Simulations for Incremental Deployment of L2-CI

Tanmay Gupta
Manoj Wadekar
Gary McAlpine

Intel Corporation
November 2004

Purpose

- To study the effect of incremental deployment of L2-CI on system performance
- Mixed environments refer to networks with only some Switches and/or End-Stations upgraded to support L2-CI

Target Topologies

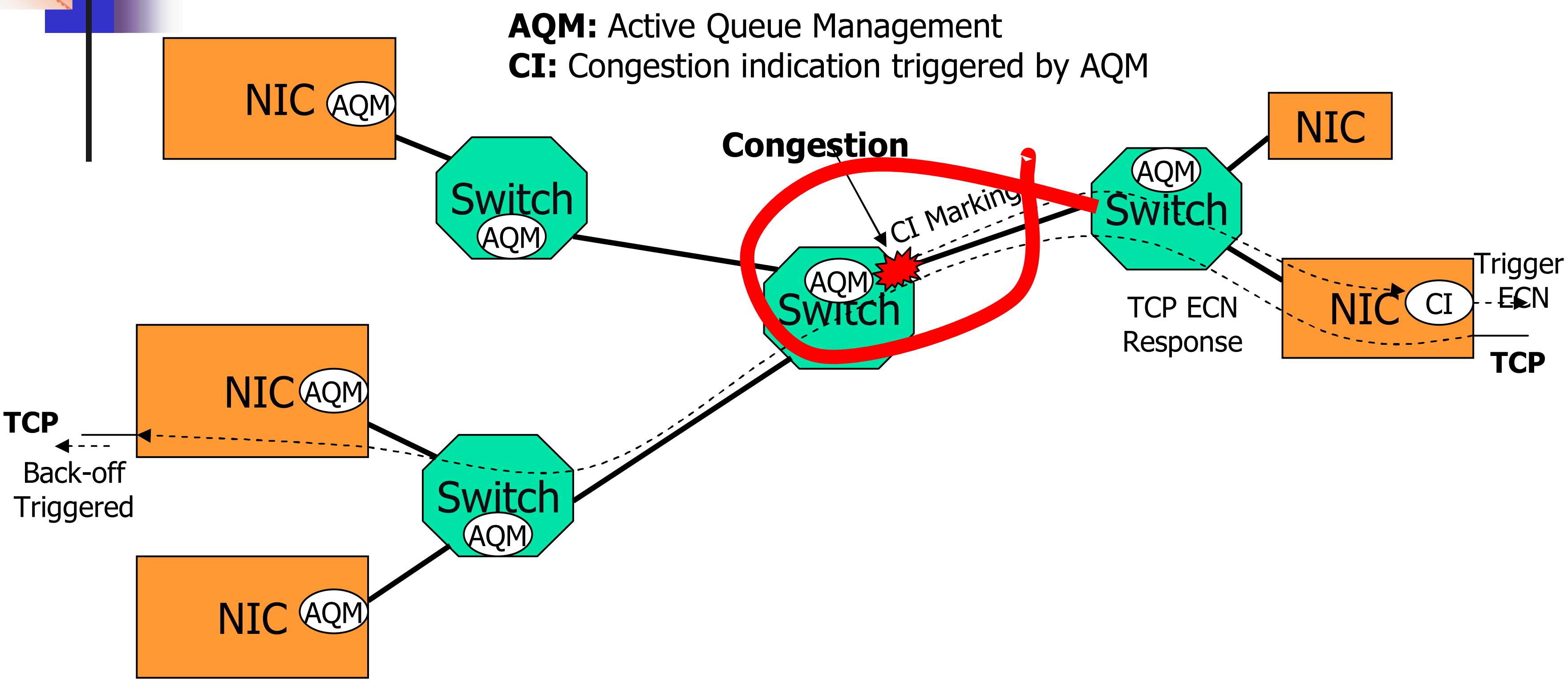
- Case 1: Incremental deployment of L2-CI in Ethernet Switches
  - Only some Ethernet Switches in the network are capable for marking L2-CI
- Case 2: Incremental deployment of L2-CI in End-Stations
  - Only some End-Stations convey L2-CI information to IP layer
- Case 3: Interaction between responsive and non-responsive traffic
  - Mix of TCP (responsive) and UDP (non-responsive) traffic

Only Case 1 and Case 2 studied so far

Assumptions

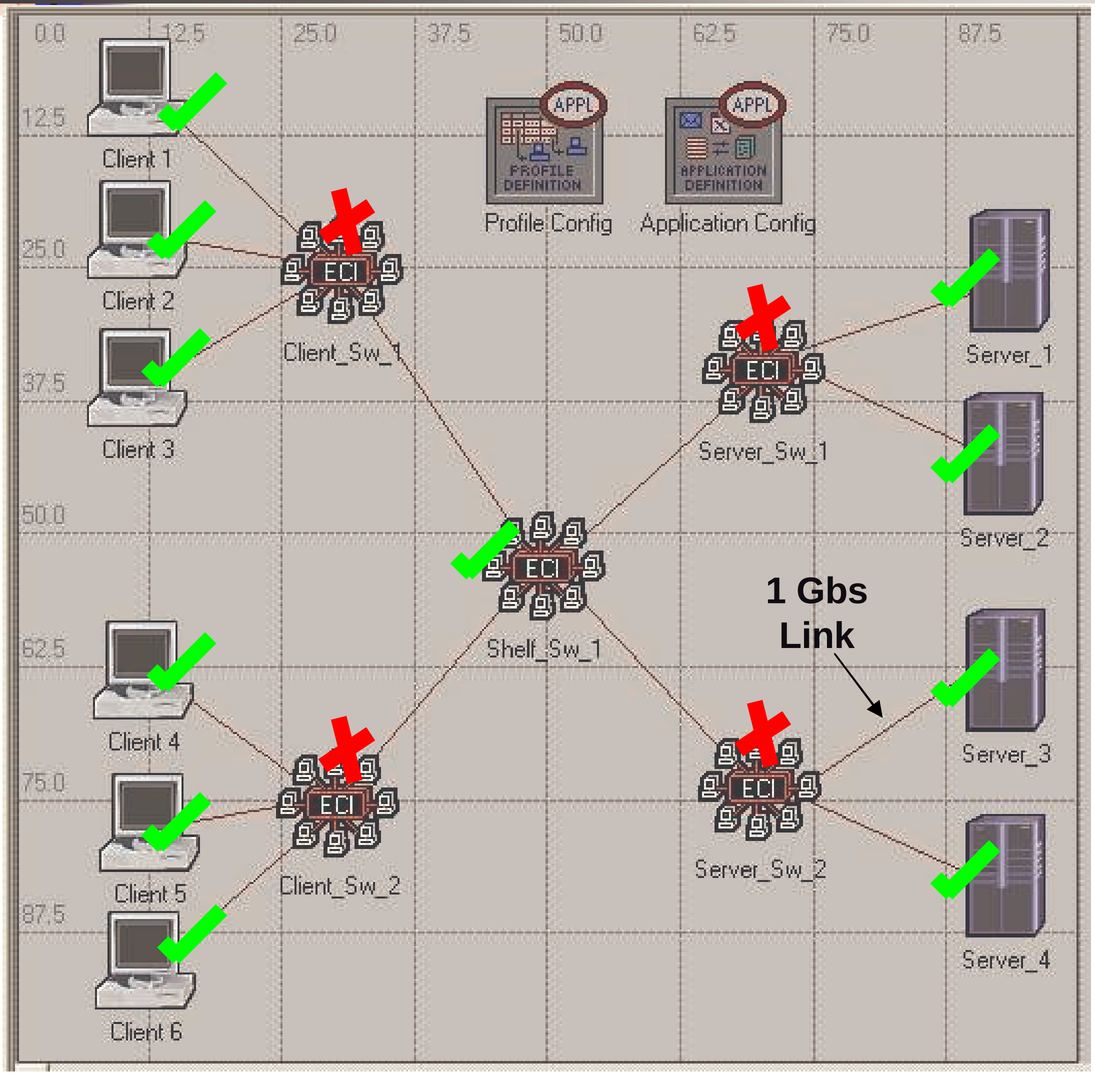
- It is assumed that **non**-ECN capable Switches and End-Stations are transparent to L2-CI information in the frames. Specifically:
  - the switches that do not support L2-CI can still handle (and forward unchanged) the frames with L2-CI marking
  - the end-stations that do not support L2-CI either do not receive L2-CI marked frames (device/switch sending to it removes L2-CI information) or ignore the L2-CI marking in the frame

Case 1: Switch support for L2-CI



- An L2-CI capable switch marks Ethernet frames instead of dropping them based on RED (or any other AQM)

Case1: Topology



All Links except one = 10 Gbs

Application = Database access

Peak Throughput possible = ~2.2 GB/s DB Entry + ~2.2 GB/s DB Query

Workload distribution = Exponential (8000)

ULP Packet Sizes = 1 Byte to ~85KB

TCP Window size = 64KB

✓ **L2-CI capable**

✗ **Legacy**

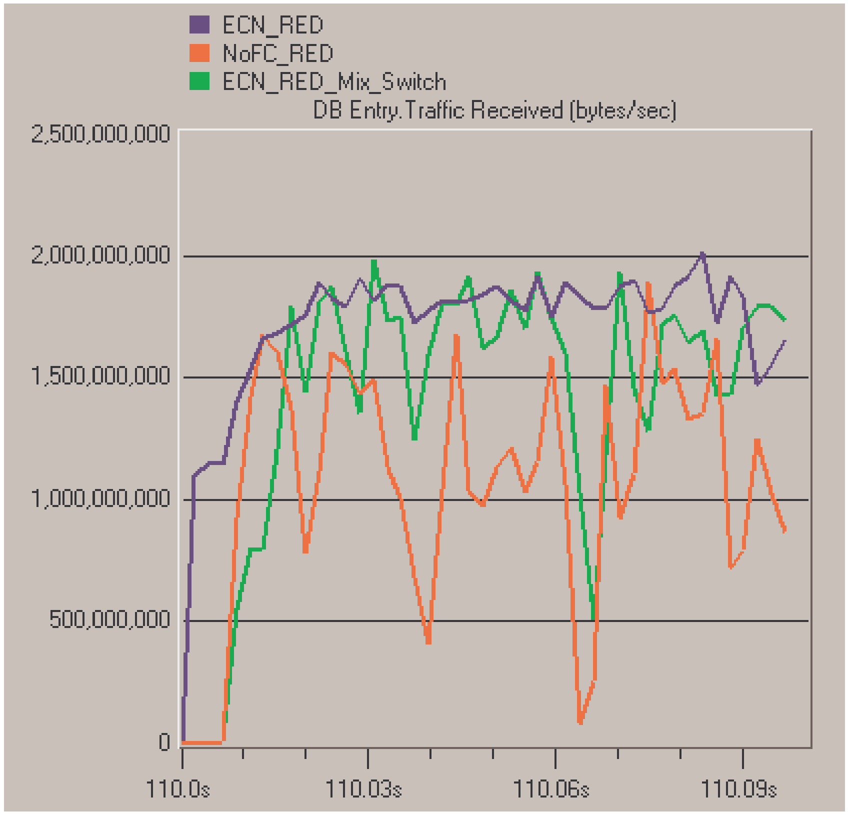
Case1: Scenarios

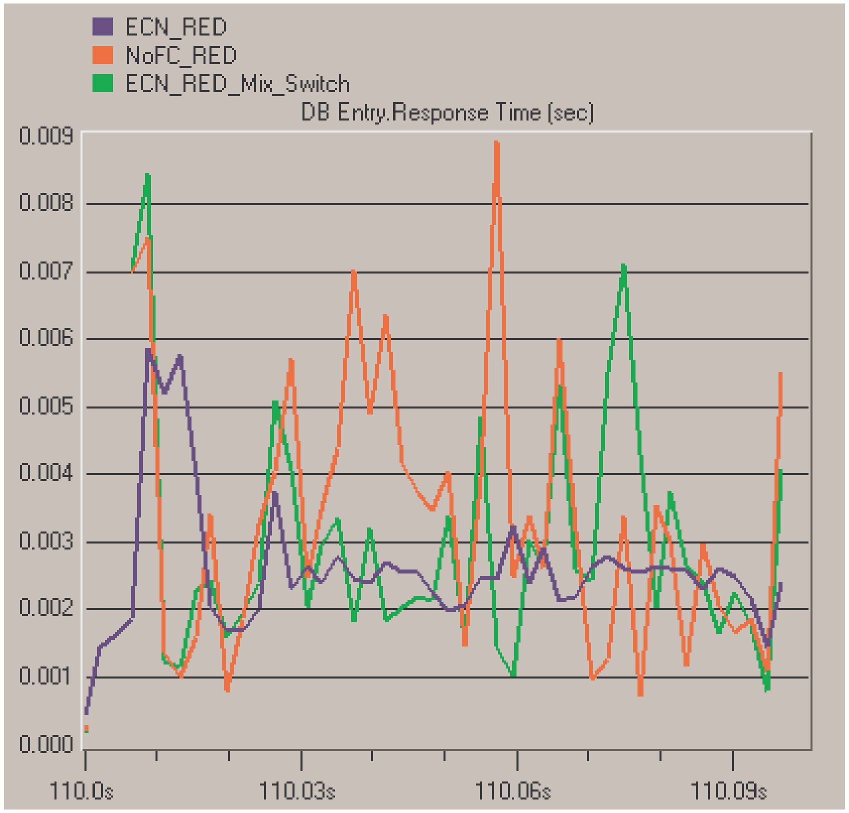
- NoFC_RED
  - No Switches are L2-CI capable
- ECN_RED_Mix_Switch
  - Only Shelf_Sw_1 L2-CI capable
- ECN_RED
  - All Switches are L2-CI capable

Note:

1. All End-Stations are L2-CI capable in all the above scenarios

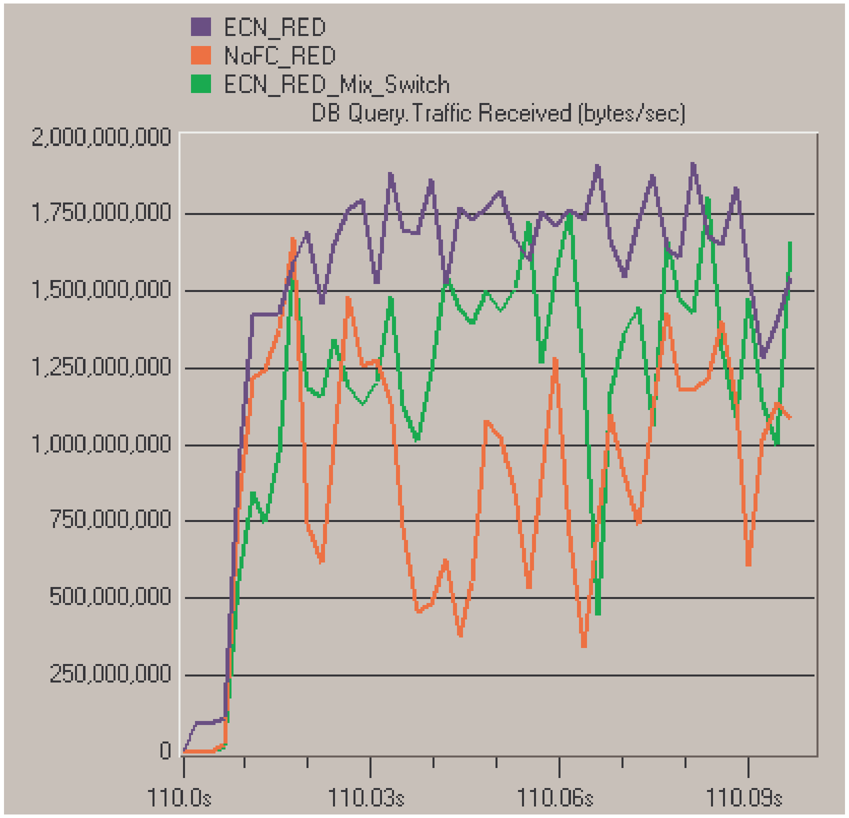
Application DB Entry Throughput & Response Time (Buffer = 32 KB per Switch Port)

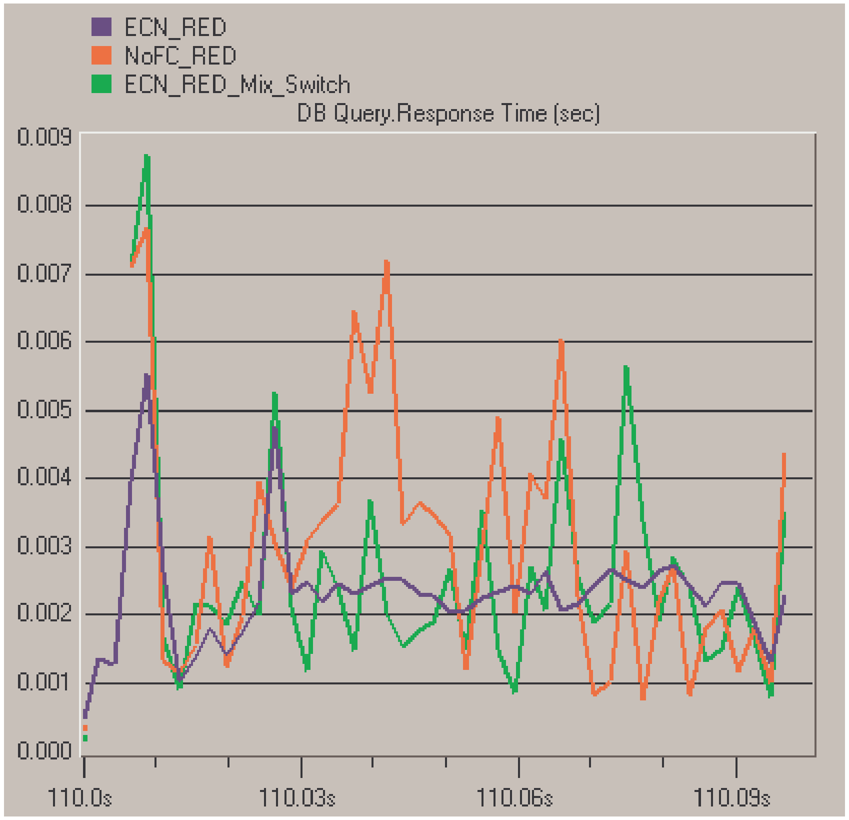




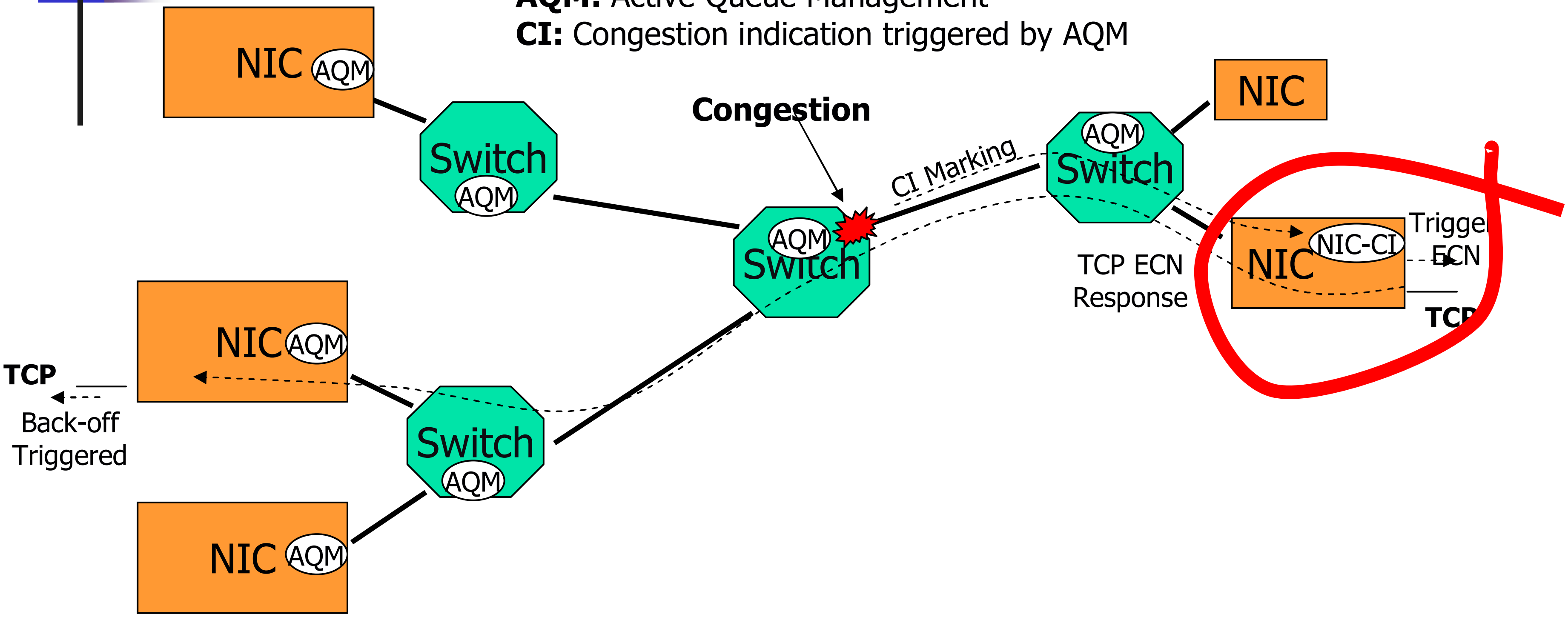
Incremental network upgrade shows proportional performance improvement

Application DB Query Throughput & Response Time (Buffer = 32 KB per Switch Port)





Case 2: End-Station support for L2-CI



AQM: Active Queue Management
CI: Congestion indication triggered by AQM

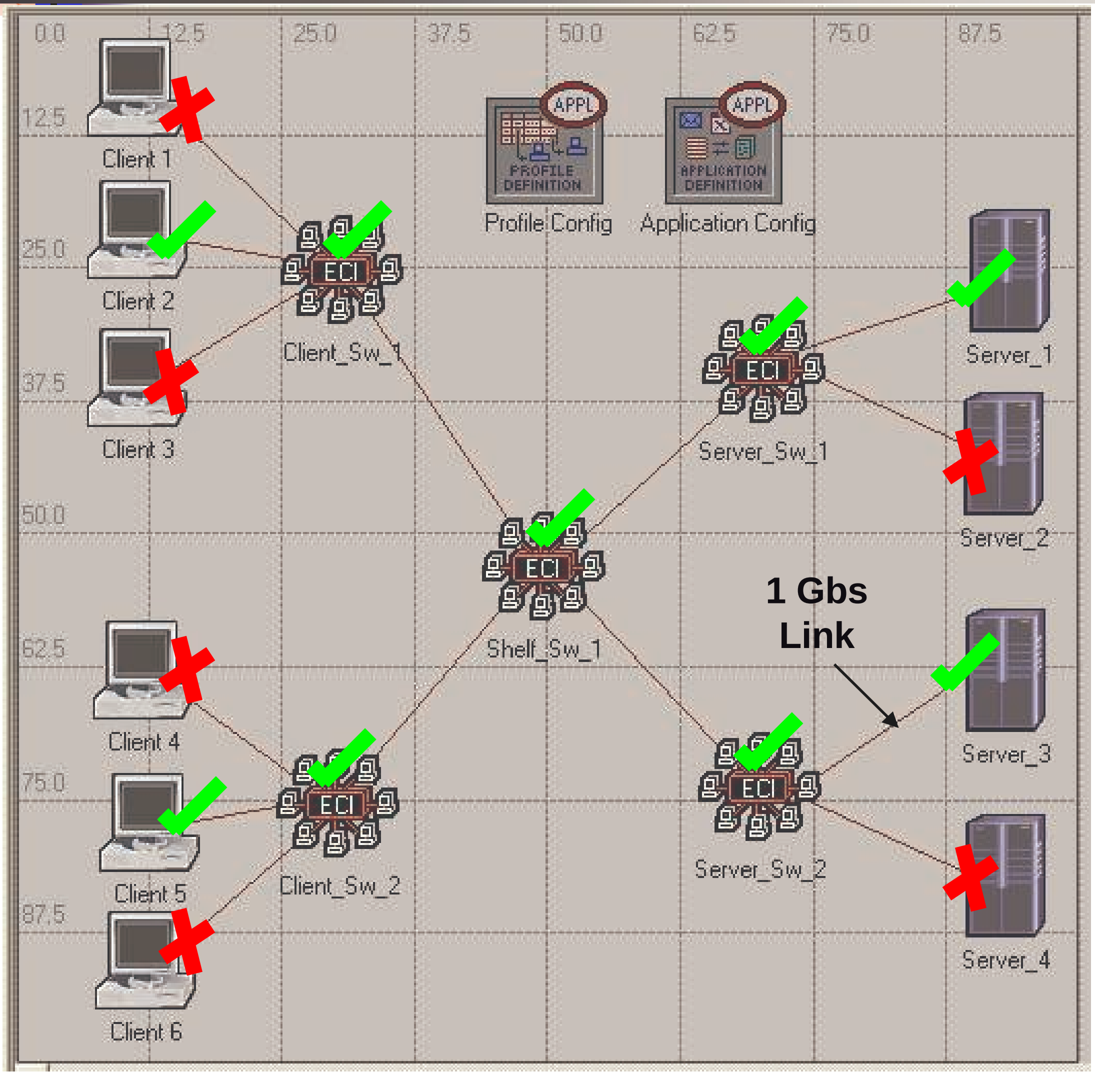
- L2-CI capable destination End-Stations convey L2-CI information to the IP layer

CCT(CI Capable Transport) Bit

- In our simulations, we used a new bit to convey the ECN capability with each Ethernet frame
- Used by Ethernet switches to determine if it should do
    - Random Early Detection and mark CI (CCT=1)

    OR
    - Random Early Discard (CCT=0)
- In our simulations, the IP ECT bit is mapped to L2 CCT

Case 2: Topology



All Links except one = 10 Gbs

Application = Database access

Peak Throughput possible = ~2.2 GB/s DB Entry + ~2.2 GB/s DB Query

Workload distribution = Exponential (8000)

ULP Packet Sizes = 1 Byte to ~85KB

TCP Window size = 64KB

✓ **L2-CI capable**

✗ **Legacy**

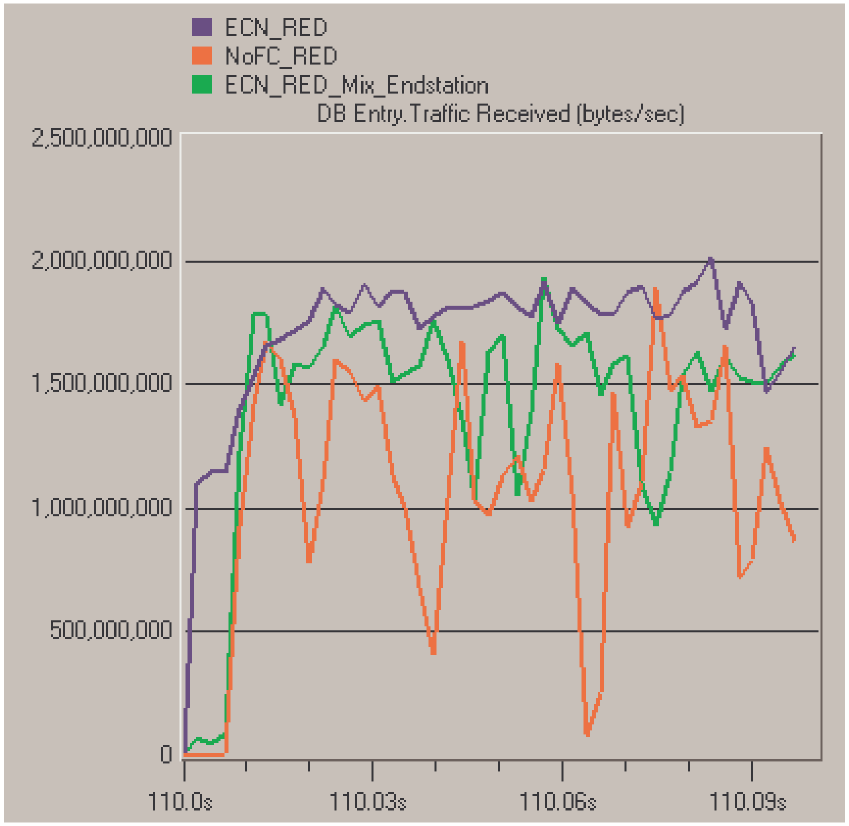
Case 2: Scenarios

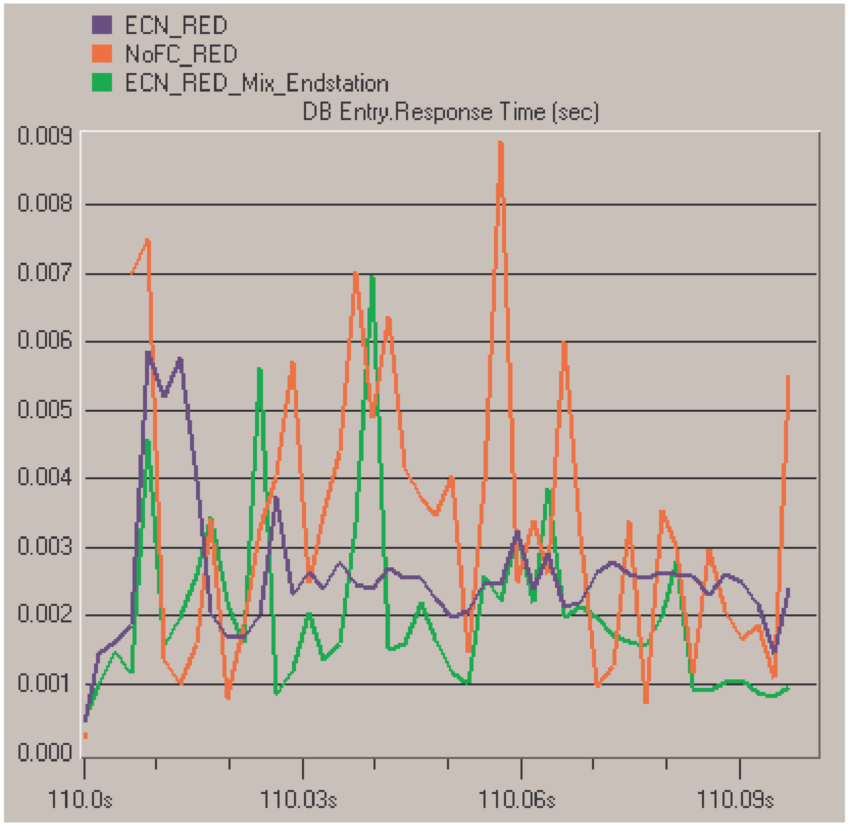
- NoFC_RED
  - No End-Station is L2-CI capable
- ECN_RED_Mix_Endstation
  - Only Server_1, Server_3, Client_2, Client_5 are L2-CI capable
- ECN_RED
  - All End-Stations are L2-CI capable

Note:

1. All switches are L2-CI capable in all the above scenarios
2. In modeling, it is assumed that the TCP ECN capability negotiation takes into account L2-CI capability of the End-Station

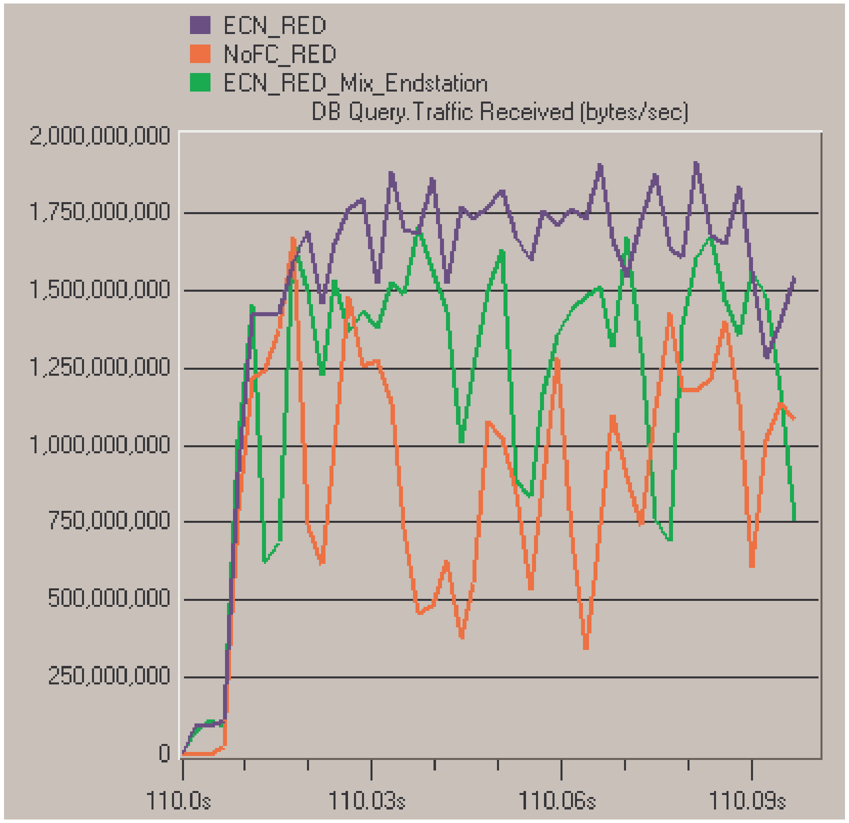
Application DB Entry Throughput & Response Time (Buffer = 32 KB per Switch Port)

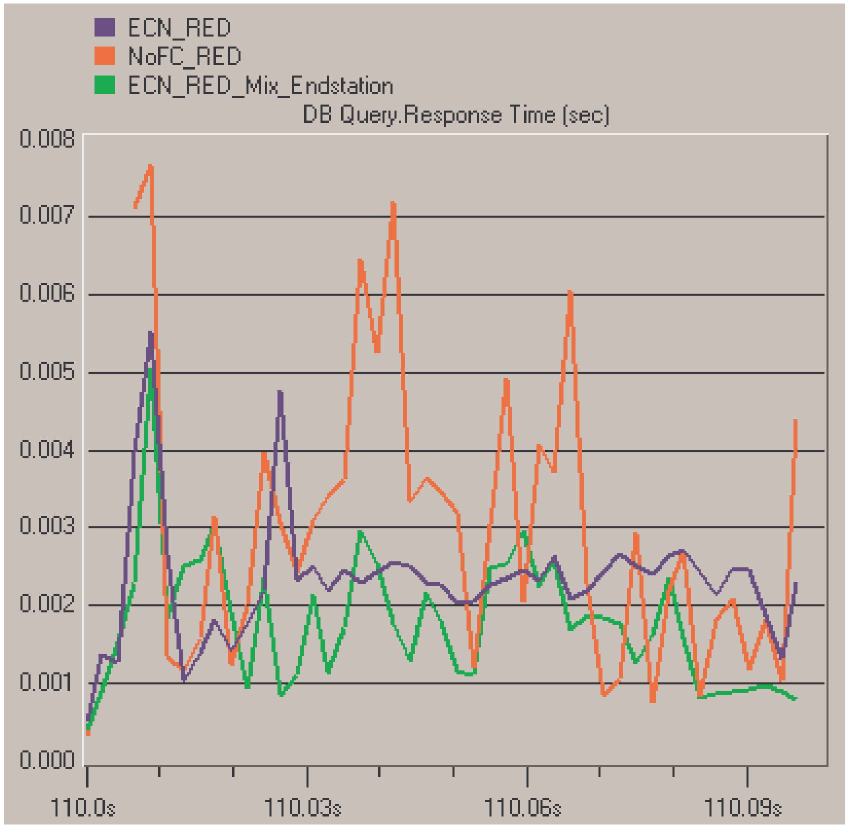




Incremental upgrade of End-Stations shows proportional performance improvement

Application DB Query Throughput & Response Time (Buffer = 32 KB per Switch Port)





Summary

- Simulations show that incremental deployment of ECN and L2-CI
  - should not adversely affect performance
  - shows improvement in performance

Next Steps

- Simulations with mix of TCP and UDP traffic flows
- Study need for CCT and the conditions under which it should be set
- Compare relative performance of messaging vs. marking
- Study use of L2-CI (and similar) mechanisms for non-TCP protocols