The Impact of Criminalization of Illegal Downloads was Limited

To examine the impact of legal enforcement measures, we had paid close attention to the impact of the criminalization of illegal downloads on October 1, 2012.

Traffic decreased for about three months following this change coming into effect, but after that the trend curves returned to previous levels, so it appears the criminalization of illegal downloads had only a temporary psychological effect.

3.1 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. Here 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 for broadband as a whole over the past six years. The drop in traffic in January 2010 is believed to be 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. After the Great East Japan Earthquake in March 2011, a decrease in traffic was observed in affected prefectures due to damage to equipment and circuits, as well as power outages. However, the overall impact on a nationwide level was minimal. In October 2012, an amended Copyright Act that incorporated criminal punishment for illegal downloads came into effect. Fluctuations in traffic were observed around this time, but impact was limited, and there was no change in long-term trends, differing from when the download of copyright infringing content was made illegal in 2010. Over the past year IN traffic has increased by 8%, while OUT traffic has increased by 16%.

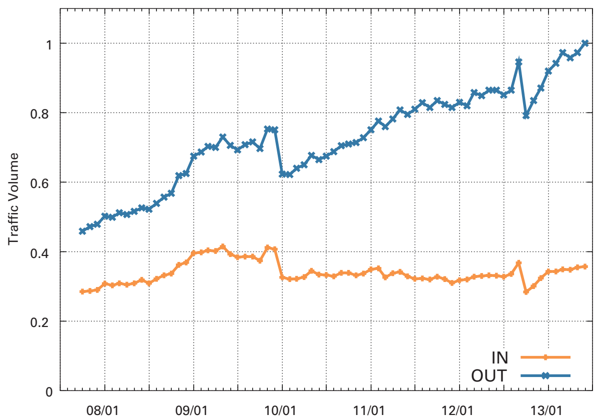

Figure 1: Broadband Traffic Volume Trends for the Past 6 Years

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. This time we compare data for the week spanning June 3 to June 9, 2013 with the data we analyzed in the previous report for the week spanning May 28 to June 3, 2012.

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

*1 Kenjiro Cho. Broadband Traffic Report: Traffic Trends over the Past Year. Internet Infrastructure Review. Vol.16. pp33-37. August 2012.

*2 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.

*3 Kenjiro Cho. Broadband Traffic Report: Traffic Shifting away from P2P File Sharing to Web Services. Internet Infrastructure Review. Vol.8. pp25-30. August 2010.

*4 Kenjiro Cho. Broadband Traffic: Increasing Traffic for General Users. Internet Infrastructure Review. Vol.4. pp 18-23. August 2009.

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 at 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 93% of users observed in 2013 having a fiber-optic connection, and these connections accounting for 96% 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.

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 2012 and 2013 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 scope of 100 GB (10^{11}).

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 2012 and 2013, the peak distribution for both IN and OUT traffic has moved slightly to the right, demonstrating that overall user traffic volumes are increasing.

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. The peak for peer-type users has grown smaller over the past few years, and this indicates that the ratio of heavy users is decreasing. Small spikes appear on the left side of the graph, but these are 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 2012 and in 2013, IN rose from 14 MB to 18 MB, and OUT rose from 282 MB to 355 MB. This demonstrates that, particularly in the case of downloads, the traffic volume for each user has increased. Meanwhile, because average

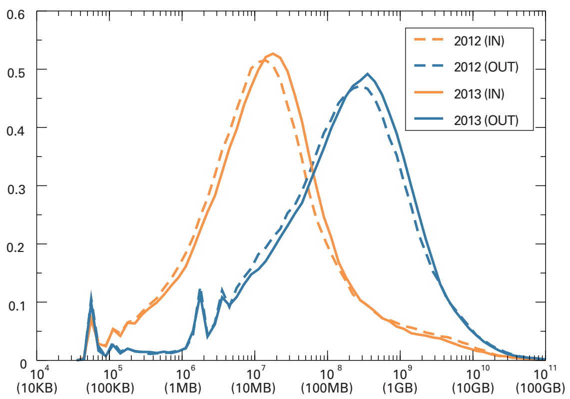

Figure 2: Daily User Traffic Volume Distribution Comparison of 2012 and 2013

	IN (MB/day)		OUT (MB/day)	
Year	Most Frequent Value	Average Value	Most Frequent Value	Average 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

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

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 397 MB and the average OUT value 1,038 MB in 2013. The average values for 2012 were 410 MB and 1,026 MB, respectively, indicating the trend of average IN value decreasing and average OUT value increasing remains unchanged.

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 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 2012.

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 2012, the tail to the right of the graph has extended slightly further to the right. Additionally, the number of extremely heavy users off to the right of the straight line, which fell last year, was once again observed. In any case, it can be said that heavy users are distributed statistically, and are by no means a special class of user.

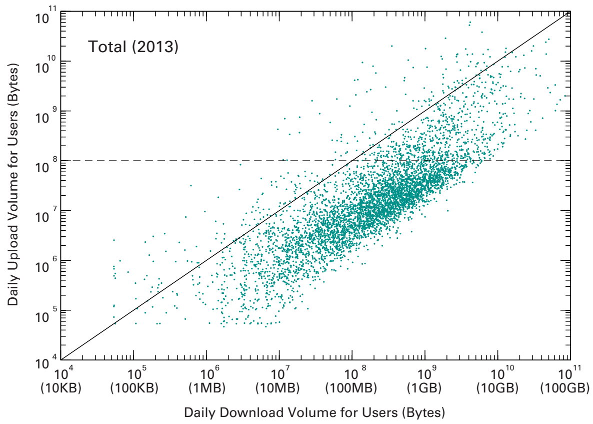

Figure 3: IN/OUT Usage for Each User

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 71% of the total OUT traffic, and 94% of the total IN traffic. Furthermore, the top 1% of users make up 32% of the total OUT traffic, and 65% of the total IN traffic.

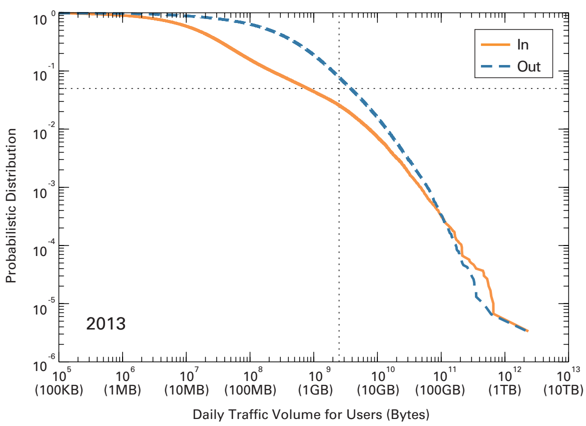

Figure 4: Complementary Cumulative Distribution of the Daily Traffic Volume for Users

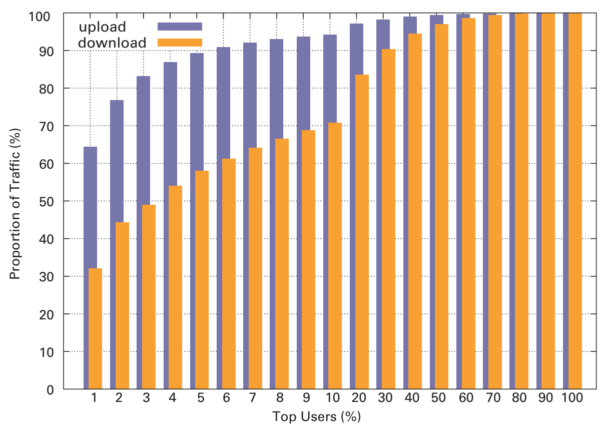

Figure 5: Traffic Usage Deviation Between Users

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 2012 and 2013. Table 2 shows detailed numeric values for this figure.

80% of traffic in 2013 is TCP based. Furthermore, looking at overall traffic, TCP dynamic ports that accounted for 41% of the total in 2012 have dropped to 30% in 2013. 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 36% in 2012 to 43% in 2013. Most non-TCP traffic is related to VPN, and this is on the rise.

Looking exclusively at client-type users, port 80 traffic that accounted for 79% of the total in 2012 has increased to 82% in 2013. The next highest is port 443 used for HTTPS, which increased from 3% in 2012 to 5%. In contrast, the ratio of dynamic ports decreased from 10% to 9%.

From this data, we can confirm that the upward trend in TCP port 80 traffic continues to affect heavy users as well as general 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.

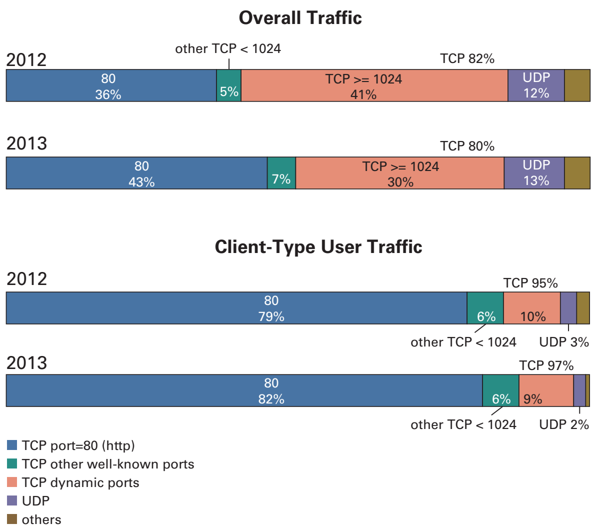

Figure 6: Usage Level Overview by Port

protocol port	2012 total (%)	2012 client type	2013 total (%)	2013 client type
TCP	**81.86**	**95.09**	**79.79**	**96.91**
(<1024)	41.23	85.25	49.57	88.15
80 (http)	36.22	79.39	43.44	81.61
443 (https)	2.45	3.43	3.90	4.80
554 (rtsp)	0.77	1.01	0.51	0.58
22(ssh)	0.22	0.06	0.24	0.04
(>=1024)	40.63	9.84	30.22	8.76
1935 (rtmp)	2.12	3.91	2.39	3.60
7144 (peercast)	0.44	0.04	0.40	0.04
8080	0.30	0.17	0.34	0.19
UDP	**12.38**	**2.94**	**13.21**	**2.12**
ESP	**5.29**	**1.79**	**6.54**	**0.88**
GRE	**0.16**	**0.14**	**0.20**	**0.06**
IP-ENCAP	**0.09**	**0.01**	**0.13**	**0.00**
L2TP	**0.14**	**0.00**	**0.09**	**0.00**

Table 2: Usage Level Details by Port

Figure 7 compares trends in TCP port usage over a week for overall traffic in 2012 and 2013. 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 2012, we can see that the overall ratio of port 80 usage has increased further, and is now growing larger than 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.

In the same way, Figure 8 and Figure 9 compare weekly TCP port usage trends for client-type and peer-type users in 2012 and 2013. Port 80 usage accounts for the vast majority among client-type users, with peak hours between 21:00 and 23:00, showing very little change from last year. Meanwhile, for peer-type users, the usage ratio for dynamic ports has decreased, with dynamic port figures just slightly higher than those for port 80 usage in 2013.

■ Port 80
■ Other Well-Known Ports
■ Dynamic Ports

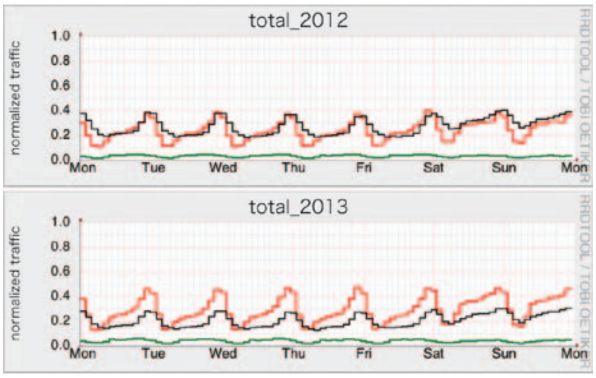

Figure 7: Weekly TCP Port Usage Trends 2012 (top) and 2013 (bottom)

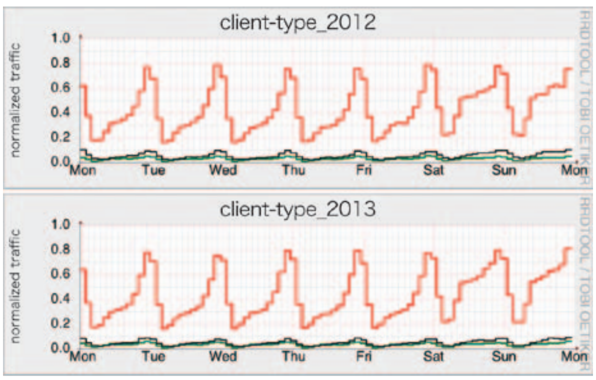

Figure 8: Weekly TCP Port Usage Trends for Client-Type Users 2012 (top) and 2013 (bottom)

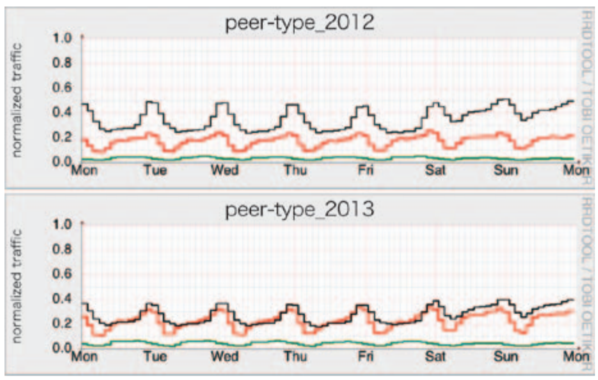

Figure 9: Weekly TCP Port Usage Trends for Peer-Type Users 2012 (top) and 2013 (bottom)

3.5 Conclusion

These results demonstrate that despite fluctuations around October 2012, there were no significant changes in the overall trend for broadband traffic over the past year. All in all, download volumes increased by 16%. Upload volumes also rose by 8%, breaking the flat trend seen since 2010. Additionally, with the ratio of TCP port 80 traffic showing further gains, we can confirm that the migration to Web services we previously reported has gained even more traction.

The amended Copyright Act incorporating criminal punishment for illegal downloads, which was enacted on October 1, 2012, had limited impact. When the download of copyright infringing content was made illegal in January 2010, it had a clear impact on long-term traffic trends. However, as discussed in a previous report, this could also be seen as merely a trigger that accelerated an existing trend. To examine the impact of legal enforcement measures, we paid close attention to the impact of the incorporation of criminal punishment for illegal downloads.

As it turned out, this time there was an increase in traffic thought to represent last-minute downloads before the amendments came into effect. Traffic was lower for three months after the changes became law, but then the trend curve returned to previous levels. It is not possible to calculate fluctuations in illegal downloads simply from macro traffic trends, but it is reasonable to assume that the fluctuations before and after enactment of the amended Copyright Law reflect the illegal download behavior of some users. However, considering the trend curve returned to previous levels in three months, and no significant changes in trends seen in other figures, it appears that the introduction of criminal punishment for illegal downloads had only a temporary psychological effect.

One aspect of the role of legislation is to navigate a social change in which user behavior is shifting according to technological and social developments. With this in mind, we believe the amendments to the Copyright Act in 2010 had an effect because it paralleled the trend of migration to Web services, and was also accepted by the target users. Meanwhile, the amendments in 2012 do not seem to have been accepted by the target users.

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

Author:

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