IIR

Internet
Infrastructure
Review

Jun.2024
Vol. 62

Periodic Observation Report

SOC Report

Focused Research (1)

Research on a Method of Constructing Sender Reputation

Focused Research (2)

IIJ and the Evolution of Data Centers —Commemorating 30 Years

IIJ
Internet Initiative Japan

Internet Infrastructure Review

June 2024 Vol.62

Executive Summary

We present IIR Vol. 62, the first edition for 2024. For Japan, the year began with a huge earthquake that hit the Noto Peninsula. We offer our deepest sympathies to everyone affected by the 2024 Noto Peninsula earthquake.

On New Year's Day afternoon, a magnitude 7.6 earthquake with a maximum seismic intensity of 7 destroyed local residents' infrastructure in the blink of an eye. Telecommunications infrastructure was no exception. The four mobile carriers held a joint press conference to disclose the damage caused by the earthquake, including power outages to buildings and base stations and the disconnection of trunk and backhaul circuits. As has been pointed out during previous disasters, the ability to use the Internet via smartphone as one would during normal times is crucial during emergencies. Reports coming out of the press conference, too, highlighted the fact that mobile communications are essential to users, and that electrical power and optical fiber constitute the foundations on which such communications depend.

Physical steps, such as deploying mobile power supply vehicles and generators, are needed to furnish electric power, but in this case, satellite communications stepped in as an alternative to backhaul circuits. Satellite mobile phones have also been used during past disasters, but in the case of the recent earthquake, high-speed data communications using low-orbit satellites were made available to function as base station backhaul circuits and Wi-Fi hotspots, and reports indicate this to have been extremely useful. This experience showed that, even in the broadband era, satellite-based wireless communications can be used to supplement optical fiber and terrestrial wireless communications.

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.

Chapter 1 presents our SOC Report, our periodic observation report for this edition. As usual, this report looks at security incidents on which IIJ's SOC focused its attention from among those that occurred in 2023. This is followed by two discussions looking at the SOC's efforts to address certain challenges. The first looks at challenges with the Data Analytics Platform that the SOC operates and the introduction of a tool called dbt. The second discusses a method of enabling the visualization of C&C communications with a focus on time interval variability. I think you will find both to be interesting accounts of some of the actual work that IIJ's SOC carries out.

Our focused research report in Chapter 2 discusses a method of constructing sender reputation based on a paper titled "Sender Reputation Construction Method And Feedback Loop Using Sender Authentication," presented at the Information Processing Society of Japan. Email has been in widespread use since the Internet's early days, and that remains true today, and so there are strong social reasons for maintaining the security of email. The method proposed here only uses the results of sender authentication. There is no need to look at the content of email, which makes the method valuable from a privacy perspective as well. The team behind this method obtained good results when testing it on real-world emails too, which seems to validate the effectiveness of send authentication technology here.

The focused research report in Chapter 3 continues our series commemorating IIJ's 30-year history, this time with a look at data centers. The article walks through IIJ's data center initiatives through the years, from the 1990s and the dawn of the Internet era through to the 2020s and the rise of demands for carbon neutrality and AI in the data center. Technology has evolved substantially over the past 30 years, and the Internet and society at large have undergone huge changes. Data centers are an important piece of the infrastructure that has underpinned those changes, and we look back on those 30 years with the strong sense that they will continue to play that role in the future. I encourage you to read through.

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's MVNO Council, stepping down from that post in May 2023. In June 2021, he also became a vice-chairman of the association.

SOC Report

1.1 Introduction

IIJ maintains the wizSafe security brand and works constantly to create an environment in which the Internet can be used safely. One of its initiatives in this regard is the regular dissemination of information on security issues in blog form via wizSafe Security Signal[*1]. IIJ also uses its position as an ISP to perform data analysis on its Data Analytics Platform[*2], which aggregates backbone traffic as well as security appliance logs, focusing its efforts on preventive measures and swift follow-up responses against the increasingly sophisticated cyberattacks out there.

Section 1.2 of this report looks back at security incidents that occurred in 2023 and provides a security summary covering key incidents of note in calendar format. Section 1.3 then discusses our latest efforts to deal with emerging challenges faced when performing data analysis and operating the Data Analytics Platform under the headings "Challenges in operating and performing analysis on the Data Analytics Platform and the deployment of dbt" and "Visualizing C&C communications with a focus on time interval variability". As part of this discussion, Section 1.3.1 describes the events that led us to deploy dbt to address the issues of data freshness and data mart creation and maintenance. And in Section 1.3.2, we share the results of our evaluation of the use of a statistical measure called the moving coefficient of variation to deal with three issues related to the evaluation of communications time interval variability.

1.2 2023 Security Summary

Tables 1 and 2 show the security incidents that the SOC focused on from among those that rose to prominence in 2023.

*1 wizSafe Security Signal (https://wizsafe.iij.ad.jp/).

*2 Internet Infrastructure Review (IIR) Vol. 38 (https://www.iij.ad.jp/en/dev/iir/038.html).

Table 1: Incident Calendar (January–June)

Month	Summary
January	Access to a local government's official website was disrupted, and it was revealed that a group calling itself Anonymous may have been to blame for the disruption (DDoS attack). A user claiming to be Anonymous posted content on Twitter (now X) hinting at the attack.
January	IIJ's SOC observed an increase in attack emails with Microsoft OneNote attachments. Multiple malware infection campaigns using OneNote files were seen from January onward.
February	Multiple organizations reported ransomware campaigns (ESXiArgs) exploiting an OpenSLP heap-overflow vulnerability (CVE-2021-21974) in VMware ESXi products. IIJ's SOC also confirmed there to be an increase in scanning on port 427, which is used by OpenSLP, around the same time.
February	A crypto asset exchange disclosed that its employees had been subject to a social engineering attack. The attacker targeted multiple employees via smishing and attempted to log into the exchange's internal network using stolen employee credentials but was blocked by multi-factor authentication (MFA). The attacker attempted to keep the attack going by calling the employees and posing as an IT staff member, but the exchange's CSIRT was able to intervene and foil the attack.
February	Ransomware group Clop claimed to have stolen data from over 130 organizations by exploiting a zero-day vulnerability in the file transfer tool GoAnywhere MFT. The vulnerability is identified as CVE-2023-0669. Thereafter, several organizations, including Japanese companies, announced that they have been impacted.
March	The JPCERT Coordination Center (JPCERT/CC) issued an alert on mailouts aiming to infect systems with Emotet, which had not been observed since November 2022. In some cases, the attached ZIP files contained a Word file that exceeded 500MB once extracted, which is likely a contrivance designed to avoid detection by security products(Note 1).
March	A foreign company that develops business communications software revealed there to be a risk of information-stealing malware infections from the official installer of software it provides. An investigating security vendor subsequently reported that the company had suffered a supply chain attack, compromising its software development environment, which led to other supply chain attacks.
March	A telecommunications carrier disclosed that customer information associated with personal internet services and video streaming services had been leaked from one of its contractors. A follow-up report revealed that the cause of the breach was not a malware infection but that a former temp employee who worked on the services at the contractor had exfiltrated the information.
March	A company that provides information communications and system integration services disclosed that an incident occurred in which a local government certificate issuing service it provides was issuing certificates to residents other than the certificate applicants. The service was temporarily suspended, but in the subsequent process of fixing the system program that caused the incident, it was found that the patch had not been applied at some local governments, so the company said it was carrying out inspections and moving quickly to apply the patch.
April	An IT company that works on public services as a contractor disclosed that servers used in solution services it provides for local government meetings had been subject to unauthorized access. The unauthorized access prompted several local governments to report a temporary suspension of Internet-based meeting live-streaming services.
May	A major automaker's business strategy firm disclosed that customer information, including vehicle data and footage captured by drive recorders, had been made externally accessible due to a cloud misconfiguration. A subsequent investigation of other cloud environments revealed new cases of some customer information having been made accessible.
May	Progress Software disclosed that the web application of its MOVEit Transfer file transfer service contained an SQL injection vulnerability (CVE-2023-34362), and that the vulnerability had already been exploited. The company initially advised customers to check for indicators of unauthorized access over "at least the past 30 days", but other security organizations subsequently disclosed that exploitation dated back to more than 30 days prior.
June	Fortinet disclosed that the SSL-VPN functionality of FortiOS and FortiProxy contained a heap-based buffer overflow vulnerability (CVE-2023-27997), and that it may have been exploited in some limited cases. Multiple security vendors reported that the attack group Volt Typhoon uses a zero-day vulnerability in Fortinet products to gain initial access, but it was revealed that no evidence of the heap-based buffer overflow vulnerability being used had been found.
June	From June, there were ongoing reports of people who had made bookings through a travel booking site receiving messages directing them to phishing sites. This was traced back to unauthorized access to the accommodation booking information management system, and it was revealed that devices used to manage the system at some accommodation facilities were infected with malware. Multiple security vendors reported that the attacker was posing as a customer and sending enquiry emails and the like to get operators to open attachments that would infect the devices with information-stealing malware.

Note 1: Alert Regarding Re-emergence of Emotet Malware Infection Activities (https://www.jpcert.or.jp/english/at/2022/at220006.html).

Table 2: Incident Calendar (July–December)

Month	Summary
July	A port transportation group announced that its terminal system was down due to a ransomware infection and that it was working to restore it. After the outage, media outlets reported that the group had received threats from an attack group going by the name Lockbit. Operations were resumed in short order, over the course of about two and a half days.
July	Microsoft disclosed that the attack group Storm-0558 had gained unauthorized access to Exchange Online and Outlook.com and was able to access emails and personal accounts of around 25 organizations, including government agencies. It revealed that the group had used a Microsoft Service Account (MSA) signing key to forge authentication tokens that enabled it to access the services. While it did not have any concrete evidence as to how the group obtained the signing key, Microsoft issued a subsequent report that described likely methods.
August	Media outlets reported that an Indonesian man had been arrested for using a phishing kit called 16shop to steal and fraudulently use other people's credit card information. This was the first case in which Japan's National Police Agency was involved in a joint cyber investigation with another country: it worked with INTERPOL and Indonesian police to arrest the suspect.
August	There were multiple incidents of login screen tampering on mobile network-compatible IoT routers used in Japan. IIJ's SOC confirmed that the affected login screens contained content protesting the discharge of treated water from nuclear power plants. An account on X thought to be involved in the tampering posted an explanation of the vulnerability exploited.
September	A security company revealed that when AI researchers published open-source AI models on GitHub, a misconfiguration of the tokens that grant access to the public data meant that 38TB of data, including researchers' computer backups, was also exposed. It also revealed that the exposed data included passwords to the AI researchers' organization's services, secret keys, and employees' internal messages.
September	It was revealed that a domain used by a remittance and payment service had been listed on a domain registrar's auction site, causing a stir, with people raising concerns of the domain potentially being acquired and misused by a third party. After speaking to the company that owned the domain, some media outlets reported that the incident was due to internal mishandling.
September	The Zero Day Initiative (ZDI), a vulnerability discovery community, disclosed a vulnerability (CVE-2023-42115) in the mail transfer agent (MTA) Exim that could allow remote code execution. The vulnerability had been reported by a security researcher in June of the previous year, and in September, ZDI informed the vendor that it intended to publish it as a zero-day advisory, and subsequently did so(Note 2).
October	Google, Cloudflare, and Amazon Web Services all announced that a large-scale DDoS attack exploiting a vulnerability in the HTTP/2 protocol (CVE-2023-44487) had been observed in August. The vulnerability exploits the HTTP/2 protocol's stream multiplexing, which enables the concurrent processing of multiple HTTP requests and responses within a single TCP connection. The attacking client sends a high volume of HEADERS frames and RST_STREAM frames to drain server resources. This is called an HTTP/2 Rapid Reset attack.
October	Cisco Systems disclosed a vulnerability (CVE-2023-20198) in the web UI feature of Cisco IOS XE. By exploiting this vulnerability, an attacker can obtain the hightest level of access without authentication and create new local users. A few days later, it disclosed another vulnerability (CVE-2023-20273) in the web UI feature that would allow local users to execute commands with root privileges. Reports of damage from attacks combining these two vulnerabilities were reported in Japan as well, and IIJ's SOC observed traffic related to this vulnerability.
October	A contact center company disclosed that a person involved in the operation and maintenance of the call center system had been exfiltrating customer information and leaking it to third parties. This breach is thought to have been ongoing for around 10 years. The root cause was that it was possible to download customer information on operation and maintenance devices and write it to external storage media, and these operations were not detectable via timely detection measures or routine log checks.
October	A company that provides identity management and authentication services disclosed that its support case management system had been subject to unauthorized access. It initially said that only files uploaded by some customers had been affected, but it subsequently revealed that all users had been affected. An investigation into the account used for unauthorized access revealed that an employee had signed in to a personal Google account via a device managed by the company, and the account information had been synced to the employee's personal device, suggesting that the account may have been stolen because the personal device was compromised.
November	It was disclosed that ownCloud, open-source software for building online storage, contained a vulnerability (CVE-2023-49103) that allows confidential information and settings in containerized environments to be leaked. Some days after this was revealed, several organizations reported that the vulnerability was being exploited.
November	Akamai's SIRT disclosed that a Mirai variant called InfectedSlurs was using a zero-day vulnerability in attacks designed to build out DDoS botnets. A follow-up report revealed that one of the products with this vulnerability is a wall-jack-compatible wireless LAN router sold in Japan, and that it is possible to exploit the authentication details in this device's factory default settings. IIJ has observed DDoS attacks emanating from such botnets as ongoing since November.
December	Microsoft and Arkose Labs announced they had seized infrastructure used by Storm-1152, a group that sells fraudulent Microsoft accounts and CAPTCHA bypass tools, identified the ringleaders, and filed criminal charges with law enforcement. It was also revealed that Storm-1152 had sold around 750 million fraudulent Microsoft accounts, and was providing tutorial videos and support chat services for the tools it was selling.
December	Security consulting firm SEC Consult disclosed an SMTP implementation vulnerability called SMTP smuggling. The SMTP protocol defines an end-of-data sequence, but many products will recognize a sequence as signaling the end of data even when an email message deviates from the standard sequence. So by adding a second email message in the data following the sequence, an attacker can, for example, bypass sender spoofing checks that would normally occur per the SPF sender authentication method.

Note 2: (0Day) Exim AUTH Out-Of-Bounds Write Remote Code Execution Vulnerability (https://www.zerodayinitiative.com/advisories/ZDI-23-1469/).

1.3 Security Topics

This section discusses the IIJ SOC's efforts to tackle some of the challenges it faces.

1.3.1 Challenges in operating and performing analysis on the Data Analytics Platform and the deployment of dbt

At IIJ, we operate a data platform that we call our Data Analytics Platform, with the objective being to implement appropriate preventive measures and post-incident responses to cybersecurity threats. For additional details on these initiatives, refer also to the SOC Report in IIR Vol. 38 (https://www.iij.ad.jp/en/dev/iir/038.html), published March 2018. We continue to operate the Data Analytics Platform, continuously rolling out software updates and expanding data sources. Here, we describe some challenges that have emerged through the operation of the platform and our efforts to introduce open-source software called dbt (data build tool) to address them.

Before delving into the challenges, an explanation of the underlying data platform operations and architecture is in order. Note that the discussion here is not limited to the Data Analytics Platform. Firstly, in general, work on the data platform is performed by team members serving in the following two types of roles.

- Data engineer
- Data analyst

While more fine-grained distinctions may be used at times, here we will say that a data engineer is someone who creates the data platform, and a data analyst is someone who uses the data platform. With IIJ's Data Analytics Platform too, we break the overall staff up into a team responsible for development and a team responsible for analysis, and those teams collaborate with each other.

The current mainstream approach is to manage the data stored on data platforms using the following layered structure, and we manage data on our Data Analytics Platform in a similar manner (Figure 1).

- Data lake layer
- Data warehouse layer
- Data mart layer

The data lake layer stores data collected from or transferred by data sources, which generate the data, and to the extent possible, keeps it in the form in which it was received. The data warehouse layer transforms data from the data lake layer into structured data to facilitate analysis. The data mart layer transform data from the data warehouse layer into specific forms tailored to specific use cases.

The data mart layer will be relevant to the challenges discussed below, so an example from the domain of security is in order. Say that data collected from an IPS/IDS on a particular data platform is imported into the

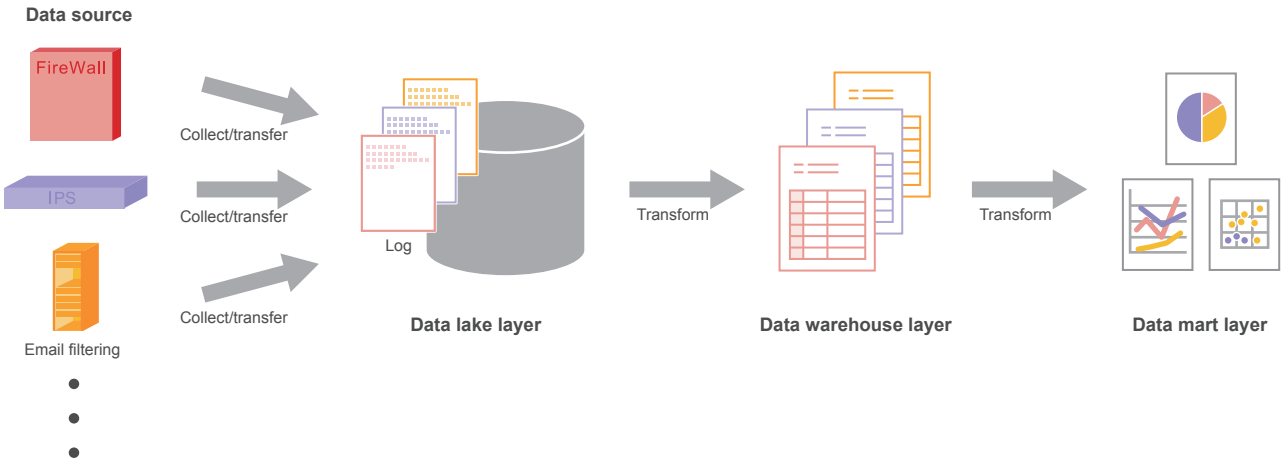


Figure 1: Illustration of How Data is Managed in a Layered Structure on a Data Platform

data warehouse layer. And let's say that a particular data point (record) contains information about the signature of and time at which a threat was detected. If you want to look at daily trends based on this data, you might consider producing daily counts of the type and number of detected signatures. In this scenario, data counts that have been generated in advance, rather than created ad hoc by a data analyst, constitute a data mart. For important information that is frequently looked up, you might also consider adding visualization features using a dashboard or the like.

Next, we look at two challenges that have emerged in the course of operating and performing analysis on the Data Analytics Platform. The first relates to data quality. For data analysts to perform appropriate analyses, data on the platform must remain in a healthy state. Yet there is no such thing as a perfect system, and on our Data Analytics Platform, we experienced the following types of events.

- Decline in freshness of certain data
- Unexpected circumstances befalling certain data

The concept of data freshness concerns whether data is imported into the data platform in a timely manner without any delays. A decline in data freshness occurs when data is imported into the data platform slower than expected or when the data flow temporarily stops. A decline in data freshness may mean that data analysts are unable to perform timely analysis[*3].

Saying that unexpected circumstances can befall data is somewhat vague, but what this means is that, from a data analyst's perspective, something appears to be amiss with the data. Easy-to-visualize examples include values that are supposed to be unique appearing multiple times and unexpected values popping up in the data[*4]. Overlooking these sorts of phenomena can affect the reliability of analysis results.

The second challenge relates to creating and maintaining data marts. Data marts are crucial when it comes to routine tasks that are performed over and over again. This is because the presence or absence of a data mart can greatly impact on work efficiency.

On the other hand, when creating data marts on the Data Analytics Platform, there was a tendency for engineers to develop them individually. With this style of development, significant resources are spent on everything from development through to maintenance. And the lead time required to actually put the data mart into use tends to be long.

To address this challenge on our Data Analytics Platform, we deployed the open-source version of dbt, called dbt Core. A SaaS version called dbt Cloud is also available. dbt is specifically geared to the T (data transformation) in ETL (extract, transform, load), the basic functionality of a data platform. One of its key traits is that most features are available simply via SQL and YAML configurations[*5]. The fact that SQL and YAML configurations are all that is needed is quite a boon for data analysts, who tend to use SQL extensively when performing analyses.

The first challenge related to data quality is addressed by testing the data using dbt. Automated testing is a best practice in software engineering. And this is also a valid approach in data engineering.

dbt includes functionality for testing data, and the tests can also be written in SQL and YAML. Tests in dbt are SELECT statements that return 0 records in a normally functioning system. In other words, if the SELECT statement for a particular test returns one or more records, the test fails. You can of course write your own SELECT statements from scratch, or you can use built-in test macros and test macros implemented by third-party plugins. Additionally, the source freshness feature can be used to detect when data freshness has declined.

*3 The degree to which a decline in freshness can be tolerated depends on the nature of the work performed by the data analyst.

*4 Situations that actually occur range from abnormalities that are obvious from the data definition to cases that are difficult to detect, such as changes in the distribution of data when viewed through summary statistics.

*5 Differences between databases (SQL dialect etc.) are handled by adapters tailored to each database.

In the past, if something unexpected had happened to data on the Data Analytics Platform, data analysts would typically notice this during the course of performing analyses. But when data analysts perform analyses, they do this with an objective in mind. That is, they usually need that data right at that moment. So discovering something unexpected with the data at that point tended to heavily impact on workflow. Deploying dbt, however, made it possible to efficiently discover and address such occurrences by incorporating data definitions and data analysts' experience-based rules into automated tests.

The second challenge relating to creating and maintaining data marts is addressed by using dbt to transform data with SQL. In dbt, objects called models can be defined using SQL SELECT statements. Results transformed by the model's SELECT statement can be accessed in the form of views and tables. The Jinja templating language can also be used to define models. So it is possible to group frequently used content into macros, and to perform looping and branching that is difficult to accomplish in pure SQL.

Deploying dbt on the Data Analytics Platform has made it possible, in many cases, to create data marts simply by writing SQL. In some cases, we have been able to create data marts for quickly searching for security IoCs (Indicators of Compromise) within a short period of time, making certain tasks dozens of times more efficient. Some processes are difficult to implement using SQL alone, however, so individual development has not been rendered entirely unnecessary. Meanwhile, it has become easier than before to divide labor between data engineers and data analysts in some respects. Specifically, data engineers can implement processes that are difficult to perform using SQL alone as SQL user-defined functions. Data analysts can then write routines for transforming the data in dbt using SQL that includes those user-defined functions. With this workflow, tasks follow an existing framework, so we can expect to reduce resource requirements and lead times from what they were before. And laying out a more concrete path for creating data marts has streamlined our operations and made it easier to come up with new ideas for analysis.

1.3.2 Visualizing C&C communications with a focus on time interval variability

When infecting a system, some types of malware use an Internet-connected command and control (C&C) server to receive instructions from an attacker. The basic behavior of the communications that take place (the C&C communications) is predetermined according to what malware program is running. For this reason, certain patterns are more likely to emerge than those seen in, for example, communications that occur when humans use a browser or other such application in an ad hoc fashion.

One key characteristic that tends to reflect particular patterns is the communication time interval. A typical example that should make this easy to see is polling that occurs at regular intervals, such as a heartbeat used in system health monitoring. Time interval variability tends to be low with these sorts of communications. To enable security analysts to efficiently capture C&C communication patterns, we therefore tried out the use of a statistical measure that focuses on time interval variability to generate visualizations of communications.

It must be noted, however, that small time interval variations can occur even with legitimate, non-malware applications. If you detect small communication time interval variations in a production environment, it's best to assume that most of that is attributable to legitimate applications. The method described here is thus intended to be used to help analysts visually understand what's going on when breaches occur.

First, let's go over the problem and the tools that can help us solve it. C&C communications can be viewed as events that occur over the passage of time. Each event

belongs to a specific session[*6], typically identified by the below elements in combination. The source is the system infected with malware, and the destination is a C&C server set up by the attacker. Note that the source port number is not included in the session identifier because it may be an ephemeral port that changes from connection to connection.

.

- Source IP address
- Destination IP address
- Transport layer protocol
- Destination port number

Next, we turn to the descriptive statistical measures of variance and standard deviation, which is based on variance, to quantify the variability in our figures. However, there are some challenges to address when it comes to assessing variability in communication time intervals with variance and standard deviation using all events in a particular session.

■ (a) Comparing values from different sessions is difficult

Consider the following situation, for example. Say that in session A, the average time interval between events is 100 seconds and the standard deviation (a measure of variability) is 10 seconds. Meanwhile, say that in session B, the average time interval between events is 1,000 seconds and the standard deviation is 50 seconds. If we only compare the absolute values, we will decide that the standard deviation is lower for session A than for session B. But if we consider the magnitude of variability relative to the average, we will find this to be smaller for session B than for session A. Hence, variance and standard deviation alone are insufficient when comparing different sessions.

■ (b) Outliers have a substantial impact

If, for example, a malware-infected computer is used for business purposes, it might only be turned on during weekday daytimes. And naturally, when it's not powered on, no C&C communications will occur. Including the times when the computer was off in the calculations would result in a long interval between events, as if the readings were outliers. The calculated variance and standard deviation are likely to be larger than expected as a result.

■ (c) Event time intervals can change partway through sessions

Some malware shortens the time interval between events when the attacker is actively issuing commands. This is possibly because a lot of malware has autonomy over when it goes to the C&C server to receive commands[*7], so if the polling interval is long, it will take time for operations to be reflected in the malware's behavior. When long and short time intervals are mixed like this, it is difficult to express the variability in terms of a single measure like variance or standard deviation.

To address these challenges, we looked at a statistical measure called the moving coefficient of variation. This takes the ordinary coefficient of variation and introduces the idea of moving it over time. The smaller the moving coefficient of variation of the communications time interval, the more we can say that the communications have low variability and are thus regular.

First, the coefficient of variation is obtained by dividing the standard deviation by the mean. As mentioned in challenge (a) above, the standard deviation alone is not enough to determine variability relative to the mean size of the observation. We therefore obtain a dimensionless

*6 For convenience, we refer to a series of exchanges between any particular nodes as a session. This is different from a TCP session.

*7 This is possibly because if the communications originate from a C&C server on the Internet, they are likely to be hindered or blocked by NAT and firewalls.

measure by dividing the standard deviation by the mean. This makes it possible to compare the variability of different sessions.

Next, in statistics, a moving measure is something that is calculated using a moving set of local (as opposed to global) data points. Such measures are used mainly for time series and other continuous data sets and are based on a predetermined number of data points. The range covered by that predetermined number of data points is called the window. The window is progressively moved along the data set so that all of the data is eventually used.

So the moving coefficient of variation is obtained by moving the calculation of the coefficient of variation along the data set. The moving coefficient of variation is calculated by dividing the moving standard deviation by the moving average. This reduces the impact of the issues discussed in challenges (b) and (c) above. As regards (b), outliers only affect the windows in which they occur. And as for (c), even if the time intervals at which events occur change during the observation period, this should show up as changes in the statistical measure over time.

We will now look at how malware-generated C&C communications from past infection incidents can be visualized using the moving coefficient of variation. The first case deals with malware communicating with multiple C&C servers at the same time. In recent times, malware programs often attempt to communicate with multiple C&C servers at the same time in the manner discussed here.

In Figure 3, time elapsed since the reference time is plotted on the horizontal axis, and the moving coefficient of variation is plotted on the vertical axis. Each line represents a session, all of which constituted communications to a C&C

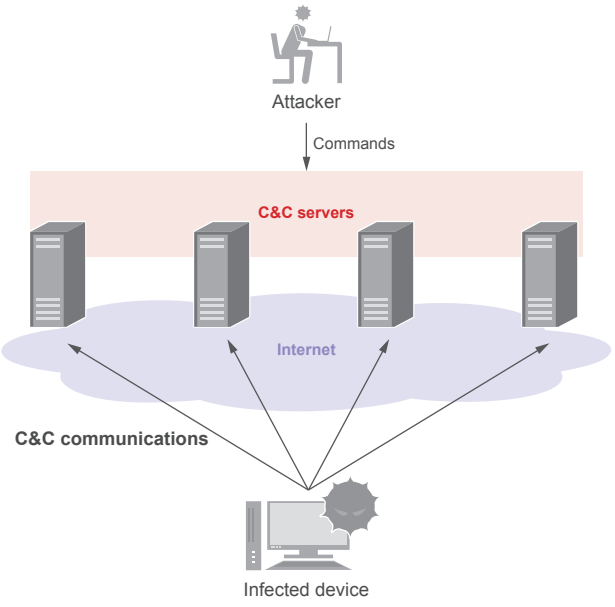


Figure 2: Malware Communicating with Multiple C&C Servers at the Same Time

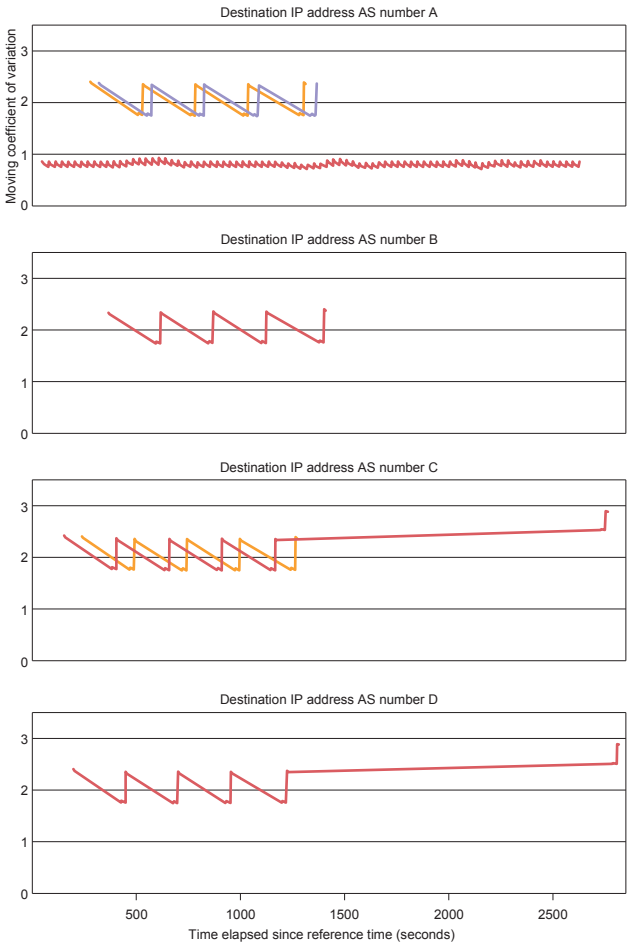


Figure 3: Typical Sawtooth Wave Pattern of Malware Communications to C&C Servers

server. In this case, each session used the same source IP address and transport layer protocol but had different destination IP addresses and port numbers. For ease of viewing, the graph is divided into panels by the AS number of the destination IP address. The window size used to calculate the moving coefficient of variation was 10 data points.

The graphs in each panel of Figure 3 show the typical sawtooth wave pattern, oscillating around a value of 2, in the moving coefficient of variation. The waves are not perfectly aligned with each other, but they are close. This means that attempts to communicate with C&C servers happened at similar times, even across different C&C servers.

The second example shows malware falling back to another C&C server when the C&C server it was originally communicating with becomes unavailable. While this malware only communicates with one C&C server at a time, it likely has the ability to switch to a different C&C server if it loses contact with its current C&C server for some set amount of time.

In Figure 5, as before, time elapsed since the reference time is plotted on the horizontal axis, and the moving coefficient of variation is plotted on the vertical axis. Each line again represents a session, and the window size used to calculate the moving coefficient of variation is 10 data points here also. Note, however, that elapsed time is calculated in minutes here, as opposed to seconds in the previous graphs.

Two sessions are plotted on the graph, and you can see times during which the moving coefficient of variation remained stable at close to zero in both cases. In other words, there are times when communication was occuring at regular intervals such that there was very little variability in the communication time interval. In fact, during these times, the malware was repeatedly communicating with a C&C server at intervals of around 317–320 seconds.

In the session indicated by the orange line, the moving coefficient of variation increases from around the 200-second mark[*8] and the line subsequently cuts out. A little while after this, the purple line appears. This is

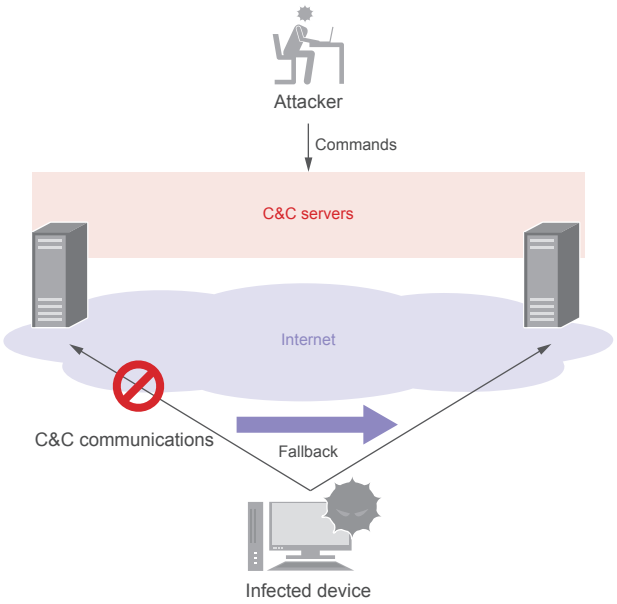


Figure 4: Malware Falling Back to Another C&C Server

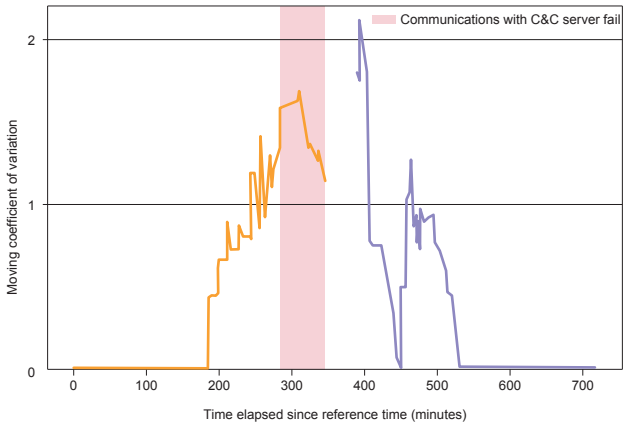


Figure 5: Plot of Communications When Malware Falls Back to Another C&C Server

*8 An increase in the moving coefficient of variation indicates a breakdown in the regularity of communication intervals partway through the session. This corresponds to challenge (c) above.

what it looks like when the malware falls back to another C&C server. In the orange session, communications with the C&C server had failed about an hour before the line cut out. The malware, it seems, therefore switched to sending its communications to a different C&C server.

Note that the examples covered here are typical ones and that there is a whole range of malware communication patterns in the wild. It is not uncommon, for instance, for malware to vary communication intervals depending on which C&C server it is communicating with, or to introduce some degree of randomness. These sorts of behaviors are likely attempts to avoid detection based on communications time interval variability. We have observed that, even in such cases, the shape of the graphs of the moving coefficient of variation exhibit similarities, and that the coefficient will move within a particular range.

Here, we have described our work to enable visualizations of communications traffic using a statistical measure that focuses on time interval variability. IIJ's SOC will continue to study techniques for efficient data-based security analysis.

1.4 Conclusion

In this article, we discussed several incidents that our SOC focused on out of those that we observed in 2023. The annual summary in Section 1.2 indicated that software vulnerabilities continue to come to light, and that ransomware attacks as well as information breaches due to configuration errors and external attacks continue to occur. Section 1.3.1 described how we used dbt to address issues related to data freshness and quality and the creation and maintenance of data marts through automated testing and the use of SQL for data conversion in a way that does not require individual development. Finally, Section 1.3.2 discussed how we arrived at the idea of using the moving coefficient of variation, along with some actual use cases, as a means of addressing challenges that arise when assessing the variability in communications time intervals based on variance and standard deviation—namely, the challenges of comparing different sessions, addressing the impact of outliers, and addressing changes in time intervals.

IIJ will continue to publish information to address the ever-changing array of threats out there. We hope that you will continue to turn to the IIR and wizSafe Security Signal for such information and that it will prove useful to you in your security responses and operations.

Junya Yamaguchi
Data Analytics Section, Security Operation Department, Advanced Security Division, IIJ

Hiroyuki Kamogawa
Data Analytics Section, Security Operation Department, Advanced Security Division, IIJ

Satoshi Kobayashi
Data Analytics Section, Security Operation Department, Advanced Security Division, IIJ

Research on a Method of Constructing Sender Reputation

2.1 Introduction

Twenty years ago, in January 2004, IIJ joined MAAWG (Messaging Anti-Abuse Working Group), which works on measures to combat spam globally. I participated in the group's founding meeting in April 2004, and I have continued to attend the general meetings since. The group's name has been changed slightly to M3 AAWG[*1], and the scope of its activities has been expanded. The 60th general meeting, marking the group's 20th anniversary, took place in February 2024.

MAAWG technical discussions were initially focused on evaluating and popularizing sender authentication technologies, particularly SPF[*2] and DKIM[*3], to address the fact that it was not possible to accurately identify the sender of an email, which really should be regarded as a flaw in the email system. Those technical discussions continued over the years, with members of M3 AAWG playing a central role in creating technical specifications like DMARC[*4], ARC[*5], and BIMI[*6]. From the start, technical discussions about sender authentication were premised on the idea that, as a next step, we would need to be able to determine if we should accept email using authenticated domain names—in other words, sender reputation. Indeed, the first SPF specification, RFC4408, mentioned domain reputation, and Google and Yahoo in the US have recently been pushing heavily for email senders to get on board with sender authentication to tighten up the way incoming email is handled. Indeed, even in Japan the number of domains supporting DMARC rose sharply after the companies announced these tougher measures. As a visiting researcher at IAjapan, I study domain jp names, and as of February 2024, around a quarter of the domain names used for email had a DMARC record set, indicating a roughly three-fold rise in the proportion of domains with such a record.

The IIJ Research Laboratory conducts research on methods for building sender reputation. This article discusses a paper[*7] published in the journal of the Information Processing Society of Japan. The paper describes a sender reputation construction method and feedback loop. In this article, I focus on the sender reputation aspect of the paper. The paper was also recognized as a specially selected paper by the Information Processing Society of Japan.

2.2 Sender Reputation

DNSBL (DNS Blocklist) has long been used as a mechanism for determining whether to accept email based on sender information. It uses a DNS query to look up the source host's IP address. While the source IP address is not an appropriate way to identify the sender of an email, DNSBL has so far been used because the email address specified in the email headers and during the email delivery process to indicate who the sender is cannot be relied on. With the spread of SPF and DKIM for sender authentication, there are moves to use domain names as authenticated, trustworthy information for determining whether to accept or reject email—this is the concept of domain-name sender reputation.

In addition to domains with a negative reputation from which email should not be accepted, one can also imagine there to be legitimate domain names from which email should be accepted. When quantified, the factors behind this determination result in a reputation score. In simpler terms, reputation can be thought of as the basis for a Block List and Allow List of domain names.

Alongside the rise of sender authentication technology, we have also seen an increase in the volume of spam from actors who have registered their own domain names and properly configured SPF and DKIM. The domain names registered for this type of spam are used as throwaways, so building a Block List based on domain reputation is, unfortunately, not all that effective. It may thus be more effective to use an approach that involves building an

*1 Messaging, Malware and Mobile Anti-Abuse Working Group.
*2 Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1 (RFC7208).
*3 DomainKeys Identified Mail (DKIM) Signatures (RFC6376).
*4 Domain-based Message Authentication, Reporting, and Conformance (RFC7489).
*5 Authenticated Received Chain.
*6 Brand Indicators for Message Identification (Internet-Draft).
*7 Shuji Sakuraba et al., Sender Reputation Construction Method And Feedback Loop Using Sender Authentication, Journal of the Information Processing Society of Japan, Vol. 64, No. 1, pp. 13–23 (2023).

Allow List and combining this with email filters and the like to assess the content of emails not covered by the Allow List. Spam as a proportion of all email has declined from what it was in the past, and so if the bulk of email can be assessed using the simpler and easier methods of sender authentication and sender reputation, then more computing resources can be devoted to assessing the content of email messages.

With this background now laid out, this article describes a method for collecting the domain names of legitimate emails and building reputations on that basis.

2.3 Characteristics of Sender Authentication Technologies

Detailed descriptions of the SPF and DKIM sender authentication technologies can be found in sources such as the Sender Authentication Technology Deployment Manual[*8]. Here, therefore, I focus on parts of the paper related to the method of constructing sender reputation.

SPF authenticates the domain name of an email address representing the sender of an email over the Simple Mail Transfer Protocol (SMTP). On the sending side, an SPF record listing email sender IP addresses and the like is published on the domain's DNS server, and on the receiving end, when an email is received, the receiving server looks up the IP address to determine whether the email is from the correct sender. This mechanism means that implementing SPF on sending servers is relatively easy as all the administrator needs to do is publish an SPF DNS record, and use of SPF thus continues to spread. One problem, however, is that recipients cannot properly authenticate emails when they were sent by an entity other than the original email sender.

With DKIM, the server creates a digital signature from the email header and body for each email sent and affixes this along with other relevant information to the email header. Because it uses an authentication method that does not depend on the route by which an email was delivered, DKIM is not subject to the problems inherent in SPF, such as the inability to properly authenticate forwarded emails. But because sending mail servers need to perform the additional steps of creating the digital signature and adding the DKIM signature information to emails, DKIM has not become as widespread as SPF.

2.4 A Method of Constructing Sender Reputation

Here, I describe a method that uses sender authentication to collect SPF-authenticated domain names from which email should be accepted. In general, it is relatively easy to collect this information in the case of spam because this sort of email is itself unwanted and shouldn't be accepted. The approach has been to collect data for spam filters and block lists by extracting salient characteristics and sender information from the collected spam. When it comes to legitimate emails, however, a challenge is that the messages may contain highly confidential information, making it generally difficult to collect the desired data. And because the information is used to determine whether or not an email should be accepted, erroneously recording an email sender as a spam emitter can cause substantial damage, so accuracy is required when making such entries.

Here, we take forwarded emails to be emails that should be accepted, and describe a method for extracting the senders of forwarded emails and creating a list on that basis.

2.4.1 The Nature of Forwarded Email

Email forwarding is often used as a way of consolidating emails, such as when you use multiple email accounts and want to view them in a single place. This mechanism has long been used in email systems such as the opensource Sendmail, which can be configured to automatically redirect received emails by adding the forwarding address to the .forward file in the user's home directory. Hence, the forwarder's email forwarding settings point to the recipient of the forwarded emails, and so from the forwarded email recipient's perspective, the forwarding email sender can be regarded as an email sender from whom email should be accepted.

*8 Anti-Spam Consultation Center, Japan Data Communications Association, Sender Authentication Technology Deployment Manual (https://www.dekyo.or.jp/soudan/aspc/report.html#dam, in Japanese).

If you can collect a list of such email forwarders, you should be able to construct reputations for email senders from which email should be accepted.

2.4.2 Forwarded Email and Sender Authentication

In basic email forwarding, the email address set by the original sender is used in the envelope-from field[*9], which corresponds to the domain authenticated by SPF. Because of this mechanism, SPF authentication at the forwarded email destination fails. DKIM, meanwhile, does not use the sender's IP address for authentication, so emails to which a DKIM signature has been added can be DKIM-authenticated at the forwarding destination. The results are illustrated in Figure 1. The SDID (Signing Domain Identifier) in the figure is the domain name authenticated by DKIM.

Recently, an increasing number of email-receiving servers are refusing to accept emails that cannot be SPF-authenticated as a means of tightening defenses against email spoofing. For this reason, when forwarding email, some forwarding sources rewrite the envelope-from field to contain the domain name of the forwarding source. When this is done, emails will pass both SPF and DKIM authentication at the forwarding destination. However, the domain names authenticated in this case usually differ. The results are illustrated in Figure 2.

2.4.3 Assessing Forwarded Email Source using Sender Authentication Results

I have explained that there are two email forwarding methods and that the SPF and DKIM authentication results differ across those methods. Thus, we use the sender authentication results to determine whether an email has been forwarded, and collect this as reputational information on the forwarded email senders. First, we identify forwarding sources that do not rewrite the RFC5321.From field when

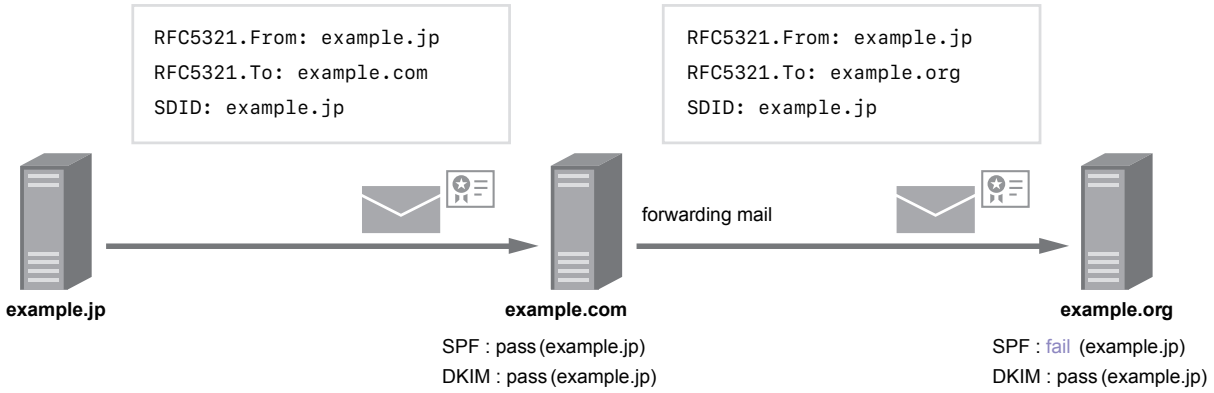


Figure 1: Results of Sender Authentication of Forwarded Email

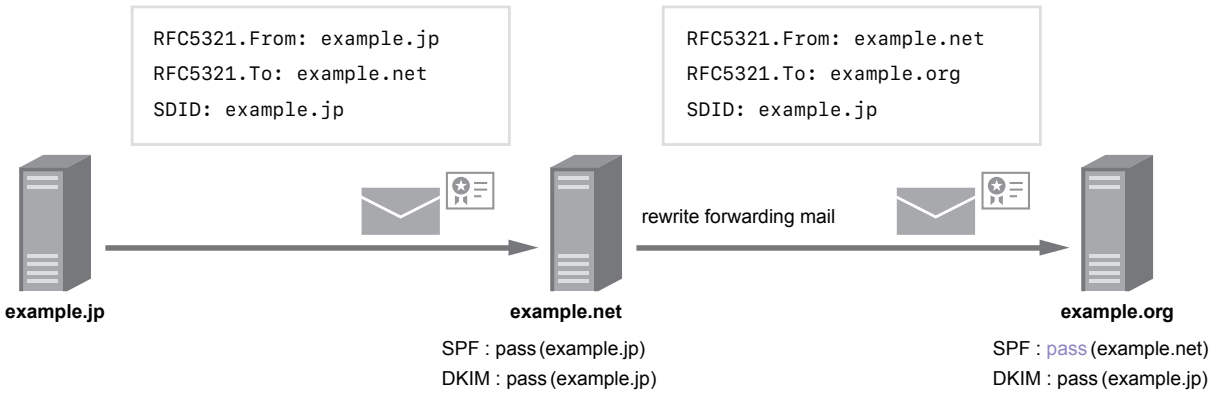


Figure 2: Sender Authentication of Forwarded Email with Sender Information Rewritten

*9 The sender email address according to the email delivery protocol (SMTP). May appear as RFC5321.From, referring to the SMTP RFC.

forwarding emails based on the following authentication criteria, and thus build a list of SPF-authenticated domains from which email should be accepted.

- Sending IP addresses for which SPF authentication fails and DKIM authentication passes
- SPF-authenticated domains that send email from the above IP address and for which SPF authentication passes

With the first criterion, we collect the source IP address of forwarded emails. There can be multiple outgoing mail servers, so to capture a broader range of legitimate email senders, we collect a list of SPF-authenticated domains for which email is sent from the forwarded email's source IP address and for which SPF authentication passes, the objective being to identify the administrative origin of sent emails. This is the second criterion. Both the forwarding IP address and the SPF-authenticated domain name sent from that address constitute sender reputation. Such IP addresses are legitimate sources from which email should be accepted, so any email sent from such addresses, including email that has not been forwarded, is considered acceptable. This makes it possible to use not only IP addresses but also SPF-authenticated domain names for sender reputation.

Next, we identify forwarding sources that rewrite the envelope-from field when forwarding emails according to the following criteria and again collect a list of SPF-authenticated domains from which emails should be accepted.

- Source IP addresses for which both SPF and DKIM authentication passes and the domain names are unrelated
- Of those above, sending IP addresses for which multiple DKIM-authenticated domain names can be obtained via DKIM authentication
- SPF-authenticated domains sent from the above IP addresses for which SPF authentication passes

Outgoing emails that have not been forwarded and that support both SPF and DKIM are expected to be closely related—for instance, usually they will have the same domain name or the same upper domain name. For example, DMARC defines an organizational domain name and assumes that the SPF- or DKIM-authenticated domain name and the sending domain name in the header are the same or have the same organizational domain name. Given this specification, the SPF- and DKIM-authenticated domain names are closely related, even for ordinary email. If the original email sender supports DKIM, and the SPF-authenticated domain name is rewritten when the email is forwarded, it is common for there to be no relation between the original DKIM-authenticated domain name and the SPF-authenticated domain name at the forwarding destination. To identify forwarding sources that rewrite the envelope-from field when forwarding email, we focus on the relationship between SPF- and DKIM-authenticated domain names. To automatically collect these sender IP addresses, we look at emails sent from the same IP address, which pass SPF authentication, and for which the email sender IP address yields multiple DKIM-authenticated domain names. These are determined to be the email forwarding sources. These email forwarding source IP addresses and the SPF-authenticated domain names that they send constitute sender reputation indicating that emails should be accepted.

2.5 Constructing and Verifying Sender Reputationn

To assess the effectiveness of these methods, we constructed sender reputations and applied this to incoming emails. We used the incoming email logs from a real-world email service. This email service performs SPF and DKIM sender authentication when receiving emails, and applies a spam filter to all emails, so the results of these operations are available in the logs. We used the results of this spam filter as the basis for evaluating the judgements made according to sender reputation.

That is, we construct sender reputation from the SPF and DKIM authentication results. Next, we check received email against the sender reputations, compare it with the results of the spam filter, and measure the volume of email classified as non-spam (ham) and as spam.

We constructed sender reputations from around 340 million incoming email log entries for the month of September 2019. At the time, spam accounted for 11.7% of email, the SPF authentication pass rate was 71.1%, and the DKIM authentication pass rate was 38.1%. From these data, we were able to extract 15,169 forwarding IP addresses, 744,660 SPF domain names sent from ordinary forwarding sources, and 11,164 domain names that rewrite the sender domain name when forwarding email.

We applied these sender reputations (indicating that email should be accepted) to the roughly 36 million emails received in the week of October 2019 immediately following the week from which we collected the reputation data (Table 1). We used the same incoming email logs when doing this. Differences between two reputation types are shown below.

(1) Ordinary forwarding sources (IPs) and SPF-authenticated domain names that do not rewrite sender information when forwarding
(2) In addition to (1), sendes (IPs) and SPF-authenticated domain names that rewrite sender information when forwarding

In Table 1, the ham column indicates the proportion of email determined not to be spam by the spam filter for which sender reputation was successfully applied. In other words, this is the true positive rate (TPR). The spam column indicates the proportion of email determined to be spam by the spam filter for which sender reputation was misapplied. That is, this is the false positive rate (FPR). This differs from the meaning of a positive result from an email filter's assessment of spam. Here, a positive result means that email should be accepted, so it is important to be aware of the relationship between true positives and false positives based on reputation.

2.6 Discussion

By assessing email forwarding source using sender authentication technology and constructing sender reputations based on this, we were able to correctly identify around 58% of acceptable email (ham). At the time, the SPF authentication rate was around 70%, so a large portion of that can be attributed to the use of sender reputation. Detecting forwarding sources that rewrite the sender information when forwarding emails and using this to add to sender reputation increased the effectiveness of our method. We were able to increase TPR by over 10 points while holding down the increase in FPR to only 0.25pt. During the period to which we applied our sender reputations, spam accounted for around 9% of received email, so the actual number of false positives was quite low. We also understand, to an extent, why these false positives occurred, so we think it will be possible to further reduce FPR.

This method of constructing sender reputation only uses the results of sender authentication and does not look at the content of email. Even though it is simple

Table 1: Results of Applying Sender Reputation

Reputation	ham(%)	spam(%)
(1)	47.45	3.01
(2)	58.01	3.26

in comparison with common email filtering methods, we still achieved high spam-detection accuracy. We do use the results of DKIM authentication, which does not have a high uptake rate, in identifying the sender of forwarded emails, but the sender does not necessarily have to support DKIM for reputation to be useful, and we only use a few DKIM-authenticated emails to determine forwarding source. So even though DKIM uptake is low, it is possible to construct adequate sender reputations. In constructing and applying sender reputations, we used SPF-authenticated domain names, which are widespread, and if SPF uptake increases further, this should enable accurate determinations about even more emails. If DKIM or DMARC uptake were to increase, we could also consider using those authenticated domain names for the purposes of sender reputation.

The fact that SPF authentication fails at email forwarding destinations has until now been considered a shortcoming of SPF authentication. However, we believe the favorable results we obtain using the method for constructing sender reputation that we describe here—using the features of SPF as a network-based method and DKIM as a digital-signing method—are actually positive for the uptake of SPF.

2.7 Conclusion

With phishing and other forms of spam becoming more sophisticated these days, such that it is hard to tell legitimate emails from malicious ones, we believe that our method, which uses information on sender trustworthiness to determine whether email should be accepted or not, is a significant contribution. The fact that our method of constructing sender reputation does not involve looking at the content of emails also makes it valuable from a privacy perspective. Further, as demonstrated when we tested this method, the fact that it makes it possible to construct sender reputation using, for example, incoming email logs means that it is possible to produce sender reputations geared to the emails that your organization receives, opening up the prospect of greater sorting accuracy. While this method may be inapplicable to a small number of incoming emails, it should make more computing resources available, which could then be used to perform deeper assessments based on the content of those email and so forth.

It has long been the case that emails would be delivered even if you had not deployed some form of sender authentication, and relatively new technologies like DMARC have thus struggled to gain traction. Yet the recent announcement of new countermeasures for incoming email from the likes of Google and Yahoo in the US has prompted more uptake of DMARC, as well as SPF and DKIM, on which it is based. This will make it possible to combat email spoofing while also increasing opportunities to apply measures of domain reputation. We will continue to pursue research relevant to achieving more accurate measures of domain reputation.

Shuji Sakuraba, Ph.D.

Senior Research Engineer, Technology Coordination Office, IIJ Research Laboratory
Dr.Sakuraba is engaged in research and development related to messaging security. He is also involved in various activities in collaboration with related external organizations aimed at bringing about safe and secure messaging environments.He has been a member of M3 AAWG (Messaging, Malware and Mobile Anti-Abuse Working Group) since its establishment. He is the chair of JPAAWG (Japan Anti-Abuse Working Group). He is acting chairperson of the Anti-Spam mail Promotion Council (ASPC) and a member of its administrative group, as well as chief examiner for the Technology Workgroup. He is chairman of the Internet Association Japan's Anti-Spam Measures Committee and a visiting researcher at the association. He is a cooperating researcher at the University of Electro-Communications.

IIJ and the Evolution of Data Centers —Commemorating 30 Years

3.1 1990s—It Started with the Effective Use of Space

Even before the privatization of the Nippon Telegraph and Telephone Public Corporation in 1985 (an event referred to as Japan's telecommunications deregulation), system integrators were engaged in the business of housing other companies' computers in their computing centers in space freed up by computer downsizing. International telecommunications companies were also in the business of housing dealer phone systems and private telephone exchanges for foreign financial institutions in space freed up in their telecommunications spaces by the downsizing of exchange and transmission equipment. Both types of buildings were sturdier than ordinary buildings, and people recall them as the predecessor to today's data center facilities.

Before Japan's telecommunications deregulation, international communications were handled exclusively by Kokusai Denshin Denwa Corporation (KDD, now KDDI), but post deregulation, several new communications carriers (NCCs) entered the market, and computer communications also started to spread, driven by Type 2 Communications Carriers (General, Special), which do not have their own transmission equipment.

IIJ was founded in 1992 and, after much effort, was licensed as a Special Type 2 Communications Carrier and thus permitted to provide Internet services. IIJ's first installation of communications equipment was in the KDD Otemachi Building (now the KDDI Otemachi Building) in Otemachi, within Tokyo's Chiyoda Ward. That was followed by the installation of NSPIXP-1 in the basement of Iwanami Shoten, Publishers' premises in Kanda Jimbocho, Chiyoda Ward, and when Internet exchange (IX) interconnectivity tests began, IIJ set up routers on dedicated lines. Once NSPIXP-2 was subsequently installed at the KDD Otemachi Building, Internet service providers (ISPs) began to congregate at that location.

ISPs use a huge number of dedicated lines and telephone lines, so it makes sense to install network equipment within communication stations and connect them to transmission equipment and exchanges on-site, but NTT's station facilities were not yet made broadly available, and because ISPs needed to connect both domestically and internationally, they naturally decided to make use of the communication stations of KDD and the NCCs. NSPIXP-3 was opened in Osaka, and 1997 saw Japan Internet Exchange Co., Ltd. and Internet Multifeed Co., Ltd. established, marking the beginning of commercial IX services. This was the era in which Japan's underlying Internet framework was formed.

The term data center was still not in use at this point. The services were called "co-location," because equipment was located alongside transmission equipment, and "housing," because they involved housing customer equipment. The rooms were heavily cooled and seemed cold and inorganic, and at the time, a supply of 1–2KVA per rack was quite sufficient.

3.2 2000s—Internet Data Centers

With the explosive growth of Internet connectivity services in the late 1990s, all at once many ISPs began leasing communications station buildings and providing services. Late in the era, in October 1998, IIJ established a joint venture with Sony and Toyota Motor Corporation called Crosswave Communications Inc. (CWC), which set up a nationwide network as an NCC. The following year, it opened access points across Japan.

At the same time, we selected buildings with a fairly large floor load capacity in major Japanese cities (Sapporo, Sendai, Tokyo, Nagoya, Osaka, and Fukuoka) and set up data centers equipped with uninterruptible power supplies and emergency generators. Later, we built our own data centers—encompassing land and structure—in Kawaguchi

Table 1: Differing Features of Computing Centers and Communications Stations

Feature	Computing center	Communications station (telephone & telegraph offices)
Air conditioning	Water cooled (large, general-purpose coolers)	Air cooled (transmission equipment, telephone exchanges)
Power supply	Main supply: AC (3-phase 200V)Generator backup	Main supply: DC (48V DC) CVCF also allows AC supply
Building structure	Designed on par with office buildings Free-access flooring	Designed as a telegraph and telephone office Slab floor
Provisioning	Mainly space leasing	Mainly rack leasing

City, Saitama Prefecture, and Yokohama City, Kanagawa Prefecture. Our guiding concept at the time was to create human-friendly data centers, with integrated people flow lines, cafeterias for break times, conference spaces, kitting rooms, and more. This marked a shift away from inorganic computing centers and communications stations to intelligent buildings for the Internet era.

While it was opening data centers, CWC was also providing wide area LAN services to which customers could connect via Ethernet, which was the standard LAN interface. This communications service (called a "virtual building") allowed users to connect as if on a LAN even when physically distant and use the network as if they were all in the one building. Combined with a fee structure that was independent of physical distance, this service took the world by storm at the time, and it is fair to say that this setup has now become the de facto standard for L2 data services.

As networks and data centers evolved, companies began installing the equipment needed for email services, web services, firewalls for security, and remote access. The facilities attracted not only ISPs but also OTT (over-the-top) businesses such as content providers, and with the advent of rental server services and hosting services, the facilities became not just places to locate equipment but places for providing the functionality to connect such systems to the Internet.

And thus they came to be called Internet data centers. In this era, floor load capacities exceeded one ton per square meter, multiple power feeds were installed, and the power supply and air conditioning equipment was designed to facilitate 4–6KVA per rack. The Internet has long been likened to a cloud, and it was from around this time that Internet data centers would transform into cloud service and connectivity hubs.

3.3 2010s—Accommodating Cloud Services

With cloud services, which provide computing resources over a network, service providers and the like own and operate the servers and other IT equipment essentially on behalf of users. To achieve high equipment densities and efficiently run the huge amount of IT equipment involved, in the 2010s, operators began building hyperscale data centers capable of adequately cooling the equipment and supplying 10KVA or more per rack.

Power receiving capacities were now on the order of 50MW, equivalent to some 10,000 ordinary households (on 100V 50A contracts). The facilities were not only large in scale but also designed to be energy efficient. The rise of hyperscale data centers also saw improvements in an indicator of energy efficiency called PUE (power usage effectiveness; found by dividing the power consumption of the entire data center by the total power consumption of its IT equipment, with a score of 1 being optimal and the Japan average being 1.7). To build data centers that would accommodate cloud service platforms, IIJ developed the IZmo containerized modules capable of providing 10KVA per rack in a high-density environment with outside-air cooling for high energy efficiency, and this culminated in IIJ becoming the first in Japan to operate a commercial outside-air-cooled containerized data center facility in 2011. At the time, IIJ's business model involved reselling data center capacity rented from other companies. We had no experience building our own data centers or developing electrical and air conditioning equipment, so we had to start from scratch. We visited the data centers of industry leaders (GAFA and the like) in North America and determined that outside-air cooling would be the most effective way to reduce power consumption, and we settled on the concept of combining this with containerized modules that would allow us to add capacity in stages. Working with our partners, we designed and built test facilities, and after a year

Table 2: Key Considerations when Constructing a Data Center

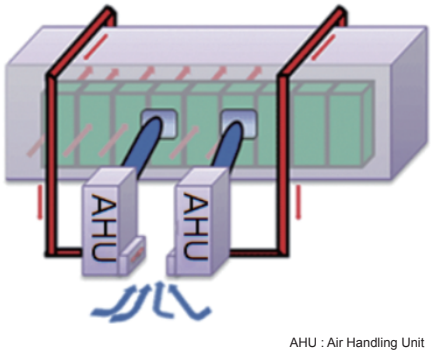
Consideration	Details
Building structure	Earthquake-resistant structure, seismic-isolation structure, seismic-isolation flooring
Power receiving system	3-spot network, receipt of power from substations
Power supply	UPS redundancy (standby redundancy, parallel redundancy, common standby method)
Communication circuits	Multi-carrier, carrier neutral, conduit redundancy
Disaster preparedness	Distance from active faults, avoidance of locations directly under air routes, hazard mapping
Ground	Ground strength (N value), liquefaction potential index
Indicators	PUE, WUE, PLM value, Tier level classification

of testing, we were able to build the Matsue Data Center Park (Matsue DCP).

We named IZmo after the Izumo region, where it was first implemented, and the kanji for "cloud" (which is part of the Japanese spelling of Izumo). IZmo's excellent energy-saving performance and ease of installation saw it not only used in Japan but also exported overseas for use at a hydroelectric power plant in Russia, a national data center in Laos, and elsewhere. And a project is currently underway to provide data centers to Uzbekistan's national telecommunications carrier. With edge computing becoming more widespread, we now offer the modules both in Japan and internationally under the DX edge moniker as a solution for easily and rapidly building edge computing environments and digital/IT infrastructure.

Matsue DCP is a medium-sized center with a power receiving capacity of 4MW, but from the late 2010s, the construction of multiple hyperscale data centers for foreign cloud providers began in earnest. These were initially clustered in the Inzai district around Chiba New Town Chuo Station. This area is so well known as a data center hotspot that the name Inzai is recognized internationally. This clustering can be attributed to a confluence of factors, including the area's solid ground, the fact that it is within 30km of Otemachi, easy access to a submarine cable landing station, and the fact that infrastructure is well developed by virtue of the area having long been a site for computing centers. Even overseas, though, data centers do tend to cluster in certain areas, so Japan is no exception here. Other cluster sites include Osaka's Saito region and the Keihanna area, and with multiple large-scale development plans underway in Japan, this trend can be expected to persist.

As the scale of IIJ's business expanded, in 2019, we also built and began operating a hyperscale data center, dubbed Shiroi Data Center Campus (Shiroi DCC), in the Inzai district (Shiroi City is next to Inzai City). Located on a 40,000sqm site, the facility can be expanded to a maximum power receiving capacity of 50MW. Matsue DCP can be expanded in container increments (nine racks), but to allow for the expansion of Shiroi DCC's service infrastructure in larger increments, we first constructed a system module building with a 1,000-rack capacity, and this is divided into four modules to allow for expansion in stages. The use of outside-air cooling systems on a large scale also means that the facility achieves excellent energy savings on par with Matsue DCP. In addition to outside-air cooling, Shiroi DCC also uses side-flow systems, which are often used in hyperscale data centers. The low-speed, high-capacity airflow of this system is more energy efficient than the conventional method of vigorously blowing cold air out from under the double floor and creates a comfortable working environment.



AHU : Air Handling Unit

Outside-air cooled container module concept at initial planning stage in 2009 (excerpt from internal documents)

Shiroi Data Center Campus, launched in 2019

Figure 1: Creating a New Form of Data Center, Something from Nothing

In the 2010s, as cloud services became widespread, the data center industry saw substantial change, including the ongoing build up of hyperscale data centers and the entry of foreign players into the market. In response, IIJ developed container modules from scratch (Figure 1) and was eventually able to build its own hyperscale data center. In 2023, the Shiroi DCC Phase 2 building went into operation, and IIJ is currently working on plans for the facility's Phase 3 building with a view to putting it into operation in 2026. We continue to explore new data center forms as we look further into the 2020s and the prospect of even more drastic change than we have seen so far.

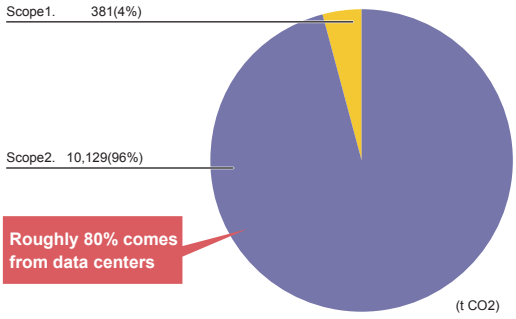
3.4 2020s—Next-generation Data Centers

After all the changes so far, what will we ask of data centers going forward? The data center equipment life cycle is over 10 years, longer than the 3–5 years for IT equipment, so societal and technological trends must be assessed from a medium- to long-term perspective. As calls for sustainability intensify and technological innovations like generative AI, large language models (LLMs), and post-5G move forward, IIJ will be focusing on the following three themes around next-generation data centers.

Carbon Neutrality

The use of data centers helps to save energy across society as a whole through the centralized, efficient operation of IT equipment, but even so, the energy consumption of the facilities themselves is a focus of attention. Japan's revised Energy Saving Act was enacted in 2023, introducing a benchmark system for the data center industry (PUE of 1.4 or less), and requiring operators to submit a plan for converting to non-fossil energy. Companies listed on the Tokyo Stock Exchange's Prime market are also now effectively required to disclose information on climate change risks based on the TCFD recommendations, and thus carbon neutrality is an issue of some urgency for data centers. At IIJ, data centers account for around 80% of greenhouse gas emissions (Scope 1 and 2)*1, and we have set and are working toward goals concerning the use of renewable energy and improvements in energy-efficiency (Figure 2).

The use of renewable energy is a new initiative for IIJ, and takes into account not only the quantity of power but also types of power generation that do not produce CO2 emissions. When it comes to IIJ data centers, we evaluate power procurement methods based on delivery times, cost trends, and the



FY2022 unconsolidated data for IIJ

Calculation method: "Basic Guidelines for Calculating Greenhouse Gas Emissions Throughout the Supply Chain (Ver. 2.3)" (Ministry of the Environment, Ministry of Economy, Trade and Industry)

- **Usage of renewable energy*1**
  The target is to increase the renewable energy usage rate of data centers (Scope 1 and 2) to 85% in FY2030
- **Improvements in energy conservation**
  The target is to keep the PUE of data centers at or below the industry's highest level (1.4)*2 until FY2030 through continuous technological innovation.

Disclosures based on TCFD (Task Force on Climate-related Financial Disclosures) recommendations

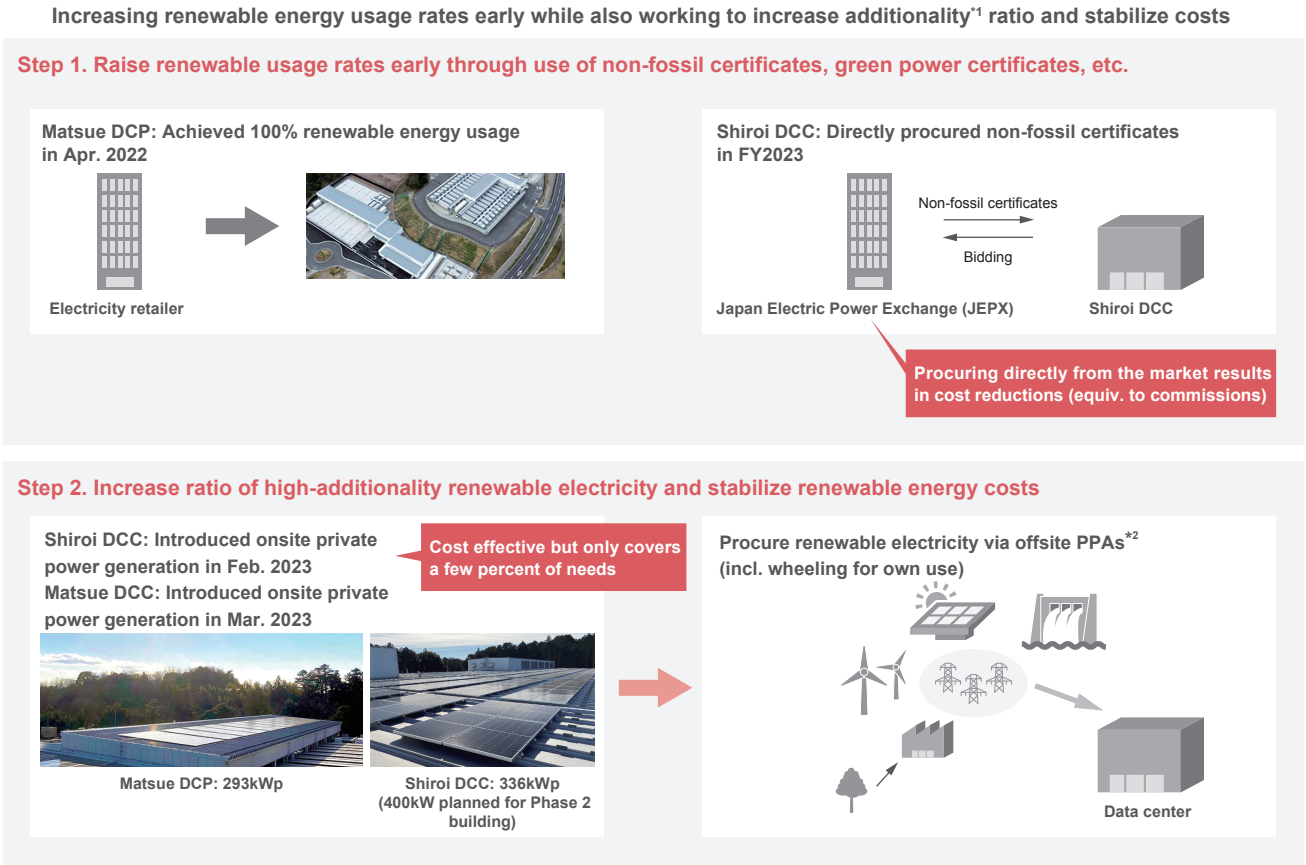
*1. Renewable energy usage includes substantial renewable energy through the use of non-fossil fuel certificates.
*2. As of April 2022, the Agency for Natural Resources and Energy set the benchmark index and target level for the data center industry to be a PUE of 1.4 or lower; businesses that achieve this are regarded as energy-saving leaders.

Figure 2: IIJ's Carbon Neutrality Initiatives

*1 Scope 1, 2 (company's own greenhouse gas emissions): Direct emissions from fuel used or industrial processes, and indirect emissions from energy/heat purchased (GHG Protocol definition).

like, and we are combining multiple methods in an effort to increase our renewable energy usage rates, increase our proportional usage of power with high “additionality,” and stabilize our procurement costs (Figure 3). We have joined the Japan Electric Power Exchange (JEPX)’s renewable energy value trading market, obtaining certificates directly from the market, and we have installed solar power generation systems on our data center premises, reducing our renewable energy procurement costs. We aim to achieve carbon neutrality by combining our traditional approach to energy savings with offsite power purchasing and the like. We have also launched a new service that utilizes these resources to provide environmental value to data center users (Figure 4).

We are also working on initiatives to develop the environmental value trading business, with plans to start providing digital assets (tokens) representing environmental value in July 2024 (Figure 5). And we are developing new roles for the data center to help solve societal issues, such as power grid stabilization and resilience bolstering. For instance, we are part of a VPP (virtual power plant) project to use storage batteries in data centers (launched in FY2023), and we are working with Matsue City, Shimane Prefecture as a co-proposer in the Ministry of the Environment’s Decarbonization Leading Regions program, with plans to use storage batteries to supply power to local communities in the event of disasters.



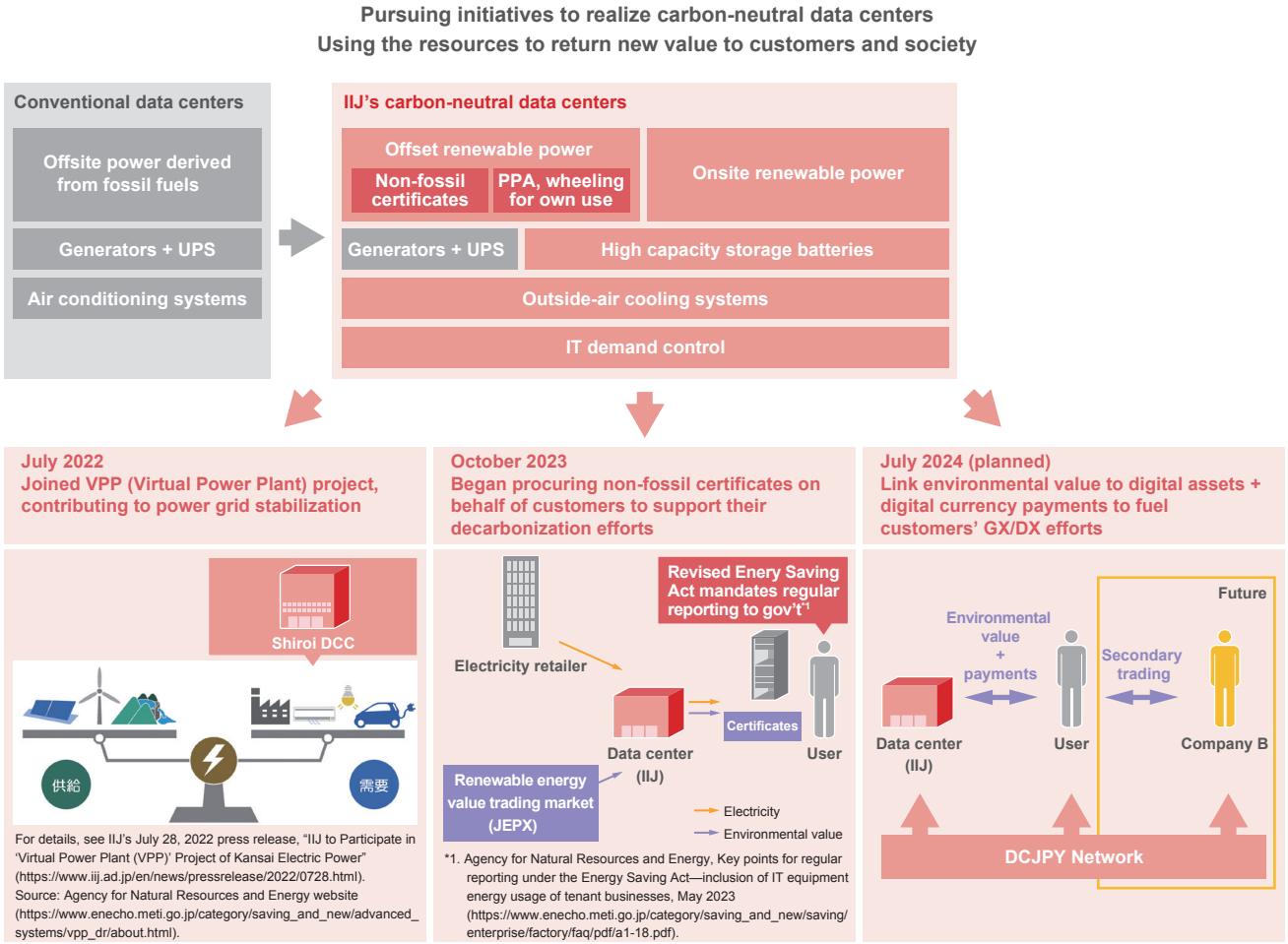
*1. Has the effect of encouraging growth and investment in new renewable energy equipment.

*2. Power Purchase Agreement. Electricity sales agreement between an electricity user (consumer) and an electric power company (PPA operator) that sells electricity to the consumer.

Figure 3: IIJ’s Efforts to Use Renewable Energy

AI-based Control and Automation

Reducing energy usage is crucial to achieving carbon neutrality in the data center. The air conditioning systems used to cool IT equipment account for the greatest share of energy consumed by equipment other than IT equipment like servers. To achieve data center PUEs close to the optimal value of 1.0, we are actively using outside air to cool IT equipment, and we select highly efficient equipment. At Shiroi DCC, we are also using AI technologies to further improve energy-saving performance. We aim to achieve more effective control of the overall air conditioning system by training AI on data from sensors and IT equipment. In addition to automating air conditioning systems, we are also taking steps to reduce operator workloads at data centers. We are, for instance, working to automate the various procedures and tasks that are crucial to providing data center services, a key example being the use of automated reception systems, which automate the procedure of admitting people into data center buildings. We are moving in this direction because, amid the decline in Japan's working population and the country's work-style reforms, we expect it to become more and more difficult to maintain the quality of our product offerings as the scale of our data centers expands if we continue to rely on solely on human resources through the hiring and training of operations personnel. We are looking to progressively expand automation to other areas. For example, with respect to our



For details, see IIJ's July 28, 2022 press release, "IIJ to Participate in 'Virtual Power Plant (VPP)' Project of Kansai Electric Power" (https://www.iij.ad.jp/en/news/pressrelease/2022/0728.html).
Source: Agency for Natural Resources and Energy website (https://www.enecho.meti.go.jp/category/saving_and_new/advanced_systems/vpp_dr/about.html).

*1. Agency for Natural Resources and Energy, Key points for regular reporting under the Energy Saving Act—inclusion of IT equipment energy usage of tenant businesses, May 2023 (https://www.enecho.meti.go.jp/category/saving_and_new/saving/enterprise/factory/faq/pdf/a1-18.pdf).

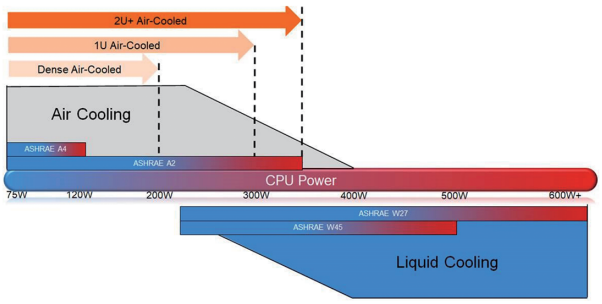
Figure 4: Beyond Carbon-Neutral Data Centers

plans to supply environmental value-linked power to data center users, as described in the previous section, we hope to go beyond simple automation to provide more sophisticated, higher-quality offerings by developing an electricity supply/demand matching platform, introducing digital currency payments, and the like.

Supporting Higher Server Densities and Water Cooling

As the use of technologies like generative AI and LLMs in a range of fields continues to advance and the processing power of CPUs and GPUs increases, next-generation data centers will need to be able to efficiently accommodate huge numbers of these CPUs and GPUs. Ahead, data center CPUs are set to come out with TDP (thermal design power) ratings in excess of 300W, and this is only expected to rise ahead in response to demand for AI. According to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASRAE), once TDP exceeds 300W, facilities will need to adopt water-cooled systems instead of conventional air-cooled systems (Figure 6).

Because they will also house network devices and other such IT equipment that is air coolable, future data centers will need to adopt hybrid cooling systems that combine air and water cooling while also providing excellent energy-saving performance in order to achieve carbon neutrality. IT equipment



Source: Ashrae Emergence and Expansion of Liquid Cooling in Mainstream Data Centers (https://www.ashrae.org/file%20library/technical%20resources/bookstore/emergence-and-expansion-of-liquid-cooling-in-mainstream-data-centers_wp.pdf).

Figure 6: TDP (CPU Power) and Cooling Methods

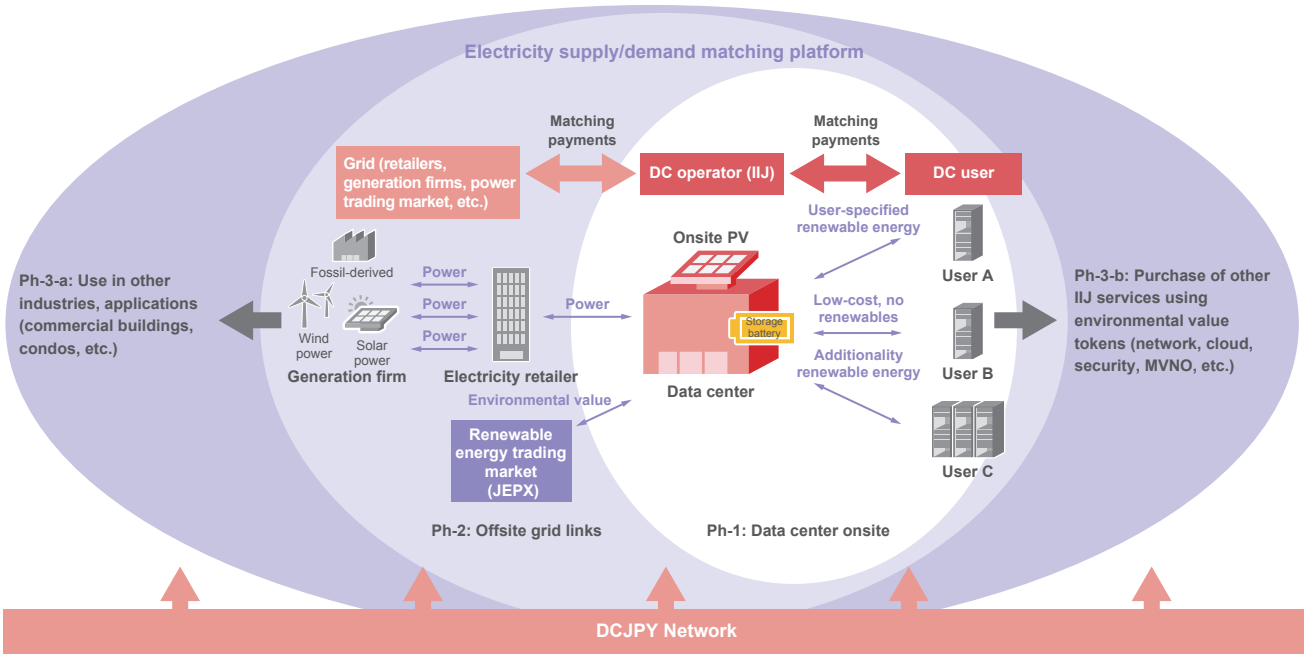


Figure 5: Addressing the Rising Need for Decarbonization

and facilities will become more closely intertwined, and we believe this will be an area in which IIJ will be able to leverage its strengths, given the company's experience building and operating large-scale data centers and cloud services. IIJ, Preferred Networks Inc., and the Japan Advanced Institute of Science and Technology (JAIST) jointly put forward a project, which was subsequently accepted, for research and development into ultra-high-efficiency AI computation infrastructure for inclusion in the Ministry of Economy, Trade and Industry and the New Energy and Industrial Technology Development Organization (NEDO)'s program of research and development of enhanced infrastructure for post-5G information and communication systems, specifically addressing the "Development of post-5G information and communication systems (commissioned project)" category. IIJ is responsible for research and development of basic technology for high-density data centers. We will be working on the development of a high-density data center reference model, establishing hybrid cooling methods that combine air-cooling and water-cooling technologies, and formulating and developing evaluation methods for energy-saving indicators geared to AI computing platforms.

3.5 Conclusion

While the users of IT systems may not often give much thought to data centers, there is growing recognition that data centers are a key component of the infrastructure underpinning a digitalized society. Drawing on our data center journey so far, from computing centers and communications stations, through to colocation services, the cloud, and hyperscaling, we will continue to pursue new forms of data centers and as we go about operating the infrastructure that supports our society.

Yoshikazu Yamai
Managing Executive Officer and Director of the Infrastructure Engineering Division, IIJ

Isao Kubo
General Manager, Infrastructure Services Department, Infrastructure Engineering Division, IIJ
Mr. Kubo joined IIJ in 2008. He oversees the data center business and the construction of Matsue DCP and Shiroi DCC. His aim is to achieve carbon neutrality as soon as possible.

Yusuke Tsutsumi
Data Center Infrastructure Engineering Section, Infrastructure Services Department, Infrastructure Engineering Division, IIJ
Mr. Tsutsumi joined IIJ in 2015 and works on the construction of data centers in Japan and abroad. He is engaged in the development of technologies for next-generation data centers, which includes studying new technologies in the field of electric power.

Takahiro Mimura
Data Center Infrastructure Engineering Section, Infrastructure Services Department, Infrastructure Engineering Division, IIJ

June 2024 Vol.62

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-MKTG020-0060

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/