Joint Tutorial

IEEE 802.3br

TF Interspersing express traffic (IET)

Ludwig Winkel, Siemens AG

and

IEEE 802.1 Time sensitive Networking (TSN)

Michael Johas Teener, Broadcom

Joint IEEE 802.1 / IEEE 802.3 Tutorial #2,
March, 2015
Berlin, Germany

Title and presenters

TITLE OF TUTORIAL:
Real-time Ethernet on IEEE 802.3 Networks

NAME OF PRESENTERS, THEIR AFFLIATIONS AND CONTACT INFO:

Presenter(s) Name:	Affiliation:	Email Address:
Ludwig Winkel	Siemens	Ludwig.Winkel@Siemens.com
Michael J.Teener	Broadcom	mike@JOHASTEENER.COM
Norm Finn	Cisco	nfinn@CISCO.COM
Pat Thaler	Broadcom	pthaler@broadcom.com
Panelists		
Albert Tretter	Siemens	albert.tretter@siemens.com
Stephan Kehrer	Hirschmann (Belden)	Stephan.Kehrer@belden.com
Christian Boiger	b-plus GmbH	christian.boiger@b-plus.com
David Brandt	Rockwell Automation	ddbrandt@ra.rockwell.com
Helge Zinner	Bosch	Helge.Zinner@de.bosch.com

Abstract

There have been multiple networks based on propriety technology or specialized standards developed to support carrying highly time sensitive traffic for applications such industrial automation and automotive control. Some of these are modified Ethernet networks. The efforts in IEEE 802.1 Time Sensitive Networking and P802.3br Interspersing Express Traffic provide an example of bringing together the requirements of those applications to provide a standard network that can support traffic requiring deterministic delivery time for real-time communication along with traditional traffic. This tutorial will cover the fundamentals of the projects and how they work together to fulfill the requirements of the various verticals.

Agenda

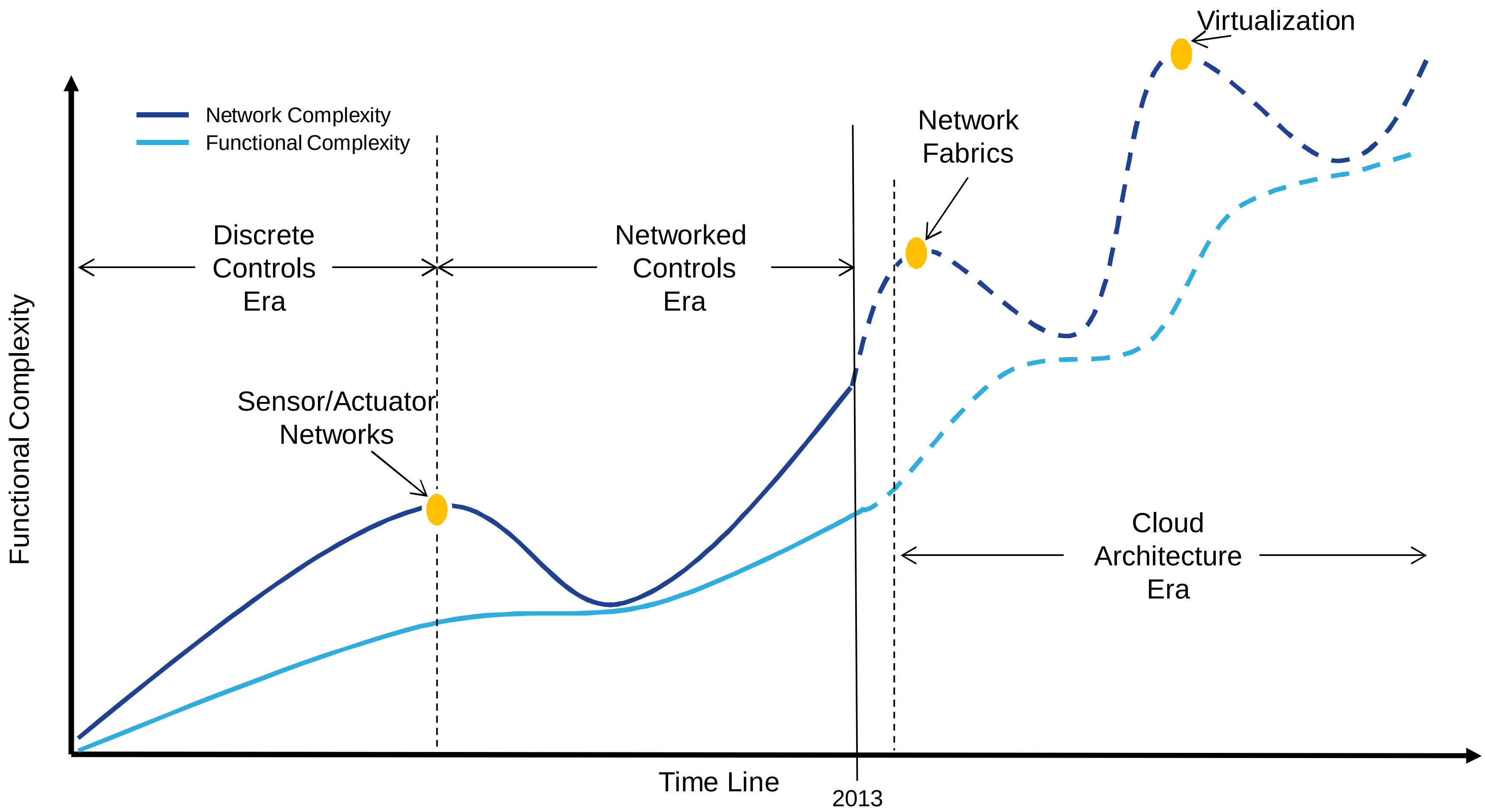
- Welcome, Introduction (Michael, Ludwig) 5min
- Recap of Geneva Tutorial (Ludwig) 5min
- Project time lines (Michael, Ludwig) 5min
- Architectural Options/System Overview (Norm) 30min
  - Enhancing IEEE 802.1Q tool set
- Interspersing Express Traffic (Pat) 30min
  - Preemption for Ethernet
- Conclusion,Q&A/Discussion (Michael, Ludwig) 15min

Recap of Geneva Tutorial

- Presented by Ludwig or Mike

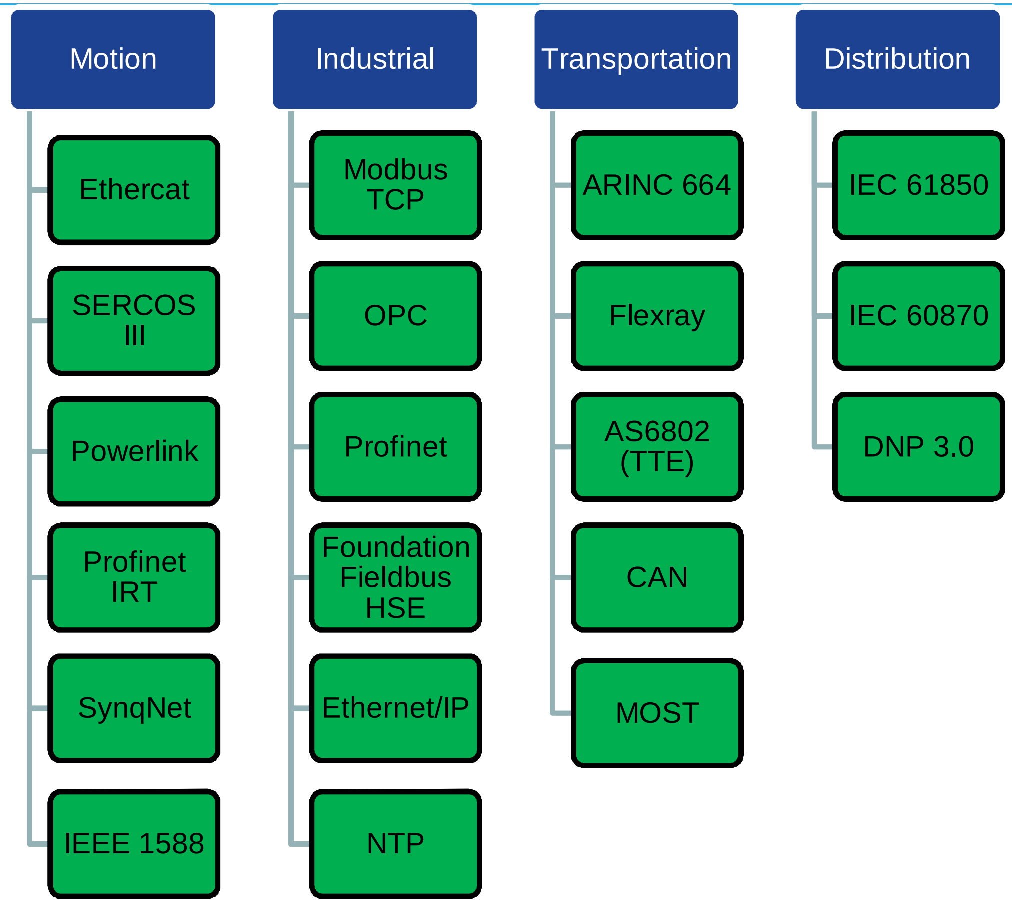
Potential Markets Served by IET

Control Function and Network Complexity Progression



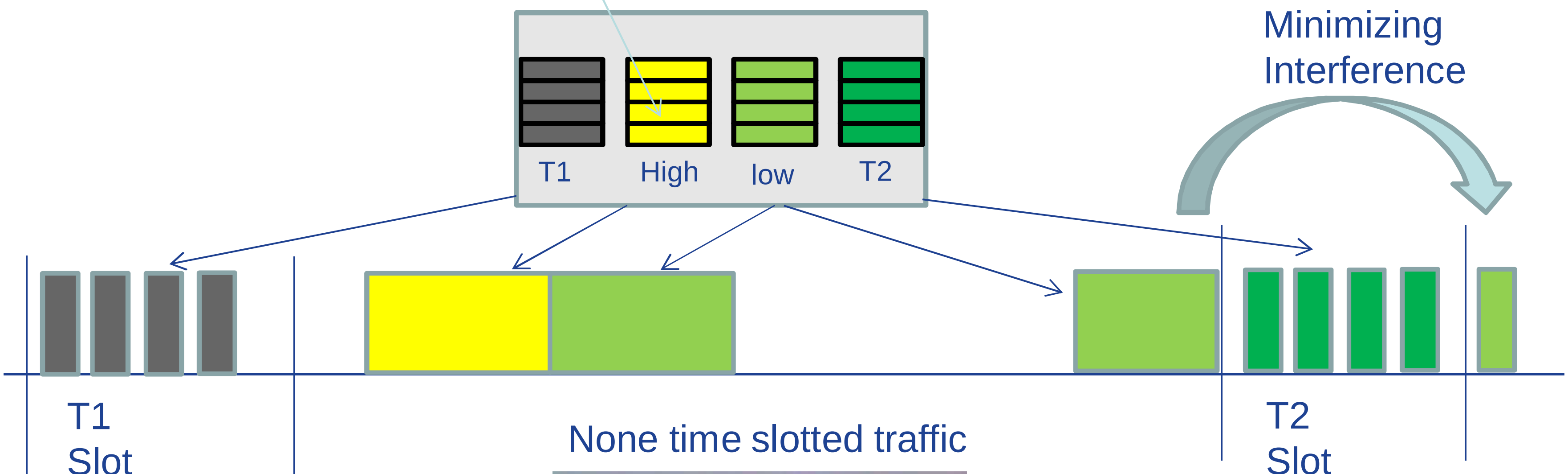
Control Systems in all market sectors perpetually increase in functional complexity.
Communications complexity limits functional capability.
Advanced communications architectures enable advances in controls.

Application Protocols for Control



Note: There are many other proprietary protocols not on this list

Why Converged Traffic Networks



Time Cyclic
Control Traffic

None Real time Traffic
- Logging
- Alerting

Real time – non cyclic Traffic
- Critical Alarms
- Discrete/event control

Why one single Network for all Communication Services

Only one network means:

- Reduced possibility of network failures
  - wire breaks, reduced confusion in case of maintenance
- Reduced installation costs
  - fewer cables and connectors, lower installed costs and faster startups
- Enables smaller devices
  - reduced space for connectors, lower power consumption (only half the number of PHYs needed)
- Reduced maintenance costs
  - easier to understand and to maintain, less personnel training
- Only one interface in the devices
  - only one MAC address, only one IP address, easier to understand and to maintain, easier coordination of the communication relations in the stack and application layer in the devices, more direct access to data.

Summary: Industrial Requirements for Interspersed Traffic

Performance requirements for Interspersed Traffic:

– **Minimum latency: < 3µsec max per hop accumulated latency (GE – min frame)**

– Guaranteed latency, low jitter

– Topology independent

– Typical data size (payload size): 40 - 300 bytes

– Range of transmission period: 31.25µs – 100ms and aperiodic

– Scheduled Traffic & Alarm has higher priority than Reserved Traffic and Best Effort Traffic

– Low cost, Low power, Low complexity

* These are our best estimates derived from multiple use cases of the current and future industrial applications.

Main Benefits of IET

- Better network utilization for scheduled traffic (More capacity).
- Lower latency for High Priority, critical asynchronous (non-scheduled) traffic.
- Lower cost and power consumption (for equivalent performance).
- Better environmental characteristics.

Project time lines, Work plan

- Work plan IEEE 802-3br:
  - TF review in Dec 2014 done
  - WG ballot in March 2015
  - Publication in 1Q/2016
- Work plan IEEE 802-1Qbu:
  - TG review in Sep 2014
  - WG ballot in Jan 2015
  - Publication in Sep 2015

Architectural options / system overview

Who, What, Why, How

Norman Finn

What is Time-Sensitive Networking?

Same as normal networking, but with the following features for **critical data streams**:

1. **Time synchronization** for network nodes and hosts to better than 1 µs.
2. Software for **resource reservation** for critical data streams (buffers and schedulers in network nodes and bandwidth on links), via configuration, management, and/or protocol action.
3. Software and hardware to ensure **extraordinarily low packet loss ratios**, starting at 10^{-6} and extending to 10^{-10} or better, and as a consequence, a **guaranteed end-to-end latency** for a reserved flow.
4. **Convergence** of critical data streams and other QoS features (including ordinary best-effort) on a single network, even when critical data streams are 75% of the bandwidth.

Who needs Time-Sensitive Networking?

- Two classes of bleeding-edge customers, Industrial (including in-automobile) and Audio/Video. Both have moved into the digital world, and some are using packets, but now they all realize they must move to Ethernet, and most will move to the Internet Protocols.

1. **Industrial:** process control, machine control, and vehicles.
   - At Layer 2, this is IEEE 802.1 **Time-Sensitive Networking (TSN)**.
   - Data rate per stream very low, but can be large numbers of streams.
   - Latency critical to meeting control loop frequency requirements.
2. **Audio/video:** streams in live production studios.
   - At Layer 2, this is IEEE 802.1 **Audio Video Bridging (AVB)**.
   - Not so many flows, but one flow is 3 Gb/s now, 12 Gb/s tomorrow.
   - Latency and jitter are important, as buffers are scarce at these speeds.

- (You won’t find any more market justification in this deck.)

Why such a low packet loss ratio?

Back-of-the-envelope calculations for big networks:

1. **Industrial:**
   - Automotive factory floor: 1000 networks • 10000 packets/s/network • 100,000 s/day = 10^{12} packets/day.
   - Machine fails safe when 2 consecutive packets of a stream are lost.
   - At a random loss ratio of **10^{-6}**, 10^{-12} is chance of 2 consecutive losses.
   - 10^{12} packets/day • 10^{-12} 2-loss ratio = **1 production line halt/day**.
   - In extreme cases, lost packets can damage equipment or require expensive measures to protect people.
2. **Audio video production:** (not distribution)
   - 10^{10} b/s • 10 processing steps • 1000 s/show = 10^{14} bits = 10^{10} packets.
   - Waiting for ACKs and retries = too many buffers, too much latency.
   - Lost packets result in a **flawed master recording**, which is the user's end product.

How such a low packet loss ratio?

1. **Zero congestion loss.**
   - This requires reserving resources along the path. (Think, "IntServ" and "RSVP") You cannot guarantee anything if you cannot say, "No."
   - This requires hardware in the form of buffers, shapers, and schedulers. Overprovisioning not useful: its packet loss curve has a tail.
   - Circuits only scale by aggregation in to larger circuits. (MPLS? Others?)
   - 0 congestion loss goes hand-in-hand with finite guaranteed latency.
2. **Seamless redundancy.**
   - 1+1 redundancy: Serialize packets, send on 2 (or more) fixed paths, then combine and delete extras. Paths are seldom automatically rerouted.
   - 0 congestion loss means packet loss is failed equipment or cosmic rays.
   - Zero congestion loss satisfies some customers without seamless redundancy. The reverse is not true in a converged network—if there is congestion on one path, congestion is likely on the other path, as well.

Why all the fuss? You could just …

- Old-timers remember the fuss 1983-1995 about Ethernet vs. Token Bus, Token Ring, and other "more deterministic" versions of IEEE 802 wired media. **Ethernet won.** One could argue that this TSN stuff sounds like the same argument. So, what's different besides, "That was then, this is now"?

1. TSN stays within the 802.1/802.3 paradigm.
2. Applications are more demanding of the network, now.
3. No IEEE 802 medium entirely captured the real-time control applications that drive the present effort—they went to non-802 (including non-packet) answers.
4. Yes, Voice over IP works pretty well—except when it doesn't. That's a non-starter for these users.
5. Too much data to overprovision.

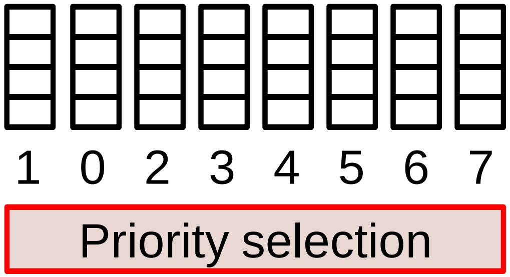
Queuing models

The IEEE 802.1Q Queuing Model

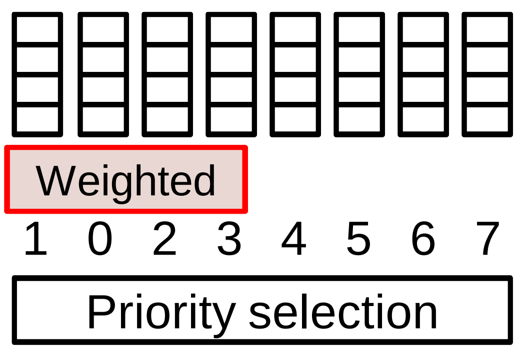
- IEEE 802.1 has an integrated set of queuing capabilities.
- There are several capabilities, most familiar to all.
- The “integrated” part is important—**the interactions among these capabilities are well-characterized and mathematically sound**.

Priority queuing and weighted queuing

- 802.1Q-1998: **Strict Priority**

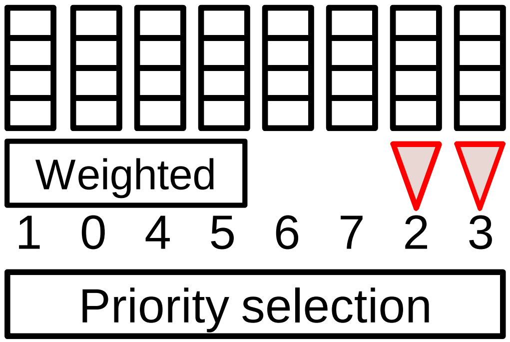


- 802.1Q-2012 (802.1Qaz) adds **weighted queues**. This standard provides standard management hooks for weighted priority queues without over-specifying the details.



AVB shapers

- 802.1Q-2012 (802.1Qat) adds credit-based **shapers**. Shaped queues have higher priority than unshaped queues. The shaping still guarantees bandwidth to the highest unshaped priority (7).

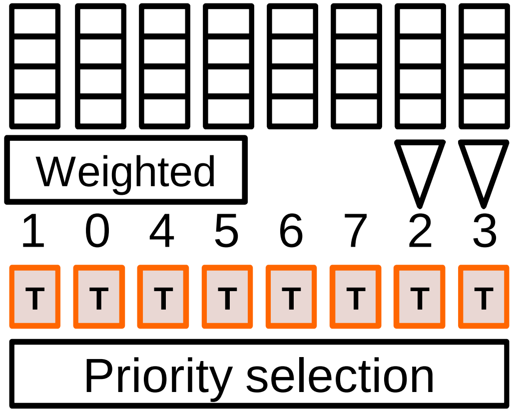


← Highest priority for shaped queues

- The AVB shaper is similar to the typical run rate/burst rate shaper, but with really useful mathematical properties.
  - Only parameter = bandwidth.
  - The impact on other queues of any number of adjacent shapers Is the same as the impact of one shaper with the same total bandwidth.

Time-gated queues

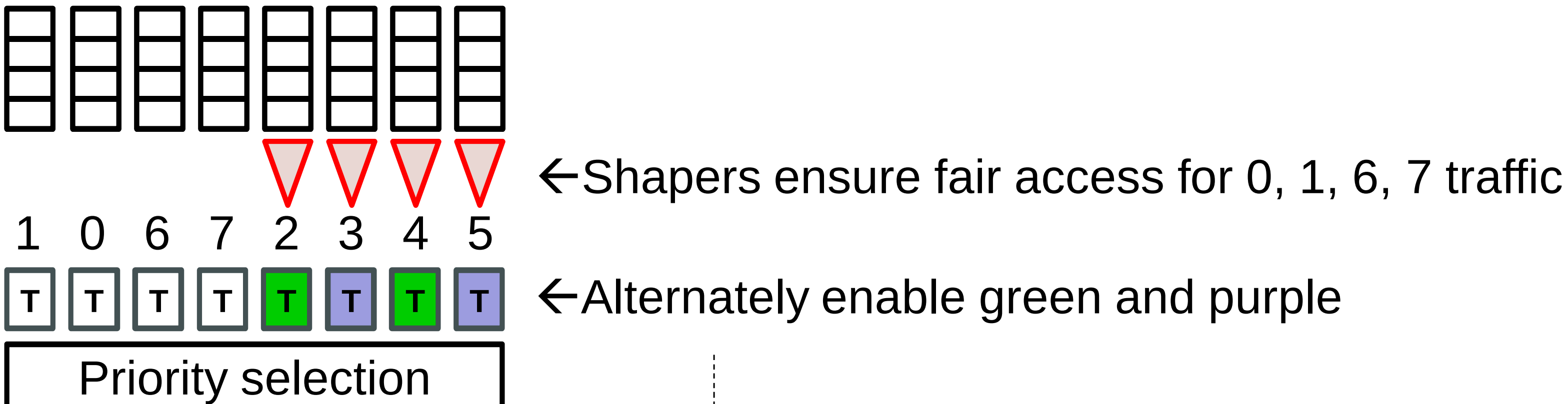
- 802.1Qbv: A circular **schedule** of {time, 8-bit mask} pairs controls gates between each queue and the priority selection function.



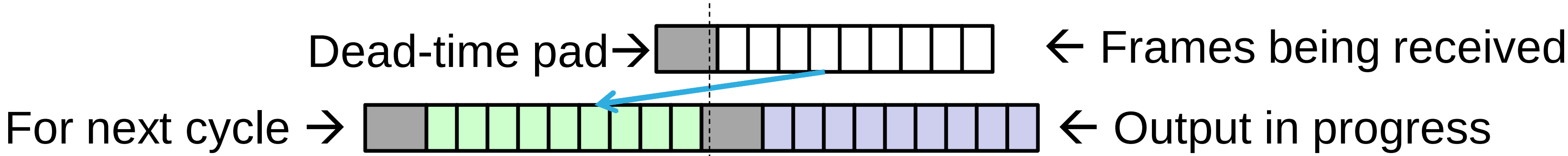
←Operated by a repeating schedule

Cyclic Queuing and Forwarding

- 802.1Qch: The 1Qbv time gated queues are used to create **double buffers** (two pairs, 2–3 and 4–5, shown in this example)



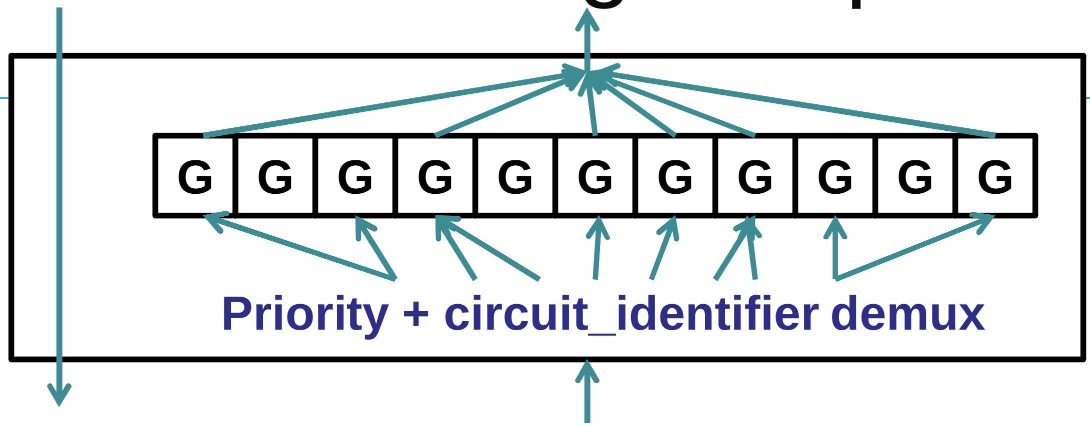
  – If the wire length and bridge transit time are negligible compared to the cycle time, double buffers are sufficient.



Security and misbehavior

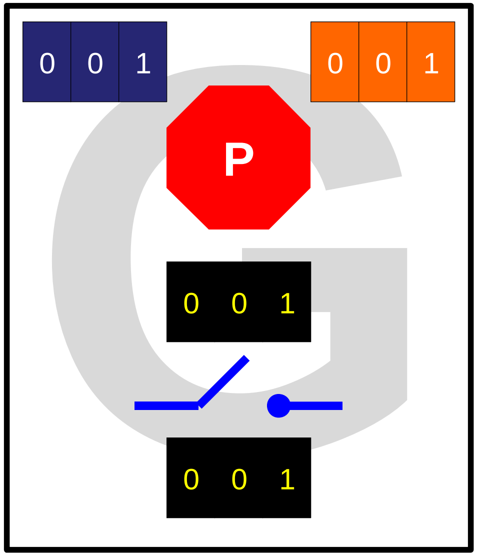
- Security has traditionally been concerned with
  - **Privacy**: Hiding the data from intruders
  - **Authentication**: Ensuring that the data is not altered.
- But now, proper operation depends upon the transmission **timing**, as well as the **contents**, of a packet.
- The only difference between a malicious intruder and a software bug, misconfiguration, or hardware failure is **intent**, not **result**.
- For example, a “babbling idiot” sending extra data on a TSN priority can cause the loss of packets from properly-behaving flows that share the same output queue.
- Therefore, **defense in depth** is required to protect the network.

Per-stream filtering and policing



Applies to frames coming **up** the stack, **not down**.

- The priority and packet flow ID (circuit_identifier) select to which Gate a frame is directed in P802.1Qci.
- Each Gate can have:
  1. A pass / don't pass **switch**. (May be time scheduled)
  2. A standard 802.1Q **policing** function.
  3. **Counters** of frames: e.g. passed, marked down, and discarded.
  4. A Service Class or priority output specifyer (TBD)
  5. Filters, e.g. max frame size.



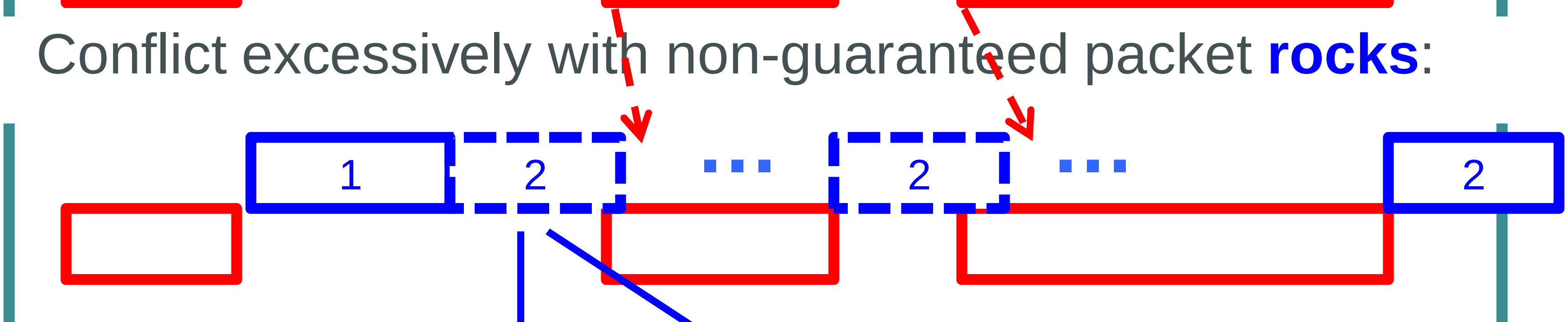
Interspersed Express Traffic

- Preempting a non-time-critical frame with a low-latency frame does get the low-latency frame out, sooner.
- But, in many networks of interest, there are many conflicting low-latency frames—and the preemption of the non-time-critical frame only helps the first one.
- Scheduling the time-critical frames' transmission (P802.1Qbv) gives almost 0 jitter and guarantees end-to-end latency. These scheduled transmissions are the "rocks" around which a time-critical application is built.

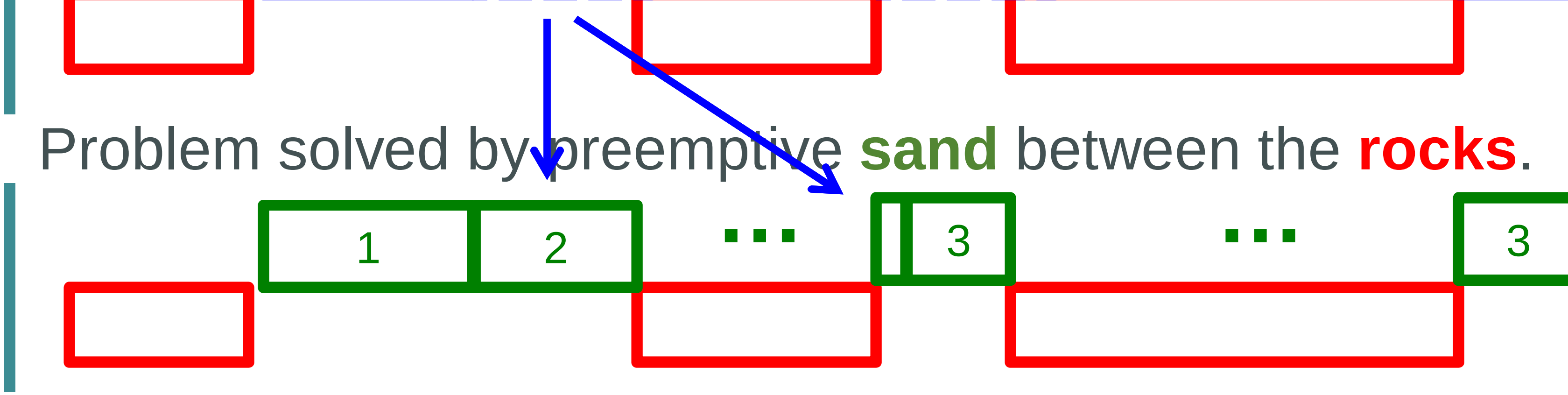
Interspersed Express Traffic

- IET is critical for convergence; non-scheduled does not mean “unimportant”.
- Scheduled **rocks** of critical packets in each **cycle**:

- Conflict excessively with non-guaranteed packet **rocks**:



- Problem solved by preemptive **sand** between the **rocks**.



But wait! There's more!

- As a consequence of the above, you also get …
- **Cut-through forwarding:** The scheduling tools mentioned, above, allow one to guarantee scheduled cut-through forwarding opportunities for predictable ultra-low-latency packets.
- **Intentional buffering delays:** Time-scheduled transmissions can intentionally delaying transmissions in order to guarantee both a minimum and a maximum latency, thus minimizing jitter for the critical traffic. Industrial systems that trigger events based on packet reception require this.

Current IEEE 802 Status

IEEE 802 standards now and coming

802.1 Audio Video Bridging is now the Time-Sensitive Networking TG.

- **Time:** A plug-and-play Precision Time Protocol (PTP) profile that allow bridges, routers, or multi-homed end stations to serve as "time relays" in a physical network, regardless of L2/L3 boundaries. (1AS complete, 1ASbt improvements in TG ballot)
- **Reservation:** A protocol (MSRP) to reserve bandwidth along an L2 path determined by L2 topology protocol, e.g. ISIS. (1Qat complete, 1Qcc enhancements in TG ballot)
- **Execution:** Several kinds of resources (shapers, schedulers, etc.) that can be allocated to realize the promises made by the reservation. (See next slides.)
- **Path distribution:** ISIS TLVs to compute and distribute multiple paths through a network. (1Qca in sponsor ballot)
- **Seamless Redundancy:** 1+1 duplication for reliability. (1CB in TG ballot)

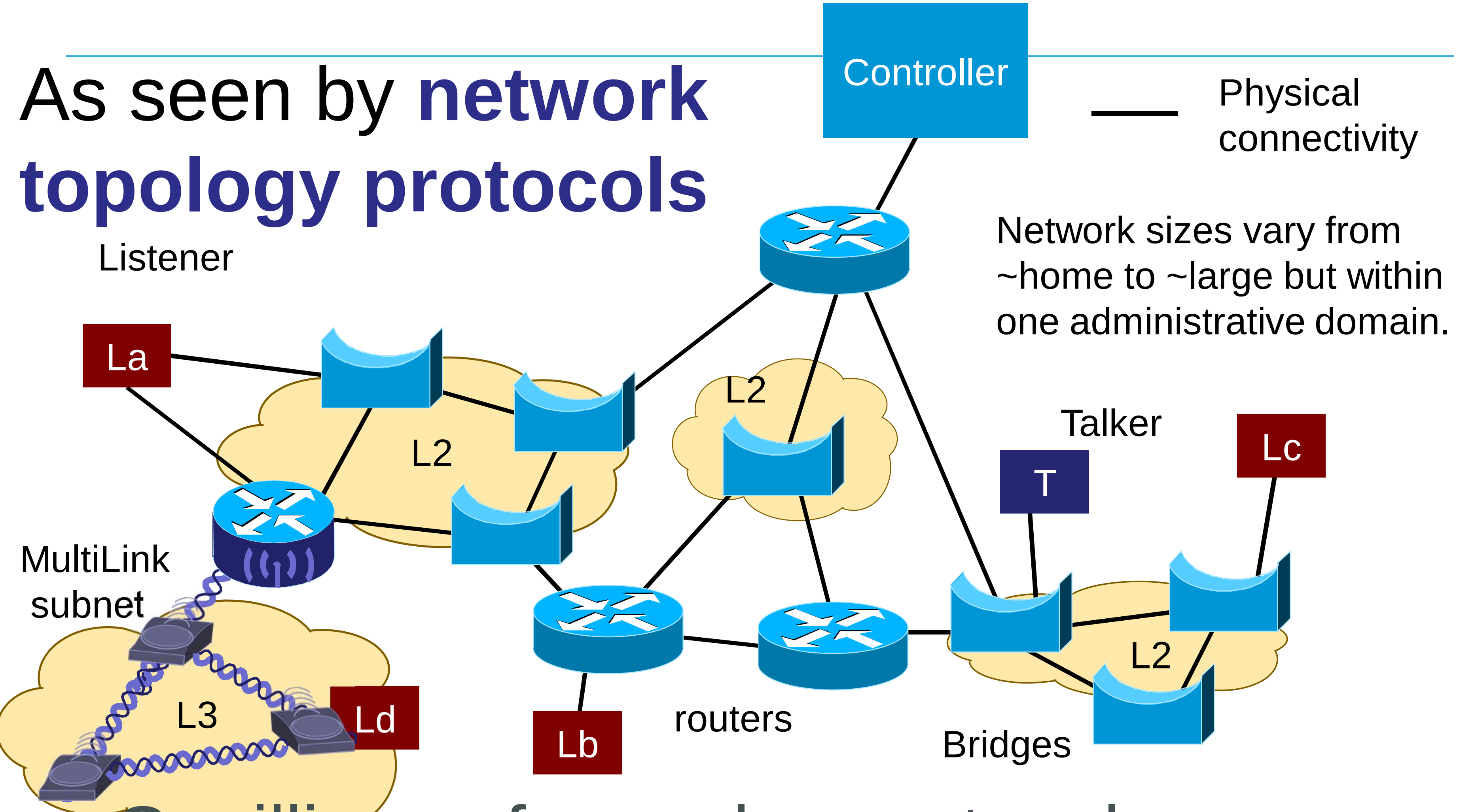
IEEE 802 schedulers and shapers

- **AVB Credit-Based Shaper:** Similar to the typical run rate/burst rate shaper, but with really useful mathematical properties. (**1Qat done**)
  - Only parameter = bandwidth.
  - The impact of any number of shapers = the impact of one shaper with the same total bandwidth.
- **Transmission preemption / express forwarding:** Interrupt (1 level only) transmission of an Ethernet frame with a frame with tight latency requirements, then resume the interrupted frame. (3br, 1Qbu TG ballot)
- **Time scheduled:** Every bridge port runs a synchronized, repeating schedule that turns on and off each of the 8 queues with up to nanosecond precision. (1Qbv WG ballot)
- **Synchronized Queuing and Forwarding:** Every flow proceeds in lock-stepped transmission cycles, like arterial blood. (1Qch PAR approval)
- **Per-Stream Filtering and Policing:** Packets accepted only from the right port only at the right time or at the right rate. (1Qci PAR approval)

Mixed L2/L3 need

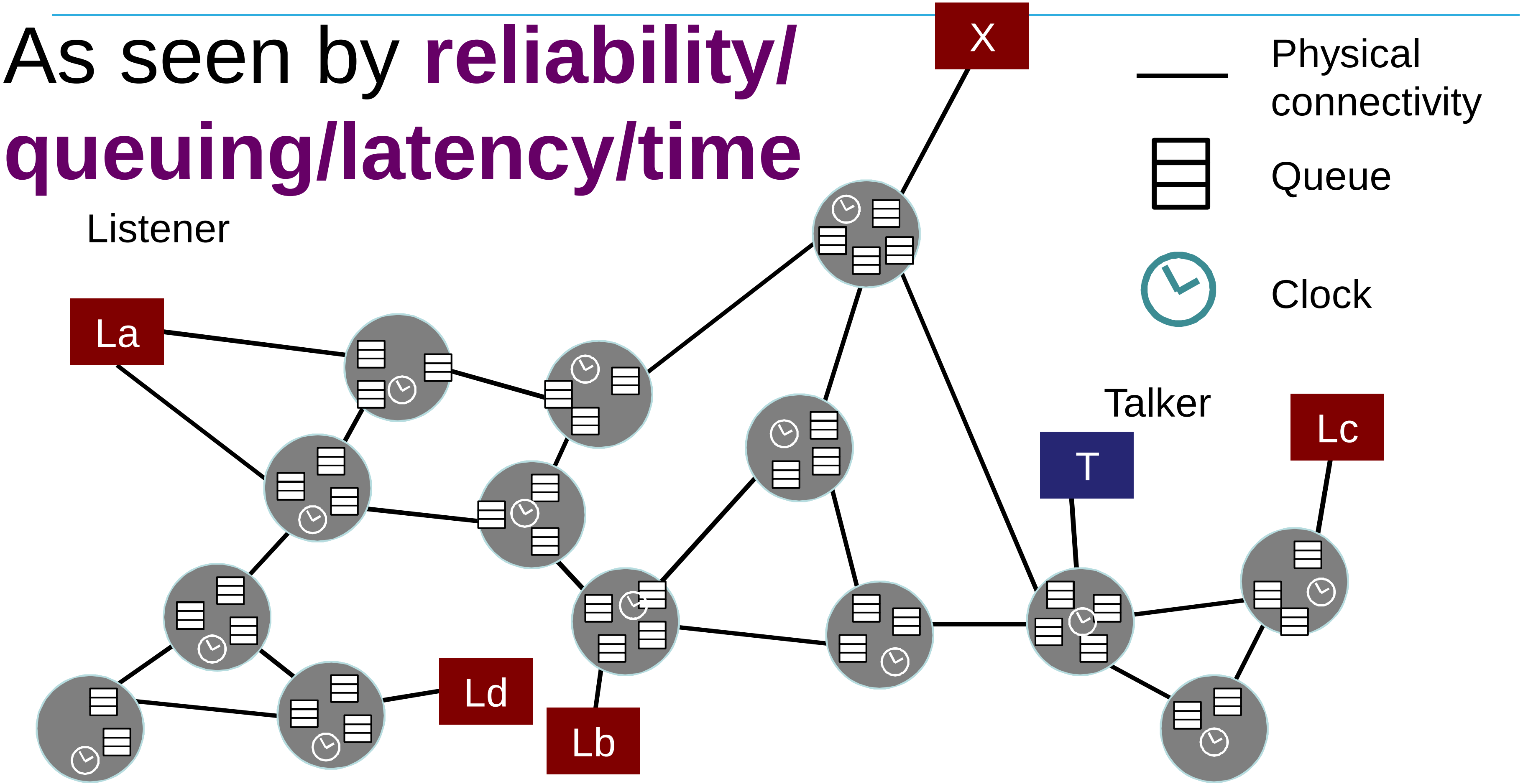
Reference network

As seen by **network topology protocols**



- Gazillions of complex protocols

Reference network



- Just nodes, queues, clocks, and wires!!

Summary

- By means of resource reservation, via protocol, configuration, or net management, time-critical traffic can be guaranteed a low, finite end-to-end latency and extraordinarily low loss rate.
- Preemption enables these guarantees to be made without sacrificing the ability of the network to carry “ordinary” traffic, and without compromising the promises made to time-critical traffic.
- These features can, and should, work irrespective of L2/L3 boundaries, though of course, proper layering must be respected.

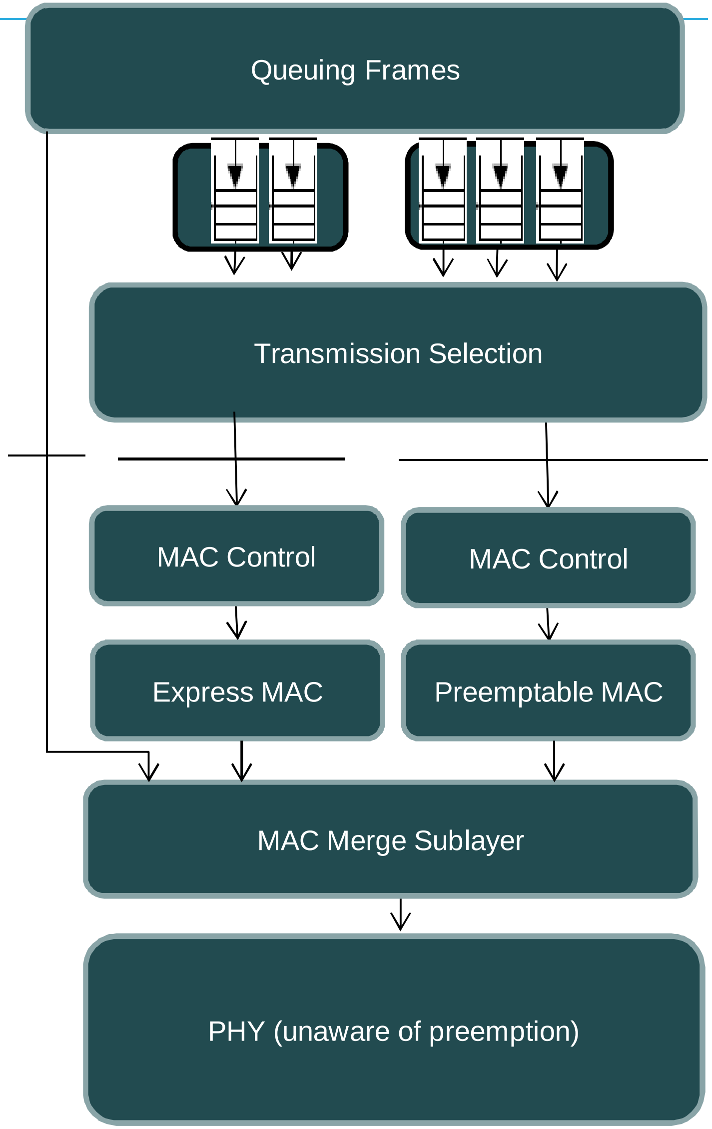
Interspersing Express Traffic

Preemption for Ethernet

Pat Thaler

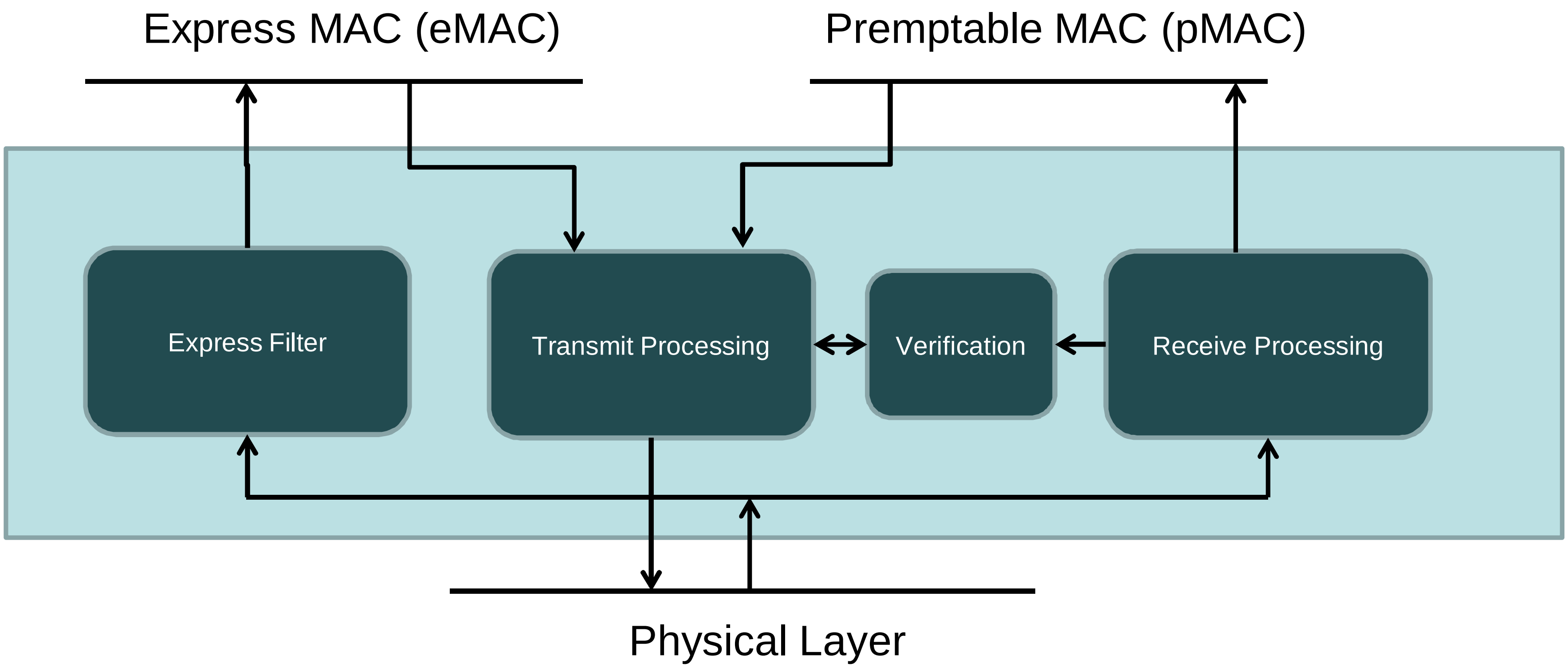
IET Architecture

- MAC Merge sublayer
  - Capability discovery without negotiation
  - Preserves frame integrity
  - Is transparent to existing non-deprecated PHYs above 10 Mb/s
  - Doesn’t change MAC operation
  - Minimizes impact on throughput
  - Provides lower latency for express traffic
  - Provides cut-through for scheduled traffic



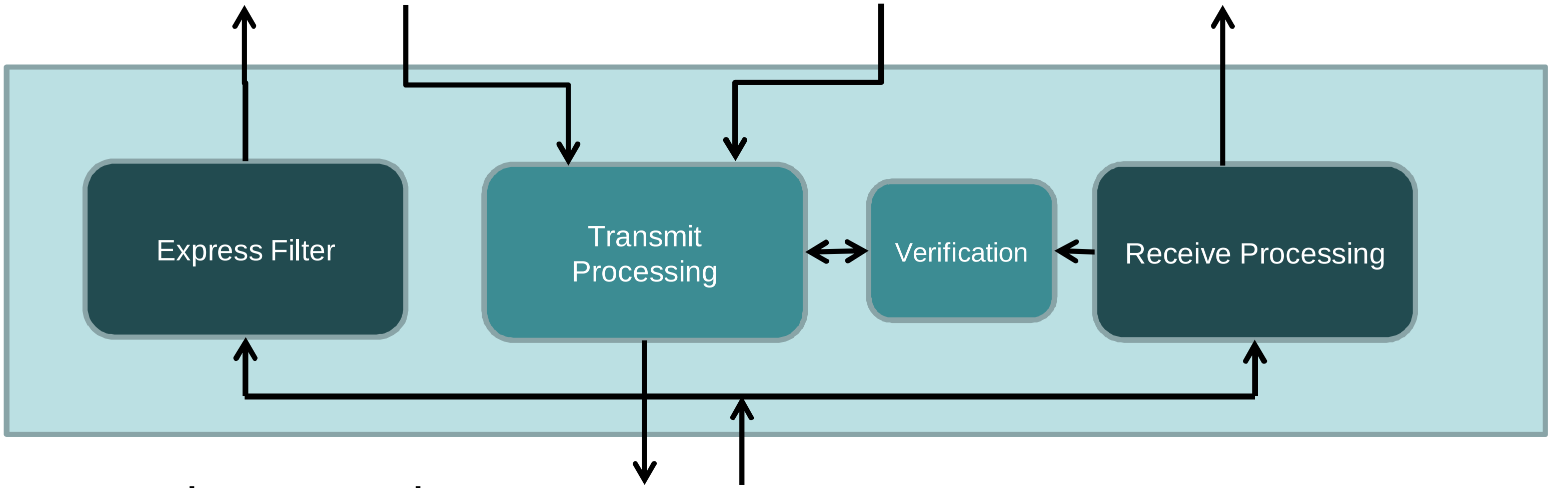
MAC Merge Sublayer

MAC Merge sublayer



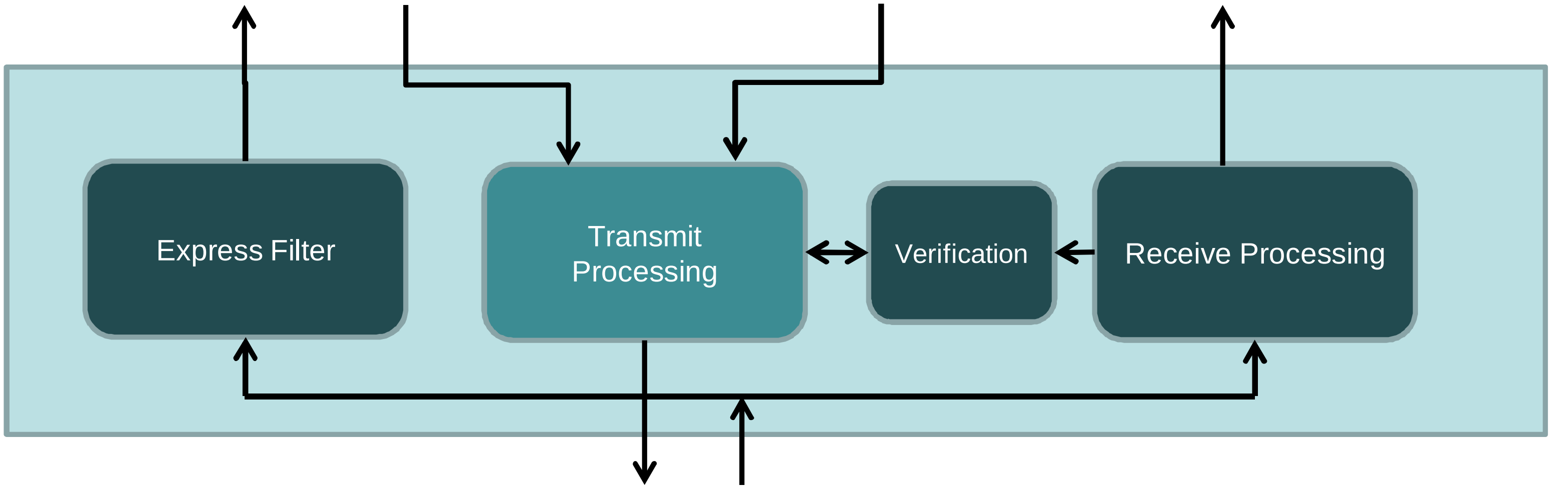
- Transmit processing arbitrates between eMAC and pMAC transmit packets and preempts if preemption capability is active.
- Express filter sends express packets to eMAC
- Receive Processing handles mPacket formats, checks fragments and sends to pMAC
- Verification tests that the link can support preemption before preemption is activated

Preemption capability disabled



- Transmit processing
  - eMAC packets have priority over pMAC packets
    - They don't preempt but if both have a frame ready to start, eMAC packet is sent.
  - Preemptable mPacket formats aren't used
- Able to receive preemptable mPackets from link partner
  - If link partner preemption capability isn't active, all packets received by eMAC
- Verification will respond to verify request from link partner

Preemption enabled, not active

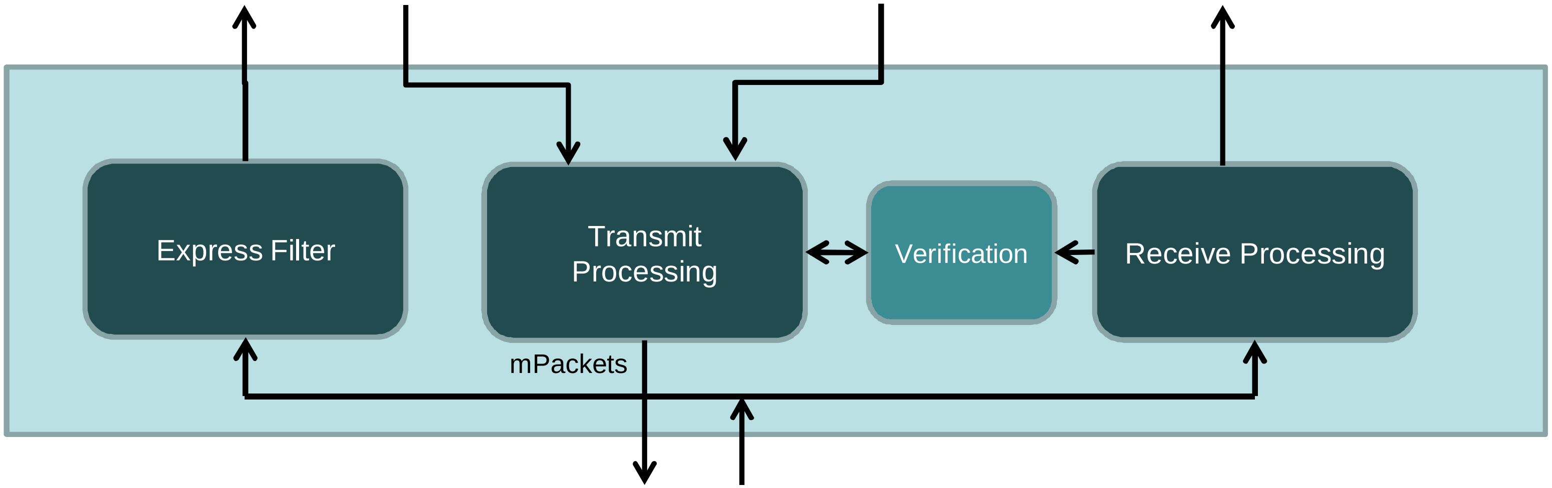


- Verification function attempts to verify link preemption capability
  - Transmits a verify mPacket
  - Receipt of a response mPacket verifies the link and preemption capability can go active.
- No change to Express Filter, Receive Processing or Transmit Processing

Why verify?

- A link partner's preemption capability is discovered through LLDP,
- IEEE 802.1Q bridges don't forward if the SA is nearest bridge group address, but …
- Some non-standard devices (e.g. buffered repeaters) don't block the address.
- If such a device is between two ports, it may drop or alter the preemptable mPackets.
- Verify tests that the link between to ports is able to carry preemptable mPacket formats.
- Networks that are fixed by design (e.g. automotive networks) can disable verification.

Preemption Active



- Transmit processing
  - Uses mPacket formats
  - Preempts preemptable packets if eMAC has a packet to send or for a HOLD request.
- Verification responds to verify requests

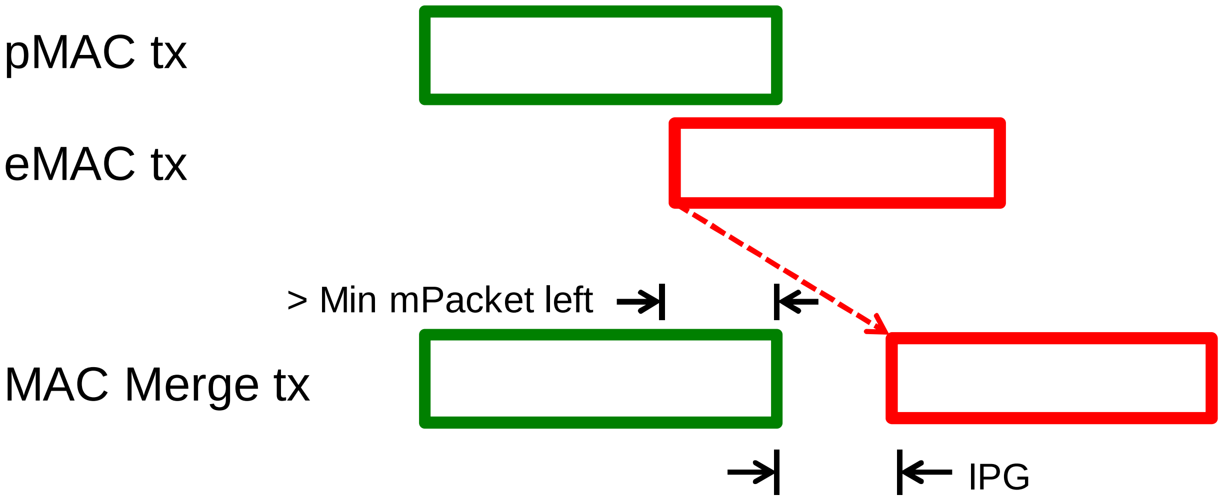
Discovery and verification summary

- Preemption capability independently activated on each end.
  - Capability discovery – not negotiation.
- Receiver is always ready for preemption
  - Receive Processing and Express Filter behavior is the same regardless of whether preemption capability is active.
- Link ability to support preemption is verified

MAC Merge Service Interface

Minimizing latency for scheduled traffic

Without Hold and Release

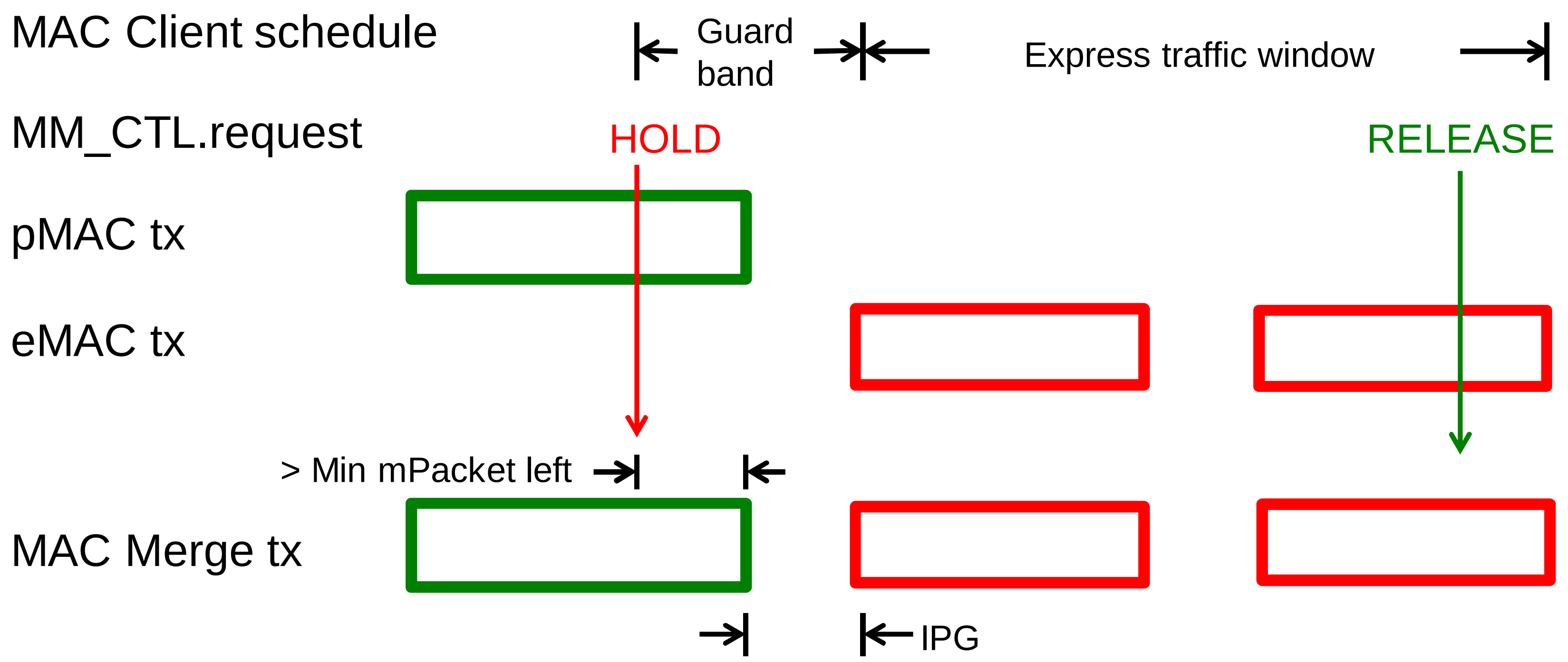


- Preemption isn't instantaneous.
- Packets with less than min packet size left to transmit or packets less than 123 octets can't be preempted.
- In many use cases, this delay is short enough but not in all cases.

MMSI Hold and Release

- MAC Merge Service Interface primitive:
  - Primitive from the MAC Client to MAC Merge sublayer
  - MM_CTL.request (hold_req)
  - hold_req takes one of 2 values: HOLD, RELEASE
  - hold stops transmission from the pMAC – preempting if preemption capability is active
  - release allows pMAC transmission.

With Hold and Release



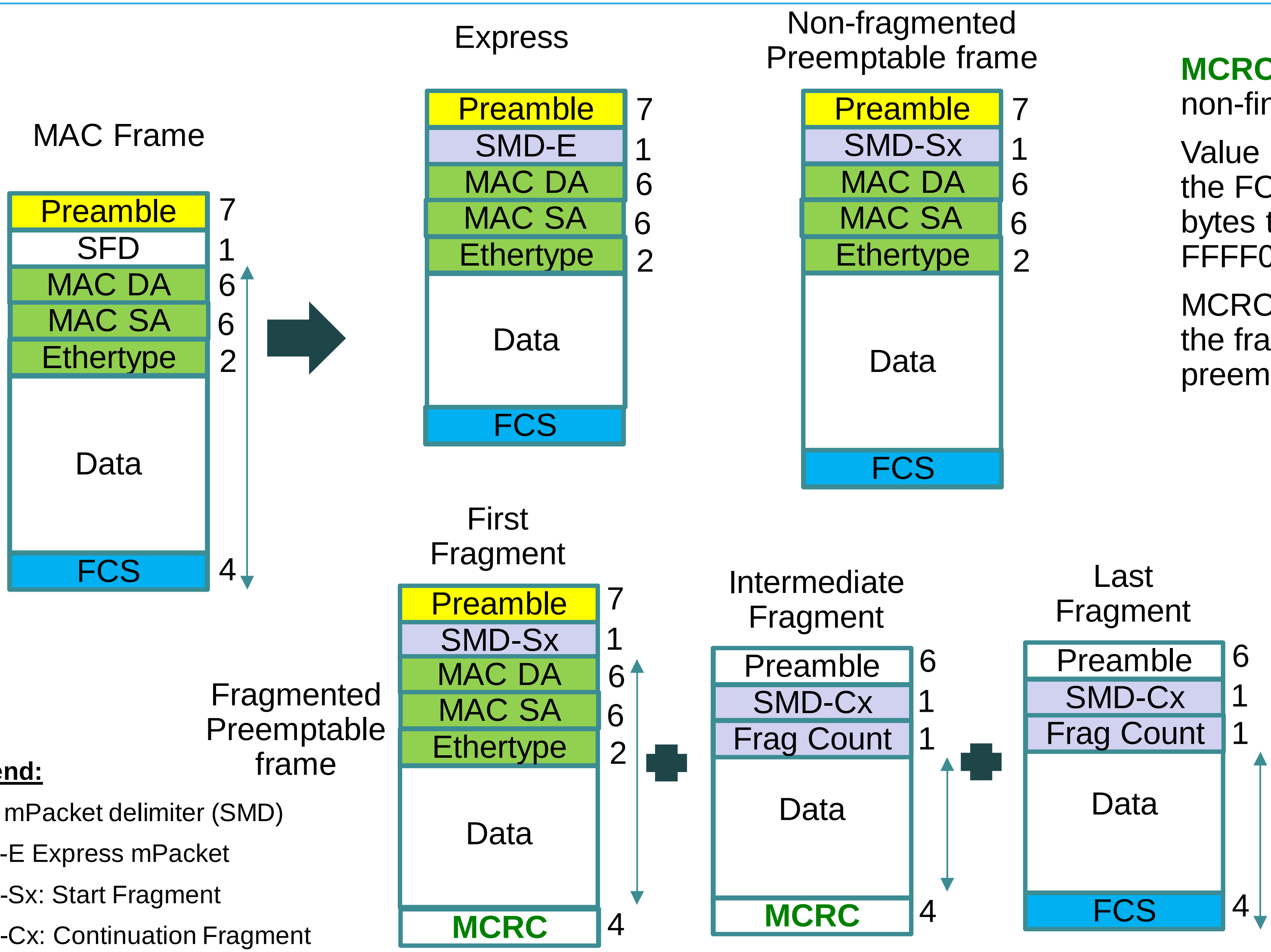
- Asserting MM_CTL.request (HOLD) a guardband in advance of a scheduled express traffic window ensures minimal latency (cut-through) for express traffic

mPacket Formats

Reassembly error protection

- Maintain Ethernet’s robust protection against false packet acceptance
- Detect any errors due to:
  - Up to 3 bit errors in mPacket format
  - Up to 3 lost fragments in a frame
  - Loss of last fragment of one frame and start of the next frame.
- By providing
  - Hamming distance of 4 between mPacket start delimiters
  - Mod 4 fragment count
  - Mod 4 frame count

mPacket Format



MCRC is the CRC of a non-final fragment.

Value is the same as the FCS of the frame bytes transmitted XOR FFFF0000

MCRC indicates that the frame has been preempted

Payload of each fragment (DATA plus CRC) ≥ min packet size

Legend:

Start mPacket delimiter (SMD)

SMD-E Express mPacket

SMD-Sx: Start Fragment

SMD-Cx: Continuation Fragment

SMD and Frag Count encoding

mPacket type	Frame #	SMD
SFD (express)	NA	0xD5
SMD-Sx Premptable frame start	0	0xE6
	1	0x4C
	2	0x7F
	3	0xB3
SMD-Cx Non-initial fragment	0	0x61
	1	0x52
	2	0x9E
	3	0xAD
Verify		0x07
Respond		0x19

Frag Count	Frag
0	0xE6
1	0x4C
2	0x7F
3	0xB3

mPacket summary

- Protects against reassembly errors
- Minimum impact on throughput
  - No extra overhead for un-preempted traffic
- Maintains Ethernet IPG and minimum packet size for compatibility with PHYs
- Compatible with all Ethernet full-duplex PHY standards operating at greater than 10 Mb/s

IET Summary

- IET
  - Supports preemption without change to the Ethernet MAC and PHYs
  - Maintains data integrity
  - Provides for capability discovery and verification
  - Supplies a primitive to further reduce latency for scheduled traffic

MACsec and Preemption

MACsec and preemption

- A port may have one Secure Channel (SC) serving both the express and preemptable traffic
- Preemption may alter the arrival of the packets
  - Not the only case where this happens, e.g. a Secure Channel running between Provider Bridging customer ports may reorder between priorities
- SCs transition from one Secure Association (SA) to another changing keys
  - A preemptable packet sent with the old key may complete after express frames with a new key.
  - Not a problem – SAs are designed to overlap and the MACsec header Association Number identifies the SA for the frame.
- MACsec header contains a Packet Number (PN) to provide replay protection
  - Default is strict replay protection
  - Out of order arrivals will be dropped
  - That would be a problem

MACsec/Preemption Solution Space

- Non-zero replayWindow parameter
  - Packets are tested for
    PN ≥ nextPN – replayWindow
  - If the test fails, packet is discarded
  - replayWindow default is 0 but it can be set higher to allow for some out of order arrival.
  - However it isn't always possible to predict how large replayWindow is needed to allow for the reordering and non-zero replayWindow slightly reduces security
- Use 2 Secure Connections
  - One for preemptable traffic and one for express
  - No reordering occurs within an SC and strict replay protection can be used.
  - Per traffic class SC is being considered in 802.1AEcg

Conclusion

- IEEE 802.1 Time Sensitive Networking and IEEE 802.3br Interspersing Express Traffic together enable real time traffic on Ethernet
- This supports applications such as
  - Industrial control systems
  - Automotive networks

Thus these applications can share a single network with traditional Ethernet traffic

THANK YOU

for your attention