Simulation Analysis of Congestion Isolation (CI)

Kevin Shen

kevin.shenli@huawei.com

Sam Sun

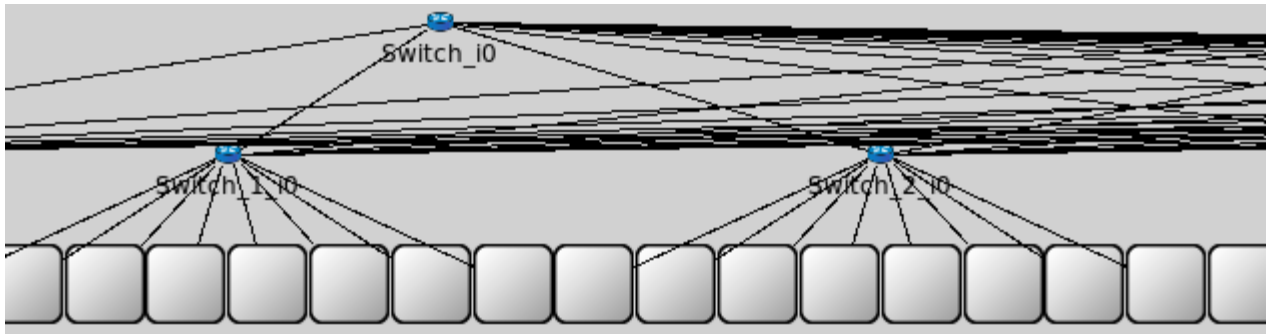
sam.sunwenhao@huawei.com

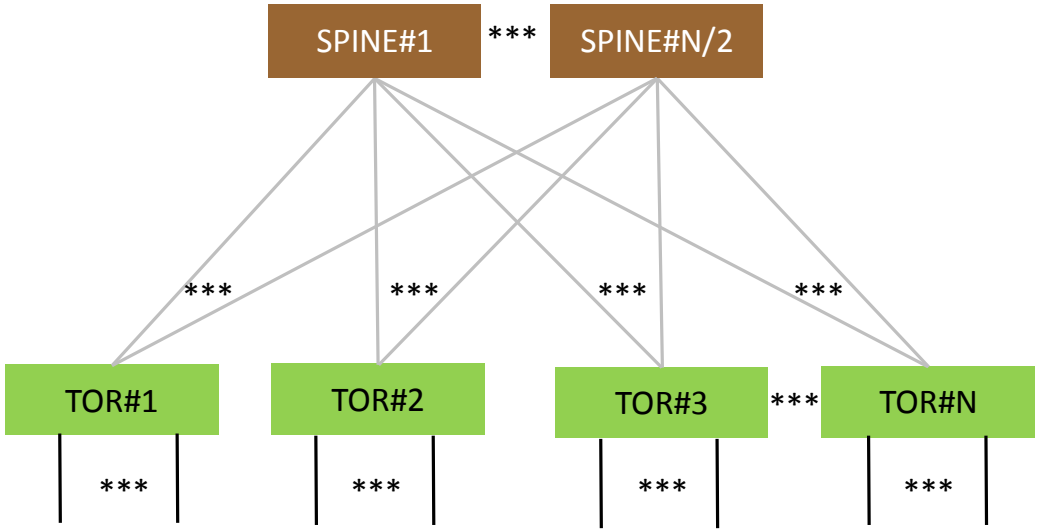
IEEE 802.1 DCB, Chicago, March 2018

Objectives of the Analysis

- **Investigate the XON threshold impact**
  - Keep other configuration unchanged, and compare the performances under different XON threshold settings
- **Find out the best combination of local CI, signaling and PFC**
  - Compare the performances under different combinations of local CI, signaling and PFC

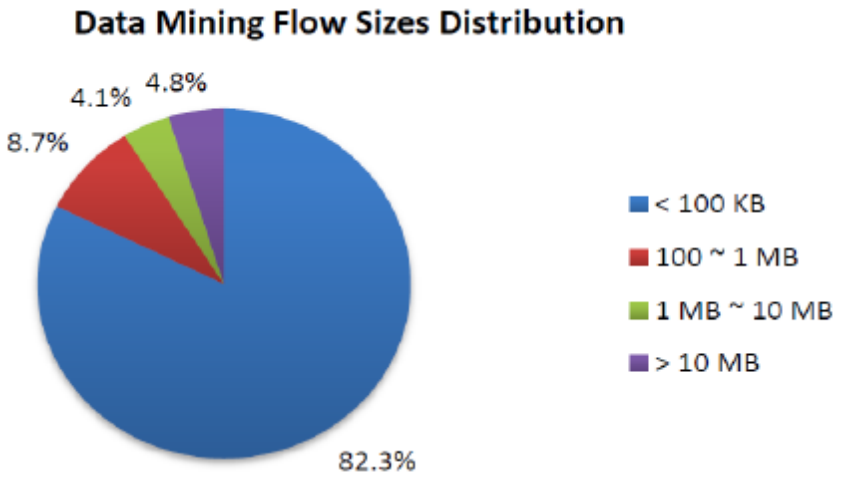
Simulation Set-up

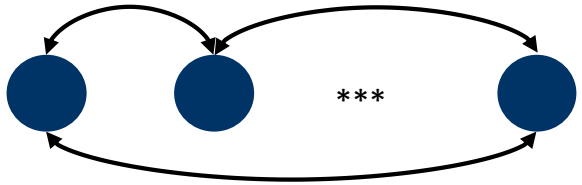




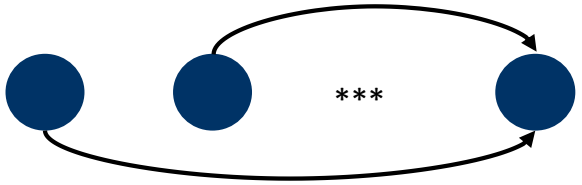
- **Platform:** OMNET++
- **2 Tier CLOS:** 100GbE interface with 200ns of link latency (about 40 meters)
- **Scale:** 1152 servers, 72 switches

Simulation Set-up





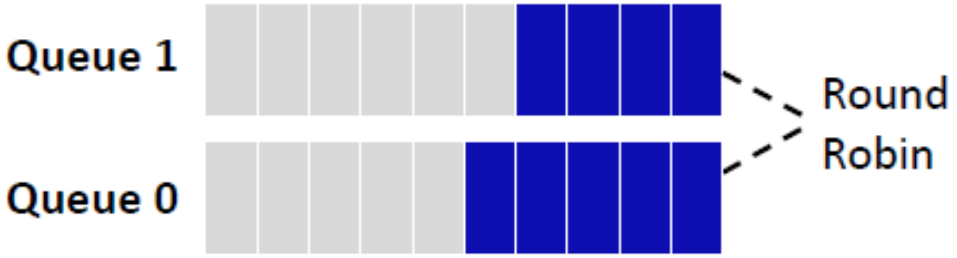
Many to many traffic



Many to one incast traffic

- **Traffic Pattern**
  - Data mining applications with different flow size distributions
  - Randomly select 21 servers as a small cluster for many to many traffic, 50 that kind of small clusters in all.
  - Randomly select 20:1 permanent many to one incast traffic, 4 that kind of many to one incasts in all.

Compared Solutions for Objective 1

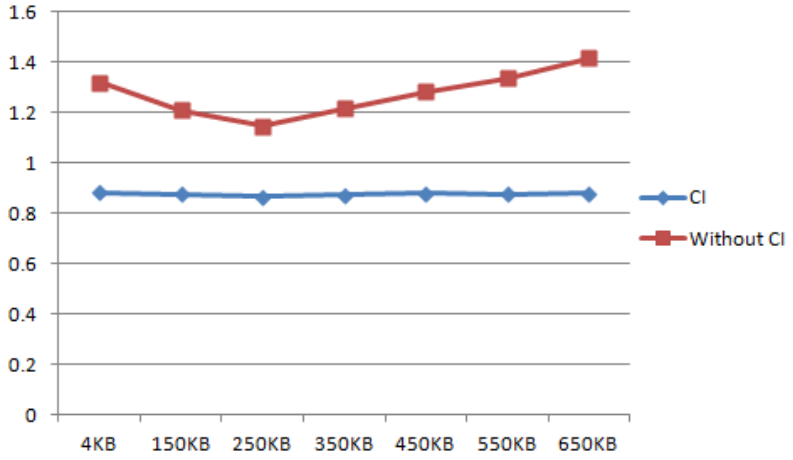


- Solution "**Without CI**" means PFC + ECN without CI.
- Flows are mapped to one of the two queues by hash of destination IP.
- PFC and ECN are enabled on both queues.
- Queue setting:
  - Queue size: 1 MB;
  - PFC threshold: XOFF 750 KB;
  - ECN: Low 10 KB, High 300 KB, Max Probability 1%.

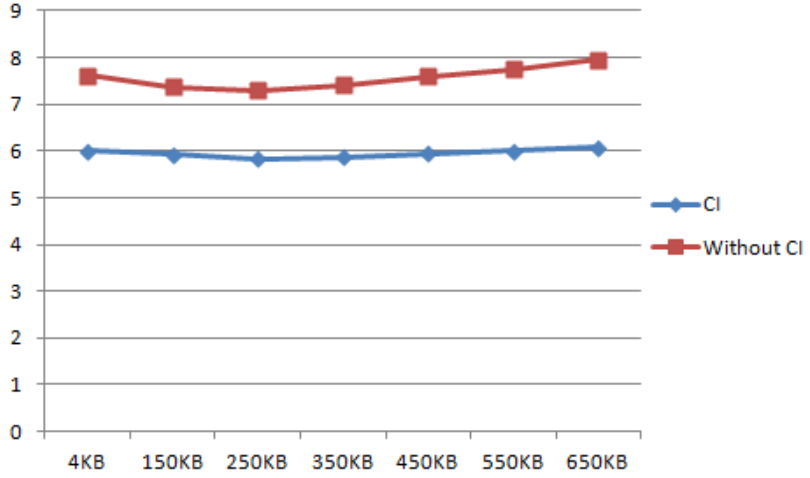
- Solution "**CI**" means PFC + ECN with CI.
- Flows go through the non-congested queue by default, and congested flows are dynamically isolated to the congested queue based on congestion.
- ECN is marked once a packet is isolated.
- Queue setting:
  - Queue size: 1 MB;
  - PFC threshold: XOFF 750 KB;
  - CI: Low 10 KB, High 300 KB, Max Probability 1%.

XON threshold impact

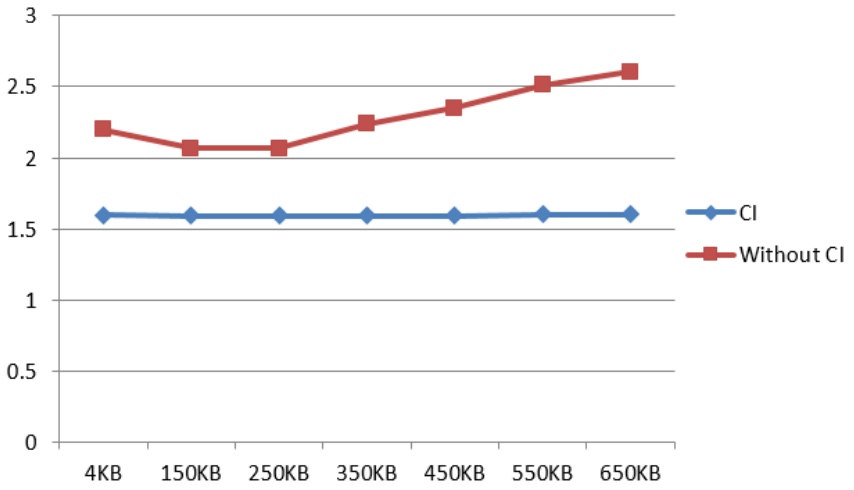
Average flow completion time(ms)
(all flows)



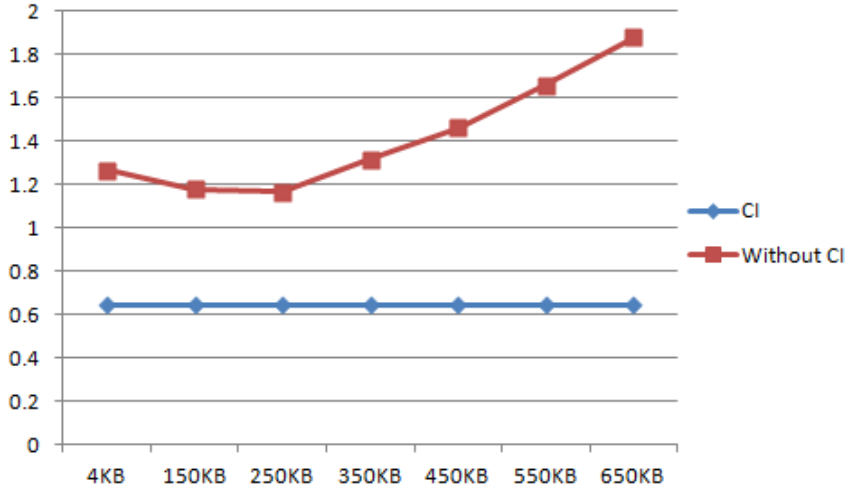
Average flow completion time(ms)
(>10MB flows)



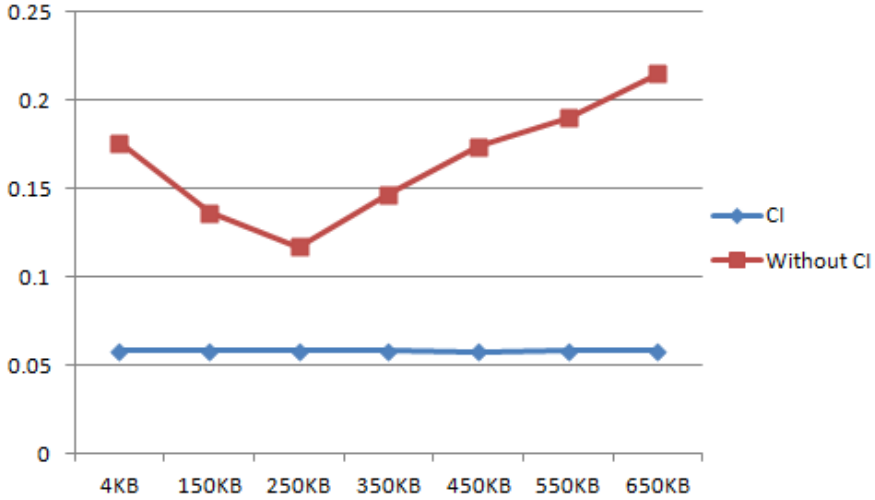
Average flow completion time(ms)
(1MB~10MB flows)



Average flow completion time(ms)
(100KB~1MB flows)



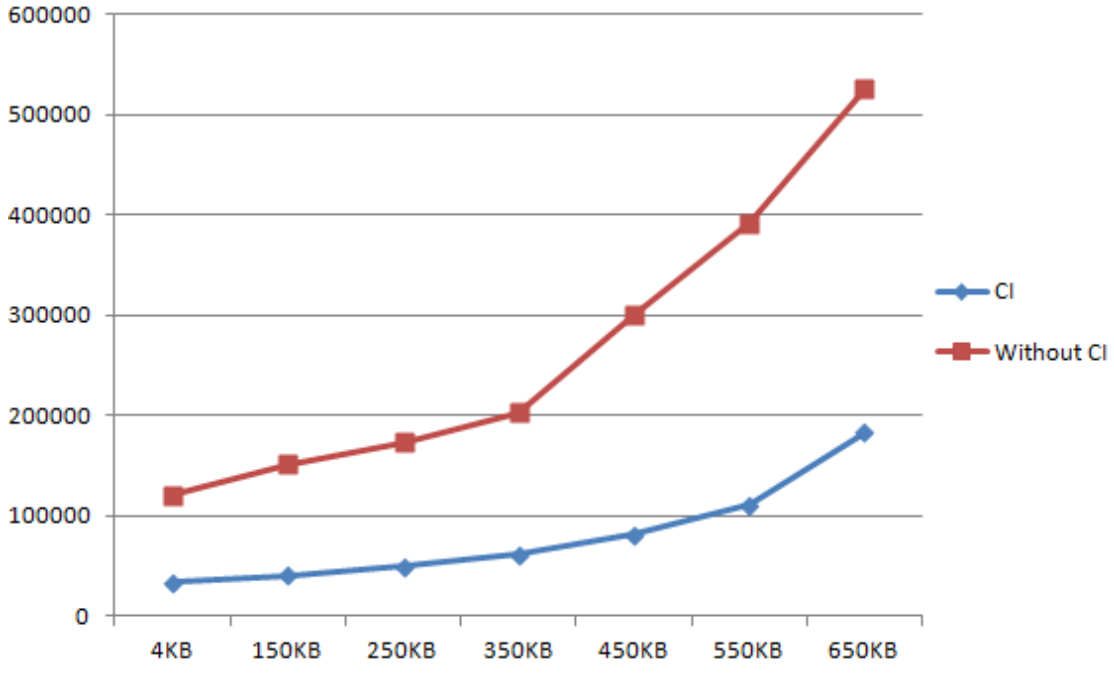
Average flow completion time(ms)
(<100KB flows)



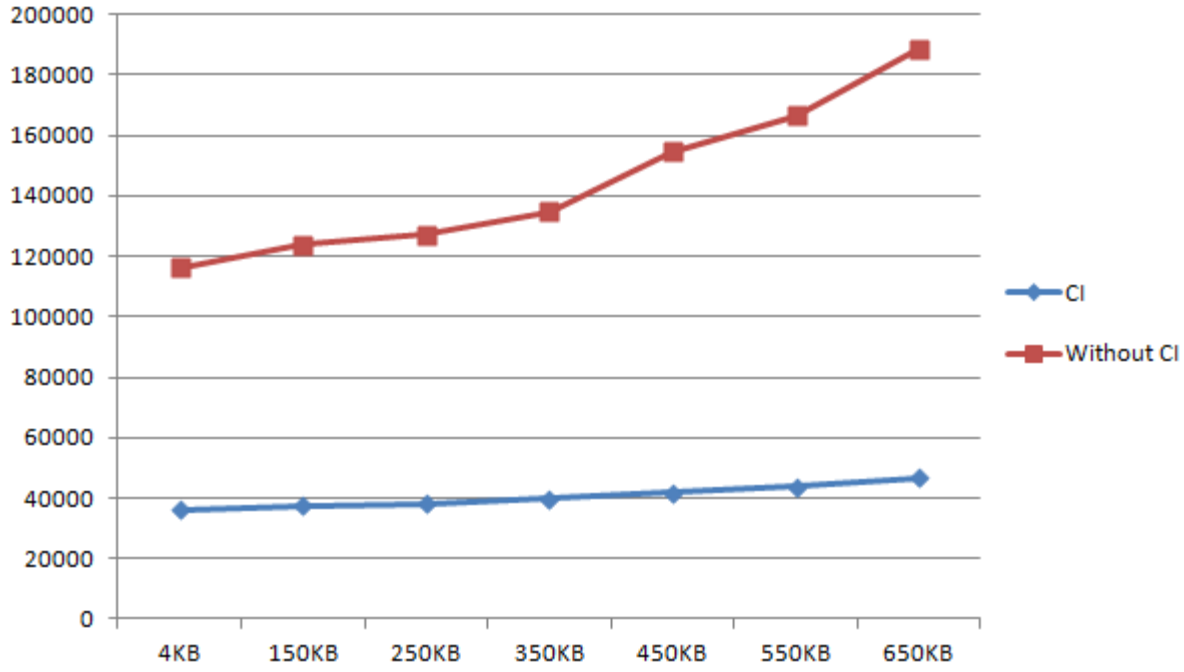
- For solution “Without CI”, XON threshold is critical.
- But for CI, XON threshold is not so important, because PFC only impact the congested flow.
- Even at the best configuration of XON threshold, CI has a big performance improvement compared to “Without CI”.

XON threshold impact

Pause Frame Count Receive by Servers



CNP Count Received by Servers

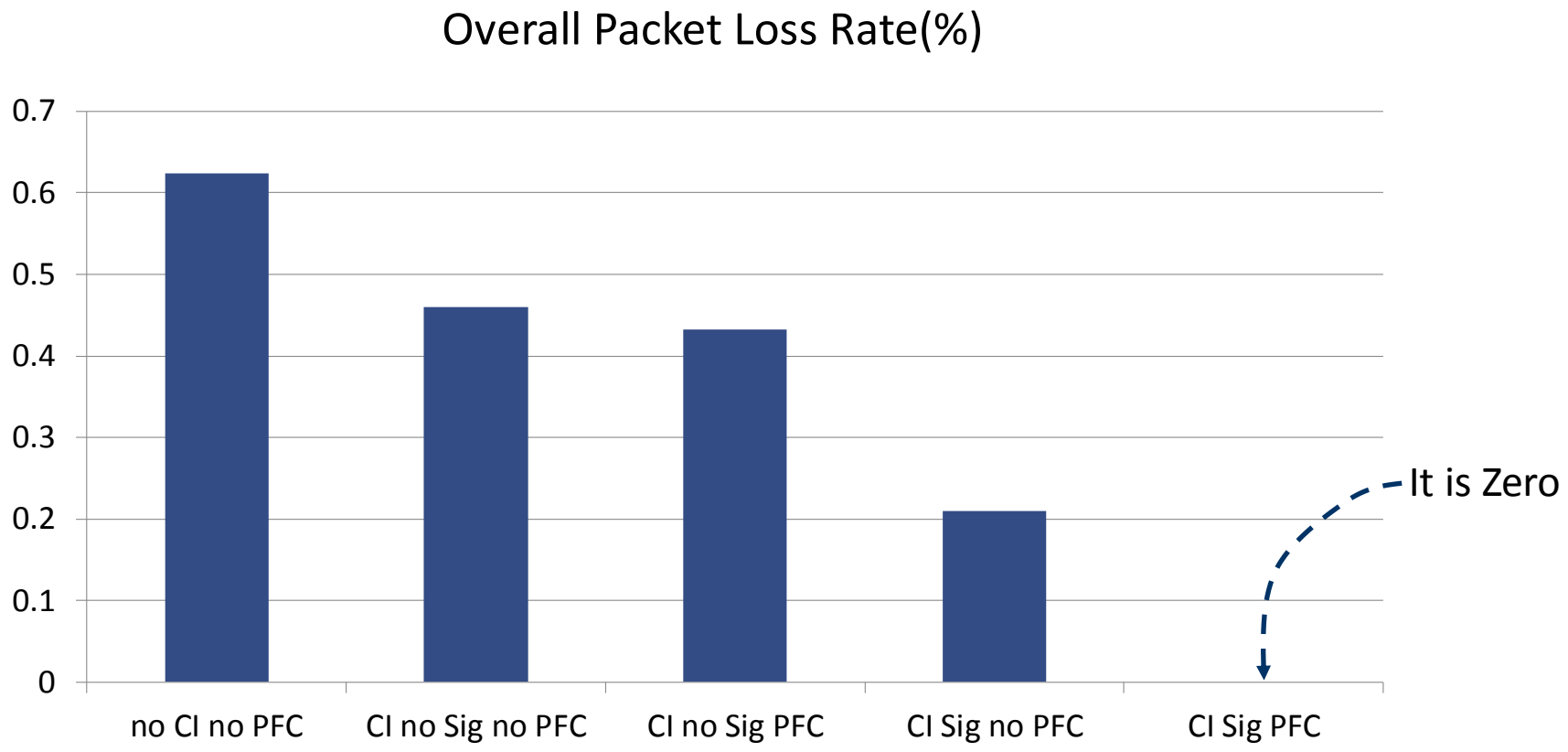


- “CI” can reduce Pause frame count and CNP count significantly at all XON threshold setting.

Compared Solutions for Objective 2

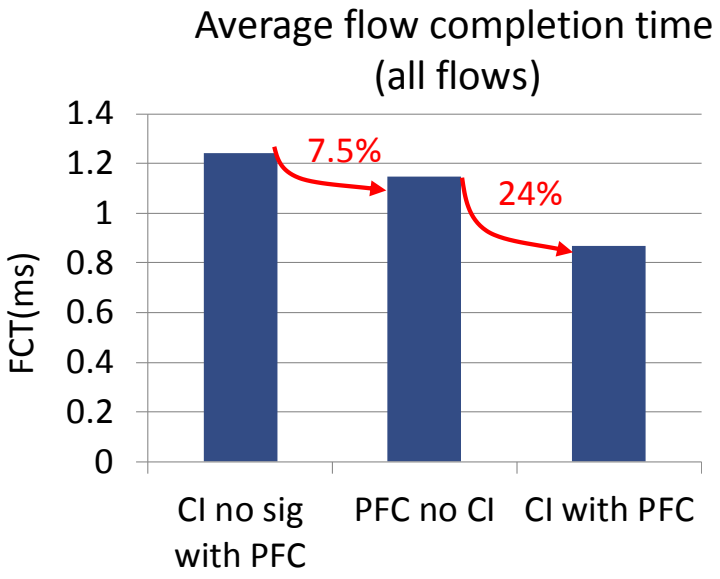
- Compared Solutions:
  - "**no CI no PFC**": Just ECN with neither CI nor PFC.
  - "**CI no Sig no PFC**": Local CI with ECN, but no signaling to upstream to isolate the congested flow and no PFC.
  - "**CI no Sig PFC**": Local CI with ECN and PFC, but no signaling to upstream to isolate the congested flow.
  - "**CI Sig no PFC**": Intact CI with ECN but without PFC.
  - "**CI Sig PFC**": Intact CI with ECN and PFC.

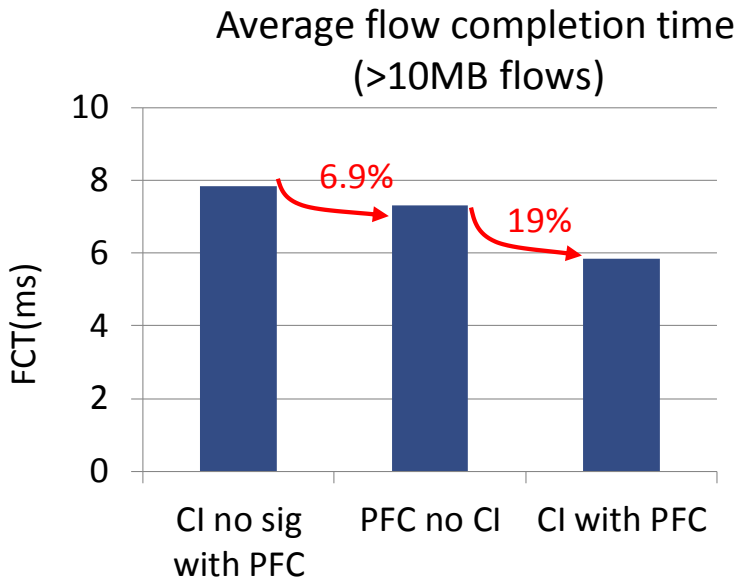
Packet Loss Rate Comparison

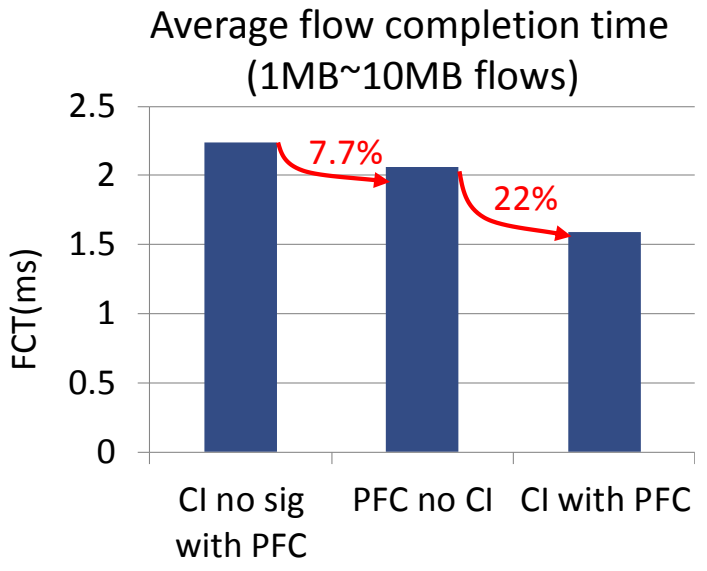


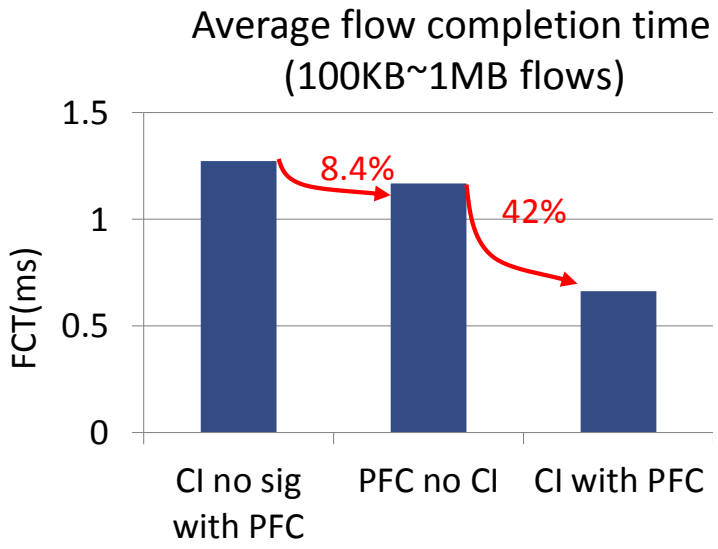
- Without signaling or PFC, CI solutions cannot prevent packet loss, only intact CI with PFC can.

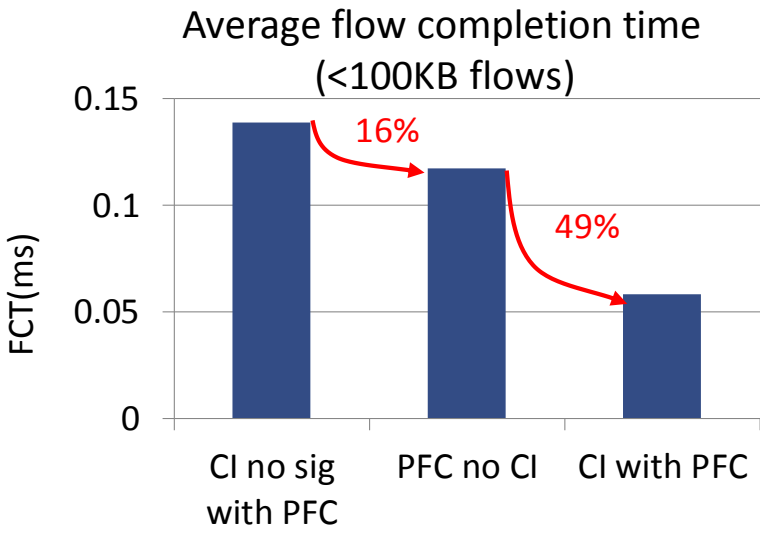
FCT Comparison between Solutions with PFC











- We make “CI no sig with PFC” lossless (Pause the non-congested queue and pause both queue as last resort), but it performs even worse than “PFC no CI”.
- Because under “CI no sig with PFC”, it plays just like one queue model with PFC enabled, which involves more HOLB.

Summary

- **Investigate the XON threshold impact**
  - There is a best XON threshold setting around 250KB.
  - Compared with “Without CI”, “CI” gets much less impacts from the XON threshold.
  - Even at the best configuration of XON threshold, “CI” has a big performance improvement compared to “Without CI”.
- **Find out the best combination of local CI, signaling and PFC**
  - The intact “CI” with PFC has best performance.

Questions?