An Inline Measurement Mechanism for High-Speed Networks

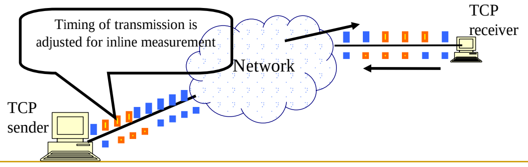
Osaka University
Cao Le Thanh Man, Go Hasegawa and Masayuki Murata
mlt-cao@ist.osaka-u.ac.jp

Content

- Available bandwidth & bandwidth measurement
- Inline network measurement
- Problems with measurement in high-speed networks
- ICIM: Interrupt Coalescence –aware inline measurement
  - Works well in high-speed networks
- Simulation results

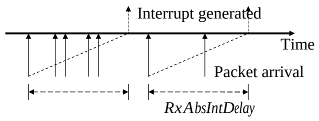
Our measurement approach: Inline network measurement

- **Performing active measurement without probe traffic**
- Inline measurement
  - Using data packets in a TCP connection as probe packets
- Our previous work: ImTCP
  - Adjusting the transmission intervals of some data packets
  - Measuring the available bandwidth from arrival intervals of ACK packets



Bandwidth measurement in high-speed network

- Interrupt Coalescence (IC)
  - Deployed in most high-bandwidth Network Interface Cards (NICs)
  - Multiple packets are grouped and passed to the kernel in a single interrupt
  - Absolute timer (default) [1]

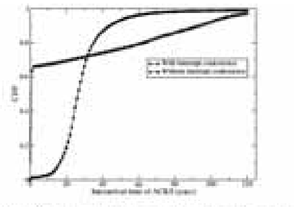


[1]: Intel, "Interrupt moderation using Intel Gigabit Ethernet Controllers" http://www.intel.com/design/network/applnots/ap450.pdf (2003)

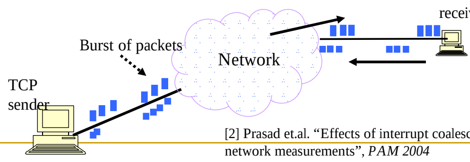
- **Why existing tools can not work in high-speed network?**
  - Limitation in packet pacing
    - Active measurement tools: Utilizing packet intervals for measurement
    - High-speed network measurement requires small packet sending/receiving intervals: 1Gps: 0.012ms (packet size 1500B)
    - For a general-purpose machine, such small intervals causes high CPU overhead
  - Effects of Interrupt Coalescence
    - Inter-arrival intervals of packets are changed

Effect of Interrupt Coalescence on TCP

- Bursty transmission of TCP packets
  - The bursty arrival of packets at the receiver causes bursty transmission of ACKs, and subsequently bursty transmission of more data packets from the sender.
  - With IC, 65% of ACKs arrive with intervals of less than 1 μs[2]



CDF of ACK interarrivals in a 10-second TCP connection over a GigE path.



[2] Prasad et.al. "Effects of interrupt coalescence on network measurements", *PAM 2004*

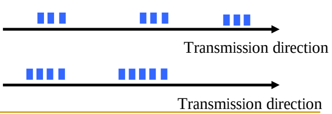
Proposed algorithm: ICIM

Interrupt Coalescence-aware Inline Measurement

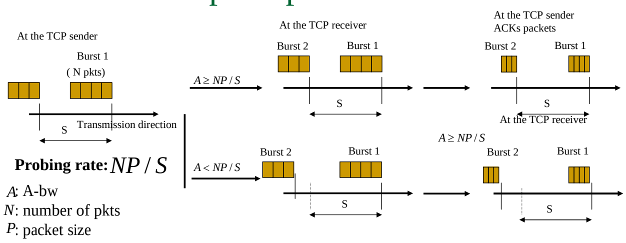
- Objective
  - A new inline measurement algorithm
  - Measuring available bandwidth when IC is enabled
  - Not requiring packet pacing
- Approach
  - Packet-burst based measurement
  - Making use of bursty transmission of TCP when IC is enabled: Adjusting the number of packets in a burst for the measurement
  - Adjustment of number of packets: The TCP sender store one whole burst then add packets to the next bursts

Transmission without adjustment in packet number of bursts

After adjustment



Measurement principle



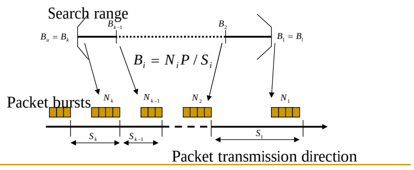
Probing rate: NP / S

A: A-bw
N: number of pkts
P: packet size
S: transmission interval of two bursts

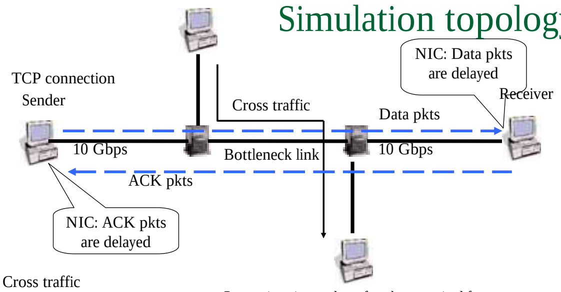
- Measurement principle
  - If probing rate NP/S of a burst is larger than A-bw, then the arrival interval of the two bursts is enlarged by the network, otherwise, it is not.
    - S is estimated from the size of Burst 1 and the average throughput of TCP

Steps in a measurement

- Decide a search range
  - Statistical information from previous measurement results
  - Faster measurement
- Adjust packets in K burst to probe K points in the search range
  - $K = 4$ in the simulations
- Infer the available bandwidth from the probing results
- Wait for Q seconds before starting a new measurement
  - Q must be long enough for TCP transmission to recover from the packet store-and-forward process, otherwise the estimation of S is not good



Simulation topology



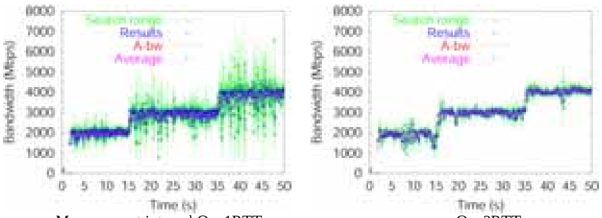
- Cross traffic
  - UDP flows
  - Packet size according to the monitored results of the Internet reported in NLANR

Comparison in number of packets required for a measurement

$A - bw$	ICIM	IC-aware Pathload	Ratio ICIM:Pathload
2 Gbps	110	200 · 12 · 8 = 19 200	0.006
3 Gbps	130	200 · 12 · 9 = 21 600	0.006
4 Gbps	154	200 · 12 · 10 = 24 000	0.006

- ICIM uses far fewer packets than Pathload

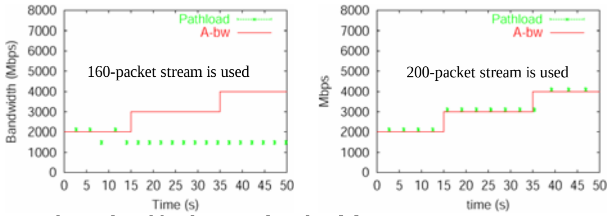
Measurement results for ICIM



Measurement interval Q= 1RTT   Q= 2RTTs

- **ICIM can deliver good measurement results in high-speed network**
- Q=2RTTs has better accuracy
  - The estimation of S is better because TCP transmission can recover from the last measurement
  - Measurement frequency is only a half: 16.7 results/s vs. 34.2 results/s

Measurement results of Pathload



- Enhanced Pathload is introduced in [2]
  - The last packet of a burst is supposed to be delayed shortly
  - Use only the last packet for Pathload algorithm
- Results are yielded in rather long intervals
  - 0.28 results/s (when 200-packet stream is used)
- Long packet streams are required

[2] Prasad et.al. “Effects of interrupt coalescence on network measurements”, *PAM 2004*