# Traffic Trends over the Past Year

**There have been no major shifts in broadband traffic over the past year, and it is in a stable trend. Close attention will be paid to how the amended Copyright Act that comes into effect on October 1 of this year affects broadband traffic.**

## 3.1 Introduction

In this report we analyze traffic over the broadband access services operated by IIJ and present the results[*1][*2][*3]. We once again report on changes in traffic trends over the past year based on daily user traffic and usage by port.

## 3.2 About the Data

As with our previous reports, the survey data utilized here was collected using Sampled NetFlow from the routers accommodating fiber-optic and DSL broadband customers of our personal and enterprise broadband access services. Because broadband traffic trends differ between weekdays and weekends, we analyze a full week of traffic. We will compare data for the week spanning May 28 to June 3, 2012 with the data we analyzed in the previous report for the week spanning May 30 to June 5, 2011.

The usage volume for each user was obtained by matching the IP address assigned to users with the IP addresses observed. We collected statistical information by sampling packets using NetFlow. The sampling rate was set to 1/8192, taking into account router performance and load. We estimated overall usage volumes by multiplying observed volumes by the reciprocal of the sampling rate. Due to the sampling method used there are slight estimation errors in data for low-volume users. However, for users with usage above a certain level we were able to obtain statistically meaningful data.

Over the past few years the migration from DSL to fiber-optic connections has continued, with 91% of users observed in 2012 having a fiber-optic connection, and these connections accounting for 95% of overall traffic volumes.

The IN/OUT traffic presented in this report indicates directions from an ISP's perspective. IN represents uploads from users, and OUT represents user downloads.

Figure 1 shows average monthly traffic for broadband as a whole over the past five years. As we reported the year before last, traffic decreased in January 2010. It is thought this was caused by the amended Copyright Act that came into effect in January 2010, making the download of copyright infringing content illegal. Since then, download volumes (OUT) have continued to rise, while upload volumes (IN) have remained mostly level, indicating that the ratio of P2P file sharing traffic has decreased. Over the past year IN traffic has increased a tiny 0.2%, while OUT traffic has increased 8.8%.

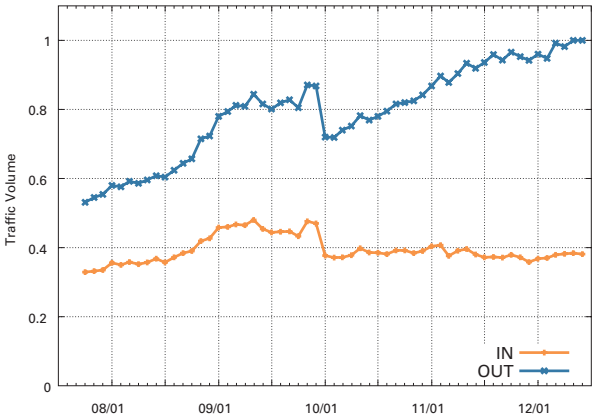


**Figure 1: Broadband Traffic Volume Trends for the Past 5 Years**

*1 See IIR Vol.12 "Broadband Traffic Report: Examining the Impact of the Earthquake on Traffic on a Macro Level" (http://www.iij.ad.jp/en/company/development/iir/pdf/iir_vol12_bband_EN.pdf).

*2 See IIR Vol.8 "Traffic Shifting away from P2P File Sharing to Web Services" (http://www.iij.ad.jp/en/company/development/iir/pdf/iir_vol08_report_EN.pdf)

*3 See IIR Vol.4 "Broadband Traffic: Increasing Traffic for General Users" (http://www.iij.ad.jp/en/company/development/iir/pdf/iir_vol04_bband_EN.pdf)

## 3.3 Daily Usage Levels for Users

First, we will examine the daily usage volumes for broadband users from several perspectives. Daily usage indicates the average daily usage calculated from a week's worth of data for each user.

Figure 2 shows the average daily usage distribution (probability density function) per user. It compares data for 2011 and 2012 divided into IN (upload) and OUT (download), with user traffic volume on the X axis, and probability density of users on the Y axis. The X axis shows volumes between 10 KB ($10^4$) and 100 GB ($10^{11}$) using a logarithmic scale. Some users are outside the scope of the graph, with the highest user usage levels climbing to 800 GB, but most fall within the scope of 100 GB ($10^{11}$).

The IN and OUT distribution shows almost log-normal distribution, which is the normal distribution in a semi-log graph. A linear graph would show a long-tailed distribution, with the peak close to the left end and a slow decay towards the right. The OUT distribution is further to the right than the IN distribution, indicating that the download volume is an order of magnitude larger than the upload volume. Comparing 2011 and 2012, the peak distribution for both IN and OUT traffic has moved slightly to the right, demonstrating that overall user traffic volumes are increasing.

The tail to the right of the IN distribution has grown longer. Previously, both IN and OUT showed a clearer peak here, indicating heavy users with symmetrical IN/OUT volumes. For convenience, we labeled users with asymmetrical IN/OUT traffic distribution that make up the majority "client-type users", and the distribution of heavy users with symmetrical IN/OUT traffic that make up the minority on the right side "peer-type users". In this report we will continue to use these conventions. Comparing 2011 and 2012, we can see that the peak for peer-type users is now smaller, indicating that the proportion of heavy users has decreased. A slight spike appears on the left side of the graph, but this is just noise caused by the sampling rate.

Table 1 shows trends in the average value and most frequent value that represents peak distribution. Comparing the most frequent values in 2011 and in 2012, IN rose from 8.5 MB to 14 MB, and OUT rose from 223 MB to 282 MB. This demonstrates that, particularly in the case of downloads, the traffic volume for each user has increased. Meanwhile, because average values are pulled up by the heavy users to the right of the graph, they are significantly higher than the most frequent values, with the average IN value 410 MB and the average OUT value 1,026 MB in 2012. The average values for 2011 were 432 MB and 1,001 MB, respectively, indicating that the average IN value decreased and the average OUT value increased.

Figure 3 plots the IN/OUT usage volumes for 5,000 randomly sampled users. The X axis shows OUT (download volume) and the Y axis shows IN (upload volume), with both using a logarithmic scale. Users with identical IN/OUT values are plotted on the diagonal line.

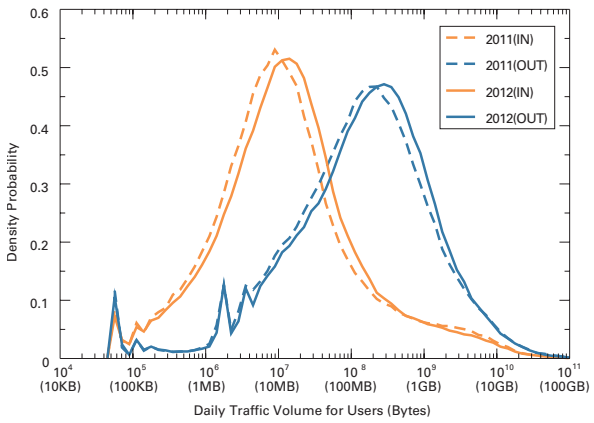


**Figure 2: Daily User Traffic Volume Distribution Comparison of 2011 and 2012**

| | IN (MB/day) | | OUT (MB/day) | |
|---|---|---|---|---|
| Year | Average Value | Most Frequent Value | Average Value | Most Frequent Value |
| 2005 | 430 | 3.5 | 447 | 32 |
| 2007 | 433 | 4 | 712 | 66 |
| 2008 | 483 | 5 | 797 | 94 |
| 2009 | 556 | 6 | 971 | 114 |
| 2010 | 469 | 7 | 910 | 145 |
| 2011 | 432 | 8.5 | 1,001 | 223 |
| 2012 | 410 | 14 | 1,026 | 282 |

**Table 1: Trends in Average Daily Traffic Volume for Users and Most Frequent Values**

The cluster below the diagonal line and spread out parallel to it represents general client-type users with download volumes an order of magnitude higher than upload volumes. Previously there was a clearly-recognizable cluster of peer-type heavy users spread out thinly on the upper right of the diagonal line, but this is now more difficult to identify. Though we have separated client-type and peer-type users for convenience, in actual fact client-type general users also use peer-type applications such as Skype, and peer-type heavy users also use download-based applications on the Web, blurring the boundary between them. In other words, many users use both types of applications in varying ratios. There are also differences in the usage levels and IN/OUT ratio for each user, pointing to the existence of diverse forms of usage. In this respect, almost no difference can be discerned between current data and 2011.

Figure 4 shows the complementary cumulative distribution of the daily traffic volume for users. This indicates the percentage of users with daily usage levels greater than the X axis value on the Y axis in a log-log scale, which is an effective way of examining the distribution of heavy users. The right side of the graph falls linearly, showing a long-tailed distribution close to power-law distribution. Compared to 2011, the tail to the right of the graph has extended slightly, and traffic volumes for heavy users have increased, but the number of extremely heavy users off to the right of the straight line has actually decreased. In any case, it can be said that heavy users are distributed statistically, and are by no means a special class of user.

Figure 5 shows the deviation in traffic usage levels between users. It indicates that users with the top X% of usage levels account for Y% of the total traffic volume. There is a great deal of deviation in usage levels, and as a result traffic volume for a small portion of users accounts for the majority of the overall traffic. For example, the top 10% of users make up 73% of the total OUT traffic, and 95% of the total IN traffic. Furthermore, the top 1% of users make up 33% of the total OUT traffic, and 59% of the total IN traffic.

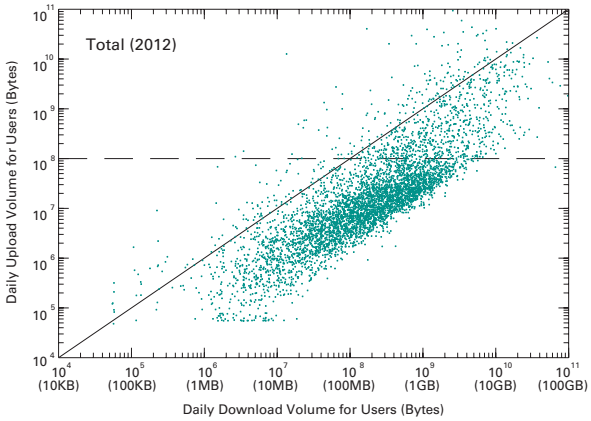


**Figure 3: IN/OUT Usage for Each User**

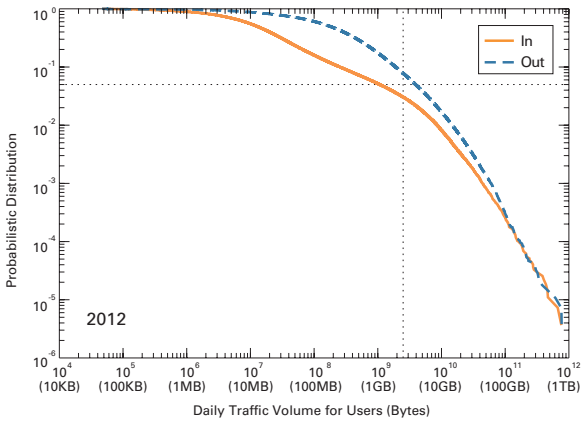


**Figure 4: Complementary Cumulative Distribution of the Daily Traffic Volume for Users**

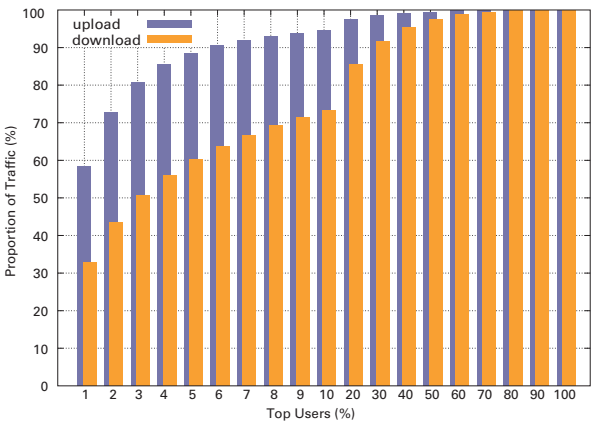


**Figure 5: Traffic Usage Deviation Between Users**

## 3.4 Usage by Port

Next, we will look at a breakdown of traffic and examine usage levels by port. Recently, it has been difficult to identify applications by port number. Many P2P applications use dynamic ports on both ends, and a large number of client/server applications utilize port 80 assigned to HTTP to avoid firewalls. To broadly categorize, when both parties use a dynamic port higher than port 1024, there is a high possibility of it being a P2P application, and when one party uses a well-known port lower than port 1024, there is a high possibility of it being a client/server application. In light of this, here we will look at usage levels for TCP and UDP connections by taking the lower port number of the source and destination ports.

As overall traffic is dominated by peer-type heavy user traffic, to examine trends for client-type general users, we have taken the rough approach of extracting data for users with a daily upload volume of less than 100 MB, and treating them as client-type users. This corresponds to users below the horizontal line at the IN=100 MB point in Figure 3.

Figure 6 shows an overview of port usage, comparing all users and client-type users for 2011 and 2012. Table 2 shows detailed numeric values for this figure.

82% of traffic in 2012 is TCP based. Furthermore, looking at overall traffic, TCP dynamic ports that accounted for 50% of the total in 2011 have dropped to 41% in 2012. The ratio of individual dynamic port numbers is tiny, with port 1935 used by Flash Player the highest at 2% of the total, and the next highest under 0.5%. Meanwhile, the use of port 80 has increased from 32% in 2011 to 36% in 2012.

Looking exclusively at client-type users, port 80 traffic that accounted for 67% of the total in 2011 has increased to 79% in 2012. The next highest is port 443 used for HTTPS, which increased from 2% in 2011 to 3%. Last year port 554 used for the Real-Time Streaming Protocol (RTSP) was next highest, but this has dropped from 7% in 2011 to 1%. The ratio of dynamic ports also decreased from 11% to 10%.

From this data, we can confirm that the upward trend in TCP port 80 traffic continues to spread from to general users to heavy users. Port 80 traffic is also used for data such as video content and software updates, so we cannot identify the type of content this is attributed to, but it is clear that client/server communications are on the rise.

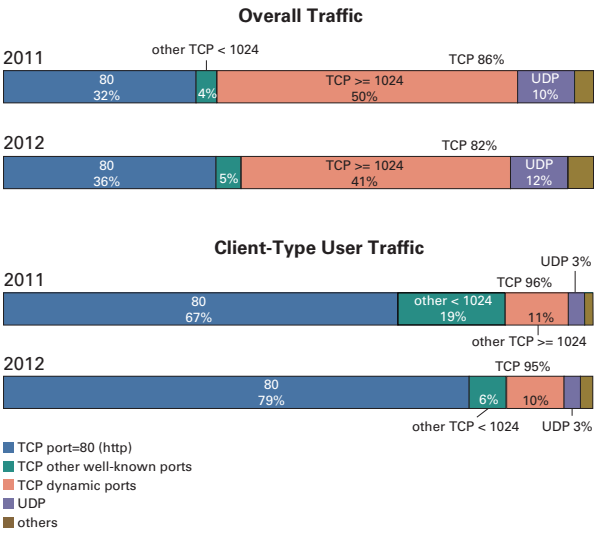


**Figure 6: Usage Level Overview by Port**

| protocol port | 2011 | | 2012 | |
|---|---|---|---|---|
| | total (%) | client type | total (%) | client type |
| **TCP** * | **85.95** | **96.28** | **81.86** | **95.09** |
| (<1024) | 36.24 | 85.69 | 41.23 | 85.25 |
| 80 (http) | 32.10 | 67.30 | 36.22 | 79.39 |
| 443 (https) | 1.33 | 1.91 | 2.45 | 3.43 |
| 554 (rtsp) | 1.33 | 6.89 | 0.77 | 1.01 |
| 22 (ssh) | 0.27 | 0.17 | 0.22 | 0.06 |
| (>=1024) | 49.71 | 10.59 | 40.63 | 9.84 |
| 1935 (rtmp) | 1.58 | 1.51 | 2.12 | 3.91 |
| 7144 (peercast) | 0.38 | 0.00 | 0.44 | 0.04 |
| 6346 (gnutella) | 0.68 | 0.60 | 0.37 | 0.09 |
| 8080 | 0.26 | 0.14 | 0.30 | 0.17 |
| **UDP** | **10.01** | **2.61** | **12.38** | **2.94** |
| **ESP** | **3.56** | **1.02** | **5.29** | **1.79** |
| **GRE** | **0.15** | **0.05** | **0.16** | **0.14** |
| **L2TP** | **0.13** | **0.00** | **0.14** | **0.00** |
| **IP-ENCAP** | **0.10** | **0.01** | **0.09** | **0.01** |

**Table 2: Usage Level Details by Port**

Figure 7 compares trends in TCP port usage over a week for overall traffic in 2011 and 2012. Trends in TCP port usage are shown for three categories: port 80, other well-known ports, and dynamic ports. Traffic is normalized by the total peak traffic volume. Compared with 2011, we can see that the overall ratio of port 80 usage has increased further, and is now close to the ratio for dynamic ports. The overall peak is between 21:00 and 1:00, and traffic also increases in the daytime on Saturday and Sunday, reflecting times when the Internet is used at home.

Figure 8 shows weekly TCP port usage trends for client-type and peer-type users. Port 80 usage accounts for the vast majority among client-type users, with peak hours between 21:00 and 23:00. We can confirm that dynamic port usage is still prevalent among peer-type users.

## 3.5 Conclusion

These results demonstrate that there have been no major shifts in broadband traffic over the past year, and it is in a stable trend. Overall, download volumes increased about 9%, but upload volumes remained about the same, and the ratio of TCP port 80 usage has climbed higher. As we reported previously, it is clear that the migration to Web services has gained further traction.

However, it is not clear whether this stable trend will continue in the future. On June 20 of this year, an amended Copyright Act that incorporates criminal punishments for illegal downloads and illegalization of ripping copy-protected DVDs was passed by the Diet, and will come into effect from October 1. Traffic dropped sharply after the download of copyright infringing content became illegal in January 2010, and it is possible that the same thing will occur this time. On the other hand, as we discussed in the previous report, if making the download of copyright infringing content illegal was merely a trigger that accelerated a process that was already occurring, there is a chance that the new amendment will have little effect. Close attention will be paid to how broadband traffic changes in October of this year, in order to verify the effects of legal enforcement measures.

IIJ monitors traffic levels on an ongoing basis so we can respond swiftly to changes in user behavior. We will continue to publish reports such as this periodically.

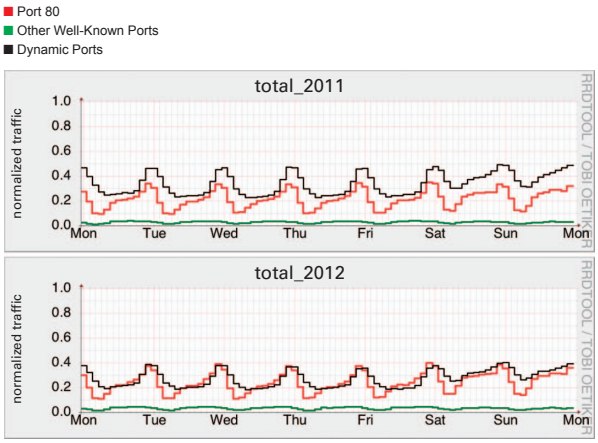


**Figure 7: Weekly TCP Port Usage Trends for 2011 (Top) and 2012 (Bottom)**

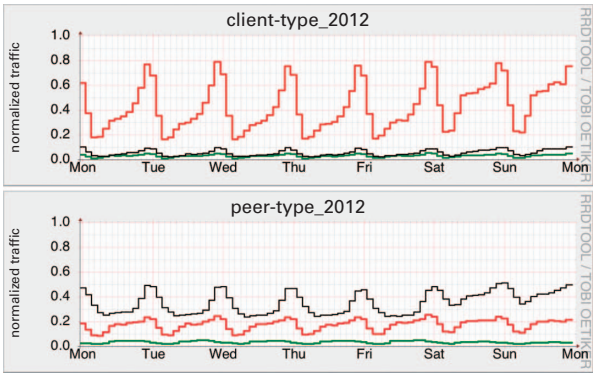


**Figure 8: Weekly TCP Port Usage Trends for Client-Type (Top) and Peer-Type (Bottom) Users**

Author:
**Kenjiro Cho**
Dr. Cho is Research Director of Research Laboratory at IIJ Innovation Institute Inc. He is engaged in Internet research such as traffic measurement and data analysis. He is a guest professor at Keio University, Faculty of Environmental Studies. He is also an adjunct professor of Information Science at Japan Advanced Science and Technology.