QCN Hardware Evaluation

Abdul Kader Kabbani, Berk Atikoglu,
Jianying Luo, Balaji Prabhakar
(Stanford University)

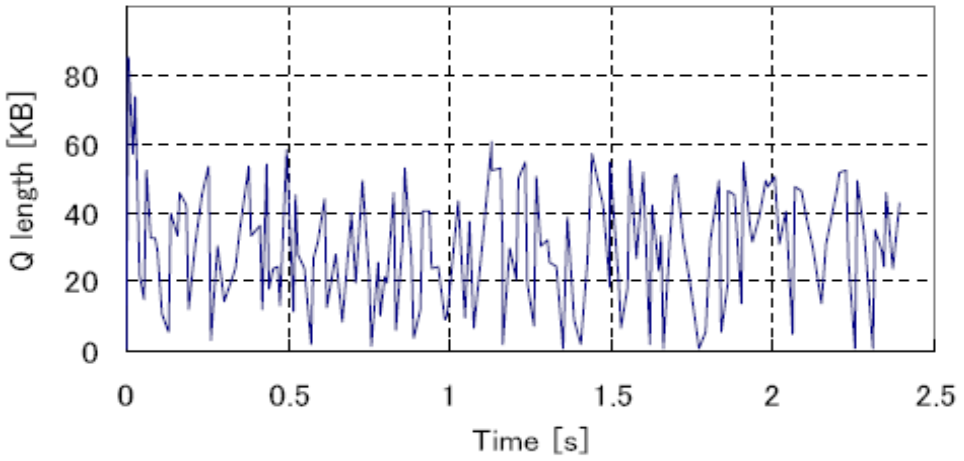
Masato Yasuda (NEC)

Overview

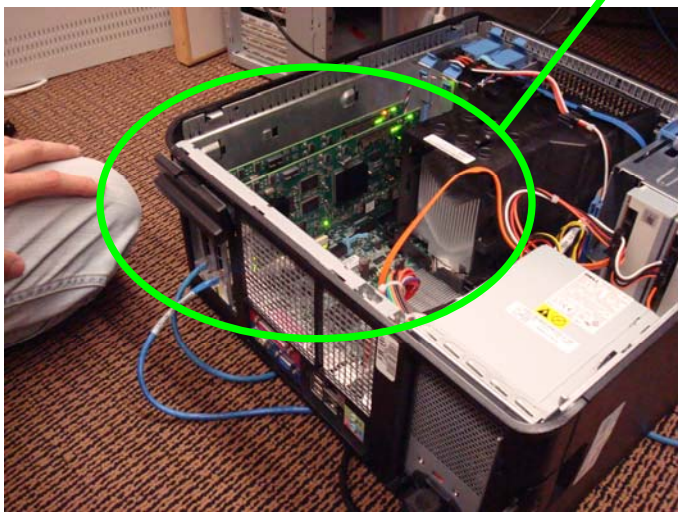
- Summary of the New Orleans implementation results
- QCN setup in a a real network with
  - Up to 8 RPs on 2 QCN NICs (implemented at Stanford)
  - 1 QCN switch (implemented at Stanford)
- Hardware comparison with Omnet++ results
  - Throughput
  - Queue length
  - Other details
- Live demo

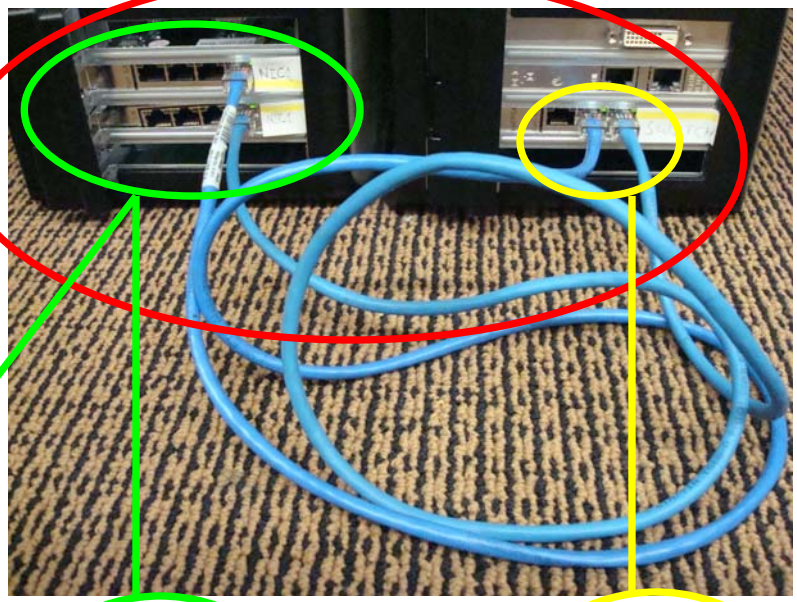
Previous Implementation

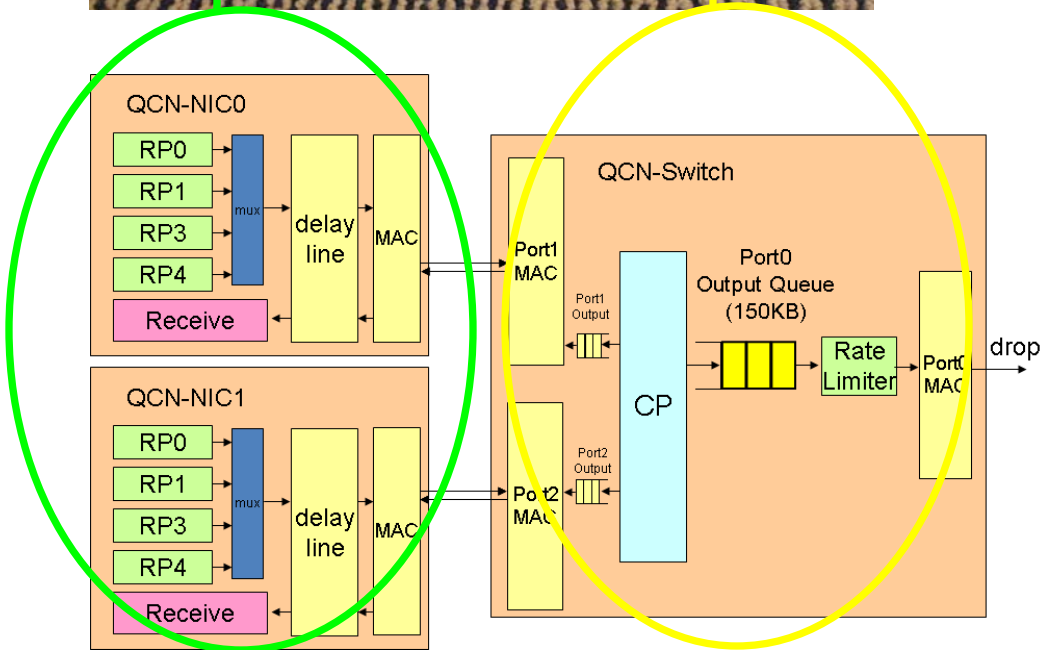
- Results presented in January by Nobuharu Kami (NEC)
- A Single NIC (1 RP) connected to a 10Gbps switch with 300ns latency
  - Queue length sampled every 1KB and sent to the NIC, which computes its own Fb value and adjusts its rate accordingly.
  - Queues are wigglier than expected when 1Gbps OG hotspot occurs:



Hardware Setup

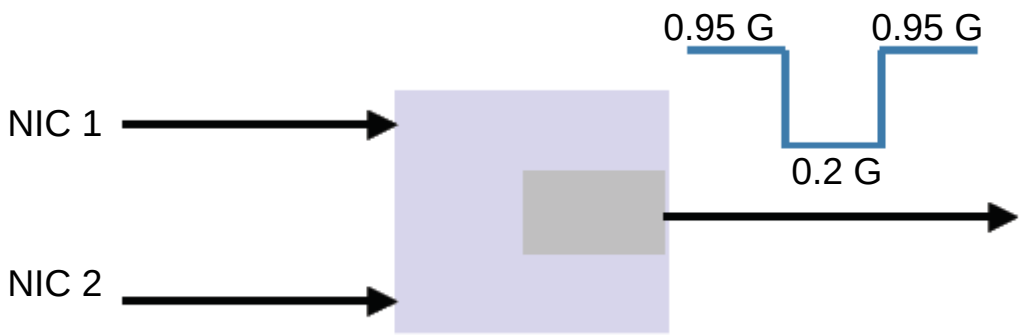






Experiment & Simulation Parameters

- Consider the Baseline Scenario
  - Single output queue
  - Vary the number of active RPs: 1 to 8
  - OG hotspot; hotspot severity: 0.2Gbps, hotspot duration ~3.5sec
  - Vary RTT: 100us to 1000us
  - Compare the stability and response time with that of Omnet++



Heads up! 1G – 10G Differences

- 1G can tolerate larger RTTs than 10G. Roughly speaking, 1G and 10G scenarios with comparable BW-Delay product have similar stability margins.
- Since per-flow BW share is an order-of-magnitude lower with 1G, byte counting goes slower in this case. The timer should be increased to make sure enough pkts have been sent before it expires excessively.
- 10G AI and HAI increments are too aggressive in the case of 1G and need to be decreased therefore.
- For the above two reasons, bandwidth recovery will take longer in 1G.

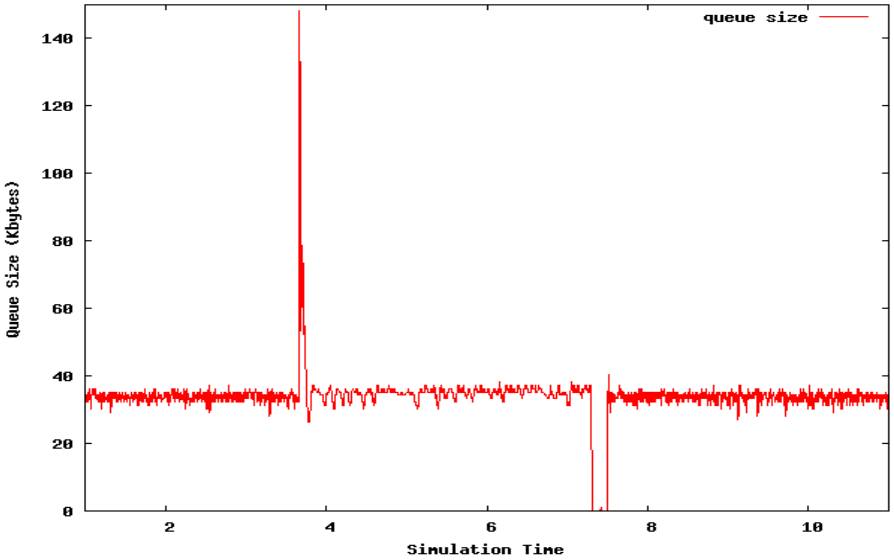
QCN Parameters

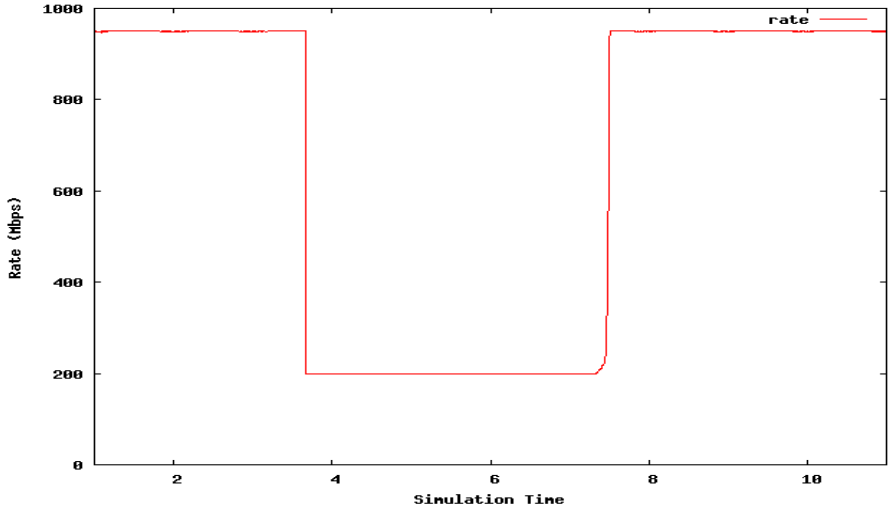
- NIC
  - FAST_RECOVERY_THRESHOLD = 5
  - AI_INC = 0.5 Mbps
  - HAI_INC = 5 Mbps
  - BC_LIMIT = 150 KB (30% randomness)
  - TIMER_PERIOD = 25 ms (30% randomness)
  - MIN_RATE = 0.5 Mbps
  - GD = 1/128
- Switch
  - Quantized_Fb: 6 bits
  - Q_EQ = 33 KB
  - W = 2
  - Base marking = 150 KB, and varies according to the lookup table in the pseudo code (30% randomness)

QCN Evaluation Result

1 source, RTT = 100us

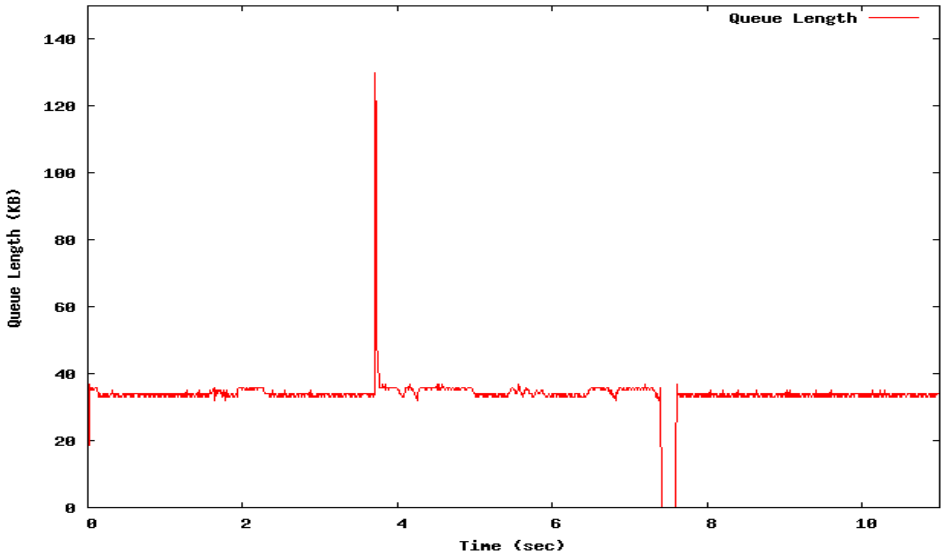
Hardware

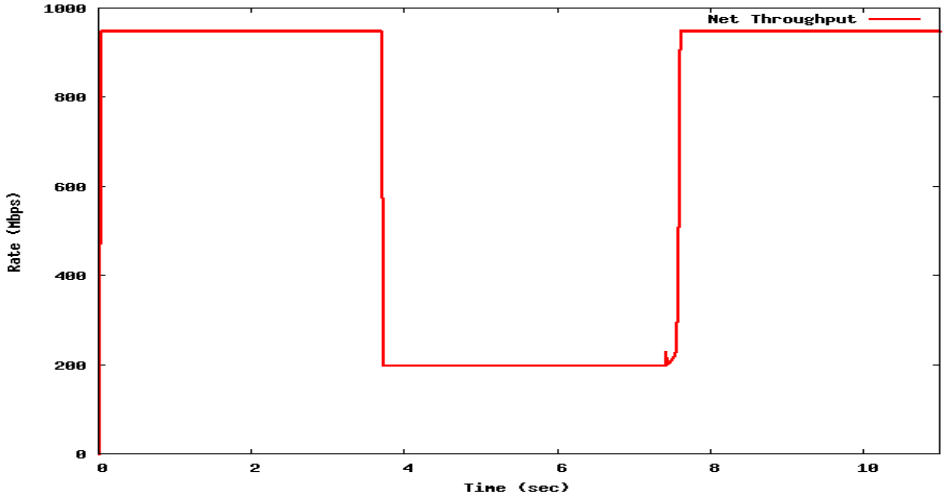




Recovery time = 179ms

OMNET++

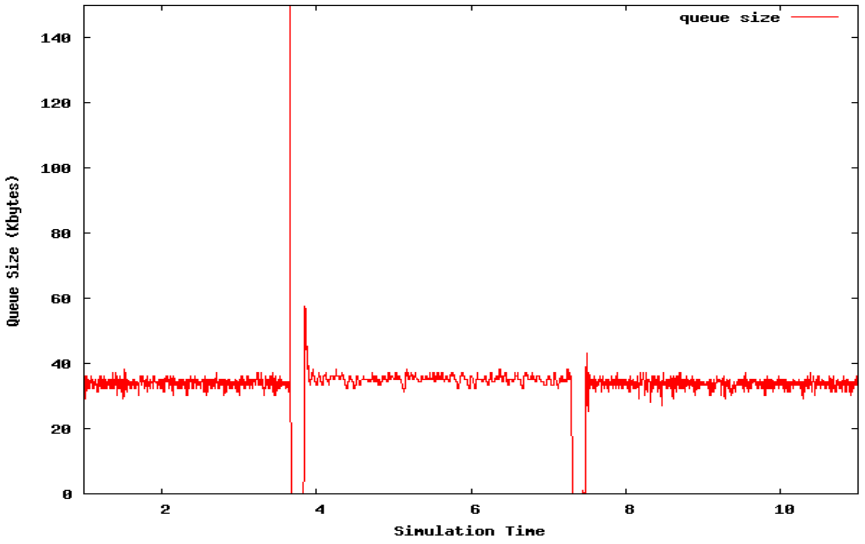


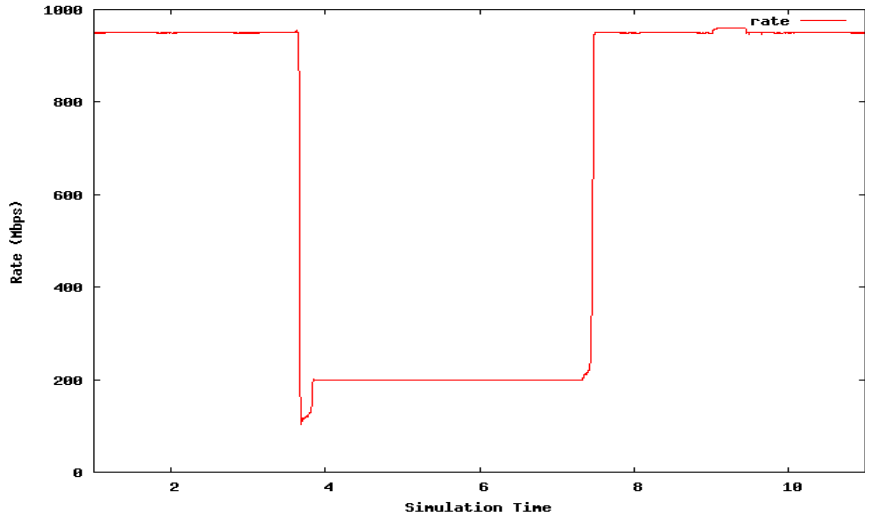


Recovery time = 200ms (80ms in 10G previous talks)

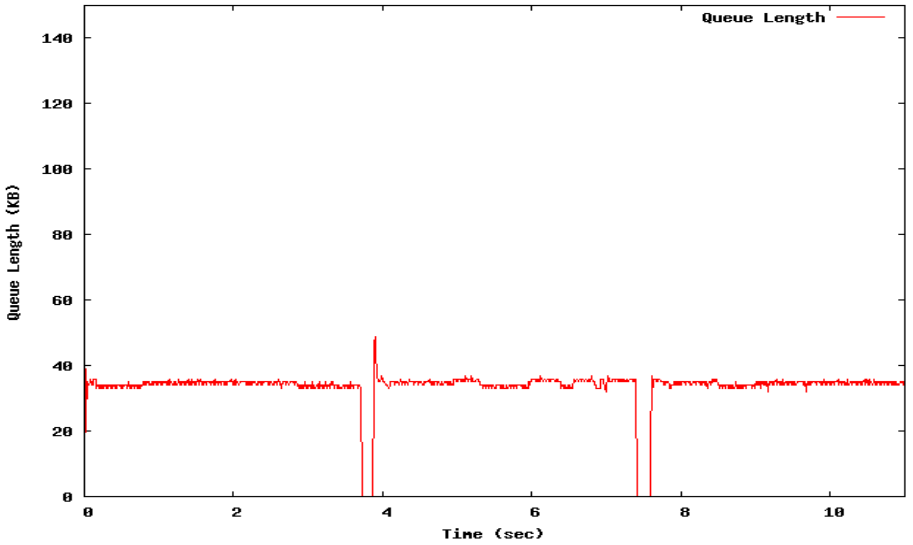
1 source, RTT = 500us

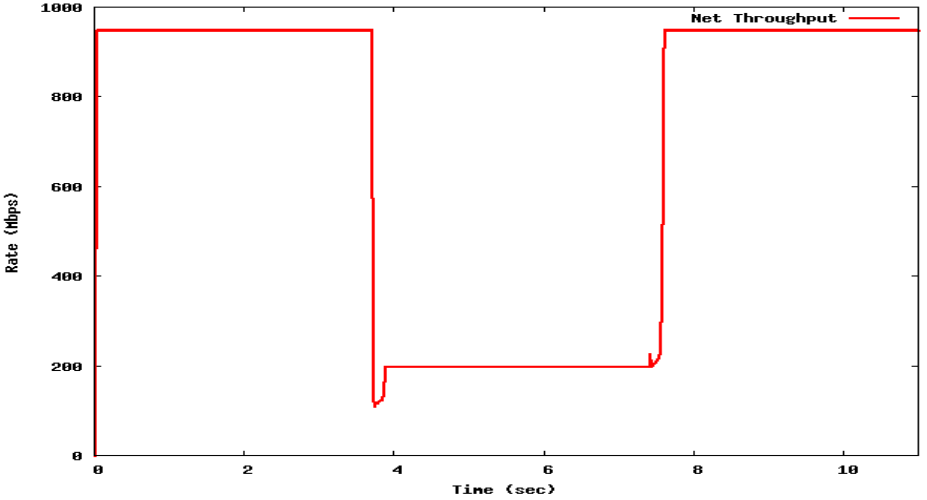
Hardware





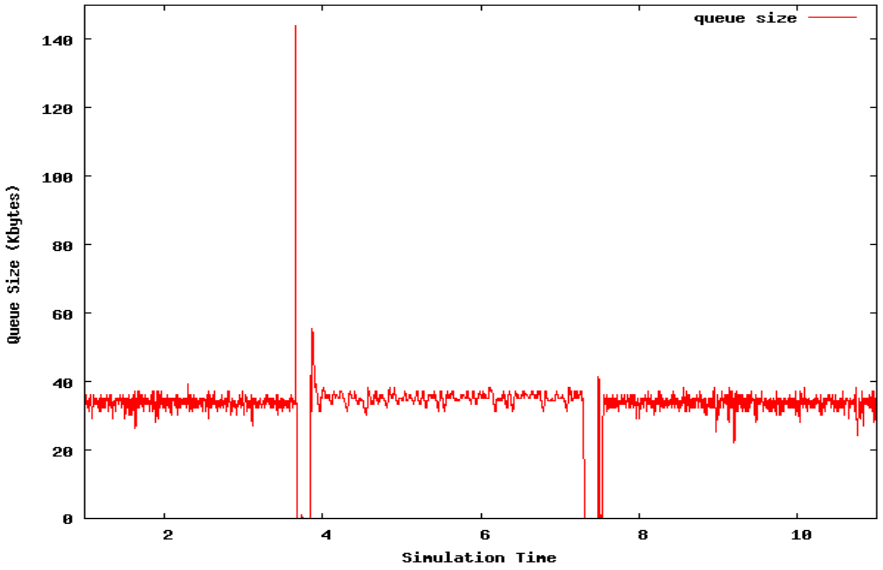
OMNET++

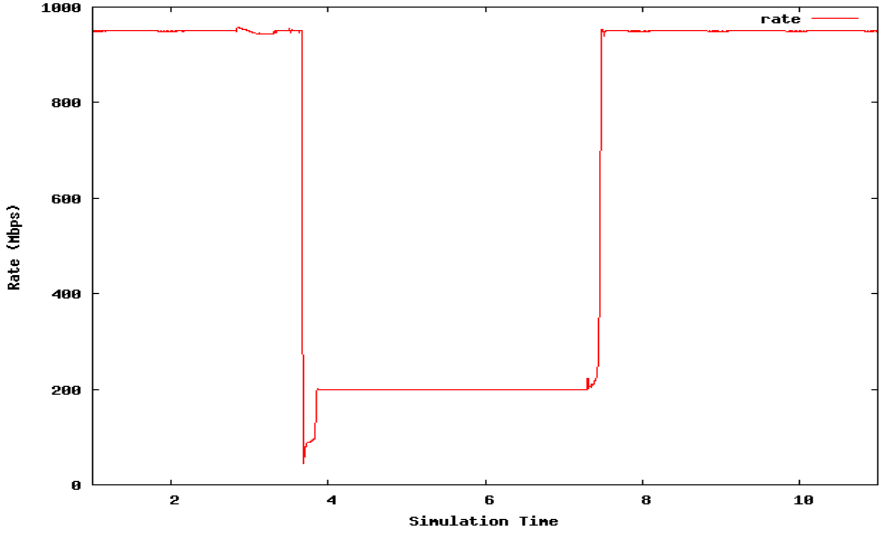




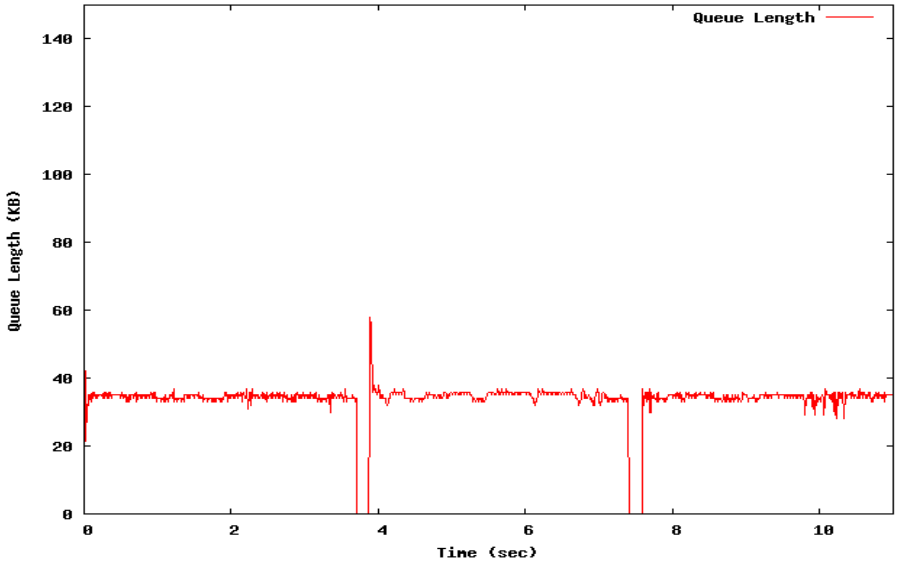
1 source, RTT = 1000us

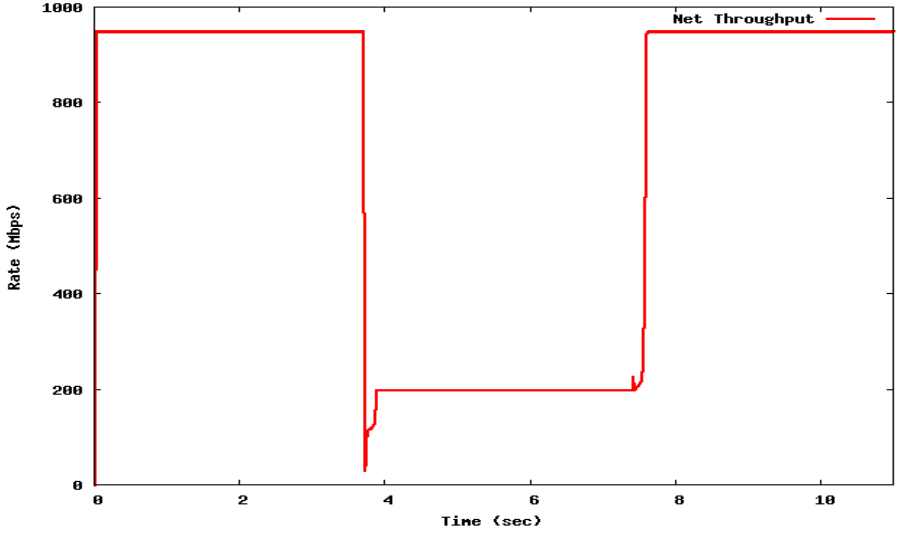
Hardware





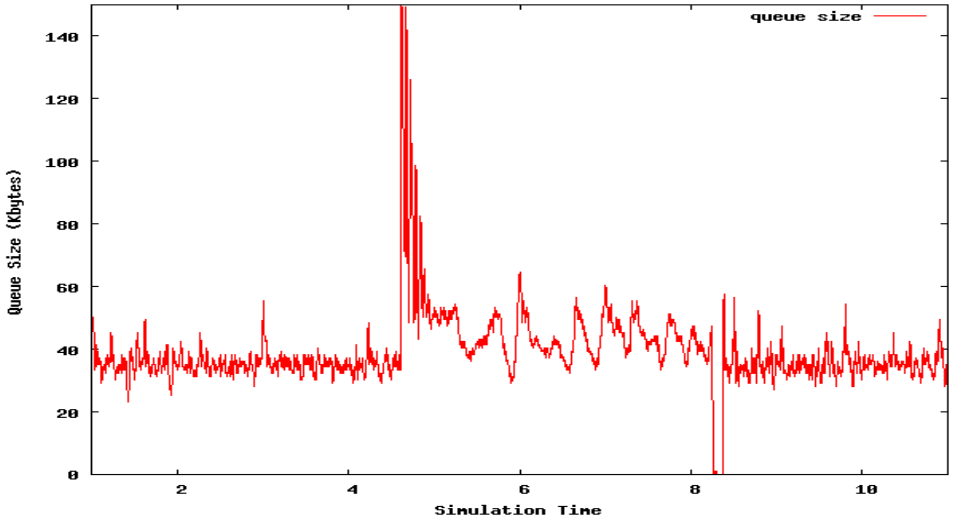
OMNET++

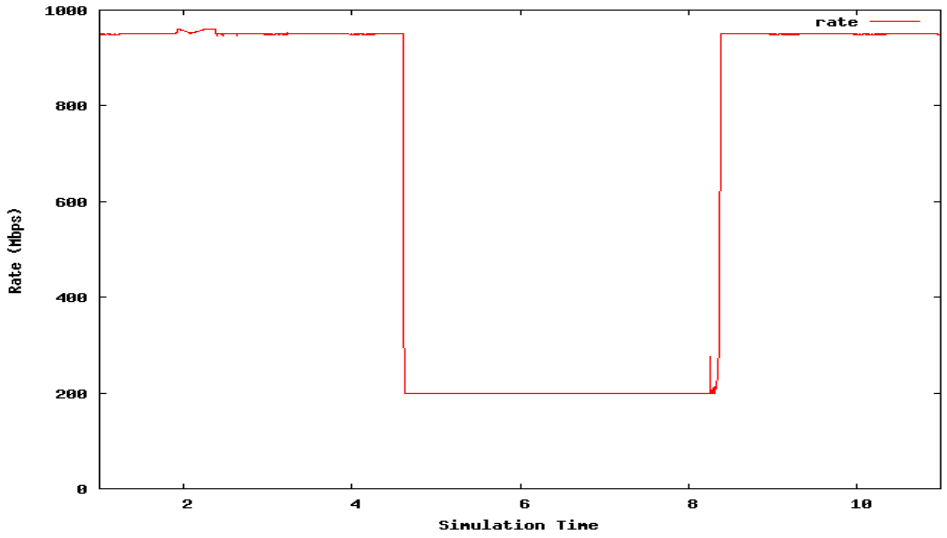




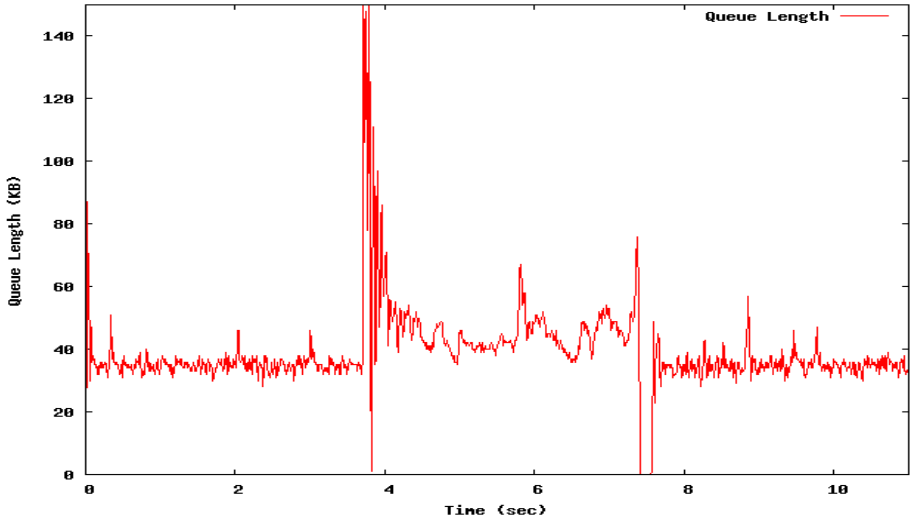
8 sources, RTT = 100us

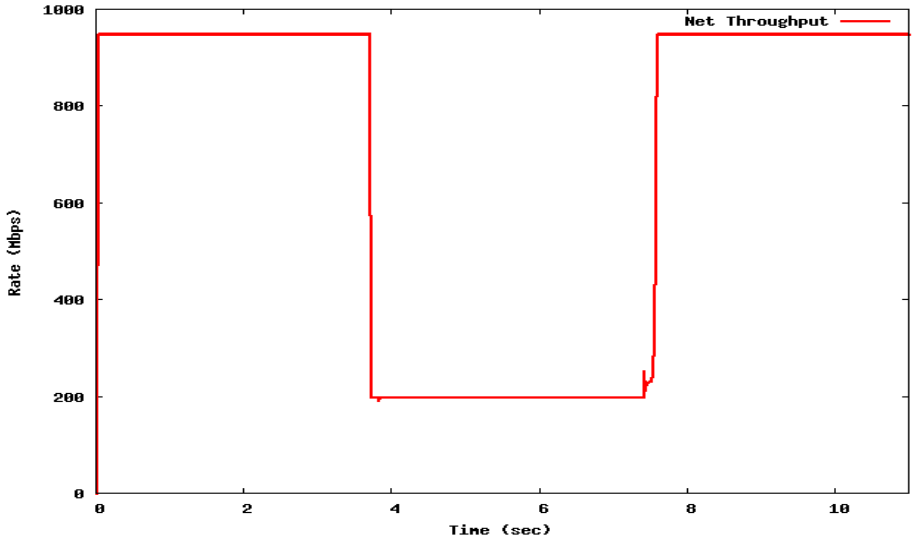
Hardware





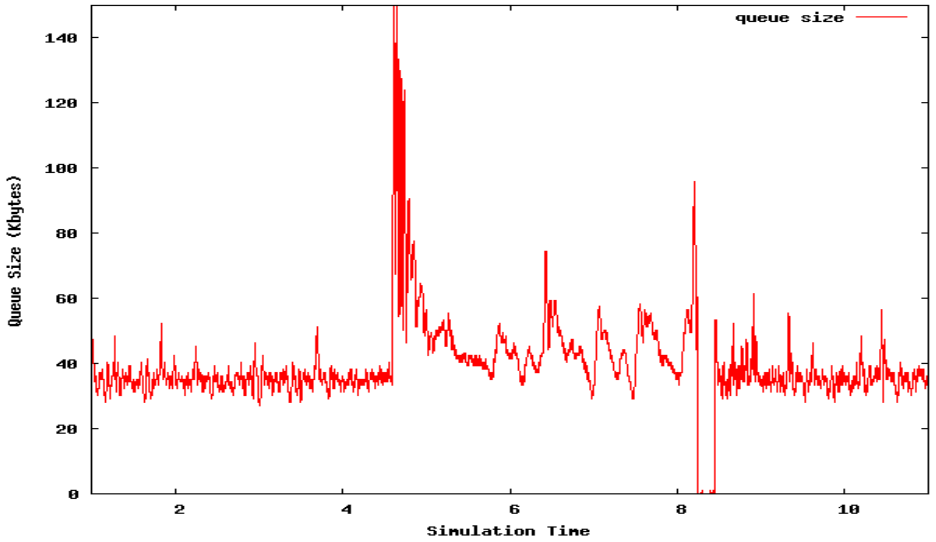
OMNET++

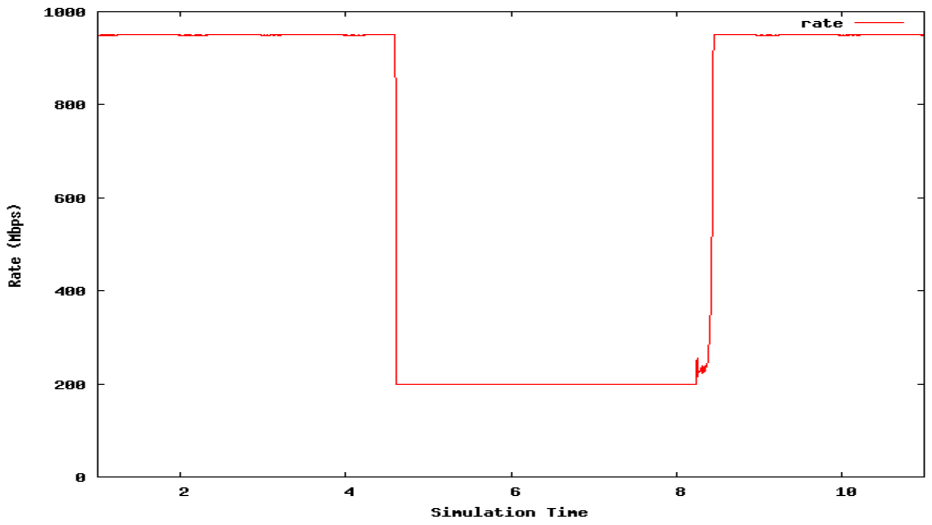




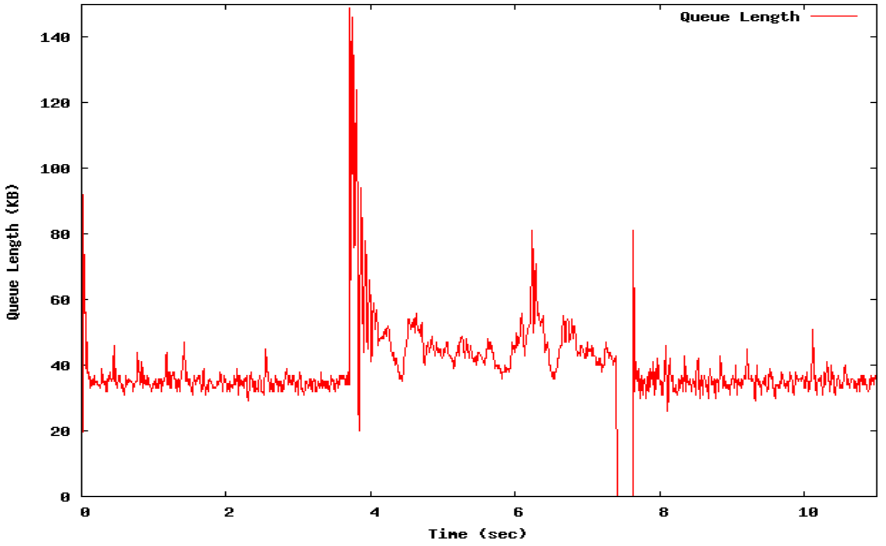
8 sources, RTT = 500us

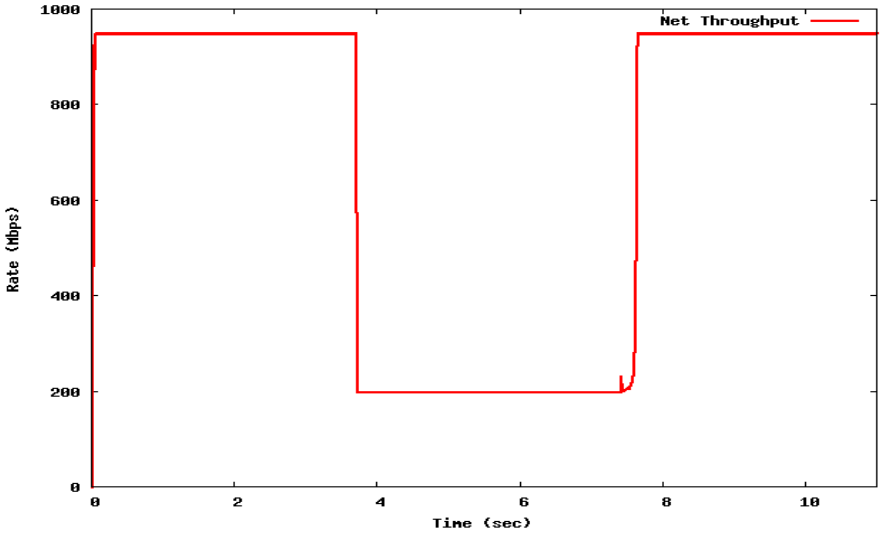
Hardware





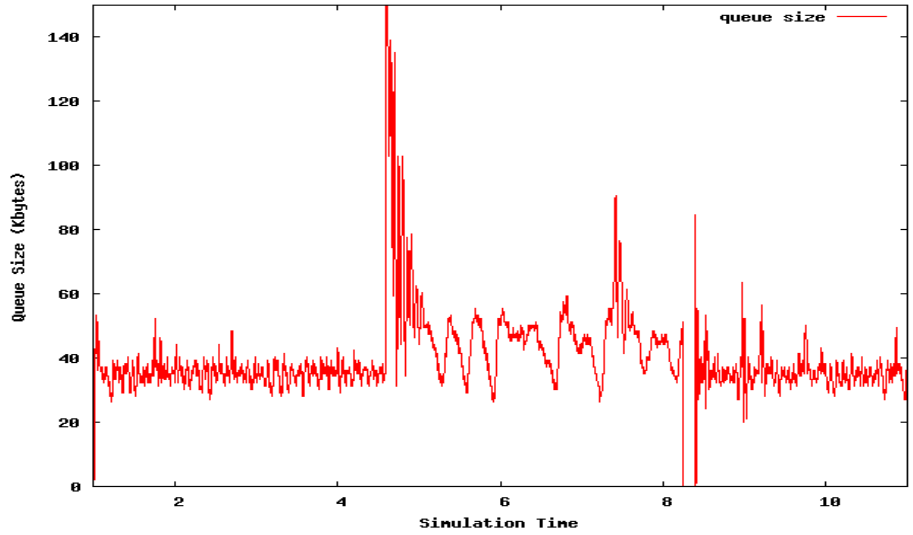
OMNET++

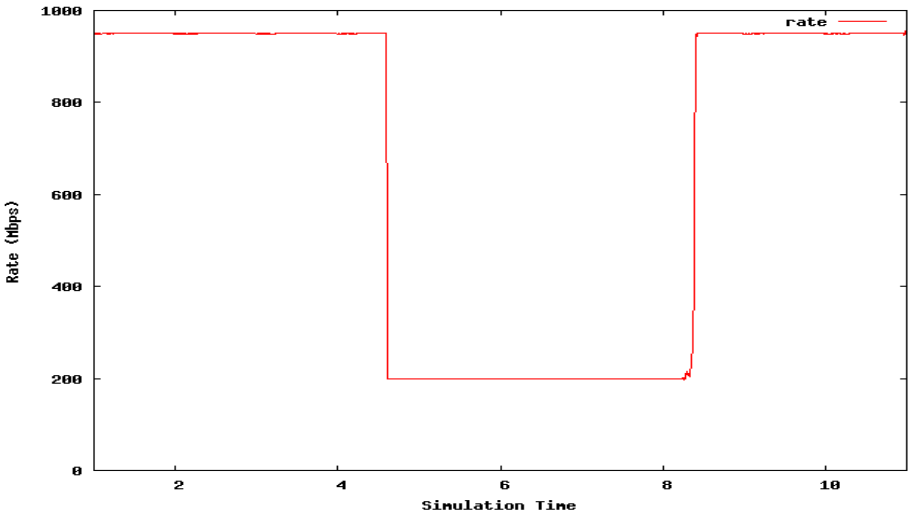




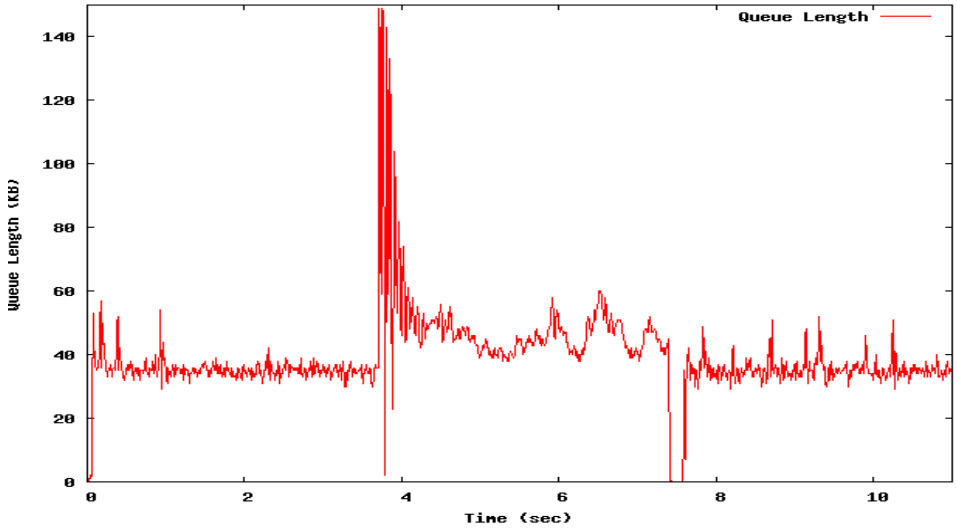
8 sources, RTT = 1000us

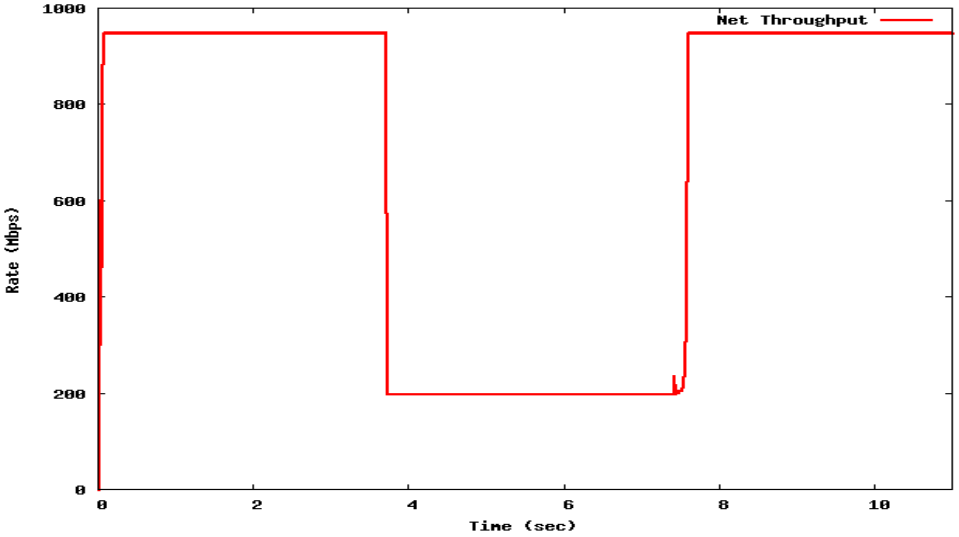
Hardware



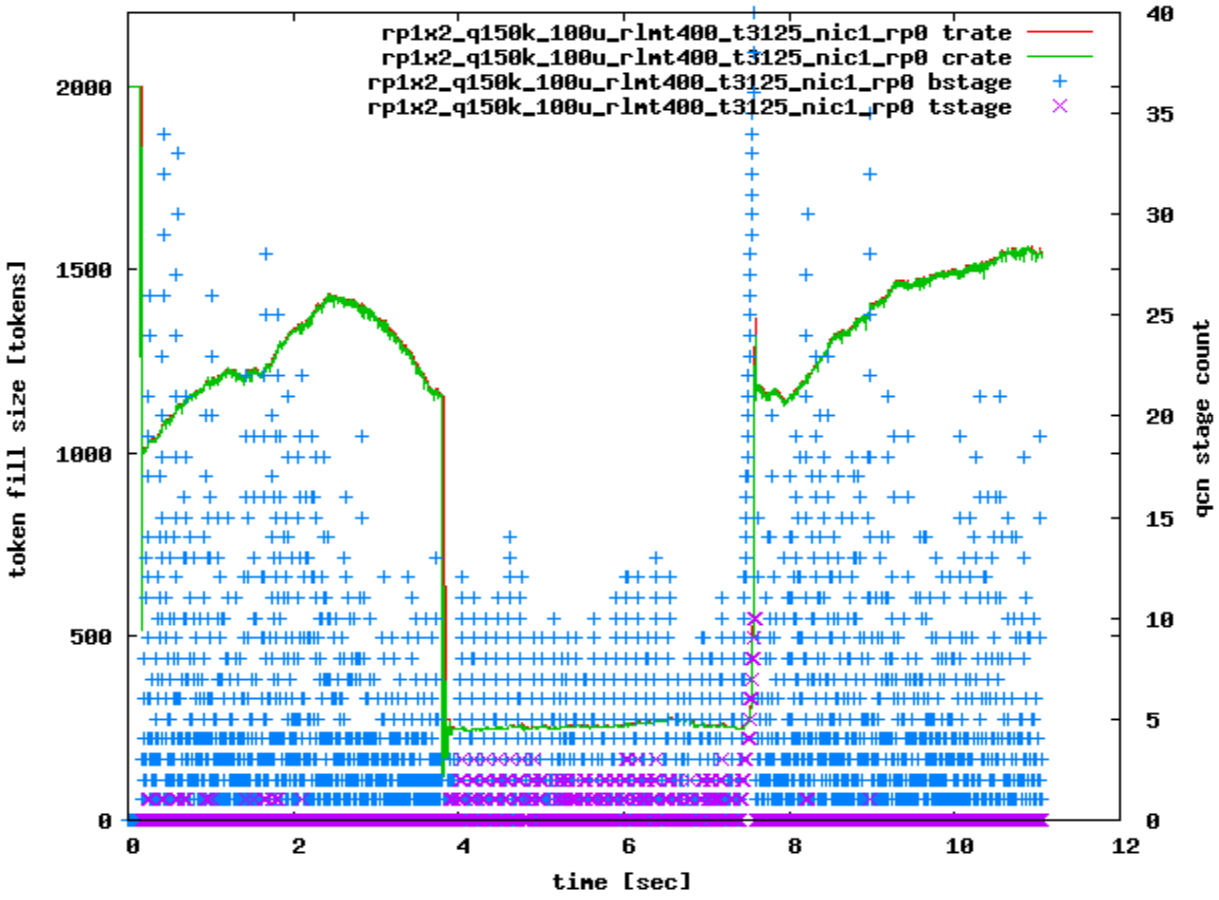


OMNET++





Statistic for Debugging & Evaluation



Byte_counter_stage, timer_stage, TR, and CR values for one of two RPs at 200us RTT

Demo

Summary

- Demonstrated successful QCN system operation
- Experiments and simulations match very well
- Built a test-bed for conducting further experiments
  - Scalable topologies for data centers (4 RPs and 4 CPs per FPGA board)
  - Tunable:
    - output switch buffer sizes
    - rate-limiter queue sizes
    - link RTTs
    - link capacities
    - QCN parameters
  - Ability to trace and analyze all major QCN variables
- Easily portable to 10Gbps hardware
- Will study the TCP-QCN interaction further in the future