Summary of Recent BCN Simulation Results

Bruce Kwan

November 15, 2006

Overview

- Key Observations

- Scaling Number of Flows

Severe Congestion Management

- BCN(0,0)
  - When Q_inst>Qsc, set current rate R to 0
  - Set random timer [0, Tmax]
  - When timer expires, current rate R set to Rmin
  - If BCN(0,0) occurs again before timer expiration, Tmax doubled and Rmin halved (exponential backoff)
- BCN(MAX):
  - When Q>Qsc, send BCN(MAX) to decrease the rate by maximum amount
  - BCN(MAX) → (Qoff = Qeq, Qdelta = 2Qeq)
- Drift (or Self-Increase)
  - Increase rate at fixed time intervals Ti, the current rate is additively incremented by a configurable amount
  - Always active (except during BCN(0,0) timeout)

BROADCOM.
Connecting everything®

Key Observations

- BCN + PAUSE
  - Avoids frame loss
  - Minimizes throughput degradation on innocent flows due to head of line blocking (for contending long lived flows)
  - Improves fairness across contending flows
  - *See au-sim-ding-bcn-pause-w-innocent_20061019.pdf*

- Severe Congestion Management
  - Throughput variations are negligible between BCN(0,0) and BCN(MAX) for the single hop scenario examined
  - With PAUSE, BCN(0,0) provided better RMS Fairness than BCN(MAX)
  - Drift (or Self Increase) aids in improving fairness
  - *See au-sim-ding-bcn-pause-102606.pdf*

- Scaling Number of Flows
  - Parameters require adjustment as a function of number of contending flows at a congestion point.
    - Decrease gain
    - Sampling rate at congestion detection point
  - *See au-sim-thaler-bcn-large-topo-110206.pdf*

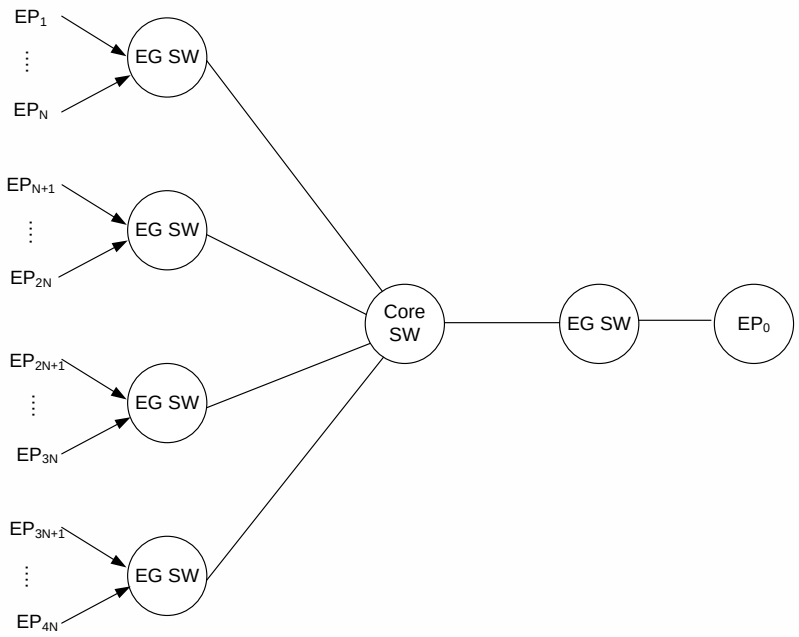
BROADCOM.
Connecting everything

Overview

- Key Observations

- Scaling Number of Flows

Scaling Number of Flows Simulation Topology



- Link Capacity = 10 Gbps
- Egress Port Buffer Size = 150 KB
- Switch Latency = 1 us
- Link Length = 100 m (.5 us propagation delay)
- Endpoint response time = 1 us

BROADCOM.
Connecting everything®

Scaling Number of Flows Workload

- Traffic Type: 100% UDP (or Raw Ethernet) Traffic
- Frame Size Distribution: Fixed length (1500 bytes) frames
- Arrival Distribution: Bernoulli temporal distribution
- Offered Load/Endpoint = 2%
- N=25, Destination Distribution: EP_1 – EP_{100} send to EP_0
- Simulation Time
  - Each source starts at 5ms, and simulation stops at 200ms

BROADCOM.
Connecting everything®

BCN Parameters

- Qeq
  - 16 (1500-byte frames)
  - 375 * 64 byte pages
- Frame Sampling
  - Frames are sampled on average 150 KB received to the egress queue
- W = 2
- Gi = 12.42
  - Computed as (Linerate/10) * [1/((1+2*W)*Q_eq)]
  - Gi = 5.3 x 10^{-1} * (1500/64) = 12.42
- Maximum rate decrease
  - 0.5, computed as 1/2*[1/((1+2*W)*Q_eq)]
  - 0.95, computed as 0.95*[1/((1+2*W)*Q_eq)]
- Ru = 1 Mbps
- Drift (Self-Increase)
  - At fixed time intervals Ti, the current rate is incremented by a unit
  - Never stop drifting
  - Drift = 1 Mbps every 100us
- *PAUSE is not active*

BROADCOM.
Connecting everything

Scaling Number of Flows - Results

Severe Congestion Behavior	Max Rate Decrease Percentage	Sampling Rate	CS Packet Loss
BCN(MAX)	50%	150KB	137099
BCN(MAX)	95%	150KB	12363
BCN(MAX)	95%	25KB	851