1. Cloud desires hyper-speed networking

Today, clouds have:
- Bigger data to compute & store
- Faster compute & storage devices
- More types of compute/storage resources

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

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

Resource disaggregation
- More network load
- Need ultra-lower latency: 3-5us, > 40Gbps
- (Gao et.al. OSDI’16)

2. Challenges in current CC in RDMA networks

Operation challenge-1: Threat from PFC
- Buffer overflow happens during incast or failures
- PFC is used to prevent packet loss → PFC storm & deadlock
  - Disabling PFC causes bad performance!!!

Operation challenge-2: running multiple applications
- Bandwidth intensive applications need large in-network queues
- Latency sensitive applications need small in-network queues
  - QoS queues are scarce resources!!!

Operation challenge-3: complex parameter tuning
- DCQCN has at least 15 parameters to tune

3. HPCC: High Precision Congestion Control

New commodity ASICs have in-band network telemetry (INT) ability
Use INT as precise feedback for congestion control (CC)

- Fast convergence
  - Sender knows the precise rate to adjust to, on every ACK
  - Near-zero queue
  - Feedback does not rely on queue
  - Few parameters

Precise feedback, so no need for heuristics which requires many parameters

4. Experimental study with real production traffic

- Challenging for CC
  - Under 80% traffic load
  - Extremely light-tailed
    ▶ 22.9% of queries are 4KB
    ▶ 90.5% of queries are <= 32KB

- HPCC reduces QCT for short queries
  - 4KB queries: 96% reduction in TCP and 81% reduction in RDMA
  - HPCC has small queues