Low Latency Discussion for Ethernet Networking

Huajie Bao (Huawei)
Xiaomin Yang (Huawei)

Nov 2021, IEEE 802

Background

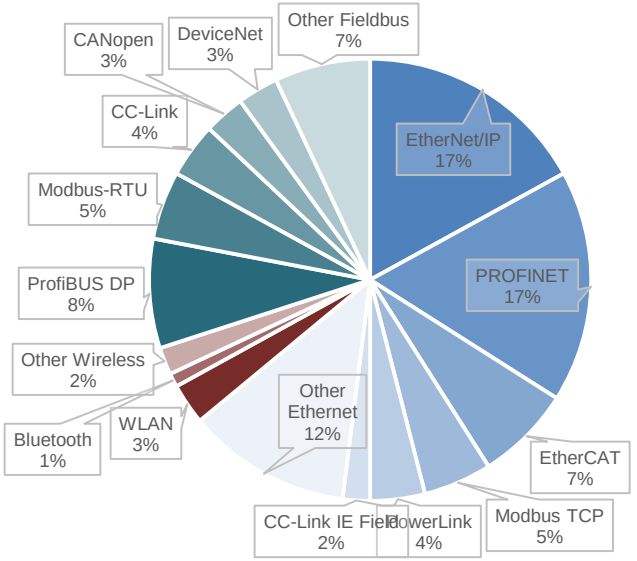
- Low latency has been discussed quite a lot in data center networks.
  - iLossless DCN whitepaper
  - PFC, sPFC
- However, low latency is not only required in DCN, industrial also has critical latency requirement in some scenarios.
  - Industrial has different environment compared with data center, including:
    - Topology
    - Bandwidth
    - Traffic pattern
    - Etc.
- The presentation intends to discuss
  - Industrial low latency requirement
  - Ethernet networking gaps to meet the requirements
  - And try to compare the possible aspects of low latency technology about the DCN and industrial

Connectivity is the One Word Summary for Industry 4.0 Revolution, but…

Industry 4.0 revolution can be summarized with one word: '**Connectivity**'. Connectivity will **enable intelligent production**, and smart devices can **collect various data** …, which will be **used for complex task coordination**, decision making, and remote access to machinery etc.

-- Source: Industrial IoT Challenges, Design Principles, Applications, and Security
Springer Nature Switzerland AG 2020

Fragmented Industrial Networking Capability

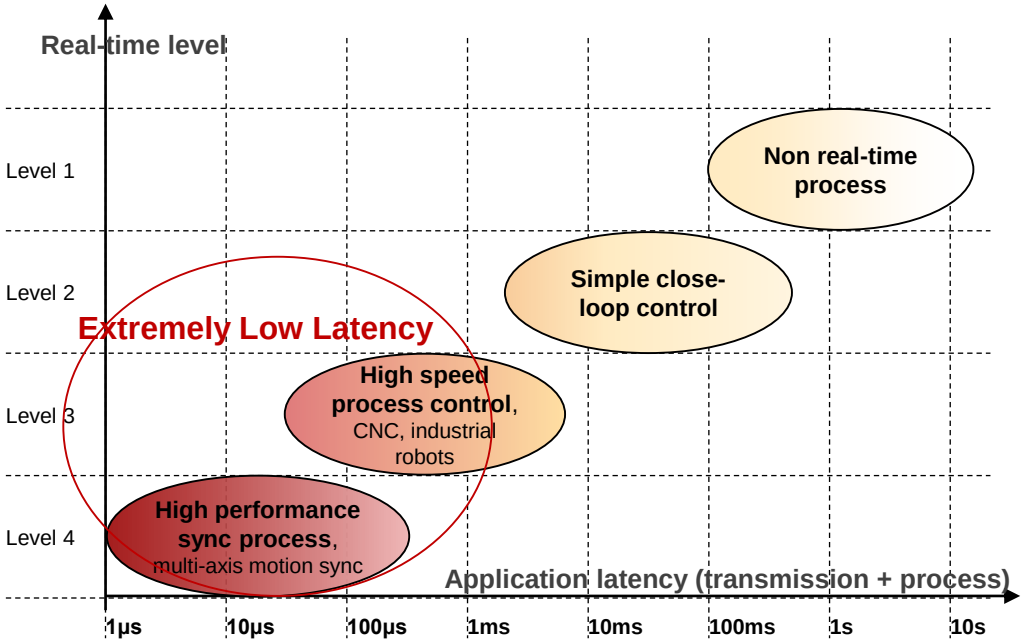


Source: Industrial network market shares 2020 according to HMS Networks

- **Convergence of the industrial networking is the trend.**
- **The Ethernet Networking is playing an important role in industrial domain.**

Low Latency Requirement in Industrial Scenarios

As increasing of control complexity and precision of the control system, the real-time performance is also increasing.



*Source: "Research on Collaborative Scheduling and Optimization in Reliable Real-time EPA Network" From China National Knowledge Infrastructure

The Industrial Ethernet Networking is required to pursue extremely low latency capability continuously.

Extremely Low Latency

- Smaller application cycle time: 31.25μs(1/32ms)
  - ➢ Corresponding transmission low latency
- Low jitter: 20ns (< 0.1% cycle time)

* Reference from ProfiNET, EtherCAT KPI.

The Extremely Low Latency is a challenge for Ethernet Networking in industrial scenarios.

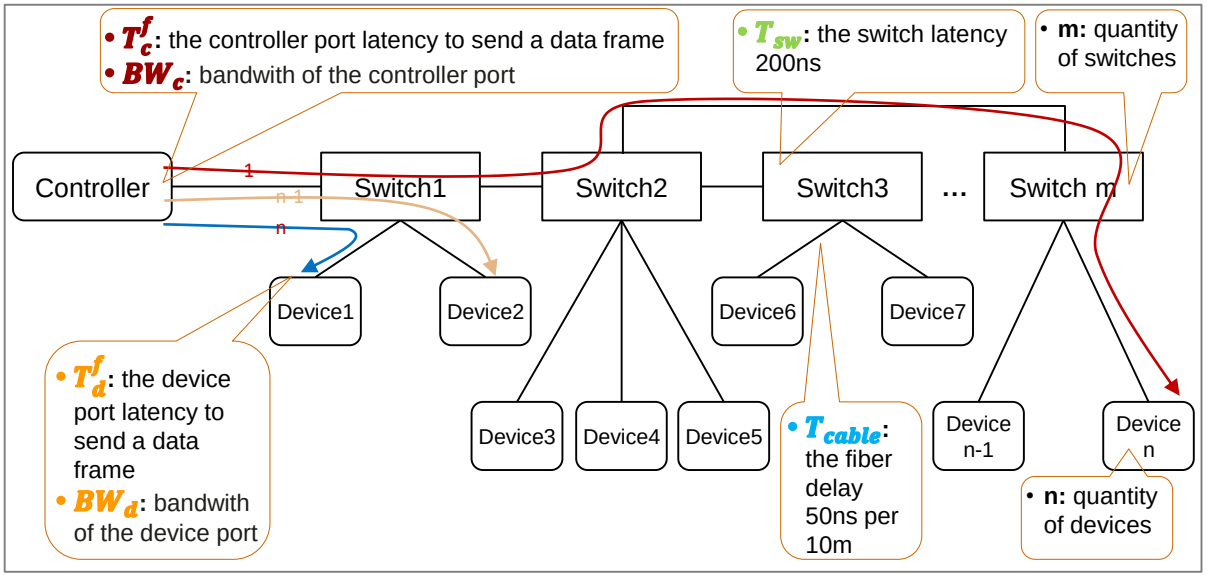
Latency Analysis for Industrial Ethernet Networking 1/2

Industrial application characteristics:

- Short frame message
- Frequently sending & receiving frames
- Message flow direction C➔D, D ➔ C
- Increasing networking scale-size

Assumptions:

- The controller sends frames to all devices sequentially, from the furthest one to the closest one. (**the controller sends frames to Device 1 at the n^{th}**)
- Assume that device 1 is the closest one and device n is the furthest one.
- Each industrial ethernet networking switch has 8 ports, **m =** $\frac{n}{8}$.

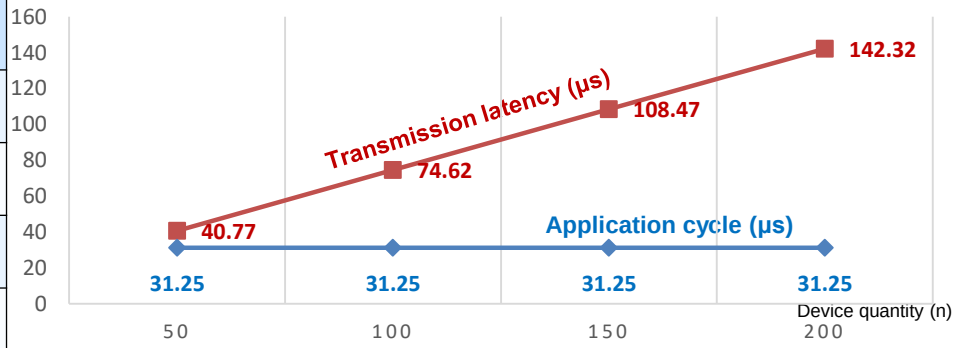


- $T_{latency}$ is the industrial transmission latency not including application latency.
- $T_{latency}$ = maximum [$T_{Controller-Device1}$, $T_{Controller-Device\ n}$] = maximum [nT_c^f + T_{sw} + T_{cable} +T_d^f, T_c^f + mT_{sw} + T_{cable} + T_d^f]
- $T_c^f = \frac{frame\ size}{BW_c}$, $T_d^f = \frac{frame\ size}{BW_d}$, frame size = 84Bytes(minimum frame length 64 Bytes, payload 8Bytes, preamble etc.)

Latency Analysis for Industrial Ethernet Networking 2/2

- Controller bandwidth 1Gbps，device bandwidth 100Mbps.
- The industrial ethernet networking transmission latency increases according to the networking scale-size / device quantity increasing.

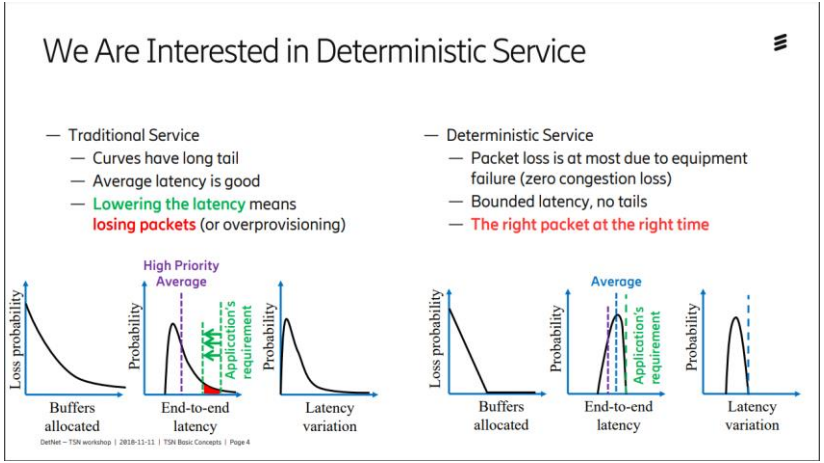
Device quantity (n)	Transmission latency (μs)
50	**40.77**
100	**74.62**
150	**108.47**
200	**142.32**



- In large-scale industrial ethernet network (e.g. steel industry, nuclear energy industry, might be 10+K devices), Industrial Ethernet Networking **transmission latency** **cannot meet** the **application cycle (including application processing time)** requirement.

TSN Focus on ‘Deterministic’ Instead of Absolute Low Latency

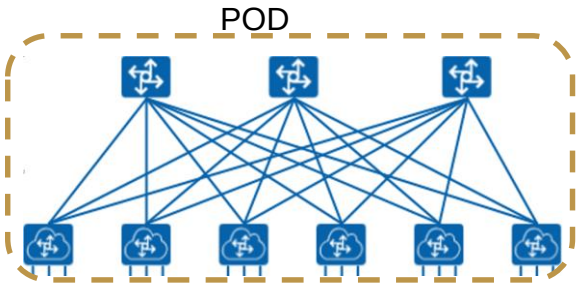
- The main purpose of TSN is to guarantee the boundary
  - ‘Bounded latency, no tails’ is more important, instead of pursuing absolute low latency
  - Technologies like shaping, scheduling etc. are used to achieve the target. ‘Time’ can be scarified in order to gain the ‘deterministic’.
    - For example, CQF



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

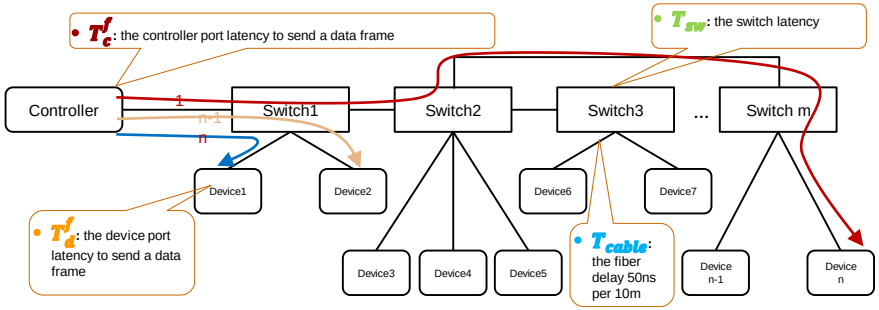
- However, there are scenarios which are sensitive to absolute low latency.
  - New technologies need to be considered to address low latency issues.
  - Maybe, DCN low latency technologies could be referred to?

Low Latency Ethernet Networking for DC and Industrial



Data Center

- **Dynamic latency** (mainly caused by congestion)
  - Congestion control, flow control, load balance etc.
- **Static latency** consisted of switch delay and cable delay
  - CTF
  - Topology-based forwarding (Sigcomm 2008: A Scalable, Commodity Data Center Network Architecture)



Industrial

- **Dynamic latency**
  - Congestion control, flow control, load balance ×
  - Traffic pattern based orchestration +
- **Static latency** consisted of switch delay and cable delay
  - CTF ✓
  - Topology-based forwarding **?**
  - Bandwidth & frame size ($T_c^f = \frac{frame\ size}{BW_c}$, $T_d^f = \frac{frame\ size}{BW_d}$) +

- Some technologies can be used in both scenarios.
- Other technologies specifically for Industrial scenarios needs to be considered.

Next Steps

- Further exploring low latency requirements in industrial and potential technologies
  - Anything in common with data center?
  - Any data center low latency Ethernet networking mindset/technologies can be leveraged
  - Any new technologies may help industrial case
- Consider to initiate a new study item/work item for further discussion.

Thank you