Choosing the Right TSN Tools to meet a Bounded Latency

Don Pannell
Fellow
Automotive Ethernet Networking

Sep 2019 | IP Tech Day - Detroit
The Need

- Ethernet’s high speeds saves wires in Zonal networks
- And Zonal networks bring new requirements that (TSN) solves
  - Multiple Domains using the same wire
  - Yet each Domain needs to know it’s data will get delivered in the needed maximum time – as it no longer has its own dedicated wire!
  - How to guarantee & plan the maximum bounded latency for each flow is the focus of this presentation
Overview

➢ This presentation focuses on the TSN standards that affect bounded latency of flows through the Automotive Ethernet network

➢ It briefly lists the unique problems each of these Time Sensitive Networking (TSN) standard solves & the relative ‘costs’ of using each tool

➢ Based on these numbers, a per-hop metric is proposed, to help determine which TSN tool should be used and when

➢ This priority of tool usage, makes the job of “Engineering” the network easier via to the step-by-step process described
List of Available TSN Tools for Controlling Latency

<table>
<thead>
<tr>
<th>Standard’s Name:</th>
<th>Also Known As:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strict Priority</td>
<td>802.1p-1998 / QoS</td>
</tr>
<tr>
<td>Forwarding &amp; Queueing for Time-Sensitive Streams</td>
<td>802.1Qav-2009 / Credit Based Shaper or FQTSS</td>
</tr>
<tr>
<td>Enhancements for Scheduled Traffic</td>
<td>802.1Qbv-2015 / Time Aware Shaper</td>
</tr>
<tr>
<td>Frame Preemption</td>
<td>802.1Qbu-2016 &amp; 802.3br-2016</td>
</tr>
<tr>
<td>Cyclic Queueing &amp; Forwarding</td>
<td>802.1Qch-2017 / Peristaltic Shaper</td>
</tr>
</tbody>
</table>

Note: New tools will become available, e.g., the Asynchronous Traffic Shaping, 802.1Qcr
The Shaper Standards:

What Problems the Standards Solve &
How They were Envisioned to be Used
### Strict Priority Shaper (Strict) – 802.1p-1998

- Priority solves the problem that some frames are more important than others.
- It was needed so Network Management could work.
- Management frames had to get through the Network in order to fix any Network problem.
- The Strict hardware selector is defined as: “Frames are selected from the corresponding queue for transmission only if all queues corresponding to numerically higher values of traffic class … are empty at the time of selection.”

#### 8 Traffic Class (TC) Queues

<table>
<thead>
<tr>
<th>Default Priority Code Point (PCP) to TC Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

#### Strict Priority Selector

- **Tx**: Transmission
- **xMII**: Ethernet MAC Interface
- **MAC**: Media Access Control
Credit Based Shaper (CBS) – 802.1Qav-2009

- CBS solves the problem that long bursts of data are really bad for the Bridges
- It was needed so Reserved frames are not dropped
- It caps the bandwidth a queue can transmit with hardware
- It de-bursts flows in hardware so that optimized software stacks that try to burst can be used (for streams that are not self-shaping)
  - I.e., audio from a USB drive vs. audio from a microphone or radio
  - It allows very small bursts of data to ‘catch-up’ due to momentary interference so the Reserved data rate can be maintained
- In AVB, PCP 2,3 are re-mapped above Mgmt since they don’t use 100% of the wire
Time Aware Shaper (TAS) – 802.1Qbv-2015

➢ TAS delivers the theoretically lowest possible latency for scheduled periodic data
➢ It uses significant bandwidth, so is best used as a last resort
➢ Transmission Gates are added for ALL queues just before the Strict Priority Selector
   ➢ Following a defined periodic schedule, the gates on the queues are opened or closed for a period of time – allowing critical traffic to pass without interference
➢ ALL queues are time-gated, but really only 1 or 2 queues are actually “Scheduled” and the “non-Scheduled” queues are left open during the remainder of the time
➢ Any TC can be used for “critical” scheduled traffic (TC 2 in the figure)
Preemption – 802.1Qbu-2016 & 802.3br-2016

- Preemption delivers very low latency for a limited set of non-scheduled data
- Preemption gains the most on slow data links (100 Mb/s)
- Two 802.3 MACs are used, one for “normal” preemptable traffic (pMAC) and one for “express” preemting traffic (eMAC)
- Only 1 level of preemption is supported & frames < 126 bytes can’t be preempted
- 802.1 allows connection of each TC queues to either MAC – if more than one queue connects to a MAC, the Strict selector algorithm is assumed
- In the figure, TC 1 is effectively above all the other TC’s since it can preempt them!
Cyclic Queuing & Forwarding (CQF) – 802.1Qch-2017

- A fully accurate worst case equation for the Credit Based Shaper (CBS) does not exist.
- CQF is a replacement for the Credit Based Shaper for very large, higher speed Networks.
- CQF receives data in an Even time window and transmits it in an Odd time window and visa versa using parallel buffers/queues.
- Can be built using two Traffic Class queues with Even/Odd alternate Time Aware Shaper gating to select only one queue at a time, as long as the input mapping to the two queues is synchronously time switched as well.
The Shaper Standards: Their Metrics
# Latency TSN Tool Comparison

<table>
<thead>
<tr>
<th>TSN Tool</th>
<th>Silicon Complexity</th>
<th>Engineering Complexity</th>
<th>Wire Efficiency</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strict Priority</td>
<td>Low</td>
<td>Easy</td>
<td>100%</td>
<td>Needed component, but it is not deterministic by itself</td>
</tr>
<tr>
<td>Credit Based Shaper</td>
<td>Medium</td>
<td>Easy</td>
<td>100%</td>
<td>All CBS queues are deterministic + next highest TC (for Mgmt)</td>
</tr>
<tr>
<td>Time Aware Shaper</td>
<td>Medium</td>
<td>Hard (&gt;1 TC) Medium (1 TC)</td>
<td>- Guard band - IdleOpens</td>
<td>All TAS queues are deterministic</td>
</tr>
<tr>
<td>Frame Preemption</td>
<td>High</td>
<td>Medium but only 1 level deep</td>
<td>- Fragment overhead</td>
<td>Fragmentation may affect determinism on the other flows</td>
</tr>
<tr>
<td>Cyclic Queueing</td>
<td>Medium</td>
<td>Medium</td>
<td>50% - FE 90% - GE</td>
<td>Effects are not well know % numbers due to Guard band</td>
</tr>
</tbody>
</table>

Engineering Complexity is the expected user difficulty or effort, needed to get proper results
Wire Efficiency is how much data can go down the wire – this includes critical data and background data
Note: Preemption is the only standard that requires support on both sides of the wire
Per Hop Latency – Credit Based Shaper

➢ Class A ≈ tInterval + tMaxFrameSize
  ➢ tInterval = observation interval of the Class (125 uSec for AVB – but can be changed)
  ➢ tMaxFrameSize = the maximum size of an interfering frame + gaps, etc.
  ➢ This is a good rule-of-thumb equation that results in slightly larger numbers than the equation in 802.1BA-2011 subclause 6.5

➢ Class B ≈ tInterval + tMaxFrameSize + tTimeForAllHigherFrames
  ➢ tTimeForAllHigherFrames = the time to transmit all Class A frames (+ gaps, etc.) for the duration of Class B’s tInterval (which is typically multiple Class A tIntervals)

➢ Class C ≈ tInterval + tMaxFrameSize + tTimeForAllHigherFrames
  ➢ Where tTimeForAllHigherFrames includes Class A & Class B frames

➢ Etc.
Per Hop Latency – Credit Based Shaper – part 2

- **Class A** ≈ \( t_{\text{Interval}} + t_{\text{MaxFrameSize}} \)
  - For a 64 byte frame in a 125uSec observation interval the worst case # is:
  - On a 100BASE link ≈ 125 uSec + 124 uSec = 249 uSec per hop
  - On a 1000BASE link ≈ 125 uSec + 13 uSec = 138 uSec per hop

- The observation interval is a significant portion of these latencies
  - Lower worst case latency numbers are possible on 1000BASE links by using shorter observation intervals, but don’t go below the time of \( t_{\text{MaxFrameSize}} \)
  - But lowering this number reduces latency at the cost of capacity
    - 1000 vs 100 is either 10x lower latency or 10x the capacity or somewhere in between

Note: The simplified equation on the previous page is useful for calculating the worst case latency range for a fully loaded (i.e., 75% bandwidth allocation) on a Class A link. A scheduling tool needs to use the equation that is in IEEE 802.1BA. Also see: http://www.ieee802.org/1/files/public/docs2011/ba-boiger-per-hop-class-a-wc-latency-0311.pdf
Per Hop Latency – Time Aware Shaper

➢ Store & Forward with Gate Open ≈ $t_{Device} + t_{FrameSize}$
  ➢ $t_{Device} =$ the delay through a Store & Forward bridge
  ➢ Good Rule-of-Thumb is 2 x 512 bit times + Cable delay
  ➢ or 10.5 uSec for 100BASE & 1.5 uSec for 1GBASE
  ➢ $t_{FrameSize} =$ the size of the frame passing through the bridge

➢ For a 64 byte frame the worst case # is:
  ➢ On a 100BASE link ≈ 10.5 uSec + 5.2 uSec = 15.7 uSec per hop
  ➢ On a 1000BASE link ≈ 1.5 uSec + 0.5 uSec = 2.0 uSec per hop
Per Hop Latency – Frame Preemption

➢ Store & Forward w/ Preemption ≈ \( t_{Device} + t_{FrameSize} + t_{Framelet} \)
  ➢ \( t_{Device} \) = the delay through a Store & Forward bridge
  ➢ Good Rule-of-Thumb is 2 × 512 bit times + Cable delay
  ➢ or 10.5 uSec for 100BASE & 1.5 uSec for 1GBASE
  ➢ \( t_{FrameSize} \) = the size of the frame passing through the bridge
  ➢ \( t_{Framelet} \) = 126 bytes + overhead, as this size interfering frame can’t be preempted

➢ For a 64 byte frame the worst case # is:
  ➢ On a 100BASE link ≈ 10.5 uSec + 5.2 uSec + 11.7 uSec = 27.4 uSec per hop
  ➢ On a 1000BASE link ≈ 1.5 uSec + 0.5 uSec + 1.2 uSec = 3.2 uSec per hop

Note: Preemption requires support on both sides of the wire
Per Hop Latency – Cyclic Queueing

➢ Even/Odd Buffering $\approx t_{\text{Interval}} \times 2$
  ➢ $t_{\text{Interval}}$ = observation interval of the Class (must be at least $2 \times t_{\text{MaxFrameSize}}$)
  ➢ $t_{\text{MaxFrameSize}}$ = the maximum size of an interfering frame + gaps, etc. which is needed because, when using Qbv for the Even/Odd queue selection, a guard band is needed
  ➢ The 1st $t_{\text{Interval}}$ is the reception of the data
  ➢ The 2nd $t_{\text{Interval}}$ is the transmission of the data where the worst case frame is at the end

➢ For a 64 byte frame the worst case # is:
  ➢ On a 100BASE link $\approx 250 \text{ uSec} \times 2 = 500 \text{ uSec per hop}$
  ➢ On a 1000BASE link $\approx 125 \text{ uSec} \times 2 = 250 \text{ uSec per hop}$

Note: The $t_{\text{Interval}}$ for 1000BASE could be smaller, but the limit needs to be studied
## Latency TSN Tool Comparison in Lowest Latency Order

<table>
<thead>
<tr>
<th>TSN Tool</th>
<th>Engineering Complexity</th>
<th>Wire Efficiency</th>
<th>Worst Case Latency for the Examples</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Aware Shaper</td>
<td>Hard (&gt;1 TC) Medium (1 TC)</td>
<td>- Guard band - Idle Opens</td>
<td>15.7 uSec FE Hop 2.0 uSec GE Hop</td>
<td>2</td>
</tr>
<tr>
<td>Frame Preemption</td>
<td>Medium but only 1 level deep</td>
<td>- Fragment overhead</td>
<td>27.4 uSec FE Hop 3.2 uSec GE Hop</td>
<td>3</td>
</tr>
<tr>
<td>Credit Based Shaper</td>
<td>Easy</td>
<td>100%</td>
<td>249 uSec FE Hop 138 uSec GE Hop</td>
<td>1</td>
</tr>
<tr>
<td>Cyclic Queueing</td>
<td>Medium</td>
<td>50% - FE 90% - GE</td>
<td>500 uSec FE Hop 250 uSec GE Hop</td>
<td>4</td>
</tr>
<tr>
<td>Strict Priority</td>
<td>Easy</td>
<td>100%</td>
<td>Can’t determine</td>
<td>N/A</td>
</tr>
</tbody>
</table>

1 = Multiple queues can be used with different observation intervals/latencies
2 & 3 = Assuming only 1 TC is used for very limited, very critical traffic, 2 is more available with lower latencies
4 = Effectively requires GE or faster, it’s not generally available, and its effects over different link speeds is not well understood
The Shaper Standards: Which Tool to Use First
Proposed Tool Usage Order

➢ Process the critical flows in the lowest to highest latency order

➢ First insure the total bandwidth through any link is not more that 75% loaded with these flows
  ➢ This # could go a bit higher, but 60% to 75% is a good place to start

➢ Start with the Credit Based Shaper
  ➢ Select an Observation Interval that is as large as possible that delivers the required latency over the path(s) the flow uses
  ➢ If the default 125 μSec Observation Interval is too long, reduce it, but don’t go < 125 μSec on 100BASE links
  ➢ If that doesn’t work, use Time Aware Shaping &/or Preemption as last resorts
    ➢ As these are limited resource that are less wire efficient
    ➢ Subtract any wire efficiency loss as used bandwidth toward the 75% critical flow limit
Proposed Tool Usage Order

- Multiple Credit Based Shaper’s with increasing Observation Intervals can be used – More than two Classes can be used if needed!
- Start by loading each Class no more than 20% of the link’s bandwidth
- Keep in mind that the sum total of ALL Reserved flows, & their frame (IFG, etc.) & scheduler overhead (Qbv & Qbu), must not exceed 75% of any one link’s bandwidth
- If this happens, try an alternate path for the flow
- 60% may be a better starting number so that new flows can fit in easier
  - CAN network loading is typically started at 50% so new messages can be added
    - To fix bugs & oversites
    - And to add new features

- Network Mgmt must be the highest non-CBS Traffic Class
- The remaining “non-Reserved” flows will use the remainder of the unused bandwidth in a Best Effort fashion
Summary & Proposed Queueing Model

➢ These standards are designed to work together
➢ Multiple different data delivery requirements/latencies can be supported on the same wire
➢ The Credit Based Shaper is not limited to just Audio & Video data & it is not limited to the AVP Profile’s plug-&-play parameters
➢ There is a current limit of 8 Priority Code Points (PCP) that are effectively used to indicate the “type of service” a flow needs
➢ Automotive networks are Engineered, but let the hardware enforce the needed guarantees to make the job much simpler
Disclaimers

➢ This is a really hard concept that has been simplified so that an easy starting point on which shaper to use for a target flow can be made.

➢ The listed latency numbers are in the correct range but they are still estimates. For example:

➢ A generic bridge delay is used vs. the actual delay in the specific bridge you are using.

➢ All latency numbers use 64 byte data frames. In most cases, larger data frames will impact the latency numbers.

➢ Cable delay is mostly ignored – which is approximately 80ns for 15 meters.

➢ Look at the referenced presentations & others on the same subject in the same areas.

➢ As a rule-of-thumb for link speed conversion in a bridge:

➢ For `MaxFrameSize` & `Framelet` use the egress link speed, for `FrameSize` use the ingress link speed and for `Device` use the faster link speed of the two.
### IEEE 802.1 Automotive AVB and TSN Standards Handout

<table>
<thead>
<tr>
<th>Standards</th>
<th>Transport</th>
<th>Synchronization</th>
<th>Stream Reservation</th>
<th>Quality of Service</th>
<th>Redundancy</th>
<th>Security</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AVB</strong></td>
<td>1722-2011</td>
<td>802.1AS-2011 gPTP</td>
<td>802.1Qat-2010 SRP (now Q clause 35)</td>
<td>802.1Qav-2009 Credit Based Shaper (now Q clause 34)</td>
<td>-</td>
<td>802.1X-2010 802.1Xbx-2014 802.1Xck-2018</td>
</tr>
<tr>
<td><strong>TSN</strong></td>
<td>1722-2016</td>
<td>802.1AS-Rev Redundant gPTP</td>
<td>-</td>
<td>802.1CB-2017 Frame Replication &amp; Elimination</td>
<td>-</td>
<td>802.1Qci-2017 Policing</td>
</tr>
<tr>
<td>adds CAN, FlexRay, LIN, + more Audio/Video Transports</td>
<td></td>
<td></td>
<td></td>
<td>802.1AS-Rev Redundant gPTP</td>
<td></td>
<td>802.1Qci-2017 Policing</td>
</tr>
</tbody>
</table>

Standards without an appended year are not completed yet.

Updated 9-2019
<table>
<thead>
<tr>
<th>Media Interface</th>
<th>10 Mbit/s</th>
<th>100 Mbit/s</th>
<th>1000 Mbit/s</th>
<th>2500 Mbit/s</th>
<th>10 Gbit/s</th>
<th>Next</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC Interface</td>
<td>Digital/SERDES</td>
<td>SNI, xMII/SGMII</td>
<td>xMII/SGMII</td>
<td>xGMI/SGMII</td>
<td>OC-SGMII</td>
<td>USXGMII XFI</td>
</tr>
<tr>
<td>Media Interface</td>
<td>Single Twisted Pair</td>
<td>802.3cg 10BASE-T1S 15m Point to Point 25m Multi-Drop 10BASE-T1L 1000m Point to Point</td>
<td>802.3bw-2015 100BASE-T1 15m Point to Point</td>
<td>802.3bp-2016 1000BASE-T1 15m Point to Point</td>
<td>802.3ch 2500BASE-T1 15m Point to Point</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Media Interface (PHY) Standards without an appended year are not completed yet. Updated 9-2019