Tomography-based Overlay Network Monitoring

Yan Chen, David Bindel, and Randy H. Katz

UC Berkeley
Motivation

• Infrastructure ossification led to thrust of overlay and P2P applications
• Such applications flexible on paths and targets, thus can benefit from E2E distance monitoring
  - Overlay routing/location
  - VPN management/provisioning
  - Service redirection/placement ...
• Requirements for E2E monitoring system
  - Scalable & efficient: small amount of probing traffic
  - Accurate: capture congestion/failures
  - Incrementally deployable
  - Easy to use
Existing Work

• **General Metrics:** RON ($r^2$ measurement)

• **Latency Estimation**
  - Clustering-based: IDMaps, Internet Isobar, etc.
  - Coordinate-based: GNP, ICS, Virtual Landmarks

• **Network tomography**
  - Focusing on inferring the characteristics of physical links rather than E2E paths
  - Limited measurements $\rightarrow$ under-constrained system, unidentifiable links
Problem Formulation

Given an overlay of $n$ end hosts and $O(n^2)$ paths, how to select a minimal subset of paths to monitor so that the loss rates/latency of all other paths can be inferred.

Assumptions:

- Topology measurable
- Can only measure the E2E path, **not** the link
Our Approach

Select a basis set of $k$ paths that fully describe $O(n^2)$ paths ($k \ll O(n^2)$)

- Monitor the loss rates of $k$ paths, and infer the loss rates of all other paths
- Applicable for any additive metrics, like latency
Modeling of Path Space

Path loss rate \( p \), link loss rate \( l \)

\[
1 - p_1 = (1 - l_1)(1 - l_2)
\]

\[
\log(1 - p_1) = \log(1 - l_1) + \log(1 - l_2) = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \log(1 - l_1) \\ \log(1 - l_2) \\ \log(1 - l_3) \end{bmatrix}
\]

\[
\begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = b_1
\]
Putting All Paths Together

Totally \( r = O(n^2) \) paths, \( s \) links, \( s \ll r \)

\[ Gx = b, \text{ where path matrix } G \in \{0 \mid 1\}^{r \times s} \]

link loss rate vector \( x \in \mathbb{R}^{s \times 1} \), path loss rate vector \( b \in \mathbb{R}^{r \times 1} \)
Sample Path Matrix

- $x_1 - x_2$ unknown => cannot compute $x_1$, $x_2$
- Set of vectors $\alpha[1 \ -1 \ 0]^T$ form null space
- To separate identifiable vs. unidentifiable components: $x = x_G + x_N$

$$G = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$G \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$x_G = \frac{(x_1 + x_2)}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} b_1/2 \\ b_1/2 \end{bmatrix}$$

$$x_N = \frac{(x_1 - x_2)}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
Intuition through Topology Virtualization

Virtual links:
- Minimal path segments whose loss rates uniquely identified
- Can fully describe all paths
- $x_G$ is composed of virtual links

All E2E paths are in path space, i.e., $Gx_N = 0$

$$b = Gx = Gx_G + Gx_N = Gx_G$$
More Examples

Real links (solid) and all of the overlay paths (dotted) traversing them

Virtualization

Virtual links

$G = \begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}$

$\text{Rank}(G)=2$

$G = \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 \\
\end{bmatrix}$

$\text{Rank}(G)=3$
Algorithms

\[ G \mathbf{x}_G = b \]

- Select \( k = \text{rank}(G) \) linearly independent paths to monitor
  - Use QR decomposition
  - Leverage sparse matrix: time \( O(rk^2) \) and memory \( O(k^2) \)
    - E.g., 10 minutes for \( n = 350 \) (\( r = 61075 \)) and \( k = 2958 \)

- Compute the loss rates of other paths
  - Time \( O(k^2) \) and memory \( O(k^2) \)
How many measurements saved?

\[ k \ll O(n^2) \]?

For a power-law Internet topology

- When the majority of end hosts are on the overlay
  \[ k = O(n) \] (with proof)

- When a small portion of end hosts are on overlay
  - If Internet a pure hierarchical structure (tree): \( k = O(n) \)
  - If Internet no hierarchy at all (worst case, clique):
    \[ k = O(n^2) \]
  - Internet has moderate hierarchical structure [TGYJ+02]

For reasonably large \( n \), (e.g., 100), \( k = O(n \log n) \)
(extensive linear regression tests on both synthetic and real topologies)
Practical Issues

• Topology measurement errors tolerance

• Measurement load balancing on end hosts
  - Randomized algorithm

• Adaptive to topology changes
  - Add/remove end hosts and routing changes
  - Efficient algorithms for incrementally update of selected paths
Evaluation

• Extensive Simulations
• Experiments on PlanetLab
  - 51 hosts, each from different organizations
  - $51 \times 50 = 2,550$ paths
  - On average $k = 872$

• Results Highlight
  - Avg real loss rate: 0.023
  - Absolute error mean: 0.0027
    90% $< 0.014$
  - Relative error mean: 1.1
    90% $< 2.0$
  - On average 248 out of 2550 paths have no or incomplete routing information
  - No router aliases resolved

<table>
<thead>
<tr>
<th>Areas and Domains</th>
<th># of hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>US (40)</td>
<td></td>
</tr>
<tr>
<td>.edu</td>
<td>33</td>
</tr>
<tr>
<td>.org</td>
<td>3</td>
</tr>
<tr>
<td>.net</td>
<td>2</td>
</tr>
<tr>
<td>.gov</td>
<td>1</td>
</tr>
<tr>
<td>.us</td>
<td>1</td>
</tr>
<tr>
<td>Europe (6)</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>1</td>
</tr>
<tr>
<td>Sweden</td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
</tr>
<tr>
<td>UK</td>
<td>2</td>
</tr>
<tr>
<td>Asia (2)</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>1</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1</td>
</tr>
<tr>
<td>Canada</td>
<td>2</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
</tr>
</tbody>
</table>
Conclusions

• A tomography-based overlay network monitoring system
  - Given $n$ end hosts, characterize $O(n^2)$ paths with a basis set of $O(n \log n)$ paths
  - Selectively monitor the basis set for their loss rates, then infer the loss rates of all other paths

• Both simulation and PlanetLab experiments show promising results