Deterministic Networking

Norman Finn, Pascal Thubert, Cisco Systems

Jouni Korhonen, Broadcom

This presentation represents the opinions of the authors. Its contents have not been vetted by their companies, or by IEEE 802. See http://www.ieee802.org/1/files/public/docs2014/liaison-ietf-deterministic-networking-0714.txt for the official liaison message from IEEE 802.1 to IETF that triggered the creation of this presentation.

Why are we here?

- IEEE 802 has a suite of useful protocols for real-time processes, but it is limited to L2.
- A number of vendors and users want a mixed L2/L3 solution.
- **How do we get there?**
- The outline:
  - What, Who, Why, and How. A 4-slide summary of Deterministic Networking
  - History, and current state of IEEE 802 Deterministic Networking standards.
  - The need for a mixed L2/L3 solution.
  - The call for advice on what to do, next.

Who, What, Why, and How

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: the 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.
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.

IEEE 802 standards

completed and under development

IEEE 802 standards now and coming

802.1 Audio Video Bridging 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)

IEEE 802 schedulers and shapers

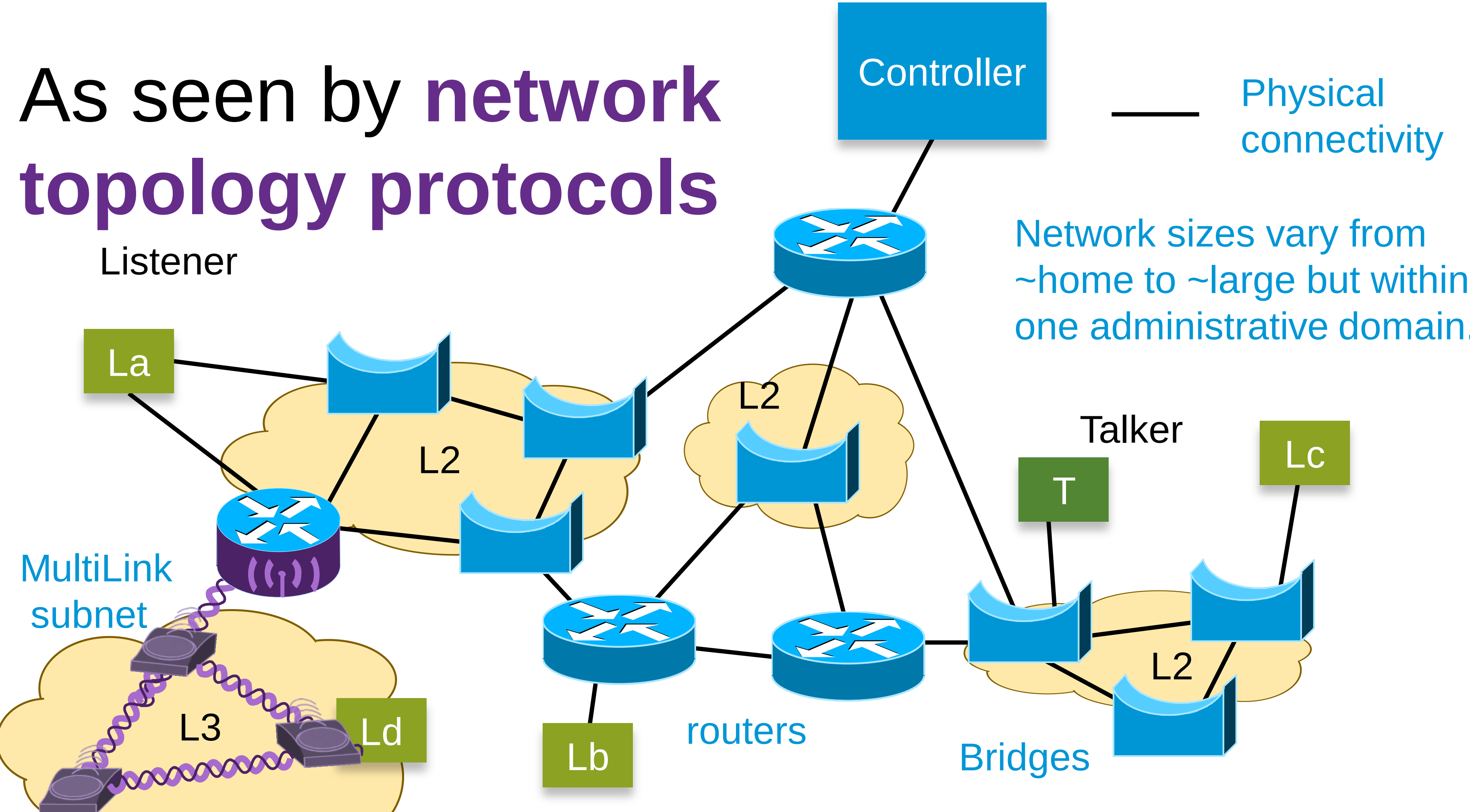
- **AVB Credit-Based Shaper:** Similar to the typical run rate/burst rate shaper, but with really useful mathematical properties. (Done)
  - Only parameter = bandwidth.
  - The impact of any number of shapers = the impact of one shaper with the same total bandwidth.
- **Transmission preemption:** Interrupt (1 level only) transmission of an Ethernet frame with a frame with tight latency requirements, then resume the interrupted frame. (In progress.)
- **Time scheduled:** Every bridge port runs a synchronized, repeating schedule that turns on and off each of the 8 queues with up to nanosecond precision. (In progress.)
- **Synchronized Queuing and Forwarding:** Every flow proceeds in lock-stepped transmission cycles, like arterial blood. (In progress).

IEEE AVB standards success

- The AVB TG gained some slight traction for the home – every Apple laptop supports AVB – but this is not widely used. Yet.
- The biggest use case today is audio and video studios and their equipment suppliers, e.g. Dolby, Gibson, Harman, Riedel, Extreme, Arista.
- The AVnu Alliance, an industry consortium, was created to promote the AVB standards (as Ethernet Alliance supports IEEE 802.3 standards). Originally oriented towards audio and video in the home and studio, focus now approximately even between AV and industrial/automotive.

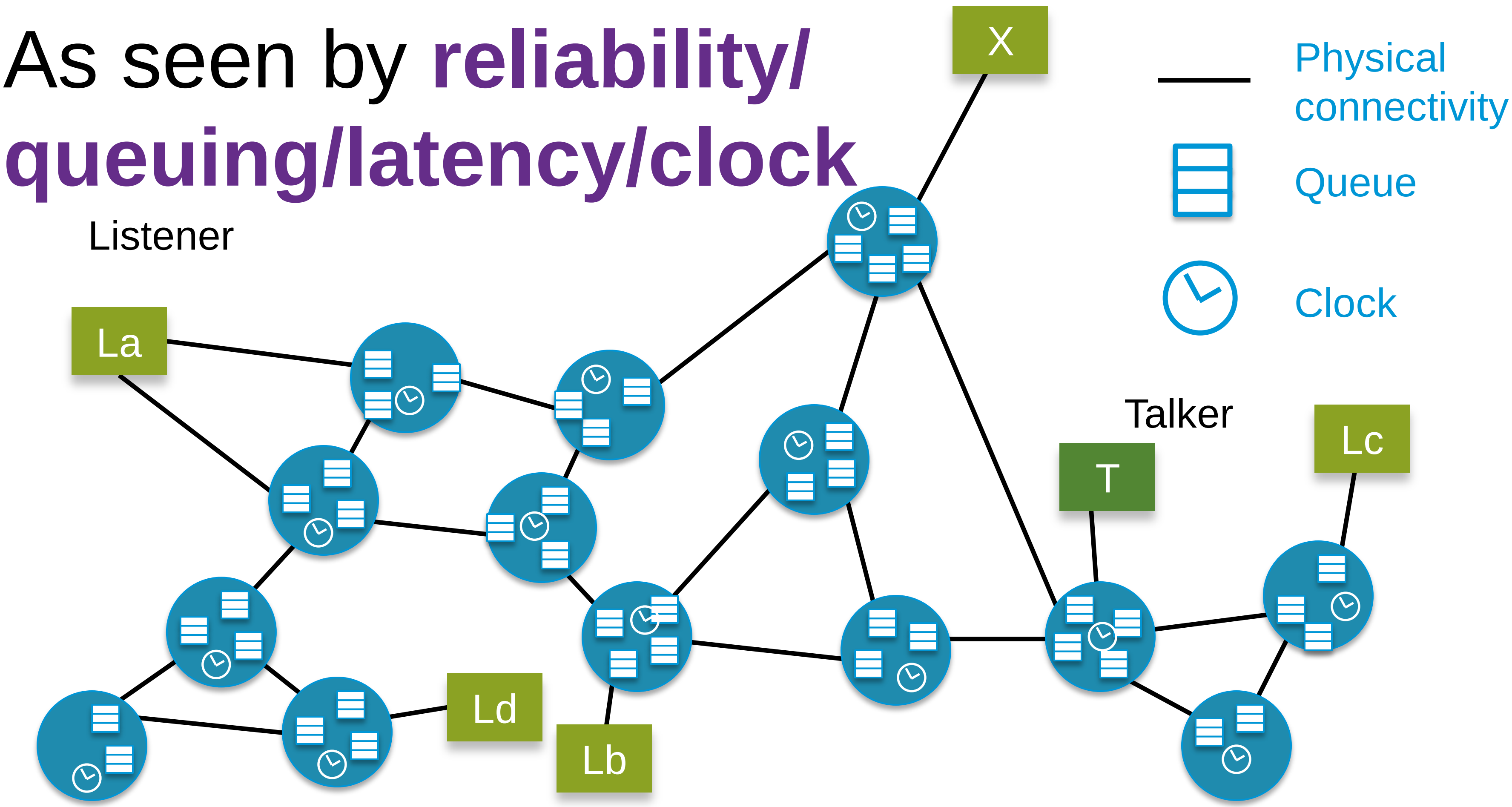
Mixed L2/L3 need

Reference network



- Gazillions of complex protocols

Reference network

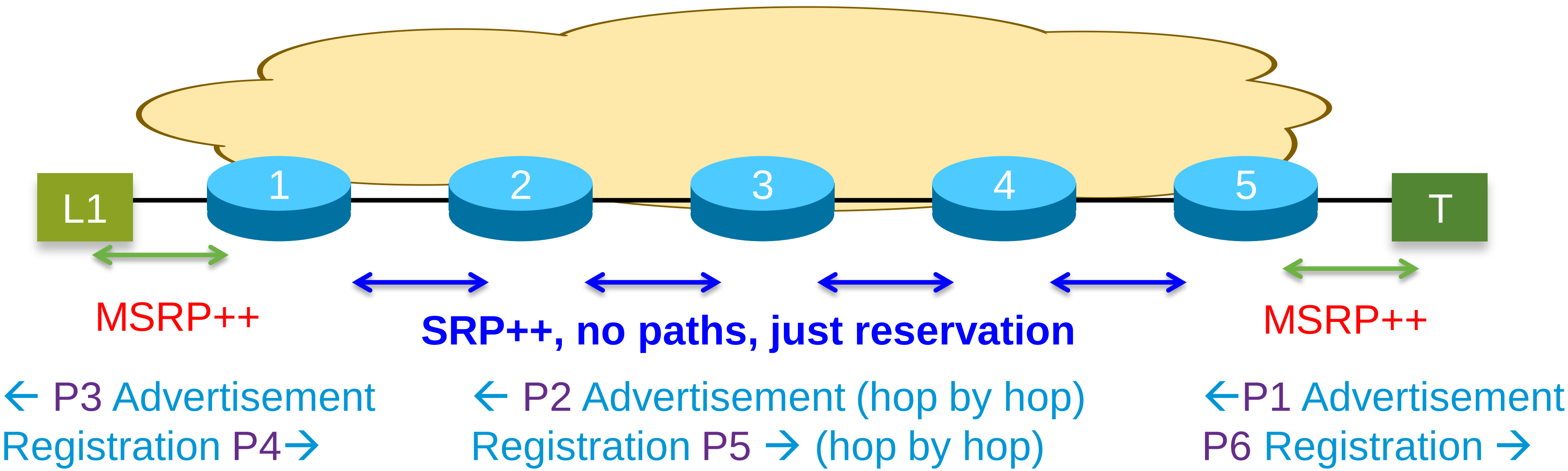


- Just nodes, queues, clocks, and wires!!

Mixed L2/L3 = IEEE/IETF cooperation

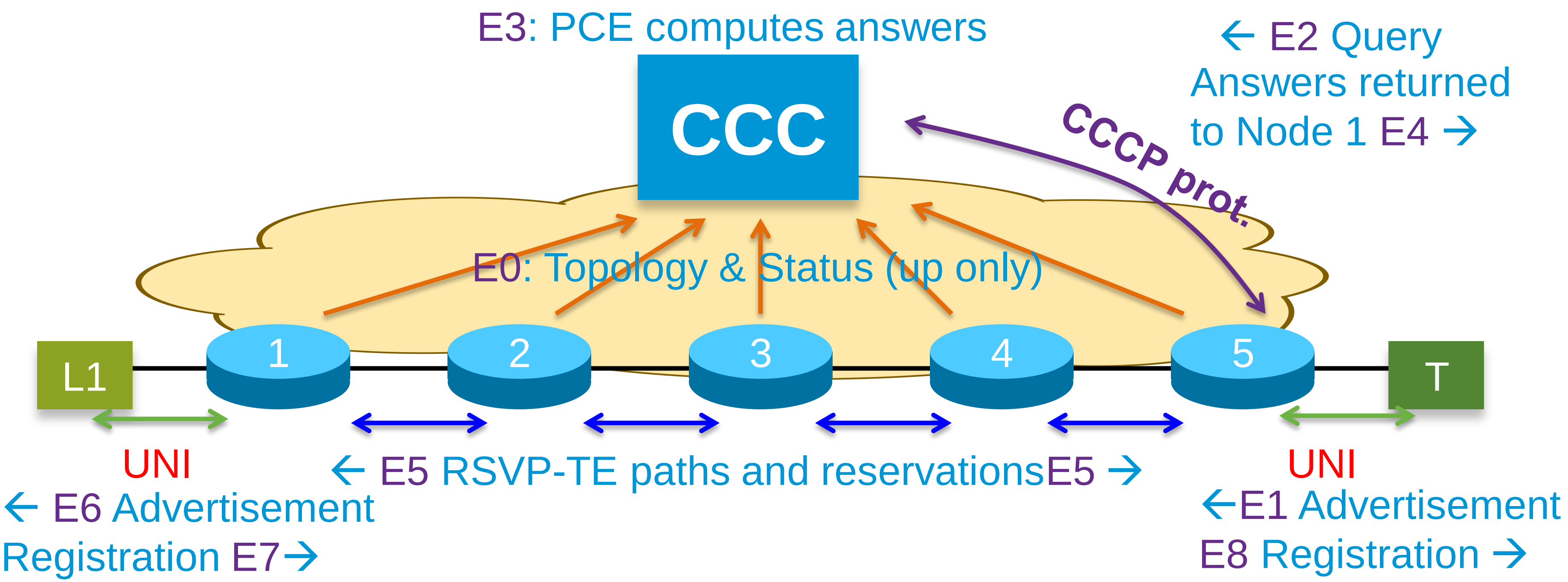
- Bridges are an important part of these networks. They are not going away.
- Routers will be an important part of these networks. They need to be introduced.
- Every box along the path must reserve resources, whether a bridge or a router.
- Every box along the path must participate in reservation protocols, whether a bridge or a router.
- Reservations from pre-configuration, management, or protocol.
- End stations = hosts = applications 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.

Control plane: Peer-to-peer



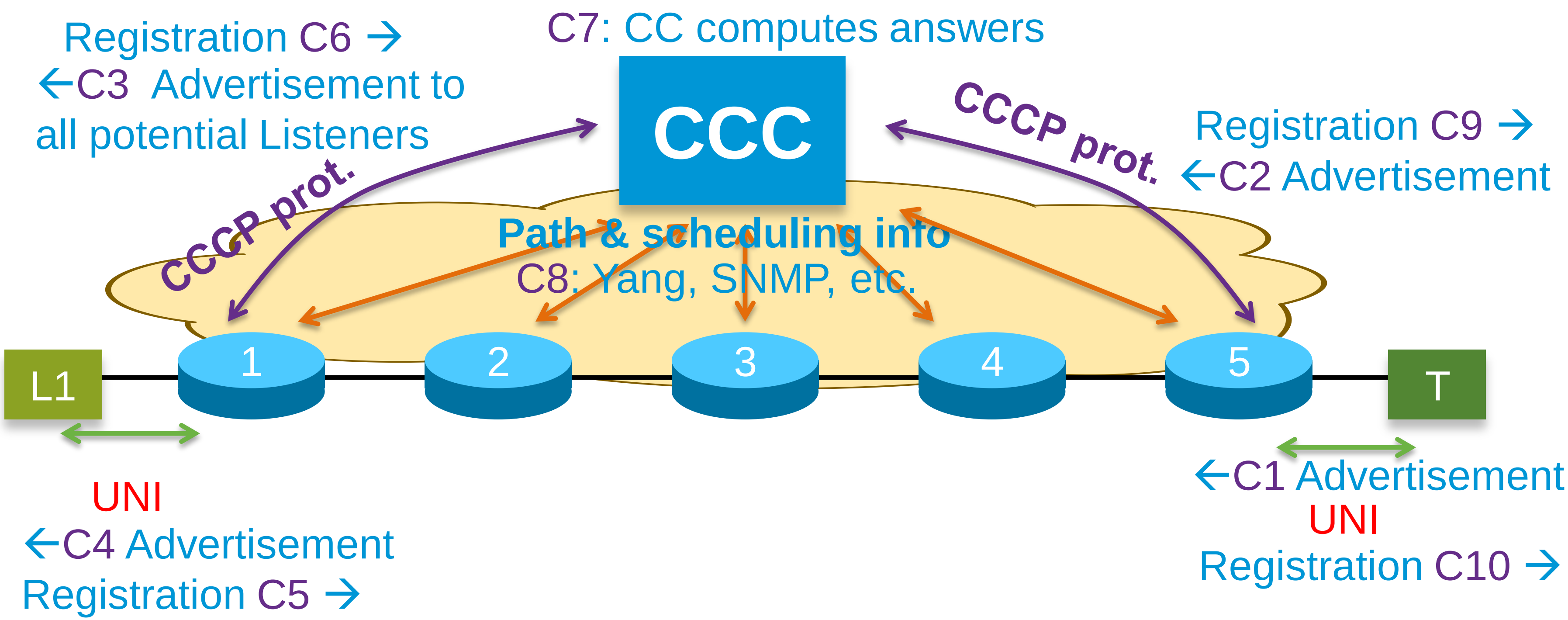
- A peer-to-peer control paradigm is used by IEEE MSRP (and RSVP).
- This paradigm is adequate for some scheduler/shaper methods, but not for all. (Some require a central brain.)

Control plane: Central control RSVP-style



- Edge node turns user request into query/response with central server, then propagates the answer peer-to-peer through the network.
- This is the current IETF PCE model, with the addition of hosts and UNI.

Control plane: Central control hub-spoke



- A central server communicating radially with network nodes can support all schedulers/shapers with the minimum amount of standards writing, and maximum velocity of features.
- Several existing IETF solutions available as the basis for “CCCP” and transferring “Path & scheduling info”.

Next steps

How do we get this?

- Have we, in fact, convinced you that this is a good thing?
- Do we have a BOF at IETF91?
- Do we create a Working Group and do it all there?
- Do we create a Working Group that does some of it, but farms out bits to other groups (e.g. pcewg, aqmwg)?
- Do we bombard 8 WGs with drafts for mixed L2/L3 features?
- Do we work with IEEE 802.1 formally? Via shared membership?