Use of Cut-Through for Industrial and Automotive Markets

IEEE 802.1 Plenary
March 2017

Avnu Industrial and Automotive Groups

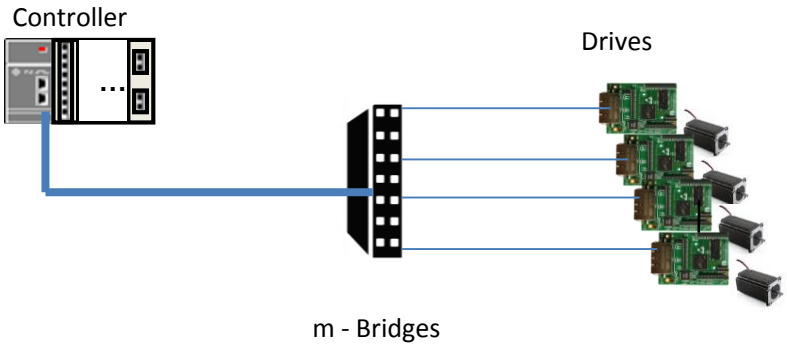
RELIABLE SYNCHRONIZED NETWORKING

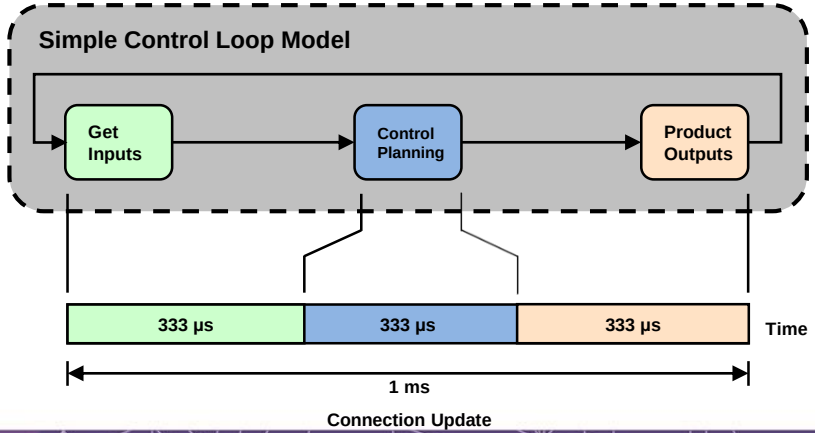
Background:

- During the November, 2016 IEEE 802 Plenary. AVnu presented a liaison requesting guidance regarding the use of cut-through with IEEE802 technologies
  - http://www.ieee802.org/1/files/public/docs2016/liaison-woods-Avnurequest-1116-v00.pdf
- IEEE Responded with a request for contributions
  - http://www.ieee802.org/1/files/public/docs2016/liaison-response-avnu-1116-v01.pdf
  - Unfortunately, AVnu did not receive this request for contributions at the January IEEE 802.1 Interim meeting until the meeting was underway. Therefore, we were not prepared to contribute to the discussion.
- However, a contributions outlining some concerns regarding the use of cut-through technologies was made at that meeting. (Thank you, Pat Thaler).
    - http://ieee802.org/1/files/public/docs2017/new-tsn-thaler-cut-through-issues-0117-v01.pdf
- This contribution is intended to continue the dialog and hopefully provide context for the discussion.

A Simple Control Model

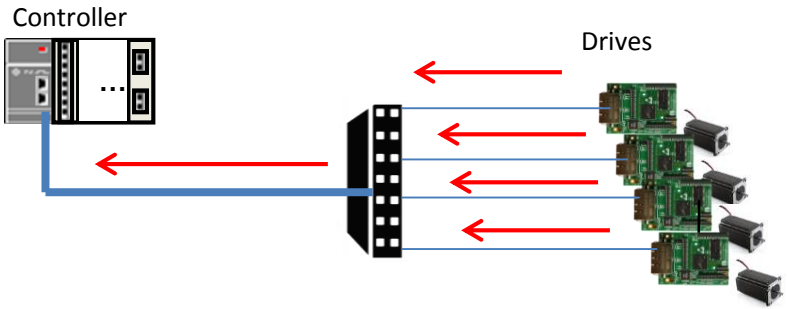
- Input data must arrive at Controller before the end of the input interval
- Planned data outputs should be transmitted before the end of the planner interval
- Output data must arrive at the drive before the end of the output interval
- Of course, this all assumes the drives and controller have a common understanding of time.

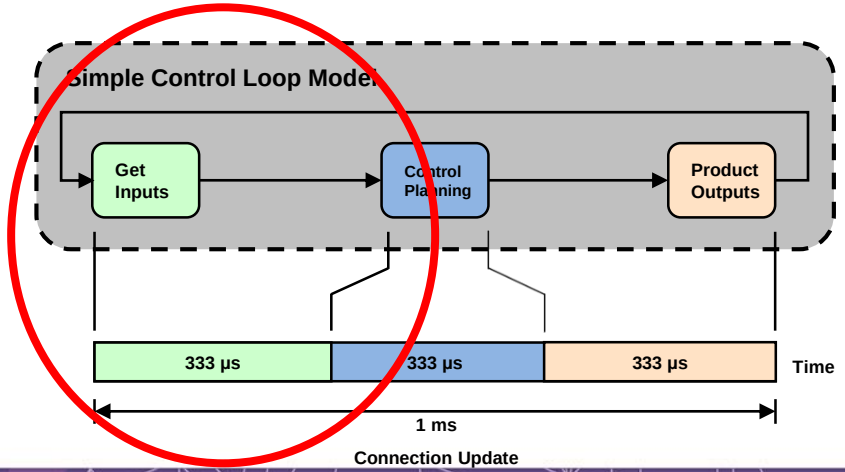




A Simple Motion Control Model

- We'll focus on a part of the problem associated with network performance
- Ideally, we'd like all of the drives to transmit their output data simultaneously
- In this way the link between the controller and bridge is optimally utilized





A Simple Motion Control Model

- Assumes all network elements are time-aware
- Assumes standard QoS/priority throughout.
- Assume cut-through switch (cut-through latency ~2usec @ 100 Mbs; ~1usec @ 1 Gbs)
  - Important for upcoming line topology discussion
- Assumes some control of traffic volume and the size of interfering traffic on the network

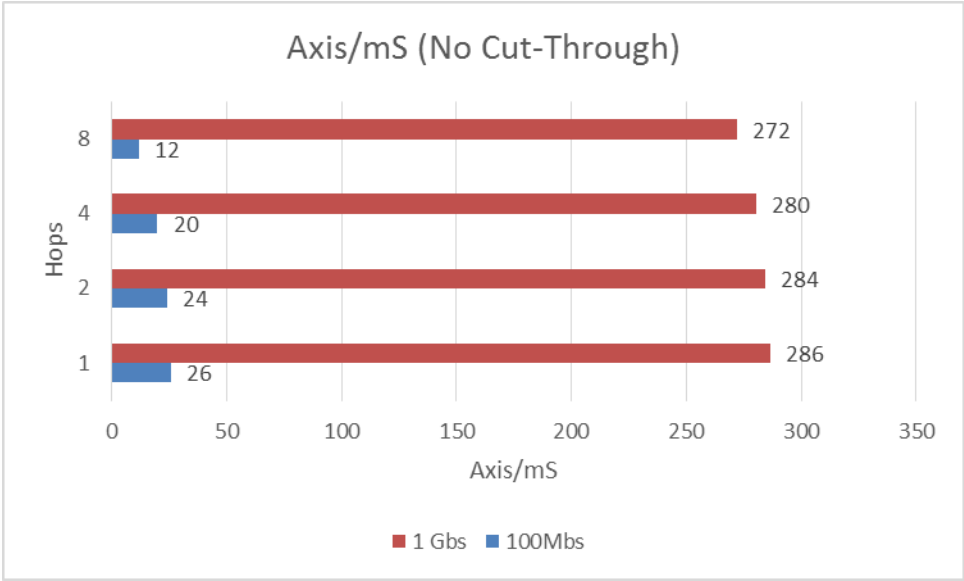
A Simple Motion Control Model

- *Max Axis = 1 + {1/3 * Connection Update Period – (Drive Transmission Delay + (m + 1) * Ethernet Transmission Time + m * Switch Latency + NIC Packet Processing Delay + Bus Interface Delay)}/NIC Packet Processing Delay*

  - *(Where m = # of hops)*
  - *Drive Transmission Delay:* We'll assume all drives have outputs queued prior to transmission, so this is contribution is small with respect to other operands, effectively 0 *usec*
  - *Assume update packets are fairly small(124 bytes), so Ethernet Transmission Time is (124+20)*80ns/byte = 11.52 usec (at 100 Mbs)*
  - *Switch Latency = (interfering packet size+20)*80ns/byte*
  - *NIC Packet Processing Delay - clever things can be done to ensure the network is the bottleneck (e.g. 2 cycle processing):* 11.5 usec for 100 Mbs, 1.15 for Gig.
  - *Bus Interface Delay: has a lot to do with the overall system architecture. could go effectively to 0 (given good bus structure, DMA/ etc.). We'll assume 0 for this analysis.*

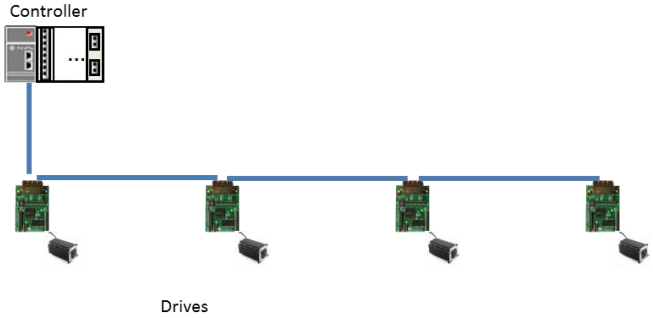
100 Mbs Baseline

- Performance strongly influenced by interfering traffic and thus, the number of hops
- In practice, control systems will engineer the network to limit the size of interfering packets (this simple example assumes no interfering traffic so the effects of switch latency can be easily examined



So what's the problem?

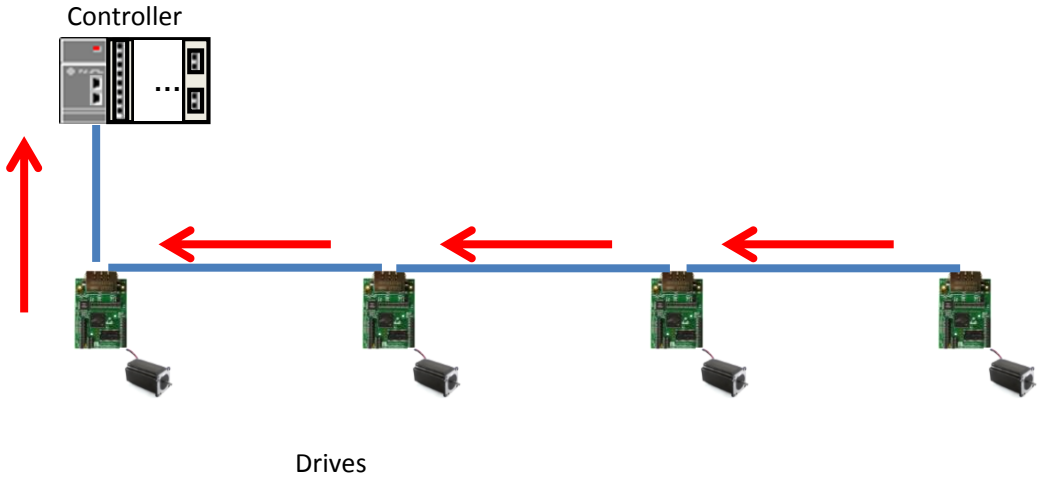
- Utilization of line topologies is prevalent in motion applications utilizing embedded switch technology
- There can be many hops along the line (64 hops or greater)
- As indicated in the model, switch latency along these hops accumulates, eating into the time available for updates.



Why line topologies?

- Physical constraints make cabling for star topologies impractical
- The construction of the application naturally lends itself to point-to-point connectivity

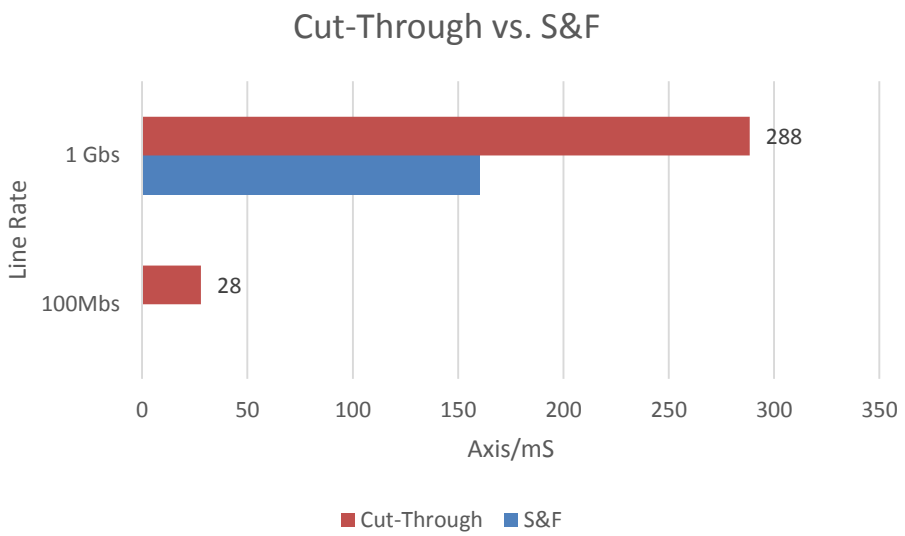
Typical Motion Topology



- The schedule of drives can be individually adjusted to compensate for drive transmission delay and switch latency (**NOTE: Schedule does not necessarily refer to .1Qbv, scheduling may take place in the application**).
- However, the effects of these delays are cumulative. Each consume part of the time available during the cycle.
- This is really a question of the accumulated latency per hop.

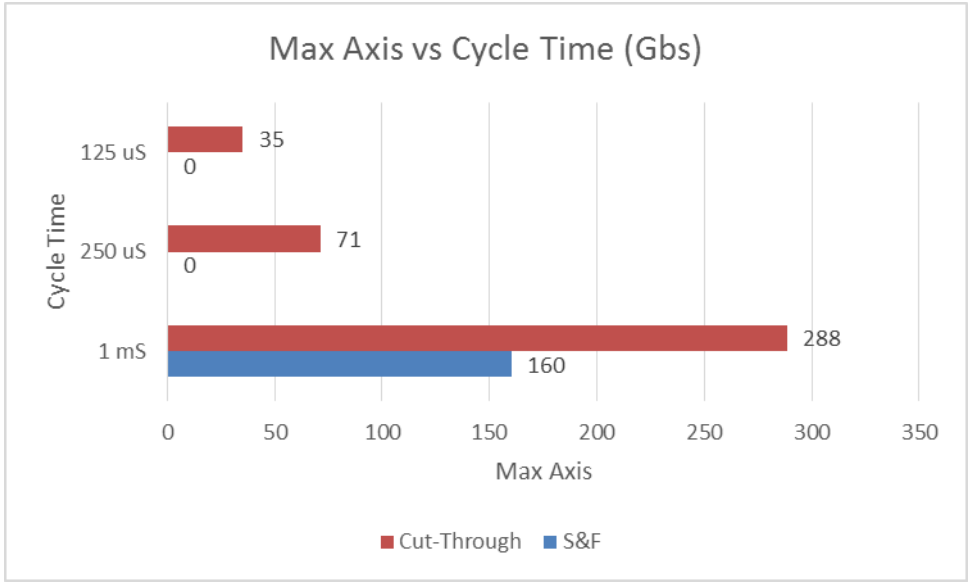
Cut-Through vs Store and Forward (line topology – 64 hops, 1 mS update)

- At 100 Mbs, a 1 mS update rate cannot be maintained using S&F.
- Why not use Gigabit?
  - Power, power and power
  - Robustness (radiated emissions and susceptibility)
  - Cost



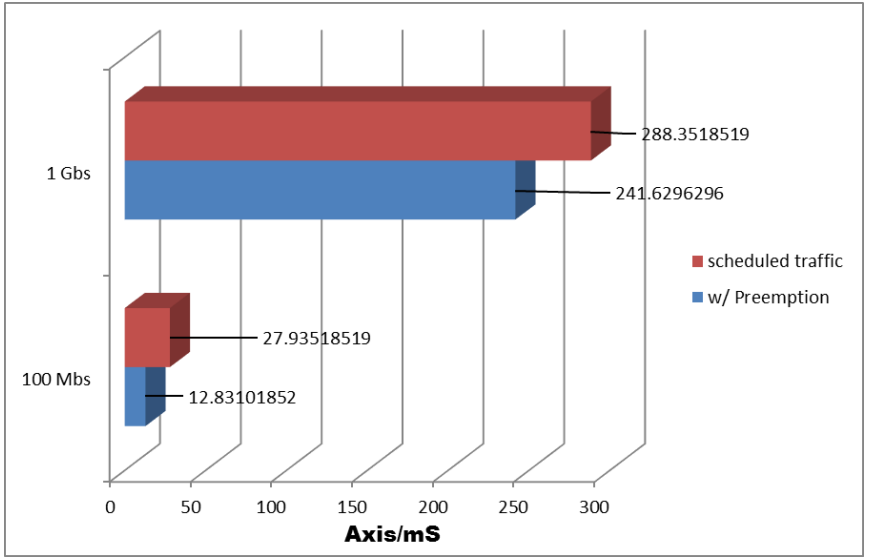
Gigabit Performance vs. Update rate (line topology – 64 hops)

- At faster update rates even Gigabit performance becomes problematic.



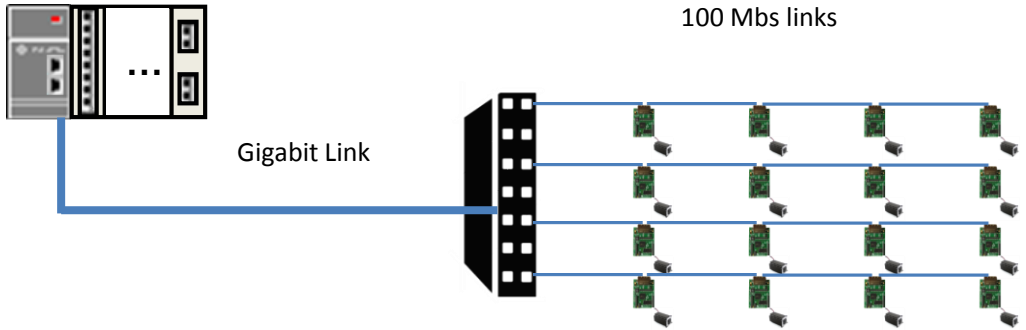
Does cut-through only apply to .1Qbv?

- No, preemption offers a means to limit the effect of interfering traffic on the control traffic without the added complexity of scheduled traffic.
- At the moment that an express frame preempts a best-effort frame, the conditions for cut-through apply, meaning that you know that the express frame can cut-through.
- Properly engineered, line topology limits the effects of interfering traffic to a single hop (i.e. control traffic is transmitted in a burst)
- With preemption, the effects of interfering traffic are minimal with respect to a 1 mS update cycle
- As we phase in 802.1Qcr (async traffic shaping), similar benefits will apply.



IEEE Cut-Through Concerns

- Limited applicability: Cut through is only useful when going between two ports running at a similar data rate.
  - True, but the use cases outlined earlier, clearly involve applications in which utilization of the data rate is engineered into the network
  - Likely use case for Gigabit:
    - Maximum utilization of controller link
    - Low EMI/Power on Drive links



IEEE Cut-Through Concerns

- Broad Market Potential Impact - Given the limitations on cut-through in speed diverse networks, is there broad market potential for cut-through?
  - Absolutely:
    - “The industrial robotics market is expected to grow at CAGR of 11.92% between 2016 and 2022, and reach USD 79.58 Billion by 2022” - http://www.marketsandmarkets.com/PressReleases/industrial-robotics.asp

IEEE Cut-Through Concerns

- Bit errors in headers
  - Bit errors in headers can change fields including address, VLAN, and priority fields
  - Cut-through occurs before CRC is checked.
  - Therefore data may be forwarded to the wrong link and/or in the wrong class of service
- Impact of incorrect forwarding
  - Deterministic latency disruption
    - Incorrectly forwarded packet may use bandwidth that has been reserved for other traffic disrupting deterministic latency.
- *True. Generally, these applications rely upon verification at the listener including CRC and application-layer mitigations. For instance, motion applications have some built-in tolerance for missed packets and specified limits for safety considerations.*

IEEE Cut-Through Concerns

- Bad BER Link identification
  - If CRC errors on incoming packets become CRC errors on outgoing cut-through packets, it may be hard to identify and correct the links that have high BER.
- *Again, these applications rely upon application-layer mitigations. Generally there is a some talker-to-listener ID which would provide visibility/traceability to the misbehaving device.*

IEEE Cut-Through Concerns

- Impact of incorrect forwarding
  - Security / Privacy
    - Packet payload may become visible on links where it shouldn't be seen.
- *Agreed. However, we submit that security mitigations should be tailored to the application space:*
  - *Is confidentiality the primary consideration in these applications? Do we care if an attacker can determine the robot's position 5 minutes (or even 5 seconds) ago?*
  - *Authentication and Integrity can be established between the talker and listener rather than hop-to-hop.*
  - *In general, mitigation strategies in highly-constrained devices are an area of concern and need to be addressed.*

IEEE Cut-Through Concerns

- Service interface
  - 802 MAC service interface is a packet interface.
    - MAC client receives and sends whole frames, not bytes or words.
  - Therefore, we have no way of formally specifying cut-through behavior.
- We assume this comment refers to the following

2.3.2.3 When generated

The MA_DATA.indication is passed from the MAC sublayer entity (through the optional MAC Control sublayer, if implemented) to the MAC client entity or entities to indicate the arrival of a frame to the local MAC sublayer entity that is destined for the MAC client. Such frames are reported only if they are validly formed, received without error, and their destination address designates the local MAC entity. Frames destined for the optional MAC Control sublayer are not passed to the MAC client if the MAC Control sublayer is implemented.

IEEE Cut-Through Concerns

- Service interface
  - We understand this interpretation, but it does not invalidate the use cases previously presented.
  - These use cases represent a large market and large installed base
  - Failure to address these use cases has the potential to affect adoption of TSN technologies

Summary

- There are significant broad market applications in which the use of cut-through technologies is required.
- Providing standards-based management of cut-through is critical to interoperability of applications supporting cut-through
- It is understood that the use of cut-through carries certain risks that must be mitigated/managed
  - Incorrect forwarding, bit errors, diagnostic, security.
- We remain convinced that standardized management of cut-through is in the best interest of our membership and associated markets.

Thank you!