Observing Slow Crustal Movement in Residential User Traffic

(PaperID:1569138435 12pages)

Kenjiro Cho
IIJ

Kensuke Fukuda
NII

Hiroshi Esaki
Univ. of Tokyo

Akira Kato
Keio Univ.

ABSTRACT

It is often argued that rapidly increasing video content along with the penetration of high-speed access is leading to explosive growth in the Internet traffic. Contrary to this popular claim, technically solid reports show only modest traffic growth worldwide. This paper sheds light on the causes of the apparently slow growth trends by analyzing commercial residential traffic in Japan where the fiber access rate is much higher than other countries. We first report that Japanese residential traffic also has modest growth rates using aggregated measurements from six ISPs. Then, we investigate residential per-customer traffic in one ISP by comparing traffic in 2005 and 2008, before and after the advent of YouTube and other similar services. Although at first glance a small segment of peer-to-peer users still dictate the overall volume, they are slightly decreasing in population share. Meanwhile, the rest of the users are steadily moving towards rich media content with increased diversity. Surely, a huge amount of online data and abundant headroom in access capacity can conceivably lead to a massive traffic growth at some point in the future. The observed trends, however, suggest that video content is unlikely to disastrously overflow the Internet, at least not anytime soon.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network monitoring*

General Terms

Measurement, Management

Keywords

traffic growth, ISP traffic, residential broadband

1. INTRODUCTION

For the last few years, video content becomes increasingly popular among Internet users. Today's Internet users casually view and share video content, benefited from the penetration of broadband access and easy-to-use video services on the Net. On the other hand, video and other rich media content are by orders of magnitude larger in amount than traditional web content. Thus, it is often projected that rapidly increasing video content is leading to explosive growth in the Internet traffic.

Despite of this popular claim, technical sources report only modest traffic growth worldwide. Odlyzko estimates the Internet traffic growth rate in 2007 to be about 50% to 60% in the U.S. and worldwide[19]. Cisco's recent report also estimates that worldwide Internet traffic growth has been around 50% per year over the last few years[6].

The growth of the Internet traffic volume is one of the key factors driving research, development and investment in data communication technologies and infrastructure. With the annual growth rate of 100%, it grows 1000-fold in 10 years, while with 50%, it grows only 58-fold. Hence, a difference in growth rate has a considerable impact in the long run. In order to accommodate innovations brought by empowered end users, crucial is not just growth rate but the balance among demands, technological advances and investment in infrastructure. If the growth is underestimated, we may not have capacity enough to handle new demands. If overestimated, investment in technologies and capacity may be only wasted.

We have been studying commercial backbone traffic in Japan with support from six ISPs covering 42% of the Japanese backbone traffic. (We started with seven ISPs but two were merged in 2006.) In our previous work[3], we reported that the backbone in 2005 was dominated by residential user-to-user traffic.

Japan is one of the highest fiber access penetration countries in the world[20], and the number of Fiber-To-The-Home (FTTH) subscribers is about to exceed that of DSL subscribers as shown in Figure 1[29]. Meanwhile, the total number of broadband subscribers has reached 56% of households and increased only 5% in 2007[29]. However, even with the abundant access capacity, the traffic growth rate at Japanese major IXes has remained under 50% for the last few years as depicted in Figure 2. The plot shows aggregated peak traffic at major IXes, JPIX[15], JPNAP[16] and NSPIXP[18], and illustrates multiplicative growth of traffic where the

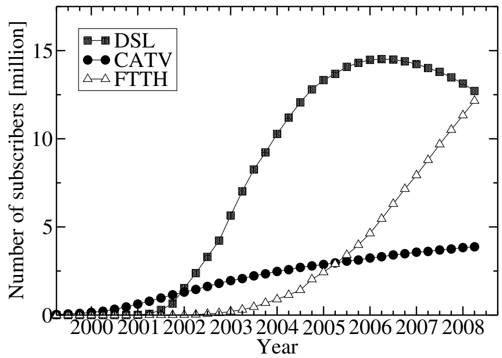

Figure 1: Number of residential broadband subscribers in Japan: 28.7 million total broadband subscribers, 12.7 million for DSL, 12.2 million for FTTH and 3.3 million for CATV as of March 2008.

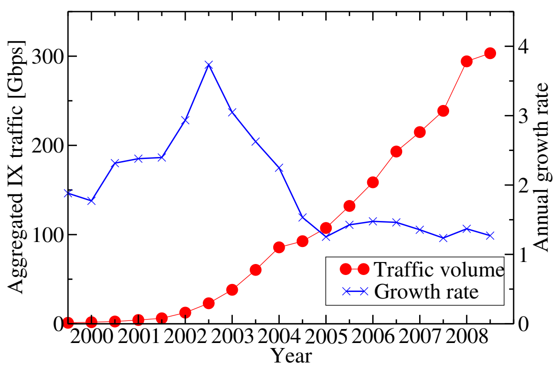

Figure 2: Traffic growth of the aggregated peak rate at the major Japanese IXes

volume is doubling roughly every two years since 2005.

In this paper, we try to answer a key question: what is the macro level impact of video and other rich media content on the traffic growth at the moment? We shed light on the causes of the apparently slow growth trend by detailed analyses of commercial residential traffic.

We followed the methodology used in our previous work[3], and used two different data sets. The first set was collected by aggregating interface counters of edge routers from the six ISPs. The other set was collected by Sampled NetFlow[4] from one of the ISPs for detailed per-customer analysis.

We first report that Japanese residential traffic also has a modest average annual growth rate of 27% over the last three years by means of aggregated measurements from the six ISPs. Then, we further investigate residential per-customer traffic in one of the ISPs by comparing traffic in 2005 and 2008 that is before and after the advent of YouTube and other similar video services. Although at first glance a small segment of peer-to-peer users still dictate the overall traffic volume, it is decreasing in both population share and volume share. At the same time, the rest of the users are steadily moving towards rich media content with increased diversity.

The current traffic is heavily affected by an eruption of peer-to-peer applications but the crust underneath is also slowly rising with video and other rich media content. The crustal movement is slow at the macro level so that it is unlikely to cause a major quake in the near future.

2. RELATED WORK

This is a sequel of our previous work[3] in which we compared traffic usage between fiber and DSL users, and between heavy-hitters and normal users. The focus was on heavy-hitters in residential traffic but this paper focuses on ordinary residential users. One of the contributions of this paper is to illustrate the impacts of peer-to-peer and rich media traffic on the current traffic growth. Another contribution is to quantify changes in traffic usage of ordinary residential users.

The traffic growth rate is reported to be slowing down in recent years. Odlyzko monitors over 100 publicly available traffic statistics, and estimates the Internet traffic growth rate[19]: the U.S. Internet traffic grew at around 100% in early 1990, jumped up to 1,000% in 1995-1996 driven by world wide web and user expansion, went back to 70-150% in 1997-2003, and settled down to around 50% in 2004-2007. Cisco's reports[6, 5] are based on traffic data from its provider customers combined with analyst projections, and also estimates that worldwide Internet traffic growth has been around 50% per year over the last few years. Furthermore, Japanese IX traffic data as shown in Figure 2 clearly showed that the growth rate (around 30%) is stable (i.e., multiplicative growth) over the last four years. These estimates do not have details on user distributions so that our detailed analyses are complementary to these reports.

There is little solid work in literature that tries to measure commercial residential traffic. Studies on ADSL networks[23, 27] are similar to our per-customer analysis in monitoring access lines and comparing traffic volumes among data sets but their focus is on file-sharing traffic. It is reported that the average daily traffic volume of heavy-hitters is 470MB for inbound and 760MB for outbound, and that of non heavy-hitters is 9MB for inbound and 27MB for outbound in 2006[27]. Another study on academic backbone traffic[14] reports that peer-to-peer traffic account for 86-93% of the total traffic volume.

There are numerous studies on traditional web traffic, peer-to-peer traffic and video traffic. Many of them try to characterize traffic at the flow level. Peer-to-peer traffic is highly variable and skewed among participating nodes[24, 26, 12], and exhibit behavior considerably different from traditional web traffic[11, 1]. A recent study on YouTube traffic[10] by monitoring an academic network compares characteristics of video content with traditional web content. The mean and median size of video content are 10MB and 8MB, and the mean and median transfer rate is 394kbps and 328kbps. Another study[17] characterizes streaming media files stored online by crawling web sites, and compares different types of video content. Many studies report that flow size and duration follow heavy-tailed distributions (e.g., [25, 31, 2]), and discuss elephants and mice in flows. Daily traffic volume per user is the summation of flow size transferred to a user but there exits few previous work focusing on the evolution of traffic volume per user, especially for residential users.

Regarding a shift in traffic mix, Cisco's report[6] has statistics about a traffic mix shift in consumer traffic; peer-to-peer file-sharing traffic grew 29% in volume but the volume share dropped from 60% in 2006 to 51% in 2007. The volume share of video content not including peer-to-peer file-sharing traffic grew from 12% in 2006 to 22% in 2007, and is projected to be 50% in 2012. The estimated peer-to-peer traffic share is smaller than our results. The projection is calculated by projected online hours per user multiplied by the average MB per hour for different traffic types. Our finding in traffic distributions could provide new insights into these traditional projection methods.

3. AGGREGATED TRAFFIC ANALYSIS

This section examines the traffic growth trends in Japan by means of aggregated measurements collected from the six ISPs covering 42% of the Japanese backbone traffic. The results show that Japanese commercial traffic has modest growth rates over the last three years. In particular, the average annual growth rate of residential traffic has been 27% from May 2005 through May 2008.

3.1 Measurement Methodology

The ISPs we collaborate with collect interface counter values of almost all routers in their service networks via SNMP, and archive per-interface traffic logs using MRTG [22] or RRDtool [21]. 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 provider traffic such as peering and transit. For practical purposes, we selected the five traffic groups in Figure 3 for data collection.

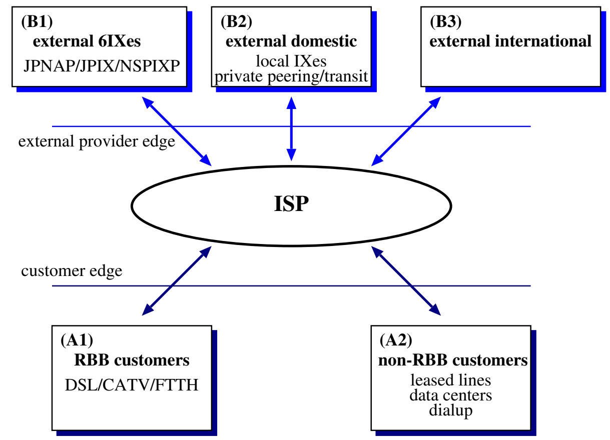

Figure 3: Five traffic groups for data collection at ISP customer and external boundaries

(A1) RBB customers represent residential broadband customer lines. This group also includes small business customers using residential broadband access.

(A2) non-RBB customers represent customer lines other than RBB customers, including leased lines, data centers, and dialup lines. This group includes RBB customers behind leased lines, e.g., second or third level ISPs, since ISPs do not distinguish them from other leased lines.

(B1) external 6IXes represent links for 6 major IXes, namely JPIX, JPNAP and NSPIXP in both Tokyo and Osaka in order to compare measurements at these IXes as well as to know the traffic share of our measurement.

(B2) external domestic represents domestic external provider links other than the 6IXes, including regional IXes, private peering and transit. We used the term "domestic" to indicate that both ends of a link are located in Japan. This group also includes domestic peering with global ASes.

(B3) external international represents international external provider links with one end point outside of Japan.

These groups are chosen by the existing operational practices of the participating ISPs because it is impossible to draw a strict line for grouping, e.g., residential/business and domestic/international, on the global Internet. We re-aggregate each ISP's aggregated logs, and only the resulting aggregated traffic is used in our study so as to not reveal the share of each ISP. Although the peak rate is often used for operational purposes, only the mean rate is collected since the peak rate is not summable.

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 provider traffic are overestimated since a packet could be counted multiple times if it travels across intermediate ISPs. Nevertheless, the error should be negligible in this particular result since most of the ISPs in our data sets are peering, and thus, not providing transit to each other. Although one ISP is identified to have double counts for the transit link, it is compensated in the results by data provided separately for the double count adjustment.

Each ISP provided month-long traffic logs aggregated for each traffic group by a log aggregation tool we developed. This allows ISPs not to disclose the internal structure of their network or unneeded details of their traffic. The final results were obtained by aggregating all traffic logs provided by the six ISPs.

The time resolution of the logs is 2 hours since it was the highest common factor for month-long data in MRTG and RRDtool. 2-hour boundaries in UTC fall on odd hours in Japanese Standard Time (UTC+9). Throughout the paper, *inbound* and *outbound* are presented from the ISPs' point of view.

The data for each month was separately collected, and consistency such as each ISP's share, differences from the previous measurements, the coverage of the IX traffic was examined. Then, the aggregated results were provided to the ISPs so that each ISP can compare and check its own data against the aggregated results. In addition, a face-to-face meeting with representatives from the participating ISPs is held after each data collection to check and discuss the results.

Monthly traffic logs with two-hour resolution allow us to identify major changes in each ISP's traffic. When such changes are found, we contact the ISP to confirm the cause of the change, e.g., a network reconfiguration, an outage, missing SNMP data, or a mis-classification of interface counter logs. Afterwards, if necessary, we ask the ISP for corrected data.

3.2 Data Sets

We analyzed month-long traffic logs from six major Japanese ISPs over five years; September, October, November in 2004, May and November from 2005 through 2008. After the initial trials over three months, we decided to collect data only twice a year to reduce the workload of the participating ISPs.

3.3 Growth of Traffic

The monthly average rates in bits/second of the traffic groups are shown in Tables 1 and Figure 4. RBB customers (A1) consist of DSL/FTTH/CATV residential users. Between May 2005 and May 2008, the average annual growth rate of (A1) was 26% for inbound, 28% for outbound, and 27% for the combined volume. The difference between inbound and outbound has widened over the years.

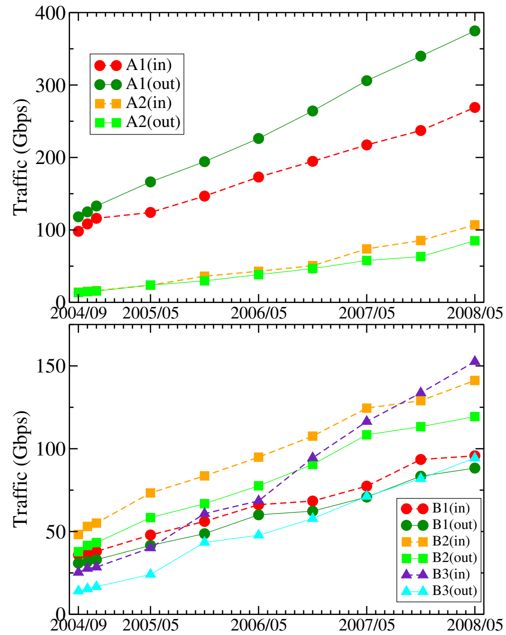

Figure 4: Traffic growth: customer traffic (upper) and external provider traffic (bottom)

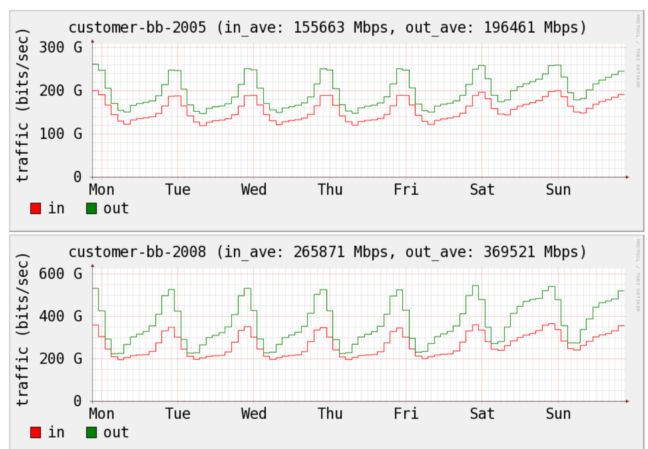

Figure 5: Residential broadband customer traffic in May 2005 (upper) and May 2008 (bottom)

Figure 5 compares weekly RBB customer traffic (A1) in May 2005 and in May 2008. For weekly data analysis, we took the averages of the same weekdays in a month. We excluded holidays from the weekly analysis since holiday traffic patterns are closer to those of weekends. The inbound and outbound traffic volumes were almost equal in 2005, and about 120Gbps is constantly flowing in both directions, probably due to peer-to-peer appli-

Table 1: Monthly average rates of aggregated customer and external traffic (bps)

		(A1)customer-RBB (6 ISPs)		(A2)customer-non-RBB (4 ISPs)		(B1)ext-6ix (6 ISPs)		(B2)ext-dom (6 ISPs)		(B3)ext-intl (6 ISPs)	
		inbound	outbound	inbound	outbound	inbound	outbound	inbound	outbound	inbound	outbound
2004	Sep	98.1G	111.8G	14.0G	13.6G	35.9G	30.9G	48.2G	37.8G	25.3G	14.1G
	Oct	108.3G	124.9G	15.0G	14.9G	36.3G	31.8G	53.1G	41.6G	27.7G	15.4G
	Nov	116.0G	133.0G	16.2G	15.6G	38.0G	33.0G	55.1G	43.3G	28.5G	16.7G
2005	May	134.5G	178.3G	23.7G	23.9G	47.9G	41.6G	73.3G	58.4G	40.1G	24.1G
	Nov	146.7G	194.2G	36.1G	29.7G	54.0G	48.1G	80.9G	68.1G	57.1G	39.8G
2006	May	173.0G	226.2G	42.9G	38.3G	66.2G	60.1G	94.9G	77.6G	68.5G	47.8G
	Nov	194.5G	264.2G	50.7G	46.7G	68.4G	62.3G	107.6G	90.5G	94.5G	57.8G
2007	May	217.3G	306.0G	73.8G	57.8G	77.4G	70.8G	124.5G	108.4G	116.4G	71.2G
	Nov	237.2G	339.8G	85.4G	63.2G	93.5G	83.4G	129.0G	113.3G	133.7G	81.8G
2008	May	269.0G	374.7G	107.0G	85.0G	95.7G	88.3G	141.2G	119.4G	152.6G	94.4G

cations which generate traffic independent of daily user activities. The ratio between the peak and bottom (i.e., constant flows) of the traffic volume decreased from 0.63 to 0.56 for inbound, and from 0.59 to 0.43 for outbound. The diurnal pattern reflects home user activities, i.e., the traffic increases in the evening, and the peak hours are from 21:00 to 23:00. Weekends can be identified by larger daytime traffic although the peak rates are similar to weekdays. The outbound traffic to customers is slightly larger than the inbound in 2005, and becomes much larger in 2008 with larger fluctuations and with clearer peaks. We believe that peer-to-peer applications contribute significantly to the upstream traffic from residential users.

The data for non-RBB customer traffic (A2) was obtained only from four ISPs; it is difficult for the other ISPs to distinguish external links from other links due to historical reasons. Since (A2) from these other ISPs is missing, it is not possible to directly compare (A1) with (A2). Hence, we estimated the ratio of (A1) to (A) using only data from the four ISPs with both (A1) and (A2). The estimated ratio (A1)/(A1+A2) was 59% for inbound and 64% for outbound in November 2005, and was 43% for inbound and 56% for outbound in May 2008. The ratio has decreased, especially for inbound, probably due to a decrease in peer-to-peer traffic share.

The total volume of external domestic traffic (B2), mainly private peering, exceeds the volume for the six major IXes (B1) and the difference is widening, probably because major ISPs are shifting from public peering to private peering to handle increasing demands. This result underlines a possible deviation in estimating nation-wide traffic only from IX traffic. However, it is possible that the ratio of private peering is overestimated in our results as private peering is usually exercised only between large ISPs.

Another noticeable change is the increase in international traffic, especially for inbound. This is probably due to popular video services such as YouTube which do not have servers in Japan.

We examined the relationship between our IX traffic data (B1) and the total input rate of the six major IXes, as obtained directly from these IXes [29]. In comparison with the published total incoming traffic of these IXes, our data consistently represents about 42% of the total traffic as shown in Table 2. If we assume this ratio to be the traffic share of the six ISPs, the total amount of residential broadband traffic in Japan in May 2008 can roughly be estimated to be 632Gbps for inbound and 880Gbps for outbound, or 205PB/month for inbound and 285PB/month for outbound. These numbers are in line with Cisco's estimates of 226PB/month for consumer traffic in Japan in 2008[6].

Table 2: Share of (B1) outbound to the IXes against the inbound measured on the IX side (%)

2004			2005		2006		2007		2008
Sep	Oct	Nov	May	Nov	May	Nov	May	Nov	May
41.5	41.8	41.7	42.0	41.5	43.2	41.5	42.4	41.8	42.6

In summary, by comparing traffic in 2008 with 2005, we observed the larger download volume as well as larger evening volume in the residential traffic, the decreased share of residential traffic in the customer traffic, and the increased inbound volume in the international traffic. All of these, along with the volume comparable in size to peer-to-peer traffic, indicate an increase in video service content but the growth has been modest. A plausible explanation is that residential traffic is shifting from peer-to-peer file-sharing to video services. We will examine this hypothesis in the next section using per-customer measurements from one of the ISPs.

4. PER-CUSTOMER TRAFFIC ANALYSIS

Sets of Sampled NetFlow data in 2005 and 2008 were obtained from one of the participating ISPs to further analyze the behavior of residential traffic. Here, only residential broadband customer traffic is analyzed. In the data, one end of a flow is always the residential customer of the ISP but the other end is generally a customer of another ISP. This ISP has residential broadband customers over DSL and fiber but not over CATV.

By comparing the aggregated residential traffic graphs in Figure 5 with the ISP's corresponding graphs, the traffic characteristics are consistent. Hence, we believe that the results represent Japanese residential traffic even though the data sets are from only one ISP.

4.1 Measurement Methodology

Data was collected from all edge routers accommodating residential broadband customers. Traffic volume is derived by dividing the measured volume by the sampling rate. In 2005, the sampling rate was 1/2048 unanimously but, in 2008, 45% of the routers, notably busy ones, have 1/8192, 45% have 1/2048, and 10% have 1/1024, adjusted according to the load the routers. We believe it is still enough for analyzing user behaviors but there is a certain amount of sampling errors, especially for lightweight users.

4.2 Data Sets

Two types of week-long data sets are used for analysis. One type is daily inbound and outbound traffic volume of each customer obtained by matching customer IDs with the assigned IP addresses. This data also includes customer's line type (DSL or fiber), and is used to analyze per-customer behavior in Section 4.2 through 4.4. The data was collected in February 21-27 in 2005 and June 2-8 in 2008. The other type is raw NetFlow data, and used to analyze protocol usage in Section 4.5. The data was collected in July 4-10 in 2005 and June 2-8 in 2008.

By comparing the 2008 data with the 2005 data, the total number of active users increased by 94.7%, and the total traffic volume increased by 187.0%. The growth rate of the total traffic volume is larger than the aggregated result, probably because the ISP has expanded business in the residential broadband services. The daily traffic volume per active user increased from 430MB to 483MB for inbound by 12.3% and from 447MB to 797MB for outbound by 78.3%; a significant difference between inbound and outbound.

Table 3 shows the number of active unique users for fiber and DSL in 2005 and 2008. The fiber user ratio increased from 46% in 2005 to 79% in 2008 so that a large majority of the active users become fiber users. Accordingly, the total volume share of the fiber users increased from 79% in 2005 to 87% in 2008.

Table 3: Ratio of fiber and DSL active users and total traffic volumes

		active users (%)	total volume (%)
2005	fiber	46	79
	DSL	54	21
2008	fiber	79	87
	DSL	21	13

4.3 Distribution of per-customer traffic

Figure 6 shows the probability density functions of daily traffic per customer in log-linear scale. The daily traffic is the average for the week, and the distribution is computed independently for inbound and outbound traffic. The vertical lines in the plots are at 100MB/day and 2.5GB/day.

When we look at the fiber users in 2005 (top middle), there appears to be two roughly lognormal distributions where the logarithm of the variable is normally distributed. It is less clear in the other plots but the hypothesis still holds.

The one group on the left includes majority of users, and the outbound volume (download for users) is about ten times larger than the inbound volume. The other group on the right is heavy-hitters with symmetric inbound and outbound volumes. The two distinct distributions suggest that they have different mechanisms; most likely, the symmetric high-volume distribution is driven by the symmetric and aggressive nature of peer-to-peer file-sharing. For convenience, we call this group *peer-type*, and the other group *client-type*. The two distributions overlap, and the range of the overlap increased in 2008, especially for outbound. The spikes at the left edge are due to the increased sampling rates.

The distribution of the client-type group has moved towards higher volume, especially for outbound. The mode for inbound shifted from 3.5MB/day in 2005 to 5MB/day in 2008, and the mode for outbound shifted from 32MB/day in 2005 to 94MB/day in 2008. The modes of the distributions are similar in both fiber and DSL plots. The distribution of the outbound volume becomes wider, suggesting increased diversity.

On the other hand, the distribution of the peer-type group is not growing from 2005 to 2008. The mode of the distribution stays at around 2GB/day, and the covered area has slightly decreased. In fact, if we subtract the overlap of the client-type users from the plots, the population share and volume share of the peer-type users slightly decreased. The outbound distribution is almost absorbed by the client-type distribution, and becomes harder to distinguish in the 2008 plots. When focused on the tail at the right edge, the outbound volume slightly grew but the inbound volume did not.

The corresponding (complementary) cumulative distributions of daily traffic per user for the total users in 2005 and 2008 are shown on a log-log scale in Figure 7. The distributions are heavy tailed but there is a knee in the slope, at the top 4% of heavy-hitters using more than 2.5GB/day (or 230kbits/sec) for the total users in 2005. The distribution for inbound did not change much in 2008 but that for outbound has moved to the right, suggesting visible growth only in the download volume. The inbound and outbound volumes of the fiber heavy-hitters become symmetric, probably because the majority of the heavy-hitters become fiber users so that they do not need to compensate for the shortage of upstream bandwidth of DSL heavy-hitters.

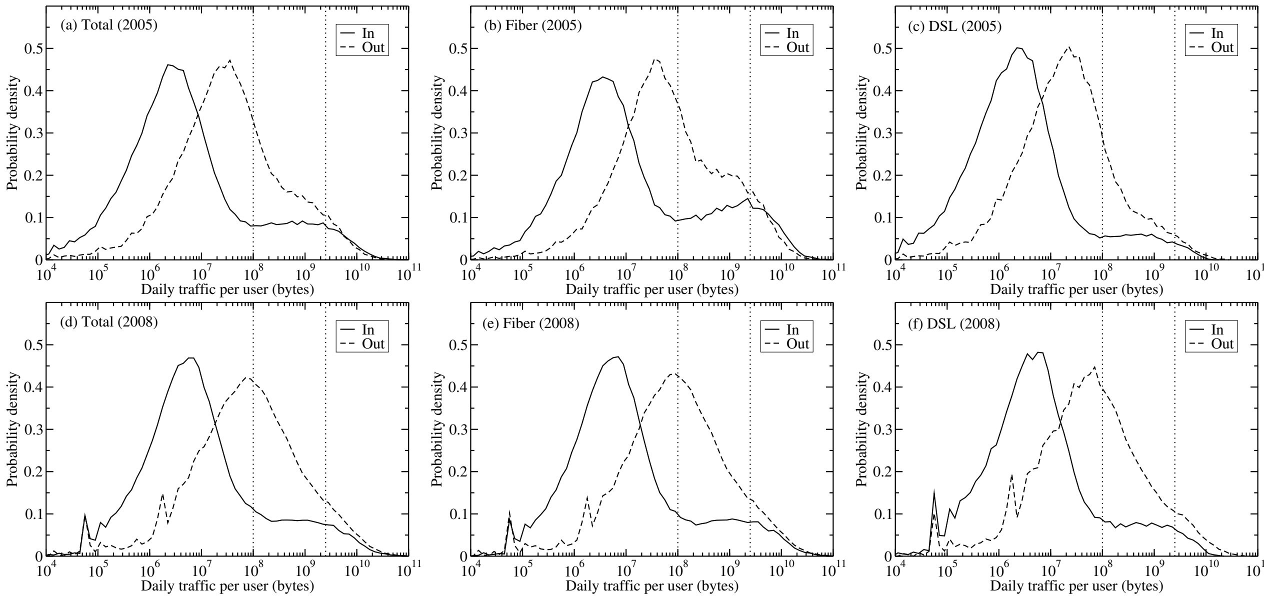

Figure 6: Probability density function of daily traffic per user: total (left), fiber users (middle), and DSL users (bottom) in 2005 (top) and 2008 (bottom).

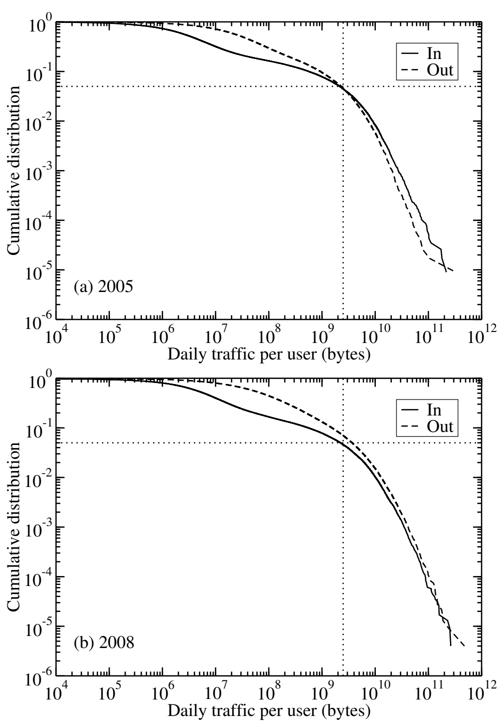

Figure 7: Cumulative distribution of daily traffic per user: 2005 (top) and 2008 (bottom).

Figure 8 shows the cumulative distribution of traffic volume consumed by top ranking heavy-hitters, computed independently for inbound and outbound traffic.

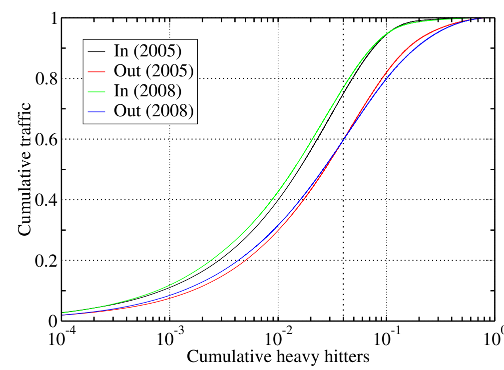

Figure 8: Cumulative distribution of traffic volume of heavy-hitters in decreasing order of volume: 2005 and 2008

The plot reveals a skewed traffic distribution among users; the top N% of heavy-hitters use X% of the total traffic. For example, the top 4% use 75% of the total inbound traffic, and 60% of the outbound. The distribution has not changed much from 2005 to 2008; a small group of heavy-hitters similarly represent a significant part of the total traffic. The overall traffic is still dictated by the heavy-hitter traffic even with the increased traffic by the client-type users, probably because the client-type users also have a long-tailed distribution.

4.4 Inbound and Outbound Correlation

To observe the ratio of inbound and outbound vol-

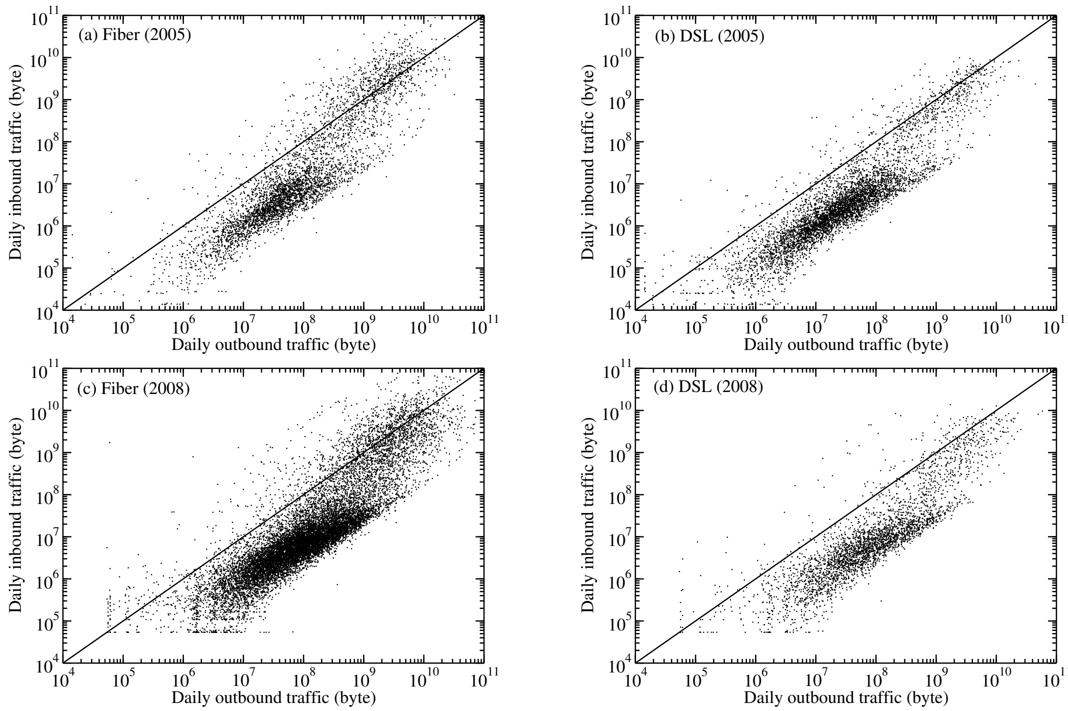

Figure 9: Correlation of inbound and outbound traffic volumes per user in one metropolitan prefecture: fiber (left) and DSL (right) in 2005 (top) and 2008 (bottom)

umes of each user, daily inbound and outbound volumes per user are shown in log-log scatter plots in Figure 9. These are taken from a metropolitan prefecture but the characteristics are common to all the prefectures. The number of plotted users is about 4300 for fiber and about 5400 for DSL in 2005 and about 14700 for fiber and about 3400 for DSL in 2008; a clear shift from DSL to fiber in 2008.

In all the plots, the highest density cluster is below and parallel to the unity line where outbound volume (download for users) is about ten times larger than that of inbound, which corresponds to the client-type users. In a higher volume region, a different cluster appears to exist around the unity line, which corresponds to the peer-type users.

The difference between the fiber and DSL plots is basically heavy-hitters' population share. It is also observed that there are heavy-hitters in the client-type group whose outbound volumes are comparable to that of heavy-hitters in the peer-type group. It is especially evident in the plot of the fiber users in 2008.

However, the boundary between the client-type users and the peer-type users as well as the boundary between the heavy-hitters and the normal users are not very clear. It can be also observed that, across the entire traffic volume range, the inbound/outbound ratio varies greatly, up to 4 orders of magnitude in all the plots. It suggests that there exit diverse users with a different traffic mix of client-type and peer-type applications. That is, although it is possible to classify users into client-type and peer-type at the macro level, actual individual users have mixed traffic of both types.

4.5 Protocol and Port Usage

Table 4 shows the ranking of protocols and ports for the total users, the client-type users, and the peer-type users in 2005 and 2008. The ranking is similar for all the groups with minor differences in dynamic ports so that the table is ordered by the ranking of the total users in 2008.

To rank port numbers in TCP and UDP, we take the smaller of the source and destination ports of a flow. TCP ports are further divided into well-known ports that are smaller than 1024, and dynamic ports that are equal to or larger than 1024. We do not distinguish registered ports from dynamic ports since many implementations use the registered port range from 1024 through 49151 for dynamic ports. We believe that majority of dynamic port traffic represent peer-to-peer applications so that dynamic port traffic can be used as a rough estimate of peer-to-peer traffic.

When the total traffic in 2005 is compared with that in 2008, port 80 (http) accounted only for 9% in 2005, and increased to 14% in 2008. TCP dynamic ports accounted for 83% in 2005 and decreased to 78% in 2008. That is, 5% shifted from dynamic ports to port 80. Still,

Table 4: Protocol breakdown: 2005 and 2008

protocol	port	2005			2008		
		total (%)	client type	peer type	**total** (%)	client type	peer type
TCP	*	**97.43**	**94.93**	**97.66**	**96.00**	**95.51**	**96.06**
	(< 1024)	*13.99*	*58.93*	*8.66*	*17.98*	*76.16*	*11.35*
	80 (http)	9.32	50.78	5.54	14.06	64.96	8.26
	554 (rtsp)	0.38	2.44	0.19	1.36	8.21	0.58
	443 (https)	0.30	1.45	0.19	0.58	1.63	0.46
	20 (ftp-data)	0.93	1.25	0.90	0.24	0.17	0.25
	81 (-)	0.15	0.04	0.16	0.23	0.04	0.25
	82 (-)	0.05	0.01	0.06	0.19	0.01	0.21
	110 (pop3)	0.17	1.00	0.10	0.14	0.46	0.10
	22 (ssh)	0.09	0.17	0.10	0.10	0.16	0.09
	25 (smtp)	0.14	0.51	0.11	0.07	0.11	0.07
	1000 (-)	0.03	0.02	0.04	0.07	0.01	0.08
	others	2.52	1.26	1.27	0.94	0.40	1.00
	(>= 1024)	*83.44*	*36.00*	*89.00*	*78.02*	*19.35*	*84.71*
	6346 (gnutella)	0.92	0.84	0.93	0.94	0.67	0.97
	6699 (winmx)	1.40	1.14	1.43	0.68	0.24	0.73
	7743 (winny)	0.48	0.15	0.51	0.30	0.04	0.33
	1935 (rtmp)	0.20	0.81	0.14	0.22	0.73	0.16
	6881 (bittorrent)	0.25	0.06	0.27	0.22	0.02	0.24
	7144 (-)	-	0.02	0.03	0.19	0.03	0.21
	8080 (-)	0.11	0.15	0.14	0.11	0.09	0.15
	4662 (edonkey)	0.12	0.02	0.13	0.13	0.00	0.14
	11560 (-)	0.03	0.02	0.10	0.04	0.01	0.11
	3074 (-)	0.02	0.03	0.10	0.02	0.32	0.07
	others	79.91	32.76	85.22	75.17	17.20	81.60
UDP	*	**1.38**	**3.41**	**1.19**	**1.94**	**2.50**	**1.88**
	53 (dns)	0.03	0.14	0.02	0.04	0.12	0.03
	others	1.35	3.27	1.17	1.90	2.38	1.85
ESP		**1.09**	**1.35**	**1.06**	**1.93**	**1.85**	**1.94**
GRE		**0.07**	**0.12**	**0.06**	**0.09**	**0.08**	**0.09**
ICMP		**0.01**	**0.05**	**0.01**	**0.02**	**0.05**	**0.02**
others		**0.02**	**0.14**	**0.02**	**0.02**	**0.01**	**0.01**

78% of the total traffic in 2008 are between dynamic ports and most of them are considered to be peer-to-peer traffic. The usage of each dynamic port is small because recent peer-to-peer applications use arbitrary ports. The largest one, port 6699 in 2005 is only 1.4% and port 6346 in 2008 is only 0.9%. It is evident that it is no longer possible to make use of port numbers for identifying applications.

To observe differences in peer-type and client-type users, the users are classified simply by average daily inbound volume with the threshold of 100MB/day. This threshold is roughly at the center of the two distributions in the inbound probability density functions in Figure 6. In our previous work, we used 2.5GB/day to distinguish heavy-hitters to focus on heavy-hitters but 100MB/day is used here to focus on client-type users.

For the client-type users, port 80 accounted for 51% in 2005 and increased to 65%. Dynamic ports accounted for 36% in 2005 and decreased to 19% in 2008. Port 80 is much larger in this group, and 17% shifted from dynamic ports to port 80 and other well-known ports. Also, there is a noticeable increase in port 554 used for the Real Time Streaming Protocol (RTSP), from 2% to 8%. RSTP is a control protocol for streaming content employed by major streaming services. RTSP has the interleaved mode where the video data stream is interleaved on the original TCP control connection. Thus, the increase in port 554 is an evidence of increased video content. Many of video content are also transported over port 80 so that considerable video volume is supposed to be included in port 80. It is almost impossible to quantify the exact video volume in port 80 but, given the heavy-tailed size distribution of video content which is much larger than traditional web content[10], it is likely that large part of the traffic on port 80 is already video and other rich media content.

Figure 10 compares temporal behaviors of three port groups: port 80, well-known ports other than port 80, and dynamic ports for the total users, the client-type users and the peer-type users in 2005 and 2008. The volume is normalized to the peak value of the total traffic size not to reveal the absolute traffic volume of the ISP.

The total traffic is heavily affected by the peer-type user group as expected. For the client-type users, it is clear that port 80 has increased and dynamic ports have decreased. Port 554 has a temporal pattern similar to port 80 but not shown in the plots.

For the peer-type users, most traffic is in dynamic ports, and an increase in port 80 is also observed. The daily fluctuation is speculated to be caused by stopping file-sharing applications after files are downloaded rather than leaving them running all the time. This can be observed from the traffic pattern of the dynamic ports that gradually decreases after midnight towards morning with the bottom at around 8am. It is in contrast to the port 80 traffic that drops quickly after midnight with the bottom at around 4am. The difference in traffic patterns suggests that many file-sharing users stop the application when they wake up in the morning.

This limited usage pattern of peer-to-peer file-sharing is partly due to increasing number of ISPs imposing limits on bandwidth usage, and partly due to users' risk awareness about leaving file-sharing applications as many incidents caused by compromised file-sharing applications have been reported in Japan.

For ordinary users, peer-to-peer file-sharing becomes less attractive with the advent of rich content services and their easy-to-use applications. At the same time, peer-to-peer traffic will not go away anytime soon. Also, the peer-to-peer mechanism itself could evolve into a powerful engine to drive content disribution in large scale, if it becomes more friendly to users and ISPs[30].

4.6 Summary

The overall traffic is still dominated by heavy-hitters, mainly using peer-to-peer applications. However, their traffic decreased slightly in both population share and volume share.

The current slow growth rate is attributed to the fact that the dominant aggressive peer-to-peer traffic is not growing much. On the other hand, the client-type traffic is slowly moving towards high volume usage. The circumstantial evidence indicates that it is driven by

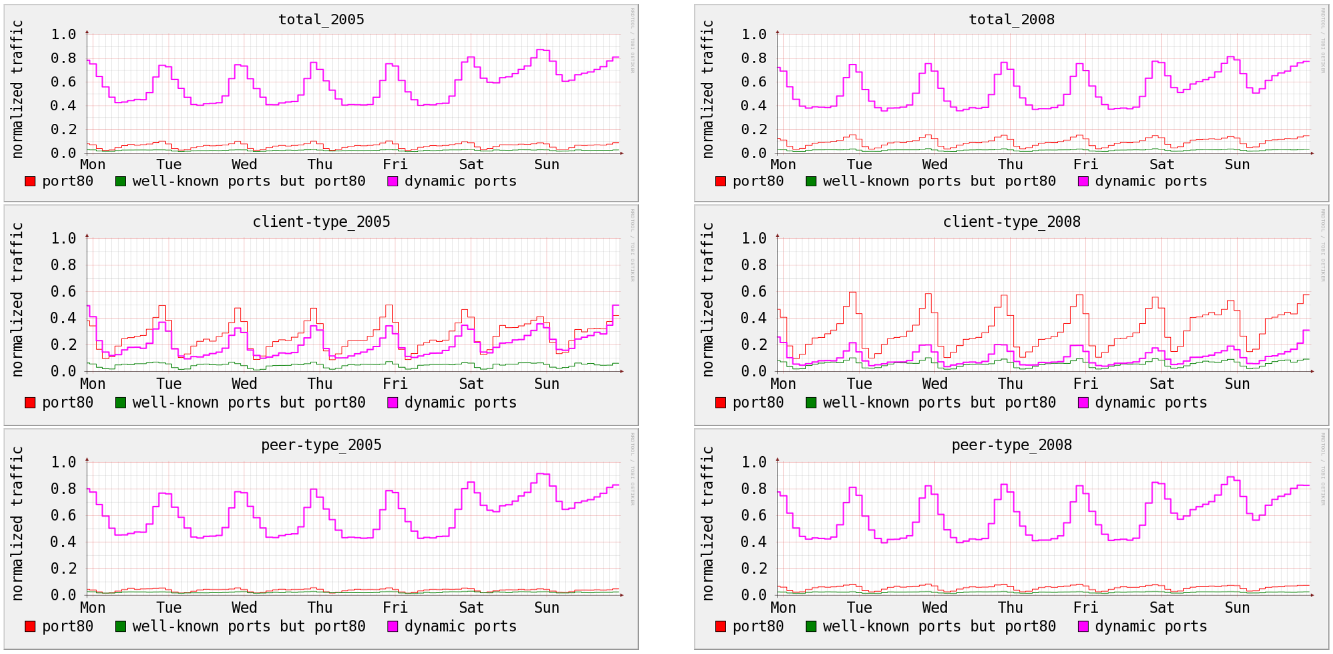

Figure 10: Temporal behavior of port usage: total users (top), client-type users (middle) and peer-type users (bottom) in 2005 (left) and 2008 (right)

video and other rich media content. The increase of video content is, however, not yet very visible in the total residential traffic volume at the macro level. We will examine the growth of this client-type user traffic in the next section.

Meanwhile, the capacity of the Internet, both access networks and core networks, will continue to grow. The annual growth of the access capacity is reported and projected to be 50% per year by the FTTH Council Europe[7] and, in fact, Japanese residential users are shifting from DSL to fiber access. The backbone networks also have been increasing capacity fast enough to handle the demands. TeleGeography[28] reports that annual international Internet capacity growth was 45% in 2004-2006 and 68% in 2007.

If these trends continue, the traffic growth will stay at a modest rate comparable to the growth of network capacity. Video content is steadily growing but unlikely to disastrously overflow the global Internet in the short term.

5. DISCUSSION

Video and rich media content in residential traffic will have a significant impact on the traffic growth in the future[5]. Hence, we focus on the client-type users and analyze their ourbound traffic as this group is characterized by download volume.

From the probability density function in Figure 6, traffic volume per user for the client-type group is roughly lognormally distributed. Because the distribution has a long tail, its simple arithmetic mean is heavily affected by the tail of the distribution. The mean is of interest as it is directly relates to the total traffic volume. We examine the traffic characteristics using the properties of the lognormal distribution.

It is not surprising that per-customer traffic volume follows a lognormal distribution because a multiplicative stochastic process leads to a lognormal distribution [9], and traffic growth is well modeled by a multiplicative stochastic process; each user increases traffic following a growth rate which is an independent identically distributed random process. Similar observations were reported in the distribution of traffic volume at an organization level[8] as well as in a growth model of web pages[13].

The probability density function of a log-normal distribution is given by

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(\frac{-(\ln x - \mu)^2}{2\sigma^2}) \quad (1)$$

where μ and σ are the mean and standard deviation of $\ln(x)$. The mean of the original distribution, $p(x)$, is

$$E(x) = \exp(\mu + \sigma^2/2) \quad (2)$$

and the variance of $p(x)$ is $\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$.

By fitting a lognormal distribution to the outbound of the total traffic in 2005 and 2008 with nonlinear least-square Marquardt-Levenberg algorithm, we estimated $\mu = 17.29, \sigma = 1.813$ for 2005 and $\mu = 18.36, \sigma =$

2.105 for 2008. Thus, the mode and mean of $p(x)$ are computed from μ and σ as 32.2MB and 166.7MB for 2005, and 93.8MB and 860.0MB for 2008. The mode of 93.8MB/day seems reasonable considering popular web sites with rich media content (e.g., the mean file size of YouTube videos is 10MB[10], and popular applications such as Google Maps proactively fetch images in background). The mean of 860MB/day, however, seems much larger than one would expect but its annual growth rate is still within a reasonable range.

The compound annual growth rate (CAGR) is 1.39 for the mode and 1.65 for the mean, computed by 3.28 years that is between February 21-27 2005 and June 2-8 2008. The growth rate of the mean is much larger than that of the mode as characterized by Equation 2 where the mean grows much faster than the mode. Also, σ slightly increased from 2005 to 2008, which means larger contributions from the tail of the distribution.

Table 5 shows simplistic growth projections using both additive and multiplicative growth. Traffic growth has been proved to be multiplicative but additive growth is shown as the lower bounds. The simple arithmetic mean of the original data is also listed in the table as *total mean*. With multiplicative growth, the mode and the median will be 252MB/day and 3.9GB/day in 2011.

Table 5: Simplistic growth projections for outbound traffic per user (MB/day)

	additive growth			multiplicative growth		
	total	lognormal		total	lognormal	
	mean	mode	mean	mean	mode	mean
2005 Feb	446.6	32.2	166.7	446.6	32.2	166.7
2008 Jun	796.6	93.8	860.0	796.6	93.8	860.0
growth/yr	106.7	18.8	211.4	1.19	1.39	1.65
2009 Jun	903	113	1071	948	130	1419
2010 Jun	1010	131	1281	1128	181	2341
2011 Jun	1117	150	1494	1342	252	3863

Since the distribution is long-tailed, the mean is dictated by a small segment of users at the tail and becomes less predictable than the mode. We do not think that the mean keeps growing with this rate because 2005 is before popular video services took off so that this index could be overestimated. Also, a larger bias against heavy-hitters will take place as traffic volume approaches to the access link capacity. However, at least, there is a potential that traffic of ordinary users could grow to a substantial volume.

The growth looks significant but it is similar to how the web traffic growth looked in late 1990 when people were still using dial-up access. From an optimistic view, technical advances in access and core networks are likely to offset the traffic growth of this level.

6. CONCLUSION

As the Internet becomes a social communication platform, individual users of all ages start actively participating in digital communications with diverse content, notably with much richer images and videos than traditional web content. The traffic mix and volume are changing accordingly.

It is difficult to predict future traffic. As we have witnessed in the past, the advent of new applications would change the traffic mix and, accordingly, growth rate. However, at the macro level, traffic will keep growing.

There will be many challenges posed by unprecedented traffic volume. A variety of technical solutions exist to mitigate possible problems such as content caching, content distribution networks and preferential quality of service. It also involves with economic factors such as the access service cost and the costs of backbone capacity and equipment as well as political and social factors such as net-neutrality and content management. Technical and solid observations on traffic are essential to understand changes and adapt to new demands.

In this paper, we have shown that the apparent slow traffic growth is due primarily to the stalled growth of peer-to-peer traffic that is still dominant in the current traffic. At the same time, the usage of ordinary users is slowly swelling with increased diversity, driven by video and other rich media content. At the macro level, the impact of video content is still small compared with peer-to-peer traffic, and continue to be so in the next few years.

As for the generality of our results, some aspects are specific to Japanese traffic such as the high penetration of fiber access, geographic concentration and language barriers. Nevertheless, our key findings seem to be common to other countries, although the exact ratio of traffic mix and their growth should be different from country to country, and probably, from ISP to ISP.

Persistent and consistent data collection is essential for this type of study. We will continue monitoring traffic by collecting aggregated traffic logs from the participating ISPs, and by analyzing per-customer traffic, hopefully from more ISPs.

Acknowledgments

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

K.F. is partially supported by the Grant-in-Aid for Young Scientists (B) from MEXT Japan.

7. REFERENCES

[1] N. Basher, A. Mahanti, A. Mahanti,

C. Willamson, and M. Arlitt. A comparative analysis of web and peer-to-peer traffic. In *WWW'08*, pages 287–296, Beijing, China, Apr. 2008.

[2] N. Brownlee and kc claffy. Understanding internet traffic streams: Dragonflies and tortoises. *IEEE Communications Magazine*, 40:110–117, 2002.

[3] K. Cho, K. Fukuda, H. Esaki, and A. Kato. The impact and implications of the growth in residential user-to-user traffic. In *SIGCOMM2006*, pages 207–218, Pisa, Italy, Aug. 2006.

[4] Cisco Sampled NetFlow. http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120limit/120s/120s11/12s_sanf.htm.

[5] Cisco. Approaching the zettabyte era. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374_ns827_Networking_Solutions_White_Paper.html, June 2008.

[6] Cisco visual networking index – forecast and methodology, 2007-2012. http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html, June 2008.

[7] FTTH Council Europe. True broadband in Europe. http://www.emailbrain.com/eb_members/3301/ftp/Ventura_Team__Press_Release-UK_expected_to_lag_Europe_4-10_years_in_Fibre_Broadband.pdf, Mar. 2008.

[8] K. Fukuda. Towards modeling of traffic demand of node in large-scale network. In *IEEE ICC2008*, pages 214–218, Beijing, China, May 2008.

[9] R. Gibrat. *Les inégalitś économiques*. Paris, Sirey, 1931.

[10] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: A view from the edge. In *IMC2007*, pages 15–28, San Diego, CA, Oct. 2007.

[11] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In *SOSP-19*, pages 314–329, Bolton Landing, NY, Oct. 2003.

[12] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang. Measurements, analysis, and modeling of bittorrent-like systems. In *IMC2005*, pages 35–48, Berkeley, CA, Oct. 2005.

[13] B. A. Huberman and L. A. Adamic. Growth dynamics of the world-wide web. *Nature*, 401:131, 1999.

[14] W. John, S. Tafvelin, and T. Olovsson. Trends and differences in connection-behavior within classes of internet backbone traffic. In *PAM2008 (LNCS 4979)*, pages 192–201, Cleveland, OH, Apr. 2008.

[15] Japan Internet Exchange Co., Ltd. http://www.jpix.co.jp.

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

[17] M. Li, M. Claypool, R. Kinicki, and J. Nichols. Characteristics of streaming media stored on the web. *ACM Transactions on Internet Technologies*, 5(4):601–626, Nov. 2005.

[18] NSPIXP. http://nspixp.wide.ad.jp.

[19] A. M. Odlyzko. Minnesota Internet traffic studies. http://www.dtc.umn.edu/mints/home.html.

[20] OECD Broadband Statistics. http://oecd.org/sti/ict/broadband, Dec. 2007.

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

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

[23] L. Plissonneau, J.-L. Costeux, and P. Brown. Analysis of peer-to-peer traffic on ADSL. In *PAM2005 (LNCS3431)*, pages 69–82, Boston, MA, Mar. 2005.

[24] S. Saroiu, P. K. Gummadi, and S. D. Gribble. A measurement study of peer-to-peer file sharing systems. In *MMCN'02*, pages 156–170, San Jose, CA, Jan. 2002.

[25] S. Sarvotham, R. Riedi, and R. Baraniuk. Connection-level analysis and modeling of network traffic. In *IMW'01*, pages 99–103, San Francisco, CA, Nov. 2001.

[26] S. Sen and J. Wang. Analyzing peer-to-peer traffic across large networks. In *IMW'02*, pages 137–150, Marseille, France, Nov. 2002.

[27] M. Siekkinen, D. Collange, G. Urvoy-Keller, and E. W. Biersack. Performance limitations of adsl users: A case study. In *PAM2007 (LNCS 4427)*, pages 145–154, Louvain-la-neuve, Belgium, Apr. 2007.

[28] TeleGeography Research. Globel Internet Geography. , 2007.

[29] The Ministry of Internal Affairs and Communications of Japan. MIC Statistics on data communications. http://www.soumu.go.jp/joho_tsusin/eng/statistics.html.

[30] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz. P4P: Provider portal for applications. In *SIGCOMM 2008*, Seattle, WA, Aug. 2008.

[31] Y. Zhang, L. Breslau, V. Paxson, and S. Shenker. On the characteristics and origins of internet flow rates. In *SIGCOMM 2002*, pages 309–322, Pittsburgh, PA, Aug. 2002.