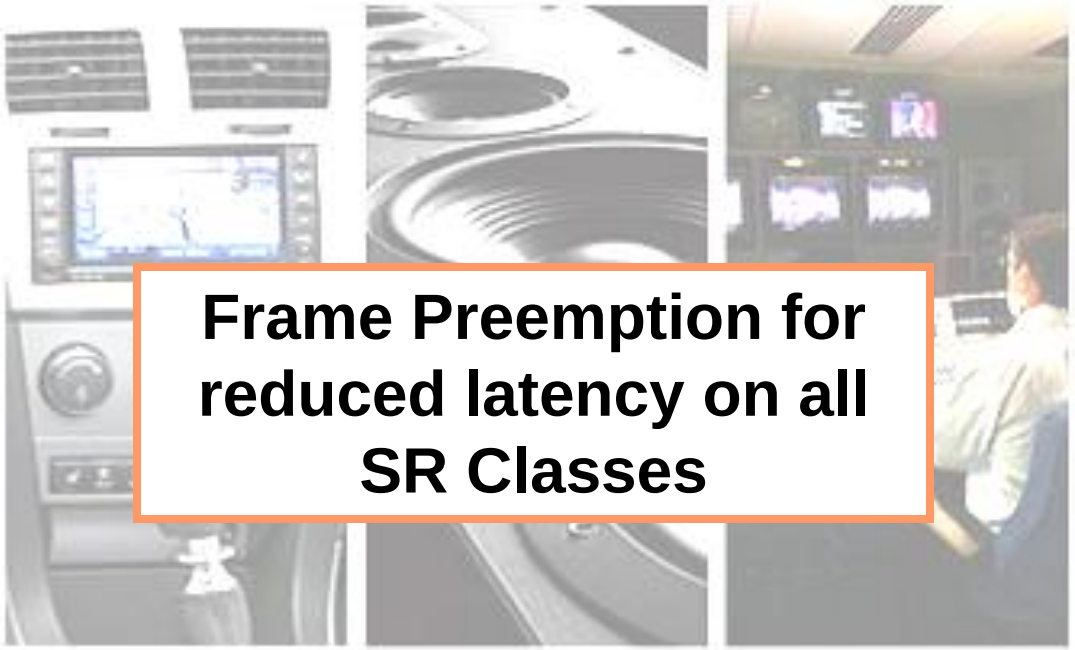


Frame Preemption for reduced latency on all SR Classes

Craig Gunther (craig.gunther@harman.com)

July 2011 – San Francisco plenary

Acknowledgements

Reference materials:

1. new-kim+goetz-Ultra-Low-Latency-Switching-v5.pdf
2. ba-kw-stream-latency-Improvements-0311.pdf
3. BA-pannell-latency-math-1110-v5.pdf
4. ba-boiger-per-hop-class-a-wc-latency-0311.pdf
5. IEEE P802.1BA/D2.5

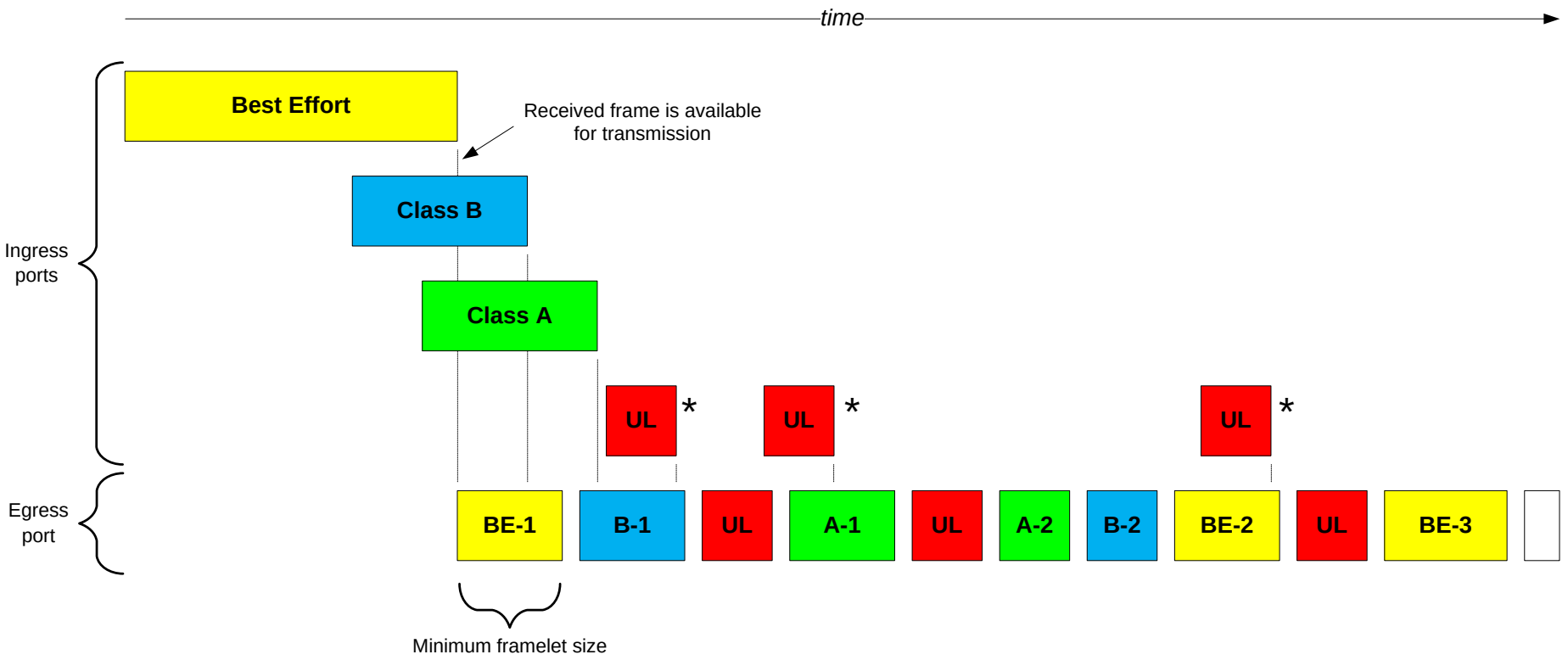
Introduction

- Ultra-low latency[1] & Preemption[1] are two separate topics
- The focus of this presentation is not ultra-low latency, but to explore the benefits of preemption for existing AVB Classes
- **Goal**: Make preemption available to all AVB shapers

Definitions

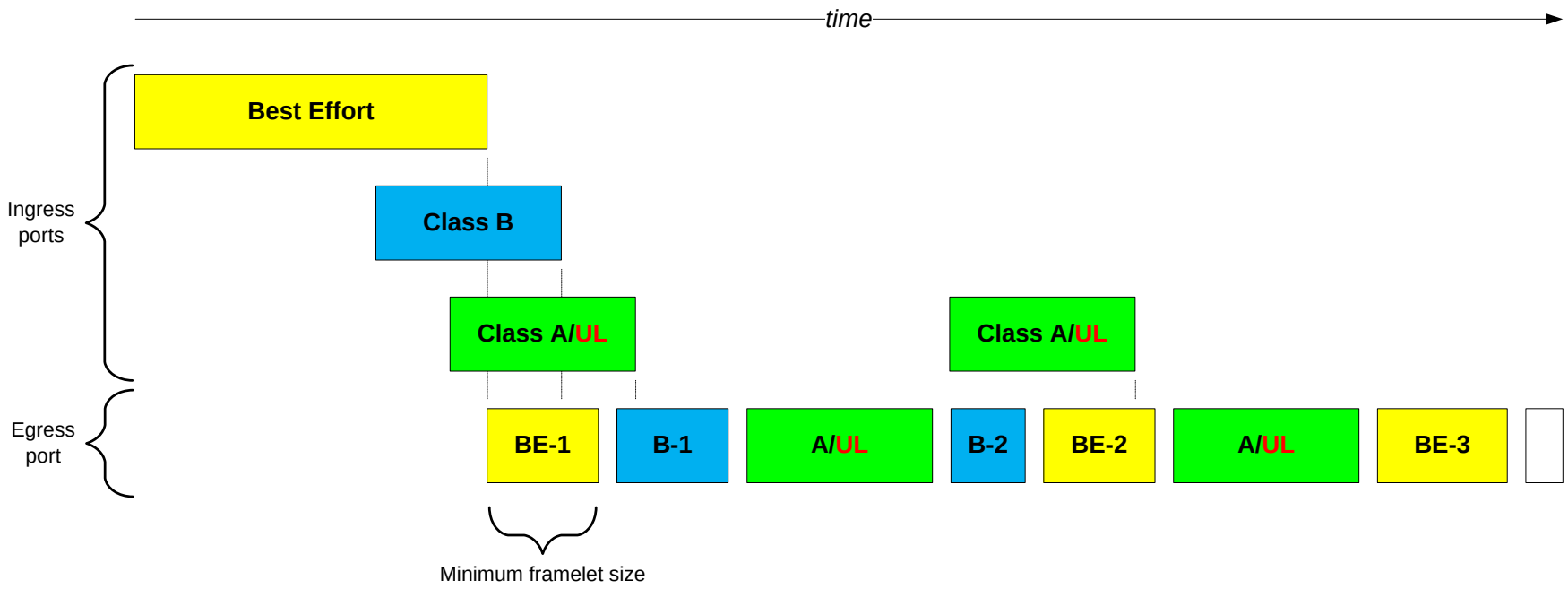
- **Interfering Traffic (IT):** frames of a lower priority which cause delays to transmission of higher priority frames.
- **Preemption:** Suspending transmission of a lower priority frame so a higher priority frame can be transmitted, followed by resumption of the lower priority frame. This can occur more than once to a large low priority frame.

Multi-Class Preemption (with separate UL* frames)



*Note: This slide assumes Ultra-Low latency (UL) frames are separate from SR Class A & B frames

Multi-Class Preemption (with UL in SR Class A)



Possible Preemption Marking

- Assume neighboring Gen-2 devices agree they can do preemption (negotiate via SRP Domain packet?). Here's two possible implementations:
  - Option 1: neighboring devices know every packet, including AVB streams, sent between them has a new header for preemption flags. This introduces wasted bandwidth and increased AVB latency if preemption is not applicable because of new flags in every packet.
  - Option 2: use a new "preempted packet" tag and only those packets have preemption flags.

New 32-bit Preemption Tag Fields

- 16 bits of EtherType
- 8 bits of control/command
  - 2-bit flag field: begin, previousEnd
  - 2-bit SR Class ID
    - Three preemption classes requires 2 bits to identify
    - Two preemption classes would require a 1-bit field
  - 4-bit sequence number per preemption class to detect dropped frames
- 8 bits unused

Reassembling the Framelets[1]

- If “previousEnd” bit is set then previous frame has been completely reassembled; pass it on
  - Sequence numbers can be used to detect missing framelets. Note that there are only 16 sequence numbers so this can fail if there are 16 missing framelets in a row.
- If “begin” bit is set then reset the class reassembly buffer pointer to the beginning
- Append framelet to per-class reassembly buffer

Multi-Class Preemption Concerns

- One Ingress buffer for each preemption class
  - Class A and Class B buffers are limited size
    - Note: Class A buffer is only needed if Class A can be preempted by UL
  - Best Effort buffer must support Jumbo frames
- MACsec, etc, concerns?
- Effects on PHY/MAC/CAM?
- How much latency added to Best Effort traffic?

Multi-Class Preemption Benefits

- Jumbo frames are back!
- Talker burst limit of two back-to-back frames[4]
  - Can we now define a latency formula?
- Gen-1 and Gen-2 switches can co-exist between Talkers and Listeners
  - Obviously preemption (and reduced latency) can only occur between Gen-2 devices
- Reduced latency for higher priority frames

Jumbo frames vs. MSRPDUs

(*"Do we need preemption for AVB BPDUs?"*)

- Wikipedia (http://en.wikipedia.org/wiki/Jumbo_frame)
  - Jumbo frame maximum size of 9000 bytes
  - No support on Fast Ethernet, only Gigabit or better
  - Super Jumbo frames ~ 64K bytes
- MSRPDUs and Jumbo frames
  - LeaveTime = 600-1000ms (802.1Q Table 10-7)
  - 9000 bytes @ FE (.08µs/byte) = 720µs, @ GE=72µs
  - 64K bytes @ FE = 5.12ms, @ GE=512µs
- MSRPDUs on Jumbo frame network are okay

(*"No, we do not need preemption for AVB BPDUs"*)

Bridge Port Latency Math with Preemption[5]

Max Latency = t_{Device} + $t_{MaxPacketSize+IPG}$ + ($t_{AllStreams}$ - $t_{StreamPacket+IPG}$) * Rate/MaxAllocBand + $t_{StreamPacket}$

t_{Device} = 5.12µs

$t_{MaxPacketSize+IPG}$ = 6.72µs (for 64 bytes preempt), 9.28µs (for 96 bytes preempt)

$t_{AllStreams}$ = 93.75µs

$t_{StreamPacket+IPG}$ = 6.72µs (assuming 64-byte stream frames)

$t_{StreamPacket}$ = 5.76µs (assuming 64-byte stream frames)

Max $Latency_{100\ MB/s}$ = 5.12µs + 6.72µs + (93.75µs – 6.72µs) * 100/75 + 5.76µs = **133.64µs**

Max $Latency_{1000\ MB/s}$ = 0.512µs + 0.672µs + (93.75µs - 0.672µs) * 100/75 + 0.576µs = **125.864µs**

	100 MB/s [x7]	1000 MB/s [x7]
Without preemption (Jumbo)	***848.52µs [5940µs]***	***197.36µs [1382µs]***
Without preemption (Super Jumbo)	***5248.52µs [36740µs]***	***637.36µs [4462µs]***
Without preemption (1522 byte)	**250.28µs [1752µs]**	**137.53µs [963µs]**
With 64-byte preempted packets	**133.64µs [936µs]**	**125.87µs [882µs]**
With 96-byte preempted packets	**136.20µs [954µs]**	**126.12µs [883µs]**

Recommendations

- New EtherType for preemption tag.
- Only SR Classes (A, B, UL) can preempt. AVB BPDUs do not appear to need to preempt.
- Update proposed PAR to support SR Class A & B preemption. It is currently limited to UL.
- No recommendation regarding combining SR Class A & UL.

Thanks

Change history

- v01 original presentation at Santa Fe
- v02 presented on 18May2011 AVB weekly call
  - Slide 2: Update presentation references
  - Slide 3: Fragmentation should only be used by UL & SR Classes
  - Slide 5: Fix packet diagram on ingress ports
  - Slide 6: Sequence #s are not optional
  - Slide 8: Address reassembly buffer requirements
  - Slide 9: Jumbo packets are more important than originally thought
- V03 changes suggested on 18May2011 AVB weekly call
  - Don't say "fragmentation" or "SaR", use "preemption" to reduce confusion
  - Slide 2: Number Acknowledgements as footnotes
  - Slide 3: Make goal of presentation more obvious
  - Slide 6: Insert new diagram for UL contained in SR Class A
  - Slide 7: Add note about a 1-bit field for preemption class ID
  - Slide 9: Add note about possible 32-bit header w/EtherType

Change history (continued)

- v04 presented at San Francisco plenary
  - Slide 7: Rework
  - Slide 8: Rework
  - Slide 10: Move 32-bit header idea onto slides 7 & 8
  - Slide 12: Insert slide on effect of Jumbo frames on MSRPDUs
  - Slide 13: Use new latency formula from IEEE P802.1BA/D2.4 (corrected). Add latencies for Jumbo (9000 bytes) frames and Super Jumbo (64000 bytes) frames.
  - Slide 14: New recommendations slide
- v05 modifications suggested during San Francisco plenary presentation
  - Slide 2: Reference 802.1BA/D2.5 for corrected latency formulas. Add hyperlinks.
  - Slide 5: Introduce term “framelets”. Add note showing when received frame is available for transmission.
  - Slide 7: Determine peer preemption capability via SRP Domain negotiation instead of LLDP . Refer to new 32-bit preemption tag (not header).
  - Slide 8: Refer to new 32-bit preemption tag (not header).
  - Slide 10: How much latency added to Best Effort traffic because of preemption?
  - Slide 11: Use term “framelets” for preempted frame pieces.