QCN: Quantized Congestion Notification

Rong Pan, Balaji Prabhakar, Ashvin Laxmikantha

IEEE 802.1@Geneva
May 29, 2007

Overview

- Functional description of the QCN
  - Pseudo-code available with Rong Pan: ropan@cisco.com

- Basic simulations
  - Infinitely long-lived flows: stability of control loop
  - Dynamic flows: FCT
  - Baseline Simulations

QCN

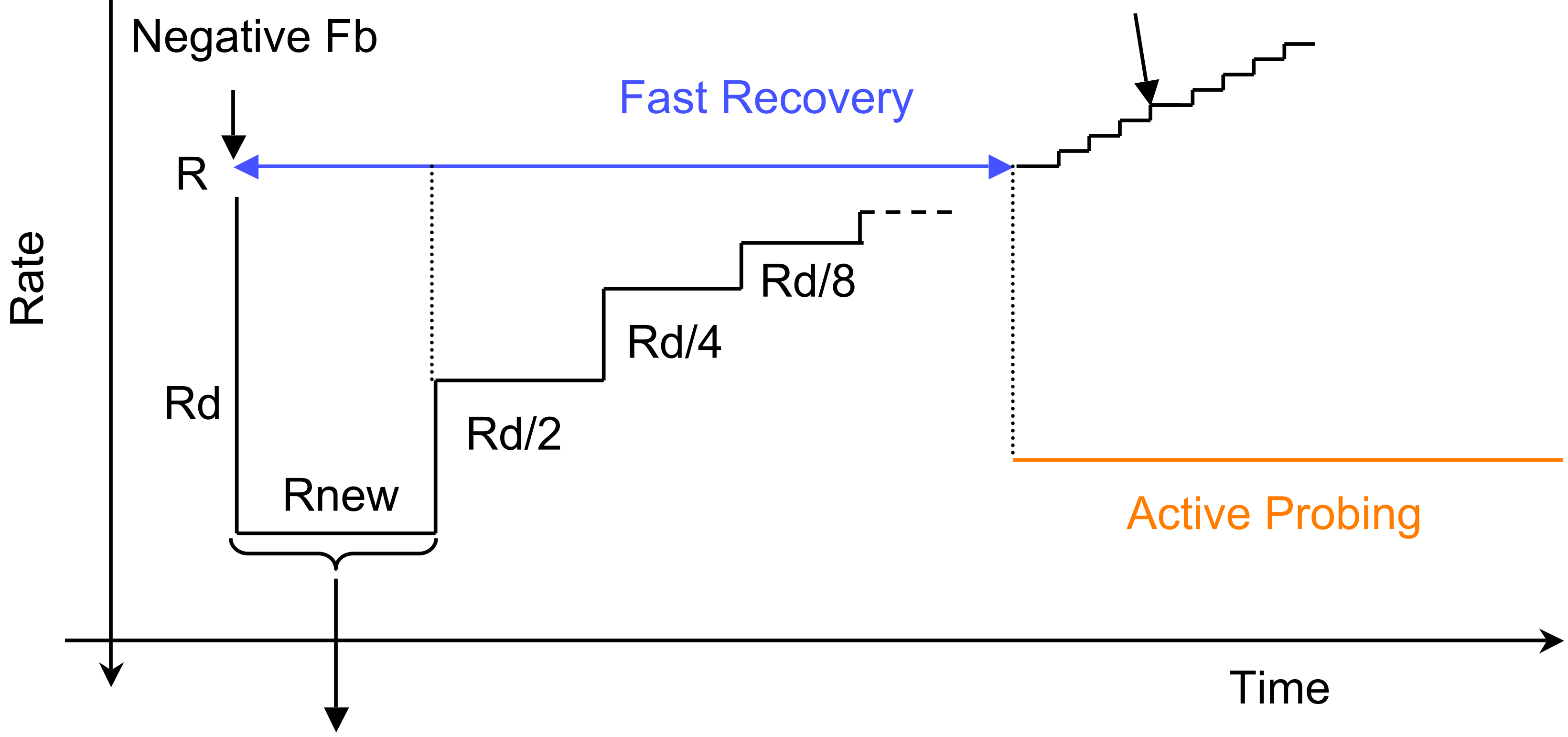
- Reaction Point:
  - Multiplicative decrease based on the value of Fb (same as BCN)
  - In the absence of negative signals, perform self-clocked fast recovery and active probing

- Congestion Point:
  - Compute Fb
  - If Fb < 0, reflect Fb back to the sources with probability depending on |Fb|
  - For 3-point architecture, set the “frame-reflected” bit (or the DE bit)

- Reflection Point:
  - For 2-point architecture, do nothing
  - For 3-point architecture: if a packet has not already been reflected, send Fb=0 signals back to reaction point with a fixed probability (e.g. 1%)

Reaction Point: Fast Recovery and Active Probing

After fast recovery, if there are no negative signals, source makes frequent small probes for extra bandwidth



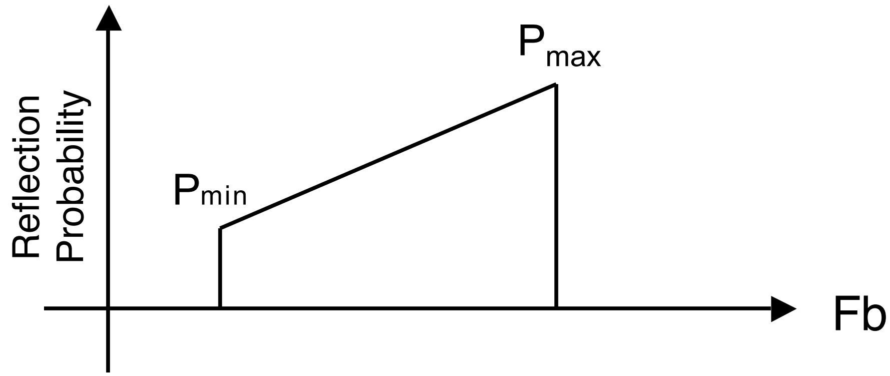
Self clocking: under congestion, a negative signal should be received after roughly 100 packets (because minimum sampling probability equals 1%);
If no negative signal, then there is likely extra capacity, ok to increase

Reaction Point: Fast Recovery and Active Probing

- Fast Recovery: Intuition
  - Amount of increase following a rate decrease is proportional to the amount of decrease; by recovering less than it was dinged, the source ensures stable behavior
  - By doing binary search for the next rate, we get fast convergence time

- Active Probing: Intuition
  - Make small rate increases frequently to probe for the extra bandwidth

Congestion Point



- At the CP
  - Compute: $Fb = -[q_{off} + w\, q_{delta}]$
  - Reflect
    - If Fb < 0, then reflect Fb value probabilistically back to the source with a bias which increases with Fb
    - Set the “frame-reflected” bit / DE bit

Reflection Point

- The end reflection point (in the 3-point architecture)
  - If the incoming frame has the “frame-reflected” bit set, do nothing
  - Else Fb=0; reflect this back to source with some small probability, say 1%

Basic Simulations

Outline

1. Infinitely long-lived flows
   - Simultaneous starts
   - Staggered starts

2. Dynamic heavy-tailed flows
   - Flow completion times for long flows, short flows
   - Losses

3. Baseline simulation, scenario #1

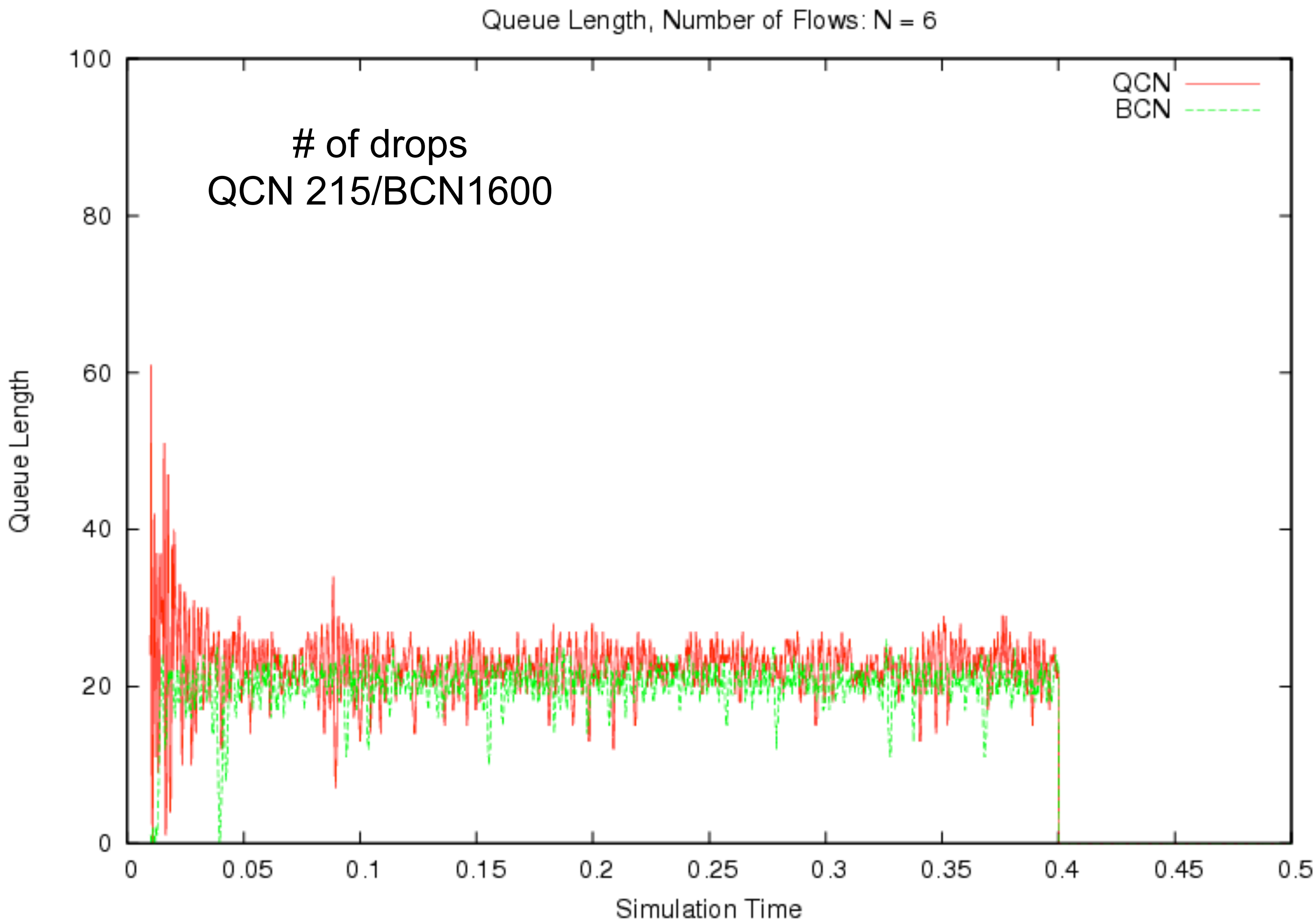
Parameters and settings

- Infinitely long-lived flows: simultaneous starts
  - Single link, 6 flows on at 10 Gbps at time 0
  - Link delay (RTT): 40 microseconds
  - Gd = 1/128
  - w = 2
  - Ri = 12 Mbps
  - Sampling function = linearly increases with |Fb| from 1--10%

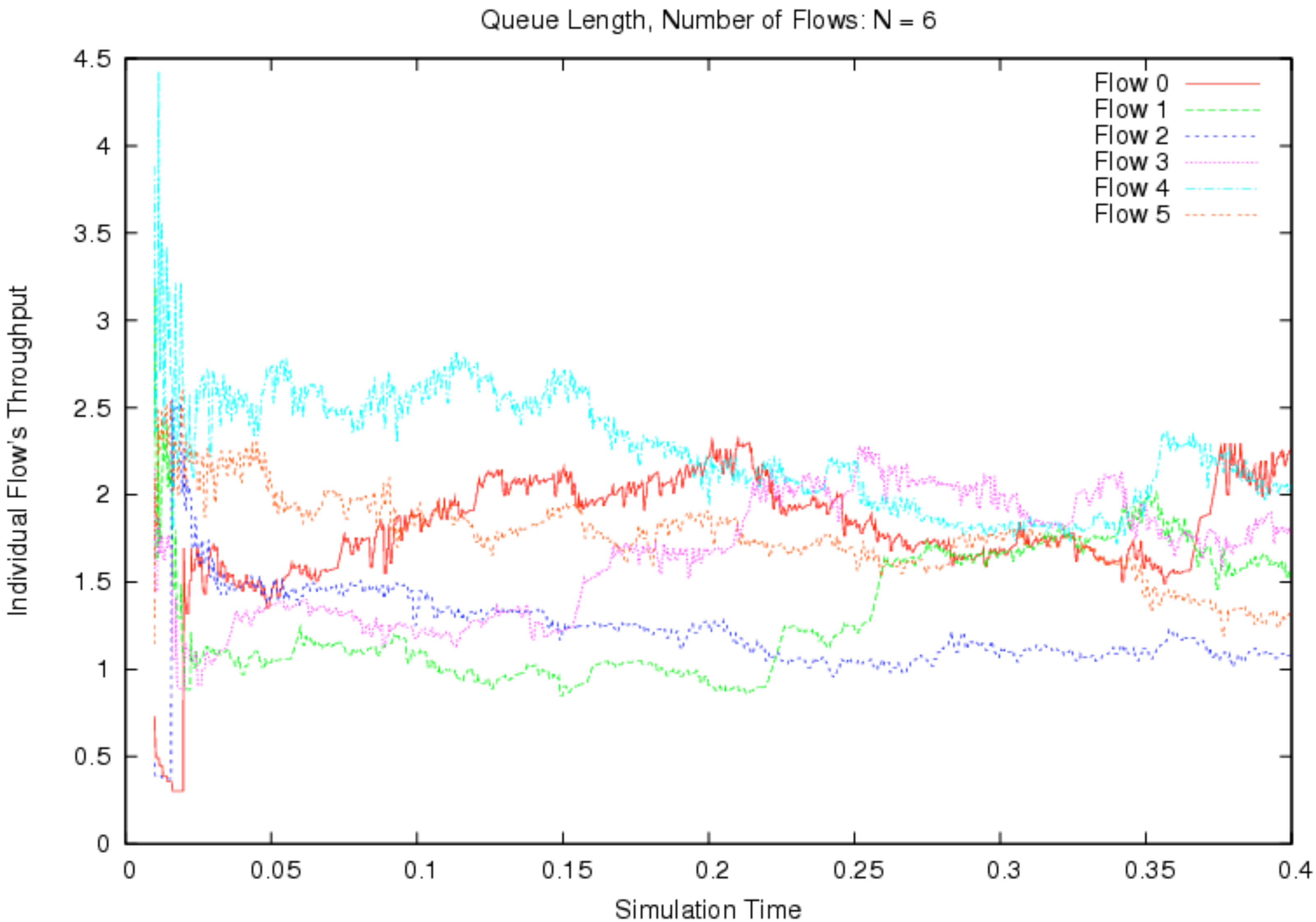
- BCN parameters
  - Gd = 1/128
  - Gi = 2.0
  - w = 2.0
  - Sampling Probability = 1%

- Staggered starts: staggered starts
  - Single link, 6 flows on 500 microseconds apart
  - Same parameters as above

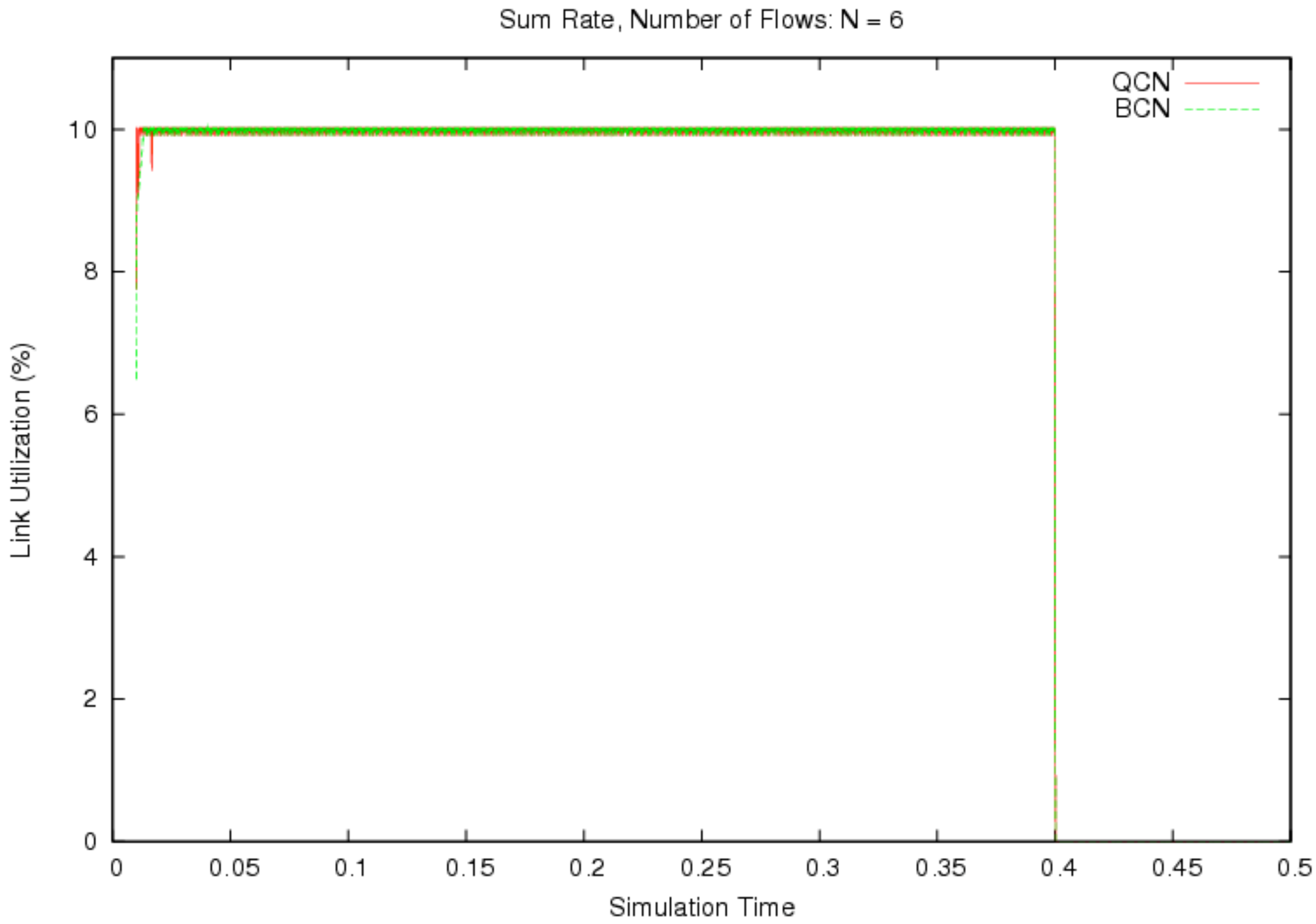
Queue Length
- Simultaneous, 2-point architecture



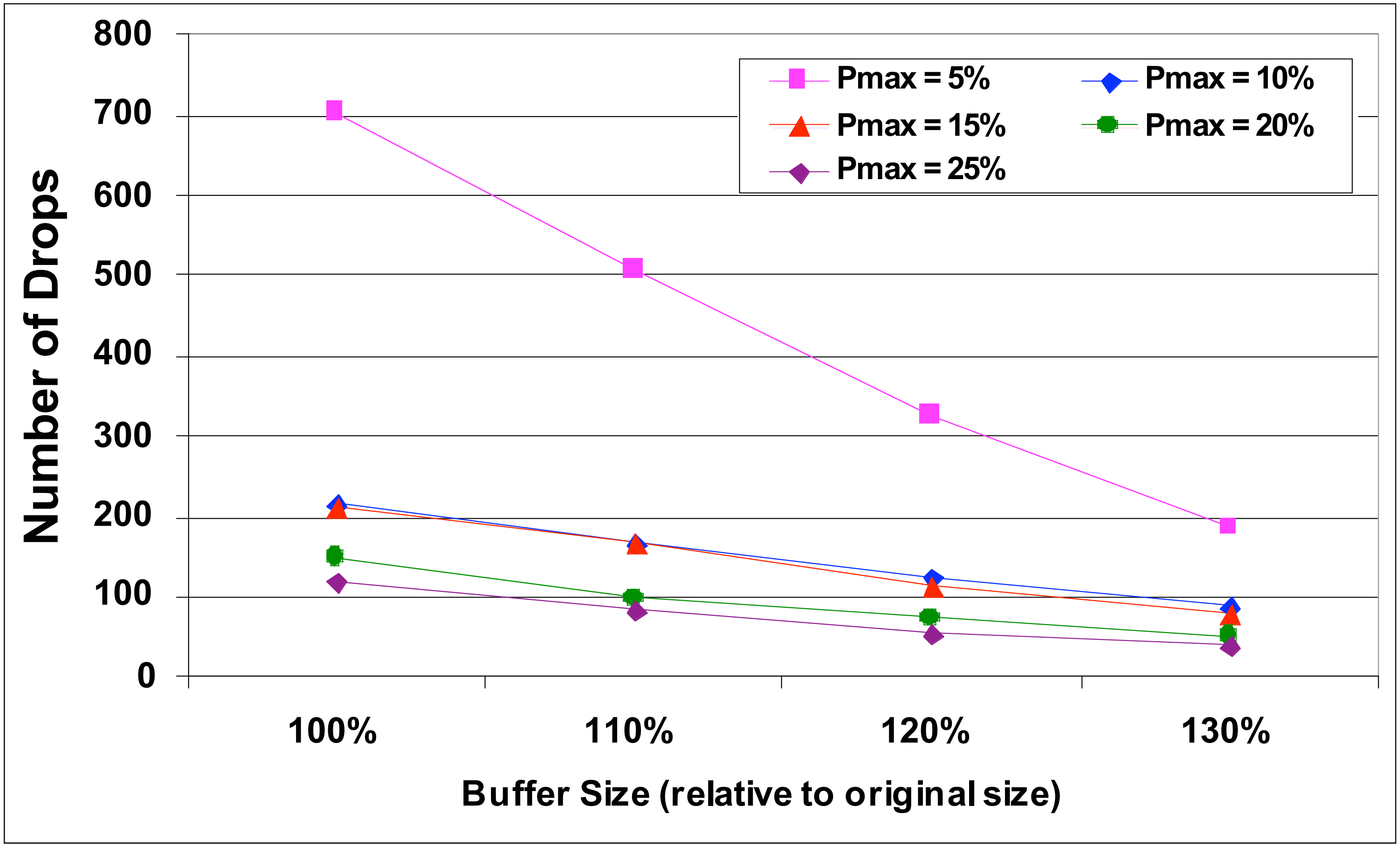
Fairness
- Simultaneous, 2-point architecture



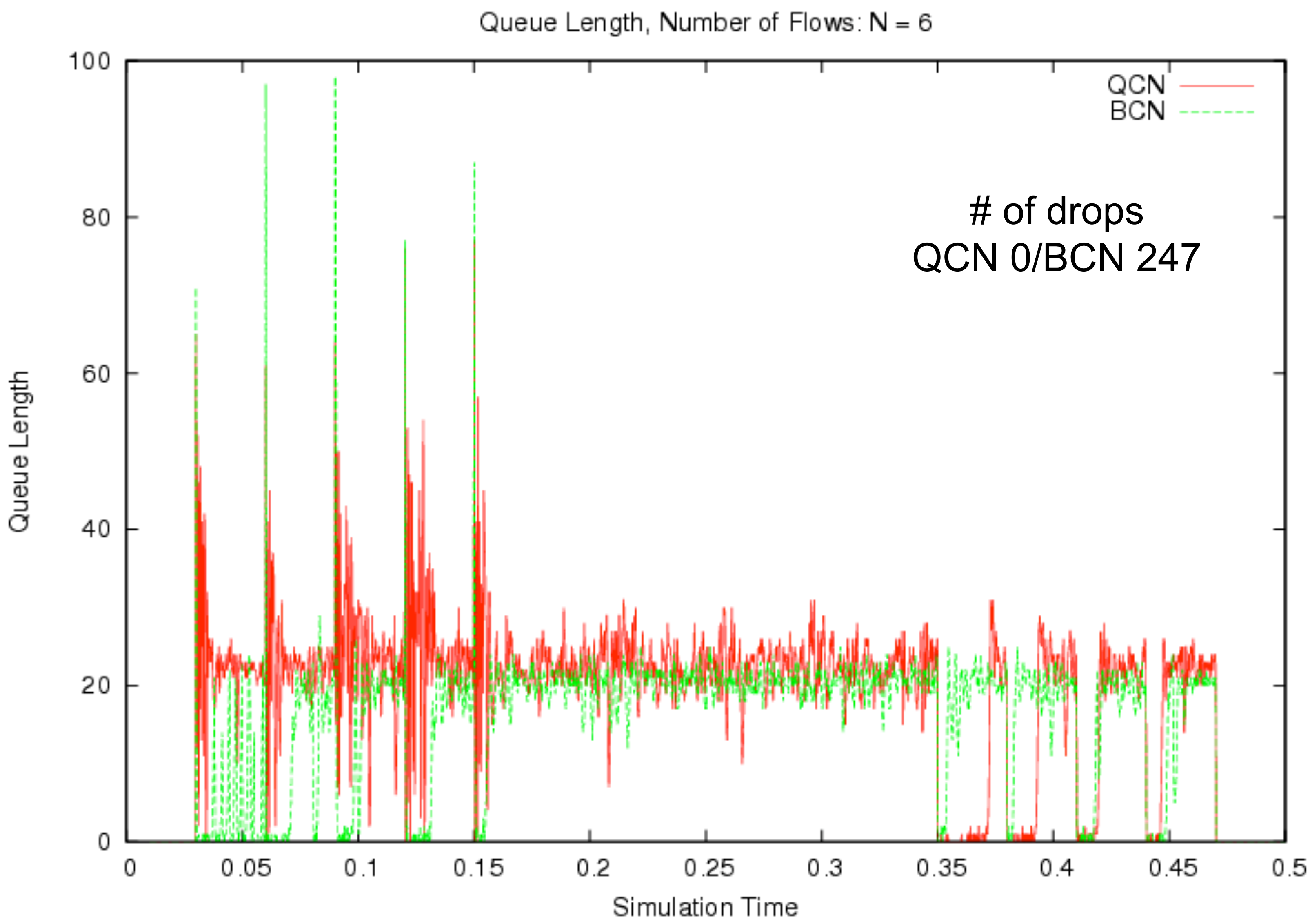
Throughput
- Simultaneous, 2-point architecture



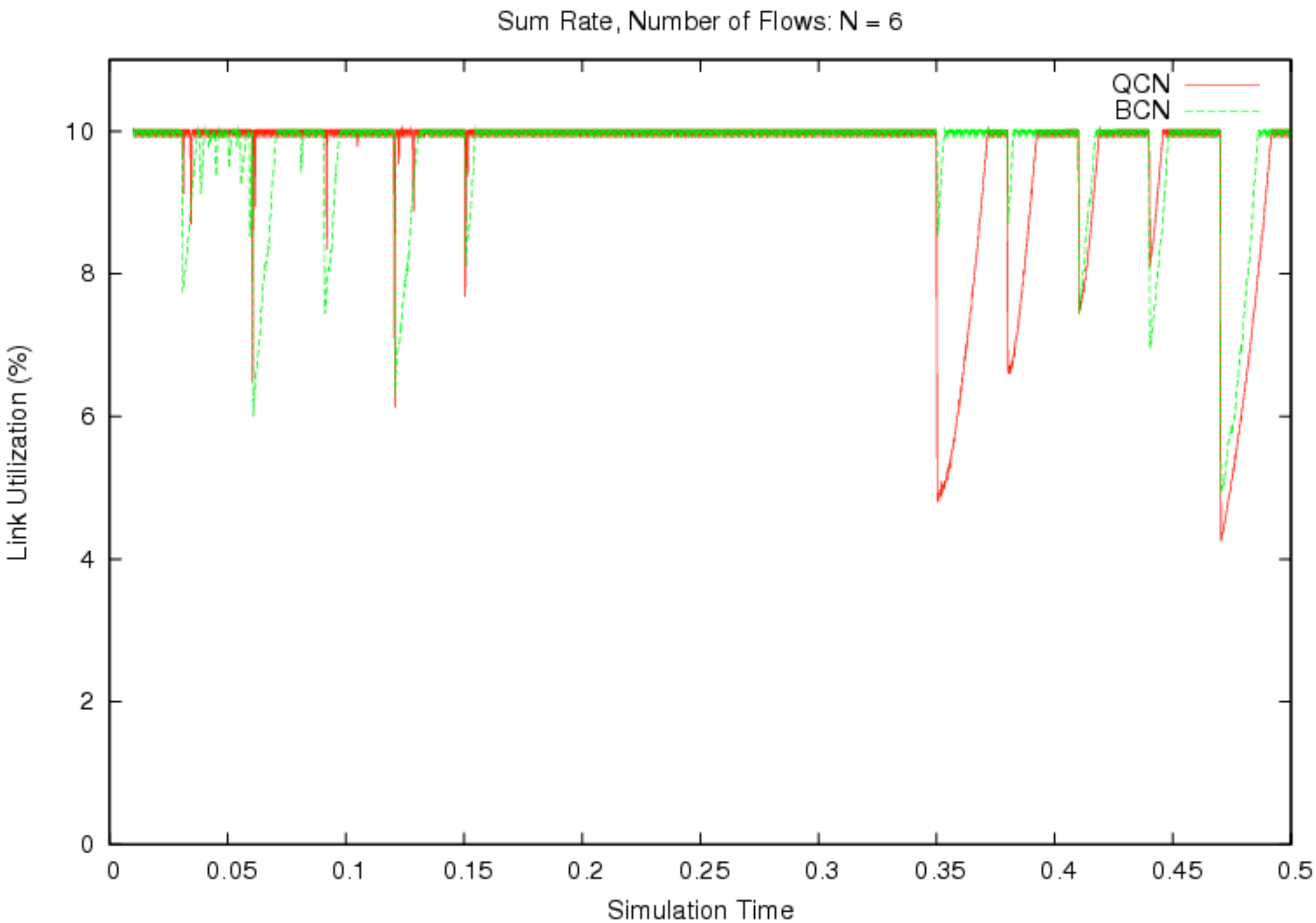
Number of drops vs. Buffer size for 6 flows starting simultaneously



Queue Length
- Staggered, 2-point Architecture



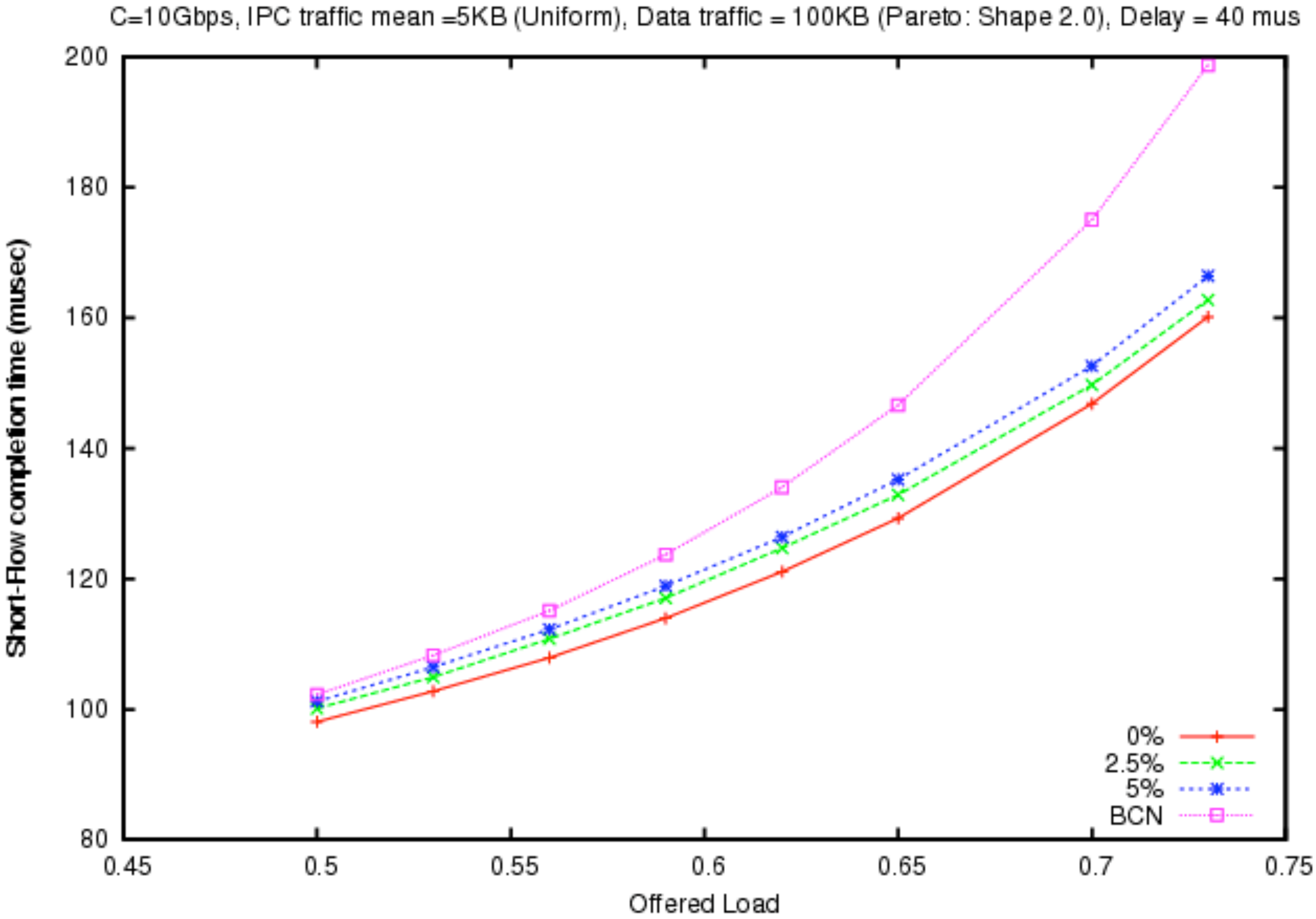
Link throughput
- Staggered, 2-point architecture



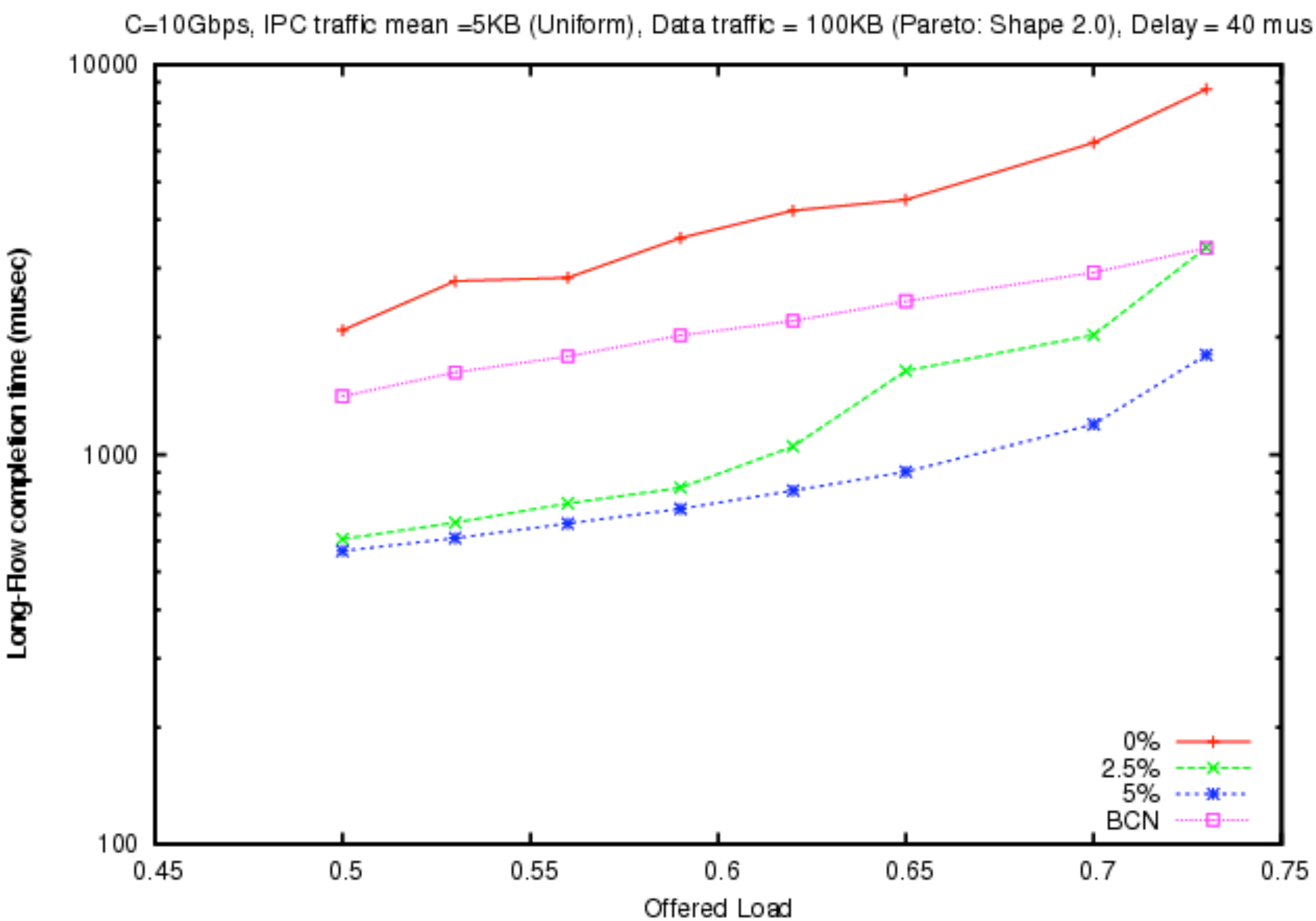
Dynamic flows: FCT and Drops

- Workload
  - IPC traffic: Mean = 5 KB (uniform distribution)
  - Data traffic: Pareto, shape 2, mean 100 KB
  - Parameters (Gd, w, etc): same as before
  - Reflection probability = 0, 2.5 and 5%

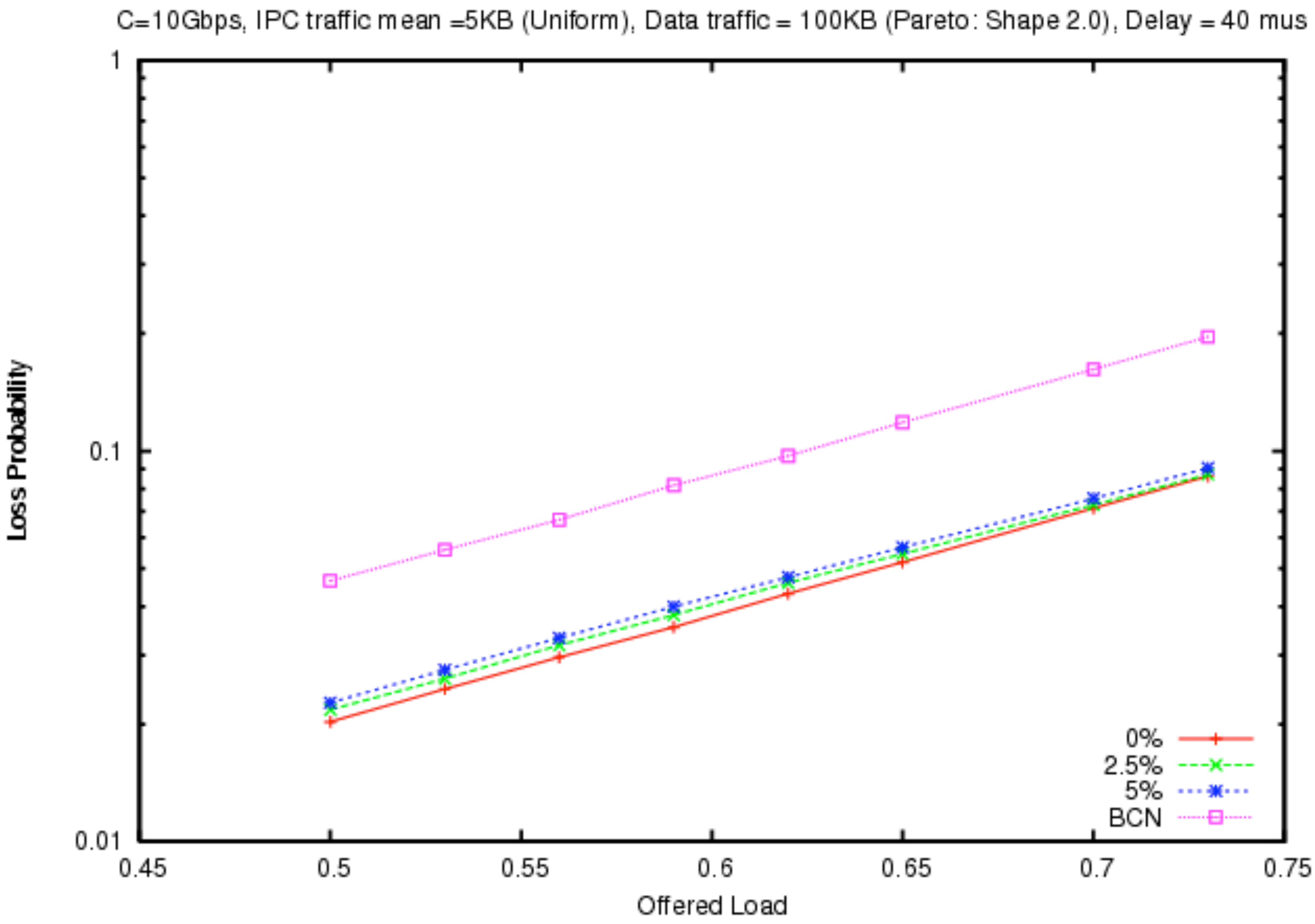
Completion Time of Short Flows
BCN sampling probability = 1%



Completion Time of Long Flows
BCN sampling probability = 1%

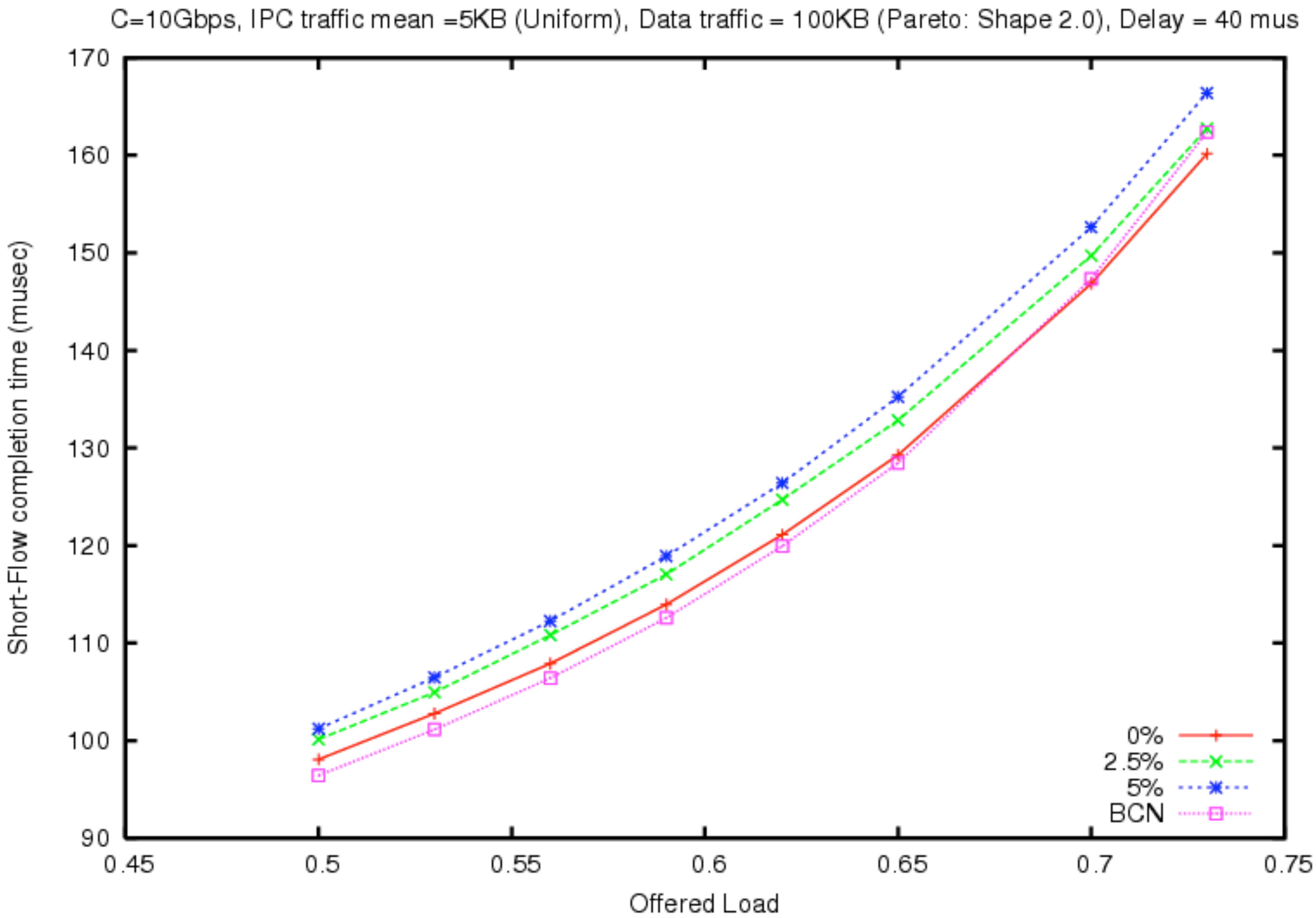


Drops
BCN sampling probability = 1%

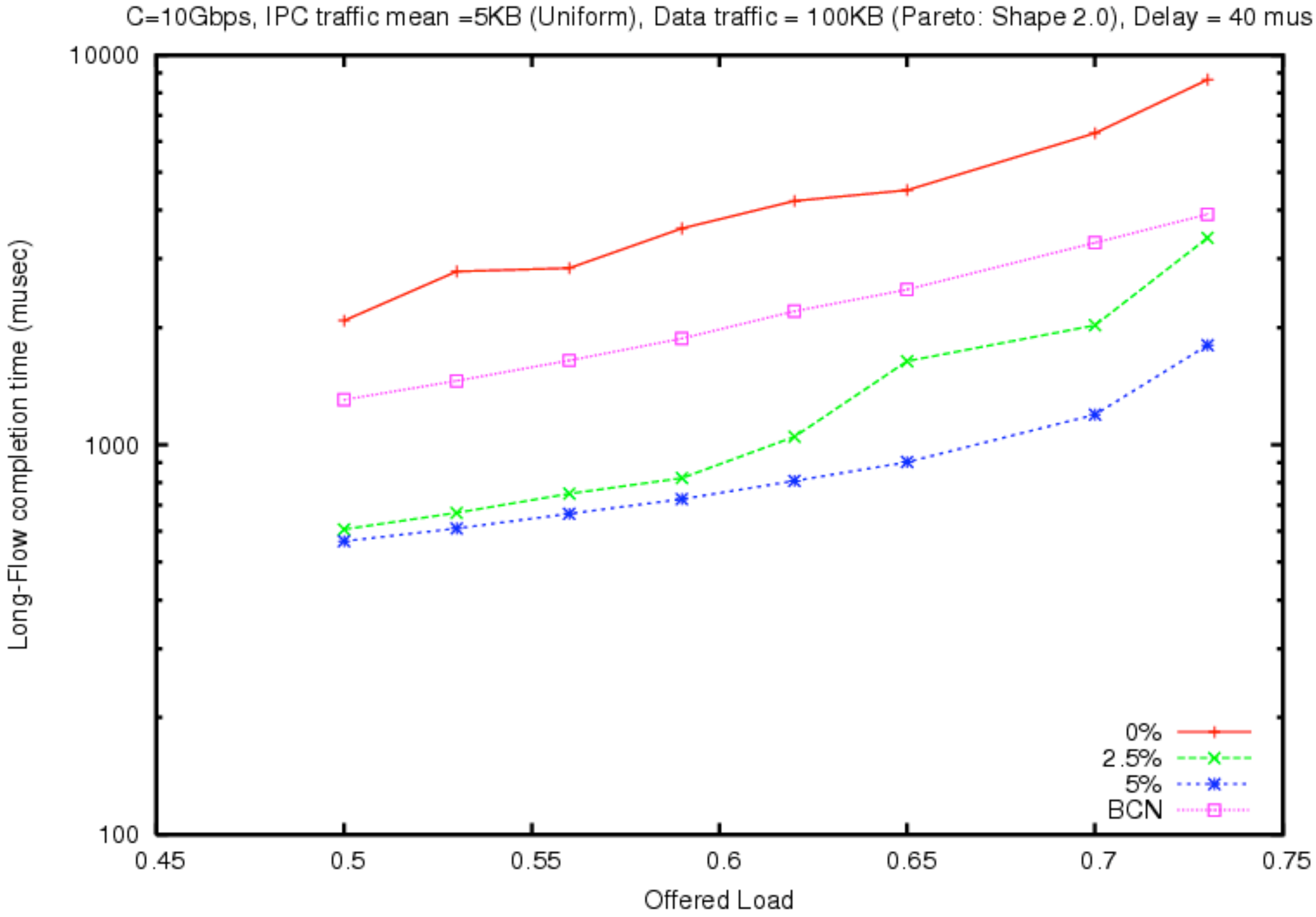


Completion Time of Short Flows

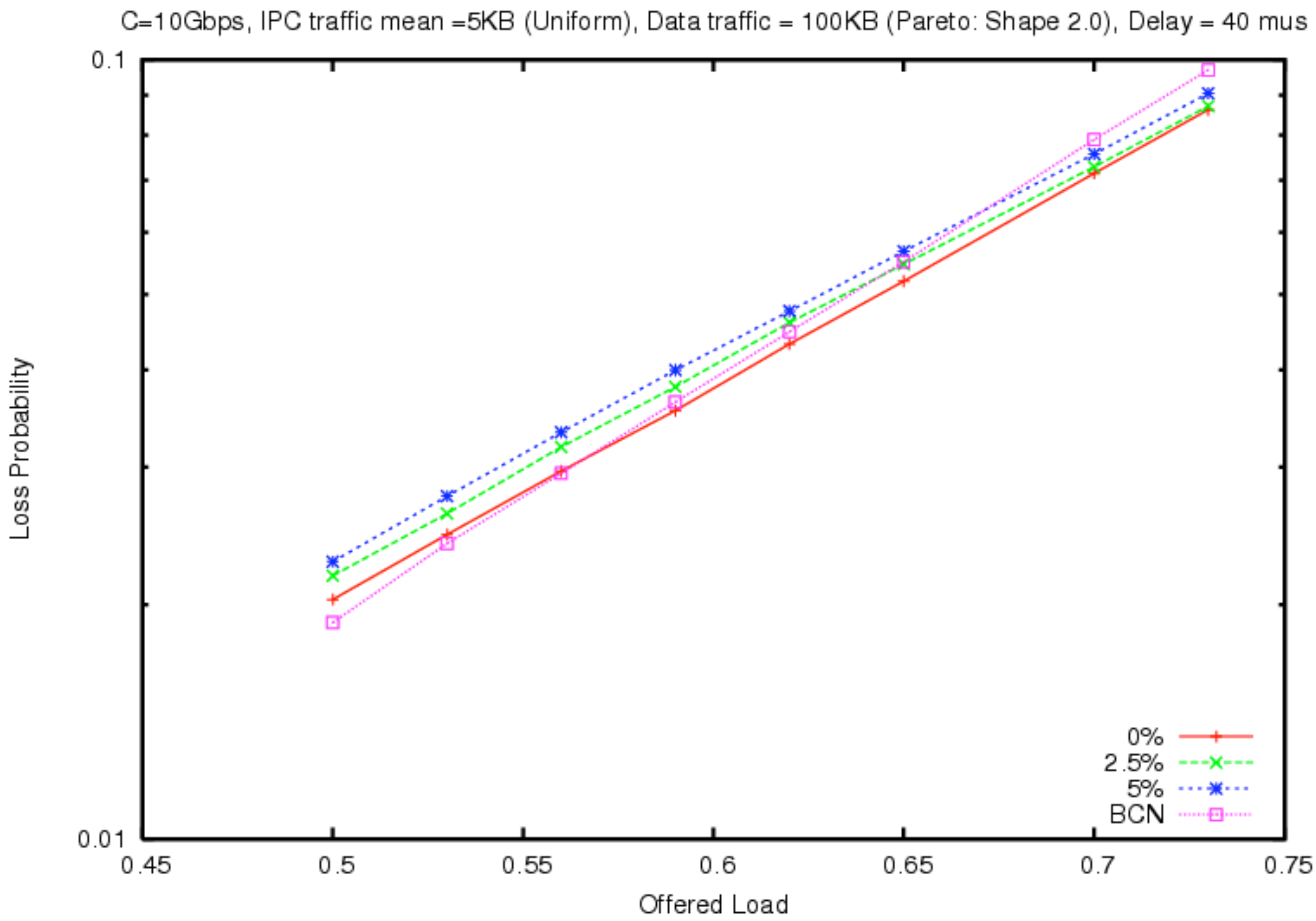
3% sampling probability with BCN



Completion Time of Long Flows
- compared with 3% sampling probability with BCN

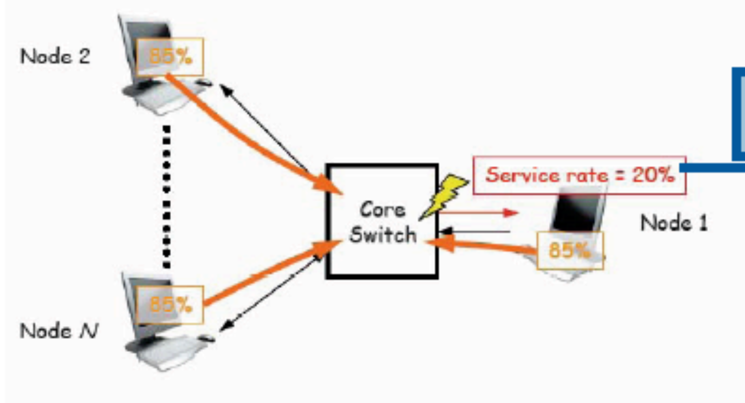


Drops
3% sampling probability with BCN



Baseline Setup

1. Output Generated Hot Spot Single Stage



Changed to 10%

Workload:

- All Nodes (10) : Uniform Distribution, load = 85% (8.5Gbps)
- Node 1 Service Rate = 10%
- One Congestion Point
  - Hotspot:
    - Degree: 9, Severity = 8.5:1,
    - Duration: 80 mS from ti=10 to 90 mS
  - All Flows affected

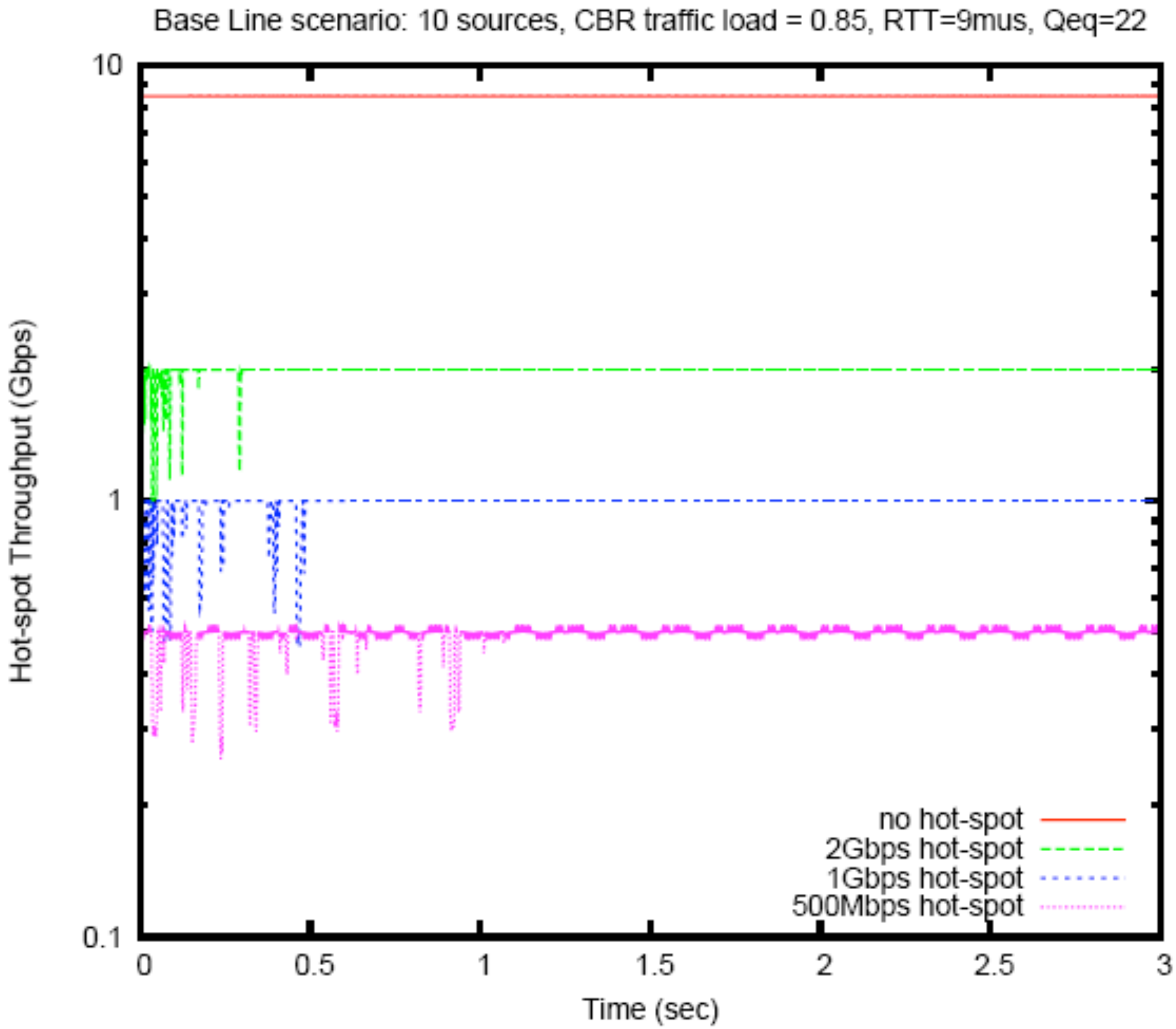
Required

Verdana regular 7pt.
Legal text goes here

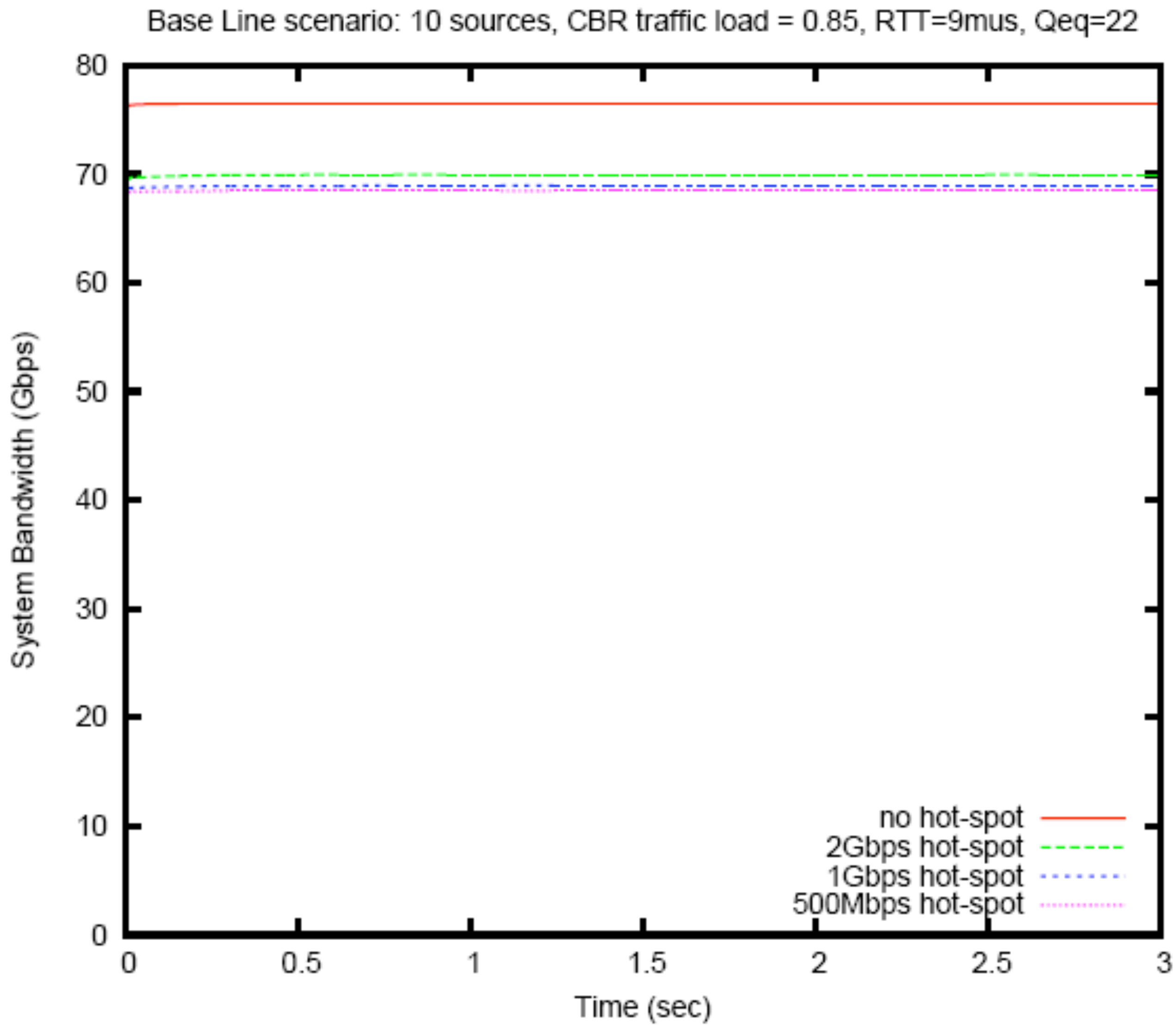
intel

http://www.ieee802.org/1/files/public/docs2007/au-sim-wadekar-reqd-extended-sim-list020107.pdf

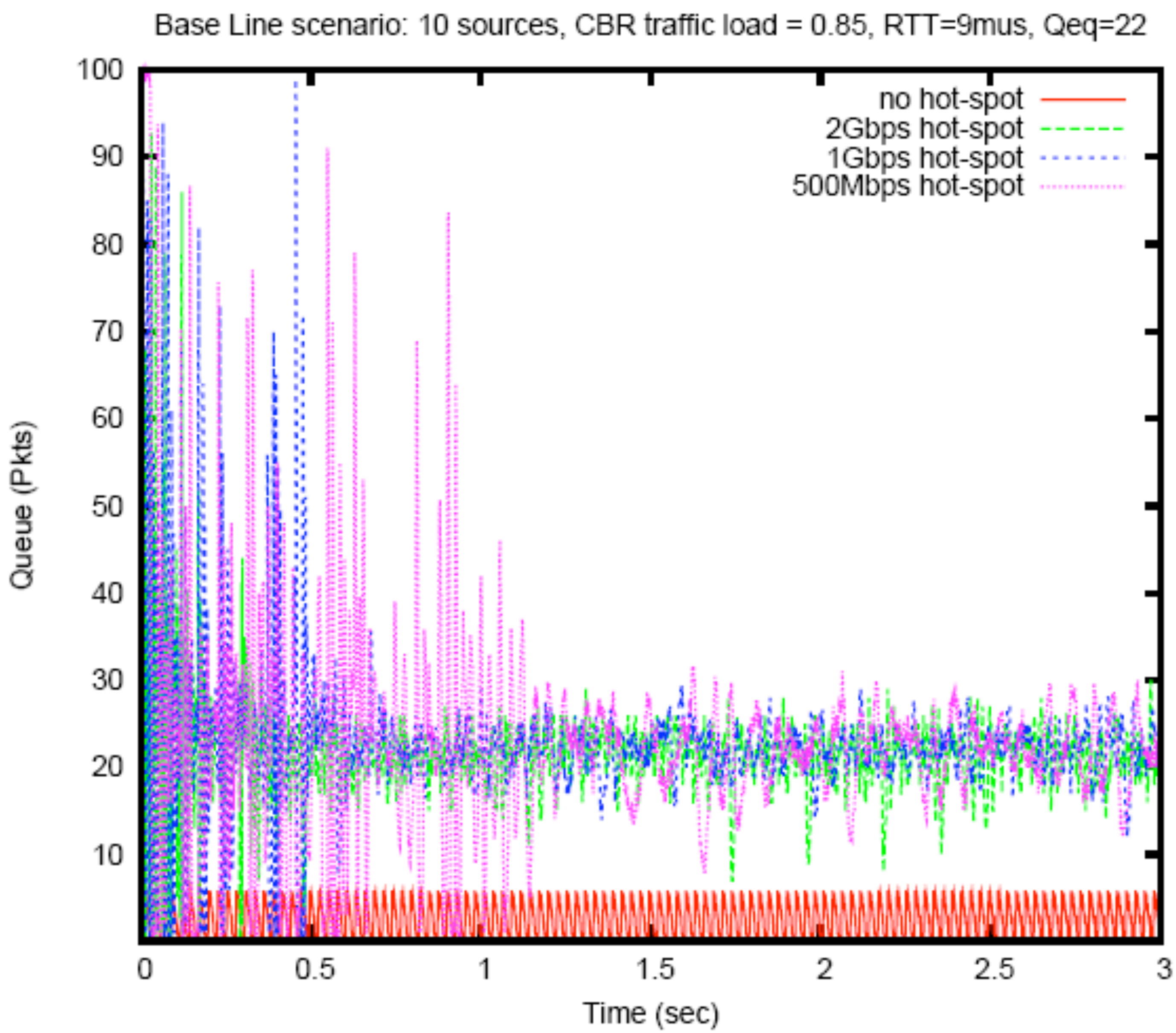
Baseline Setup
- Hot Spot Throughput



Baseline Setup
- System Throughput



Baseline Setup



Summary

- The main features of QCN are
  - Quantization of feedback removes the need for hardware-based computations; therefore, it is easily reconfigurable
  - 2-point architecture makes it easy to deploy incrementally
  - 3-point architecture involves only 1 bit in the header, which is very simple, yet enables really good performance both in terms of drops and underflows
- Further work
  - More extensive simulations, more general topologies
  - Decide crucial aspects: e.g. to use probe packets or use packet headers
  - Explore connections with other proposals
  - Your continued feedback would be appreciated!