Low-Latency Bridged Network Requirements

Presented by
Yong Kim, Broadcom

Content Supported by
Franz Goetz, Siemens
Oliver Kleineberg, Belden/Hirschmann
Karl Weber, ZHAW, Zurich Univ.
Christian Boiger, Hochschule Deggendorf Univ.
Objectives

Provide ultra low latency switched paths for automotive and industrial applications.
   1. automotive control loops at 100 Mb/s and above
   2. industrial control loop application over 40+ daisy chained switches at 1000 Mb/s and above
   3. datacenter?

Must provide guaranteed arrival when there is no adverse condition (network, link, or port failures).

Should be consistent with AVB architecture such that this new class could be accounted within AVB Class A and B rules.
Problem Statement

◆ “Head of line blocking”. Loosely put, a worst case is a non-low-latency packets scheduled ahead of the maximum sized frame at every scheduling point, and reduce the effect of legacy traffic on frame delivery latency bounds.
  ■ Reference: new-goetz-avb-ext-industrcom-0113-v01.pdf on hop-by-hop fragmentation on this need.
  ■ A problem w/ 100 Mb/s system (may be popular in automotive) where a max size frame is around 0.12 msec (very close to popular 8 KHz cycle) of 0.125 msec) over some number of bridges.
  ■ A problem w/ 1000 Mb/s system (may be popular in industrial control) where a number of cascaded bridges is expected exceed 40+ bridges but still need to meet end-to-end control loop latencies.

◆ This presentation explores a possible solution of: Preemption of packet-in-transmission with the support of suspend-and-resume of packet-in-transmission, and allow low-latency packets to be transmitted after a suspend.
Preemption Objectives

◆ Proper place for this work is in 802.1 – PHY agnostic low-latency services within IEEE 802 networks.
  ▪ Starts with AVB with Ethernet PHY, but may extend to other media/MAC/PHY.

◆ Between abort-and-retransmit versus suspend-and-resume, suspend-and-resume is desirable, if not required (not to be presumptuous).
  ▪ For 100 Mb/s network, 0.12 mS maximum sized frame, if preempted, may never get through if abort-and-retransmit is used and popular 8 kHz control-loop timers are used in the application.
  ▪ In general, useable BW efficiency is vastly improved (abort- may result in more than double BW usage).
  ▪ Note: “Pause” has special meeting in 802.1 and 802.3, thus ‘suspend’ is used to convey “pause” in “pause-and-resume” going forward.

◆ Consequence: AVB class A (and class B) to fully account for the worst case fragmentation overhead in its reservation.
  ▪ Preempting class packets require reservation, but not necessarily AVB shaping, and AVB class A and B should account for added low-latency BW requirements.
  ▪ A bounded default % bandwidth to be specified in similar fashion as AVB class(es).
Preemption considerations

◆“802.1 Bridge Preemption” preferred (L2-only) and Suspend-&-resume.
◆Main rationale:
  ■ No PHY changes (i.e. no revision in 802.3, 802.11, etc). This requires any L2-fragments to be well-formed L2 frame from PHY perspective. And 802.1AE (MACSEC) encryption works on fragments transparently. Ditto for FCS generation/checks.
  ■ Little or no MAC changes (i.e. any revision to 802.3 or other MACs)
  ■ Changes in 802.1, adding pre-emption services near ISS and between ISS and EISS service interfaces.
◆Pro
  ■ No PHY changes – all existing PHYs should be used as is.
  ■ TX/RX handling changes in Bridges only (to simplify standard development, solves the majority of the low-latency networking application issues)
  ■ Use of this in end-point is out of scope for standard (but *may be in scope* for products through “embedded bridge” model).
◆Con
  ■ Payload extra framing overhead – IPG/IFS added per fragmented.
  ■ Added complexity in receiver and transmitter handing.
Functional Implementation considerations

- Not a suggested implementation, but showing the major new functions on TX and RX.
Baggie pants model

The position of a VLAN-aware Bridge’s MAC Relay Entity (8.2) within the MAC Sublayer is shown in Figure 6-1.

NOTE—The notation “IEEE Std 802.n” in this figure indicates that the specifications for these functions can be found in the relevant standard for the media access method concerned; for example, n would be 3 (IEEE Std 802.3) in the case of Ethernet.

Figure 6-1—Internal organization of the MAC sublayer

Suggested preemption Q-Rev work
Preemption Transmitter (functional)

 Transmit Behavior

- If preemption capable link, and preempt-eligible frame (assumption),
  a) if scheduled to transmit, transmit with L2-fragment header, with the initial fragment sequence #.
  b) While **transmitting** (note: this method allows only one level of preemption).
    i. If preempt class frame(s) becomes available, suspend transmission at the next preempt point (i.e. min_frag_size but less than max_frag_size, e.g. every 64 byte (min_frag_size, but less than 128 byte (or reasonable max_frag_size to limit max latency) boundary, or end-of-frame) [and append valid FCS (MAC function)], and
       1) Transmit preempt class frame(s) until queue empty.
       2) Resume transmission with the L2-fragment header with the next fragment sequence #. And mark last fragment if the remaining is less that 128 bytes.
  c) If preempt sequence # is the same as the initial # (i.e. no preemption occurred with this preempt-eligible frame) at the end-of-frame transmission (pre-empt eligible but not fragmented, then transmit a null fragment with a valid next fragment sequence #.
    - Alternative to null frame transmit, a transmitter *may* create a fragment near the end of frame transmission that meets min. and max. fragment size.

Preempt-eligible: CoS mapped, or drop-eligible in pre-emption capable link all TBD.
Note: Last fragment status is set if a fragment is either null (no preemption occurred), or fragment of size between min_frag_size and max_frag_size, inclusive. The frame size need not be communicated.
Preemption Receiver (functional)

◆ Receiver Behavior
  ■ Parse RX for L2-Fragmentation header and if the header found,
    a) if new and resume-status is false, then store fragmentation packet context and initialize reassembly buffer and resume-status, and store the received payload, or
    b) elseif new and resume-status is true, then do what’s in a) above and log error (previous L2-fragmented frame aborted or lost)
    c) elseif resumed (pre-emption packet context same has previous) and resume-error-status is ok, then check fragmentation-sequence # and
      i. If the fragmentation-sequence # is right (next # from previous) then, append received payload to reassembly buffer, and if last-fragment-sequence is set in the frame (with proper null fragment handling, if null), also present reassembly buffer content to the bridge ISS, or
      ii. elseif the fragmentation-sequence # is wrong, [may not know this but] or FCS and other error(s) is wrong, set resume-error-status to error, discard the rx frame and log error.
    d) Elseif resumed, and resume-error-status is error, then discard and log discard.

New – Marked as first fragment, sequence # = initial (1),
Note:
FCS Handling: Check layer model - FCS checked and not stored – frame discard below ISS layer of bridge.
Fragmented frames may have been lost due to drop-eligibility.
Summary and next steps.

Summary
◆ Preemption meets low-latency bridging requirements for automotive (100 Mbps and above) and industrial control (1000 Mbps and above @ 40+ bridges in daisy-chain (or ring w/ RSTP or MSTP)).
◆ MAC & PHY agnostic approach preferred, if not required. to not violate layering (OSI religion).
◆ Proposed TX and RX preemption behavior is reasonable optimization to date – would welcome further improvements.

Next Steps
◆ Discussions in 802.1 to validate the baggie pants model and place within relative space between ISS and EISS.
◆ Pending above, any further consideration on preemption-eligibility and drop-eligibility, if desired.
◆ Pending above, prepare and submit PAR and 5 criteria by July plenary, and complete the work in May Interim (w/ associated motion at this meeting)
MAC Services and Bridge models in 802.1Q
(easy reference purposes)
More 802.1 Services Models

Figure 6-2—MAC entities, the MAC Service, and MAC Service users (clients)

Note: Preserve MAC services model, i.e. no changes.
Provider Backbone Baggie Pants Model

See subclause 26.8 for optional CFM stacks

Figure 25-4—Port-based service interface

Suggested preemption Q-Rev work
Provider Backbone Services Frames

Figure 25-6 illustrates the information passed over each of the ISS interfaces of a BEB.

Suggested preemption Q-Rev work