Current state of IEEE 802.1 Time-Sensitive Networking Task Group

Norman Finn, Cisco Systems

What is Deterministic Networking?

Same as normal networking, but with the following features for **critical data streams**:

1. **Time synchronization** for network nodes and hosts to better than 1 µs.
2. Software for **resource reservation** for critical data streams (buffers and schedulers in network nodes and bandwidth on links), via configuration, management, and/or protocol action.
3. Software and hardware to ensure **extraordinarily low packet loss ratios**, starting at 10^{-6} and extending to 10^{-10} or better, and as a consequence, a **guaranteed end-to-end latency** for a reserved flow.
4. **Convergence** of critical data streams and other QoS features (including ordinary best-effort) on a single network, even when critical data streams are 75% of the bandwidth.

Who needs Deterministic Networking?

- Two classes of bleeding-edge customers, Industrial and Audio/Video. Both have moved into the digital world, and some are using packets, but now they all realize they must move to Ethernet, and most will move to the Internet Protocols.

1. **Industrial:** process control, machine control, and vehicles.
   - At Layer 2, this is IEEE 802.1 **Time-Sensitive Networking (TSN)**.
   - Data rate per stream very low, but can be large numbers of streams.
   - Latency critical to meeting control loop frequency requirements.
2. **Audio/video:** streams in live production studios.
   - At Layer 2, this is IEEE 802.1 **Audio Video Bridging (AVB)**.
   - Not so many flows, but one flow is 3 Gb/s now, 12 Gb/s tomorrow.
   - Latency and jitter are important, as buffers are scarce at these speeds.

- (You won't find any more market justification in this deck.)

Why such a low packet loss ratio?

Back-of-the-envelope calculations:

1. **Industrial:**
   - Automotive factory floor: 1000 networks • 1000 packets/s/network • 100,000 s/day = 10^{11} packets/day.
   - Machine fails safe when 2 consecutive packets are lost.
   - At a random loss ratio of **10^{-5}**, 10^{-10} is chance of 2 consecutive losses.
   - 10^{11} packets/day • 10^{-10} 2-loss ratio = **10 production line halts/day**.
   - In extreme cases, lost packets can damage equipment or kill people.
2. **Audio video production:** (not distribution)
   - 10^{10} b/s • 10 processing steps • 1000 s/show = 10^{14} bits = 10^{10} packets.
   - Waiting for ACKs and retries = too many buffers, too much latency.
   - Lost packets result in a **flawed master recording**, which is the user's end product.

How such a low packet loss ratio?

1. **Zero congestion loss**.
   - This requires **reserving** resources along the path. (Think, “IntServ” and “RSVP”) You cannot guarantee anything if you cannot say, “No.”
   - This requires **hardware** in the form of buffers, shapers, and schedulers. Overprovisioning not useful: its packet loss curve has a tail.
   - Circuits only scale by aggregation in to larger circuits. (MPLS? Others?)
   - 0 congestion loss goes hand-in-hand with finite **guaranteed latency**, also of importance to the users.
2. **Seamless redundancy**.
   - 1+1 redundancy: Serialize packets, send on 2 (or more) fixed paths, then combine and delete extras. Paths are seldom automatically rerouted.
   - 0 congestion loss means packet loss is failed equipment or cosmic rays.
   - Zero congestion loss satisfies some customers without seamless redundancy. The reverse is not true in a converged network—if there is congestion on one path, congestion is likely on the other path, as well.

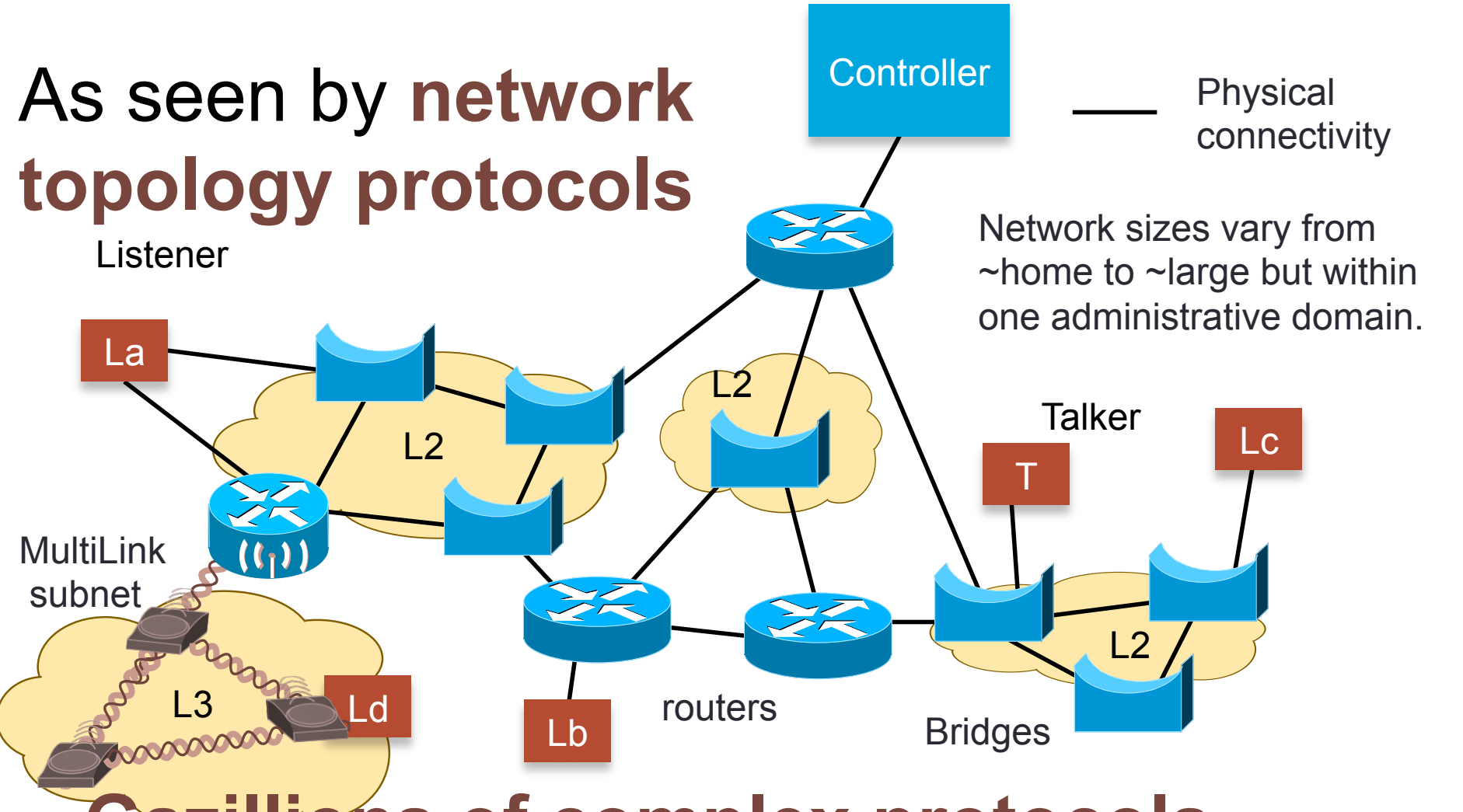
Why all the fuss? You could just …

- Old-timers remember the fuss 1983-1995 about Ethernet vs. Token Bus, Token Ring, and other "more deterministic" versions of IEEE 802 wired media. **Ethernet won.** One could argue that this TSN stuff sounds like the same argument. So, what's different besides, "That was them, this is us"?

1. Neither Ethernet nor any other IEEE 802 medium captured the business of the industrial control, vehicle control, or video studios that drive the present effort—they went to non-802 (including non-packet) answers.
2. Yes, Voice over IP works pretty well—except when it doesn't. The "except when it doesn't" is a non-starter for these users.
3. Too much data to overprovision.

Reference network

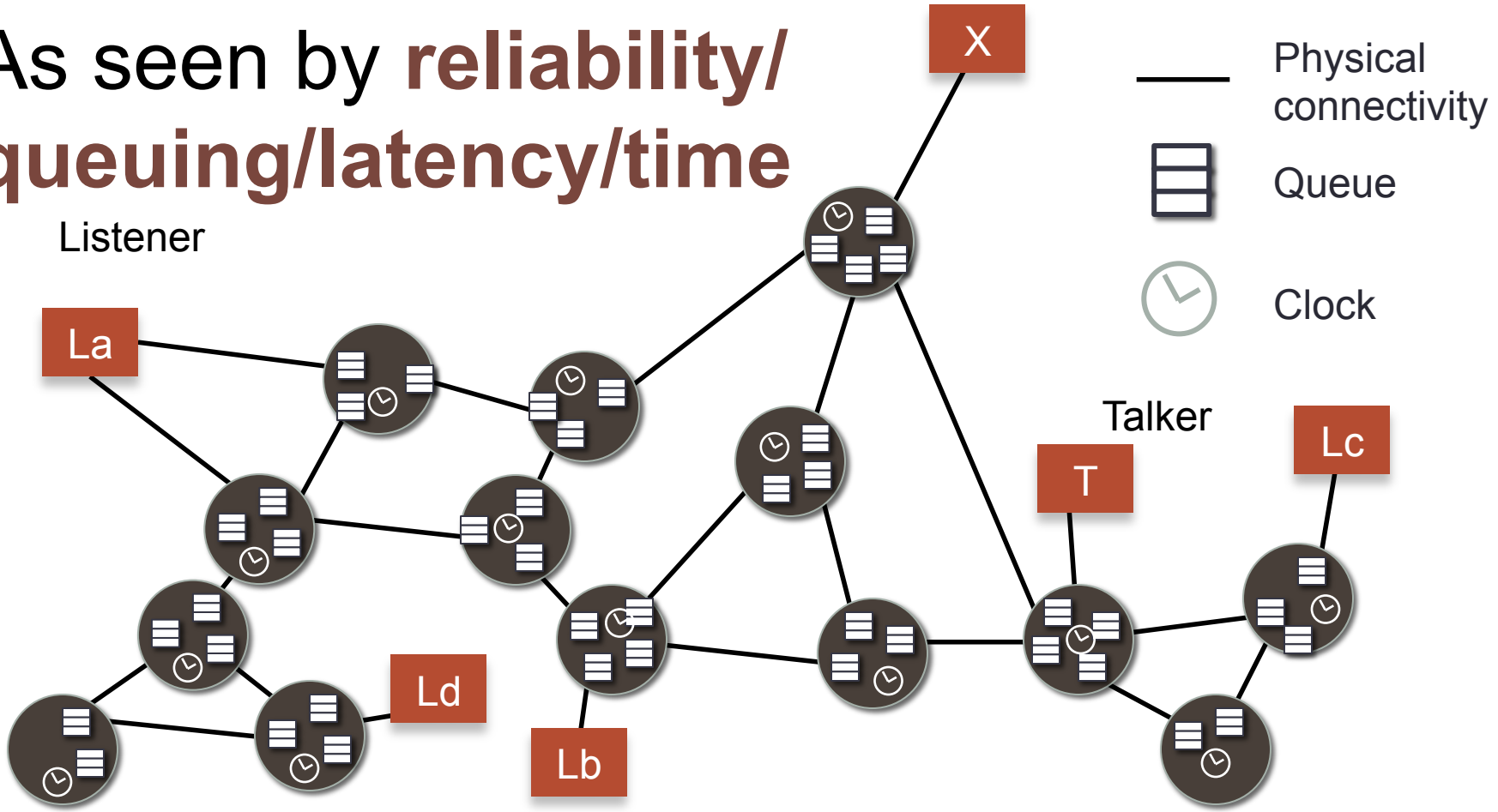
As seen by **network topology protocols**



- **Gazillions of complex protocols**

Reference network

As seen by **reliability/ queuing/latency/time**



- **Just nodes, queues, clocks, and wires!!**

Mixed L2/L3 = IEEE/IETF cooperation

- Both bridges and routers are important parts of these networks. Neither is going away.
- Every box along the path must reserve resources, and participate in the reservation protocols, whether a bridge or a router.
- Reservations from pre-configuration, management, or protocol.
- Hosts = applications can participate in the protocols.
- Hosts and operations managers don't know or care whether network is bridged or routed. One Host UNI, one operator view.
- There are valid use cases for application-driven peer-to-peer control flow models, for centrally controlled models, and for mixed scenarios.

IEEE 802 standards complete and in-progress

802.1 Audio Video Bridging TG is now the Time-Sensitive Networking TG.

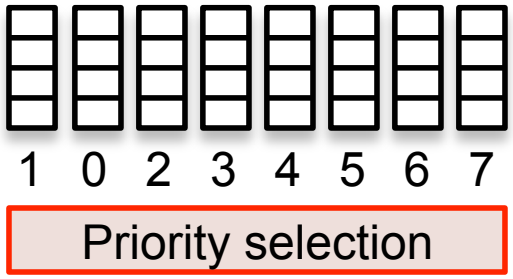
- **Time:** A plug-and-play Precision Time Protocol (PTP) profile that allow bridges, routers, or multi-homed end stations to serve as "time relays" in a physical network, regardless of L2/L3 boundaries. (Complete. Enhancements in progress.)
- **Reservation:** A protocol (MSRP) to reserve bandwidth along an L2 path determined by L2 topology protocol, e.g. ISIS. (Complete. Enhancements in progress.)
- **Execution:** Several kinds of resources (shapers, schedulers, etc.) that can be allocated to realize the promises made by the reservation. (See next slide.)
- **Path distribution:** ISIS TLVs to compute and distribute multiple paths through a network. (In progress)
- **Seamless Redundancy:** 1+1 duplication for reliability. (In progress)

The IEEE 802.1Q Queuing Model

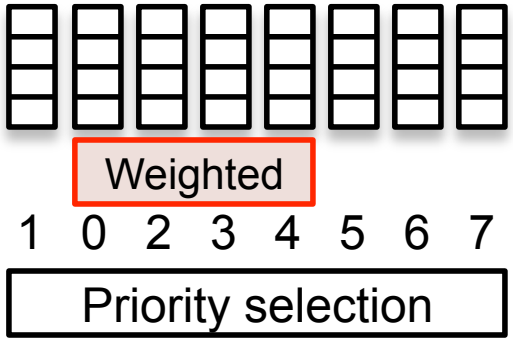
- IEEE 802.1 has an integrated set of queuing capabilities.
- There are several capabilities, most familiar to all.
- The “integrated” part is important—the interactions among these capabilities are well-characterized and mathematically sound.

Priority queuing and weighted queuing

- 802.1Q-1998: **Priority** (including weighted round robin)



- 802.1Q-2012 (802.1Qaz) adds **weighted queues**. This standard provides standard management hooks for weighted priority queues without over-specifying the details.



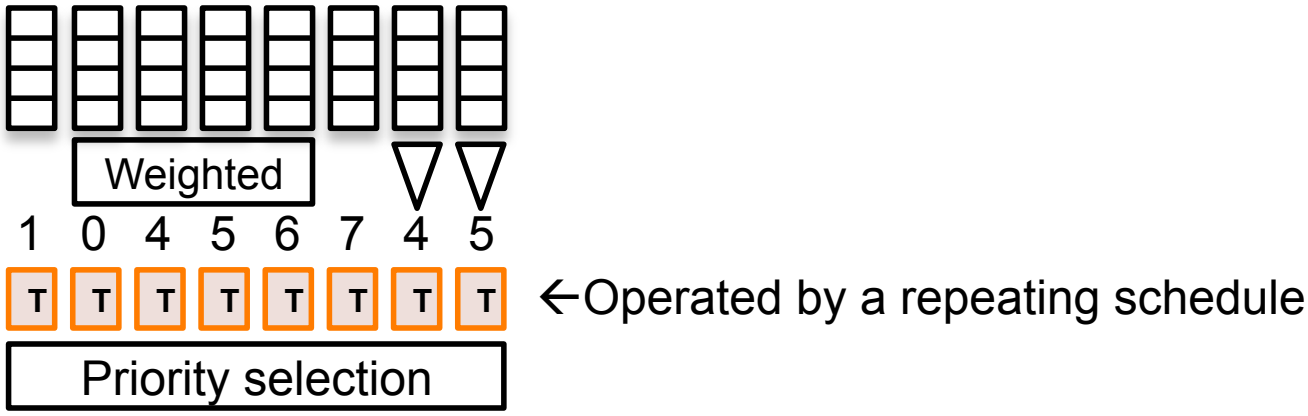
AVB shapers

- 802.1Q-2012 (802.1Qat) adds **shapers** . Shaped queues have higher priority than unshaped queues. The shaping still guarantees bandwidth to the highest unshaped priority (7).

- The AVB shaper is similar to the typical run rate/burst rate shaper, but with really useful mathematical properties.
  - Only parameter = bandwidth.
  - The impact on other queues of any number of adjacent shapers Is the same as the impact of one shaper with the same total bandwidth.

Time-gated queues

- 802.1Qbv: A circular **schedule** of {time, 8-bit mask} pairs controls gates between each queue and the priority selection function.



- These features, plus others in progress, support **guaranteed zero congestion loss** and **guaranteed finite latency** for reserved flows, and those guarantees are maintained as more reservations are made (or refused).

But wait! There's more!

- **Transmission preemption:** Interrupt (1 level only) transmission of an Ethernet frame with a frame with tight latency requirements, then resume the interrupted frame. (With a calculable impact on the other queuing mechanisms.)
  - But remember, if everyone is "special", then no one is special.
- **Cut-through forwarding:** The scheduling tools mentioned, above, allow one to guarantee scheduled cut-through forwarding opportunities for predictable ultra-low-latency packets.
- **Intentional buffering delays:** Time-scheduled transmissions can intentionally delaying transmissions in order to guarantee both a minimum and a maximum latency, thus minimizing jitter for the critical traffic. Industrial systems that trigger events based on packet reception require this.

Details and pointers

Complete IEEE AVB Standards

- IEEE Std 802.1BA-2011 "Audio Video Bridging (AVB) Systems"
  - A profile of a number of standards, picking out required options, special initialization parameters, etc., required for AVB-compliant bridges and end stations.
  - **An "AVB Device" is a device conforming to 802.1BA.**
- IEEE Std 802.1AS-2011 "Timing and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks"
  - A plug-and-play profile of IEEE 1588, including master clock selection, link discovery, and automatic creation of a tree to distribute the clock signal.
- IEEE Std 802.1Qat-2010 "Stream Reservation Protocol (SRP)"
  - The software protocol for making stream reservations
  - Has been rolled into 802.1Q as Clauses 34 and 35 of IEEE Std 802.1Q-2011.
- IEEE Std 802.1Qav-2009 "Forwarding and Queuing Enhancements for Time-Sensitive Streams"
  - A special credit-based hardware shaper for bridges and end stations that gives better latency guarantees than the usual shapers.
  - Has been rolled into 802.1Q as Clause 34 of IEEE Std 802.1Q-2011.
- IEEE Std 1722-2011 "Layer 2 Transport Protocol for Time Sensitive Applications in a Bridged Local Area Network"
  - A Layer 2 transport protocol carrying a time-to-display stamp on each packet.
- (All 802 standards (not 1722) are free 6 months after publication at http://standards.ieee.org/about/get.)

IEEE 802.1 TSN standards under way

- P802.1AS-REV* “Timing and Synchronization”
  - Revision of 802.1AS making it clear that it can run on a router as easily as on a bridge.
- P802.1Qbu* “Frame Preemption”
  - Amends 802.1Q to support 802.3br
- P802.3br “Interspersed Express Traffic”
  - One level of transmission preemption – interrupts transmission of an ordinary frame to transmit an “express” frame, then resumes the ordinary.
  - 802.3 document, not an 802.1 document.
- P802.1Qbv* “Enhancements for Scheduled Traffic”
  - Runs the 8 port output queues of a bridge on a rotating schedule.
- P802.1Qca* “Path Control and Reservation”
  - Enhances 802.1 ISIS to create multiple paths through a network.
- P802.1CB* “Seamless Redundancy”
  - Defines the sequence-split-recombine method for reliability improvement.
  - Stand-alone document. NOT an amendment to 802.1Q.
- P802.1Qcc* “Stream Reservation Protocol (SRP) Enhancements and Performance Improvements”
  - For more streams, faster convergence, less chattiness, and maybe more.

* For the necessary password, Google “p802.1 username password”

Other TSN-related standards

- Two IEEE 802 standards are underway at present, P802.1Qbz* and P802.11ak. These standards make it legal to:
  - Integrate a bridge into a Wi-Fi Access Point; and
  - Use a Wi-Fi station as a port on a bridge.
  - And thus, use an 802.11 link interior to a network, instead of only at the edge.
  - Upon completion, the entirety of TSN is available for Wi-Fi links.
- In IETF, the 6tisch Working Group (under Cisco's Pascal Thubert) is defining Deterministic Networking for the wireless space, particularly for IEEE 802.15.4e equipment.
- IEEE hopes to adapt the IETF Path Control Element into Deterministic Networking. That effort is just now starting.
- IEC 62439-3 defines the High-availability Seamless Redundancy (HSR) and Parallel Redundancy Protocol (PRP).

* For the necessary password, Google "p802.1 username password"

0 congestion loss: AVB class shaper

- 802.1Qav defines a hardware credit-based traffic shaper, one of which is applied to each AVB priority queue (typically two). All traffic on one class goes through the same shaper.
- Shaper can start transmitting whenever a) its queue is not empty and b) it has ≥ 0 credit.
- There is no configured “burst limit”, but the configuration of bandwidth of this and higher-priority queues limits the maximum credit that can be accrued.
- Shaper acquires credit at its programmed bandwidth whenever a) it’s transmitting, b) its credit is below 0, or c) its queue is not empty. Programmed bandwidth = sum of all flows’ reservations using the queue.
- Shaper loses credit at line rate whenever it is transmitting. The net credit rate is therefore a loss of (line rate) – (configured rate) during a transmission.
- Shaper’s credit is forced to zero whenever a) its credit is ≥ 0, b) it is not transmitting, and c) the queue is empty.
  - This inability to save up for the future never reduces the configured bandwidth, but does reduce the receiver’s worst-case buffer requirements.

0 congestion loss: AVB class shaper

- The AVB traffic shaper operating on classes (it does not operate on individual streams) cannot guarantee 0 congestion loss without knowledge of the network topology and intense calculations.
- But, it does give good enough results, in practice, to drive a growing market.

0 congestion loss: TSN Time-aware queues

- 802.1Q defines a maximum of 8 queues per output port, one per priority level.
- In the 802.1Q architecture, the 8 queues all feed a “transmission selection” function that selects among the queues presenting the “not empty” flag. This function operates by priority, modified by an optional weighted round-robin algorithm. Of course, AVB shaped queues go ahead of any priority queue.
- P802.1Qbv introduces a gate between each queue’s “not empty” flag and the transmission selection function. The 8 gates are controlled by a repeating schedule that can be synchronized over all ports in all bridges in a network.
- This simple mechanism can be configured to support a wide range of behaviors:
  - Create a window when a queue has the wire to itself.
  - TDM (time domain multiplexing) allocation for queues or groups of queues.
  - Time slots that can guarantee that cut-through forwarding is available.

0 congestion loss: Urgency Based Scheduler (UBS) (TSN)

- UBS uses a large array of AVB traffic shapers, all prioritized at one level with each other, and with the best-effort queues, with one shaper per stream, instead of one per class.
- A central server is required to set the priorities for each flow at each bridge port (and host port).
- The AVB shaper makes calculation of the worst-case buffer dwell time very easy as these queues are stacked up; six queues behave exactly like one queue with the same total bandwidth.
- By juggling the priorities at each node, any physically realizable set of latency requirements for intersecting streams can be met. This author believes that the required calculation is NP-complete.
- UBS has been proposed in the 802.1 TSN TG, but no project has yet started. (There are a number of presentations by Johannes Specht in the IEEE 802.1 2013 and 2014 public folders.)

0 congestion loss: Cyclical Queuing and Forwarding (CQF) (TSN)

- Formerly called "peristaltic shaping," proposed by Michael Johas Teener (Broadcom).
- Each bridge runs a clock at a certain frequency. All clocks are synchronized. 1/ frequency = "cycle time"
- Reservations made via SRP define the maximum number of bytes (on the wire) allowed to each flow per clock cycle. Minimum reservation is one frame/cycle.
- Each port requires up to 3 buffers, each big enough to hold one cycle of data.
- The key is to match the (delayed) receive cycle on each port to the next cycle on the output port. As long as no frame jumps to the wrong cycle, there is 0 congestion loss.
- User must make a tradeoff when configuring:
  - Small cycles = low latency.
  - Large cycles = more flows and higher ratio between maximum and minimum bandwidth reservation and more buffers.
- TSN TG is just starting work on this project. The PAR is not approved.

Seamless Redundancy (TSN)

- Serializes packets, splits them among several paths, then recombines the streams.
- May be done by multi-homed hosts or by network.
- Two very different scenarios:
  - “Industrial” = interval between transmissions on one flow slower than delivery times. Recombination is trivial: remember last-received serial number and discard repeats.
  - “Video” = many packets in flight, receive out-of sequence at recombination point. Recombination requires remembering a bit vector of recently-received sequence numbers and optional buffering to restore order.
- There is a challenge when combining bridging and routing in a single network, e.g. data center virtual controllers for a factory floor: how is the sequence number encapsulated?
- P802.1CB is underway in TSN TG. Since this author is the editor, it will have a framework suitable for mixed L2/L3 networks.

Control Plane details: 802.1AS Time Synchronization

- 802.1AS does not support “transparent clocks”.
  - A “transparent clock” in the 802.1 context would be a bridge that forwards an L2 PTP packet as an ordinary packet, based on its destination MAC address and VLAN, while updating a field in the packet that totals the time spent in bridges along the path waiting for forwarding.
- 802.1AS adds TLVs to the normal PTP packets to automatically elect a Grand Master and construct a delivery tree for the time signals. These tasks are combined into what is an addition to, not just a profile of, IEEE Std 1588, because:
  - Links that are measured to have variable delays (typically due to the presence of non-time-aware relay devices) or overly-long delays are removed from the active time topology. That means that the time topology does not equal the data topology. (This helps explain “no transparent clocks.”)
  - There is no point in electing a single “best” Grand Master using the algorithm in IEEE Std 1588 if that GM has no path to some of the users; every user needs a Grand Master.

Control Plane details: IEEE Std 802.1Qat Stream Reservation Protocol

- IEEE Std 802.1Qat-2010 was first published separately, as an amendment to IEEE Std 802.1Q-2005, and has since been rolled into IEEE 802.1Q-2011.
- SRP is based on the 802.1Q Multiple Reservation Protocol (MRP), just like the 802.1Q VLAN and multicast pruning protocols, MVRP and MMRP.
  - “Talker Declarations” of streams are distributed throughout the network along the path (paths for multicasts) to the Listeners.
  - Each Declaration creates a “Registration” on the receiving port, indicating the direction back to the device issuing the Declaration.
  - Listeners issue Listener Declarations that run back towards the Talker and actually reserve the resources.
- Because MRP is chatty, and is optimized for the case that all data fits in one data frame (often **not** the case for SRP), it tends to be a CPU pig, especially if implemented naively.

Control Plane details: P802.1Qca Path Control and Reservation

- Defines extensions to ISIS to allow the multiple paths required by P802.1CB Seamless Redundancy to be computed and distributed throughout the network.
  - Paths can be restrained by metrics other than the usual used for forwarding.
  - Paths can be pinned down completely, or pinned only to certain points.
  - Algorithms have been included for computing multiple paths that are maximally disjoint, according to various criteria.

Control Plane details: P802.1Qcc SRP Enhancements

- New project for Stream Reservation Protocol (SRP) Enhancements and Performance Improvements
- Charter includes improving the chattiness of SRP and its current difficulty with handling 1000s of streams.
  - May use the same mechanism as ISIS LSPs to transmit and acknowledge Declarations.
  - Will use the same “context forwarding” scheme as the existing SRP, so that the data follows the same path and gets to the same places; only the bits on the wire will change.
- 802.1Qcc will likely not cover the L3 needs for a Deterministic Networking UNI,

Control Plane details: P802.1CB Seamless Redundancy

- New project for 1+1 redundancy
- Standalone document, not an amendment to the 802.1Q Bridge specification.
- Current draft includes useful view of the overall TSN architecture.
- Supports a variety of sequence number marking methods including:
  1. A new L2 sequence number tag for Ethernet frames.
  2. HSR or PRP sequence numbers.
  3. Pseudowire sequence numbers.