Outline of proposed new work item: “The Intelligent Lossless Data Center Network”

Liang Guo

Paul Congdon

Liyang Sun

IEEE 802 Nendica
Atlanta USA
March 2020

CONTRIBUTORS/SUPPORTERS

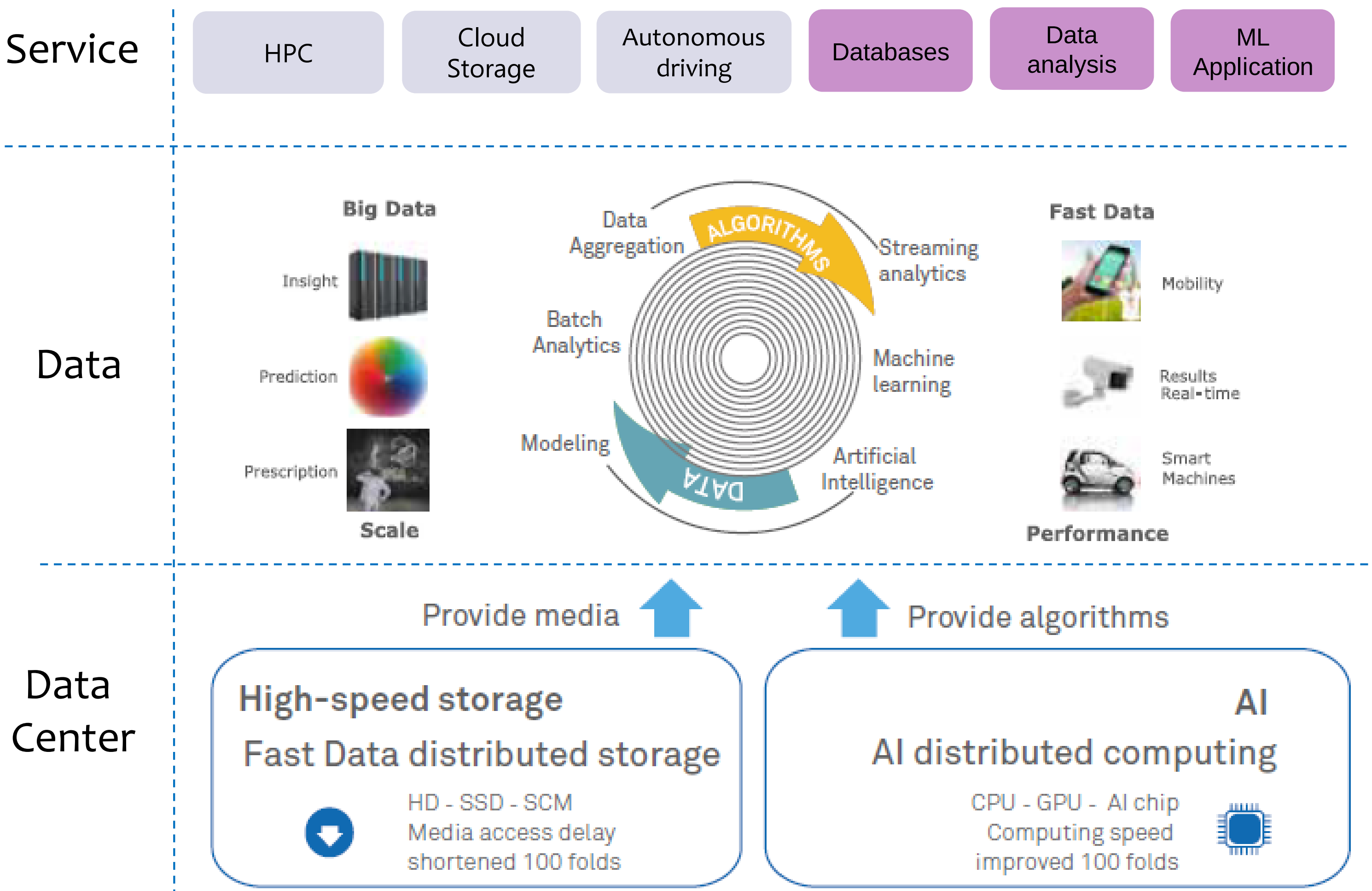
- Jie Li (CAICT/ODCC)
- Feng Gao (Baidu)
- Rong Gu (China Mobile)
- Jizhuang Zhao (China Telecom)
- Chuansheng Chen (Tencent)
- Guangming Tang (Meituan)
- Yue Yin (Huawei)
- Qingchun Song (Mellanox)
- Jun Liu (Cisco)
- Zongying He (Broadcom)
- Paul Congdon (Tallac)
- Bai Zhang (Huawei)

Background

- Nendica work item initiation at the March Plenary meeting
  - *Several discussions and contributions to the work item*
  - *A motion was proposed: "To forward 802.1-20-0002-02 (IEEE 802 Nendica Work Item Proposal: Revision of 'The Lossless Network for Data Centers') for March approval by 802.1."*
  - *Result: Approved, without objection at teleconference on 11th February*
- See:
  - https://1.ieee802.org/802-nendica/ieee-802-nendica-procedures/
  - https://mentor.ieee.org/802.1/dcn/20/1-20-0002-02-ICne.pptx
  - https://mentor.ieee.org/802.1/dcn/20/1-20-0014-00-ICne.docx
  - https://mentor.ieee.org/802.1/dcn/20/1-20-0016-00-ICne-draft-minutes-of-the-nendica-meeting-of-2020-02-11.docx
- This contribution proposes an outline of the future Nendica data center report

Bringing the Data Center To Life

Today's Data Center enables our digital real time world



Evolving data center technologies

- Cloud-based AI platforms
  - *Combines CPUs, storage and networking to simulate cognitive functions such as problem-solving, learning, reasoning, social intelligence.*
  - *Data Center resource planning and utilization are critical to success.*
- SmartNICs
  - *Improvements in network bandwidth exceed improvements in compute capacity, so various network functions have been offloaded onto network interface controller (NIC) hardware.*
  - *Two design choices: (a) fully programmable network processors, (b) FPGAs connected directly to the NIC ASIC over a high-speed interconnect.*
- Distributed Storage
  - *Storage performance needs to improve by an order of magnitude to achieve more than 1 million input/output operations per second (IOPS)* [1].
  - *Communication latency has recently increased from 10% to 60% of storage E2E latency* [2].
- Distributed Computing
  - *Computing speeds of Google's machine translation reaches 105 ExaFlops(10^{18})*[3].
  - *Using a traditional architecture, one AI training task can take half a year.*
  - *The waiting time for GPU communication exceeds 50% of the job completion time (JCT)*[4].

Communication Between Distributed Nodes Becomes a Bottleneck!

[1] Jim Handy, Thomas Coughlin. SNIA Survey: Users Share Their Storage Performance Needs. 2014 SNIA.
[2] AI, This Is the Intelligent and Lossless Data Center Network You Want. https://www.cio.com/article/3347337/ai-this-is-the-intelligent-and-lossless-data-center-network-you-want.html
[3] Ettikan Kandasamy Karuppiah. REAL WORLD PROBLEM SIMPLIFICATION USING DEEP LEARNING / AI.2017
[4] Omar Cardona. Towards Hyperscale High Performance Computing with RDMA. NANOG 76, 2019

Network for high performance applications

- Applications have impressive improvements by adopting RDMA
  - *Fast startup, maximizing the bandwidth usage.*
  - *One copy operation, effectively reducing the kernel latency.*
  - *Zero CPU resources consumed with network adapter offloading.*
- High I/O throughput with low-latency storage network
  - *As media access speeds increase, network latency becomes the bottleneck.*
  - *Storage interface protocols evolve from Serial Attached SCSI (SAS) to Non-Volatile Memory Express (NVMe).*
  - *Reducing dynamic latency (latency from queuing and packet loss) is key to reducing the NVMe over Fabric latency.*
- Ultra-low latency network for distributed computing [1]
  - *Important ultra-low latency applications include:*
    - *High-frequency trading*
    - *HPC/ AI Training*
  - *Controlling the tail latency of these applications is critical. It must be measured in microseconds, not milliseconds.*

[1] M. Alizadeh, A. Kabbani, T. Edsall, B. Prabhakar, A. Vahdat, and M. Yasuda. Less is more: Trading a little bandwidth for ultra-low latency in the data center. In Proc. USENIX NSDI, 2012..

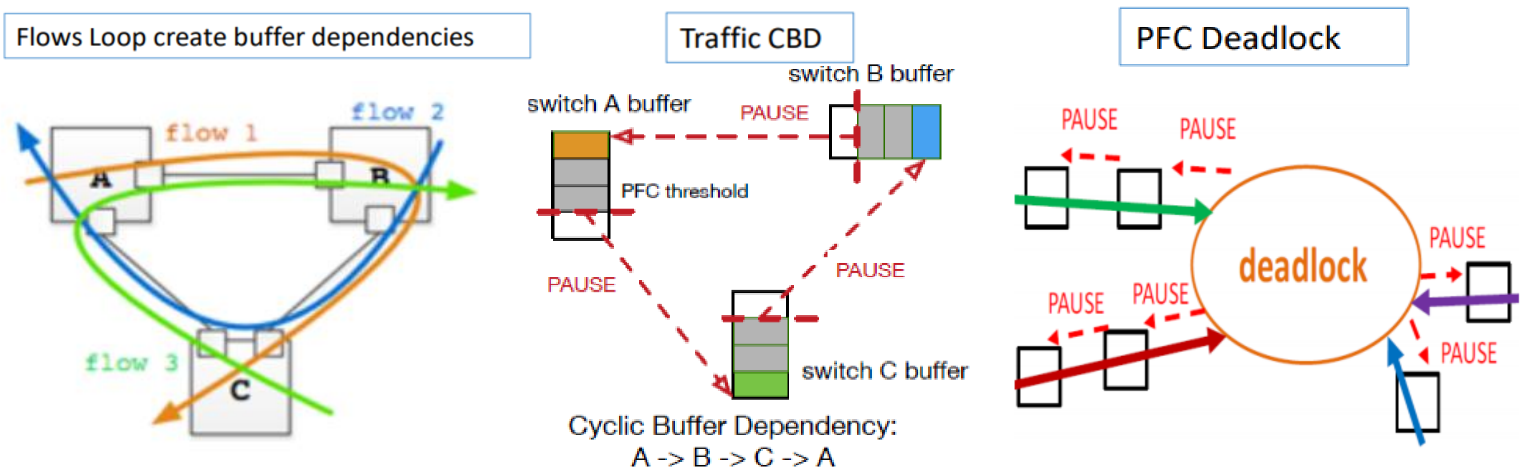
Challenges with today's network solutions

- Bandwidth vs. Latency tradeoff
  - *It's difficult to achieve high bandwidth and low latency simultaneously.*
  - *For consistently low latency, the network needs to maintain small queues (which means low ECN marking thresholds), while high bandwidth benefits from larger queues and higher ECN thresholds.*
  - *Experimentation shows the tradeoff still exists after varying algorithms, parameters, traffic patterns and link loads* [1]
- Congestion Control issues in large-scale networks
  - *High-performance RDMA requires zero packet loss. Zero packet loss Ethernet requires Priority-based Flow Control (PFC). PFC, however, makes the network prone to deadlocks* [2,3].
  - *QoS cannot be guaranteed when TCP and RDMA over Converged Ethernet (RoCE) flows coexist in a network using Smart Buffering.*
  - *Tuning RDMA networks is an important factor to achieving high-performance, but it can be a complex operation.*
- Some other problems?

[1] Yuliang Li, Rui Miao, Hongqiang Harry Liu, et al. 2019. HPCC: High Precision Congestion Control.
[2] Chuanxiong Guo, Haitao Wu, Zhong Deng, et,al. RDMA over Commodity Ethernet at Scale. SIGCOMM, 2016.
[3] RoCE v2 Considerations. https://community.mellanox.com/s/article/roce-v2-considerations#jive_content_id_If_I_run_RoCE_v2_should_I_use_PFC_or_global_pause_for_lossless_L2_subnet

Challenges with today's network solutions

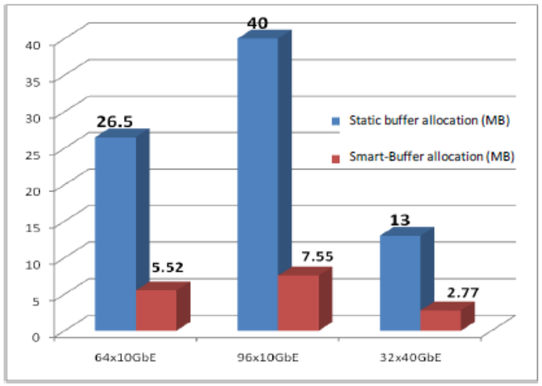
- PFC can cause a severe deadlock in the data center network
  - *How does PFC deadlock form?*
    - *Cyclic Buffer Dependency (CBD) is a necessary condition for deadlock formation.*
    - *Multiple flows in a loop is a necessary condition for CBD.*
  - *Example deadlock problem in a CLOS network*
    - *Reproduce the PFC deadlock in both level 2 CLOS and level 3 CLOS network.*
    - *Although a CLOS network does not have routing loops, when a link fails, a flow loop can happen, and CBD appears. When CBD appears, PFC deadlocks may happen.*



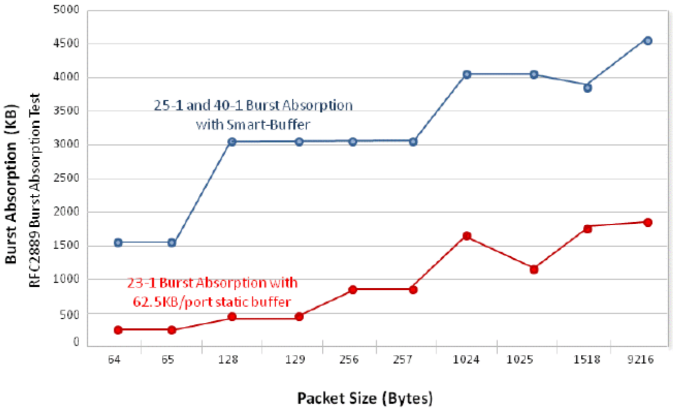
[1] http://www.ieee802.org/1/files/public/docs2019/new-yu-pdf-0919-v00.pdf

Challenges with today's network solutions

- Smart-buffer mechanisms in mainstream switch chips[1]
  - *Switch packet buffer performance and cost tradeoff*
    - *To prevent packet loss caused by microbursts, each queue on each port of the switch should be configured with enough buffers to absorb the burst.*
    - *The cost is too high for a switch to implement purely static per-port buffer allocation schemes.*
  - *Smart-buffer mechanisms attempt to optimize buffer utilization and burst absorption*
    - *Dynamic sharing and self-tuning is transparently enabled across all ports.*
    - *Optimized for specific traffic scenarios to maximize overall throughput and lossless behavior.*



Smart-buffer delivers up to five times better packet buffer utilization

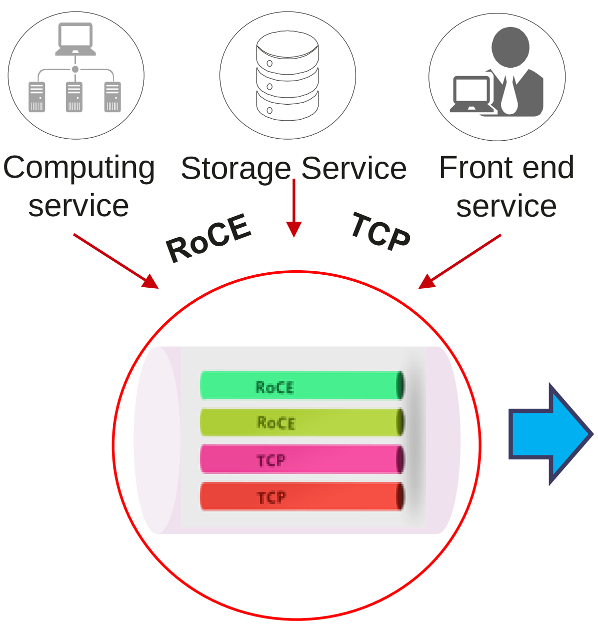


Burst absorption capacity is 3 to 6 times better than per-port static buffer architecture

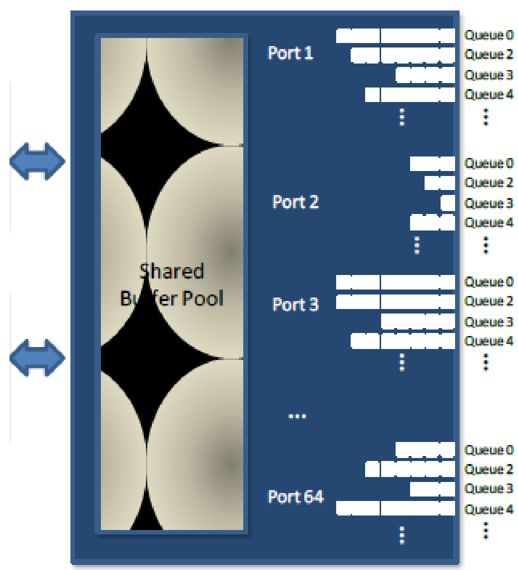
[1] Sujal Das, Rochan Sankar. 2012. Broadcom smart-buffer technology in data center switches for cost-effective performance scaling of cloud applications.

Challenges with today's network solutions

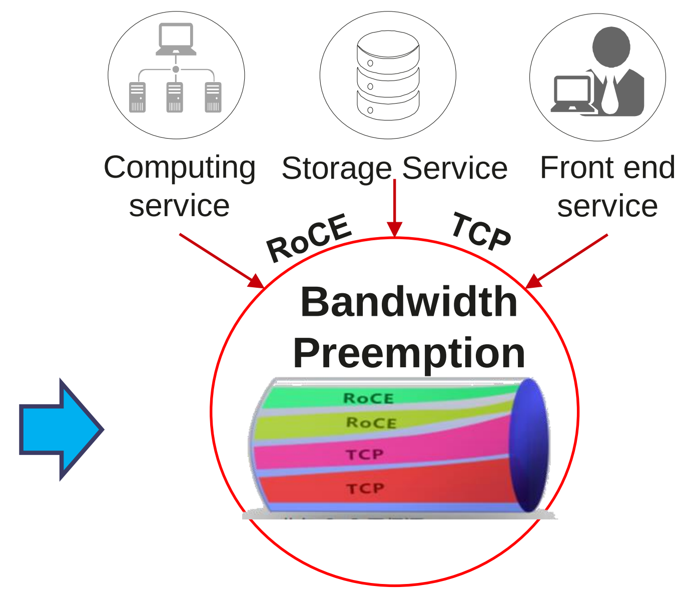
- SLAs cannot be guarantee when TCP and RoCE traffic coexists
  - *TCP and RoCE traffic proportions can have an unexpected mix*
    - *TCP and RoCE have different congestion control mechanisms and TCP is more aggressive.*
    - *The ratio of TCP to RoCE can vary from initial settings due to smart shared buffering.*
    - *TCP flows can preempt the bandwidth of RoCE flows, even when using separate traffic classes.*
    - *Other popular transport protocols, like QUIC, have been shown to NOT mix fairly with TCP.*



Initialization phase: bandwidth is allocated to different flows based on service requirements



Shared buffer architecture

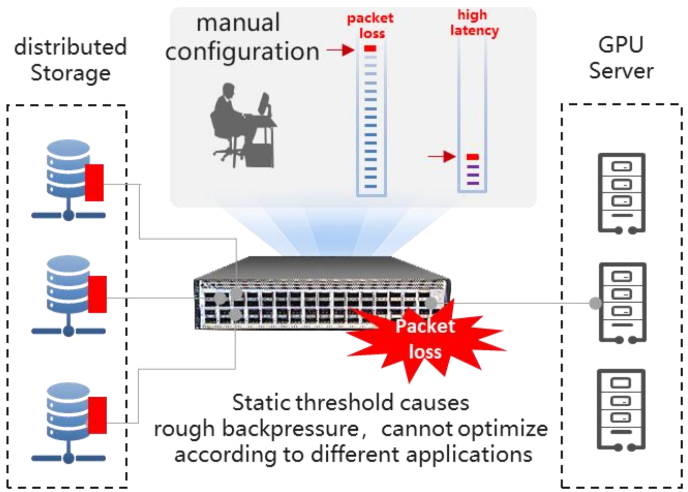


TCP preempts the RoCE bandwidth
The RoCE flow delay increases by 100 times (ms level)

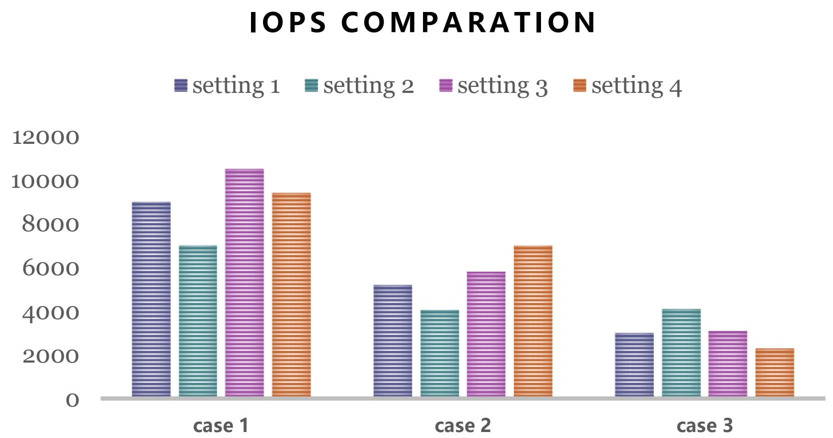
[1] Paul Congdon. Towards Hyperscale HPC & RDMA. IETF-104 HotRFC. https://mentor.ieee.org/802.1/dcn/19/1-19-0035-00-ICne-ietf-104-dcn-hyperscale-rdma-hpc-report.pdf

Challenges with today's network solutions

- Operational complexity of congestion control algorithm configuration
  - *Congestion control algorithms usually requires collaboration between the NIC and switch*
    - *Each node needs to be configured with dozens of parameters, and the parameter combination of the entire network can reach in the hundreds of thousands.*
    - *Historically, parameters are configured manually according to experience:*
      - *Difficult to adapt to real-time network traffic and workload changes.*
      - *Expensive operation that can result in low throughput and high latency.*
    - *Static configuration cannot ensure the optimal performance of most service scenarios in the customer environment*



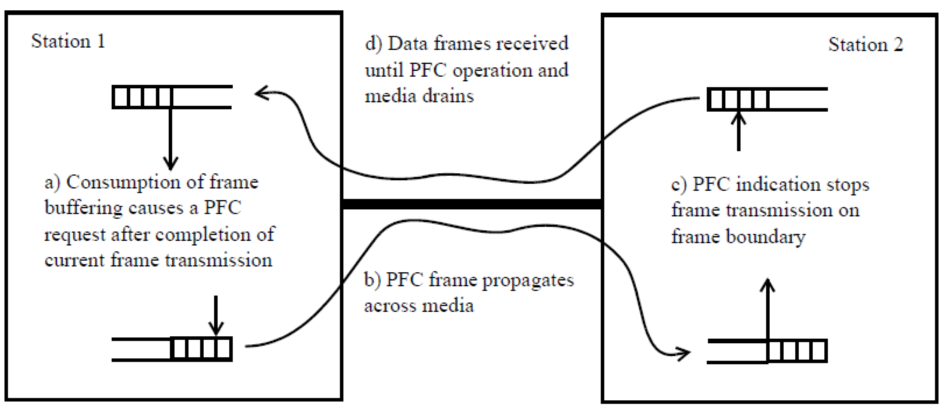
Packet loss occurs due to incast
idle waiting time of the GPU >50%

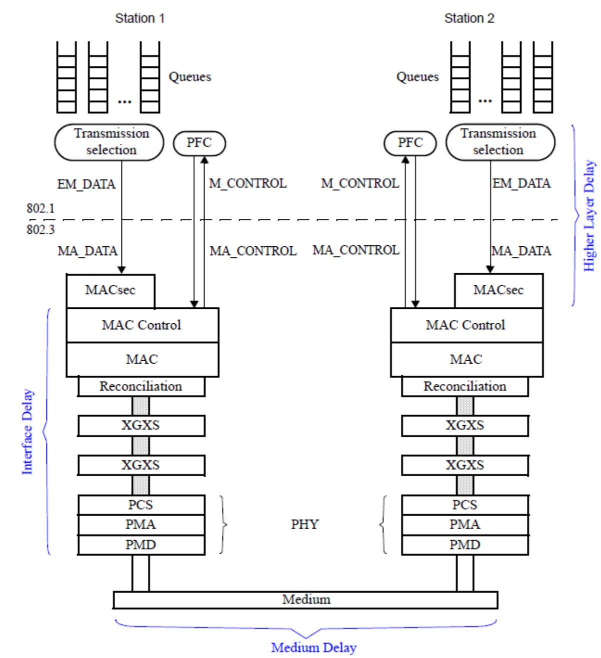


Lab tests show that different congestion control algorithms produce different effects in the same application scenario

Challenges with today's network solutions

- Complexity of PFC headroom configuration
  - *RoCE needs PFC mechanism to achieve lossless Ethernet*
    - *Each lossless queue needs to be configured with enough headroom buffer [1]*
    - *Historically done by manual configuration; Complex calculation with lots of parameters (Buffer structure and unit size, switching delay, cable delay and interface delay) [2]*
    - *Excessive headroom leads to reduced number of lossless queues; too little headroom leads to packet loss [3].*





1. Configuration Guide - Low Latency Network, https://support.huawei.com/enterprise/en/doc/EDOC1100040243/c28a82e4/buffer-optimization-of-lossless-queues
2. 802.1Qbb-2011 Amendment 17: Priority-based Flow Control, https://standards.ieee.org/standard/802_1Qbb-2011.html
3. C, Guo et all. RDMA over Commodity Ethernet at Scale, https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/rdma_sigcomm2016.pdf

Technical considerations to address some of today's problems

- Approaches to PFC storm elimination
  - *Deadlock detection*
  - *Deadlock elimination*
- Improving Congestion Notification
  - *Improved Explicit Congestion Notification*
  - *Enhanced version of Quantized Congestion Notification (originally IEEE 802.1Qau)*
  - *Methods of improving QoS support in mixed traffic environments.*
- Congestion parameter optimization
  - *Heuristic algorithms for identifying congestion parameters*
  - *Methods for dynamic optimization based on services*
- Buffer Optimization of Lossless Queues
  - *Self-adaptive headroom configuration*

Some other contents

- Standardization considerations
  - *related standardization work*
  - *potential upcoming standardization work*
- Conclusions
  - *Summary to Nendica report*
  - *Some new technical development discussion*
- Citations

References

- IEEE 802 Nendica Procedures
  - https://1.ieee802.org/802-nendica/ieee-802-nendica-procedures
- IEEE 802 Nendica ICAID (March 2019 – March 2021)
  - https://standards.ieee.org/content/dam/ieee-standards/standards/web/governance/iccom/IC17-001-IE.pdf
- Nendica Work Item: Lossless Data Center Networks [LLDCN]
  - https://1.ieee802.org/802-nendica/nendica-lldcn
- IEEE 802 Nendica Report: The Lossless Network for Data Centers
  - https://mentor.ieee.org/802.1/dcn/18/1-18-0042-00-ICne.pdf
- IEEE 802/IETF Data Center Workshop – Bangkok, 2018-11-10
  - https://1.ieee802.org/802-nendica/802-ietf-workshop-data-center-bangkok
- IETF Side meeting: Data Center Congestion Control – Where's the best fit in IETF/IRTF?
  - https://datatracker.ietf.org/doc/draft-zhuang-tsvwg-ai-ecn-for-dcn/
  - https://mentor.ieee.org/802.1/dcn/19/1-19-0087-00-ICne-ietf-106-sidemeeting.pdf
  - https://datatracker.ietf.org/doc/draft-zhuang-tsvwg-open-cc-architecture/
  - https://mentor.ieee.org/802.1/dcn/19/1-19-0087-00-ICne-ietf-106-sidemeeting.pdf