Examining the Impact of the Earthquake on Traffic on a Macro Level

The Great East Japan Earthquake in March 2011 also caused extensive damage to communications infrastructure. However, when examining this on a macro level, the earthquake had only a limited impact on traffic. Here we report on the impact that the Great Earth Japan Earthquake had on traffic, and then go on to investigate the changes in traffic compared to 2010 using traffic and port usage levels for the week starting May 30, 2011.

3.1 Introduction

In this report we analyze traffic over the broadband access services operated by IIJ and present the results. Last year, in IIR Vol.8, we reported that the existing trend of migration from P2P file sharing applications to web services had also spread to heavy users as a result of the amended Copyright Act that made the download of copyright infringing content illegal. In this report we will first examine the impact of the Great East Japan Earthquake on broadband traffic. Following this, we will present our regular report on changes in traffic over the past year based on daily user traffic and usage by port.

3.2 The Impact of the Earthquake

The Great East Japan Earthquake that struck on March 11 2011 also caused extensive damage to communications infrastructure, severing lines and damaging facilities and equipment. However, the impact on the Internet in Japan as a whole was limited. Because communications over land lines and mobile phones were difficult immediately after the earthquake, the Internet played a significant role in the exchange of information.

In this report we will first take a look at the impact of the earthquake on broadband traffic. Figure 1 shows traffic for the month of March 2011 for Miyagi prefecture and the whole country on the broadband access service operated by IIJ. Because we cannot disclose absolute values for traffic volumes on the Y axis, we have presented data normalized by the peak value.

There was very little damage to equipment at IIJ's Sendai data center that provides services to the Touhoku district, but the redundant configuration of backbone lines that link Tokyo and Sendai were both severed. However, one of these backbone lines was restored to service by the morning of March 12, the day after the quake. At this point most areas in Miyagi prefecture were without power, and although the Sendai data center was operating using in-house power generation, there was almost no traffic. Traffic slowly began to return to normal as power and communication lines were repaired, recovering to 85% 10 days after the earthquake struck, after which it slowed down. The transmission of power from Tohoku Electric Power Co., Inc. and communications services by NTT East also recovered quickly to 90% of normal 10 days after the quake, with service restored to all but a few areas by the end of April.

Looking at traffic on a nationwide level, it fell by about 20% immediately after the earthquake. It recovered slightly overnight, but traffic was affected by power outages on the day of the quake, and large numbers of people being stranded in the Tokyo area without a means to return home. On the following day, a Saturday, traffic had recovered

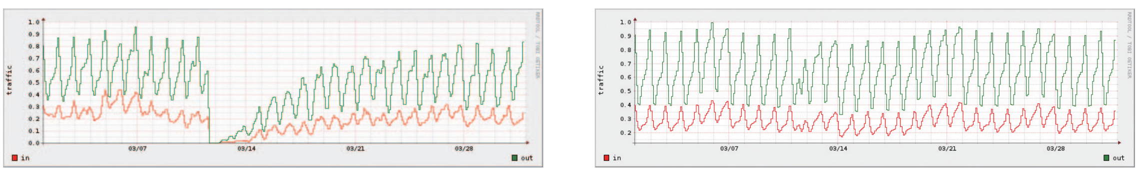

Figure 1: Broadband Traffic in March 2011 (Left: Miyagi prefecture, Right: Nationwide)

to about 85% of the previous Saturday. For a period of about a week following this, the impact of rolling blackouts caused traffic to drop by a few percent, but after this it recovered to levels similar to before the quake. We hear that other ISPs experienced similar situations, so these circumstances were not specific to the IIJ broadband access service.

Figure 2 shows average monthly traffic for broadband as a whole over the past four years. This also demonstrates that the impact of the earthquake was limited. It could be said that the impact was less than that of the amended Copyright Act we reported on last year*1 that came into effect in January 2010, making the download of copyright infringing content illegal. Since January 2010 OUT (download) has continued to rise while IN (upload) has remained flat, so it would seem that use of file sharing has declined. While it is not clear to what extent this has affected traffic, there have also been reports of more users sharing information about the earthquake over P2P networks, making use of its resistance to failures in such circumstances.

There are several reasons that the impact on traffic was limited. First, because there were no major facilities in the areas hit most hard by the earthquake and tsunami, critical services escaped mostly unscathed. This refers to the impact when examined on a macro level, of course. At the time of writing, now four months after the earthquake, service has still not been restored to many of the affected areas. Although many lines have been severed in the frequent aftershocks since March 11, none have caused major disruption thanks to swift on-site restoration work. The impact of the earthquake was kept to a minimum through the tireless efforts of those working to restore services.

There was also little impact on traffic due to power shortages or energy saving measures after the earthquake. Initially it was thought that power shortages would be limited to the service areas of Tohoku Electric Power Company and Tokyo Electric Power Company. That said, these areas account for over a half of the total broadband traffic. However, in reality the reduced usage due to power saving measures was offset by increased use of the Internet to search for information. It is likely that few users cut back on their use of the Internet for power savings in the first place. Many organizations took steps such as shutting down as many in-house servers and PCs as possible. However, even if internal traffic dropped significantly in this case, it is difficult to cut back on the inter-organization traffic seen by providers like us due to the need to deal with clients and customers. Furthermore, due to the earthquake and introduction of rolling blackouts, there was an increase in the migration of work and home servers to cloud services and backup to remote locations, and this may have boosted traffic levels.

Examining the situation on a macro level like this, traffic does not appear to have changed significantly due to the earthquake. In the following section we will use data from the end of May 2010 to analyze changes in traffic trends in more detail.

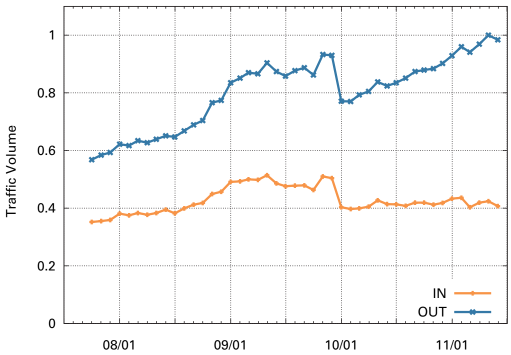

Figure 2: Broadband Traffic Trends for the Past 4 Years (Normalized to the OUT value for May 2011)

*1 "Broadband Traffic Report: Traffic Shifting away from P2P File Sharing to Web Services" (Kenjiro Cho, Internet Infrastructure Review Vol.8 pp. 25 to 30) http://www.iij.ad.jp/development/iir/pdf/iir_vol08_report.pdf

3.3 About the Data

Similar to the previous report, the survey data utilized here was collected using Sampled NetFlow from the routers accommodating fiber-optic and DSL users of our personal and enterprise broadband access services. Because broadband traffic trends differ between weekdays and weekends, we analyzed a full week of data. In this case, we used data for the week spanning May 30 to June 5, 2011. For comparison, we used the data we analyzed in the previous report for the week spanning May 24 to May 30, 2010.

The usage volumes of each user were obtained by matching the IP address assigned to each user with the IP addresses observed. We collected statistical information by sampling packets using Sampled 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. Because this kind of sampling method was used, there is a risk of slight estimation errors in data for low-volume users. However, for users with reasonable usage levels we were able to obtain statistically meaningful data.

Over the past few years the migration from DSL to fiber-optic connections has continued, with 88% of users observed in 2011 having a fiber-optic connection, and these connections accounting for 93% 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.

3.4 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 by dividing each user's data for the period of a week by seven.

Figure 3 shows the average daily usage distribution (probability density function) per user. This is 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 left side of Figure 3 shows a comparison between 2010 and 2011, while the right side shows a comparison between 2005 and 2011. The X axis indicates volumes between 10^4 (10 KB) and 10^{11} (100 GB) using a logarithmic scale. The traffic volume for users with the highest usage levels climbed to 1.2 TB, and some users are outside the scope of the graph, but most fall within the scope of 10^{11} (100 GB). A slight spike appears on the left side of the graph, but this is just noise caused by the sampling rate.

The distribution for IN (upload) and OUT (download) 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.

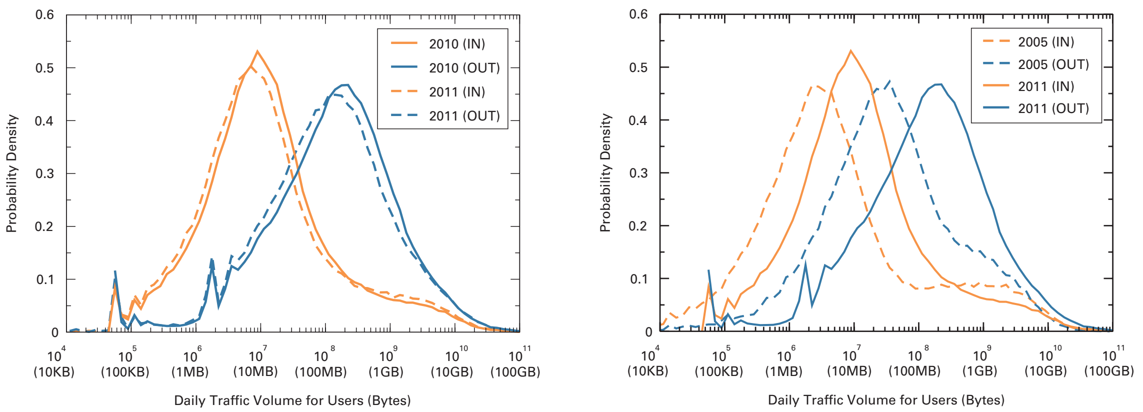

Figure 3: Daily User Traffic Volume Distribution (Left side: 2010 and 2011 comparison, Right side: 2005 and 2011 comparison)

Looking at the right end of the IN (upload) distribution, you will notice another small peak in the distribution. In fact, a similar peak can be seen on the OUT (download) side, overlapping with the main distribution. These distributions have IN and OUT traffic at about the same position, indicating heavy users with symmetrical IN/OUT traffic. For convenience, we will call the asymmetrical IN/OUT traffic distribution that makes up the vast majority "client-type users," and the distribution of heavy users with symmetrical IN/OUT traffic making up a minority on the right side "peer-type users."

Table 1 shows shifts in the average traffic volume and most frequent value that represents peak distribution. The average traffic is affected by the usage of heavy users in the graph to the right, so the average for 2011 was 432 MB for IN (upload) and 1,001 MB for OUT (download). In 2010 these averages were 469 MB and 910 MB, respectively, indicating a drop for IN and an increase for OUT.

Comparing the most frequent values for client-type users in 2010 and in 2011, IN (upload) volume rose from 7 MB to 8.5 MB, and OUT (download) volume rose from 145 MB to 223 MB. This demonstrates that, particularly in the case of downloads, the traffic volume has increased dramatically.

Looking at the right side of Figure 3, which compares values for 2005 and 2011, we can see that usage levels for general users have increased steadily, while those for heavy users that account for the majority of traffic are flat, and make up a smaller percentage of the total.

Figure 4 plots the IN/OUT volumes for 5,000 randomly sampled users. The X axis shows OUT (download) and the Y axis IN (upload), expressed as a log-log graph. When the IN/OUT traffic volumes for a user are identical, they are plotted on the diagonal line in the graph.

Two clusters can be seen here. The cluster below the diagonal line and spread out parallel to it is client-type users with OUT (download) volumes an order of magnitude higher than their IN (upload) volumes. Meanwhile, the cluster spread out thinly above the diagonal line in the upper right is peer-type users. However, the boundary between these two clusters is ambiguous. This is because client-type users also use peer-type applications such as Skype, and peer-type users also use download-based applications on the web. In other words, many users use both types of applications in varying ratios. There are also significant differences in the usage levels and IN/OUT ratio for each user, pointing to the existence of diverse forms of usage. These trends showed almost no change compared to those for 2010.

3.5 Usage by Port

Next, we will look at a breakdown of traffic from the perspective of 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 use port 80 assigned for HTTP to avoid firewalls. To broadly categorize, when both parties

	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

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

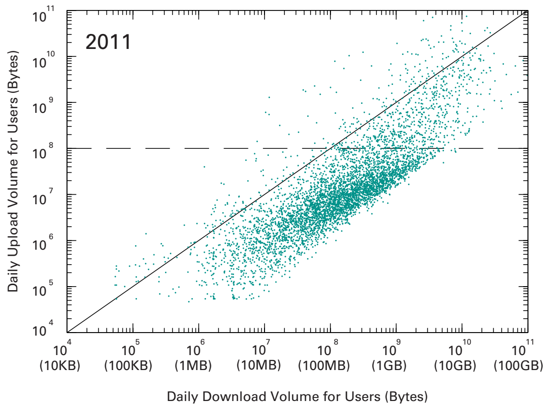

Figure 4: IN/OUT Usage for Each User

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.

Overall traffic is greatly affected by peer-type heavy user traffic. Because of this, to examine trends for client-type 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 the intermediate point between the two IN (upload) distributions in Figure 3, and users below the horizontal line at the IN = 10^8 (100 MB) point in Figure 4.

Figure 5 shows an overview of port usage, comparing all users and client-type users for 2010 and 2011. Table 2 shows detailed numeric values for this figure.

86% of traffic in 2011 is TCP based. However, looking at overall traffic, although the ratio of TCP dynamic ports higher than 1024 was 64% of the total volume in 2010, this has decreased to 50% in 2011. Individual ports in the dynamic port range make up only a small percentage, comprising just 1% of the total traffic at most. Meanwhile, the use of port 80 has increased from 23% in 2010 to 32% in 2011.

Looking at client-type users, port 80 traffic that accounted for 75% of the total in 2010 has decreased to 67% in 2011. Port 554, which accounted for the second largest portion, was about 7% of the total before 2009, and fell to 2% in 2010 before recovering to 7% again in the latest figures. This port is used in RTSP (Real-Time Streaming Protocol), and relates to the increase in video content. Meanwhile, the ratio of dynamic ports has decreased from 15% to 11%.

From this data, we can see that the trend of increased TCP traffic over port 80 spreading from general users to heavy users that we reported last year is still continuing. 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 shows weekly trends in TCP port usage for overall traffic. TCP port usage is divided into three categories: port 80, other well-known ports, and dynamic ports. Traffic is normalized by the total peak traffic volume. The proportion of overall traffic using dynamic ports remains large, peaking between 11:00 PM and 1:00 AM. Traffic also increases in the daytime on Saturday and Sunday, reflecting times when the Internet is used at home. However, comparing the data for 2010 and 2011, in 2011 the port 80 ratio has increased to a point where it now rivals the ratio for dynamic ports.

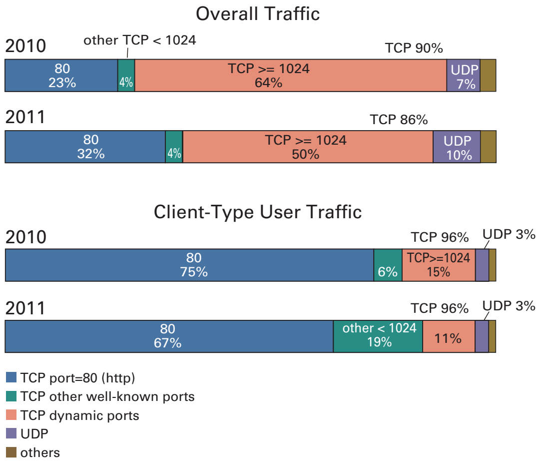

Figure 5: Usage by Port

protocol port	2010 total (%)	2010 client type	2011 total (%)	2011 client type
TCP *	90.09	95.82	85.95	96.28
(<1024)	26.46	80.87	36.24	85.69
80 (http)	23.00	75.12	32.10	67.30
443 (https)	0.98	2.28	1.33	1.91
554 (rtsp)	1.15	2.45	1.33	6.89
22 (ssh)	0.14	0.10	0.27	0.17
(>=1024)	63.63	14.95	49.71	10.59
1935 (rtmp)	1.04	2.91	1.58	1.51
6346 (gnutella)	0.86	0.33	0.68	0.60
6699 (winmx)	0.65	0.17	0.40	0.24
7144 (peercast)	0.34	0.04	0.38	0.00
UDP	6.79	2.76	10.01	2.61
ESP	2.91	1.30	3.56	1.02
GRE	0.14	0.06	0.15	0.05
L2TP	0.00	0.00	0.13	0.00
ICMP	0.02	0.04	0.10	0.04

Table 2: Usage Level Details by Port

Figure 7, similarly to Figure 6, shows weekly trends in TCP port usage by client-type users. In 2011 the ratio for port 80 has increased above the 2010 level. The peak hours are between 9:00 PM and 11:00 PM, which is slightly earlier than those in Figure 6, and usage is on the rise from morning on Saturday and Sunday.

3.6 Conclusion

In the previous report we noted that the migration of general users from P2P file sharing applications to web services that was until then only evident among general users was spreading to heavy users. We also showed that usage levels for general users were steadily increasing due to video content and web 2.0 rich content. In the current survey there were no changes in these trends, and we identified that the migration to web services is accelerating.

In this report we also noted that the impact of the Great East Japan Earthquake was limited when examined on a macro level. The fact that there was little impact on Internet traffic even in a disaster that had a significant effect on society demonstrates that the Internet is becoming an indispensable infrastructure in our daily lives.

Incidentally, I expected that the earthquake would cause a larger drop in traffic. This is because I believed that more users would shut down the home servers they use for activities such as P2P file sharing because of rolling blackouts and energy saving measures.

There are two ways of looking at the reasons behind the fact that there was less impact on traffic compared to the introduction of the amended Copyright Act in January 2010. One is that the activity of some heavy users will not change unless more enforcement measures such as making the download of copyright infringing content illegal are introduced. As mentioned in our previous report, another way of looking at this is that making the download of copyright infringing content illegal was merely a trigger that accelerated a process that was already occurring, and because a shift in usage had already taken place due to this first trigger, the second trigger had a lesser effect. Even if the amended Copyright Act and the earthquake had occurred in reverse order, migration may have occurred in a similar manner due to the first trigger.
If in the near future other enforcement measures are introduced regarding the use of P2P file sharing, we should be able to surmise that the former perspective is correct if traffic drops, or that the latter perspective is correct if it does not.

IIJ monitors traffic levels on an ongoing basis so we can respond promptly to events such as the recent earthquake or changes in user behavior. We will continue to publish reports such as this periodically.

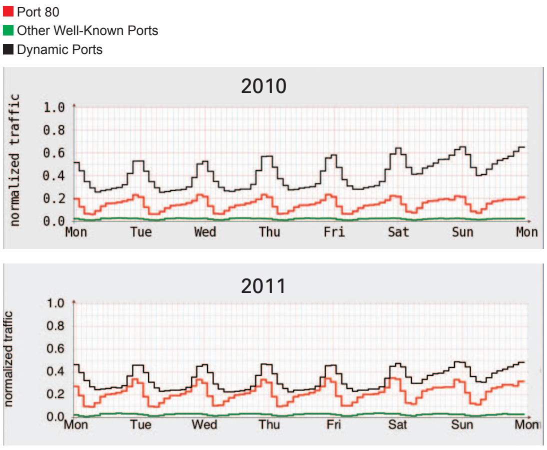

Figure 6: Weekly TCP Port Usage Trends

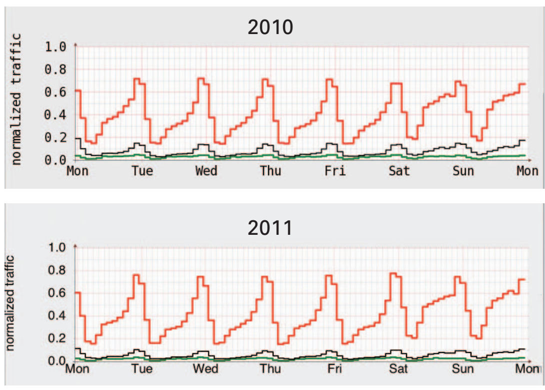

Figure 7: Weekly TCP Port Usage Trends for Client-Type

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