High Performance Computing Ethernet

Yolanda Yu (Xiang)

yolanda.yu@huawei.com

NENDICA, Waikoloa, Hawaii November, 2019

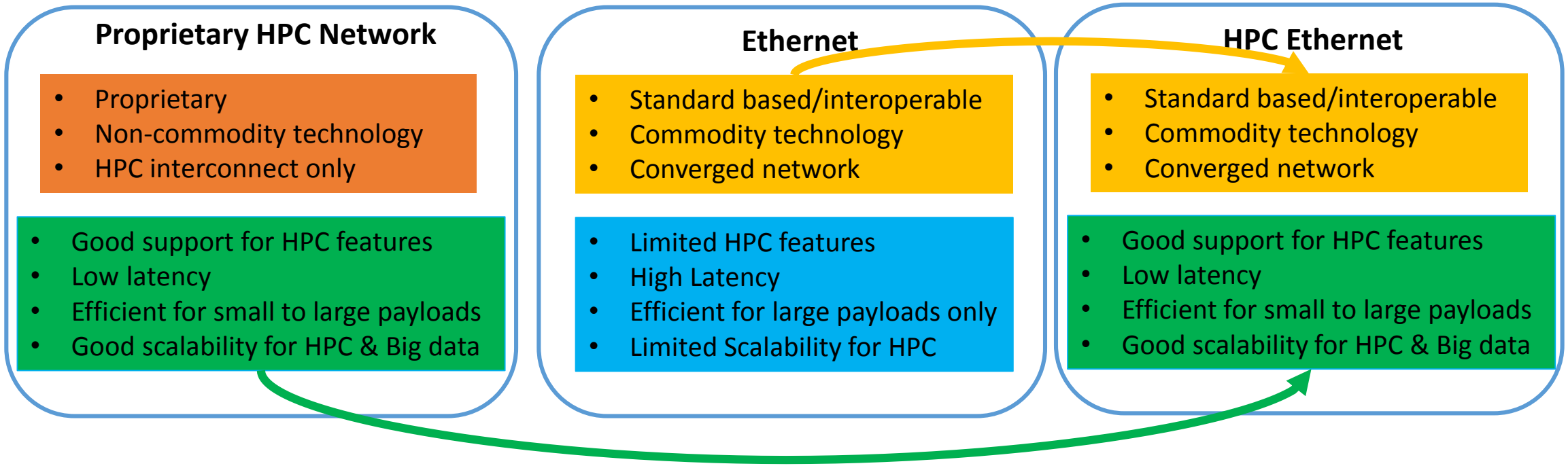
The New Demand for Converged Network

- Currently, HPC Data Center(DC) customers may maintain 2 different networks:
  - ✓ Proprietary network for HPC (For example Infiniband)
  - ✓ Ethernet for front end, storage, management, etc
  - ✓ There is a very strong voice to converge the networks they have to reduce CAPEX/OPEX
- Concurrent supercomputing with different scales (intra/inter-organizational communication) WANT TO utilize universal Ethernet bearers:
  - ✓ Across different buildings within one campus or different campus (less than 100 KMs)
  - ✓ Across different cities (hundreds KMs)
  - ✓ Across the globe (trans-continental)
- The emerging technology of lossless network make converged network possible within one POD[1] or one data center(short-distance)
- Most of the proprietary network protocols are designed for short distance transport. For example, the congestion control mechanism design architecture.

How to support emerging HPC resources collaboration with converged universal network? For example: How to realize the long-distance lossless network ? Or else?

1 Point Of Delivery, "a module of network, compute, storage, and application components that work together to deliver networking services. The PoD is a repeatable design pattern, and its components maximize the modularity, scalability, and manageability of data centers."

Improve Ethernet to support HPC features



- Current HPC network is based on proprietary Link Layer. HPC Ethernet realize the convergence of HPC and Ethernet.
- RDMA over Ethernet (For example, RoCEv2) enhanced the transport layer(for example the congestion control mechanisms) and use as-is Ethernet.
- HPC Ethernet will enhance Ethernet to support HPC features.

* The original picture is from https://www.nextplatform.com/2019/08/16/how-cray-makes-ethernet-suited-for-hpc-and-ai-with-slingshot/

The Value of HPC Ethernet

- Realize the converged network in HPC data centers
- Realize Ultra-low latency based on the low latency link layer, simple protocol and encapsulation
- Support Cloud Native protocol stack and open ecosystem
- Enable HPC Cloudification, HPC support multi-tenant and IP bearer
- Make Cloud DC/MEC[1]/IoT[2] support HPC

1 Mobile Edge Computing, "is an ETSI-defined network architecture concept that enables cloud computing capabilities and an IT service environment at the edge of the cellular network and, more in general at the edge of any network.“

2 The Internet of Things, “is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.”

The Challenge for HPC Ethernet

- Ultra-low latency link layer: low latency FEC[1], etc
- Simple Encapsulation: reduce the packet overhead
- Low latency switching: Tableless switching, in-flight preemption, intelligent cut-through to reduce the latency
- Multi-path: adaptive routing (Link layer multipath)

Call for Interest

- Collect more requirements or scenario to validate that HPC Ethernet is the trend
- Contribute on the HPC Ethernet techniques
- Cooperate on the standardization work in the future

Thanks