PLCA Multidrop Latency, Jitter, and Scalability Enhancements

Call for Interest (CFI) Opening Report

Yong Kim, General Motors, and
Markus Jochim, General Motors.

IEEE 802.3 July Plenary, Montreal, Quebec, Canada

CFI Objectives

- To gauge the interest in starting a Study Group for **PLCA Multidrop Latency, Jitter, and Scalability Enhancements**
- This meeting will **NOT**
  - Fully explore the problem
  - Debate possible solutions
  - Choose a solution
  - Create a project documentation (PAR, CSD, or Objectives)
  - Create a standard or specification.
- Anyone in the room may vote or speak

Agenda

- **Presentation**
  - Terminology: “10BASE-T1S/M”
  - PLCA - Fair and Equal Access
  - Applications and Requirements
  - Latency, Determinism, and Scalability
  - Simulation Results
  - Problem Statement / CFI-Statement
  - Technical Feasibility
  - Market Needs – Why now?
- **Q&A Panel**
  - Yong Kim (GM), Markus Jochim (GM), Tim Bagget (Microchip), Karl Budweiser (BMW, remote)
- **Straw Polls**

Terminology

Terminology: 10BASE-T1S/M

- In this presentation the term 10BASE-T1S/M denotes both 10BASE-T1S operated in multidrop mode and 10BASE-T1M
- On some slides (e.g., slides that discuss automotive use cases) we use the term 10BASE-T1S:
  - In such cases, 10BASE-T1S in multidrop mode is implied.
  - And conclusions and simulation results on those slides apply to 10BASE-T1S/M as defined above.

PLCA (Physical Layer Collision Avoidance)

Fair and Equal Access

But First, What is PLCA?

noting only the relevancies to this presentation…

- It's entirely in Reconciliation Sublayer (RS)
- It works with CSMA/CD MAC (Clause 4)
- Signals bus cycle start ('beacon') and then all stations send its packets in round-robin. by every station by ID. Packet burst by configuration.
- When all stations are operating in PLCA, max latency is bounded.
  - Theoretical: 255 stations, 1536 octets, 256 packet burst = 80.2 s
  - Practical: 8 stations, 1536 octets, 1 packet burst = 9.83 ms
    Practical: 8 stations, 256 octets, 1 packet burst = 1.76 ms
  - Single packet transit latency (1536 octets): 1.22 ms

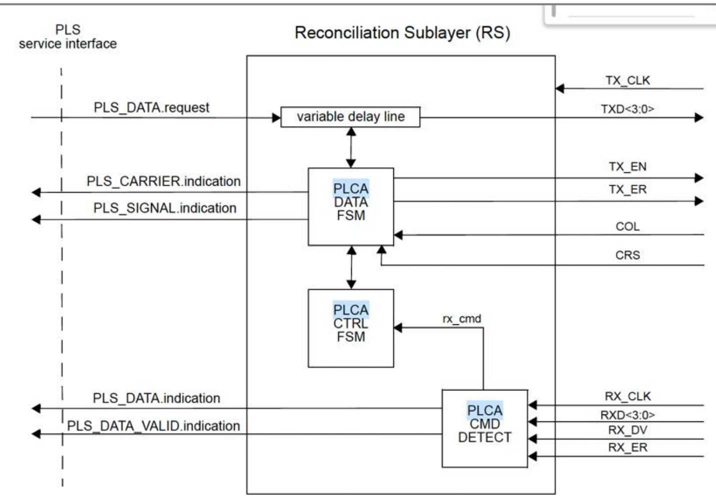
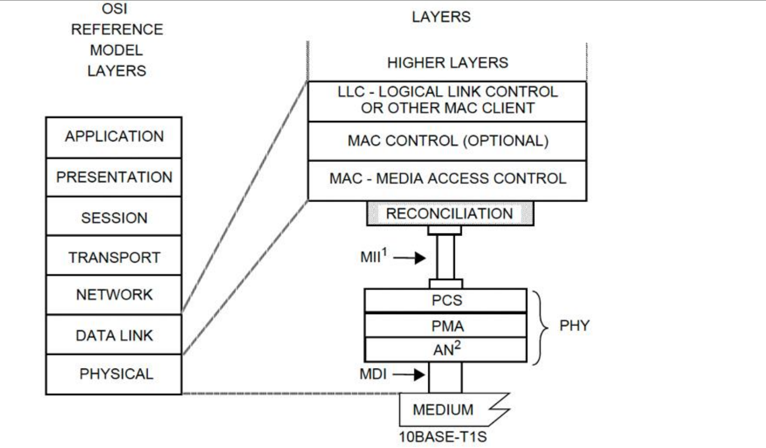


Figure 148–2—PLCA functions within the Reconciliation Sublayer (RS)



MDI = MEDIUM DEPENDENT INTERFACE
MII = MEDIA INDEPENDENT INTERFACE
NOTE 1—MII is optional
NOTE 2—Auto-Negotiation is optional

PCS = PHYSICAL CODING SUBLAYER
PMA = PHYSICAL MEDIUM ATTACHMENT
PHY = PHYSICAL LAYER DEVICE
AN = AUTO-NEGOTIATION

Figure 148–1—Relationship of PLCA Reconciliation Sublayer to the ISO/IEC OSI reference model and the IEEE 802.3 Ethernet model

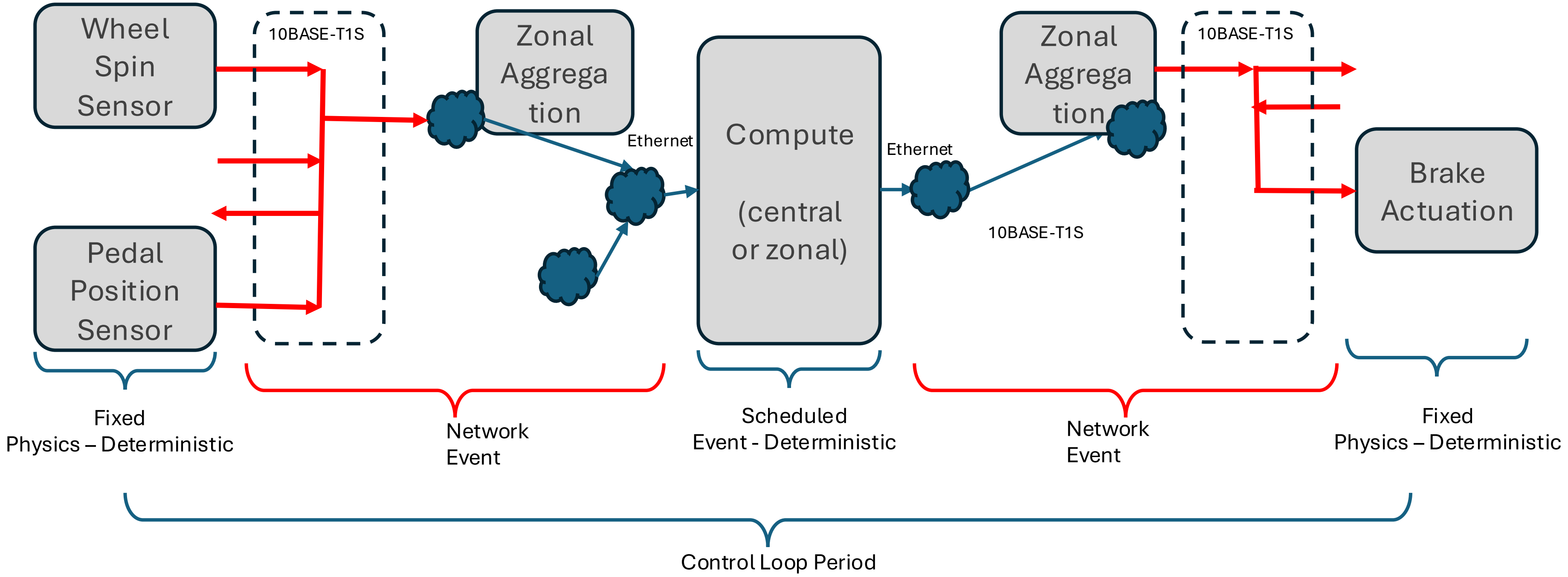
PLCA – Fair and Equal Access

- **PLCA provides fair and equal access to all stations.**
  - Not always desirable
  - A station N1 may have low latency traffic and another station N2 may have best effort traffic.
  - And in practical applications, stations frequently have mixed criticality traffic (= traffic with varying degrees of latency requirements)
- **How to be “unfair”?**
  - Only provision for treating stations differently are bursts.
  - Bursts discriminate on a ‘per station’ basis.
  - Bursts can help to empty TX queues.
  - Bursts don’t change the fact that a station that just finished transmitting needs to wait a full cycle before it can transmit again (full cycle latency).
  - Bursts can make cycles longer.
- **No mechanism for respecting frame priorities on the mixing segment**
  - The PLCA cycle defines which station gets to transmit next.
  - It is not necessarily the station that holds the highest priority frame that gets to transmit next.

Applications and Requirements

Automotive Controls Applications

- 10BASE-T1S in multidrop mode is widely adopted in automotive applications.
- [Sense] → [Compute] → [Actuate] control loop.
- The worst-case end-to-end latency often determine the robustness and feature of the control applications.



Automotive application requirements

- Typical periodic task loop times range from 1ms to 1s.
  (E.g.: 1, 2, 5, 10, 20, 50, 100, 500, 1000ms)
- Majority of loop times range from ≈ 2ms to 100ms.
- Latency and jitter requirements are application specific.
  There is no such thing as a single latency requirement number!
- For many control loops the loop time of a periodic task can serve as a rough rule of thumb frame latency requirement

Therefore:

- *Different periodic frames on the mixing segment can have very different latency requirements.*
- *Fair and equal bus access is not necessarily always desirable.*

Latency, Determinism, and Scalability

Latency

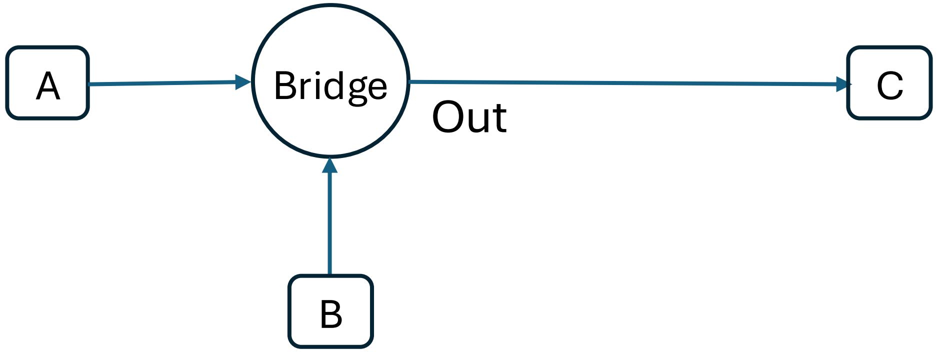
- 10BASE-T1S/M latency is bounded by number of stations, max packet sizes, and burst size.
- But 10BASE-T1S/M **latency is too long** for many control loop applications
- **Latency jitter is high**, the behavior is **not deterministic enough.**

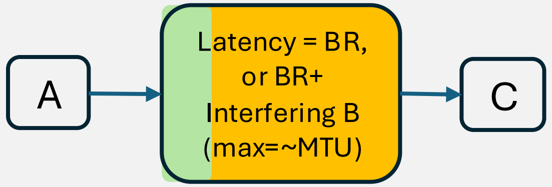
We will show simulation results that quantitative latency and jitter data in later slides.
First, let's explore the qualitative root cause of high latency and jitter.

Let’s illustrate through HoLB latency analogy

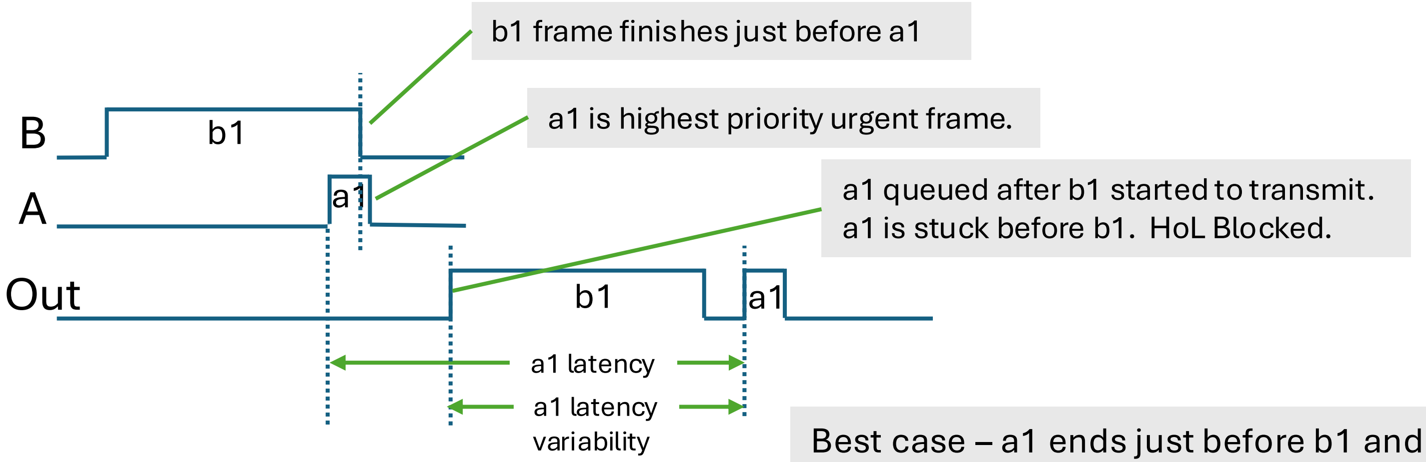
(HoLB = Head of Line Blocking)

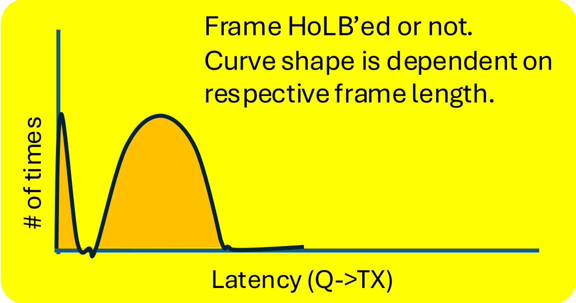
Showing a single 802.1Q Bridge hop, typical of point-to-point PHYs





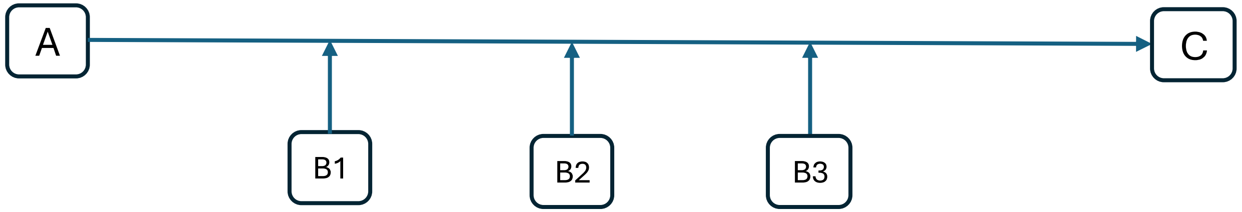
Either the minimum bridge transit delay, or additional max MTU interferer frame length delay.

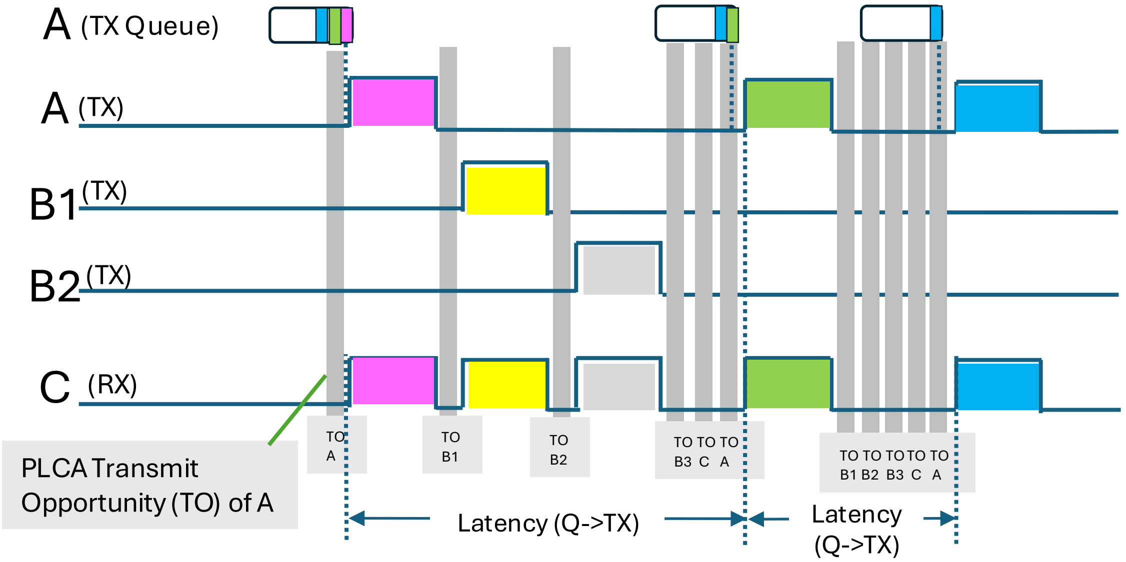


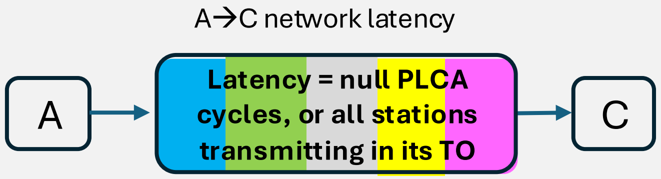


Best case – a1 ends just before b1 and queued.
Worst case – a1 ends just after b1 and HoL Bloked

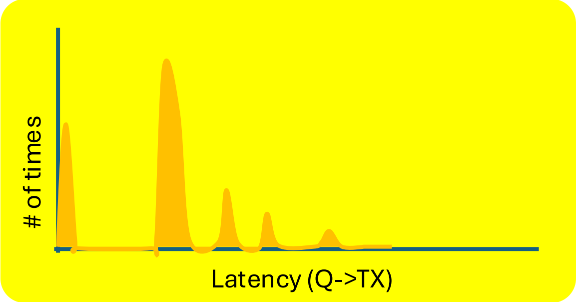
PLCA Segment latency, A→C





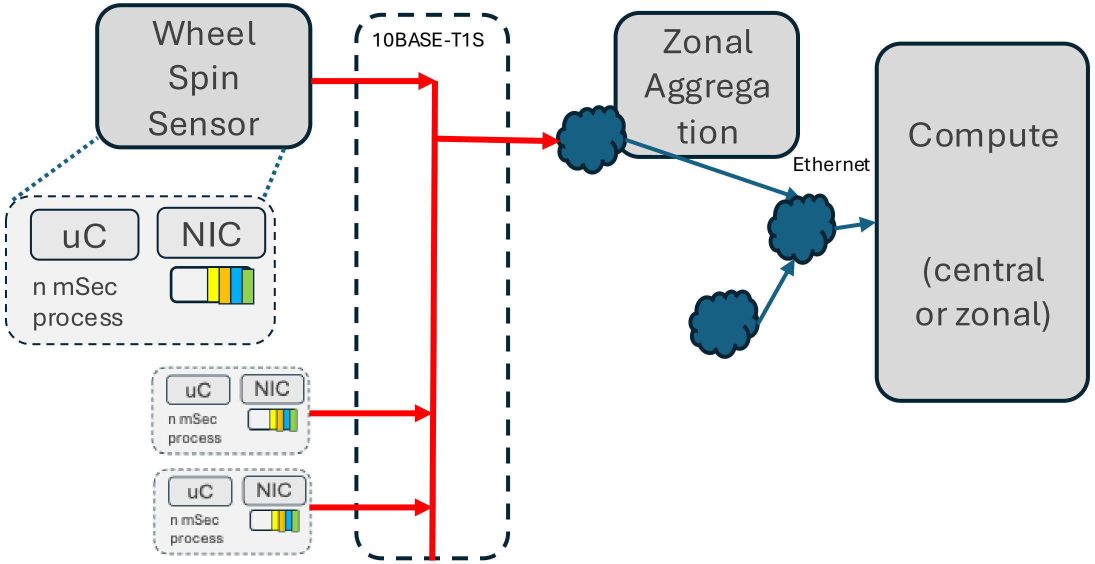


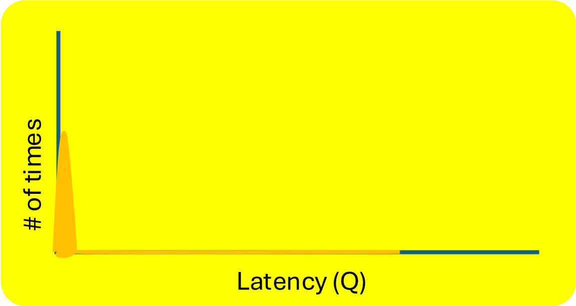
Either the minimum PLCA cycle delay, or additional frame transmissions from all the other stations.

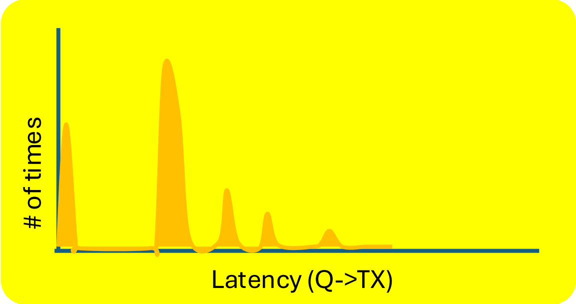


Latency in the context of control loops

- End-station operates on a control application cycle.
- Variable-length PLCA cycles (due to other stations activity and frame sizes) create discrete blocking delays in the control node’s latency.

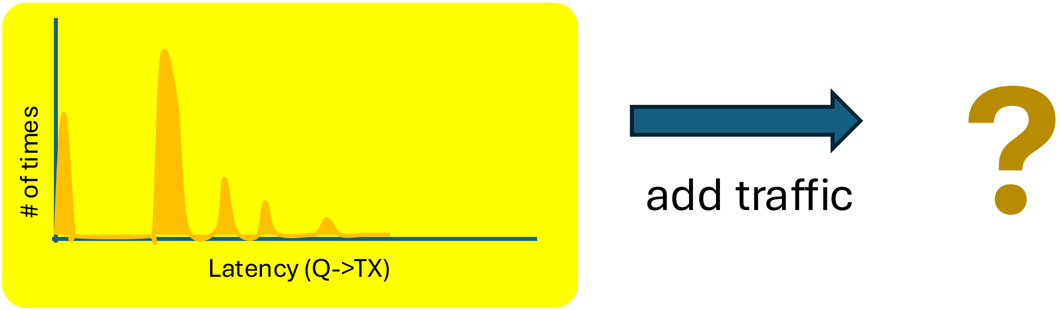






Scalability

- For our purposes, ‘scalability’ means we can introduce *additional traffic* on the 10BASE-T1S mixing segment without causing the *existing control traffic* to violate its latency and jitter requirements.



Subsequent slides show simulation results that highlight the Scalability issue we are facing with 10BASE-T1S/M

Simulation Results

Simulation setup

- We will show OMNEST simulation results of a 10BASE-T1S segment with four stations N0, N1, N2, N3, and realistic automotive network traffic.
- The next slide describes the network traffic
  - **Traffic A:** Baseline control traffic
  - **Traffic A+B:** Baseline control traffic + Additional traffic used to explore the **scalability** property

Network Traffic

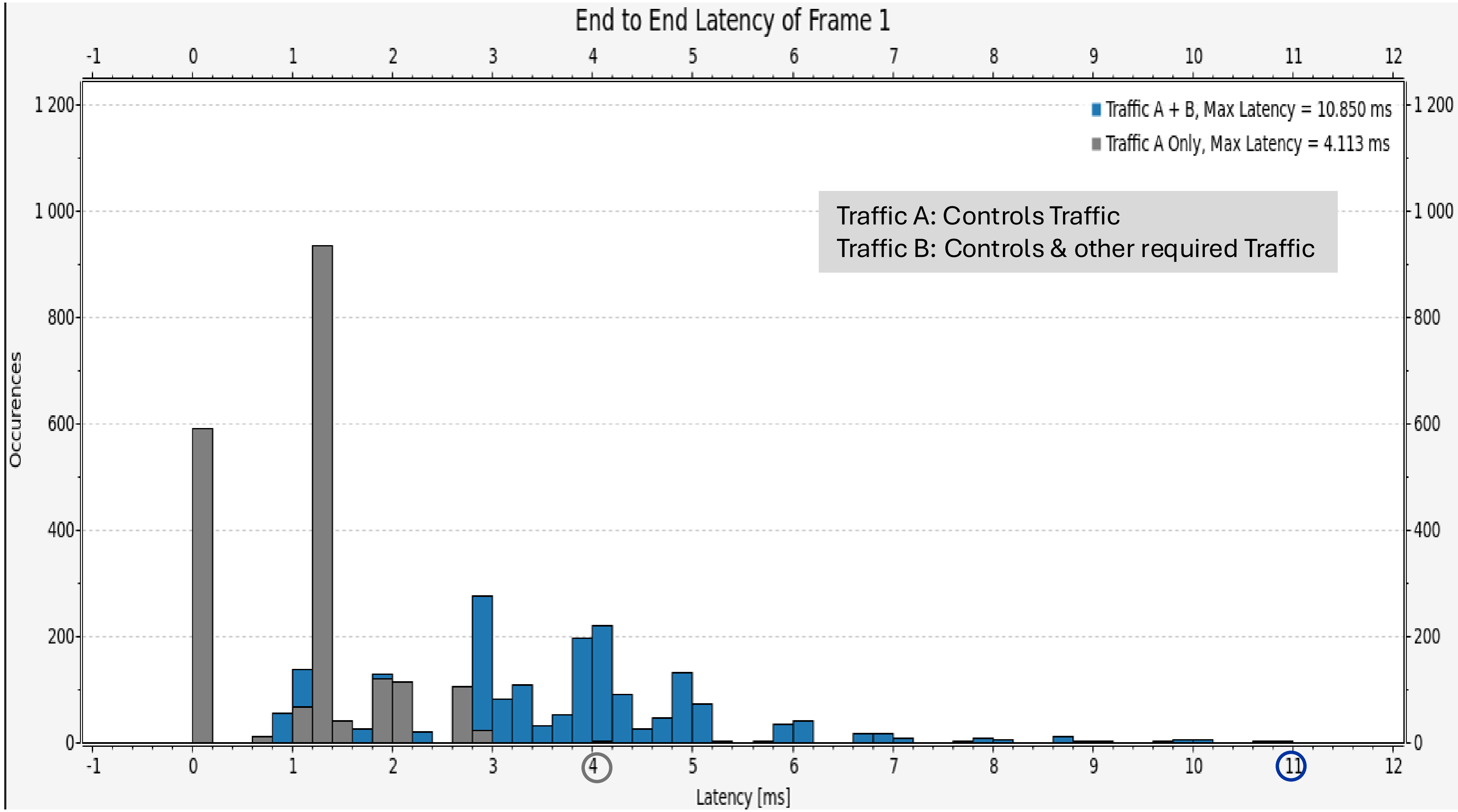
OMNEST simulations of a 10BASE-T1S segment with 4 stations with realistic automotive network traffic.

		# periodic frames	min period (ms)	max period (ms)	min length (bytes)	max length (bytes)
Traffic A:						
	N0	22	1	1000	64	617
	N1	14	5	1000	93	1305
	N2	11	10	1000	70	98
	N3	11	10	1000	70	105
Traffic B:						
	N0	2	10	25	133	311
	N1	9	2	100	93	128
	N2	4	25	50	94	106
	N3	2	25	50	99	106

Total # of messages = 75

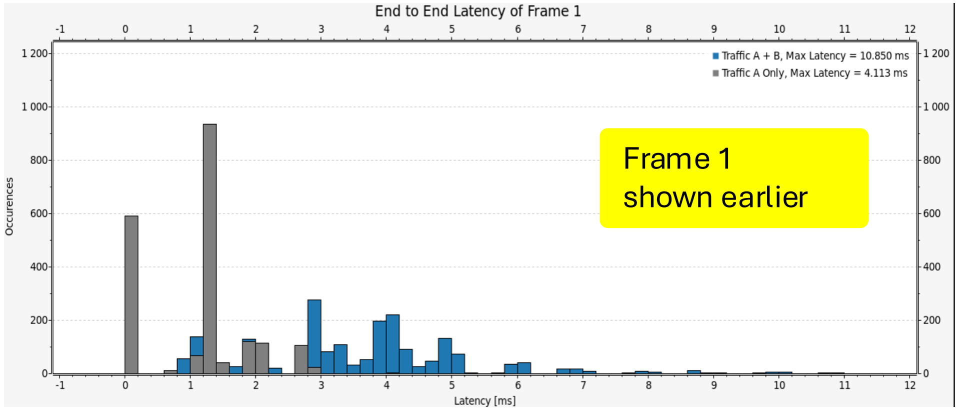
Simulation of 10BASE-T1S network

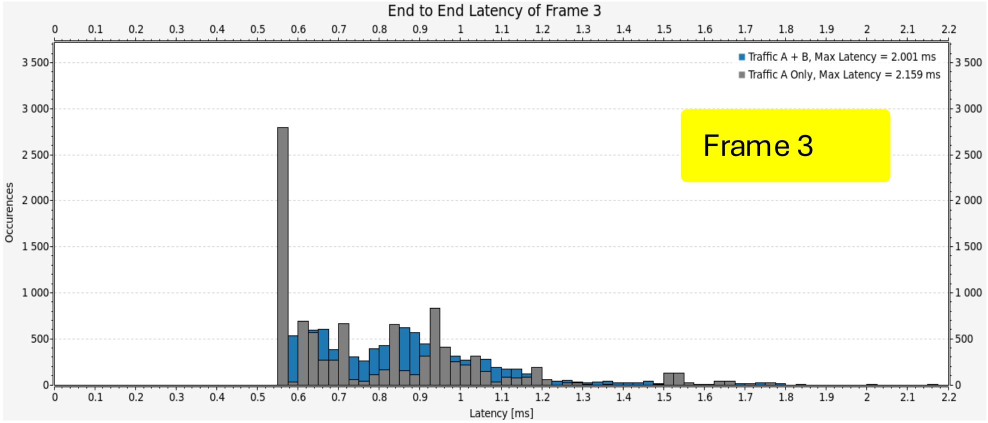
- Discrete latency distribution caused by variable length PLCA cycles.
- Challenging to engineer network for to meet latency requirements for controls traffic A.
- After adding traffic B: Near impossible to assure traffic A frames still meet latency requirements!
- Any change to the engineered network requires re-validation.

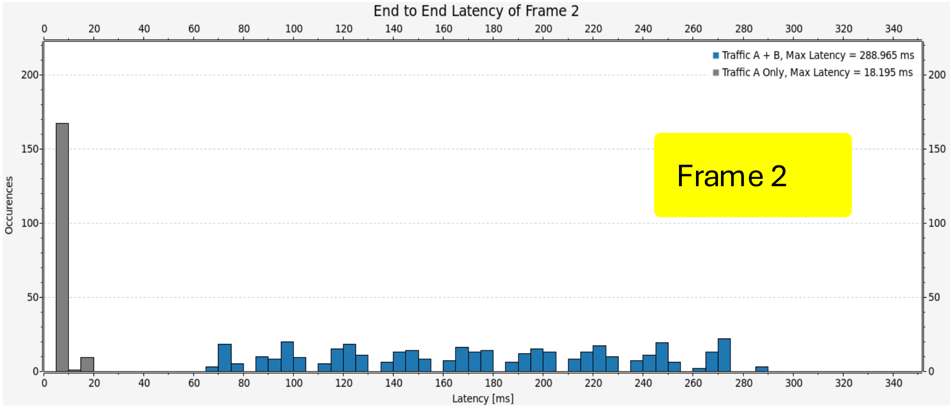


Adding traffic affects frame latencies in unpredictable ways

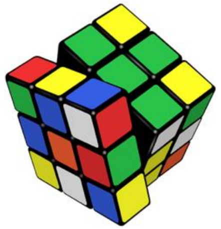
Grey:	Traffic Profile A,	58periodic frames
Blue:	Traffic Profile A+B,	58 + 17 periodic frames







	Latency impact of added traffic
Frame 1	2.6x
Frame 2	15.9x
Frame 3	1.0x



Summary of observations

- Controls application (sensor → compute, and compute → actuator) have their own cyclic periods.
- 10BASE-T1S/M PLCA cycles insert variable latencies ahead of the message on hand (top of the queue).
- Observed latency over 10BASE-T1S PLCA network is highly dependent on the other traffic from the other stations.
- **Wide range of worst-case latencies** were observed in simulation (ranging from 2.1ms to very **high latencies** such as 288.9ms)
- **High latency jitter** for periodic frames of up to 288.9 – 65 = 223.9ms was observed.
- Adding traffic affected frame latencies in unpredictable way, **making it difficult to scale** the network (= add traffic) while still meeting latency and jitter requirements for control frames.

In our experience it is possible to engineer a message latency to a desired range by limiting the number of stations, limiting traffic, and limiting MTU size.
However, these design measures are undesirable, and it is exceedingly difficult to hold all message latencies in desired ranges when traffic is added.

Problem Statement & CFI Statement

Problem Statement (1 of 2)

Mixed criticality traffic:

- Different devices in automotive networks frequently have traffic with different application specific latency requirements for their traffic.

  Example: Closure modules (door, trunk) with best effort traffic, Propulsion related control loops with low latency requirements.
- Even on a single station we often have traffic with different latency requirements.

PLCA characteristics:

- PLCA provides **fair and equal access** to all stations (1)
- The resulting PLCA **worst-case latency is too high** for control applications with low-latency requirements.
- PLCA's variable cycle length leads to **high latency jitter** (lack of sufficient determinism).
- PLCA has **no mechanisms that enable tailoring latency and jitter** to meet traffic requirements.

(1) As stated before, bursts are the only provision for 'unequal' access.

Problem Statement (2 of 2)

<u>Scalability:</u>

- Automotive networks are engineered, but over time networks evolve. Traffic changes due to over-the-air updates or because new features are added from model year to model year.

<u>PLCA characteristics:</u>

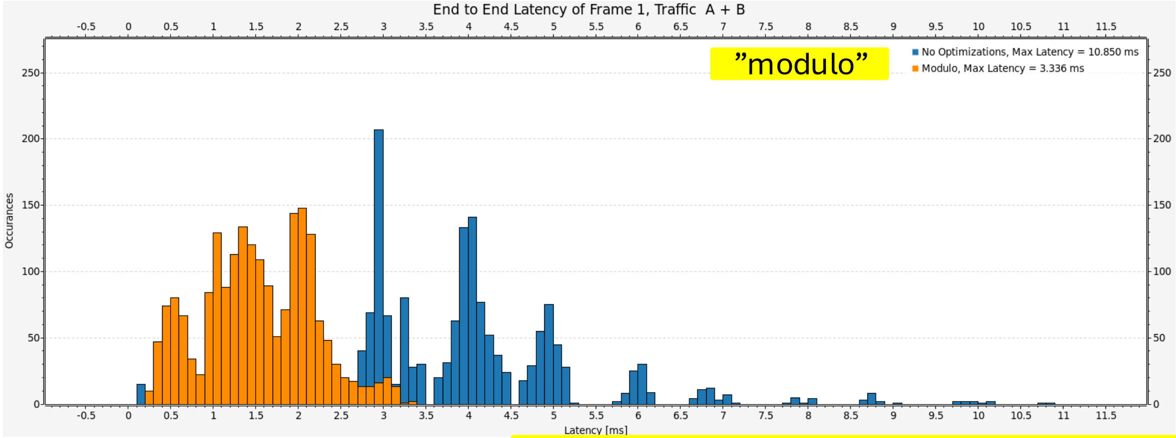
- 10BASE-T1S/M segments can be engineered to meet latency requirements (2)
- But **when traffic is added** to a previously engineered segment, it **impacts latencies and jitter in unpredictable ways** (scalability concern)

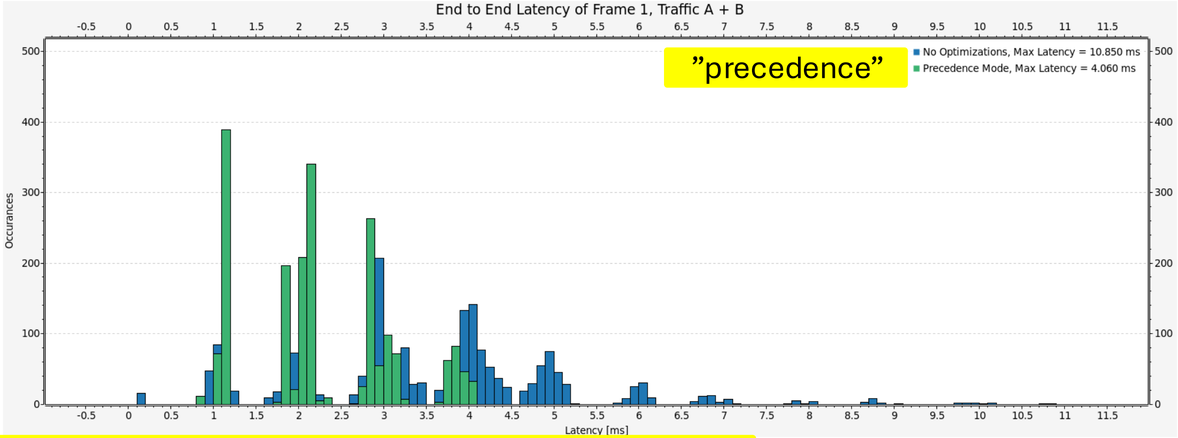
We need to study the latency, jitter and scalability related concerns to get a deeper understanding of the issues that enables us to define scope and objectives of a subsequent project.

Technical Feasibility

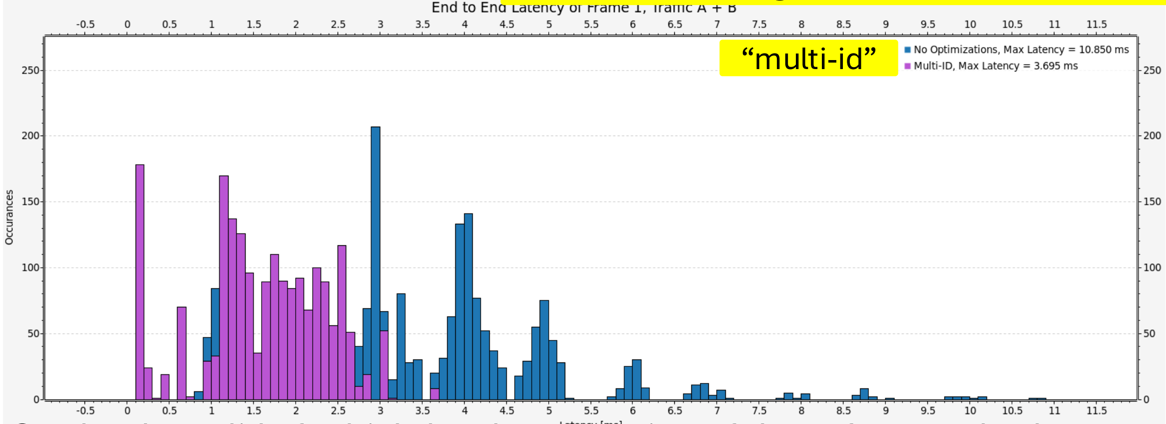
Technical Feasibility – existence there of

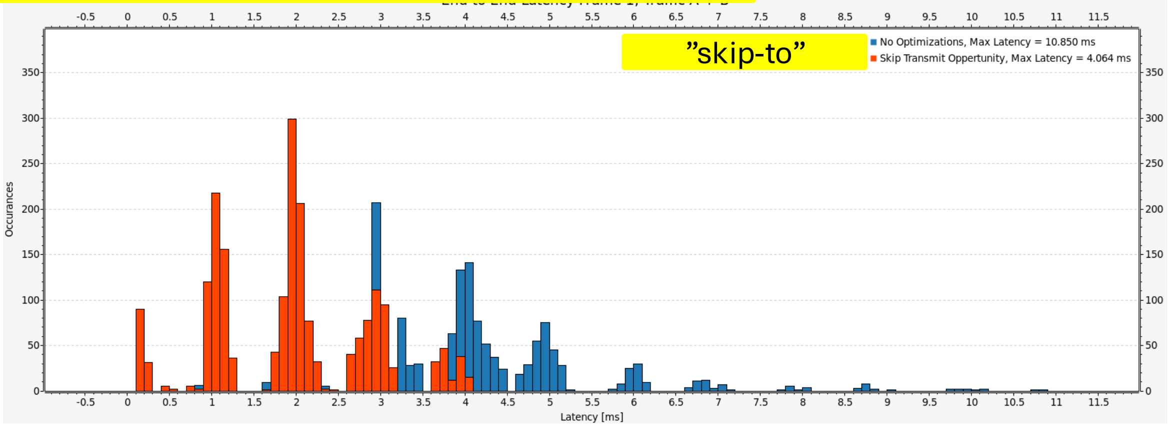
Several proprietary methods that were designed to address latency and jitter concerns exist.
One or more of these method may be used to enhance determinism and latencies of critical traffic.





Caveat. These diagrams are for the latencies of one message. What about the other 74 messages?



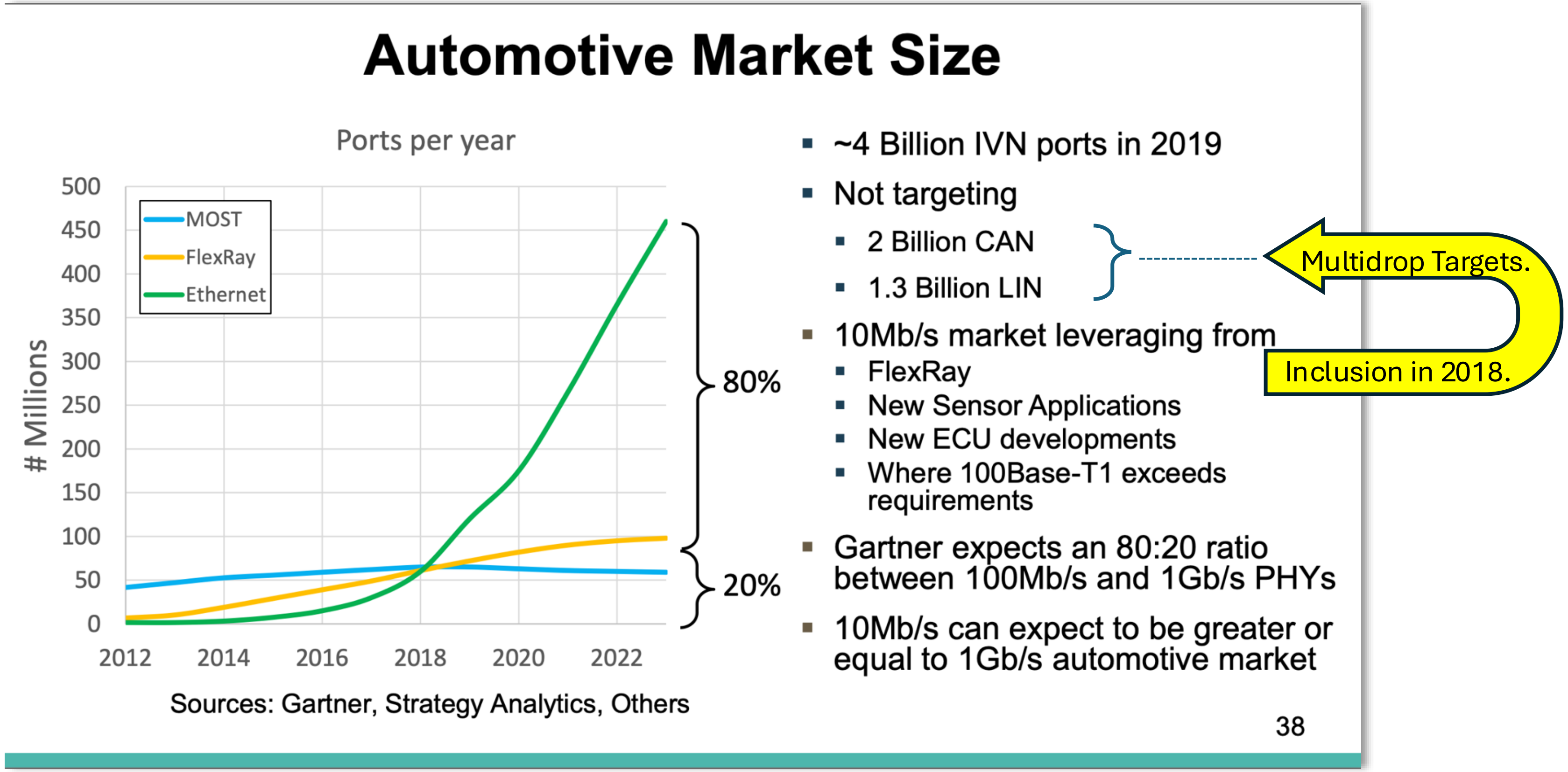


See backup slide for high-level explanation of these four methods

Market Needs - Why now?

Automotive Market Size

- From 802.3cg CFI (did not explicitly mention multidrop), 2016
  https://www.ieee802.org/3/cfi/0716_1/CFI_01_0716.pdf, slide 38.



IVN Market Evolved Around Single Pair Multidrop

IVN = in-vehicle network

- 2016 • Key market application for the Single Pair Ethernet operating at 10 Mb/s short-reach was for automotive networks when 802.3cg task force was formed.
- 2018 • 2018 – Active work on single pair multidrop (PLCA) in 802.3.cg to serve automotive Ethernet needs.
- 2020 • The automotive market for 10BASE-T1S has evolved around multidrop networks. Automotive application is primarily in mechatronics controls applications, body, chassis, etc controls, often replacing or augmenting CAN and LIN buses. Appropriate for the low-speed aspect of 10 Mbps Ethernet. OPEN Alliance helped to accelerate adoption.
- 2022 • Very few 10BASE-T1S full-duplex point-to-point installations exist due to component availability.
- 2022 • Vehicles with 10BASE-T1S networks are in development and in production planning in very near future.
- 2022 • Many early adopters of production systems ran into non-deterministic latencies
- 2024 • Many proprietary scheme are already available that expands the ‘toolkit’ for the engineered network design work. Some of these schemes address determinism, not just relative priority.
- 2026 • **Time to start a study group to address this need and provide interoperable standardized solution to this market.**

Q&A

Expert Panel:

- Yong Kim, General Motors
- Markus Jochim, General Motors
- Tim Baggett, Microchip
- Karl Budweiser, BMW   (remotely)

Supporters

Amir BarNiv, Infineon
Ariel Lasry, Qualcomm
Ben Mecklenburg, Huawei
Brian Macfarlane, NXP
Christian Boiger, Infineon
Christoph Arndt, AUMOVIO
Dance Wu, Infineon
Daniel Hopf, AUMOVIO
David Bollati, C&S Group
Dieter Pröll, Siemens
Geet Modi, Texas Instruments
George Zimmerman, CME Consulting
Hiroshi Ono, ADI
Joe Triggs, ADI
Jörg Kock, NXP
Max Turner, Ethernovia
Natalie Wienckowski, IVN Solutions LLC
Piergiorgio Beruto, onsemi.com
Pusik Park, KETI

Shen Jack, Texas Instruments
Thomas Hogenmueller, Bosch
Thorsten Hoffleit, Renesas
Tim Baggett, Microchip
Tingting Zhang, Huawei
TJ. Houck, Infineon
Tobias Belitz, Renesas
William Lo, Axonne
Yutao(Tony) Wang. Rockwell Automation
Zhe Lou, Huawei
Zhu Yue, Yinwang Intelligent Tech.

Amrit Gopal, Ford
Anton Schedl, BMW
Helge Zinner, Cariad / VW
Hideki Goto, Toyota
Karl Budweiser, BMW
Olaf Krieger, Cariad / VW
Steve DiBella, General Motors

Call For Interest

10BASE-T1-S multidrop are being widely adopted in automotive Ethernet applications across e.g., body, chassis, and powertrain control applications, whose functions span a broad range of latency and jitter requirements, from tightly bounded control loops to more timing-tolerant comfort and convenience functions.

The Physical Layer Collision Avoidance (PLCA) layer developed as a part of 802.3cg and refined in 802.3da provides a fair and equal access to stations on the mixing segment while providing bounded maximum access latency. The maximum access latency is not sufficiently bounded for many of the tightly bounded control loops. Likewise, the variable length of the PLCA cycle results in high jitter (lack of determinism) that such applications cannot tolerate. PLCA as defined today is cannot treat traffic differently based on latency and jitter requirements. Additionally, adding traffic to multidrop networks engineered to meet latency and jitter requirements can change the latency and jitter of existing traffic in ways that are difficult to predict resulting in scalability issues and the need to frequently reengineer networks.

Proprietary schemes are emerging in the market to address these needs.

This is a call for interest to initiate a Study Group to explore the market requirements for an interoperable solution and to develop a PAR and CSD for PLCA Multidrop Latency, Jitter, and Scalability Enhancements.

https://www.ieee802.org/3/cfi/request_0726_1.html

Straw Polls

Straw Polls

- Should a study group be formed to develop a PAR, CSD responses and objectives for “PLCA Multidrop Latency, Jitter, and Scalability Enhancements”
  Y: ___ N: ___ A: ___

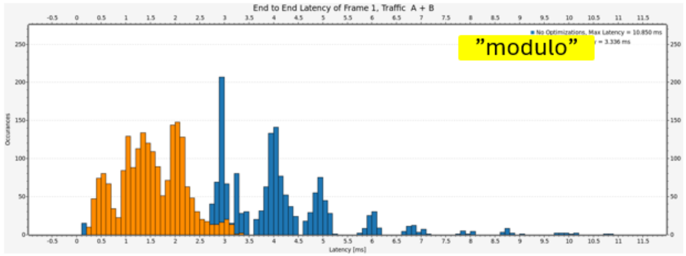
- If formed, will you participate in this Study Group
  Y (Individuals): ____

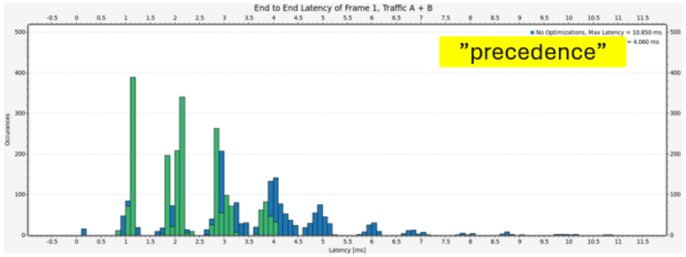
- Unique affiliations for those who indicated they’d participate (post-processed after this meeting): ____

Thank You for Joining

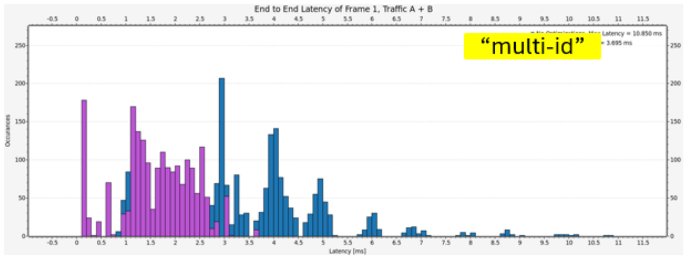
Backup

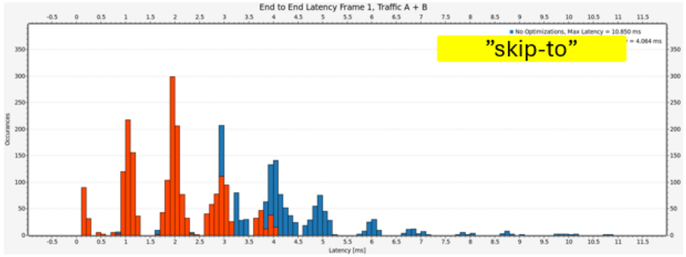
High-level description of publicly available proprietary methods for addressing latency and jitter concerns





Caveat. These diagrams are for the latencies of one message. What about the other 74 messages?





- **Modulo:**
  Establishes a schedule that determines which PLCA ID can transmit in which cycle. See link for details.

- **Precedence:**
  In each cycle, only one PLCA ID can transmit. The first PLCA ID that leverages it TO ends the cycle.

- **Multi-ID:**
  A node can have one or more PLCA IDs. Nodes with multiple PLCA IDs can transmit more than once per cycle.

- **Skip-TO:**
  Nodes can be configured to remain silent for n cycles once it leveraged a TO and transmitted a frame.