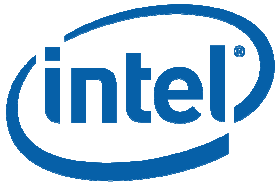


QCN Serial-HAI Simulation Benchmarks and Qeq

Eric Geisler

Manoj Wadekar

14 February 2008

Goals

- Present required benchmarks for QCN using Opnet
- Study effects of varying Qeq
- Study effects of more severe congestion

Simulation Parameters

- Traffic
  - I.i.d. Bernoulli arrivals
  - Uniform destination distribution (to all nodes except self)
  - Fixed frame size = 1500 B
- Switch
  - VOQ with 1.5MB shared mem
  - Partitioned memory per input, shared among all outputs
  - No limit on per-output memory usage
  - PAUSE enabled
    - ✓ Applied on a per input basis based on local high/low watermarks
    - ✓ $watermark_{high}$ = 130 KB
    - ✓ $watermark_{low}$ = 110 KB
- Adapter
  - RLT: VOQ and single; RR service
  - One rate limiter per destination, limited to 16
  - Egress buffer size = 150 KB,
  - Ingress buffer size = Unlimited
  - PAUSE enabled
    - ✓ $watermark_{high}$ = 150 – rtt*bw KB
  - $watermark_{low}$ = $watermark_{high}$ - 20 KB

- QCN and ECM base
  - W = 2.0
  - M = 150 KB
  - Q_{eq} = 26 KB
  - R_{unit} = R_{min} = 10 Mb/s
  - ECM_{MAX} enabled, Q_{mc} = M
  - $ECM_{0,0}$ disabled
- QCN-SHAI
  - G_d = 0.0078125 (1 / 128)
  - BC_LIMIT = 150 KB
  - SI timer period = 15 ms
  - A_{ai} = 5 Mb/s
  - A_{hai} = 50 Mb/s
  - Fast Recovery Threshold = 5
  - 6-bit quantization
  - Jitter at RP (bytes and timer) = 30%
  - Jitter at CP (packet marking) = 30%

Switch & Adapter Parameters

- Switch parameters
  - M = 150 KB per port
  - Dedicated per input, shared across all outputs
  - Configurable OQ limit; frames are dropped when OQ length exceeds limit
  - PAUSE enabled or disabled
    - Applied on a per input basis based on local high/low watermarks
    - $watermark_{high}$ = M – rtt*bw KB
    - $watermark_{low}$ = M – rtt*bw – 10 KB
    - If disabled, frames dropped when input partition full

- Adapter parameters
  - Virtual output queuing, round-robin VOQ service
  - Input buffer size IB = 1.5 MB, partitioned per VOQ
    - Drop when VOQ full
  - Output buffer size OB = 150 KB
  - Limit of 16 rate limiters
  - PAUSE enabled
    - $watermark_{high}$ = OB – rtt*bw KB
    - $watermark_{low}$ = $watermark_{high}$ - 10 KB

Topology & Workload
Benchmark 1

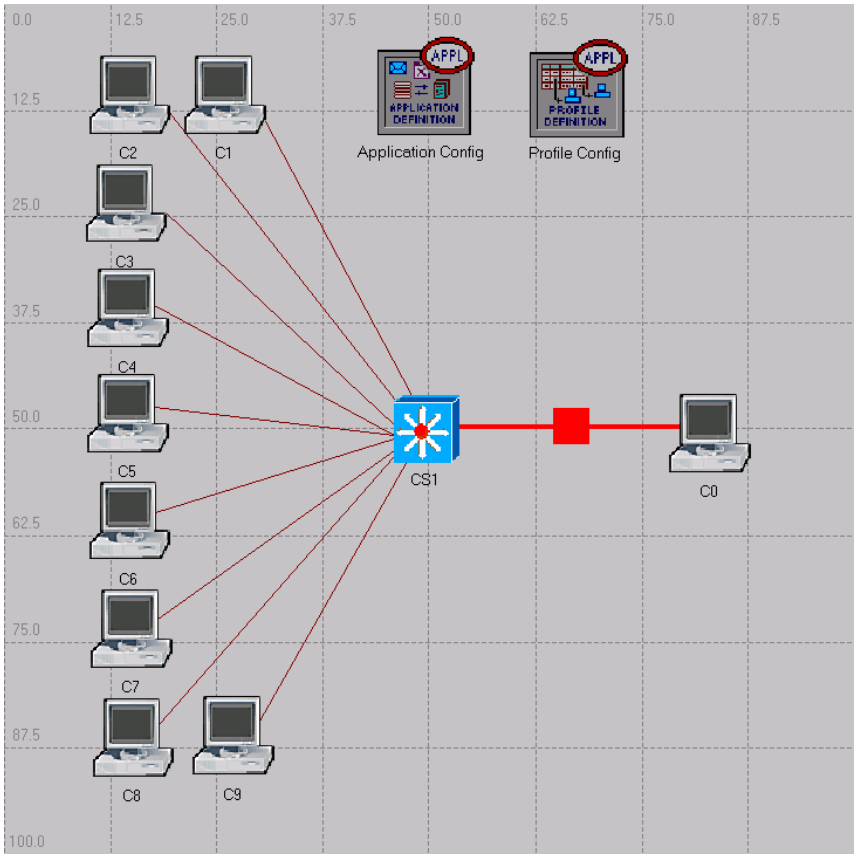
OG Hotspot, Single Hop

2 Gbps OG hotspot for 80ms @ CS1→C0

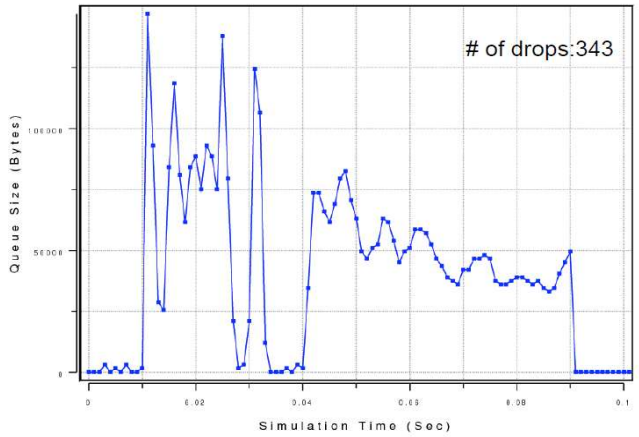
802.3x PAUSE disabled

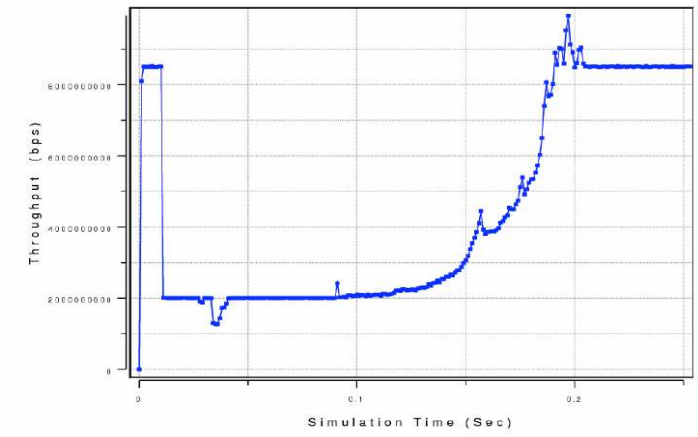
Traffic pattern:

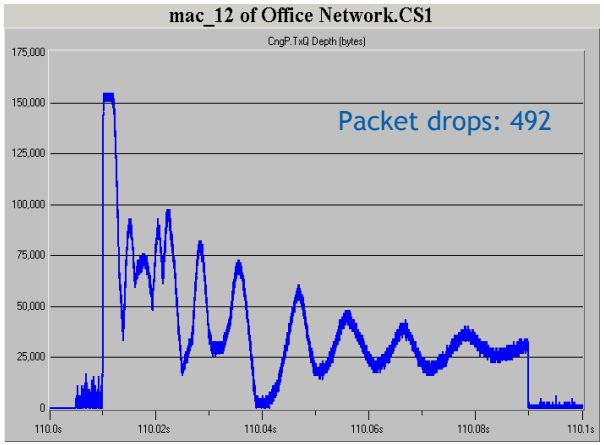
- 10 Gbps links, 500ns link latency
- All 10 hosts C0-C9 @ 85% loading
- Spatially uniform (except self)
- Temporally Bernoulli

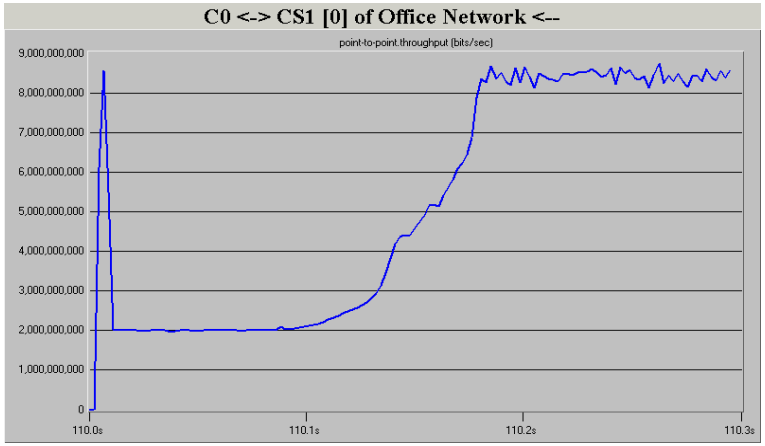


Queue Depth and Hotspot Throughput









Topology & Workload
Benchmark 3

OG Hotspot, Multi Hop: Selected Victims

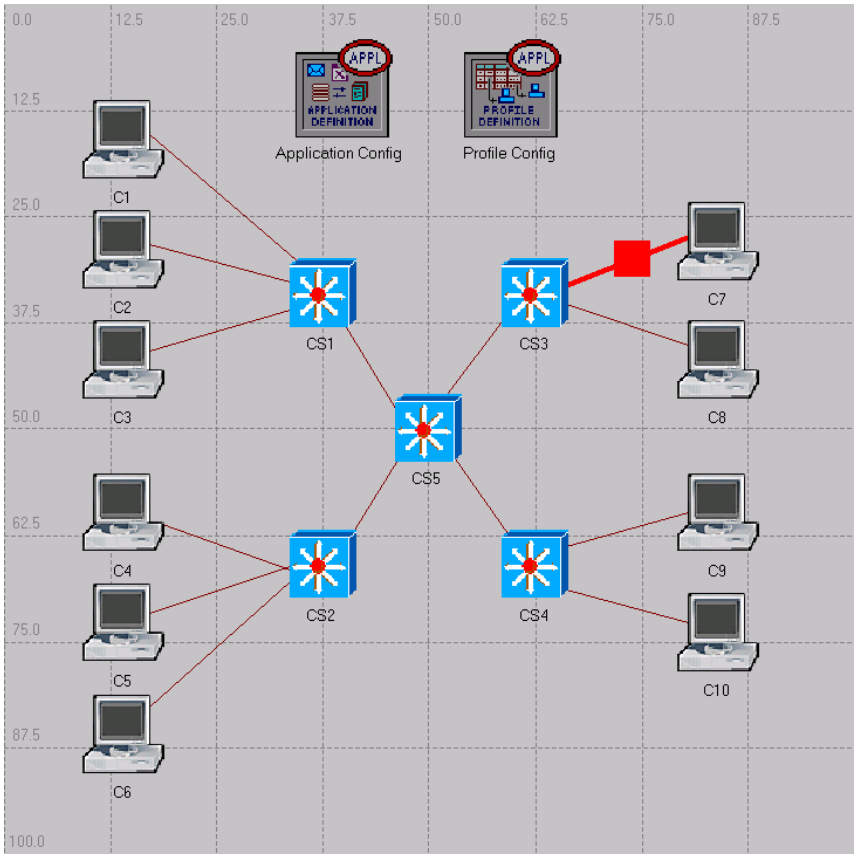
2 Gbps OG hotspot for 80ms @ CS3→C7

802.3x PAUSE enabled

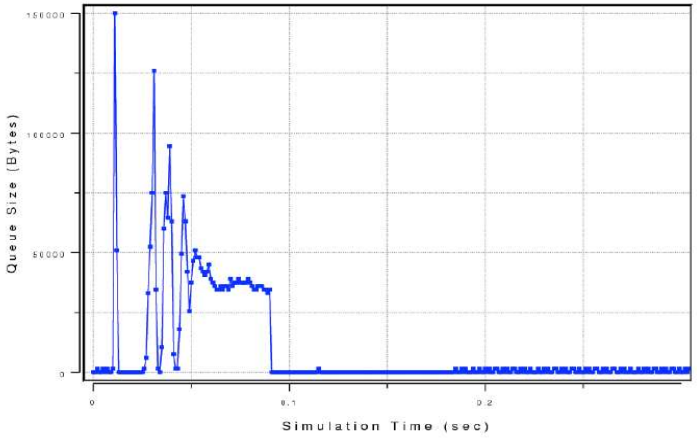
Fair allocation provides 0.5Gb/s to culprit flows and 7Gb/s to victim flows

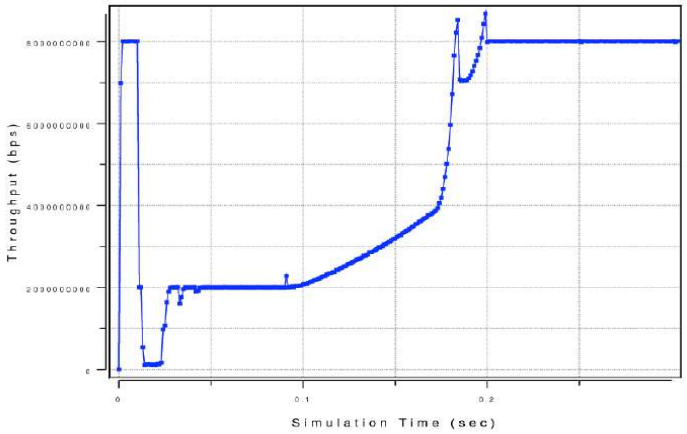
Traffic pattern:

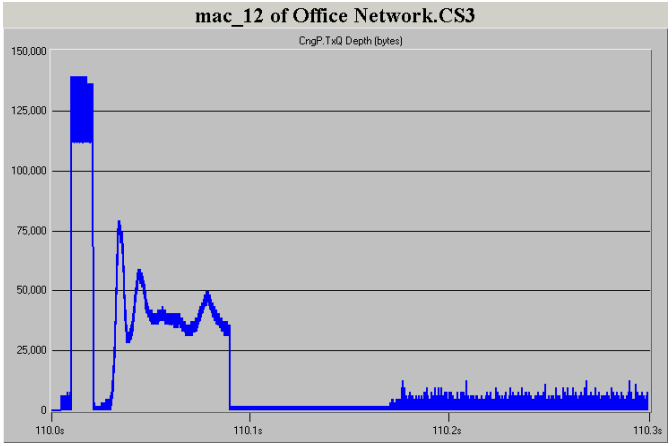
- 10 Gbps links, 500ns link latency
- 4 culprit flows: 1, 4, 8, 9 @ 70% → 7
- 3 victim flows: 2 → 9, 5 → 3, 10 → 6 @ 20%
- Hosts 3, 6, 7 are only receiving
- Temporally Bernoulli

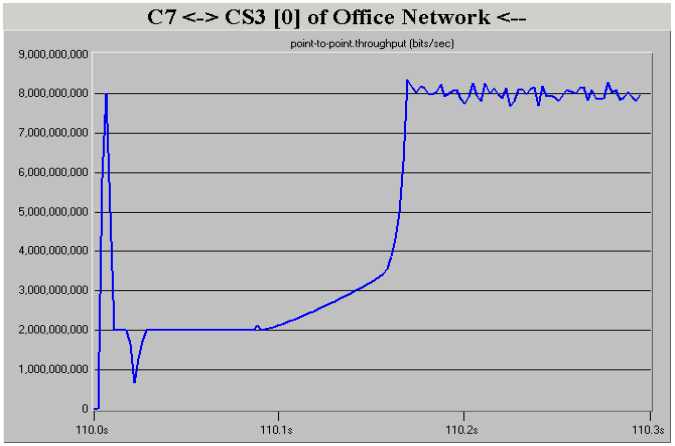


Queue Depth and Hotspot Throughput









Topology & Workload
Benchmark 5

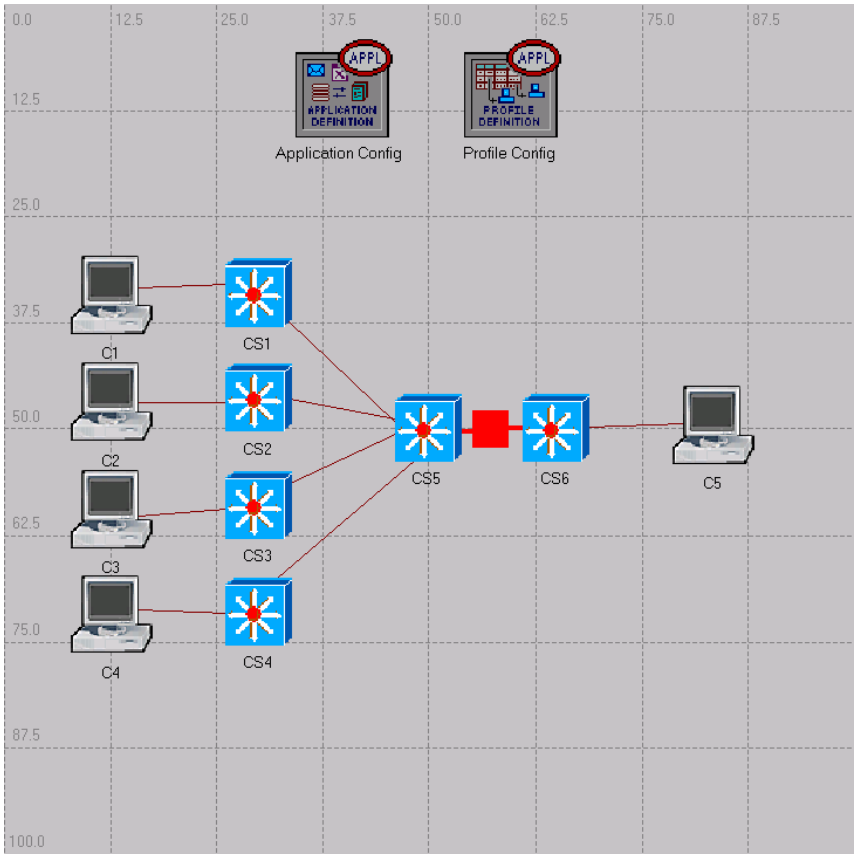
Symmetric Topology, Single HS: Bursty

Congestion point typically occurs at CS5→CS6

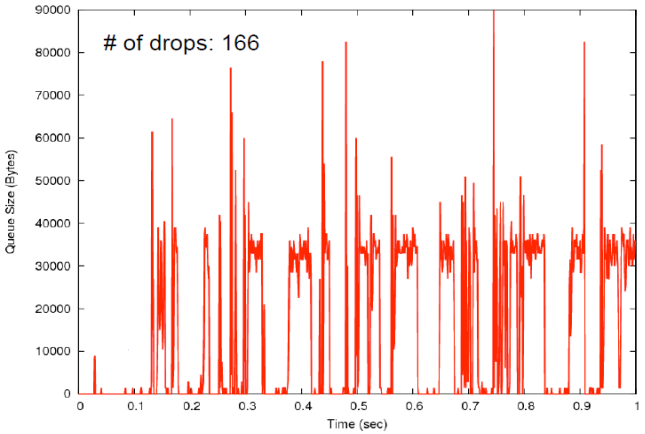
802.3x PAUSE disabled

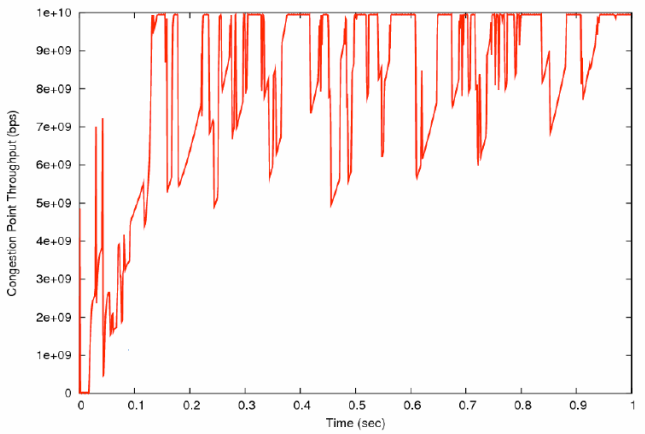
Traffic pattern:

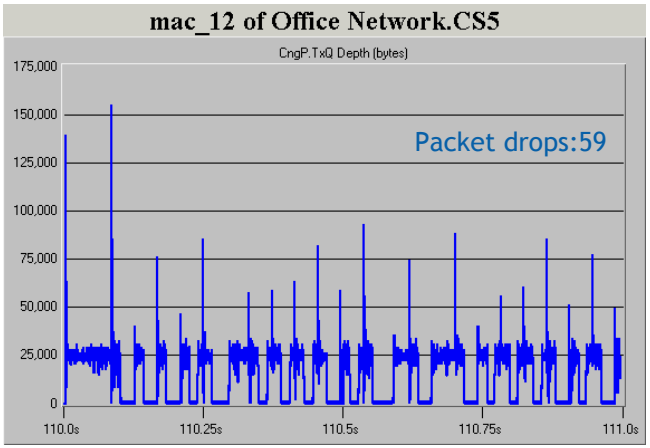
- 10 Gbps links, 500ns link latency
- Point-to-point from C1-C4 to C5
- Nodes 1-4 @ 100% → C5
- Nodes 1-2 have bursty load (Ton = Toff = 20ms)
- On/off period exponential distribution

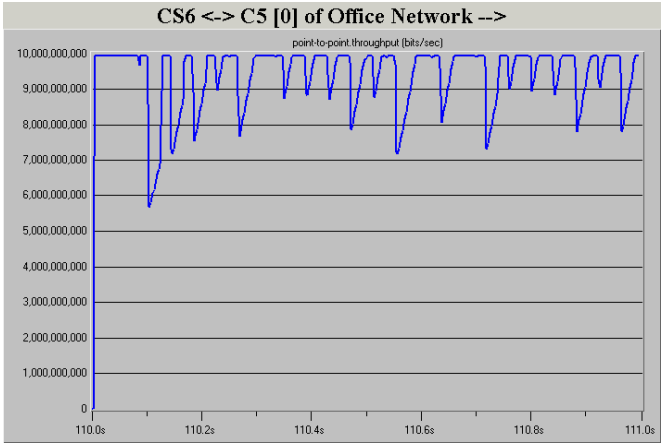


Queue Depth and Hotspot Throughput







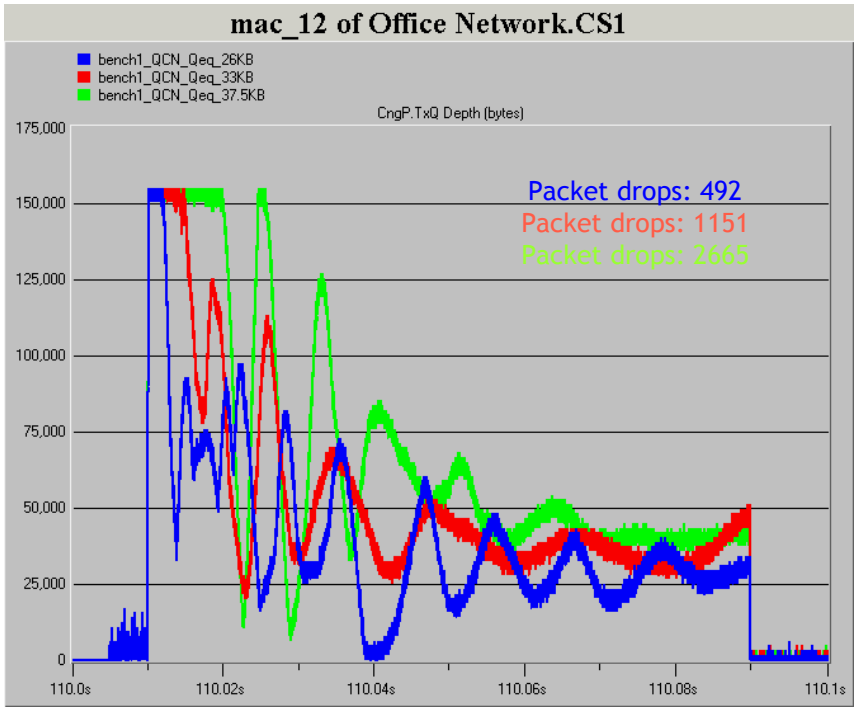


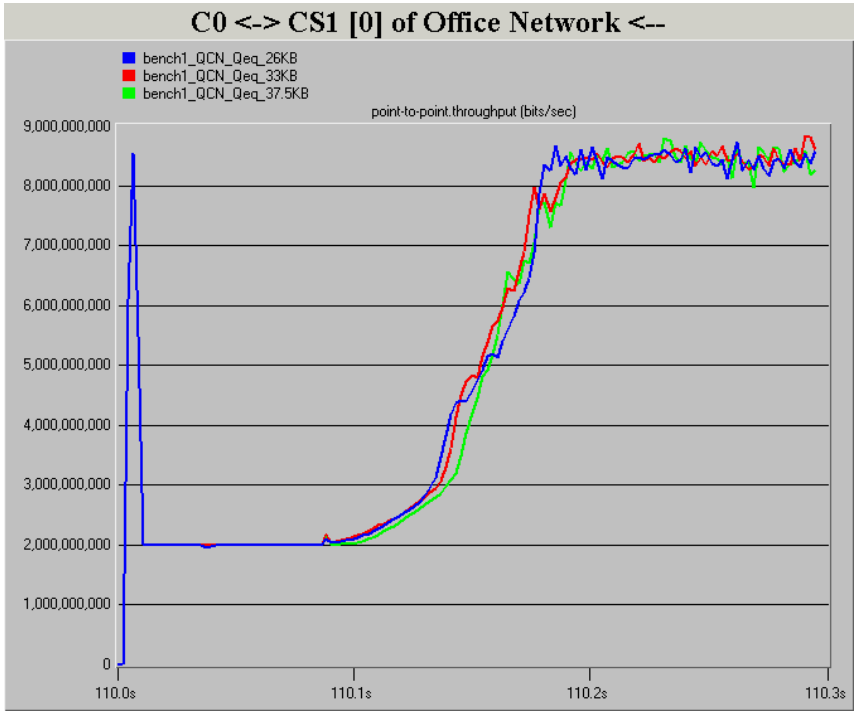
Next Goal

- Present required benchmarks for QCN using Opnet
- **Study effects of varying Qeq**
- Study effects of more severe congestion

Queue Depth and Hotspot Throughput

Benchmark 1, Qeq = 26, 33, 37.5 KB





Qeq and quantized Fb

Formulas:

Fb = (Qeq - Qlen) – W · (Qlen - Qlen_old)

Max_Fb = Qeq · (2 · W + 1)

Quantized_Fb = (Fb / Max_Fb) · 64;

Analysis:

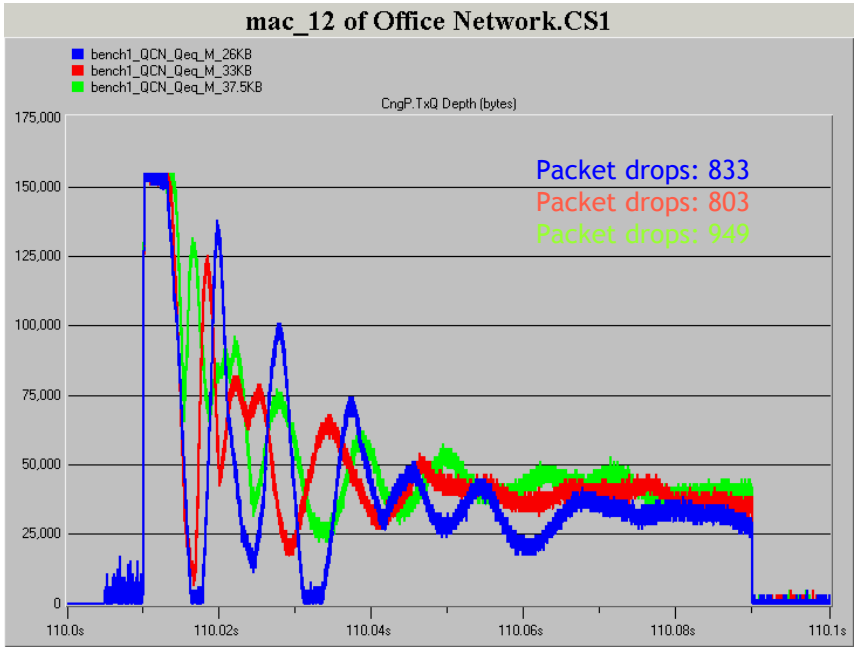
- Qeq scales the quantization of Fb for a given congestion and max queue size.
- As Qeq ↑, the negative transient slope ↓ and the queue remains full longer.
- To remove the impact of Qeq on Fb quantization, Max_Fb could be set to a constant. The Max_Fb formula above is approximately the size of the egress buffer.
- Using Max_Fb = M may avoid any Qeq tuning to improve negative transient performance.

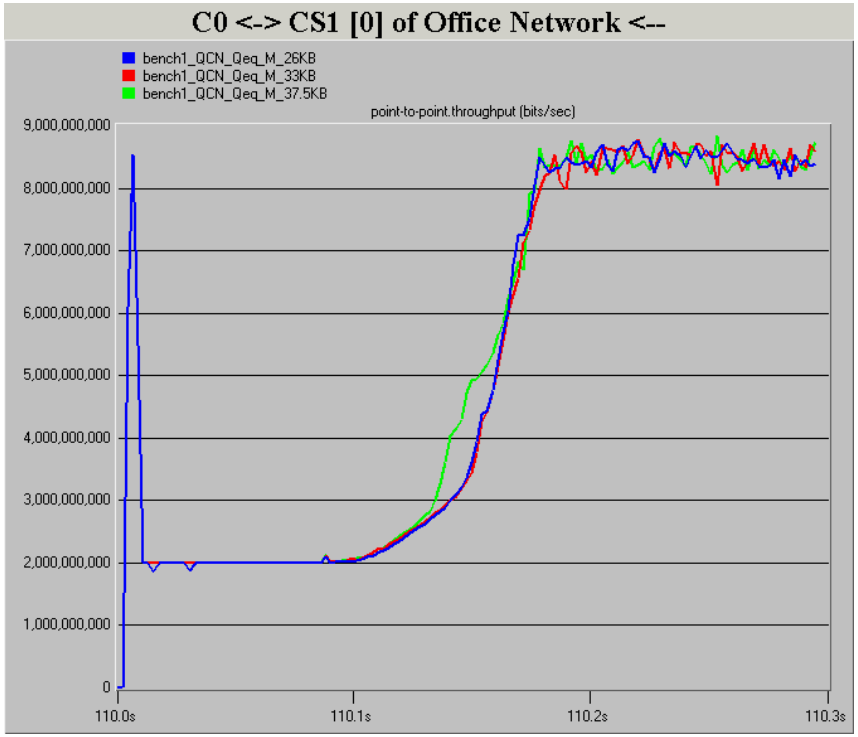
Experiment:

Use Max_Fb = M and repeat benchmark 1 to compare the queue depth and throughput.

Queue Depth and Hotspot Throughput

Benchmark 1, Qeq = 26, 33, 37.5 KB, Max_Fb = M





Next Goal

- Present required benchmarks for QCN using Opnet
- Study effects of varying Qeq
- **Study effects of more severe congestion**

Topology & Workload
Benchmark 3

OG Hotspot, Multi Hop: Selected Victims

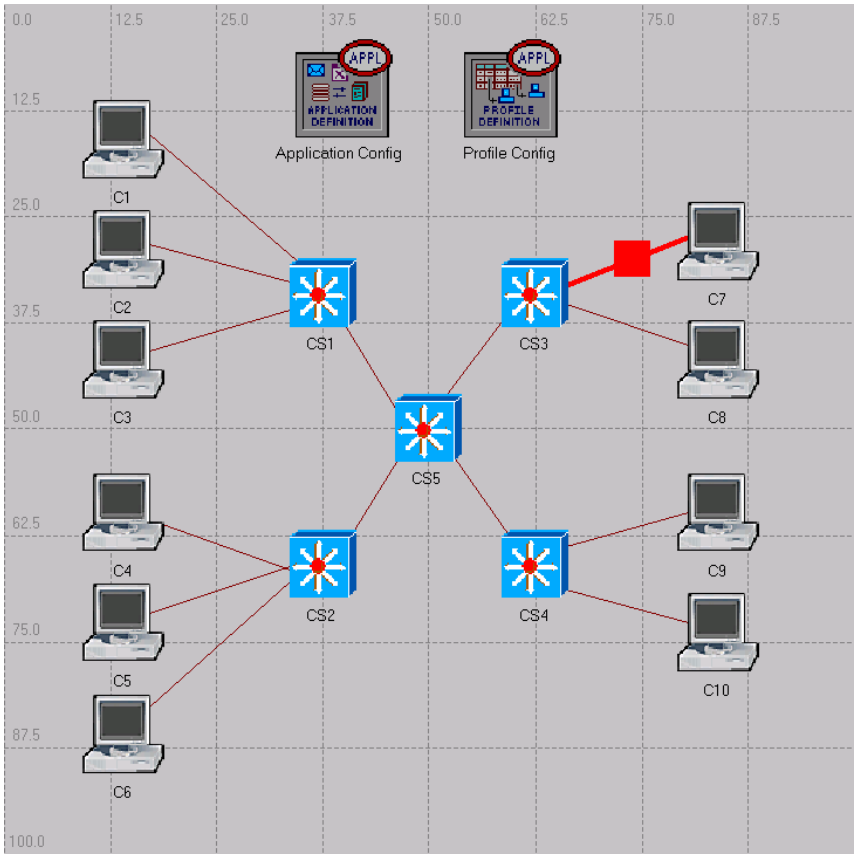
100 Mbps OG hotspot for 80ms @ CS3→C7

802.3x PAUSE enabled

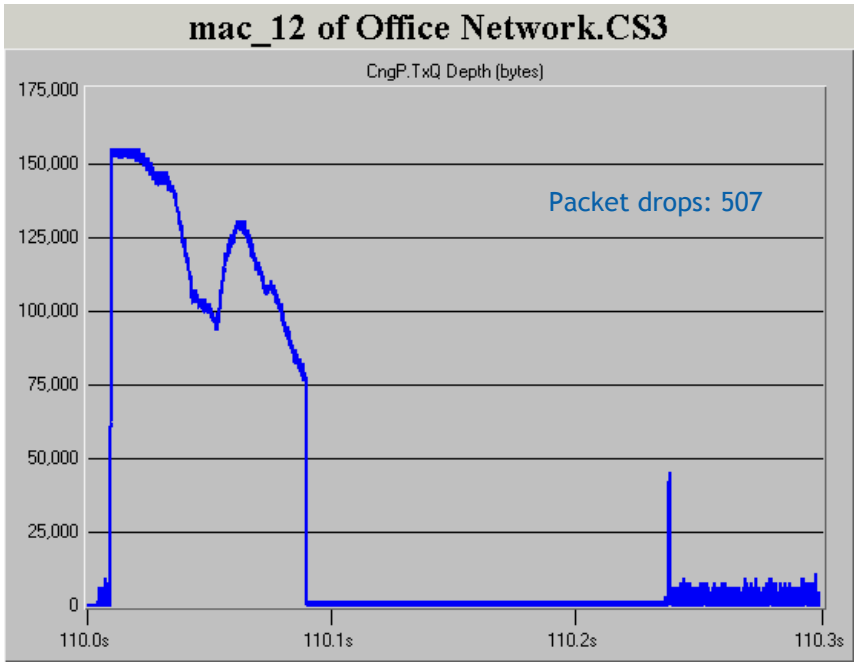
Fair allocation provides 50Mb/s to culprit flows and 7Gb/s to victim flows

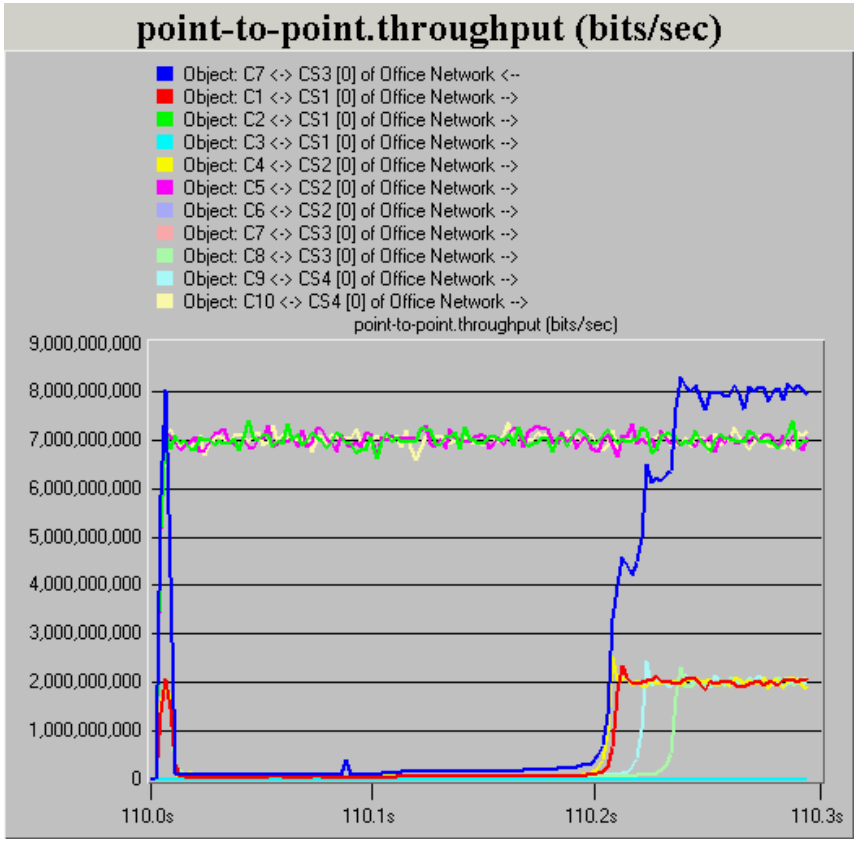
Traffic pattern:

- 10 Gbps links, 500ns link latency
- 4 culprit flows: 1, 4, 8, 9 @ 70% → 7
- 3 victim flows: 2 → 9, 5 → 3, 10 → 6 @ 20%
- Hosts 3, 6, 7 are only receiving
- Temporally Bernoulli



Queue Depth and Hotspot Throughput
Benchmark 3, 1% service rate





Next Steps

- Discuss simulation goals of each team.
  - Identify complimentary tasks.
- Heterogeneous link speeds (1/10/100 Gbps)
- TCP


intel
Leap ahead™