HPCC++: Enhanced High Precision Congestion Control

draft-miao-ccwg-hpcc

draft-miao-ccwg-hpcc-info

Rui Miao, Rong Pan, Jeongkeun Lee, Barak Gafni, Yuval Shpigelman, Jeff Tantsura, Guy Caspary, Surendra Anubolu, Allister Alemania, Jiaqi Gao

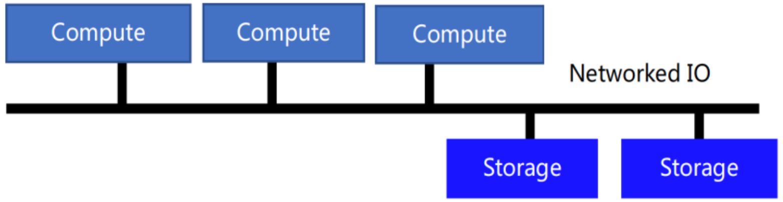
IETF-117 ccwg

July 2023

Cloud desires hyper-speed networking

Today, clouds have
- bigger data to compute & store
- faster compute & storage devices
- more types of compute and storage resources

High-performance storage

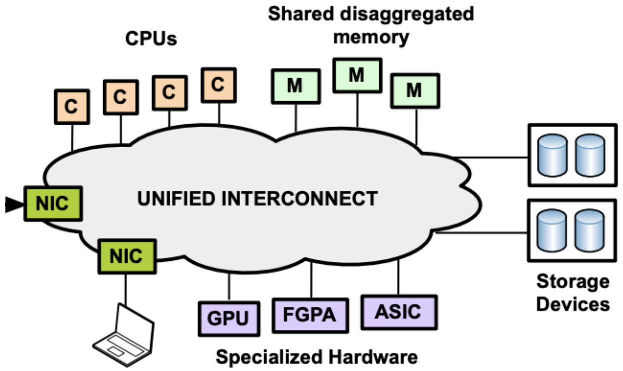


- Storage-compute separation is norm
- HDD→SSD→NVMe
- Higher-throughput, lower latency
- 1M IOPS / 50~100us

High-performance computation

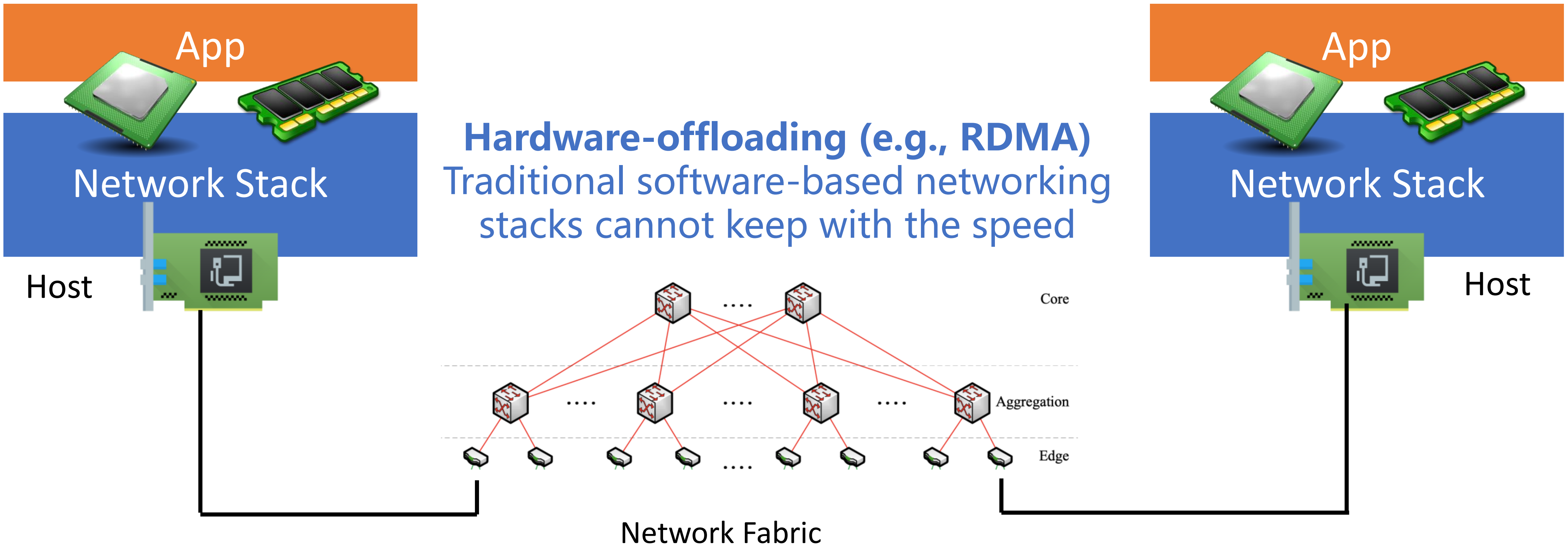
- Distributed deep learning, HPC
- CPU→GPU, FPGA, ASIC
- Faster compute, lower latency
- E.g. latency <10us

Resource disaggregation



- More network load
- Need ultra-low latency: 3-5us, > 40Gbps (Gao Et.al. OSDI'16)

Hyper-speed network chips to form hyper-speed networking



Hardware-offloading (e.g., RDMA)
Traditional software-based networking stacks cannot keep with the speed

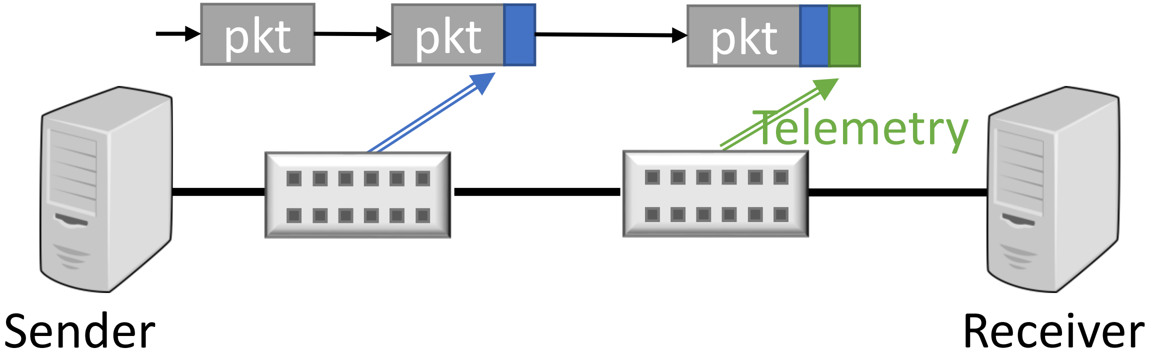
Real-time Congestion Control (CC)
Lots of data and communication => more pressure on the network

Challenges in some CC suites in high-speed networks

- Convergence upon congestion
- Running multiple applications over converged network
  - Queues and buffers are scarce resources
- Parameter tuning

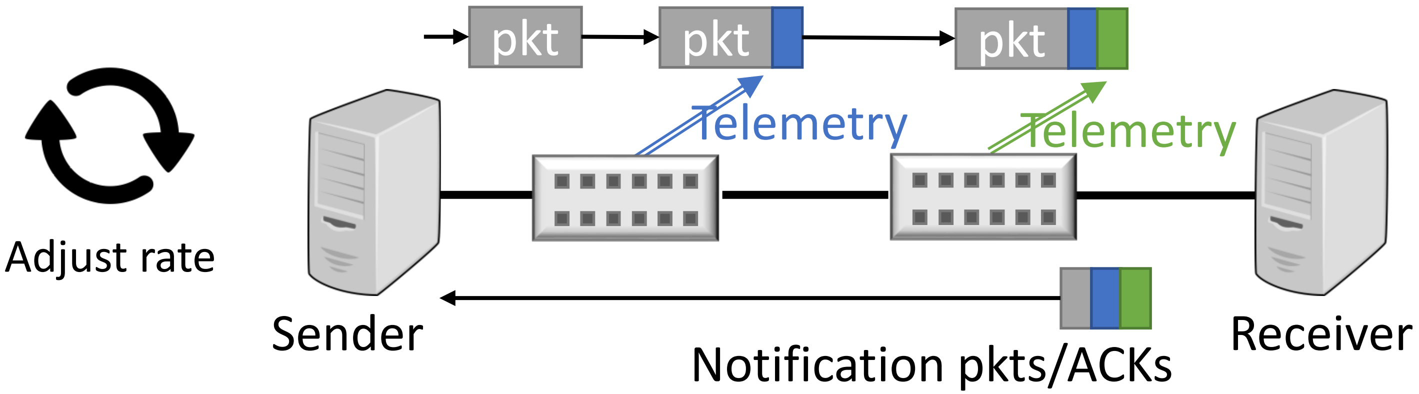
In-band Telemetry

- New networking ASICs have in-band telemetry capabilities
- Packets can collect telemetry on their route
- Various efforts to define inband telemetry
  - IETF IOAM
  - INT/P4.org
  - IFA



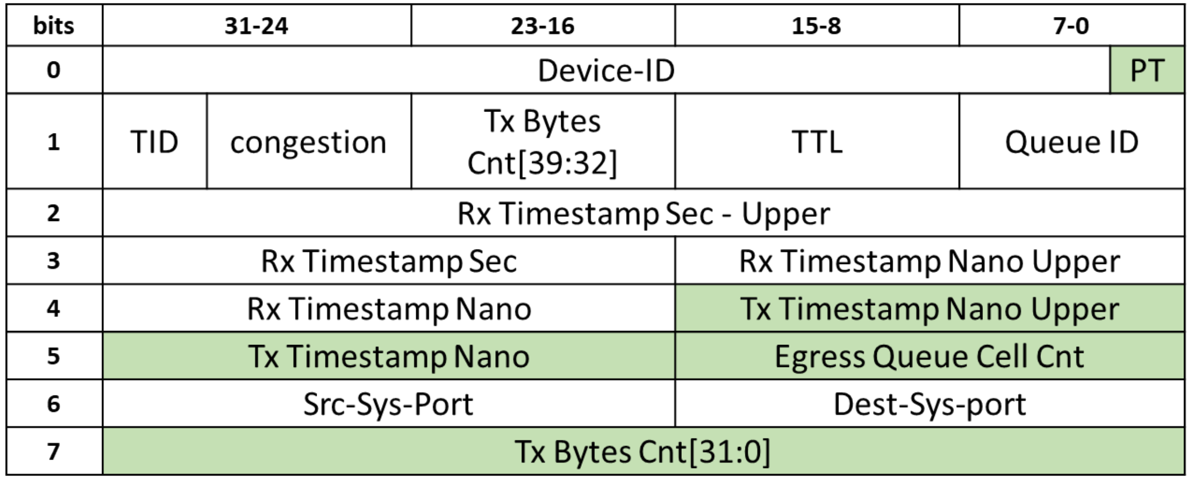
HPCC++: Enhanced High Precision Congestion Control

Can we use inband telemetry as more precise/richer feedback for congestion control?



In-band telemetry format

- HPCC++ defines the algorithm of using telemetry information
  - including queue length, transmitted bytes, timestamp, link capacity, etc.
  - draft-miao-ccwg-hpcc
- Yet, packet format is up to the environment
  - draft-miao-ccwg-hpcc-info provides examples of different telemetry encodings



bits	31-24		23-16	15-8	7-0
0	Device-ID				PT
1	TID	congestion	Tx Bytes Cnt[39:32]	TTL	Queue ID
2	Rx Timestamp Sec - Upper				
3	Rx Timestamp Sec			Rx Timestamp Nano Upper	
4	Rx Timestamp Nano			Tx Timestamp Nano Upper	
5	Tx Timestamp Nano			Egress Queue Cell Cnt	
6	Src-Sys-Port			Dest-Sys-port	
7	Tx Bytes Cnt[31:0]				

Example format of in-band telemetry used by HPCC++

HPCC++ Addresses all the discussed challenges

Using in-band telemetry as the precise feedback enables

- Faster convergence
  - ➢ Sender knows the precise rate to adjust to
- Near-zero queue
  - ➢ Feedback does not only rely on queue
- Fewer parameters
  - ➢ Rich and precise feedback, reduces heuristics which requires more parameters

So, What HPCC++ Actually Is?

- **It is a service**
- This service can be utilized by a given transport
- This service can also be utilized by a routing engine

Additional work

- Multi-queue considerations
- Consider additional receiver feedback
- Extend on encoding examples

Your Feedback is Appreciated!

Thank You