Simulation Modeling of BCN V2.0
Phase 1: Model Validation

J. Jiang and Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63131
Jain@cse.wustl.edu

IEEE 802.1 Congestion Group Meeting, Denver, March 8, 2006
These slides are available on-line at:
http://www.cse.wustl.edu/~jain/ieee/bcn603.htm
Overview

- Congestion Management Components
- BCN Mechanism
- Simulation Results
- Observations
- Parameter Selection
- Near Future Steps
Congestion Management Components

1. **Signaling**: Users need to tell/negotiate their QoS requirements with the network
2. **Admission Control**: Network can deny requests that it can not meet
3. **Shaping**: Traffic is smoothed out so that it is easier to handle
4. **Policing**: Ensuring that the users are sending at the rate they agreed to.
5. **Marking/Classification**: Packets are classified based on the source, destination, TCP ports (application)
6. **Scheduling**: Different flows get appropriate treatment. **Priority Scheduling**.
7. **Drop Policies**: Low priority packets are dropped. **Per priority Pause**
8. **Routing**: Packets are sent over paths that can meet the QoS
9. **Traffic Monitoring and Feedback**: Sources may be asked to reduce their rates to meet the loss rate and delay guarantees
BCN Mechanism

- Backward Congestion Notification - Closed loop feedback
  - **Detection**: Monitor the buffer utilization at possible congestion point (Core Switch, etc)
  - **Signaling**: Generate proper BCN message based on the status and variation of queue buffer
  - **Reaction**: At the source side, adjust the rate limiter setting according to the received BCN messages
    - Additive Increase Multiplicative Decrease (AIMD)

Parameters for BCN

- **Key Parameters**
  - Threshold for buffer:
    - $Q_{eq}$ (Equilibrium),
    - $Q_{sc}$ (Severe Congestion),
  - **Queue Variation**: $Q_{off}$, $Q_{delta}$
    - Queue is sampled randomly with 0.01 probability
    - $Q_{len}$ (current length)
    - $Q_{off} = Q_{eq} - Q_{len}$, [-$Q_{eq}$, +$Q_{eq}$]
    - $Q_{delta} = \# pktArrival - \# pktDeparture$, [-2$Q_{eq}$, +2$Q_{eq}$]
AIMD Algorithm

- Source Rate \( R \)
- Feedback
  - \( Fb = (Q_{off} - W \times Q_{delta}) \)
- Additive Increase \( (Fb > 0) \)
  - \( R = R + Gi \times Fb \times Ru \)
- Multiplicative Decrease \( (Fb < 0) \)
  - \( R = R \times (1 - Gd \times Fb) \)
- Parameters used in AIMD:
  1. Derivative weight \( W \)
  2. Additive Increase gain \( Gi \)
  3. Multiplicative Decrease Gain \( Gd \)
  4. Rate Unit \( Ru \)
Configuration

Congestion point
Configuration Parameters

- Configuration same as in Davide, IEEE 802.1, May 05
- Link Capacity = 10 Gbps (all links)
- Switch latency = 1 us (all switches)
- Propagation delay = 0.5 us (all links)
- TCP only
  - ST1-ST4: 10 parallel connections transferring 1MB each and repeat
  - SR1: 1 connection transferring 10 KB (wait 16 us after finishing, then repeat)
  - SR2: 1 connection transferring 10 KB (wait 1us after finishing, then repeat)
- Our simulation Platform: NS2 simulator
AIMD parameters

\[ F_b = (Q_{off} - W \times Q_{delta}) \]
\[ R = R + G_i \times F_b \times R_u \]
\[ R = R \times (1 - G_d \times F_b) \]

- Cisco’s settings
  - Derivative weight: \( W = 2 \)
  - Increase Gain: \( G_i = 4 \)
  - Decrease Gain: \( G_d = \frac{1}{64} \)
  - Rate Unit: \( R_u = 8 \text{ Mbps} \)

- Our settings
  - \( W, G_i, \) and \( R_u \) are same with Cisco
  - Decrease Gain: \( G_d = 0.0124 \)
  - Since \( F_b \)’s range is \([-80, 80]\)
    \( R \) becomes negative with \( G_d = 1/64 \)
  - In our simulation, \( G_d = 0.0124 \) to make sure \( R \) is always positive
Simulation Results: Throughput

- **Cisco’s results with BCN v1.0**

<table>
<thead>
<tr>
<th>CM</th>
<th>Reference Flow 1</th>
<th>Reference Flow 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Throughput(Tps)</td>
<td>Throughput(Gbps)</td>
</tr>
<tr>
<td>None</td>
<td>609</td>
<td>0.05245</td>
</tr>
<tr>
<td>BCN</td>
<td>4491</td>
<td>0.3868</td>
</tr>
</tbody>
</table>

  - **Bulk Traffic:**

- **Our Results with BCN v2.0**

<table>
<thead>
<tr>
<th>CM</th>
<th>Reference Flow 1</th>
<th>Reference Flow 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Throughput(Tps)</td>
<td>Throughput(Gbps)</td>
</tr>
<tr>
<td>None</td>
<td>501</td>
<td>0.0442</td>
</tr>
<tr>
<td>BCN</td>
<td>8697</td>
<td>0.7532</td>
</tr>
</tbody>
</table>

  - **Bulk Traffic:**

<table>
<thead>
<tr>
<th>CM</th>
<th>Average Source Throughput</th>
<th>Standard Deviation/Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>2.5484</td>
<td>4.44</td>
</tr>
<tr>
<td>BCN</td>
<td>2.2022</td>
<td>11.49</td>
</tr>
</tbody>
</table>
Observations

- For reference flow, BCNv2 in our simulation performs better than BCNv1 (by Cisco), nearly double the rate of BCNv1;
- For bulk flow, BCNv2 in our simulation performs similar to BCNv1 (by Cisco). Maybe it is because Reference Flows have higher data rates;
- Fairness: Our current results always have larger deviation reported by Cisco. Even with None-CM, we have larger standard deviation. Time to fairness is longer.
Symmetric Topology-Buffer Utilization

- Compared with Cisco’s result, the equilibrium is almost the same. However, in our results, there are larger variations. (Reasons: Tradeoff between oscillation size and time to fairness)
Parameter Selection

\[ R = R + Gi \times Fb \times Ru \]
\[ R = R \times (1 - Gd \times Fb) \]

- \( Q_{off}, Q_{delta} \) are \#packets per observation, then \( Fb \) is \#packets per observation (sampling time gap)
- \( Ru \) is 8 Mbps
- \( Gi \) and \( Gd \) are not dimensionless \( \Rightarrow \) Link rate dependent
  \( \Rightarrow \) \( Fb \) should be normalized to be dimensionless
- Our preliminary simulation results show that optimal parameter values depend upon link speeds.
  \( \Rightarrow \) Need to simulate mixed 1G and 10G environments
- AIMD parameters should be carefully chosen to optimize BCN performance
Near Future Steps

- Fix the dimensioning problem
- Asymmetric Topology
- Multi-bottleneck case
- Larger/smaller Bandwidth×Delay product networks
- Bursty Traffic
- Non-TCP traffic
- Interaction with TCP congestion mechanism
- Effect of BCN/Tag messages getting lost
1. BCN V2 simulation validate Cisco’s results on throughput
2. Time to Fairness and oscillation trade-off needs to be studied further
3. Parameter setting needs more work
   Need to modify formula so that parameters are dimensionless
4. Need to simulate more configurations:
   asymmetric, larger bandwidth delay, and multi-bottleneck cases
References

- new-bergamasco-backward-congestion-notification-0505.pdf
- new-bergamasco-bcn-july-plenary-0705.ppt
- new-bergamasco-bcn-september-interim-rev-final-0905.ppt
- new-cm-five-criteria-03-1105.pdf
- new-cm-hazarika-gopi-cm-par-bkgnd-1105.pdf
- new-cm-hazarika-gopi-cm-par-rev-0-8-1105.pdf

http://ieee802.org/1/files/public/docs2006/
- new-barrass-cm-constraints-0106.pdf
- new-barrass-cm-overview-0106.pdf
- new-cm-capabilities-of-various-fabrics-0106.pdf
- new-cm-nfinn-1Q-placement-0106-02.pdf
- new-cm-nfinn-1Q-placement-0106-03.pdf
- new-seaman-cm-congestion-notification-0206-01.pdf
- new-seaman-cm-interim-constraints-doc-structure-0106.pdf
- new-seaman-cm-interim_dot1_integration-0106.pdf
- new-seaman-cm-interim_dot1_operation-0106.pdf