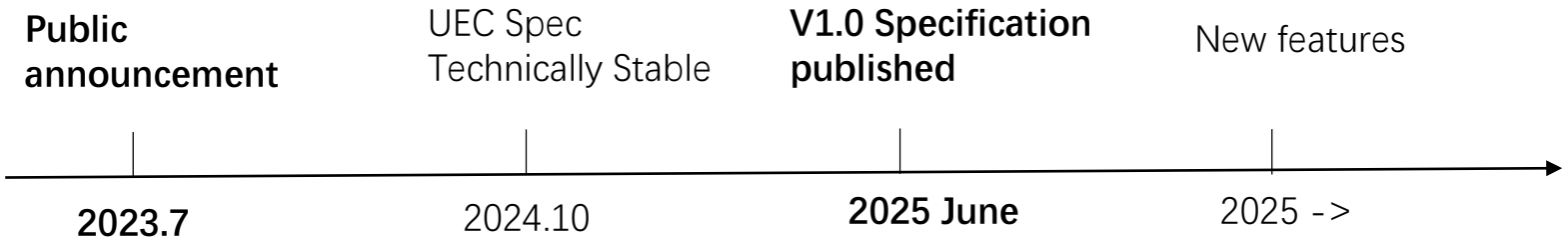
Discussion on Liasion Draft from IEEE802.1 to UEC

Lily Lyu (Huawei)

Background

Given the critical role of networks in AI cluster, the industry has seen rapid development. .

Ultra Ethernet Consortium(UEC) was established in July 2023 to address the emerging challenges in AI/HPC networking. Last month UEC spec1.0 has been publicly released.



IEEE802.1 has relevent work. It's necessary to collaborate with UEC.

- AI computing network (AICN) study item in Nendica
- P802.1Qdw source flow control(SFC) project
- P802.1Qdt PFC enhancement project

Liasion draft from IEEE802.1 to UEC has been vetted in Nendica.

Current Liaision Draft

The IEEE 802.1 Working Group would like to inform UEC and its members of its ongoing activities related to high-performance datacenter networking： AI Computing Network (AICN) study item and IEEE P802.1Qdw Source Flow Control project.

The AICN study item analyzes key requirements and challenges for AI training and inference networks. It also identifies potential areas for future IEEE 802 standardization efforts. This work is performed within IEEE 802 "Network Enhancements for the Next Decade" Industry Connections Activity (Nendica). AICN's website is https://1.ieee802.org/nendica-aicn/. Nendica addresses emerging requirements for all IEEE 802 networks and facilitate industry consensus towards proposals to initiate new standards development.

Project IEEE P802.1Qdw is specifying enhancements to flow control mechanisms, addressing known limitations of Priority-based Flow Control. It is conducted under IEEE 802.1 Time-Sensitive Networking(TSN) task group. The project website is https://1.ieee802.org/tsn/802-1qdw/.

Introduce IEEE802.1 relevent work

Both Nendica and TSN task group hold teleconference meetings, which are without registration fees, and meet during bi-monthly sessions, which require registration fee. Participation and contributions from interested individuals are welcome and encouraged. Nendica's website is https://1.ieee802.org/802-nendica/, and TSN's website is http://1.ieee802.org/tsn/.

Introduce IEEE802.1 logistics

IEEE WG 802.1 appreciates the public availability of Ultra Ethernet™ Specification v1.0, and notes with interest the announced forward-looking statements for future enhancements. We would like to request information and material regarding ongoing and future work (e.g., improved telemetry and congestion control, UE bindings for storage protocols, and in-network collectives). Please feel free to contact us with any questions.

Express IEEE802.1 request to UEC

Link to liasion draft: https://mentor.ieee.org/802.1/dcn/25/1-25-0008-06-ICne-proposed-draft-liaison-statement.pdf

New Received Comments

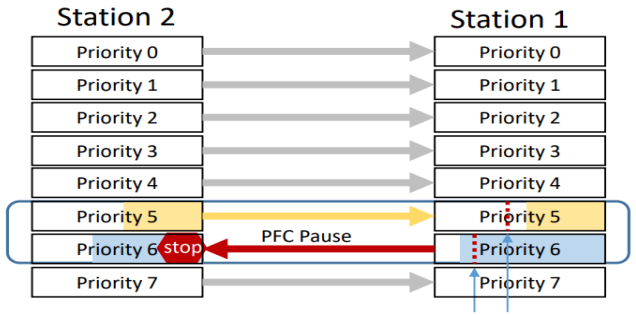
Start with addressing the callouts from UEC spec to 802.1 work:

- CBFC <----> PFC
- Trimming <----> SFC
- UE future technologies <----> AICN study item

PFC and CBFC

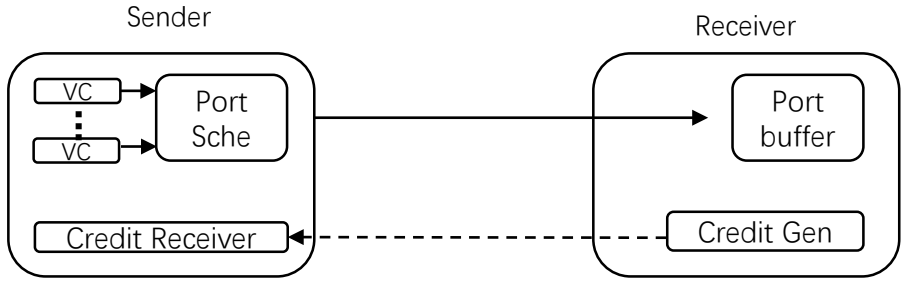
Priorty-Based Flow Control(PFC):

- 8 priorities
- ON/OFF control
- Passive way to control buffer usage



Credit-Based Flow Control (CBFC) – optional feature in UEC

- Up to 32 VCs (UEC)
- Credit allocation
- Proactive way to control buffer usage



PFC	CBFC
Passive control	Proactive control
Receiver	Sender
ON/OFF pattern	Credit allocation
Simple to implement, but sensitive to cable length/frame size/response time	Complex to implement
Full use of link bandwidth	Introduce messaging overhead
2BDP, but better buffer sharing across ports	1BDP, but no buffer sharing across ports

How can PFC and CBFC operate simultaneously?

- CBFC and PFC can be deployed in a mixed way in network?
- CBFC and PFC can be activated on same port(priority/VC)?

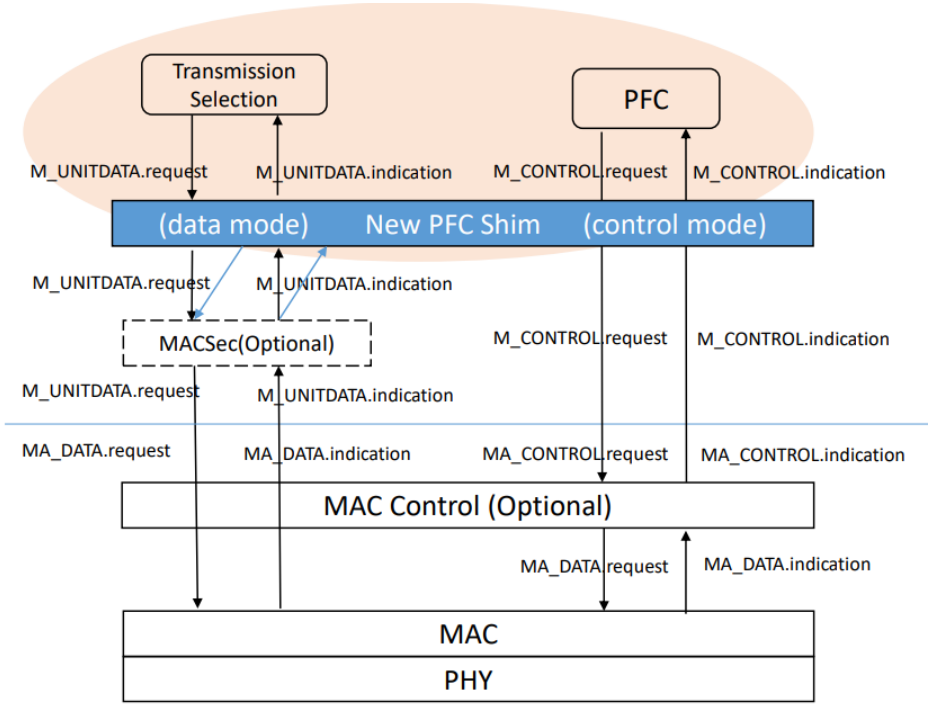
Will be any issue considering the fundamental differences between CBFC and PFC?

- Fixed buffer size for each port vs. Shared buffer among ports
- VC(up to 32) vs. Priority(8)

PFC and LLR

Link Layer Retry(LLR) -- optional feature in UEC

- LLR function is located between MAC client and MAC control
- Two additional fields are added to the interface from MAC client to LLR

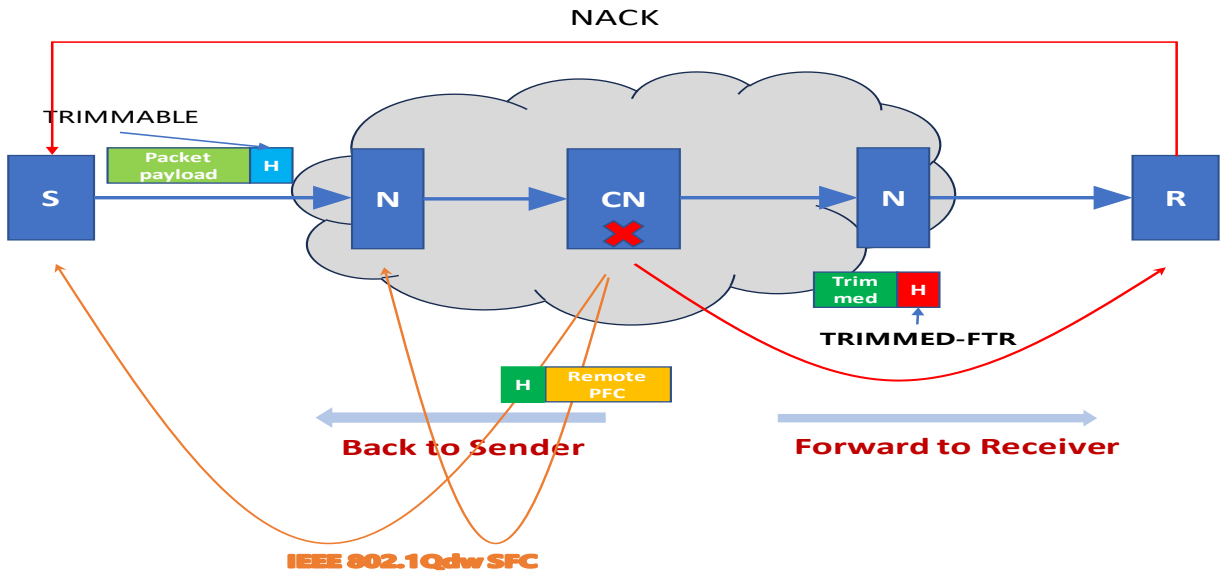


How does PFC interact with LLR?

- Control-mode PFC does not appear to be affected by LLR, as it interacts directly with MAC control, bypassing LLR.
- Data-mode PFC (introduced by Qdt project) requires the additional two fields to interface with LLR?
- LLR is above MACsec of below MACsec?

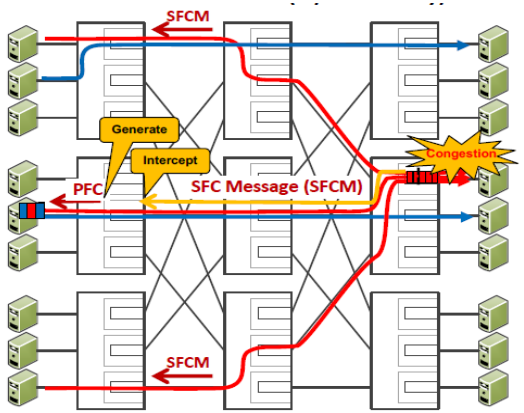
https://www.ieee802.org/1/files/public/docs2021/new-congdon-PFC-Headroom-and-Enhancements-0921-v01.pdf

SFC and Trimming



Packet trimming -- feautre in UEC

- Trimmed packets forwarded to receiver
- NACK sent from receiver to sender



Source Flow Control (SFC):

- Proxy mode
- Host mode

Can SFC augment trimming?

- SFC is a back-to-sender(BTS) approach, which differs from current forward-to-receiver trimming, but intuitively provides faster notification to the sender.
- Could SFC mechanism collaborate with, or leverage, trimming or other features in Congestion Management Sublayer(CMS) defined/will be defined by UEC?

Conclusion

- CBFC and LLR may have impact on PFC.
- SFC may augment trimming or other CMS features.
- Request to UEC:
  - Further clarification on simultaneous operation of CBFC and PFC
  - Re-evaluation of the impact of LLR on PFC
  - Share ongoing and future work of UEC

Proposed text in Liasion Draft(1/2)

Part 1: outline the relevence between IEEE802.1 and UEC and present IEEE802.1's requests

- IEEE WG 802.1 appreciates the public availability of Ultra Ethernet™ Specification v1.0. Several new features are recognized as being related to current activities within IEEE WG 802.1, including CBFC, LLR and packet trimming.
- CBFC is an alternative to PFC, but is also designed for possible simultaneous operation with. We would appreciate further clarification on how PFC and CBFC can be used together in practice, given their fundamental design differences (e.g. priority queue vs. VC, shared buffer vs. fixed buffer).
- LLR is supposed to be interact with PFC. Currently, the IEEE 802.1 Working Group is developing project P802.1Qt (PFC enhancement), which enables PFC pause frames to be transmitted via the MAC data path. Under this design, the PFC frame would pass through the LLR layer. Clarification on how LLR affects or interacts with PFC behavior in this context would be appreciated.
- Packet trimming is a potentially relevent feature. P802.1 Qdw(SFC) project is mentioned in UEC spec 1.0 as a Back-to-sender approach for congestion control, but not addressed in UEC spec 1.0. We would like to understand the concerns from UEC to not specify BTS and woud like to explore if SFC could augment or extend with packet trimming.
- IEEE WG 802.1 has another ongoing activity - AI Computing Network (AICN) study item. It analyzes key requirements and challenges for AI training and inference networks. It also identifies potential areas for future IEEE 802 standardization efforts. In this context, we have noted with interest the announced forward-looking statements for future enhancements in UEC. We would like to request information and material regarding ongoing and future work (e.g., improved telemetry and congestion control, UE bindings for storage protocols, and in-network collectives).

Proposed text in Liasion Draft(2/2)

Part 2: introduce IEEE802.1 logistics to facilitate individual's participation

- Above activities have been conducted respectively under IEEE 802.1 Time-Sensitive Networking(TSN) task group, Security task group and IEEE 802 "Network Enhancements for the Next Decade" Industry Connections Activity (Nendica) which addresses emerging requirements for all IEEE 802 networks and facilitate industry consensus towards proposals to initiate new standards development.
  - P802.1 Qdw project website is https://1.ieee802.org/tsn/802-1qdw/.
  - P802.1 Qdt project website is https://1.ieee802.org/tsn/802-1qdt/.
  - AICN's website is https://1.ieee802.org/nendica-aicn/.
- Both Nendica and task group hold teleconference meetings, which are without registration fees, and meet during bi-monthly sessions, which require registration fee. Participation and contributions from interested individuals are welcome and encouraged.
  - TSN's website is http://1.ieee802.org/tsn/.
  - Security's website is https://1.ieee802.org/security/. '
  - Nendica's website is https://1.ieee802.org/802-nendica/,