Broadband Traffic Report: COVID-19's Impact in its 2nd Year

1.1 Overview

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

As in 2020, home Internet usage again increased under the COVID-19 pandemic, with broadband traffic staying in an uptrend. Meanwhile, with people venturing outdoors less, mobile usage has been largely range-bound.

Figure 1 graphs 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 OUT observations for January 2020 for both services to 1.

Broadband services traffic surged from March to May 2020, when COVID-19 cases were starting to ramp up in Japan. It fell slightly in June after Japan's state of emergency was lifted but turned up again from August. Over the past year, broadband IN traffic increased 20% and OUT traffic increased 23%. While these are smaller increases than the year-earlier figures of 43% and 34%, the growth rates do appear to have returned to their former levels. Mobile services traffic, meanwhile, remained range-bound overall during this period amid lower rates of use outside the home/office, despite an increase in the use of services for remote work. Over the past year, mobile IN traffic increased 39% and OUT traffic fell 1%. A year earlier, IN was up 28% and OUT down 7%.

The broadband figures include IPv6 IPoE traffic. IPv6 traffic on IIJ's broadband services comprises both IPoE and PPPoE traffic[*5]. As of June 2021, IPoE accounted for almost a third of all traffic, at 31% of IN and 30% of OUT broadband traffic overall, year-on-year increases of 7 and 10 percentage points, respectively. With PPPoE congestion having become quite noticeable amid COVID-19, users are increasingly shifting to IPoE, and use of IPoE thus continues to rise.

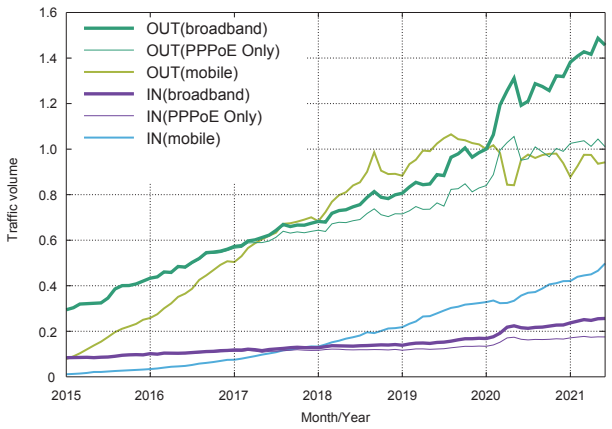

Figure 1: Monthly Broadband and Mobile Traffic

*1 Kenjiro Cho. Broadband Traffic Report: The Impact of COVID-19. Vol.48. pp4-9. September 2020.
*2 Kenjiro Cho. Broadband Traffic Report: Moderate Growth in Traffic Volume Ongoing. Vol. 44. pp4-9. September 2019.
*3 Kenjiro Cho. Broadband Traffic Report: Download Growth Slows for a Second Year Running. Vol. 40. pp4-9. September 2018.
*4 Kenjiro Cho. Broadband Traffic Report: Traffic Growth Slows to a Degree. Internet Infrastructure Review. Vol. 36. pp4-9. September 2017.
*5 Akimichi Ogawa and Satoshi Kubota. Tettei Kaisetsu v6 Plus. Lambda Note. January 2020 (https://www.jpne.co.jp/books/v6plus/, in Japanese).

We now look at broadband traffic by time of day on weekdays and weekends amid COVID-19. Traffic volume here is the sum of PPPoE and IPoE. Figures 2 and 3 show traffic for the following seven weeks: the week of February 25, 2020, before Japan's school closures; the week of April 20, 2020, corresponding to Japan's first state of emergency; the week of June 22, 2020, after the state of emergency was lifted; the week of August 31, 2020, when the second COVID-19 wave was easing; the week of January 18, 2021, the second state of emergency; the week of March 22, 2021, after Tokyo's state of emergency was lifted; and the week of July 5, which marked the start of the fifth wave of COVID-19, involving variant strains. We graph hourly average traffic volume figures for each of these weeks, partitioned into weekdays (Monday–Friday, excluding public holidays) and weekends (Saturday/Sunday). The lines in the lower part of each plot represent uploads, but here we focus on download volume.

First, we look at weekday traffic. Comparing February and April 2020 to see the impact of the first state of emergency, we see that traffic was up substantially in the daytime and that it also increased during evening peak hours. When the state of emergency was lifted in June, the additional daytime traffic fell to less than half what it had been, but peak hours saw almost no decline. Daytime traffic subsequently edged upward but did not return to its April 2020 levels until March of this year. Daytime traffic fell a little in July, which seems to reflect that schools were in session and remote work had declined a bit. Focusing on the 20:00–22:00 peak hours, we see a fairly consistent increase.

Weekend traffic, meanwhile, shows less variability than weekdays. Weather has a greater impact than school or remote work on the proportion of people at home on weekends. For example, poor weather on January 23 and 24 this year pushed home Internet usage upward; and on March 27 and 28, the cherry blossoms were in full bloom west of the Kanto region, resulting in more people going out and thus lower traffic levels. Traffic during nighttime peak hours is roughly the same as on weekdays.

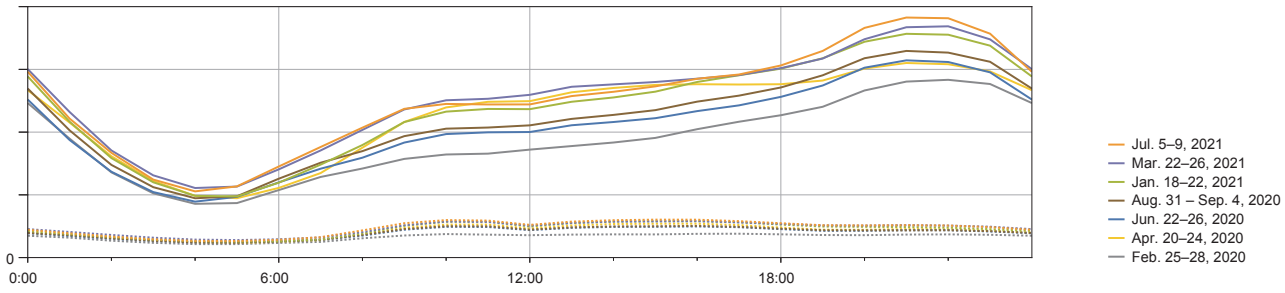

Figure 2: Hourly Average Broadband Traffic on Weekdays

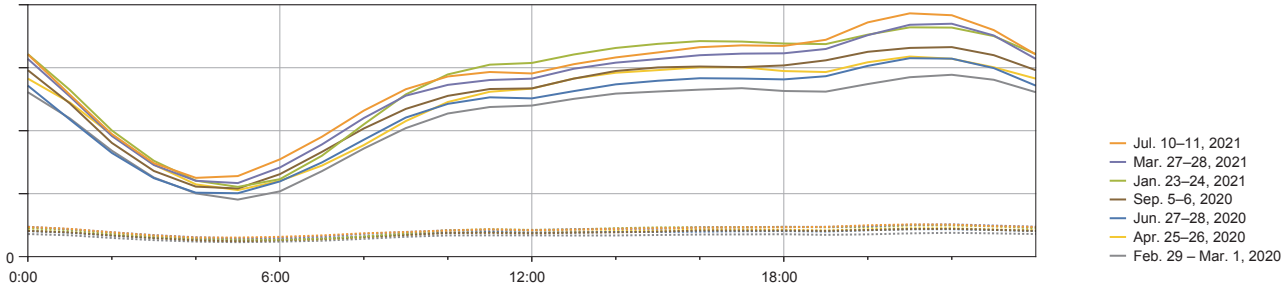

Figure 3: Hourly Average Broadband Traffic on Weekends

Note that IPoE traffic is not included in the following analysis, as detailed data is not available because we use Internet Multifeed Co.'s transix service for IPoE.

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 31 – June 6, 2021, and compare those data with data for the week of June 1–7, 2020, which we analyzed in the previous edition of this report.

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 to around 1/8,192, 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 2021, 99% were using fiber-optic connections.

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.

Since our 2019 report, we use daily usage data only on services provided to individuals. The distribution is heavily distorted if we include enterprise services, where usage patterns are highly varied. So to form a picture of overall usage trends, we determined that using only the individual data would yield more generally applicable, easily interpretable conclusions. Note that the analysis of usage by port in the next section does include enterprise data because of the difficulty of distinguishing between individual and enterprise usage.

Figures 4 and 5 show the average daily usage distributions (probability density functions) for broadband and mobile users. Each compares data for 2020 and 2021 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

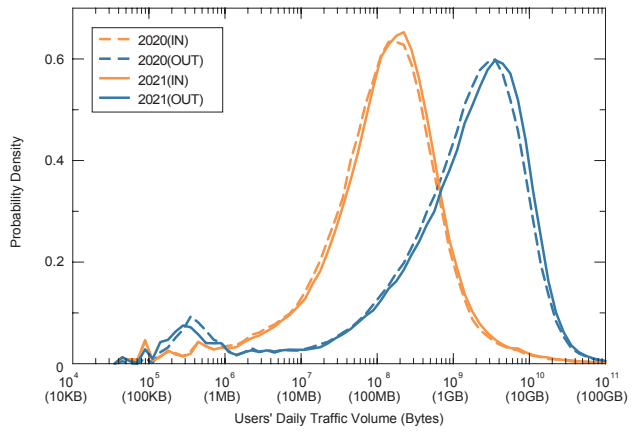

Figure 4: Daily Broadband User Traffic Volume Distribution Comparison of 2020 and 2021

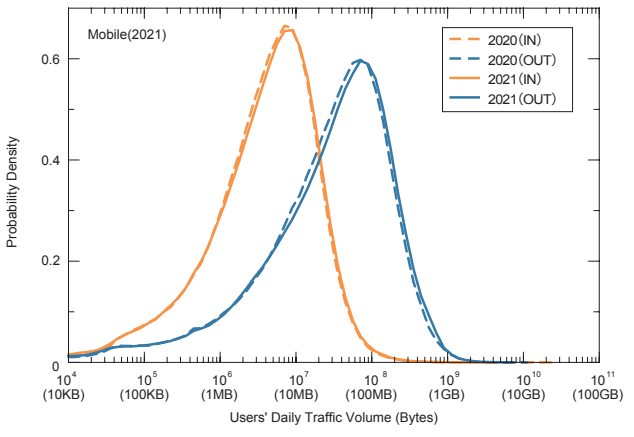

Figure 5: Daily Mobile User Traffic Volume Distribution Comparison of 2020 and 2021

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 2021 are further to the right than the peaks of the 2020 distributions, indicating that overall user traffic volumes are increasing. Compared with what we have seen in the past, there is almost no change in the distributions this time around, which is also evident from the lack of growth in total PPPoE traffic.

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

Similarly, the peaks of the mobile distributions in Figure 5 have moved slightly to the right, indicating that overall traffic has increased, albeit only slightly. Mobile 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.

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). When the peak is slightly off the center of the distribution, the distribution is adjusted to bring the mode toward the center. All of the values increased this time around. Comparing 2020 and 2021, the IN mode rose from 158MB

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
2020	609	122	158	3810	1638	3162
2021	684	136	200	4225	1875	3981

to 200MB and the OUT mode rose from 3,162MB to 3,981MB, translating into growth factors 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 are significantly higher than the corresponding modes, with the IN mean at 684MB and the OUT mean at 4,225MB in 2021. The 2020 means were 609MB and 3,810MB, respectively.

For mobile traffic, the mean and mode are close owing to the lack of heavy users. As Table 2 shows, the IN mean has fallen slightly while all other values are up. In 2021, the IN mode was 8MB and the OUT mode was 71MB, while the means were IN 10MB and OUT 86MB. The 2020 modes were IN 7MB and OUT 63MB, and the means were IN 10MB and OUT 79MB.

Figures 6 and 7 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. 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. For both

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
2020	10.4	4.5	7.1	79.4	35.1	63.1
2021	9.9	4.7	7.9	85.9	37.9	70.8

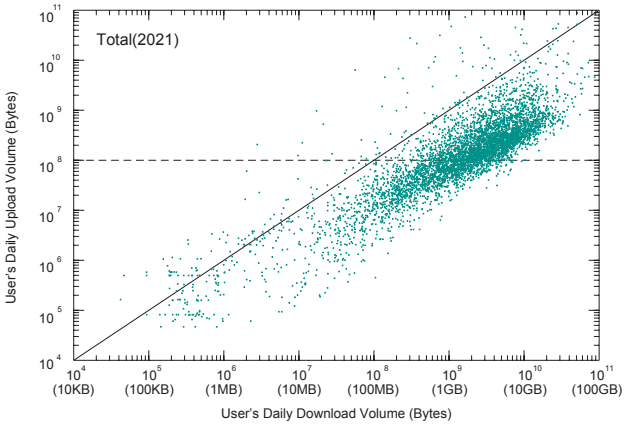

Figure 6: IN/OUT Usage for Each Broadband User

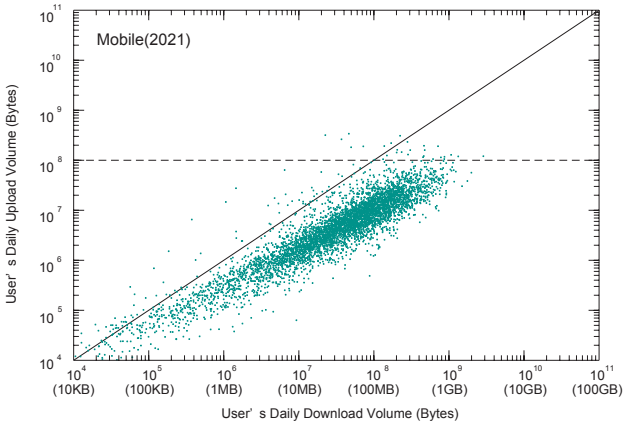

Figure 7: IN/OUT Usage for Each Mobile User

broadband and mobile, there is almost no difference between these plots and those for 2020.

Figures 8 and 9 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.

The broadband distributions are largely unchanged from last year. But on mobile, the bump observed at the bottom right of the IN distribution last year due to a heavy volume of uploads has disappeared, and the slope is now fairly linear.

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 account for 48% of total OUT and 76% of total IN traffic, while the top 1% of users account for 15% of OUT and 50% of IN traffic. The skew has not changed much from last year. As for mobile, the top 10% of users account for 48% of OUT and 49% of IN traffic, while the top 1% account for 12% of OUT and 16% of IN traffic. The skew here is also mostly unchanged from last year's report.

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.

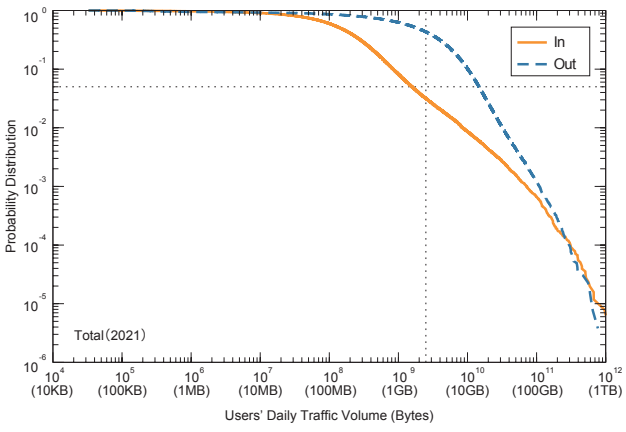

Figure 8: Complementary Cumulative Distribution of Broadband Users' Daily Traffic Volume

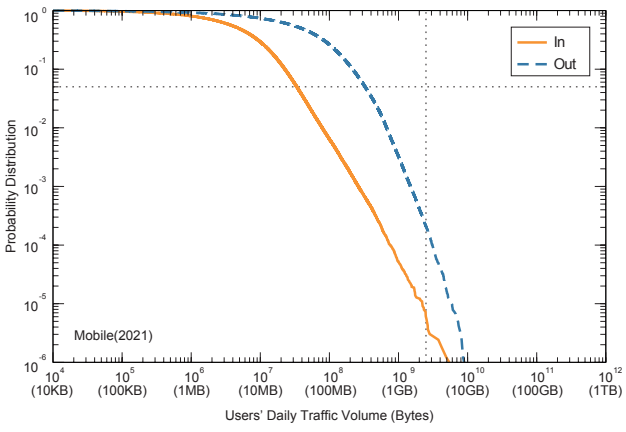

Figure 9: Complementary Cumulative Distribution of Mobile Users' Daily Traffic Volume

Table 3 shows the percentage breakdown of broadband users' usage by port over the past five years. In 2021, 72% of all traffic was over TCP connections, down 5 percentage points vs. 2020. The proportion of traffic over port 443 (HTTPS) was 54%, a 2-point increase from last year. The proportion of traffic over port 80 (HTTP) fell from 17% to 12%. The figure for UDP port 443, which is used by the QUIC protocol, rose from 11% to 16%, so HTTP declined by roughly the amount that QUIC increased.

TCP dynamic port traffic, which has been in decline, fell to 6% in 2021. Individual dynamic port numbers account for only a tiny portion, with the most commonly used port 31000 only making up 0.6%. Port 1935, which is used by Flash Player, makes up 0.2%, but almost all other traffic 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. This is likely because apps similar to those for PC platforms are now also used on smartphones, and because the proportion of broadband usage on smartphones is rising.

Figure 10 compares overall broadband traffic for key port categories across the course of the week from which observations were drawn in 2020 and 2021. 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. Comparing 2020 and 2021, we see that UDP port 443 has become more prominent than TCP port 80 in 2021. The increase in weekday daytime traffic observed in 2020 is down a little. The overall peak is between 19:00 and 23:00.

Figure 11 shows the trend for TCP ports 80 and 443 and UDP port 443, which account for the bulk of mobile traffic.

Table 3: Broadband Users' Usage by Port

year	2017	2018	2019	2020	2021
protocol port	(%)	(%)	(%)	(%)	(%)
TCP	**83.9**	**78.5**	**81.2**	**77.2**	**71.9**
(< 1024)	72.9	68.5	73.3	70.5	65.8
443(https)	43.3	40.7	51.9	52.4	53.5
80(http)	28.4	26.5	20.4	17.2	11.6
22(ssh)	0.1	0.1	0.2	0.2	0.2
993(imaps)	0.2	0.2	0.3	0.2	0.1
(>= 1024)	11.0	10.0	7.9	6.7	6.1
31000	0.1	0.1	0.2	0.4	0.6
8080	0.3	0.3	0.5	0.4	0.4
1935(rtmp)	1.1	0.7	0.3	0.4	0.2
UDP	**10.5**	**16.4**	**14.1**	**19.4**	**24.5**
443(https)	3.8	10.0	7.8	10.5	15.9
8801	0.0	0.0	0.0	1.1	0.9
4500(nat-t)	0.2	0.2	0.3	0.6	0.8
ESP	**5.1**	**4.8**	**4.4**	**3.2**	**3.3**
GRE	**0.1**	**0.1**	**0.1**	**0.1**	**0.2**
IP-ENCAP	**0.3**	**0.2**	**0.2**	**0.1**	**0.1**
ICMP	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**

Table 4: Mobile Users' Usage by Port

year	2017	2018	2019	2020	2021
protocol port	(%)	(%)	(%)	(%)	(%)
TCP	**84.4**	**76.6**	**76.9**	**75.5**	**70.3**
443(https)	53.0	52.8	55.6	50.7	44.4
80(http)	27.0	16.7	10.3	7.4	5.0
993(imaps)	0.4	0.3	0.3	0.2	0.2
1935(rtmp)	0.2	0.1	0.1	0.1	0.1
UDP	**11.4**	**19.4**	**17.3**	**18.0**	**23.8**
443(https)	7.5	10.6	8.3	9.3	16.3
4500(nat-t)	0.2	4.5	3.0	1.8	3.7
8801	0.0	0.0	0.0	1.4	0.7
3480	0.0	0.0	0.0	0.4	0.3
12222	0.1	2.3	3.4	0.8	0.2
ESP	**0.4**	**3.9**	**5.8**	**6.4**	**5.8**
GRE	**0.1**	**0.1**	**0.0**	**0.1**	**0.1**
ICMP	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**

Comparing the figures with 2020, UDP port 443 has risen further here also, and the lunchtime peak sticks out a bit more. 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.

1.5 Conclusion

Looking back on the situation during the COVID-19 pandemic over the past year and a half, weekday daytime broadband traffic increased substantially from May to March of 2020 as the brakes were put on human movement and a greater proportion of people stayed at home. But excluding this period, traffic volume looks to be rising steadily largely in line with an underlying growth curve. So, although traffic has seen an annual increase of around 40% due to COVID-19, the increase has not been so dramatic as initially feared, and while we do observe ups and downs due to changes in stay-at-home rates associated with COVID case numbers, traffic continues to grow steadily overall.

Also, in contrast with PPPoE, which is subject to bottlenecks and other constraints, IPoE traffic is growing apace and driving growth in broadband traffic overall. The use of IPoE can be expected to continue rising ahead.

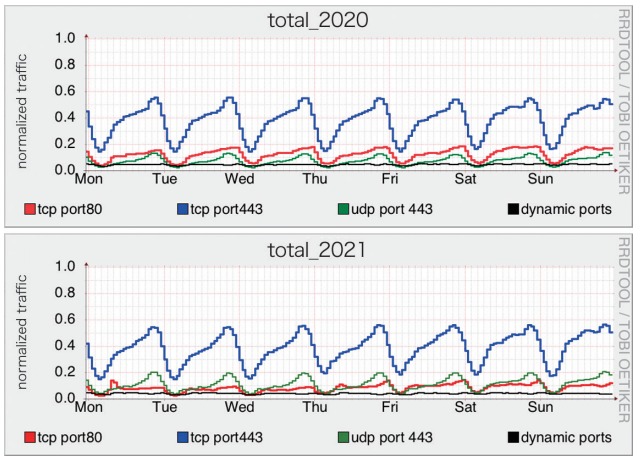

Figure 10: Broadband Users' Port Usage Over a Week
2020 (top) and 2021 (bottom)

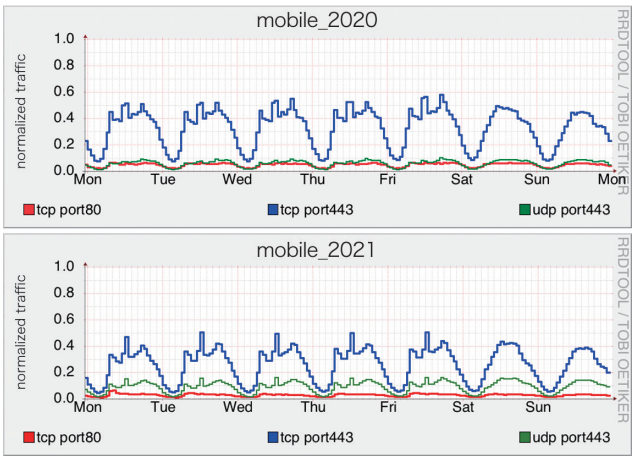

Figure 11: Mobile Users' Port Usage Over a Week
2020 (top) and 2021 (bottom)

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