Broadband Traffic Report: Moderate Growth in Traffic Volume Ongoing

1.1 Overview

In this report, we analyze traffic over the broadband access services operated by IIJ and present the results each yearr[*1][*2][*3][*4][*5][*6][*7][*8][*9][*10]. Here, we again report on changes in traffic trends over the past year, based on daily user traffic and usage by port.

Figure 1 shows the overall average monthly traffic trends for IIJ's fixed broadband services and mobile services. IN/OUT indicates the direction from the ISP perspective. IN represents uploads from users, and OUT represents user downloads. Because we cannot disclose specific traffic numbers, we have normalized the data, setting the latest OUT observation in each dataset to 1.

Since the previous edition of this report, the broadband data have included IPv6 IPoE traffic. The thin line labeled "broadband-IPoE" excludes IPv6 IPoE traffic. IPv6 traffic on IIJ's broadband services comprises both IPoE and PPPoE traffic[*11], but IPoE traffic does not pass directly through IIJ's network as we use Internet Multifeed Co.'s transix service for IPoE, and IPoE is therefore excluded from the analysis that follows here. As of June 2019, IPoE accounted for 19% of IN and 14% of OUT broadband traffic overall. Respectively, these are year-on-year increases of 7 and 6 points, so use of IPoE is rising.

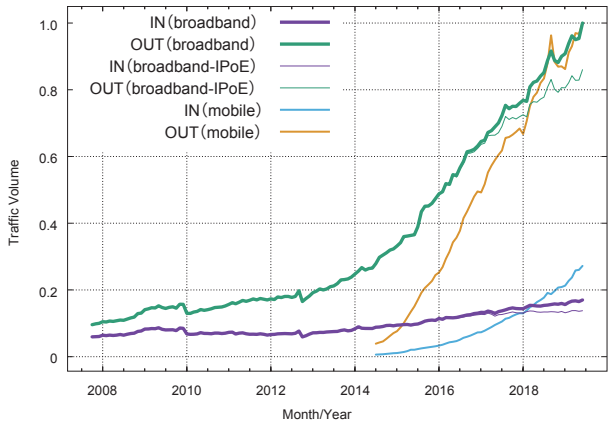

Figure 1: Monthly Broadband and Mobile Traffic over Time

Growth in both broadband and mobile traffic has risen, with some ups and downs, over the past two years or so. These fluctuations in broadband and mobile have been mostly synchronous with each other, suggesting that the underlying factors are the same.

Over the past year, broadband IN traffic increased 12% and broadband OUT traffic increased 19%, virtually the same growth rates as the year-earlier figures of 12% and 20%. Growth in mobile traffic has slowed, IN from 69% a year earlier to 60% this year, and OUT from 36% to 22%. And the total volume of mobile traffic remains an order of magnitude lower than broadband traffic.

1.2 About the Data

As with previous reports, for broadband traffic, our analysis uses data sampled using Sampled NetFlow from the routers that accommodate the fiber-optic and DSL broadband customers of our personal and enterprise broadband access services. For mobile traffic, we use access gateway billing information to determine usage volumes for personal and enterprise mobile services, and we use Sampled NetFlow data from the routers used to accommodate these services to determine the ports used.

Because traffic trends differ between weekdays and weekends, we analyze traffic in one-week chunks. In this report, we look at data for the week of May 27 through June 2, 2019, and compare those data with data for the week of May 28 through June 3, 2018, which we analyzed in the previous edition of this report.

*1 Kenjiro Cho. Broadband Traffic Report: Download Growth Slows for a Second Year Running. Vol.40. pp4-9. September 2018.
*2 Kenjiro Cho. Broadband Traffic Report: Traffic Growth Slows to a Degree. Internet Infrastructure Review. Vol.36. pp4-9. September 2017.
*3 Kenjiro Cho. Broadband Traffic Report: Traffic Growth is Accelerating. Internet Infrastructure Review. Vol.32. pp28-33. August 2016.
*4 Kenjiro Cho. Broadband Traffic Report: Comparing Broadband and Mobile Traffic. Internet Infrastructure Review. Vol.28. pp28-33. August 2015.
*5 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.
*6 Kenjiro Cho. Broadband Traffic Report: The Impact of Criminalization of Illegal Downloads was Limited. Internet Infrastructure Review. Vol.20. pp32-37. August 2013.
*7 Kenjiro Cho. Broadband Traffic Report: Traffic Trends over the Past Year. Internet Infrastructure Review. Vol.16. pp33-37. August 2012.
*8 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.
*9 Kenjiro Cho. Broadband Traffic Report: Traffic Shifting away from P2P File Sharing to Web Services. Vol.8. pp25-30. August 2010.
*10 Kenjiro Cho. Broadband Traffic Report: Increasing Traffic for General Users. Internet Infrastructure Review. Vol.4. pp18-23. August 2009.
*11 Akimichi Ogawa. Appendix A.3 "IPv6 PPPoE and IPv6 IPoE" in Professional IPv6 (in Japanese). Lambda Note. July 2018.

© Internet Initiative Japan Inc.

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. Sampling rates were set between 1/8,192 and 1/16,382, taking into account router performance and load. We estimated overall usage volumes by multiplying observed volumes with the reciprocal of the sampling rate.

IIJ provides both fiber-optic and DSL broadband services, but fiber-optic access now accounts for the vast majority of use. Of users observed in 2019, 98% were using fiber-optic connections and accounted for 99% of overall broadband traffic volume.

1.3 Users' Daily Usage

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

Starting with this edition, we use daily usage data only on services provided to individuals. Enterprise usage varies widely and is readily influenced by the usage patterns of a subset of users, such that the overall distribution is clearly distorted. Individual usage, by contrast, shows a smooth distribution that remains stable. So we determined that using only the individual data to ascertain usage patterns would yield more generally applicable and easily interpretable conclusions. Note that because of the difficulty of distinguishing between individual and enterprise usage, the analysis of usage by port in the next section does include enterprise data.

Figure 2 and Figure 3 show the average daily usage distributions (probability density functions) for broadband and mobile users. Each compares data for 2018 and 2019 split into IN (upload) and OUT (download), with user traffic volume plotted along the X-axis and user frequency along the Y-axis. The X-axis shows volumes between 10KB (104) and 100GB (1011) using a logarithmic scale. Most users fall within the 100GB (1011) range, with a few exceptions.

The IN and OUT broadband traffic distributions are close to a log-normal distribution, which looks like a normal distribution on a semi-log plot. A linear plot would show a long-tailed distribution, with the peak close to the left and a slow gradual decrease toward the right.

The OUT distribution is further to the right than the IN distribution, indicating that download volume is more than an order of magnitude larger than upload volume. The peaks of both the IN and OUT distributions for 2019 are slightly further to the right than the peaks of the 2018 distributions, indicating that overall user traffic volumes are increasing. But that rightward shift in the distribution in 2019 was smaller than it had been in the past few years.

The peak of the OUT distribution, which appears toward the right in the plot, has steadily been moving rightward over

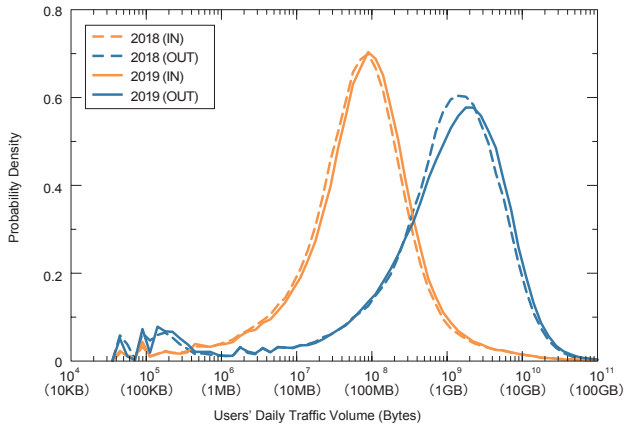

Figure 2: Daily Broadband User Traffic Volume Distribution Comparison of 2018 and 2019

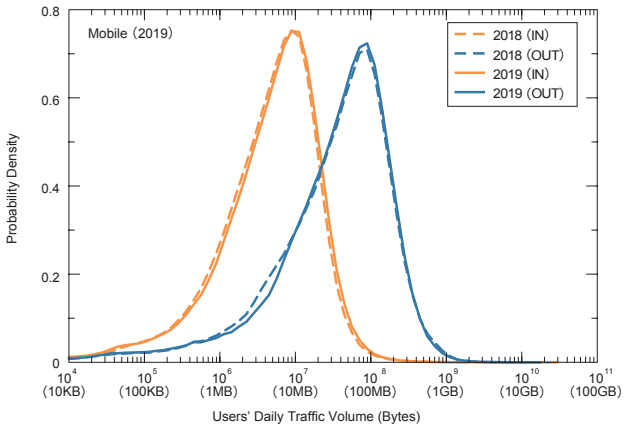

Figure 3: Daily Mobile User Traffic Volume Distribution Comparison of 2018 and 2019

© Internet Initiative Japan Inc.

the past few years, but heavy-user usage levels have not increased much, and as a result, the distribution is becoming less symmetric. The IN distribution on the left, meanwhile, is generally symmetric and closer to a log-normal distribution.

The data for mobile traffic indicate that usage volumes are significantly lower than for broadband. And limits on mobile data usage mean that heavy users, which fall on the right-hand side of the distribution, account for only a small proportion of the total, so the distribution is asymmetric. There are also no extremely heavy users. The variability in each user's daily usage volume is higher for mobile than for broadband owing to there being users who only use mobile data when out of the home/office as well as limits on mobile data. Hence, the daily average for a week's worth of data shows less variability between users than the data for individual days. Plotting the distributions for individual days in the same way results in slightly lower peaks and correspondingly higher tails on both sides, but the basic shape and modal values of the distribution remain largely unchanged. The difference in the mobile distributions versus last year is also minimal.

Table 1 shows trends in the mean and median daily traffic values for broadband users as well as the mode (the most frequent value, which represents the peak of the distribution).

Table 1: Trends in Mean and Mode of Broadband Users' Daily Traffic Volume

	IN (MB/day)			OUT (MB/day)		
Year	Mean	Median	Mode	Mean	Median	Mode
2007	436	5	5	718	59	56
2008	490	6	6	807	75	79
2009	561	6	6	973	91	100
2010	442	7	7	878	111	126
2011	398	9	9	931	144	200
2012	364	11	13	945	176	251
2013	320	13	16	928	208	355
2014	348	21	28	1124	311	501
2015	351	32	45	1399	443	708
2016	361	48	63	1808	726	1000
2017	391	63	79	2285	900	1259
2018	428	66	79	2664	1083	1585
2019	479	75	89	2986	1187	1995

Table 2: Trends in Mean and Mode of Mobile Users' Daily Traffic Volume

	IN (MB/day)			OUT (MB/day)		
Year	Mean	Median	Mode	Mean	Median	Mode
2015	6.2	3.2	4.5	49.2	23.5	44.7
2016	7.6	4.1	7.1	66.5	32.7	63.1
2017	9.3	4.9	7.9	79.9	41.2	79.4
2018	10.5	5.4	8.9	83.8	44.3	79.4
2019	11.2	5.9	8.9	84.9	46.4	79.4

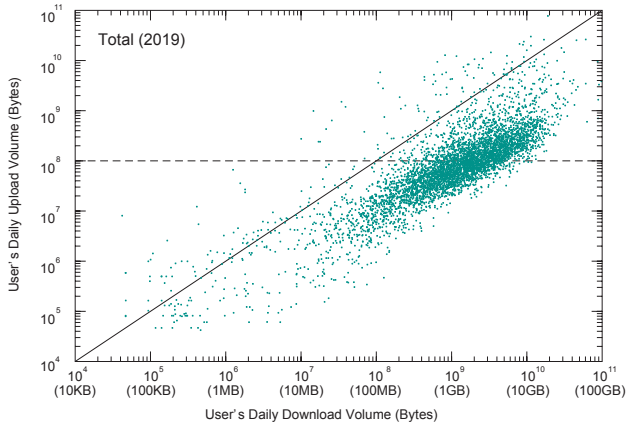

Figure 4: IN/OUT Usage for Each Broadband User

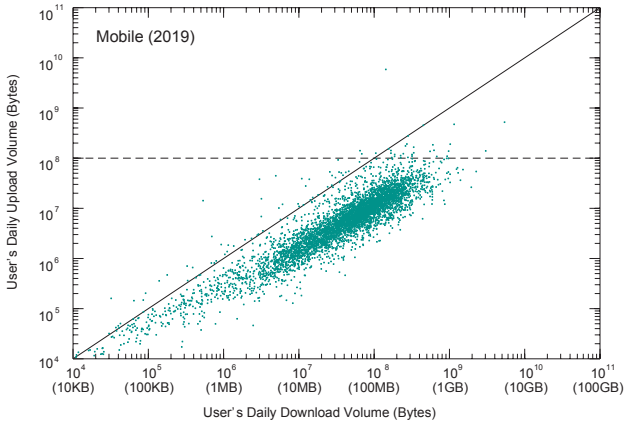

Figure 5: IN/OUT Usage for Each Mobile User

© Internet Initiative Japan Inc.

The peak was slightly off from the center of the distribution, so the distribution was adjusted to bring the mode toward the center. Comparing the values for 2018 and 2019, the IN mode rose from 79MB to 89MB and the OUT mode rose from 1,585MB to 1,995MB, translating into a growth factor of 1.3 for both IN and OUT. Meanwhile, because the means are influenced by heavy users (on the right-hand side of the distribution), they were significantly higher than the corresponding modes, with the IN mean being 479MB and the OUT mean being 2,986MB in 2019. The 2018 means were 428MB and 2,664MB, respectively.

For mobile traffic (Table 2), the mean and modal values are close owing to the lack of heavy users. In 2019, the IN mode was 9MB and the OUT mode was 79MB, while the means were IN 11MB and OUT 85MB. The modes for both IN and OUT traffic were identical to those for the previous year. The modes were unchanged but the means and medians increased, which reflects a slight decrease in the proportion of light users, corresponding to the part of the distribution to the left of the peak in Figure 2.

Figure 4 and Figure 5 plot per-user IN/OUT usage volumes for random samples of 5,000 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 fall on the diagonal.

The cluster spread out below and parallel to the diagonal in each of these plots represents typical 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 about the upper right of the diagonal, but this is now no longer discernible. Variability between users in terms of usage levels and IN/OUT ratios is wide, indicating that there is a diverse range of usage styles. There is almost no difference between these plots and those for 2018, too.

For mobile traffic, the pattern of OUT being an order of magnitude larger also applies, but usage volumes are lower than for broadband, and there is less variability between IN and OUT.

Figure 6 and Figure 7 show the complementary cumulative distribution of users' daily traffic volume. On these log-log plots, the Y-axis values represent the proportion of users with daily usage levels greater than the corresponding X-axis values. These plots are an effective way of examining the distribution of heavy users. The linear-like decline toward the right-hand side of the plots indicates that the distributions are long-tailed and close to a power-law distribution. Heavy users appear to be distributed statistically and do not appear to constitute a separate, special class of user.

On mobile, heavy users appear to be distributed according to a power-law for OUT traffic, but the linear-like slope breaks down somewhat for IN traffic, with a large proportion of users uploading large volumes of data.

Traffic is heavily skewed across users, such that a small proportion of users accounts for the majority of overall traffic volume. For example, the top 10% of broadband users

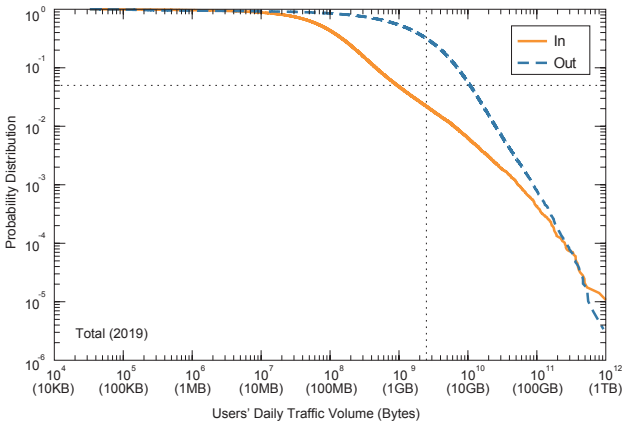

Figure 6: Complementary Cumulative Distribution of Broadband Users' Daily Traffic Volume

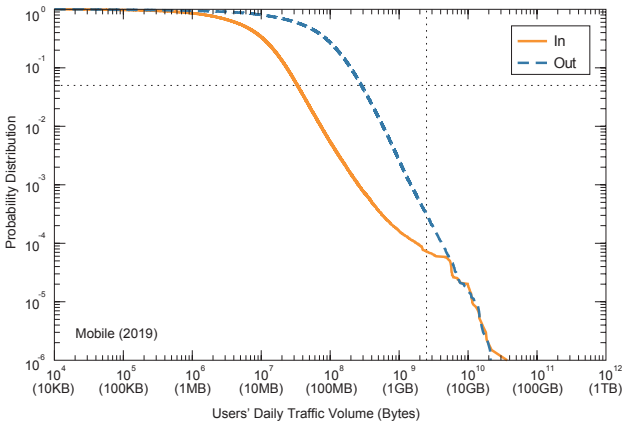

Figure 7: Complementary Cumulative Distribution of Mobile Users' Daily Traffic Volume

© Internet Initiative Japan Inc.

account for 52% of total OUT and 82% of total IN traffic, while the top 1% of users account for 17% of OUT and 58% of IN traffic. As the proportion of heavy users has declined over the past few years, the skew has also decreased, albeit only slightly. As for mobile, the top 10% of users account for 43% of OUT and 47% of IN traffic, while the top 1% account for 12% of OUT and 18% of IN traffic. The skew is less pronounced than indicated in our reports up to last year because we are now looking only at data on individuals.

1.4 Usage by Port

Next, we look at a breakdown of traffic and examine usage levels by port. Recently, it has become 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, which is assigned to HTTP, to avoid firewalls. Hence, generally speaking, when both parties are using a dynamic port numbered 1024 or higher, the traffic is likely to be from a P2P application, and when one of the parties is using a well-known port lower than 1024, the traffic is likely to be from a client/server application. In light of this, we take the lower of the source and destination port numbers when breaking down TCP and UDP usage volumes by port.

Table 3 shows the percentage breakdown of broadband users' usage by port over the past five years. In 2019, 81% of all traffic was over TCP connections. The proportion of traffic over port 443 (HTTPS), which was down a little last time, rose sharply from 41% to 52%. The proportion of traffic over port 80 (HTTP) fell from 27% to 20% here, and the figure for UDP port 443, which is used by Google's QUIC protocol, fell from 10% to 8% after having risen up until the previous edition of this report. These figures demonstrate that the shift from HTTP to HTTPS is ongoing, while QUIC has tapped the brakes on growth a little.

TCP dynamic port traffic, which has been on the decline, fell to 8% in 2019. The proportion accounted for by individual dynamic port numbers is tiny, with the most commonly used port 8080 only accounting for 0.5%. Port 1935, which is used by Flash Player and has also been in decline, fell to around 0.3%. Almost all other traffic here is VPN related.

Table 4 shows the percentage breakdown by port for mobile users. The figures are close to those for broadband on the whole, suggesting that mobile users use applications in a manner similar to broadband users.

Table 3: Broadband Users' Usage by Port

year	2015	2016	2017	2018	2019
protocol port	(%)	(%)	(%)	(%)	(%)
TCP	**80.8**	**82.8**	**83.9**	**78.5**	**81.2**
(< 1024)	63.3	69.1	72.9	68.5	73.3
443 (https)	23.3	30.5	43.3	40.7	51.9
80 (http)	37.9	37.1	28.4	26.5	20.4
993 (imaps)	0.1	0.1	0.2	0.2	0.3
22 (ssh)	0.2	0.2	0.1	0.1	0.2
182	0.4	0.3	0.3	0.3	0.2
(>= 1024)	17.5	13.7	11.0	10.0	7.89
8080	0.3	0.2	0.3	0.3	0.5
1935 (rtmp)	1.8	1.5	1.1	0.7	0.3
UDP	**11.4**	**11.1**	**10.5**	**16.4**	**14.1**
443 (https)	0.9	2.4	3.8	10.0	7.8
4500 (nat-t)	0.2	0.2	0.2	0.2	0.3
ESP	**7.4**	**5.8**	**5.1**	**4.8**	**4.4**
IP-ENCAP	**0.2**	**0.2**	**0.3**	**0.2**	**0.2**
GRE	**0.2**	**0.1**	**0.1**	**0.1**	**0.1**
ICMP	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**

Table 4: Mobile Users' Usage by Port

year	2015	2016	2017	2018	2019
protocol port	(%)	(%)	(%)	(%)	(%)
TCP	**93.8**	**94.4**	**84.4**	**76.6**	**76.9**
443 (https)	37.4	43.7	53.0	52.8	55.6
80 (http)	52.5	46.8	27.0	16.7	10.3
31000	0.0	0.2	1.8	2.9	6.4
993 (imaps)	0.5	0.5	0.4	0.3	0.3
1935 (rtmp)	0.5	0.3	0.2	0.1	0.1
UDP	**5.2**	**5.0**	**11.4**	**19.4**	**17.3**
443 (https)	1.0	1.5	7.5	10.6	8.3
12222	0.0	0.1	0.1	2.3	3.4
4500 (nat-t)	0.3	0.2	0.2	4.5	3.0
53 (dns)	0.1	0.2	0.1	0.1	0.1
ESP	**0.7**	**0.4**	**0.4**	**3.9**	**5.8**
GRE	**0.3**	**0.1**	**0.1**	**0.1**	**0.0**
ICMP	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**

© Internet Initiative Japan Inc.

Figure 8 compares overall broadband traffic for key port categories across the course of the week from which observations were drawn in 2018 and 2019. We break the data into four port buckets: TCP ports 80 and 443, dynamic ports (1024 and up), and UDP port 443. The data are normalized so that peak overall traffic volume on the plot is 1. By comparison with 2018, the proportion of traffic over TCP port 443 has risen even further whereas TCP port 80 has seen a decrease. The overall peak is between 19:00 and 23:00 hours. Traffic increases during the daytime on Saturday and Sunday, reflecting household Internet usage times.

Figure 9 shows the trend for TCP ports 80 and 443 and UDP port 443, which account for the bulk of mobile traffic. In the mobile space as well, TCP port 443 has seen an increase while the proportion of traffic over TCP port 80 has declined. When compared with broadband, we note that mobile traffic levels remain high throughout the day, from morning through night. The plot shows that usage times differ from those for broadband, with three separate mobile traffic peaks occurring on weekdays: morning commute, lunch break, and evening from 17:00 to 22:00 hours.

1.5 Conclusion

Traffic volume has been growing moderately over the past few years. Although I say “moderately”, it’s only moderate in relation to past growth. At its annual rate of 20%, traffic is set to more than double over a four-year period. Both broadband and mobile traffic have been increasing, albeit with some ups and downs. The fact that both tend to rise and fall around the same time suggests that common factors are at play, but we have not been able to pinpoint what the specific factors are.

Both broadband and mobile usage volume by user have not changed much in the past few years. No new services that would drive traffic upward have appeared over that time, and it is clear that users’ Internet usage has not changed much as a result. Video resolutions are definitely on the rise, but it looks like the accompanying rise in codec compression rates is keeping total traffic growth in check.

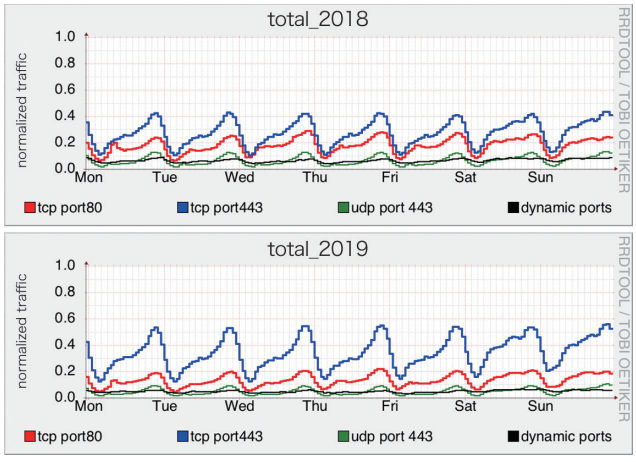

Figure 8: Broadband Users’ Port Usage Over a Week
2018 (top) and 2019 (bottom)

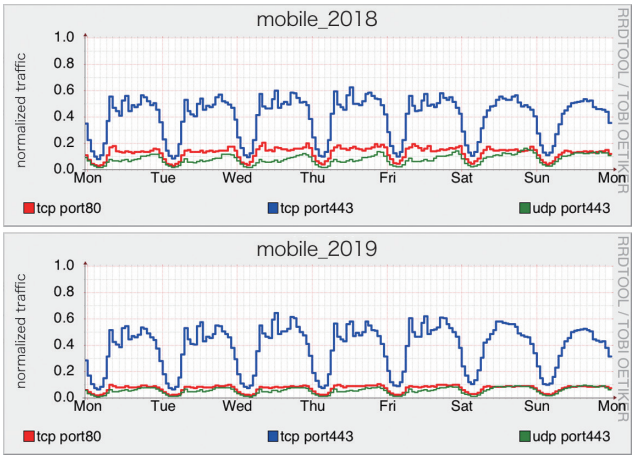

Figure 9: Mobile Users’ Port Usage Over a Week
2018 (top) and 2019 (bottom)

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

© Internet Initiative Japan Inc.