Inferring available bandwidth of overlay network paths based on inline network measurement

M. Cao, G. Hasegawa and M. Murata
Osaka University, Japan

ICIMP 2007

Background

- Overlay networks
  - Allow end nodes to decide the overlay route of traffic
  - Quickly detect and recover from path outages and periods of degraded performance
- Overlay routing requires network resources information: bandwidth, delay, loss ratio…
  - Up-to-date
  - of all overlay paths
  - for every overlay nodes

Content

- Background
- Our previous research
  - Inline network measurement
- Proposed bandwidth monitoring systems
  - ImSystem
  - ImSystemPlus
- Evaluation results
- Conclusions

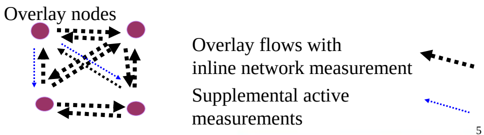
Monitoring available bandwidth

Large measurement overhead

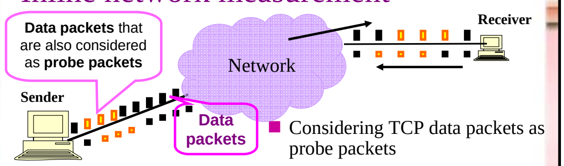
- Measure the full mesh (RON)
  - Overhead $O(N^2)$ *N*: number of overlay nodes
  - 25 GByte for just one snap shoot
    - ✓ One measurement: 2.5 MByte (Pathload)
    - ✓ For network of 100 nodes
- Exploit IP topology to reduce measurements
  - Exploit overlapping parts of overlay paths
    - ✓ C. Tan et al., (ICNP'03), Y. Chen et al., (SIGCOMM'04)
    - ✓ Overhead: $O(N \log(N))$
  - Exploit common bottleneck links
    - ✓ N. Hu et al. (IMC'05)
    - ✓ Overhead: $O(N)$

Proposed bandwidth monitoring approach

- What is the advantage?
  - Measurement overhead is small
- Basic idea
  - Deploying measurements in overlay traffic flows
    - ✓ Inline network measurement
  - Supplemental active measurements
    - ✓ For overlay paths where there is no overlay traffic



Our previous work: Inline network measurement



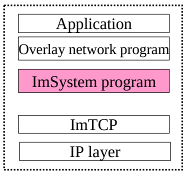
Inline **m**easurement **TCP**

Measures available bandwidth without extra probe traffic

- Considering TCP data packets as probe packets
  - Adjusting the transmission intervals of some data packets
  - Measuring the network characteristics from arrival intervals of ACK packets
  - Deliver results in every 2-4 RTTs

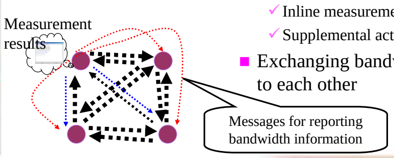
[4] M. Cao, G. Hasegawa and M. Murata, *ImTCP: TCP with an inline measurement mechanism for available bandwidth*, Computer Communications, vol. 29, no. 10, pp. 1614-1626, 2006

Proposed system: ImSystem

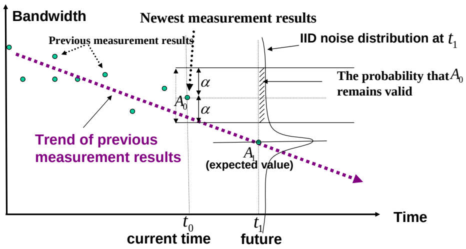


Placement of ImSystem program

- Overview
  - Distributed in every node of overlay network
  - Monitoring available bandwidth of all overlay network paths
- Work of an ImSystem node
  - Collecting bandwidth information of overlay paths that starts from it
    - ✓ Inline measurements in ImTCP
    - ✓ Supplemental active measurements
  - Exchanging bandwidth information to each other



Supplemental active measurement



- Active measurement will start if there is no overlay traffic after a certain period (from t_0 to t_1)
  - The length of the period is based on the slope of the trend
    - ✓ Bandwidth changes fast: active measurement will start soon

Exchanging bandwidth information

- ImSystem nodes report bandwidth information to the others
  - Report only big changes in measurement results
    ⇒ Reduce number of exchanged messages
- Equation for abrupt change detection from [6]

$$g_k = (1-\omega)g_{k-1} + \omega(y_k - \mu)^2, g_0 = 0$$

The newest inline measurement result (y_k)

Average of the measurement results (μ)

g_k : indicator of an abrupt change at the newest measurement

$g_k \geq h$: There is abrupt change in bandwidth values

$g_k < h$: There is no abrupt change

In the following simulations, we set $\omega = 0.5$ and $h = 120$

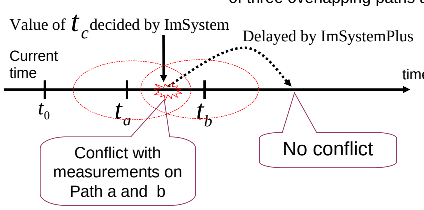
[6] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc., 1993.

ImSystemPlus

- Conflicts of active available bandwidth measurements
  - Degrade the measurement performance
  - Cause located congestion
- ImSystemPlus
  - Based on ImSystem
  - Decrease conflicts of active measurements in overlapping paths
  - Deploy IP network topology
    - ✓ Can be inferred by network tools like *traceroute*

Measurement conflict avoidance (1)

t_a t_b t_c : Scheduled active measurement time of three overlapping paths a, b, c…



- Basic idea
  - Delay active measurement on a path to reduce the probability of conflict with other measurements on the overlapping paths

Measurement conflict avoidance (2)

- Probability of conflict at timing t_y

$$Q(t_y) = \begin{cases} S(t_y) & S(t_y) < 1 \\ 1 & S(t_y) \geqq 1 \end{cases}$$

where

$$S(t_y) = \sum_{x; t_y - T < t_x < t_y + T} P_x \cdot joint(x, y)$$

$$P_x = e^{-\lambda_x \cdot (t_x - t_0)}$$

Probability that active measurement is performed at t_x

$$Joint(a,b) = \frac{Latency(G)}{min(Latency(a), Latency(b))}$$

Overlap index of paths a and b

- Probability of delaying measurement time *k.T* (seconds)

$$H^0 = 1 - Q(t_y)$$

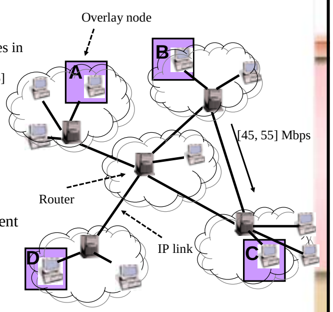
$$H^{k \cdot T} = \prod_{h=0..k-1} Q(t_y + h \cdot T) \cdot (1 - Q(t_y + k \cdot T))$$

(k = 0, 1…)

Delay time depends on
- How much the paths are overlapping
- The probability that the measurement on the paths are performed

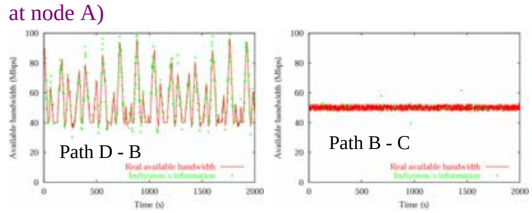
Simulation 1

- IP network
  - Link capacity: 100 Mbps
  - Available bandwidth: fluctuates in [0, 60] Mbps
    - For path from B to C: [45, 55] Mbps
  - Routing: Shortest path
- 4-node overlay network
  - Overlay flows
    - Arrival: Poisson process
    - Duration: Exponential distribution, average 20 s
    - Rate: [100 Mbps, 1Mbps]
- Supplemental active measurement
  - Supposed to be Pathload
  - Relative error: 10%
  - Duration: 10 s
  - Average rate: 2.5 Mbps
- Inline network measurement
  - Supposed to be ImTCP
  - Relative error: 20%
  - Duration: 1 s



Simulation 1

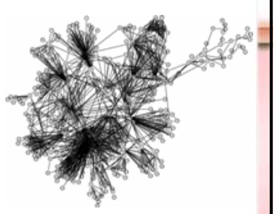
Bandwidth information collected by ImSystem (observed at node A)



- Path D-B: Bandwidth changes dramatically over time
  - ImSystem updates the information frequently
- Path B-C: Bandwidth of the path does not fluctuate significantly
  - ImSystem does not update the value very often to reduce the number of exchanged messages

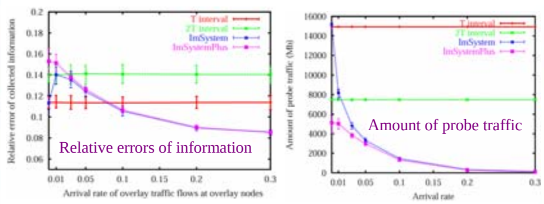
Simulation 2

- IP network topology
  - Sprint backbone network topology
  - 467 nodes, 1280 links
- Overlay network
  - 10-node overlay network
  - Randomly distributed in the IP network
  - Overlay traffic flows
    - Arrival: Poisson process
    - Duration: Exponential distribution, average 20 s
    - Rate: [100 Mbps, 1Mbps]
- Evaluation
  - 10 simulations with different distributions of overlay network
  - Average of the following values in 10 simulations are examined:
    - Relative error of the collected information
    - Amount of probe traffic



Sprint topology *(inferred by **Rocketfuel**)*

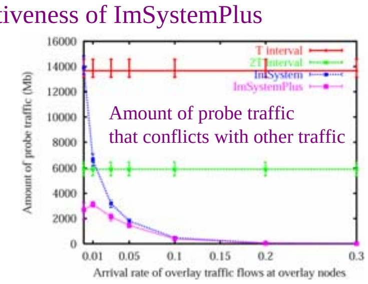
Simulation 2



- For comparison, the case when active measurement results are performed in fixed intervals (*T*, *2T*) are also shown
  - *T*=15(s), the time required for one active measurement
- ImSystem and ImSystemPlus
  - show small relative error,
  - send less active probe traffic when the overlay traffic increases

Simulation 2

Effectiveness of ImSystemPlus



- ImSystemPlus can decrease a large amount of conflict probe traffic
  - Decrease at most 80% of the conflict probe traffic that exists in ImSystem

Conclusions

- Distributed systems for bandwidth monitoring
  - Deploying inline network measurement
  - Small overhead
- Sending active measurement when there is not enough overlay traffic is a drawback
  - When there is little overlay traffic, probe traffic may cause no problems
- Future works
  - Implementation and evaluation in real environment