# Broadband Traffic Report: The Impact of COVID-19

## 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]. Here, we again report on changes in traffic trends over the past year, based on daily user traffic and usage by port. Home Internet usage increased substantially amid the spread of COVID-19, and broadband traffic was thus up this time around. Mobile usage, meanwhile, has declined with people venturing outdoors less.

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 June 2019, a year earlier, for both services to 1.

Broadband services traffic surged from March to May, when COVID-19 cases were really starting to ramp up in Japan, and decreased slightly in June after Japan's state of emergency was lifted. We gave a detailed account of this period in the last issue[*4]. Over the past year, broadband IN traffic increased 43% and OUT traffic increased 34%. These are large increases vs. the corresponding year-earlier figures of 12% and 19%. Mobile services traffic, meanwhile, declined overall during this period, reflecting a large drop in usage out of the home/office, despite an increase in the use of remote-work services. Traffic subsequently made a slight comeback in June here as well. Over the past year, mobile IN traffic increased 28% and OUT traffic fell 7%, marking the first ever decline for downloads. A year earlier, IN was up 60% and OUT up 22%.

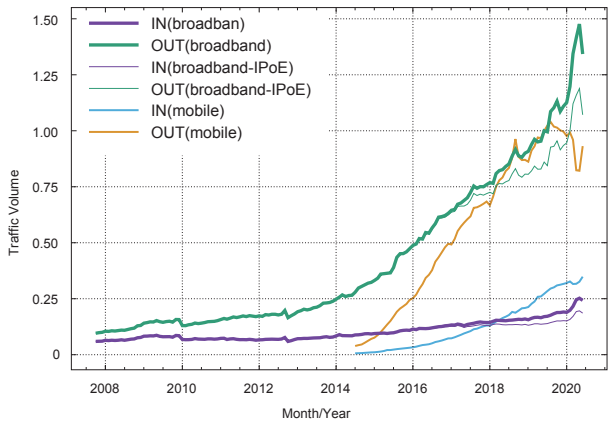


**Figure 1: Monthly Broadband and Mobile Traffic**

The broadband figures include IPv6 IPoE traffic. IPv6 traffic on IIJ's broadband services comprises both IPoE and PPPoE traffic[*5], 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 2020, IPoE accounted for 24% of IN and 20% of OUT broadband traffic overall, year-on-year increases of 5 and 6 points, respectively. PPPoE congestion has become quite noticeable since March in particular, and the use of IPoE is accelerating as users shift to IPoE to avoid this.

## 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 June 1–7, 2020, and compare those data with data for the week of May 27 – June 2, 2019, 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

*1 Kenjiro Cho. Broadband Traffic Report: Moderate Growth in Traffic Volume Ongoing. Vol.44. pp4-9. September 2019.
*2 Kenjiro Cho. Broadband Traffic Report: Download Growth Slows for a Second Year Running. Vol.40. pp4-9. September 2018.
*3 Kenjiro Cho. Broadband Traffic Report: Traffic Growth Slows to a Degree. Internet Infrastructure Review. Vol.36. pp4-9. September 2017.
*4 Kenjiro Cho. COVID-19's Impact on FLET'S Traffic, Internet Infrastructure Review. Vol.47. pp18-23. June 2020.
*5 Akimichi Ogawa and Satoshi Kubota. Tetei Kaisetsu v6 Plus. January 2020 (https://www.jpne.co.jp/books/v6plus/, in Japanese).

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 1/16,384, 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 2020, 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.

Since last edition, 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.

Figure 2 and Figure 3 show the average daily usage distributions (probability density functions) for broadband and mobile users. Each compares data for 2019 and 2020 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 ($10^4$) and 100GB ($10^{11}$) using a logarithmic scale. Most users fall within the 100GB ($10^{11}$) 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 2020 are further to the right than the peaks of the 2019 distributions, indicating that overall user traffic volumes are increasing. The increase in volume this time around was greater than it was in 2019.

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

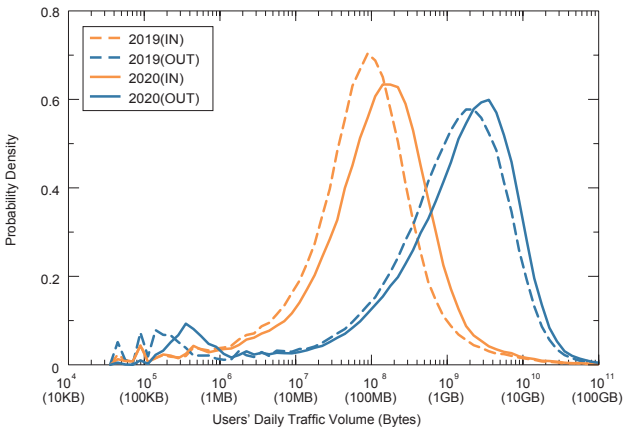


Figure 2: Daily Broadband User Traffic Volume Distribution Comparison of 2019 and 2020

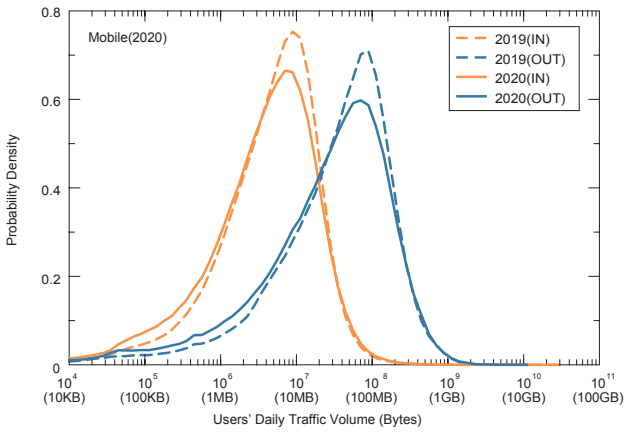


Figure 3: Daily Mobile User Traffic Volume Distribution Comparison of 2019 and 2020

Figure 3 shows the peaks in the mobile distributions have actually moved left and gotten lower, while the left tails have risen. This indicates that the proportion of high-volume users to total has not changed much, while the proportion of mid-volume users has fallen and the proportion of low-volume users has increased.

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 from the center of the distribution, the distribution is adjusted to bring the mode toward the center. All of the values grew substantially this time around. Comparing the values for 2019 and 2020, the IN mode rose from 89MB to 158MB and the OUT mode rose from 1,995MB to 3,162MB, translating into growth factors of 1.8 for IN and 1.6 for OUT. Meanwhile, because the means are influenced by heavy users (on the right-hand side of the distribution), they are significantly higher than

**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 |

**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 |

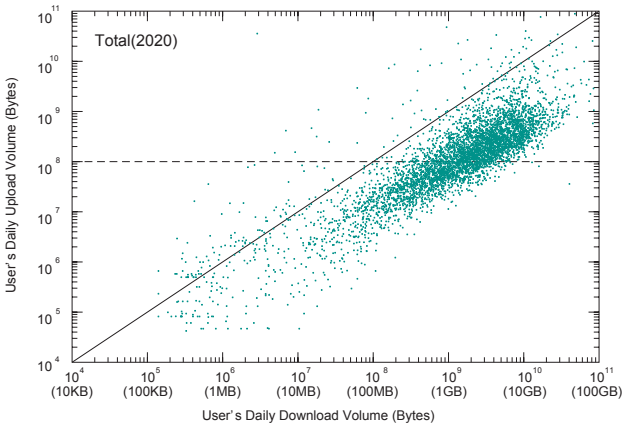


**Figure 4: IN/OUT Usage for Each Broadband User**

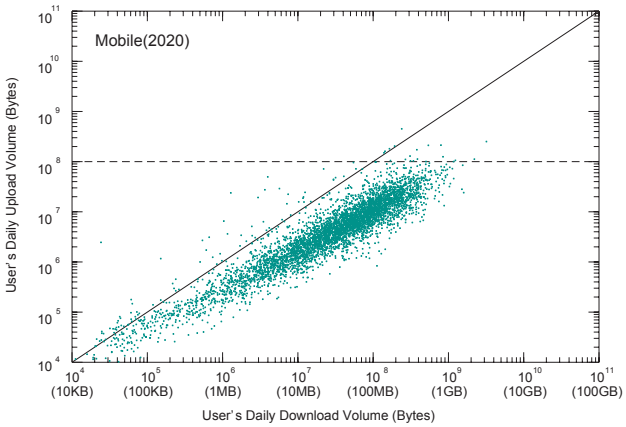


**Figure 5: IN/OUT Usage for Each Mobile User**

the corresponding modes, with the IN mean at 609MB and the OUT mean at 3,810MB in 2020. The 2019 means were 479MB and 2,986MB, respectively.

For mobile traffic, the mean and modal values are close owing to the lack of heavy users. As Table 2 shows, all of the values have fallen. In 2020, the IN mode was 7MB and the OUT mode was 63MB, while the means were IN 10MB and OUT 79MB. The 2019 modes were IN 9MB and OUT 79MB, and the means were IN 11MB and OUT 85MB.

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. 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 broadband and mobile, there is almost no difference between these plots and those for 2019.

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 larger proportion of users uploading large volumes of data. This year, the right edge of the distribution has shifted further out to the right, indicating a further increase in upload volume from some high-volume uploaders.

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 50% of total OUT and 76% of total IN traffic, while the top 1% of users account for 16% of OUT and 50% of IN traffic. The skew has decreased a little compared with last year. As for mobile, the top 10% of users account for 48% of OUT and 53% of IN traffic, while the top 1% account for 13% of OUT and 23% of IN traffic. The skew is a little larger than that in last year's report.

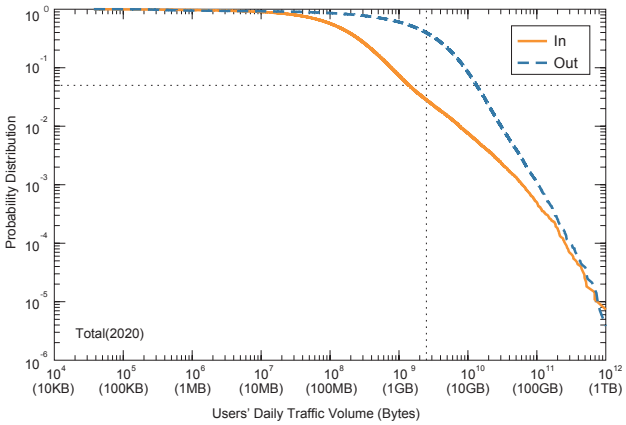


Figure 6: Complementary Cumulative Distribution of Broadband Users' Daily Traffic Volume

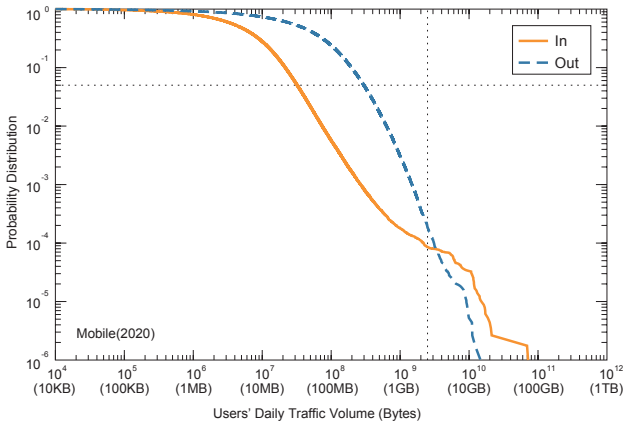


Figure 7: Complementary Cumulative Distribution of Mobile Users' Daily Traffic Volume

## 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 2020, 77% of all traffic was over TCP connections. The proportion of traffic over port 443 (HTTPS) was 52%, the same as last year. The proportion of traffic over port 80 (HTTP) fell from 20% to 17%. The figure for UDP port 443, which is used by the QUIC protocol, rose from 8% to 11%, so HTTP declined by roughly the amount that QUIC increased.

TCP dynamic port traffic, which has been in decline, fell to 7% in 2020. Individual dynamic port numbers account for only a tiny portion, with the most commonly used port 8080 only making up 0.4%. Port 1935, which is used by Flash Player and has also been in decline, makes up 0.4%, but 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. 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 8 compares overall broadband traffic for key port categories across the course of the week from which observations were drawn in 2019 and 2020. 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 2019, weekday daytime traffic is up substantially with people spending more time at home amid the COVID-19 situation. The overall peak is between 19:00 and 23:00.

Table 3: Broadband Users' Usage by Port

| year | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|
| protocol port | (%) | (%) | (%) | (%) | (%) |
| **TCP** | **82.8** | **83.9** | **78.5** | **81.2** | **77.2** |
| (< 1024) | 69.1 | 72.9 | 68.5 | 73.3 | 70.5 |
| 443(https) | 30.5 | 43.3 | 40.7 | 51.9 | 52.4 |
| 80(http) | 37.1 | 28.4 | 26.5 | 20.4 | 17.2 |
| 993(imaps) | 0.1 | 0.2 | 0.2 | 0.3 | 0.2 |
| 22(ssh) | 0.2 | 0.1 | 0.1 | 0.2 | 0.2 |
| 182 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 |
| (>= 1024) | 13.7 | 11.0 | 10.0 | 7.9 | 6.7 |
| 8080 | 0.2 | 0.3 | 0.3 | 0.5 | 0.4 |
| 1935(rtmp) | 1.5 | 1.1 | 0.7 | 0.3 | 0.4 |
| **UDP** | **11.4** | **10.5** | **16.4** | **14.1** | **19.4** |
| 443(https) | 2.4 | 3.8 | 10.0 | 7.8 | 10.5 |
| 8801 | 0.0 | 0.0 | 0.0 | 0.0 | 1.1 |
| 4500(nat-t) | 0.2 | 0.2 | 0.2 | 0.3 | 0.6 |
| **ESP** | **5.8** | **5.1** | **4.8** | **4.4** | **3.2** |
| **GRE** | **0.1** | **0.1** | **0.1** | **0.1** | **0.1** |
| **IP-ENCAP** | **0.2** | **0.3** | **0.2** | **0.2** | **0.1** |
| **ICMP** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** |

Table 4: Mobile Users' Usage by Port

| year | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|
| protocol port | (%) | (%) | (%) | (%) | (%) |
| **TCP** | **94.4** | **84.4** | **76.6** | **76.9** | **75.5** |
| 443(https) | 43.7 | 53.0 | 52.8 | 55.6 | 50.7 |
| 80(http) | 46.8 | 27.0 | 16.7 | 10.3 | 7.4 |
| 993(imaps) | 0.5 | 0.4 | 0.3 | 0.3 | 0.2 |
| 1935(rtmp) | 0.3 | 0.2 | 0.1 | 0.1 | 0.1 |
| **UDP** | **5.0** | **11.4** | **19.4** | **17.3** | **18.0** |
| 443(https) | 1.5 | 7.5 | 10.6 | 8.3 | 9.3 |
| 4500(nat-t) | 0.2 | 0.2 | 4.5 | 3.0 | 1.8 |
| 8801 | 0.0 | 0.0 | 0.0 | 0.0 | 1.4 |
| 1701(12tp) | 1.0 | 0.0 | 0.0 | 0.4 | 0.9 |
| 12222 | 0.1 | 0.1 | 2.3 | 3.4 | 0.8 |
| **ESP** | **0.4** | **0.4** | **3.9** | **5.8** | **6.4** |
| **GRE** | **0.1** | **0.1** | **0.1** | **0.0** | **0.1** |
| **ICMP** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** |

Figure 9 shows the trend for TCP ports 80 and 443 and UDP port 443, which account for the bulk of mobile traffic. Mobile is virtually unchanged from 2019. 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

Traffic volume has been growing moderately over the past few years, and the data this time around bear out major changes in Internet usage caused by the spread of COVID-19. Weekday daytime traffic volume has increased substantially amid the rapid rise of remote work and the shift to online learning. Web conferencing tools have also proliferated, and are even being used for parties and other social gatherings as well as children's tutoring sessions and club activities.

COVID-19 had the biggest impact on traffic volumes in May, and although the effects had settled down somewhat by the period in June that we report on here, broadband downloads were still up 34% vs. the corresponding period last year. Data on volumes per user show substantial growth in broadband, with people spending more time at home, and a decline in mobile, as they refrain from going out.

We had expected net services rolled in conjunction with the Olympic and Paralympic Games to have altered usage trends this time around, but it was the spread of COVID-19 that ultimately produced changes in Internet usage in a way we had not foreseen. While traffic volumes have settled down somewhat since June, they are unlikely to return to past levels given that the likes of remote work and video conferencing are now entrenched. The outlook remains murky at present, so the situation will continue to bear close watching.

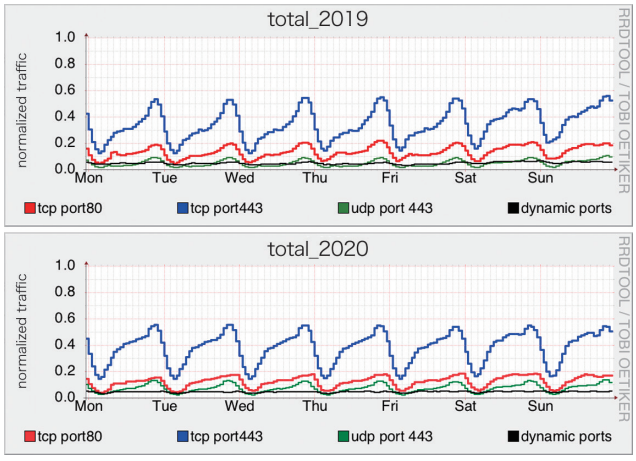


Figure 8: Broadband Users' Port Usage Over a Week
2019 (top) and 2020 (bottom)

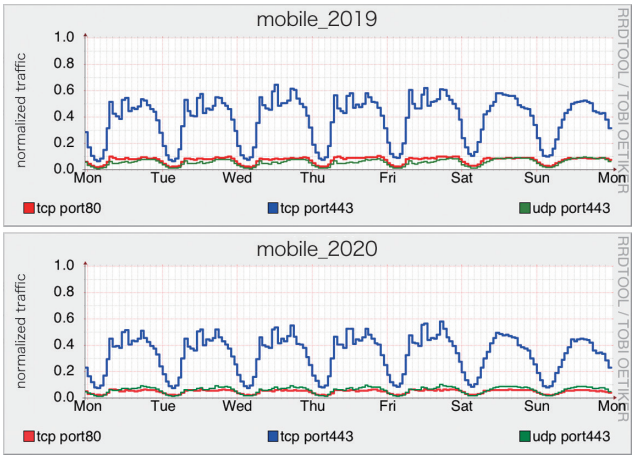


Figure 9: Mobile Users' Port Usage Over a Week
2019 (top) and 2020 (bottom)

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