The Impact of Residential Broadband Traffic on Japanese ISP Backbones

Kensuke Fukuda
NTT/WIDE
fukuda@t.ecl.net

Kenjiro Cho
IIJ/WIDE
kjc@iijlab.net

Hiroshi Esaki
U.Tokyo/WIDE
hiroshi@wide.ad.jp

ABSTRACT

In this paper, we investigate the effects of the rapidly-growing residential broadband traffic on commercial ISP backbone networks. We collected month-long aggregated traffic logs for different traffic groups from seven major ISPs in Japan in order to analyze the macro-level impact of residential broadband traffic.

Our results show that (1) the aggregated residential broadband customer traffic in our data exceeds 100Gbps on average. Our data is considered to cover about 40% of the total Japanese customer traffic so that we estimate the current residential broadband traffic in Japan is about 250Gbps in total. (2) About 70% of the residential broadband traffic is constant all the time. The rest of the traffic has a daily fluctuation pattern with the peak in the evening hours. The behavior of residential broadband traffic has considerable deviations from academic or office traffic. (3) The total traffic volume of the residential users is much higher than the office users so that the backbone traffic is dominated by the behavior of the residential user traffic. (4) The traffic volume exchanged through domestic private peering is comparable with the volume exchanged through the major IXes. (5) Within external tarffic of ISPs, international traffic is about 23% for inbound and about 17% for outbound. (6) The distribution of the regional broadband traffic is roughly proportional to the regional population.

We expect other countries will experience similar traffic patterns as residential broadband access becomes widespread.

1. INTRODUCTION

The availability of residential broadband access has made tremendous advances over the past few years, especially in Korea and Japan where both the penetration rate and the average line speed are much higher than other countries. A government survey shows that there are 14.5 million broadband subscribers in Japan as of February 2004; 11 million DSL subscribers, 2.5 million CATV Internet subscribers, and 1 million FTTH subscribers [1]. The number of broadband access subscribers is still increasing as shown in Figure 1 [1]. At the same time, broadband access technologies are shifting to higher speed such as 50Mbps DSL and 100Mbps FTTH.

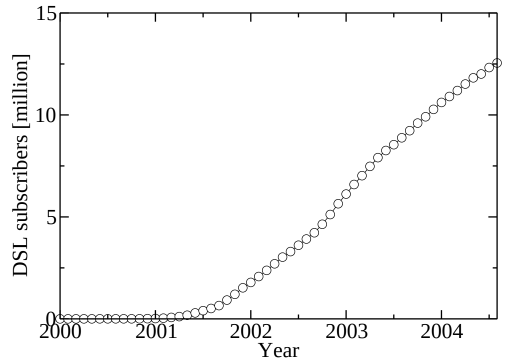

Figure 1: Increase of DSL subscribers in Japan

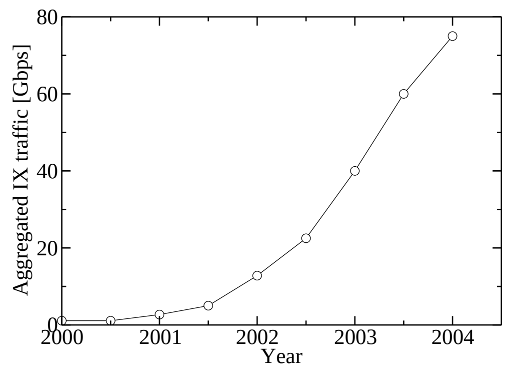

Figure 2: Traffic growth at the major Japanese IXes

As residential broadband access becomes widespread, we are observing an unprecedented traffic increase on commercial backbone networks. Figure 2 shows the aggregated traffic at major IXes (JPNAP[7], JPIX[6], and NSPIXP[9]) in Japan and illustrates the exponential growth in backbone traffic [1]. The impact of residential broadband traffic is not only in volume but also in usage patterns. The peak hours has shifted from office ours to evening hours, and emerging file sharing or other peer-to-peer communications with audio/video contents exhibit behavior considerably different from traditional world wide web [?, ?]. There are striking differences in traffic patterns from earlier observations [?, 4, 8, 3, 5].

Although a drastic change in backbone traffic has already been observed, it is difficult to plan for the future because residential broadband traffic is undergoing a transformation; access network technologies are under continual innovation, new applications and their usage are emerging to take advantage of low-cost high-speed connectivity.

There is a strong concern that, if this trend continues, Internet backbone technologies will not be able to catch up with the rapidly growing residential traffic. Moreover, commercial ISPs will not be able to invest in backbone networks simply for low-profit residential traffic.

It is critical to ISPs and policy makers to understand the effects of growing residential broadband traffic but it is difficult both technically and politically to obtain traffic data from commercial ISPs. Most ISPs are collecting traffic information for their internal use but such data contains sensitive information and is seldom made available to others. In addition, measurement methods and policies differ from ISP to ISP so that it is in general not possible to compare a data set with another set obtained from different ISPs.

In order to seek out a practical way to investigate the impact of residential broadband traffic on commercial backbone networks, we have formed an unofficial study group with specialists including members from 7 major commercial ISPs in Japan.

Our goal is to identify the macro-level impact of residential broadband traffic on ISP backbones. More specifically, we are trying to obtain a clearer grasp of the ratio of residential broadband traffic to other traffic, changes in traffic patterns, and regional differences across different ISPs. As the first step, we have collected aggregated bandwidth usage logs for different traffic groups. Such statistics will provide reference points for further detailed analysis, most likely by sampling methods. In this paper, we report findings in our data sets that residential broadband traffic presents a significant impact on ISP backbones.

2. METHODOLOGY

There are several requirements in order to solicit ISPs to provide traffic information. We need to find a common data set which all the participating ISPs are able to provide. The workload for ISPs to provide the data set should not be high. The data set should be coarse enough not to reveal sensitive information of the ISP but is meaningful enough so that the behavior of residential broadband traffic can be analyzed. It is also desirable to be able to cross-check the consistency of the results with other data sets. The data sets should be summable in order to aggregate them with those provided by other ISPs.

We found that most ISPs collect interface counter values of almost all routers in their service networks via SNMP, and archive per-interface traffic logs using MRTG [11] or RRDtool [10]. Thus, it is possible for the ISPs to provide aggregated traffic information if they can classify router interfaces into a common set.

Our focus is on traffic crossing ISP boundaries which can be roughly divided into customer traffic, and external traffic such as peering and transit. We selected the 5 traffic groups shown in Table 1 for data collection for practical purposes. It is impossible to draw a strict line for grouping (e.g. residential/business and domestic/international) on the global Internet so that these groups are chosen by the existing operational practice of the participating ISPs. We re-aggregate each ISPs' aggregated logs, and only the resulted aggregated traffic is used in our study so as to not reveal a share of each ISP.

Our main focus is on **(A1) RBB (residential broadband) customers** but other items are used to understand the relative volume of (A1) with respect to other types of traffic as well as to cross-check the correctness of the results. **(A2) non-RBB customers** is used to obtain the ratio of the residential broadband traffic to the total customer traffic. The total customer traffic (A) is $(A) = (A1) + (A2)$. **(B1) external 6IXes** and **(B2) external domestic** are used to estimate the coverage of the collected data sets. **(B3) external international** is used to compare domestic traffic with international traffic. The total external traffic (B) is $(B) = (B1) + (B2) + (B3)$. **(C) prefectural** is to observe regional differences. This group covers only 2 major residential broadband carriers who provide aggregated links per prefecture to ISPs; other carriers' links are not based on prefectures. This group is a subset of (A1).

In general, it is meaningless to simply sum up traffic values from multiple ISPs since a packet could cross ISP boundaries multiple times. Customer traffic is, however, summable because a packet crosses customer edges only once in each direction, when entering the source ISP and exiting the destination ISP. The numbers for external traffic are overestimated since a packet could be counted multiple times if it travels across more than 2 ISPs. However, the error should be relatively small in this particular result since these ISPs are peering with each other.

We collected month-long traffic data that was sampled every two hours from the participating ISPs because a 2 hour resolution is the highest common factor for month-long data. This is because both MRTG and RRDtool aggregate old records into coarser records in order to preserve the database size. In MRTG, 2-hour resolution records are maintained for 31 days in order to draw monthly graphs. RRDtool does not have fixed aggregation intervals but most operators configure RRDtool to maintain 1-hour or 2-hour resolution records for longer than needed for monthly graphs.

We developed a perl script to read a list of MRTG and RRDtool log files, and aggregate traffic for a give period with a given resolution. It outputs "timestamp, in-rate, out-rate" for each time step. Another script makes the output into a graph using RRDtool. We provided the tools to the ISPs so that each ISP can create aggregated logs by themselves. It allows ISPs not to disclose the internal structure of their networks and their detailed traffic information.

The biggest workload for the ISPs is to classify the large number of per-interface traffic logs and create a log list for each group. For large ISPs, the total number of the existing per-interface traffic logs exceeds 100,000. To reduce the workload, ISPs are allowed to use the internal interface of a border router instead of a set of external (edge) interfaces if the traffic on the internal interface can be an approximation of the sum of the external interfaces. In such a case, we need to instruct the tool to swap "in" and "out" records since in/out records in the per-interface logs denote the routers' point of view but we need in/out records to signify the ISPs' point of view.

Table 1: Traffic groups for data collection

traffic group	description	notes
(A1) RBB customers	residential broadband customer lines	includes small business customers using RBB
(A2) non-RBB customers	includes leased lines, data centers, dialup lines	may include RBB customers behind leased lines
(B1) external 6IXes	links for 6 major IXes (JPNAP/JPIX/NSPIXP in Tokyo/Osaka)	
(B2) external domestic	external domestic links other than the 6IXes (regional IXes, private peering, transit)	domestic: both link-ends in Japan. includes domestic peering with global ASes
(B3) external international	external international links	
(C) prefectural	RBB links divided into 47 prefectures in Japan	prefectural links from 2 RBB carriers

3. RESULTS

We analyzed traffic logs for September and October in 2004 from 7 major ISPs in Japan. Each ISP provides traffic logs with 2-hour resolution for the months. The results are obtained by aggregating all the traffic logs provided by the 7 ISPs. Two-hour boundaries are computed in UTC by MRTG and RRDtool so that they fall on odd hours in Japanese Standard Time that is 9 hours ahead of UTC.

For weekly data analysis, we took the averages of the same weekdays in the month. We excluded 2 holidays in September and 1 holiday in October for weekly analysis since their traffic pattern is closer to that of weekends. We also excluded another 2 days in October for weekly analysis as one ISP failed to record traffic logs during this period.

3.1 Customer Traffic

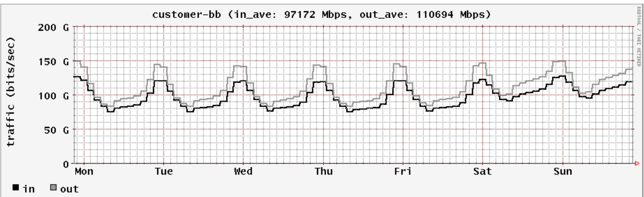

Figure 3: Aggregated RBB customer weekly traffic in September. Darker vertical dotted lines indicate the start of the day (0:00 am in localtime).

Figure 3 shows the weekly traffic of RBB (residential broadband) customers, consisting of DSL/FTTH/CATV residential users (A1). This group also includes small business customers using residential broadband access. It is apparent that the residential broadband customer traffic has already exceeded 100Gbps in total. Note that the plot is the mean rate not the peak rate, although the peak rate is often used for operational purposes. The inbound and outbound traffic are almost equal, and about 70Gbps is constant for both directions, probably due to peer-to-peer applications which generate traffic independent of daily user activities. The fluctuation pattern indicates that home user traffic is dominant, i.e. the traffic increases in the evening and the peak hours are from 21:00 to 23:00. Weekends can be easily identified by larger daytime traffic though the peak rates are close to weekdays. The outbound traffic to customers is slightly larger than the inbound, even though it is often assumed that home users' downstream traffic is much larger than upstream. We believe that so-called peer-to-peer applications significantly contribute to the residential broadband traffic.

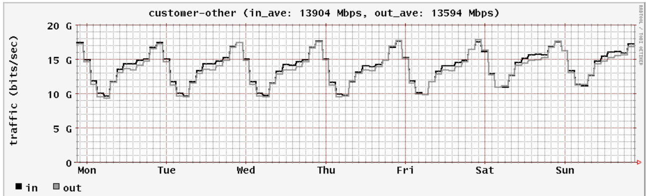

Figure 4: Aggregated non-RBB customer weekly traffic in September.

Figure 4 shows the weekly traffic of non-RBB customers (A2). This group contains leased lines, data centers, and other customers (e.g. dialup customers). It also includes leased lines used to accommodate residential broadband access within the customer networks (e.g. second or third level ISPs) since ISPs do not distinguish them from other leased lines. As a result, the traffic pattern looks still dominated by residential traffic, which is indicated by the peak hours and the differences between weekdays and weekends. However, we can also observe office hour traffic (from 8:00 to 18:00) in the daytime on weekdays but traditional office customer traffic has less impact to the total customer traffic. Note that we cannot directly compare the traffic volume of (A2) with that of (A1) because (A2) was provided by only 4 out of 7 ISPs.

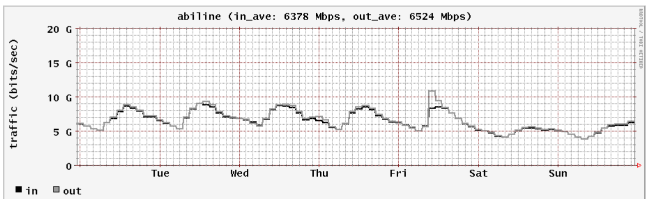

Figure 5: Aggregated total traffic from ABILENE. Time is in CDT.

The traffic patterns common to Figure 3 and 4 are quite different from academic or business usage patterns as widely known. For example, Figure 5 shows the weekly traffic of ABILENE [2] that is an Internet2 backbone network for universities and research labs. Figure 5 clearly presents that office hour traffic is dominant; the peak hour is around noon, and there is little user activities on weekend.

3.2 External Traffic

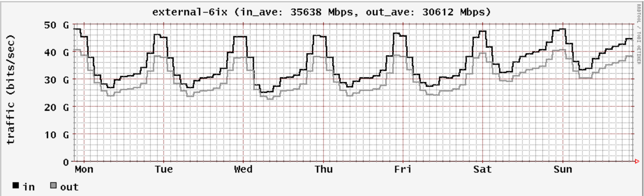

Figure 6: Weekly external traffic to/from the 6 major IXes in September.

The external traffic groups are used to understand the total traffic volume in Japan. Figure 6 shows traffic to and from the 6 major IXes (B1). It is apparent that the traffic behavior is strongly affected by residential traffic.

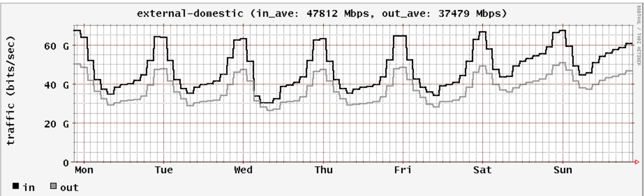

Figure 7: Weekly other domestic external traffic in September

Figure 7 shows the external domestic traffic (B2) including regional IXes, private peering and transit but not including traffic for the 6 major IXes. The traffic pattern is very similar to Figure 6.

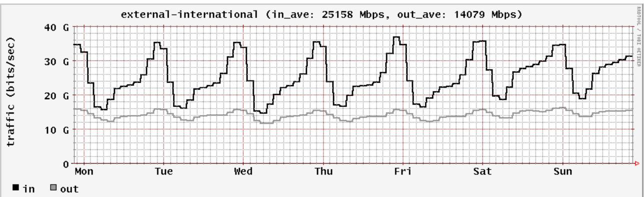

Figure 8: Weekly international external traffic in September

Figure 8 shows international traffic (B3). The inbound traffic is much larger than the outbound, and the traffic pattern is clearly different from the domestic traffic. The peak hours are still in the evening but the size of fluctuations for outbound is much smaller than that for inbound, suggesting that the traditional behavior of content downloading to Japan is still dominant.

3.3 Prefectural Traffic

In order to investigate regional differences (i.e. between metropolitan and rural areas), we collected regional traffic consisting of 47 prefectures. Figure 9 compares aggregated

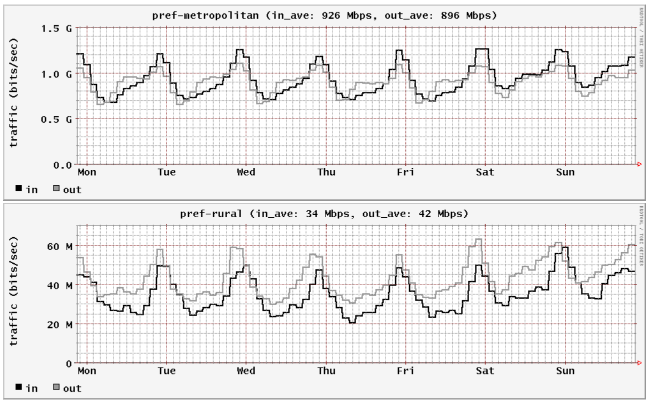

Figure 9: Example prefectural traffic: metropolitan area (upper) and rural area (lower)

traffic for a metropolitan prefecture (upper) with that for a rural prefecture (lower), as an example. It is clear that they are similar in temporal patterns such as peak positions and weekday/weekend behavior. In addition, about 70% of the average traffic is constant regardless of the traffic volume. These characteristics are common to other prefectures. One noticeable difference found in metropolitan prefecture traffic is that the metropolitan prefectures have larger office-hour traffic, probably due to business customers.

Figure 10 is a scatter plot of traffic and population for the 47 prefectures. We found that prefectural traffic is roughly proportional to the population of the prefecture. We obtained similar results when the number of the Internet users are used instead of populations. The result indicates that there is no clear regional concentration of heavy hitters of the Internet. That is, the probability of finding a heavy hitter in a given population is constant.

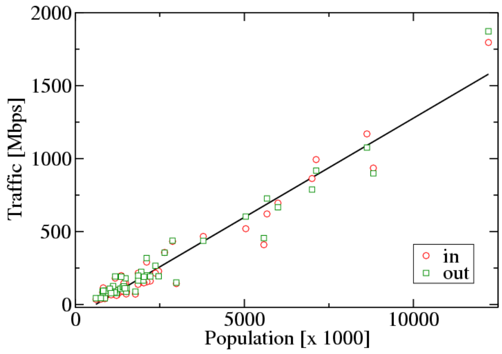

Figure 10: Relationship between population and traffic for prefectures

In order to analyze the scaling property of traffic volume —typical size of traffic volume—, we show the cumulative distribution of prefectural traffic in log-log plot in Figure 11. The plot conforms to a power law distribution with a cutoff point at 700Mbps, meaning that there is no characteristic scale of the traffic amount. In other words, most prefectures generate a small amount of traffic, still prefectures with a large amount of traffic are observable with a certain probability. It is also clear that the plots for the top 5 largest prefectures deviate from the power law. To investigate this power law decay, we show the cumulative distribution of populations in prefectures in the sub-panel. The plots reveal that the power law appeared in traffic volume is also observed in the population graph, as can be inferred from the linear relationship between traffic and populations in Figure 10. Thus, we can conclude that the probability of finding a heavy hitter in a given population is constant and the distribution of aggregated traffic volume directly depends on the population.

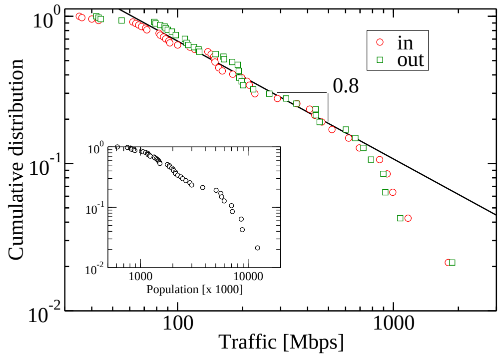

Figure 11: Cumulative distribution of prefectural traffic. Sub-panel indicates the cumulative distribution of populations for comparison.

3.4 Summary of Traffic

The monthly average rates in bits/sec of the traffic groups are shown in Tables 2 through 5.

Table 2: Average rates of aggregated customer traffic

	(A1)customer-RBB (7 ISPs)		(A2)customer-non-RBB (4 ISPs)	
	in	out	in	out
Sep	98.1G	111.8G	14.0G	13.6G
Oct	108.3G	124.9G	15.0G	14.9G

Table 2 is the average rates of aggregated customer traffic. As explained before, the non-RBB customer traffic was obtained only from the 4 ISPs so that it is difficult to directly compare (A1) and (A2). Thus, we estimated the ratio of the RBB customer traffic (A1) to the total customer traffic (A) only from the 4 ISPs' data with both (A1) and (A2). The estimated ratio (A1)/(A1+A2) is 65% for inbound and 67% for outbound.

Table 3: Average rates of aggregated external traffic

	(B1)ext-6ix (7 ISPs)		(B2)ext-dom (7 ISPs)		(B3)ext-intl (7 ISPs)	
	in	out	in	out	in	out
Sep	35.9G	30.9G	48.2G	37.8G	25.3G	14.1G
Oct	36.3G	31.8G	53.1G	41.6G	27.7G	15.4G

Table 3 summarizes the average rates of aggregated external traffic. We observe that the total volume of external domestic traffic (B2) exceeds the volume for the 6 major IXes (B1). This result points out that just looking at IXes' data might be misleading to estimate and understand nation-wide traffic, because a considerable amount of traffic is exchanged by private peering. Also, the ratio of international traffic to the total external traffic is 23% for inbound and 17% for outbound.

Table 4: Average rates of total customer traffic and total external traffic

	(A)customer(A1+A2)		(B)external(B1+B2+B3)	
	in	out	in	out
Sep	112.1G	125.4G	109.4G	82.8G
Oct	123.3G	139.8G	117.1G	88.8G

There is a relationship between the total customer traffic (A1 + A2) and the total external traffic (B1 + B2 + B3) in Table. 4. If we assume all inbound traffic from other ISPs is destined to customers, the inbound traffic volume for the total external tarffic (B) should be close to the outbound traffic volume for the total customer traffic (A), and vice versa. However, the non-RBB customer data is provided by only 4 ISPs. If we interpolate the missing ISPs in the non-RBB customer traffic using the ratio from the other 4 ISPs, the total customer traffic for inbound is estimated at 152.1Gbps, and that for outbound is at 167.8Gbps. Though these volumes are higher than those for the total external traffic volume, this is probably because the total customer traffic contains traffic whose source and destination belong to the same ISP.

Table 5: IX traffic observed from ISPs and from IXes

	(B1)ext-6ix	traffic observed by IXes
	out	in
Sep	30.9G	74.5G
Oct	31.8G	77.1G

Finally, we explain a relationship between our IX traffic data (B1) and the total input rate of the 6 major IXes obtained directly from these IXes [1]. In comparison with the published total incoming traffic of these IXes, our data occupies about 40% of the total traffic as shown in Table 5. If we assume this ratio as the traffic share of the 7 ISPs, the total amount of residential broadband traffic in Japan is roughly estimated at 250Gbps.

3.5 Distribution of per-user traffic

In order to verify our assumption that the distribution of heavy hitters is similar in different regions, we obtained per-cusomer traffic distribution for October 2004 from one of the participating ISP. The data of residential broadband customers for each prefecture was collected by means of sampled NetFlow [?] and matching customer IDs with the assigned IP addresses.

Figure 12 shows the cumulative distribution of daily traffic per user in a log-log plot, and compares all the prefectures (left) with a metropolitan prefecture (middle) and a rural prefecture (right). The daily traffic usage is the average of the month, and the distribution is computed independently for inbound and outbound. About 4% of the customers use more than 2.5GB/day (or 230kbits/sec) and, beyond this point, the slope of the distribution changes. The distribution

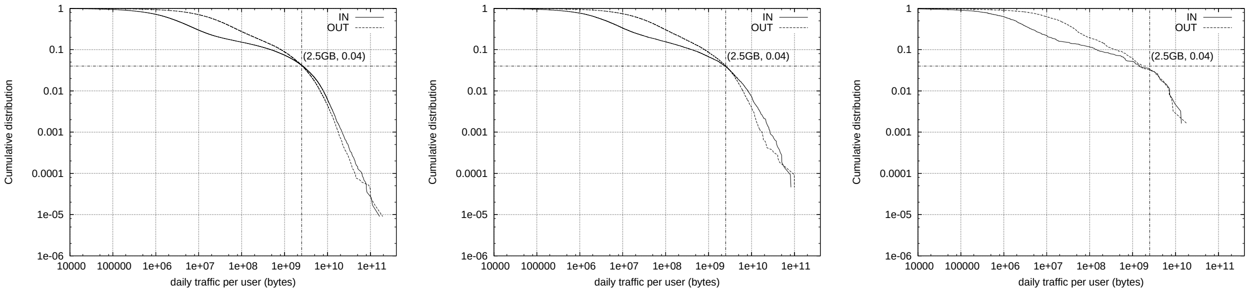

Figure 12: Cumulative distribution of daily traffic per user: all prefectures (left), a metropolitan prefecture (middle) and a rural prefecture (right)

also shows that outbound traffic is dominant for most customers but it becomes lower than inbound traffic for heavy hitters. These trends are consistent among different perfectures as shown in Figure 12 and the differences are only in the tail length due to the number of customers, which confirms that the distribution of heavy hitters is similar in different regions.

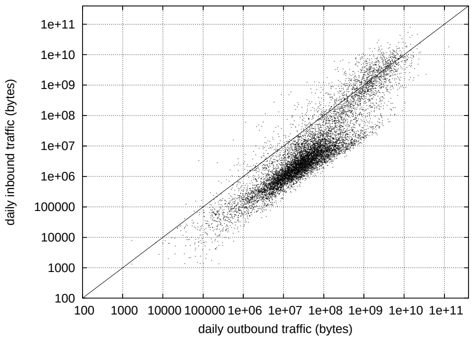

Figure 13: daily inbound and outbound traffic volume for each customer in a metropolitan prefecture

Figure 13 shows daily inbound and outbound traffic volume for each customer in a metropolitan prefecture.

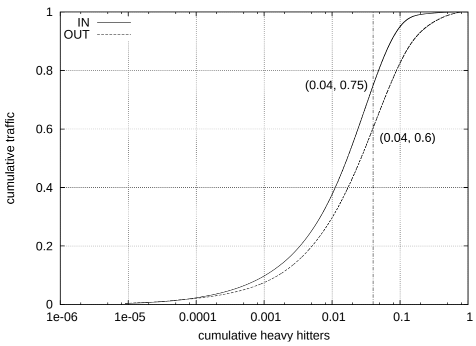

Figure 14: Cumulative distribution of traffic volume with heavy hitters in decreasing order of volume

Figure 14 shows the cumulative distribution of traffic volume with heavy hitters in decreasing order of volume for all the prefectures. It reveals that top N% heavy hitters use X% of the total traffic. Again, the distribution is computed independently for inbound and outbound. Top 4% customers use 75% of the total traffic for inbound, and 60% for outbound.

4. CONCLUSION

The widespread deployment of residential broadband access has tremendous implications to our lives. Although its effects to the Internet infrastructure are difficult to predict, it is essential for ISPs to prepare for the future to accommodate innovations brought by empowered end-users.

Residential broadband traffic has already contributed to a significant increase in commercial backbone traffic. In our study, residential broadband traffic accounts for more than 50% of the ISP backbone traffic, which would have a significant impact to pricing and cost structure of ISP business.

The properties of residential broadband traffic differ considerably from academic or office traffic often seen in literature. The constant portion in daily traffic fluctuations is about 70%, much larger than ones found in earlier reports [3, 5]. Research results obtained from campus or other academic networks may not apply to commercial traffic any more. More research efforts should be directed to measurement and analysis of residential broadband traffic.

The in/out rates are roughly symmetric throughout our data sets. Many access technologies employ asymmetric line speed for inbound and outbound based on the assumption that content-downloading is dominant for normal users. However, this assumption does not hold in our measurements.

The prefectural results show that traffic volume is roughly proportional to regional population. It indicates a unique characteristic of the cyber world in which activities are not bound by time and place. If this is the case, it would affect the design of capacity planning for the future Internet.

For future work, we will continue collecting aggregated traffic logs from ISPs. We are also planning to do more detailed analysis of residential broadband traffic by selecting a few sampling points.

Acknowledgments

We are grateful to the following ISPs for their support and assistance with our data collection; IIJ, Japan Telecom, K-Opticom, KDDI, NTT Communications, POWEREDCOM, and SOFTBANK BB. We would like to thank Ministry of

Internal Affairs and Communications, Japan for their support in coordinating our study. We would also like to thank Akira Kato of the University of Tokyo and anonymous reviewers for valuable inputs.

5. REFERENCES

[1] A study group report on the next generation IP-based infrastructure (in Japanese). , Ministry of Internal Affairs and Communications, Japan, June 2004.

[2] ABILENE aggregate traffic statistics. http://abilene.internet2.edu/observatory/summary-view.html.

[3] A. Feldmann, A. G. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True. Deriving traffic demands for operational IP networks: methodology and experience. In *SIGCOMM*, pages 257–270, 2000.

[4] M. Fomenkov, K. Keys, D. Moore, and k claffy. Longitudinal study of Internet traffic in 1998-2003. In *WISICT*, Cancun, Mexico, Jan. 2004.

[5] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot. Packet-level traffic measurements from the sprint IP backbone. *IEEE Network*, 2003.

[6] Japan Internet Exchange Co., Ltd. (JPIX). http://www.jpix.co.jp.

[7] Multifeed JPNAP service. http://www.jpnap.net.

[8] S. McCreary and k. claffy. Trends in wide area IP traffic patterns. In *ITC Specialist Seminar*, Monterey, CA, Sept. 2000.

[9] NSPIXP: Network Service Provider Internet eXchange Project. http://nspixp.wide.ad.jp.

[10] T. Oetiker. RRDtool: Round Robin Database Tool. http://ee-staff.ethz.ch/~oetiker/webtools/rrdtool/.

[11] T. Oetiker. MRTG: The multi router traffic grapher. In *USENIX LISA Conference*, pages 141–147, Boston, MA, Dec. 1998.