Traffic Growth is Accelerating

2.1 Overview

In this report we analyze traffic over the broadband access services operated by IIJ each year and present the results*1*2*3*4*5*6*7. Here we will 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 trends for IIJ's overall broadband services and mobile services. The IN/OUT traffic indicates directions from an ISP's perspective. IN represents uploads from users, and OUT represents user downloads. Because we cannot disclose traffic numbers, we have normalized the latest values for each set of OUT data to 1. For broadband, over the past year IN traffic has increased by 18%, while OUT traffic has increased by 47%. A year ago each had increased by 15% and 38%, respectively, indicating that significant growth has occurred.

For mobile, we only present figures for the past two years, but over the last year traffic continued to grow, albeit at a slightly slower pace. IN traffic increased by a factor of 2.3, and OUT traffic by a factor of 2.4, while a year earlier they increased by factors of 3.5 and 4.2, respectively. Despite this growth, the total volume of mobile traffic is still an order of magnitude lower than broadband.

2.2 About the Data

For broadband traffic, 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. For mobile traffic, access gateway billing information was applied to determine usage amounts for personal and enterprise mobile services, while Sampled NetFlow data from the routers accommodating these services was employed to determine the ports used. Because traffic trends differ between weekdays and weekends, we analyze a full week of traffic. In this case, we used data for the week spanning May 30 to June 5, 2016. For comparison, we used the data for the week spanning June 1 to June 7, 2015, which we analyzed in the previous report.

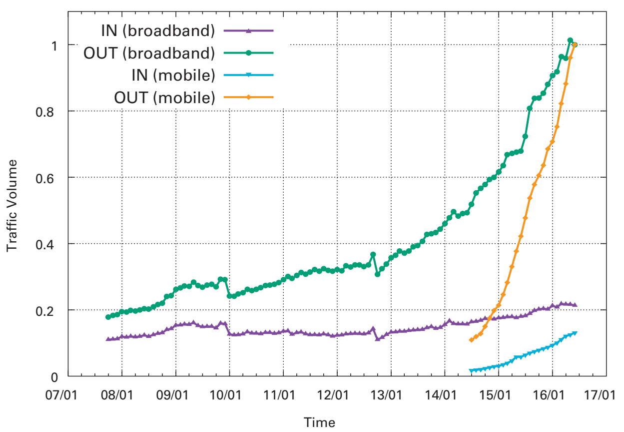

Figure 1: Trends in Monthly Traffic for Broadband and Mobile

Results are aggregated by subscription for broadband traffic, and by phone number for mobile traffic, as some subscriptions cover multiple phone numbers. The usage volume for each broadband user was obtained by matching the IP address assigned to users with the IP addresses observed. We gathered statistical information by sampling packets using NetFlow. The sampling rate was set between 1/8192 and 1/16382, taking into account router performance and load. We estimated overall usage volumes by multiplying observed volumes by the reciprocal of the sampling rate. IIJ provides both fiber-optic and DSL access for its broadband services, but fiber-optic access now accounts for the vast majority of use. 97% of users observed in 2016 used fiber-optic connections, making up 98% of the overall broadband traffic volume.

*1 Kenjiro Cho. Broadband Traffic Report: Comparing Broadband and Mobile Traffic. Internet Infrastructure Review. Vol.28. pp28-33. August 2015.
*2 Kenjiro Cho. Broadband Traffic Report: Traffic Volumes Rise Steadily Over the Past Year, and HTTPS Use Expands. Internet Infrastructure Review. Vol.24. pp28-33. August 2014.
*3 Kenjiro Cho. Broadband Traffic Report: The Impact of Criminalization of Illegal Downloads was Limited. Internet Infrastructure Review. Vol.20. pp32-37. August 2013.
*4 Kenjiro Cho. Broadband Traffic Report: Traffic Trends over the Past Year. Internet Infrastructure Review. Vol.16. pp33-37. August 2012.
*5 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.
*6 Kenjiro Cho. Broadband Traffic Report: Traffic Shifting away from P2P File Sharing to Web Services. Internet Infrastructure Review. Vol.8. pp25-30. August 2010.
*7 Kenjiro Cho. Broadband Traffic: Increasing Traffic for General Users. Internet Infrastructure Review. Vol.4. pp18-23. August 2009.

2.3 Daily Usage Levels for Users

First, we will examine the daily usage volumes for broadband and mobile users from several perspectives. Daily usage indicates the average daily usage calculated from a week's worth of data for each user. Figure 2 and Figure 3 show the average daily usage distribution (probability density function) per broadband and mobile user. They compare data for 2015 and 2016 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 for broadband 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 2015 and 2016, 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 on the right side, 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 IN distribution on the left side demonstrates left-right symmetry, and is closer to log-normal distribution.

The data for mobile traffic in Figure 3 indicates that usage volumes are significantly lower than broadband. Additionally, because there are limits on data usage, the ratio of heavy users to the right of the distribution is lower, creating left-right asymmetry. There are also no extremely heavy users. Due to those who use mobile only when going out, as well as limits on data usage, there is greater variance in daily usage volumes for each user compared to broadband. For this reason, when you look at the daily average for a week worth of data, there is less variance between users than when examining individual days. Plotting distribution for individual days in the same way results in lower peaks, and raises the tails on either side, but the basic form and modal values of the distribution remain largely unchanged.

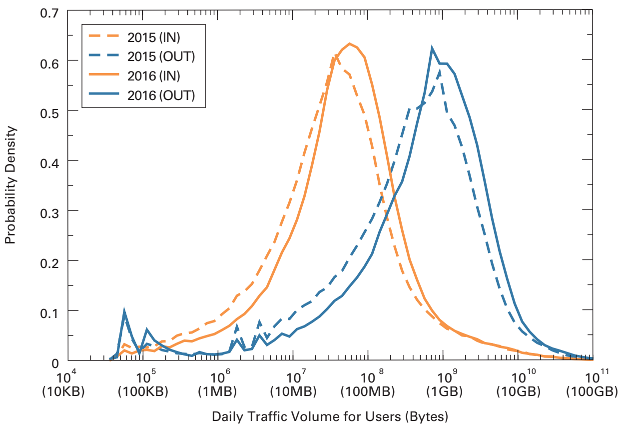

Figure 2: Daily Broadband User Traffic Volume Distribution Comparison of 2015 and 2016

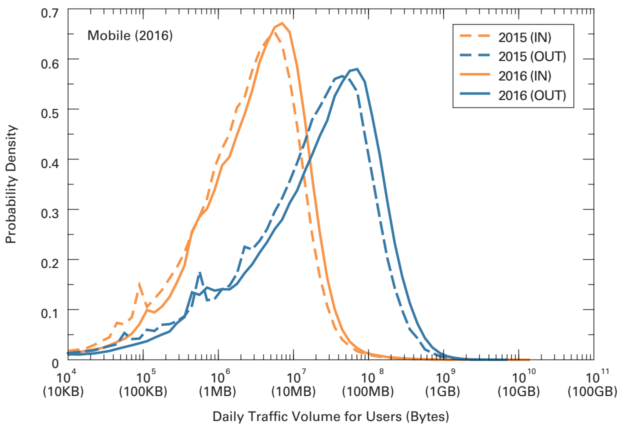

Figure 3: Daily Mobile User Traffic Volume Distribution Comparison of 2015 and 2016

Table 1 shows trends in the mean and median daily traffic values for broadband users as well as the modal value (the most frequent value that represents peak distribution). The peak position has shifted slightly from the center of the distribution, so the modal value was corrected towards the center of the distribution. Comparing the modal values in 2015 and 2016, IN rose from 40 MB to 56 MB, and OUT rose from 708 MB to 1,000 MB. This means that both IN and OUT traffic grew by a factor of 1.4. Meanwhile, because mean values are pulled up by the heavy users to the right of the graph, they are significantly higher than the modal values, with the mean IN value 475 MB and the mean OUT value 2,081 MB in 2016. In 2015, these were 467 MB and 1,621 MB, respectively.

For mobile, as shown in Table 2, the mean and modal values are virtually identical due to the lack of heavy users. In 2016 the modal values were 7 MB for IN and 63 MB for OUT, while the mean values amounted to 7.8 MB for IN and 63 MB for OUT. The modal values grew by a factor of 1.3 for IN traffic, and 1.6 for OUT traffic. Figure 4 and Figure 5 plot 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 users with download volumes an order of magnitude higher than upload volumes. For broadband traffic, there was previously a clearly-recognizable cluster of heavy users spread out thinly on the upper right of the diagonal line, but this is now no longer discernible. There are also differences in the usage levels and IN/OUT ratio for each user, pointing to the existence of diverse forms of usage. Here, almost no difference can be discerned compared to 2015.

	IN (MB/day)			OUT (MB/day)		
Year	Mean	Median	Mode	Mean	Median	Mode
2005	430	3	3.5	447	30	32
2007	433	5	4	712	58	66
2008	483	6	5	797	73	94
2009	556	7	6	971	88	114
2010	469	8	7	910	108	145
2011	432	9	8.5	1,001	142	223
2012	410	12	14	1,026	173	282
2013	397	14	18	1,038	203	355
2014	437	22	28	1,287	301	447
2015	467	33	40	1,621	430	708
2016	475	48	56	2,081	697	1,000

Table 1: Trends in Mean and Modal Values for the Daily Traffic Volume of Broadband Users

The trend for OUT traffic to be an order of magnitude larger also applies to mobile, but usage volumes are lower than broadband, and there is less variance between IN/OUT. The inclination of the cluster is also now lower than the diagonal line, indicating that download ratios are relatively higher for

	IN (MB/day)			OUT (MB/day)		
Year	Mean	Median	Mode	Mean	Median	Mode
2015	6.0	2.7	5.5	46.6	19	40
2016	7.8	3.6	7	63.0	27	63

Table 2: Trends in Mean and Modal Values for the Daily Traffic Volume of Mobile Users

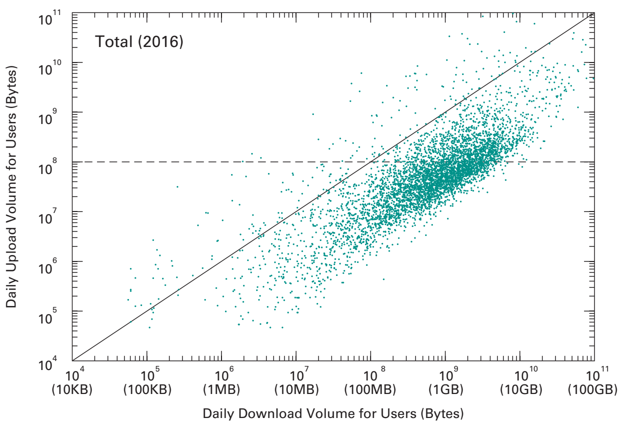

Figure 4: IN/OUT Usage for Each Broadband User

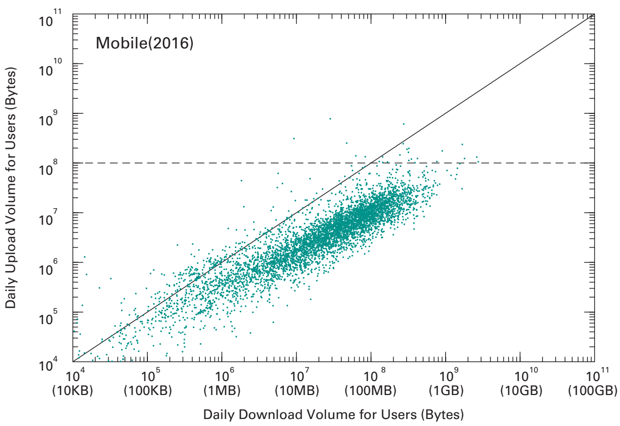

Figure 5: IN/OUT Usage for Each Mobile User

users with higher usage levels. Figure 6 and Figure 7 show 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. It can be said that heavy users are distributed statistically, and are by no means a special class of user.

Heavy users also exhibit power-law distribution for mobile, but the ratio of heavy users is lower. Last year the OUT usage level was also several times higher than the IN levels in the distribution on the right end, but this year IN and OUT are reversed at the tail of the distribution, and there are users that upload large volumes of data. There is a great deal of deviation in traffic usage levels between users, and as a result traffic volume for a small portion of users accounts for the majority of overall traffic. For example, the top 10% of broadband users make up 60% of the total OUT traffic, and 87% of the total IN traffic. Furthermore, the top 1% of users make up 26% of the total OUT traffic, and 60% 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. For mobile, the top 10% of users account for 48% of OUT traffic and 50% of IN traffic, while the top 1% make up 12% of OUT traffic and 21% of IN traffic. This also demonstrates the low ratio of heavy users among mobile users.

2.4 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, 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 heavy user traffic, to examine trends for 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 light users. This constitutes users below the horizontal line IN=100 MB point in Figure 4, which roughly corresponds to mobile user usage levels.

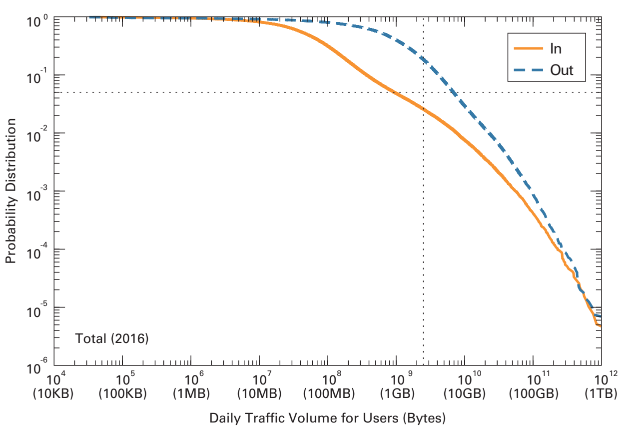

Figure 6: Complementary Cumulative Distribution of the Daily Traffic Volume for Broadband Users

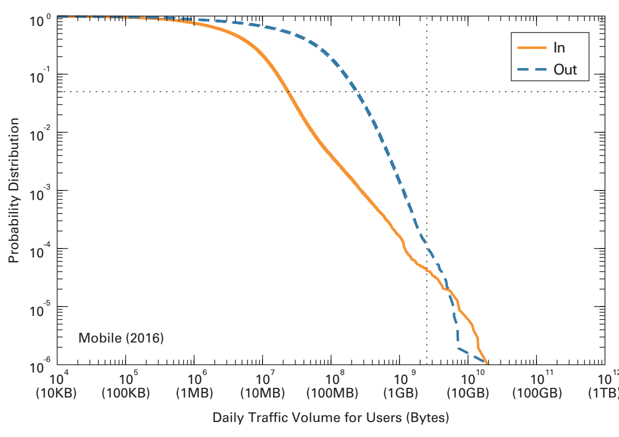

Figure 7: Complementary Cumulative Distribution of the Daily Traffic Volume for Mobile Users

Table 3 compares port usage ratios for broadband users in 2015 and 2016, showing total users and light users. 83% of overall traffic in 2016 is TCP based. The ratio of port 80 (HTTP) traffic was down from 38% in 2015 to 37% this year. The ratio of port 443 (HTTPS) traffic climbed from 23% to 31%. TCP dynamic port traffic, which has been on the decline, dropped from 18% in 2015 to 14% in 2016. 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%. Aside from TCP, there is port 443 (HTTPS) UDP traffic, which is thought to be Google's QUIC protocol. The rest is mostly VPN-related.

Looking exclusively at light users, port 80 traffic that accounted for 53% of the total in 2015 dropped 4 percentage points to 49% in 2016. Meanwhile, the next highest traffic ratio for port 443 (HTTPS) traffic climbed 5 percentage points, going from 35% in 2015 to 40% in 2016. We believe this was due to the shift of some HTTP traffic to HTTPS. The ratio of dynamic ports also decreased from 5% to 3%.

Table 4 shows port usage ratios for mobile users, which as a whole were close to the values for light users on broadband. The growth in the use of HTTPS is due to more and more services mainly based in the United States 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 2016, about 70% was related to Google, demonstrating their efforts to proactively adopt HTTPS. We also believe YouTube traffic volumes are boosting HTTPS usage.

Figure 8 compares trends in TCP port usage over a week for overall broadband traffic in 2015 and 2016. Trends in TCP port usage are shown for four categories: port 80, port 443, other well-known ports, and dynamic ports. Traffic is normalized to 1 for the total peak traffic volume. Compared with 2015, we can see that the overall ratio of port 443 usage has increased further, and the use of dynamic ports is decreasing. The overall peak is between 21:00 and 1:00 the next day, and traffic also increases in the daytime on Saturday and Sunday, reflecting times when the Internet is used at home. This time there was an increase in port 80 traffic on Wednesday morning, which we believe can be attributed to a Microsoft automatic update program. Figure 9 mobile data shows trends for port 80 and port 443, which account for the majority of traffic. Compared to broadband, high traffic levels continue from

protocol port	2015		2016	
	total (%)	light users	total (%)	light users
TCP	**80.8**	**94.7**	**82.8**	**93.3**
(< 1024)	*63.3*	*89.9*	*69.1*	*90.1*
80 (http)	37.9	53.2	37.1	49.2
443 (https)	23.3	35.1	30.5	39.6
81	0.5	0.7	0.4	0.7
182	0.4	0.3	0.3	0.2
22 (ssh)	0.2	0.0	0.2	0.0
110 (pop3)	0.1	0.1	0.1	0.1
(>= 1024)	*17.5*	*4.8*	*13.7*	*3.2*
1935 (rtmp)	1.8	2.4	1.5	1.7
8080	0.3	0.1	0.2	0.1
7144 (peercast)	0.2	0.0	0.1	0.0
UDP	**11.4**	**2.6**	**11.1**	**4.0**
443 (https)	0.9	0.9	2.4	2.8
4500 (nat-t)	0.2	0.1	0.2	0.1
1701 (12tp)	0.2	0.3	0.1	0.1
ESP	**7.4**	**2.6**	**5.8**	**2.6**
IP-ENCAP	**0.2**	**0.0**	**0.2**	**0.0**
GRE	**0.2**	**0.0**	**0.1**	**0.0**
ICMP	**0.0**	**0.0**	**0.0**	**0.0**

Table 3: Usage by Port for Broadband Users

protocol port	2015	2016
	total (%)	total (%)
TCP	**93.8**	**94.4**
80 (http)	52.5	46.8
443 (https)	37.4	43.7
81	0.5	0.5
993 (imaps)	0.5	0.5
1935 (rtmp)	0.5	0.3
UDP	**5.2**	**5.0**
443 (https)	1.0	1.5
1701 (12tp)	1.8	1.0
4500 (nat traveral)	0.3	0.2
53 (dns)	0.1	0.2
ESP	**0.7**	**0.4**
GRE	**0.3**	**0.1**
ICMP	**0.0**	**0.0**

Table 4: Usage by Port for Mobile Users

morning to night. On weekdays there are three peaks representing the morning commute hours, lunch breaks, and evening to night periods, demonstrating that usage times are different from broadband.

2.5 The Impact of the Kumamoto Earthquakes

The Kumamoto earthquakes that occurred in April of this year affected broadband traffic in the prefecture, as shown in Figure 10. Earthquakes with a seismic intensity of 7 occurred during the night of April 14 and at daybreak on April 16, and traffic dropped immediately after they struck. We can see that power outages from the 16th also caused the overall volume to decrease, then recover over a period of about a week. We did not observe such a large impact in neighboring prefectures.

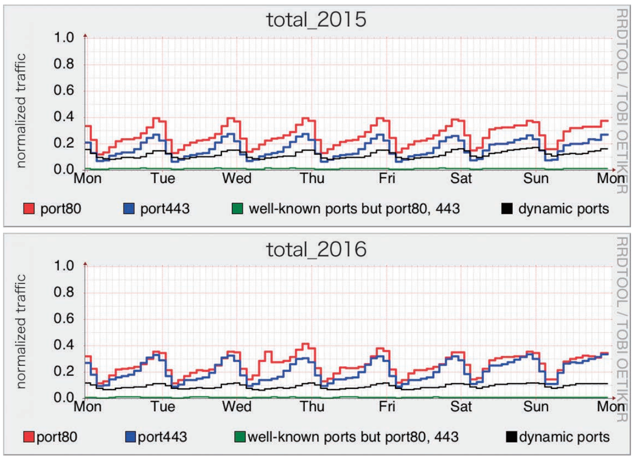

Figure 8: Weekly TCP Port Usage Trends for Broadband Users 2015 (top) and 2016 (bottom)

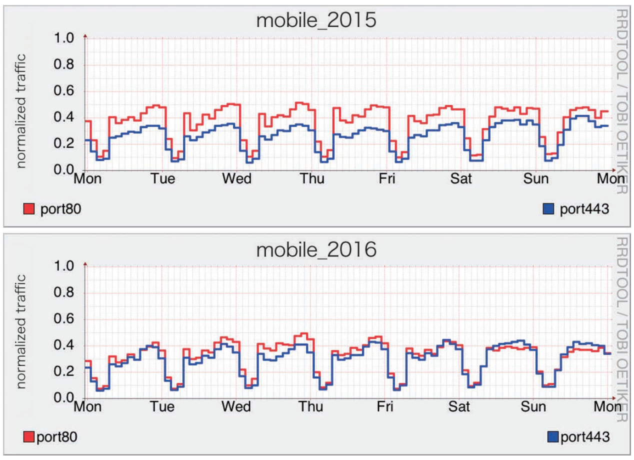

Figure 9: Weekly TCP Port Usage Trends for Mobile Users 2015 (top) and 2016 (bottom)

2.6 Conclusion

One of the broadband traffic trends for this year we can identify is that increases in the volume of traffic are accelerating. Over the last year download volumes climbed 47%, while upload volumes also grew 18%. Last year these increases were 38% and 15%, and the year before that they were 27% and 13%, so growth rates are rising. In addition, the use of HTTPS has expanded greatly over the past two years, with traffic volumes now almost on par with HTTP. Reasons for the increase in traffic include more frequent and larger software updates, and the widespread use of flat-rate music and video streaming services. Another important point is that migration to the FLET'S NGN network is progressing, and this access network infrastructure is supporting increased traffic. The mobile traffic that we added since the last report has also grown significantly over the past two years. This is different from broadband traffic in several ways, such as there being fewer heavy users, and higher levels of usage during weekday commute hours and lunch breaks.

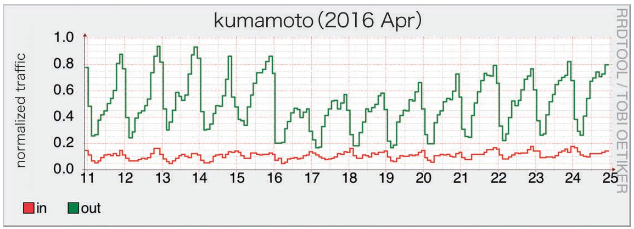

Figure 10: Kumamoto Prefecture Broadband Traffic April 11 to 24, 2016

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