Internet Inter-Domain Traffic

Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, Farnam Jahanian

Presented by: Mario Sanchez
Instructor: Fabian Bustamante
Date: 01/10/2011
Goals

• Examine changes in Internet inter-domain traffic demands and interconnection policies
• Longitudinal observations of Internet traffic
Methodology

- Focused on inter-domain traffic, not application layer (web hits/tweets/VPN/etc)
- Exported coarse-grain traffic statistics about ASNs, ASPaths, protocols, ports, etc. via anonymous XML forwarded to central servers
- Leverage commercial probes within given ISPs, with limited visibility into payload-based classification
- Incorporated informal and formal discussions with providers, and information about known traffic volumes
- Validated predictions based on a ground-truth based on 12 known ISP traffic demands (Known peak Tbps)
Methodology

- Covered 110 ISPs/content providers
- 3k edge routers
- 100k interfaces
- ~25% of all inter-domain traffic
- Waited for 2 years

- Calculated percentages per category then weighted averages using number of routers in each deployment

\[
W_{d,i} = \frac{R_{d,i}}{\sum_{x=1}^{N} R_{d,x}} \quad P_{d}(A) = \sum_{x=1}^{N} W_{d,x} \times \frac{M_{d,x}(A)}{T_{d,x}} \times 100
\]
Internet Evolution

• Most of the past 15 years of commercial Internet:
  • 10 to 12 large transit providers, interconnecting:
    • Tier-2,
    • Regional providers,
    • Consumer networks
    • Content/hosting companies

• Last five years saw a shift in Internet inter-domain traffic demands and peering policies
  • Content providers build their own global backbones
  • Cable Internet service providers offer wholesale national transit
  • Transit ISPs offer CDN and cloud/content hosting services
Traditional Internet logical topology

Diagram showing the logical topology of the Internet, starting with National Backbone Operators at the top, followed by Regional Access Providers, then Local Access Providers, and finally Customer IP Networks at the bottom, leading to Consumers and business customers.
Emerging new Internet logical topology

2009: 65% of study participants use direct links with Google, 52% with Microsoft, 49% with Limeligh, 49% with Yahoo
ASN Traffic Analysis

• Calculate 10 largest contributors of inter-domain traffic using weighted average percentage (either originating or transiting each ASN)
• Aggregate all ASNs which are managed by the same Internet commercial entity
• Exclude stub ASNs from the aggregation step which only observed downstream from other corporate ASN
Impact of commercial policy and traffic engineering changes

- 2007: Largest Internet providers correlate with 12 largest transit networks (Tier 1)
- 2009: Includes the addition of non-transit companies to the list: Google and Comcast
The case of Google

- Google inter-domain traffic enjoyed the largest growth (gaining 4% of all-inter domain traffic)
- Google’s traffic share increase came through the post-acquisition migration of YouTube inter-domain traffic to Google’s ASNs
- Google the fastest growing ASN group
The case of Comcast – In/Out peering ratio

- Weighted average percentage of inter-domain traffic into all Comcast ASNs vs outbound
- Comcast began offering wholesale transit (GigE and 10GigE IP), cellular backhaul and IP video distribution
Inter-domain traffic consolidation

- 2007: 150 ASNs contribute 30% of all inter-domain traffic
- 2009: 150 ASNs originate more than 50% of all inter-domain traffic
- Majority of traffic by volume flows directly between large content providers, datacenter / CDNs and consumer networks
Application Traffic Analysis

• Methodology
  • Applications are classified by protocol and TCP/UDP port
  • The appliances follow heuristics to select a single probable application (each flow record may contain multiple ports)

• Limited
  • If application uses non-standard ports or ephemeral port numbers
  • Does not identify tunneled applications (video over HTTP) or encrypted traffic (P2P)
  • Port-based heuristics could not identify a probable application in more than 25% of all observed inter-domain traffic

• Augment
  • Using small set of DPI appliances (payload classification)
The winners: HTTP & video

<table>
<thead>
<tr>
<th>Rank</th>
<th>Application</th>
<th>2007</th>
<th>2009</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Web</td>
<td>41.68</td>
<td>52.00</td>
<td>+10.31</td>
</tr>
<tr>
<td>2</td>
<td>Video</td>
<td>1.58</td>
<td>2.64</td>
<td>+1.05</td>
</tr>
<tr>
<td>3</td>
<td>VPN</td>
<td>1.04</td>
<td>1.41</td>
<td>+0.38</td>
</tr>
<tr>
<td>4</td>
<td>Email</td>
<td>1.41</td>
<td>1.38</td>
<td>-0.03</td>
</tr>
<tr>
<td>5</td>
<td>News</td>
<td>1.75</td>
<td>0.97</td>
<td>-0.78</td>
</tr>
<tr>
<td>6</td>
<td>P2P</td>
<td>2.96</td>
<td>0.85</td>
<td>-2.11</td>
</tr>
<tr>
<td>7</td>
<td>Games</td>
<td>0.38</td>
<td>0.49</td>
<td>+0.12</td>
</tr>
<tr>
<td>8</td>
<td>SSH</td>
<td>0.19</td>
<td>0.28</td>
<td>-0.08</td>
</tr>
<tr>
<td>9</td>
<td>DNS</td>
<td>0.20</td>
<td>0.17</td>
<td>-0.04</td>
</tr>
<tr>
<td>10</td>
<td>FTP</td>
<td>0.21</td>
<td>0.14</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>2.56</td>
<td>2.67</td>
<td>+0.11</td>
</tr>
<tr>
<td></td>
<td>Unclassified</td>
<td>46.03</td>
<td>37.00</td>
<td>-9.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web</td>
</tr>
<tr>
<td>Video</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>VPN</td>
</tr>
<tr>
<td>News</td>
</tr>
<tr>
<td>P2P</td>
</tr>
<tr>
<td>Games</td>
</tr>
<tr>
<td>SSH</td>
</tr>
<tr>
<td>DNS</td>
</tr>
<tr>
<td>FTP</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Unclassified</td>
</tr>
</tbody>
</table>

(a) Port / Protocol
(b) Payload Matching

Video represents the second largest and second fastest growing application class
The losers: P2P

P2P saw the largest decline with a drop of 2.8% percentage between 2007 and 2009.

DPI: in 2007 shows P2P at 40% of all traffic, in 2009 shows P2P at 20%.
P2P decline

• Reasons?
  • Improvements in P2P client and algorithm efficiency
  • Stealthier P2P clients and algorithm
  • Migration to Tunneled overlays (IPV6)
  • Provider traffic management policies
  • Increased use of P2P encryption
  • Migration to other distribution alternatives: direct download and streaming video
  • Payload analysis also suggests encrypted P2P / other ports represent another 10-15% of uncategorized traffic

• So we don’t really know whether 20% decline is really that high?
2007: 52 ports contributed 60% of the traffic
2009: 25 ports contributed 60% of inter-domain traffic!
Discussion

• Significant evolution of provider interconnection strategies and resultant inter-domain traffic demands
• Rapid transition to a more densely interconnected and less hierarchical inter-domain Internet topology