Traffic Volumes Rise Steadily Over the Past Year, and HTTPS Use Expands

Looking at broadband traffic over the past year, we can see that traffic volumes have increased steadily, with download volumes up 27%, and upload volumes up 13%.
Also, due to elevated awareness of the importance of protecting privacy there was a shift toward the use of HTTPS in Web traffic that accounts for the majority of traffic. We expect the ratio of HTTPS usage to continue to increase in the future.

2.1 Overview

In this report we analyze traffic over the broadband access services operated by IIJ every year and present the results[*1][*2][*3][*4][*5]. We once again report on changes in traffic trends over the past year based on daily user traffic and usage by port.

Figure 1 shows average monthly traffic across IIJ's entire suite of broadband services for the past seven years, with the maximum value normalized as 1. The drop in traffic in January 2010 is believed to be caused by the amended Copyright Act that came into effect that month, 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. In October 2012, a slight increase followed by a decrease was observed when an amended Copyright Act that incorporated criminal punishment for illegal downloads came into effect. Subsequently, download volumes have grown to higher levels than before, and upload volumes are also increasing slowly but steadily. Over the past year IN traffic has increased by 13%, while OUT traffic has increased by 27%.

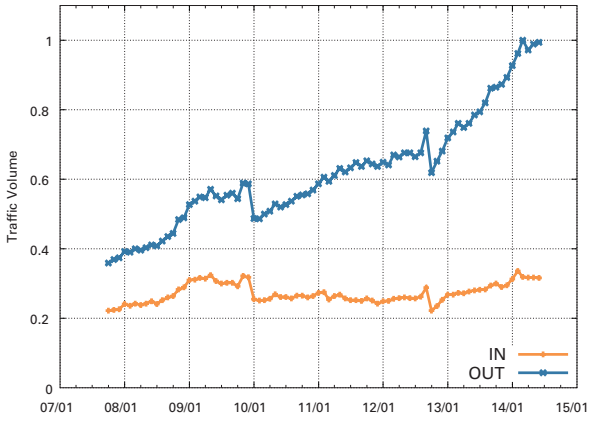

Figure 1: Broadband Traffic Volume Trends for the Past 7 Years

2.2 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. In this case, we compare data for the week spanning May 26 to June 1, 2014 with the data we analyzed in the previous report for the week spanning June 3 to June 9, 2013.

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

*1 Kenjiro Cho. Broadband Traffic Report: The Impact of Criminalization of Illegal Downloads was Limited. Internet Infrastructure Review. Vol.20. pp32-37. August 2013.
*2 Kenjiro Cho. Broadband Traffic Report: Traffic Trends over the Past Year. Internet Infrastructure Review. Vol.16. pp33-37. August 2012.
*3 Kenjiro Cho. Broadband Traffic Report: Examining the Impact of the Earthquake on Traffic on a Macro Level. Internet Infrastructure Review. Vol.12. pp25-30. August 2011.
*4 Kenjiro Cho. Broadband Traffic Report: Traffic Shifting away from P2P File Sharing to Web Services. Internet Infrastructure Review. Vol.8. pp25-30. August 2010.
*5 Kenjiro Cho. Broadband Traffic: Increasing Traffic for General Users. Internet Infrastructure Review. Vol.4. pp18-23. August 2009.

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.

IIJ provides both fiber-optic and DSL access for its broadband services. However, fiber-optic access now makes up the vast majority of use, with 95% of users observed in 2014 using fiber-optic connections, accounting for 97% 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.

2.3 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 2013 and 2014 divided into IN (upload) and OUT (download), with user traffic volume on the X axis, and user frequency 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, but most fall within the 100 GB (10^{11}) range.

The IN and OUT distribution shows almost log-normal distribution, which looks like a 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 more than an order of magnitude larger than the upload volume. Comparing 2013 and 2014, the peak distribution for both IN and OUT traffic has moved slightly to the right, demonstrating that overall user traffic volumes are increasing. The shift is greater than when 2012 and 2013 were compared last year, indicating that traffic volumes have increased at a higher rate.

Looking at OUT distribution, the peak has been steadily moving to the right over the past few years. However, the usage levels of heavy users on the right end have not increased much, and the distribution is beginning to lose its symmetry. Meanwhile, the tail of the IN distribution to the right 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. Over the past few years, the peak for peer-type users has shrunk to the point where it can hardly be distinguished. This indicates that the ratio of heavy users is decreasing. The small spikes on the left of the graph are noise caused by the sampling rate. These correspond to the minimum and maximum packet sizes when only one packet is observed.

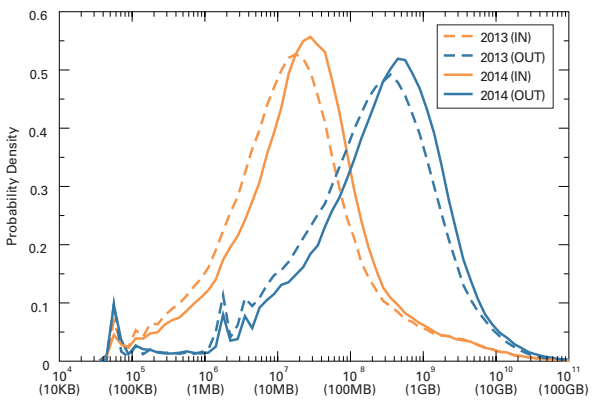

Figure 2: Daily User Traffic Volume Distribution Comparison of 2013 and 2014

	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
2013	397	18	1,038	355
2014	437	28	1,287	447

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

Table 1 shows trends in the average value and most frequent value that represents peak distribution. Comparing the most frequent values in 2013 and in 2014, IN rose from 18 MB to 28 MB, and OUT rose from 355 MB to 447 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 437 MB and the average OUT value 1,287 MB in 2014. The average values for 2013 were 397 MB and 1,038 MB, respectively. The IN value that had been falling since 2010 has begun to recover, and it seems that the migration from P2P file sharing applications to Web services has settled down.

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.

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 no longer discernible. 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 seen between the current data and that for 2013.

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. 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

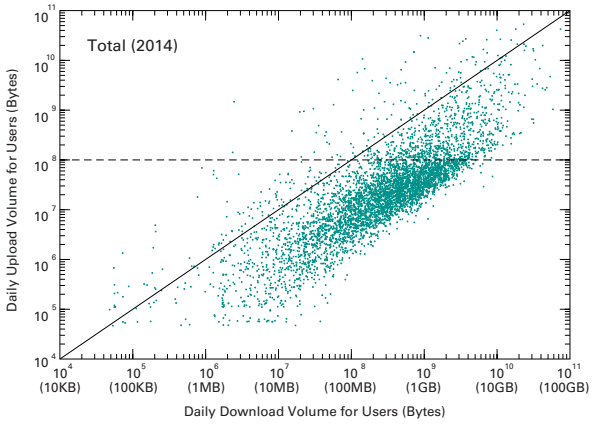

Figure 3: IN/OUT Usage for Each User

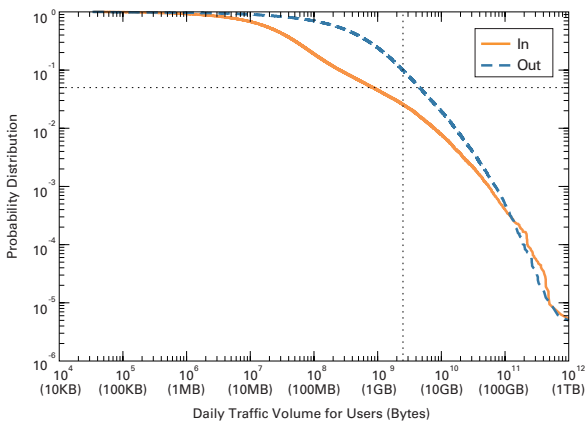

Figure 4: Complementary Cumulative Distribution of the Daily Traffic Volume for Users

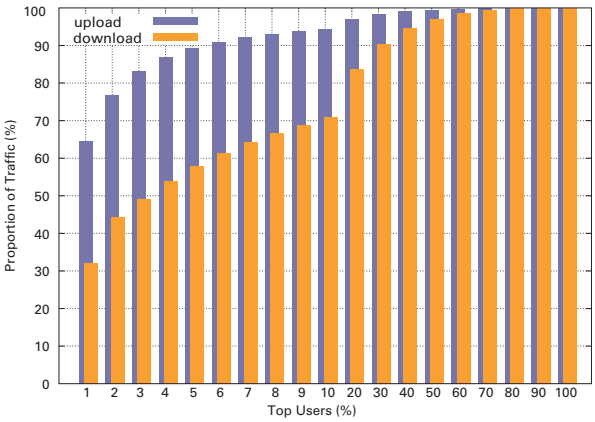

Figure 5: Traffic Usage Deviation between Users

users make up 68% of the total OUT traffic, and 93% of the total IN traffic. Furthermore, the top 1% of users make up 30% of the total OUT traffic, and 65% of the total IN traffic. Along with the decrease in the ratio of heavy users over the past few years, the distribution bias is also dropping slightly.

2.4 Usage by Port Overview

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, it is likely to be 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 2013 and 2014. Table 2 shows detailed numeric values for this figure.

80% of traffic in 2014 is TCP based. The ratio of port 80 HTTP traffic was up slightly from 43% in 2013 to 45% this year. The ratio of port 443 HTTPS traffic has also climbed from 4% to 9%. TCP dynamic ports, which have been on the decline, fell from 30% in 2013 to 24% in 2014. 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%. Almost all traffic other than TCP is related to VPN.

Looking exclusively at client-type users, port 80 traffic that accounted for 82% of the total in 2013 has fallen for the first time, dropping to 75% in 2014. Instead, the ratio for port 443 HTTPS traffic, which is the second highest, rose from 5% in 2013 to 14%. The ratio of dynamic ports also decreased from 9% to 7%.

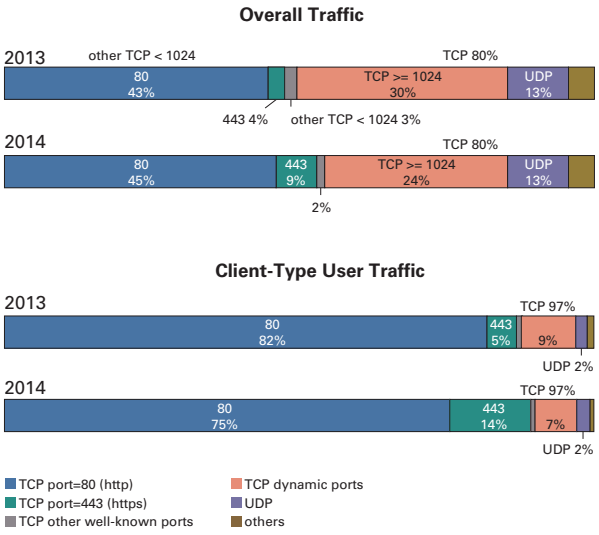

Figure 6: Usage by Port Overview

protocol port	2013		2014	
	total (%)	client type	total (%)	client type
TCP	**79.79**	**96.91**	**80.15**	**97.38**
(< 1024)	49.57	88.15	56.33	90.08
80(http)	43.44	81.61	44.87	74.81
443(https)	3.90	4.80	9.25	13.78
554(rtsp)	0.51	0.58	0.36	0.25
22(ssh)	0.24	0.04	0.31	0.03
(>= 1024)	30.22	8.76	23.82	7.30
1935(rtmp)	2.39	3.60	2.48	4.00
8080	0.34	0.19	0.40	0.17
7144(peercast)	0.40	0.04	0.32	0.02
UDP	**13.21**	**2.12**	**12.51**	**1.81**
ESP	**6.54**	**0.88**	**6.86**	**0.74**
IP-ENCAP	**0.13**	**0.00**	**0.24**	**0.00**
GRE	**0.20**	**0.06**	**0.20**	**0.04**
ICMP	**0.02**	**0.02**	**0.02**	**0.02**
IPv6	**0.01**	**0.01**	**0.01**	**0.00**
L2TP	**0.09**	**0.00**	**0.00**	**0.00**

Table 2: Usage by Port Details

The growth in the use of HTTPS is due to more and more services making regular use of encrypted HTTPS communications since the existence of a U.S. National Security Agency (NSA) program for intercepting communications stirred up controversy in June 2013. Looking at HTTPS traffic volumes broken down by provider for 2014, 59% of the total (67% when isolating client-type users) is related to Google, demonstrating their efforts to proactively adopt HTTPS. Other companies such as Akamai, Amazon, Facebook, Microsoft, and Twitter have followed suit, and the use of HTTPS is expected to continue to grow in the future.

Figure 7 compares trends in TCP port usage over a week for overall traffic in 2013 and 2014. 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 2013, we can see that the overall ratio of port 80 usage has increased further, and the use of dynamic ports is decreasing. 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.

In the same way, Figure 8 and Figure 9 compare weekly TCP port usage trends for client-type and peer-type users in 2013 and 2014. Most client-type user traffic is port 80, but the ratio of other well-known ports including the port for HTTPS is increasing. Peak times are between 21:00 and 23:00. Additionally, for peer-type users, the ratio of port 80 traffic exceeded the ratio for dynamic ports for the first time.

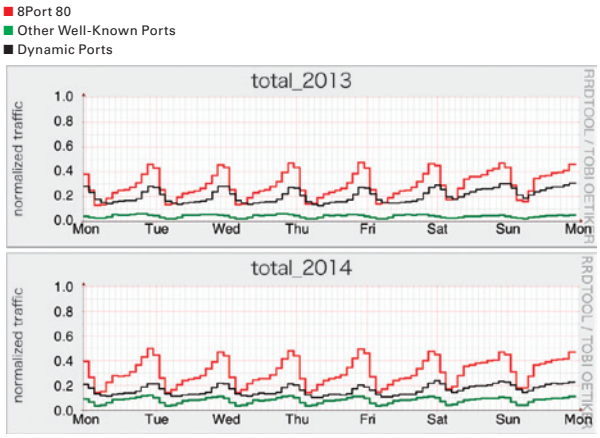

Figure 7: Weekly TCP Port Usage Trends 2013 (top) and 2014 (bottom)

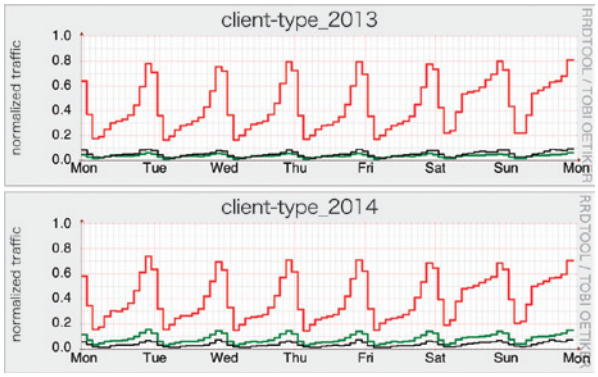

Figure 8: Weekly TCP Port Usage Trends for Client-Type Users 2013 (top) and 2014 (bottom)

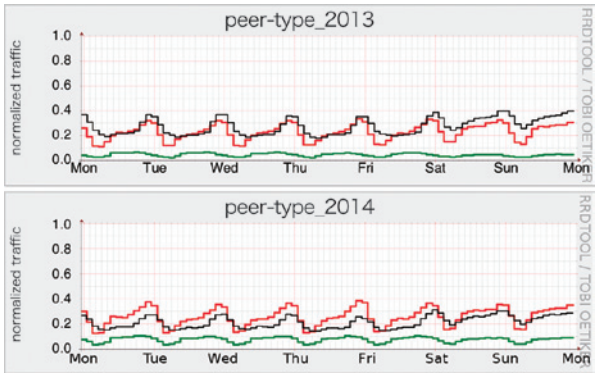

Figure 9: Weekly TCP Port Usage Trends for Peer-Type Users 2013 (top) and 2014 (bottom)

2.5 Conclusion

These results demonstrate that there were no large changes in the overall trends for broadband traffic over the past year. Traffic volumes are growing steadily, with overall download volumes up 27%, and uploads also increasing 13%.

For Web traffic, which now makes up the majority of traffic, we saw that the use of HTTPS is increasing. Google currently seems to be leading the pack in this regard, but other companies are also starting to make regular use of HTTPS in response to increased awareness of the importance of protecting privacy, and we expect the ratio of HTTPS traffic to continue to grow in the coming years.

In other news, NTT East finally launched its 1 Gbps broadband services in July 2014. NTT West has provided 1 Gbps services since 2010. As of now, the traffic for users of 1 Gbps services in Western Japan does not seem to differ significantly from other users, so there may not be many users that require a gigabit of bandwidth yet. However, the same was said when 100 Mbps fiber-optic access services began in 2001. I believe that broadband services are now facing a very important turning point, 13 years after fiber-optic access services were introduced.

Regarding protocols, HTTP/2 is currently in the planning phase, with major revisions expected to be made for the first time in 15 years. HTTP/2 will incorporate upgrades such as improved performance and more efficient use of network resources. As these kinds of generational shifts in infrastructure and protocols progress we will see a new environment emerge, paving the way for next-generation applications and services. That is why all eyes will be on the upcoming adoption of 1 Gbps services, and the resulting traffic increases and shifts in content. IIJ plans to continue to observe traffic on an ongoing basis, and provide periodic reports.

Author:

Kenjiro Cho
Research Director, Research Laboratory, IIJ Innovation Institute Inc.