Inline bandwidth measurements: Implementation difficulties and their solutions

Tomoaki Tsugawa, Cao Le Thanh Man,
Go Hasegawa, and Masayuki Murata

Osaka University, Japan

Contents

Background

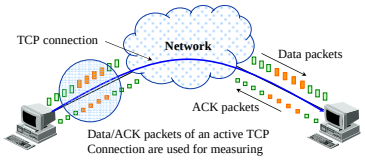
- **Varied service-oriented networks have emerged**
  - e.g., CDNs, P2P networks, Grid networks, IP-VPN
- **Acquiring the bandwidth information is important**
  - To use the resource of bandwidth effectively
  - To improve the quality of the network services
- **Many measurement mechanisms have been proposed**
- **Problems in existing measurement mechanisms**
  - Send extra traffic onto the network for measuring
  - Require a long time to obtain one measurement result

inline network measurement

The concept we have proposed to counter the above problems

Inline network measurement

- **Concept of ``plugging'' an active bandwidth measurement into an active TCP connection**



- **Advantages**
  - Require no extra traffic for measuring
  - Yield measurement results continuously and quickly
  - Require only modification of the sender end-host

Objective

- **Inline network measurement mechanisms have advantages**
- **Some problems occur when implementing the mechanisms**

- **Clarify the implementation difficulties**
  - Clock resolution of kernel system
  - IC: Interrupt Coalescence
  - Behavior of TCP receiver
- **Consider the solutions against their problems**
- **Evaluate the effectiveness of inline network measurement**
  - Implement the mechanisms in FreeBSD 4.10 kernel system
  - Test the mechanisms in the experimental network environments

Basic idea for measuring available bandwidth

Inter-arrival intervals of packets are increased when the transmission rate is higher than available bandwidth

Measurement principle

- Adjust the intervals of packets at the sender end-host
- Observe the intervals of packets at the receiver end-host
- Estimate the available bandwidth based on the sending/receiving intervals

Implementation difficulties (1/3)

Clock resolution of kernel system

- **Clock resolution of kernel system is coarse**
- **Reduce the accuracy of measurement results**
- **Clock resolution is determined by *HZ* in FreeBSD**
  - Upper bound of the measurable bandwidth depends on *HZ*

HZ	Clock resolution (1 tick) [μsec]	Measured bandwidth [Mbps] 1 tick	2 ticks	3 ticks	4 ticks	5 ticks
100	10,000	1.2	0.6	0.4	0.3	0.24
1,000	1,000	12	6	4	3	2.4
10,000	100	120	60	40	30	24
20,000	50	240	120	80	60	48
50,000	20	600	300	200	150	120
100,000	10	1,200	600	400	300	240

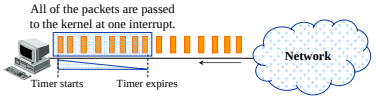
  - High *HZ* affects the performance of system

We should consider the trade-off relationship between the measurement accuracy and the performance of system

Implementation difficulties (2/3)

IC: Interrupt Coalescence

- **IC is deployed in most of gigabit NICs**
  - Group multiple packets arriving in a short period of time
  - Pass the packets to the kernel system in a single interrupt
- **IC is important for reducing the CPU overhead**

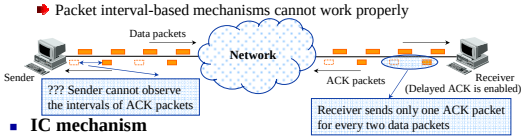


- **IC has an impact on packet interval-based mechanisms**
  - Inter-arrival intervals of packets observed by the kernel are changed
    - Intervals of packets in a single interrupt become almost zero
  - Accuracy of measurement results is degraded

Implementation difficulties (3/3)

Behavior of TCP receiver

- **Delayed ACK option**
  - TCP receiver does not generate an ACK packet for each data packet when delayed ACK option is enabled
    - TCP sender fails to observe the intervals of corresponding ACK packets
  - Packet interval-based mechanisms cannot work properly



- **IC mechanism**
  - Timeout value of IC at the receiver is larger than that at the sender
  - Inter-arrival intervals of packets are influenced of IC at the receiver
  - IC mechanism at the receiver end-host also becomes a problem

Implementations and experiments of inline network measurement mechanisms

- **Implement two inline network measurement mechanisms**
  - ImTCP: Inline measurement TCP [9]
    - Packet interval-based measurement mechanism
  - ICIM: Interrupt Coalescence-aware Inline Measurement [10]
    - Packet-burst interval-based measurement mechanism
    - Measurement mechanism for gigabit network environment
- **Test the mechanisms in experimental networks**

[9] Cao Le Thanh Man, Go Hasegawa, and Masayuki Murata, ``Available bandwidth measurement via TCP connection,'' in Proceedings of IFIP/IEEE E2EMON 2004, Oct. 2004.
[10] Cao Le Thanh Man, Go Hasegawa, and Masayuki Murata, ``ICIM: An inline network measurement Mechanism for high-speed networks,'' in Proceedings of IFIP/IEEE E2EMON 2006, Apr. 2006.

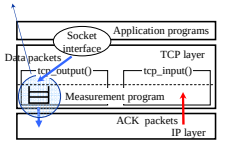
ImTCP: Inline measurement TCP

- **Packet interval-based mechanism**
  - procedure
    - Adjust the intervals of data packets
    - Observe the intervals of ACK packets
    - Estimate the available bandwidth based on the intervals of data/ACK packets
  - Use ``search range'' for reducing the number of probing packets

Build the FIFO buffer at the bottom of TCP layer

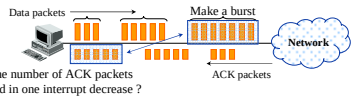
Implementation

- **tcp_output() function**
  - -- Store data packets in the FIFO buffer
  - Pass the packets to IP layer in the intervals based on ImTCP algorithm.
  - -- Record the sending time of data packets
- **tcp_input() function**
  - -- Record the receiving time of ACK packets
  - -- Estimate the available bandwidth based on the sending/receiving time



ICIM: Interrupt Coalescence-aware Inline Measurement

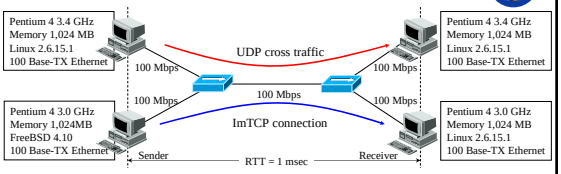
- **Packet-burst interval-based mechanism**
  - Adjust the number of data packets in a probing burst
  - Observe the number of ACK packets received in one interrupt



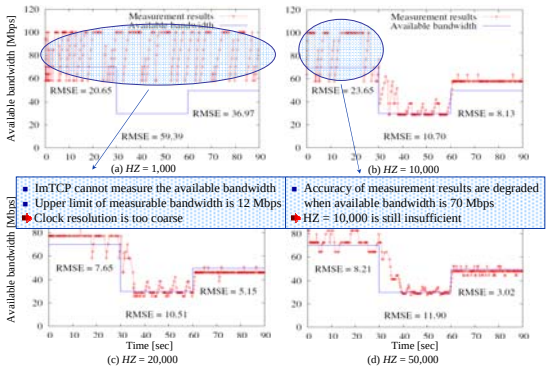
Implementation

- **Add some modification to the ImTCP implementation**
  - –Adjust the number of packets in the FIFO buffer
  - –Record the number of data packets in a probing burst
  - –Record the number of ACK packets received in one interrupt

Experiments in low-speed network



- **Available bandwidth**
  - 0-30 sec: 70 Mbps, 30-60 sec: 30Mbps, 60-90 sec: 50 Mbps
- **Evaluate the following index when changing *HZ***
  - Accuracy of measurement results of ImTCP
  - Performance of the system



- ImTCP cannot measure the available bandwidth
- Upper limit of measurable bandwidth is 12 Mbps
- Clock resolution is too coarse

- Accuracy of measurement results are degraded when available bandwidth is 70 Mbps
- HZ = 10,000 is still insufficient

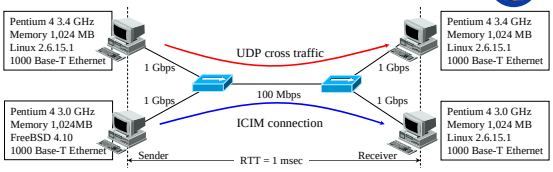
Average CPU Utilization

HZ	ImTCP [%]	TCP Reno [%]
1,000	3.07	11.28
10,000	13.21	12.22
20,000	14.19	14.01
50,000	24.42	22.86

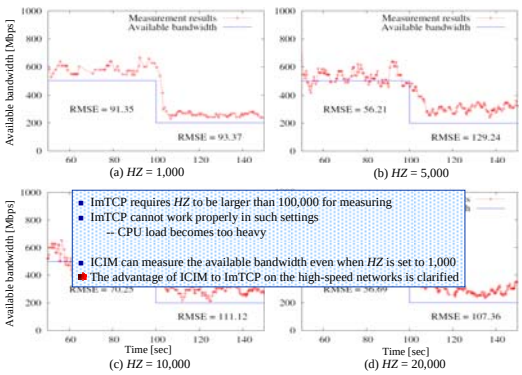
- Clock resolution becomes 1 msec when HZ is set to 1,000
- Intervals of probe packets becomes more than 1 msec
- Degrades the data transmission throughput

- **Required value of *HZ* when using ImTCP**
  - More than 10,000 not to degrade the transmission throughput
  - More than 20,000 to measure the available bandwidth accurately
  - As small as possible to reduce the CPU overhead
- ***HZ* = 20,000 is good choice in this network environment**

Experiments in gigabit network



- **Available bandwidth**
  - 50-100 sec: 500 Mbps, 100-150 sec: 200 Mbps
- **Confirm that ICIM can work well in gigabit networks**
- **Evaluate the measurement accuracy when changing *HZ***



- ImTCP requires *HZ* to be larger than 100,000 for measuring
- ImTCP cannot work properly in such settings
  -- CPU load becomes too heavy

- ICIM can measure the available bandwidth even when *HZ* is set to 1,000
- The advantage of ICIM to ImTCP on the high-speed networks is clarified

Conclusions

- **We clarified the implementation difficulties of measurement mechanisms**
  - Clock resolution of kernel system
  - IC: Interrupt Coalescence
  - Behavior of TCP receiver
- **We showed the current solutions against their problems**
- **We implemented the ImTCP and ICIM in FreeBSD**
  - Measurement mechanisms based on the inline network concept
  - Source codes of ImTCP and ICIM can be found at our web site:
    - http://www.anarg.jp/imtcp/
- **Experimental results showed the effectiveness of our inline network concept on actual networks**

Future studies

- **Evaluate the effectiveness of ICIM on various network**
  - e.g., the Internet and the more high-speed networks
- **Proposed mechanisms for measuring other bandwidth information based on the inline network concept**
  - Evaluate the mechanisms through simulations and in actual networks