Effects of Pause and QCN on TCP Sources: NewReno

Berk Atikoglu, Balaji Prabhakar

Apr 24, 2008

Overview

- Quantify performance of QCN + TCP interactions in terms of Innocent Flow and Hot Spot Flow throughput
  - 500ms congestion scenario
  - Periodic congestion scenario
  - (These have been studied by Kwan et al; we use it to validate ourselves.)

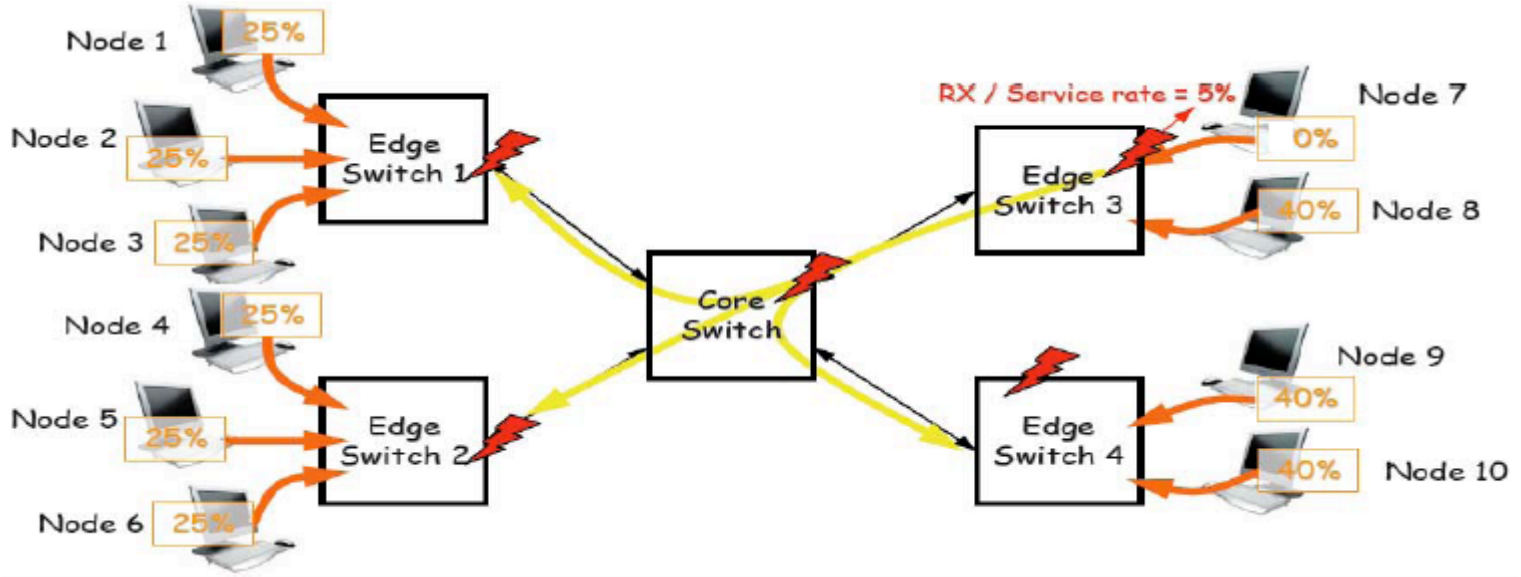
- Analyze the effects of Pause and QCN on TCP source

System Parameters

- **Congestion Management Schemes**
  - TCP Only
  - TCP + QCN
  - TCP + QCN + PAUSE
  - QCN + PAUSE
- **Switch Parameters**
  - PAUSE Disabled
    - Output queue limit of 150kbytes
  - PAUSE Enabled
    - No output queue limit
      - Applied on a per input basis based on watermarks
      - Watermark_hi = 130kbytes
      - Watermark_lo = 110kbytes
- **QCN Parameters**
  - W = 2.0
  - Q_EQ = 26kbytes
  - Gd = 1/128 = 0.0078125
  - Base marking: once every 150kbytes
  - Jitter on marking: 30%
  - Runit = 1Mb/s
  - MIN_RATE = 10Mb/s
  - BC_LIMIT = 150kbytes
  - TIMER_PERIOD = 15ms
  - R_AI = 5Mbps
  - R_HAI = 50Mbps
  - FAST_RECOVERY_TH = 5
  - Quantized_Fb: 6 bits
  - Jitter at RP: 30% (byte counter and timer)
- **TCP Version → NewReno**

Data is taken form: http://www.ieee802.org/1/files/public/docs2008/au-sim-kwan-qcn-tcp-0308.pdf

Topology and Workload: 500ms Congestion Event

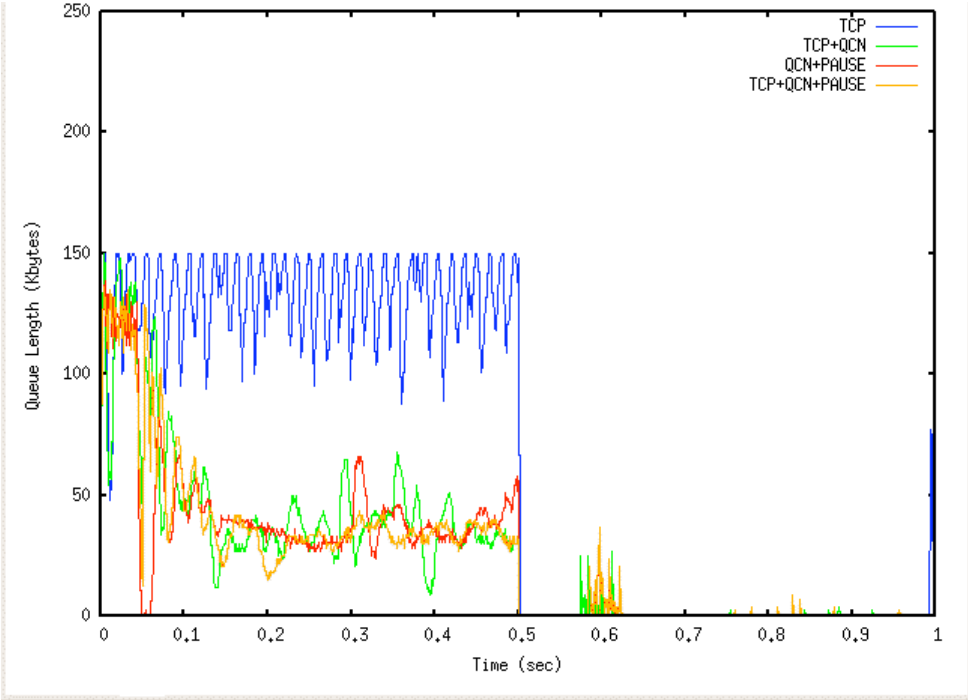


- Multi-stage Output-Generated Hotspot Scenario
  - Link Speed = 10Gbps for all links
  - Loop Latency = 16us
- Traffic Pattern
  - 9k byte transactions arriving with a Bernoulli distribution
  - Transport layer is either UDP or TCP
  - Destination Distribution: Uniform distribution to all nodes (except self)
  - Frame Size Distribution: Fixed length (1500bytes) frames
  - Offered Load
- Nodes 1-6 = 25% (2.5Gbps)
- Nodes 8-10 = 40% (4Gbps)
- Congestion Scenario
  - Node 7 temporarily reduces its service rate from 10Gbps to 500Mbps between [0-500ms]

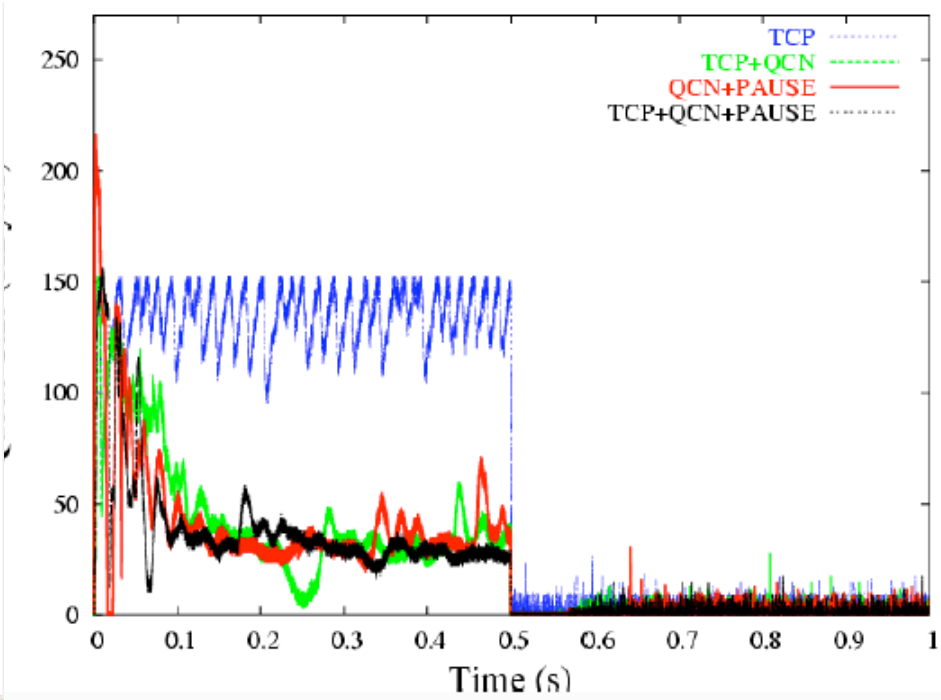
Data is taken form: http://www.ieee802.org/1/files/public/docs2008/au-sim-kwan-qcn-tcp-0308.pdf

500ms congestion scenario

Our simulation

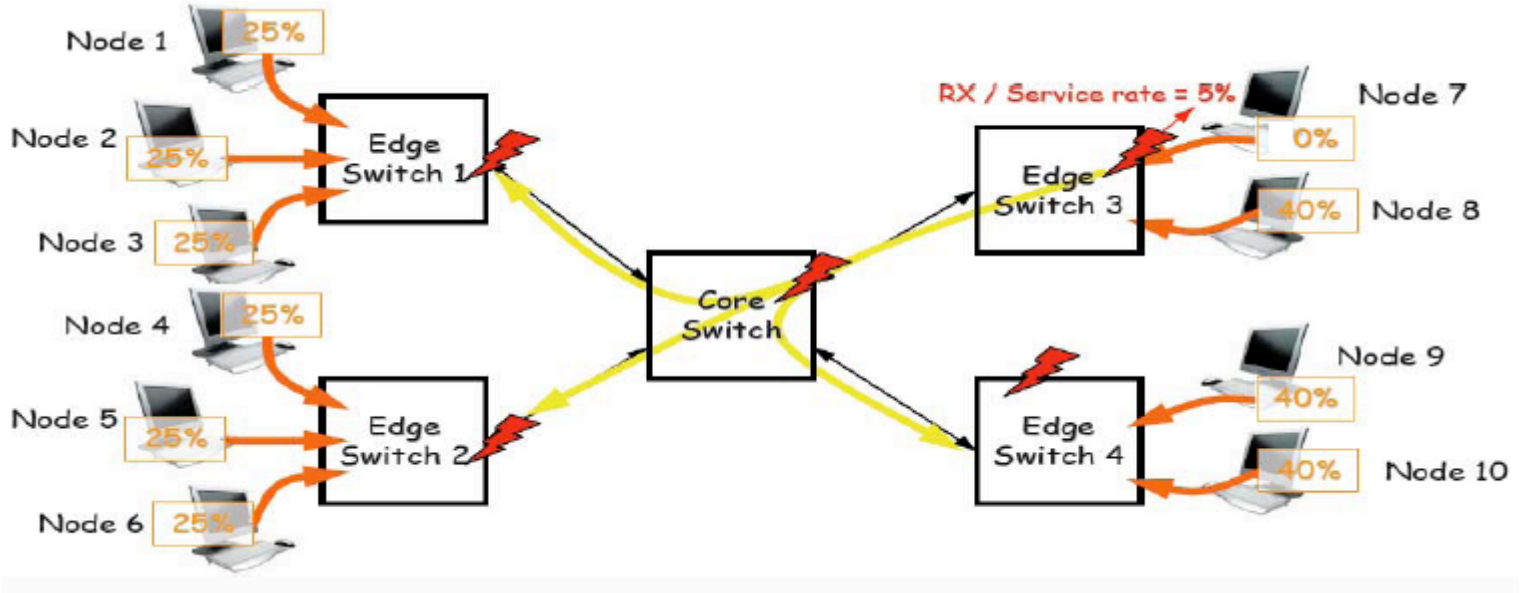


Bruce Kwan's simulation



Right picture is taken form: http://www.ieee802.org/1/files/public/docs2008/au-sim-kwan-qcn-tcp-0308.pdf

Topology and Workload: Periodic Congestion Events

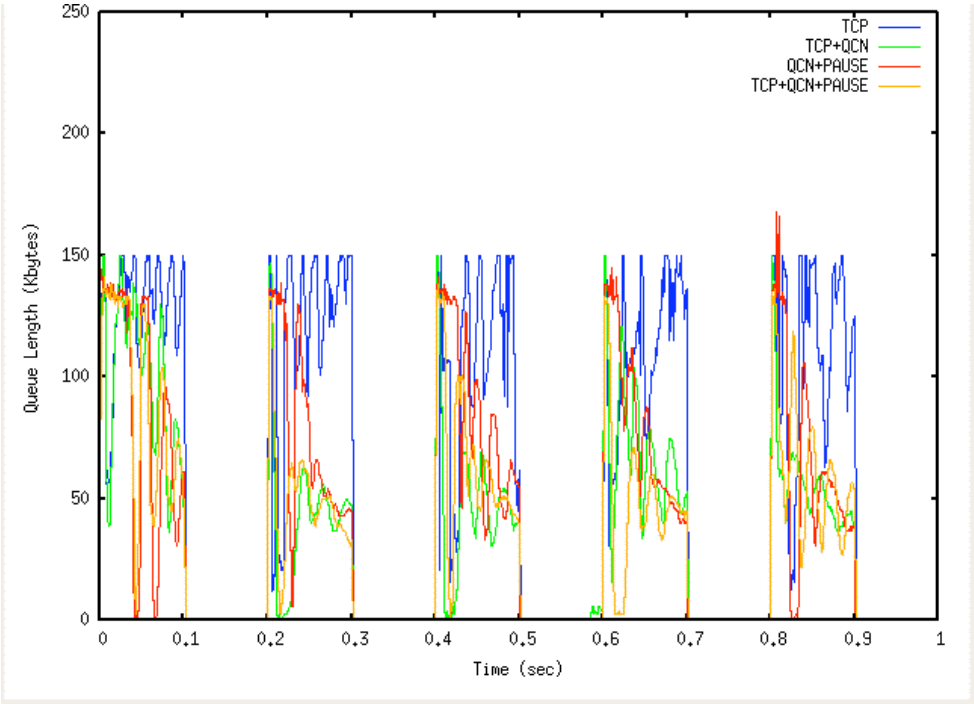


- Traffic Pattern
  - Same as before
- Congestion Scenario
  - Node 7 periodically reduces its service rate from 10Gbps to 500Mbps
  - Congestion Duration: 100ms
- Duty Cycle = 1/2
- Simulation Duration: 1 second
- Performance Metric: Aggregate Throughput
  - Ideal Aggregate Innocent Flow Throughput: 24Gbps
  - Ideal Aggregate Victim Flow Throughput: 500Mbps or 3Gbps (Avg = 1.75Gbps)

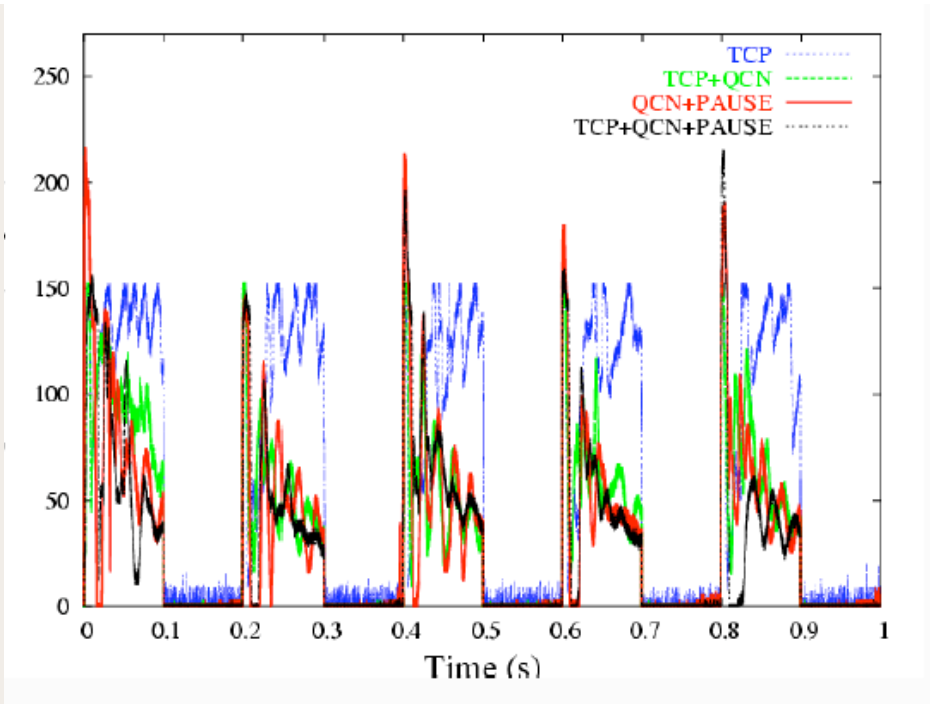
Data is taken form: http://www.ieee802.org/1/files/public/docs2008/au-sim-kwan-qcn-tcp-0308.pdf

Periodic congestion scenario

Our simulation



Bruce Kwan's simulation



Right picture is taken form: http://www.ieee802.org/1/files/public/docs2008/au-sim-kwan-qcn-tcp-0308.pdf

Analyze the Effects of Pause and QCN on TCP Sources: System Parameters

- **Congestion Management Schemes**
  - TCP Only
  - TCP + PAUSE
  - TCP + QCN + PAUSE

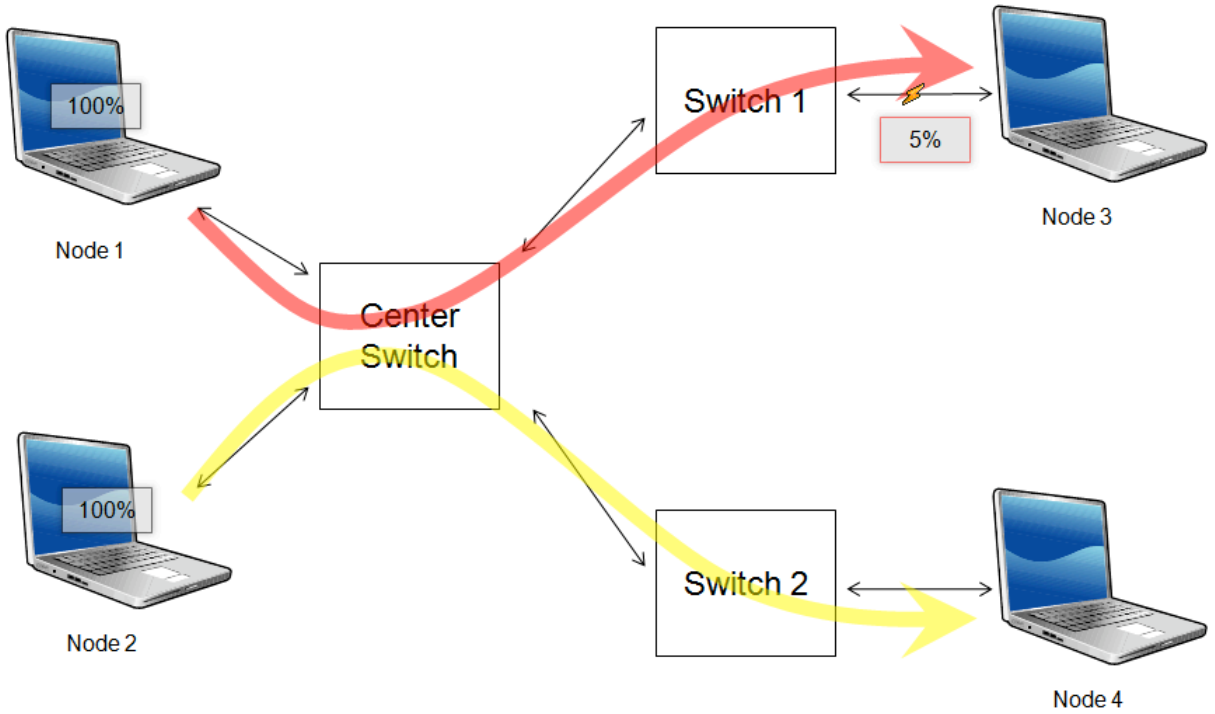
- **Switch Parameters**
  - PAUSE Disabled
    - Output queue limit of 150kbytes
  - PAUSE Enabled
  - No output queue limit
    - Applied on a per input basis based on watermarks
    - Watermark_hi = 130kbytes
    - Watermark_lo = 110kbytes

- **QCN Parameters**
  - W = 2.0
  - Q_EQ = 26kbytes
  - Gd = 1/128 = 0.0078125
  - Base marking: once every 150kbytes
  - Jitter on marking: 30%
  - Runit = 1Mb/s
  - MIN_RATE = 10Mb/s
  - BC_LIMIT = 150kbytes
  - TIMER_PERIOD = 15ms
  - R_AI = 5Mbps
  - R_HAI = 50Mbps
  - FAST_RECOVERY_TH = 5
  - Quantized_Fb: 6 bits
  - Jitter at RP: 30% (byte counter and timer)

- **TCP Version → NewReno**

Data is taken form: http://www.ieee802.org/1/files/public/docs2008/au-sim-kwan-qcn-tcp-0308.pdf

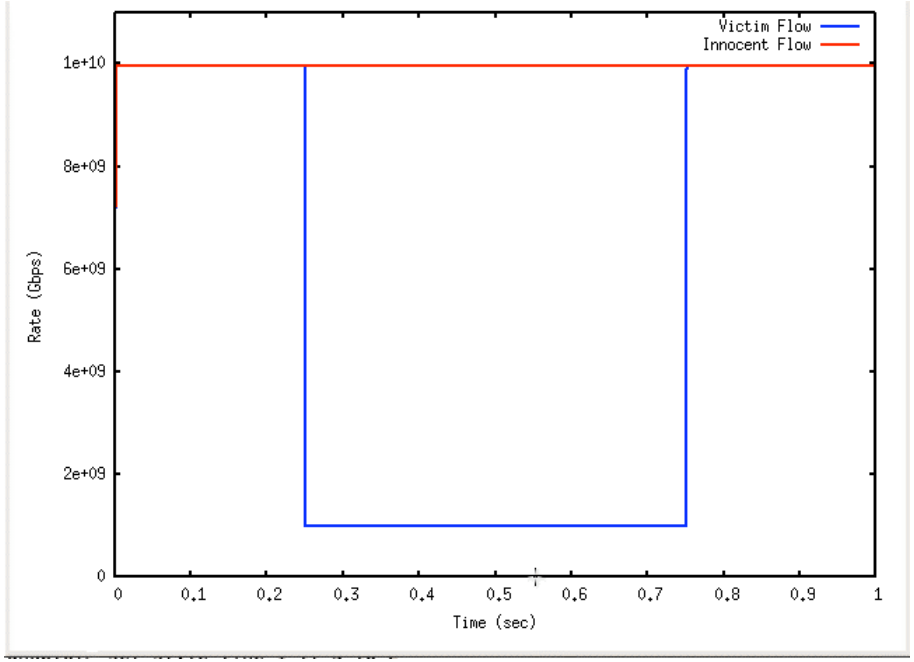
Topology and Workload



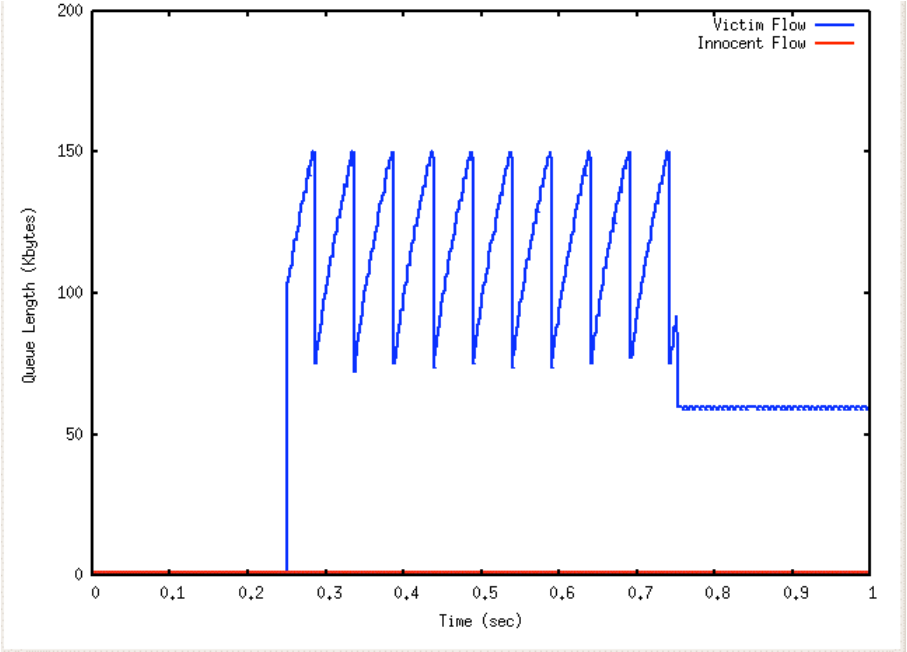
- Topology
  - Link Speed → 10Gbps for all links
  - Loop Latency → 24us
- Traffic Pattern
  - 9k byte transactions arriving with a Bernoulli distribution
  - Node 1 sends to Node 3 at 10Gbs (100%)
  - Node 2 sends data to Node 4 at 10Gbs (100%)
- Congestion Scenario
  - Node 3 temporarily reduces its service rate from 10Gbps to 1Gbps between [250-750ms]; congestion propagates to Center Switch, which pauses both incoming links

TCP Only

Net Throughput at Switch 1 and 2

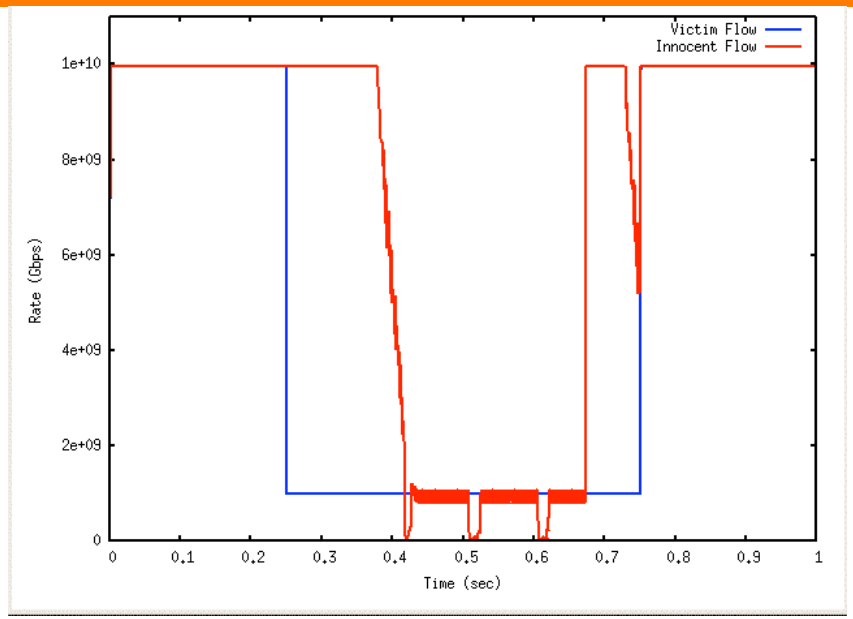


Queue Length at Switch 1 and 2

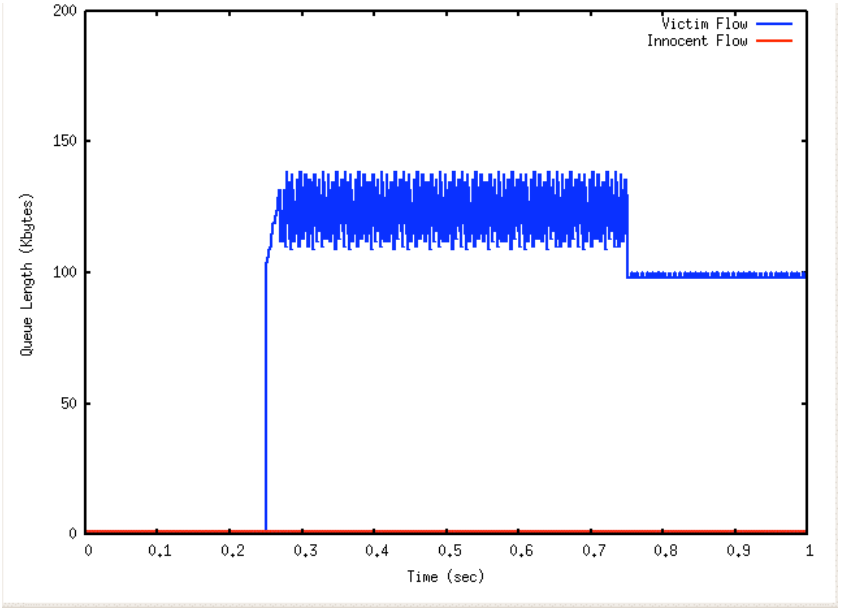


TCP and PAUSE

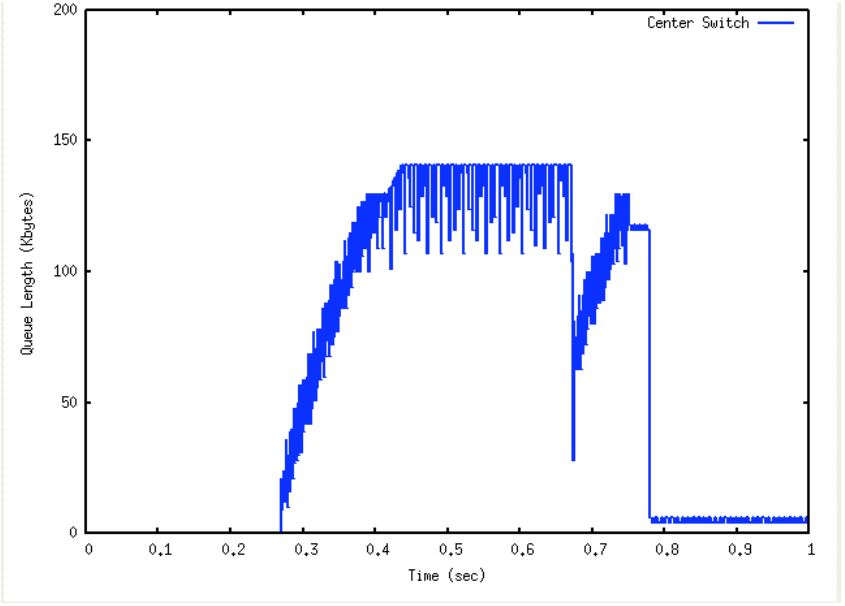
Net Throughput at Switch 1 and 2



Queue Length at Switch 1 and 2

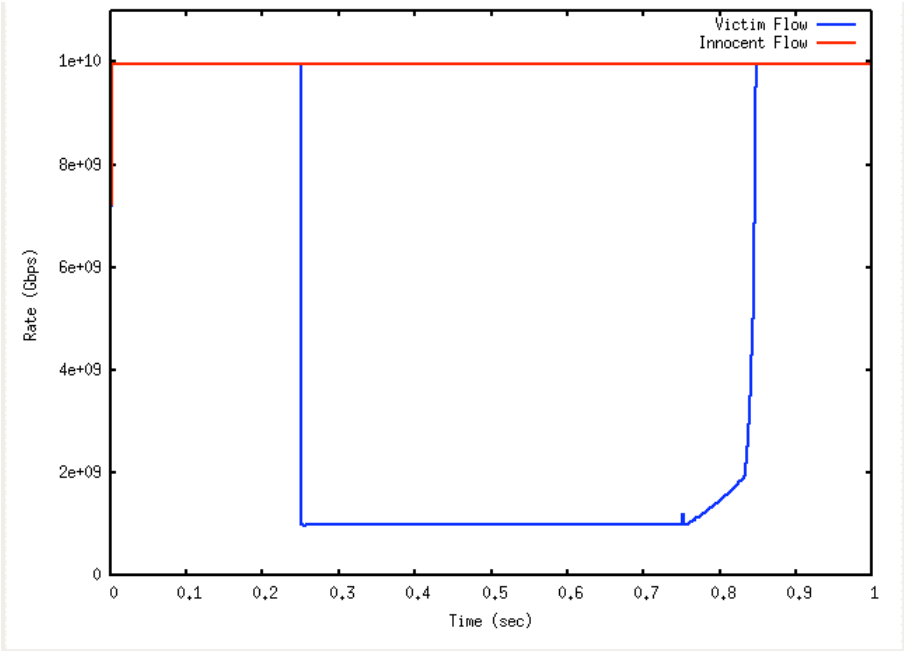


Queue Length at Center Switch

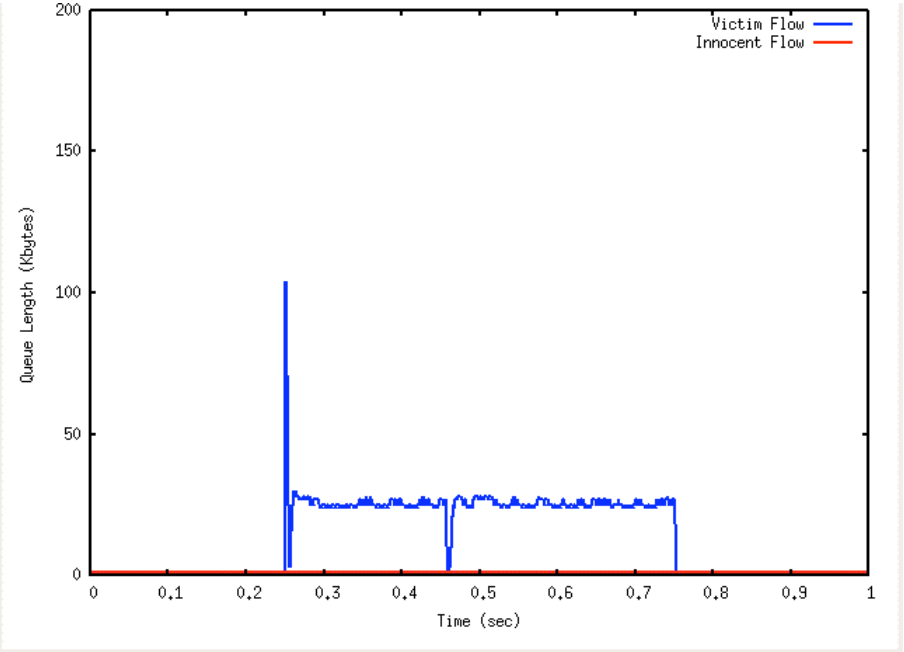


TCP, QCN and PAUSE

Net Throughput at Switch 1 and 2



Queue Length at Switch 1 and 2



Conclusion

- When pause is used, TCP alone has no knowledge of the ultimate bottleneck
  - That is L2 information which QCN is aware of

- QCN moves the bottleneck rate to the appropriate rate limiter at the edge, which TCP can then adapt to