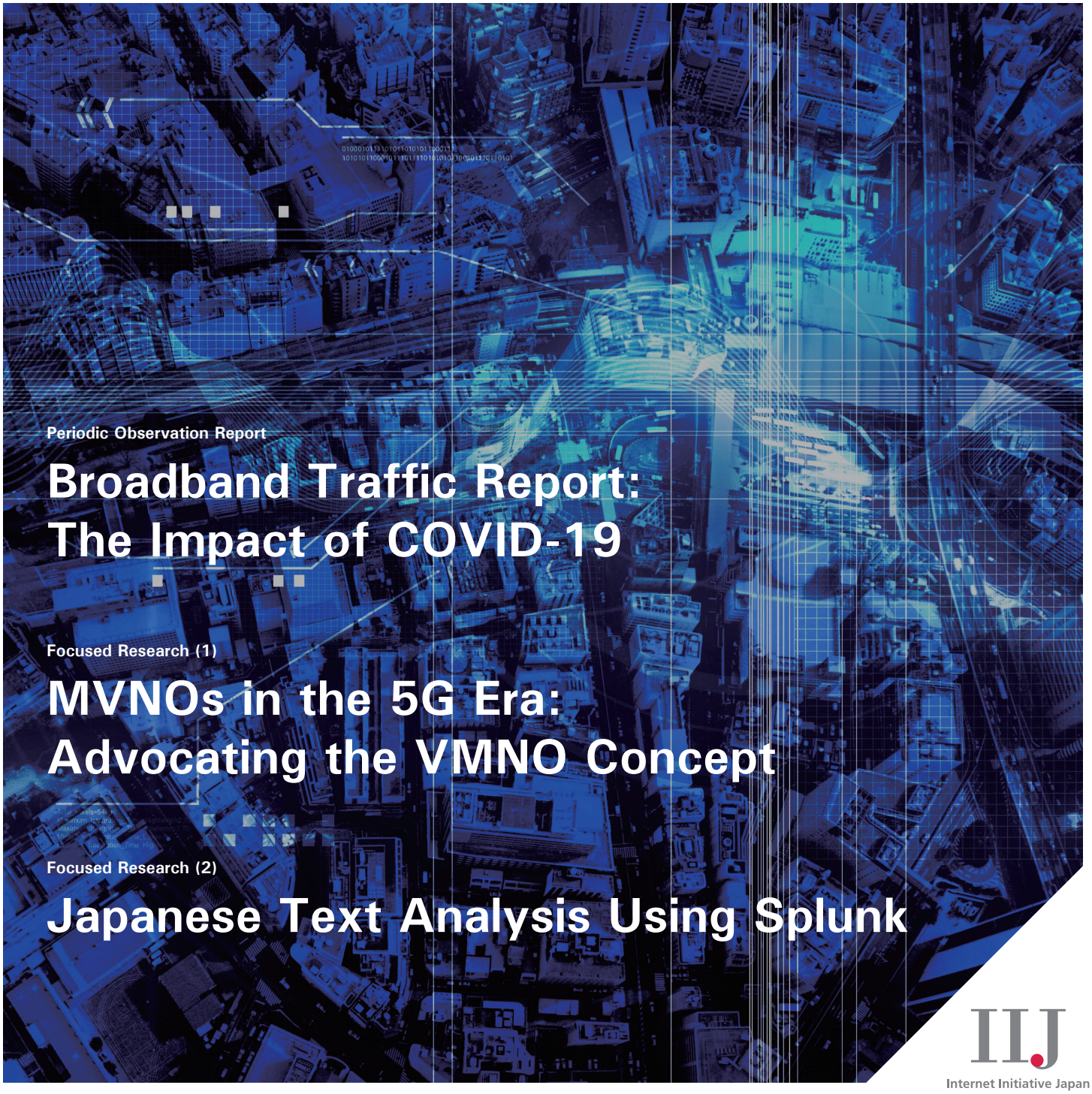
IIR

Internet
Infrastructure
Review

Nov.2020
Vol.48



Periodic Observation Report

Broadband Traffic Report: The Impact of COVID-19

Focused Research (1)

MVNOs in the 5G Era: Advocating the VMNO Concept

Focused Research (2)

Japanese Text Analysis Using Splunk

IIJ
Internet Initiative Japan

Internet Infrastructure Review

November 2020 Vol.48

Executive Summary

To understand the traffic conditions experienced by fixed broadband subscribers in Japan, the Ministry of Internal Affairs and Communications (MIC) collects and estimates traffic with cooperation from major Internet service providers, Internet exchanges, and researchers. IIJ also participates in these studies, which have been ongoing since 2004 and provide one of the most valuable datasets for chronicling the development of the Internet. The latest results, based on data for May 2020, were recently released[*1]. As reported in the media, and as you no doubt know, Internet traffic is on the rise worldwide because of the COVID-19 situation. The data period this time around coincides with when people's movements were most heavily restricted, with Japan having announced a state of emergency. As a result, fixed broadband subscribers' total download traffic was up a hefty 57.4% year on year, versus the 17.5% year-on-year increase in May 2019. Total upload traffic also saw strong growth of 48.5%. The numbers once again bear out the impact of COVID-19 on Internet traffic, and these data are also valuable in that they evidence the huge role that the Internet plays in times of emergency.

This role is not limited to the current COVID-19 situation. The Internet has long played a major role during natural disasters that have heavily impacted on society, including earthquakes and typhoons. And we will continue working to ensure that the Internet is able to fulfil its anticipated role as social infrastructure.

The IIR introduces the wide range of technology that IIJ researches and develops, comprising periodic observation reports that provide an outline of various data IIJ obtains through the daily operation of services, as well as focused research examining specific areas of technology.

Our periodic observation report for this issue, in Chapter 1, looks at our analysis of IIJ's fixed broadband and mobile traffic. As I mentioned, the MIC's recent traffic data bear out the impact of COVID-19, and our analysis here provides an even more detailed view of the impact. Our results clearly show that with people's movements being restricted this year, total traffic reached a peak in May once Japan declared a state of emergency, and subsequently started to ease off in June. The distribution of traffic per user for early June shows an increase in fixed broadband and a decline in mobile, so the traffic data also back up the observation that people were more active at home once their movements were restricted.

Our focused research report in Chapter 2 explains the VMNO concept that we are advocating in relation to how MVNOs should be set up in the 5G era. 5G services began rolling out last year in a number of countries, and MNO services launched in Japan last year as well. The current 5G services, however, are NSA (non-standalone) deployments that use the existing 4G setup with 5G introduced only on the wireless component to provide ultra-high-speed communications. In NSA deployments, the relationship between MNOs and MVNOs is unchanged from the 4G era. To achieve the 5G goals of massive machine type communications and ultra-low latency, we will need to migrate to SA (standalone) deployments that use a full 5G setup. The VMNO concept puts forward an approach that will allow MVNOs to provide services that take advantage of 5G's characteristics under SA deployments.

In the focused research report in Chapter 3, we describe our efforts as an operator of large-scale email services to automate the detection of spam and streamline service operations using machine learning technologies. We use Splunk as part of these efforts, but Splunk's NLP (Natural Language Processing) tools did not previously support Japanese, so we created our own NLP extension to make Japanese text mining possible. We hope this report also serves as a useful example of a business use case for text mining.

Through activities such as these, IIJ strives to improve and develop its services on a daily basis while maintaining the stability of the Internet. We will continue to provide a variety of services and solutions that our customers can take full advantage of as infrastructure for their corporate activities.

Junichi Shimagami

Mr. Shimagami is a Senior Executive Officer and the CTO of IIJ. His interest in the Internet led to him joining IIJ in September 1996. After engaging in the design and construction of the A-Bone Asia region network spearheaded by IIJ, as well as IIJ's backbone network, he was put in charge of IIJ network services. Since 2015, he has been responsible for network, cloud, and security technology across the board as CTO. In April 2017, he became chairman of the Telecom Services Association of Japan MVNO Council.

*1 Ministry of Internal Affairs and Communications, "Japanese Internet traffic data and estimates" (https://www.soumu.go.jp/menu_news/s-news/01kiban04_02000171.html, in Japanese).

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%.

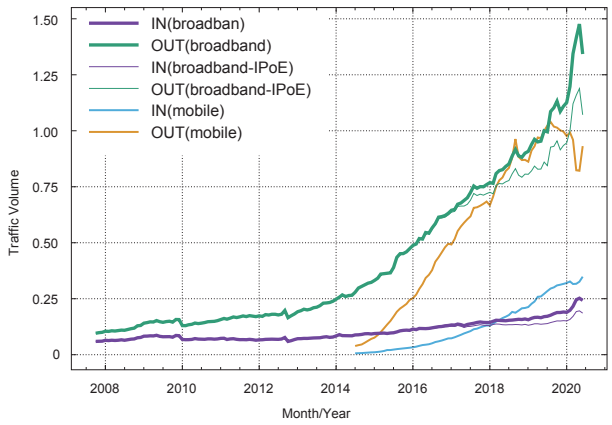


Figure 1: Monthly Broadband and Mobile Traffic

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.

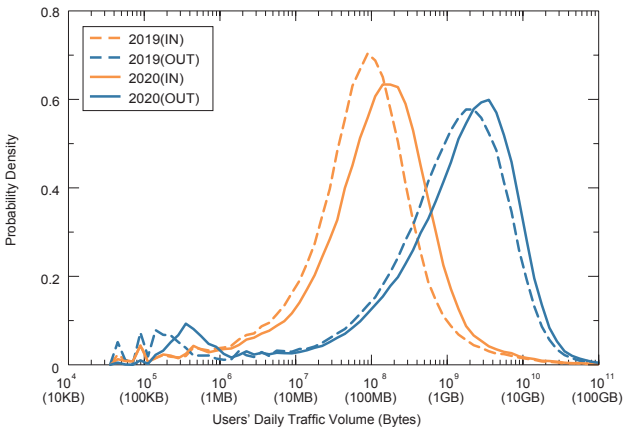


Figure 2: Daily Broadband User Traffic Volume Distribution Comparison of 2019 and 2020

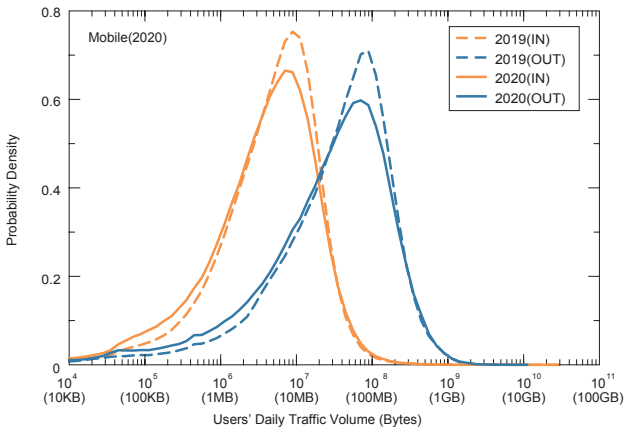


Figure 3: Daily Mobile User Traffic Volume Distribution Comparison of 2019 and 2020

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

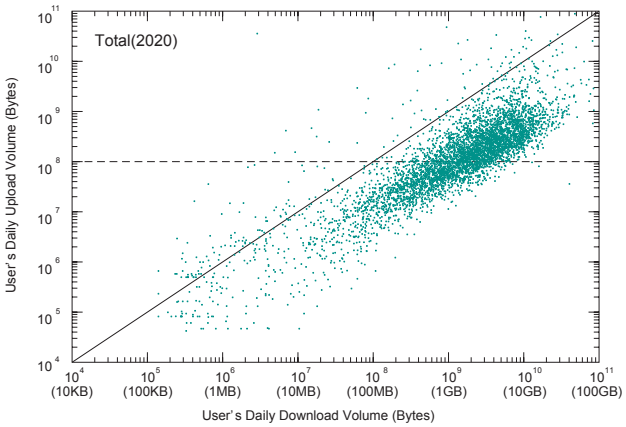


Figure 4: IN/OUT Usage for Each Broadband User

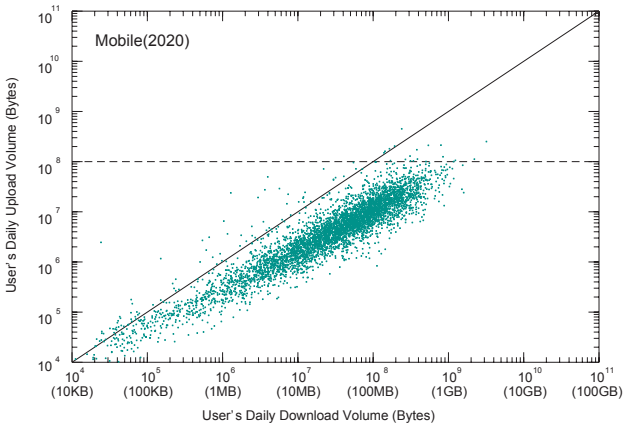


Figure 5: IN/OUT Usage for Each Mobile User

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.

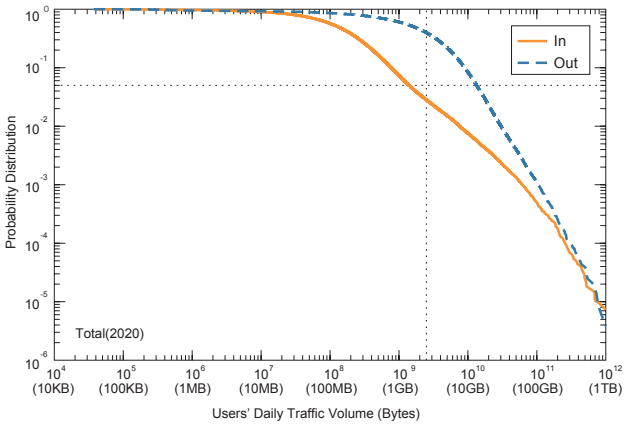


Figure 6: Complementary Cumulative Distribution of Broadband Users' Daily Traffic Volume

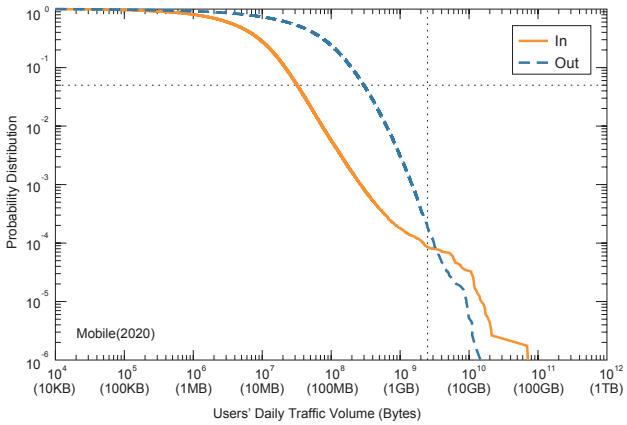


Figure 7: Complementary Cumulative Distribution of Mobile Users' Daily Traffic Volume

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.

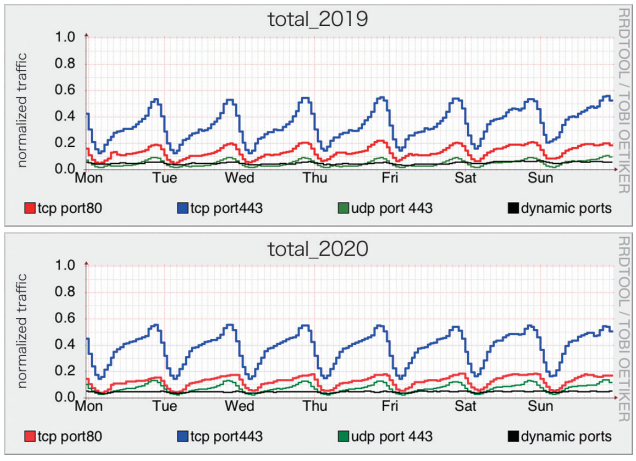


Figure 8: Broadband Users' Port Usage Over a Week
2019 (top) and 2020 (bottom)

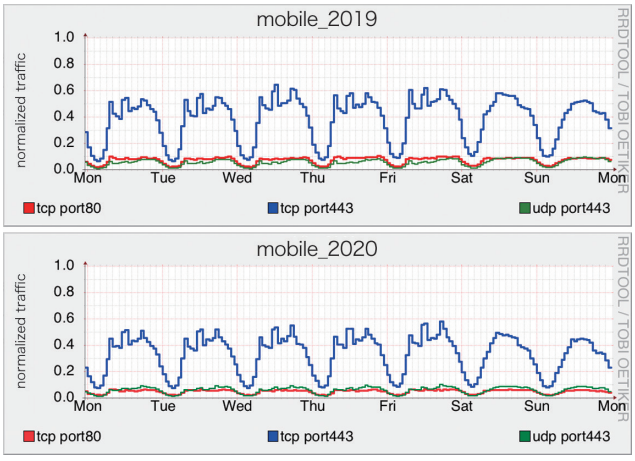


Figure 9: Mobile Users' Port Usage Over a Week
2019 (top) and 2020 (bottom)

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

MVNOs in the 5G Era: Advocating the VMNO Concept

2.1 The Runup to 5G

IIJ has consistently been a front runner in this field since launching its MVNO business in 2008 (initially deployed on a W-CDMA network, on LTE since 2012). The market environment facing MVNOs has changed significantly over that time, and IIJ has developed a diverse range of advanced MVNO businesses that serve many users, including business and consumer services, MVNE services, IoT/M2M, and full MVNO operations. The total number of subscriptions under these services now exceeds three million, and that number continues to grow, making IIJ Japan's biggest MVNO in both name and substance.

Against that backdrop, competition in the MVNO space grows more intense by the day. With direct regulations on sales of smartphones, in particular, being tightened every year, the vertical market structure consisting of MNOs—offering high-end devices on expensive rate plans premised on generous cashbacks and two-year contracts—and MVNOs—focusing on middle-class and low-end devices on "no-frills" plans—has crumbled, giving way to multifaceted competition. As MNO sub-brands and Rakuten Mobile, Japan's fourth MNO, continue to rise, some MVNOs are already struggling to earn a profit. And some of those MVNOs have no choice but to withdraw from the market. Why is this happening?

An MVNO business can only provide the limited mobile services that its host MNO provides. Based on considerations like profitability and differentiation versus peers, MNOs are relatively free to choose what type of services they provide from among all of the feasible technologies. In contrast, so long as it uses an MNO's network, an MVNO's choices are constrained by that. As successive generations of cellular communications technology rolled out, from 2G in the 1990s to 3G and 4G LTE, so too have mobile services evolved, from initially only being pay-as-you-go voice services to packet data communications, VoLTE, flat-rate and packet-based voice plans, carrier aggregation, and LPWA. Yet these services are only provided to an MVNO pursuant to the technological and economic conditions between it and the MNO, so it is fundamentally difficult for MVNOs to differentiate themselves.

There are avenues open to MVNOs, however, if they can unbundle part of the MNO's network and operate it themselves to provide their own mobile services to the extent the equipment permits. The rise of this practice of unbundling is synonymous with the history of MVNOs. Japan's Ministry of Internal Affairs and Communications approved the unbundling of packet gateway[*1], known as "Layer 2" type of MVNO in Japan, in 2008, and this has since become mandatory for Japan's three MNOs. Yet the three MNOs are not

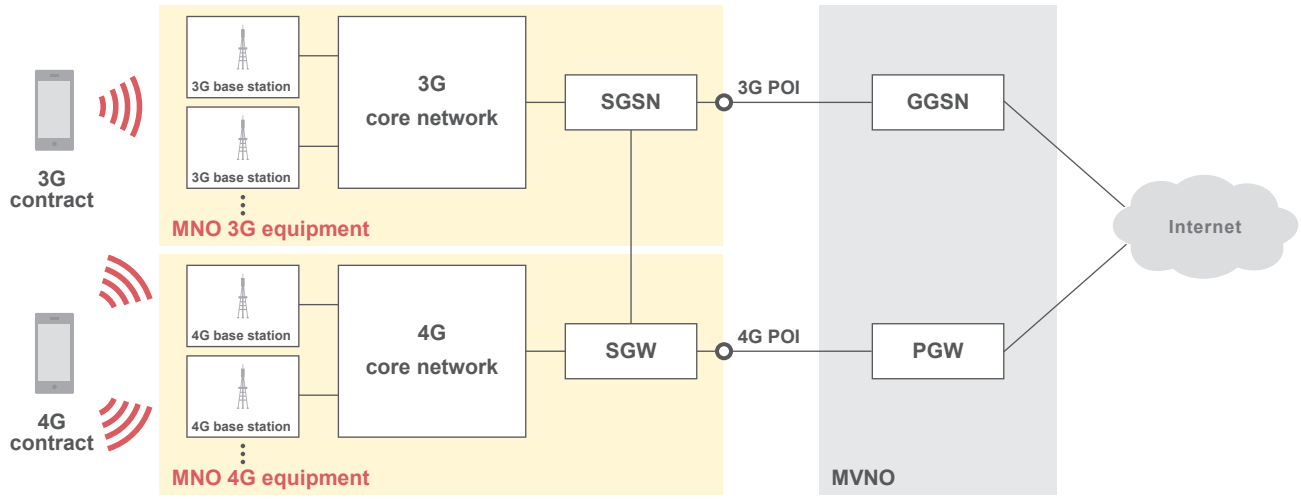


Figure 1: Illustration of Packet Exchanger Unbundling (Layer 2 Connection)

*1 The GGSN (Gateway GPRS Support Node) in 3G, the PGW (Packet Gateway) in 4G LTE.

obligated to unbundle HLR/HSS[*2], which is the equipment used to manage SIM cards, but the government guidelines say it is desirable to do so, and so full MVNOs—which take on the operation of the unbundled HLR/HSS and offer a range of innovative services that other MVNOs are unable to—are now appearing in Japan, starting with IIJ in 2018. Please see IIR Vol. 38[*3] for an overview of IIJ's efforts to develop new services as a full MVNO.

But with the era of 5G approaching in earnest, we find ourselves at a new turning point facing a trajectory that is not simply an extension of business to date. The early stages of the 5G era sees NSA[*4] implementations relying on existing 4G infrastructure, with very little changing on the infrastructure front relative to 4G. SA[*5] implementations that do not rely on 4G infrastructure are set to follow, and MNOs' 5G networks are expected to have a high degree of virtualization by the time these implementations roll out. Efficiently achieving the broad end-to-end QoS[*6] goals of 5G—namely enhanced mobile broadband, massive machine type communications, and ultra-reliable, low-latency communications—requires the introduction of virtualization technology and the horizontal layering of networks based on this, or in other words, network slicing.

From an MVNO perspective, however, a big question remains unanswered. Will an unbundling strategy continue to work on virtual networks in the 5G SA era? If not, how will MVNOs be able to differentiate themselves?

2.2 5G and MVNOs

Two major issues present themselves when we consider the possibilities for unbundling in the 5G SA era. One is network segmentation. Unbundling is a way of dividing a network vertically at a point of interface (POI[*7]), but this does not appear to work well with network slicing. In short, network slicing (horizontal division of a core network into slices) is set to be introduced to achieve the broad end-to-end QoS goals for 5G, but if MVNOs further physically separate out only part of the core network, this could hinder efforts to achieve the required QoS levels.

The other issue relates to operational aspects. Having generally standardized specifications for technical interfaces between operators at POIs is desirable. That's not the only consideration, though. The fact that autonomous operators are on either side of a POI creates very heavy operating restrictions. Even if the technical specifications for the POI are met, neither operator can make configuration changes or add new functions unless both operators are in agreement. In the 3G and 4G LTE world, once a POI was built, its configuration did not need to be changed all that frequently, and this applies in the case of Layer 2 MVNOs as well as full MVNO arrangements. Hence, the operating restrictions did not really hinder the smoothness of business. With 5G, however, slices (virtualized core network) need to be operated dynamically in order to achieve the various QoS goals for providing communications services that meet users' needs. Achieving this level of flexibility using the conventional method of unbundling would be awfully difficult in the 5G SA era.

*2 Called the HLR (Home Location Register) in 3G and the HSS (Home Subscriber Server) in 4G.
*3 Internet Infrastructure Review (IIR) Vol. 38, Focused Research (1) "Why IIJ Seeks to Become a Full MVNO" (https://www.iij.ad.jp/en/dev/iir/pdf/iir_vol38_EN.pdf).
*4 Non-standalone
*5 Standalone
*6 Quality of service
*7 Point of interface

2.3 The VMNO Concept

With the aim of addressing these two issues facing the 5G SA era, IIJ and the Telecom Services Association, an MVNO industry organization, are advocating the concept of VMNOs as a new kind of virtualized telecommunications operator for the 5G era. This original idea arises from a European report.

In a March 2017 report[*8], European think tank CERRE put forward two scenarios laying out a path to European leadership in the 5G space. The first it dubs the "Evolution" image, in which the approach used up until 4G continues in the 5G era. The second, dubbed the "Revolution" image, involves a major break from the conventional approach. Central to the Revolution scenario is the idea of VMNOs, or Virtual MNOs. The report points out that there are too few MNOs to achieve the 5G mission of providing dedicated communications services to a wide and varied array of industries, and that because they are constrained by the physical interface, MVNOs will not contribute with the same level of flexibility over their business. In the Revolution scenario, the MNOs open up an adequate set of APIs for controlling comprehensive 5G network slicing, allowing the market entry of a multitude of VMNOs, which have the same degree of flexibility as MNOs to roll out 5G solutions tailored to specific industries.

CERRE went a step further in a September 2019 white paper[*9], saying that the sort of full MVNO arrangements used up until 4G may no longer be possible on 5G virtualized networks, and it is thus calling for the VMNO concept to be pursued. Figure 2 shows the anticipated structure of the relationship between host MNO and VMNO (we simply refer to this as a "light VMNO" in the figure and this report).

Unbundling under the current generation splits the core network at POIs into an MNO side and an MVNO side. A major difference with the light VMNO setup is that the MNO operates the combined core network integrally itself. The light VMNO only has the OSS/BSS[*10] systems that control operations and business, which access a slice using an API on the MNO's network.

Adopting this structure means that the light VMNO can manage the virtualized core network (i.e., slice) on the MNO's virtualization infrastructure via the API provided by the MNO. Two sets of APIs will be required. One is for managing the QoS of the core network embodied by each slice, meaning the QoS of communications services provided to users. The other is for managing slices themselves, which includes, for example, adding new slices and deleting unneeded slices.

One other VMNO model that IIJ and the Telecom Services Association are advocating is that of the full VMNO. Light VMNOs run their businesses atop the virtualized infrastructure provided by the host MNO. The major difference with full VMNOs, on the other hand, is that they own the virtualized infrastructure in parallel with the host MNO. Figure 3 shows the anticipated structure of a full VMNO.

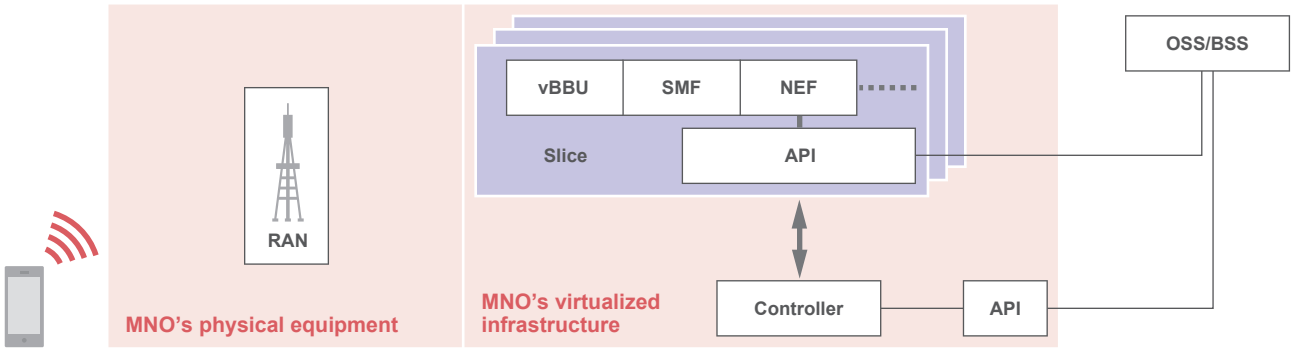


Figure 2: Anticipated structure of a light VMNO

*8 "Towards the successful deployment of 5G in Europe" (https://cerre.eu/wp-content/uploads/2020/06/170330_CERRE_5GReport_Final.pdf).
*9 Ambitions For Europe 2024 (https://cerre.eu/wp-content/uploads/2020/05/cerre_whitepaper_ambitionsforeurope2024.pdf).
*10 Operation Support System / Business Support System

The difference between light and full VMNOs lies in the ownership of the virtualized infrastructure. Light VMNOs rely on the host MNO's equipment except for the OSS/BSS, whereas a full VMNO is independent of the MNO's equipment except for the wireless part. This difference means that full VMNOs have an additional degree of technical and operational independence from the host MNO, making it possible to collaborate with other wireless operators. This is the sort of independence full MVNOs in the current generation have. Full VMNOs are likely to collaborate with multiple 5G wireless networks with their own virtualized core networks.

The Telecom Services Association's MVNO Committee put these VMNO concepts to the Ministry of Internal Affairs and Communications' study group on the competitive environment in the mobile market, which subsequently said in a February 2020 report that both of these VMNO models should be considered as concepts for virtual telecommunications operators in the coming 5G SA era. The VMNO concept has thus become the most prominent option for how virtual telecommunications operators will be set up in the future.

2.4 Benefits of the VMNO Concept

So the VMNO concept is making steady headway, but what benefits will it bring?

In its white paper, CERRE claims that the new market structure brought about by VMNOs has the potential to deliver a vibrant level of competition in both the B2B and B2C markets alike. This is because a large number of VMNOs, relative to MNOs, can be expected to appear as the number of MNOs is limited because of the finite availability of wireless resources and, to take a more macro view, in gradual decline due to industry consolidation. VMNOs, like the current generation of MVNOs, are virtual telecommunications operators that do not themselves receive radio spectrum allocations, so market entry is not restricted by natural conditions such as the scarce availability of spectrum resources. And unlike the current generation of MVNOs, the flexibility of their business will not be restrained by the conditions under which MNOs provide functionality or the depth of unbundling, so they will be able to assemble the functions that their customers need from a broad range of options to provide communications services with the required QoS levels. The presence of VMNOs like this in the market will naturally stimulate competition

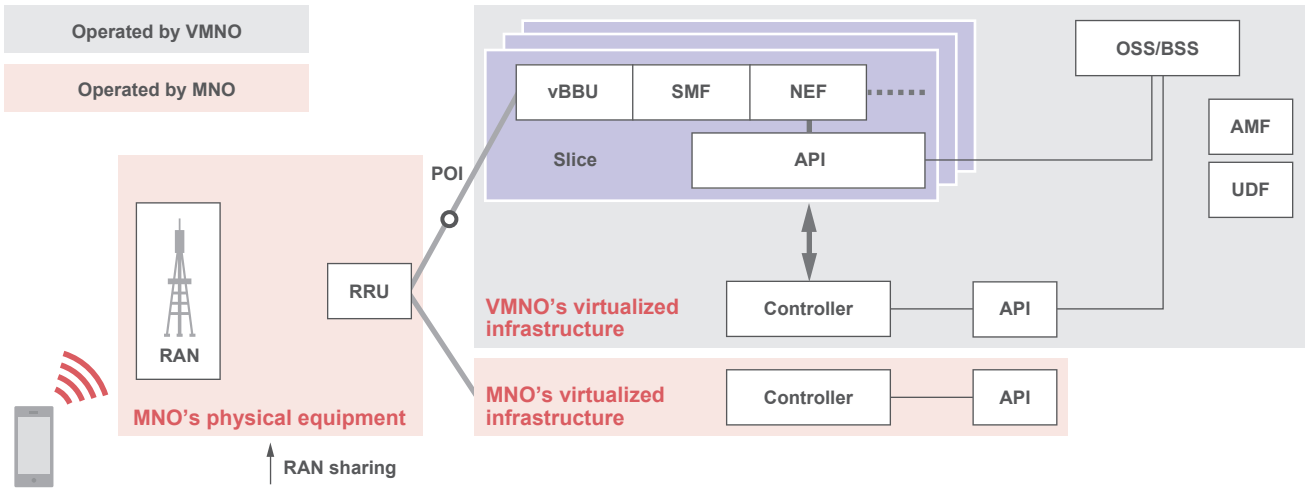


Figure 3: Anticipated structure of a full VMNO

and likely make it even easier for users to obtain the services they need in the 5G SA era.

In Japan, too, the Telecom Services Association's MVNO Committee says that the presence of VMNOs with a high degree of flexibility would accelerate the creation of innovative solutions. This benefit of VMNOs is likely to solve the problem of 5G adoption in markets/regions where 5G adoption rates are expected to be relatively slow, such as among SMEs and in rural areas.

Moreover, full VMNOs with core networks that do not rely on a specific MNO's wireless infrastructure can be expected to play a key role in driving the rollout of "local 5G"[*11], which Japan's Ministry of Internal Affairs and Communications is pushing heavily for. Full VMNOs have all components required by a local 5G operator, including SIMs, devices, virtualized infrastructure, the core network, and the OSS/BSS. And because these are independent of the operation of any specific wireless network, full VMNOs are in a position business-wise to use a whole range of wireless networks without hindrance, so they can meet the requirements of local 5G operators without worrying about sticking with any specific wireless network, which puts them in an unrivaled position. At IIJ, we believe that by being local 5G enablers, full VMNOs will be able to create completely new types of communications businesses between themselves and local 5G operators that want high-quality, low-cost private cellular connectivity on their own sites, prime examples being the owners of stadiums, hospitals, hotels, factories, and the like.

2.5 Challenges on the Path to VMNOs

Still, many challenges exist on the path to making VMNOs a reality. It will no doubt require action on the technical, business, and regulatory fronts. We take a closer look at each below.

Different technical hurdles present themselves for light and full VMNOs. API standardization is an issue for light VMNOs. The creation of VMNOs can easily be facilitated by standardizing the technical interface criteria for the APIs that light VMNOs need. If this standardization is not done, or is lacking, light VMNOs will have to ask MNOs every time they need an API or functionality developed, which would likely pose a tough impediment to VMNOs. An issue for full VMNOs, meanwhile, is that of enabling smooth RAN sharing. RAN sharing, which enables the sharing of a single wireless network across multiple core networks, is already used by some MNOs in Japan and is set to play a key role on the cost front as 5G rolls out ahead. RAN sharing is currently still confined to within MNO groups, but if RAN sharing across MNO boundaries takes off in the lead up to 5G, this could present a good opportunity for full VMNOs, which are likely to participate in that framework. Since these sorts of standardization efforts will not take place in one country, it will also be necessary to develop a globally shared awareness of the issues to be addressed. IIJ is an associate member of Study Group 3 (Tariff and accounting principles including related telecommunication economic and policy issues) within ITU-T, part of the United Nations' specialized agency focused on standardization in the telecommunications sector. In that capacity, we

*11 5G systems that a range of entities, including local companies and local governments, can install in arbitrary locations, such as within buildings or other premises, in accord with the individual needs of the region or industry. 100MHz of bandwidth in the 28GHz band (millimeter wave) has already been introduced, and work is underway with the aim of formalizing arrangements for the remaining 800MHz in the 28GHz band, along with 300MHz in the lower-frequency 4.6GHz band, which makes it easier to construct coverage areas, by the end of 2020.

have already submitted a contribution to the study group that includes the VMNO concept, and we expect the discussion to evolve toward even better international recognition and understanding of the issues ahead.

On the business front, there is a need to reconcile the interests of both MNOs and VMNOs. On the one hand, VMNOs can be seen as partners to MNOs in that they increase the profitability of the 5G infrastructure (base stations, core networks) built by MNOs and help popularize 5G by developing new solutions, but on the other, they are competitors when it comes to marketing their solutions. These sorts of conflicts are something MVNOs have long faced with respect to MNOs, and the players in this space will need to continue working, both in the public eye and behind closed doors, to ensure that good partnerships remain in place in the lead up to 5G.

The biggest challenge on the regulatory front is making a major shift in the operator interconnection model for using other operators' equipment, a system in place since Japan liberalized telecommunications in 1985. The Telecommunications Business Act currently provides two models for the use of other operators' equipment: the interconnections model and the wholesale services model. In the context of MVNO data communications, in particular, the base model is interconnections, which places heavy obligations on the MNO side. The data network rental charges (in other words, the connection fees) calculated based on Ministry of Internal Affairs and Communications ordinance are applied in the case of wholesales services as well, which has meant that MVNOs are able to use an MNO's equipment under the same conditions in both the interconnection and wholesale services models. But in the case of light VMNOs, in particular, there is no POI, that is, no physical point of connection between the operators. And even in the case of full VMNOs, where there will be a POI, there remains a discussion to be had about how the arrangement of such interconnections should be treated under the Telecommunications Business Act and how to think about the connection fees. Issues that will depend on future discussions and debate include whether connection fees should be left up to private-sector negotiations with wholesale services being the only consideration, or whether regulatory intervention should take place with respect to the fees, including how they are calculated and what the upper limits should be.

2.6 Conclusion

The prospects of the VMNO business model hinge on network virtualization in the 5G SA era, and as such it is still an idea for the future and not set to arrive anytime soon. But the time needed to build a completely new business model is on the order of years, as was the case for IIJ with the full MVNO model, so we believe it is crucial to get the discussion started at an early stage. IIJ is doing what it can to move things forward, not just through industry groups but through other initiatives as well. Creating a completely new business format will be no easy task, but we will continue working toward the implementation of the VMNO concept.

Futoshi Sasaki

Deputy General Manager, Business Development, MVNO, IIJ.
Since joining IIJ in 2000, Mr. Sasaki has been engaged in the operation, development, and planning of network services.
He was one of the founding members of IIJ's MVNO project in 2007 and has been in charge of corporate and personal MVNO services ever since.
He is a member of the MVNO Committee of the Telecom Services Association, an MVNO industry group.

Japanese Text Analysis Using Splunk

3.1 Introduction

We adopted Splunk[*1] on IIJ xSP Platform Service/Mail, a large-scale email service with millions of accounts, to extract useful information from the huge amount of logs generated, perform systems analysis, and to protect users from spammers.

We initially used it mainly for searching logs, but our wide uses for Splunk now include automated spam detection using the Splunk Machine Learning Toolkit (Figure 1)[*2] and the streamlining of services operations.

Below, we start by giving some background to our adoption of Splunk. We then go over the Japanese language processing extension we built for NLP in the Splunk Deep Learning Toolkit, which Splunk merged into the toolkit, and describe text mining using it.

3.2 Background to Adopting Splunk

On the IIJ xSP Platform Service/Mail service, customer support center staff perform log analyses in response to end-user requests, including email delivery searches, the display of individual email delivery routes, and Web mail and POP/IMAP/SMTP authentication log searches. The functionality to do this was implemented in ElasticSearch. We were also using ElasticSearch and Kibana for other service operating tools within IIJ. Back when the service was launched, for instance, we used them to identify users generating large levels of input, detect errors, and create reports for customers.

Partly because we sensed certain limitations with ElasticSearch and Kibana, we settled on Splunk when looking to introduce machine learning algorithms to improve the accuracy of spam detection in the aim of further improving

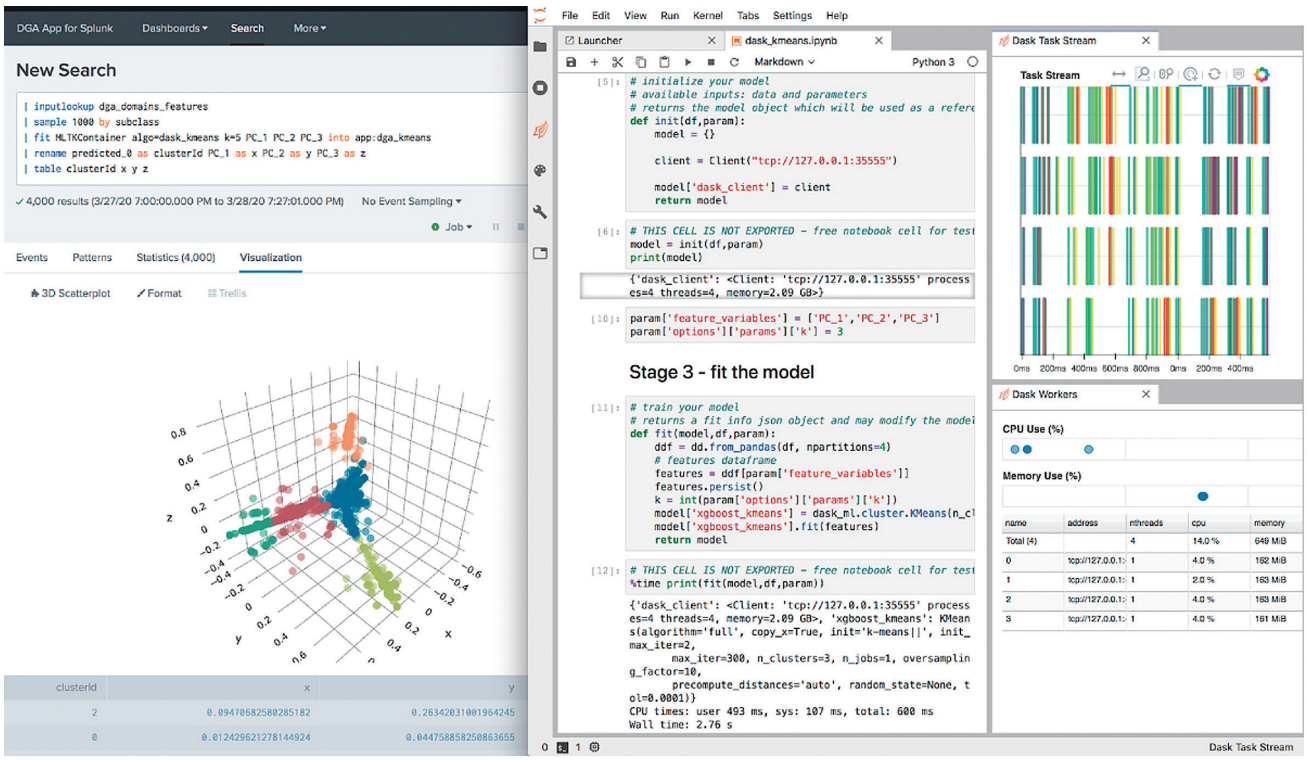


Figure 1: Overview of the Splunk Machine Learning Toolkit

*1 Splunk Enterprise: Find out what is happening in your business and take meaningful action quickly (https://www.splunk.com/en_us/software/splunk-enterprise.html).

*2 Splunk Machine Learning Toolkit (https://www.splunk.com/en_us/blog/machine-learning/deep-learning-toolkit-3-1-examples-for-prophet-graphs-gpus-and-dask.html).

the quality of service on IIJ xSP Platform Service/Mail. Our reasons were that Splunk has a wealth of plugin and visualization apps (both free and paid) optimized for a range of purposes as well as the prospect of speedy development, its offers outstanding system stability and maintainability in comparison with ElasticSearch, and the free Machine Learning Toolkit and Deep Learning Toolkit were appealing.

3.3 Using Splunk for Spam Detection

To improve accuracy using machine learning, in addition to selecting an algorithm, you also need to select axes for analysis, adjust the algorithm parameters, run the learning process, and repeatedly test the model. The Splunk Machine Learning Toolkit and Deep Learning Toolkit provide a user interface that lets you do this seamlessly, and we were able to evaluate algorithms and improve model accuracy in a short period of time.

Spam uses a variety of techniques to blend in among legitimate users. And because activity attributes differ depending on the spam, you need to take an overall view when detecting it (Figure 2).

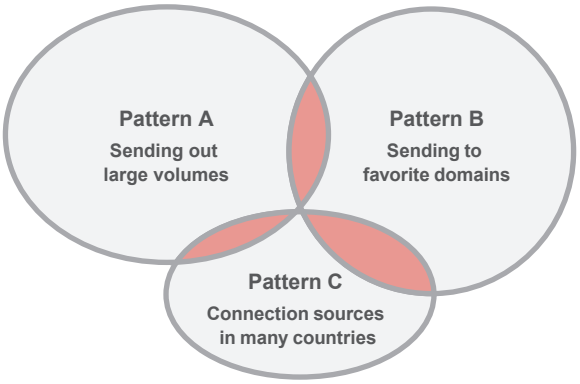
On IIJ xSP Platform Service/Mail, we evaluated a number of algorithms and combinations of variables, including number of source IPs, number of source countries, number of emails sent within a certain time frame, number of unique destinations, whether emails are being sent mainly to domains that spammers like to target or whether emails are also being sent to other domains in a similar manner, and frequency of sending errors. We obtained good results with SVMs[*3]. SVMs are supervised learning models that exhibit robust prediction performance and allow the use of n-dimensional hyperplanes. They also use margin maximization to find the boundary that represents the largest separation from each of the classes being considered.

3.4 The Need for Japanese Text Analysis and NLP (Natural Language Processing)

We have been working to create value-added by analyzing the various logs generated by our services to obtain data useful in the operation and running of those services. Beyond the feature analysis of spam sampled from specific points, we have also heard from other teams here at IIJ that they are, for example, having trouble dealing with abuse or reading in Redmine tickets for analysis, so evidently there is also a need to analyze Japanese text data within IIJ itself.

Applying NLP to text data from abuse emails and Redmine tickets and performing analysis along axes such as “people” and “equipment” makes possible the early discovery of, for example, where loads and problems are concentrated.

Splunk can do morphological analysis using MeCab, but processing large amounts of text data and performing advanced



Example: The actual spammers are those that match several patterns, as indicated by the colored-in regions in the figure

Figure 2: Spammer Activity

*3 SVM: Support Vector Machine, a type of machine learning algorithm.

text mining with this alone is difficult. So we thought about using NLP from the Splunk Deep Learning Toolkit. Using NLP makes it possible to do sentence structure analysis, extract named entities, and so forth, and we were heavily drawn to the prospect of being able to read in large amounts of text data for mining. Named entity extraction is a technique that seeks to identify named entities (objects that can be denoted with a proper name) in text and classify them according to predefined attributes into categories (entity types) such as people, organizations, place names, dates, and numbers (Figure 3).

At the time we started testing, the Splunk Deep Learning Toolkit did not support Japanese NLP, so we had to create our own extension for Japanese, which we published on Splunkbase, Splunk's official library. The extension has now been merged into the Splunk Deep Learning Toolkit. Adding Japanese NLP support to the Splunk Deep Learning Toolkit and thereby broadening the scope for its use in business in Japan generated a considerable response from people. I had the opportunity to give a presentation at a GOJAS (Go Japan Splunk User Group) event to an audience of over 100 (Figure 4).

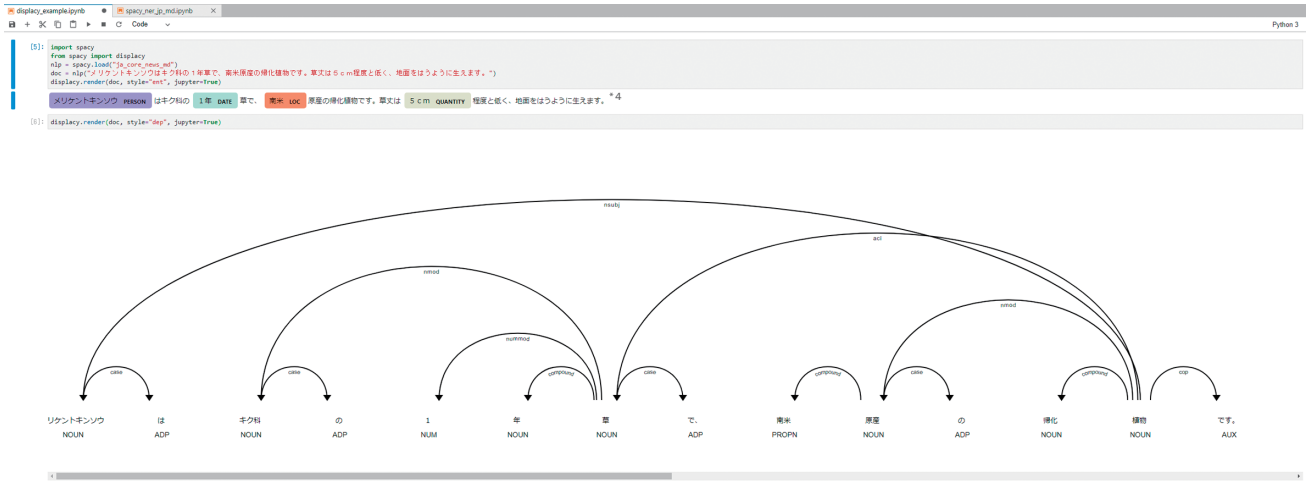
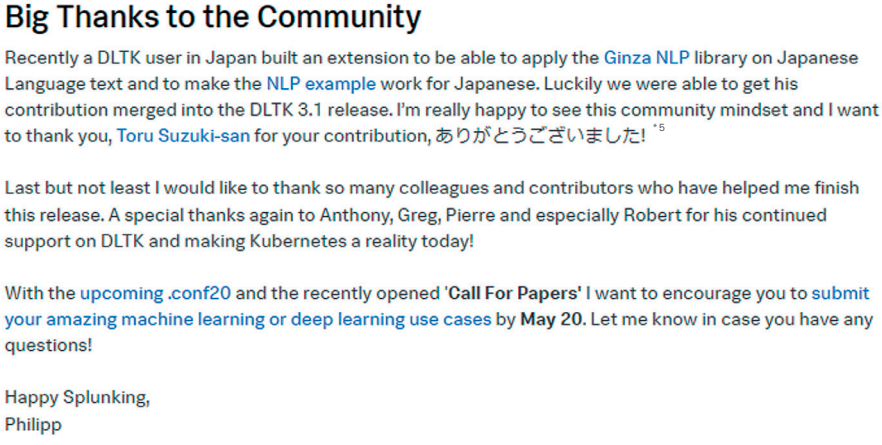


Figure 3: Example of Named Entity Extraction in Jupyter



Big Thanks to the Community

Recently a DLTK user in Japan built an extension to be able to apply the Ginza NLP library on Japanese Language text and to make the NLP example work for Japanese. Luckily we were able to get his contribution merged into the DLTK 3.1 release. I'm really happy to see this community mindset and I want to thank you, Toru Suzuki-san for your contribution, ありがとうございました! *5

Last but not least I would like to thank so many colleagues and contributors who have helped me finish this release. A special thanks again to Anthony, Greg, Pierre and especially Robert for his continued support on DLTK and making Kubernetes a reality today!

With the upcoming .conf20 and the recently opened **'Call For Papers'** I want to encourage you to submit your amazing machine learning or deep learning use cases by **May 20.** Let me know in case you have any questions!

Happy Splunking,
Philipp

Figure 4: Message from the Splunk Deep Learning Toolkit on Our Contribution*6

*4 For English-speaking readers, the output of the displacy.render function is translated from the original Japanese.

*5 The Japanese text in the message reads: Thank you very much!

*6 splunk.com, "Deep Learning Toolkit 3.1 - Examples for Prophet, Graphs, GPUs and DASK" (https://www.splunk.com/en_us/blog/machine-learning/deep-learning-toolkit-3-1-examples-for-prophet-graphs-gpus-and-dask.html).

3.5 Text Mining with NLP (Natural Language Processing)

Text mining with NLP involves getting an overall picture of a passage of text and performing feature extraction based on information obtained by analyzing the relationships between words and extracting named entities.

NLP in the Splunk Deep Learning Toolkit works in conjunction with Jupyter running in a Docker container, with the algorithms implemented in spaCy, a Python NLP library.

Entity	Entity_Count	Entity_Type	Entity_Type_Count
183万円	150	MONEY	42
1億円	96	MONEY	42
5月5日	96	DATE	55
92%	95	QUANTITY	108
日本	87	GPE	15
1万円	63	MONEY	42
9割	63	PERCENT	20
100%	56	QUANTITY	108
250万円	54	MONEY	42
100万円	52	MONEY	42
15分	49	TIME	16
1つ	43	QUANTITY	108
100人	42	QUANTITY	108
4000万円	41	MONEY	42
10分間	36	TIME	16
100%	34	PERCENT	20
火	33	DATE	55
11年	32	DATE	55
30万人	32	MONEY	42
第2267号	32	ORDINAL	10
800人	31	QUANTITY	108
橋本純樹	31	PERSON	45
3000万円	30	MONEY	42
92%	29	PERCENT	20
ワンクリックスキャル24/7完全無料公開中	28	PRODUCT	19
1割	25	PERCENT	20

Table 1: Named Entity Recognition Results for Spam Sampled at a Fixed Point on May 1, 2020 (Entity column translated from the original Japanese)

To enable the processing of Japanese text, we customize the Docker container image, upgrading to spaCy 2.3.2 and installing language models with the Japanese models added in.

The named entity recognition algorithm is written in a Jupyter notebook as so is easily customizable.

Table 1 shows the results of analyzing text data from a single day (May 1, 2020) of spam sampled at a set point using the named entity recognition algorithm we extended. We use the ja_core_news_md model (see https://spacy.io/models/ja for details). Entity denotes a named entity, Entity_Count is the number of times the named entity appears, Entity_Type is a collection of entities having similar attributes defined in the model, and Entity_Type_Count is the number of occurrences of that Entity_type.

The analysis extracts entities such as people (PERSON), monetary amounts (MONEY), place names (GPE), dates (DATE), times (TIME), and quantities (QUANTITY). It is worth noting that strings representing PRODUCT entities are extracted without being broken into separate words.

The table is sorted in descending order of Entity_Count, and the Entity_Type column shows that MONEY entities occupy top spots in the list, indicating that a lot of the spam on this day contained content relating to monetary amounts.

The names are extracted without being split into first name and surname, which is a great advantage when performing analysis along the personal name axis. Since named entity recognition lets us classify large amounts of text data by personal names or product names, it could, for example, be used to analyze operating status or turn text-based knowledge into a database.

Next, to see what differences appear between samples of spam taken in February and May 2020 at a fixed point, we graph the top 15 named entity recognition from those samples. Figures 5 and 6 show the results.

The entity English:LANGUAGE ranked at the top in February and was a clear standout in relative percentage terms as well, perhaps reflecting that overseas travel was still happening during the early stages of the COVID-19 situation. In May, once Japan had declared a state of emergency, English:LANGUAGE had fallen heavily in the ranking, being replaced by MONEY entities, which had also increased substantially in absolute terms, indicating a rise in spam activity.

The analysis of text data is hindered by the lack of classifying information and difficulty nailing down analysis axes, but using named entity recognition like this lets us classify text data using the attributes of named entities, and this makes it a highly valuable technique.

And using the combination of named entity and attribute classification makes it possible to identify overall patterns in text, which greatly opens up the possibilities for text mining.

Entity	Entity_Count	Entity_Type
橋本純樹	31	PERSON
佐々木千恵	22	PERSON
エリオット	17	PERSON
プロスペクト	17	PERSON
橋本	17	PERSON
佐々木	15	PERSON
トニー野中	9	PERSON
北条	9	PERSON
良彰	9	PERSON
アダム	8	PERSON
ロスチャイルド	8	PERSON
倉持	8	PERSON
サトー	7	PERSON
木村	7	PERSON
村岡	7	PERSON
よしあき	5	PERSON
ベール	5	PERSON
ザラ	4	PERSON
スキャルロジック	3	PERSON
たかはしよしあき	2	PERSON
カリスマ美人	1	PERSON
友宮真	1	PERSON
堀崎むつみ	1	PERSON
塚弥生	1	PERSON
大元大輝	1	PERSON

Table 2: PERSON Entities Found Using Named Entity Recognition on Spam Sampled at a Fixed Point on May 1, 2020 (Entity column translated from the original Japanese)

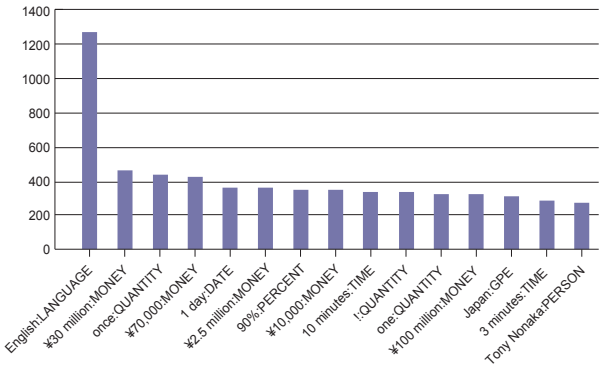


Figure 5: Graph of Top 15 Named Entity Recognition Results for February 2020

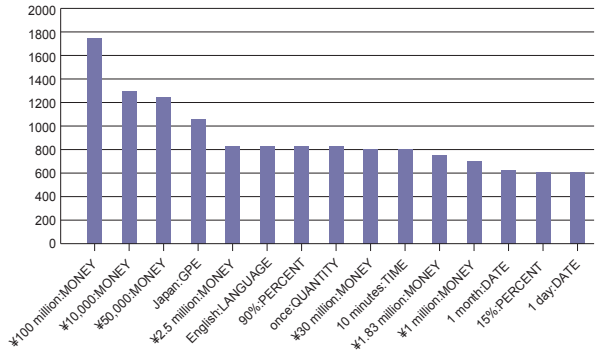


Figure 6: Graph of Top 15 Named Entity Recognition Results for May 2020

3.6 Business Use cases of NLP and Text Mining

Text mining is generally used to discover potential/latent needs based on data accumulated from various text data sources.

Voice data can also be converted to a text data source using external voice-to-text APIs, so voice data accumulated from call center operations and the like can also be used in, for example, customer insight analysis and knowledge extraction for business operations. There exist use cases that involve building a database of examples from text data and matching them by searching for similar patterns, and these approaches are used not only in needs discovery but also in applications such as performance evaluations based on content similarity.

Other companies use text mining in their service operations, an example being the use of a chatbot to serve as the primary contact in a text chat or voice chat, with the interaction being escalated to a human service representative if necessary based on an analysis of the text data generated from the chat. This service approach is used successfully in call center operations, for example, as a labor-saving measure intended to reduce costs.

3.7 Conclusion

In the past, the difficulty in making use of large amounts of text data relegated it to dark data status, but advances in the accuracy of natural language processing have now opened up a wide range of uses for text mining that make it possible to discover useful information.

There are also tools like the Splunk Deep Learning Toolkit that provide a seamless interface for performing natural language processing on text from accumulated data and doing everything from generating models through to text mining. With text mining in the spotlight of late, perhaps now is the time to start using it in your business.

Toru Suzuki
Senior Engineer, Service System Development for xSP Operations Section, Application Service Department, Network Cloud Division, IIJ.
Mr. Suzuki is an administrative member of GOJAS (Go Japan Splunk User Group).
He is engaged in efforts to use Splunk to generate value-added in services.

November 2020 Vol.48

Internet Initiative Japan

About Internet Initiative Japan Inc. (IIJ)

IIJ was established in 1992, mainly by a group of engineers who had been involved in research and development activities related to the Internet, under the concept of promoting the widespread use of the Internet in Japan.
IIJ currently operates one of the largest Internet backbones in Japan, manages Internet infrastructures, and provides comprehensive high-quality system environments (including Internet access, systems integration, and outsourcing services, etc.) to high-end business users including the government and other public offices and financial institutions.
In addition, IIJ actively shares knowledge accumulated through service development and Internet backbone operation, and is making efforts to expand the Internet used as a social infrastructure.

IIJ-MKTG020-0046

Internet Initiative Japan Inc.

Address: Iidabashi Grand Bloom, 2-10-2 Fujimi, Chiyoda-ku,
Tokyo 102-0071, Japan
Email: info@iij.ad.jp URL: https://www.iij.ad.jp/en/