Understanding the Effect of P2P Overlay on the AS-level Underlay

Amir H. Rasti, Reza Rejaie
University of Oregon

Walter Willinger
AT&T Labs - Research

1. Introduction

Why is it important?
- P2P systems have become very popular in a wide range of applications.
- Each P2P application usually hosts a large number of users with a wide geographical distribution.
- Some measurement studies performed at limited vantage points show that a large portion of the Internet traffic is associated to p2p applications.

Target questions:
1. For a given overlay, how can one capture and represent its effect on the underlying network?
2. How do changes in the overlay affect the underlay?

Factors potentially affecting the amount of P2P traffic on each AS:
1. Overlay topology and peers’ connectivity and location
2. Pattern of traffic generation and routing in the overlay.
3. Underlay topology and routing.

2. Methodology

1. Capture topology snapshots of some popular p2p application:
   - Gnutella top layer overlay taken during the years of 2004-2007 using Cruiser. (177k–1.2M peers)
2. In each snapshot, group the peers based on their originator AS (according to BGP data gathered from RouteViews) (1872-3726 ASes)
3. Obtain AS relationship snapshots from CAIDA, one for each p2p snapshot taken close (within a month of) each. (38k – 50k relations)
4. Run C-BGP over each AS relationship snapshot to determine the “valley-free” AS-path carrying traffic between each pair of connected peers (100k-400k)
5. Calculate the traffic load on each AS for the following assumptions:
   a. Equal bidirectional traffic on all p2p connections
   b. Equal traffic generation at all nodes
      - Estimate the traffic of each connection using betweenness of each edge in the overlay graph.

Underlay Simplification
- AS represented as single node
- Incomplete AS relationship information
- Simplified AS relationships
  - Customer-Provider
  - Peer-Peer
  - Sibling-Sibling
- With the pool of peers from a Gnutella snapshot, generate several random overlays with the same degree distribution as the original
- Effect of peer identity and location

3. Metrics and Results

Analyses Performed
1. Base case
   - 4 Gnutella snapshots
2. Randomized Connections
   - Randomize all connections keeping the degree of each peer fixed.
   - Any change in the underlay load pattern would show the effect of peer connectivity pattern
3. Random Peers
   - Using the pool of peers from a Gnutella snapshot, generate several random overlays with the same degree distribution as the original
   - Effect of peer identity and location
4. Regional Proportions
   - Using the same pool, generate random overlays with different regional proportions.
   - Effect of peer location

<table>
<thead>
<tr>
<th>Name</th>
<th>Tier</th>
<th>07</th>
<th>Random Con(%)</th>
<th>Random Peers(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&amp;T(US)</td>
<td>1</td>
<td>07</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AOL(US)</td>
<td>2</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Level3(US)</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tiscali(IT)</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LambdaNet(DE)</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cogent(US)</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TeliaNet(SE)</td>
<td>2</td>
<td>7</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>Sprint(CA)</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Rogers(CA)</td>
<td>2</td>
<td>9</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>FranceTelecom(FR)</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>+1</td>
</tr>
<tr>
<td>EduNet(US)</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Server Central(US)</td>
<td>2</td>
<td>12</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>Sprint(US)</td>
<td>2</td>
<td>13</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>COLT(NL)</td>
<td>2</td>
<td>14</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>QWEST(US)</td>
<td>1</td>
<td>15</td>
<td>+3</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig 2: Distribution Of Transit Traffic
Fig 3: AS Path length distribution
Tab 2: Distribution of path top tier

4. Ongoing Work

Any-Any analysis
- Gather a large pool of peers from several recent snapshots (Gnutella & Kad)
- Divide ASes to 3 main grps of NA, EU, AS
- Assume full-mesh connectivity
- Separately, perform the analysis for all P2P connection from zone X to zone Y

Sub-path analysis
- Calculate the load on all sub-paths
- Determine the value of each AS

Traffic Matrix Analysis & Modeling