler

- Simple endpoint provisioning

Introduction to TRILL

The IETF TRILL Standard

Donald E. Eastlake 3rd

Co-Chair, IETF TRILL Working Group

d3e3e3@gmail.com, +1-508-333-2270

WHAT/WHY/WHO TRILL?

- What is TRILL?
  - TRILL is a new standard protocol to perform Layer 2 bridging using IS-IS link state routing.
- Who invented TRILL?
  - Radia Perlman of Intel, the inventor of the Spanning Tree Protocol, a major contributor to link-state routing, and the inventor of DECnet Phase V from which IS-IS was copied.

WHAT/WHY/WHO TRILL?

- TRILL –
  TRansparent Interconnection of Lots of Links
  - A standard specified by the IETF (Internet Engineering Task Force) TRILL Working Group co-chaired by
    - Donald E. Eastlake 3rd
    - Erik Nordmark, Oracle
- RBridge – Routing Bridge
  - A device which implements TRILL
- RBridge Campus –
  - A network of RBridges, links, and any intervening bridges, bounded by end stations / layer 3 routers.

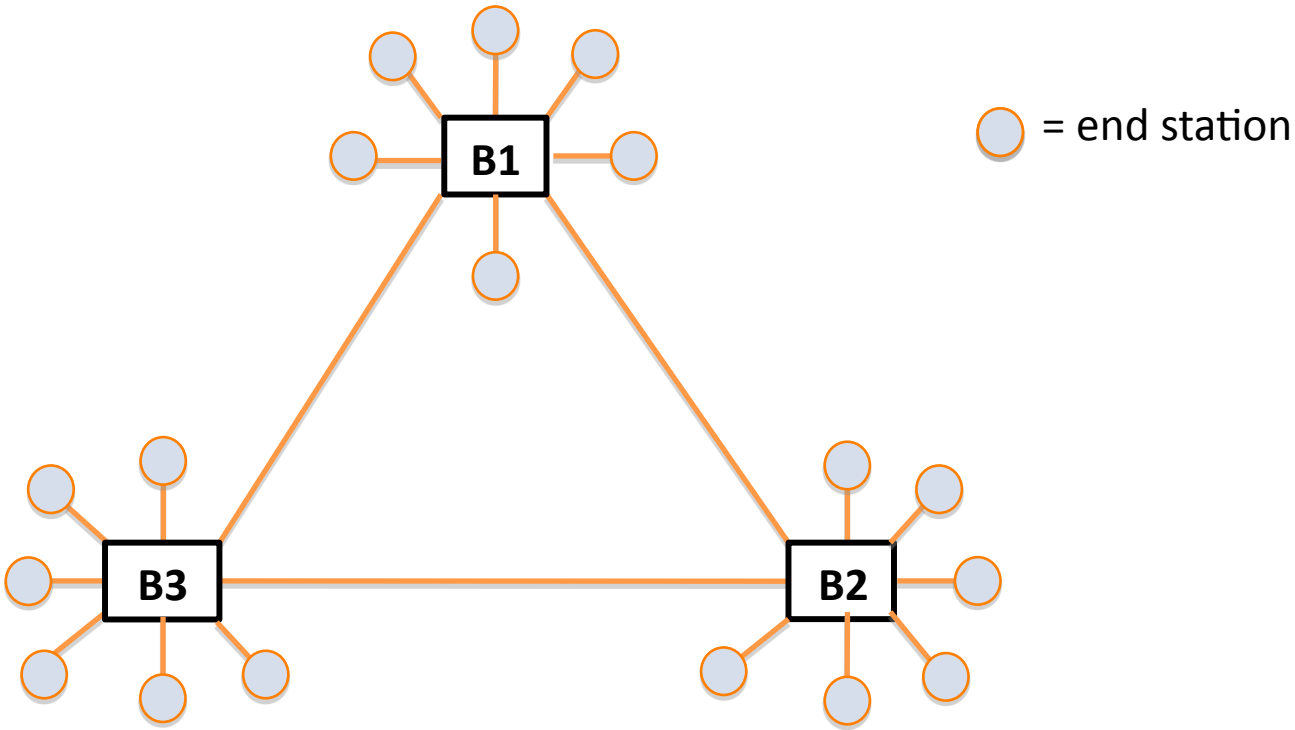
WHAT/WHY/WHO TRILL?

- Basically a simple idea:
  - Encapsulate native frames in a transport header providing a hop count.
  - Route the encapsulated frames using IS-IS.
  - Decapsulate native frames before delivery.

Why IS-IS For TRILL?

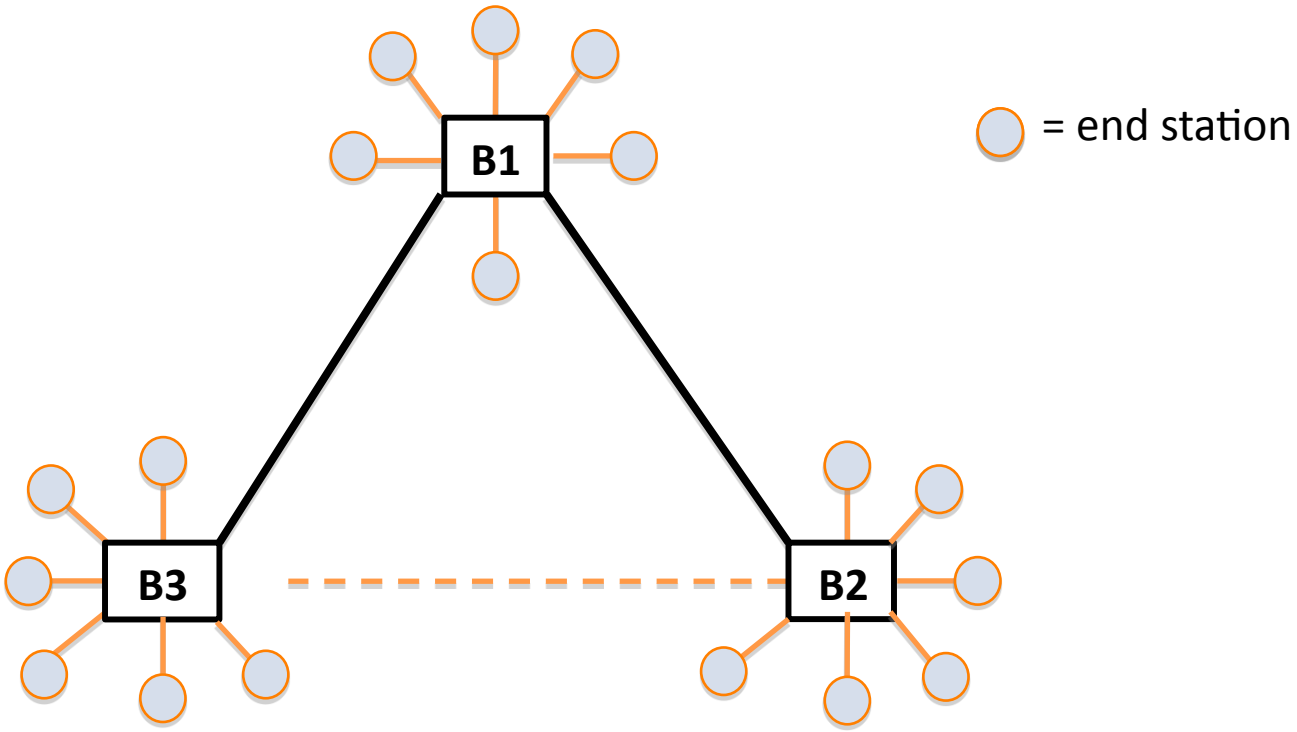
- The IS-IS (Intermediate System to Intermediate System) link state routing protocol was chosen for TRILL over OSPF (Open Shortest Path First), the only other plausible candidate, for the following reasons:
  - IS-IS runs directly at Layer 2. Thus no IP addresses are needed, as they are for OSPF, and IS-IS can run with zero configuration.
  - IS-IS uses a TLV (type, length, value) encoding which makes it easy to define and carry new types of data.

Optimum Point-to-Point Forwarding



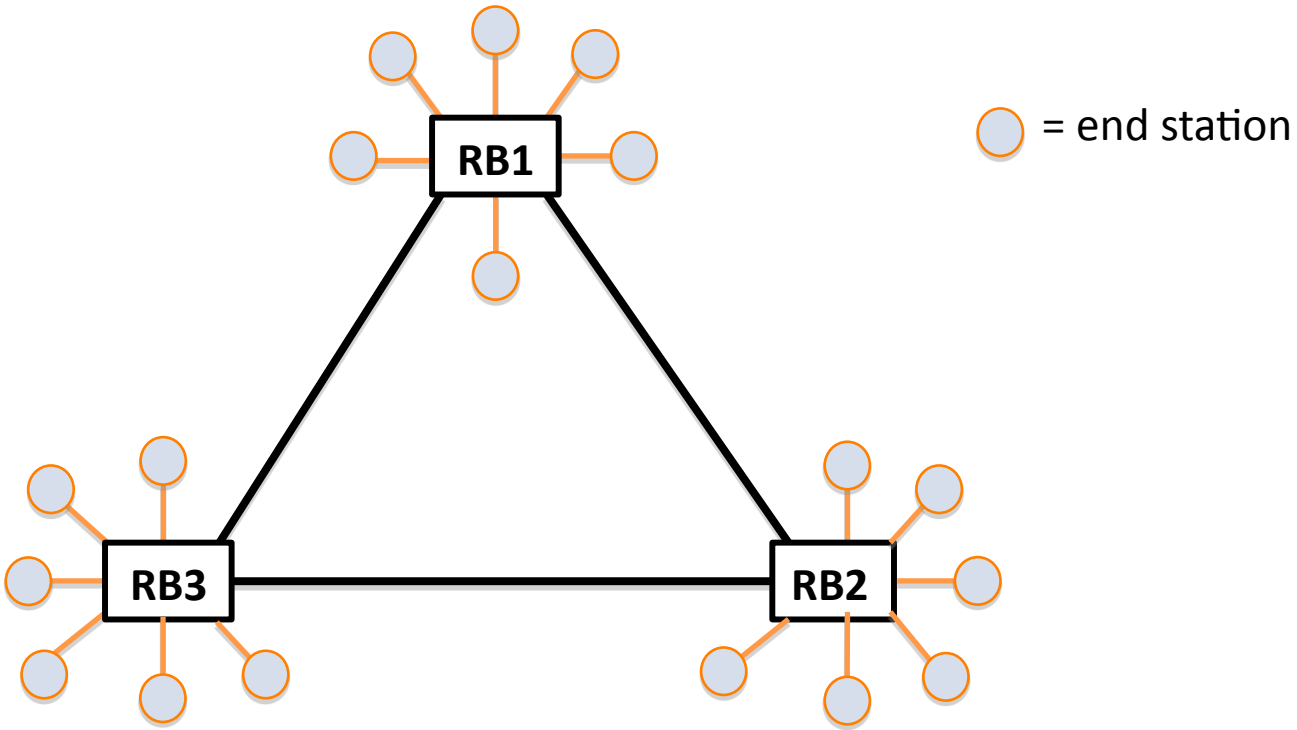
A three bridge network

OPTIMUM POINT-TO-POINT FORWARDING



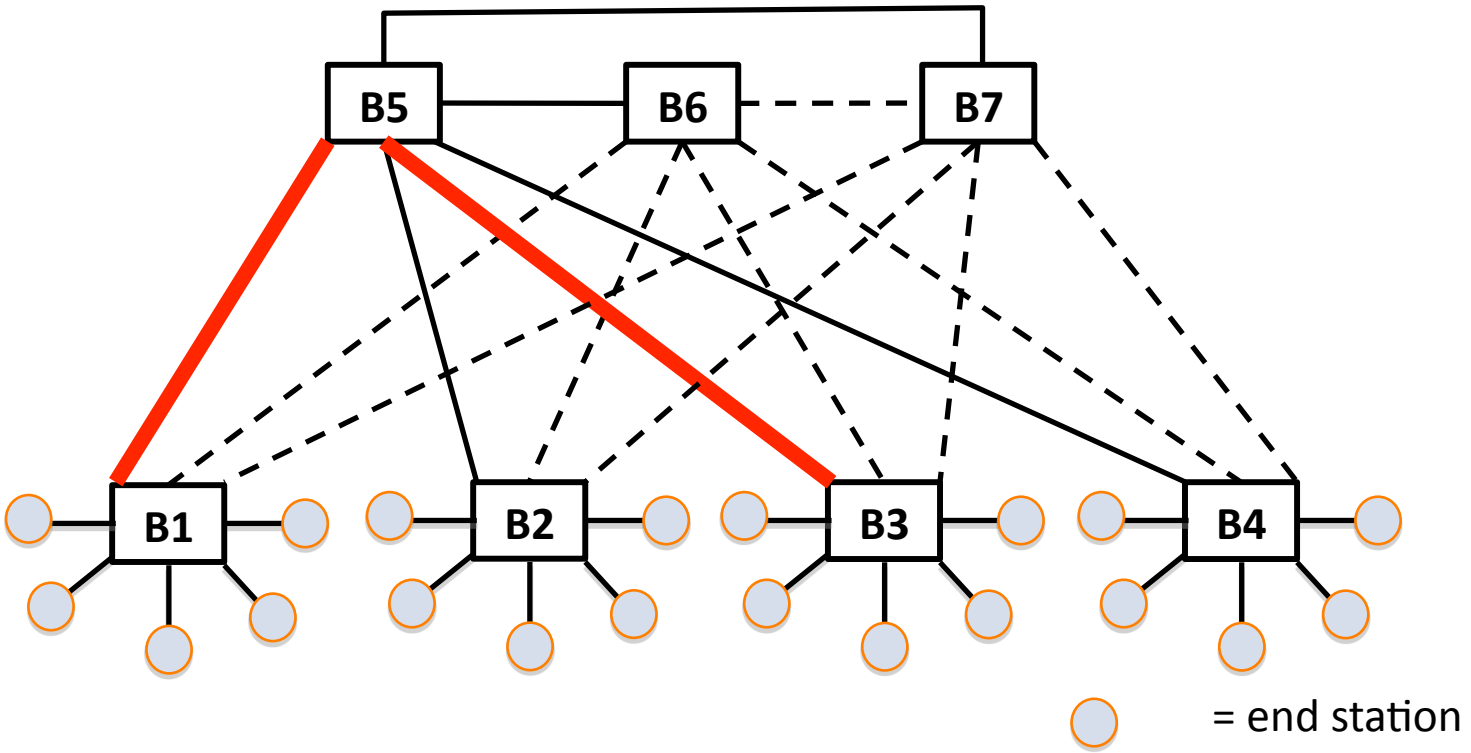
Spanning tree eliminates loops by disabling ports

Optimum Point-to-Point Forwarding



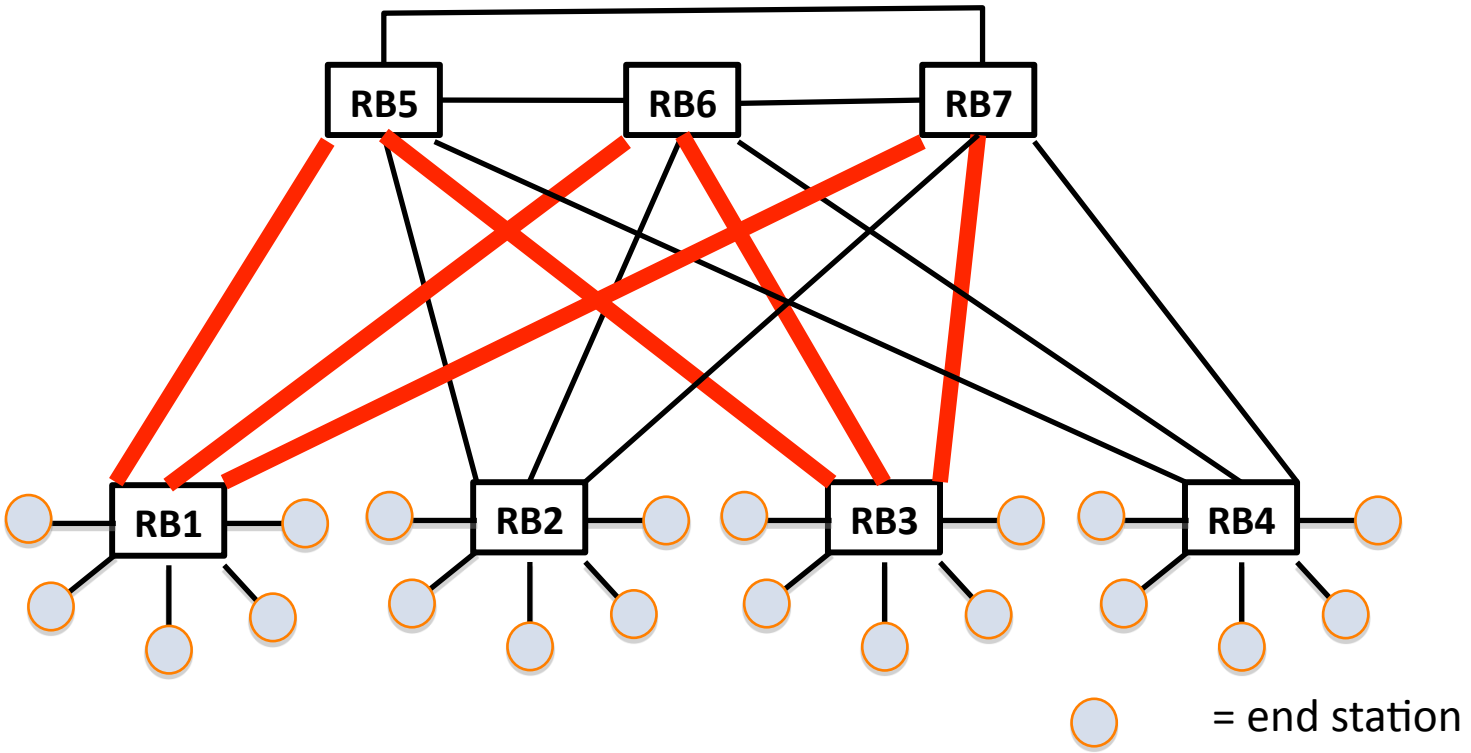
A three RBridge network: better performance using all facilities

MULTI-PATHING



Bridges limit traffic to one path

MULTI-PATHING



RBridges support
multi-path for higher throughput

Some Other TRILL Features

- Compatible with classic bridges. RBridges can be incrementally deployed into a bridged LAN.
- Unicast forwarding tables at transit RBridges scale with the number of RBridges, not the number of end stations. Transit RBridges do not learn end station addresses.
- A flexible options feature. RBridges know what options other RBridges support.
- Globally optimized distribution of IP derived multicast.

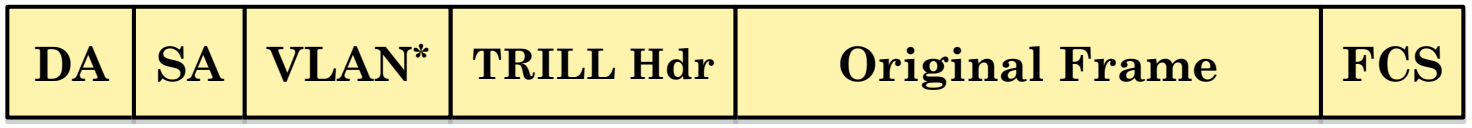
TRILL FEATURES

- Transparency
- Plug & Play
- Virtual LANs
- Frame Priorities
- Data Center Bridging
- Virtualization Support
- Multi-pathing
- Optimal Paths
- Rapid Fail Over
- The safety of a TTL
- Options

Frame Types

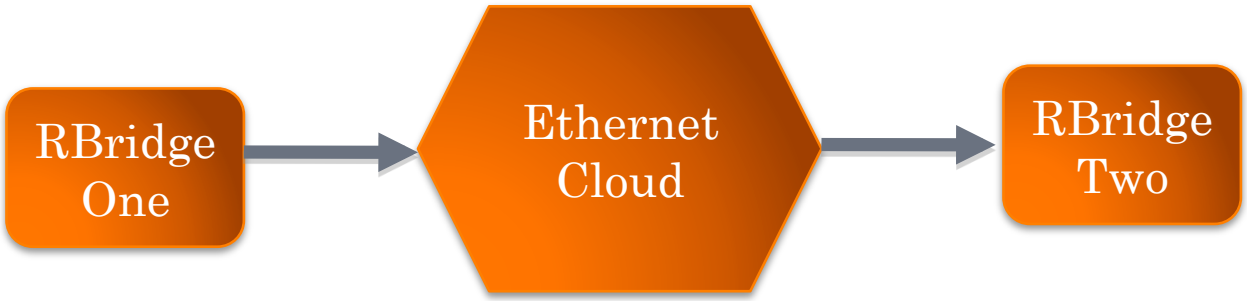
- Frame Type Names Used in TRILL
  - TRILL IS-IS Frames – Used for control between RBridges.
  - TRILL Data Frames – Used for encapsulated native frames.
  - Layer 2 Control Frames – Bridging control, LLDP, MACSEC, etc. Never forwarded by RBridges.
  - Native Frames – All frames that are not TRILL or Layer 2 Control Frames.

TRILL Encapsulation and Header



Link Transport Header

Original Frame with original VLAN



* Link Transport VLAN need only be present if RBridges are connected by a bridged LAN or carrier Ethernet requiring a VLAN tag or the like.

TRILL Encapsulation and Header

- TRILL Data frames between RBridges are encapsulated in a local link header and TRILL Header.
  - The local link header is addressed from the local source RBridge to the next hop RBridge for known unicast frames or to the All-RBridges multicast address for multidestination frames.
  - The TRILL header specifies the first/ingress RBridge and either the last/egress RBridge for known unicast frames or the distribution tree for multidestination frames.

TRILL Encapsulation and Header

- TRILL Header – 8 bytes

TRILL Ethertype	V	R	M	OpLng	Hop
Egress RBridge Nickname	Ingress RBridge Nickname				

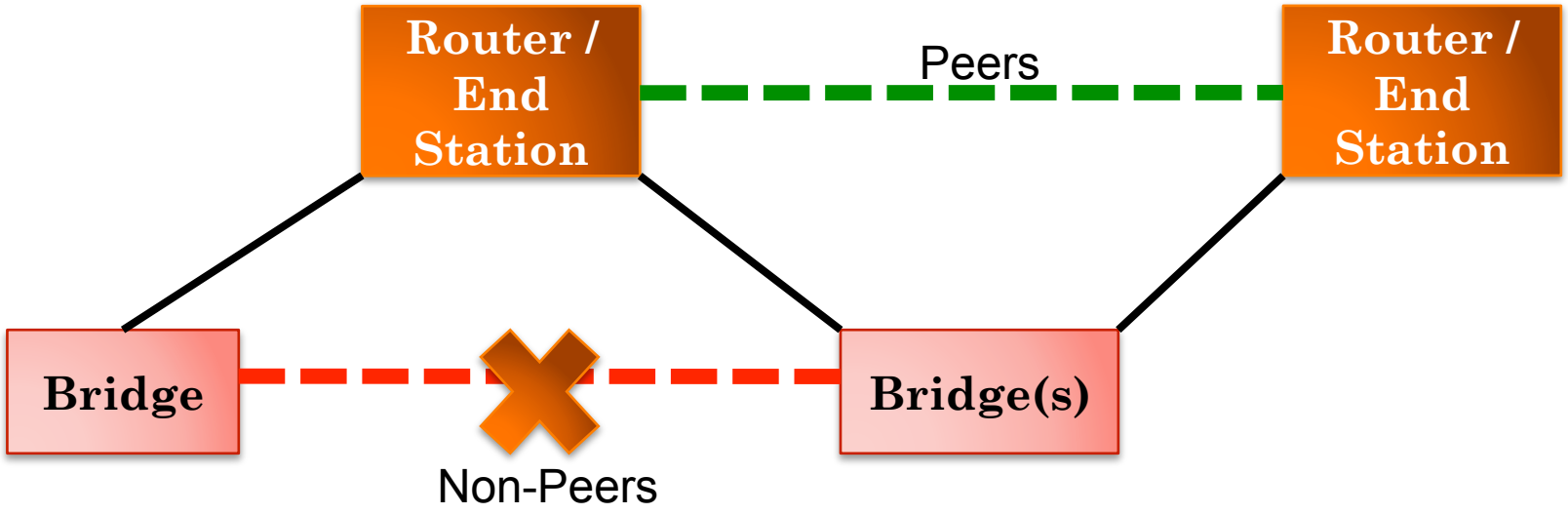
  - Nicknames – auto-configured 16-bit campus local names for RBridges
  - V = Version (2 bits)
  - R = Reserved (2 bits)
  - M = Multi-Destination (1 bit)
  - OpLng = Length of TRILL Options
  - Hop = Hop Limit (6 bits)

How RBridges Work

- TRILL Data frames
  - That have known unicast ultimate destinations are forwarded RBridge hop by RBridge hop to the egress RBridge.
  - That are multi-destination frames are forwarded on a distribution tree selected by the ingress RBridge.
    - For loop safety, a Reverse Path Forwarding Check is performed on multi-destination TRILL Data frames when received at each RBridge.
    - Distribution trees should be pruned based on VLAN and multicast group.
    - Distribution trees are shared campus-wide bi-directional trees. Each tree covers the entire campus and is not limited by VLAN or the like.

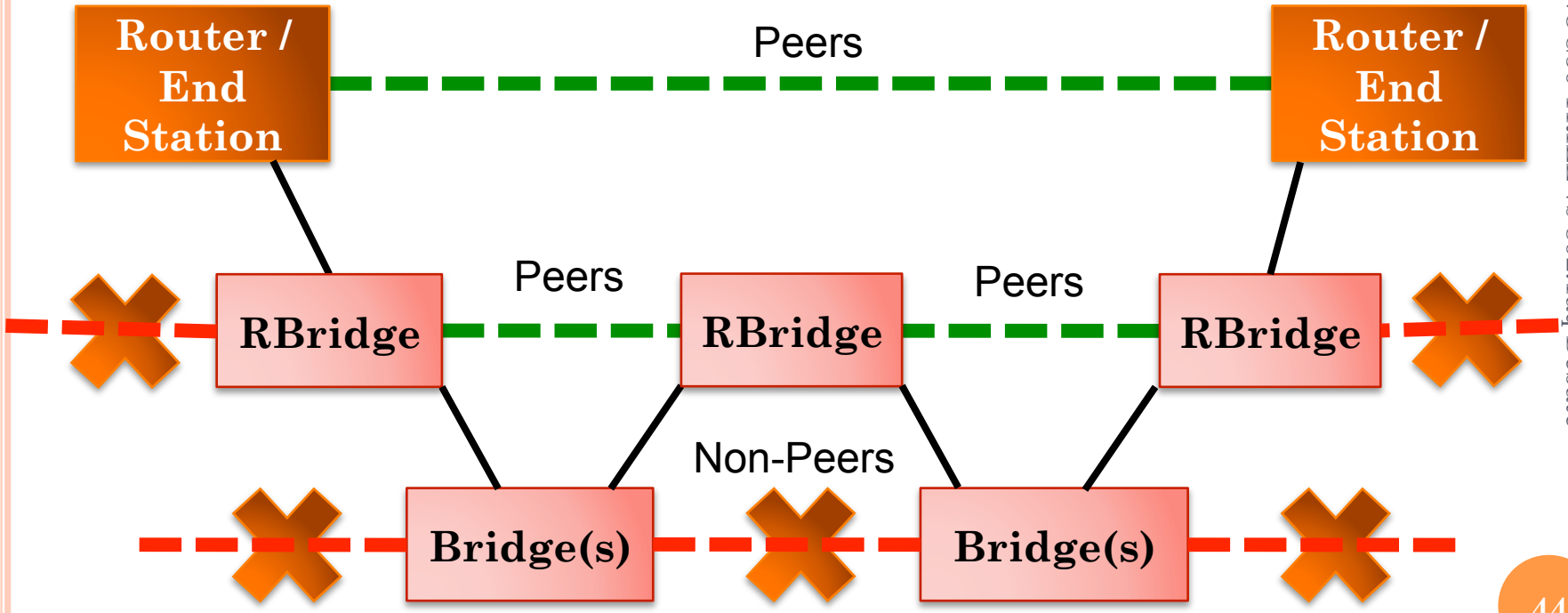
PEERING

- Former Situation



PEERING

- With RBridges



VLANS

- TRILL tries hard to glue together all end stations in a particular VLAN within the campus.
  - In an RBridge campus, any two end stations in the same VLAN that can each reach an RBridge will be able to communicate with each other.
    - I.E., TRILL glues together any VLAN islands
  - Surveys of customers have found this to be generally desirable but there are instances where you want the same VLAN ID in different parts of your RBridge campus to be different or different VLAN IDs in different parts of the campus to be connected. TRILL has an optional feature to do this.

ALGORHYME V2

- I hope that we shall one day see
- A graph more lovely than a tree.
- A graph to boost efficiency
- While still configuration-free.
- A network where RBridges can
- Route packets to their target LAN.
- The paths they find, to our elation,
- Are least cost paths to destination!
- With packet hop counts we now see,
- The network need not be loop-free!
- RBridges work transparently,
- Without a common spanning tree.
- - By Ray Perlner

RBRIDGE SUPPORT OF DATA CENTER BRIDGING

- "Data Center Ethernet"

1. Priority Based Flow Control
   - Per Priority PAUSE
2. Enhanced Transmission Selection
3. Congestion Notification

Data Center Bridging

4. TRILL

STANDARDIZATION STATUS

- Time span of effort: 5 ½ + years
- Earlier organizational meetings in late 2004
- First TRILL WG meeting: March 2005
- Base protocol draft pass up from TRILL Working Group December 2009
- Base protocol approved as a standard by the IETF March 15th 2010

STANDARDIZATION STATUS

- Non-IETF Assignments:
  - TRILL Ethertype: 0x22F3
  - L2-IS-IS Ethertype: 0x22F4
  - Block of Multicast Addresses for TRILL: 01-80-C2-00-00-40 to 01-80-C2-00-00-4F
  - TRILL NLPID: 0xC0
- Final approval of IS-IS code points and data structures pending.

STANDARDIZATION STATUS

- First open interoperability testing (plug fest) was held at the University of New Hampshire Interoperability Laboratory (UNH IOL) 3-5 August 2010:
  - http://www.iol.unh.edu/services/testing/bfc/grouptest/TRILL_plugfest.php
- Second planned for Q1, 2011.

- Some ongoing standards work:
  - RBridge MIB
  - TRILL over PPP
  - RBridge VLAN Mapping
  - RBridge Support of DCB
  - OAM
  - TRILL Header Options

Comparisons

Trill point-of-view

FRAME OVERHEAD

- For point-to-point Ethernet links with multi-pathing:
  - TRILL: 20 bytes
    - + 8 bytes TRILL Header (including Ethertype) + 12 bytes outer MAC addresses
  - SPBM: 22 bytes
    - + 18 bytes 802.1ah tag (including Ethertype) -12 bytes for MAC addresses swallowed by 802.1ah + 4 bytes B-VLAN (including Ethertype) + 12 bytes outer MAC addresses
- For complex multi-access links with multi-pathing:
  - TRILL: 24 bytes (20 + 4 for outer VLAN tag)
  - SPBM: Fails

SPBV VLAN Consumption

- SPBV consumes VLANs at a cubic rate.
- If you have N nodes, want to handle V real VLANs and do K way multipathing, SPBV consumes

  N*V*K

  VLAN IDs.
- So, for 100 nodes handling 100 real VLANs doing 10 way multipathing, you need to find 100,000 distinct VLAN IDs…

Routing Computation

- TRILL
  - For unicast, the usual Dijkstra n*(log n) to calculate shortest paths to other RBridges.
    - Arbitrary multi-pathing available by just keeping track of equal cost paths.
  - For multi-destination, k*n*(log n) to have k distribution trees available.
- SPB
  - Unicast and multi-destination unified.
  - k*n*n*(log n) for k-way multi-pathing. K currently limited to 16.

Evolution of TRILL and SPB

- Radia Perlman, inventor of spanning tree and inventor of IS-IS routing invents the concept of transparent routing.
  - Radia Perlman gives a tutorial at IEEE 802 and the ideas are rejected.
  - Radia Perlman organizes a BoF at IETF and the ideas are accepted.

Evolution of TRILL

1. Radia Perlman's idea is accepted by the IETF and the TRILL WG is formed. Basic idea is shortest path transparent frame routing using IS-IS and encapsulation with a hop count.
2. Basic idea unchanged + improved data plane address learning & VLAN support
3. Basic idea unchanged + improved data plan address learning & VLAN support + MTU robustness
4. To Come: continued additive enhancements with OAM, etc.

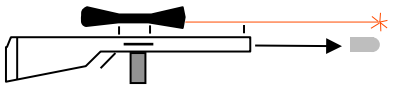
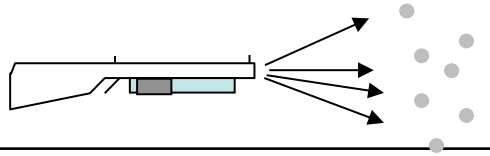
Evolution of SPB

1. Radia Perlman's idea are rejected by IEEE 802.1. They say there isn't a problem, TRILL is a terrible idea, spanning tree is good, routing sucks, and hop counts (TTLs) are evil.
2. Whoops, there is a problem. They start 802.1aq for spanning tree based shortest path bridging. Still say TRILL is terrible, routing sucks, hop counts are evil.
3. Whoops, spanning tree doesn't hack it. They copy a little of using IS-IS and nicknames from TRILL but don't actually do routing. Still say TRILL is a terrible idea and hop counts are evil.
4. Whoops, we can't multipath enough. Try to multipath more. Link agreement protocol etc. is a kludge. Try to find some way to add hop counts to SPB. Still say TRILL is a terrible idea.

Comparisons

802.1aq/SPB point-of-view

The major differences

Aspect	IEEE 802.1aq	TRILL
Encapsulation	Ethernet	New Trill
Equal Cost	16 x head end 	N x transit hash 
Multicast/Bcast	Shortest path	no
ANTI-LOOPING	Reverse path (RPF) unicast & multicast + AP	TTL for unicast RPF multicast
OA&M	Ethernet/ITU	?
Congruence	Yes	no

So the REAL debate

- Introduce a new data plane.
  - All new ASIC's/HW line cards etc**…$$$$$$**
  - All new OA&M.. Highly non trivial exercise.
  - Training costs/testing etc.

- Use Ethernet and modify slightly as required.
  - Use existing Ethernet ASIC especially tandem.
  - Continue building on 30 years of innovation
  - We now have a 24 bit service identifier v.s. 12 bit VLAN… that's incredibly useful lets not step backwards.

OAM Summary

•IEEE-802.1ag approved in 2008
• Mature implementations from several vendors currently in live deployments
• Supported by leading vendors of test equipment
• Same protocol for 802.1Q as for 802.1ah (MIM)
• Proven interoperability.
• **SPB does not require a new OAM protocol**
• **TRILL will have to define a whole new OAM protocol and reinvent years of work that is already standard.**

Debate Topics 1

- Intellectual Property
- Vendor Support
- Frame Header
- Tracking L2 TTL
- Symmetry
- ECMP Methods
- Protocol availability
- Complexity

Debate Topics 2

- Use of IS-IS
- Relationship to classic spanning-tree protocols
- Scale
- Relationship to IP Protocols
- Multi-topology

Q/A

- Why can't the IEEE and IETF work together and finalize one solution
- Any deployment experience yet in a live network?
- Open to audience

References

802.1aq / SPB References

"***IEEE 802.1aq***" : www.wikipedia.org:
http://en.wikipedia.org/wiki/IEEE_802.1aq

http://www.ietf.org/internet-drafts/draft-ietf-isis-ieee-aq-00.txt The IETF IS-IS draft (check for later version 01.. etc).

"***IEEE 802.1aq***" www.ieee802.org/1/802-1aq-d2-6.pdf

"***Shortest Path Bridging*** – Efficient Control of Larger Ethernet Networks" :
upcoming IEEE Communications Magazine – Oct 2010

"***Provider Link State Bridging***" :
IEEE Communications Magazine V46/N9– Sept 2008
http://locuhome.com/wp-content/uploads/2009/02/
ieeecommunicationsmagazinevol46no9sep2008-carrierscaleethernet.pdf

TRILL References (newer)

- Standard: “Rbridges: Base Protocol Specification”
  - http://tools.ietf.org/html/draft-ietf-trill-rbridge-protocol-16
- “Definitions of Managed Objects for RBridge”
  - https://datatracker.ietf.org/doc/draft-ietf-trill-rbridge-mib/
- “RBridges: Campus VLAN and Priority Regions”
  - https://datatracker.ietf.org/doc/draft-ietf-trill-rbridge-vlan-mapping/
- “PPP TRILL Protocol Control Protocol”
  - https://datatracker.ietf.org/doc/draft-ietf-pppext-trill-protocol/

TRILL References (older)

- "TRILL: Problem and Applicability Statement"
  - http://www.ietf.org/rfc/rfc5556.txt
- TRILL WG Charter:
  Current (out of date) and proposed
  - http://www.ietf.org/dyn/wg/charter/trill-charter.html
  - http://www.postel.org/pipermail/rbridge/2010-May/003986.html
- Original Paper by Radia Perlman:
  "Rbridges: Transparent Routing"
  - http://www.postel.org/rbridge/infocom04-paper.pdf