Preamble

The subsequent slides (not including this slide) contain **draft** material proposed for an IEEE 802 tutorial on CTF (see https://mentor.ieee.org/802.1/dcn/21/1-21-0030-00-ICne-ctf-ieee-802-tutorial-request-2021-05-21.pdf):

- The contents of the following slides are designed to show content as it would like in a final version.
- All contents in this slide set are subject to discussion, change/correction, removal and addition. Nonetheless, this slide set is intended to give a preview of the merged individual contributions for the IEEE 802 tutorial.

Discussions and contributions are welcome!

IEEE 802 Tutorial: Cut-Through Forwarding (CTF) among Ethernet networks

Johannes Specht, Jordon Woods, Paul Congdon, Lily Lv, Henning Kaltheuner, Genio Kronauer, Alon Regev

Abstract

Cut-Through Forwarding (CTF) is a known method to improve the delay performance in bridged Ethernet networks.

CTF is already implemented in many commercial products and is therefore technically feasible. Standardizing CTF in IEEE 802.1 and IEEE 802.3 would enable interoperable implementations.

The goal of this tutorial is to motivate standardizing CTF - the tutorial introduces CTF on a technical level, explains application areas, markets and use-cases for CTF, and addresses aspects of standardizing CTF.

Disclaimer

This presentation should be considered as the personal views of the presenters not as a formal position, explanation, or interpretation of IEEE.

Per IEEE-SA Standards Board Bylaws, August 2020:

At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that his or her views should be considered the personal views of that individual rather than the formal position of IEEE.

Table of Contents

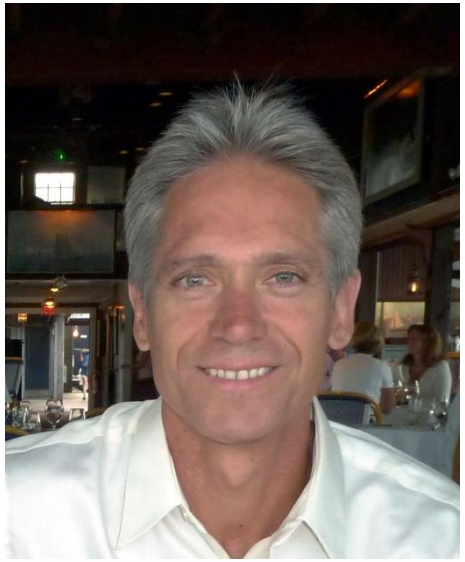
Introduction: Speakers

Speakers

Johannes Specht

Jordon Woods

Paul Congdon



Henning Kaltheuner

Alon Regev

Speakers' Biographies

Johannes Specht is a consultant and a researcher. His research area is the real-time traffic scheduling aspects of TSN. He holds a diploma in applied computer science and is currently finishing his PhD at the University of Duisburg-Essen (Germany).
He has been the technical editor of the IEEE P802.1Qcr - Asynchronous Traffic Shaping (ATS) project, and an active technical contributor to several IEEE 802.1 TSN standardization projects on real-time traffic scheduling, reliability, fault tolerance, time synchronization, and configuration since joining IEEE 802.1 nine years ago.
Johannes has been providing expert consulting to General Motors (USA) on using Ethernet for safety-critical applications since 2012. His professional career started in 2003 in the automotive industry, where he contributed to several projects on communication systems, testing, and functional safety.

Jordon Woods is a strategic technologist for Analog Devices Industrial Ethernet Technology Group (IET). IET enables seamless and secure connection of customer products across the entire landscape of Industrial IoT. Woods has 35 years of experience in the semiconductor industry. He is familiar with a variety of Ethernet-based Industrial protocols including Profinet, EtherNet/IP, as well as IEEE TSN standards. He is also a voting member of the IEEE 802 working group defining new Ethernet standards for Time Sensitive Networks and the editor of the IEC/IEEE 60802 Time-Sensitive Networking Profile for Industrial Automation.

Paul Congdon is a co-founder and is the Chief Technology officer (CTO) of Tallac Networks. He has over 34 years of experience in the networking industry and has become a widely esteemed inventor and leader in the networking industry. Prior to Tallac Networks, Paul was a Fellow at Hewlett Packard's Networking and Communications Labs with responsibility for HP's research in mobility, wireless and SDN network infrastructure. Paul has led, chaired, and is currently contributing widely to industry standards in the IEEE and IETF. Paul has a PhD in Computer Science from the University of California, Davis.

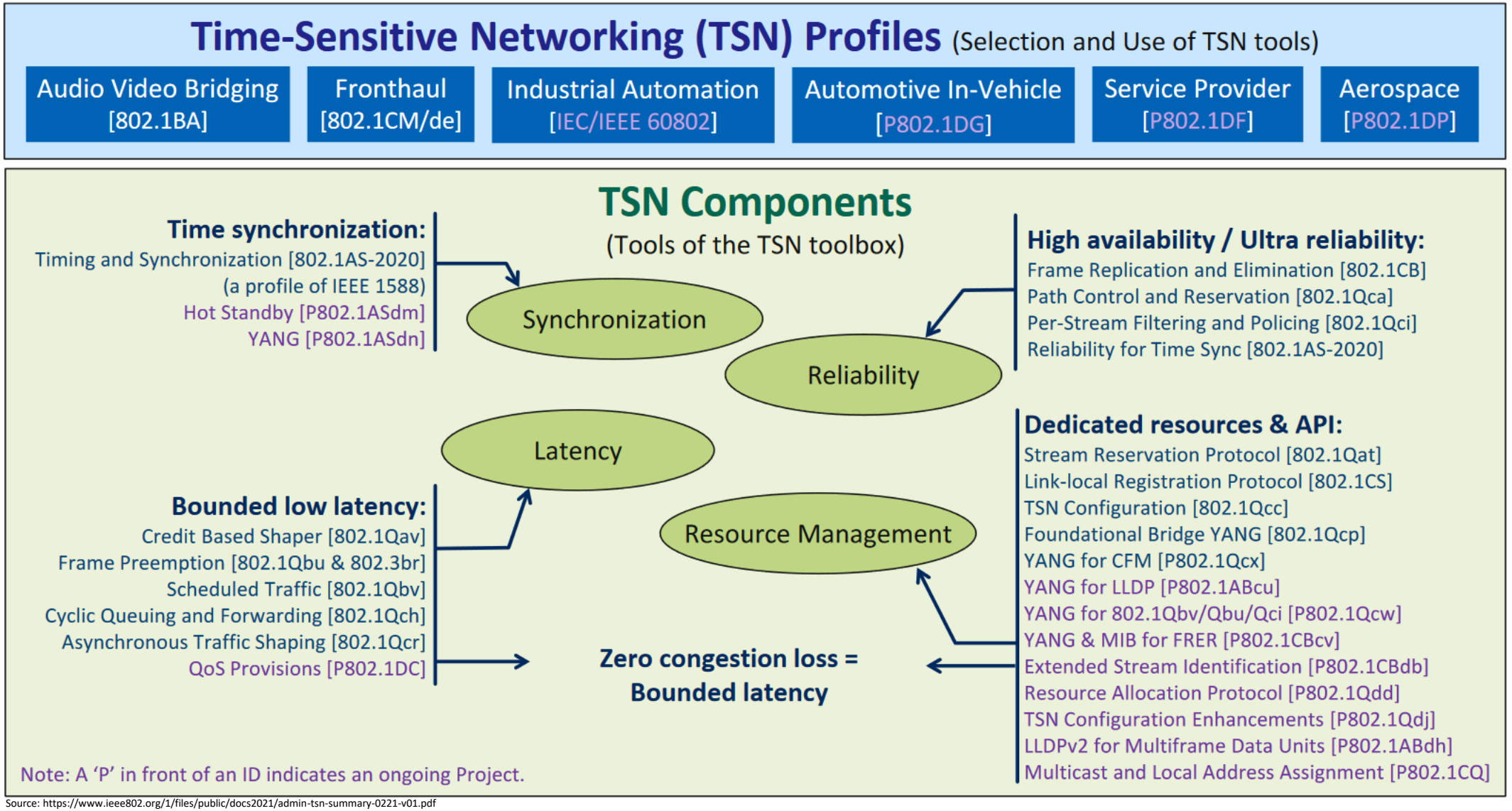
Henning Kaltheuner is Head of Business Development and Market Intelligence at d&b audiotechnik since December 2013. He knows the pro audio industry inside and out, having worked in positions ranging from Front of House to senior manufacturing roles. Before joining d&b Henning worked e.g. for brands like Riedel and Yamaha. Besides both his hands-on and managing experience in the pro audio and broadcast industry Henning's main expertise is market research based on his master degree in psychology at University of Cologne with a specialization on media psychology and qualitative research. His work has concentrated on gaining insights into market trends, brand perceptions and customer expectations in the pro audio industry. Since 2004 a special focus for him has been on trends and expectations for network systems in the field of ProAV.

Alon Regev is a system architect fluent in both the hardware and software domains who has innovated in network communications for over 30 years. For the last 20 years Alon has worked at Ixia and Keysight Technologies. Alon has founded 2 companies and has over 60 patents granted. Alon has led, participated, and is contributing to multiple standardization efforts with a focus on Industrial and Automotive network systems, Time Synchronization and Time Sensitive Networks. Alon is the chair of the Avnu testability task group, a voting member of IEEE 802.3. BSCS, California State University, Northridge (USA).

Introduction: Cut-Through Forwarding

Johannes Specht

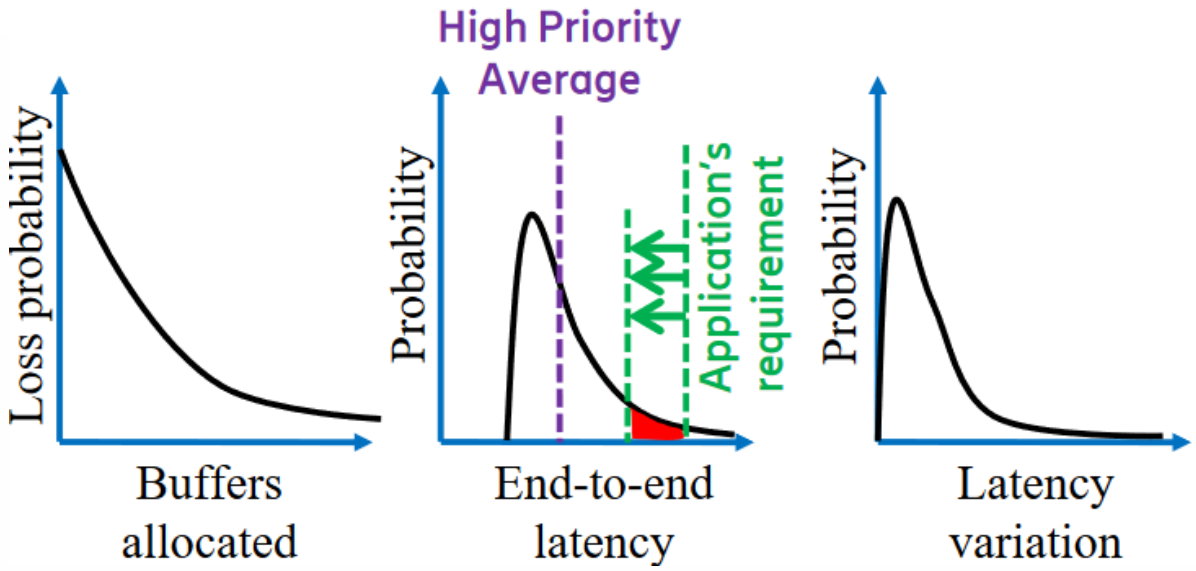
TSN Context

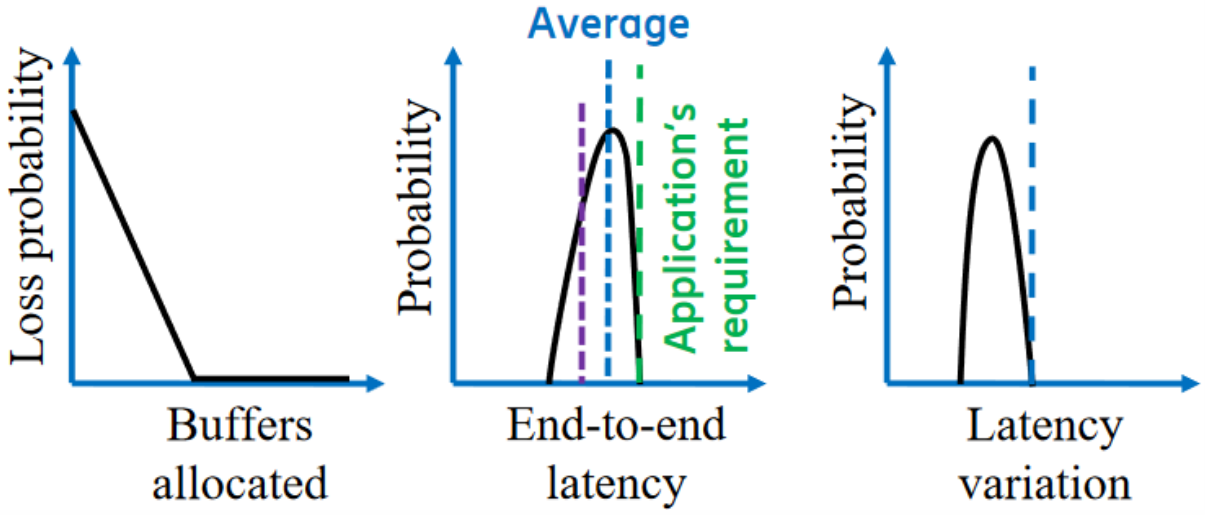


Source: https://www.ieee802.org/1/files/public/docs2021/admin-tsn-summary-0221-v01.pdf

Traditional and Deterministic Services

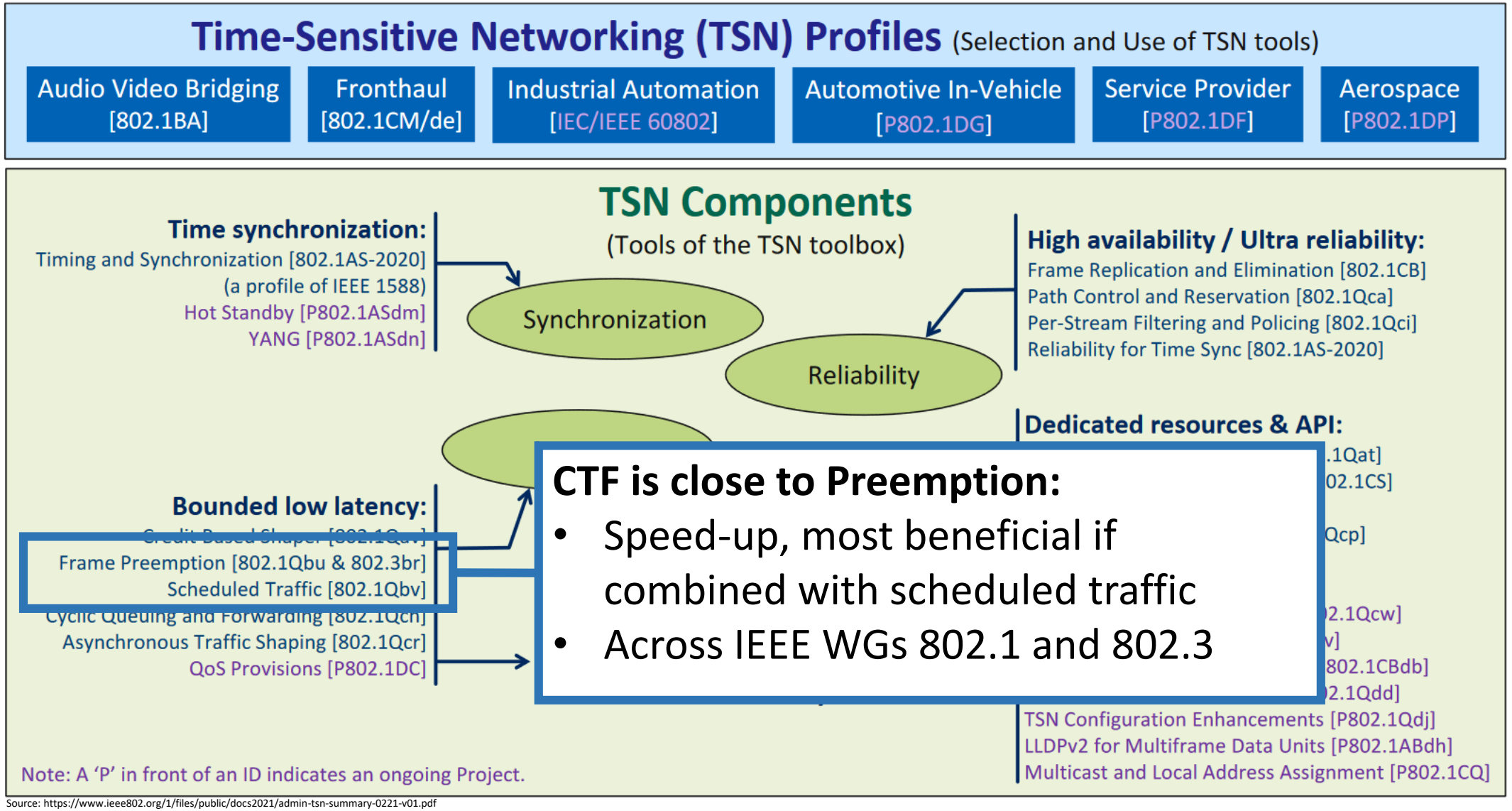
- Traditional Service
  - Curves have long tail
  - Average latency is good
  - **Lowering the latency** means **losing packets** (or overprovisioning)
- Deterministic Service
  - Packet loss is at most due to equipment failure (zero congestion loss)
  - Bounded latency, no tails
  - **The right packet at the right time**





Source: https://www.ieee802.org/1/files/public/docs2018/detnet-tsn-farkas-tsn-basic-concepts-1118-v01.pdf

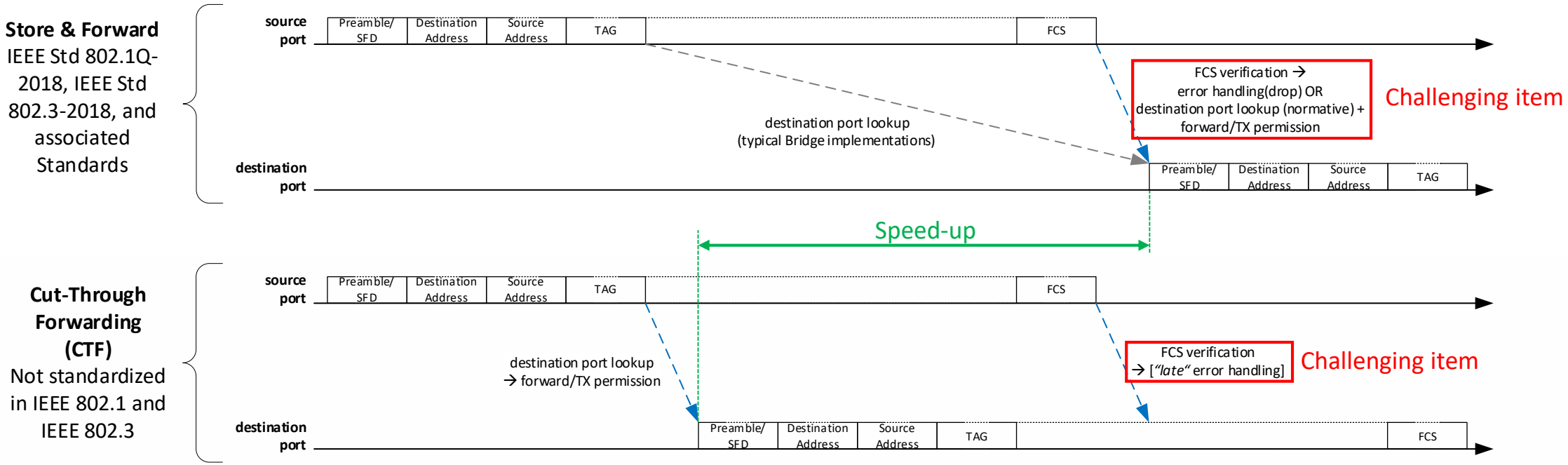
CTF in the TSN Context



Source: https://www.ieee802.org/1/files/public/docs2021/admin-tsn-summary-0221-v01.pdf

Basic CTF Operation

CTF is an alternative forwarding method to Store & Forward (S&F) in Bridges



Delay performance enhancements

- Reduced residence times of frames in Bridges ("speed-up")
- Reduced frame length dependent jitter/delay variation

(Main) Challenges

- Transmission of frames with errors discovered by FCS verification, and the associated consequences
- S&F operation "deeply" manifested in IEEE 802.1 and 802.3 Standards

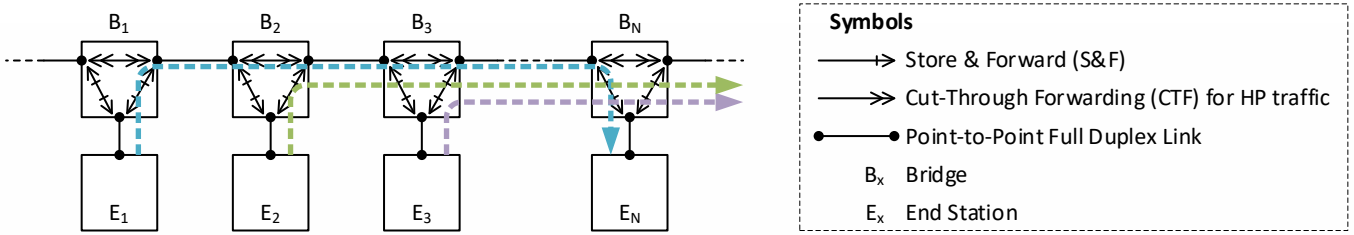
CTF Speed-up Analysis: Assumptions (1)

Purpose

- The following assumptions assemble a simplified model to focus on a simple speed-up analysis:
  - Some assumptions can be valid for some real systems, while being invalid for others.
  - The assumptions here are not intended as requirements or limitations for real systems with CTF.

Topology/Network

- Chain Network/Network segment
- Identical Link Speeds, Full-Duplex, negligible propagation delays
- CTF possible on all interconnections *except* from/to end stations (i.e., S&F at first and last hops)
- Strict Priority Transmission Selection Algorithm, optional with Enhancements for Scheduled Traffic



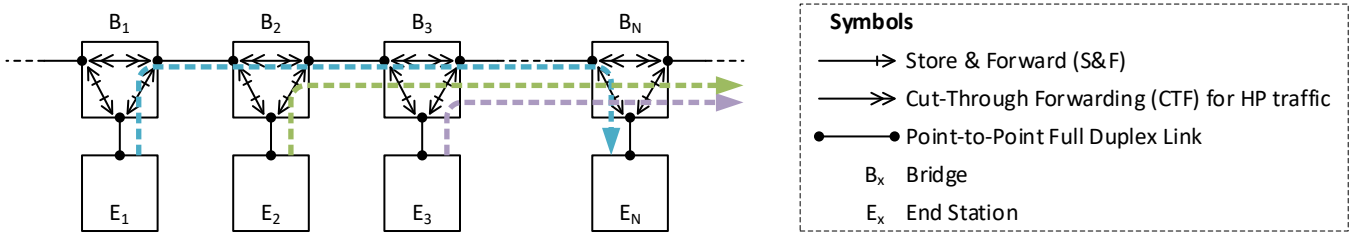
Errors

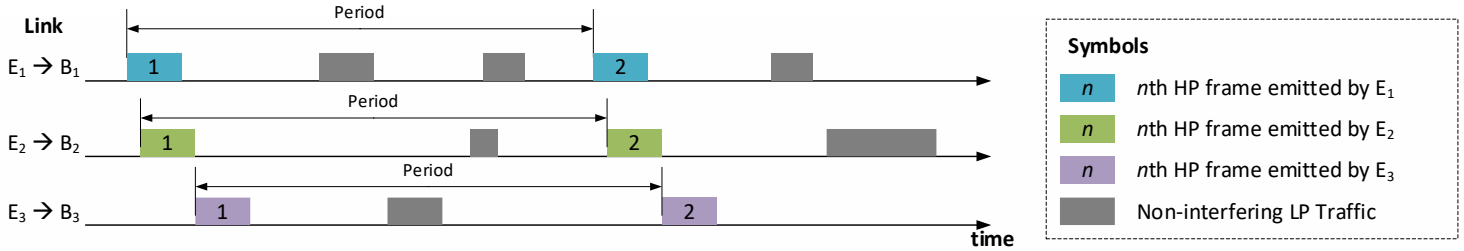
- Error free environment → no data corruption in frames
- However, errors, including late error handling, is addressed later in this tutorial

CTF Speed-up Analysis: Assumptions (2)

Traffic – Focus on Bounded Latency

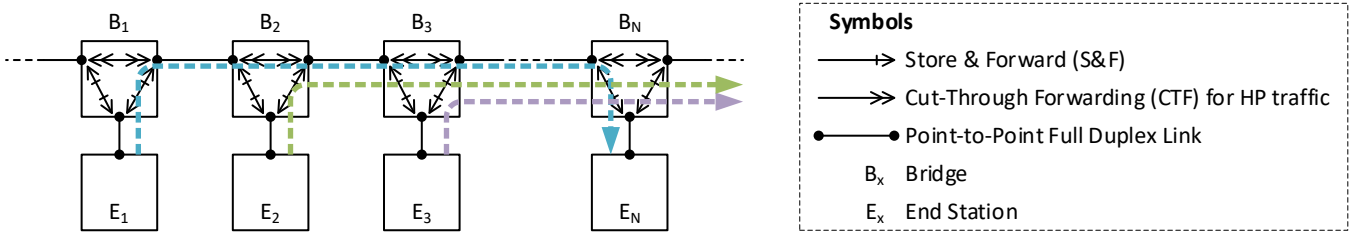
- High Priority (HP): Focus of the Analysis
  - At most one stream sent by each end station, and each end station receives HP streams from at most one direction of the chain
  - Constant frame length[1]
  - Periodic (same period for all streams)
  - Period < maximum end-to-end latency
  - Nominal transmission times at sending end stations
- Low Priority (LP): Background
  - Always Store & Forward
  - Interferes with CTF traffic
    - Without preemption: 1542 octets (max. LP frame[1,2])
    - With preemption: 155 octets (max. non-preemptible LP frame[1,3])





1) Includes all media-dependent overhead for IEEE 802.3 point-to-point full duplex media (Preamble, SFD, minimal Interpacket Gap).
2) Upper limit of 1500 octets payload in a tagged frame.
3) Defined upper limit for addFragSize=0 (cmp. 99.4.8 of IEEE Std 802.3br-2016).

CTF Speed-up Analysis: Math



Delay until forwarding to destination ports happens. Assumed that the lookup starts after l_{Hdr} octets and finishes after d_{LU} µs. Note that the lookup can finish after frame completion during reception.

$$d_{SFF}^{max} = (H+2)\left(\max\{l_{HP}d_{Oct}, l_{Hdr}d_{Oct}+d_{LU}\}+d_Q\right)+\left((H+1)l_{LP}+Hl_{HP}\right)d_{Oct}$$

Maximum interference by crossing high priority traffic (l_{HP}) and crossing low priority traffic (l_{LP}). Dependent on the subsequently introduced communication schemes, either one or both types of interference exist or not (e.g., full TDM avoids both).

$$d_{CTF}^{max} = 2\left(\max\{l_{HP}d_{Oct}, l_{Hdr}d_{Oct}+d_{LU}\}+d_Q\right)+H\left(l_{Hdr}d_{Oct}+d_{LU}+d_Q\right)+\left((H+1)l_{LP}+Hl_{HP}\right)d_{Oct}$$

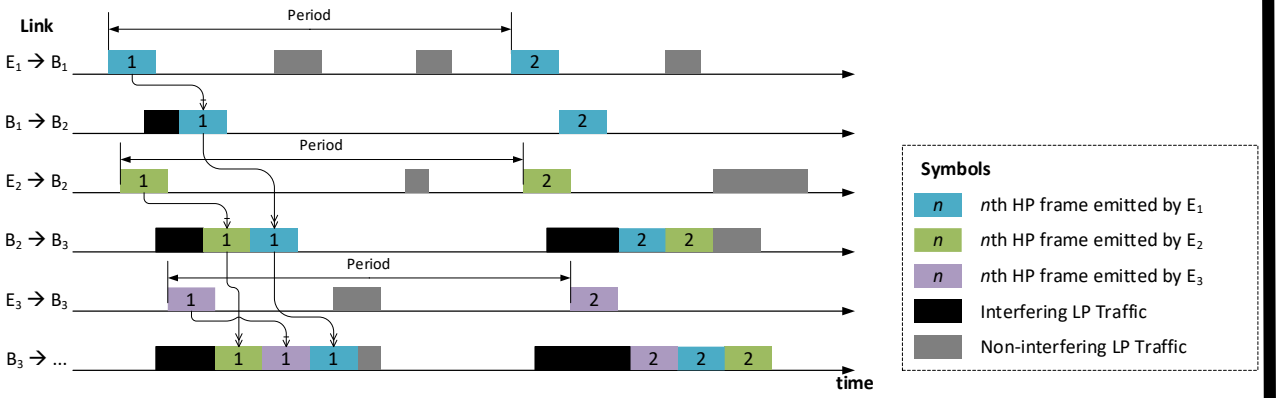
Separates the H interconnections (CTF) from the first and last ones (S&F). Note that, if the lookup finishes after frame completion during reception, then CTF provides no lower delay than S&F. The other way around, if the lookup is "fast enough", then CTF provides lower delays than S&F.

Symbol	Description
d_{SFF}^{max}	Maximum end-to-end delay without CTF of HP frames, in μs.
d_{CTF}^{max}	Maximum end-to-end delay with CTF of HP frames, in μs.
H	Number of possible CTF interconnections (e.g., N-2 for the stream of E_1).
l_{HP}	Frame size of high priority traffic (i.e., the traffic that can be subject to CTF), including all media dependent overhead, in octets.
l_{LP}	Frame size of low priority traffic (always S&F), including all media dependent overhead, in octets. Assumption: 1542 octets without preemption, 155 octets with preemption.
l_{Hdr}	Header length required for destination port lookup in Bridges, in octets. Assumption: 24 octets (preamble, start of frame delimiter, DA, SA, VLAN-Tag).
d_{Oct}	Nominal duration of an octet reflecting the link speed, in μs.
d_{LU}	Destination port lookup duration after l_{Hdr} octets were received, in μs. Assumption: 0.16 μs (e.g., 20 clock cycles @ 125 MHz).
d_Q	Interference-independent queuing delay (MAC delay, PHY delay, etc.), in μs. Assumption: 0.32 μs.

CTF Speed-up Analysis: Both Extremes

Uncoordinated

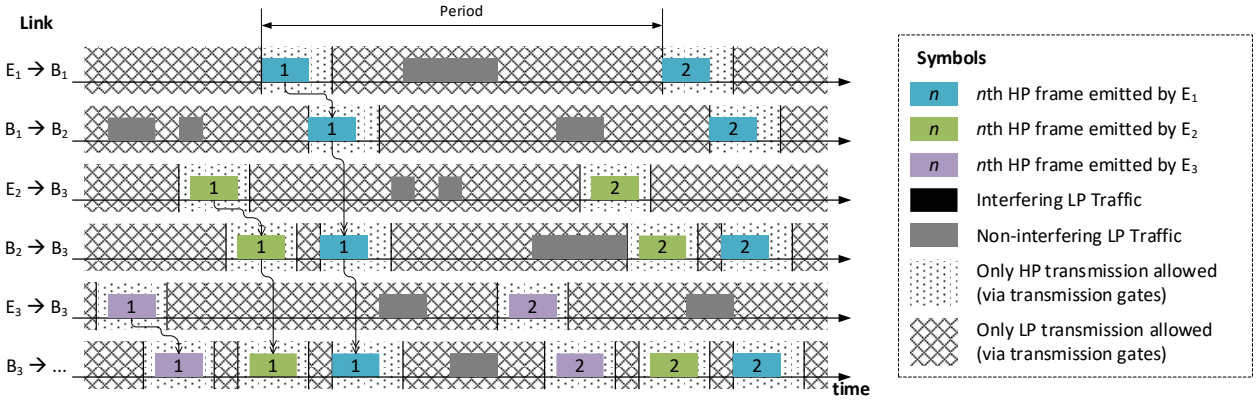
Interference by low priority and other high priority (CTF) traffic



		SFF-to-CTF ratio									
		Preemption unsupported					Preemption supported				
H	Link \ l_{HP}	128	256	512	1024	1542	128	256	512	1024	1542
2	100 Mbps	96%	93%	88%	83%	80%	85%	80%	76%	74%	73%
4	100 Mbps	96%	91%	85%	79%	75%	82%	75%	70%	67%	66%
16	100 Mbps	95%	90%	82%	74%	70%	77%	68%	62%	59%	57%
64	100 Mbps	94%	89%	81%	73%	68%	76%	66%	60%	56%	54%
2	1 Gbps	97%	94%	89%	84%	81%	89%	82%	78%	75%	74%
4	1 Gbps	96%	92%	86%	80%	76%	86%	78%	72%	68%	67%
16	1 Gbps	96%	91%	83%	75%	70%	83%	72%	65%	60%	58%
64	1 Gbps	96%	90%	82%	74%	69%	82%	71%	62%	57%	55%
2	2,5 Gbps	98%	95%	90%	84%	81%	94%	86%	80%	76%	75%
4	2,5 Gbps	98%	93%	87%	81%	77%	92%	83%	75%	70%	68%
16	2,5 Gbps	97%	92%	85%	76%	71%	90%	78%	69%	62%	60%
64	2,5 Gbps	97%	92%	84%	75%	70%	90%	77%	67%	60%	57%

Full Time Division Multiplexing

No Interference



		SFF-to-CTF ratio				
		Preemption supported or not				
H	Link \ l_{HP}	128	256	512	1024	1542
2	100 Mbps	61%	56%	53%	51%	51%
4	100 Mbps	48%	41%	37%	35%	35%
16	100 Mbps	31%	21%	16%	14%	13%
64	100 Mbps	25%	14%	9%	6%	5%
2	1 Gbps	75%	64%	58%	54%	53%
4	1 Gbps	67%	52%	43%	39%	37%
16	1 Gbps	56%	36%	25%	18%	16%
64	1 Gbps	52%	31%	18%	11%	8%
2	2,5 Gbps	88%	74%	64%	58%	55%
4	2,5 Gbps	84%	66%	52%	44%	40%
16	2,5 Gbps	79%	55%	36%	25%	21%
64	2,5 Gbps	77%	50%	31%	18%	13%

Lower percent values indicate higher end to end delay performance improvements of CTF over S&F.

Reasons for standardizing CTF in IEEE 802

Interoperable and deterministic data plane (examples)

- Distinguish CTF Traffic from S&F Traffic
  - TAGs, Addresses, Ports?
- "Late" error handling
  - Shorten/truncate erroneous frames?
  - Mark erroneous frames?
  - Do nothing?
- Behavior of existing 802.1 Bridge mechanisms for CTF traffic
  - Flow Metering (e.g. Max. SDU size filters, MEF 10.3)?
  - Transmission selection algorithms?
  - Transmission gates?
  - Link speed transitions?[1]

Unified Management

- Elements
  - Configuration Parameters (e.g., enable/disable CTF)
  - Device properties (e.g., timing)
  - Status Variables (e.g., erroneous CTF frame counters)
- Required, for example, for automated, efficient and consistent TDM configuration (e.g., centralized network controller [802.1Qcc-2018])

Application and limitations of CTF in Networks

- Quality of Service[1,2]

Limit circulating erroneous frames in topological loops; limit bandwidth loss by erroneous frames

- Security[1]

Prevent exposure of frame contents (CTF and S&F) to untrusted network segments

1) See also https://ieee802.org/1/files/public/docs2017/new-tsn-thaler-cut-through-issues-0117-v01.pdf
2) See also https://www.ieee802.org/1/files/public/docs2019/new-seaman-cut-through-scissors-0119-v01.pdf

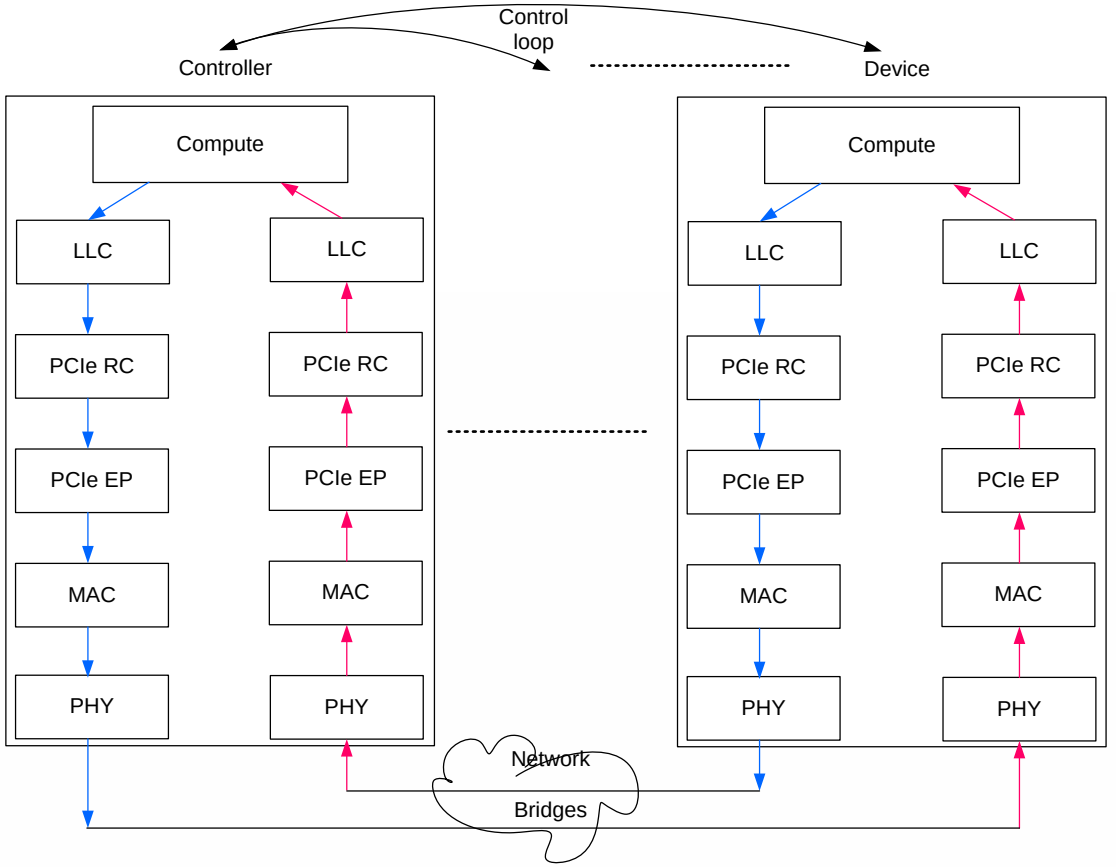
Use-Cases: Industrial Automation

Jordon Woods

Networking Requirements: Principal Data Path (Control Loop)

Principle data path between the controller and a device:

- The entities which are involved into the guaranteed latency transmission for the control loop are depicted
- Latencies for link layer control, bus interface, MAC/PHY are incurred at the controller and the device
- Combined store & forward, bridge delay and PHY delay accumulate at each hop in the network.



Networking Requirements: Summary

Industrial applications, such as machine control, are typically connected in long line configurations. For these installations, to minimize wiring cost and complexity, typical installation uses “daisy chain” where each node has (2) external switched ports and an internal port that goes to the end-node.

A common application is motion control where fast loop times are required. 125 µs cycle rate is common for 100 Mbps. Even lower rates (62.5µs/31.25µs) are desired for 1 Gbps. To support this, low latency for messages through the network is a high priority.

Even Gigabit data rates are not sufficient to solve this problem. Combined store & forward, bridge delay and PHY delay exceed timing budgets. For instance, in a line topology of 64 hops, accumulated latency would exceed a 100 µs control loop even at Gigabit speeds.

- See http://www.ieee802.org/3/ad_hoc/ngrates/public/18_01/woods_nea_01a_0118.pdf

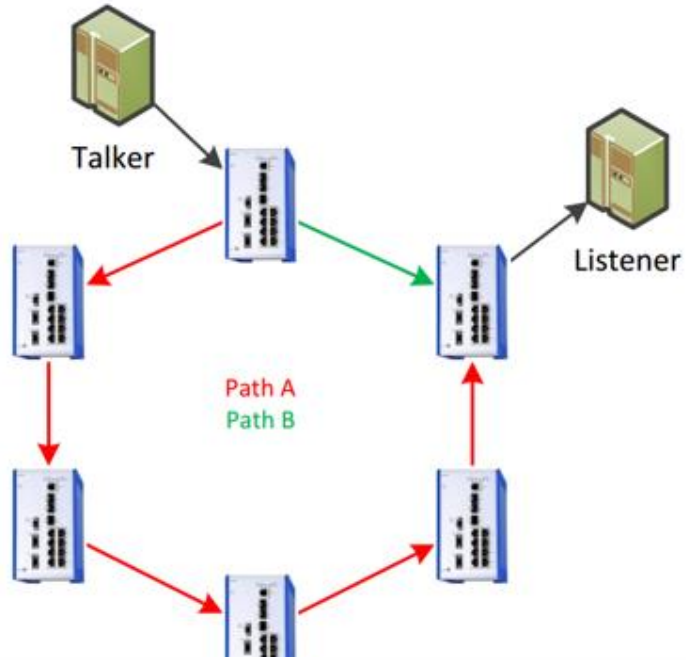
These industrial automation systems often have environmental constraints (power, space, radiated emissions, etc.) which make lower data rates desirable. There is a desire in some applications to support brown-field wiring. Often, these devices are resource, power and cost-constrained. For these applications 100Mb/s rates are desired.

Why Line Topologies?

- Physical constraints make cabling for star topologies impractical
- The construction of the application naturally lends itself to point-to-point connectivity
- They are, after all, assembly “lines”

Use Case 2 - Redundancy (ring topologies)

- Typical topology for redundancy in industrial networks is a ring:
  - Inherently different packet latency on the network along the different routes
  - Depending on the setup, packet latency on the two paths can have extreme deviation
  - Depending on the allowed reception window of redundancy mechanisms, ring size is limited
  - For instance, for a 300 byte packet and 100 us packet deviation:
    - At 100 Mbit/s: the max. tolerable difference in the path is consumed in 4 hops
    - At 1 Gbit/s: the max. tolerable difference in the path is consumed in 34 hops



Industrial Network Growth

Industrial automation market > $123B in 2019

- Source: Control Global - https://www.controlglobal.com/articles/2020/top-50-automation-companies-of-2019-under-siege/

Connectivity portion is growing

- Fieldbus (58%), 7% growth
- Ethernet (38%), 20% growth
- Limited wireless adoption

With the advent of a common layer 2 (TSN), Industrie 4.0, China 2025, etc., strong growth is expected.

- Global industrial Ethernet market valued at USD $24B in 2016
- Expected to grow to $58.98 billion by 2022
- CAGR of slightly above 16.20% (2017 and 2022)
  - Source: Zion Market Research, 2017 - https://www.zionmarketresearch.com/news/global-industrial-ethernet-market

Use-cases: Data Center Networks

Paul Congdon, Lily Lv

High Performance Applications Growth in the Data Center

High Performance Computing (HPC), AI (Artificial Intelligence)/Big Data and Cloud Computing are hot growth areas.

The convergence of these 3 areas is currently a trend in the data center.

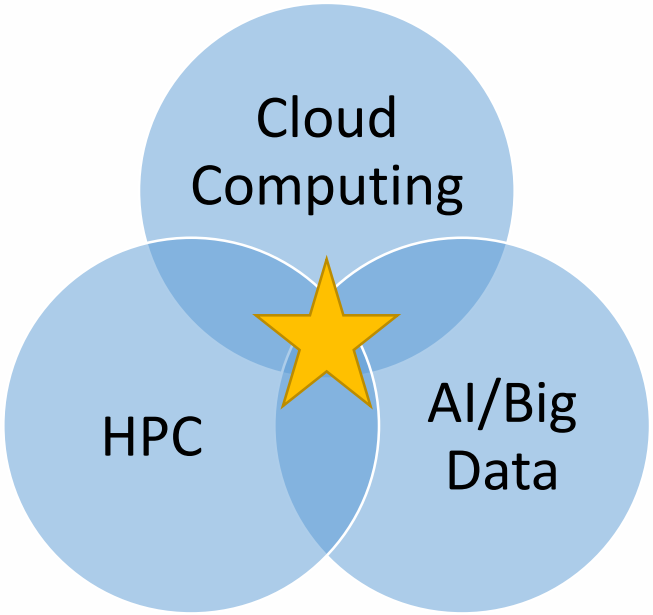
- HPC is available as a cloud service in many public offerings (AWS, Azure, Alibaba etc); growing 17.6% CAGR (Compound annual growth rate) , 2.5 times faster than on-premise HPC.
- HPDA (High performance data analytics) and HPC-based AI are fast emerging markets, with 16% and 31% CAGR respectively.

	2019	2020	2021	2022	2023	2024	CAGR
HPC cloud	$2,466	$3,910	$4,300	$5,300	$6,400	$8,800	17.6%
On-Premise HPC	$27,678	$23,981	$26,774	$31,872	$36,138	$38,214	6.7%

Source: Hyperion Research, November 2020

	2019	2020	2021	2022	2023	2024	CAGR
HPC Server Revenues	$13,713	$11,846	$13,295	$15,817	$17,942	$19,044	6.8%
HPDA Server Revenues	$3,598	$3,932	$4,737	$5,457	$6,480	$7,479	15.8%
HPC-Based AI	$918	$1,094	$1,399	$1,810	$2,745	$3,555	31.1%

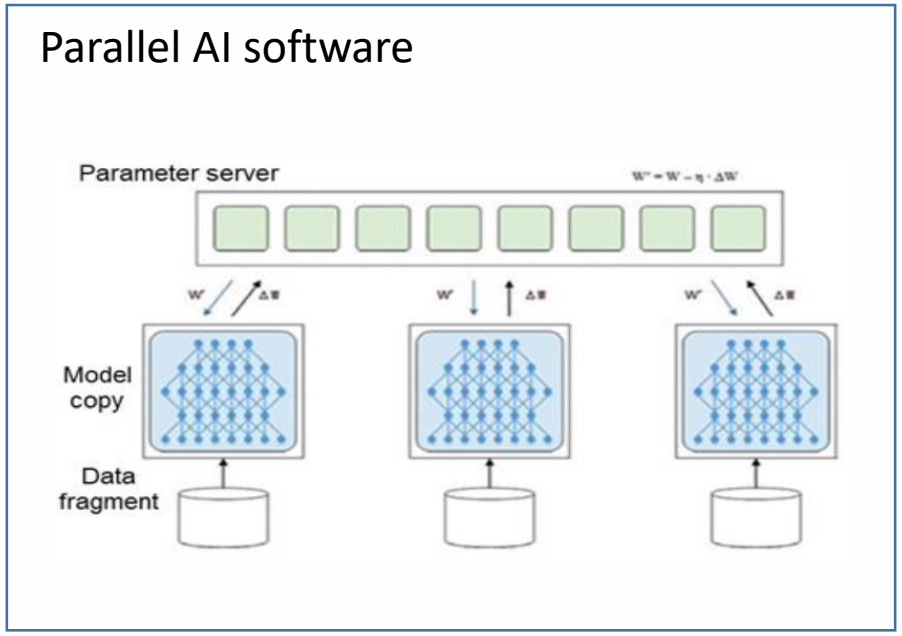
Source: Hyperion Research, November 2020

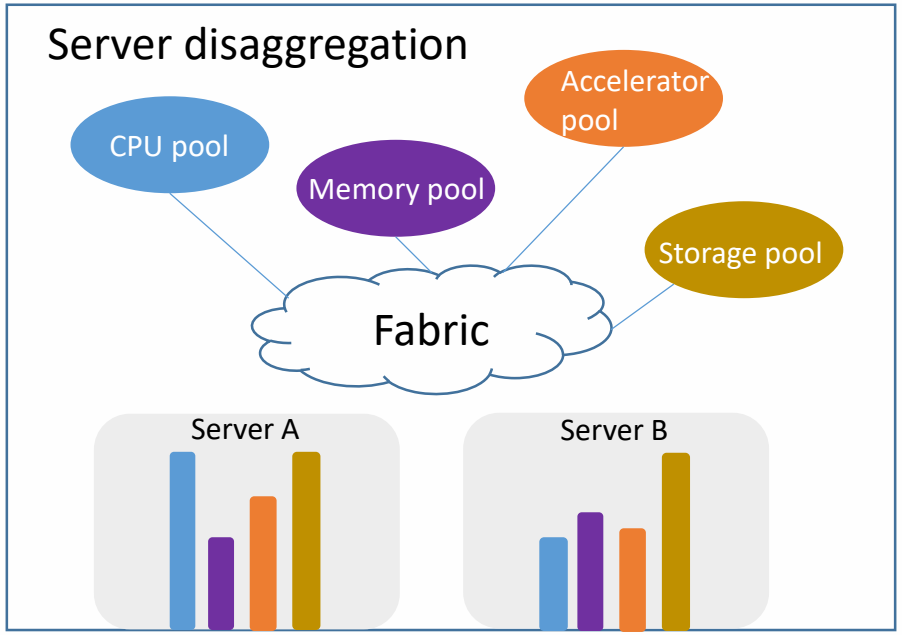


Latency is Critical in Data Center Networks (1)

High performance applications are driving change in data center, putting pressure on end-to-end latency.

- System scale is increasing significantly, with much more end points and a larger network.
- Synchronization in large parallel applications is critical to job completion time.
- New hardware architectures, such as server disaggregation, require extremely low-latency fabric.

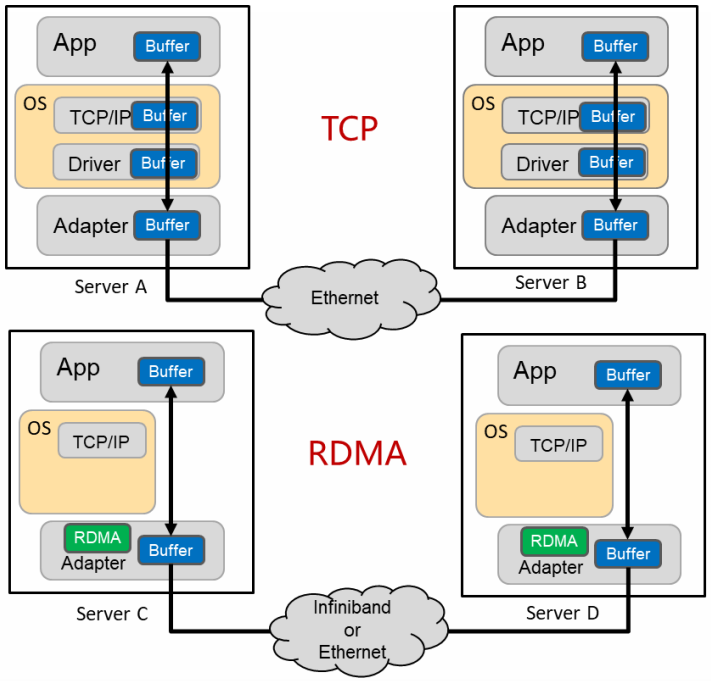


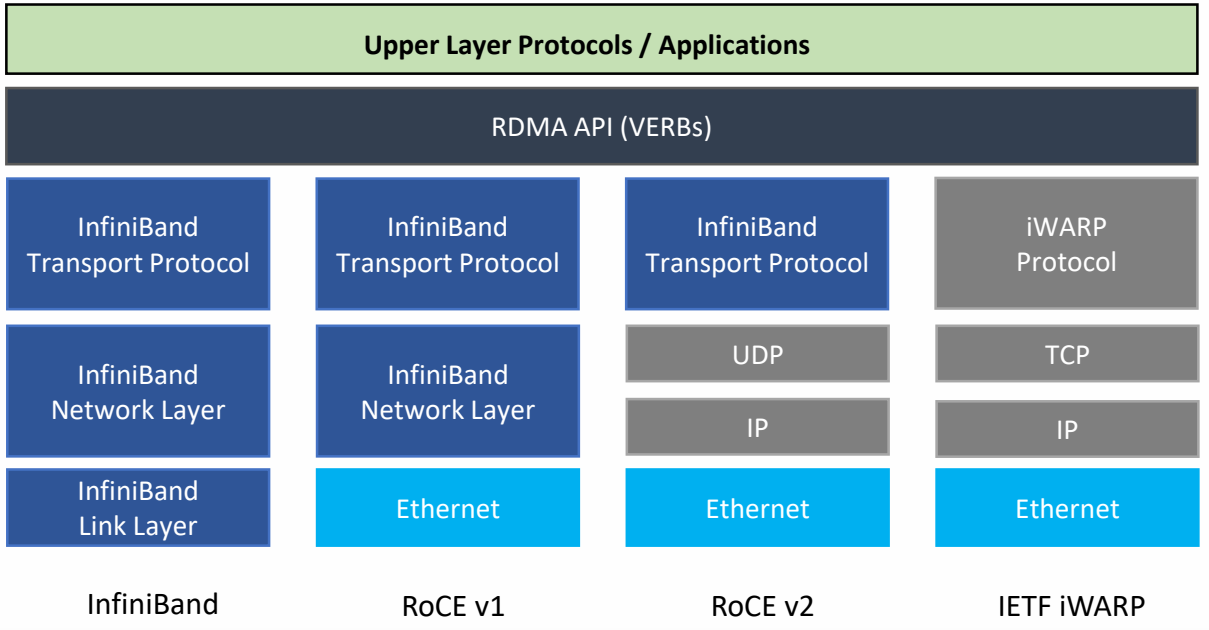


Latency is Critical in Data Center Networks (2)

New technologies are emerging to reduce system latency.

- **RDMA (Remote Direct Memory Access)**
  - RDMA enables direct memory access from one server to another, bypassing the TCP/IP stack handling in OS.
  - RDMA runs over InfiniBand or Ethernet.
    - InfiniBand, like Ethernet, is a networking technology, but customized for high throughput and low latency.
  - RDMA improves message transfer time by 5x compared with TCP/IP.

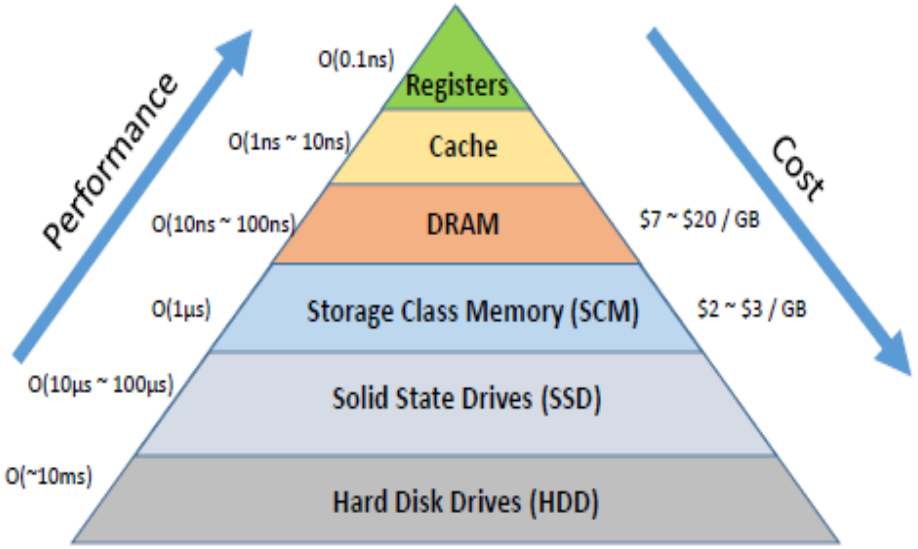




Latency is Critical in Data Center Networks (3)

New technologies are emerging to reduce storage latency.

- **Faster storage media**
  - Persistent storage latencies are approaching memory latencies with the latest Storage Class Memory (SCM) technology.
- **NVMe (Non-Volatile Memory express)**
  - NVMe is a storage interface specification defining communication between host software and PCIe SSD.
  - "The NVMe specification was designed from the ground up for SSDs. It is a much more efficient interface, providing lower latency, and is more scalable for SSDs than legacy interfaces, like serial ATA (SATA). " (https://nvmexpress.org/)
  - NVMeoF (NVMe over Fabrics) enables "networked" fast storage (SSD/SCM) however without networking enhancements, the network becomes the largest part of end-to-end latency.



Network latency becomes the bottleneck!

Latency is Critical in Data Center Networks (4)

Types of latency in data center networks: dynamic and static

Dynamic latency = queuing delay + retransmission delay

- Mainly caused by congestion
  - In-cast congestion from parallel applications.
  - In-network congestion from ineffective load balancing.

- Mainly cause by packet loss due to congestion
  - Priority-based Flow Control (PFC) guarantees no loss
  - PFC has deployment challenges: configuration, deadlocks, head-of-line blocking, congestion spreading

Static latency = switch forwarding + packet processing + link latency

- Impacted by forwarding table lookup delay, frame reception delay (if store and forwarding) and switching delay

- Impacted by header processing and packet modification

- Propagation delay impacted by distance and speed

- Dynamic latency is the major component and attracts a lot of the industry's attention: See
  - 802 Nendica - The Lossless Network for Data Centers - https://mentor.ieee.org/802.1/dcn/18/1-18-0042-00-ICne.pdf
  - 802 Nendica - Intelligent Lossless Data Center Networks - https://mentor.ieee.org/802.1/dcn/20/1-21-0004-00.pdf
- However, Static latency becomes significant in high performance scenarios, such as HPC.

Benefits of Cut-Through Forwarding in the HPC Networks

HPC network operates at the nanosecond level

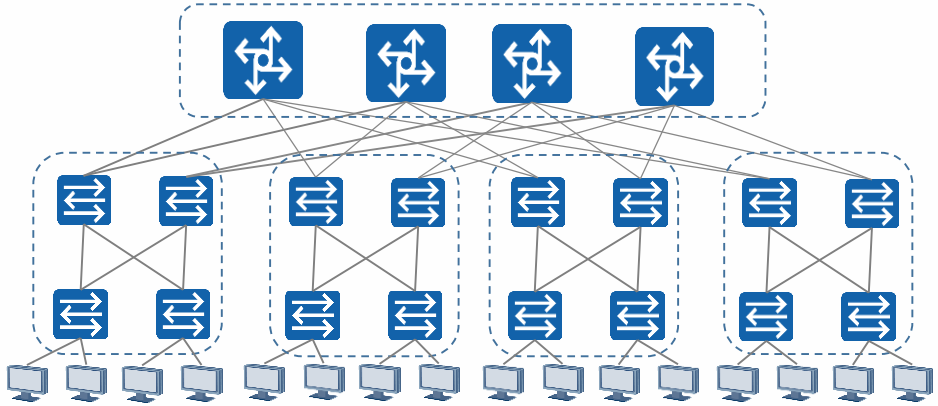
- E2E network latency is only several micro-seconds.
- Per hop latency is required as low as possible, hundreds of nano seconds, or even lower.

CTF is applicable in the HPC network

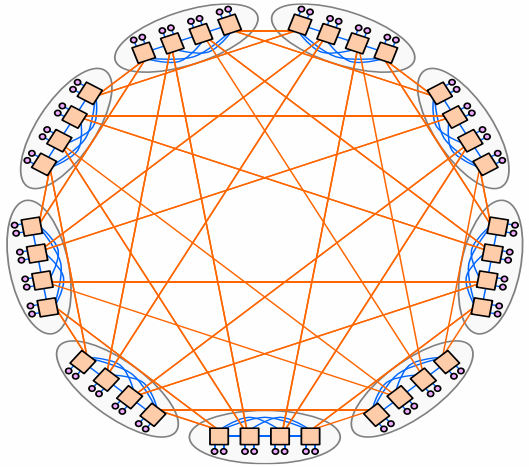
- Traffic loads can be predictable, leading to congestion avoidance techniques in switches.
- Data center topologies are well structured with similar type of switches.

Regular Topologies: Two typical HPC networks

Fat Tree is a rearrangeably non-blocking structure, which provides an oversubscription ratio of 1:1 to all servers.

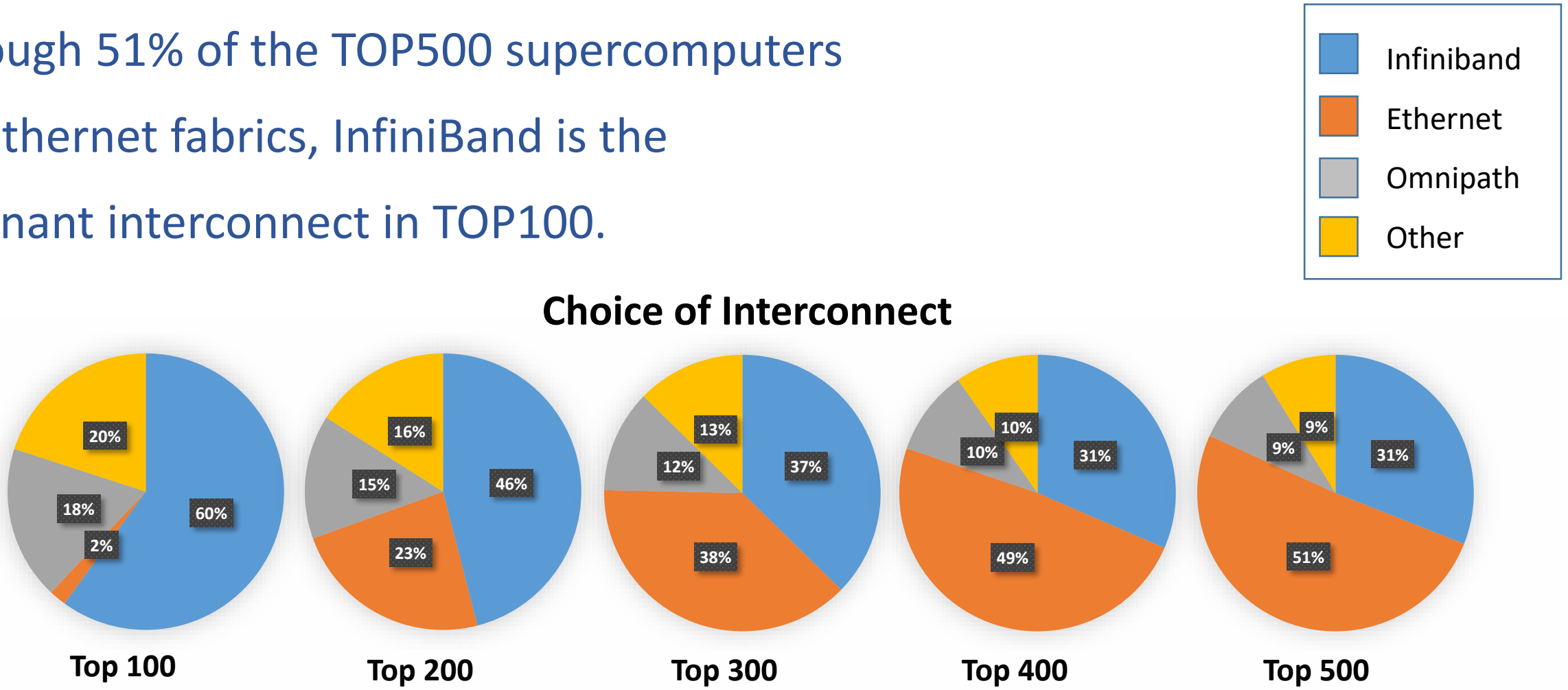


Dragonfly is a well-balanced network with no oversubscription.



InfiniBand is the ‘first-choice’ in HPC Today (1)

Although 51% of the TOP500 supercomputers use Ethernet fabrics, InfiniBand is the dominant interconnect in TOP100.



InfiniBand is the ‘first-choice’ in HPC Today (2)

InfiniBand switch per hop latency is much lower than Ethernet

- Ethernet switching chipset latency can be greater than 100s of ns.
- Latency increases with frame size using store-and-forward.
- InfiniBand switching chipset latency can be less than 100ns.
- Cut-through is an important feature for InfiniBand to keep per hop latency low.

Ethernet (non-CT)

	BRCM THK
Port	128*25G

One 25GbE Port to One 25GbE Port Test

Frame Size(Bytes)	64	128	256	512	1024	1280	1518	2176	4096	9216
Latency(ns)	511	528	556	567	717	793	872	1082	1694	3334

Source: Tolly, February 2016

IB (with CT)

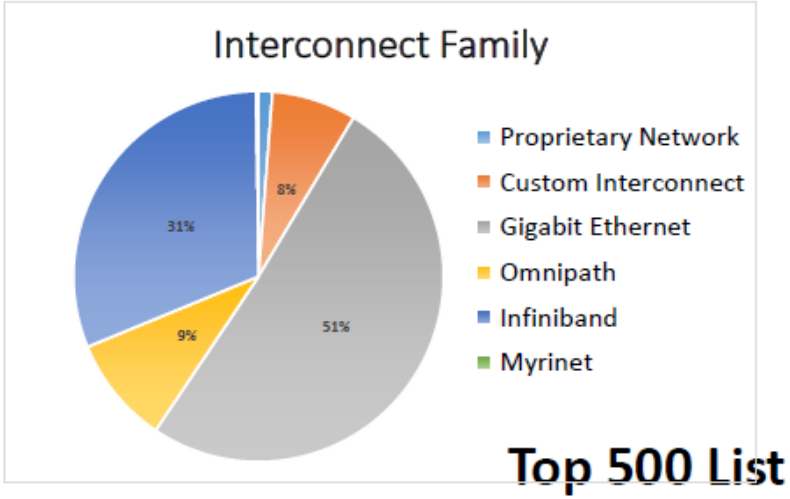
	MLNX Switch-IB	**MLNX Switch-IB2**
Port	144*25G	144*25G
Latency	90ns	90ns

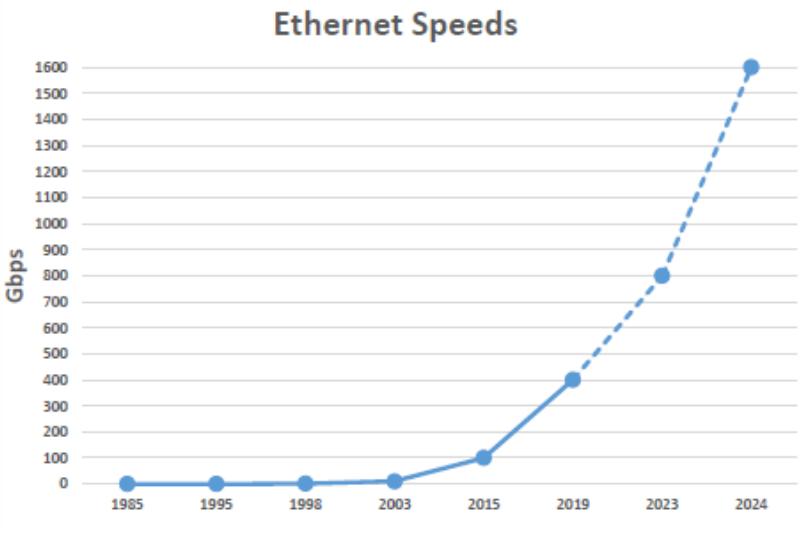
Source: https://www.mellanox.com/news/press_release/

Ethernet needs CTF to further penetrate HPC

Ethernet has great opportunity to become more competitive in HPC market.

- TOP 500 shows Ethernet Interconnects already takes the largest share (51%)
- Ethernet has its own advantages
  - Ethernet is ubiquitous technology.
    - Cost-effective solution
    - Relatively easy to deploy and manage
    - Leading technology development
  - Ethernet provides large bandwidth connectivity
    - up to 400G, 100G for single lane
    - towards 800G, 200G for single lane





The obvious gap of Ethernet is latency

- Per hop latency gap is significant compared with InfiniBand
- CTF is a good method to improve per hop latency

Use-Cases: Professional Audio/Video

Henning Kaltheuner, Genio Kronauer

Summary: Goals and Objectives

Goals and Objectives

Standardizing CTF is inevitable to ensure interoperability.

The delay performance needed by the use-cases requires Bridges to start frame transmission before complete reception (core principle of CTF) - existing IEEE 802.1/802.3 Standards do not provide this performance:

- Preemption is no alternative to CTF
  - Preemption vs. CTF
    - Preemption: Reduces delays *critical frames* experience from other interfering frames.
    - CTF: Reduces delays of the *critical frames* themselves. (regardless whether interference by other frames is present or not)
  - Nonetheless, it is desirable to combine CTF with protocols from existing IEEE 802.1/802.3 Standards, including preemption.
- Higher link speeds are no alternative to CTF
  - Inapplicable where lower link speeds are desirable → Cost, environmental constraints, brown field installations (Industrial Automation)
  - Even at high link speeds, the delay performance enabled by CTF is needed (DCN)
- Different topologies are no alternative to CTF
  - Inapplicable where daisy chain and ring topologies are inevitable → Cost, physical constraints/pre-defined structures (Industrial Automation)
  - Even in optimized topologies, the delay performance enabled by CTF is needed (DCN)

Considerations for an IEEE 802.1 Implementation

Johannes Specht

Location in IEEE 802.1 Standards

Dedicated IEEE 802.1 Standard for CTF (base Standard, not multiple amendments)

1. No distribution of CTF across multiple IEEE 802.1 Standards documents
2. Existing protocols and protocol procedures not addressed are basically “beyond specification”
3. A simple way for inclusion without adjustment is basically *“as specified in x.y.z of IEEE Std 802.1A.B.C”*
4. *If* adjustment is needed, it can apply for CTF only (i.e., limiting side effects).
   Example from 6.5.5 of IEEE Std 802.Q-20xx:

   Note that the frame is completely received before it is relayed as the Frame Check Sequence (FCS) is to be calculated and the frame discarded if in error.

Reference

- Select/import and adjust existing protocols and protocol procedures from other IEEE 802.1 Standards:
  1. IEEE Std 802.1Q-20xx
  2. IEEE Std 802.1CB-20xx
  3. IEEE Std 802.1AC-20xx
  4. [IEEE Std 802.1AE-20xx][1]

1) See later slide on security with CTF.

Main Contents

Requirements for CTF in Bridges

CTF in Networks

- Structure and elements (e.g., “CTF Bridge”)
- Application and Limitations[1]:
  - QoS Maintenance
  - Security Considerations
  - Resulting Network Requirements/Recommendations
- Usage/Performance aspects[2]

CTF in Bridges

- Bridge data plane behavior and managed objects (YANG)
  - MAC Relay Entity/Forwarding Process
  - Bridge Port Transmit and Receive[3]

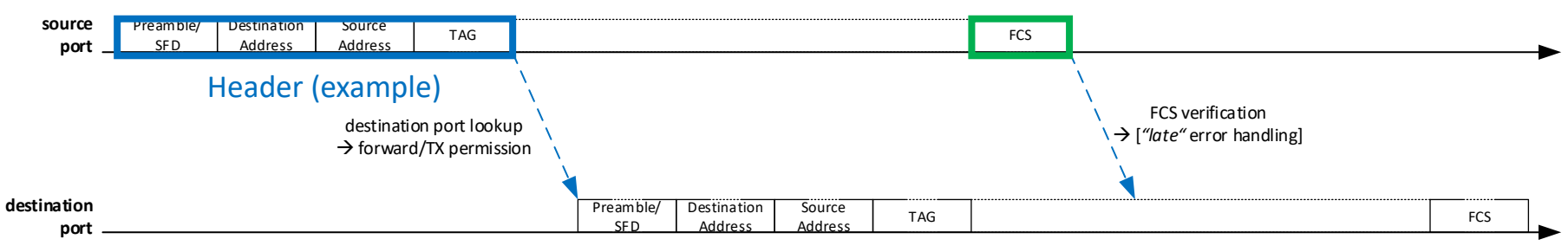
“Features” for QoS Maintenance and usage

Considered throughout the next slides
(Can be an input for in-depth considerations during a potential Stds development activity)

1) Issues introduced by CTF (cmp. https://ieee802.org/1/files/public/docs2017/new-tsn-thaler-cut-through-issues-0117-v01.pdf and ttps://www.ieee802.org/1/files/public/docs2019/new-seaman-cut-through-scissors-0119-v01.pdf) that can be addressed on a Network level.
2) See the introduction of this slide set.
3) To the extent possible in IEEE 802.1.

CTF in Networks: Application and Limitations

Cut-Through Forwarding (CTF)
Not standardized in IEEE 802.1 and IEEE 802.3



The Basic Issue

- Erroneous frame under reception by CTF Bridge are classified for CTF, and forwarded to the wrong destination Bridge Port(s), associated with the wrong traffic class, or both.
- CTF introduces this issue for frame content errors discoverable by FCS verification[1].

Option

- Wait at every hop for the FCS verification result before forwarding.
- *This would defeat the purpose of CTF[2].*

Not a Solution!

Option

- Add a "Header FCC" in frames and verify before forwarding.
- Several issues (e.g., compatibility/interoperability, frame overheads, loose header definition).

Could be analyzed during Stds activities …

Option

- Analyze the resulting issues.
- Address resulting issue individually.

1) Circulating erroneous frames
2) Additional Congestion
3) Security/frame contents exposure to unintended links

Next slides …

1) In contrast, errors that cannot be discovered by FCS verification are no issue introduced by CTF.
2) See slide <<TBD: XXX>> of this slide set.

CTF in Networks: Circulating Erroneous Frames

Issue

Erroneous frames in topological loops can circulate for “for a while”.

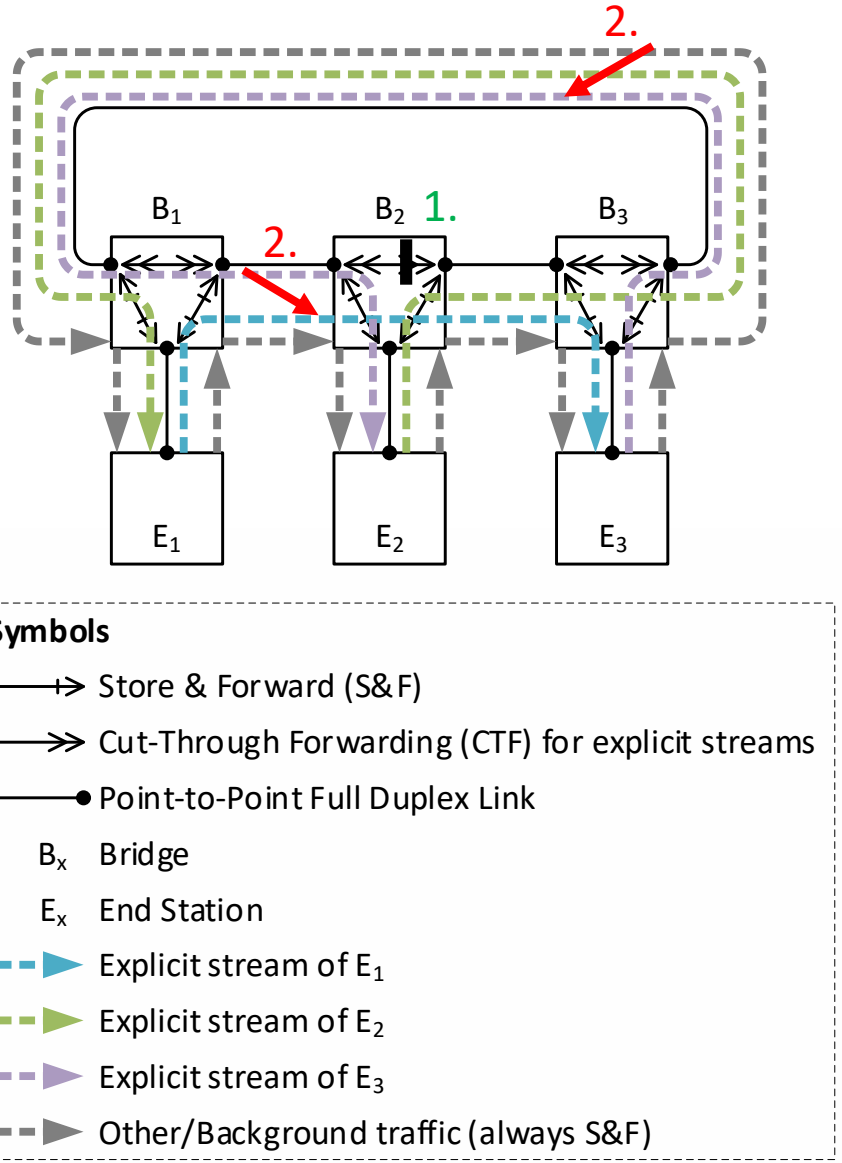
Goal

Solution(s) depend on the definition of a goal:

An erroneous frame shall circulate longer than one round in a topological loop if FCS verification can discover the error in this frame.

Solutions (Alternatives)

1. One hop with S&F-only for all traffic in each topological loop (robust/default choice).
2. Constrained FDB setups, namely explicit unicast/multicast entries only, assuming a frame experiences corruption at most on one link.
3. Long loops, combined with frame shortening in Bridges and upper bounds on frame lengths.



CTF in Networks: Additional Congestion

Issue

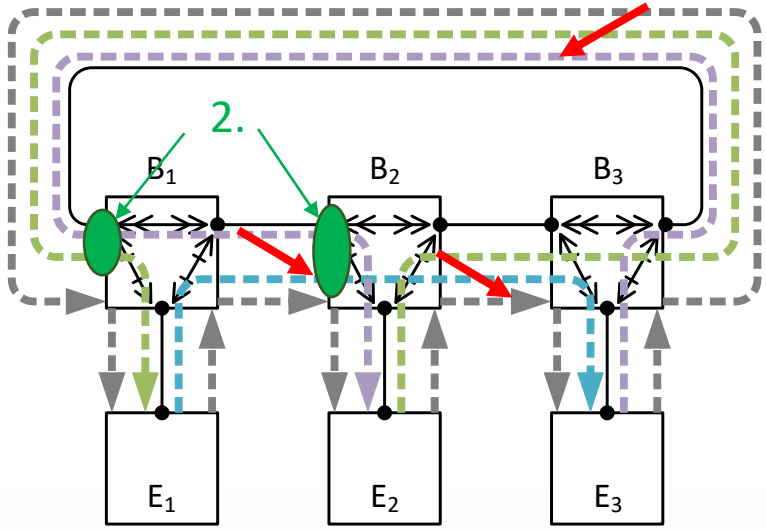
Erroneous frames queued in Bridge transmission Ports can cause additional congestion for other traffic.

→Extra delays by additional interference

→Bandwidth reduction

Solutions (Alternatives)

1. If applicable[1], usage of disjoint redundant paths via FRER[2].
2. If applicable[3], usage of Per-Stream Filtering and Policing (PSFP) functions[4].
3. Planning for additional interference/bandwidth usage.



Symbols

- Store & Forward (S&F)
- Cut-Through Forwarding (CTF) for explicit streams
- Point-to-Point Full Duplex Link
- B_x Bridge
- E_x End Station
- Explicit stream of E_1
- Explicit stream of E_2
- Explicit stream of E_3
- Other/Background traffic (always S&F)

1) Disjoint redundant paths can be inacceptable for some systems (e.g., due to cost reasons).
2) Standardized in IEEE 802.1CB-20XX.
3) The planning required to properly configure PSFP can be inacceptable for some systems.
4) Standardized in IEEE 802.1Q-20XX.

CTF in Networks: Security/Privacy

Issue

Payload of erroneous frames become visible on links where it shouldn't be seen[1].

Solution

Dependent on the security under consideration.

Security with/without cryptography

Without Cryptography

- CTF may be an issue.
- **One Possible Solution**
  - Don't use CTF on the relevant links.
  - Document the issue when using CTF (e.g., security considerations).

With Cryptography

- Closer examination needed.
- See the next column (middle) …

Security with cryptography on layer 2/above layer 2

Above Layer 2

- CTF seems to be no issue (examples)
  - Web security (TLS)
  - OPC security (UASC, TLS, PubSub Security)
  - IEC 61125 (CIPSecurity [ODVA], ProfiNet security)

On Layer 2[2]

- Closer examination needed[3].
- See the next column (right) …

Security with cryptography on layer 2 with/without confidentiality

Without Confidentiality

- CTF seems to be no issue, in absence of path assumptions.
- **Considerations for IEEE Std 802.1AE-20XX**[3]
  - Transparently forward protected frames under CTF (i.e., no special handling).
  - Limited/integrity based propagation limitation.

With confidentiality

- CTF may be an issue[3], but probably not a new one.
- **One Possible Solution**
  - Don't use CTF on the relevant links.
  - Document the issue when using CTF (e.g., security considerations).

1) See also https://ieee802.org/1/files/public/docs2017/new-tsn-thaler-cut-through-issues-0117-v01.pdf
2) Cmp. IEEE Std 802.1AE-20XX
3) See also https://www.ieee802.org/1/files/public/docs2019/new-seaman-cut-through-scissors-0119-v01.pdf

Main Contents

Requirements for CTF in Bridges

CTF in Networks

- Structure and elements (e.g., “CTF Bridge”)
- Application and Limitations[1]:
  - QoS Maintenance
  - Security Considerations
  - Resulting Network Requirements/Recommendations
- Usage/Performance aspects[2]

CTF in Bridges

- Bridge data plane behavior and managed objects (YANG)
  - MAC Relay Entity/Forwarding Process
  - Bridge Port Transmit and Receive[3]

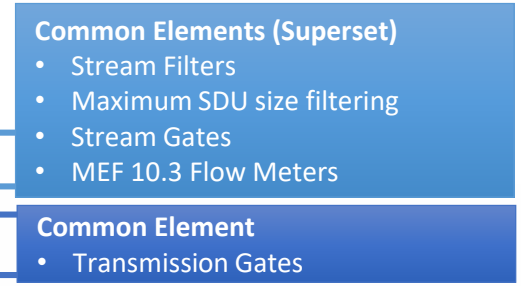
“Features” for QoS Maintenance and usage

Considered throughout the next slides
(Can be an input for in-depth considerations during a potential Stds development activity)

1) Issues introduced by CTF (cmp. https://ieee802.org/1/files/public/docs2017/new-tsn-thaler-cut-through-issues-0117-v01.pdf and ttps://www.ieee802.org/1/files/public/docs2019/new-seaman-cut-through-scissors-0119-v01.pdf) that can be addressed on a Network level.
2) See the introduction of this slide set.
3) To the extent possible in IEEE 802.1.

CTF in Bridges: Feature Set

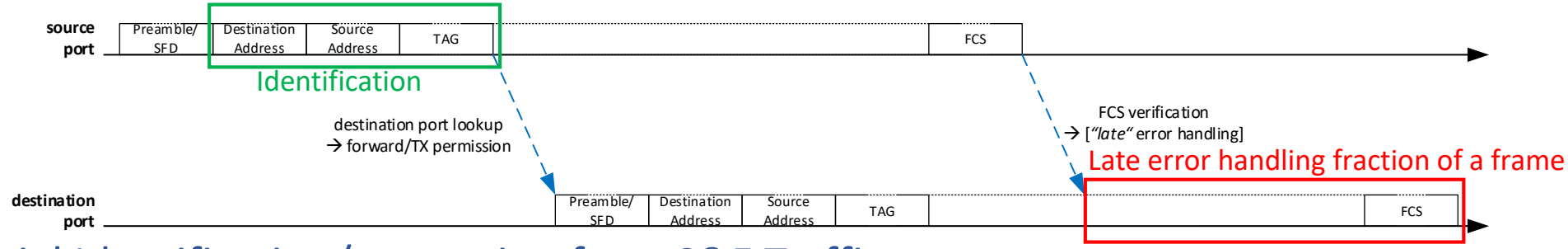
- Required:
  1. IEEE Std 802.1Q-20xx: "Basic" VLAN/MAC Bridge Operations
  2. New for CTF: Fallbacks from CTF to S&F (i.e., to behavior from existing IEEE 802.1 Stds)
  3. New for CTF: Late error handling
- Options/within specification:
  1. IEEE Std 802.1Q-20xx: Per-Stream Filtering and Policing (PSFP)
  2. IEEE Std 802.1Q-20xx: Congestion Isolation (CI)
  3. IEEE Std 802.1Q-20xx: Enhancements for Scheduled Traffic (EST)
  4. IEEE Std 802.1Q-20xx: Enhanced Transmission Selection (ETS)
  5. IEEE Std 802.1CB-20xx: Frame Replication and Elimination for Reliability (FRER)
  6. IEEE Std 802.1Q-20xx: Preemption
- For later discussion:
  1. New for CTF: Header check sequences[1]



1) Not necessarily required - header check sequences imply several challenges (interoperability with non-CTF Bridges, loose definition of headers, etc.). This topic can be considered thoroughly during a IEEE 802.1 standards development project.

CTF in Bridges: Basic path of Frames through Bridges

Cut-Through Forwarding (CTF)
Not standardized in IEEE 802.1 and IEEE 802.3



1. Reception: Initial Identification/separation from S&F Traffic

Reception on a Port for which CTF has been enabled
AND (

Priority decoded from VLAN-TAG (6.9 and 6.20 of IEEE Std 802.1Q-20xx)
OR
Stream Identification (IEEE Std 802.1CB-20xx),
used by stream filters followed by stream gates for Internal Priority Value assignments[1] (IEEE Std 802.1Q-20xx)

)

New Management Parameter(s)
- CTFReceiveEnable (Boolean, RW, default False)
- Per-Port

2. Queuing

Queuing in traffic classes (8.6.6 of IEEE Std 802.1Q-20xx) for which CTF is supported **AND** enabled

New Management Parameter(s)
- CTFTransmitEnable (Boolean, RW, default False)
- CTFTransmitSupported (Boolean, RO)
- Per-Port per traffic class

3. Transmission

- Strict priority transmission selection algorithm **OR** enhanced transmission section algorithm (if supported),
- followed by transmission gates (if supported)
- Late error handling, in case of late errors

1) The Mask-and-Match stream identification, as currently under development in IEEE P802.1CBdb, effectively enables a priority to be determined by at least the Destination Address. As one result, there are different (potentially co-existing) perceptions of a "header".

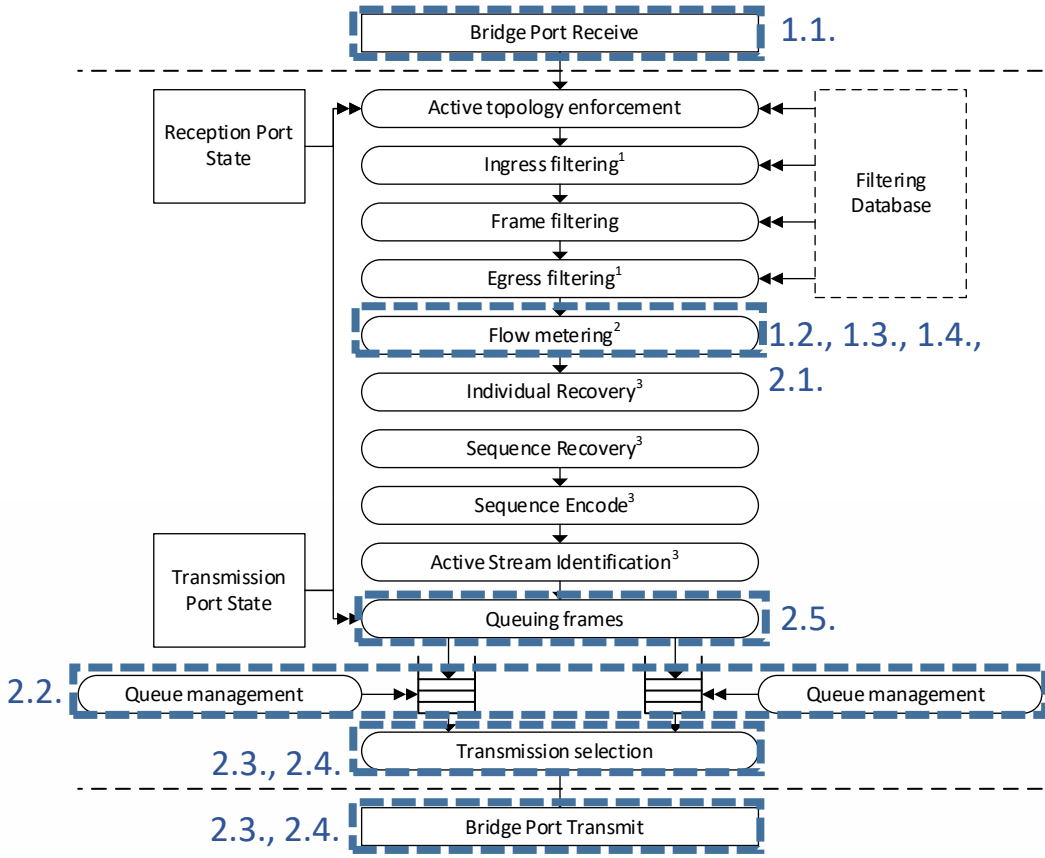
CTF in Bridges: Late Errors

1. Causes

1. Errors discovered by FCS verification
2. Maximum SDU size filtering limit reached during reception
3. Stream gates transition to closed state[1]
4. Color of flow meters (MEF 10.3) transitions to red
5. The per traffic class maximum SDU size of transmission gates is exceeded

2. Potential Handling

1. Treat as frame end by PSFP's maximum SDU size filtering, stream gates and flow meters (MEF 10.3)
2. Remove the frame from all queues
3. Shorten the end of frame by an implementation-specific amount
4. Erroneous frame marking (end of frame)



Cmp. 8.6 of IEEE Std 802.1Q-20xx and clause 8 of IEEE Std 802.1CB-20xx.
1) Not present in MAC Bridges
2) Not present if PSFP or CI is unsupported
3) Not present if FRER is unsupported

New Management Parameter(s)
- CTFTransmitShorteningMin (Integer, RO, nanoseconds)
- Per-Port

New Management Parameter(s)
- CTFReceivedErroneousMarked (Counter, RW)
- CTFReceivedErroneousUnmarked (Counter, RW)
- Per-Port

1) In contrast to stream gates, it is not intended to involve late error handling if EST transmission gates transition to a closed state during transmission for compatibility (see 8.6.8.4 of IEEE Std 802.1Q-20xx)

CTF in Bridges: Fallbacks to S&F

1. On the main relay path

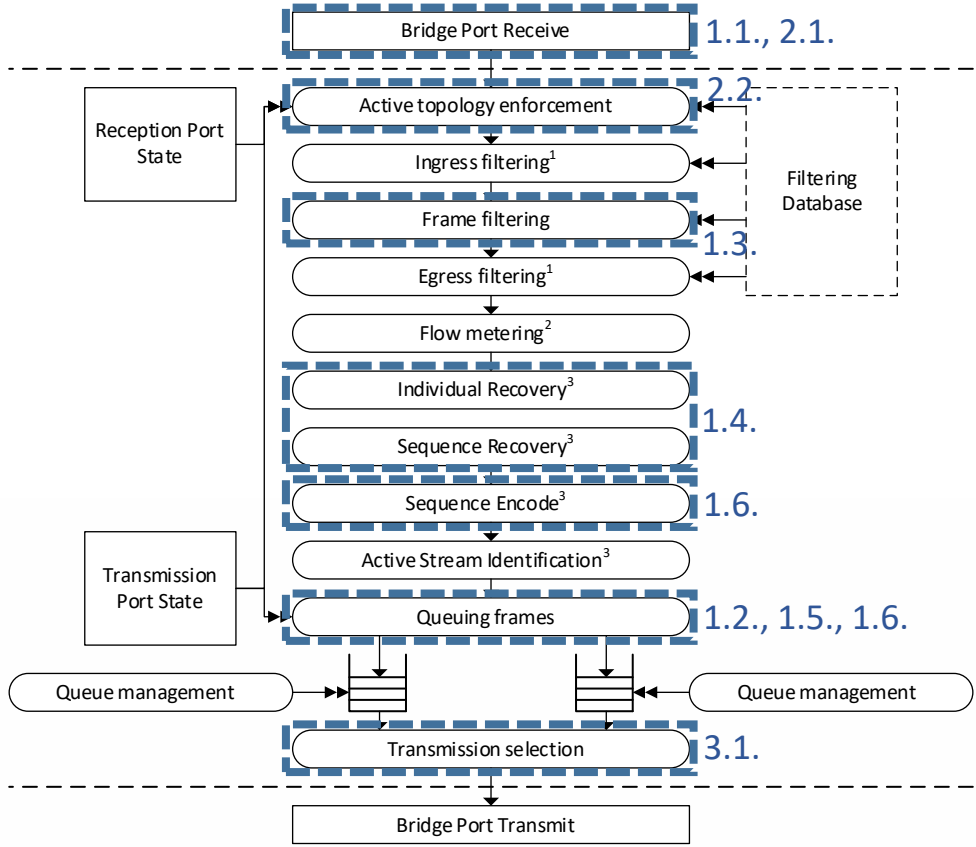
1. CTF reception is disabled on a Bridge Port
2. CTF is disabled/unsupported by a traffic class on a Bridge Port
3. No matching filtering entry in the FDB (i.e., flooding)
4. Association of a frame under reception with a FRER recovery function
5. Different link speed between reception-transmission port pairs
6. Frame length changes (e.g., TAG removal)

2. Not on the relay path, or leaving it

1. To Higher Layer Entities
2. FDB for learning

3. Implicit

1. Interfering frames during transmission

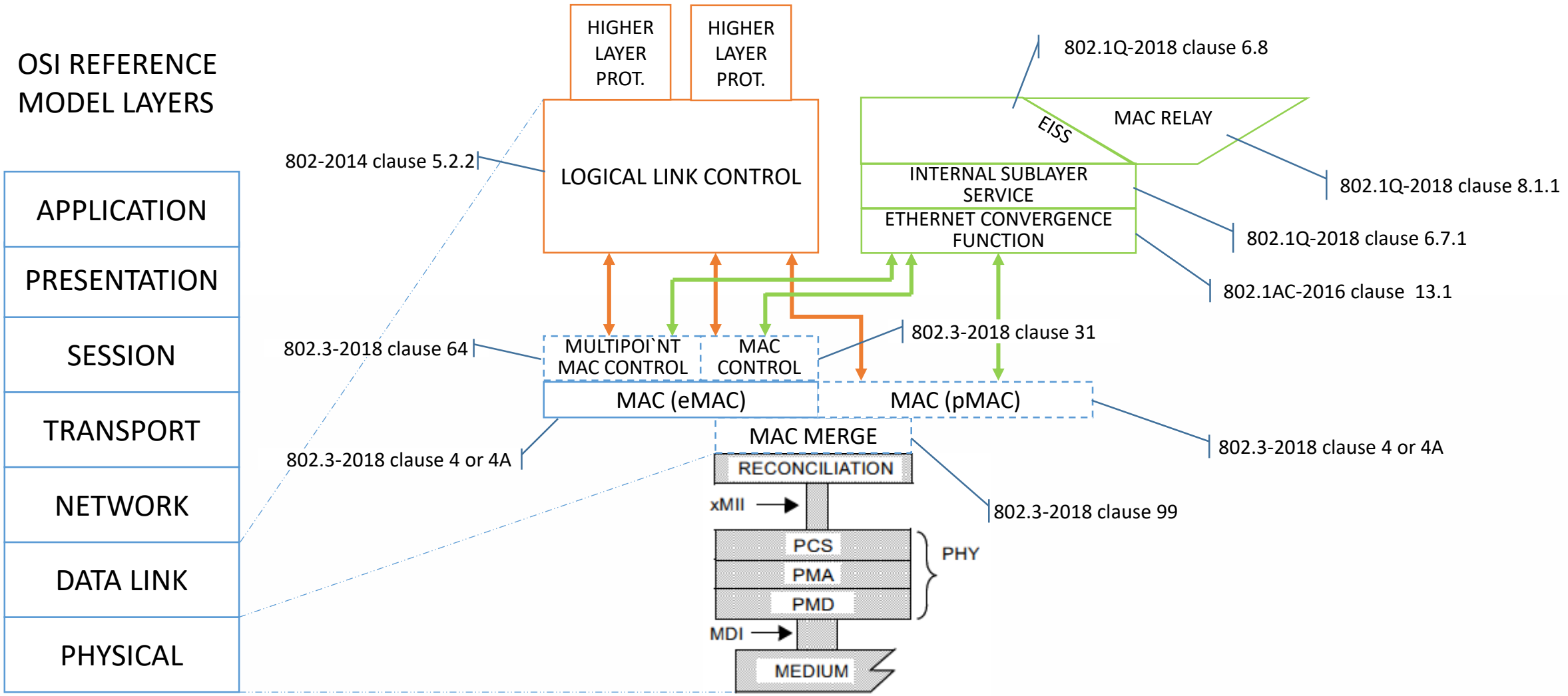


Cmp. 8.6 of IEEE Std 802.1Q-20xx and clause 8 of IEEE Std 802.1CB-20xx.
1) Not present in MAC Bridges
2) Not present if PSFP or CI is unsupported
3) Not present if FRER is unsupported

IEEE 802.3 Considerations

Alon Regev, Johannes Specht

Layering


OSI REFERENCE MODEL LAYERS
APPLICATION
PRESENTATION
SESSION
TRANSPORT
NETWORK
DATA LINK
PHYSICAL
HIGHER LAYER PROT.
HIGHER LAYER PROT.
802-2014 clause 5.2.2
LOGICAL LINK CONTROL
802.1Q-2018 clause 6.8
EISS
MAC RELAY
INTERNAL SUBLAYER SERVICE
ETHERNET CONVERGENCE FUNCTION
802.1Q-2018 clause 8.1.1
802.1Q-2018 clause 6.7.1
802.1AC-2016 clause 13.1
802.3-2018 clause 64
MULTIPOINT MAC CONTROL
MAC CONTROL
802.3-2018 clause 31
MAC (eMAC)
MAC (pMAC)
MAC MERGE
802.3-2018 clause 4 or 4A
802.3-2018 clause 4 or 4A
RECONCILIATION
802.3-2018 clause 99
xMII
PCS
PMA
PMD
PHY
MDI
MEDIUM

Problem Statements: Introduction

Background

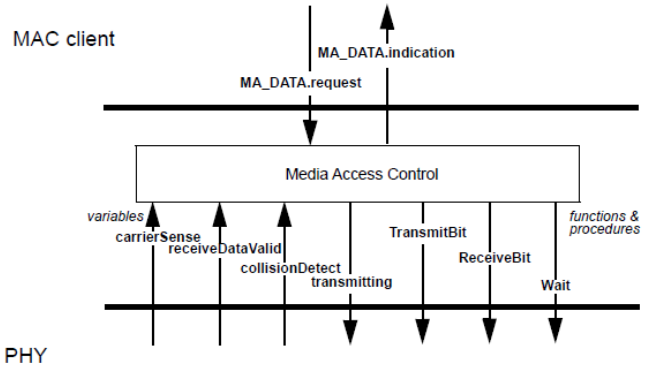
- It is intended to standardize CTF for Bridges and Bridged Networks in IEEE WG 802.1. A proposal on how this can be done has been outlined before.
- The lower layers in this tutorial were intentionally left out. In particular, details on the interface between IEEE 802.1 and IEEE 802.3 were omitted

Potential Problem Summary

- Frame-level synchronous interface at the MAC service interface
- Invalid MAC frame handling
- Handling of MAC Control frames
- Normative statements are not always clear in sections of the spec (where the style guidelines are not strictly followed)

Refinement

- The subsequent content details the problem further.
- **Remarks:**
  - **None of this presentation implies a requirement to change the standardized MAC Service Interface!**
  - **It is not intended to shift functions from IEEE 802.3 to IEEE 802.1 or vice versa!**



MA_DATA.request (
destination_address,
source_address,
mac_service_data_unit,
frame_check_sequence
)

MA_DATA.indication (
destination_address,
source_address,
mac_service_data_unit,
frame_check_sequence,
reception_status
)

Source: Clause 2 of IEEE Std 802.3-2018

IEEE 802.3 Considerations: MAC Service Interface

Background

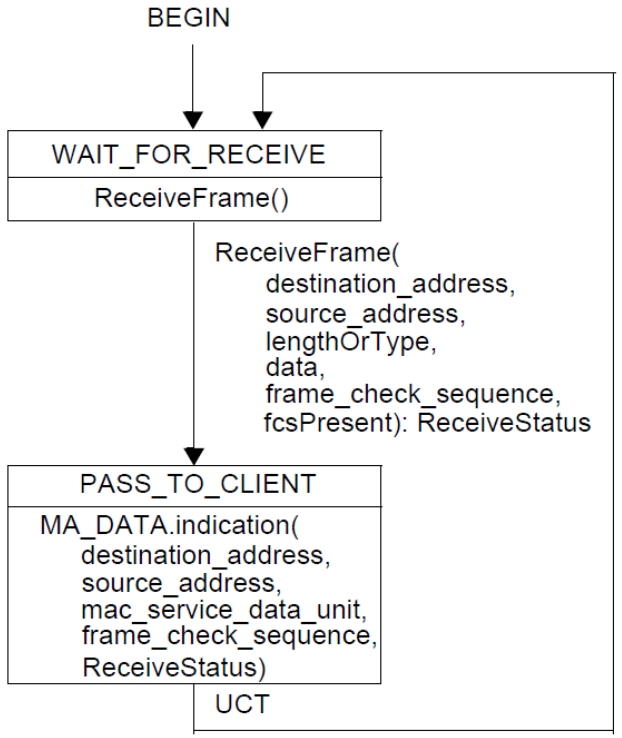
- The MAC service interface is specified in IEEE Std 802.1AC-20xx making use of ISO/IEC 10731 : 1994 [3.2 and 13.1 of IEEE Std 802.1AC-20xx].
- A service primitive is to be regarded as taking place as an *instantaneous event* [6.2, 7.2, and Figure 5 of ISO/IEC 10731 : 1994], which appears to be the case at least in IEEE Std 802.3-2018 [Figure 1-3 of IEEE Std 802.3-2018].
- Moreover, it appears that the ordering relationship between service primitive invocation and the precise specification of CSMA/CD MAC method and precise specification of MAC method of the simplified full duplex MAC suggest a sequential ordering between service primitive invocation and associated Pascal procedure call of [Figure 4-6, Figure 4-7, Figure 4A-3, Figure 4A-4 of IEEE Std 802.3-2018] based on associated state diagram conventions:

  *Labels on transitions are qualifiers that **must** be fulfilled before the transition **will** be taken* [1.2.1 of IEEE Std 802.3-2018].

  *Each primitive has a set of zero or more parameters, representing data elements that **shall** be passed to qualify the functions invoked by the primitive* [1.2.1.1 of IEEE Std 802.3-2018].

Concern

- **The MAC service interface and the relationship to the precise MAC specifications may prohibit CTF and enforce S&F.**
- **However, implementations are conformant as long as their externally visible behavior is identical to the model…**



Source: Figure 4A-4 of IEEE Std 802.3-2018.

IEEE 802.3 Considerations: Implementation vs. Model

Background

IEEE Std 802.3-2018 differentiates between an implementation and the model in state machines and the procedural models:

- *It is important to distinguish, however, between the model and a real implementation. The model is optimized for simplicity and clarity of presentation, while any realistic implementation **shall** place heavier emphasis on such constraints as efficiency and suitability to a particular implementation technology or computer architecture* [4A.2.2 of IEEE Std 802.3-2018].
- *It is the functional behavior of any unit that **must** match the standard, not its internal structure. The internal details of the model are useful only to the extent that they specify the external behavior clearly and precisely* [1.2.1 of IEEE Std 802.3-2018].
- *it is the behavior of any MAC sublayer implementations that **shall** match the standard, not their internal structure. The internal details of the procedural model are useful only to the extent that they help specify that behavior clearly and precisely* [item b) in 4.2.2.1 and 4A.2.2.1 of IEEE Std 802.3-2018].
- *The handling of incoming and outgoing frames is rather stylized in the procedural model, in the sense that frames are handled as single entities by most of the MAC sublayer and are only serialized for presentation to the Physical Layer. In reality, many implementations **will** instead handle frames serially on a bit, octet or word basis* [item c) in 4.2.2.1 and 4A.2.2.1 of IEEE Std 802.3-2018].

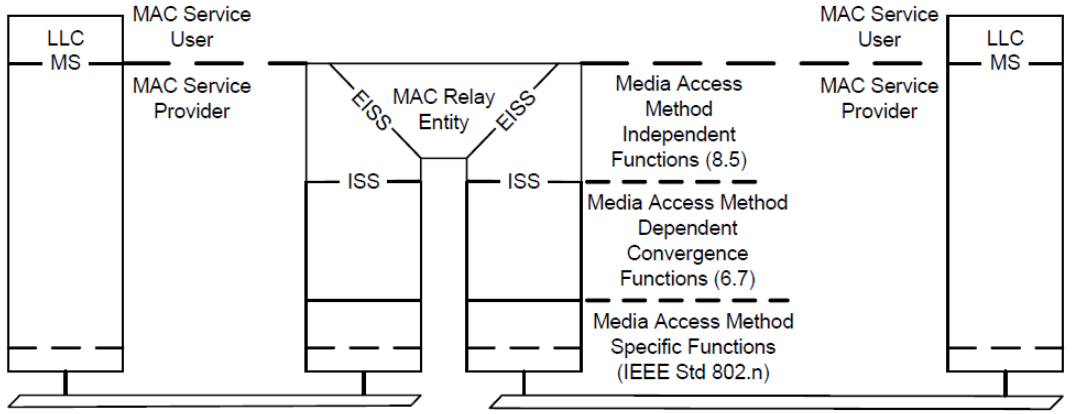
Observations

- The requirement for implementations is conformance to the externally visible behavior, not the specified structure, in certain areas of IEEE Std 802.3-2018 (but not all).
- It seems to be a statement of fact that many implementations would not be limited to the S&F operation implied by the MAC service interface.

IEEE 802.3 Considerations: Invalid MAC frames

Background

- *During reception, the contents of invalid MAC frames (e.g., frames that would fail FCS verification)* ***shall*** *not be passed to LLC and MAC control sublayers* [3.4 of IEEE Std 802.3-2018].
- There are other similar statements in IEEE Std 802.3-2018, although not using normative language (e.g., clause 2.3.2.3).
- The MAC control sublayer is optional and located between MAC client (i.e., Bridge) and MAC transparently [4.1.1 and 4A.1.1 of IEEE Std 802.3-2018]. It is thus on the path from ingress to egress in a Bridge. In contrast, the LLC is not on this path [clause 6 in IEEE Std 802.1Q-2018] and excluded from further consideration.



Source: Figure 6-1 of IEEE Std 802.1Q-2018.

Concern

- The requirement stated in 3.4 of IEEE Std 802.3-2018 is normative, but it appears to be a requirement for the model specified in IEEE Std 802.3-2018 (not for implementations).
- Implementations of the associated state machines [clause 31 of IEEE Std 802.3-2018] only need to match the external visible behavior, not the internal structure (previous slide).
- **However, in the case this requirement does actually apply for implementations, it could imply a conflict if MAC control sublayer(s) are present.**

IEEE 802.3 Considerations: Minimum Frame Size

Background

The minimum frame size of 64 octets is required for CSMA/CD operation of a CSMA/CD MAC [clause 4 IEEE Std 802.3-2018] and by the simplified full duplex MAC [clause A4 IEEE Std 802.3-2018].

Observations

- Both clause 4 and 4A MACs enforce a 64 octet minimum frame size on both Rx (smaller frames discarded) and Tx (smaller frames padded)
- The MAC merge sublayer is likewise ensuring that the minimum frame size constraint requirement is satisfied [Figure 99-4, 99.3.5 and 99.4.4 of 802.3-2018].
- The conditions that qualify invalid MAC frames [items a), b) and c) in 3.4 of IEEE Std 802.3-2018] do not directly relate to the actual frame length (i.e., a frame with less than 64 octets is not automatically an invalid MAC frame).

Concern

- The basic CTF operation is independent of the minimum frame size constraint…
- **… however, if truncation of frames under transmission after late error discovery is desired, the minimum frame size constraint may have some implications:**
  - **If minimum number of truncated octets is less than 64, there is no guarantee that truncation becomes effective (i.e., no truncation of erroneous frames with minimum size).**
  - **Delaying frames under reception for 64 octets or more prior to any transmission of these frames would be necessary for ensuring truncation of at least 64 octets, but would reduce the delay performance of CTF.**

802.3 Cut-Through Forwarding Considerations Summary

- Cut-Through Forwarding involves layers spanning both 802.3 and 802.1
- There are open questions regarding on items in 802.3 that apply to Cut-Through Forwarding including
  - The MAC service interface
  - Invalid MAC frame handling
  - MAC control frame interactions
- 802.3 and 802.1 working together will yield the best results as the requirements and interfaces will be understood by both sides

Call for Actions

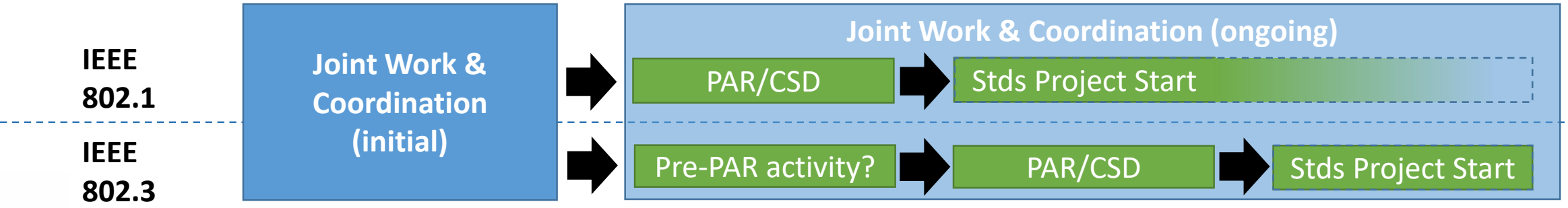
Johannes Specht

Call for Actions

Recap

- CTF is already implement in existing products – it is therefore technical feasible, but standardizing CTF is necessary for interoperability!
- Application areas, markets and use-cases for CTF, along with aspects of standardizing CTF, introduced in this tutorial.

Moving Forward Proposal



Joint work & coordination

- CTF needs the expertise of two IEEE 802 WGs.
- It appears vital to have some level of joint work & coordination on technical aspects and logistics (interfaces, meetings, etc.).
- A possible forum for initial joint work & coordination: IEEE 802 Nendica.

Administrative discussion (separate meeting)

- Topics
  - Which pre-standards activities can and should be done in a (significant) initial joint work & coordination phase?
  - What are the logistics?
  - …
- Meeting identified to plan followup: IEEE 802 Nendica weekly call on **August 5, 2021 09:00-11:00 ET** (see https://1.ieee802.org/802-nendica/)

→ ***If you are interested in this discussion you are very welcome!***

Questions & Answers

1. Introduction
   - Speakers
   - Cut-Through Forwarding (CTF)
2. Use Cases
   - Industrial Automation
   - Data Center Networks
   - ProAV
3. Summary: Goals and Objectives
4. IEEE 802.1 Considerations
5. IEEE 802.3 Considerations
6. Call for Actions
7. Q & A

Thank you for your Attention!

Questions, Opinions, Ideas?