TRILL Tutorial

Transparent Interconnection of Lots of Links

Donald E. Eastlake 3rd

Co-Chair, TRILL Working Group
Principal Engineer, Huawei
d3e3e3@gmail.com

1

HUAWEI

Donald E. Eastlake, 3rd

- Principal Engineer at Huawei Technologies
  - Previously with Cisco Systems and before that with Motorola Laboratories.
- Co-Chair of the IETF TRILL Working Group
  - Chair of the IETF PPPEXT Working Group
  - Chair of the IEEE 802.11ak Task Group
- Author of 63 IETF RFCs.

Note:

This tutorial represents my personal views, not those of the TRILL WG or Huawei. It is a high level technical overview. It is not practical to include all the details in the specification documents in a presentation of this length.

CONTENTS

WHAT IS TRILL?

- An Evolutionary Protocol
  - Attached end nodes just think it is Ethernet.

- The more bridges you convert to TRILL switches, the better your network's stability and bandwidth utilization.

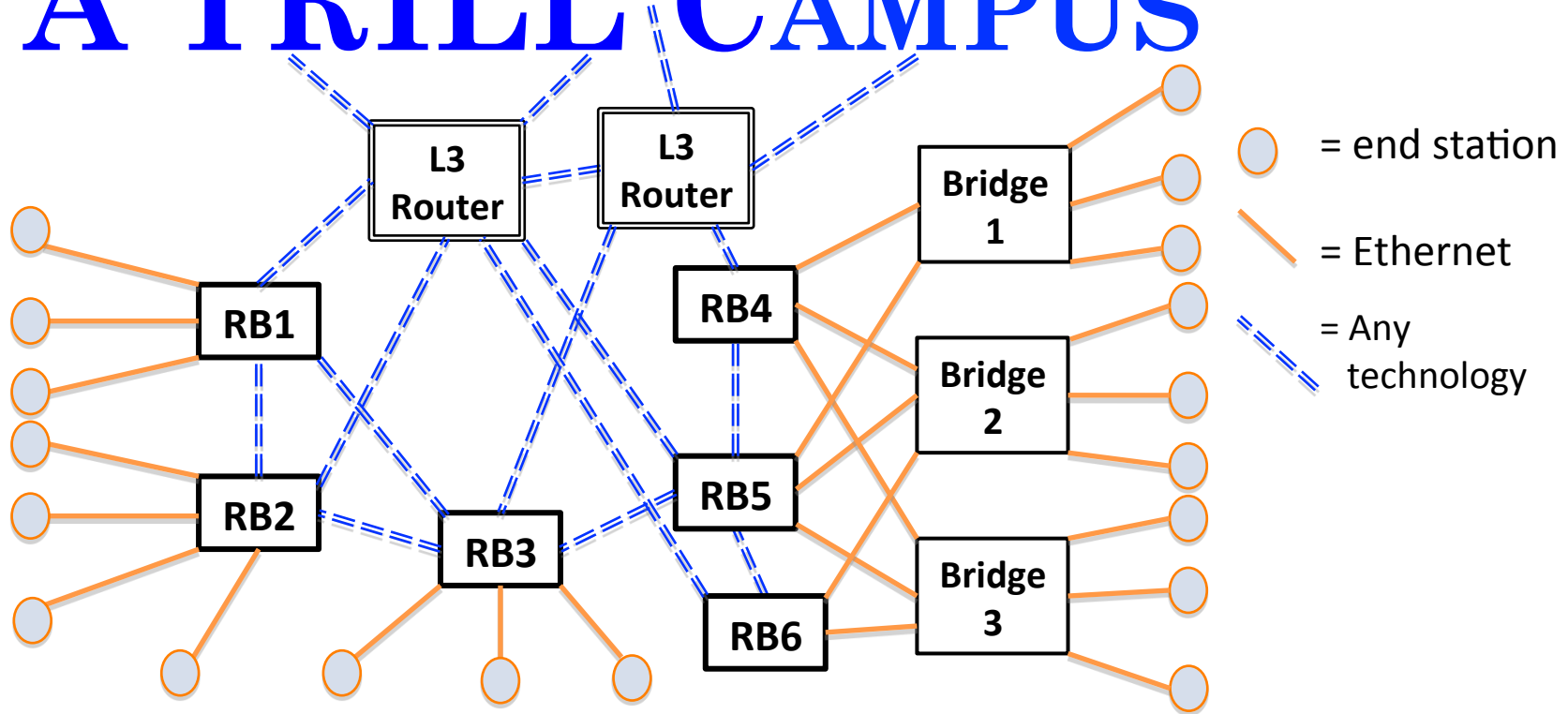- Terminates Spanning Tree Protocols

4

What is 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

- Provides
  - Least cost paths with zero/minimal configuration
  - Equal Cost Multi-Pathing of unicast
  - Multi-paths of multi-destination

What is TRILL?

- <u>TR</u>ansparent <u>I</u>nterconnection of <u>L</u>ots of <u>L</u>inks
  - TRILL WG Charter
    - http://www.ietf.org/dyn/wg/charter/trill-charter.html
  - Standardized by IETF TRILL Working Group:
    - Donald E. Eastlake 3rd (Huawei), Co-Chair
    - Jon Hudson (Brocade), Co-Chair

- <u>TRILL Switch / RBridge</u> (Routing Bridge)
  - Device that implements TRILL

- <u>TRILL/RBridge Campus</u> –
  - A network of RBridges, links, and any intervening bridges, that connects end stations and layer 3 routers.

A TRILL CAMPUS

L3 Router

L3 Router

Bridge 1

RB1

RB4

Bridge 2

RB5

RB2

RB3

Bridge 3

RB6

○ = end station

╱ = Ethernet

╱ = Any technology

- End stations and Layer 3 routers are connected to TRILL switches by Ethernet.
- TRILL switches can be connected to each other with arbitrary technology.
- In both cases, the connection can be a bridged LAN.

7

CONTENTS

- What is TRILL?
- <u>TRILL Features</u>
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

UNICAST LEAST COST PATHS
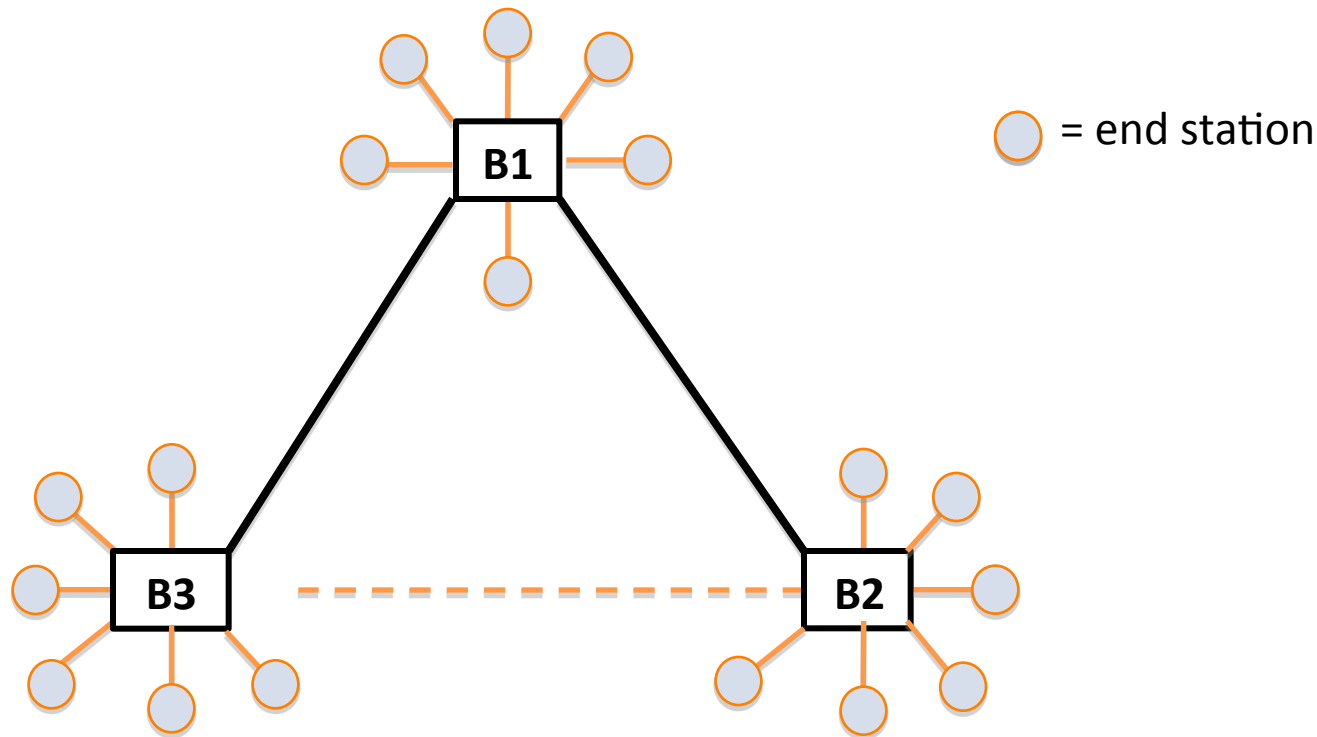
= end station
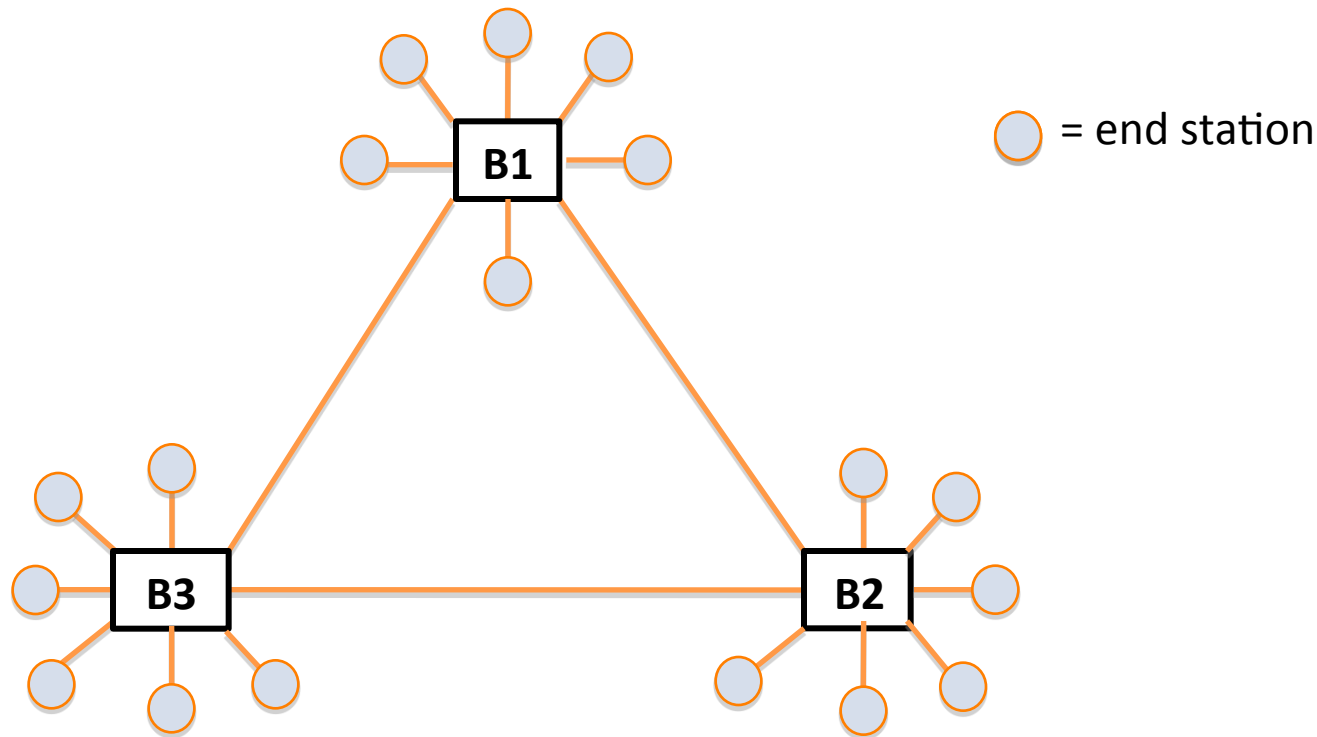
A three bridge network

UNICAST LEAST COST PATHS

= end station

Spanning tree eliminates loops
by disabling ports

UNICAST LEAST COST PATHS

◯ = end station
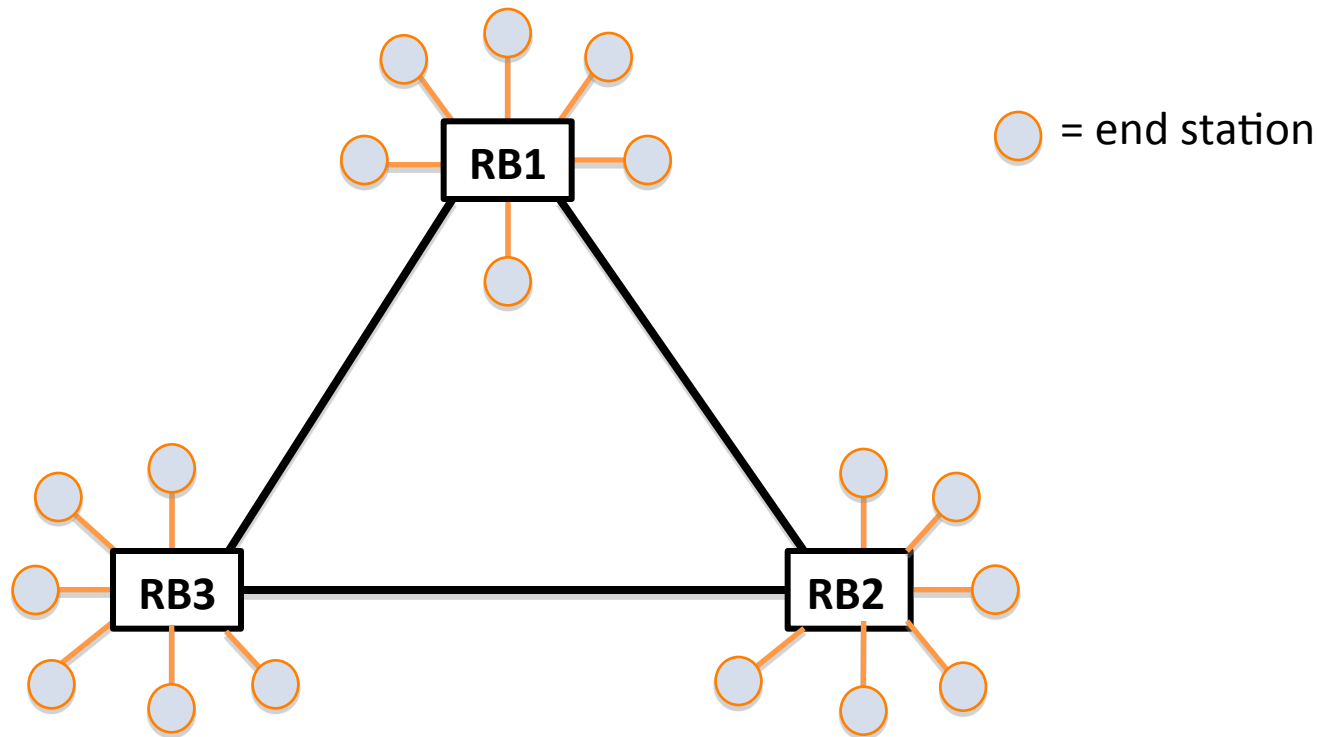
A three RBridge network: better performance using all facilities

UNICAST MULTI-PATHING

◯ = end station
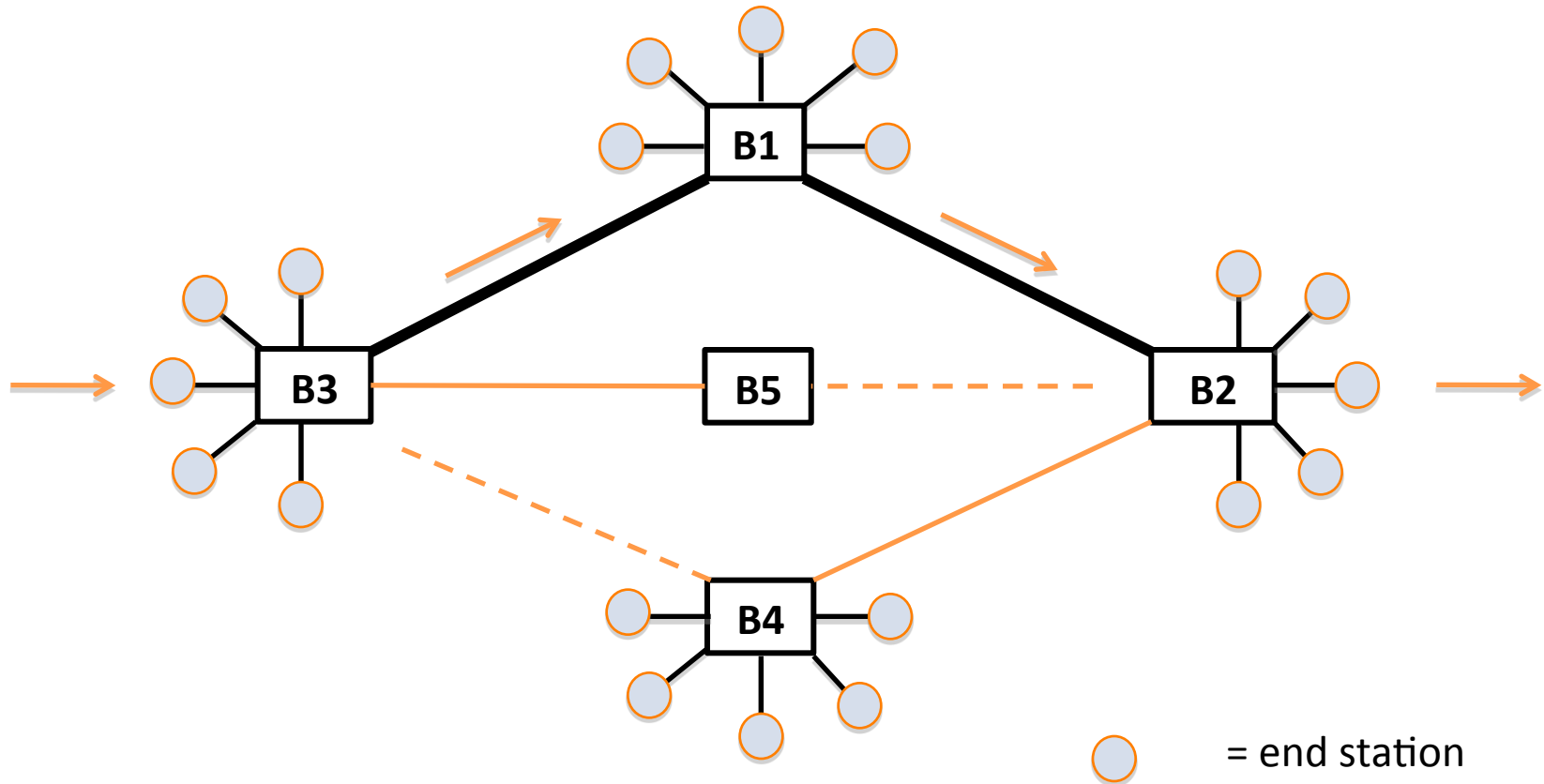
Bridges limit traffic to one path

UNICAST MULTI-PATHING

RB1

RB3   RB5   RB2

RB4

= end station
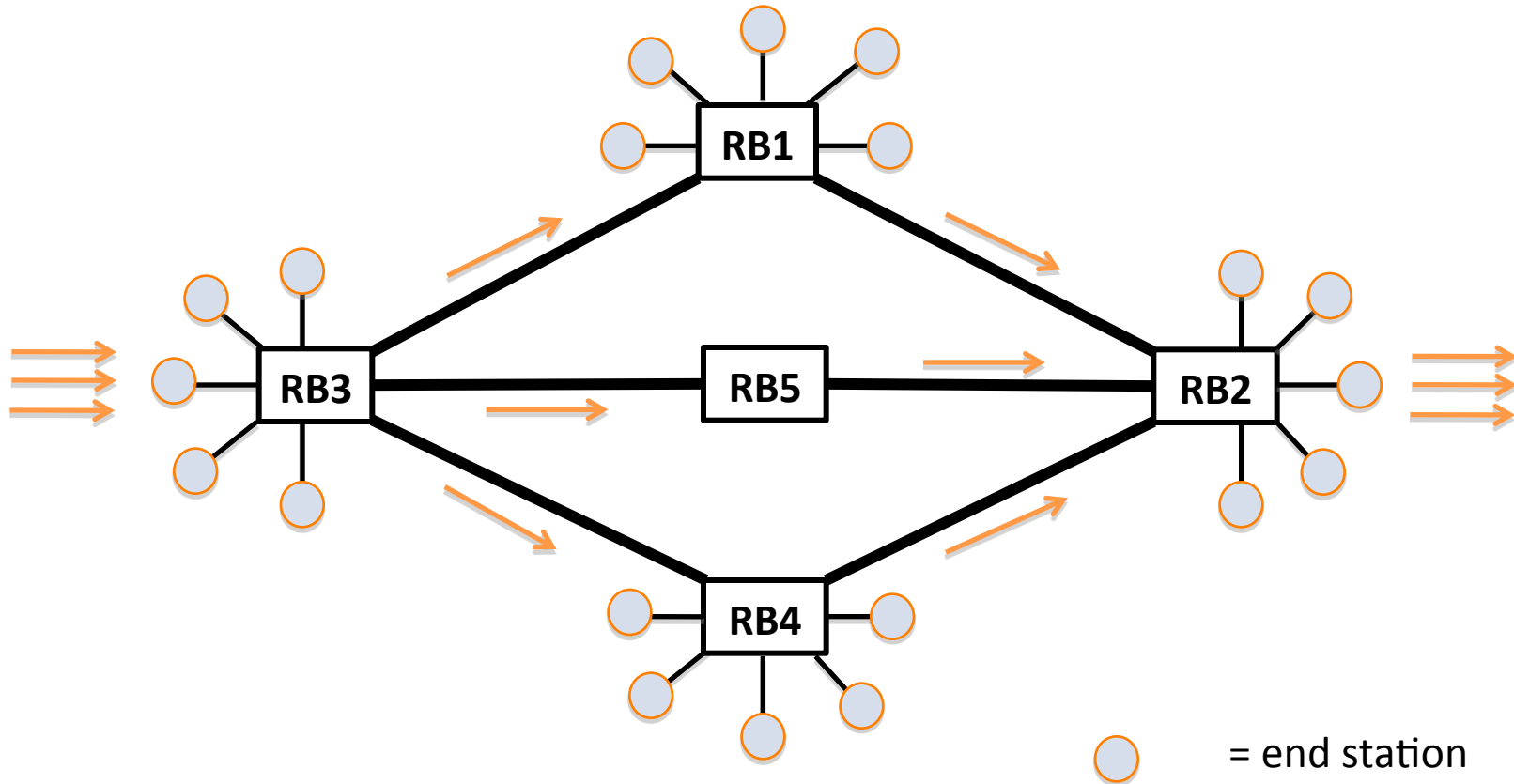
RBridges support
multi-path for higher throughput

Multi-Pathing (Unicast)

= end station
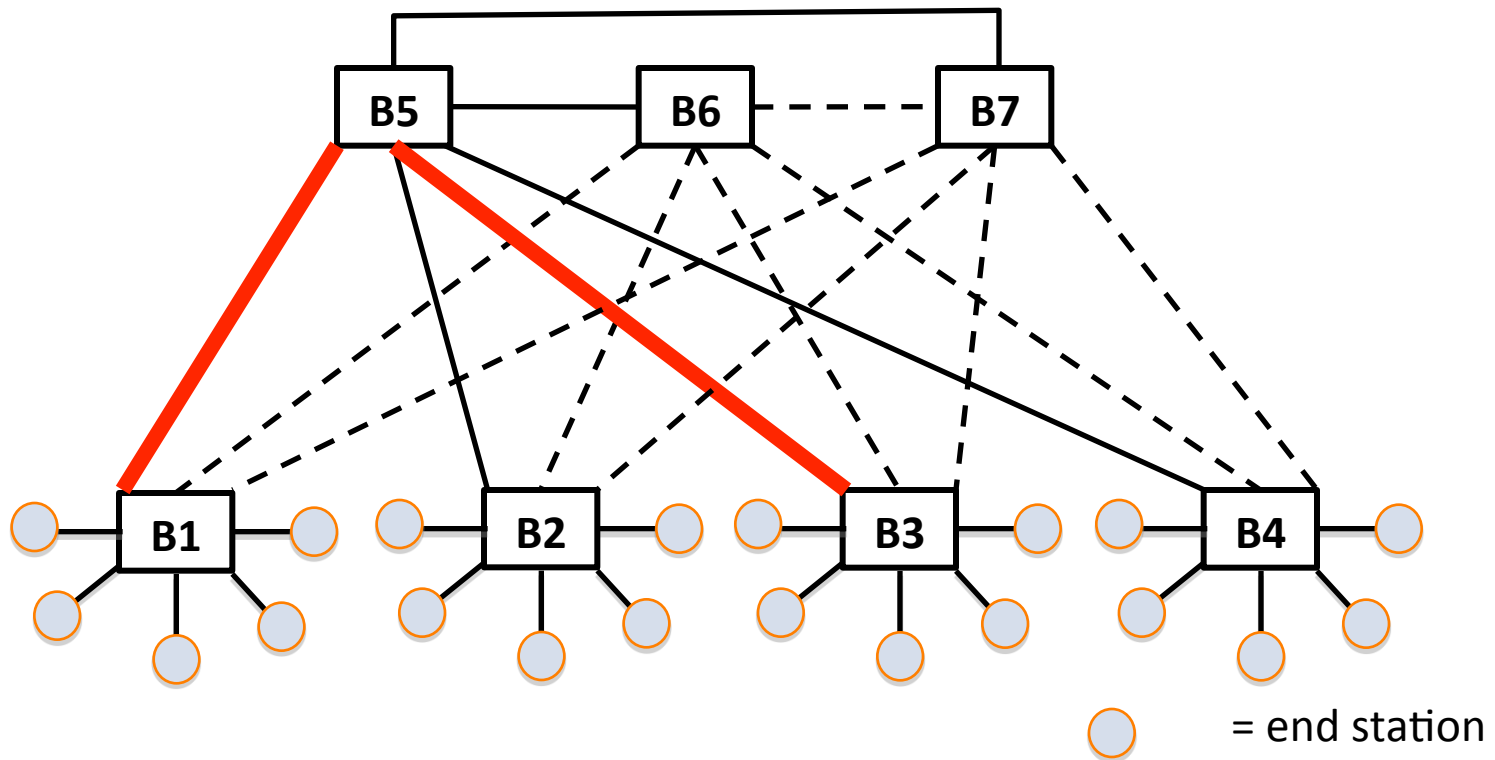
Bridges limit traffic to one path

Multi-Pathing (Unicast)

= end station
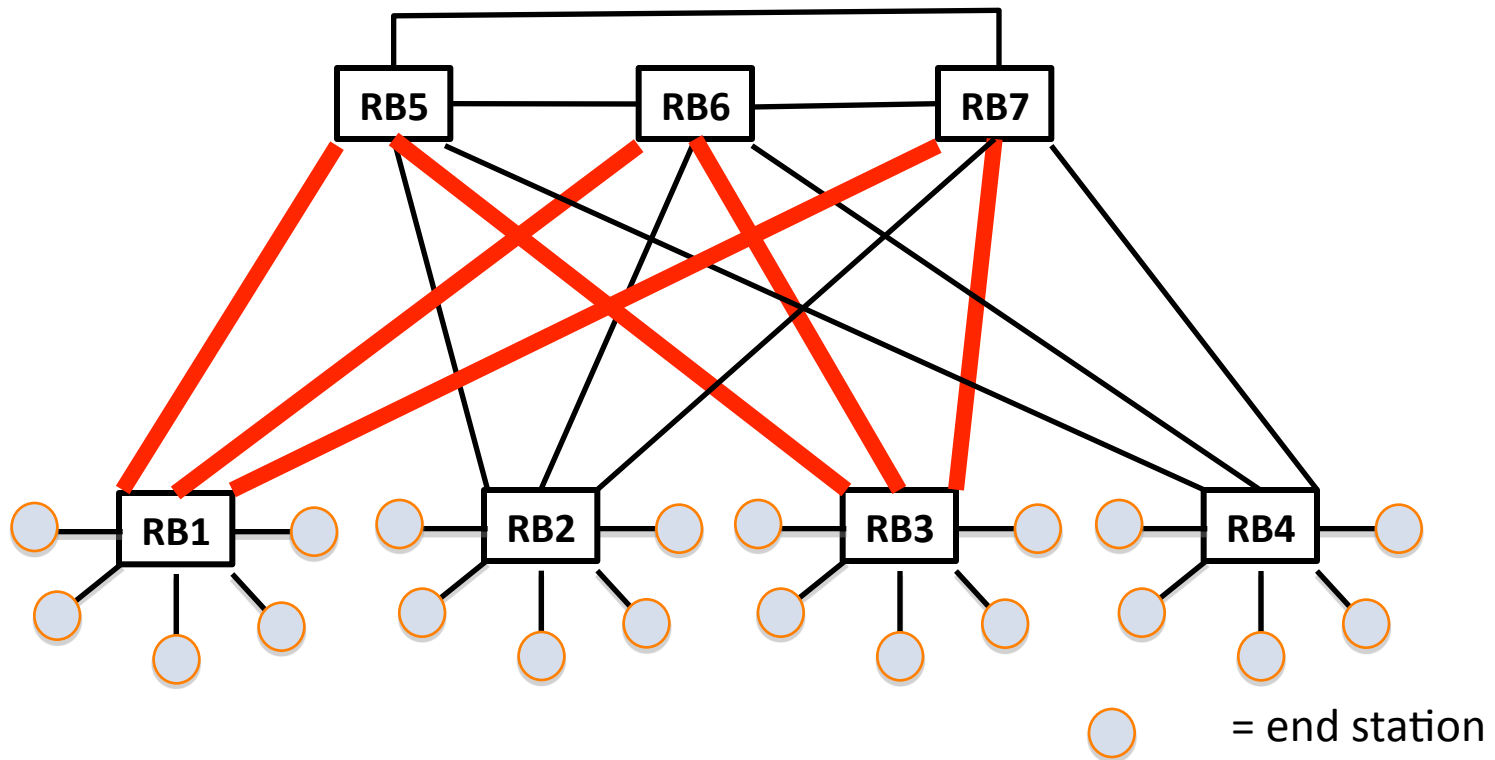
RBridges support
multi-pathing for higher throughput

Multi-Pathing (Multi-destination)

= end station
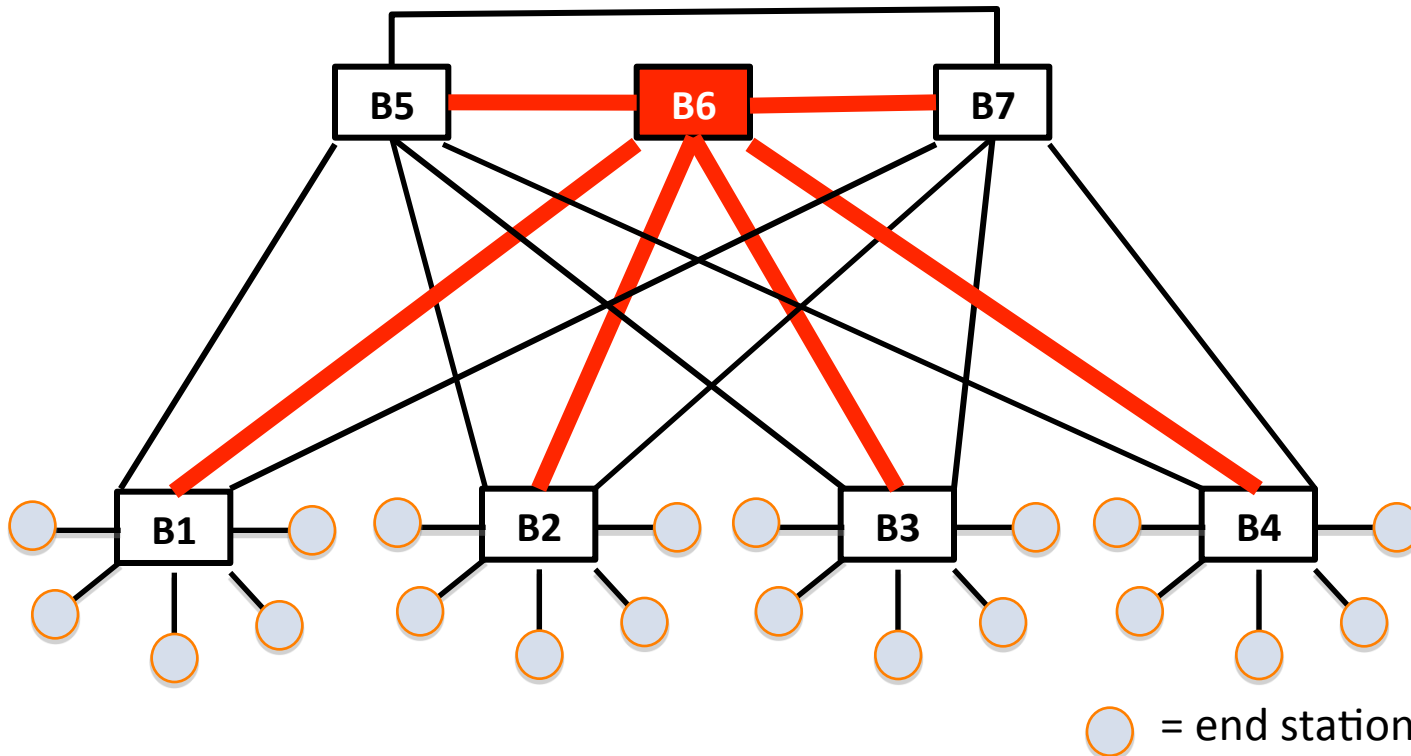
Spanning tree yields a single bi-directional tree for flooding multi-destination frames limiting bandwidth

Multi-Pathing (Multi-destination)

= end station
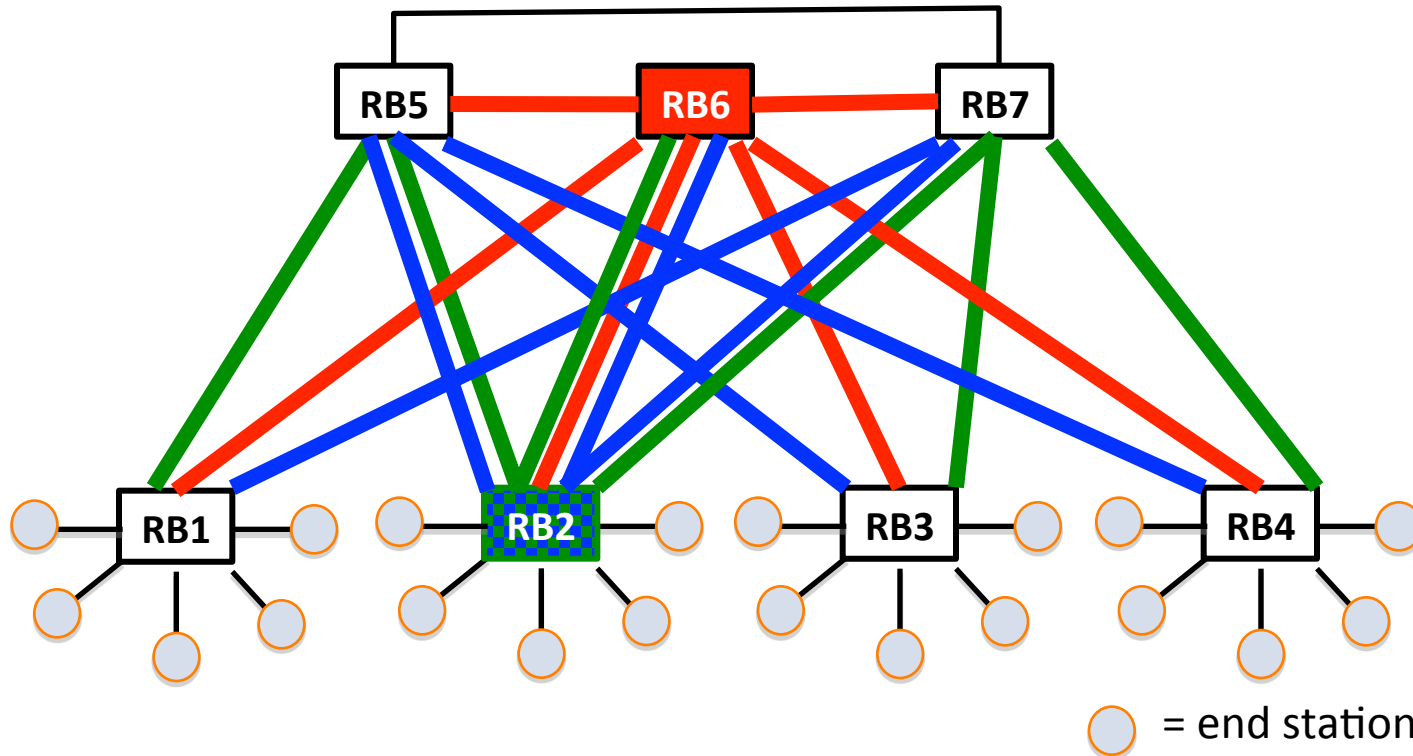
RBridges support multiple distribution trees. The encapsulating RBridge chooses which to use. RB2 can split multi-destination traffic over three trees.

Original Bridging Constraints

- When transparent bridging was initially designed, it was subject to severe constraints:
  - No modifications could be made to the frames. There was not a single spare bit that could safely be used.
  - There was a hard limit to the size. Thus, encapsulation would not have been possible.
  - End nodes must not have to do anything differently than they would on a single CSMA/CD LAN.
- Spanning Tree / Bridging was a brilliant solution, given these constraints, but it is brittle.
  - Failure to properly process control messages leads to melt down.

Routing versus Bridging

- Routing only sends data out a port when it receives control messages on that port indicating this is safe and routing has a TTL for safety.
  - If control messages are not received or not processed, it "fails safe" and does not forward data.
- Bridging (Spanning Tree Protocol) forwards data out all ports (except the one where the data was received) unless it receives control messages on that port indicate this is unsafe. There is no TTL.
  - If control messages are not received or not processed, it "fails unsafe", forwards data, and can melt down due to data loops.

TRILL Features

Bridges → **TRILL Switch** ← Routers

- Transparency
- Plug & Play
- Virtual LANs
  - Multi-tenant support
- Frame Priorities
- Data Center Bridging
- Virtualization Support

- Multi-pathing
- Optimal Paths
- Rapid Fail Over
- The safety of a TTL
  - Implemented in data plane
- Extensions

MORE TRILL FEATURES

- Breaks up and minimizes spanning tree for greater stability.

- 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.

- Compatible with existing IP Routers. TRILL switches are as transparent to IP routers as bridges are.

- Support for VLANs, frame priorities, and 24-bit data labels ("16 million VLANs").

MORE TRILL FEATURES

- MTU feature and jumbo frame support including jumbo routing frames.
- Has a poem.
  - The only other bridging or routing protocol with a poem is Spanning Tree (see Algorhyme).

Algorhyme V2 (TRILL and RBridges)

- 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
  (Radia Perlman's son)

CONTENTS

- What is TRILL?
- TRILL Features
- <u>TRILL History</u>
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

Inspired by a Real Life Incident

- In November 2002, Beth Israel Deaconess Hospital in Boston, Massachusetts, had a total network meltdown:
  - Their network took four days of heroic efforts to be restored to an operational state! In the mean time the staff was reduced to using paper and pencil.
  - Beth Israel Deaconess had grown by acquiring various clinics and just plugged all those bridged networks together.
  - The article in Boston's primary newspaper specifically mentioned "Spanning Tree Protocol" as the problem!
  - Radia Perlman, who invented spanning tree over 25 years ago, decided it was time to come up with a better way.

TRILL History up to 2008

- 1964: Packet switching/routing invented by Paul Baran.
- 1973: Ethernet invented by Robert Metcalfe
- 1979: Link State Routing invented by John McQuillan.
- 1985: Radia Perlman invents the Spanning Tree Protocol.
- 1987: DECnet Phase V / IS-IS designed by Radia Perlman.
- **2002: Beth Israel Deaconess Hospital network in Boston melts down due to deficiencies in the Spanning Tree Protocol.**
- 2004: TRILL presented by inventor Radia Perlman at Infocom.
- 2005: TRILL presented to IEEE 802 by Radia Perlman, rejected.
- **2005: TRILL presented to IETF which Charters the TRILL Working Group.**
- 2008: MTU problem delays protocol while fix is incorporated.

TRILL IN 2009/2011

- 2009: RFC 5556 "TRILL: Problem and Applicability Statement"
- 2009: TRILL Protocol passed up to IESG for Approval.
- **2010: TRILL approved IETF Standard (2010-03-15)**
  - Ethertypes, Multicast addresses & NLPID assigned
- 2010: Successful TRILL control plane interop at UNH IOL
- 2011: TRILL Protocol base document set:
  - RFC 6325: "RBridges: TRILL Base Protocol Specification"
  - RFC 6326: "TRILL Use of IS-IS"
  - RFC 6327: "RBridges: Adjacency"
  - RFC 6361: "TRILL over PPP"
  - RFC 6439: "RBridges: Appointed Forwarders"
- 2011: TRILL Working Group Re-Chartered to do further development of the TRILL protocol

TRILL IN 2012/2013

- 2012: Second Successful TRILL control plane interop at UNH IOL
- 2013: Additional TRILL documents published:
  - RFC 6847: FCoE (Fibre Channel over Ethernet) over TRILL
  - RFC 6850: RBridge MIB
  - RFC 6905: TRILL OAM Requirements
  - RFC 7067: TRILL Directory Assistance Problem and High-Level Design Proposal
- 2013: Third TRILL interop for control and data plane at UNH IOL week of May 20th
- 2013: TRILL Working Group Re-Chartered to do further development of the TRILL protocol

CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

Two TRILL Examples

1. **"Acme Power Plant" Process Control**
   - Large process control commonly uses Ethernet
   - Some process control protocols interpret network interruption >1 second as equipment failure
   - Even Rapid Spanning Tree Protocol can take >3 second to recover from root bridge failure
   - Core RBridges reduce/eliminate spanning tree

2. **"Acme Data Center"**
   - 1:1 to N:1 Backup Improvement

1. Acme Power Plant

Bridged Process Control Network

Rapid Spanning Tree Protocol Domain

A = Access Bridge
GB = aGreggation Bridge
CB = Core Bridge

31
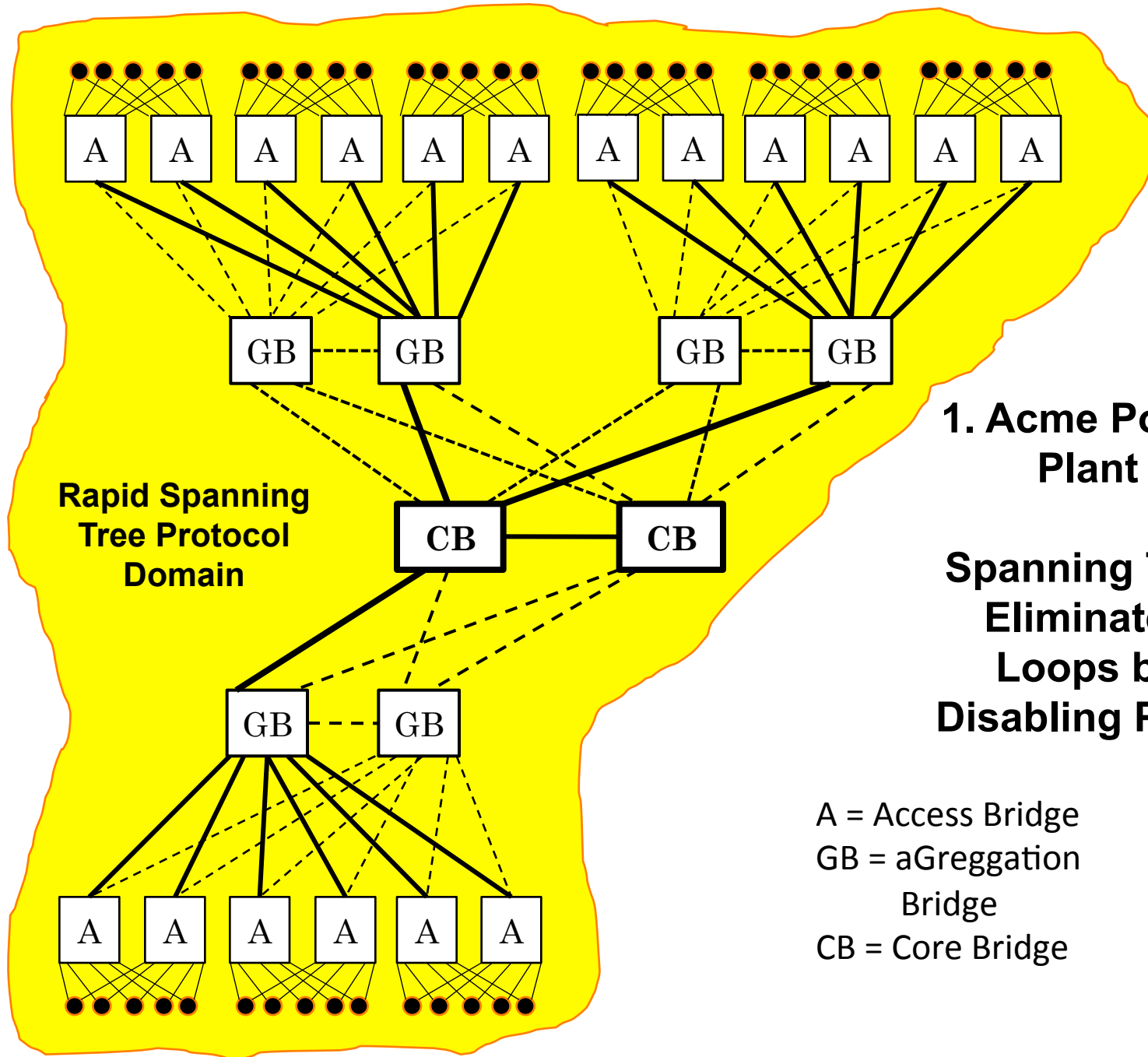
1. Acme Power Plant

Spanning Tree Eliminates Loops by Disabling Ports

A = Access Bridge
GB = aGreggation Bridge
CB = Core Bridge

Rapid Spanning Tree Protocol Domain
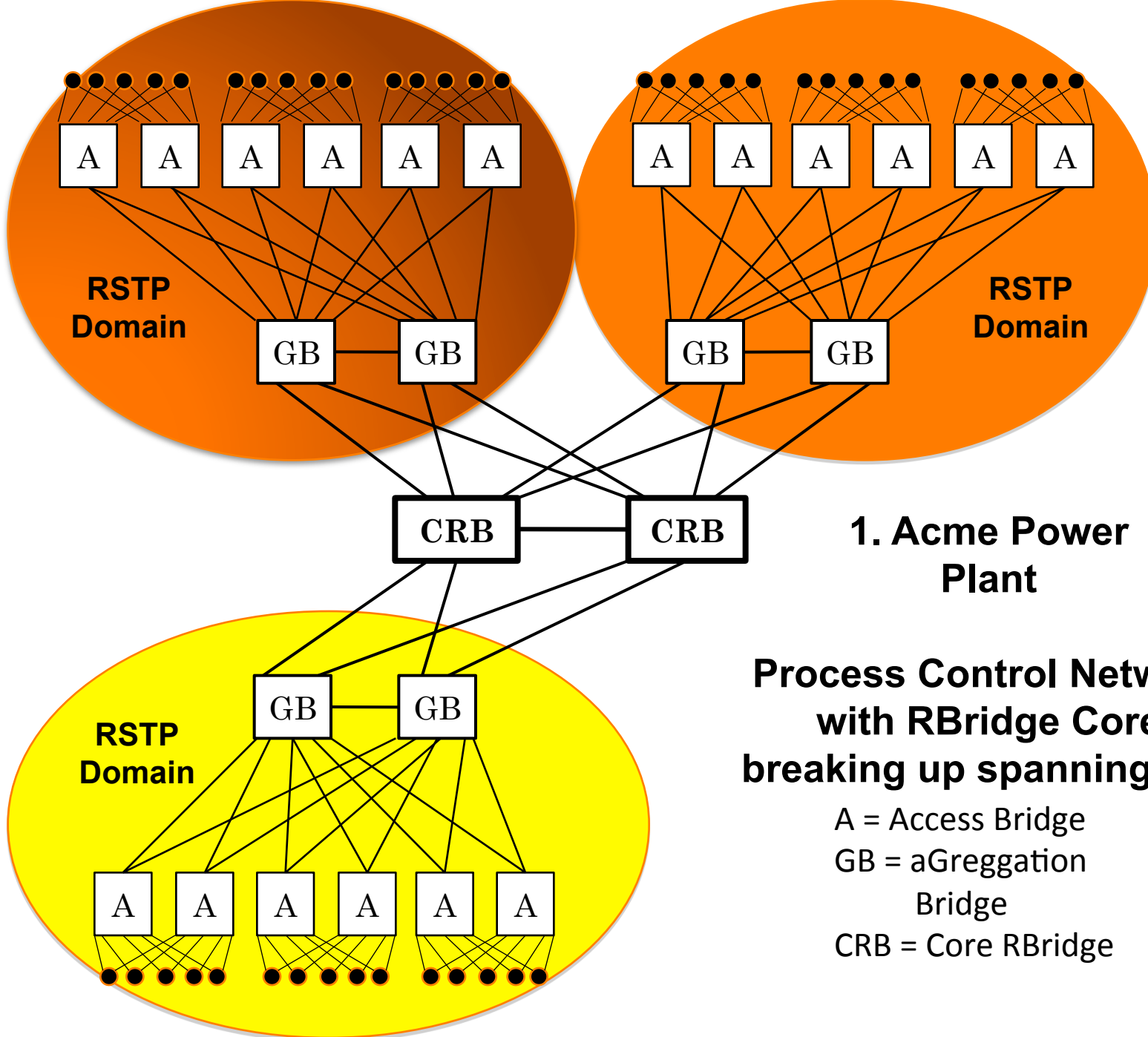
1. Acme Power Plant

Process Control Network with RBridge Core breaking up spanning tree

A = Access Bridge
GB = aGreggation Bridge
CRB = Core RBridge

1. Acme Power Plant

Process Control Network with RBridge Mesh eliminating spanning tree
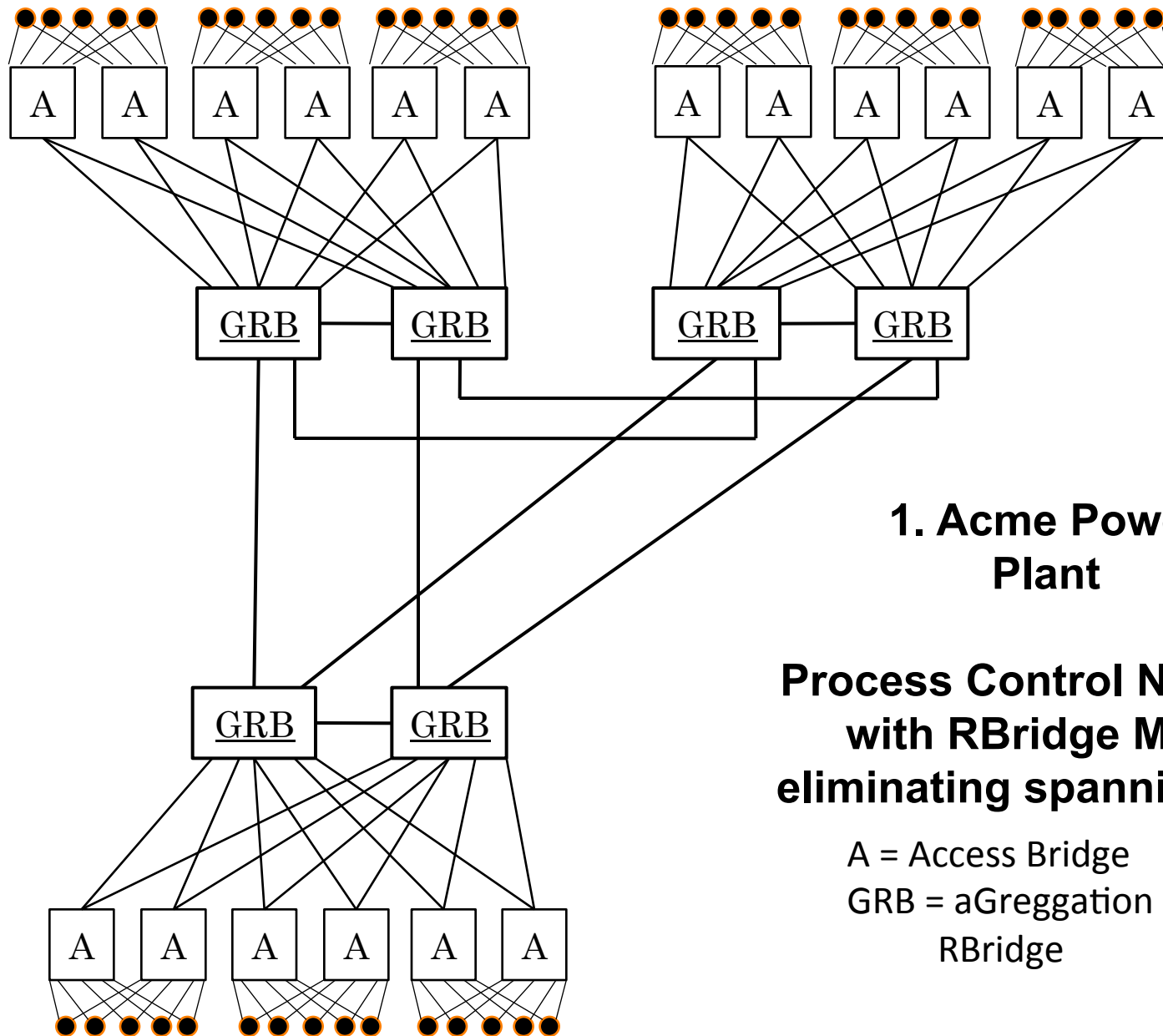
A = Access Bridge
GRB = aGreggation RBridge

34

1:1 Backup

Distribution Bridges must be able to handle 100% of the load. Only 1 path available between any pair of "B"s.

2. Acme Data Center Network with Bridges

B = Head of Rack Bridge
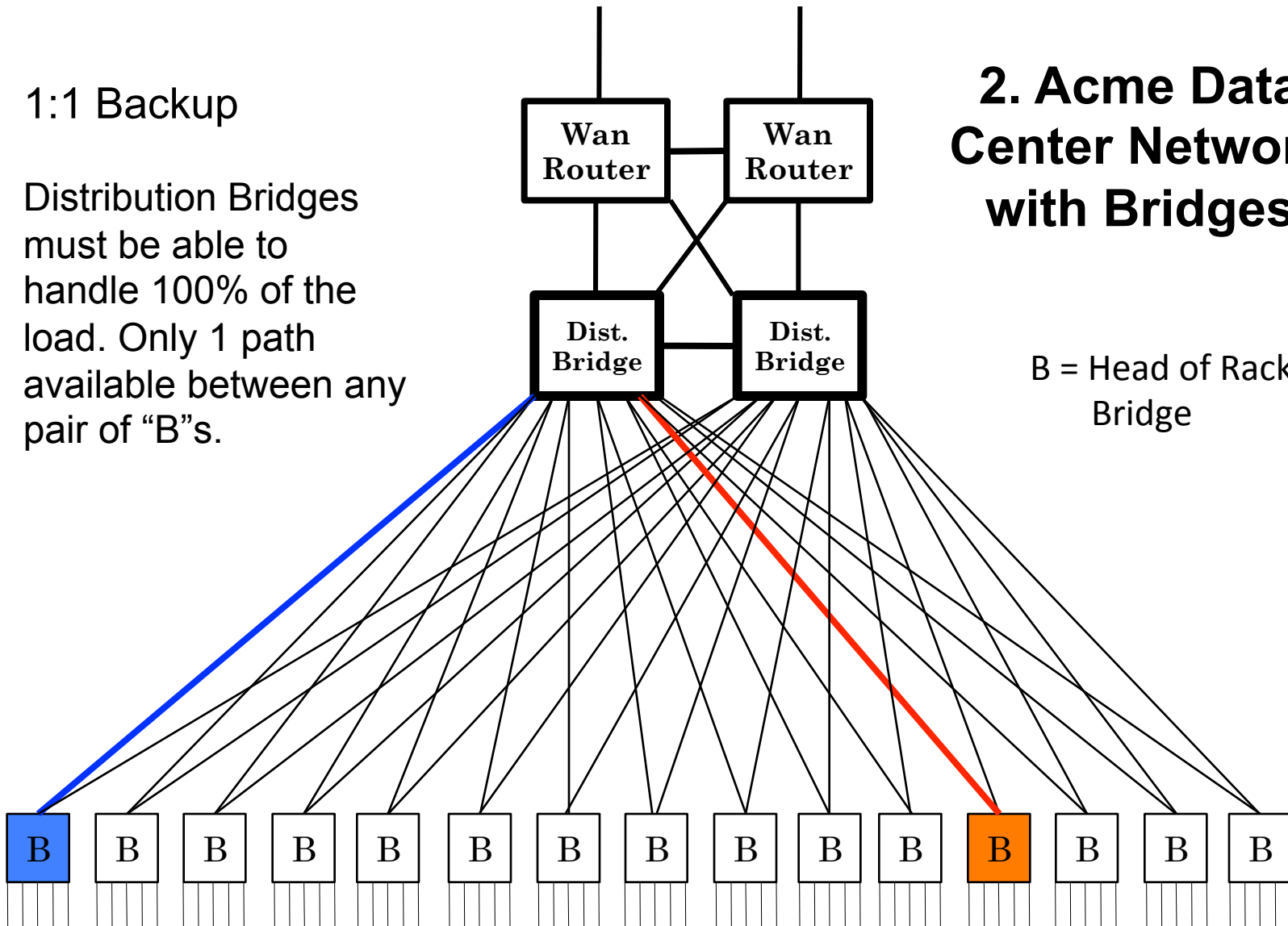
N:1 Backup

Distribution Bridges need to handle only 25% of the load. Multiple available paths between "H"s.

H = Head of Rack RBridge
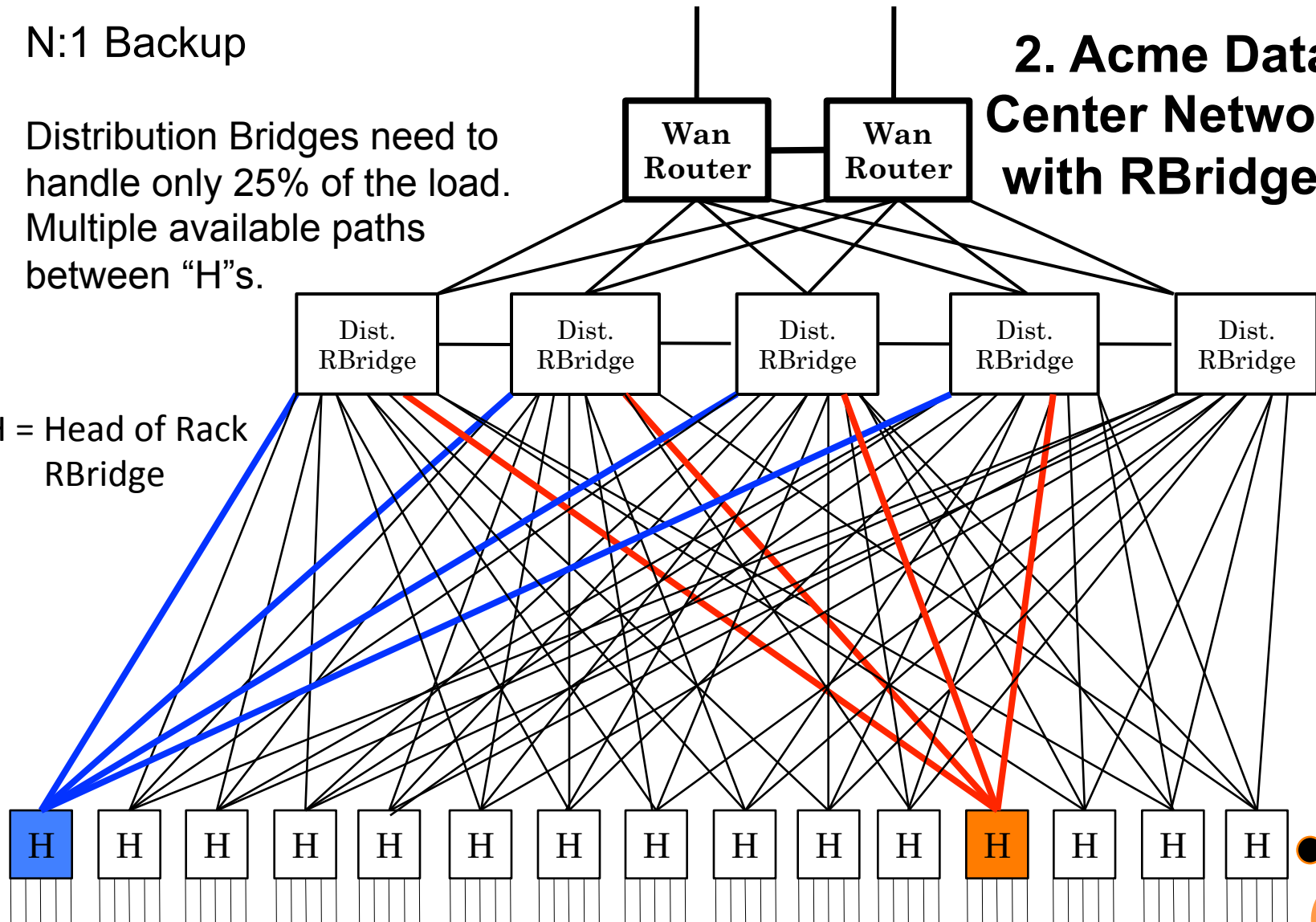
2. Acme Data Center Network with RBridges

CONTENTS

TRILL Packet Headers

- TRILL Data packets between RBridges have a <u>local link header</u> and <u>TRILL Header</u>.

  - The <u>local link header</u> on Ethernet is addressed from the local source RBridge to the next hop RBridge for known unicast frames or to the All-RBridges multicast address for multi-destination frames.

  - The <u>TRILL Header</u> specifies the first/ingress RBridge and either the last/egress RBridge for known unicast frames or the distribution tree for multi-destination frames.

TRILL Packet Headers

- Reasons for TRILL Header:
  - Provides a hop count to reduce loop issues
  - To hide the original source address to avoid confusing any bridges present as might happen if multi-pathing were in use
  - To direct unicast frames toward the egress RBridge so that forwarding tables in transit RBridges need only be sized with the number of RBridges in the campus, not the number of end stations
  - To provide a separate outer VLAN tag, when necessary, for forwarding traffic between RBridges, independent of the original VLAN of the frame

TRILL Packet Headers

○ TRILL Header

TRILL Ethertype	V	R	M	ExtLng	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)
- ExtLng = Length of TRILL Header Extensions
- Hop = Hop Limit (6 bits)

TRILL Over Ethernet
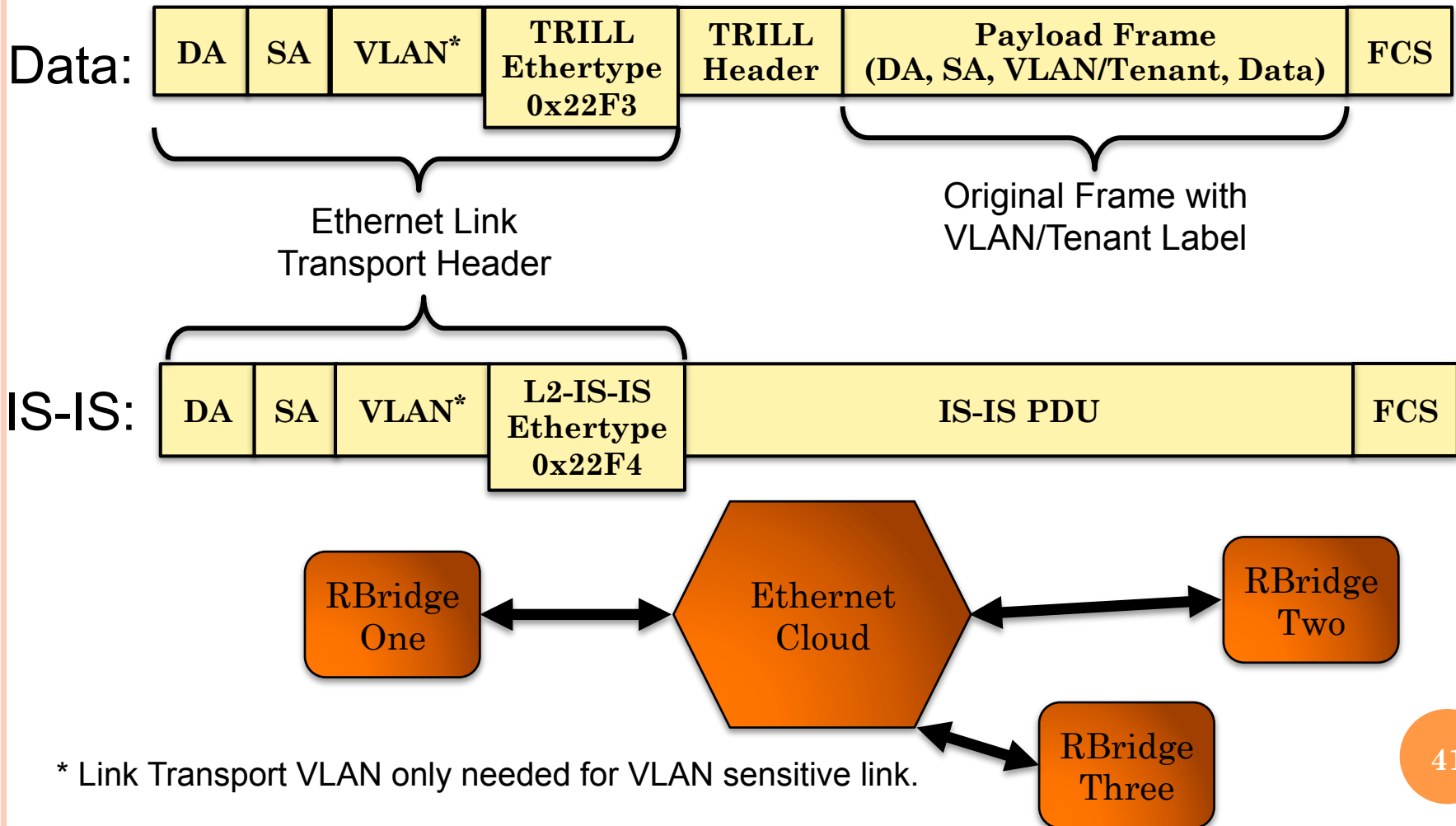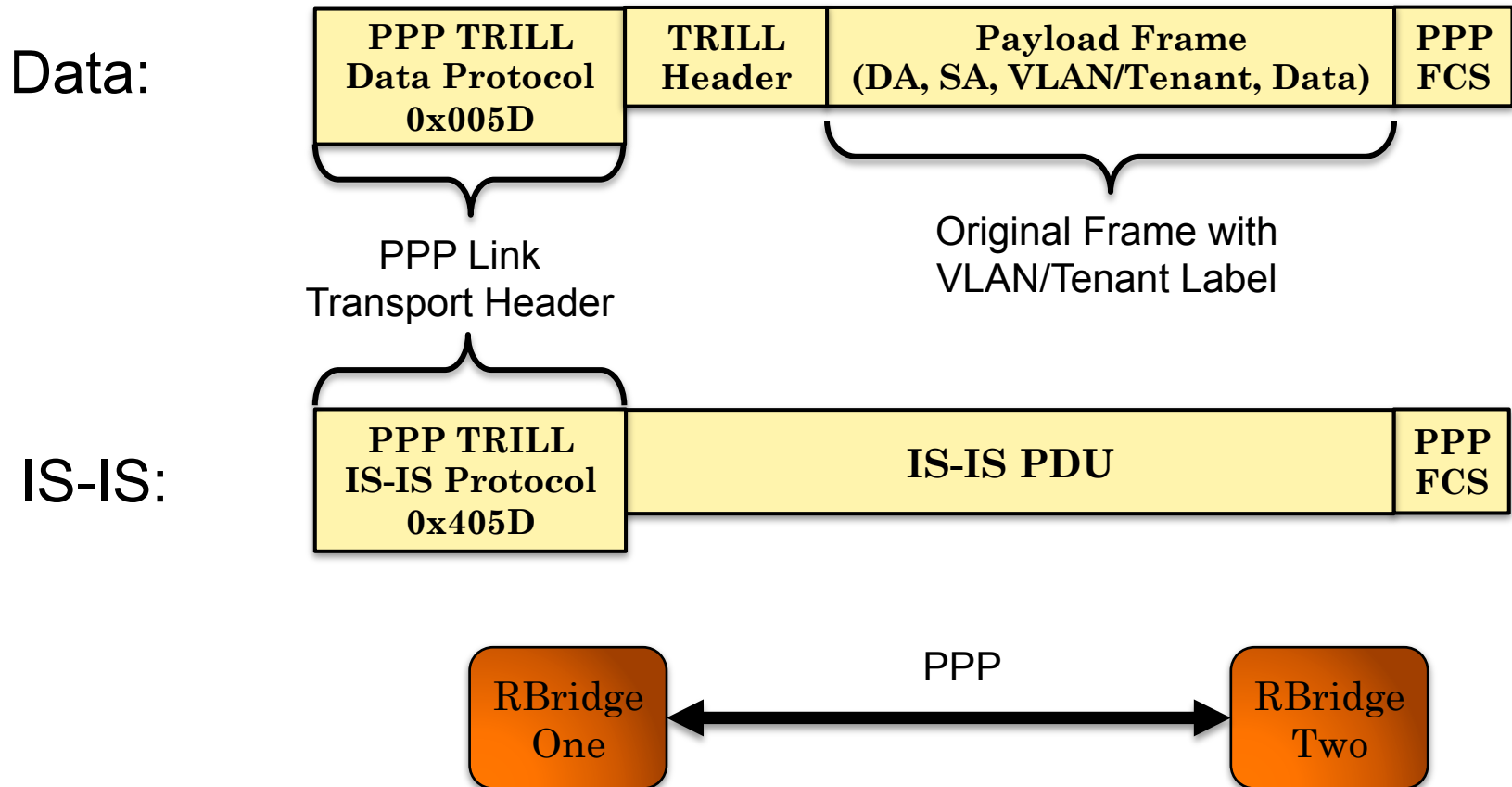
Data:

DA	SA	VLAN*	TRILL Ethertype 0x22F3	TRILL Header	Payload Frame (DA, SA, VLAN/Tenant, Data)	FCS

Ethernet Link Transport Header

Original Frame with VLAN/Tenant Label

IS-IS:

DA	SA	VLAN*	L2-IS-IS Ethertype 0x22F4	IS-IS PDU	FCS

RBridge One ↔ Ethernet Cloud ↔ RBridge Two

Ethernet Cloud ↔ RBridge Three

* Link Transport VLAN only needed for VLAN sensitive link.

41

TRILL Over PPP

Data:

PPP TRILL Data Protocol 0x005D	TRILL Header	Payload Frame (DA, SA, VLAN/Tenant, Data)	PPP FCS

PPP Link Transport Header

Original Frame with VLAN/Tenant Label

IS-IS:

PPP TRILL IS-IS Protocol 0x405D	IS-IS PDU	PPP FCS

RBridge One

PPP

RBridge Two

CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

ETHERNET UNICAST PROCESSING EXAMPLE

○ Step-by-Step on Following Slides:

- Input Port Processing
- TRILL Unicast Ingress
- TRILL Unicast Transit
- TRILL Unicast Egress
- Output Port Processing

INPUT PORT PROCESSING

- Detailed example of unicast frame TRILL routing on an Ethernet link

Input <u>Native Frame</u> on link:

Dest MAC	Src MAC		Data	FCS

<div align="center">

VLAN

</div>
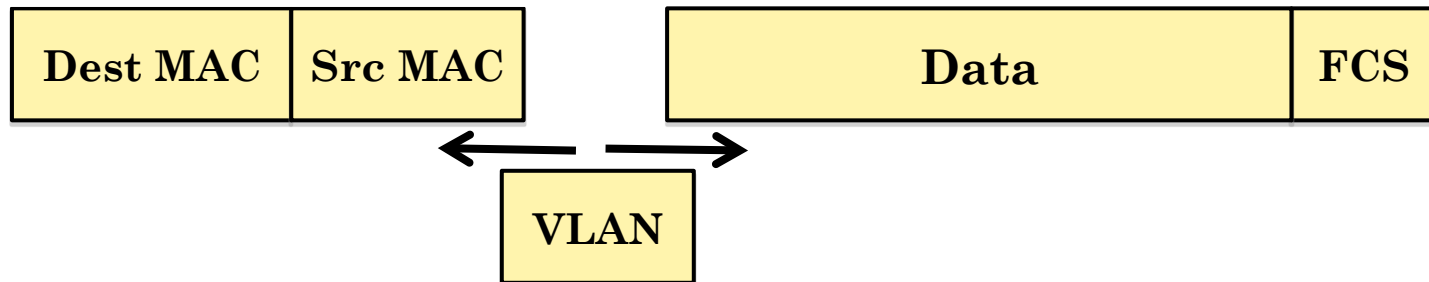
- Input port adds VLAN-ID and priority if frame untagged

Input <u>Native Frame</u> after input port:

Dest MAC	Src MAC	VLAN	Data	FCS

45

TRILL UNICAST INGRESS

Input Native Frame:

Dest MAC	Src MAC	VLAN	Data	FCS

Look Up Egress, Next Hop DA & Output Port

Output TRILL Data Frame:

DA	SA	VLAN[1]	TTL=n	Egress	Ingress	Payload Frame	New FCS
			TRILL Header				

Link Transport Header

Ingressing RBridge

Original Frame with VLAN or Tenant ID
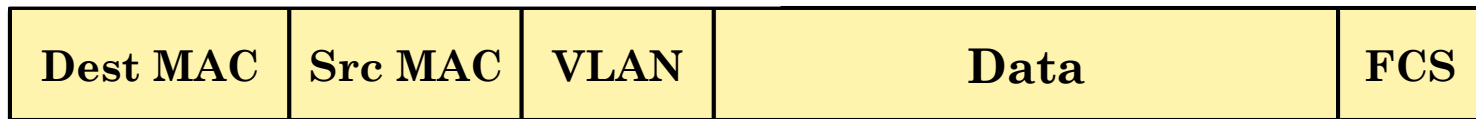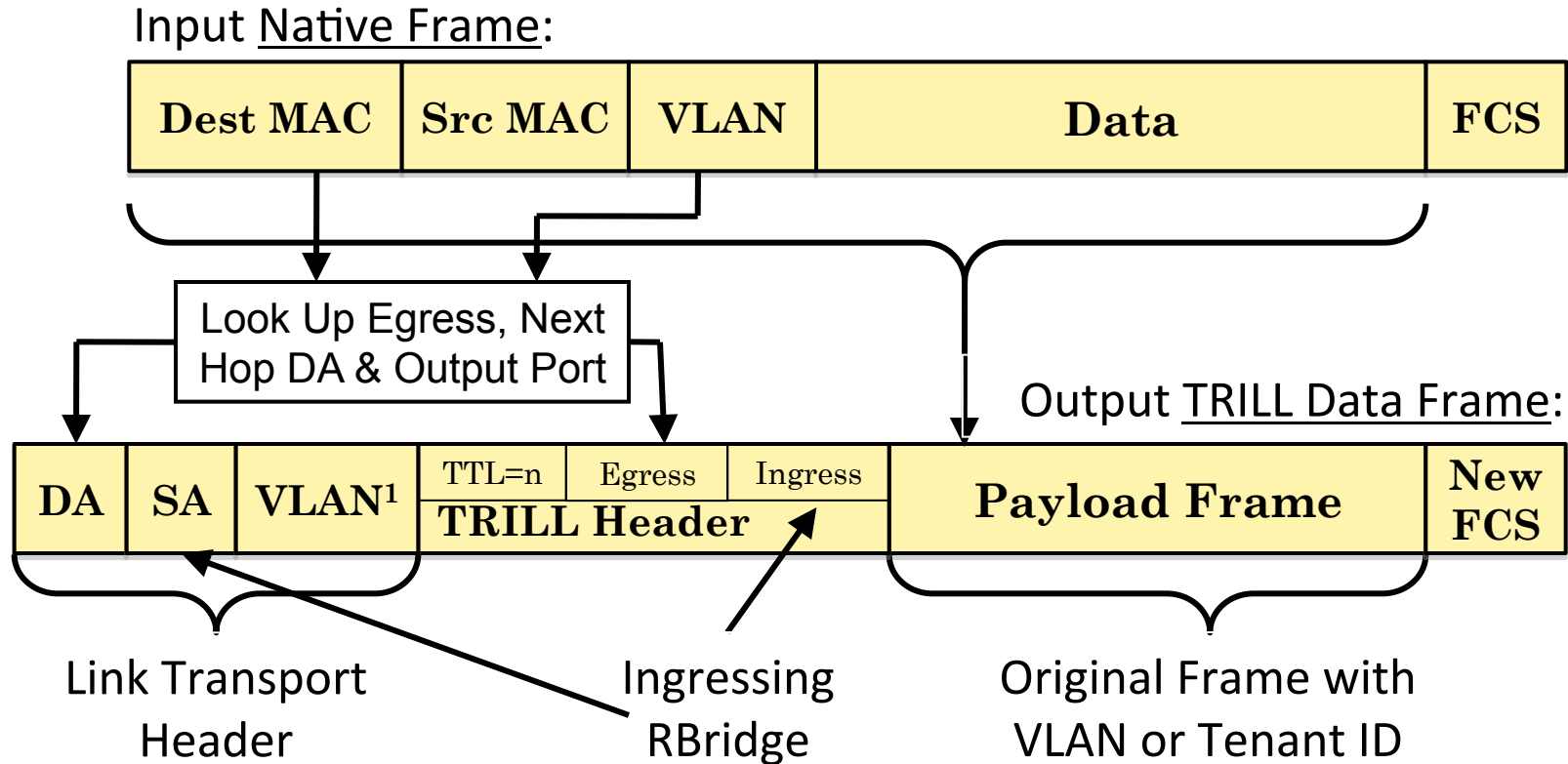
[1] Outer VLAN tag is a transport artifact and only needed if RBridges are connected by a bridged LAN or carrier Ethernet requiring a VLAN tag or the like.

46

TRILL UNICAST TRANSIT

Incoming Link
Transport Header

Input <u>TRILL Data Frame:</u>

DA	SA	VLAN[1]	TRILL Hdr	TTL=n		Payload Frame	FCS
				Egress	Ingress		

Transit RBridge

Look Up Next
DA & Output Port

DA	SA	VLAN[1]	TRILL Hdr	TTL=n-1		Payload Frame	New FCS
				Egress	Ingress		

Output <u>TRILL Data Frame:</u>

Outgoing Link
Transport Header
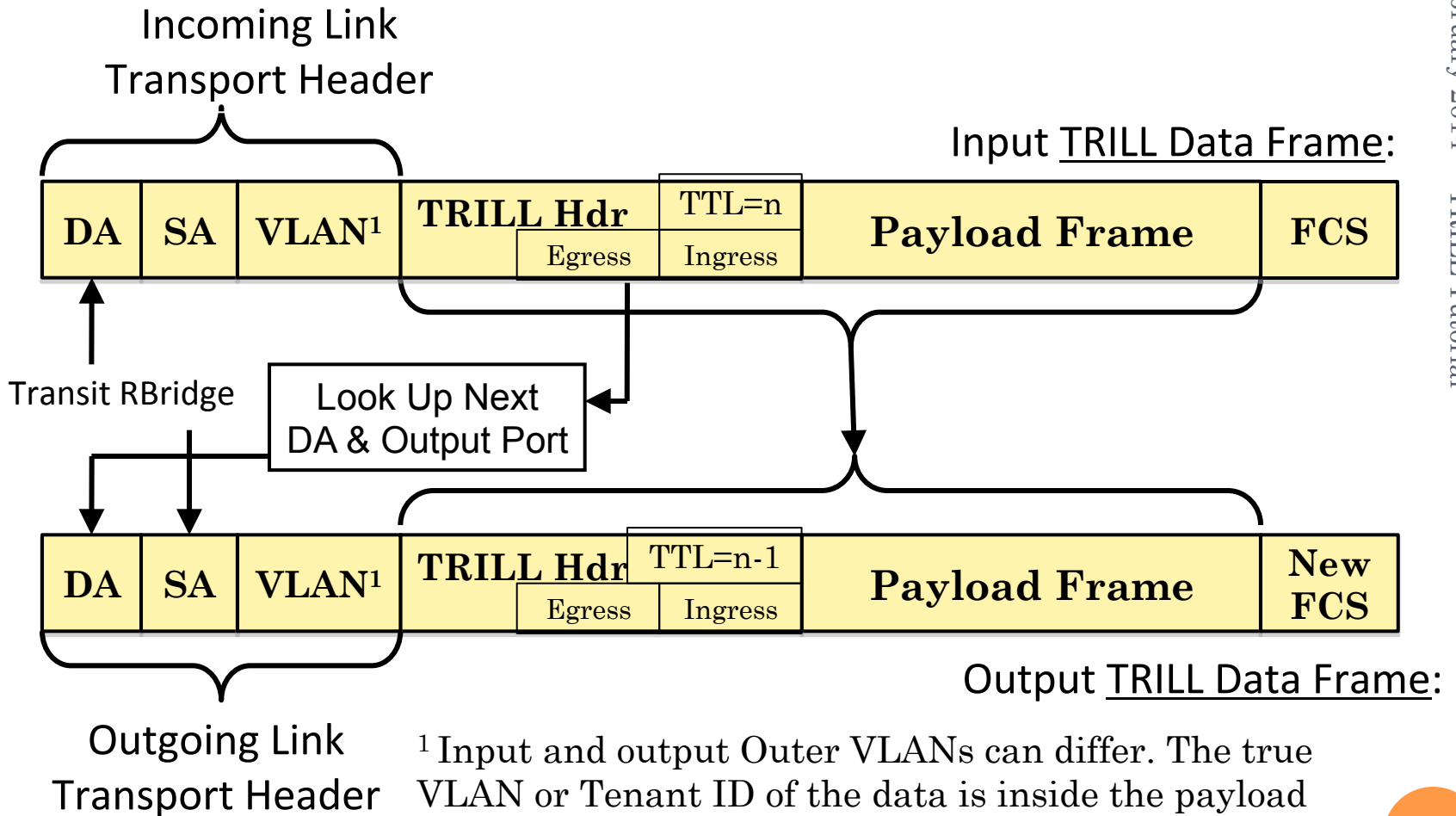
[1] Input and output Outer VLANs can differ. The true VLAN or Tenant ID of the data is inside the payload frame. Outer VLAN is only needed if link is VLAN sensitive.

TRILL UNICAST EGRESS

Link Transport Header

Egressing RBridge

Input <u>TRILL Data Frame</u>:

DA	SA	VLAN[1]	TRILL Hdr		Payload Frame	FCS
			Egress	Ingress		

Output <u>Native Frame</u>:

Dest MAC	Src MAC	VLAN[2]	Data	New FCS

Look Up Output Port

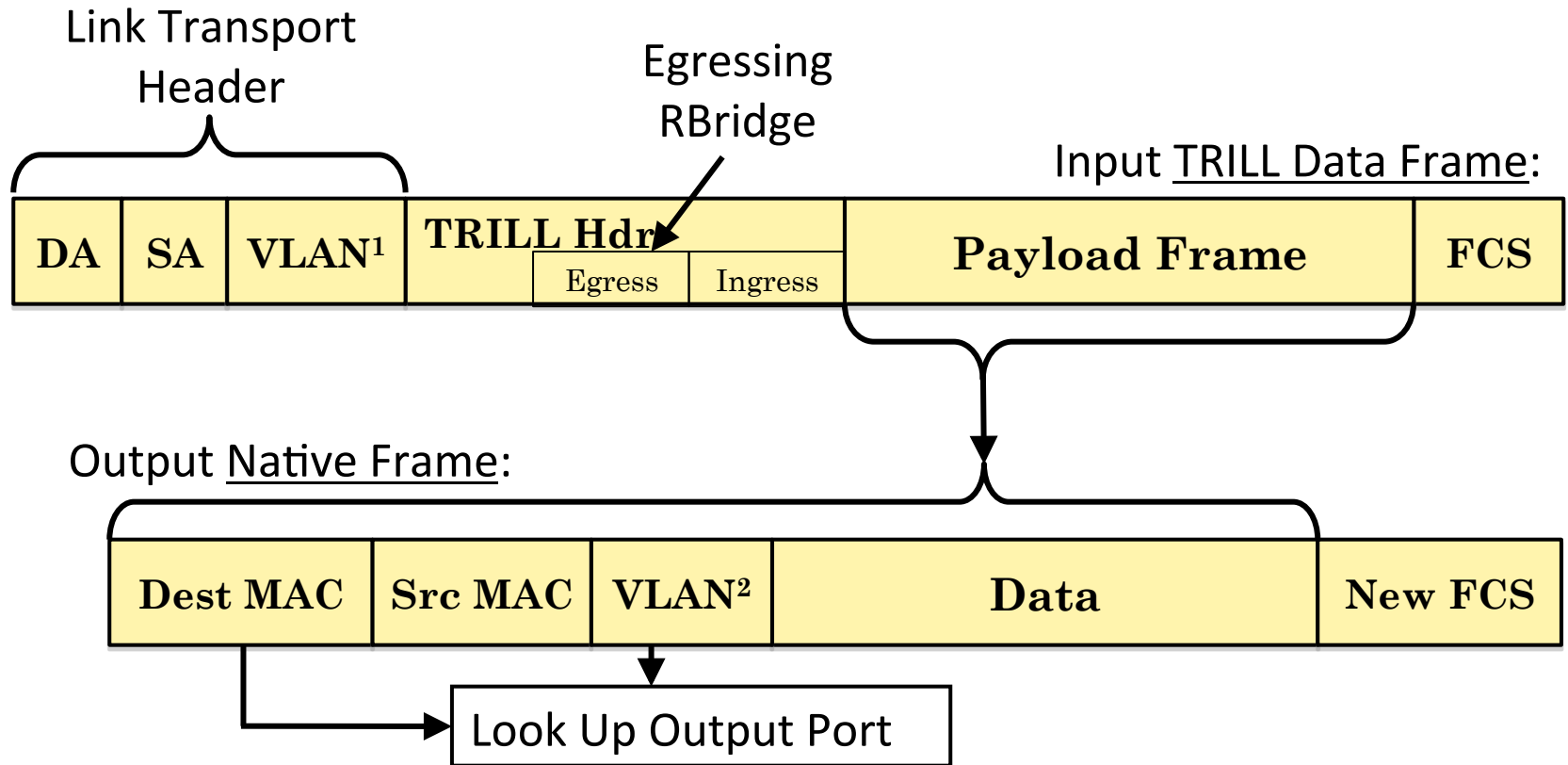[1] Outer VLAN only needed if RBridges are connected by a bridged LAN or carrier Ethernet requiring a VLAN tag or the like.
[2] Final native frame VLAN tag may be omitted depending on RBridge output port configuration.

48
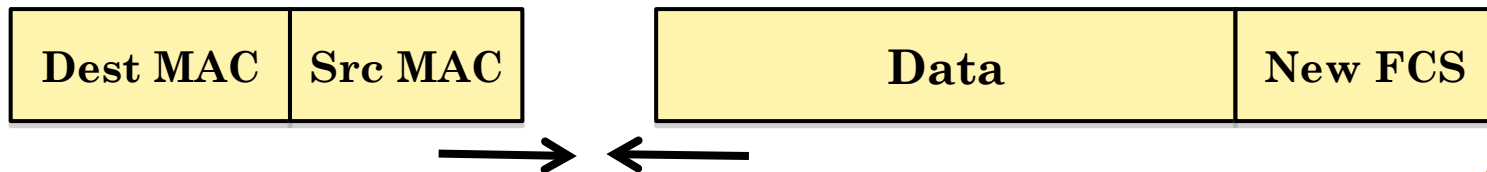
OUTPUT PORT PROCESSING

Output <u>Native Frame</u> before output port:

Dest MAC	Src MAC	VLAN	Data	New FCS

- Output port may be configured to output untagged and will do so by default for the port VLAN ID

Dest MAC	Src MAC		Data	New FCS

→ ←

CONTENTS

FINE GRAINED LABELING

- Fine Grained Labeling (FGL) provides extended 24-bit label (or tenant ID) as the inner data label.
- Ingress/egress TRILL switches that support FGL:
  - Map native frame VLAN and input port into a fine grained label on ingress and
    - do the reverse mapping on egress.
  - Remember the priority and DEI of native frames on ingress and restores them on egress.
- Fine Grained Label TRILL switches are a superset of a base protocol TRILL switches. They support VLANs as in the base standard on a port if not configured to do Fine Grained Labeling.

51

FINE GRAINED LABELING

- ○ See:
  - • https://datatracker.ietf.org/doc/draft-ietf-trill-fine-labeling/

Base protocol VLAN Labeling:

Payload

| Link Header | TRILL Header | DA SA | 8100 | Pri, DEI VID | Payload Frame Data | Link Trailer |

Data Label

Fine Grained Labeling:

| Link Header | TRILL Header | DA SA | 893B | Pri, DEI Upper | 893B | Pri, DEI Lower | Payload Frame Data | Link Trailer |

Transit Priority

24-Bit Tenant ID

Ingress/Egress Priority
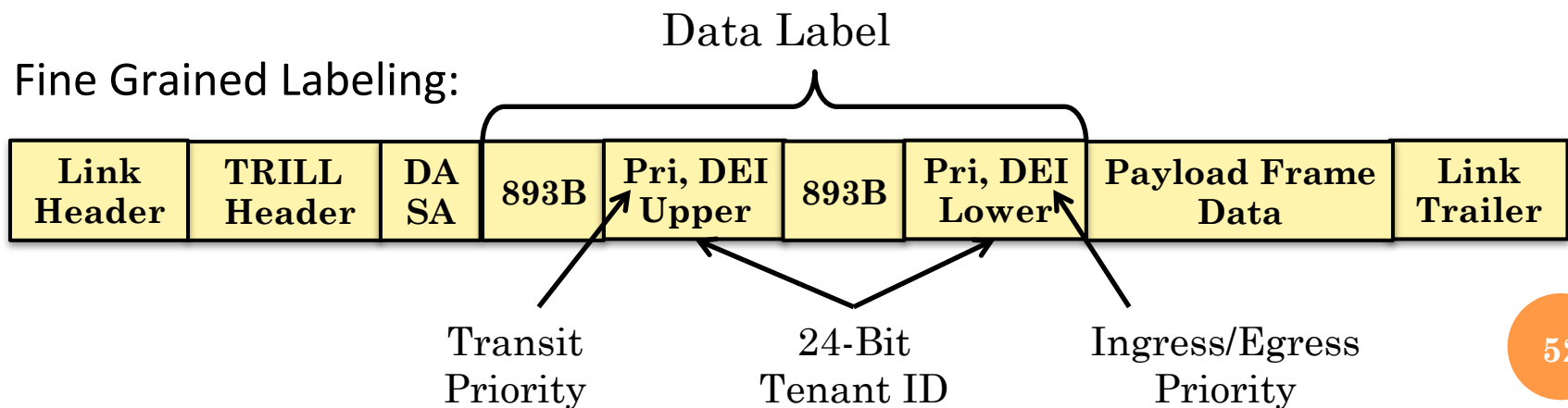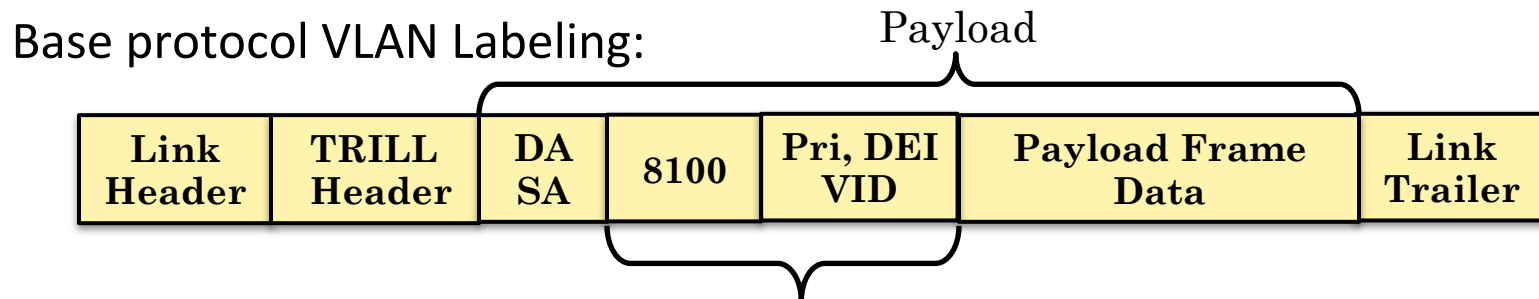
FINE GRAINED LABELING MIGRATION

- An initial deployment of VLAN labeling TRILL switches can be smoothly extended to Fine Grained Labeling
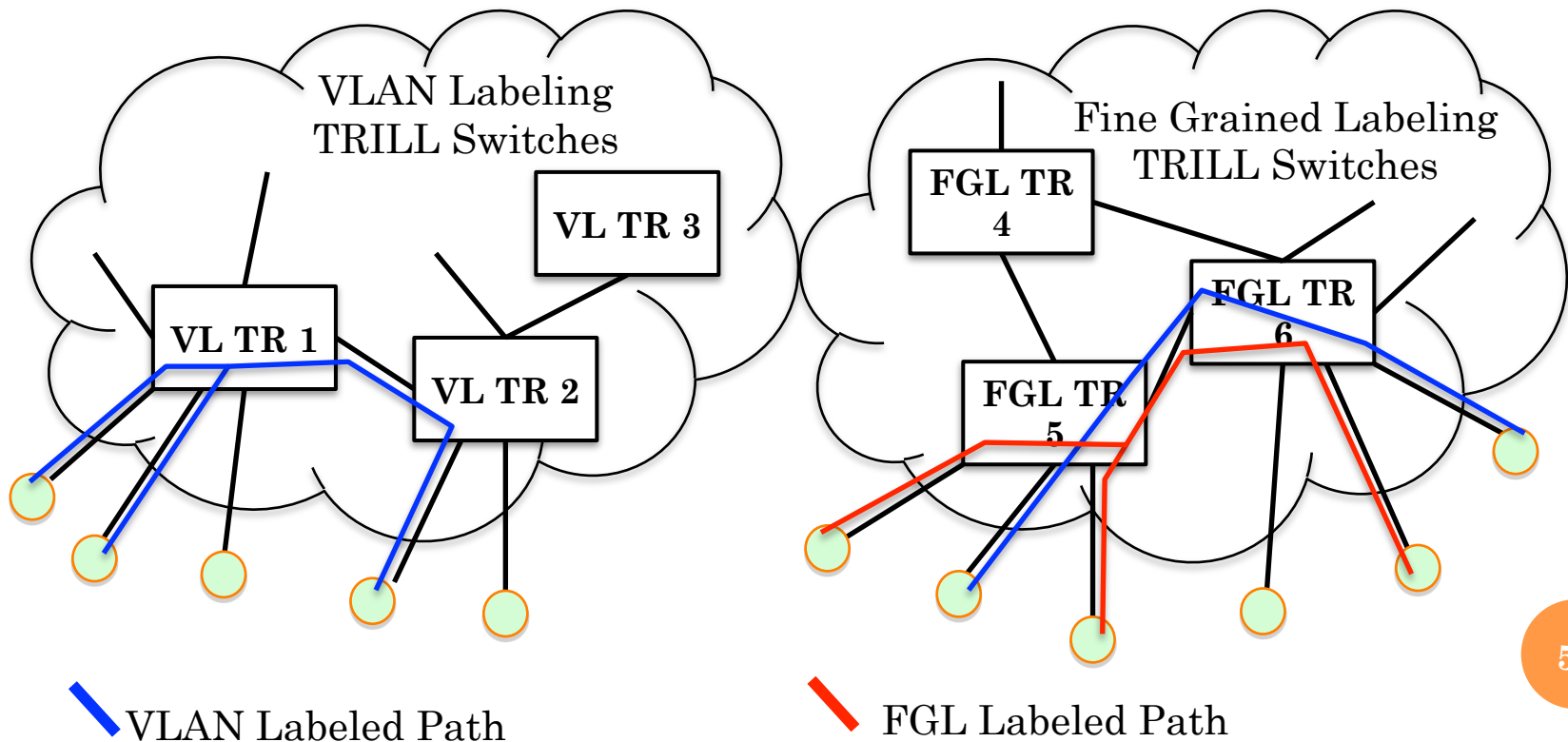
VLAN Labeling TRILL Switches

VL TR 1, VL TR 2, VL TR 3

Fine Grained Labeling TRILL Switches

FGL TR 4, FGL TR 5, FGL TR 6

VLAN Labeled Path

FGL Labeled Path

FINE GRAINED LABELING MIGRATION

- Some VL TRILL switches are convertible to FGL-safe RBridges (FGL transit only) with a software upgrade.

Fine Grained Labeling
TRILL Switches

FGL TR 3
FGL TR 4
FGL TR 1
FGL TR 2
FGL TR 5
FGL TR 6

VLAN Labeled Path

FGL Labeled Path

54

FINE GRAINED LABELING MIGRATION

- Some VL TRILL switches are convertible to FGL-safe RBridges (FGL transit only) with a software upgrade.
- Even if not upgradable, they can generally be connected.

VLAN Labeling
TRILL Switches

Fine Grained Labeling
TRILL Switches

VL TR 3

FGL TR 4

VL TR 1

FGL TR 6

VL TR 2

FGL TR 5

VLAN Labeled Path

FGL Labeled Path

55
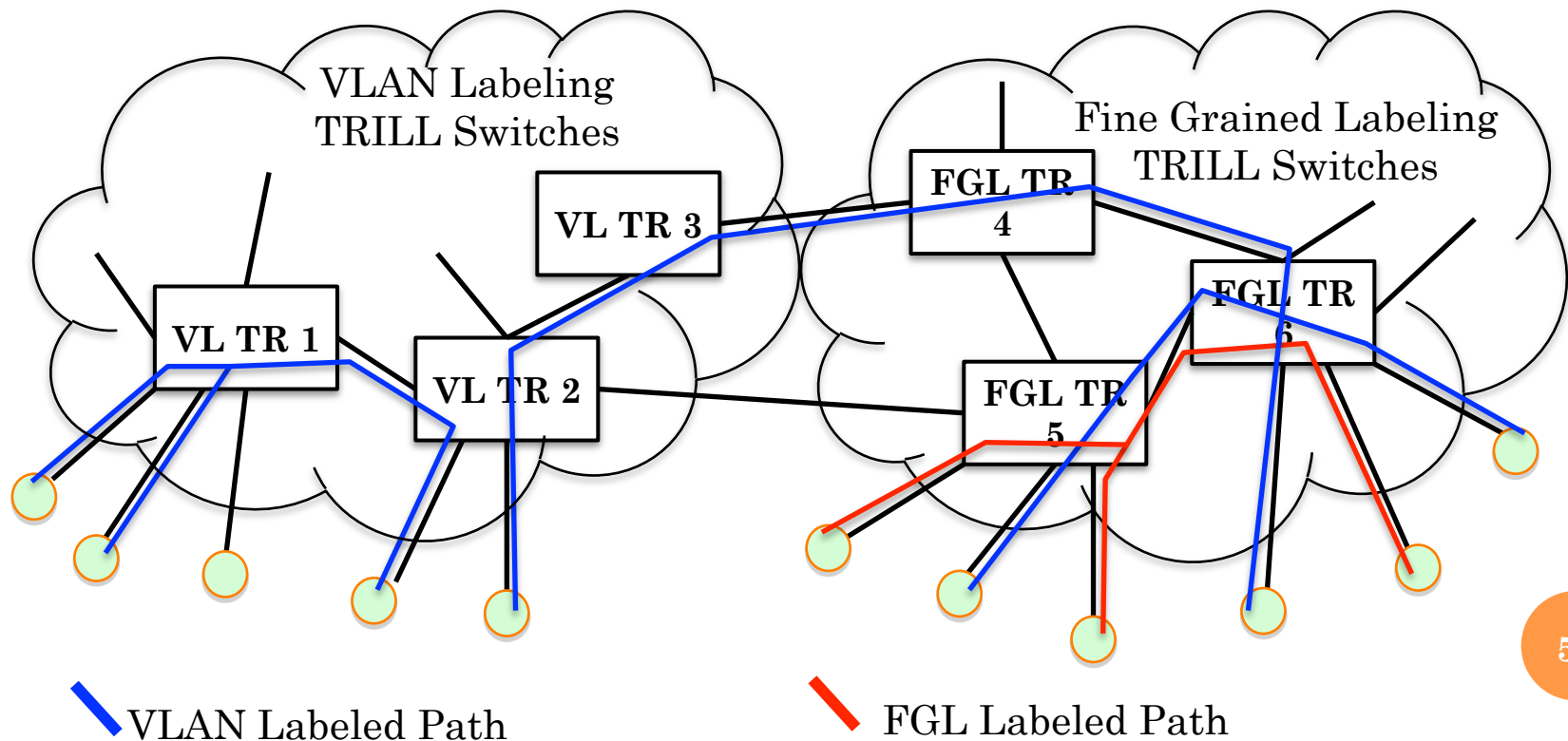
CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Encapsulation and Header
- Step-by-Step Processing Example
- Fine Grained Labeling

- <u>How TRILL Works</u>
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

How TRILL Works

- RBridges find each other by exchanging TRILL IS-IS Hello frames.
  - Like all TRILL IS-IS frames, TRILL Hellos are sent on Ethernet to the multicast address All-IS-IS-RBridges. They are transparently forwarded by bridges, dropped by end stations including routers, and are processed (but not forwarded) by RBridges.
  - TRILL Hellos are different from Layer 3 IS-IS LAN Hellos because they are small, unpadded, and support fragmentation of some information.
    - Separate MTU-probe and MTU-ack IS-IS messages are available for MTU testing and determination.
  - Using the information exchanged in the Hellos, the RBridges on each link elect the Designated RBridge for that link (the link could be a bridged LAN).

How TRILL Works

- TRILL Hellos are unpadded and a maximum of 1470 bytes long to be sure RBridges can see each other so you don't get two Designated RBridges on the same link.
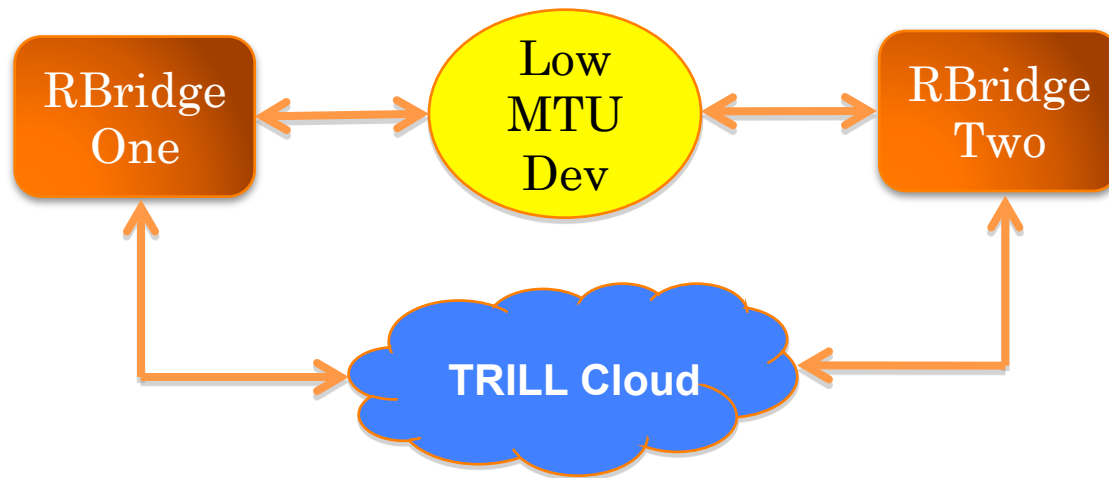
How TRILL Works

- RBridges use the IS-IS reliable flooding protocol so that each RBridge has a copy of the global "link state" database.
  - The RBridge link state includes the campus topology and link cost but also other information. Information such as VLAN/FGL connectivity, multicast listeners and multicast router attachment, claimed nickname(s), ingress-to-egress options supported, and the like.
  - The link state database is sufficient for each RBridge to independently and without further messages calculate optimal point-to-point paths for known unicast frames and the same distribution trees for multi-destination frames.

How TRILL Works

- The Designated RBridge specifies the Appointed Forwarder for each VLAN on the link (which may be itself) and the Designated VLAN for inter-RBridge communication.

- The Appointed Forwarder for VLAN-x on a link handles all native frames to/from that link in that VLAN. It is only significant if there are end station on the link.

  - It encapsulates frames from the link into a TRILL Data frame. This is the ingress RBridge function.

  - It decapsulates native frames destined for the link from TRILL Data frames. This is the egress RBridge function.

60

Why Designated VLAN?

- Ethernet links between RBridges have a Designated VLAN for inter-RBridge traffic. It is dictated by the Designated RBridge on the Link.

- For Point-to-Point links, usually no outer VLAN tag is needed on TRILL Data frames. For links configured as P2P, there is no Designated VLAN.

- However, there are cases where an outer VLAN tag with the designated VLAN ID is essential:
  - Carrier Ethernet facilities on the link restrict VLAN.
  - The link is actually a bridged LAN with VLAN restrictions.
  - The RBridge ports are configured to restrict VLANs.

How TRILL Works

- TRILL Data packets that have known unicast ultimate destinations are forwarded RBridge hop by RBridge hop toward the egress RBridge.

- TRILL Data packets that are multi-destination frames (broadcast, multicast, and unknown destination unicast) are forwarded on a distribution tree.

MULTI-DESTINATION TRAFFIC

- Multi-destination data is sent on a bi-directional distribution tree:
  - The root of a tree is a TRILL switch or a link (pseudo-node) determined by a separate election and represented by nickname.
  - The ingress RBridge picks the tree, puts the tree root nickname in the "egress nickname" slot, and sets the M bit in the TRILL Header.
- All the TRILL switches in a campus calculate the same trees.
- All trees reach every TRILL switch in the campus.

MULTI-DESTINATION TRAFFIC

- Multi-destination TRILL Data frames are more dangerous than unicast because they can multiply at fork points in the distribution tree.
  - So, in addition to the Hop Count, a Reverse Path Forwarding Check is performed. This discards the frame if, for the ingress and tree, it seems to be arriving on the wrong port.
  - To reduce the RPFC state, ingress RBridges can announce which tree or trees they will use.

MULTI-DESTINATION TRAFFIC

- As a TRILL Data frame is propagated on a distribution tree, its distribution can be pruned by VLAN and by multicast group since it is not useful to send a frame down a tree branch if
  - There are no end stations downstream in the VLAN of the frame, or
  - The frame is multicast and there is no multicast listener or multicast router downstream.

TRILL Nicknames

- TRILL switches are identified by 6-byte IS-IS System ID and by 2-bytes nicknames.
- Nicknames can be configured but by default are auto-allocated. In case of collisions, the lower priority RBridge must select a new nickname.
- Nicknames:
  - Saves space in headers.
  - An RBridge can hold more than one nickname so that
    - It can be the root of more than one different distribution tree.
    - May be used to distinguish frames following traffic engineered routes versus least cost routes.

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.

RBRIDGES & ACCESS LINKS

- You can have multiple TRILL switches on a link with one or more end stations.

- The elected Designated RBridge is in charge of the link and by default handles end station traffic. But to load split, it can assign end station VLANs to other RBridges on the link.

TRILL Cloud

RB1   RB2   RB3

B1   B2   B3
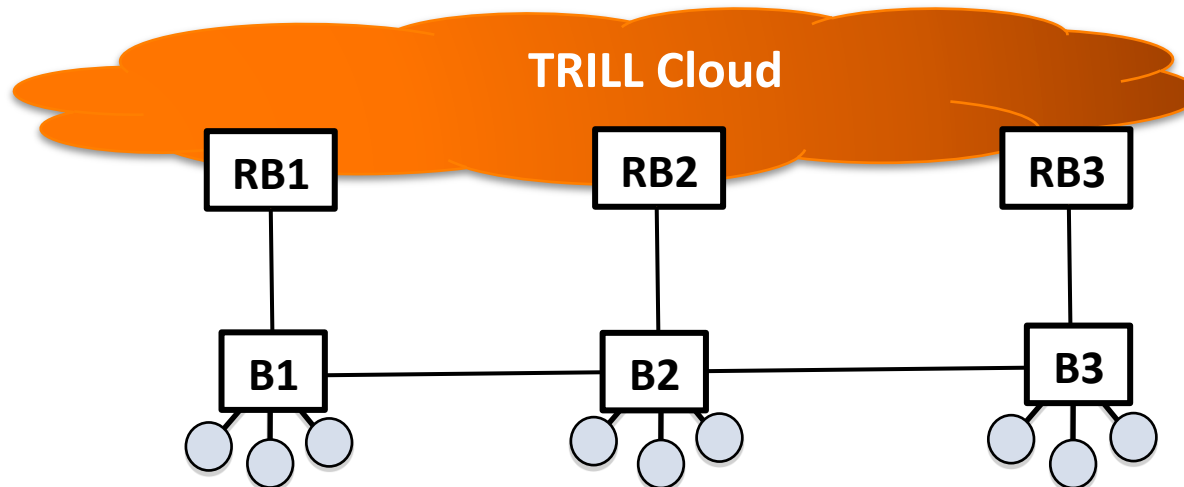
MAC ADDRESS LEARNING

- By IS-IS all TRILL switches in the campus learn about and can reach each other but what about reaching end station MAC addresses?

  - By default, TRILL switches at the edge (directly connected to end stations) learn attached VLAN/MAC addresses from data as bridges do.

  - Optionally, MAC addresses can be passed through the control plane.

  - MAC addresses can be statically configured or learned from Layer 2 registration protocols such as Wi-Fi association or 802.1X.

  - Transit TRILL switches do not learn end station addresses.

69

MAC ADDRESS LEARNING

- Data Plane Learning
  - From Locally Received Native Frames
    - { VLAN, Source Address, Port }
  - From Encapsulated Native Frames
    - { Inner VLAN, Inner Source Address, Ingress RBridge }
    - The Ingress RBridge learned is used as egress on sending
- Via

  1. Optional End Station Address Distribution Information (ESADI) control plane protocol
  2. Via Layer-2 Registration protocol(s)
  3. By manual configuration
     - { VLAN, Address, RBridge nickname }

70

ESADI

- The optional End Station Address Distribution Information (ESADI) protocol:
  - Provides a VLAN/tenant scoped way for an RBridge to distribute control plane information about attached End Stations to other RBridges.
  - Highly efficient transmission because information is tunneled through transit RBridges encapsulated as if it was normal data.
  - Intended for use for attachment data that is either secure or that changes rapidly.
    - The source RBridge selects which addresses it wants to distribute through ESADI.
    - There is no particular advantage in using ESADI for large amounts of information learned from the data plane.
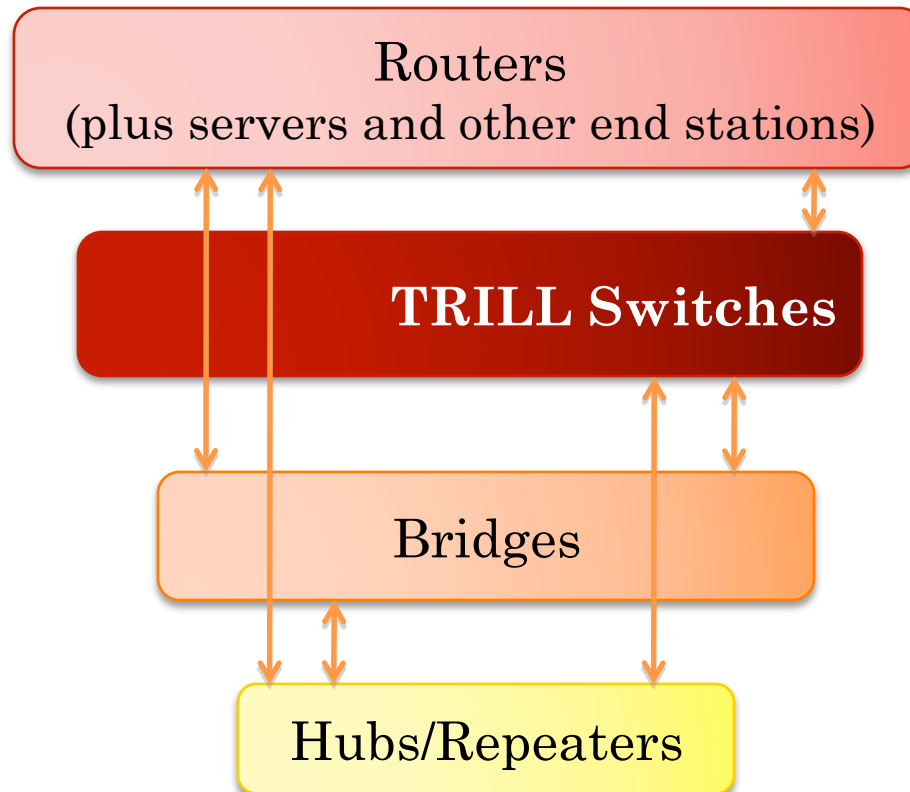
CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

Peering: Are TRILL Switches Bridges or Routers?

- Really, they are a new species, between bridges and routers:

```
┌─────────────────────────────────────────┐
│                 Routers                   │
│   (plus servers and other end stations)   │
└─────────────────────────────────────────┘

        ┌─────────────────────────────────┐
        │        TRILL Switches           │
        └─────────────────────────────────┘

        ┌─────────────────────────────────┐
        │            Bridges              │
        └─────────────────────────────────┘

             ┌───────────────────┐
             │  Hubs/Repeaters   │
             └───────────────────┘
```

Peering

- Direct Connection

Peers

Device - - - - - - - - - **Device**

Connection

Peering

- Former Situation

Router / End Station — — Peers — — **Router / End Station**
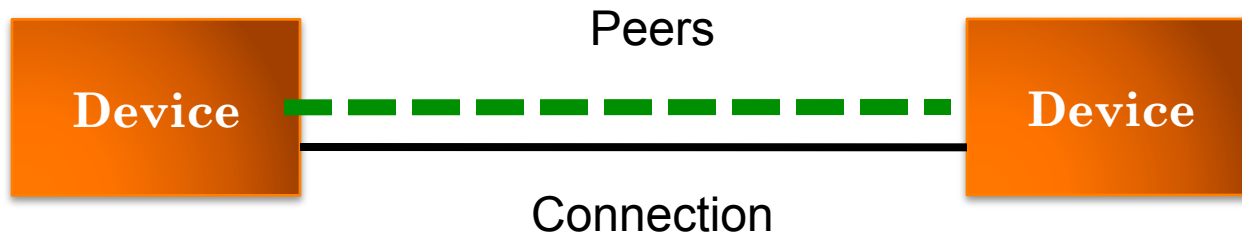
Bridge - - - ✕ - - - **Bridge(s)**

Non-Peers
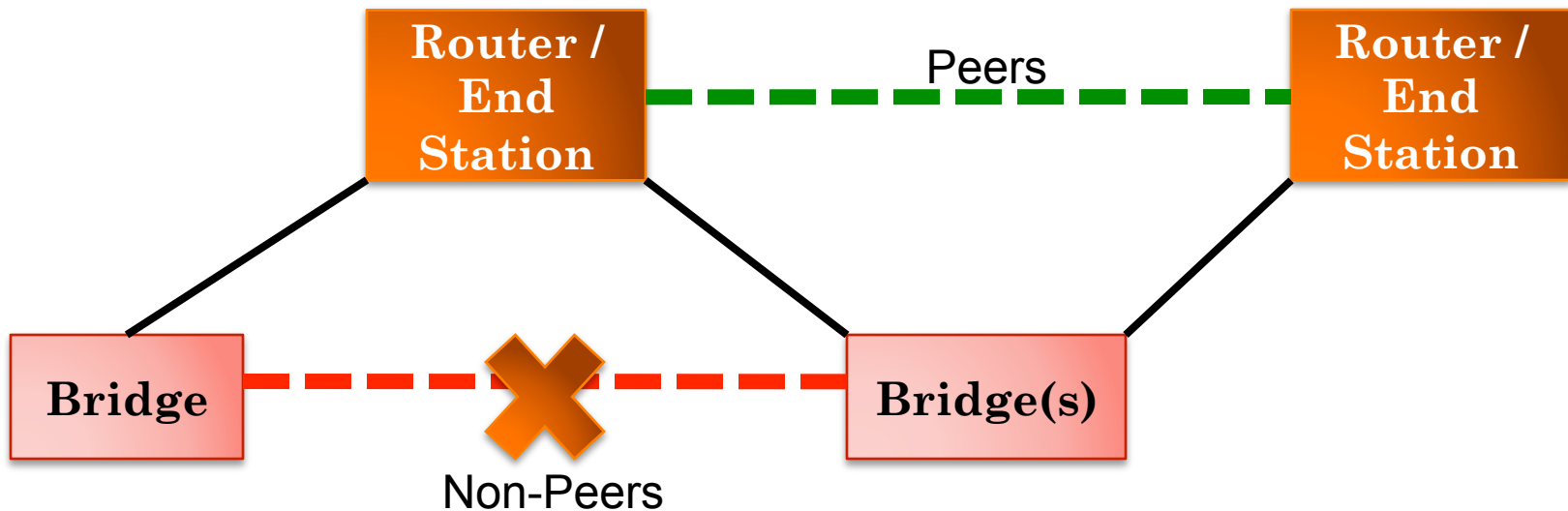
Peering

- Former Situation
  - Or perhaps

Peering

- With RBridges
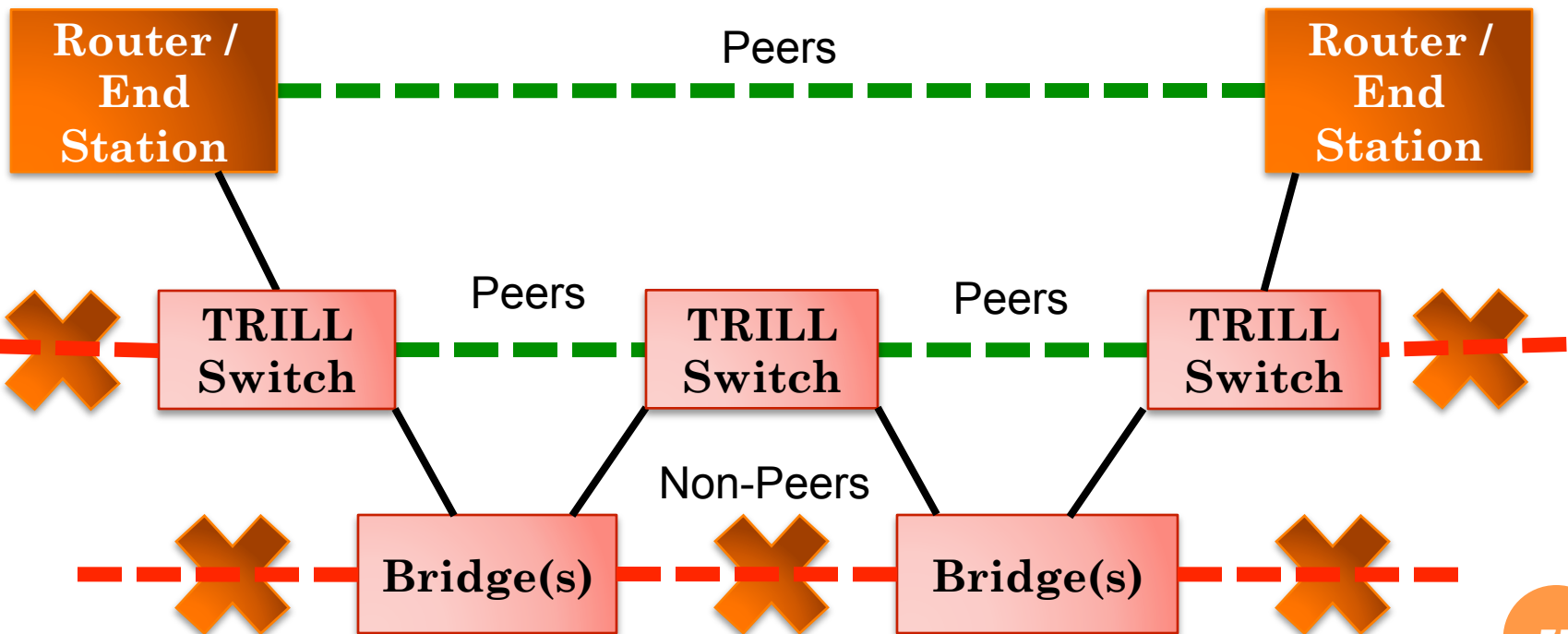
Router / End Station	— Peers —	Router / End Station

TRILL Switch — Peers — TRILL Switch — Peers — TRILL Switch

Bridge(s) — Non-Peers — Bridge(s)

CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- <u>TRILL Support of DCB</u>
- TRILL OAM
- TRILL Products
- Standardization and References

TRILL Support of DCB

- The goal is "loss-less" Ethernet. That is, no loss due to queue overflow

- Basic Ethernet PAUSE "works", but is a very blunt instrument
  - Interference with loss dependent flow control such as TCP
  - Blocking of high priority control frames
  - Congestion spreading

TRILL Support of DCB

- "Data Center Ethernet"

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

    } Data Center Bridging

    4. TRILL

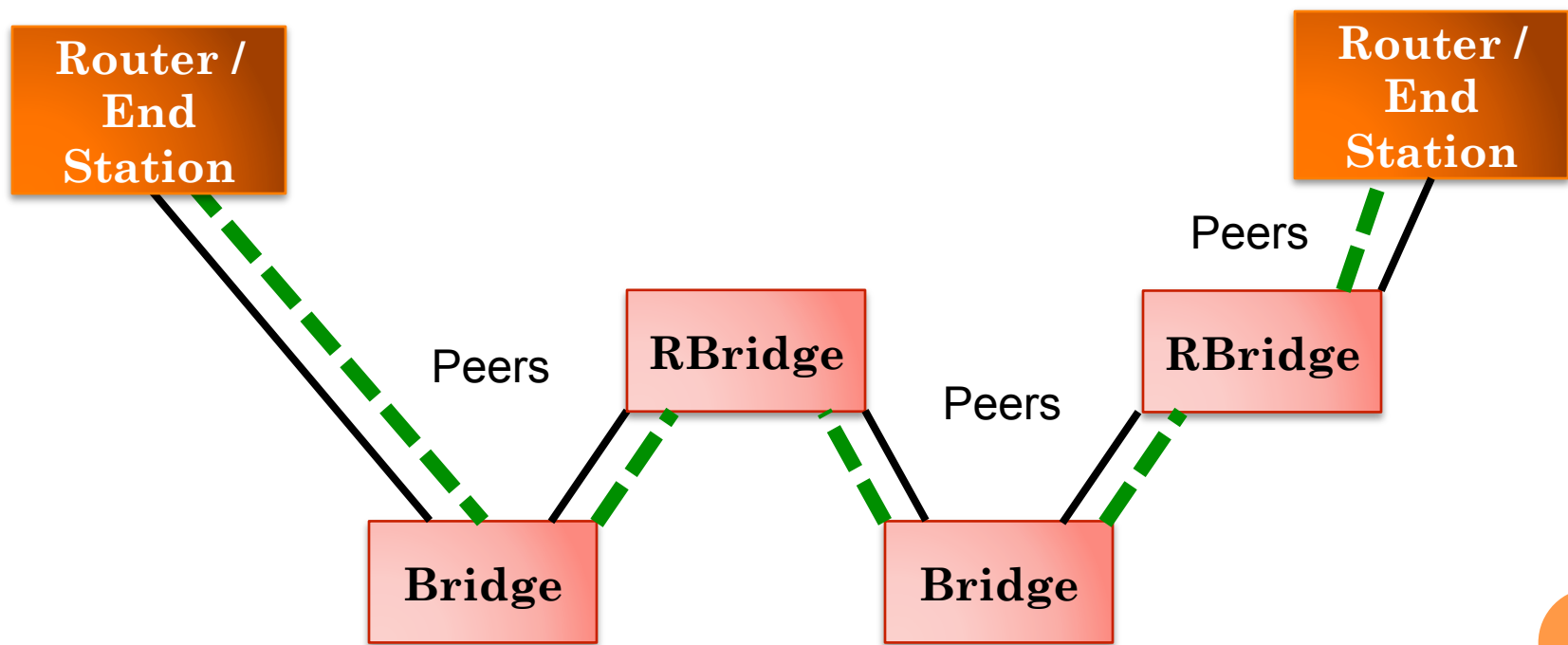# TRILL Support of DCB

- Answer 1:
  - Consider different frame priorities as different pipes
    - 802.1Qbb: Separate PAUSE per priority
    - 802.1Qaz: Ability to allocate bandwidth between these pipes
- Answer 2:
  - Provide back pressure on the origin of congesting flows
    - 802.1Qau: Congestion Notification (CN)
- 802.1Qbb is more commonly implemented and you would need it for surges in any case.

TRILL Support of DCB

- Answer 1: Consider different frame priorities as different pipes
  - 802.1Qbb: Separate PAUSE per priority
    - Don't enable for priorities where urgent control frames are sent or where loss dependent flow control is in use
    - Enable for priorities where loss-less flow is more important.
  - 802.1az: Ability to allocate bandwidth between these pipes
    - Highest priority frames not restricted
    - Remainder of bandwidth can be carved up and frames can be selected in preference to "higher priority" frames if they have not used the allocation for their pipe.
  - The above are implemented in port queuing. Can be applied to bridges, RBridges, routers, end stations.
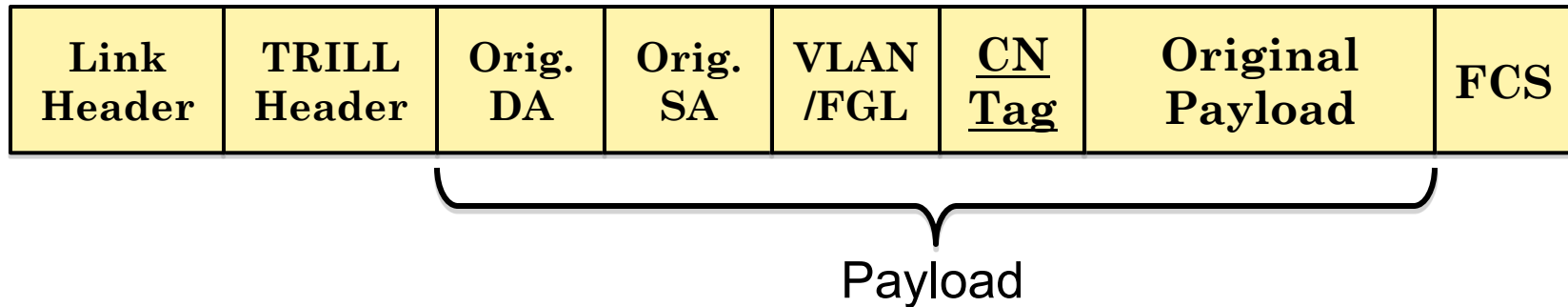
TRILL Support of DCB

- For purposes of Data Center Bridging, all nodes with queues are considered peers:

TRILL Support of DCB

- Answer 2: Congestion Notification (CN): Provide back pressure on the origin of congesting flows
  - When queue depth exceeds a bound, send a Congestion Notification Message CNM back to source MAC address in the congesting frame's VLAN
  - Enabled per priority. (CNM itself usually priority 6.)
  - Frames can be labeled with a CN tag for more fine grained flows
  - Mostly implemented in port logic
- In TRILL a CN tag, if present, goes inside the encapsulated frame and a CNM is just a native frame, except for one corner case.

84

TRILL Support of DCB

Link Header	TRILL Header	Orig. DA	Orig. SA	VLAN /FGL	CN Tag	Original Payload	FCS

Payload

- However, RBridges have to handle CNMs generated by TRILL ignorant bridges between RBridges. Such a CNM will be initially addressed to the previous hop RBridge, not the original end station.

TRILL Support of DCB

Origin → **RBridge** → **Bridge(s)** → **RBridge**

Adjusted CNM   **CNM**
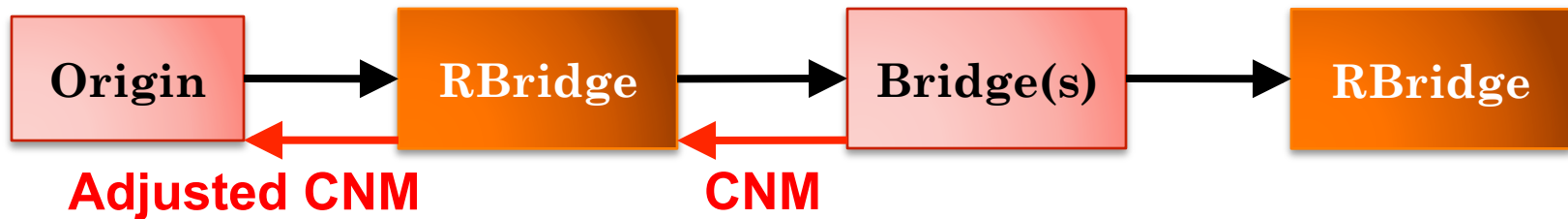
- Previous hop RBridge has to adjust the CNM so that it goes back to the origin end station.

- Note: All of the Data Center Bridging facilities depend on appropriate engineering, limited delay bandwidth product, etc., to actually provide "loss-less" service.

CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

TRILL OAM

- **SNMP**
  - Used primarily to read the status and configuration of the TRILL switches but can be used to set configuration parameters.
- **BFD over TRILL**
  - Bound to TRILL port at the transmitting TRILL router. Primarily used for rapid one-hop failure detection but multi-hop supported.
- **TRILL OAM**
  - Operates between TRILL switches and is focused on testing TRILL Data paths (both fault and performance management).

OAM DOCUMENTS STATUS

- **SNMP**
  - RFC 6850, "Definitions of Managed Objects for RBridges" (MIB)
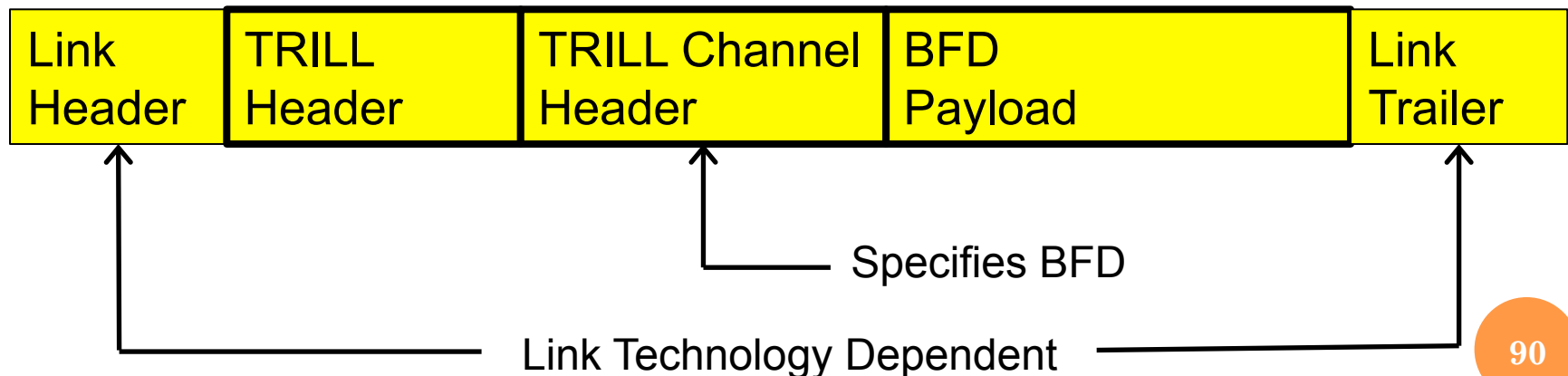- **BFD over TRILL (data center)**
  - In RFC Editor's queue:
    - draft-ietf-trill-rbridge-bfd-07.txt
    - draft-ietf-trill-rbridge-channel-08.txt
- **TRILL OAM (carrier grade)**
  - RFC 6905: TRILL OAM Requirements
  - In RFC Editor's queue:
    - draft-ietf-trill-oam-framework-04.txt (Framework)
  - draft-ietf-trill-oam-fm-02.txt (Fault Management)
  - draft-ietf-trill-loss-delay-02.txt (Performance)

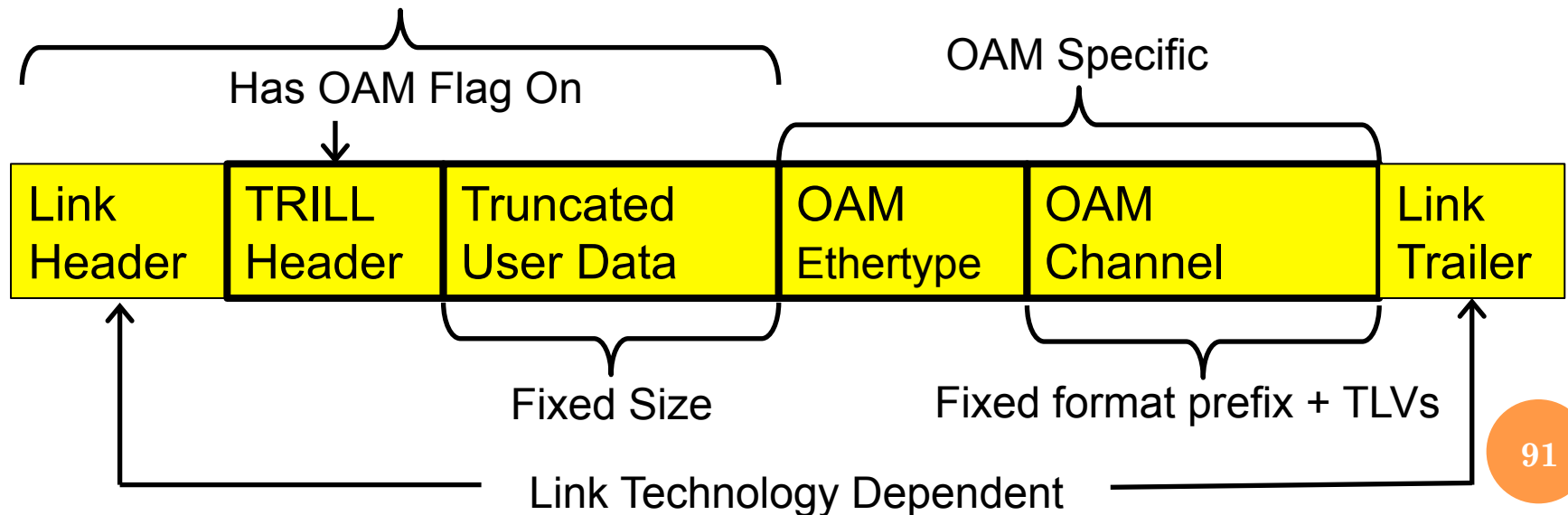# TRILL BFD Packet Format

- The BFD standard does not specify an envelope. One must be specified for each technology using BFD.

- The TRILL BFD envelope uses the RBridge Channel facility, a general method for sending typed messages between TRILL routers.

Link Header	TRILL Header	TRILL Channel Header	BFD Payload	Link Trailer

Specifies BFD

Link Technology Dependent

90

TRILL OAM Packet Format

- Because TRILL OAM packets must be able to follow the same paths and get the same processing as TRILL Data packets, their format is very similar.

Same as user TRILL Data except for OAM Flag

Has OAM Flag On

OAM Specific

Link Header	TRILL Header	Truncated User Data	OAM Ethertype	OAM Channel	Link Trailer

Fixed Size

Fixed format prefix + TLVs

Link Technology Dependent

91

CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

TRILL PRODUCTS

- Announced TRILL Standard based switches:
  - Broadcom – StrataXGS Trident (BMC5680)
  - Cisco – Nexus
  - HP/H3C – S5830V2, S5820V2
  - Huawei – Cloud Engine 5800, 6800, 12800
  - Mellanox – SwitchX
  - Ruijie Networks – RG-S12000
  - ZTE – Big Matrix 9900 Series

PRE-STANDARD PRODUCTS

- "Pre-standard" products deployed

  - Cisco FabricPath
    - Uses TRILL control plane but different data plane
    - 2,200+ installations
  - Brocade VCS
    - Uses TRILL data plane but FSPF (Fiber Shortest Path First) control plane
    - 1,600+ installations

94

TRILL PLUGFESTS

- Cumulative UNH IOL TRILL interoperability event participants:

  - Broadcom
  - Extreme Networks
  - HP/H3C Networking
  - Huawei Technologies
  - Ixia
  - JDSU
  - Oracle
  - Spirent

TRILL Silicon

- Here are six publicly known independent silicon implementations of the TRILL Fast Path. In some cases the vendor has multiple different chips supporting TRILL.

  - Broadcom – merchant silicon

  - Brocade – products

  - Cisco – products

  - Fulcrum – merchant silicon

  - Marvell – merchant silicon

  - Mellanox – merchant silicon

TRILL OPEN SOURCE

Three Open Source Implementations

1. **Oracle: TRILL for Solaris**
   - TRILL ships as part of Solaris 11

2. **VirtuOR:** www.virtuor.fr
   - http://sourceforge.net/p/opentrill/wiki/Home/

3. TRILL Port to Linux:
   National University of Sciences and Technology (NUST), Islamabad, Pakistan
   - Muhammad Mohsin Sardar
     mohsin.sardar@seecs.edu.pk
   - http://wisnet.seecs.nust.edu.pk/projects/trill/index.html

CONTENTS

- What is TRILL?
- TRILL Features
- TRILL History
- Two TRILL Examples
- TRILL Packet Headers
- Step-by-Step Processing Example
- Fine Grained Labeling

- How TRILL Works
- Peering and Layers
- TRILL Support of DCB
- TRILL OAM
- TRILL Products
- Standardization and References

STANDARDIZATION STATUS

- The TRILL protocol RFCs (**bold** = stds track)
  - RFC 5556, "TRILL Problem and Applicability"
  - **RFC 6325, "RBridges: TRILL Base Protocol Specification"**
  - **RFC 6326, "TRILL Use of IS-IS"**
  - **RFC 6327, "RBridges: Adjacency"**
  - **RFC 6361, "TRILL over PPP"**
  - **RFC 6439, "RBridges: Appointed Forwarders"**
  - RFC 6847, "FCoE over TRILL"
  - **RFC 6850, "Definitions of Managed Objects for RBridges" (MIB)**
  - RFC 6905, "TRILL OAM Requirements"
  - RFC 7067, "TRILL Directory Assistance Problem and High-Level Design Proposal"

STANDARDIZATION STATUS

- Documents that are fully approved and in the RFC Editor's Queue. These are expected to be published as RFCs within a few months:
  - "TRILL: Fine Grained Labeling:
    - https://datatracker.ietf.org/doc/draft-ietf-trill-fine-labeling/
  - "TRILL: BFD Support"
    - https://datatracker.ietf.org/doc/draft-ietf-trill-rbridge-bfd/
  - "Transport of TRILL by Pseudo-wires"
    - http://datatracker.ietf.org/doc/draft-ietf-trill-o-pw/
  - "TRILL: RBridge Channel Support"
    - https://datatracker.ietf.org/doc/draft-ietf-trill-rbridge-channel/
  - "TRILL: Edge Directory Assistance Framework"
    - https://datatracker.ietf.org/doc/draft-ietf-trill-directory-framework/
  - "TRILL: Header Extension"
    - https://datatracker.ietf.org/doc/draft-ietf-trill-rbridge-extension/

STANDARDIZATION STATUS

- Documents that are fully approved and in the RFC Editor's Queue continued (documents on this slide are all in the nature of updates):
  - "TRILL use of IS-IS"
    - http://datatracker.ietf.org/doc/draft-ietf-isis-rfc6326bis/
  - "TRILL: Adjacency"
    - http://datatracker.ietf.org/doc/draft-ietf-trill-rfc6327bis/
  - "TRILL: Clarifications, Corrections, and Updates"
    - https://datatracker.ietf.org/doc/draft-ietf-trill-clear-correct/

STANDARDIZATION STATUS

- Non-IETF Assignments:
  - Ethertypes assigned by IEEE:
    - TRILL Data: 0x22F3
    - TRILL IS-IS: 0x22F4
    - TRILL Fine Grained Labeling: 0x893B
    - RBridge Channel: 0x8946
  - Block of multicast addresses assigned to TRILL by IEEE:
    - 01-80-C2-00-00-40 to 01-80-C2-00-00-4F
  - TRILL NLPID (Network Layer Protocol ID) assigned from ISO/IEC: 0xC0

102

MORE TRILL REFERENCES

- TRILL Introductory Internet Protocol Journal Article:

  - http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_14-3/143_trill.html

- The first paper: Perlman, Radia. "Rbridges: Transparent Routing", Proceeding Infocom 2004, March 2004.

  - http://www.ieee-infocom.org/2004/Papers/26_1.PDF

SOME TRILL FUTURES

- Directory Assisted Edge:
  - In data centers, the location of all MAC and IP address and virtual machines is typically known through orchestration. This information can be provided to edge TRILL switches by a directory.
  - Directory information can be Pushed or Pulled and can reduce or eliminate ARP (IPv4), ND (IPv6), and unknown unicast MAC flooding.
- Active-active at the edge
- Multi-level and Multi-topology support
- Standard TRILL over IP specification

END

Donald E. Eastlake 3rd

Co-Chair, TRILL Working Group
Principal Engineer, Huawei

d3e3e3@gmail.com

105

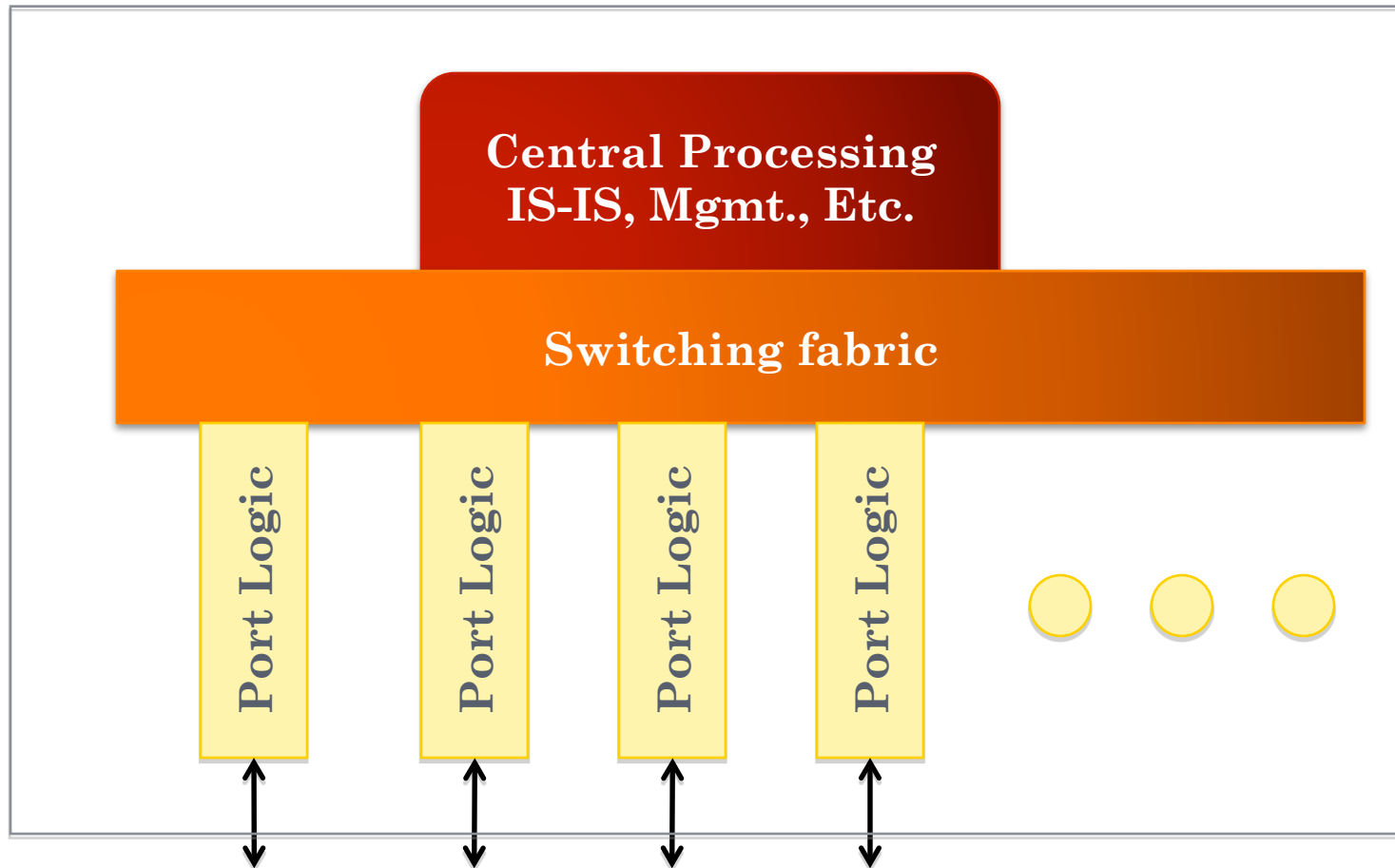# Backup Slides

Donald E. Eastlake 3rd

Co-Chair, TRILL Working Group
Principal Engineer, Huawei
d3e3e3@gmail.com

106

Algorhyme (Spanning Tree)

- I think that I shall never see
- A graph more lovely than a tree.
- A tree whose crucial property
- Is loop-free connectivity.
- A tree that must be sure to span
- So packets can reach every LAN.
- First, the root must be selected.
- By ID, it is elected.
- Least-cost paths from root are traced.
- In the tree, these paths are placed.
- A mesh is made by folks like me,
- Then bridges find a spanning tree.
- - By Radia Perlman

Structure of an RBridge

**Central Processing
IS-IS, Mgmt., Etc.**

Switching fabric

Port Logic

Port Logic

Port Logic

Port Logic

Links to other devices. Could be 802.3 (Ethernet), 802.11 (Wi-Fi), PPP, …

Structure of an RBridge Port

Assumes an Ethernet (802.3) link.

Information ⟷
Frame & Info ⟷

RBridge: Higher Level Processing (see next slide)

RBridge: High Level Control Frame Processing (BPDU, VRP)

802.1Q Port VLAN Processing

← ISS

802.1/802.3 Low Level Control Frame Processing, Port/Link Control Logic

802.3 Physical Interface

802.3 Link

Structure of an RBridge Port

Information
Frame & Info

RBridge: Inter-port Forwarding, IS-IS, Management, Etc.

TRILL Data

Encapsulation / Decapsulation Processing

TRILL IS-IS Hello Processing

Other TRILL IS-IS

TRILL Data

Appointed Forwarder and Inhibition Logic

TRILL IS-IS Hellos

Native frames

Lower Level Processing   **(see previous slide)**

110

Loop Avoidance

- All possible forms of looping behavior within a TRILL campus can be grouped into one of three classes:
  - TRILL Data Packet Loops, in which the packet would remain TRILL encapsulated.
  - Hybrid TRILL Data / Native Frame Loops, in which the frame is repeatedly encapsulated and decapsulated.
  - Native Frame Loops, in which the frame is never encapsulated.
    - Since TRILL always encapsulates data, if you have this problem, no RBridges are involved so it is not TRILL's fault.

111

Loop Avoidance

- TRILL Data Packet Loops:

  - Known unicast TRILL Data packets have a hop count and are always unicast to the next hop RBridge towards their destination.

  - Multi-destination TRILL Data packets must be received on a port which is part of their distribution tree, the ingress RBridge nickname and input port must pass a Reverse Path Forwarding Check, and they have a hop count.

Loop Avoidance

- Hybrid TRILL Data / Native Frame Loops:
    - Such a loop would require, at some point around the loop, that a TRILL Data packet be decapsulated onto a link by one RBridge and then picked up and re-encapsulated by another RBridge.
    - TRILL takes great care to minimize the probability of there being two uninhibited appointed forwarders on the same link for the same VLAN.
        - Under certain conditions, an RBridge appointed forwarder is inhibited from accepting or sending native frames. This only affects native frames. An RBridge port is never inhibited or blocked from sending or receiving TRILL Data or TRILL IS-IS frames except by very low level link flow control mechanisms such as PAUSE or if the port has been manually configured as disabled.

PDU Types

- PDU Type Names Used in TRILL
  - TRILL Packets
    - <u>TRILL IS-IS Packets</u>– Used for routing control between RBridges.
    - <u>TRILL Data Packets</u>– Used for encapsulated native frames.
  - <u>Layer 2 Control Frames</u> – Bridging control, LLDP, LACP, etc. Never forwarded by RBridges.
  - <u>Native Frames</u> – All frames that are not TRILL or Layer 2 Control Frames.

PDU Types, More Detail

- <u>TRILL Packets</u>– Ethernet local link encoding

  - TRILL Data packets – Have the TRILL Ethertype and are sent to a unicast address or, if multi-destination, to the All-RBridges multicast address.
  - TRILL IS-IS Packets – Have the L2-IS-IS Ethertype and are sent to the All-IS-IS-RBridges multicast address.
  - (TRILL Other packets– sent to a TRILL multicast address but not a TRILL Data or TRILL IS-IS frame. Not currently used.)

PDU Types, More Detail

- <u>Layer 2 Control Frames</u> – Destination Address is 01-80-C2-00-00-00 to 01-80-C2-00-00-0F or 01-80-C2-00-00-21

  - High Level – BPDU (01-80-C2-00-00-00) & VLAN Registration (01-80-C2-00-00-21)

    - RBridges handle these differently from bridges. The spanning tree protocol never runs through an RBridge but an RBridge port is not prohibited from participating in spanning tree as a leaf node.

  - Low Level – all other control frames

    - An RBridge port may implement other Layer 2 protocols such as LLDP (Link Layer Discovery Protocol), 802.1AX (Link Aggregation), 802.1X (Port Based Access Control), 802.1AE (MAC Security), and the like.