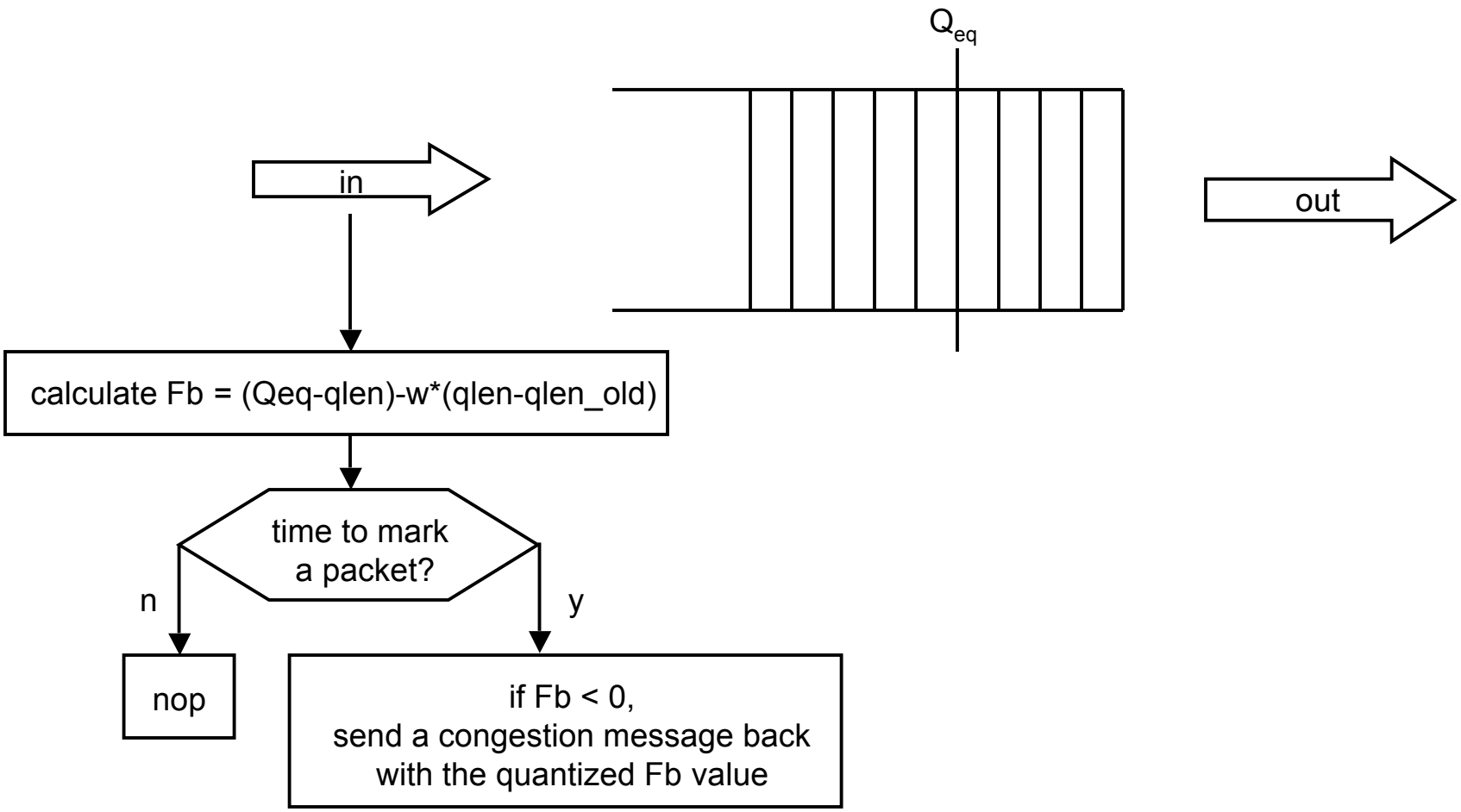
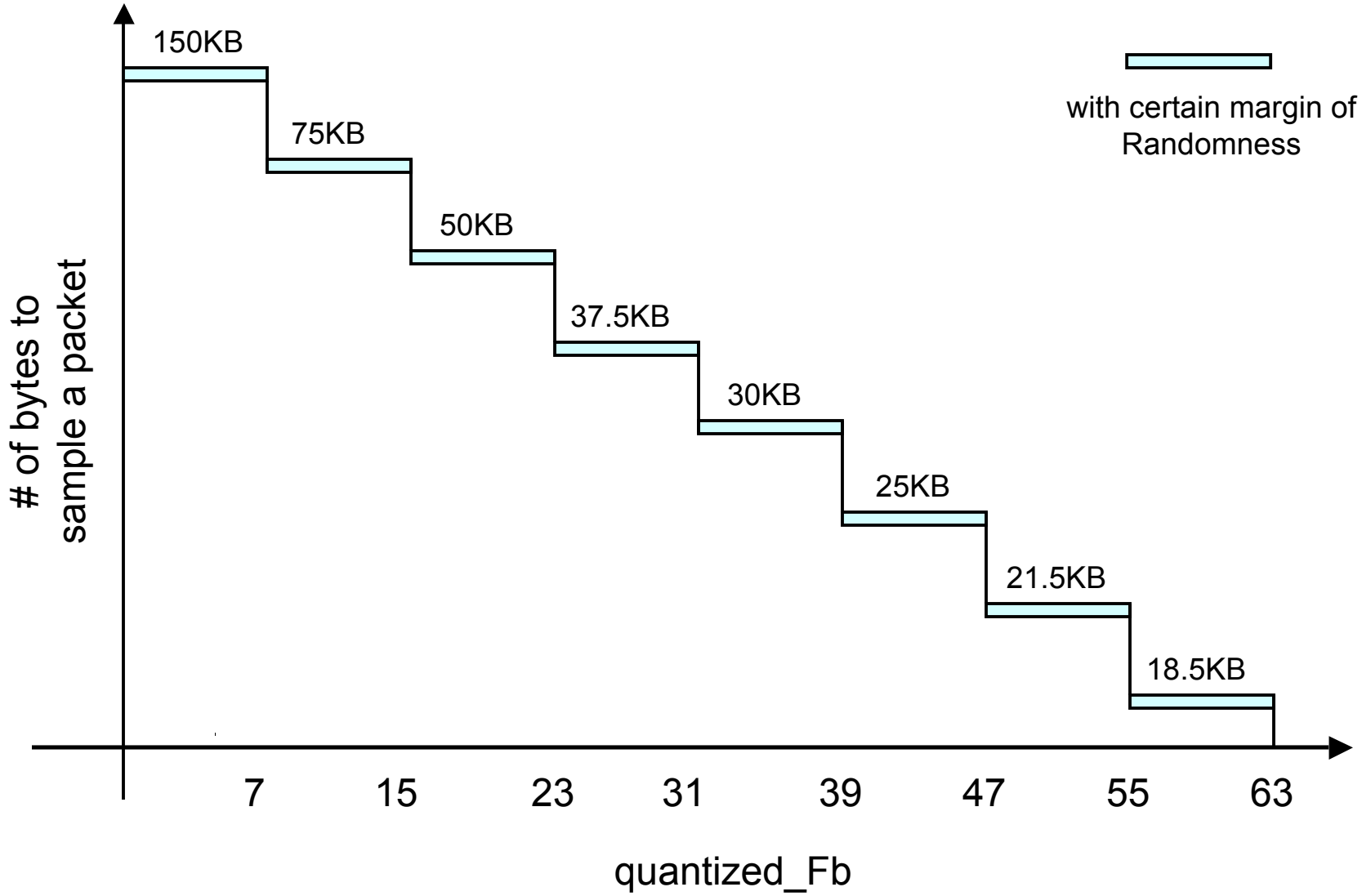
QCN: Description and Benchmark Simulations

Berk Atikoglu, Abdul Kabbani,
Rong Pan, Balaji Prabhakar

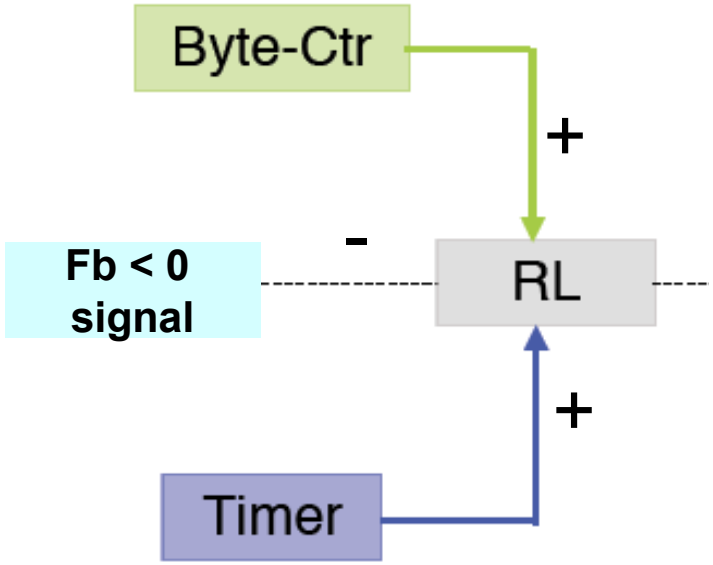
QCN (A Brief Review)
- Congestion Point



How To Sample & Mark A Packet?



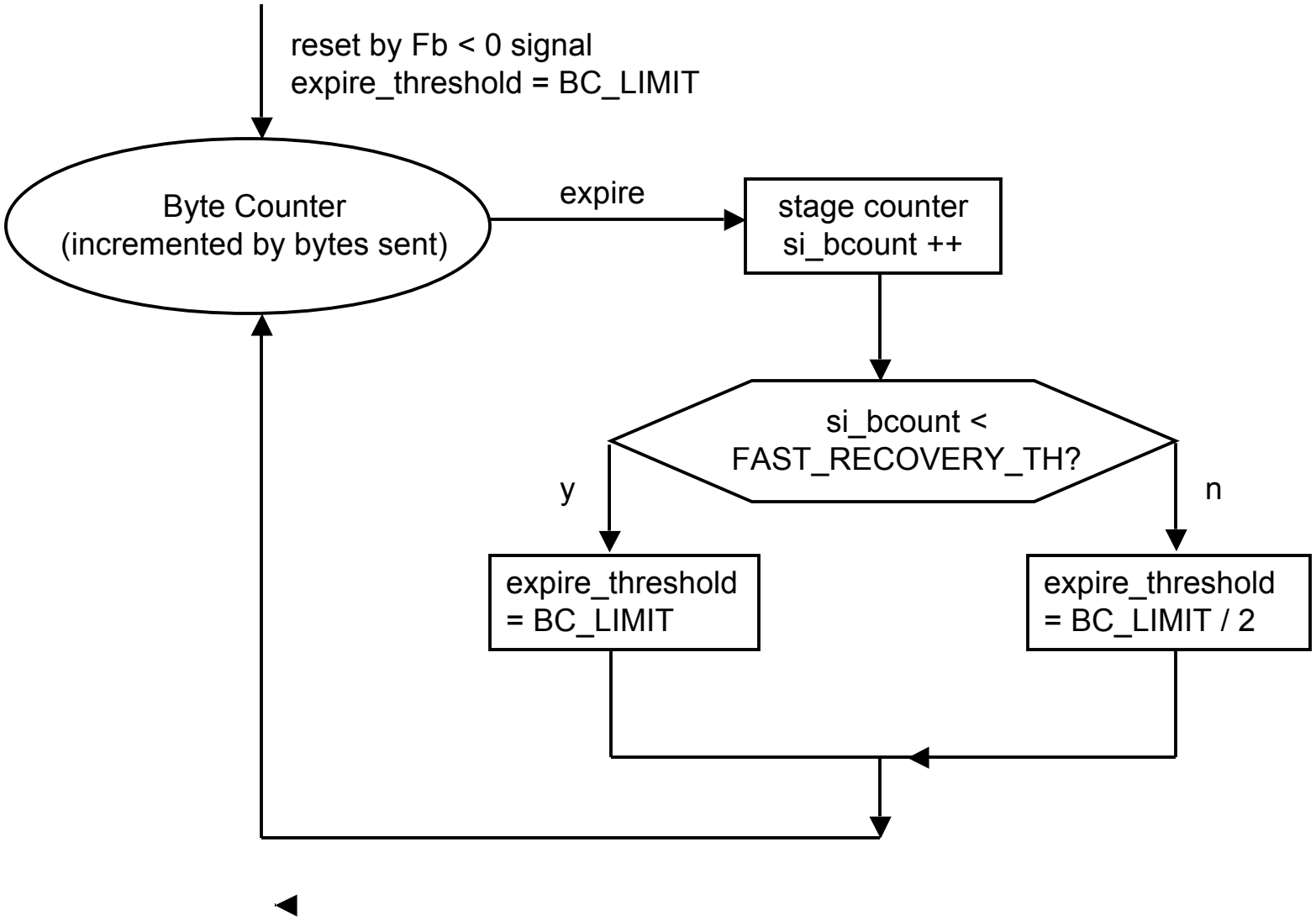
QCN (A Brief Review)
- Reaction Point



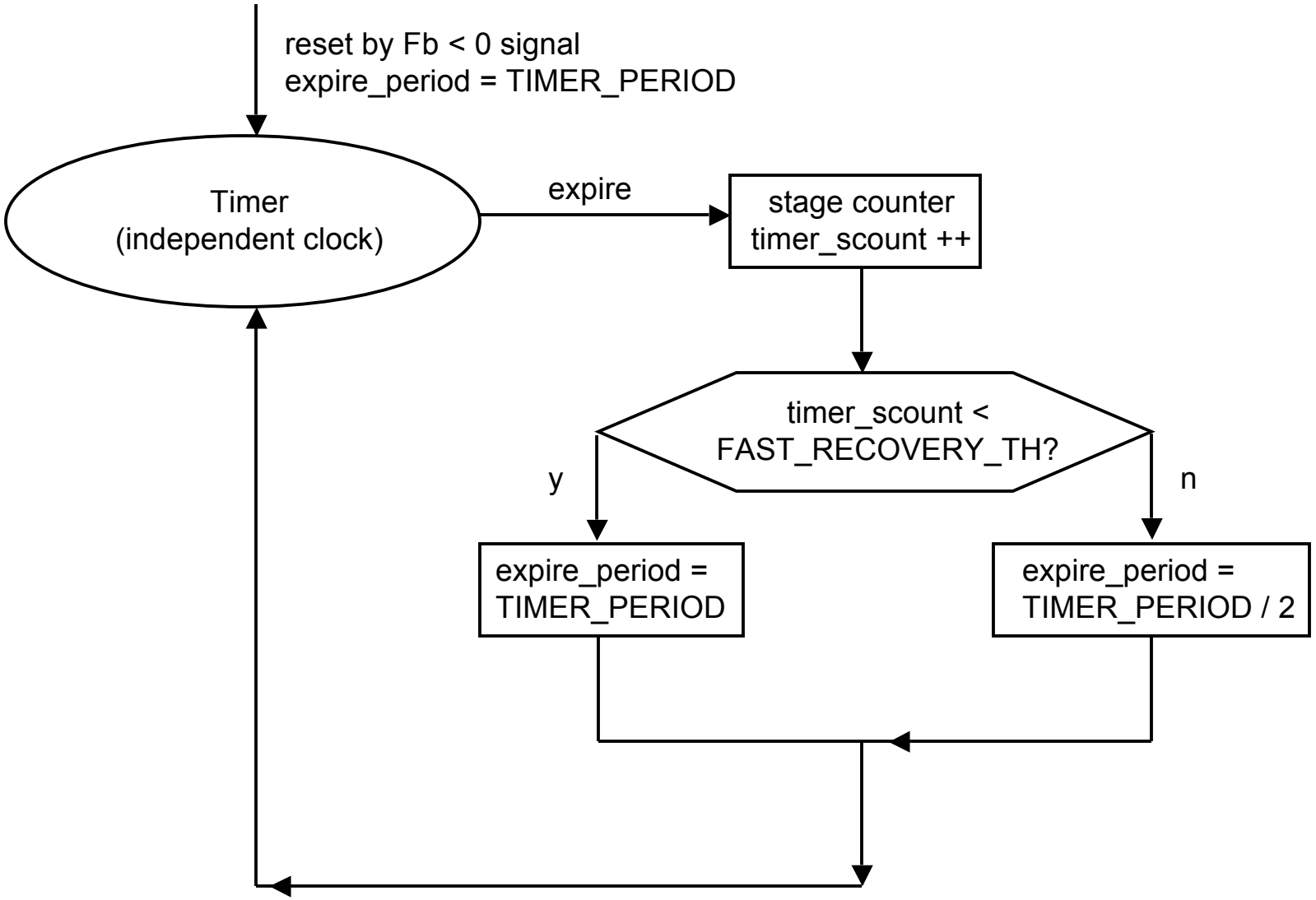
Two counters: byte-counter and timer cycle through independently
Both reset by Fb < 0 signal

- Byte-Counter
  - 5 cycles of FR (150KB/cycle)
  - AI cycles afterwards (75KB/cycle)
  - Fb < 0 sends byte-counter to FR
- Timer
  - 5 cycles of FR (TIMER_PERIOD msec/cycle)
  - AI cycles afterwards (TIMER_PERIOD/2 msec/cycle)
  - Fb < 0 sends timer to FR
- Rate Limter (RL)
  - In FR if both byte-counter and timer's stage counter is less than FAST_RECOVERY_TH
  - In AI if only one of byte_counter or timer 's stage counter is not less than FAST_RECOVERY_TH
  - In HAI if both byte_counter and timer 's stage counter are not less than FAST_RECOVERY_TH

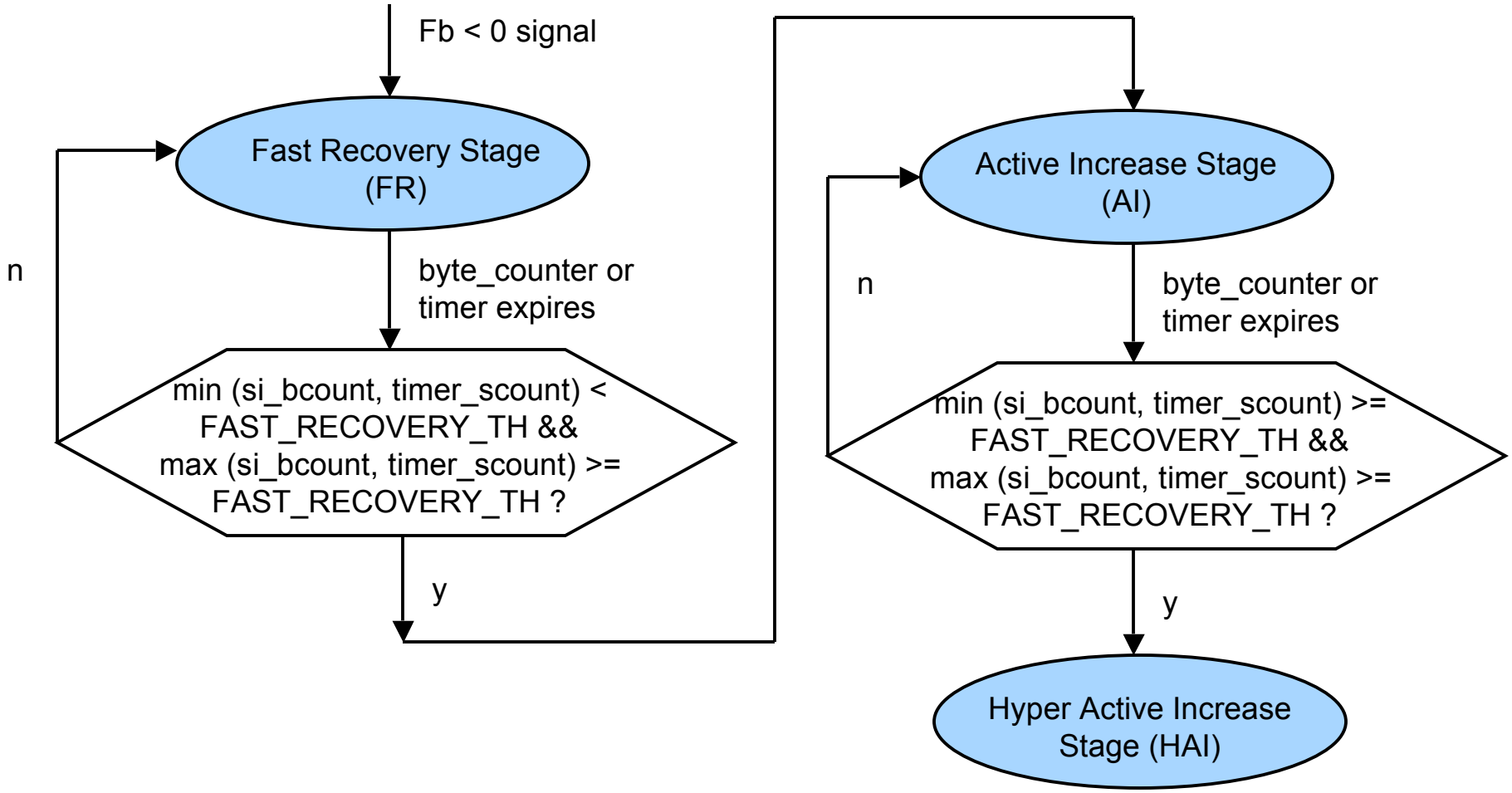
Byte Counter



Timer


reset by Fb < 0 signal
expire_period = TIMER_PERIOD
Timer
(independent clock)
expire
stage counter
timer_scount ++
timer_scount <
FAST_RECOVERY_TH?
y
n
expire_period =
TIMER_PERIOD
expire_period =
TIMER_PERIOD / 2

Rate Limiter State Diagram



Rate Changes
- Upon a Fb < 0 message

- TR is decreased implicitly
  - if si_bcount != 0
    - TR = CR

- CR = CR * (1 – Gd*|Fb|)

Rate Changes
- Upon byte_counter or timer expires

FR

- if the first cycle of byte-counter/timer expires and TR > 10*CR: TR = TR/8
- CR = (CR+TR)/2

AI

-TR = TR+R_AI; CR = (CR+TR)/2

HAI, event numbered i = 1, 2, …

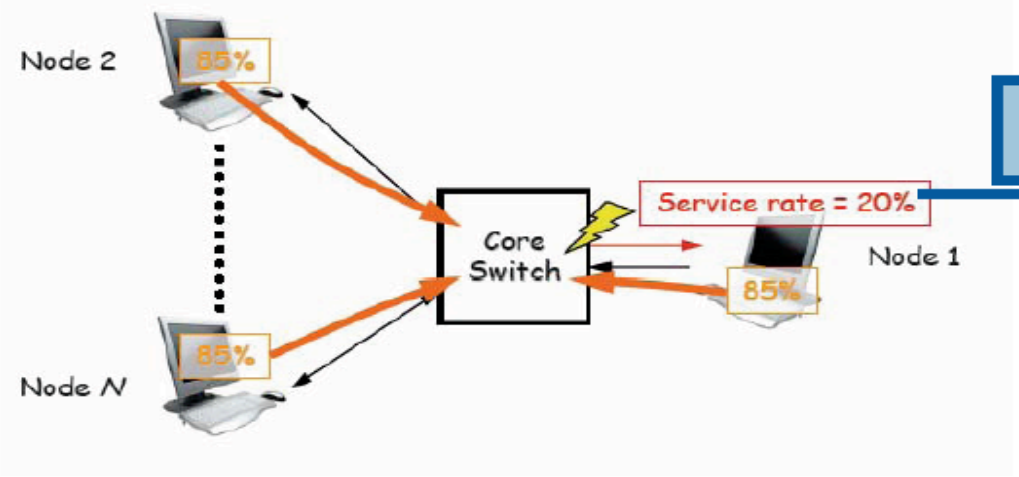
- At the end of event number i:
  TR = TR + (i*R_HAI); CR = (CR+TR)/2

Simulation Parameters

- Traffic
  - Uniform destination distribution (to all nodes except self)
  - Fixed frame size = 1500 B
- Switch
  - VOQ with 2.4MB shared memory
  - Partitioned memory per input, shared among all outputs
  - No limit on per-output memory usage
- Adapter
  - RLT: VOQ and single; RR service
  - One rate limiter per destination
  - Egress buffer size = 1500 KB,
  - Ingress buffer size = Unlimited
- QCN
  - W = 2.0
  - Q_EQ = 33 KB
  - GD = 0.0078125
  - Base marking: once every 150 KB
  - Margin of randomness: 30%
  - R_{unit} = 1 Mb/s
  - MIN_RATE = 10 Mb/s
  - BC_LIMIT = 150 KB
  - TIMER_PERIOD = 15 ms
  - R_AI = 5 Mbps
  - R_HAI = 50Mbps
  - FAST_RECOVERY_TH = 5
  - Quantized_Fb: 6 bits

Baseline #1

1. Output Generated Hot Spot Single Hop



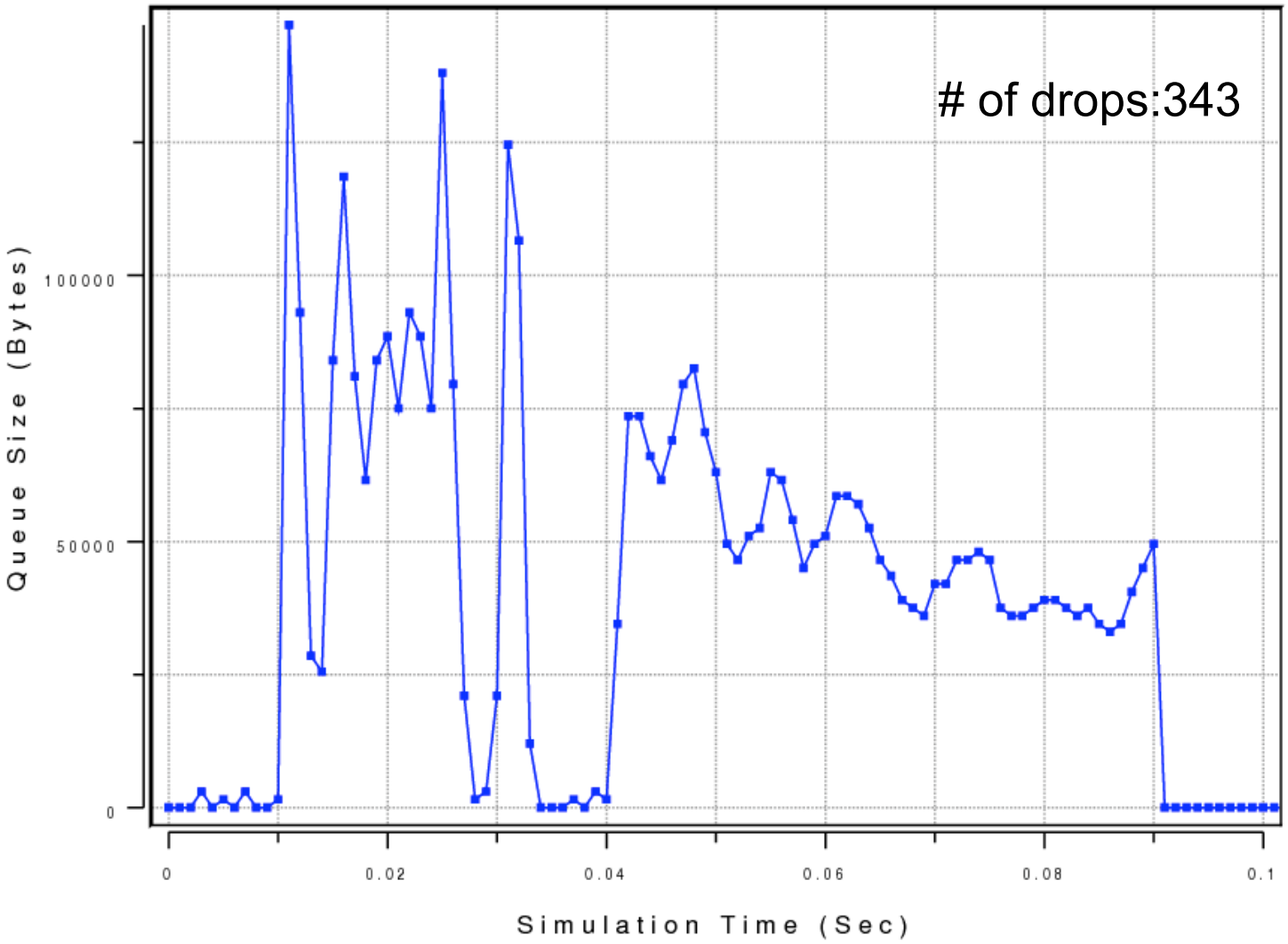
Workload:

- All Nodes (10) : Uniform Distribution, load = 8.5Gbps
- Node 1 Service Rate = 1Gbps
- One Congestion Point
  - Hotspot:
    - Degree: 9, Severity = 8.5:1,
    - Duration: 80 mS from ti=10 to 90 mS

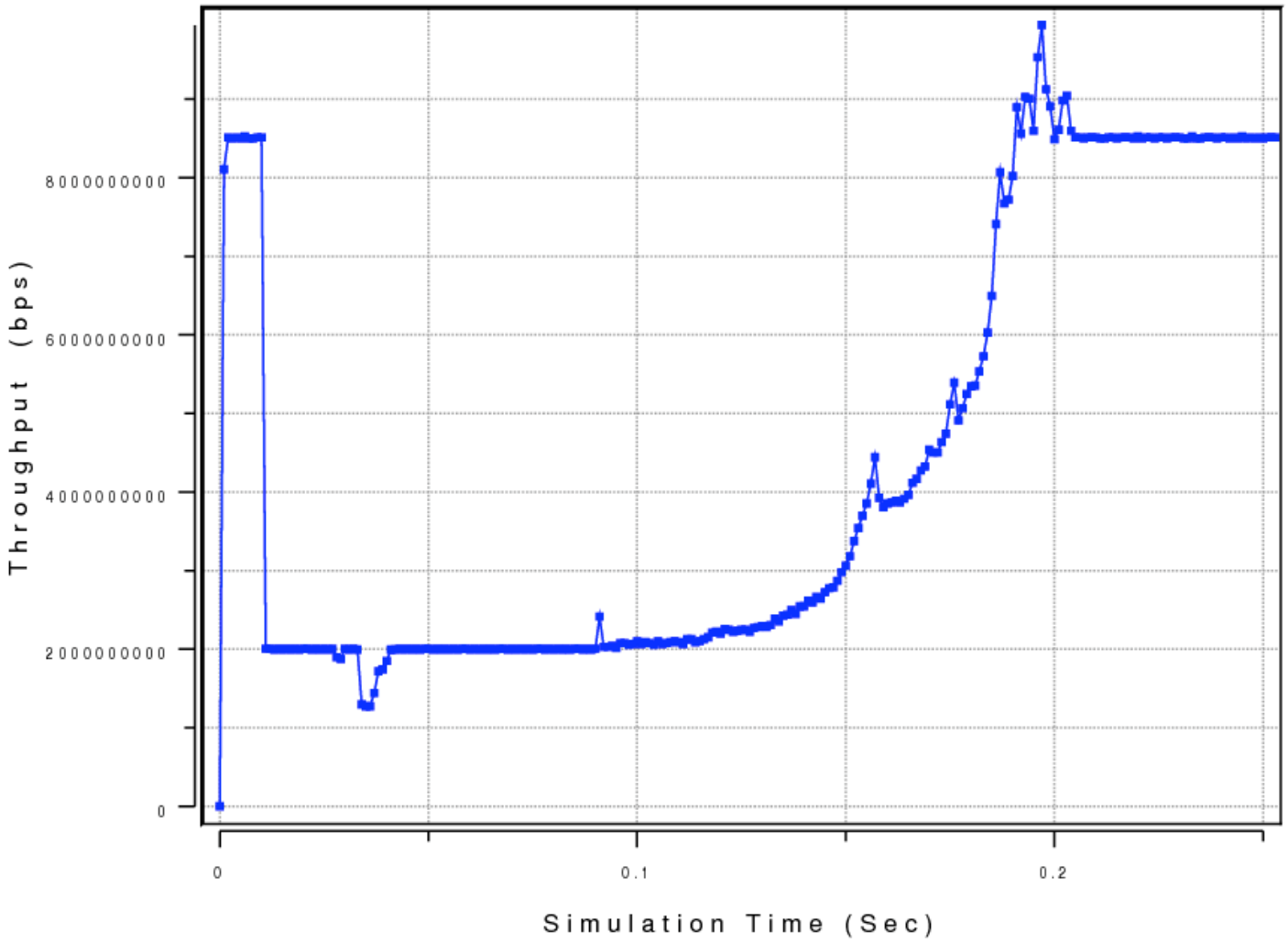
- Scenarios: 2Gbps, 1Gbps, 0.5Gbps OG service rates

Required

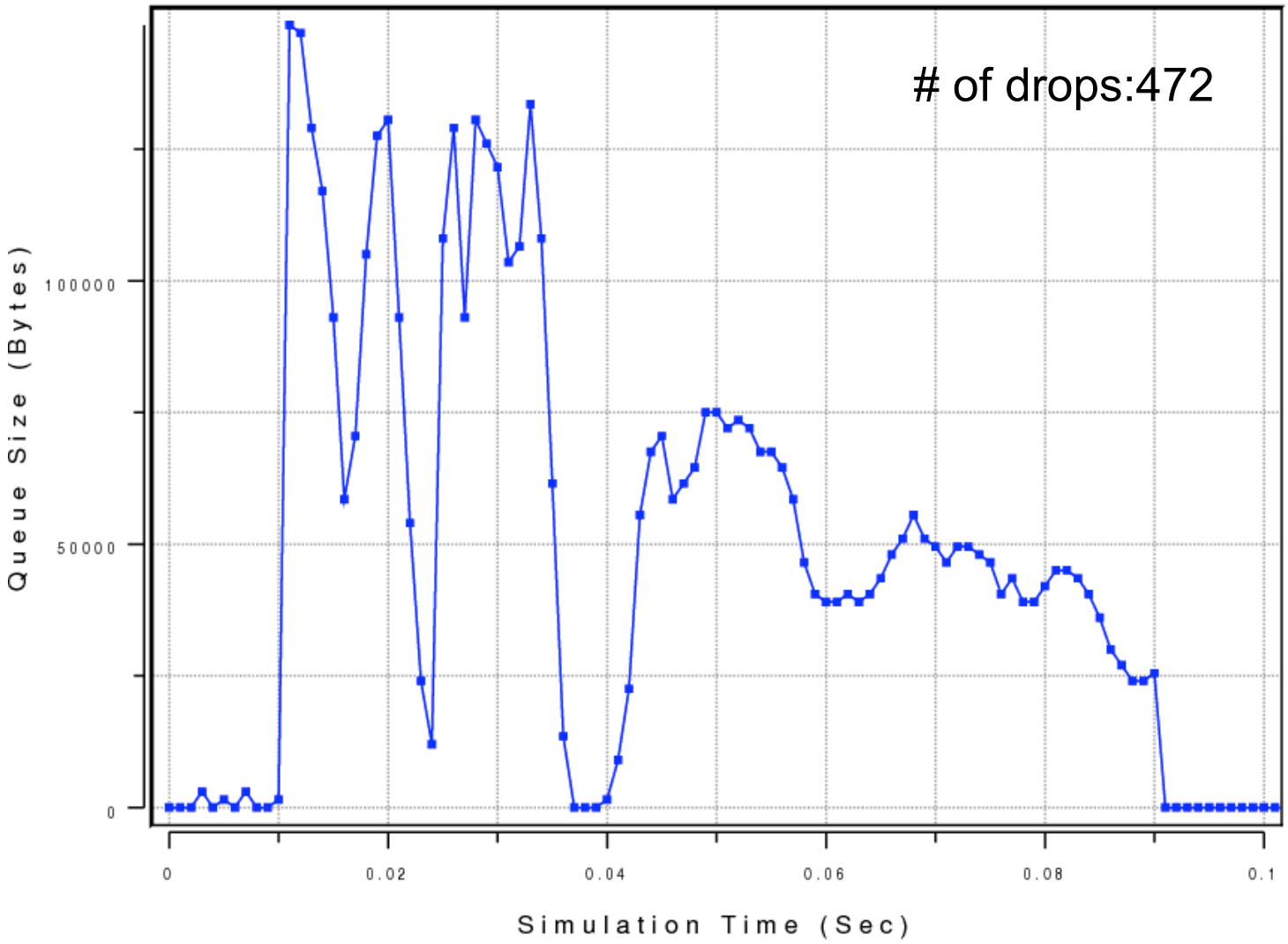
Service Rate: 2.0Gbps
- Queue Size



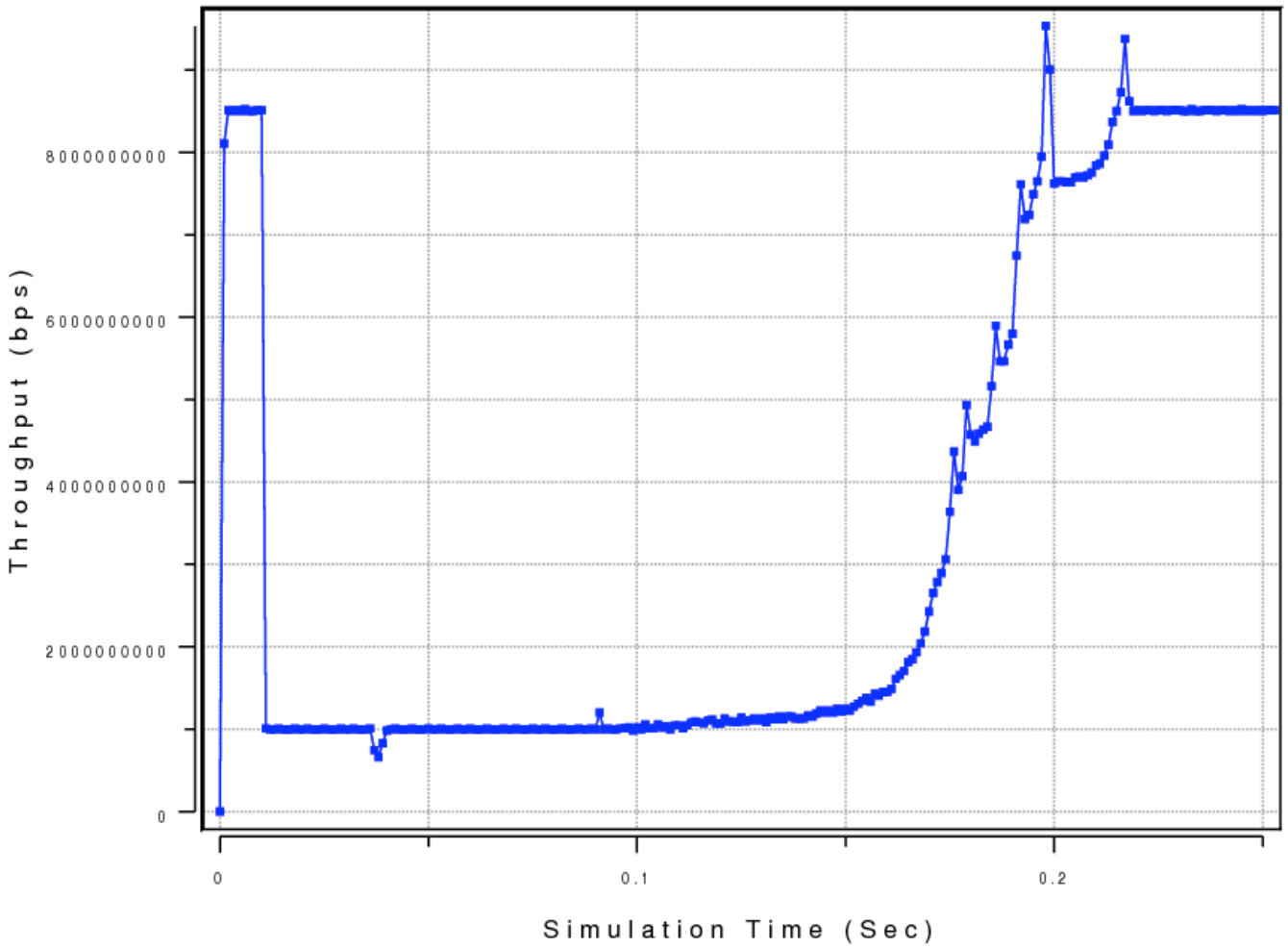
Service Rate: 2.0Gbps - Throughput



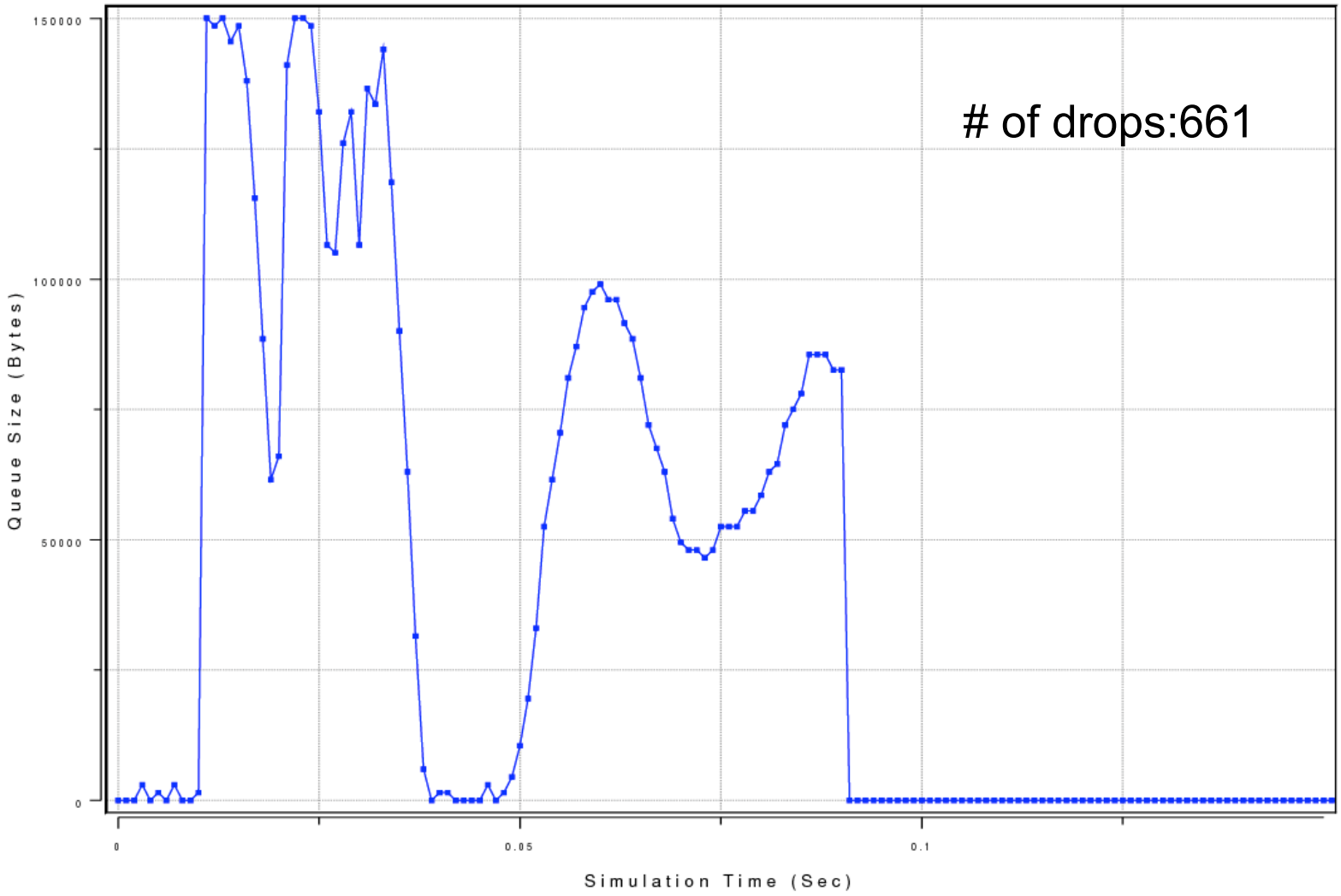
Service Rate: 1.0Gbps
- Queue Size



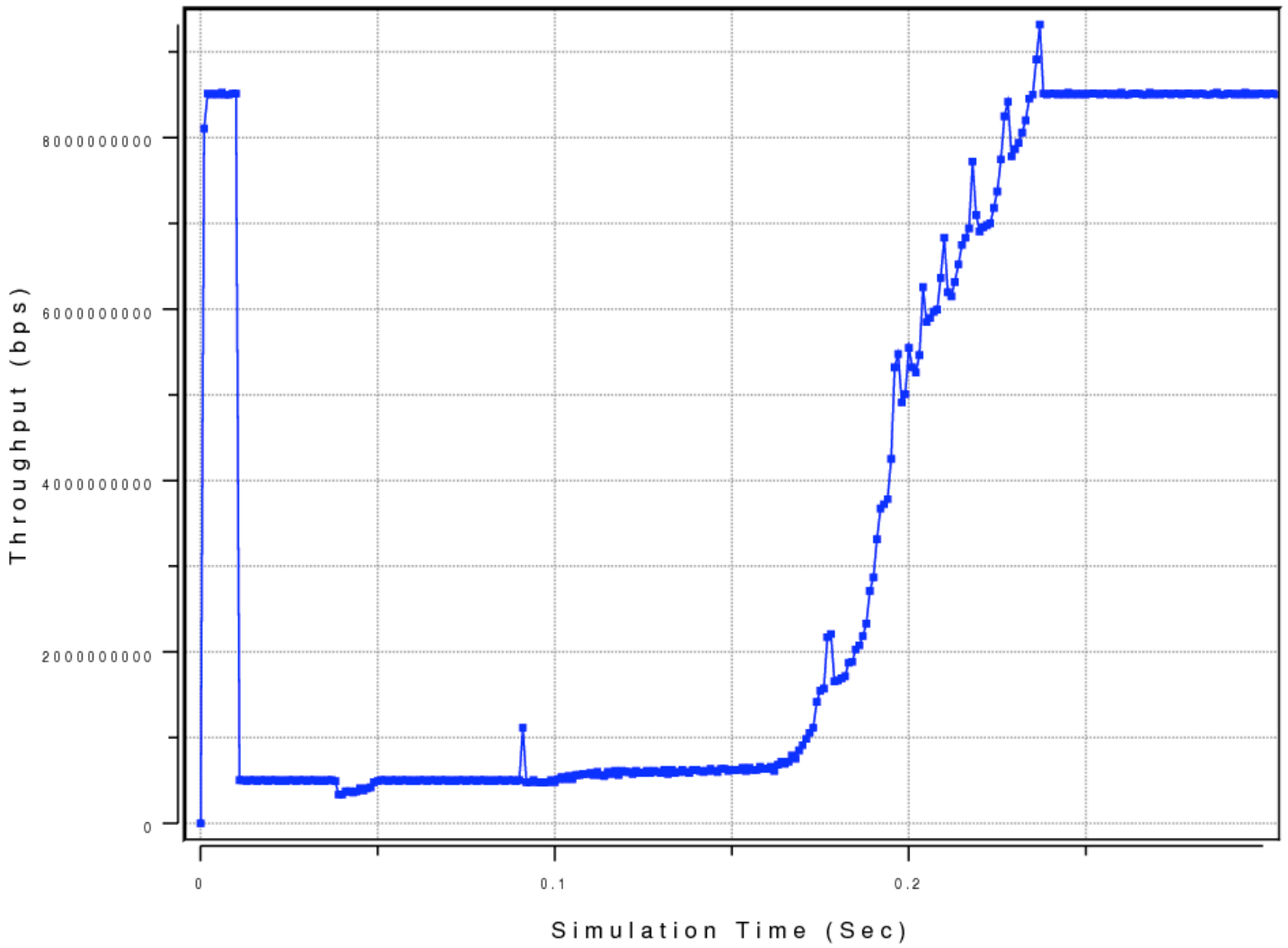
Service Rate: 1.0Gbps - Throughput



Service Rate: 0.5Gbps
- Queue Size

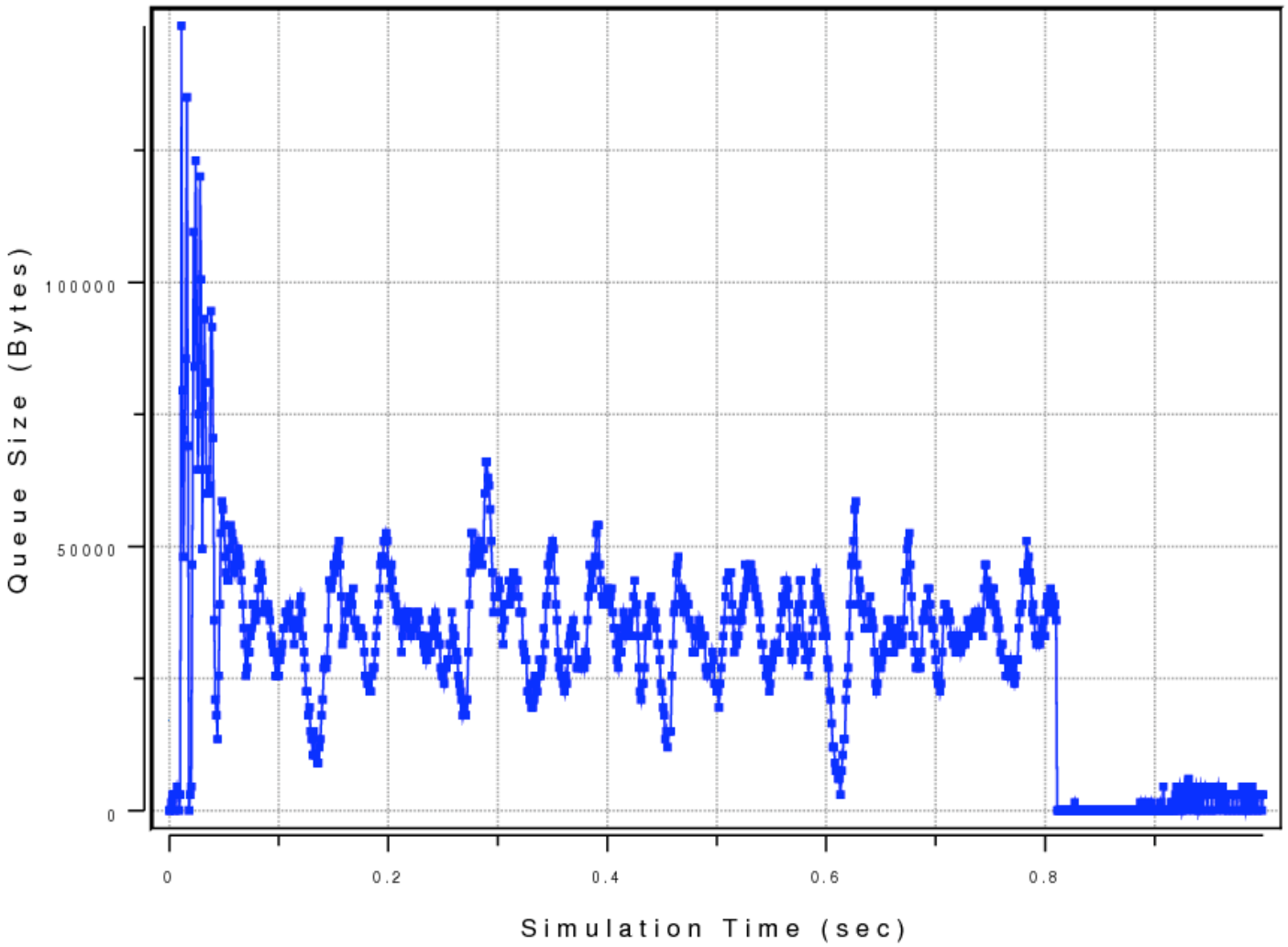


Service Rate: 0.5Gbps - Throughput

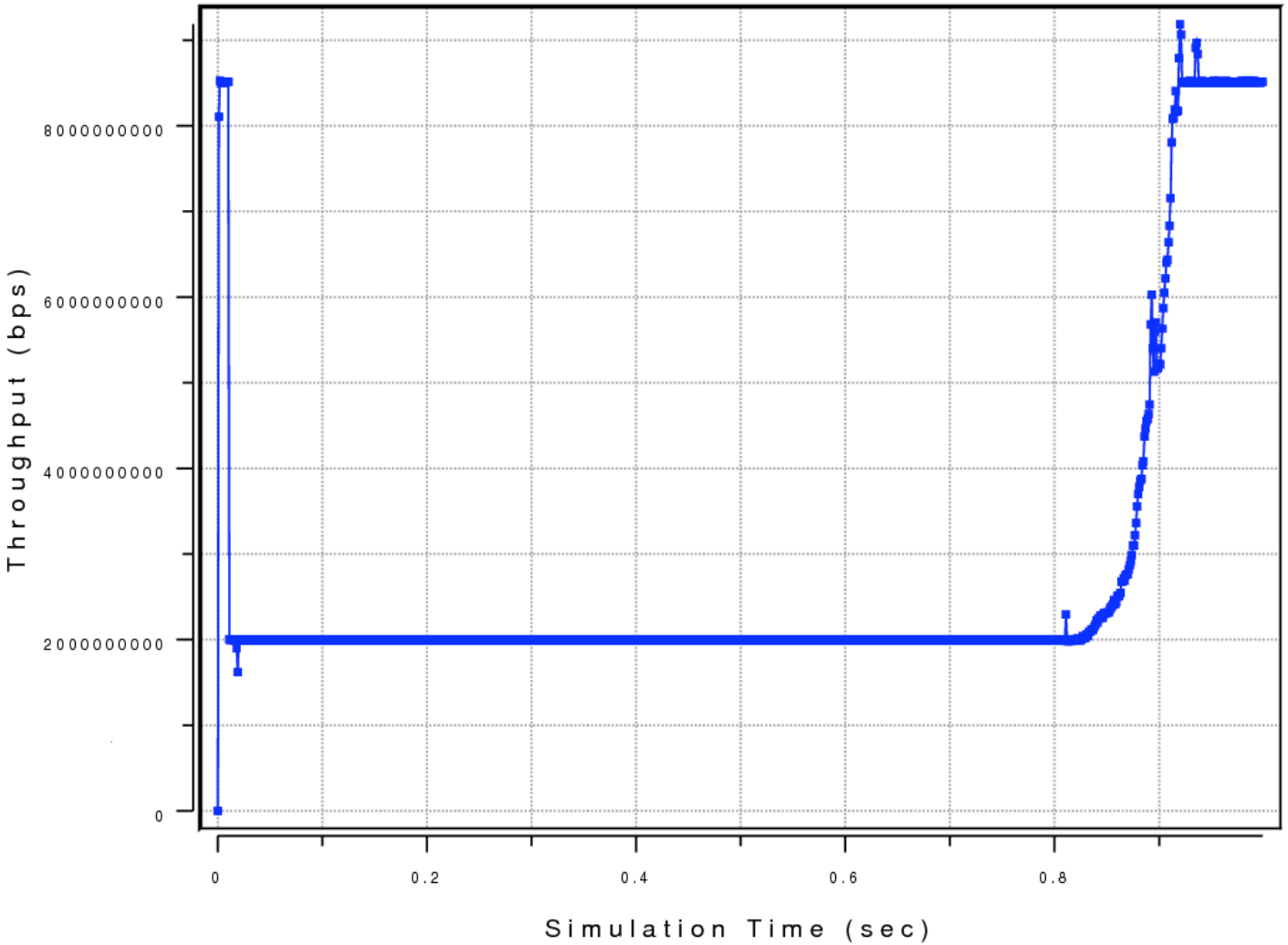


Extending The Hot Spot Duration - To Show Stability

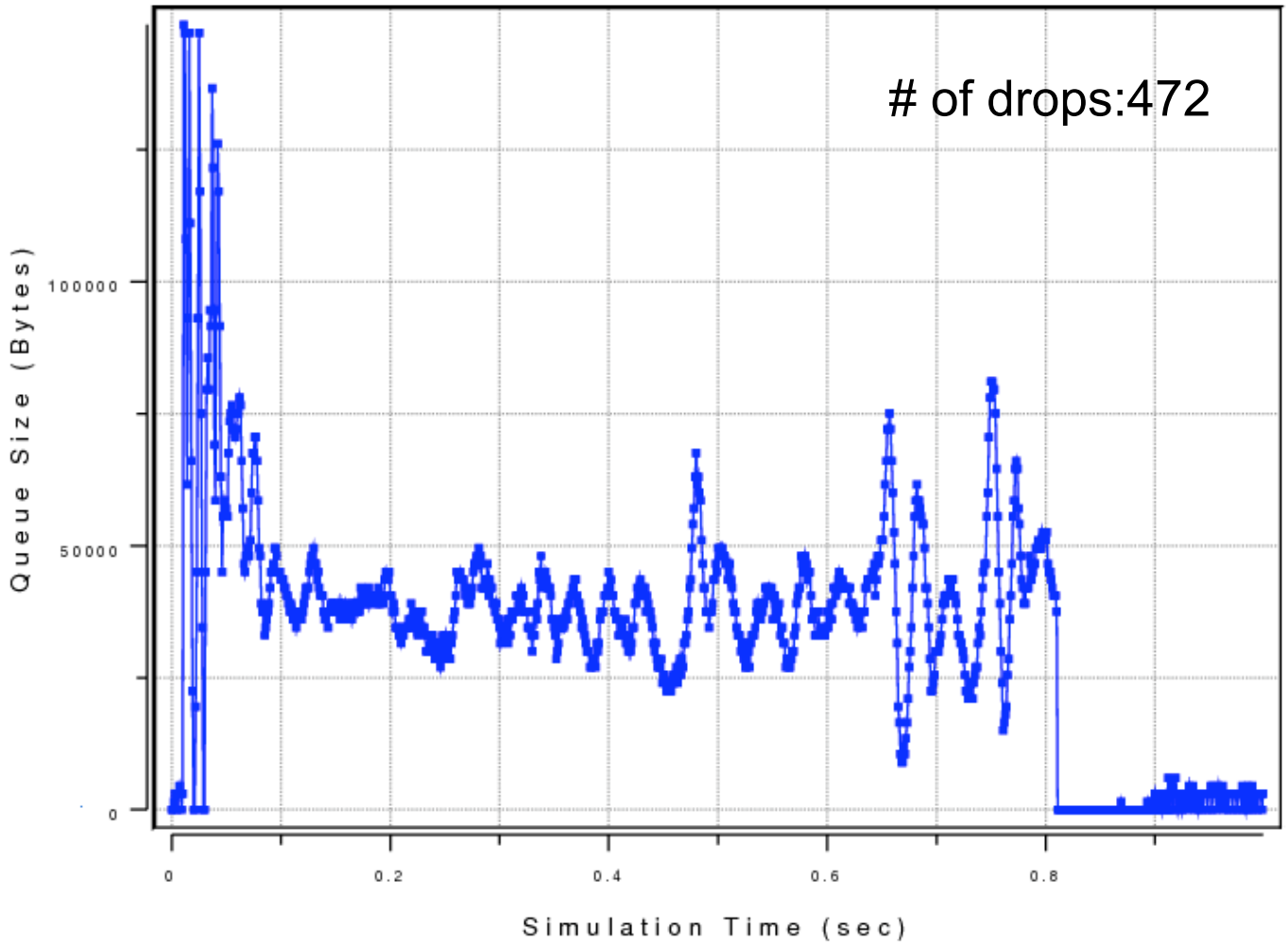
Service Rate: 2.0Gbps
- Queue Size



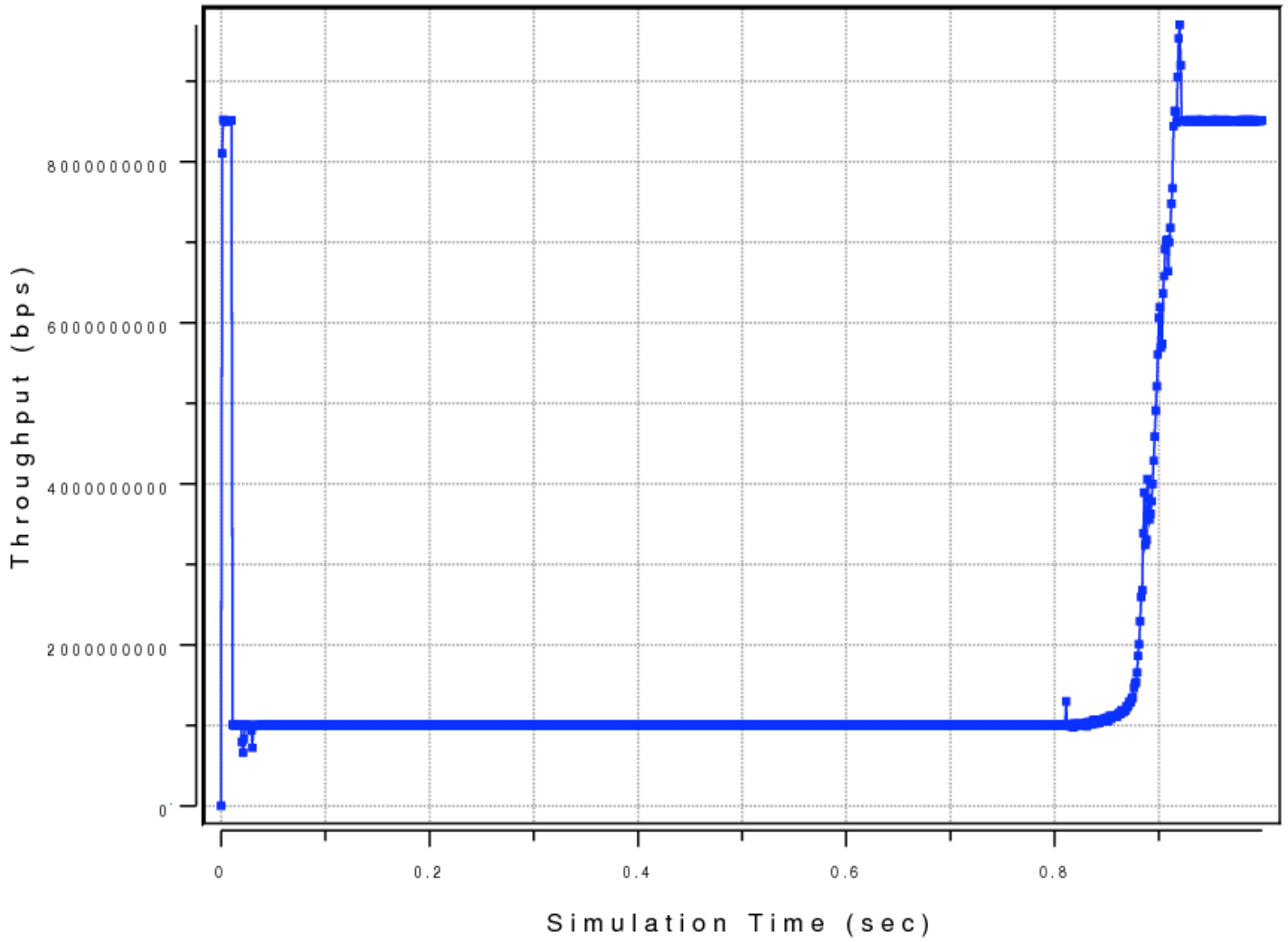
Service Rate: 2.0Gbps - Throughput



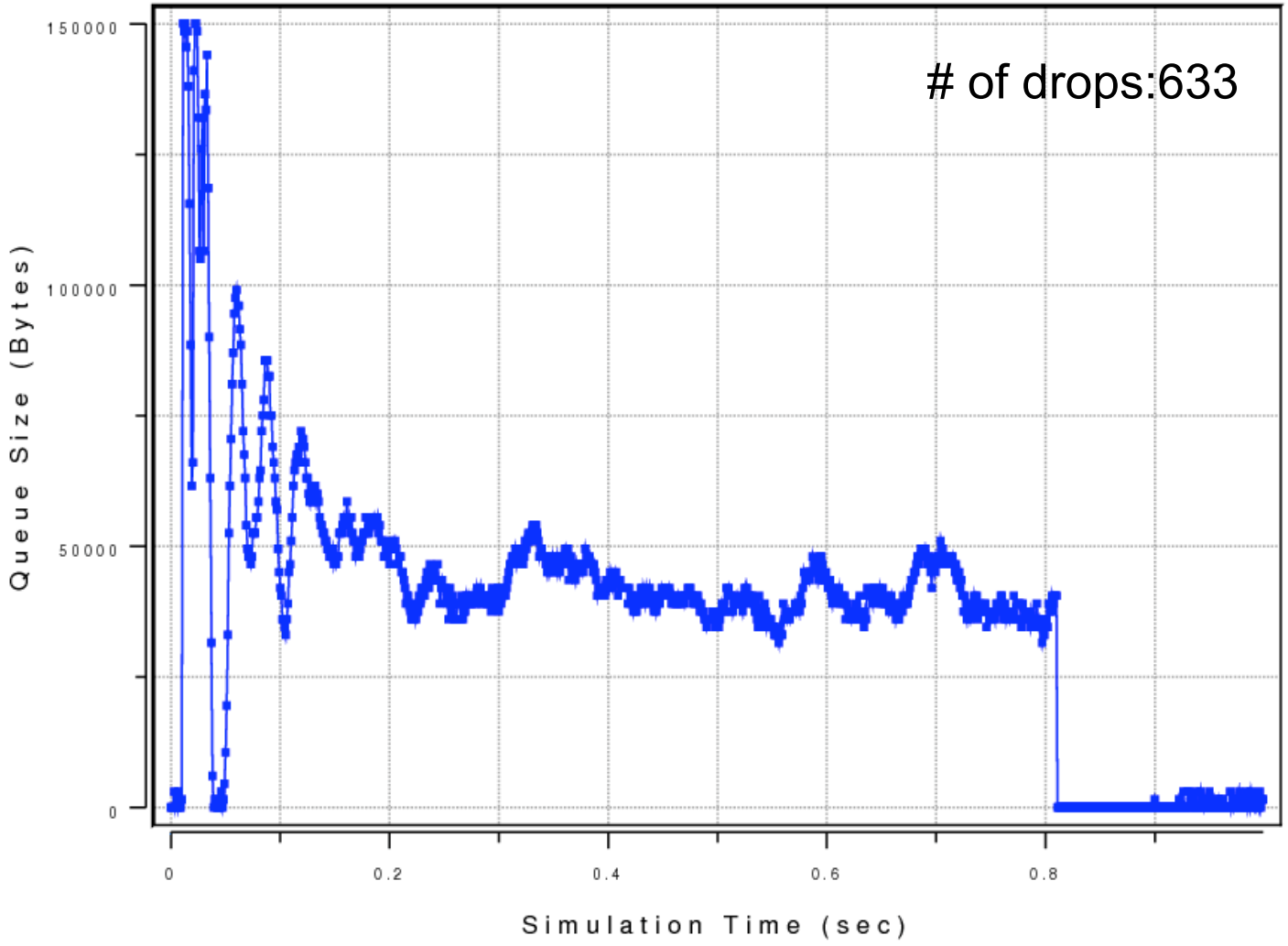
Service Rate: 1.0Gbps
- Queue Size



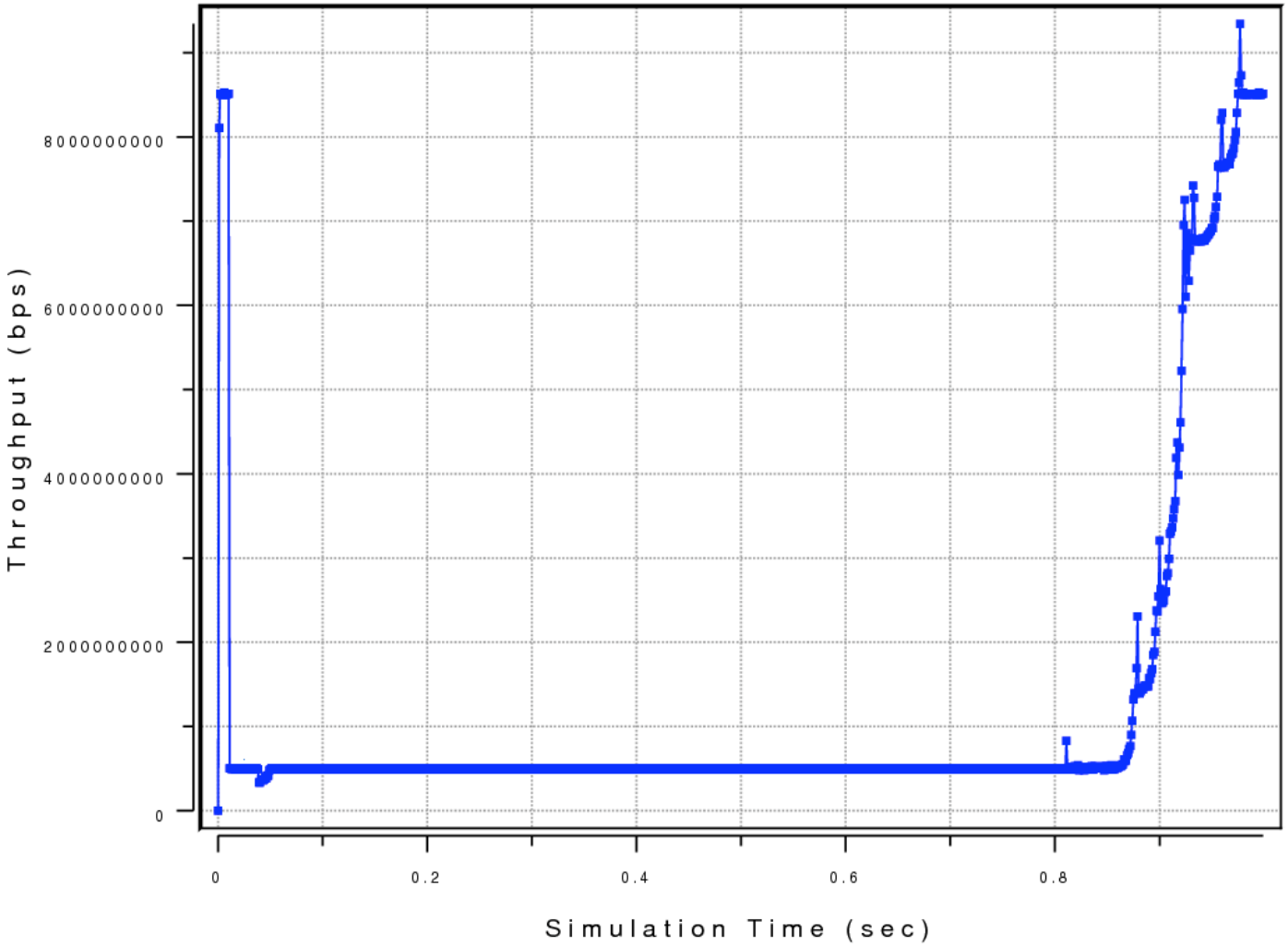
Service Rate: 1.0Gbps - Throughput



Service Rate: 0.5Gbps - Queue Size

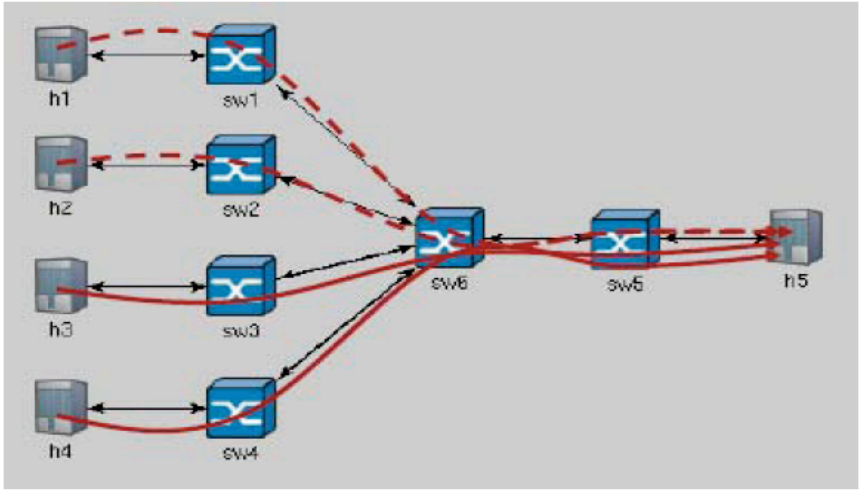


Service Rate: 0.5Gbps - Throughput



Baseline #5

5. Symmetric Topology Single HS – Bursty



Workload:

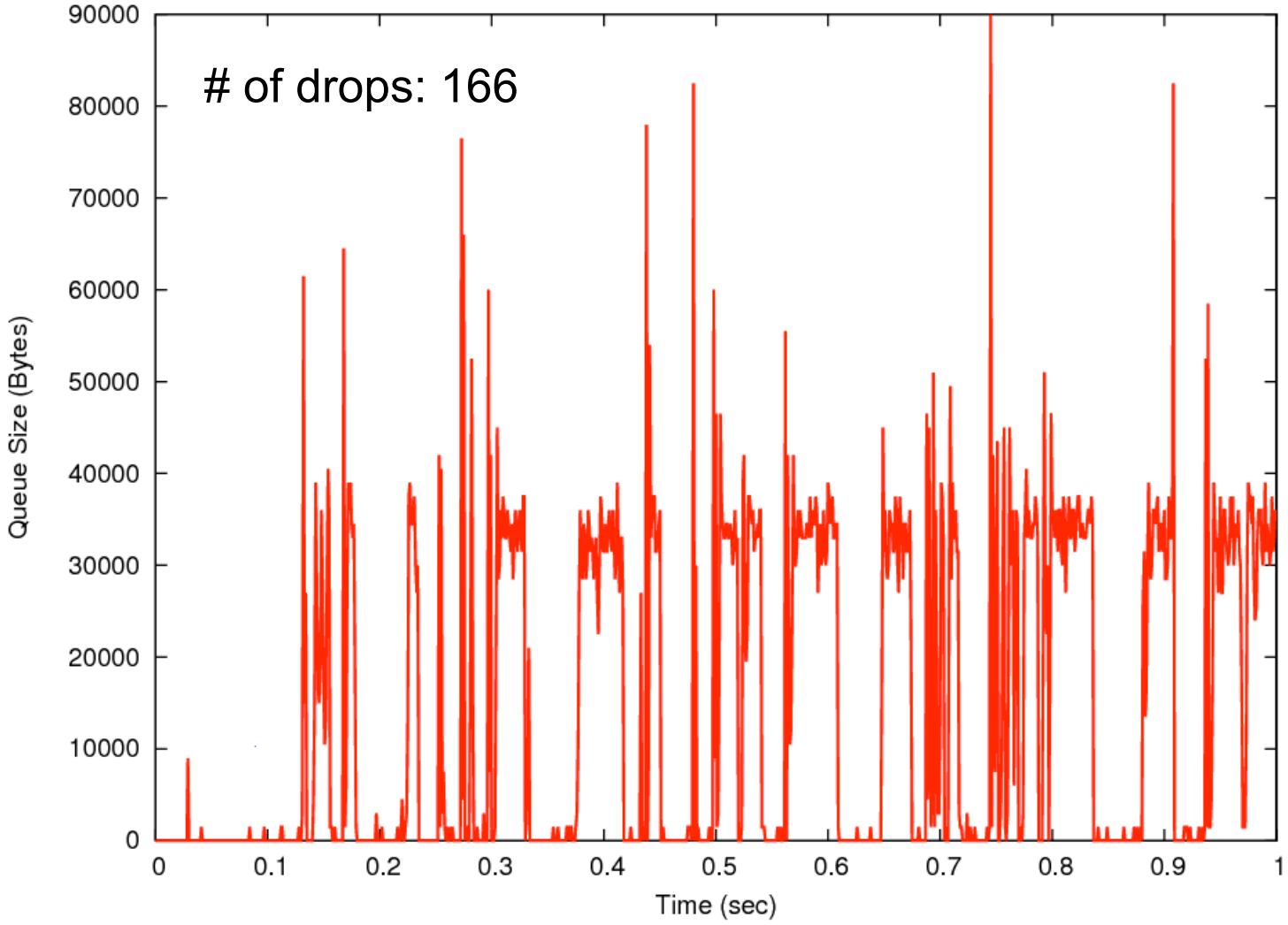
- Point-to-point from h1-4 to h5
- Load: 100%
- H1 and H2 on-off sources (Ton = Toff = 20 ms)
- On/Off period exponential distribution

Scenarios:

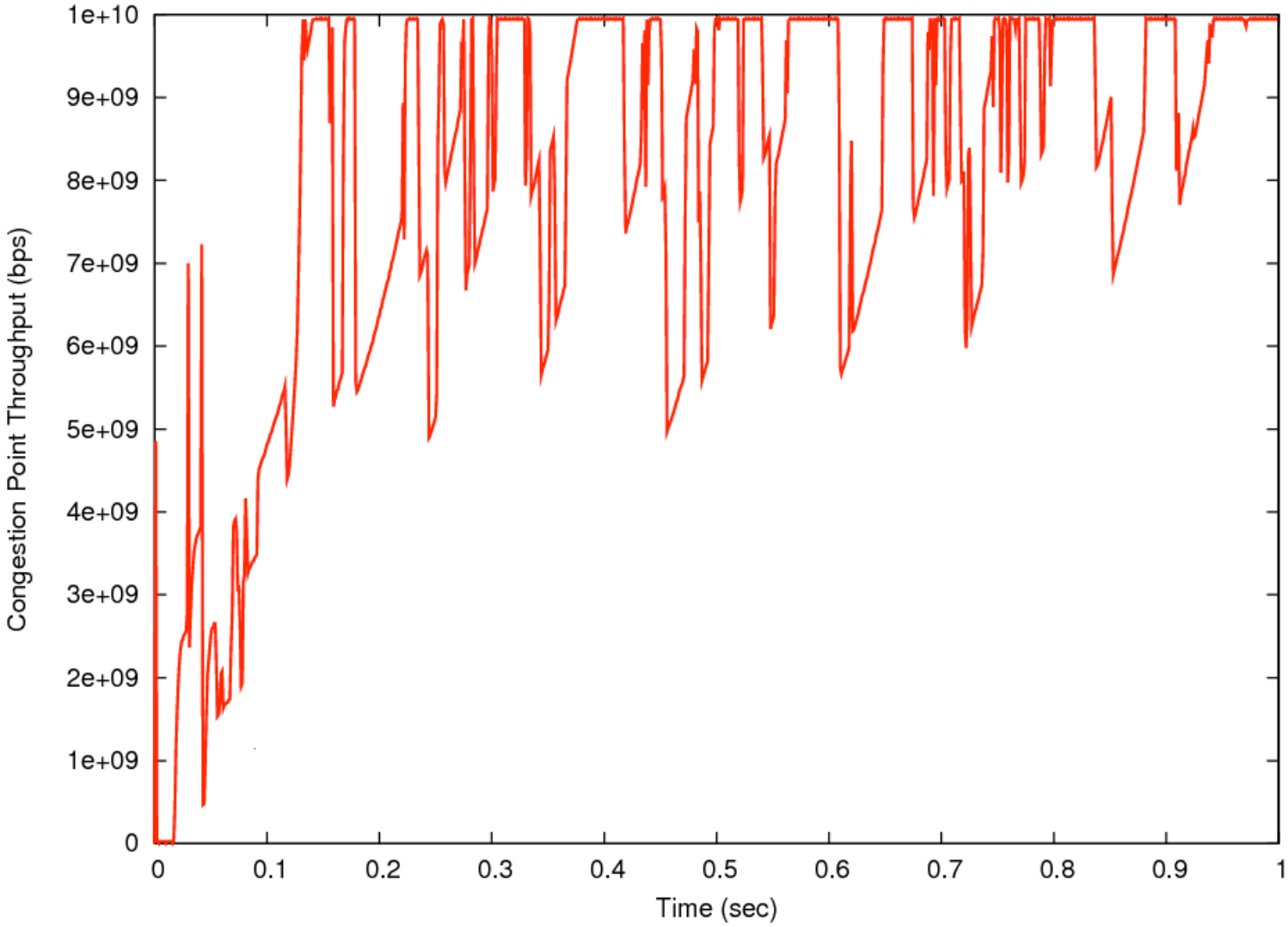
- Burst periods: 20, 10, 5mS

Required

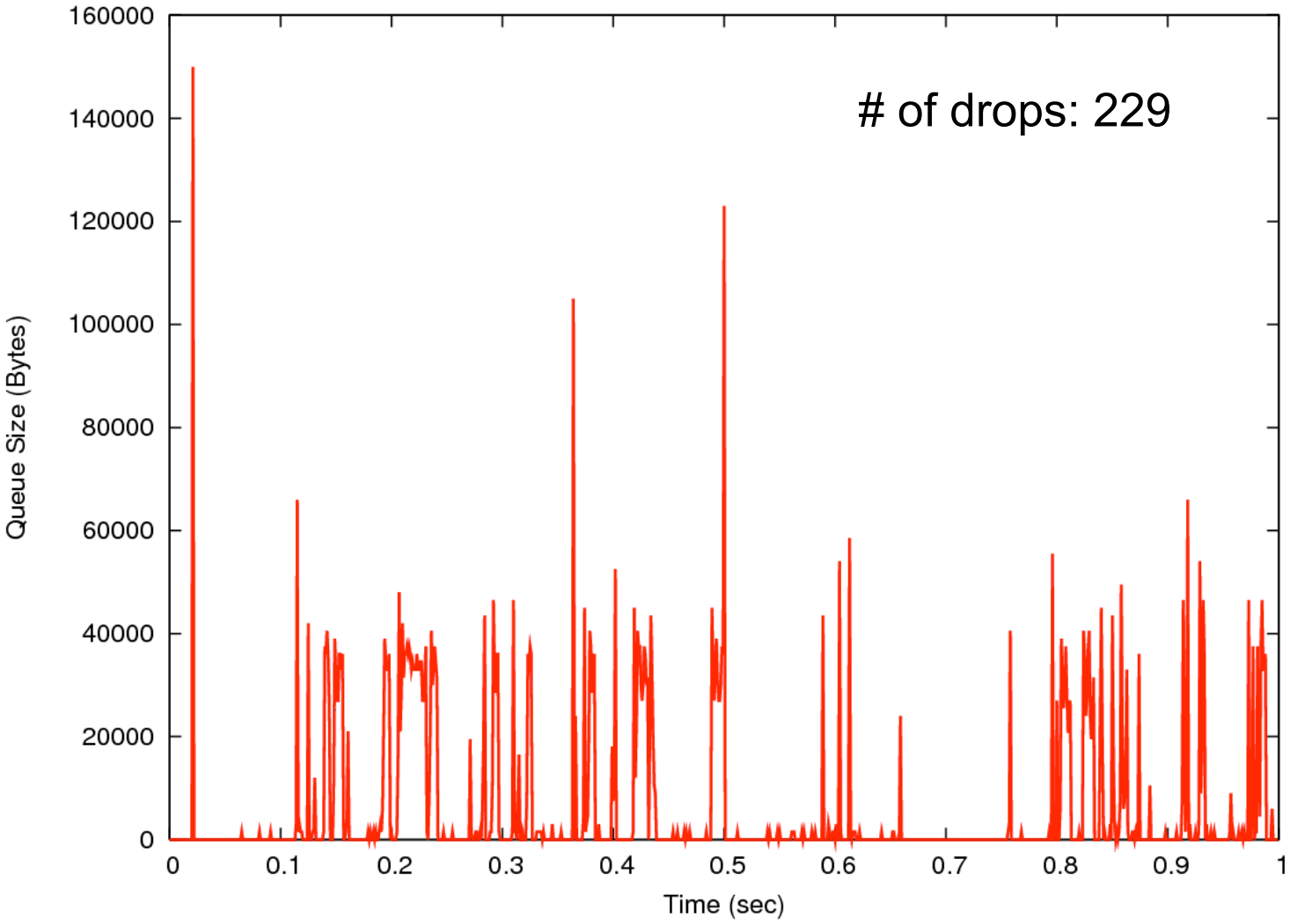
-Congest Point's Queue Size - 20ms on/off



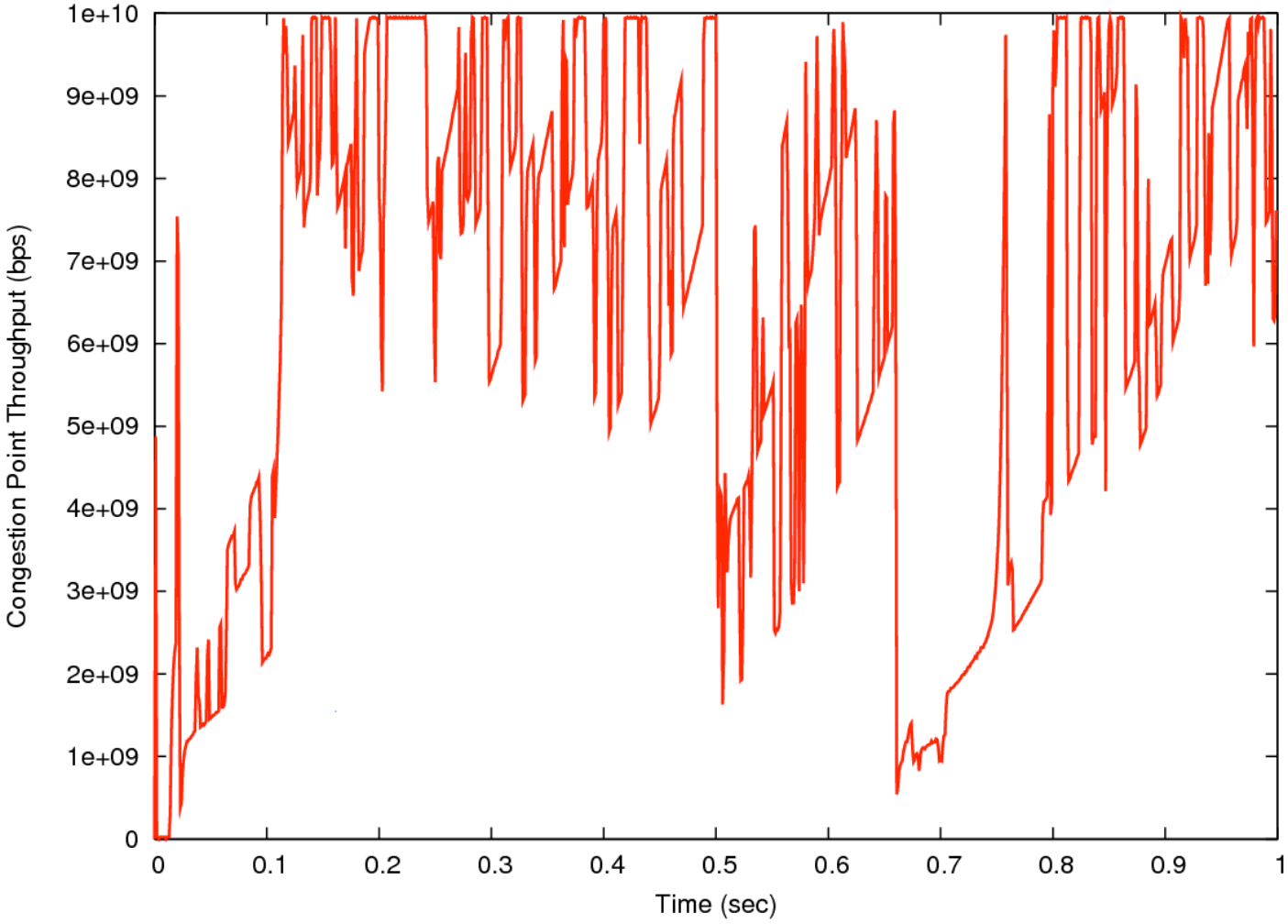
- Net Throughput @ Congestion Point - 20ms on/off



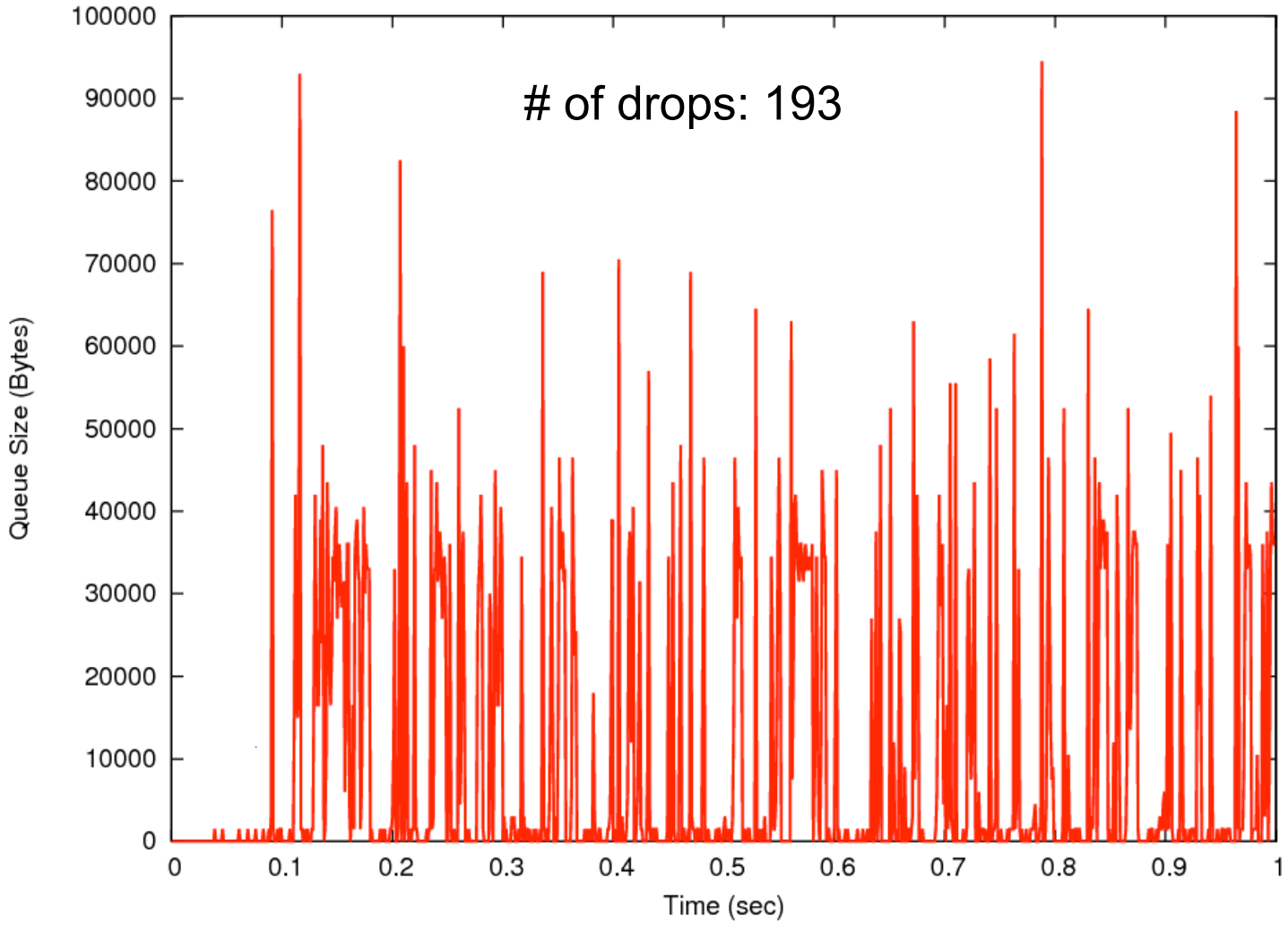
-Congest Point's Queue Size - 10ms on/off



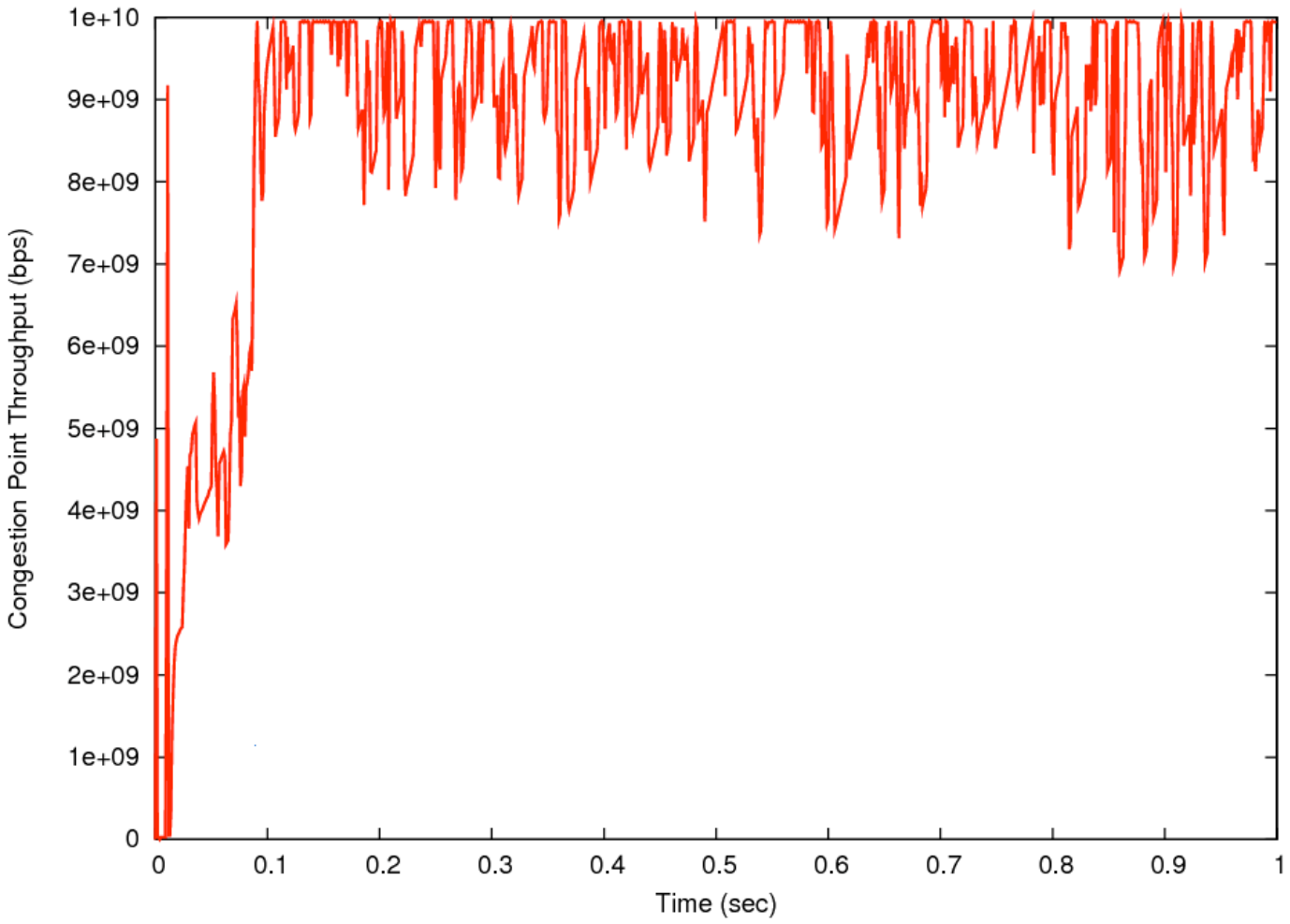
-Net Throughput @ Congestion Point - 10ms on/off



-Congest Point's Queue Size - 5ms on/off

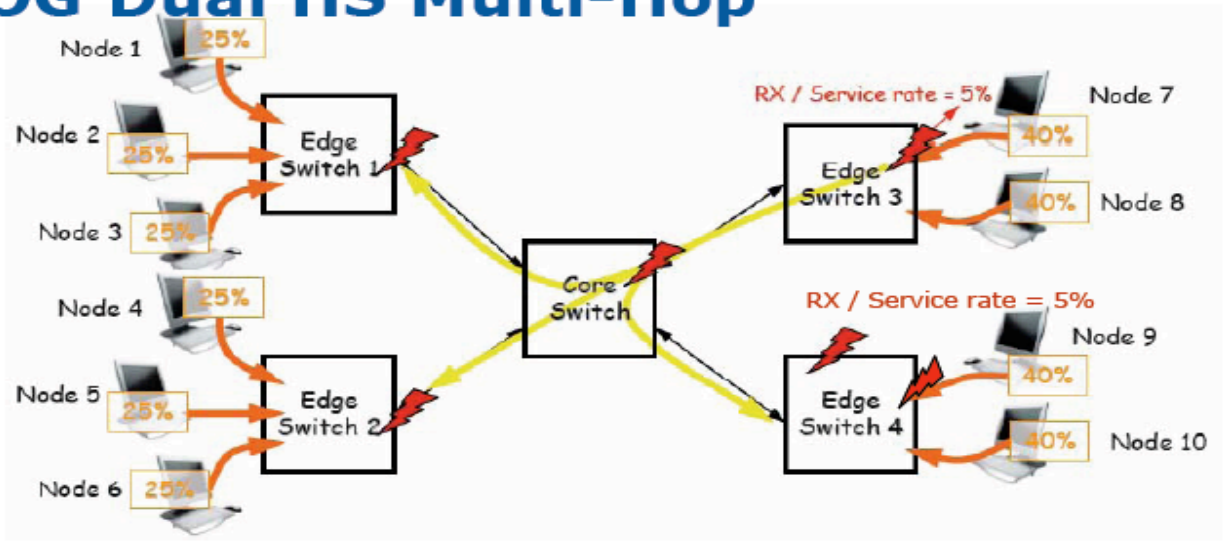


- Net Throughput @ Congestion Point - 5ms on/off



Baseline #6

2. OG Dual HS Multi-Hop



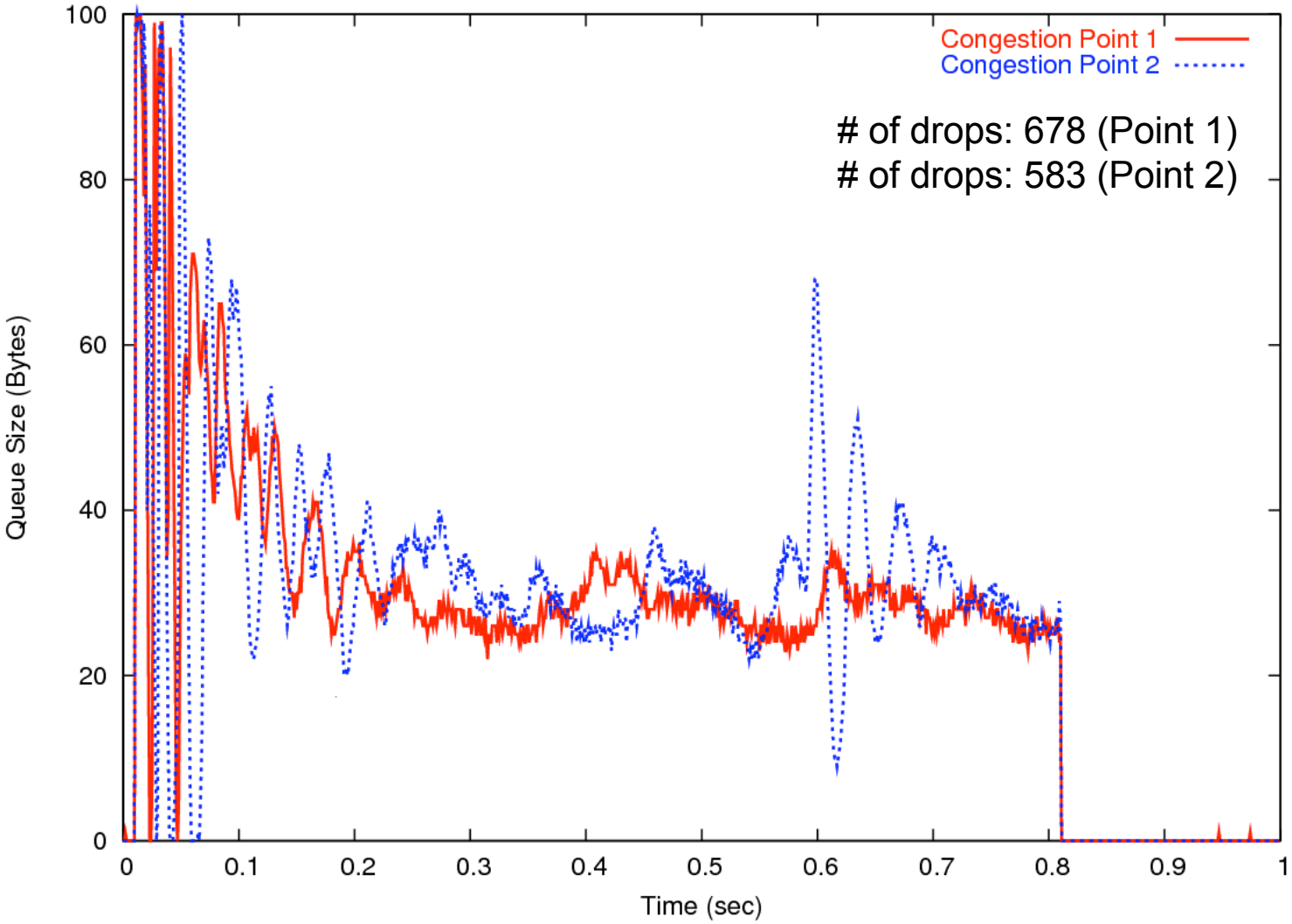
Workload:

- All: Uniform distribution traffic (background traffic)
- Nodes 1-6: 25% (2.5Gbps), Nodes 7-10: 40% (4 Gbps)
- Two Hotspots:
  - Node 7 & 9 service rate = 5% (Rx only)
  - If saturation tree spreads => 5 congestion points total

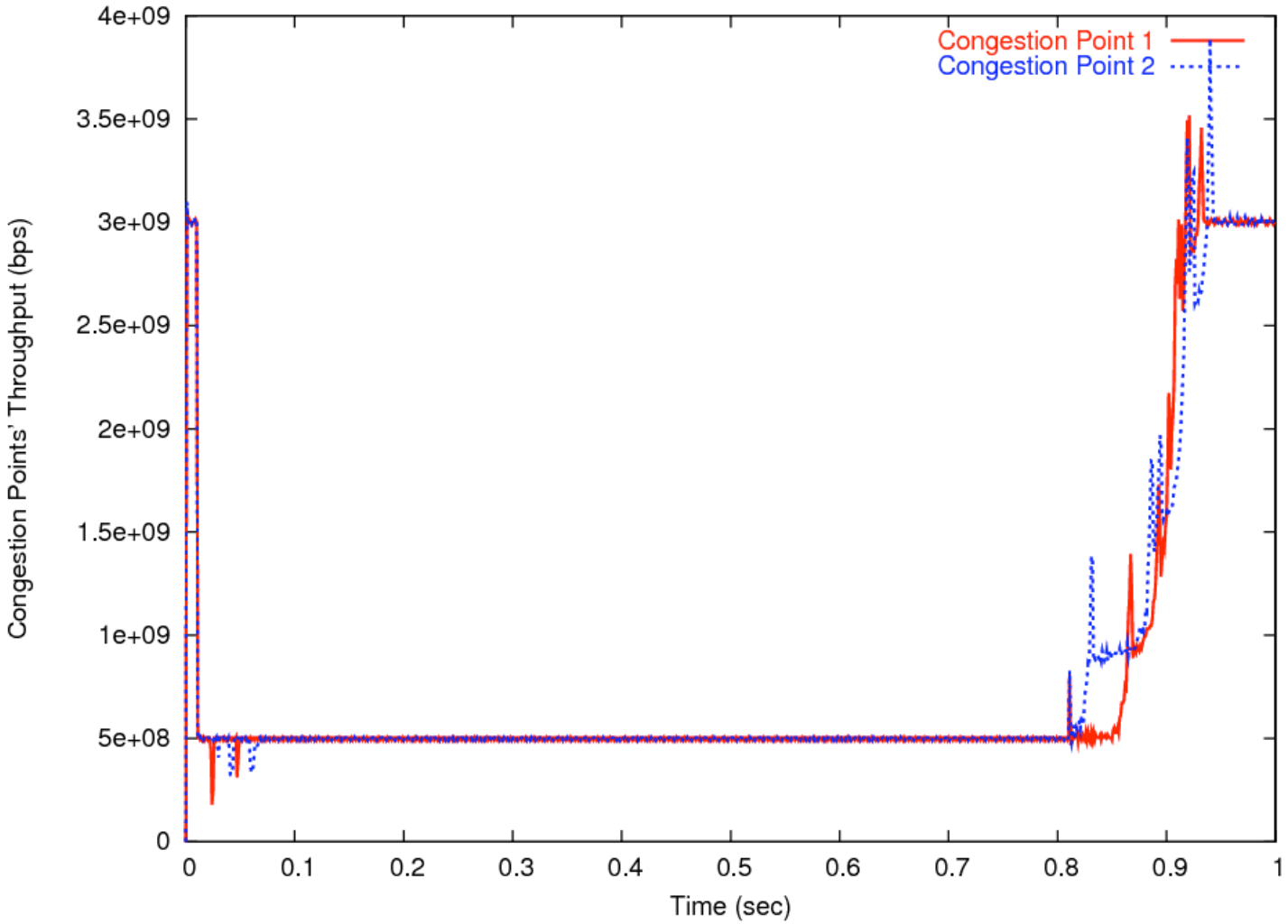
- Scenarios:
  - PAUSE: Enabled/Disabled

Extended

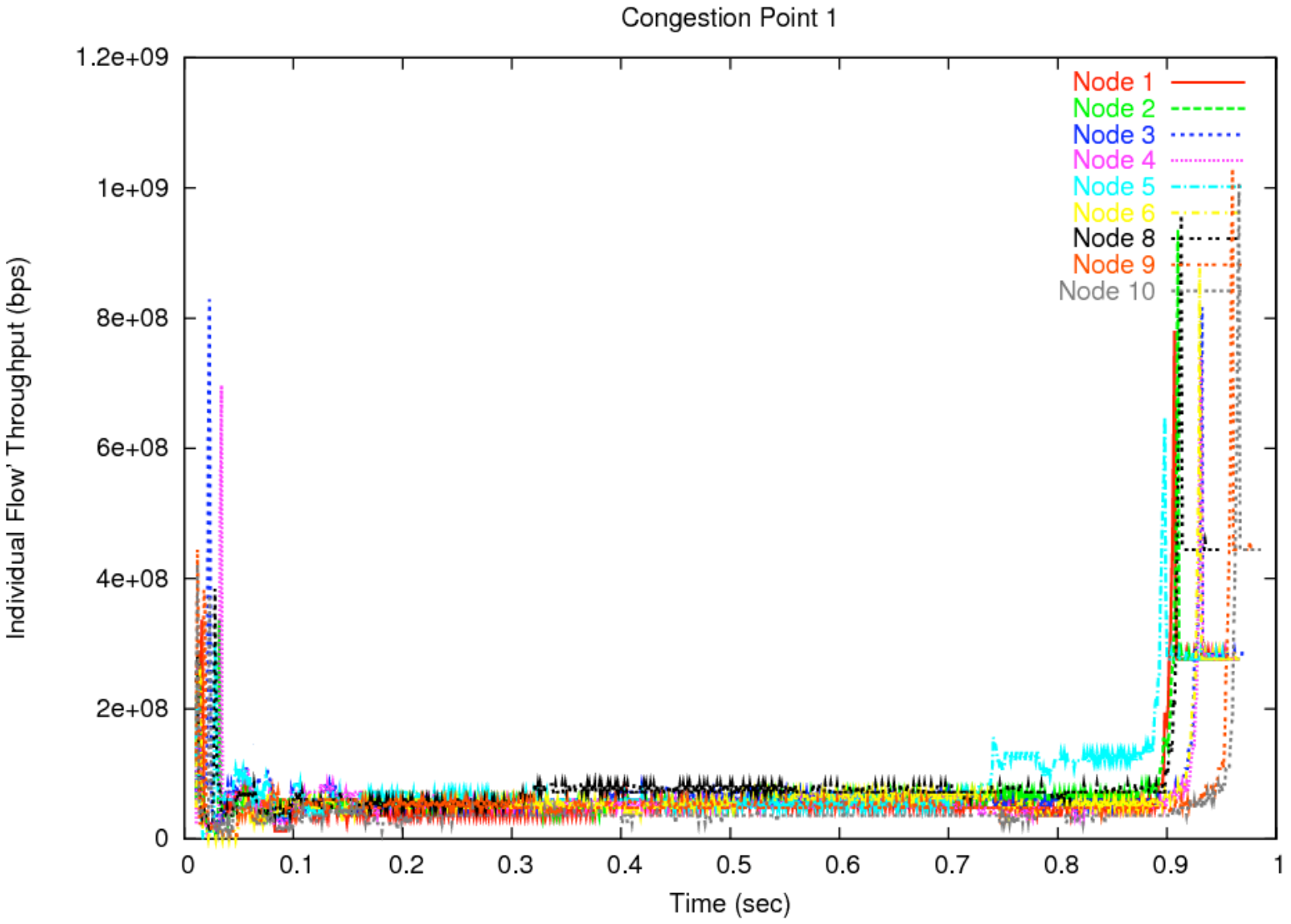
Without Pause
- Congested Queue Size (0.5Gbps)



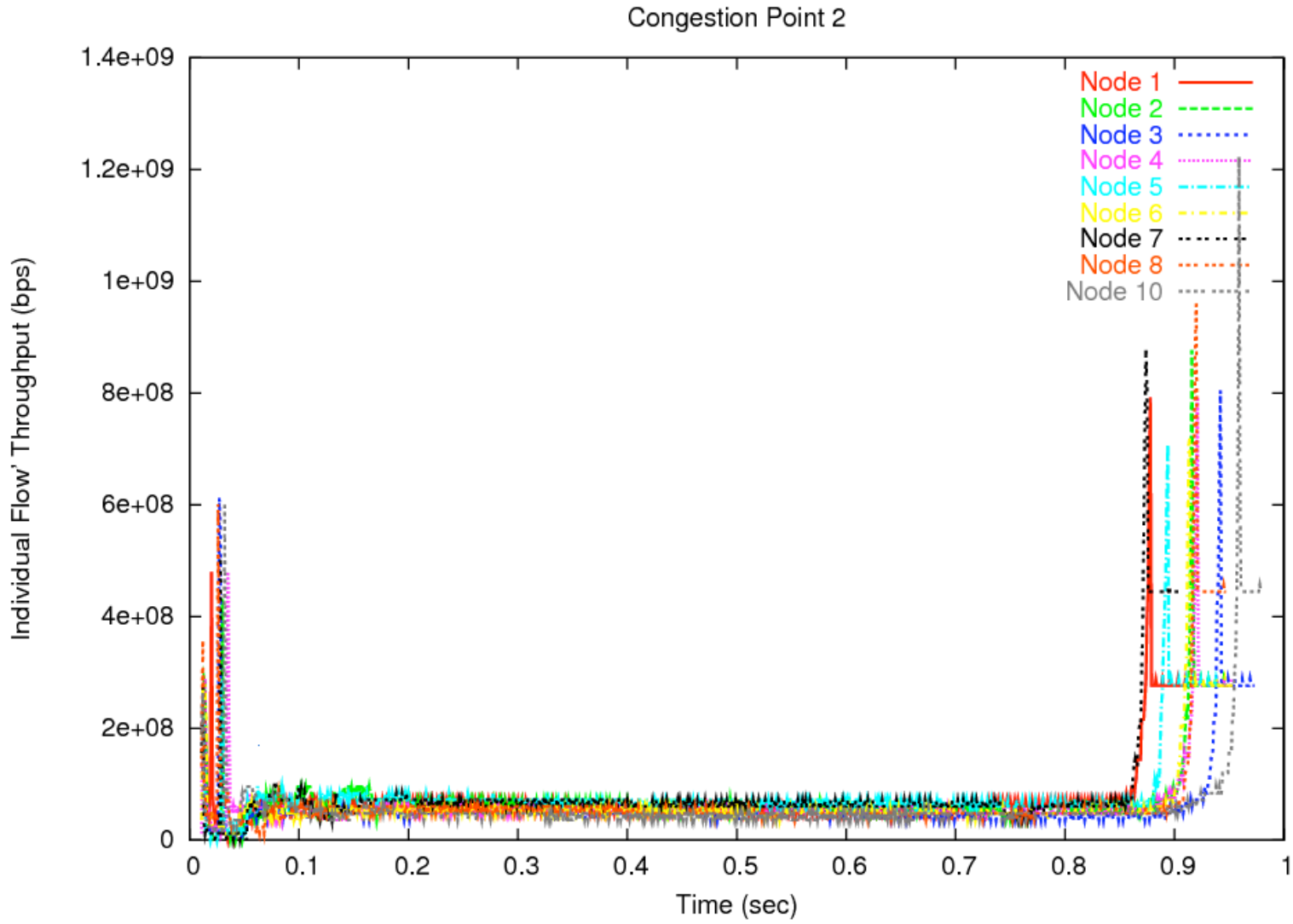
Without Pause
- Congested Points' Throughput



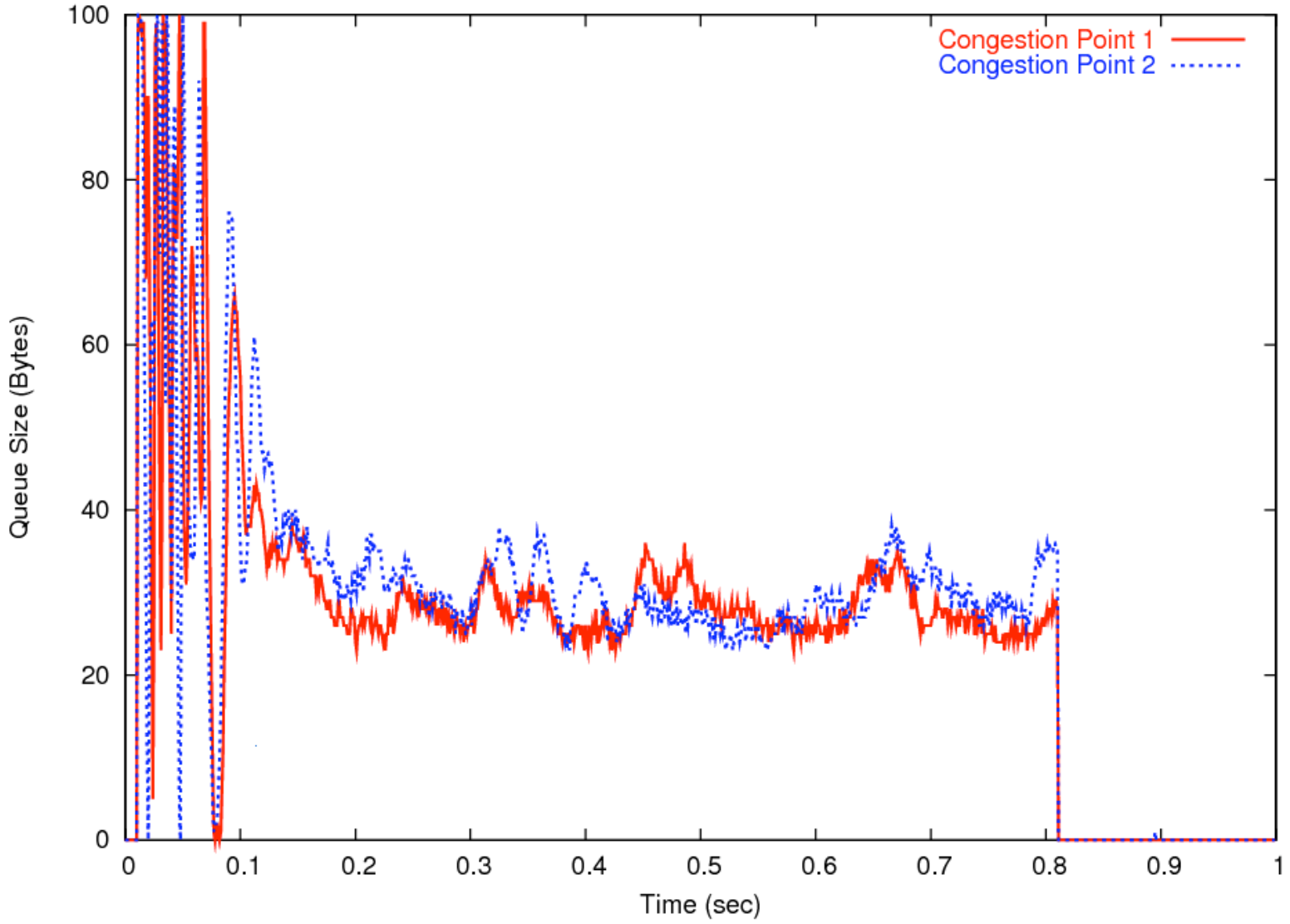
Without Pause
- Individual Rates



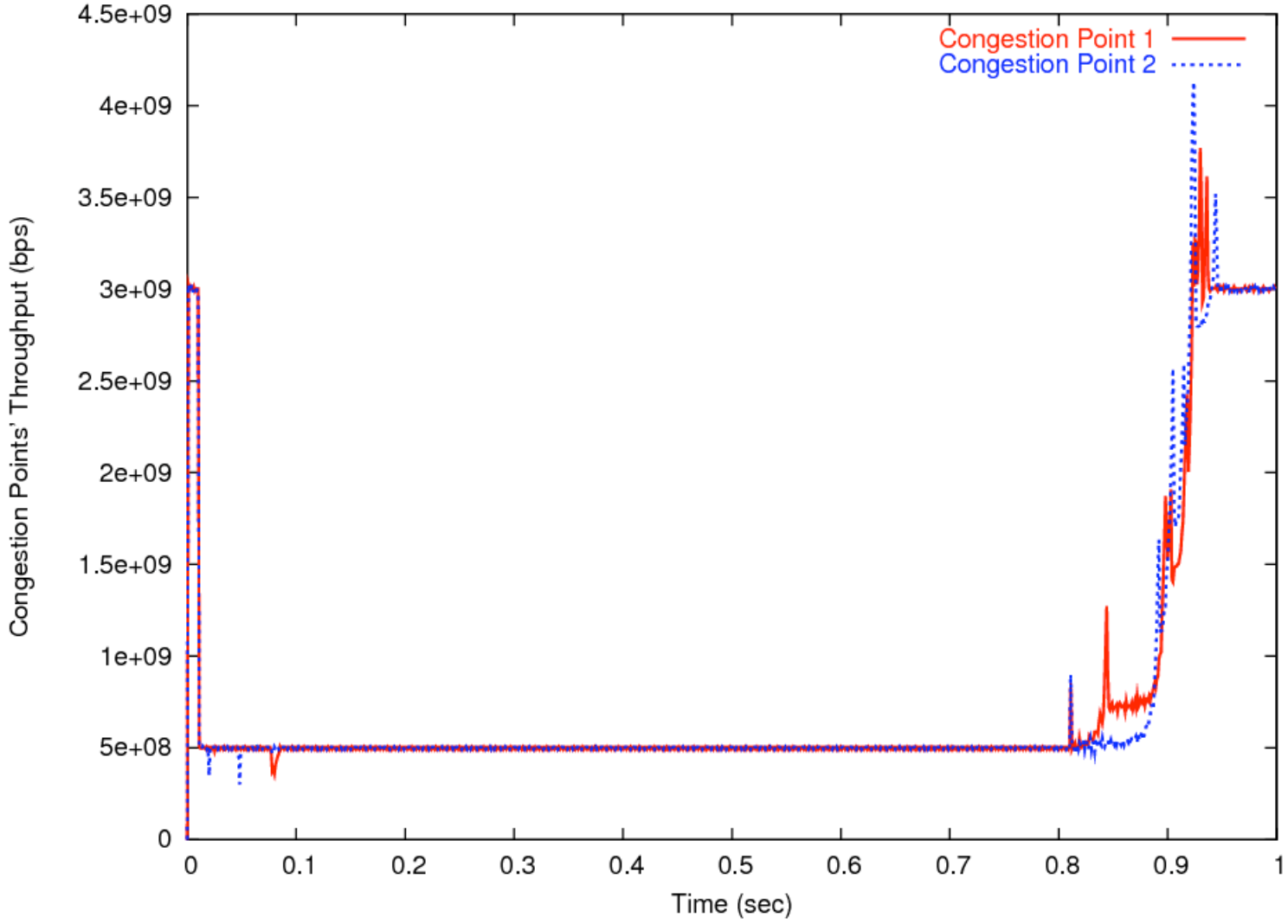
Without Pause
- Individual Rates



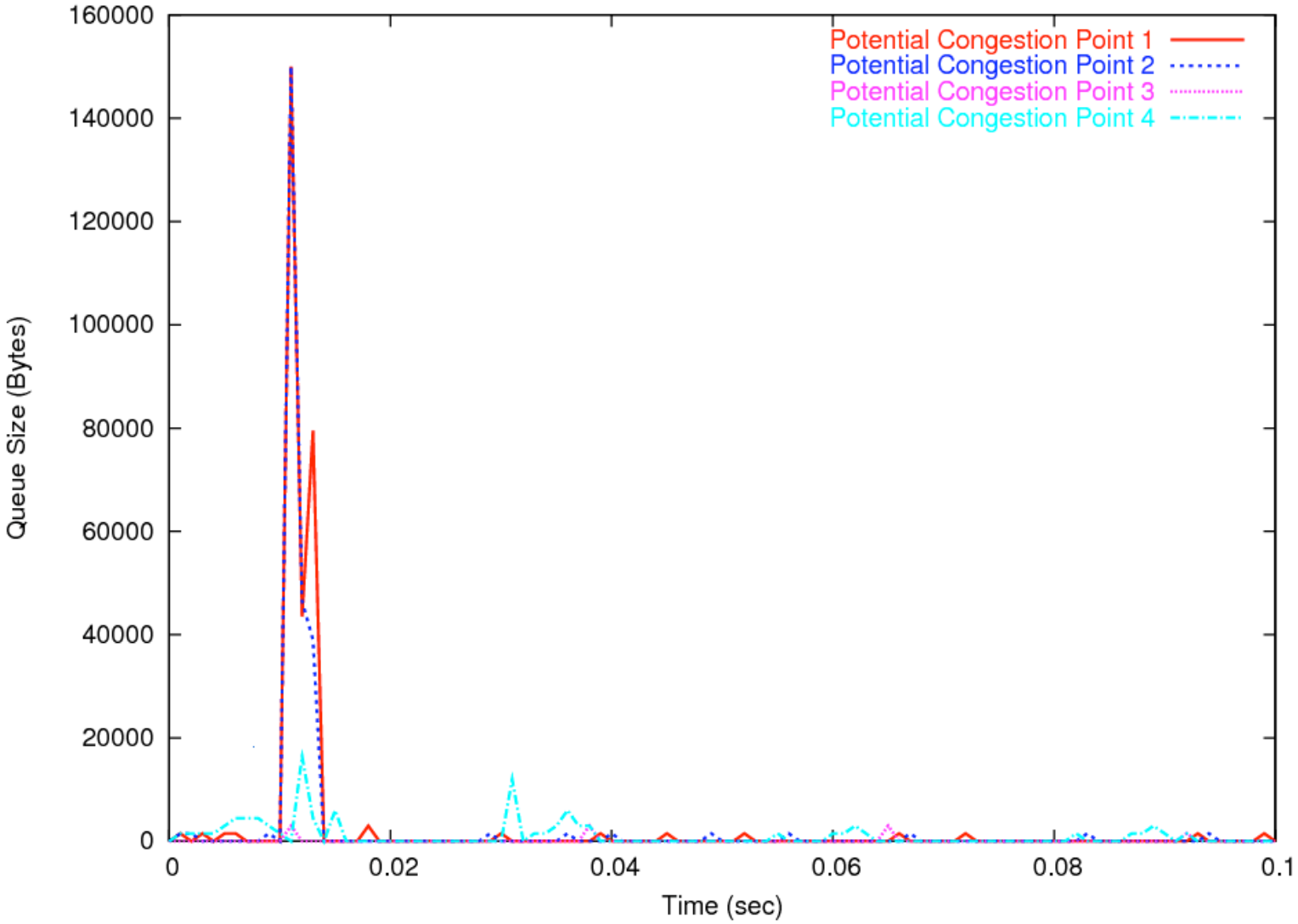
Pause
- Congested Queue Size (0.5Gbps)



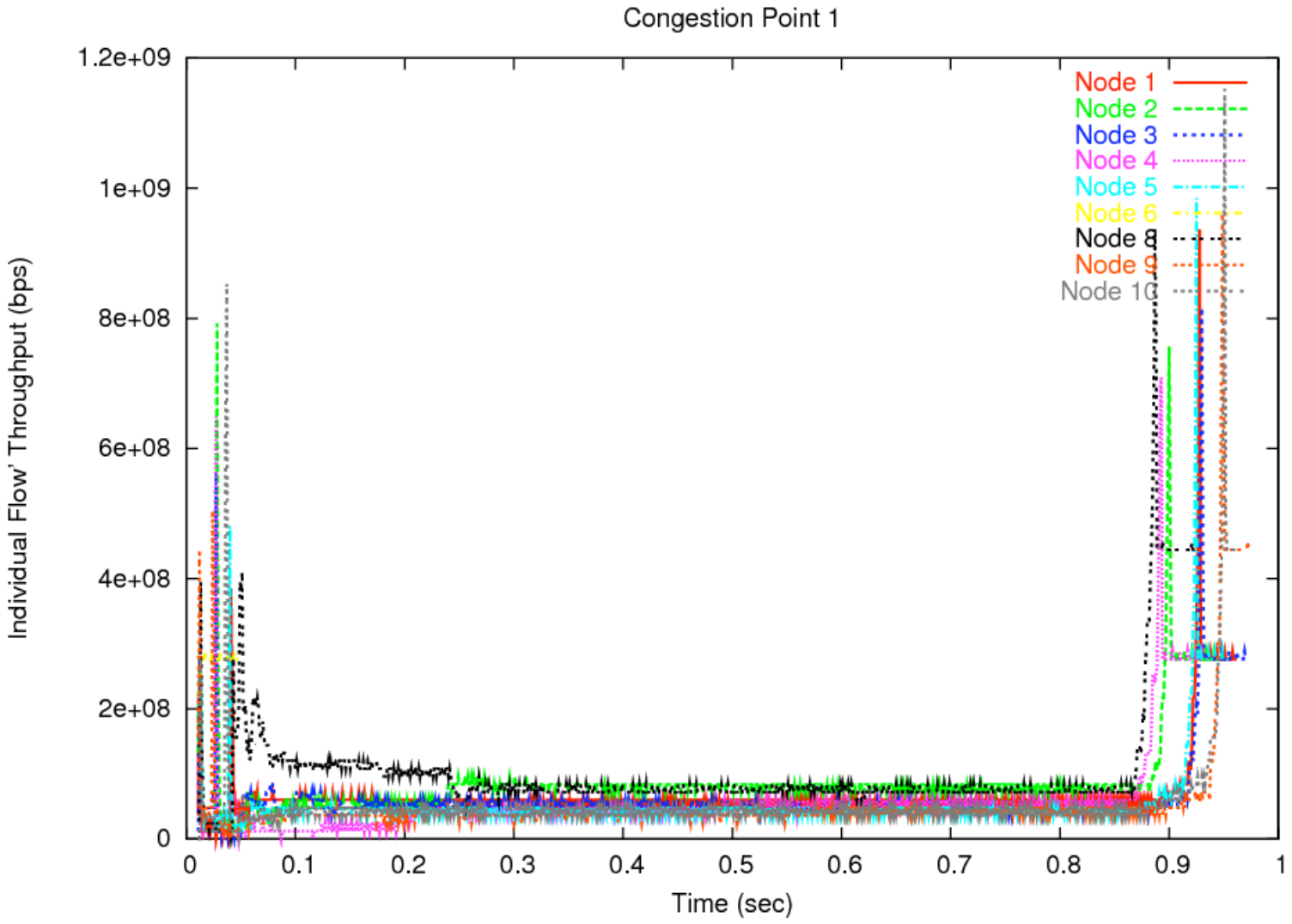
Pause
- Congested Points' Throughput



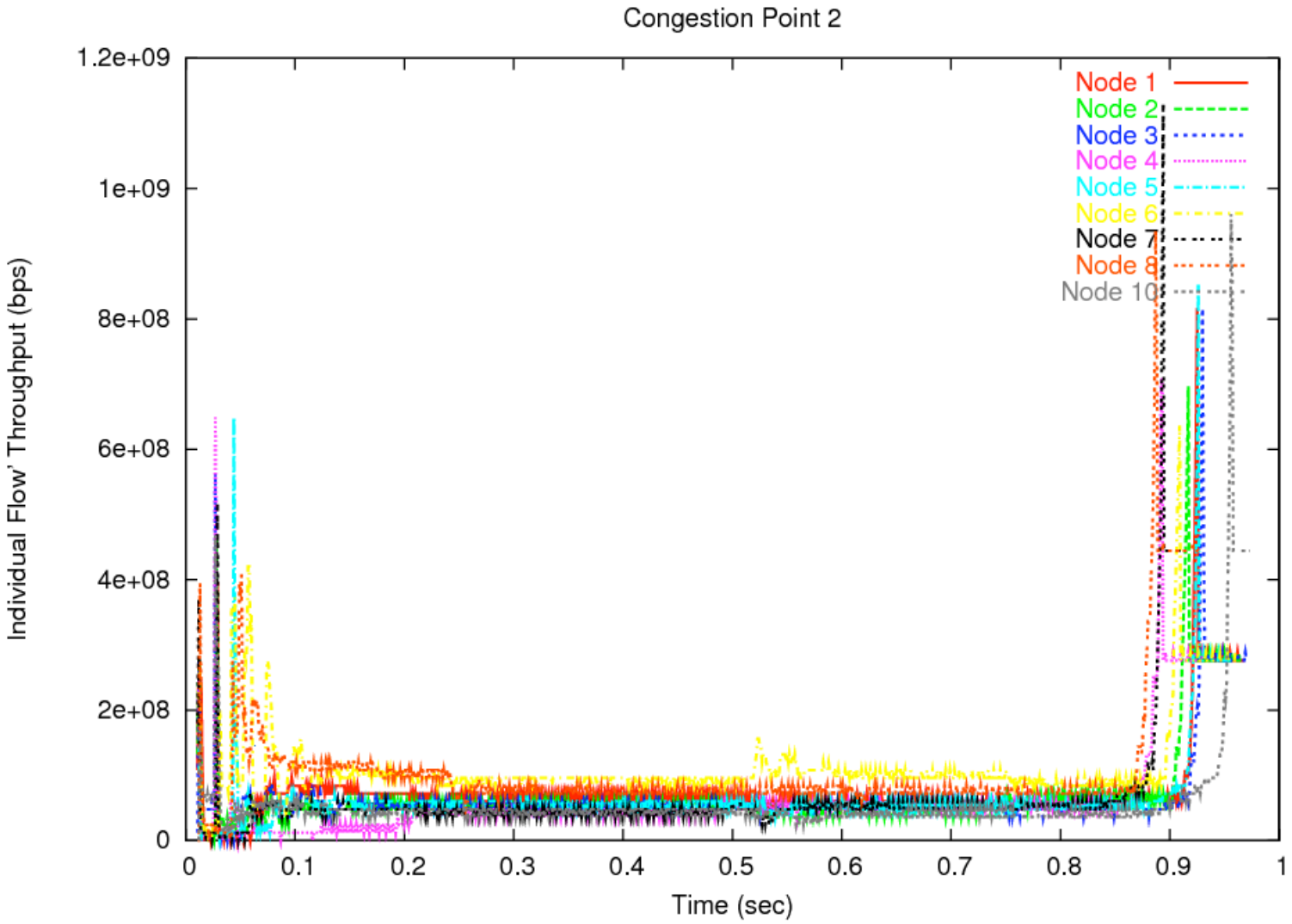
Pause
- Potential Congestion Queues



With Pause
- Individual Rates

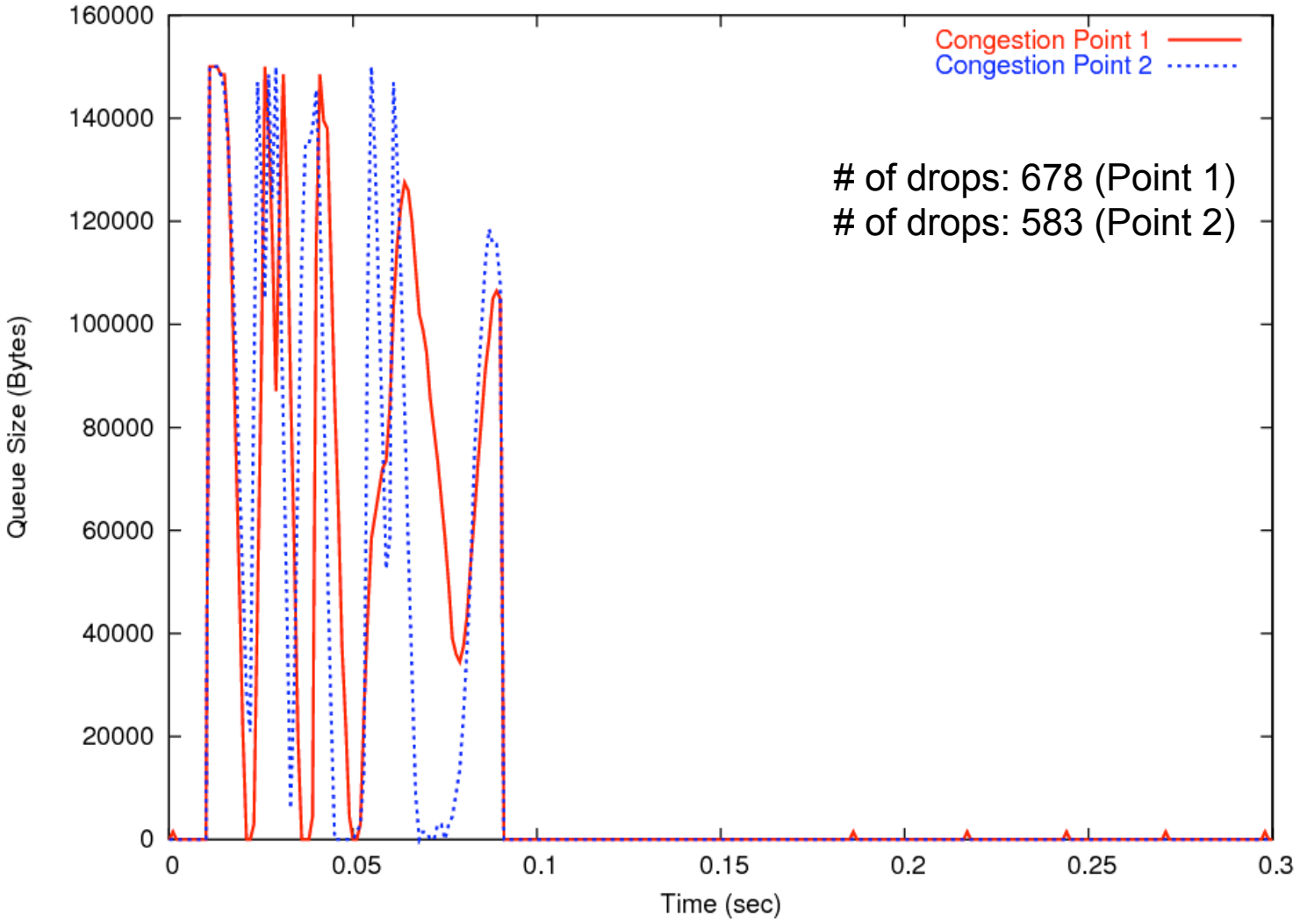


With Pause
- Individual Rates

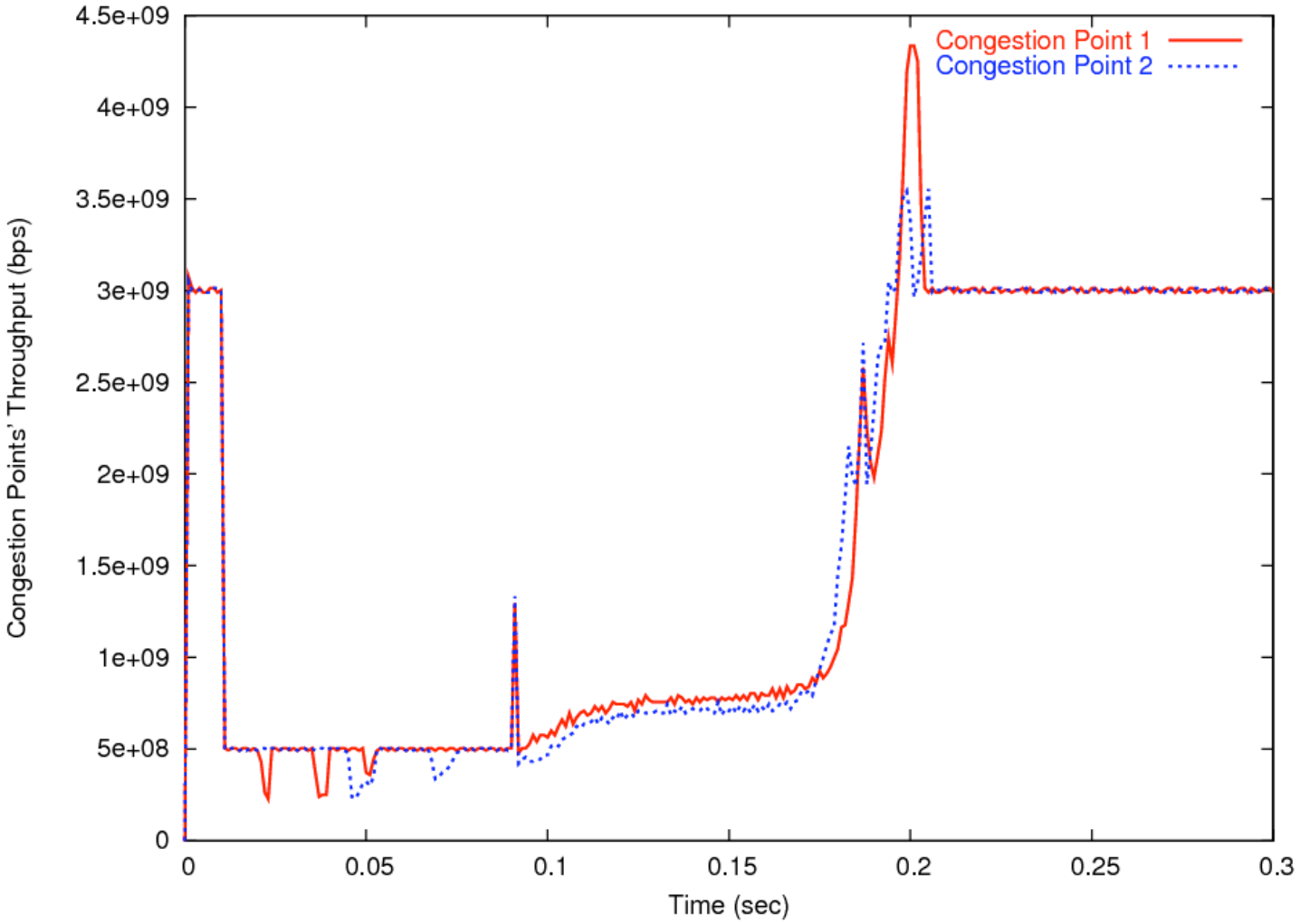


Hot Spots' Duration of 80ms

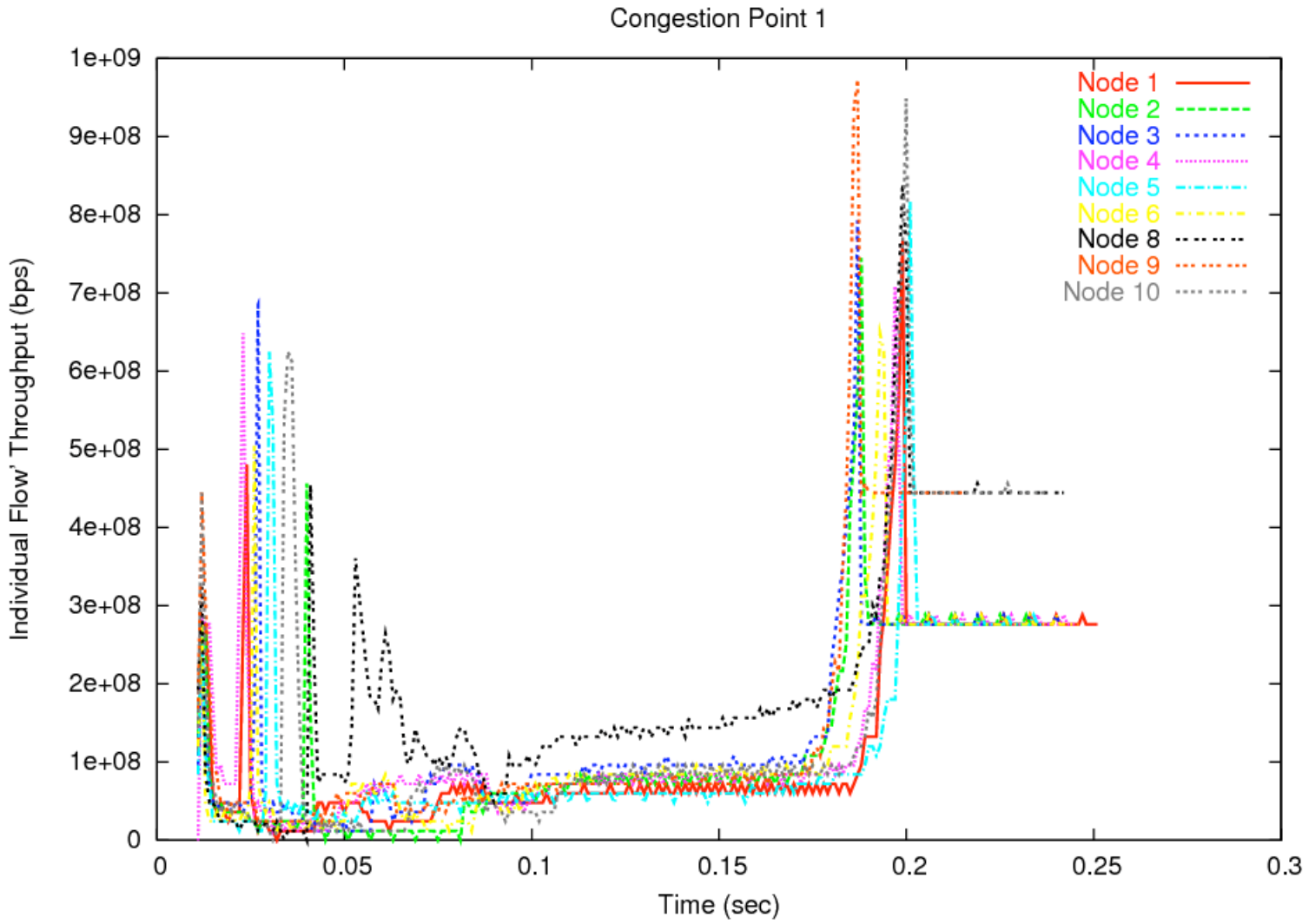
Without Pause
- Congested Queue Size (0.5Gbps)



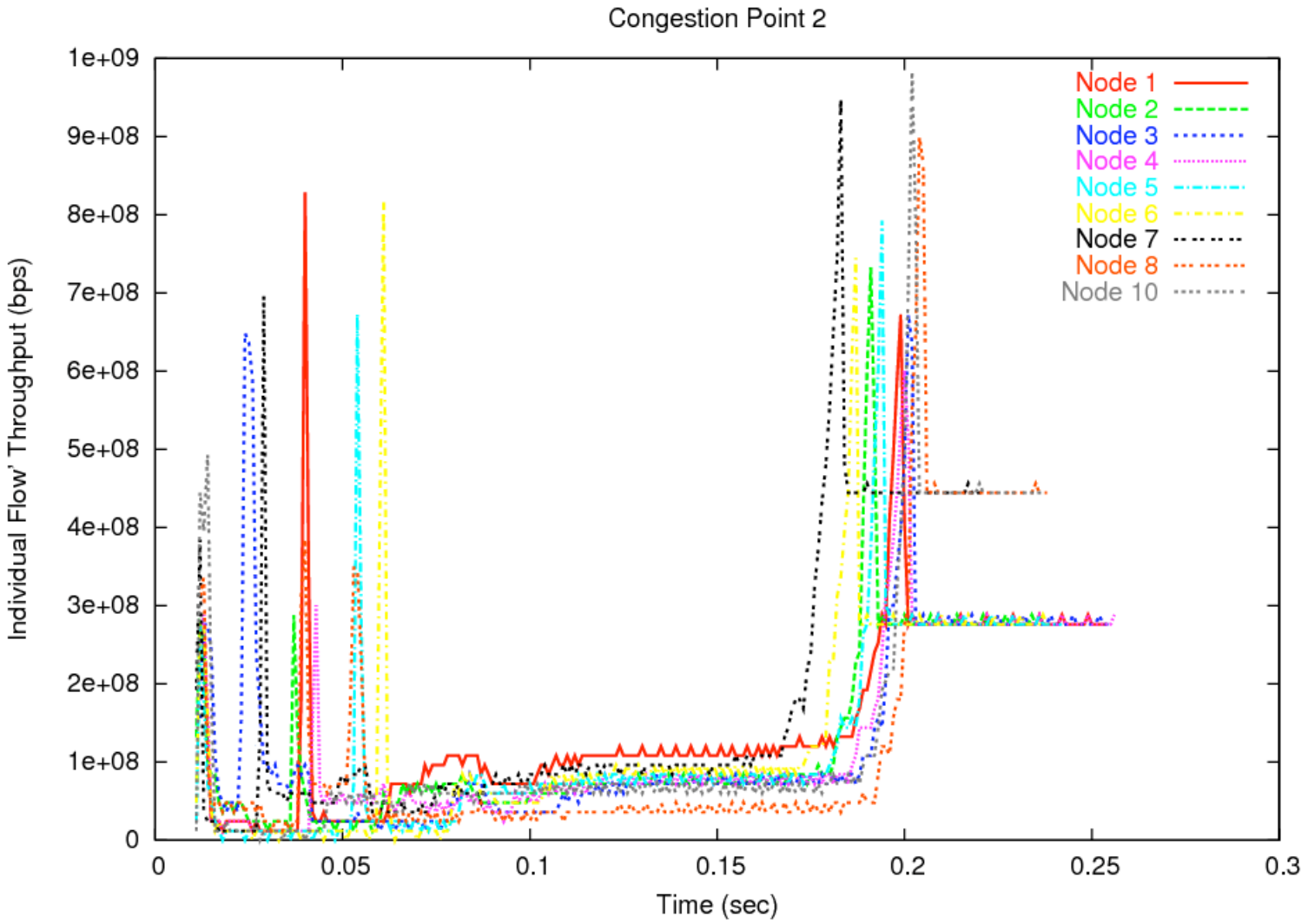
Without Pause
- Congested Points' Throughput



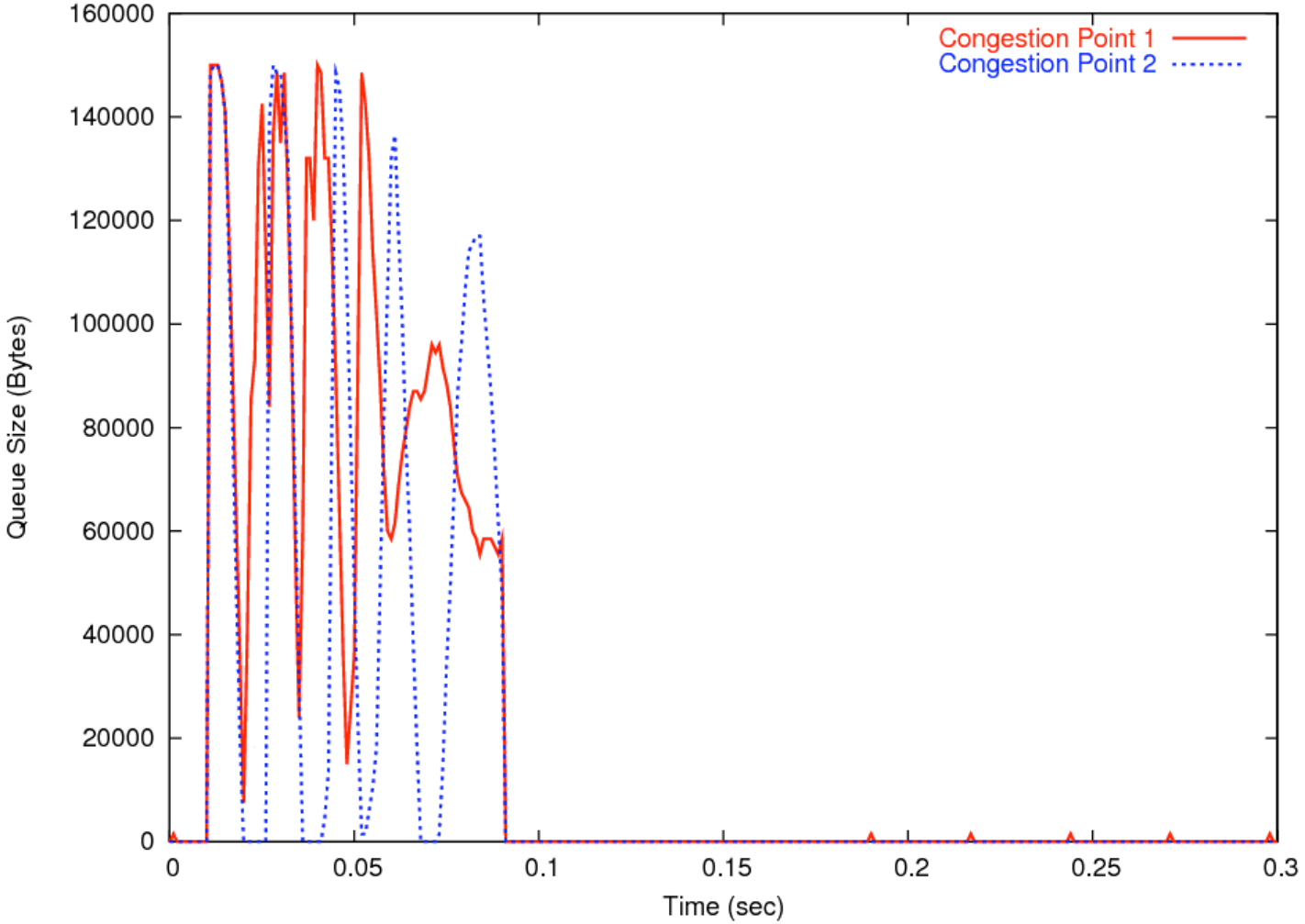
Without Pause
- Individual Rates



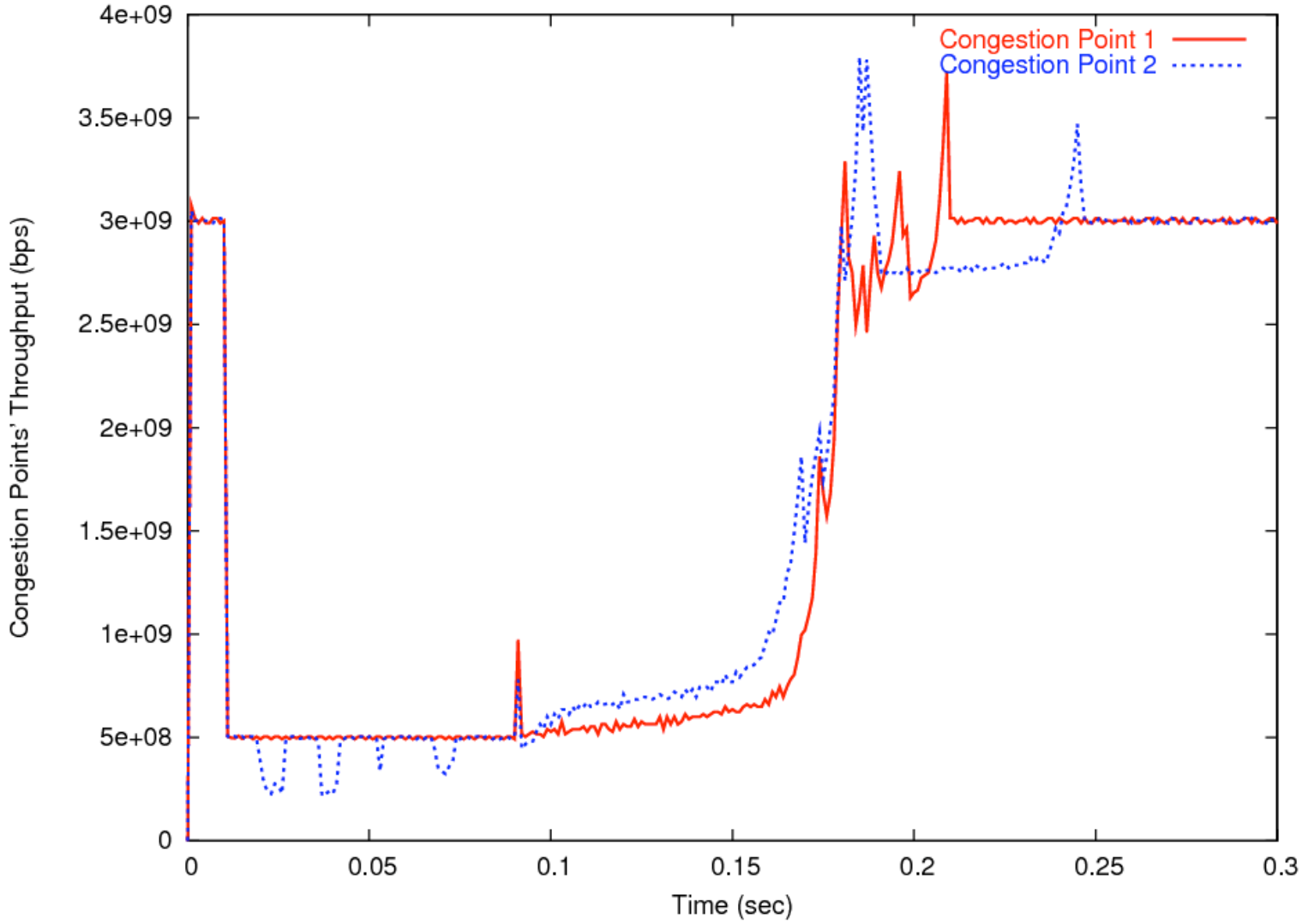
Without Pause
- Individual Rates



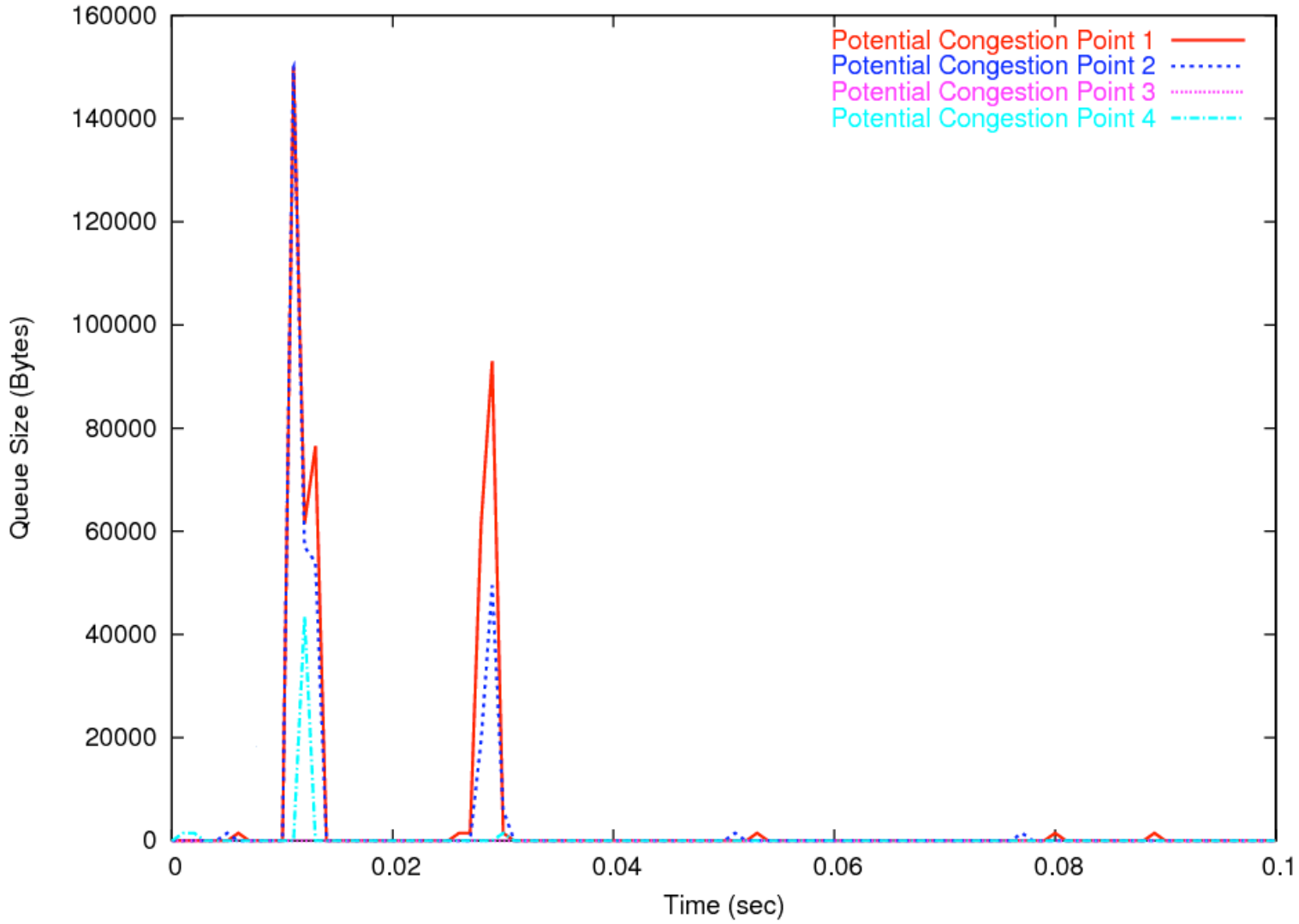
Pause
- Congested Queue Size (0.5Gbps)



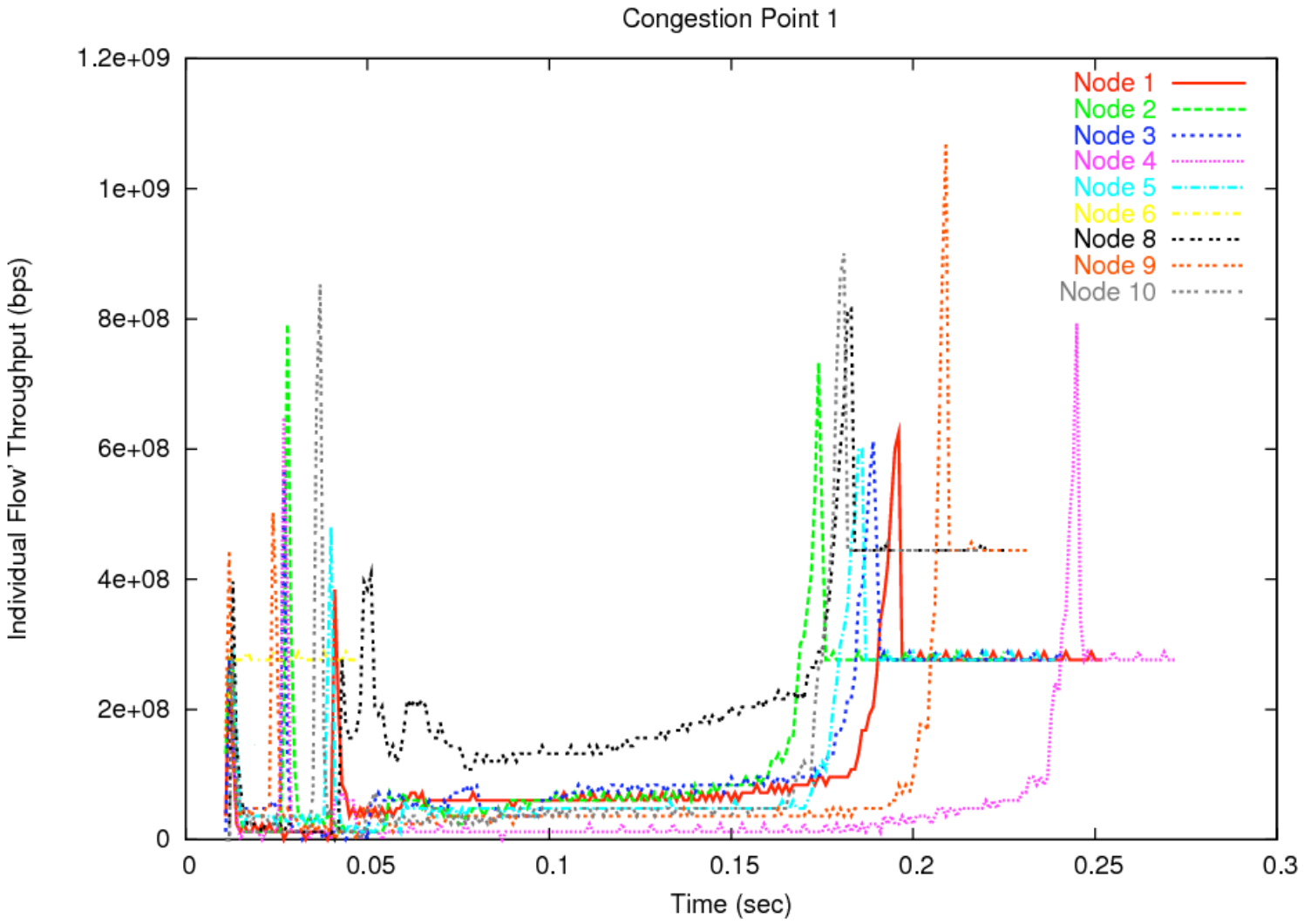
Pause
- Congested Points' Throughput



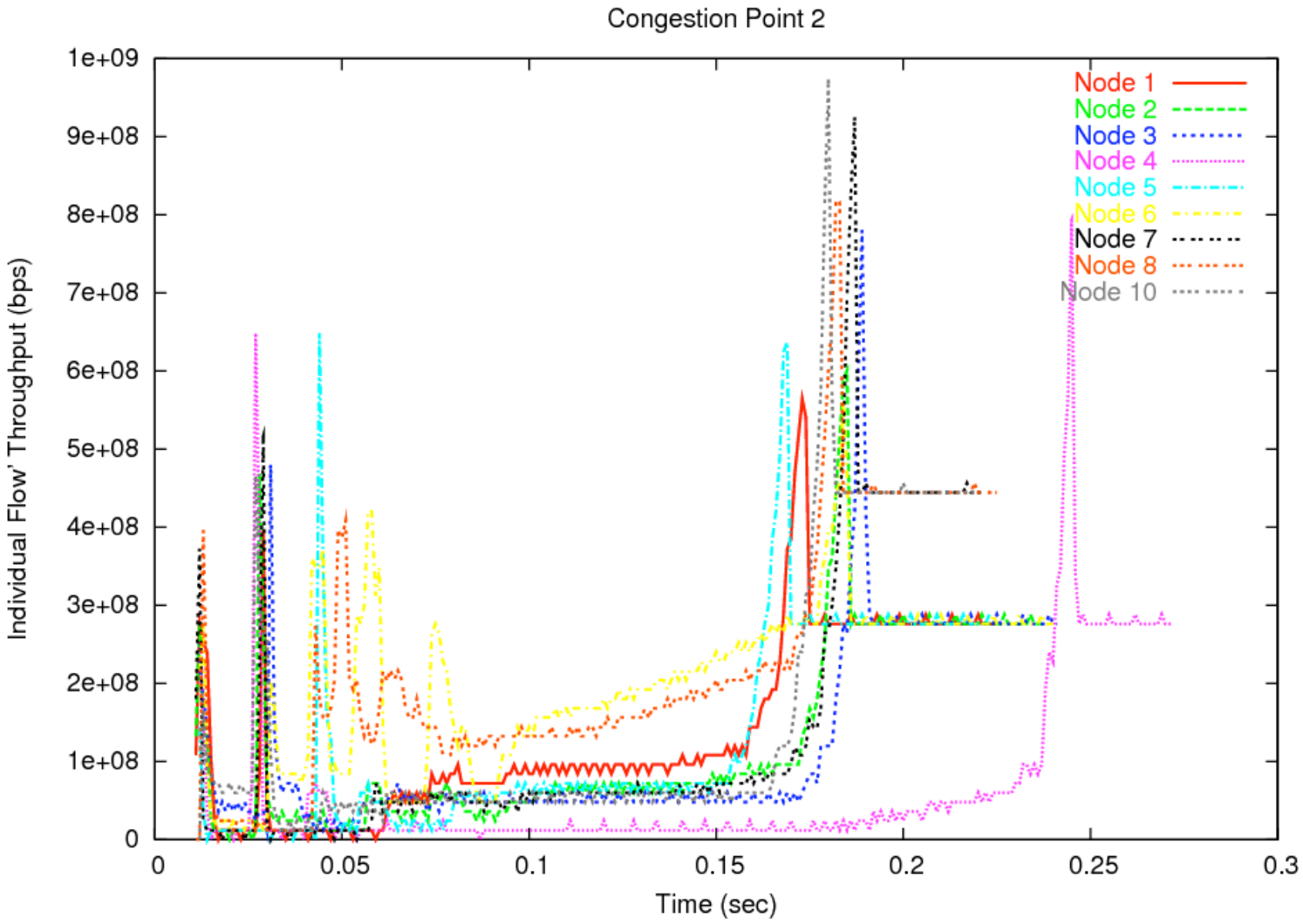
Pause
- Potential Congestion Queues



With Pause
- Individual Rates

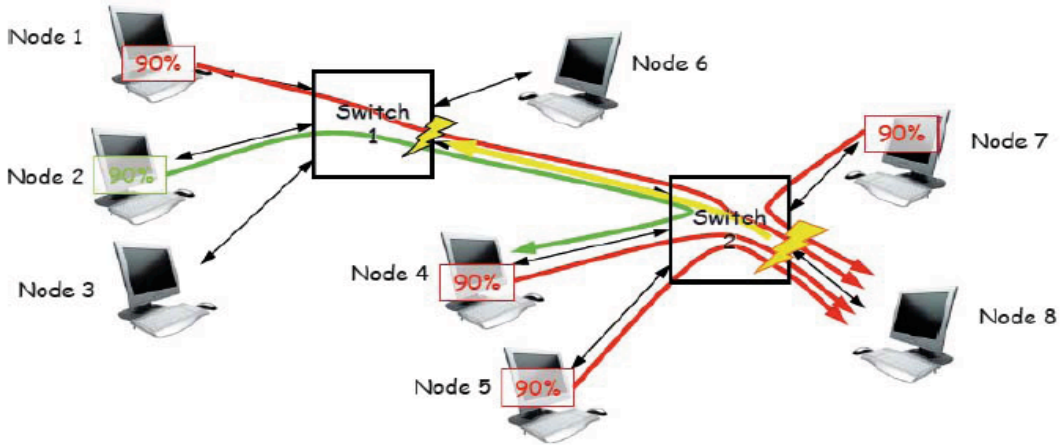


With Pause
- Individual Rates



Baseline #7

7. Multistage Dual HS (Light & Heavy)

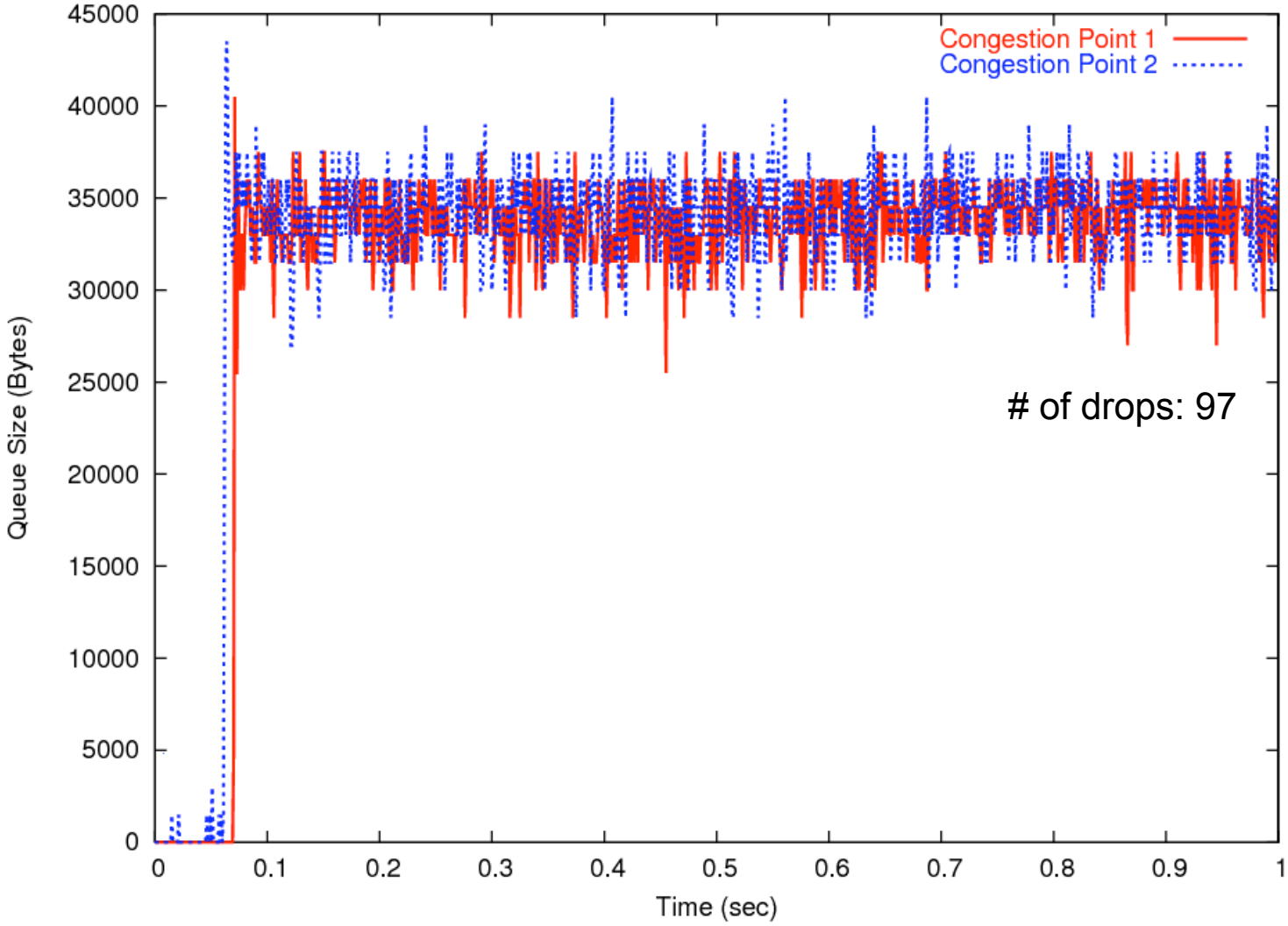


Workload:

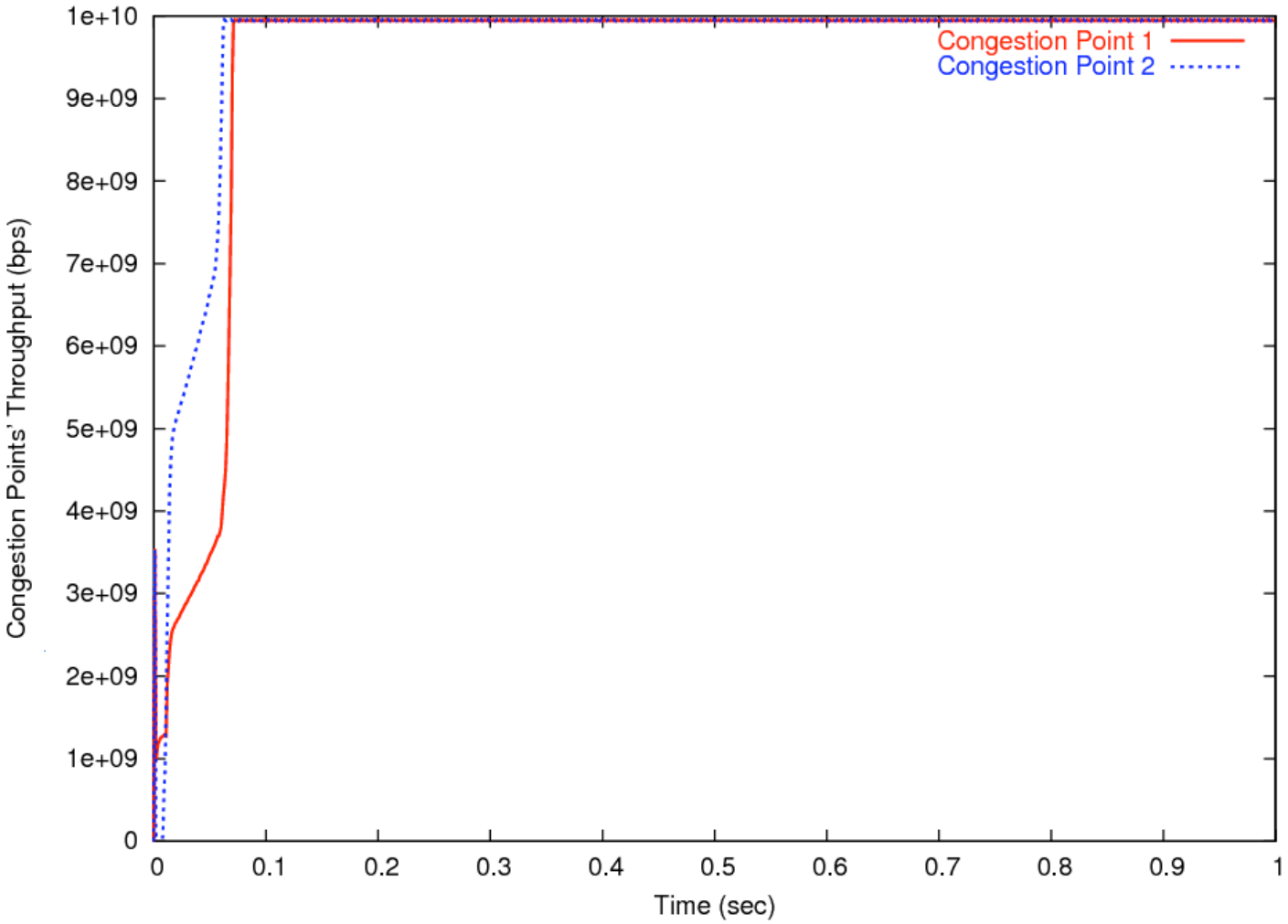
- Two switches, all links 10 Gb/s, no background traffic
- Four flows of 9 Gb/s each from nodes 1, 4, 5, 7 to node 8
- One flow of 9 Gb/s from node 2 to node 4
- Two congestion points
  - Port from switch 1 to switch 2
  - Port from switch 2 to node 8
- Fair allocation should provide 2.5 Gb/sfor all flows to node 8 and 7.5 Gb/sfor flow to node 4

Extended

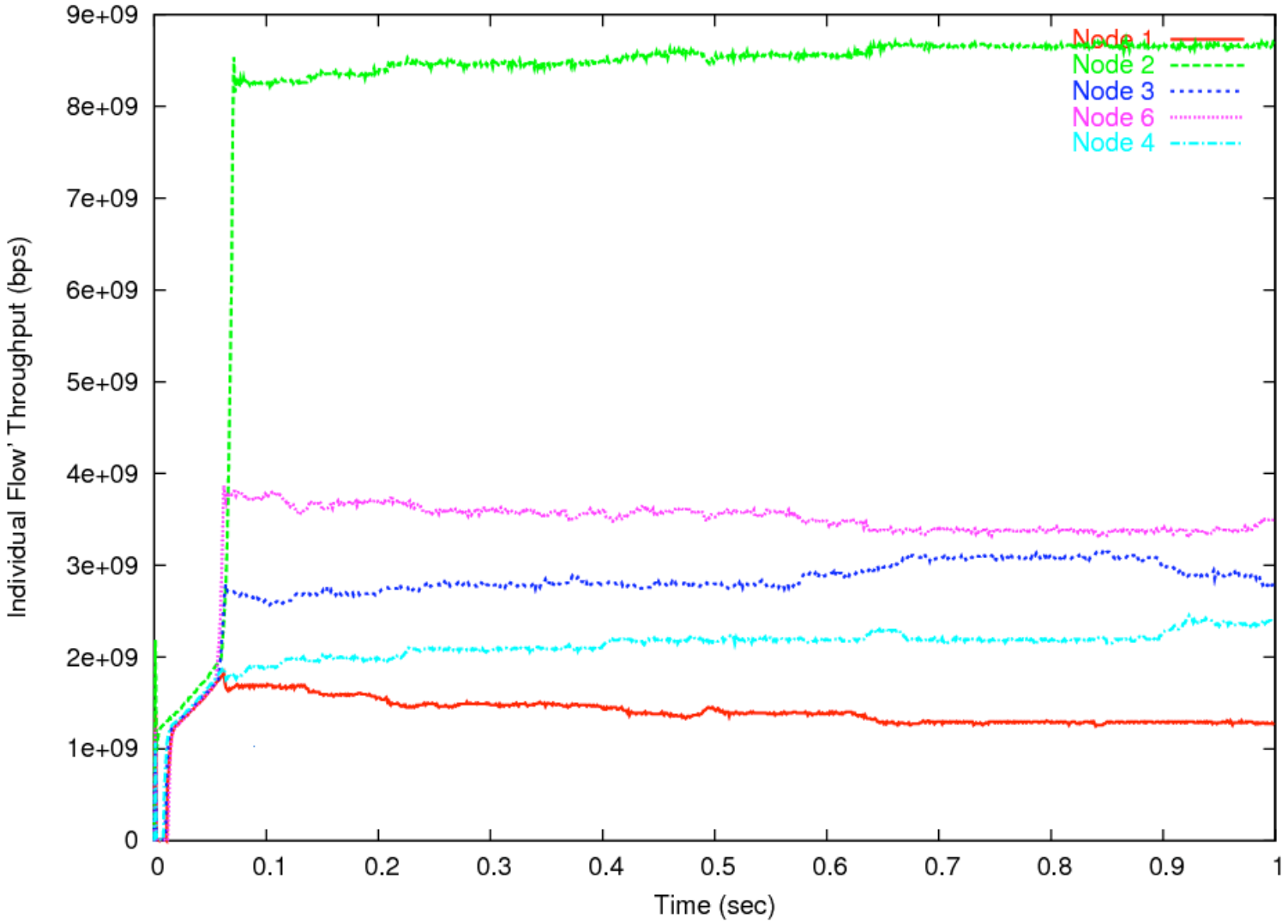
Congested Points' Queue Size



Congested Points' Throughput

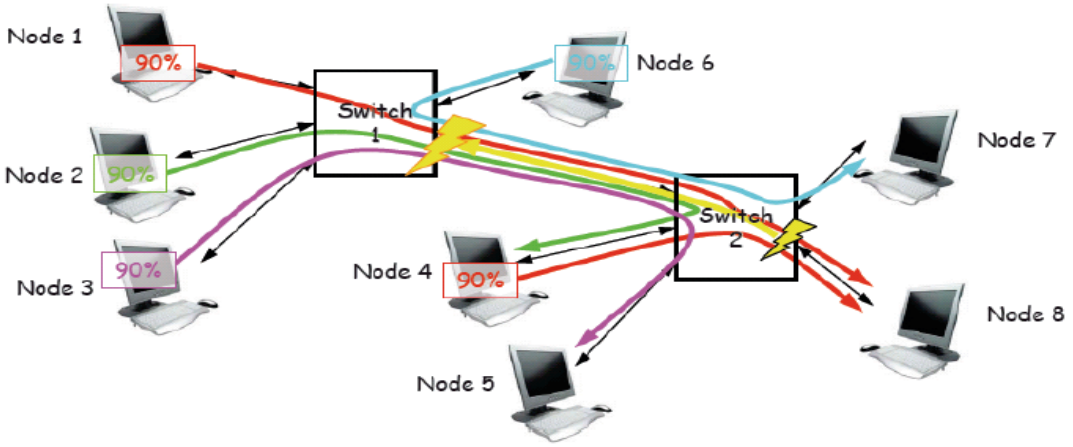


Individual Flow Rates



Baseline #8

8. Multistage Dual HS (Heavy & Light)

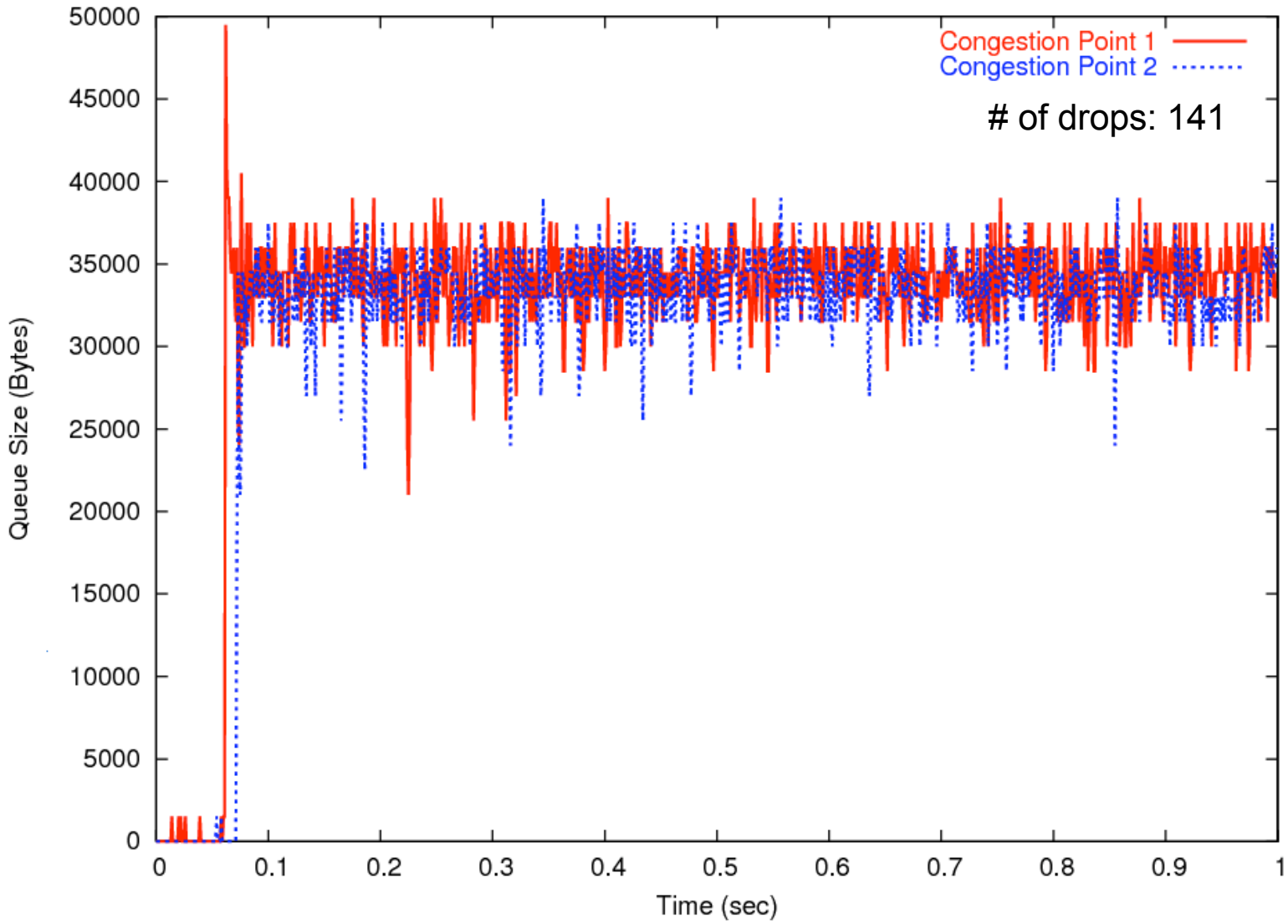


Workload:

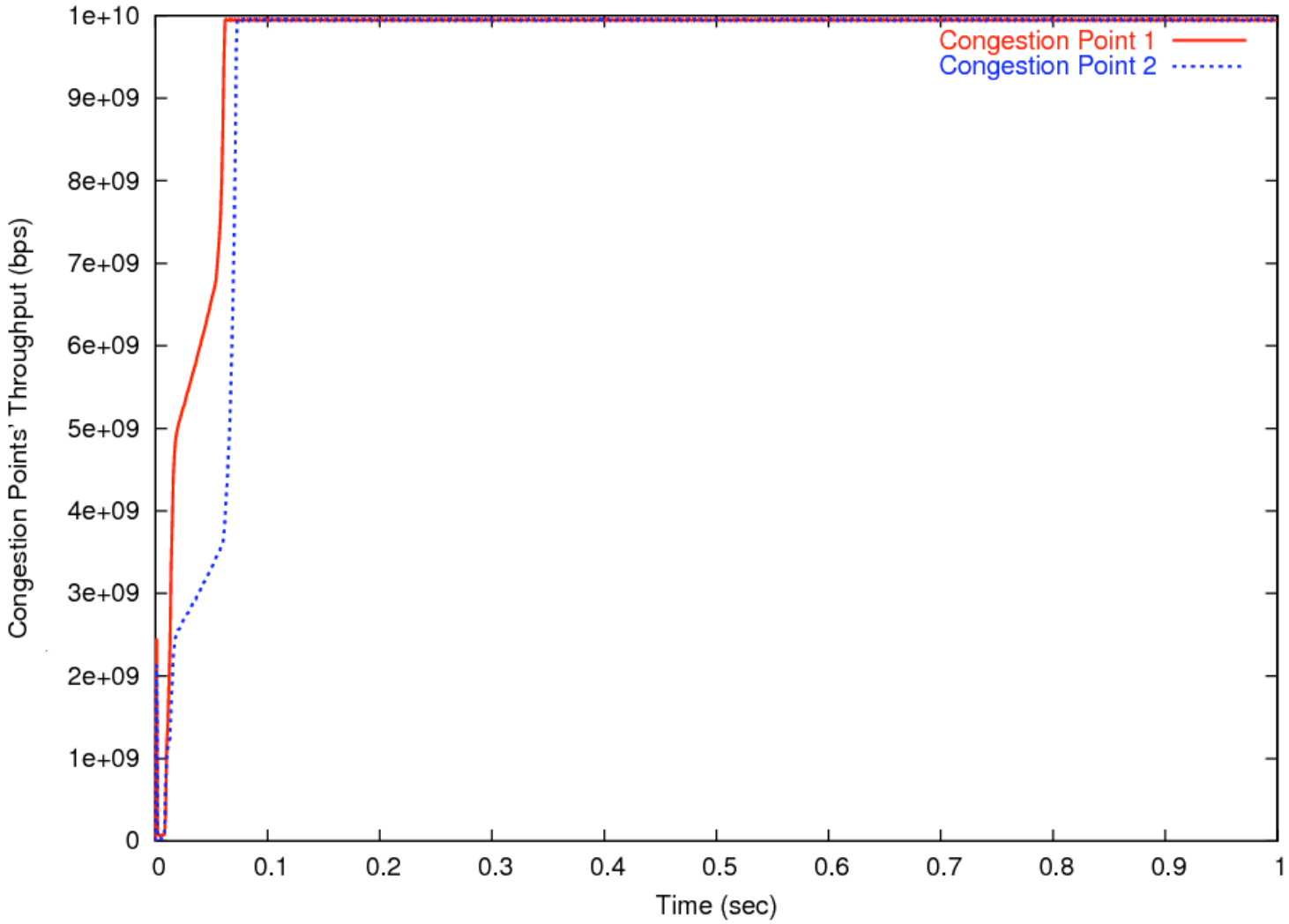
- Two switches, all links 10 Gb/s, no background traffic
- Two flows of 9 Gb/seach from nodes 1 and 4 to node 8
- Three flows of 9 Gb/seach from node 2 to node 4, 3 to 5, and 6 to 7
- Two congestion points
  - Port from switch 1 to switch 2
  - Port from switch 2 to node 8
- Fair allocation should provide 2.5 Gb/sfor all flows to switch 2 and 7.5 Gb/sfor flow from node 4 to node 8

Extended

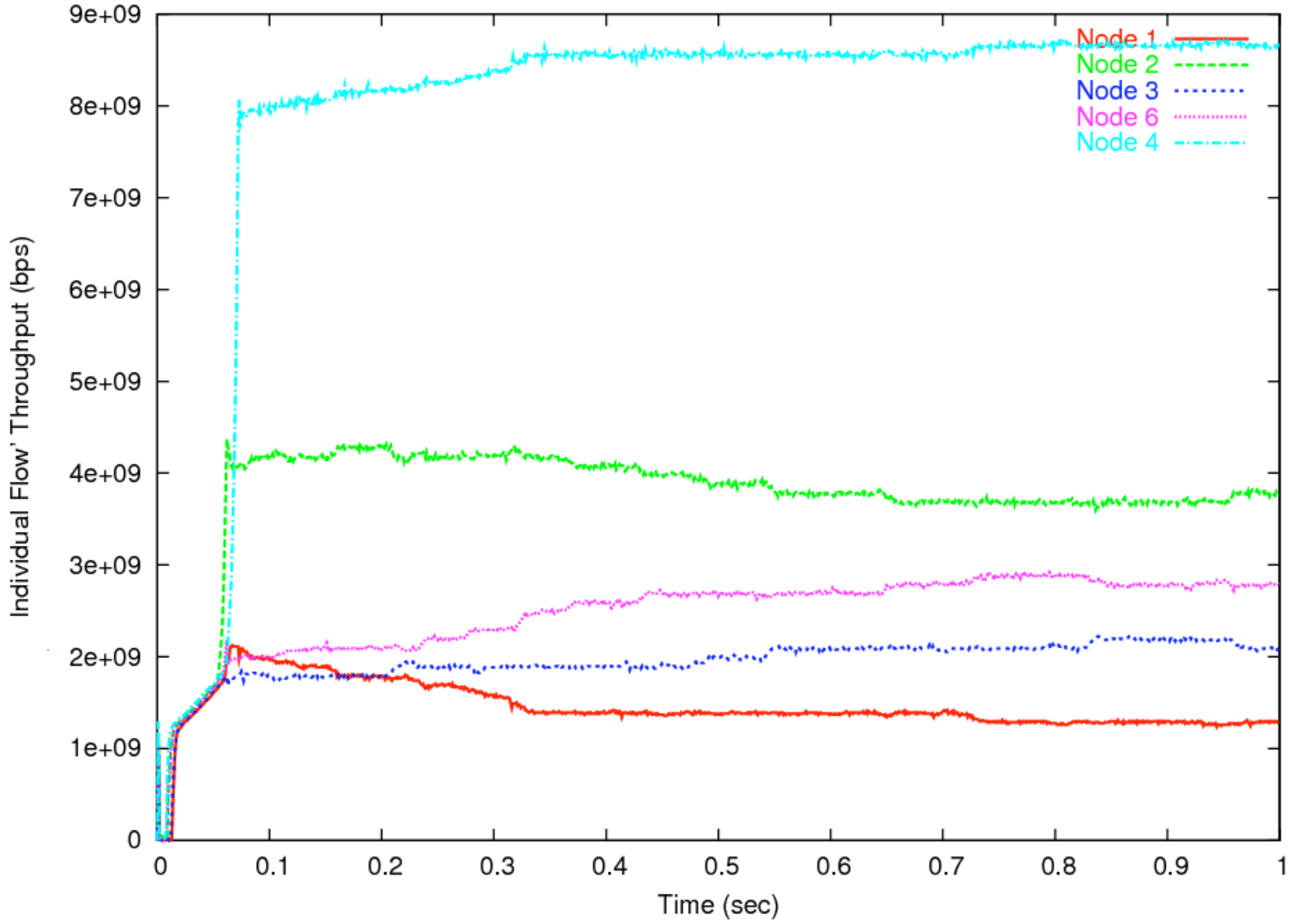
- Congested Points' Queue Size



Without Pause
- Congested Points' Throughput

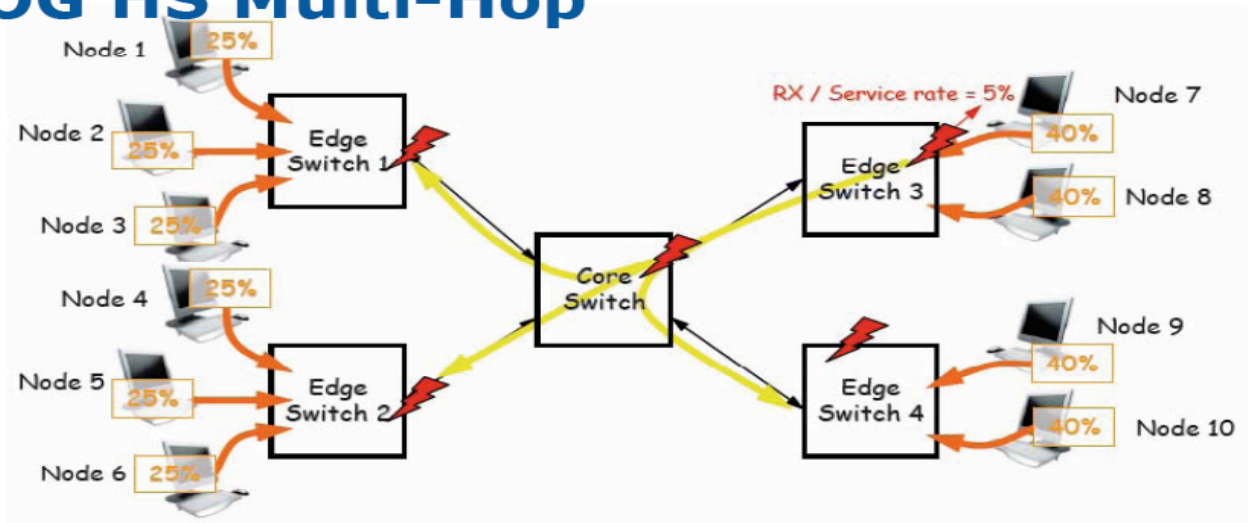


Individual Flow Rates



Baseline #2

2. OG HS Multi-Hop



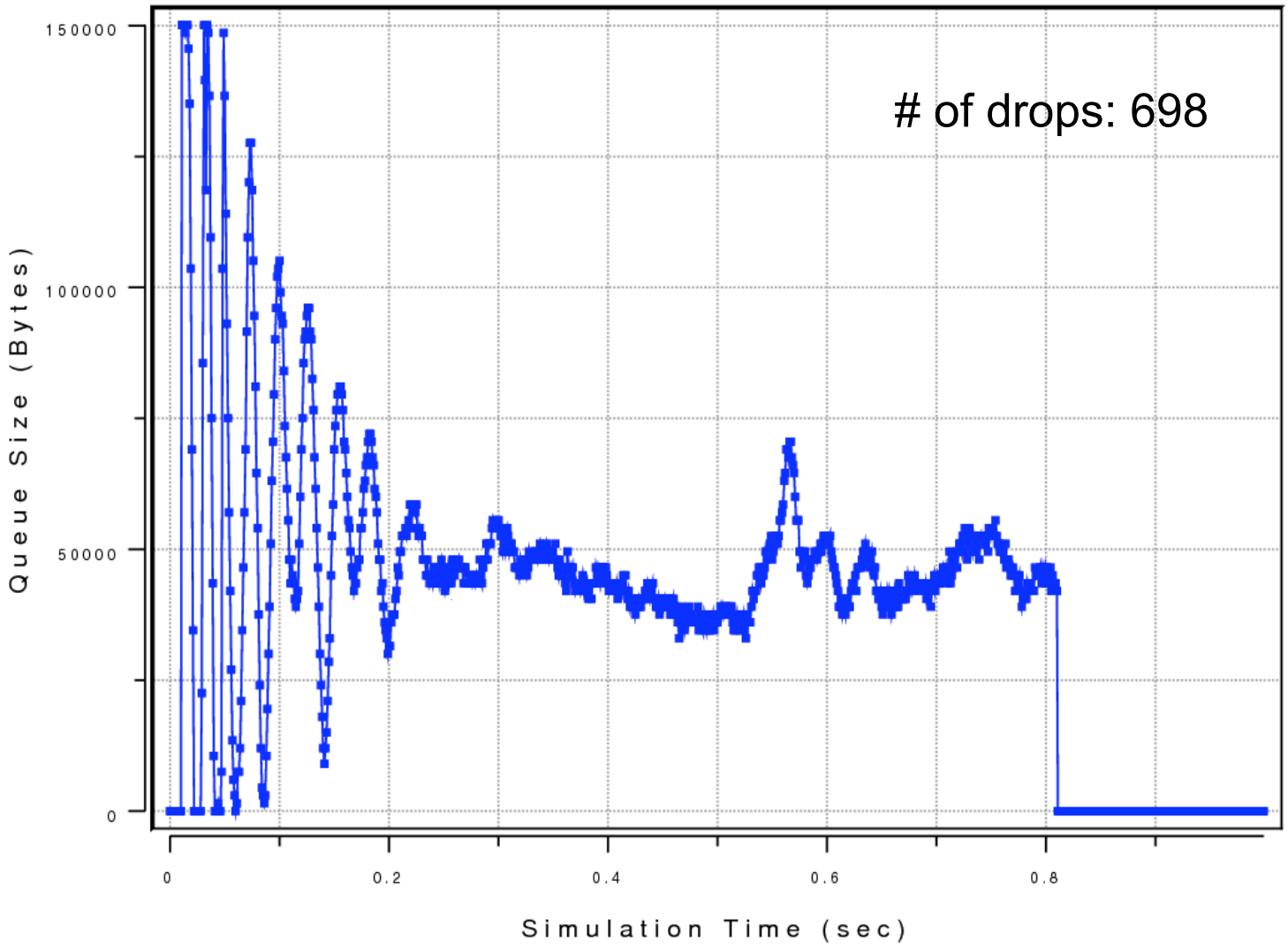
Workload:

- All: Uniform distribution traffic (background traffic)
- Nodes 1-6: 25% (2.5Gbps), Nodes 7-10: 40% (4 Gbps)
- Primary Hotspot:
  - Node 7 service rate = 5% (Rx only)
  - If saturation tree spreads => 5 congestion points total

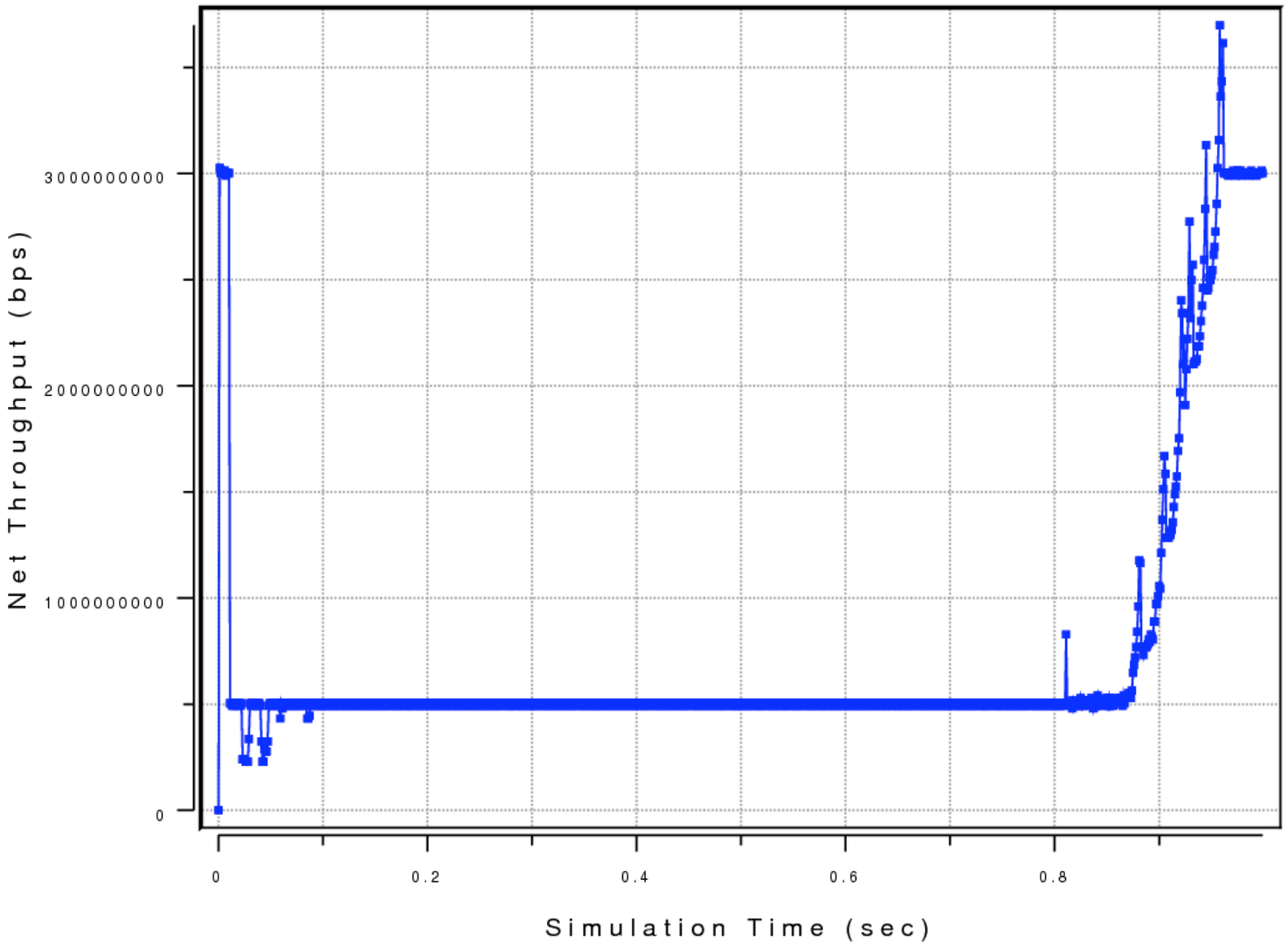
- Scenarios:
  - PAUSE: Enabled/Disabled

Required

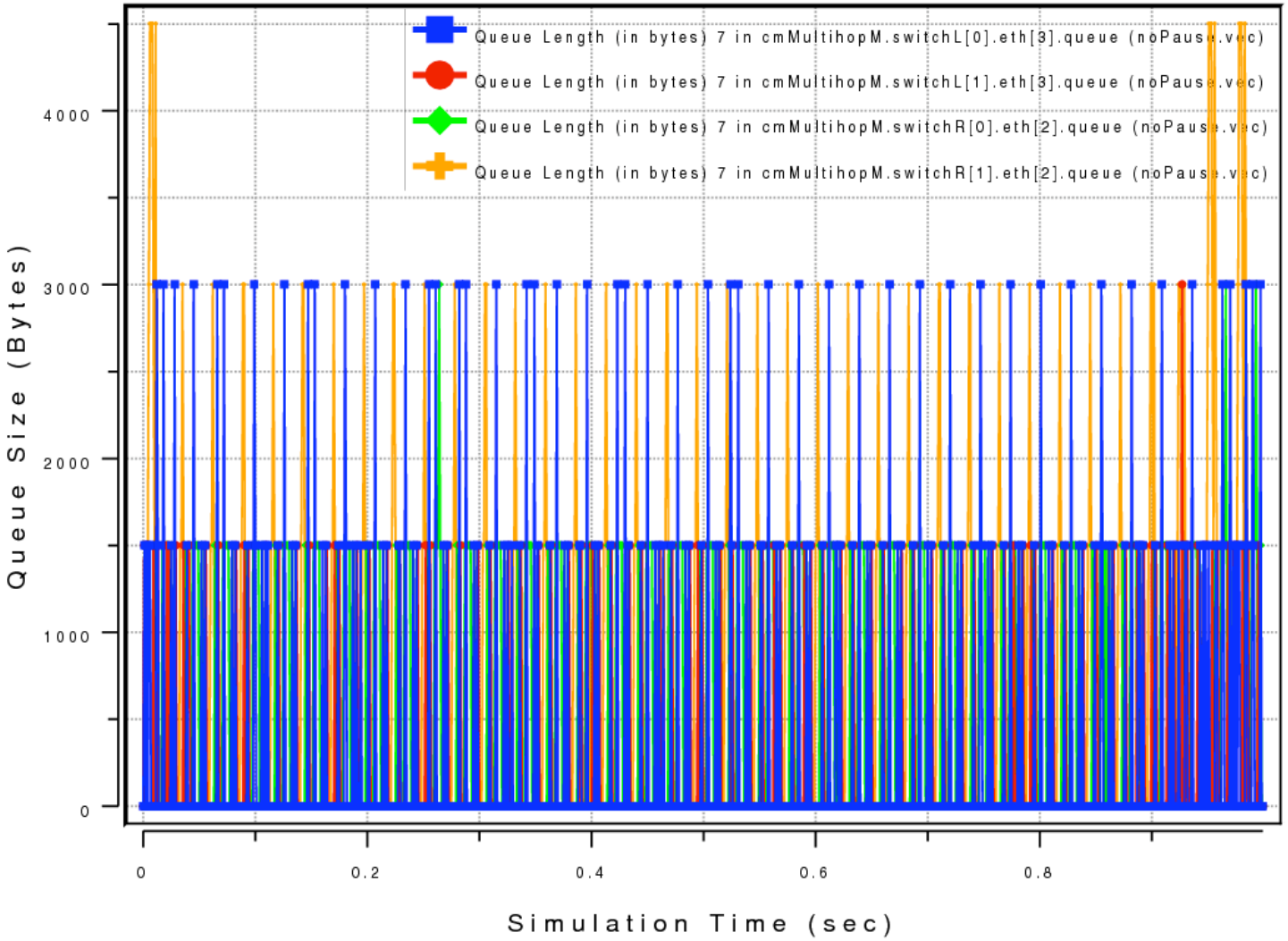
Without Pause
- Congested Queue Size (0.5Gbps)



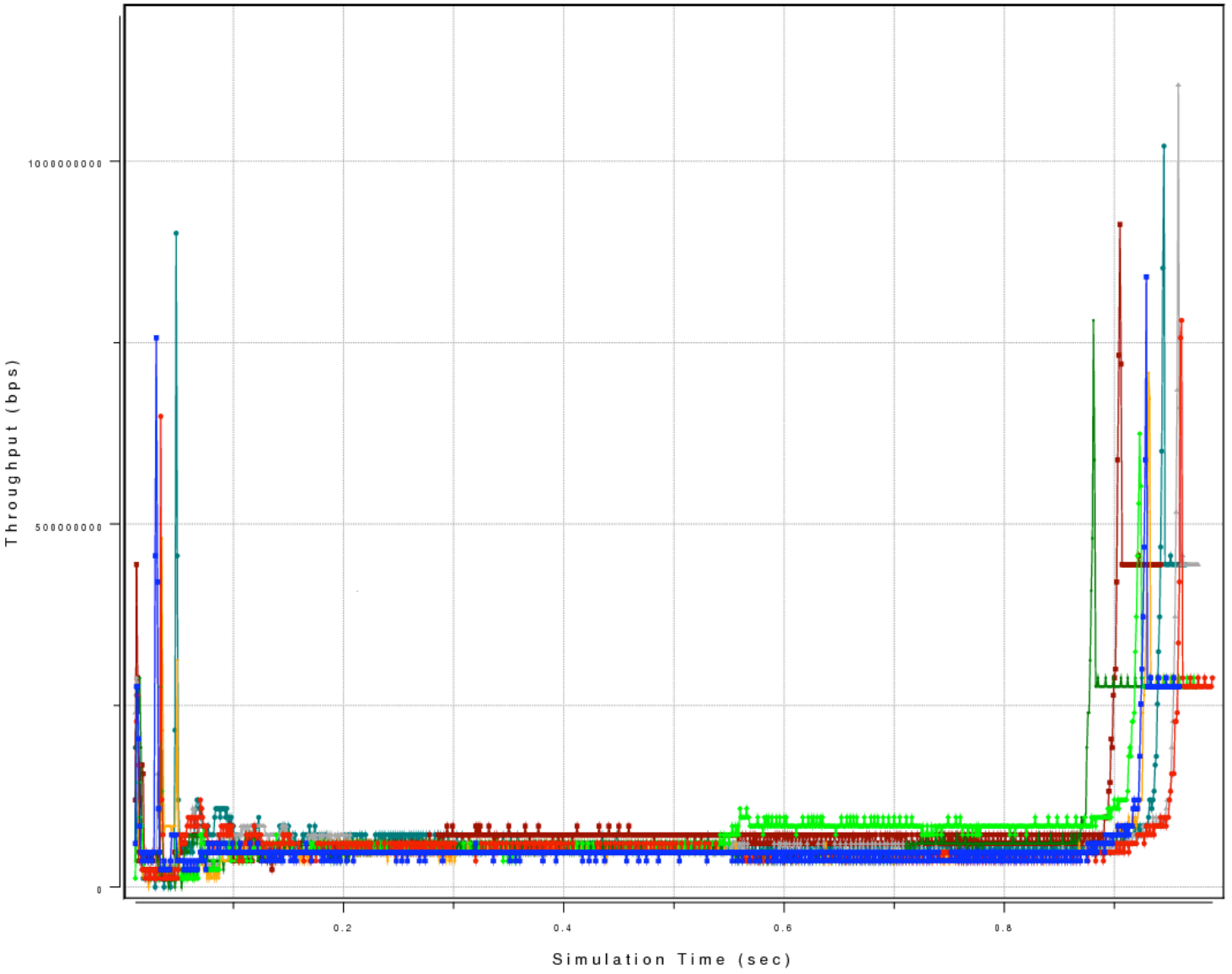
Benchmark 2 – Bottleneck Throughput



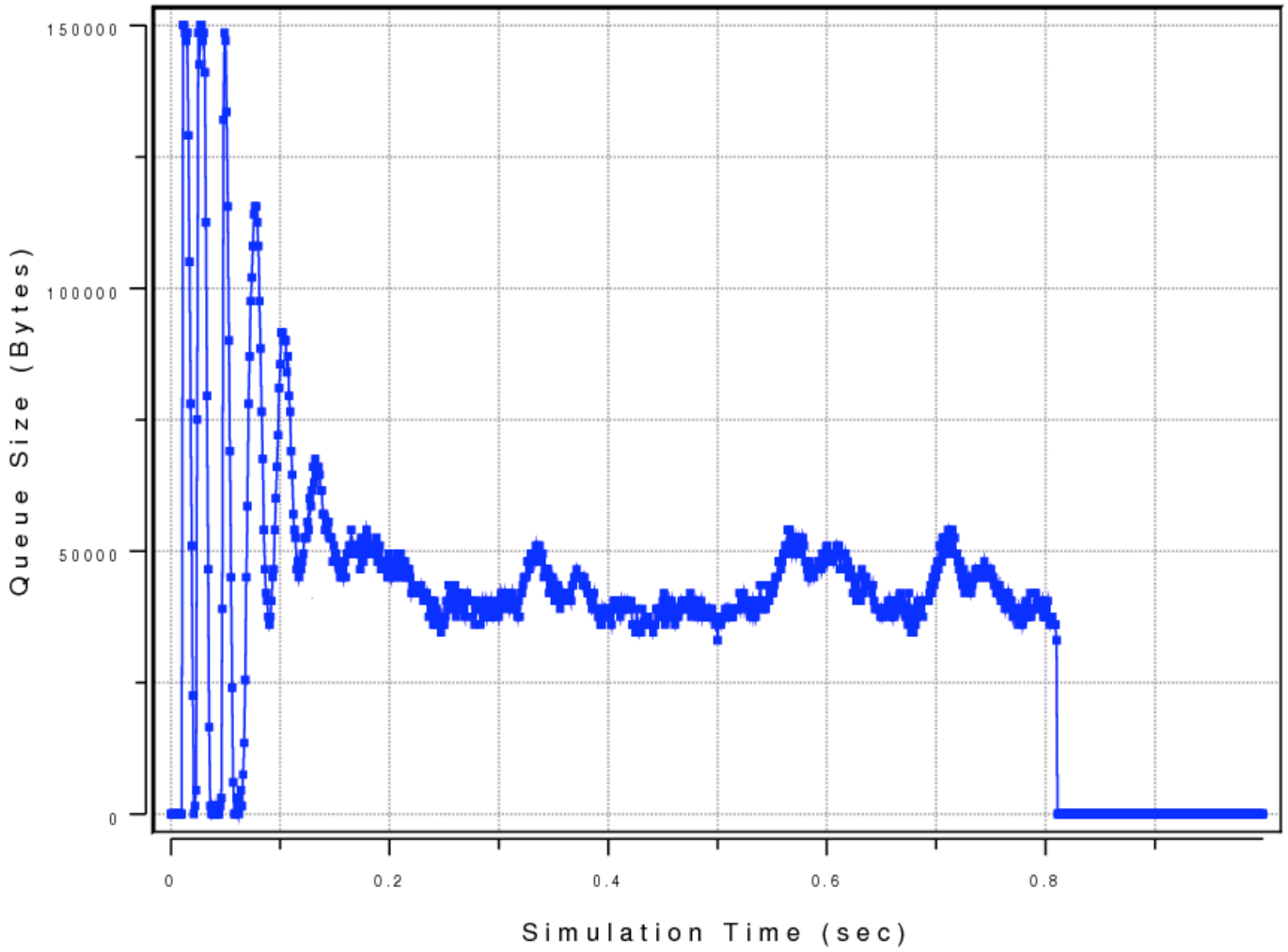
Without Pause
- Uncongested Queue Size (0.5Gbps)



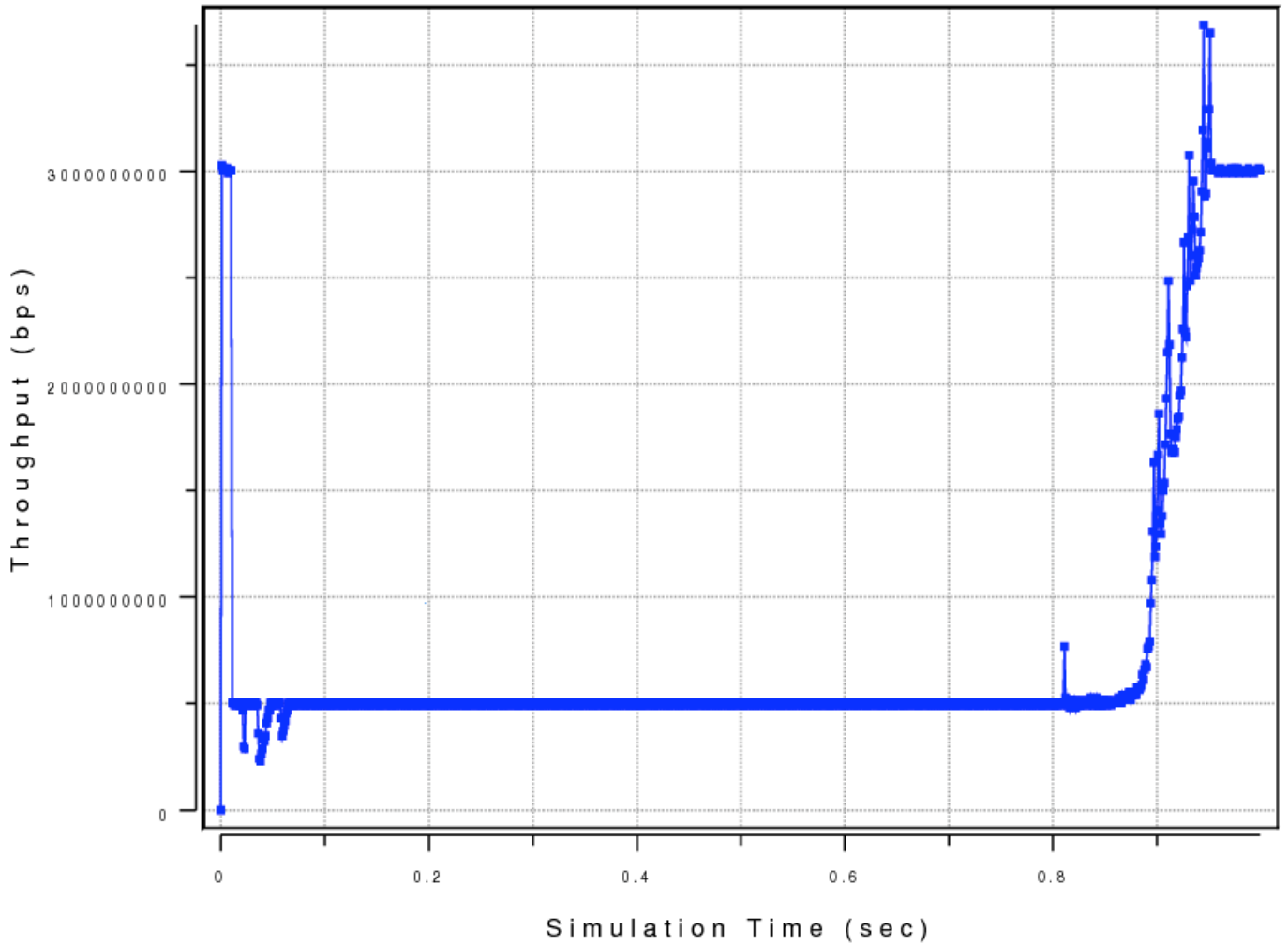
Benchmark 2 – Individual Flows' Throughput



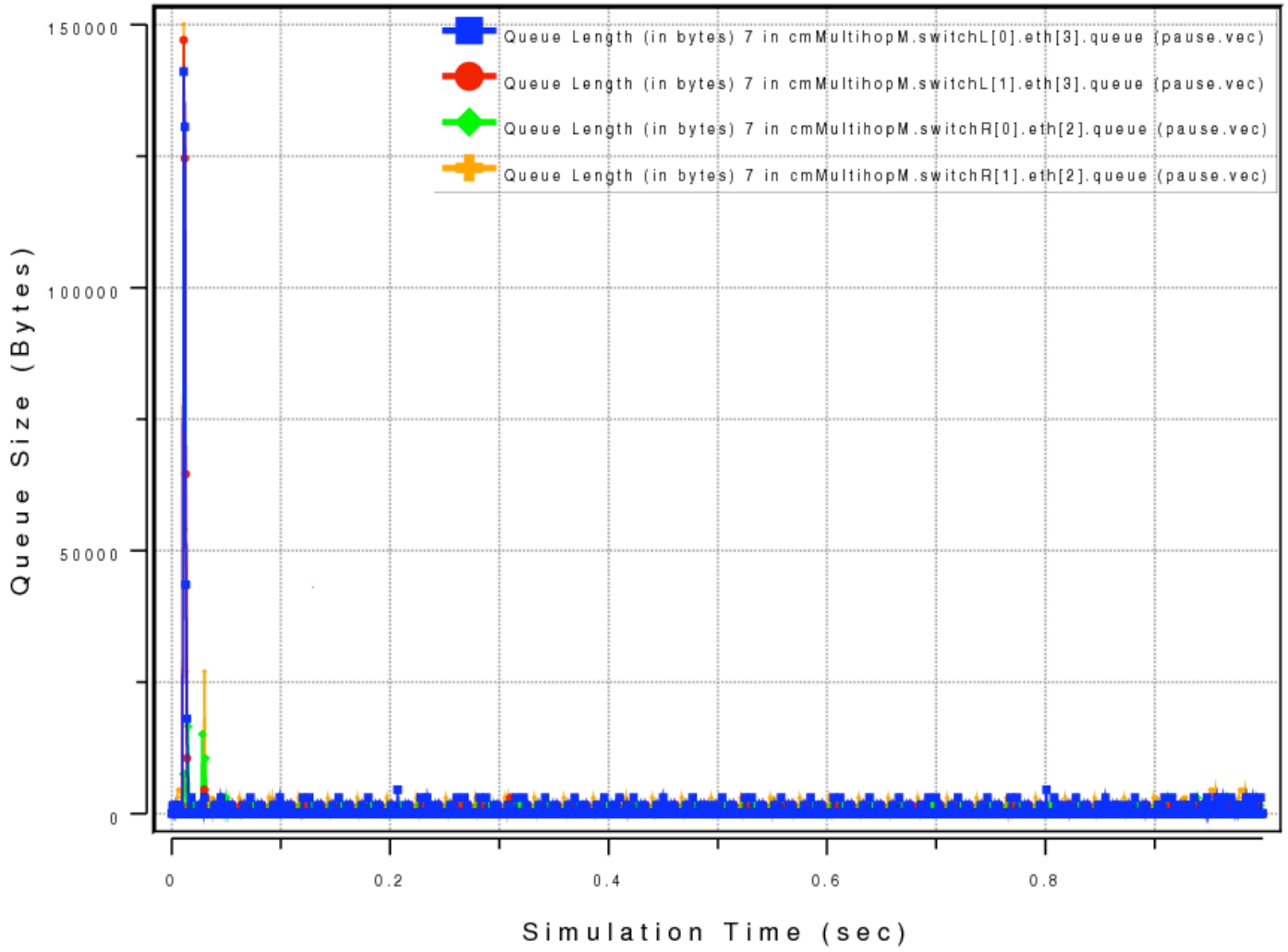
With Pause
- Bottleneck Queue Size



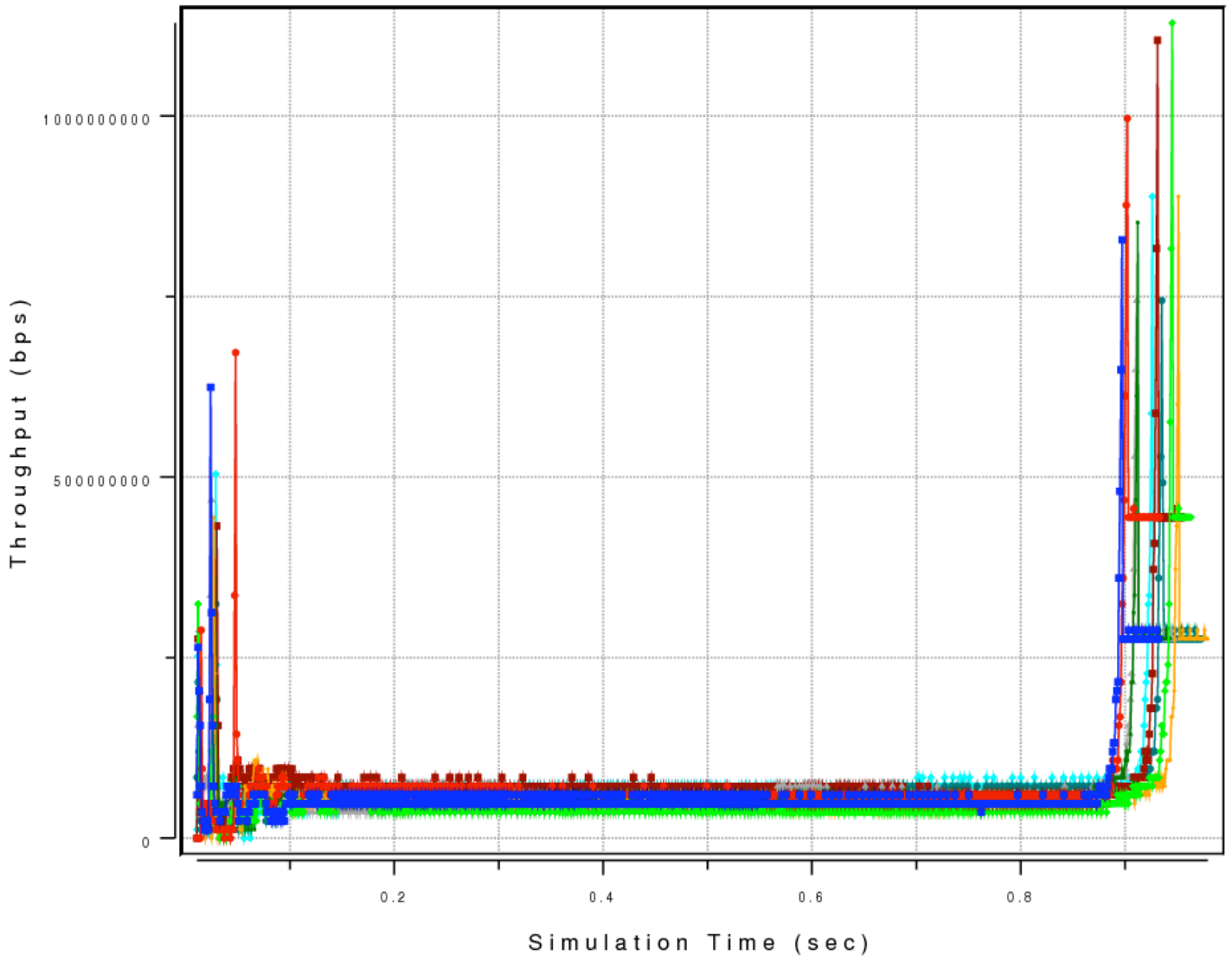
Benchmark 2 – With Pause
- Bottleneck Link Throughput (bps)



Benchmark 2 – With Pause - Uncongested Queue Size

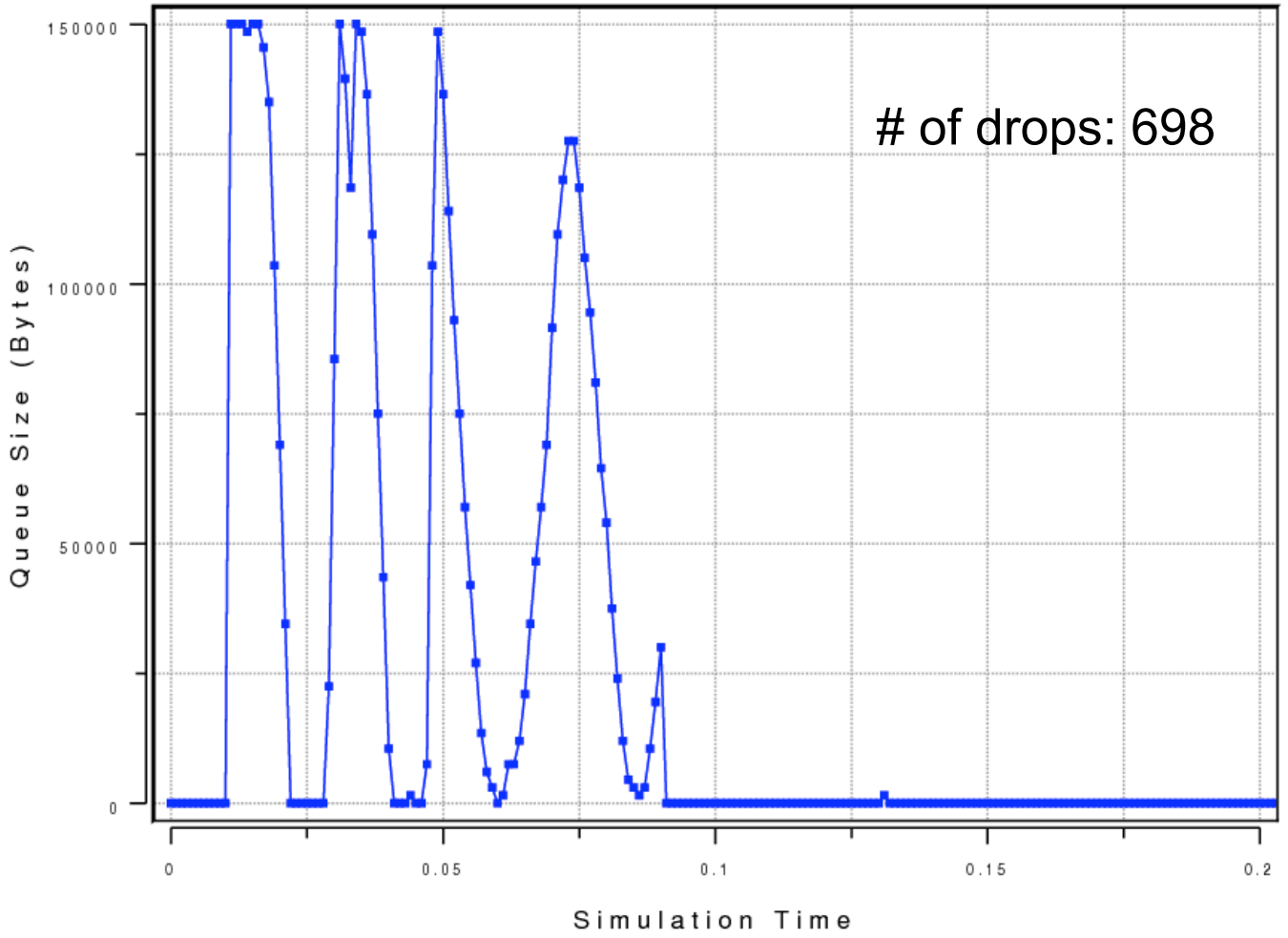


Benchmark 2 – With Pause
- Individual Flow Throughput (bps)

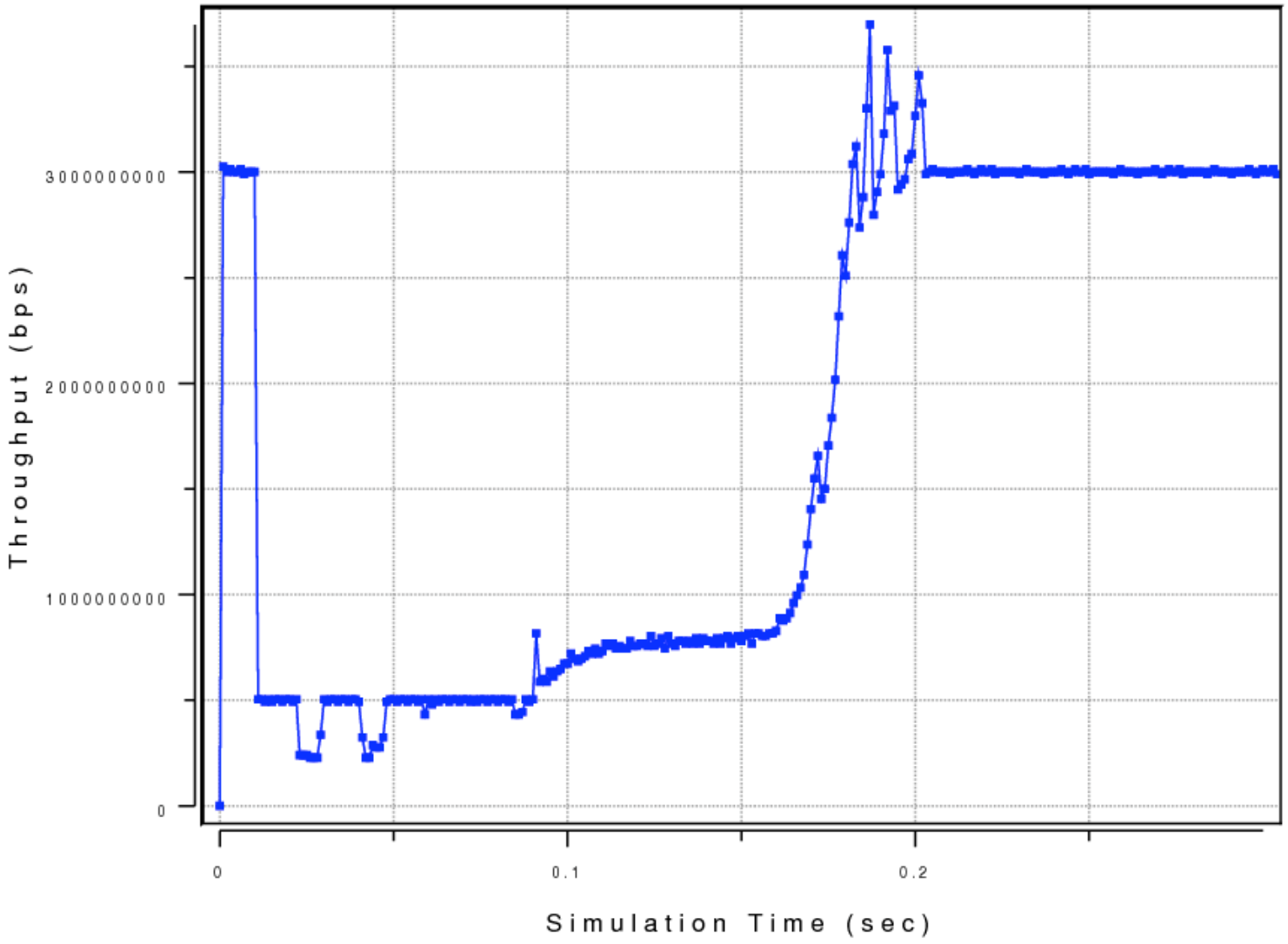


Scenario 2: 80ms Hot Spot Duration

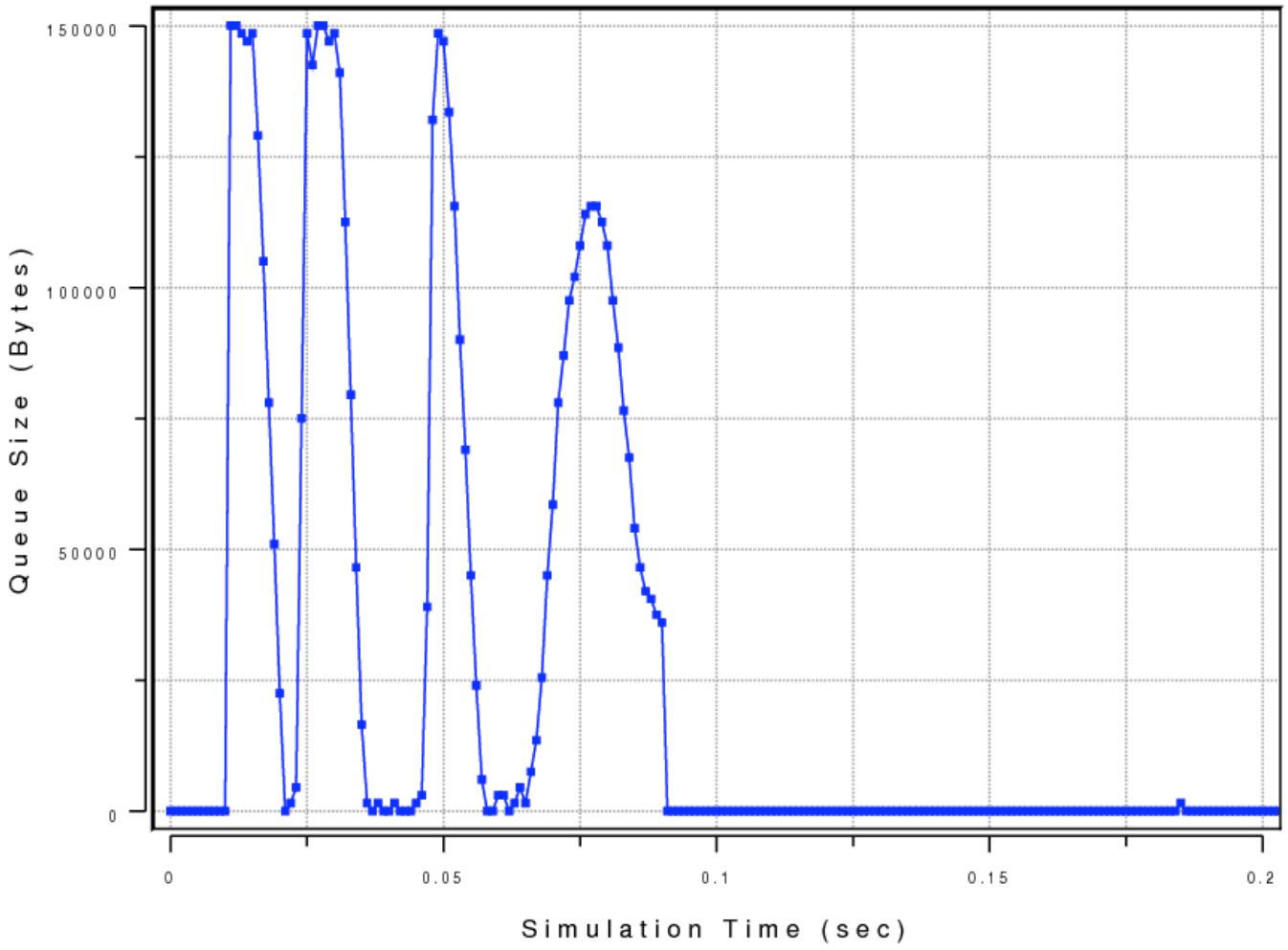
Without Pause
(80ms Hot Spot Duration)
- Congested Queue Size (0.5Gbps)



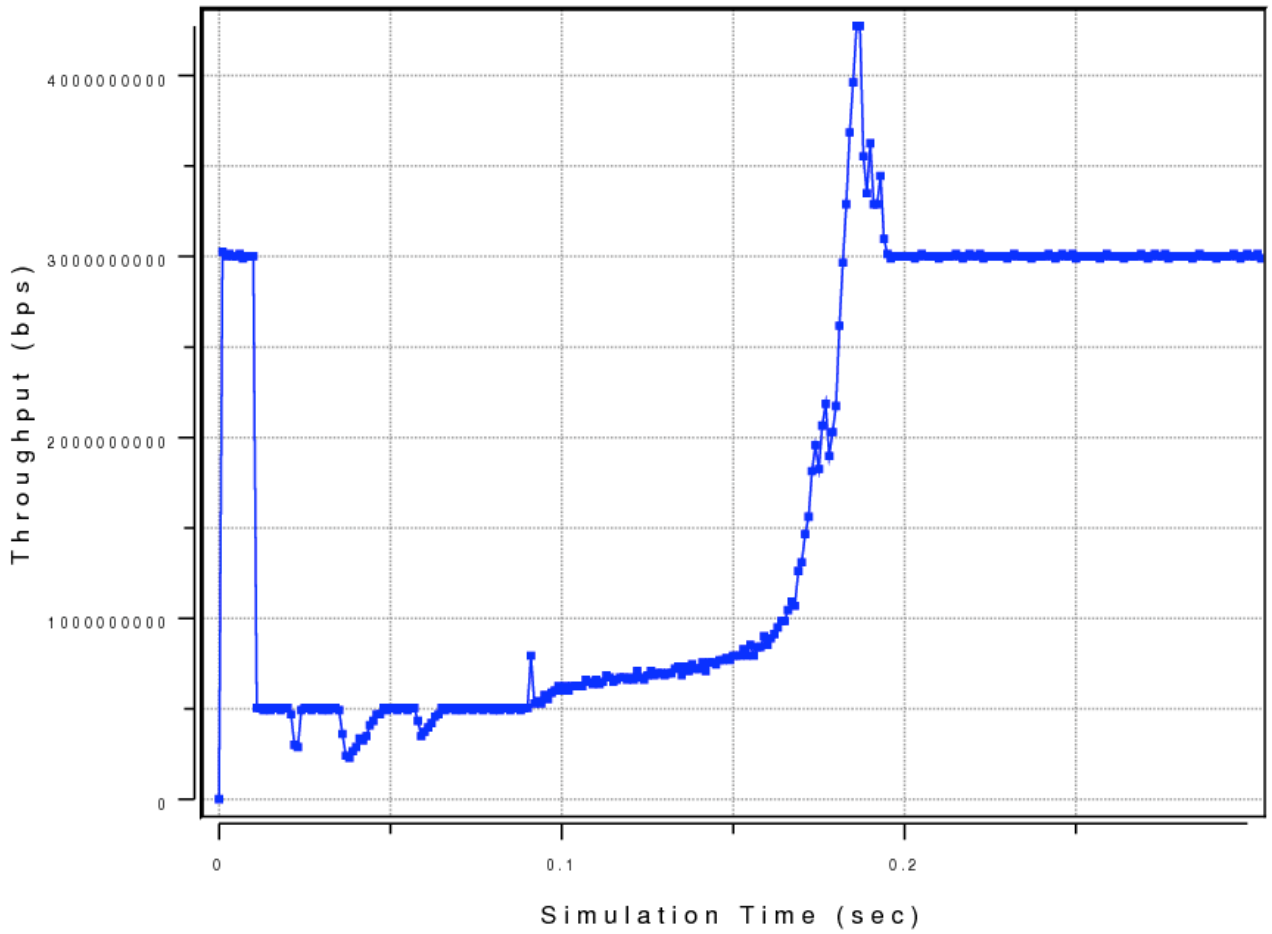
Benchmark 2 (80ms Hot Spot Duration) – Bottleneck Throughput



With Pause
(80ms Hot Spot Duration)
- Congested Queue Size (0.5Gbps)

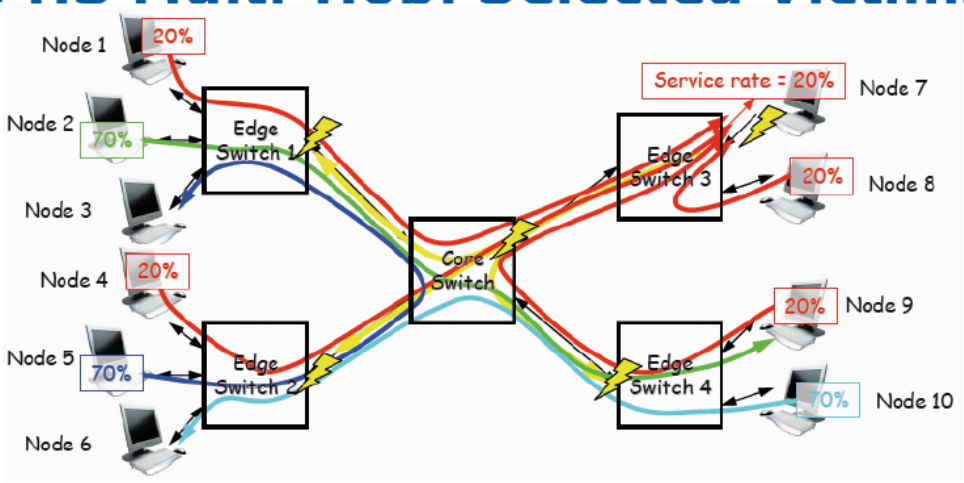


Benchmark 2 (80ms Hot Spot Duration) – Bottleneck Throughput



Benchmark 3

3. OG HS Multi-Hop: Selected Victims

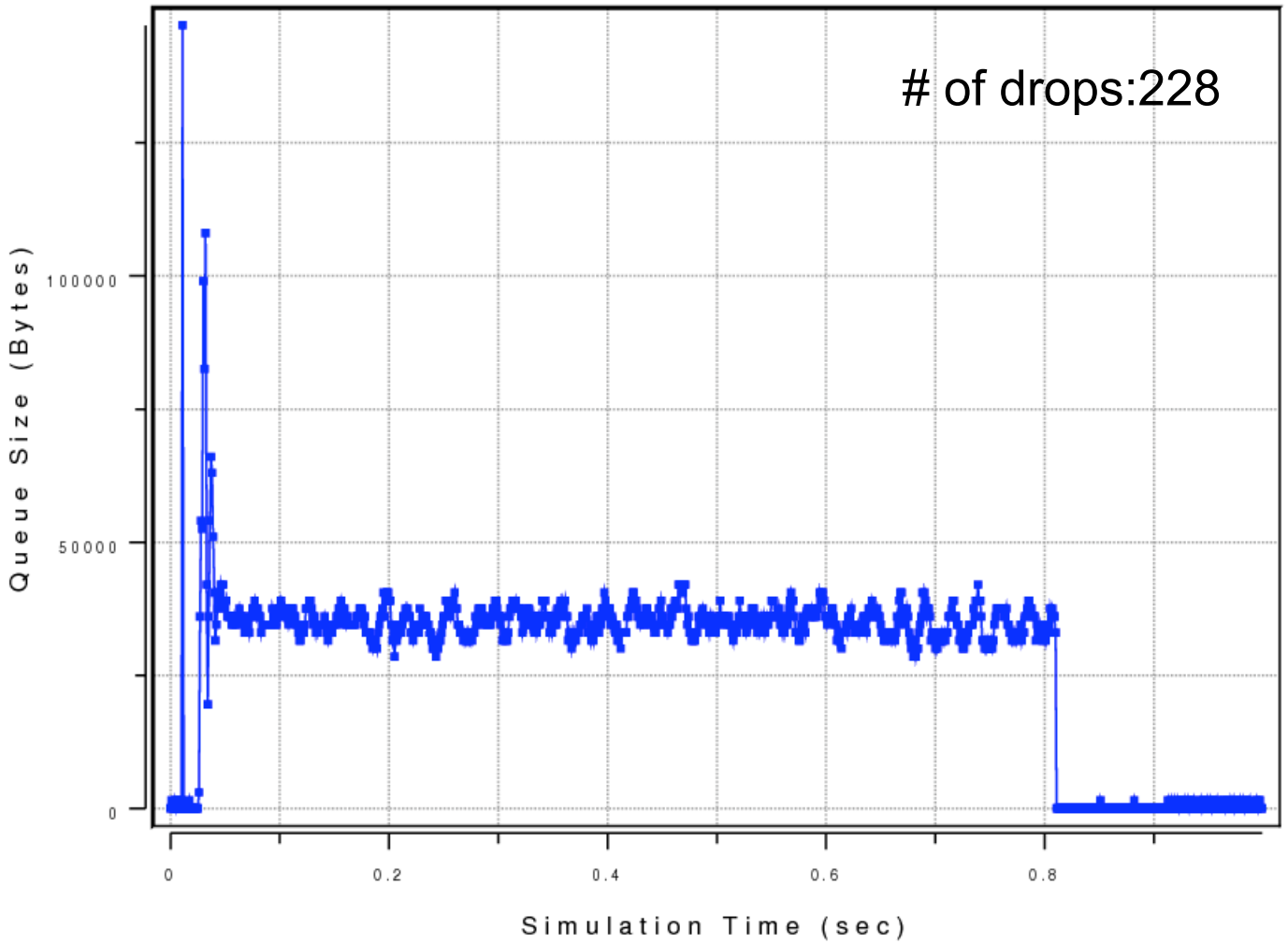


Workload:

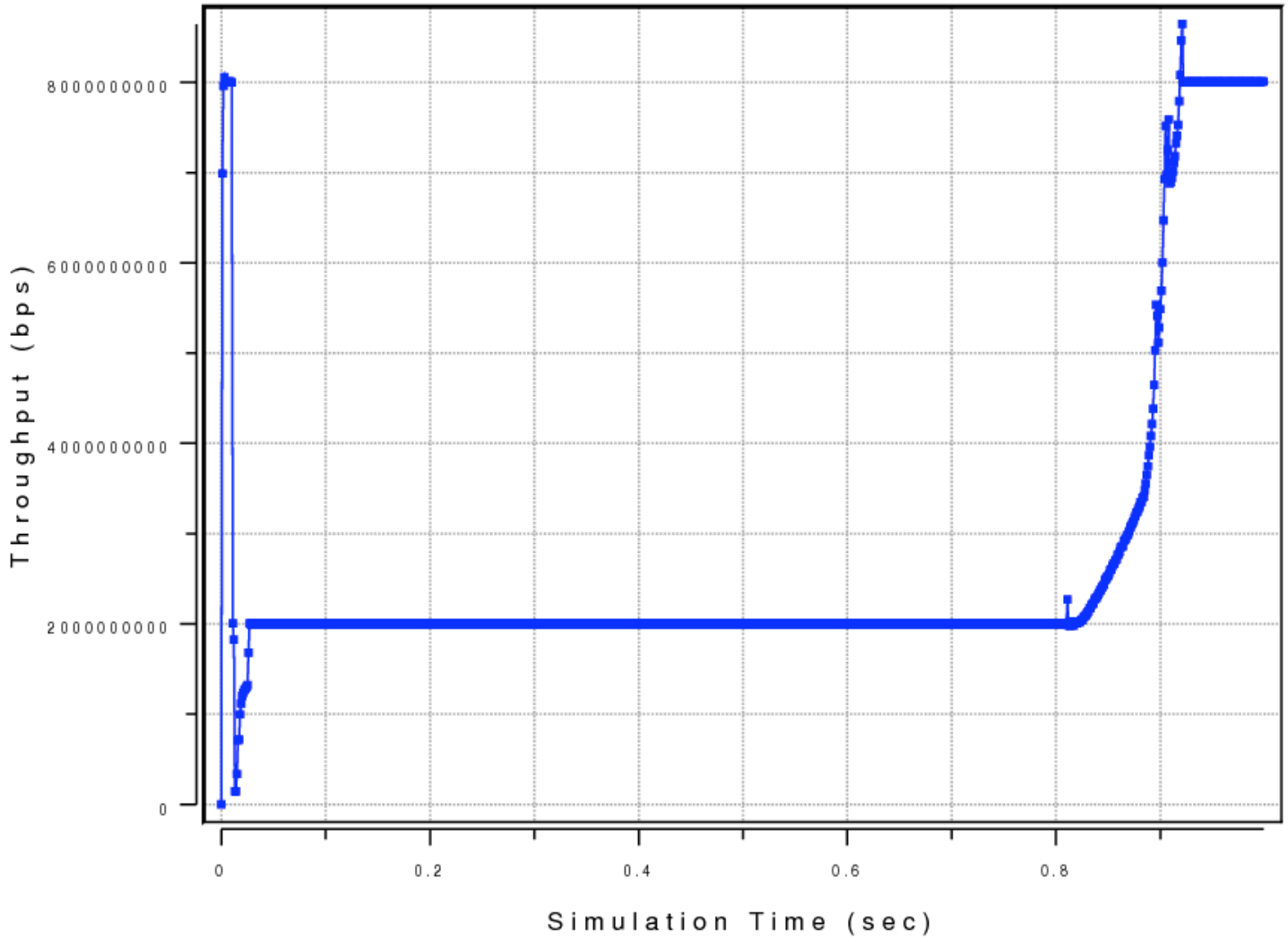
- Four culprit flows of 2 Gb/s each from nodes 1, 4, 8, 9 to node 7 (hotspot)
- Three victim flows of 7 Gb/s each: node 2 to 9, node 5 to 3, node 10 to 6
- Node 7 service rate = 20%
- Five congestion points, All switches and all flows affected
- Fair allocation provides 0.5 Gb/s to all culprits and 7 Gb/s to all victim

Required

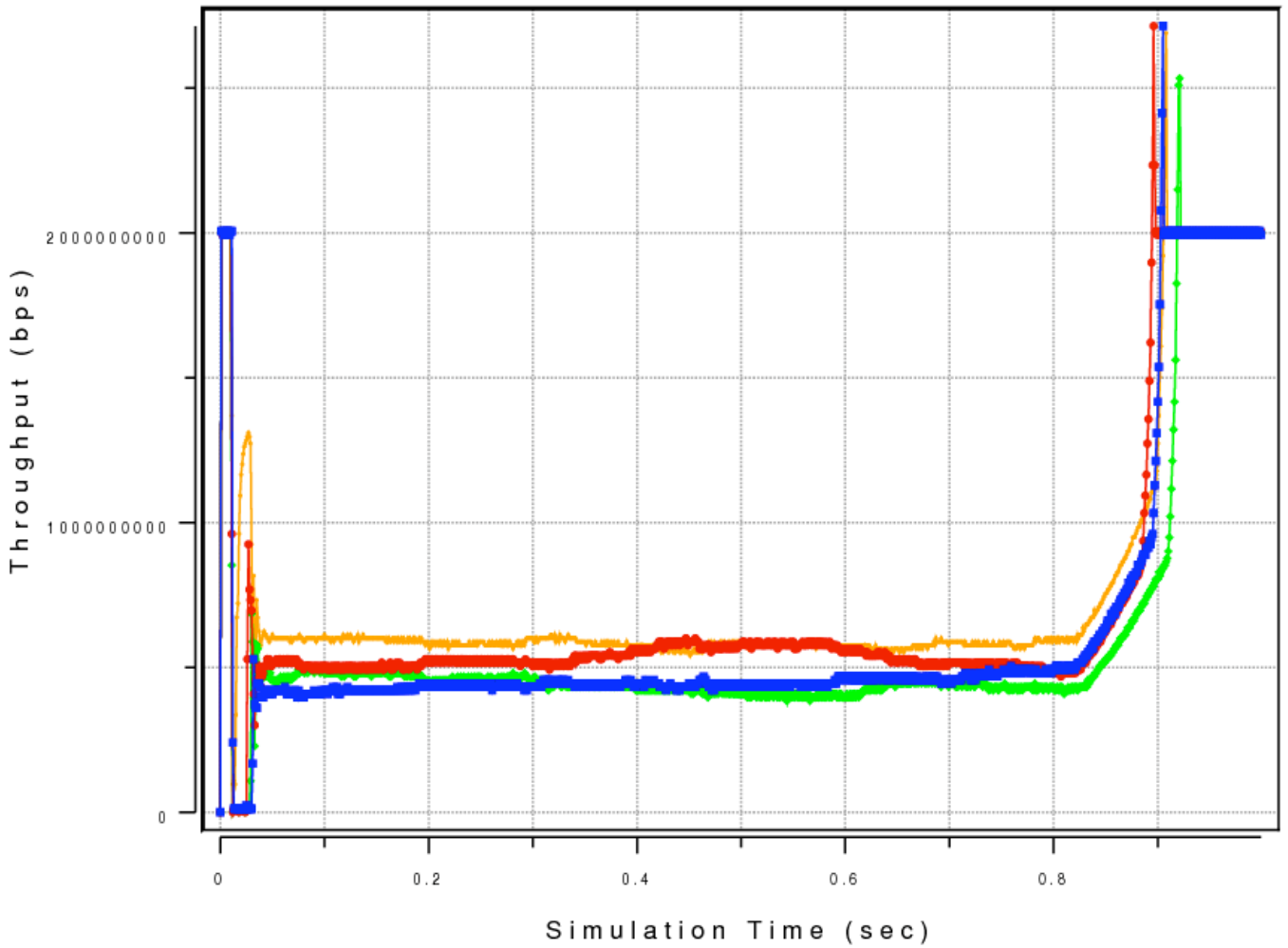
Without Pause
- Bottleneck Queue Size (2Gbps)



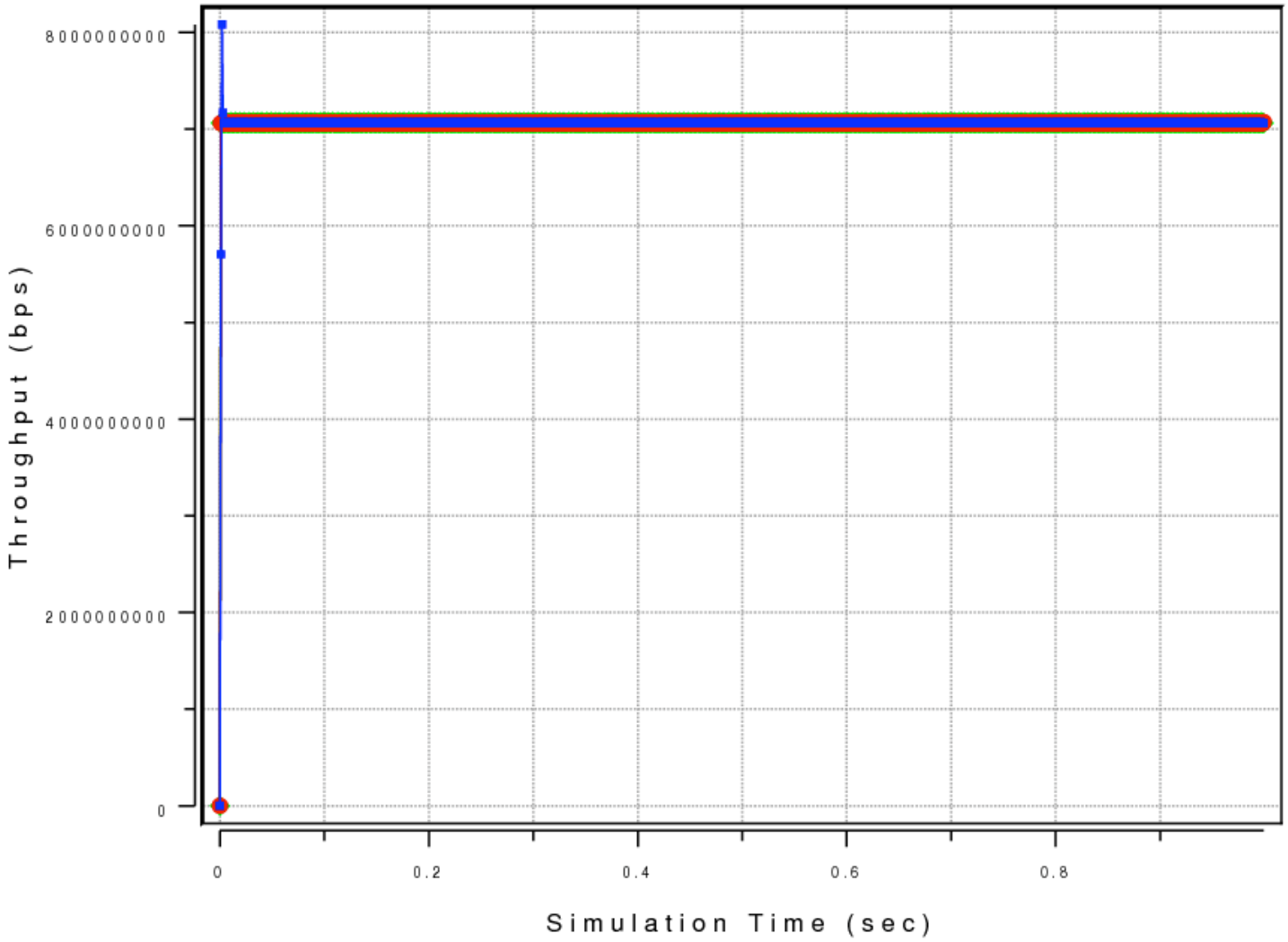
Without Pause
- Bottleneck Throughput



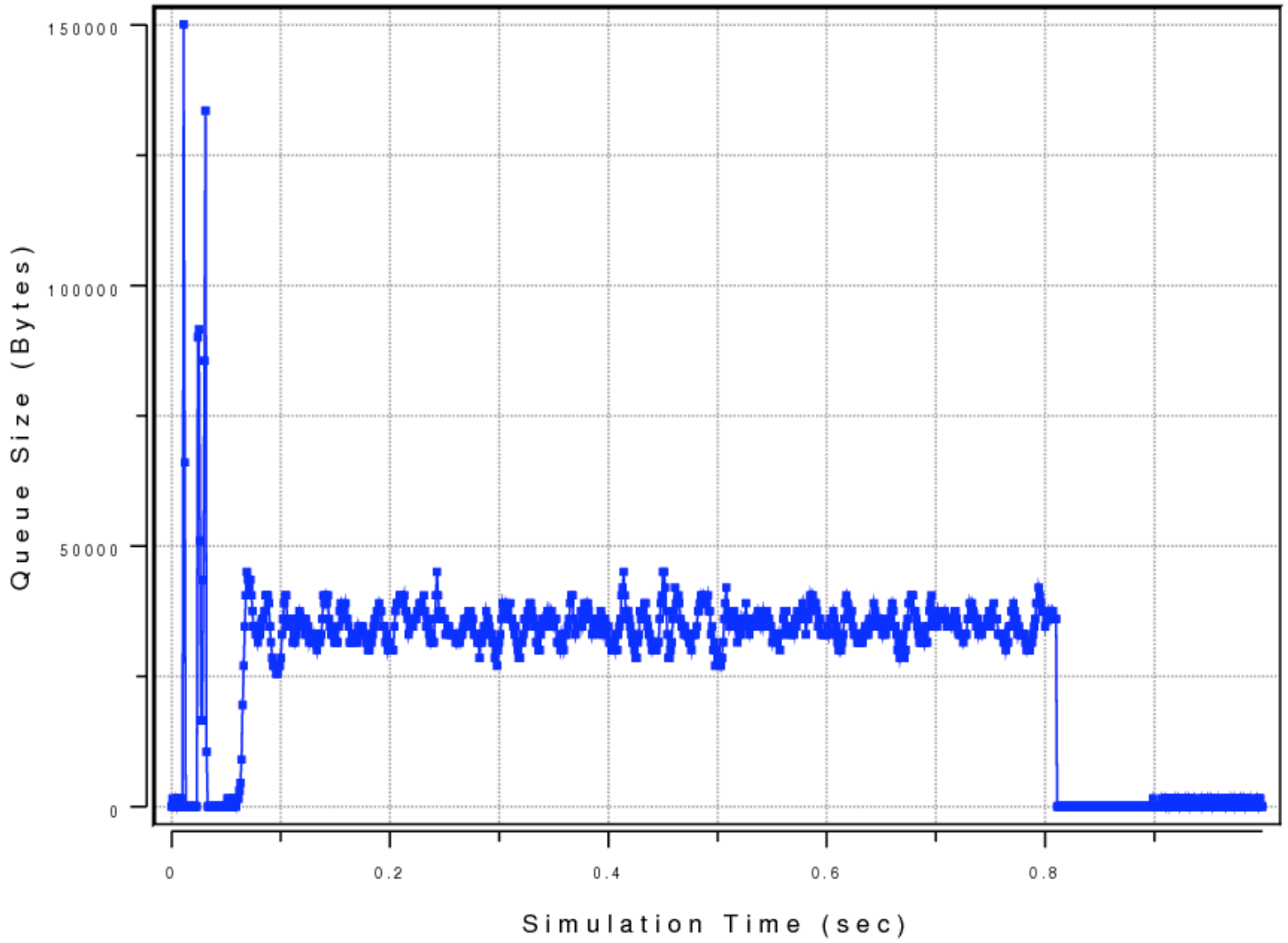
Culpit Flows - Throughput (0.5Gbps)



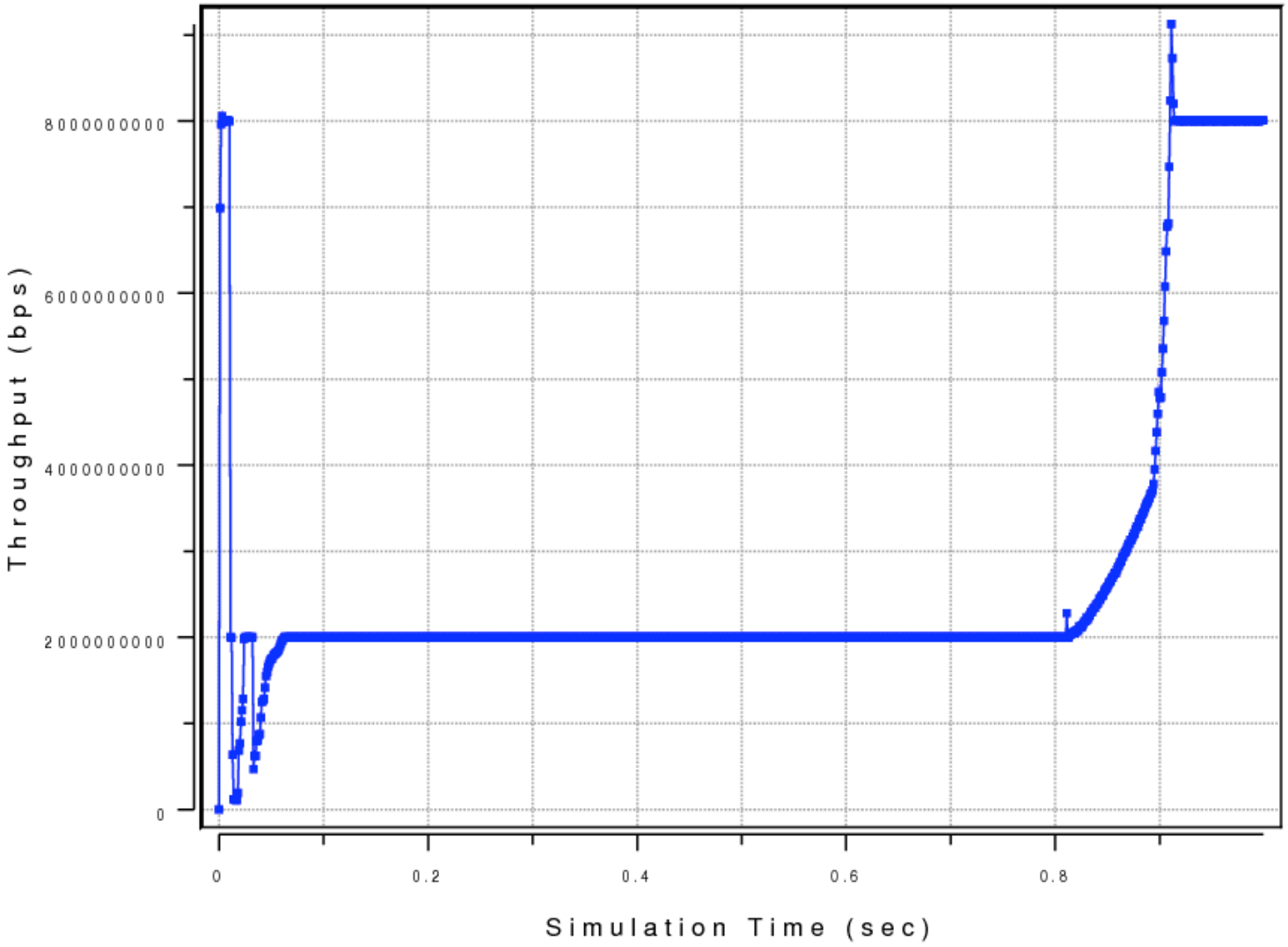
Victim Flows - Throughput (7Gbps)



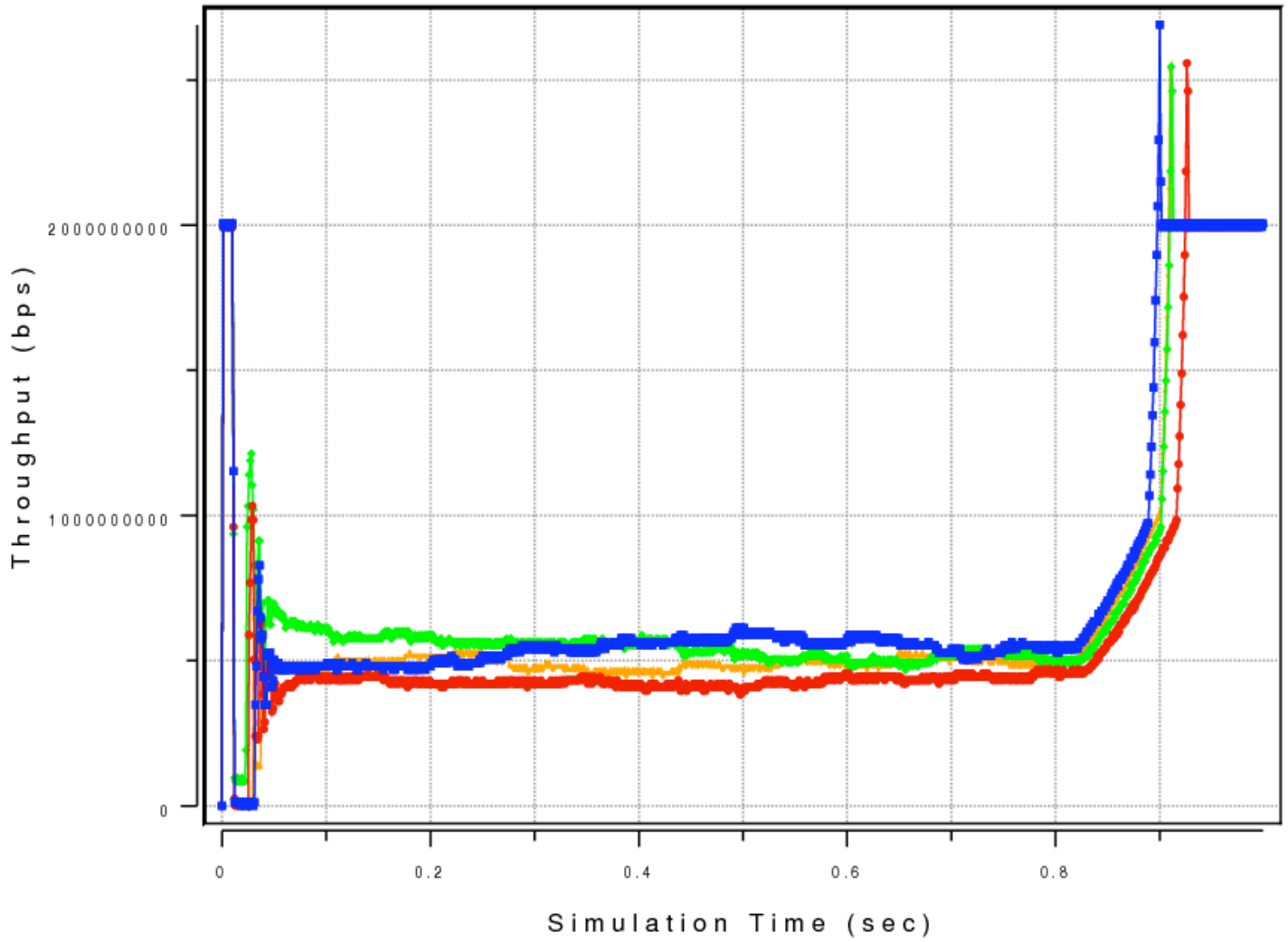
Wit Pause
- Bottleneck Queue Size (2Gbps)



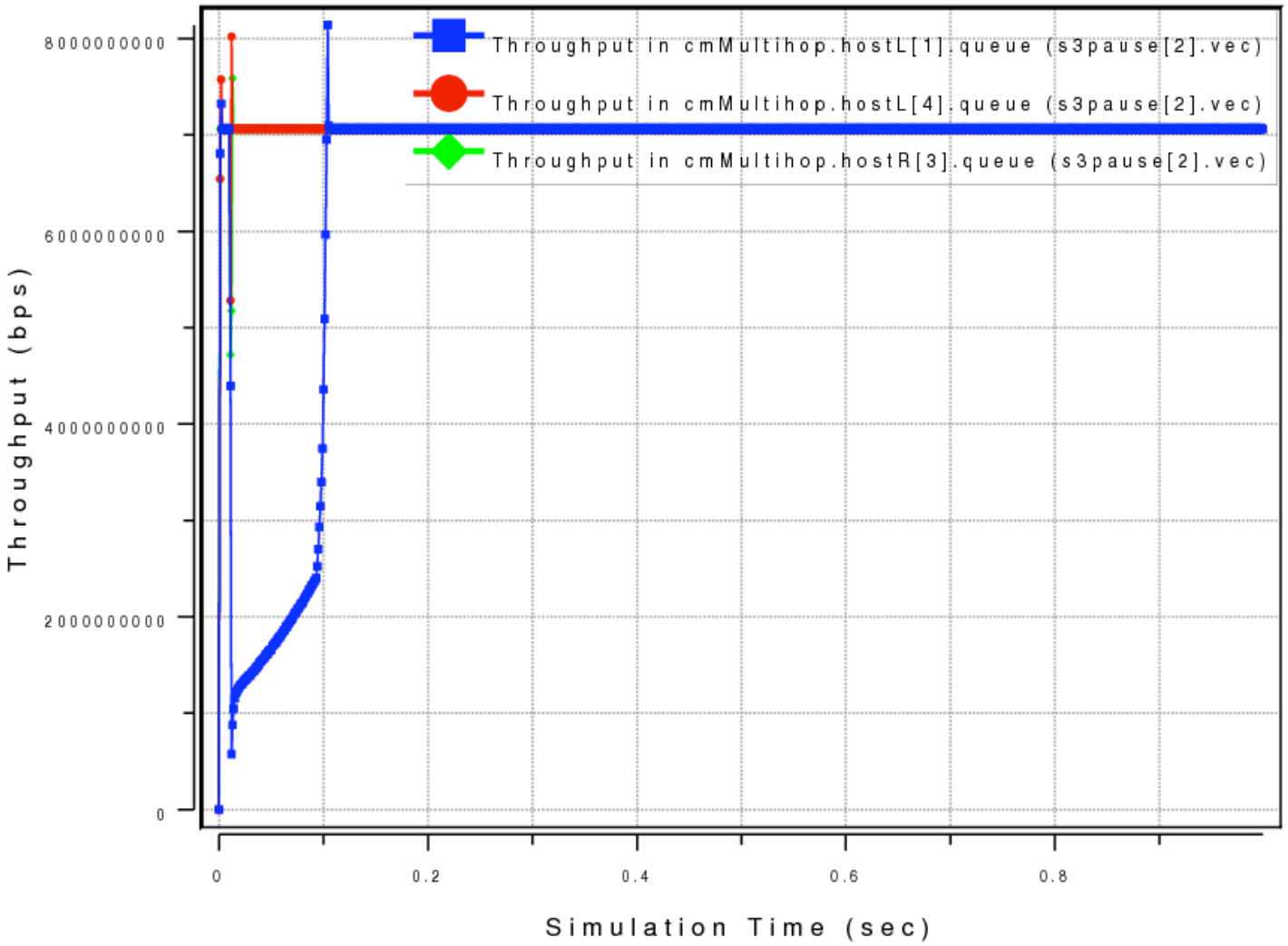
Wit Pause
- Bottleneck Throughput



Culpit Flows - Throughput (0.5Gbps)

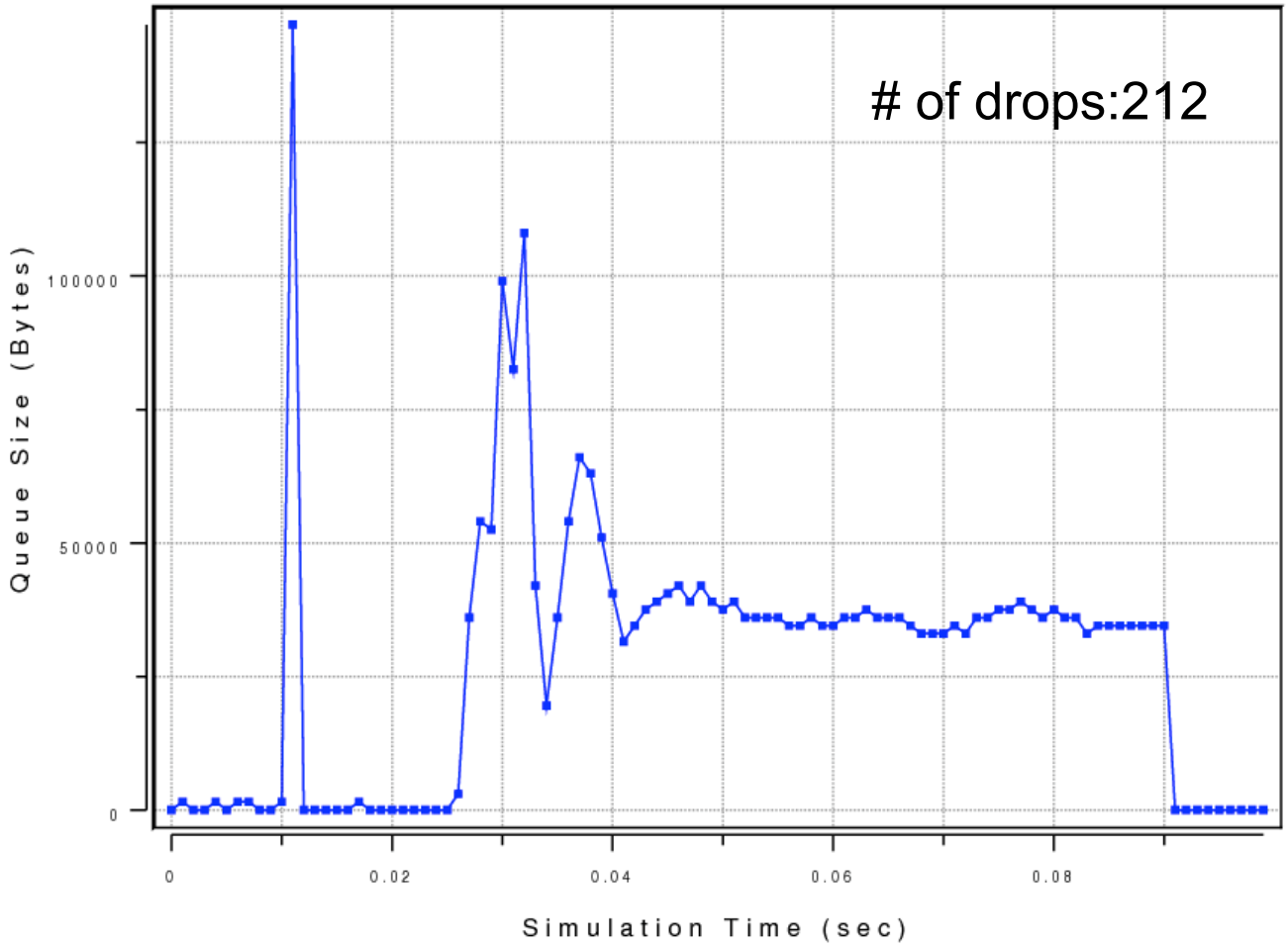


Victim Flows - Throughput (7Gbps)

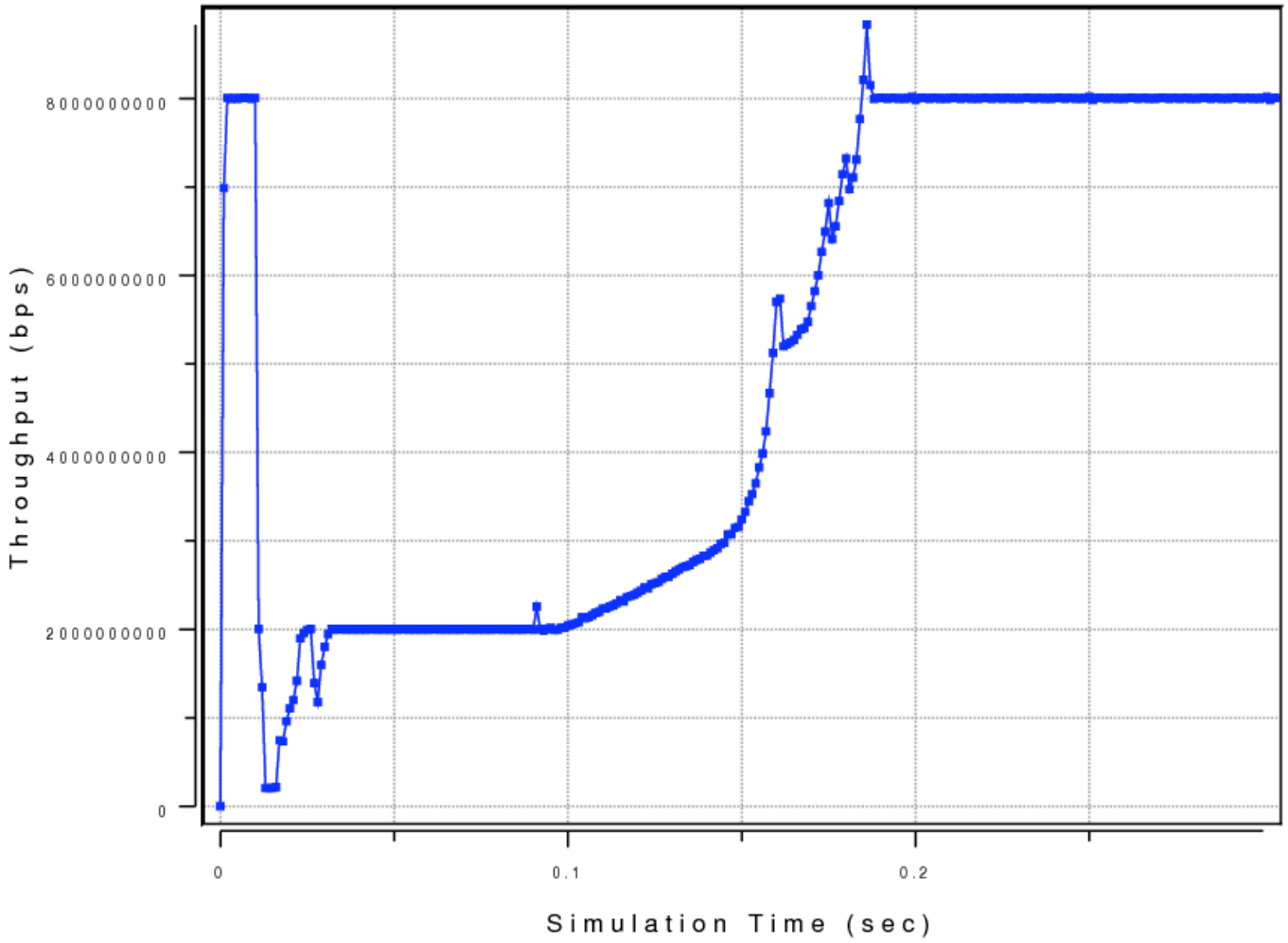


Scenario 3: 80ms Hot Spot Duration

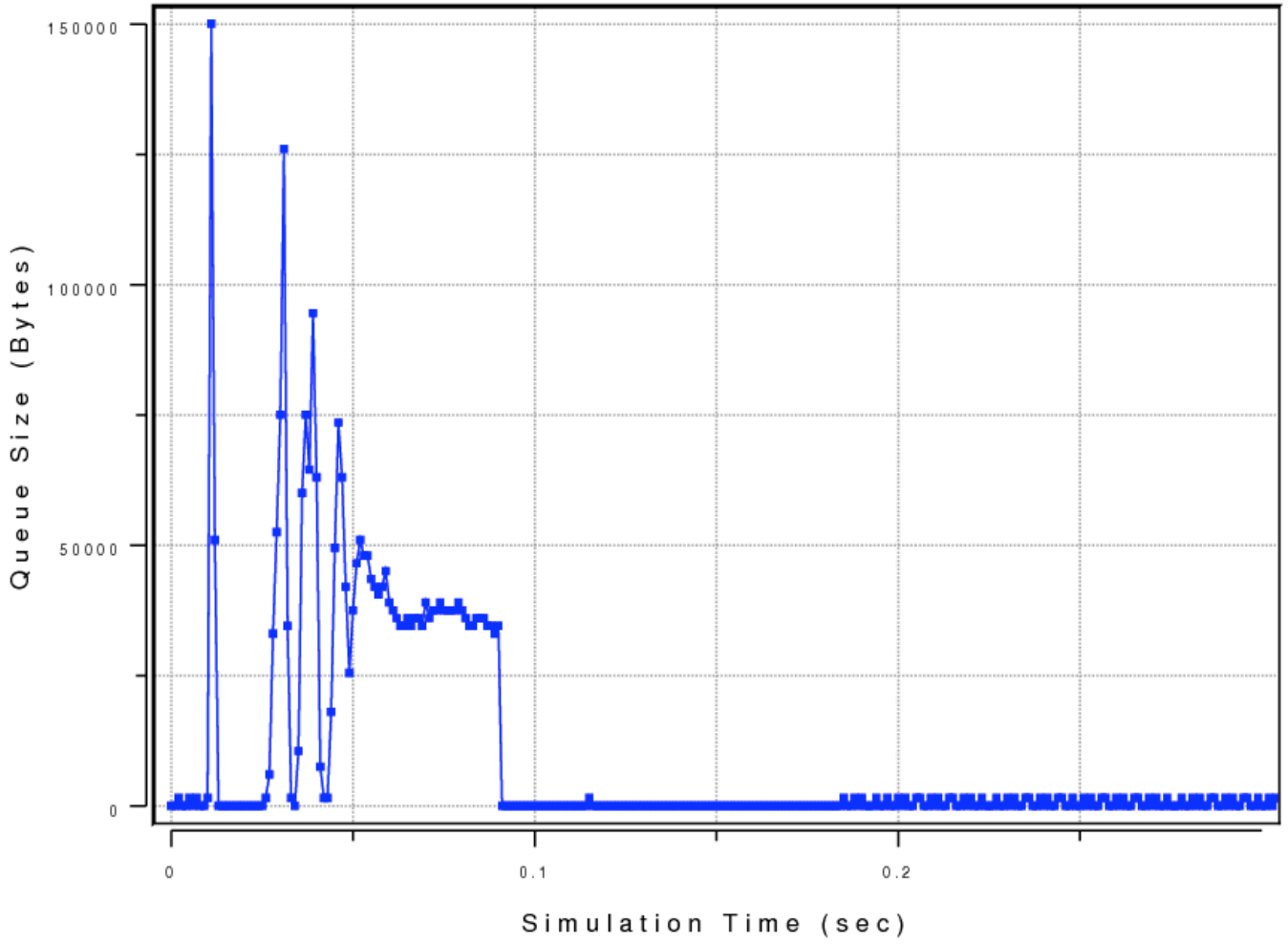
Without Pause
(80ms Hot Spot Duration)
- Congested Queue Size (2Gbps)



Benchmark 3 (80ms Hot Spot Duration) – Bottleneck Throughput



With Pause
(80ms Hot Spot Duration)
- Congested Queue Size (2 Gbps)



Benchmark 3
(80ms Hot Spot Duration)
– Bottleneck Throughput