# RBRIDGES

## LAYER 2 FORWARDING BASED ON LINK STATE ROUTING


Donald E. Eastlake 3rd

donald.eastlake@stellarswitches.com

# Contents

- Introduction
- Ethernet and Spanning Tree
- RBridge Features
- TRILL Encapsulation
- Are RBridges Bridges or Routers?
- How RBridges Work
- Structure of an RBridge
- Some Additional Details
- References

# DEFINITIONS

- RBridge – Routing Bridge
  - A device implementing the TRILL protocol, which performs Layer 2 bridging with link state routing.
- RBridge Campus –
  - A network of RBridges, links, and possibly intervening bridges bounded by end stations.
- TRILL –
  TRansparent Interconnection of Lots of Links
  - A standard being specified by the IETF (Internet Engineering Task Force) TRILL Working Group co-chaired by
    - Donald E. Eastlake 3rd, Stellar Switches
    - Erik Nordmark, Sun Microsystems

# Why/Who RBridges/TRILL?

- Why do RBridges/TRILL?
  - Provide optimum point-to-point forwarding with zero configuration.
  - Support multi-pathing of both unicast and multi-destination traffic.

- Who invented RBridges/TRILL?
  - Radia Perlman of Sun Microsystems, also the inventor of the Spanning Tree Protocol.

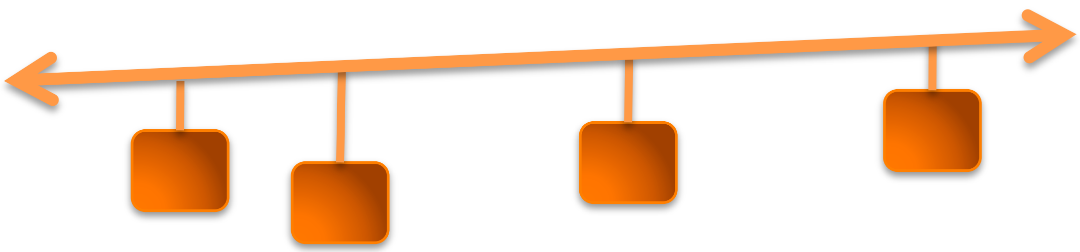
# CONTENTS

# Ethernet

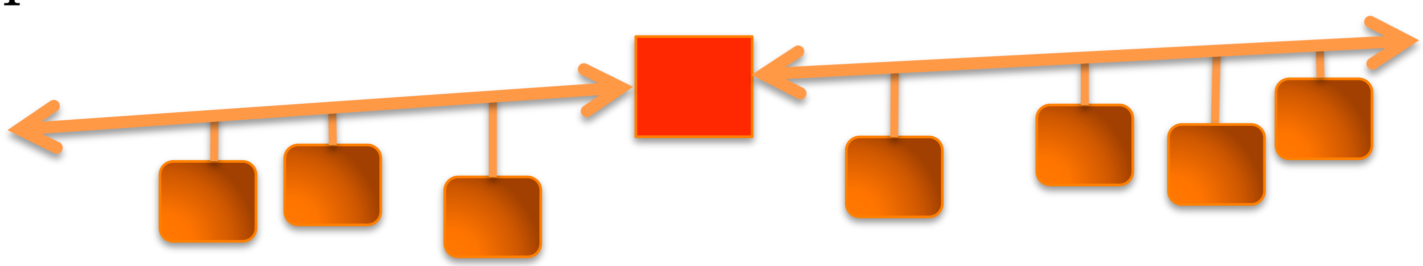
- Invented in the 1970s by Bob Metcalfe At Xerox
  - Carrier Sense Multiple Access Collision Detect (CSMA/CD)
- DIX (Digital, Intel, Xerox) agree around 1980
- IEEE Standardization started around 1983, completed in 1985
- Ever increasing speed for wired/optical-fiber:
  - <10Mbps 10Mbps
  - 100Mbps 1Gbps
  - 10Gbps
  - Under development: 40Gbps, 100Gbps

# Ethernet Local Area Network (LAN) Evolution

- Multi-access media



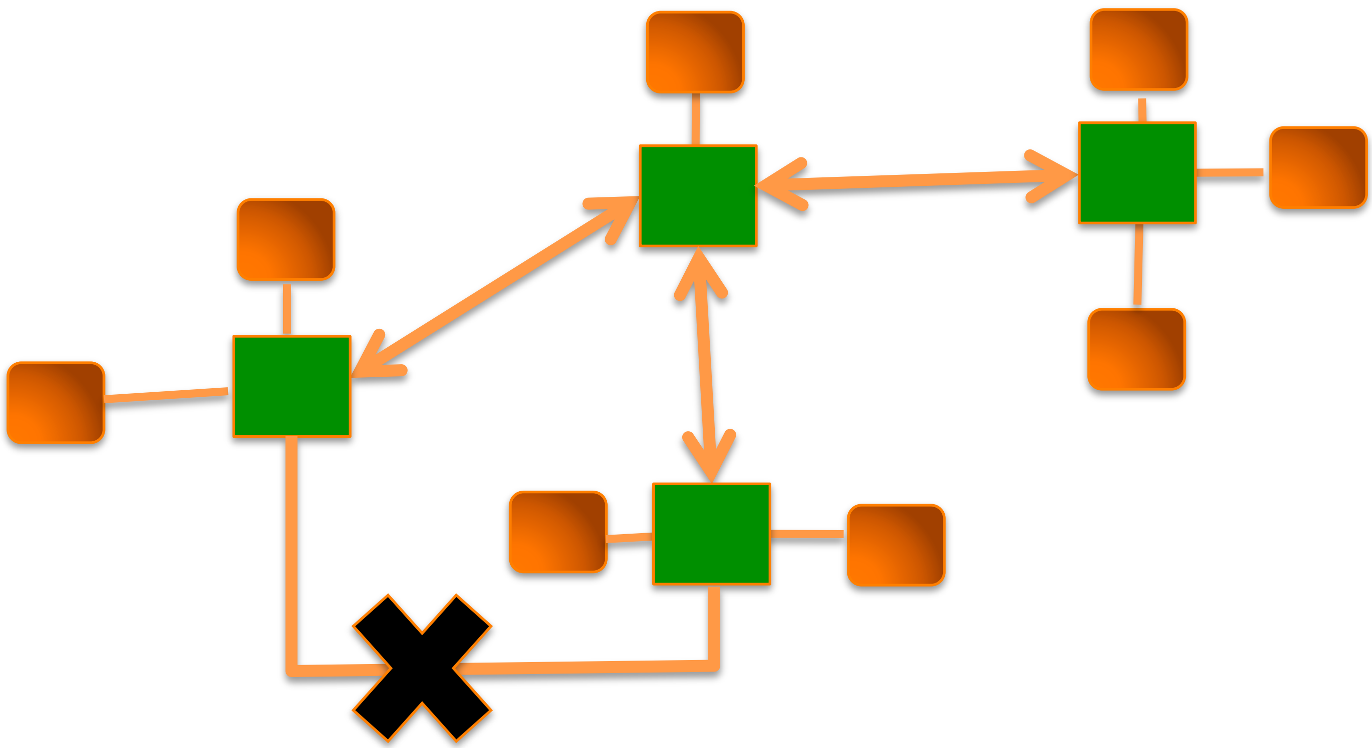
- Repeaters



- Hubs – full duplex
- Bridges, learning

# Ethernet Local Area Network (LAN) Evolution

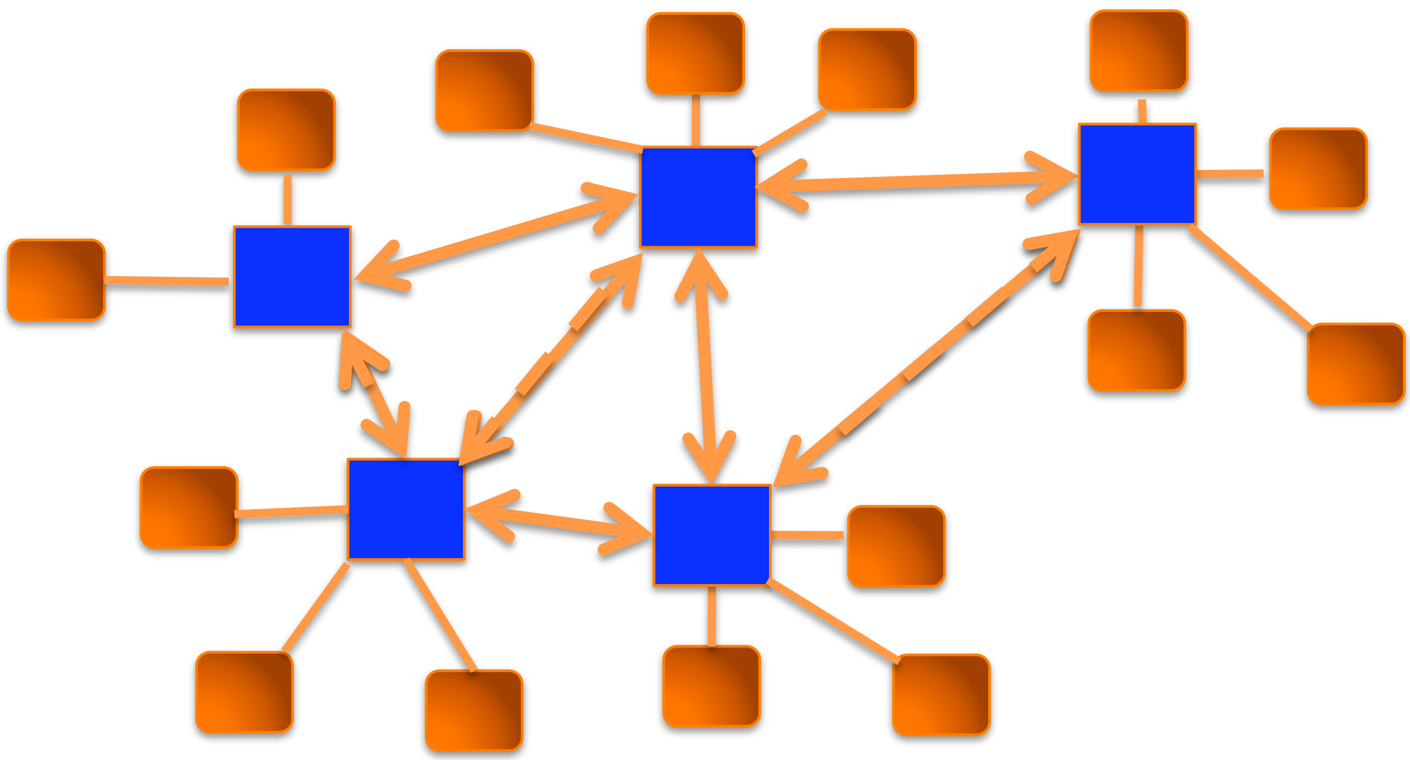
- Hubs – full duplex



- Bridges, learning

# Ethernet Local Area Network (LAN) Evolution

- Bridges
  - Spanning Tree Protocol invented by Radia in 1985
  - Address Learning and Forgetting



# Algorhyme

- 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.
- Radia Perlman

# Spanning Tree Difficulties

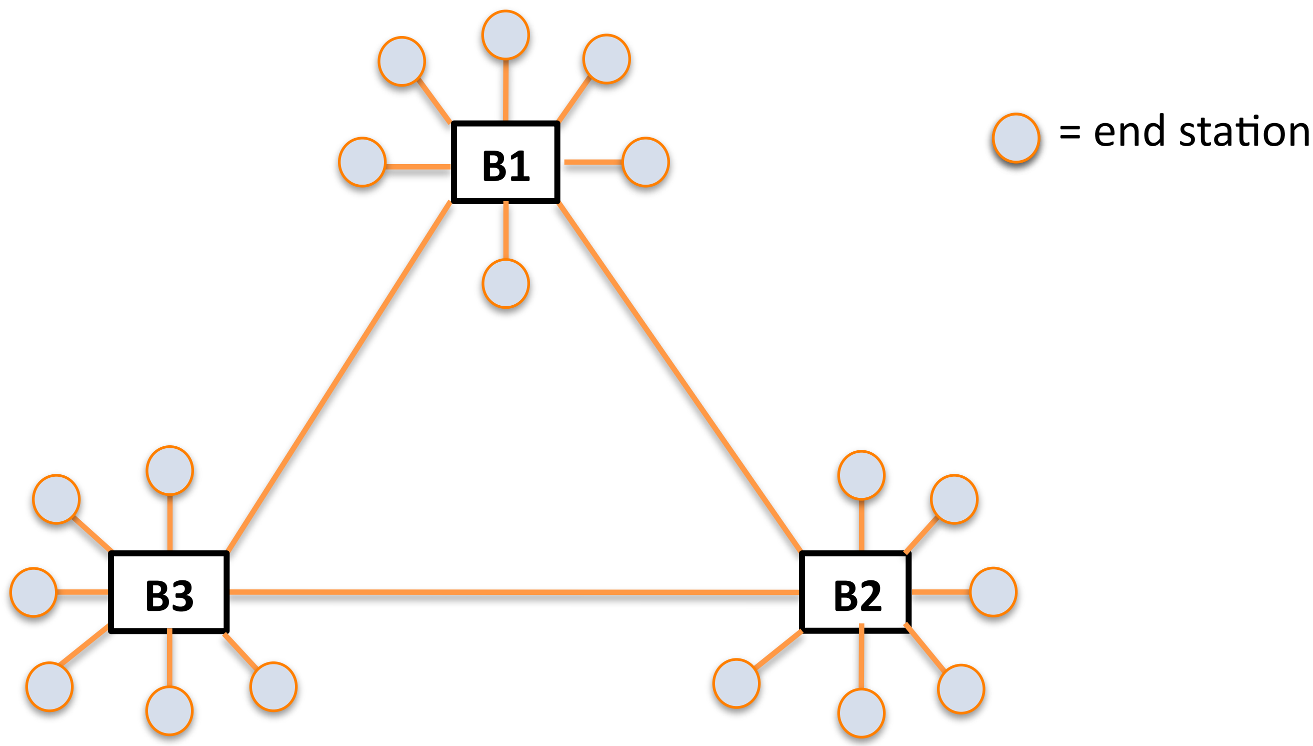
- The Spanning Tree Protocol makes a general mesh of connected bridges into a tree by disabling ports. This means that
  - traffic is concentrated on the remaining links, increasing congestion, and
  - traffic is not pair-wise shortest path but must follow whatever path is left after spanning tree blocks redundant paths.

# Spanning Tree Difficulties

- There is no hop count in Ethernet, which makes temporary loops more dangerous. Loops can appear with spanning tree due to
  - sufficient dropped spanning tree messages, or
  - the appearance of new connectivity without physical indication.
- Failover minimum time limitations for some failures.
- Connectivity changes can cause VLANs to partition.

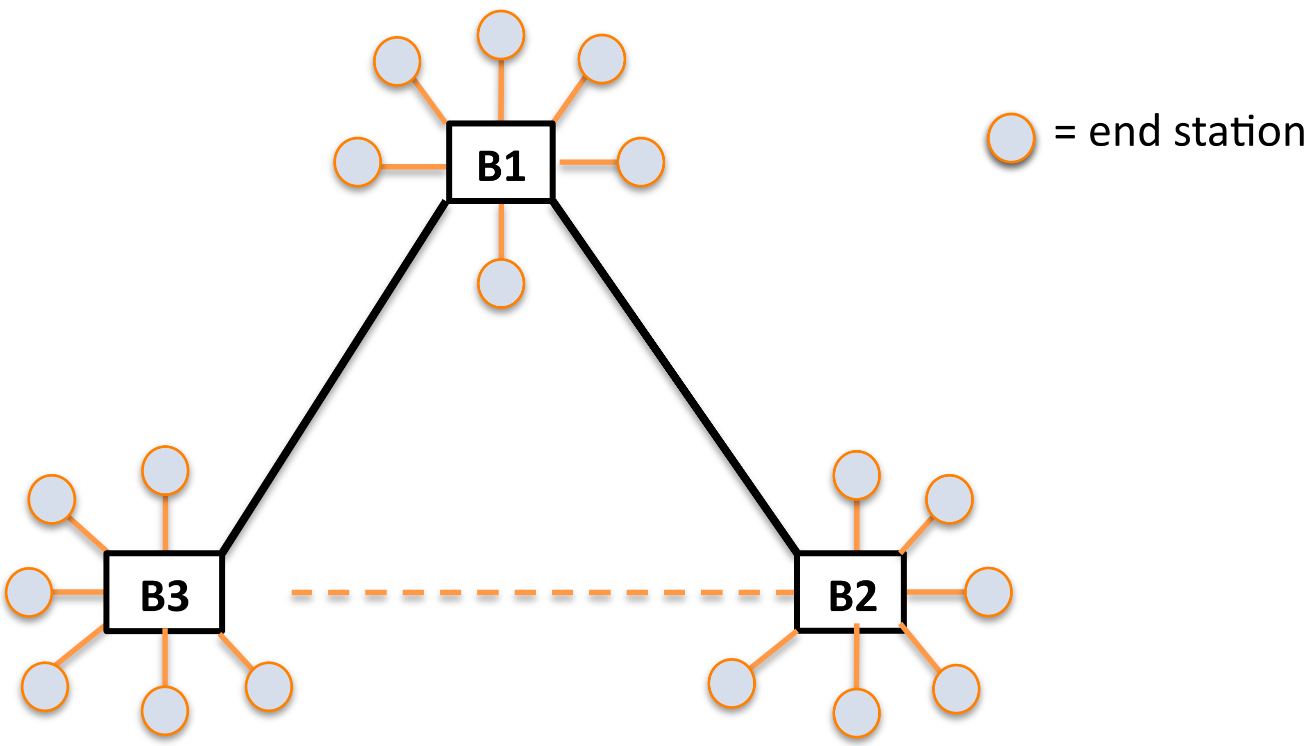
# Contents

# Optimum Point-to-Point Forwarding



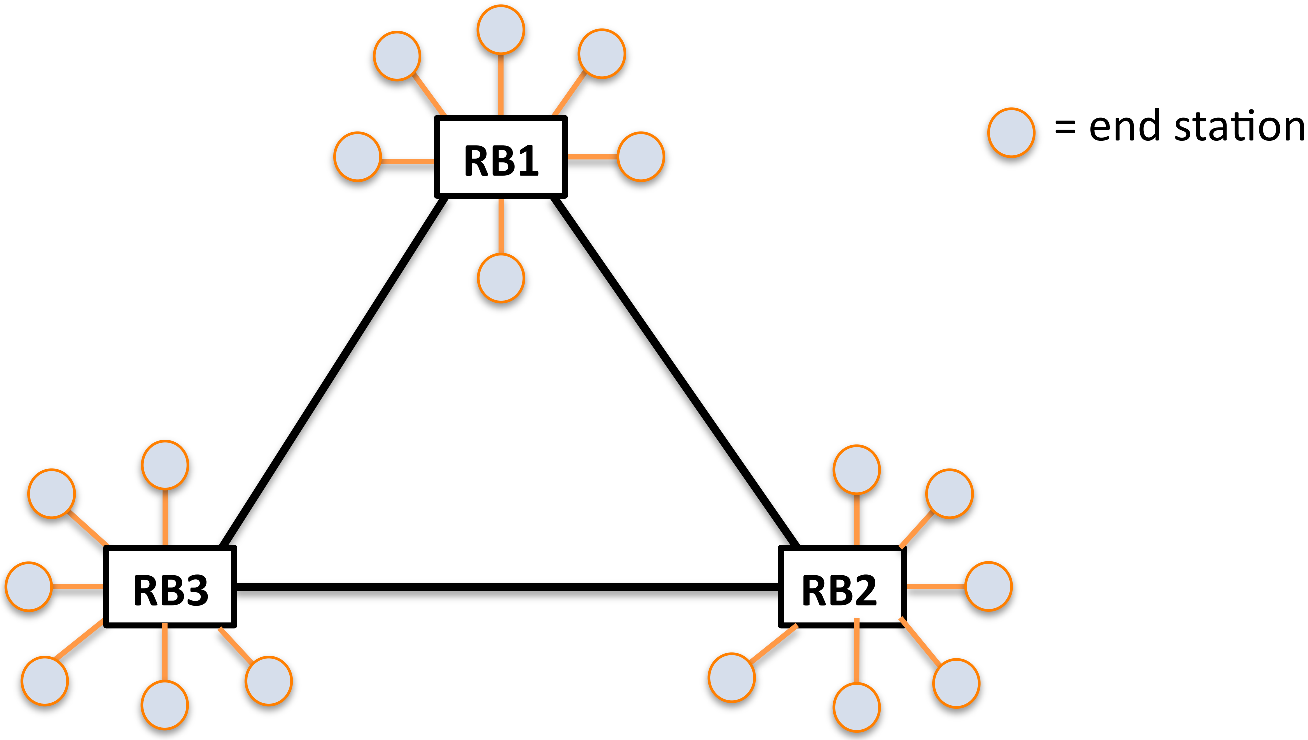

A three bridge network

# Optimum Point-to-Point Forwarding



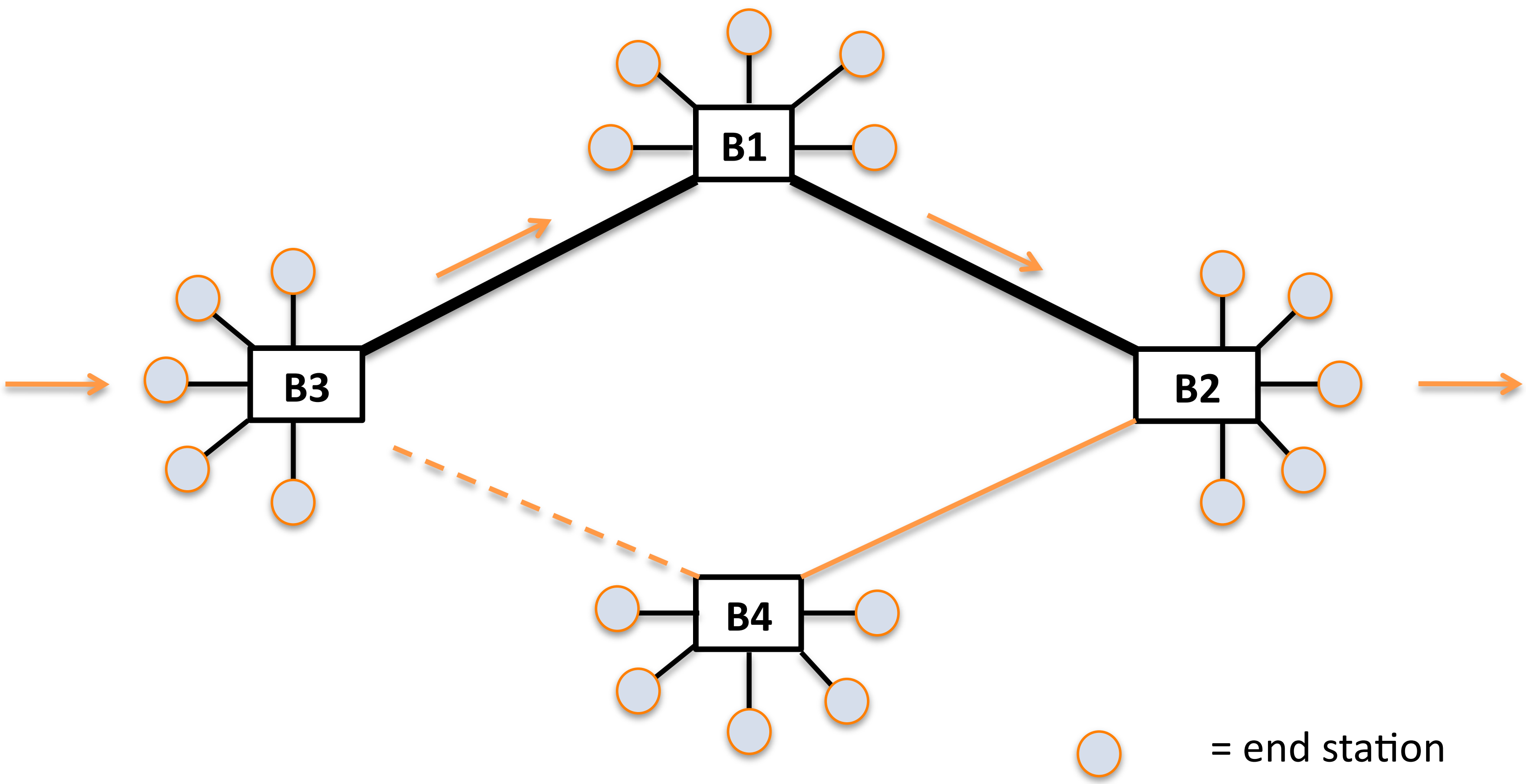

Spanning tree eliminates loops by disabling ports

# OPTIMUM POINT-TO-POINT FORWARDING



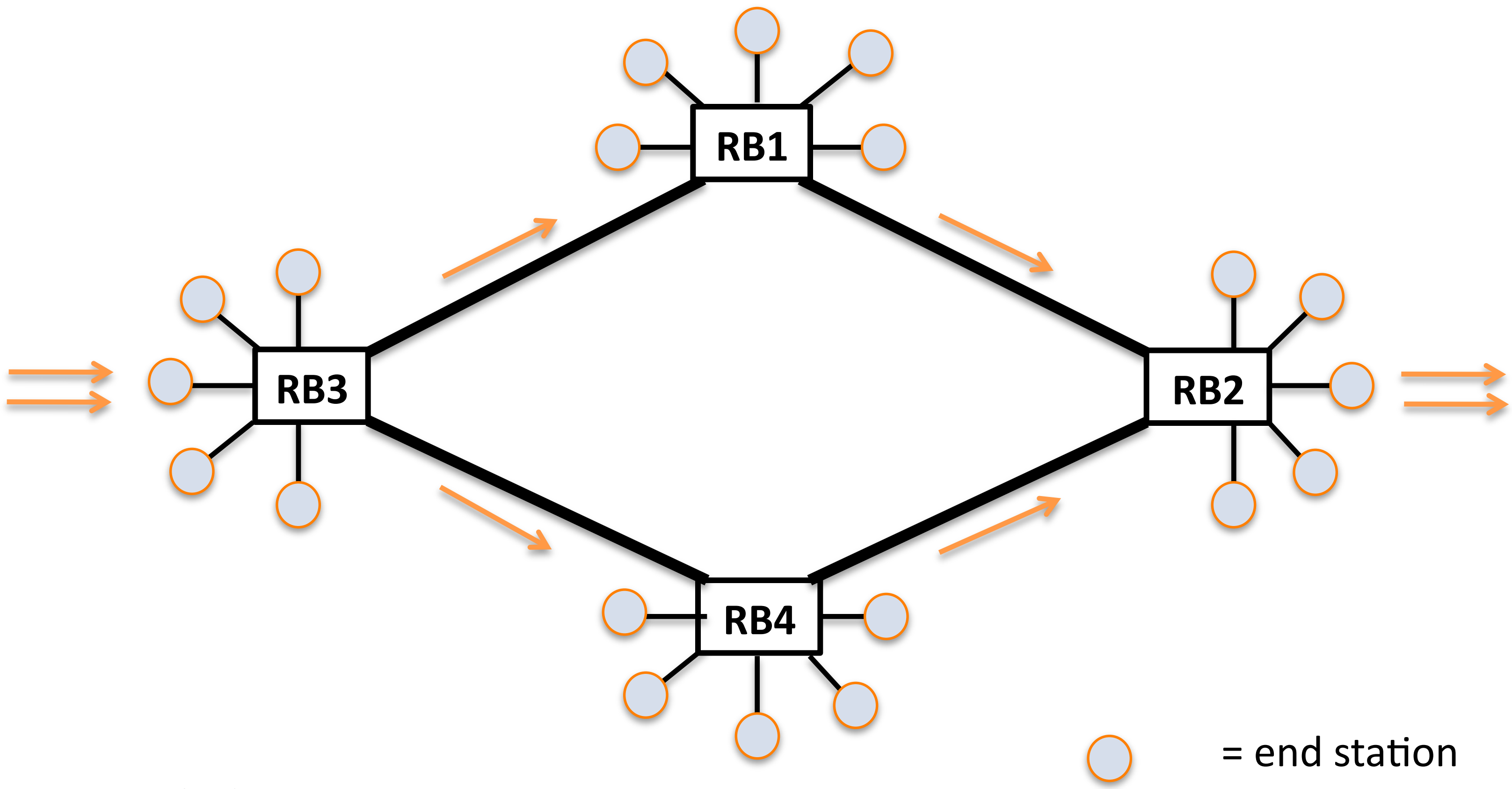

A three RBridge network: better performance using all facilities

# Multi-Pathing




Bridges limit traffic to one path

# Multi-Pathing




RBridges support
multi-path for higher throughput

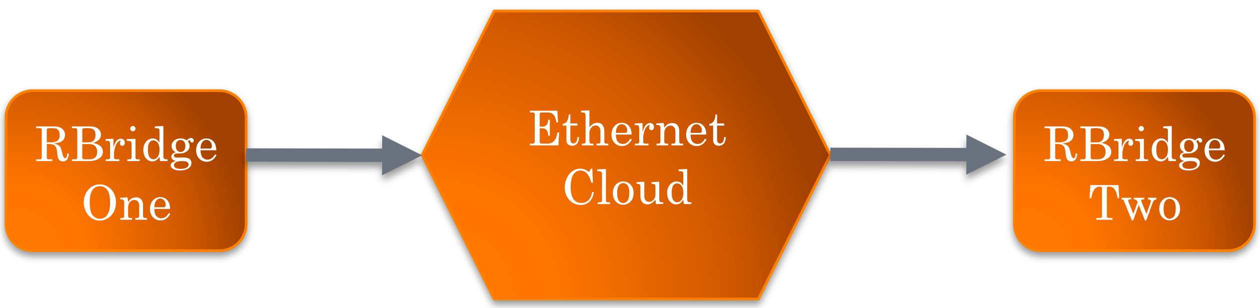
# Other RBridge Features

- Compatible with classic bridges. Can be incrementally deployed into a bridged LAN.
- 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.

# Contents

# The TRILL Encapsulation and Header

- Frames sent between RBridges are encapsulated inside a local link header, addressed from the local source RBridge to the local destination RBridge, and a TRILL header.



# The TRILL Encapsulation and Header

- Some reasons for encapsulation:
  - Provides a hop count to mitigate 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 VLAN tag for forwarding traffic between RBridges, independent of the original VLAN of the frame

# The TRILL Encapsulation and Header

- Assuming the link is Ethernet (IEEE 802.3) the encapsulation looks like:
  1. Outer Ethernet Header
     - Source RBridge One, Destination RBridge Two
  2. (Outer VLAN Tag)
  3. TRILL Header
  4. Inner Ethernet Header
     - Original Source and Destination Addresses
  5. Inner VLAN Tag
  6. Original Payload
  7. Frame Check Sequence (FCS)

# The TRILL Encapsulation and Header

- TRILL Header – 64 bits

| 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)

# CONTENTS

# Are RBridges Bridges or Routers?

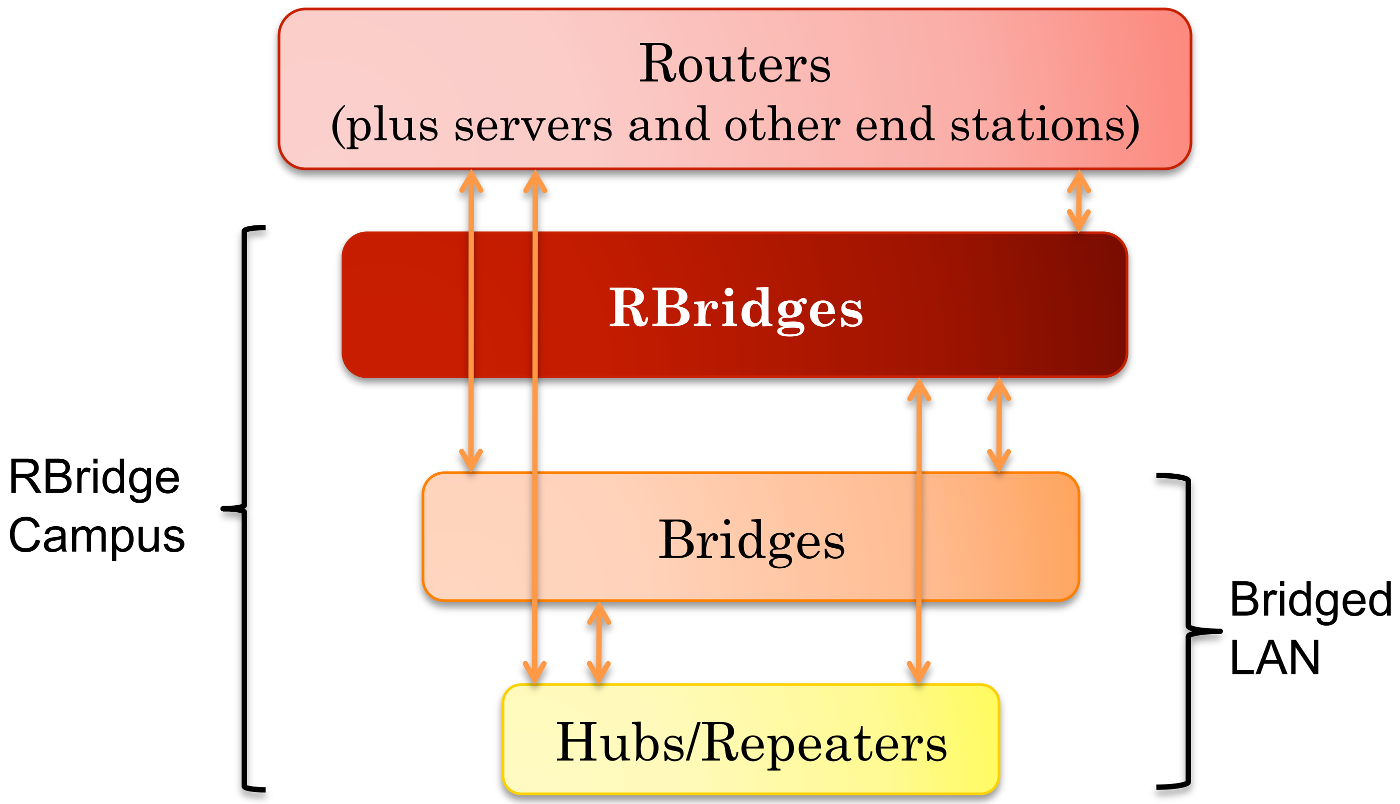
- They are obviously Bridges because
  - RBridges deliver unmodified frames from the source end station to the destination end station
  - RBridges can operate with zero configuration and auto-configure themselves
  - RBridges provide the restriction of frames to VLANs as IEEE 802.1Q bridges do
  - RBridges can support frame priorities as IEEE 802.1Q bridges do
  - RBridges, by default, learn MAC addresses from the data frames they receive

# Are RBridges Bridges or Routers?

- They are obviously Routers because
  - RBridges decrement a hop count in TRILL frames on each hop
  - RBridges swap the outer addresses on each RBridge hop from the ingress RBridge to the egress RBridge
  - RBridges use a routing protocol rather than the spanning tree protocol
  - RBridges optionally learn MAC addresses by distribution through the control messages
  - RBridges normally act based on IP multicast control messages (IGMP, MLD, and MRD) and restrict the distribution of IP derived multicast frames

# Are RBridges Bridges or Routers?

- Really, they are a new species, between IEEE 802.1 bridges and routers:



# CONTENTS

# 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 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.
- (IS-IS is the international standard which grew out of DECnet Phase V, in which Radia Perlman was heavily involved.)

# HOW RBRIDGES WORK

- RBridges find each other by exchanging TRILL IS-IS Hello frames
  - Like all TRILL IS-IS frames, TRILL Hellos are sent to the All-IS-IS-RBridges multicast address. They are transparently forwarded by bridges, dropped by end stations including routers, and are processed (but not forwarded) by RBridge ports.
  - The Hellos establish connectivity on each port.
  - Using the information exchanged in the Hellos, the RBridges on each link elect the Designated RBridge for that link

# How RBridges Work

- 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 encapsulates native frames from the link into a TRILL data frame, the ingress RBridge function.
  - It decapsulates native frames destined for the link from TRILL data frames. This is the egress RBridge function.

# How RBridges Work

- 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 information beyond connectivity and link cost. Information such as VLAN connectivity, multicast listeners and multicast router attachment, claimed nickname, options supported, and the like.
  - The 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 RBridges Work

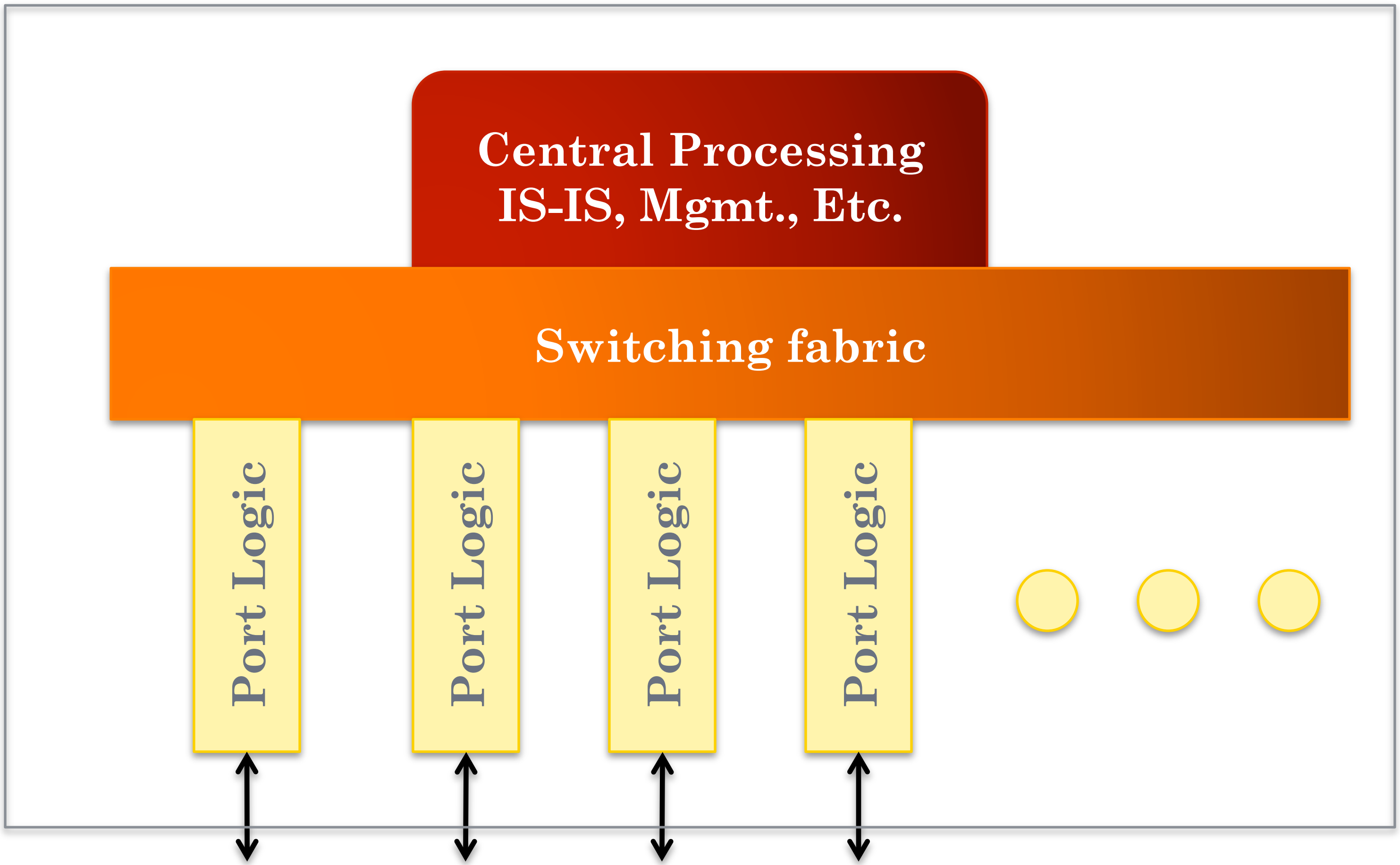
- TRILL data frames with
  - known unicast ultimate destinations are forwarded RBridge hop by RBridge hop toward the egress RBridge.
  - multi-destination frames (broadcast, multicast, and unknown destination unicast) are forwarded on a tree rooted at an RBridge selected by the ingress RBridge.

# 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.
- Ray Perlner

# CONTENTS

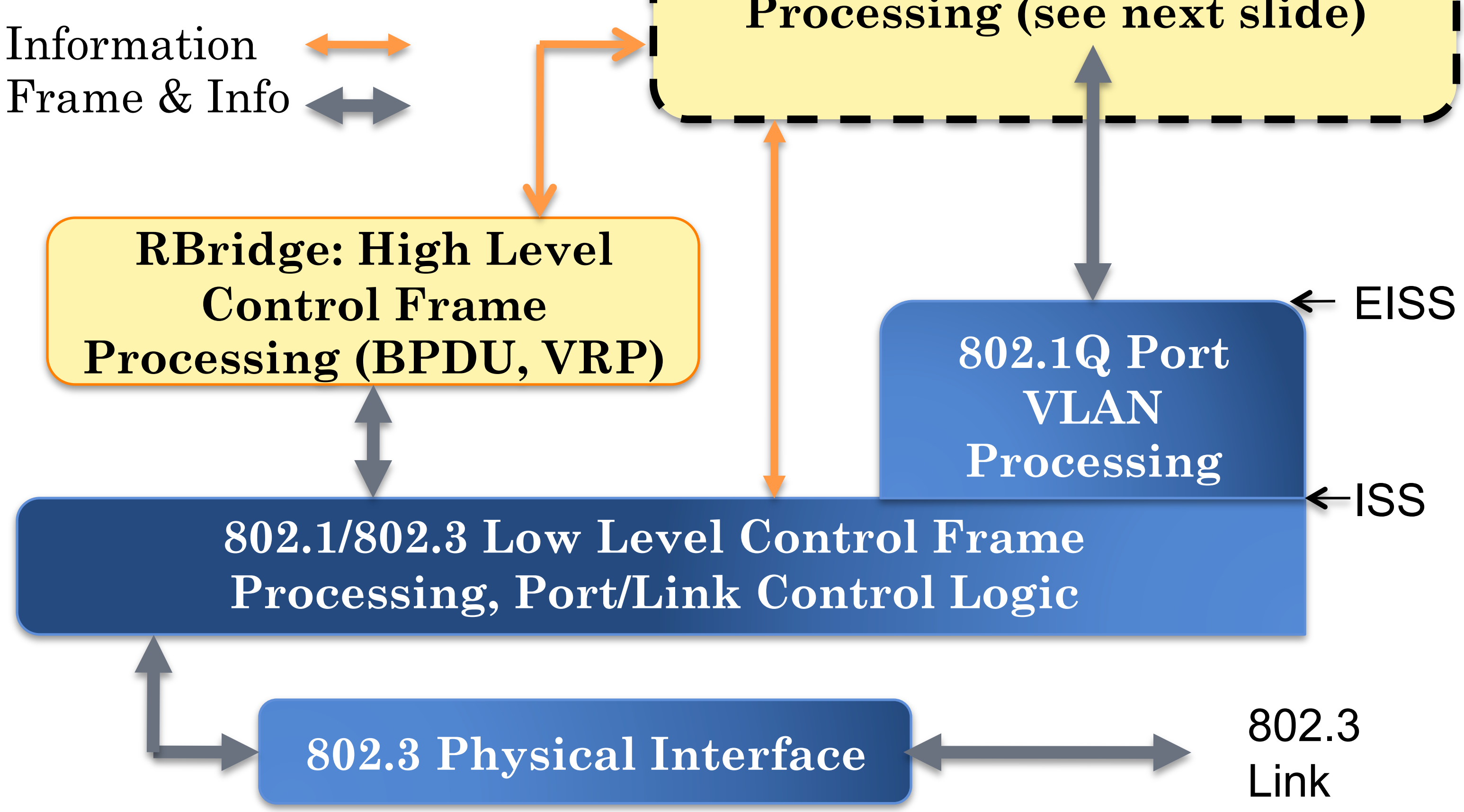
# Structure Of An RBridge



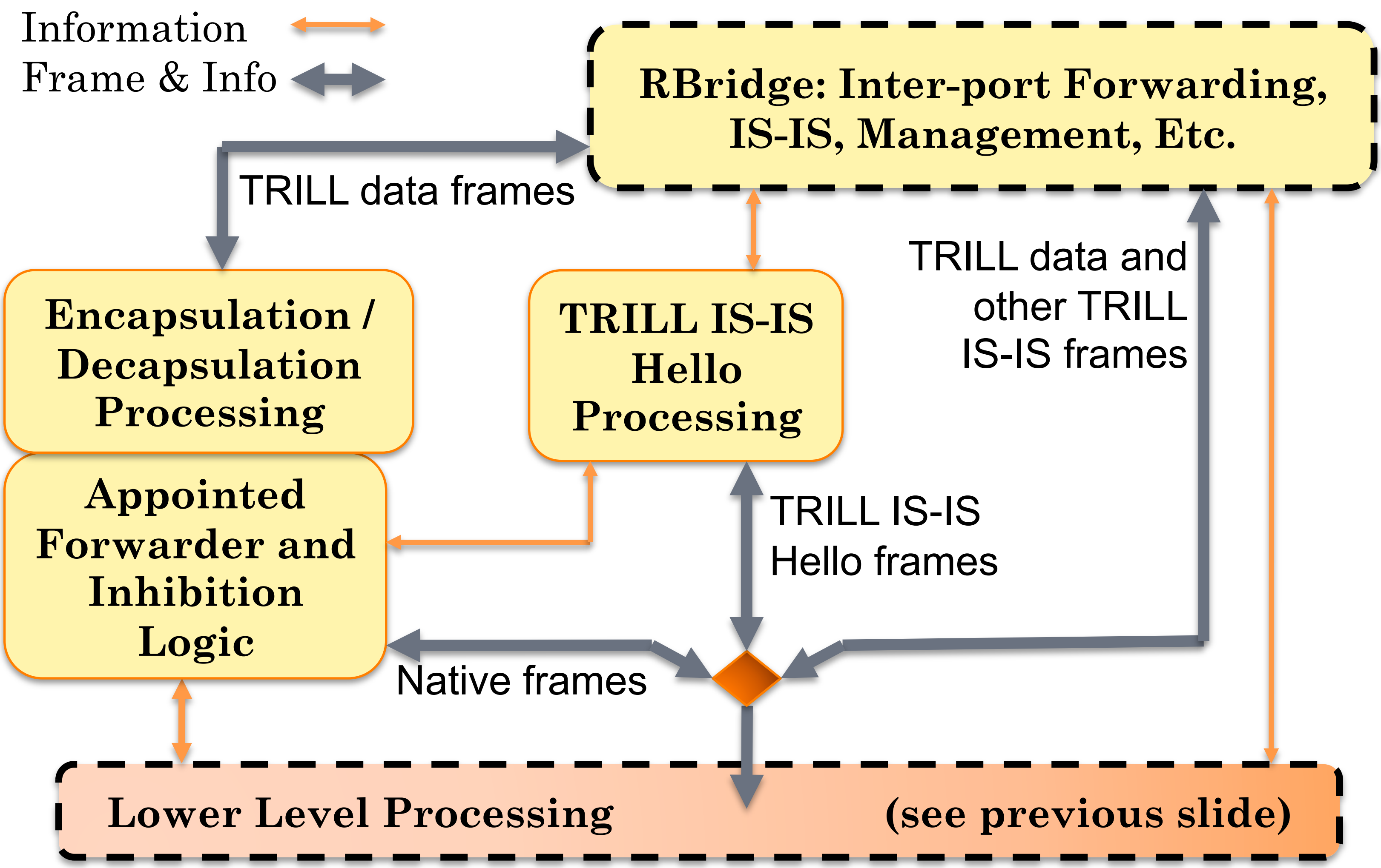

Links to other devices. Could be 802.3 (Ethernet), 802.11 (Wi-Fi), PPP, …

# Structure Of An RBridge Port

Assumes an Ethernet (802.3) link.



# Structure Of An RBridge Port



# CONTENTS

- Introduction
- Ethernet and Spanning Tree
- RBridge Features
- TRILL Encapsulation
- Are RBridges Bridges or Routers?
- How RBridges Work
- Structure of an RBridge
- <u>Some Additional Details</u>
- References

# Address Learning

- From Local Frames
  - { VLAN, Source Address, Port }
- From Decapsulated Frames
  - { Inner VLAN, Inner Source Address, Ingress RBridge }
    - The remote RBridge is learned as the proper egress RBridge for frames sent to the remote address and VLAN
- Via Optional End Station Address Distribution Information protocol
    - { VLAN, Address, RBridge nickname }

# WHAT ABOUT RE-ORDERING?

- RBridges are required to maintain frame ordering internally, modulo frame priority.
- When multi-pathing is used, all frames for an order-dependent flow must be sent on the same path if unicast or the same distribution tree if multi-destination.
- Re-ordering can occur briefly when a destination address transitions between being known and unknown or a topology change occurs.

# What About Loops?

- TRILL Data Frame Loops:
  - Known unicast frames have a hop count and are always unicast to the next hop RBridge.
  - Multi-destination frames must be received on a port which is part of their distribution tree, the ingress RBridge nickname must pass a Reverse Path Forwarding Check, and they have a hop count.
- Hybrid TRILL Data / Native Frame Loops:
  - TRILL takes great care to assure that there are almost never two uninhibited appointed forwarders on the same link for the same VLAN.
- Pure Native Frame Loops: Not TRILL's problem.

# Contents

# References

- Specification Draft:
  "Rbridges: Base Protocol Specification"
  - http://tools.ietf.org/html/draft-ietf-trill-rbridge-protocol-11
- Original Paper by Radia Perlman:
  "Rbridges: Transparent Routing"
  - http://www.postel.org/rbridge/infocom04-paper.pdf
- Current TRILL WG Charter
  - http://www.ietf.org/html.charters/trill-charter.html
- "TRILL: Problem and Applicability Statement"
  - http://www.ietf.org/internet-drafts/draft-ietf-trill-prob-05.txt

# END

Donald E. Eastlake 3rd
donald.eastlake@stellarswitches.com