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 snapshot
    - 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
- 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

Proposed system: ImSystem

- **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

- **Placement of ImSystem program**
  - Application
  - Overlay network program
  - ImSystem program
  - ImTCP
  - IP layer

- **Measurement results**
  - Messages for reporting bandwidth information

---

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]

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

- \( g_k \) : indicator of an abrupt change at the newest measurement
- \( g_k \geq h \) : There is an abrupt change in bandwidth values
- \( g_k < h \) : There is no abrupt change

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

---

ImSystemPlus

- **Conflicts of active available bandwidth measurements**
  - Degrades measurement performance
  - Causes 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)

- **Scheduled active measurement time**
  - \( t_a, t_b, t_c \)
  - Value of \( t \) decided by ImSystem
  - Delayed by ImSystemPlus

- **Conflict with measurements on Path a and b**

- **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 \( t \)**

\[ P_{\text{conflict}}(t) = \frac{2}{S(t_a, t_b, t_c) < 1} \]

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

\[ H^{k,T} = \prod_{k=0}^{\infty} \left( \frac{Q(t_a + k \cdot T)}{1 - Q(t_a + k \cdot T)} \right) \]

- **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
- **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, 1 Mbps
- **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 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, 1 Mbps]
- **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

**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