Improvements in Ethernet Standards to Further Reduce Latency and Jitter

Norman Finn, Cisco Fellow
Cisco Systems, Inc.

Technical Track

In a nutshell

Technical Track

The Problem

At present, there are a number of competing methods for building Deterministic Ethernet networks.

- o 802.1AB AVB: Some useful bits :) and Spanning Tree :(
- o IEEE 1588 and 802.1AS time sync.
- o SAE AS6802 Time-Triggered Ethernet
- o ODVA DLR: That is, Rings.
- o PROFINET: isynchronous real-time Ethernet
- o ISO/IEC 62439 and others: PRP, Rings, Traffic Engineering
- o ITU-T G.803x Protection Switching, including rings.
- o Whatever IETF may be up to, based on TRILL.

Each standards suite has advantages and weaknesses.

The market for Deterministic Ethernet is ready to grow very rapidly; in fact, to explode.

A significant obstacle to market growth is the lack of a coherent Ethernet standards story for vendors to build to.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 3
www.odva.org

What is not obvious is that the above list of protocols can largely be brought together into a coherent plan by using the right kind of glue:

- **Network cores** (VLAN tagging)
- **IS-IS protocol**

IEEE 802.1 is the basis on which any number of other organizations can build standards suitable for specific markets.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 4
www.odva.org

Current 802.1 capabilities

Standards completed

Technical Track

IEEE 802.1 AVB organization

- **Institute for Electrical and Electronic Engineers**
  - ▶ IEEE Communications Magazine
  - ▶ IEEE Standards Association
    - IEEE 802 LAN/MAN Standards Committee
      - 802.1 Higher Layers Working Group
        - » 802.1 Interworking Task Group (bridging)
        - » (IEEE Std 802.1Q VLAN bridging)
        - » 802.1 Audio Video Bridging (AVB) Task Group
        - » (IEEE Std 802.1AB Audio Video Bridging)
      - 802.3 CSMA/CD MAC/Phy (Ethernet)
      - 802.11 Wi-Fi
    - IEEE 1588 Time Synchronization

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 6
www.odva.org

802.1 goal: Converging dissimilar traffic

- **Best effort**
  - ▶ Prioritized best-effort traffic is what Ethernet does best.
  - ▶ Ability to connect legacy equipment is required.
- **Bandwidth reserved**
  - ▶ 802.1Qat/802.1Qav audio/video traffic.
  - ▶ Mission-critical traffic not time-scheduled.
  - ▶ Talkers offer streams, Listeners register to receive.
  - ▶ Switches guarantee no loss due to congestion.
- **Time-synched mission-critical**
  - ▶ Stations and switches synchronized to a few nS.
  - ▶ Network transmits only critical data in "red" time slots.
  - ▶ Zero-time failover for duplicate engineered streams.
- **And all share the same network.**

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 7
www.odva.org

802.1 AVB existing and new standards

802.1aq: Shortest Path Bridging (no more spanning tree!)

802.1AB: profile for AVB-compliant devices

802.1AS: plug-and-play version of IEEE 1588 time sync

802.1Qat: bandwidth reservation protocol

802.1Qav: queue shaper for reserved bandwidth

802.1ASbt: enhanced time incl. multiple master clocks

802.1Qbu: preemption of transmitted frames

802.1Qbv: scheduled transmission

802.1Qbz: wired/wireless bridging

802.1Qca: enhancements to IS-IS bridging

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 8
www.odva.org

802.1Qat/.1Qav Bandwidth reservation

Talkers advertise streams and state their maximum bandwidth requirements.

Listeners subscribe to advertised streams.

Listener is given, along with talker advertisement, the maximum latency that this stream can encounter, assuming no network topology change. Listener can adjust buffers accordingly.

Streams in excess of bandwidth resources are refused.

Admitted streams are virtually guaranteed to meet advertised latency with 0 frame loss due to congestion.

Unused bandwidth is available for non-reserved traffic.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 9
www.odva.org

802.1ASbt Time Sync improvements

Support for link aggregation

New media types, e.g., IEEE 1901 and Wi-Fi Direct

Interoperability with one-step clocks on receive (but with no requirement to generate one-step Sync messages)

Support for redundant paths

Longer cable lengths, new media types

Include 802.3bf time sync capabilities in 802.1AS

Improved grandmaster changeover time

Longer chains of time-aware systems

Carry multiple time scales (e.g., local time & GMT)

Management support for automatic measurement of link delay asymmetry

Additional parameter sets for non-Audio/Video applications, e.g., industrial control.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 10
www.odva.org

802.1Qbt Time-Synched Ethernet

ONE real-time network scheduling model is:
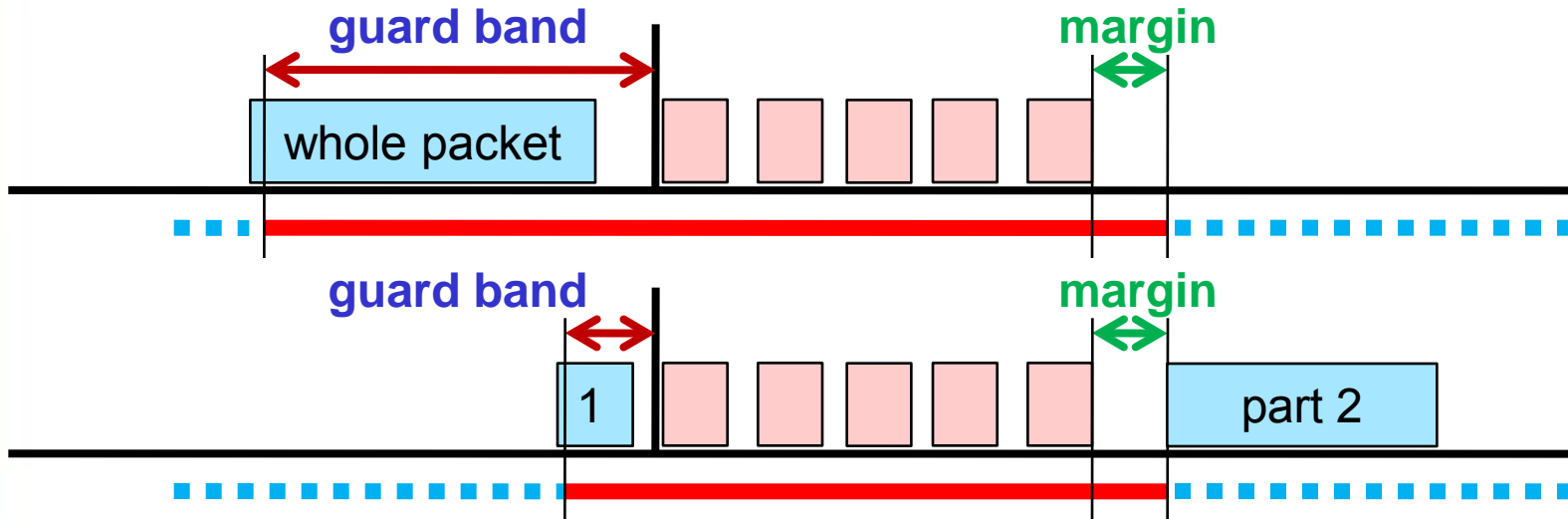**communicate**, **compute**, **communicate**, **compute**, **...**
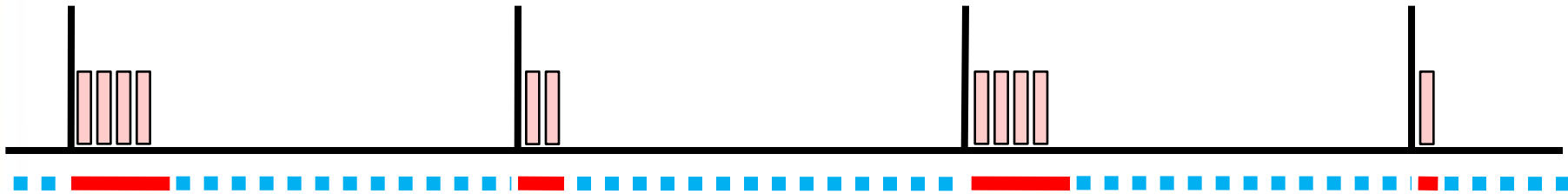
Communication occurs at specified times.

The timing is driven by the requirements of the critical application.

Strict scheduling can guarantee, no matter what happens, that we will respond to external events in a timely manner.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 11
www.odva.org

802.1Qbu Preemption shrinks guard band

guard band   margin

whole packet

guard band   margin

1   part 2

If preemption is used, the guard band need only be as large as the largest possible interfering fragment, instead of the largest possible interfering packet.

It is easy to see that the smaller the size of the time-reserved windows, the larger the impact of preemption.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 12
www.odva.org

Fault Recovery Protocols

Problems or solutions?

Technical Track

www.odva.org

Which protocol is better?

Spanning Tree: IEEE 802.1Q Rapid Spanning Tree Protocol or Multiple Spanning Tree Protocol (MSTP). One or more trees, up to one per VLAN. All data on one VLAN follows the same tree.

+ Handles absolutely any topology.

+ Can be plug-and-play.

– Often leads to very sub-optimal routing choices.

– Worst-case convergence time is several seconds.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 14
www.odva.org

Which protocol is better?

Routing Technology: SPB-V (Shortest Path Bridging V-mode) is defined by IEEE 802.1aq. Uses IS-IS to "route" at the MAC layer, rather like IETF TRILL, but without the encapsulation.

+ Handles absolutely any topology.

+ Can be plug-and-play (if 802.1 AVB defines a profile).

+ Unicasts and multicasts always routed along shortest path.

0 Worst-case convergence time is sub-second.

– (Number of VLANs) * (number of switches) < 4095

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 15
www.odva.org

Which protocol is better?

Ring Protocols: There are several candidates for protocols that assume that the network nodes are connected in a ring, and that the links are ordinary Ethernet links, including proprietary protocols, ITU-T's G.8032, and ODVA DLR rings.

+ Fast (≈10 ms) response to a link or node failure.

0 Routing typically not optimal, but separate topologies for separate VLANs can be configured by some protocols.

– Two or more failures lead to a loss of connectivity.

– If the physical topology is not, in fact, as assumed by the configuration, then the network will not operate correctly.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 16
www.odva.org

Which protocol is better?

Multiple simultaneous delivery: Either the end station or the edge switch replicates a frame and sends a separate copy along more than one path to the destination(s) outside the control of any topology recovery protocol. The receiver gets multiple copies.

+ No response required to a single link or node failure.

+ Redundancy not dependent on a device taking an action.

o Paths and flows must be configured manually or automatically.

– Bandwidth usage is at least doubled.

– If the physical topology or the configuration have a serious mismatch, then the network will not operate correctly.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 17
www.odva.org

Which protocol is better?

The answer to "which one is better," is, "It depends on your needs." Each protocol has its advantages and disadvantages.

In general, associated with each fault recovery protocol is a suite of ancillary protocols with capabilities relevant to an application. These additional capabilities can be more important than the features of the base protocols.

As we will see in this presentation, it is possible to get all of the benefits of all of these fault recovery protocols simultaneously!

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 18
www.odva.org

Network cores

Technical Track

802.1Q bridging Quality of Service features (reserved streams, scheduled queues, weighted fair queuing, strict priority) operate on priority levels, and are independent of forwarding choices that are, in part, based on VLAN.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 20
www.odva.org

Priority and VLAN are independent

The **VLAN and MAC addresses** determine on **which port(s)** a given frame is output.

The **Priority** determines **when the frame is output.**

The old standard 802.1 **forwarding plane** is suitable for almost any new topology control protocol, e.g. link state, ring, protection switched, or fixed.

The 802.1 traffic class standards (time scheduling, rate limiting, weighted fair queuing, straight priority) are **all available** to any topology control protocol.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 21
www.odva.org

If one is careful about VLAN usage, one bridge can use multiple topology selection protocols at the same time using its single Filtering Database.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
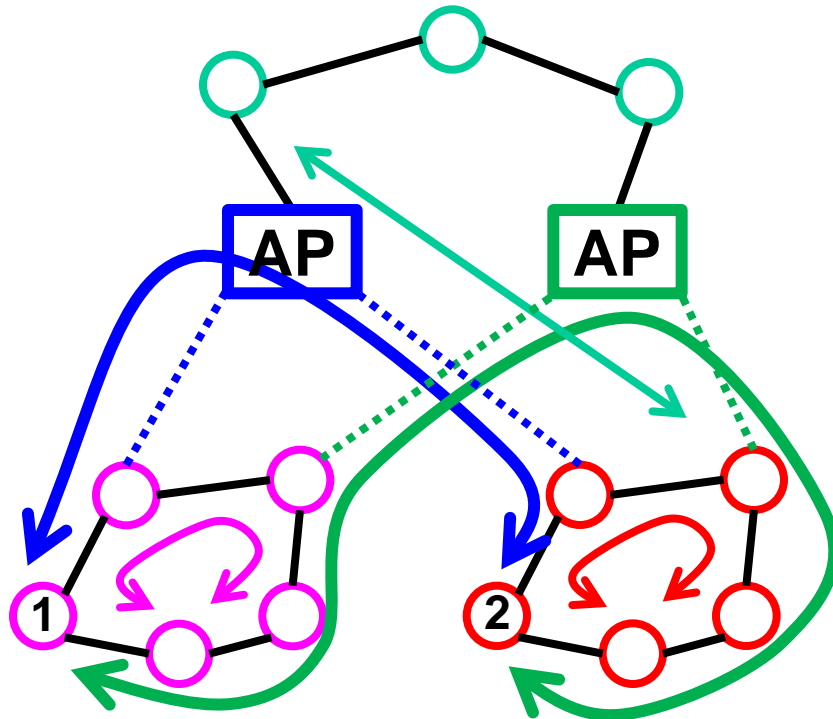All rights reserved.

page 22
www.odva.org

Multiple topology protocols by VLAN

Standard bridge forwarding using the 802.1 Filtering Information Database (FID) can be used with many topology protocols simultaneously, with the VLAN ID determining which topology a given frame follows:

- o 802.1Q MSTP
- o 802.1aq SPB-V
- o ITU-T Protection Switching
- o Duplicate delivery
- o Various Ring Protocols

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 23
www.odva.org

Which leads to: Network cores

One can imagine an industrial network with four cores:

1. SPB-V protocol runs VLAN 1, that reaches everywhere.

2. Traffic engineered paths use VLAN 8 and VLAN 9.

3. Ring protocol runs VLAN 5 for local data.

4. Ring protocol runs VLAN 6 for local data.

Frames controlled by different topology control protocols can use the same Priority values, and hence the same queues.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 24
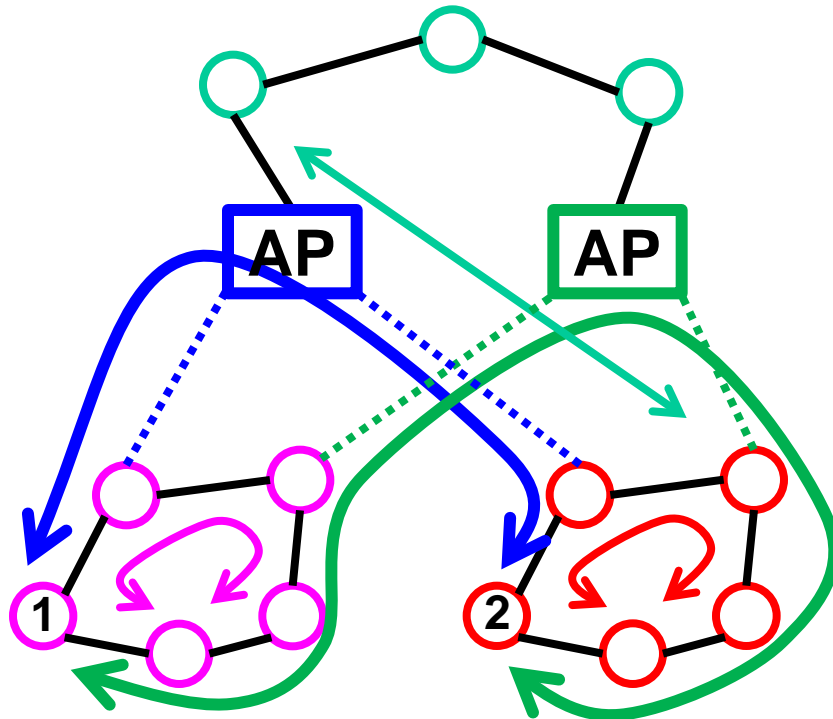www.odva.org

Network cores: Separated by VLANs

A "Network core" is a set of one or more VLANs that share a common fault recovery mechanism and a common subset (maybe all) of the nodes and links of a MAC Layer network.

Network cores can overlap each other arbitrarily, since their traffic is separated by the VLAN ID carried in every frame.

The preceding example shows a hierarchy of SPB-V above traffic engineering above ring, but hierarchy is not required.

- o SPB-V allows management traffic to reach every node as long as any connectivity is present at all, with sub-second failover times.
- o Traffic engineering supports 0-time failover for inter-ring data.
- o Rings support very fast failover times for local data.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 25
www.odva.org

Network cores

The best of all worlds

1. **SPB-V guarantees connectivity if there is a physical path.**

2. **Traffic engineered paths guarantee 0-time failover.**

3. **Ring protocols strike a balance between fast recovery and limited degree of redundancy.**

Time synchronization, scheduled transmission, bandwidth reservation, and priority levels are available for all!

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 26
www.odva.org

IS-IS under all

Technical Track

The IS-IS protocol (Intermediate System to Intermediate System) is a link state protocol.

Every switch (I use this term intentionally, as a node may be both a bridge and a router) lists what links it has to neighbor switches, along with additional information (e.g., its bridge number, what VLANs it needs to receive, and much more).

Every switch puts all this information about itself into an "advertisement" that it sends to all of its neighbor switches.

Every switch relays advertisements to and from its neighbors. The result of this flood behavior is that, eventually, every switch receives every other switches' advertisements.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 28
www.odva.org

The IS-IS advertisements allow every Switch to construct a model of the topology of the entire network.

Every switch uses that model to determine the path along which to forward any frame. Every frame is forwarded along the least-cost path, just like packets are forwarded by routers.
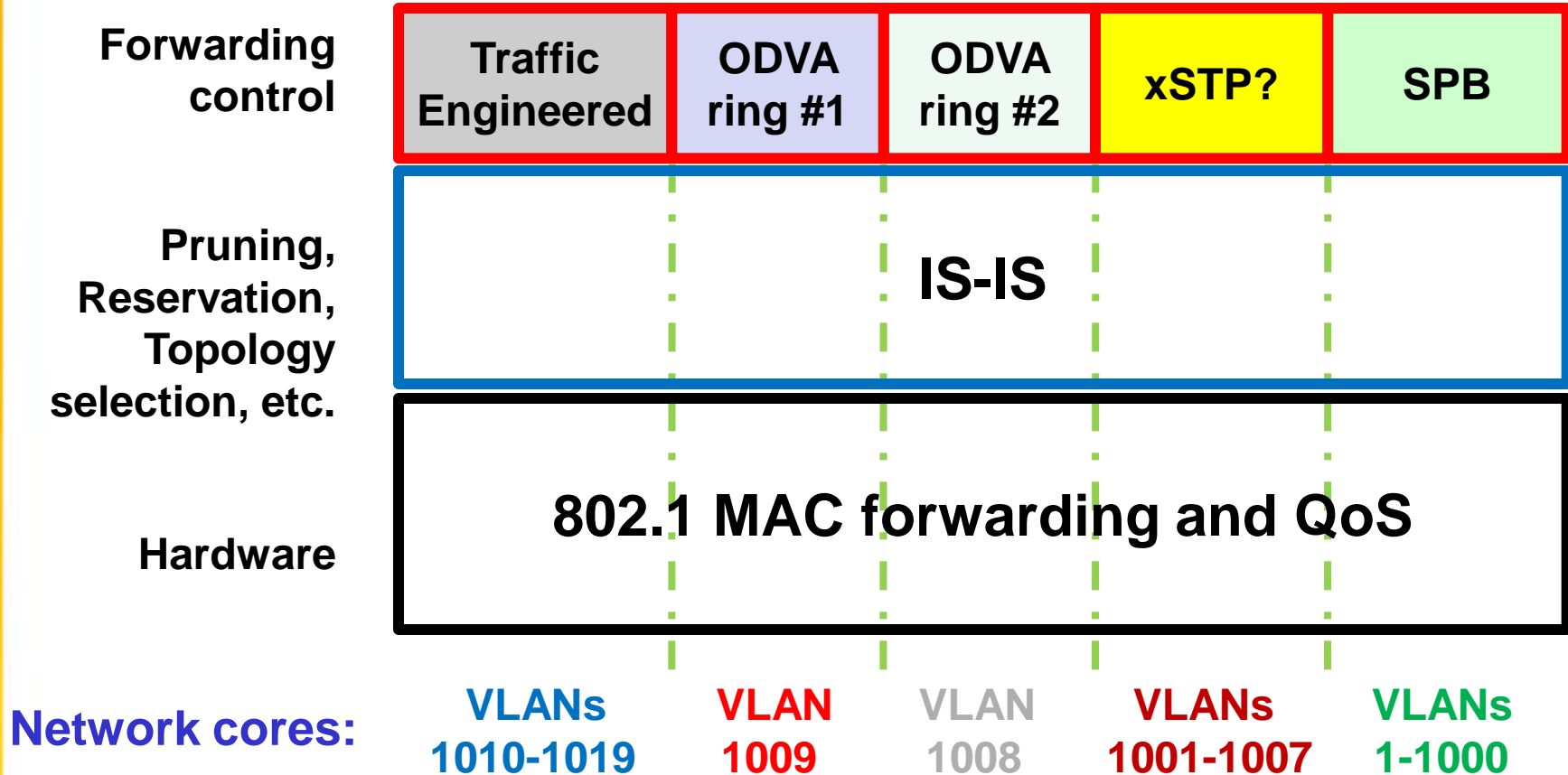
The SPB-V uses the 12-bit VID field in the 802.1Q tag to encode the tree number == the source bridge ID. One VID value is required per VLAN per bridge. Practically speaking (though not technically), the number of bridges times the number of VLANs must be less than 4095.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 29
www.odva.org

IS-IS can perform a number of protocol functions at the same time, most (or even all!) of which can be independent from the SPB-V (or any other) topology control function.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 30
www.odva.org

IS-IS outside the fault recovery context

- IS-IS now carries what VLANs are needed by each switch, thus optimizing broadcast delivery.

- IS-IS could carry information about 802.1AS Master Clocks, so that the clock distribution topology could reconverge more quickly after a failure.

- IS-IS topology information enables the members of a ring to verify that their actual connectivity matches their configured connectivity, without running a separate ring topology verification protocol.

- IS-IS topology information enables the switches on a traffic-engineered path to verify that the path is valid.

- IS-IS can support publishing engineered paths for specific flows through the network, outside the control of the topology protocol(s).

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 31
www.odva.org

Brick Wall diagram

Forwarding control	Traffic Engineered	ODVA ring #1	ODVA ring #2	xSTP?	SPB

Pruning, Reservation, Topology selection, etc.	IS-IS

Hardware	802.1 MAC forwarding and QoS

Network cores:

VLANs 1010-1019	VLAN 1009	VLAN 1008	VLANs 1001-1007	VLANs 1-1000

IS-IS and normal MAC underlie all network cores.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 32
www.odva.org

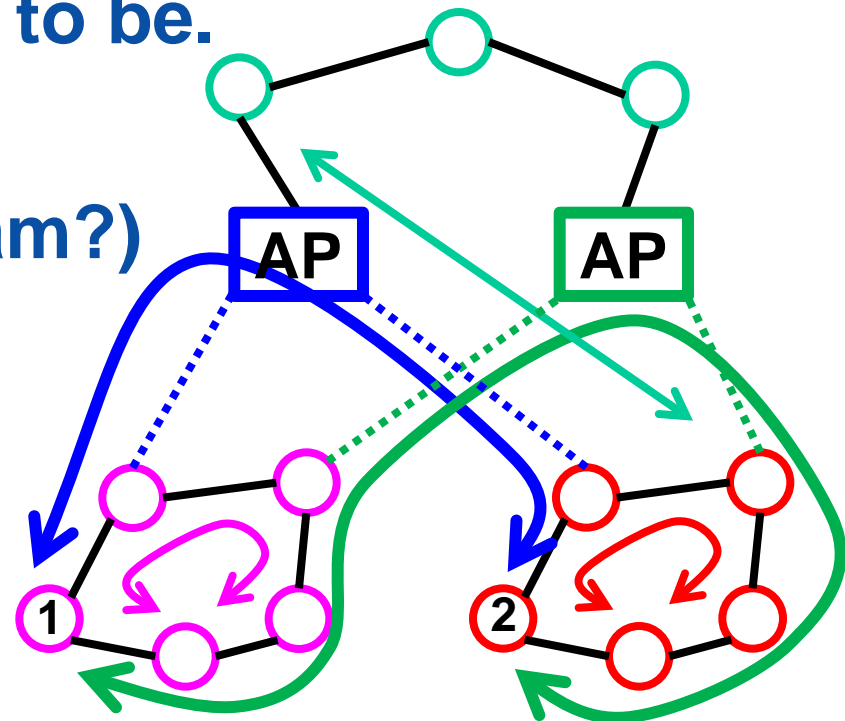# Wi-Fi is not at the edge!

Technical Track

www.odva.org

802.11 media can be in the middle of the network, not always at the edge of the network, and in Deterministic Ethernet networks often need to be.

(Remember this diagram?)

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 34
www.odva.org

IS-IS is too haaaaard!

Technical Track

www.odva.org

IS-IS is too haaaaard!

We can work this problem piecemeal, one issue and one fix at a time, or we can solve the whole problem at once.

Divide and conquer: Apply each fault recovery mechanism in its sweet spot, using all in parallel, with IS-IS as the underlying and unifying starting point.

The argument that "IS-IS is too complex" is short-sighted. What is "too complex" is trying to replicate the same capabilities across five suites of competing protocols. What is "too complex" is trying to force the one protocol to do two incompatible jobs, because the two protocols that can do those jobs are incompatible.

And, we may be able to simplify IS-IS for the smallest devices.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 36
www.odva.org

Work plan and Summary

Technical Track

- **IEEE 802 needs to work on:**
  - ▶ Our **current efforts** (802.1ASbt, Qbu, Qbv, Qbz, Qca 802.3 Single-pair 1 Gb/s Ethernet, 802.3 preemption?).
  - ▶ Fully support other media (MOST, IEEE 1901, MoCA, …)

- **Multiple standards groups need to work together to:**
  - ▶ Support **different topology control protocols** on different **VLANs**.
  - ▶ Solve particular problems in particular areas.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 38
www.odva.org

Summary

The work plan for IEEE 802 and for other Standards Development Organizations should be centered on:

- ➤ Using **VLANs** to separate **Network cores** that have separate addressing and separate forwarding topology control protocols.
- ➤ Basing QoS features on the **Q-tag Priority field**, orthogonal to VLANs and forwarding topology control protocols.
- ➤ Using **IS-IS** as a **unifying basis** for ancillary protocols for all forwarding topology control protocols.
- ➤ Putting **802.11** on par with 802.3 as a **bridging medium**.

- • **The result is that:**
  - ➤ A **maze of incompatible standards** becomes a **menu of customer choices**.

Technical Track
© 2012 ODVA, Inc.

2012 ODVA Industry Conference & 15th Annual Meeting
All rights reserved.

page 39
www.odva.org