Resiliency in Ethernet Based Transport Networks

Kari Seppänen
Kari.Seppanen@vtt.fi

Outline

- Introduction
- What is "switched" Ethernet?
- Legacy Ethernet
- Security and Reliability issues
- Rapid spanning tree protocol & others
- Multiple spanning trees
- Carrier-grade Ethernet
- Ethernet layer 2 switching
- Conclusions

Introduction

- The topic of this presentation is resiliency in native Ethernet transport networks.
  - Ethernet services (e.g., E-Line, E-LAN) can be based on virtually any transport technology like EoS, IP/MPLS VPNs, ATM...
  - Resiliency in Ethernet services depends on underlying transport network
  - In p-to-p links between IP routers or such, Ethernet is just one LL protocol
  - "Carrier-grade Ethernet" sometimes mean Ethernet over MPLS or such
- Ethernet is considered to be cheap, simple and versatile transport technology for IP NGN
- Ethernet is also considered to be easy to manage, plug-and-play solution
- Most of the Internet hosts use Ethernet i/f — so why not transport that traffic end-to-end in native form?

What is "switched" Ethernet?

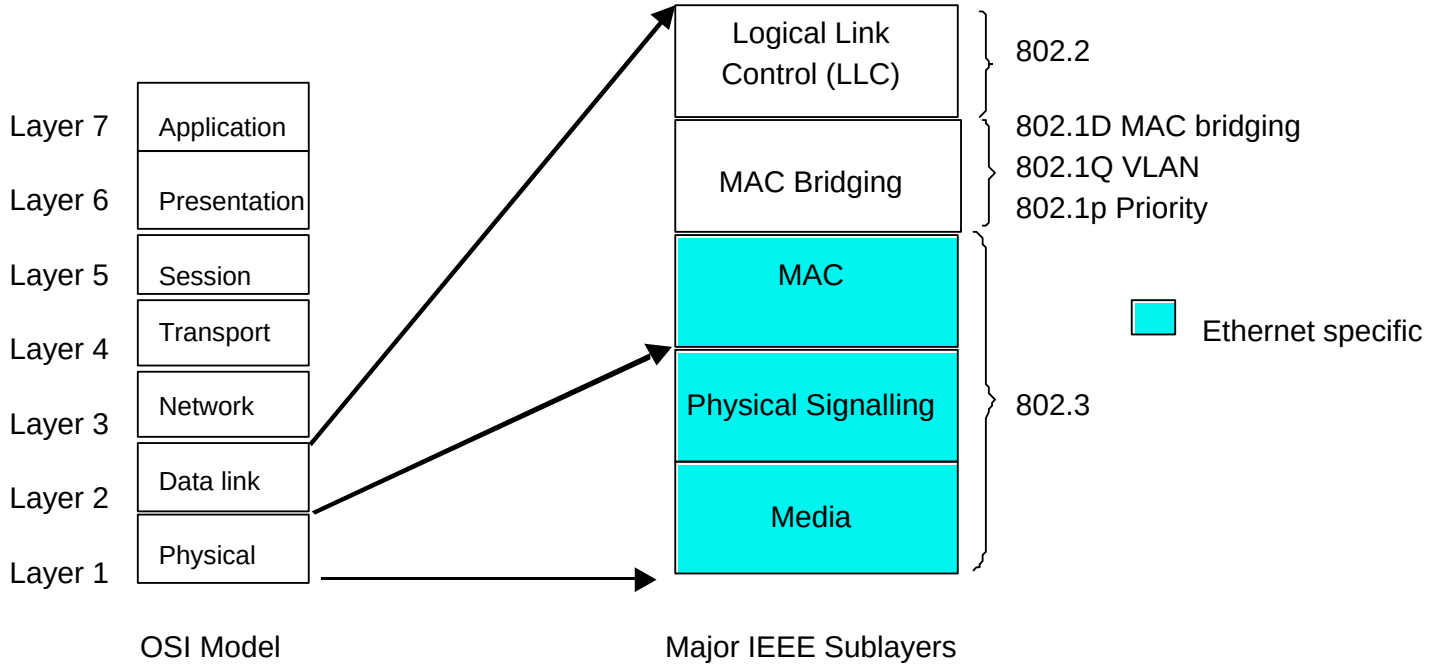
Definition of layer 2 protocol:

> Anything defined by a committee whose charter is to standardize a layer 2 protocol. *(Radia Perlman, Sun Laboratories)*

- Ether**NET** is a network — right?
  - No — it's just a link in a network.
- But there are Ethernet switches that do L2 switching?
  - Not exactly — logically an Ethernet switch is more or less a set of bridges connected by a high-speed Ethernet segment (or multiport bridge).
  - Bridged/switched Ethernet divides a single collision domain (Eth. segment) to multiple ones (rather than divides a network into multiple subnets)

Legacy Ethernet

Ethernet is a frame based multiaccess LL protocol with flat addressing.



- Bridge learning make it possible to divide a LAN into multiple collision domains automatically
- Need loop-free topology → Spanning Tree Protocol (STP)

Security, Reliability and Scalability issues

- Performance monitoring & fault detection
  - Only 10GbE WAN option (poor man's SDH) has any performance monitoring
  - Otherwise — only "loss-of-light" at Ethernet layer
  - How to get performance data for maintenance, e.g., replacing aging lasers?
- Management & monitoring
  - Relies on upper layer (IP) system like SNMP or Web application
- STP
  - Root node is a single point of failure
  - Slooooooow action in error situations (i.e., up to minutes)
  - Insecure — if user's port accepts STP packets (e.g., due to misconfig.), attacker can bring down the network or make all traffic go through attacker's machine

  - No load balancing (except for link bundles)

- Link aggregation — parallel links between two switches can be bundled to a single virtual link for STP
  - Provides link-level load sharing
  - Std. version provides "fast failover" — $\approx$ 1 s
  - faster recovery with vendor specific solutions

- Cheap shall it be...
  - 10 GbE WAN option is much like SDH/SONET but eliminates "expensive" SDH clock distribution...
  - ...actually money saved was about 5 USD per line card
  - 10 GbE transport network cannot be used to distribute clock signal to legacy networks

Rapid spanning tree protocol & others

- STP relies on timers and Bridge Protocol Data Units (BPDU) sent by root node — in RSTP (802.1w) each bridge sends hello BPDUs → faster fault detection
- All edge ports can be turned on immediately
- If a non-root bridge detects that root is unreachable, it can immediately declare itself as a new root node
- Still $\approx$ 1 s convergence
- Rapid Ring STP (RRSTP) or Ethernet Automatic Protection Switching (EAPS) for bidirectional Ethernet rings ($\neq$ RPR) — mimics bidir. SDH ring with similar performance

Multiple spanning trees

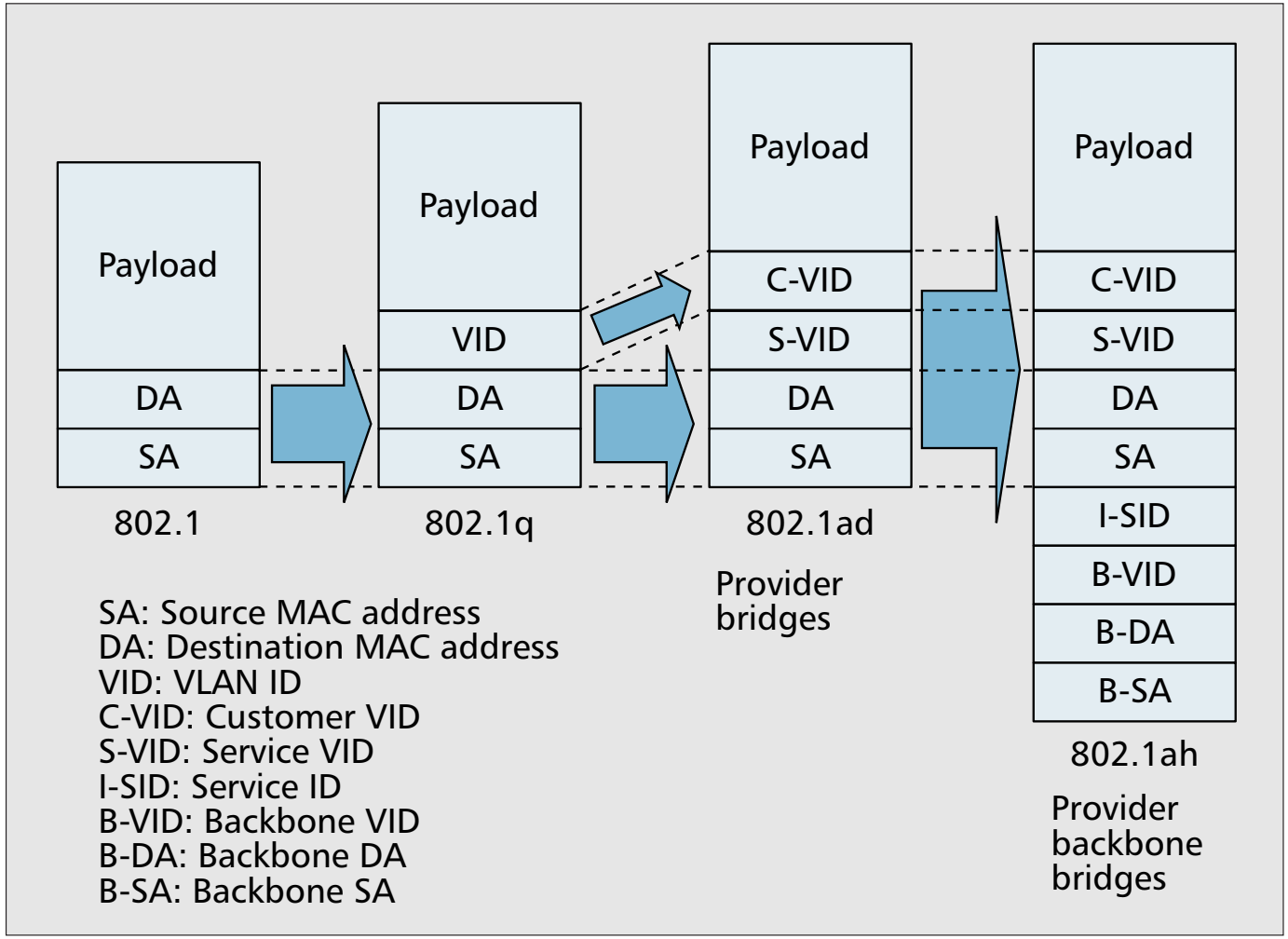
- RSTP has still only one spanning tree
  - Suboptimal "routes" for most of VLANs
  - Bottleneck links and unused resources
- MSTP (802.1s) allows for own spanning tree for each VLAN
  - Location of root node for each ST can be selected according to each structure/traffic patterns of each VLAN $\rightarrow$ better "routes"
  - Traffic is more distributed and resources are better used
  - A single fault does not (necessarily) affect all VLANs

Carrier-grade Ethernet

Major requirement: the transport network has to be protected from user's actions
$\longrightarrow$ hierarchical network is needed

- Service OAM (802.1ag connectivity fault management)
  - Continuity check, link trace, loopback and AIS messages
- Link OAM (802.3 clause 57)
  - Discovery, link monitoring, RFI, remote loopback
- L2 VPN is considered as a major application
  - Q-in-Q does not provide sufficient protection for transport network, MAC learning is also an issue
  - Limited number (4094) of VLANs
  - MAC-in-MAC is required (that is, user's frames are carried inside provider's frames)

- TDM over Ethernet to support legacy PDH services



Ethernet L2 switching

(a.k.a. Independent VLAN learning (IVL))

- Even with MSTP exact traffic engineering ("route pinning") is not possible
- L2 switching with Ethernet
  - VLAN ID (VID) and destination MAC addr. are used as a label (VID can be swapped, MAC addr. has to remain the same)
  - Ethernet LSP (or PVCCs) are created by configuring label lookup tables in switches
  - Only part of VID range can be used
- Operation: if VID is in the range of Ethernet switching
  - output port, new VID = Lookup(VID, MAC addr.)

Conclusions

So, is Ethernet up to carrier-grade transport network standards?

- With all the modifications: probably it is, but I doubt that it would be simple, cheap and easy-to-manage anymore...

- Fault-tolerance
  - Restoration: with RSTP & MSTP 1 s timescale and no single point of failure — comparable to PNNI
  - Link and ring protection: with tuned-up link bundles and RRSTP comparable to SDH
  - Path protection: ?

- "Cell-tax" was big issue — but does anyone care about efficiency anymore?
  - Voice over IPv6 over MPLS over MAC-in-MAC Ethernet (i.e. over 100 octets overhead per packet...)?

RFC 1925 (The Twelve Networking Truths)

...

(3) With sufficient thrust, pigs fly just fine. However, this is not necessarily a good idea. It is hard to be sure where they are going to land, and it could be dangerous sitting under them as they fly overhead.

...

(6) It is easier to move a problem around (for example, by moving the problem to a different part of the overall network architecture) than it is to solve it.

...

(11) Every old idea will be proposed again with a different name and a different presentation, regardless of whether it works.

...