The Impact and Implications of the Growth in Residential User-to-User Traffic

Kenjiro Cho (IIJ)
Kensuke Fukuda (NII)
Hiroshi Esaki (Univ. of Tokyo)
Akira Kato (Univ. of Tokyo)

about this talk

• extensive study on residentail broadband (RBB) traffic
  - comparison of heavy-hitters/other-users, fiber/DSL users
• results show impact of RBB to Internet usage/backbone traffic
  - research people should know
  - although each result may not be too surprising to experts
unprecedented traffic increase in backbone

- rapidly growing residential broadband access
  - low-cost high-speed services, especially in Korea and Japan
  - Japan is by far the highest in Fiber-To-The-Home (FTTH)
- traffic growth of the peak rate at major Japanese IXes
  - still keeps growth of 50% per year
  - how much is contributed by residential broadband traffic?

[Graph: Traffic growth at major Japanese IXes]

residential broadband subscribers in Japan

- 23.3 million broadband subscribers as of March 2006
  - 14.5 million for DSL, 3.3 million for CATV, 5.5 million for FTTH
- exponential increase of FTTH, expected to exceed DSL in 2007
  - 100Mbps bi-directional fiber access costs 40USD/month
  - significant impact to backbones

[Graph: RBB subscribers in Japan]
motivation

- concerns about rapid growth of RBB traffic
  - backbone technologies will not keep up with RBB traffic
  - ISPs cannot invest in backbone simply for low-profit RBB
- ISPs and policy makers need to understand the effects of RBB
  - although most ISPs internally measure their traffic
    - data are seldom made available to others
    - measurement methods and policies differ from ISP to ISP
- to identify the macro-level impact of RBB traffic on ISP backbones
  - a study group with 7 major Japanese ISPs and government
- our approach consists of 2 analyses
  - aggregated traffic analysis
    - based on aggregated SNMP data from 7 major ISPs
  - per-customer traffic analysis
    - based on Sampled NetFlow data from one of the ISPs

major findings in aggregated traffic data

- our data is considered to cover 42% of total Japanese traffic
  - total RBB traffic in Japan is estimated to be 468Gbps (2005/11)
- 70% of RBB traffic is constant, peak in the evening hours
- RBB traffic is much larger than office traffic, so backbone traffic is dominated by RBB traffic
- traffic volume exchanged via private peering is larger than volume exchanged via major IXes
- regional RBB traffic is roughly proportional to regional population
data collection across major ISPs

- focus on traffic crossing ISP boundaries (customer and external)
  - tools were developed to aggregate MRTG/RRDtool traffic logs
- only aggregated results published not to disclose individual ISP share
- challenges: mostly political or social, not technical

ISP

5 traffic groups at ISP customer and external boundaries

methodology for aggregated traffic analysis

- month-long traffic logs for the 5 traffic groups with 2-hour resolution
  - MRTG’s resolution for monthly log
- a script to read and aggregate a list of MRTG/RRDtool logs
  - each ISP creates log lists and makes aggregated logs by themselves without disclosing details
- biggest workload for ISP
  - creating lists by classifying large number of per-interface logs
    - some ISPs have more than 100,000 logs!
  - maintaining the lists
    - frequent planned and unplanned configuration changes
- data sets
  - 2-hour resolution interface counter logs
    - by re-aggregating logs provided by 7 ISPs
  - IN/OUT from ISPs’ view
traffic growth

• 26-138% increase in 2005
  - RBB: 26% increase for inbound, 46% increase for outbound
• growth has slowed down from 100% in 2002 to 50% in 2005
  - observed worldwide

RBB customer weekly traffic
  in November 2005

• DSL/CATV/FTTH customer traffic of the 7 ISPs
  - inbound and outbound are almost equal
  - almost 200Gbps on average!
  - 150Gbps is constant, probably due to automated p2p applications
  - daily fluctuations: peak from 21:00 to 23:00
comparing RBB in-volumes between 2004 and 2005

- the growth comes from the constant portion!

weekly external traffic

- external traffic is also strongly affected by RBB traffic
  - other-domestic: mainly private peering (also transit, regional IXes)
    - larger than traffic via major IXes
  - international: inbound much larger than outbound
    - traditional content downloading seems still non-negligible

External weekly traffic in November 2005
prefectural differences in RBB traffic

- similar temporal traffic pattern across different prefectures
  - e.g., peak in evening, 70% is constant, regardless the volume
  - metropolitan prefectures with larger office hour traffic

Example prefectural traffic

- prefecture traffic
  - a metropolitan prefecture
  - a rural prefecture

prefectural population and traffic

- traffic is roughly linear to population!
  - from a scatter plot of population and traffic volume
  - similar result with the number of Internet users
  - no clear difference in usage or heavy-hitter ratio

Prefectural traffic volumes are roughly linear to populations
analysis of per-customer traffic in one ISP

- one ISP provided per-customer traffic data for Feb and Jul 2005

- data sets
  - Sampled NetFlow data
  - from edge routers accommodating fiber/DSL RBB customers
  - week-long logs from Feb and Jul 2005

- heavy-hitters: denote users who upload more than 2.5GB/day
  - larger in fiber users

major findings in per-customer traffic data

- 4% of heavy-hitters account for 75% of the total inbound volume
- the fiber users account for 86% of the inbound volume
  - DSL is only 14%
  - even though the number of DSL active users is larger than fiber
- the distribution of heavy-hitters is heavy-tailed
  - no clear boundary between heavy-hitters and normal users
- dominant applications have poor locality and communicate with a wide range and number of peers
CCDF of daily traffic per user

- heavy-hitters are statistically distributed
  - over a wide range of traffic volume (heavy-tailed)
  - even up to 200GB/day (19Mbps)!
- no clear boundary between heavy-hitters and normal users
- lines at 2.5GB/day (230kbps) and the top 4% heavy-hitters
  - knee of the total users’s slope
- heavy-hitter population: 4% in total users, 10% in fiber, 2% in DSL

[prefectural comparison

- distribution similar in all prefectures
  - differences in tail length (population size)
  - probably due to universal broadband access in Japan

CCDF of daily traffic per user

CCDF of daily traffic volume per user
CDF of traffic volume of heavy-hitters

- graph: the top N% of heavy-hitters use X% of the total traffic
- highly skewed distribution in traffic usage
  - the top 4% use 75% of the total inbound traffic
  - the top 4% use 60% of the total outbound traffic

Correlation of inbound/outbound volumes per user

- 2 clusters: one below the unity line, another in high volume region
  - more heavy-hitters in fiber, more lightweight users in DSL
- no qualitative difference between fiber users and DSL users
  - except the percentage of heavy-hitters
- again, no clear boundary between heavy-hitters and normal users
**number of active users**

- numbers are normalized to the fiber/DSL combined peak
- total numbers are similar between fiber and DSL
- heavy-hitters are fairly constant, especially in DSL

![Normalized number of active users](image1)

**comparison of fiber/DSL traffic**

- again, normalized to the combined peak
- inbound: 86% is from fiber users, DSL is only 14%
- total traffic is heavily influenced by fiber heavy-hitters

![Comparison of fiber/DSL traffic](image2)
uploading behavior of top 10 heavy-hitters

- one hour average traffic over a week
  - considerable variations, suggesting differences in usage

![Graph showing traffic over time and ranking]

protocols/ports ranking

- port 80 (http) is only 9%
- 83% is TCP dynamic ports!
  - each port usage is small except port 80

<table>
<thead>
<tr>
<th>protocol</th>
<th>port name</th>
<th>(%)</th>
<th>protocol</th>
<th>port name</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt; 1024</td>
<td>13.99</td>
<td>84</td>
<td>-</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>http</td>
<td>9.32</td>
<td>25</td>
<td>smtp</td>
<td>0.14</td>
</tr>
<tr>
<td>20</td>
<td>ftp-data</td>
<td>0.93</td>
<td>119</td>
<td>nntp</td>
<td>0.13</td>
</tr>
<tr>
<td>554</td>
<td>rtp</td>
<td>0.38</td>
<td>21</td>
<td>ftp</td>
<td>0.11</td>
</tr>
<tr>
<td>443</td>
<td>http</td>
<td>0.30</td>
<td>22</td>
<td>ash</td>
<td>0.09</td>
</tr>
<tr>
<td>110</td>
<td>pop3</td>
<td>0.17</td>
<td>others</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td>(&gt;= 1024</td>
<td>13.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6099</td>
<td>winmx</td>
<td>1.40</td>
<td>1755</td>
<td>ms-streaming</td>
<td>0.20</td>
</tr>
<tr>
<td>6346</td>
<td>gnutella</td>
<td>0.92</td>
<td>2265</td>
<td>-</td>
<td>0.13</td>
</tr>
<tr>
<td>7743</td>
<td>winny</td>
<td>0.48</td>
<td>1234</td>
<td>-</td>
<td>0.12</td>
</tr>
<tr>
<td>6881</td>
<td>bittorrent</td>
<td>0.25</td>
<td>4662</td>
<td>edonkey</td>
<td>0.12</td>
</tr>
<tr>
<td>6348</td>
<td>gnutella</td>
<td>0.21</td>
<td>others</td>
<td>79.41</td>
<td></td>
</tr>
<tr>
<td>UDP</td>
<td></td>
<td>1.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6346</td>
<td>gnutella</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESP</td>
<td></td>
<td>1.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRE</td>
<td></td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICMP</td>
<td></td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>others</td>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
geographic traffic matrix of RBB traffic

- RBB (home users), DOM (other domestic), INTL (international)
  - both ends are classified by commercial geo-IP databases
- 62% of residential traffic is user-to-user
- 90% is inside Japan (among RBB and DOM)
  - possible reasons are:
    - language and cultural barriers
    - p2p super-nodes among bandwidth-rich domestic fiber users

<table>
<thead>
<tr>
<th>src\dst</th>
<th>ALL</th>
<th>RBB</th>
<th>DOM</th>
<th>INTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>100.0</td>
<td>84.8</td>
<td>11.1</td>
<td>4.1</td>
</tr>
<tr>
<td>RBB</td>
<td>77.0</td>
<td>62.2</td>
<td>9.8</td>
<td>3.9</td>
</tr>
<tr>
<td>DOM</td>
<td>18.0</td>
<td>16.7</td>
<td>1.1</td>
<td>0.2</td>
</tr>
<tr>
<td>INTL</td>
<td>5.0</td>
<td>4.8</td>
<td>0.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

prefectural traffic matrix
(src on Y-axis, dst on X-axis)

- looking into 47 prefectures
  - traffic volumes are roughly linear to prefectural populations
prefectural traffic matrix normalized to src

- the sum of columns is 100% for each row
- no clear difference among prefectures
  - similar distribution, only small locality (1-3%) is found
  - similar result when normalized to dst

implications

- we tend to attribute the skews in usage to the divide between a handful of heavy-hitters and the rest of the users
  - but there are diverse and widespread heavy-hitters
- heavy-hitters are no longer exceptional extremes
  - too many of them, statistically distributed over a wide range
    - casual users start playing with p2p applications, become heavy-hitters, and eventually shift from DŚL to fiber
    - or, sometimes users subscribe to fiber first, and then, look for applications to use the abundant bandwidth
  - these users’ behavior would be easily affected by social, economic or political factors (they don’t care about underlying technologies)
  - but surely users as a whole are shifting towards high-volume usage
- is this specific to Japan?
  - a model of widespread symmetric residential broadband access
    - with language/cultural barriers, geographic concentration
**conclusion**

- we need to prepare for the future to accommodate innovations brought by empowered end-users
- our study to understand residential broadband traffic
  - cooperation with major ISPs and government
  - detailed analysis of traffic data from one ISP
- RBB traffic accounts for 2/3 of ISP backbone traffic
  - a significant impact on pricing and cost structures of ISP business
- future work
  - we will continue collecting aggregated traffic logs from ISPs
  - plans to compare results with other Japanese ISPs, other countries

**acknowledgments**
- IIJ, Japan Telecom, K-Opticom, KDDI, NTT Communications, POWEREDCOM, SOFTBANK BB for data collection support
- Ministry of Internal Affairs and Communications for coordination