Qdt Development

--- Feedback from TSN presentation
--- Discussion of TSN comments
--- PAR & CSD updates proposal

Lily Lv

Feedback from TSN Presentation

Conclusion of TSN Presentation

- TSN TG accepts the new proposal of PFC headroom measurement.
  - It is agreed to unbind Qdt with PTP.

Summary & Next steps

- .1Qdt draft work has started.
- The development of standard draft is limited until a PFC headroom measurement method can be decided.
- A new method of PFC headroom measurement method has been proposed in the Security TG.
- Modify PAR&CSD if new proposal is adopted.
- Question: Shall we produce a new draft with the new method, or do we need further discussion?

https://www.ieee802.org/1/files/public/docs2022/dt-lv-headroom-measurement-discussion-1122-v1.pdf

- 802.1 passes the motion to modify Qdt PAR&CSD.
  - The key point is to relax the requirement for using PTP in Qdt.

Motion

- 802.1 authorizes the TSN TG to draft PAR and CSD modifications of the P802.1Qdt Priority Flow Control Enhancements project to relax the requirement for using the Precision Time Protocol (PTP) at the January 2023 interim session for pre-circulation to the EC.

- Proposed: Paul Congdon
- Second: János Farkas
- In the WG (y/n/a): 36, 4, 7

https://www.ieee802.org/1/files/public/minutes/2022-11-closing-plenary-slides.pdf

Summary of TSN Presentation Feedback

1) Agree Pdelay has different objective with Qdt. This is the main reason for concerns.

- pdelay is designed for mean link delay and provides a way to ignore the variable delays higher in the stack, but Qdt cares about the higher layer delays

2) Suggest to Provide an option to use PTP if it is present in the data center.

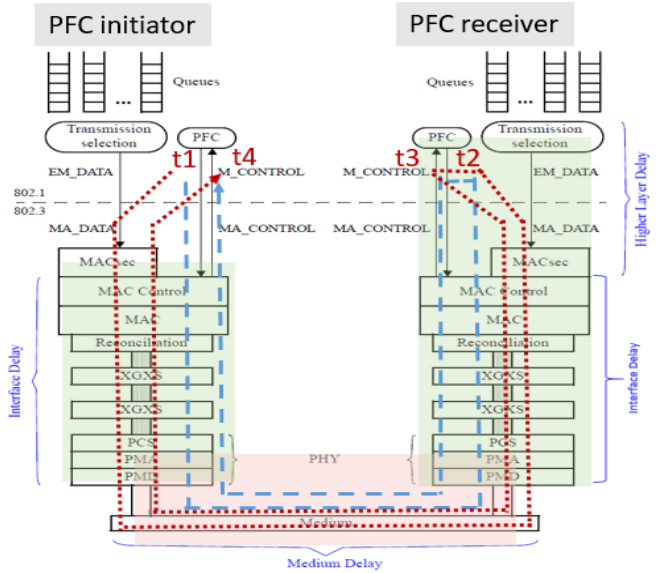
- Meta/OCP has PTP profile for DC (What parts of PTP is used in DC?)

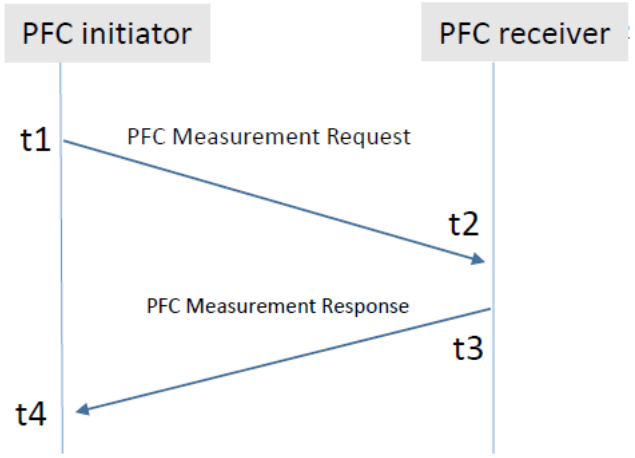
3) New Ethertype or re-use PFC ethertype for measurement frames?

4) Will (t3-t2) be impacted by queue delay?

- Answer： Since they are data frames, they will hit the queues, but we could use the high-priority queue

5) Clarify timestamp point (reference plane vs. message timestamp) in specification to let hardware vendors know how to implement it.





- DV = t4 – t1 – (t3 – t2) + HD ≈ t4 – t1

Discussion of TSN comments

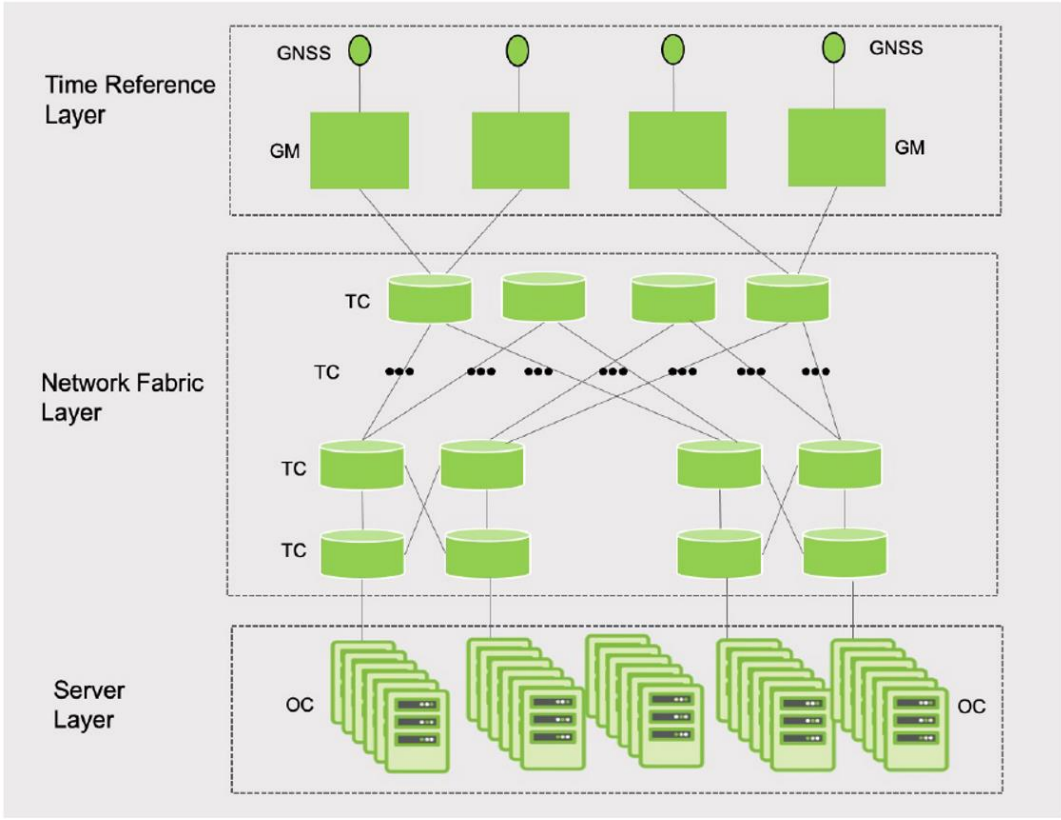
Item 2) PTP as an Option (1/5)

- What is Meta/OCP PTP profile for DC
  - Open Compute Project (OCP) https://www.opencompute.org/
    - A open-source community for datacenter innovations, initiated by Facebook in 2011.
  - Time Appliances Project (TAP) https://www.opencompute.org/wiki/Time_Appliances_Project
    - Led by Meta, aiming to synchronize elements in distributed system in order to support datacenter time-sensitive applications.
    - Data Center PTP Profile is one of its work streams

	Project	Objective
#1	Open Time Server	Development of an open time server for DC and Edge systems
#2	Data Center PTP Profile	Development of a PTP Profile tailored for data center applications
#3	Precision Time APIs	Time APIs to disseminate the time error (error bound) and bring accurate time to the user space
#4	Oscillators	Classification and measuring of oscillators
#5	PTP Servos	Design and Implement Advanced PTP Servos
#6	Instrumentation and Measurement	Open source instrumentation and measurement/testing tools for PTP

Item 2) PTP as an Option (2/5)

- What is Meta/OCP PTP profile for DC
  - PTP profile for DC
    - It uses Transparent clock model
      - "The network fabric layer consists of a chain of TCs."
      - The <meanPathDelay> computation is based on the end-to-end delay mechanism.
      - Messages allowed in this profile are Announce, Sync, Follow_Up, Delay_Req, Delay_Resp, Signaling, Management
    - Switch plays the role of **end-to-end TC**
      - TC "in this profile supports the delay request – response mechanism (i.e., end-to-end Transparent Clock)."
      - "In this profile, the TCs are assumed to be free-running. They are not syntonized either at the physical layer or via PTP"



https://github.com/opencomputeproject/Time-Appliance-Project/tree/master/DC-PTP-Profile

Item 2) PTP as an Option (3/5)

- PFC headroom measurement requires Pdelay mechanism supported by p2p TC, however, e2e TC and p2p TC cannot co-exist.
- **This OCP profile does not include p2p TC. So it does not support PFC headroom measurement.**

10.2 End-to-end Transparent Clock requirements

10.2.1 General requirements

All PTP messages shall be retransmitted in conformance with 7.3.1.

Except as noted in the following subclauses of 10.2, no changes in the PTP message common headers shall be made in the process of retransmission.

An end-to-end Transparent Clock shall not implement the peer-to-peer delay mechanism of 11.4.

10.3 Peer-to-peer Transparent Clock requirements

10.3.1 General requirements

All Announce, Sync, Follow_Up, PTP management, and Signaling messages shall be retransmitted in conformance with 7.3.1.

Except as noted in the following subclauses of 10.3, no changes in the PTP message common headers shall be made in the process of retransmission.

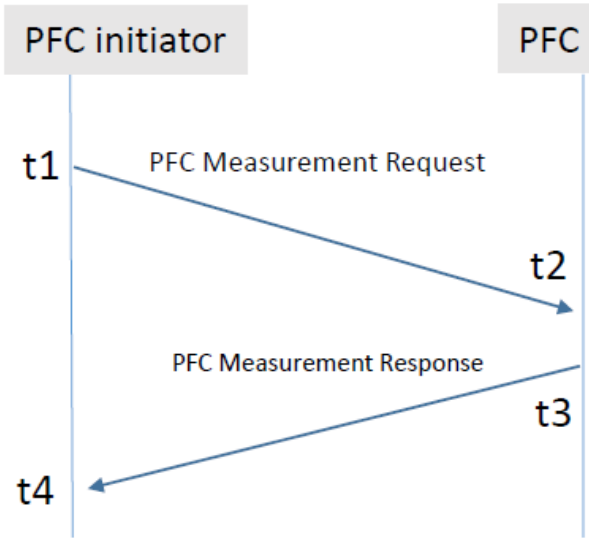
All PTP Delay_Req and Delay_Resp messages should be discarded.

The peer-to-peer delay mechanism must be implemented as specified in 11.4.

From 1588-2019

Item 2) PTP as an Option (4/5)

- The requirement of synchronization accuracy is becoming higher and higher.
- Although nowadays industry does not show interest in enabling PTP (full capability) in data center network, it might be possible to deploy “Pdelay” some day in future.
- Proposal to include PTP as an option
  - Indicate if it is a new dedicated method for PFC headroom measurement or PTP-based method in DCBX



- DV = t4-t1
  - DV = t4 – t1 – (t3 – t2) + HD ≈ t4 – t1
    - t3-t2 is the time to generate PFC headroom measurement Response.
    - HD is Higher layer delay (PFC reaction delay).
    - (t3 – t2) is similar as HD. But if the implementation is different, it should be accounted for. The response message can carry the compensation value, or DCBX can carry it.

https://www.ieee802.org/1/files/public/docs2022/dt-lv-headroom-measurement-discussion-1122-v1.pdf

New Method

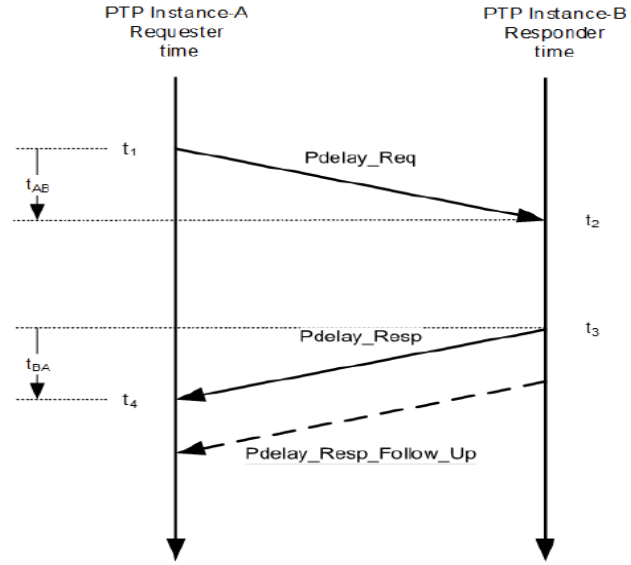
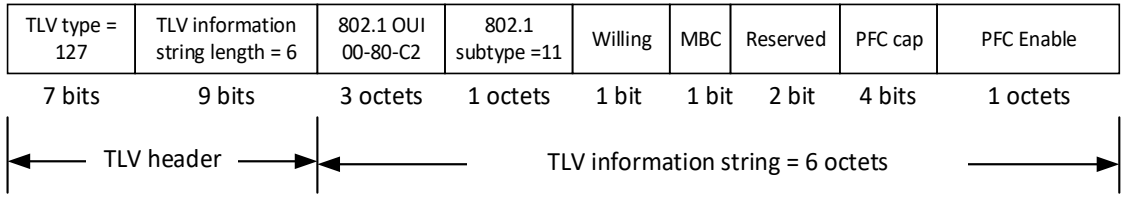


Figure 42—Peer-to-peer delay link measurement

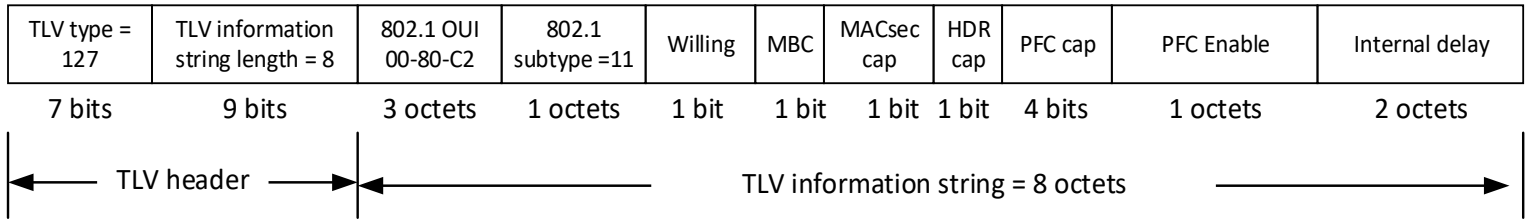
PTP-based Method

Item 2) PTP as an Option (5/5)

- Proposal to include PTP as an option
  - Indicate if it is a new dedicated method for PFC headroom measurement or PTP-based method in DCBX

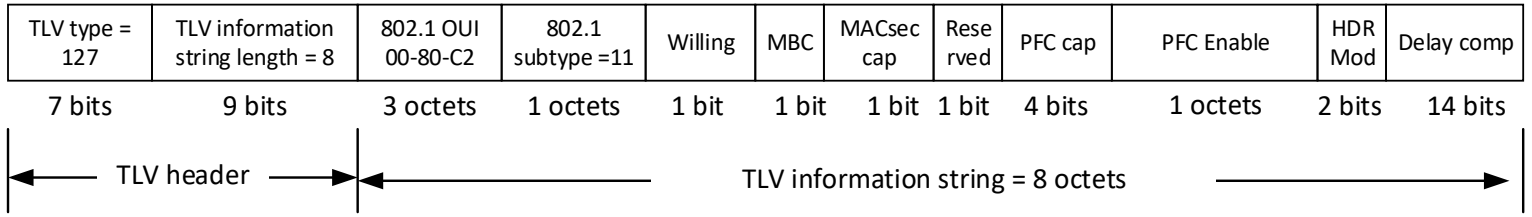


Original PFC configuration TLV format



PTP-based PFC configuration TLV format

- HDR Cap：A 1-bit unsigned integer that indicates the device support of automatic PFC headroom calculation
- A 2-octet unsigned integer contains the length of time for which the device process received PFC pause frame.

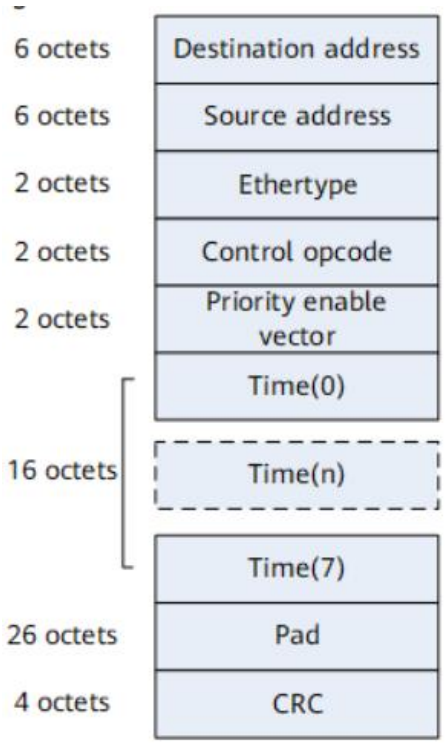


Multi-mode PFC configuration TLV format

- HDR Mod：
  - 00: do not support PFC headroom measurement
  - 01: New method
  - 10: PTP-based
  - 11: reserved
- Delay comp:
  - When HDR mod = 00 or 11: not valid
  - When HDR mode = 01: t3-t2 compensation value
  - When HDR mode = 10: internal delay

Item 3) Ethertype for Measurement Frames

- PFC frame is a MAC Control frame.



01-80-c2-00-00-01

"The globally assigned 48-bit multicast address 01-80-C2-00-00-01 has been assigned for use in MAC Control frames"

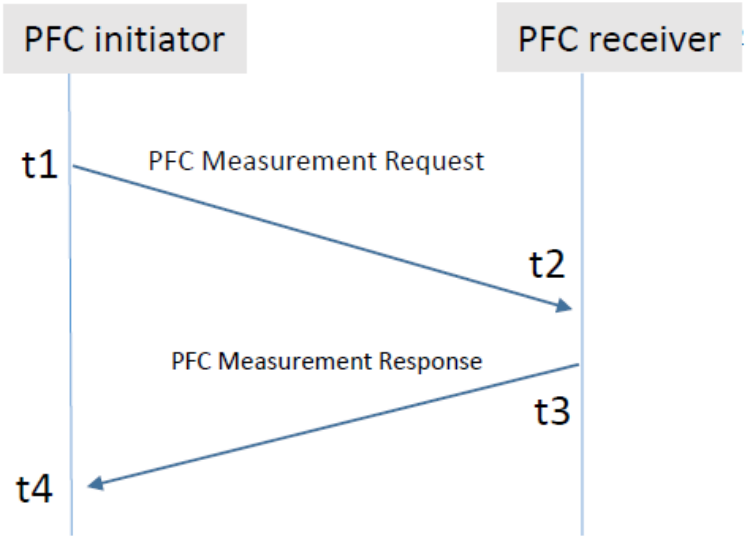
88-08

"The Length/Type field of a MAC Control frame is a 2-octet field that shall contain the hexadecimal value: 88-08. This value carries the Ethertype interpretation (see 3.2.6), and has been universally assigned for MAC Control of CSMA/CD LANs."

01-01

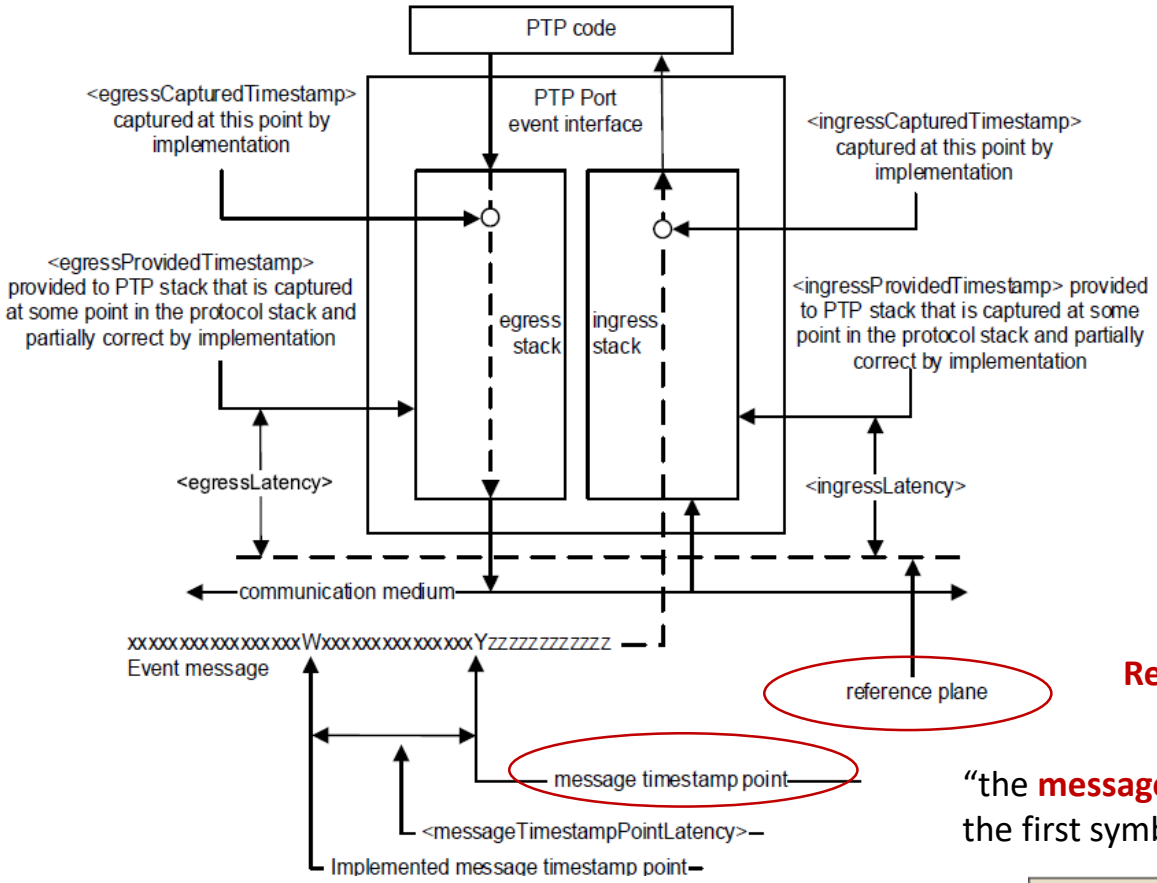
PFC

- Propose to design both request and response measurement frames as MAC data frames
  - If they are designed as MAC control frame, 802.3 would need to be involved. Better to minimize other standards dependency/involvement.
- PFC frame ethertype cannot be reused, as that is assigned for MAC control frames.
- **So measurement frames require new Ethertype.**



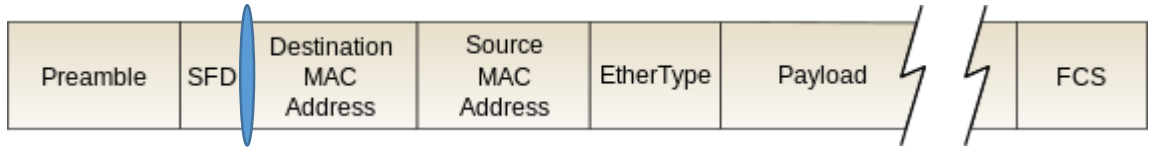
Item 5) Timestamp Point Clarification (1/2)

- 1588-2019 illustrates message timestamp point and reference plane.



Reference plane is close to medium.

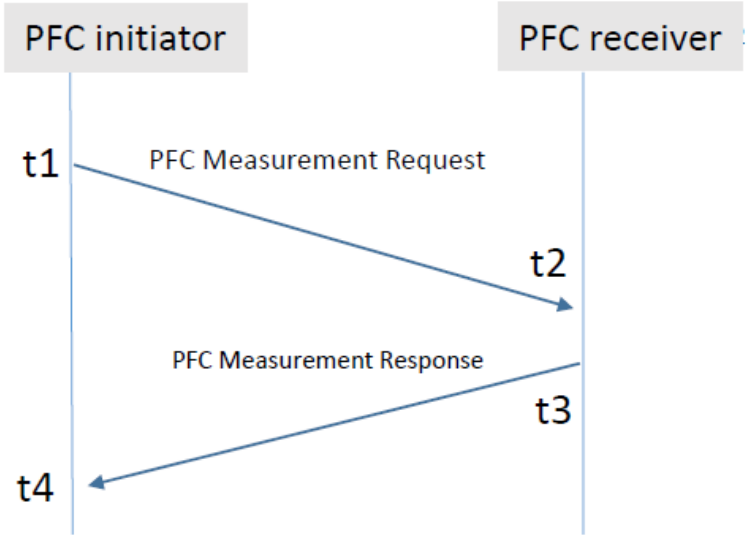
"the **message timestamp point** for a PTP event message shall be the beginning of the first symbol after the start of frame delimiter."

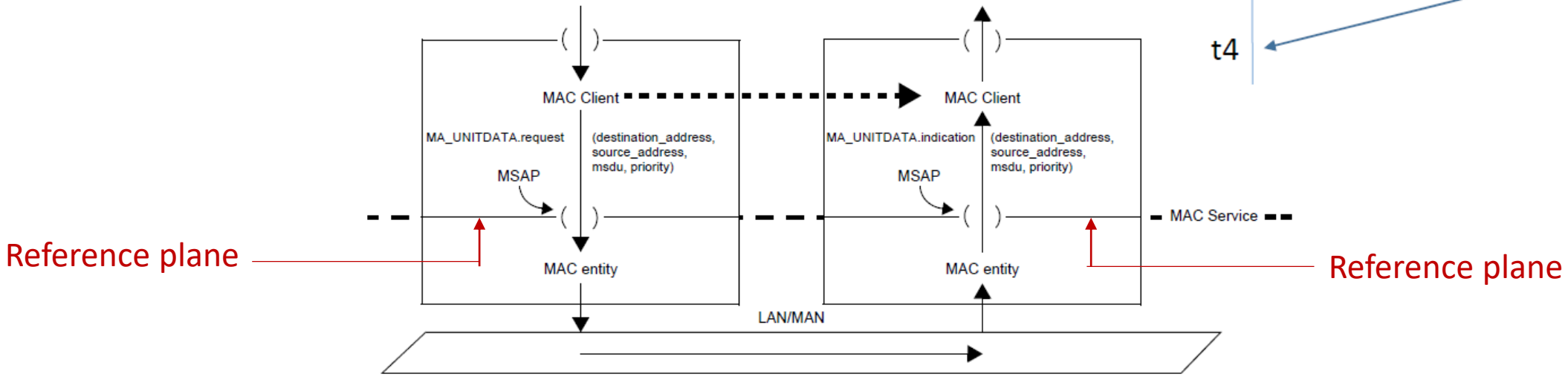


Item 5) Timestamp Point Clarification (2/2)

- PFC headroom measurement cannot reuse PTP message timestamp point. Because it is above MAC, there is no SFD.
- PFC headroom measurement uses different reference plane as PTP.

t1: the time when MA_UNITDATA.request of measurement request is invoked.
t4: the time when MA_UNITDATA.indication of measurement response is received.





PAR & CSD Updates

PAR Updates

5.2.b Scope of the project: This amendment specifies procedures and managed objects for automated Priority-based Flow Control (PFC) headroom calculation and Media Access Control Security (MACsec) protection of PFC frames, ~~using the existing Precision Time Protocol (PTP)~~ **using the point to point roundtrip measurement mechanism** and enhancements to the Data Center Bridging Capability Exchange protocol (DCBX).

This amendment places emphasis on the requirements for low latency and lossless transmission in largescale and geographically dispersed data centers.

This amendment also addresses errors of the existing IEEE Std 802.1Q functionality.

8.1 Additional Explanatory Notes: #5.2.b:
1) PFC and DCBX are specified in IEEE Std 802.1Q: IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks
~~2) PTP is specified in IEEE Std 1588: IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems~~
3) MACsec is specified in IEEE Std 802.1AE: IEEE Standard for Local and metropolitan area networks-Media Access Control (MAC) Security

CSD Updates

1.2.4 Technical Feasibility

b) Proven similar technology via testing, modeling, simulation, etc.

The proposed project enables peer nodes to advertise the new capability through the Data Center Bridging Capability Exchange (DCBX, specified in IEEE Std 802.1Q) mechanism which is widely deployed today using "Link Layer Discovery Protocol (LLDP, specified in IEEE Std 802.1AB). The principle of roundtrip delay measurement is well known and has been used in many different protocols. ~~Roundtrip delay measurements for participating systems are based on the existing Precision Time Protocol (PTP, specified in IEEE Std 1588) delay measurement mechanism.~~

1.2.5 Economic Feasibility

c) Consideration of installation costs.

A modest reduction in installation cost of new equipment is expected.

~~There are no incremental installation costs relative to the existing PTP and DCBX that will be used by the proposed standard.~~

There are no incremental installation costs by introducing roundtrip delay measurement and enhanced DCBX that will be used by the proposed standard.

Thanks