Internet Infrastructure Review

IIJ
Internet Initiative Japan

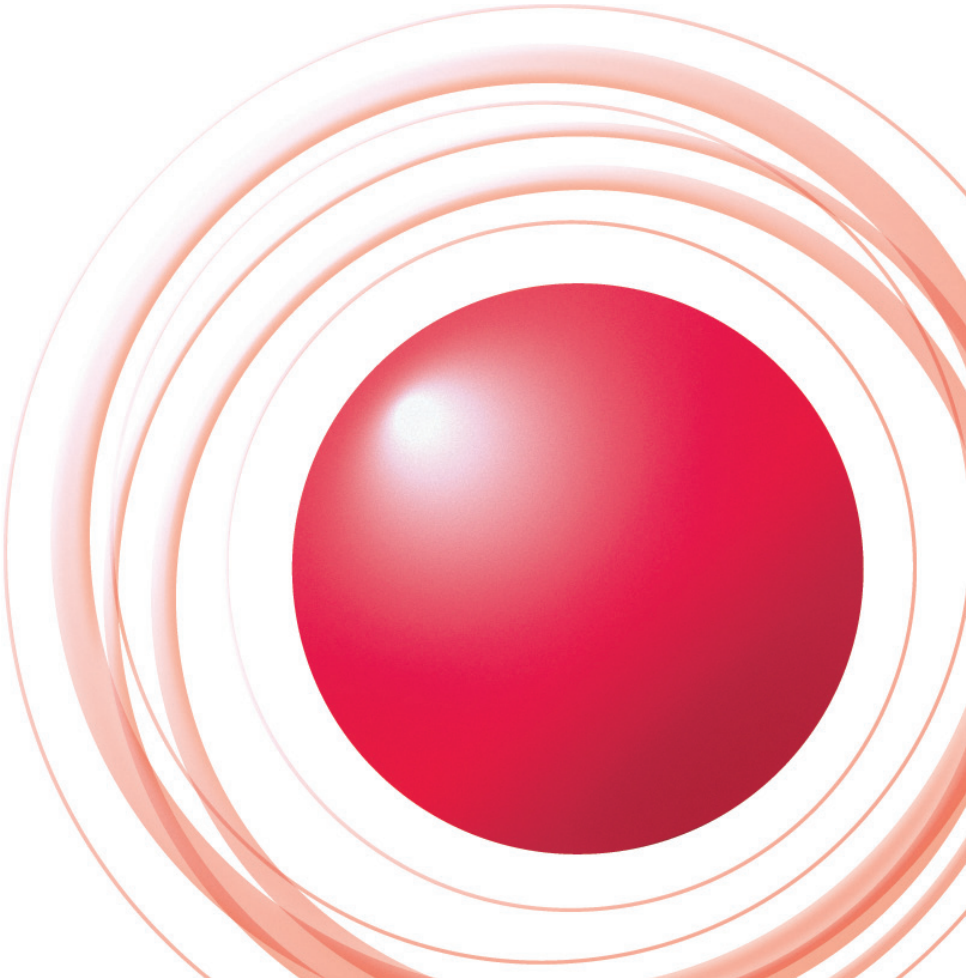
Vol.6

February
2010

Infrastructure Security
Renewed Gumblar Activity

Messaging Technology
The Need for Anti-Spam Measures Tailored
to the Regional Characteristics of the Source

Internet Backbone
Measurement Study on the Internet Reachability



Internet Infrastructure Review Vol.6 February 2010

■ To download the latest issue of the Internet Infrastructure Review, please visit (http://www.iij.ad.jp/en/development/iir/).

Executive Summary

The Internet is not controlled by a single body, as it is an autonomous distributed system that develops gradually as networks link to one another spontaneously. There are some laws governing its behavior, but there is no overall blueprint or scenario by which the Internet operates. It is affected by a number of things, such as trends in the social situation and the world economy, as well as changes in the behavior of users and forms of usage that these result in, and the demeanor and nature of the Internet is constantly changing through the interaction of these different layers.

In order to carry out the stable operation of an infrastructure like this that develops and changes autonomously, it is necessary to measure and analyze its behavior from multiple viewpoints on a constant basis, and to maintain an understanding of what developments are taking place, in order to be able to react swiftly and appropriately. When carrying this out, if appropriate methods are not used to measure and interpret data, the resulting information will be unreliable, making appropriate operation impossible.

For this reason, while it is important to develop technology for building and operating the Internet, we believe it is also crucial to measure operating status, analyze the resulting data to extract meaningful information from it, and create initiatives and systems for applying the results to everyday operation.

This report is published regularly to provide the results of a variety of measurements and analyses that IIJ carries out to maintain and develop Internet infrastructure, in addition to information about related technologies.

In the “Infrastructure Security” section, we report statistics and analyses of security incidents observed for the three months from October 1 to December 31, 2009. We also present focused research covering the details of the Gumblar malware that re-emerged in October and continues to be active, in addition to information regarding the SSL/TLS vulnerability made public in November, and an analysis of techniques for surveying P2P file sharing networks.

In the “Messaging Technology” section, we report on the state of spam trends for the entire year of 2009, and international coordination initiatives that aim to spread the adoption of anti-spam measures. We also provide an overview of the DKIM sender authentication technology that uses digital signatures.

Under “Internet Backbone” we compare results of wide scale Internet measurement using multiple methods, and identify issues with the methods that have been commonly used for measuring Internet reachability to date, in addition to presenting proposals for improvement.

IIJ will continue to publish periodic reports covering information such as this, and provide customers with a variety of solutions for the stable, secure, and innovative use of the Internet as an infrastructure for supporting corporate activities.

Author:
Toshiya Asaba
President and CEO, IIJ Innovation Institute Inc. Mr. Asaba joined IIJ in its inaugural year of 1992, becoming involved in backbone construction, route control, and interconnectivity with domestic and foreign ISPs. Asaba was named IIJ director in 1999, and as executive vice president in charge of technical development in 2004. Mr. Asaba founded the IIJ Innovation Institute Inc. in June 2008, and became president and CEO of that organization.

Renewed Gumblar Activity

In this whitepaper, we will report incidents that occurred between October and December 2009, in addition to commenting on Gumblar-related incidents that have been re-occurring since October, vulnerabilities in the SSL and TLS protocols that are widely used for encrypted communications, and techniques for surveying P2P file sharing networks.

1.1 Introduction

This whitepaper summarizes incidents to which IIJ responded, based on general information obtained by IIJ itself related to the stable operation of the Internet, information from observations of incidents, information acquired through our services, and information obtained from companies and organizations with which IIJ has cooperative relationships. This volume covers the period of time from October 1 through December 31, 2009. In this period the Gumblar malware that steals IDs and passwords re-emerged, and many website alterations related to this have been reported. A series of vulnerabilities related to Web browsers were also discovered, in addition to an issue with the SSL and TLS protocols that are widely used for encrypted communications. Besides these there was also a hijacking incident in which DNS information was manipulated without authorization, and SEO poisoning incidents that took advantage of a natural disaster. As seen above, the Internet continues to experience many security-related incidents.

1.2 Incident Summary

Here, we discuss the IIJ handling and response to incidents that occurred between October 1 and December 31, 2009. Figure 1 shows the distribution of incidents handled during this period*1.

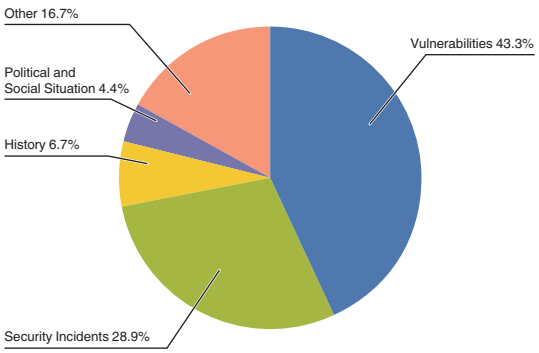


Figure 1: Incident Ratio by Category (October 1 to December 31, 2009)

*1 Incidents discussed in this whitepaper are categorized as vulnerabilities, political and social situation, history, security incident and other.
Vulnerabilities: Responses to vulnerabilities associated with network equipment, server equipment or software used over the Internet, or used commonly in user environments.
Political and Social Situations: Responses to incidents related to domestic and foreign circumstances and international events such as international conferences attended by VIPs and attacks originating in international disputes.
History: Historically significant dates; warning/alarms, detection of incidents, measures taken in response, etc., related to attacks in connection with a past historical fact.
Security Incidents: Wide propagation of network worms and other malware; DDoS attacks against certain websites. Unexpected incidents and related response.
Other: Those incidents not directly associated with security problems, including highly concentrated traffic associated with a notable event.

■ Vulnerabilities

During this period a large number of Web browser-related vulnerabilities were discovered and fixed, including Microsoft's Internet Explorer*2, Adobe Systems' Adobe Acrobat, Adobe Reader*3, Adobe Flash Player, Adobe AIR*4, and Adobe Shockwave Player*5, and Oracle's Java Runtime Environment (JRE)*6. Several of these vulnerabilities were exploited before patches were released*7.

Vulnerabilities were also discovered and fixed in widely used servers such as NTP*8 used for time synchronization, and BIND9*9 DNS servers. Additionally, a vulnerability was discovered in the SSL and TLS protocols*10 that are utilized for encrypted communication by many services. See "1.4.2 MITM Attacks Using a Vulnerability in the SSL and TLS Renegotiation" for more information about this vulnerability.

■ Political and Social Situations

IIJ pays close attention to various political and social situations related to international affairs and current events. During the period under study IIJ observed visits to Japan by foreign VIPs, such as US President Obama in November and Chinese Vice-President Xi Jinping in December, but no related attacks were noted.

■ History

The period in question included several historically significant days on which incidents such as DDoS attacks and website alterations have occurred. For this reason, close attention was paid to political and social situations. However, IIJ did not detect any direct attacks on IIJ facilities or client networks.

■ Security Incidents

Unanticipated security incidents not related to political or social situations included the discovery of malware that infects the Apple iPhone*11. There was also an incident where DNS information for the popular Twitter SNS was manipulated without authorization, causing traffic to be redirected to another website*12. Incidents of users being induced to download fake security software (scareware) through search engine results continued to occur*13.

Additionally, the Gumblar*14 malware that was active on a large scale in April resurfaced at the beginning of October. See "1.4.1 Renewed Gumblar Activity" for more information about this.

*2 Microsoft Security Bulletin MS09-072 – Critical: Cumulative Security Update for Internet Explorer (976325) (http://www.microsoft.com/technet/security/bulletin/ms09-072.mspx).

*3 Security updates available for Adobe Reader and Acrobat APSB10-02 (http://www.adobe.com/support/security/bulletins/apsb10-02.html).

*4 Security updates available for Adobe Flash Player APSB09-19 (http://www.adobe.com/support/security/bulletins/apsb09-19.html).

*5 Security updates available for Shockwave Player APSB09-16 (http://www.adobe.com/support/security/bulletins/apsb09-16.html).

*6 Oracle Corporation, "JavaTM SE 6 Update Release Notes" (http://java.sun.com/javase/6/webnotes/6u17.html).

*7 A zero-day attack is the exploitation of software vulnerabilities for which no fix is available yet. For example, during this period IIJ confirmed incidents of Gumblar and other malware exploiting vulnerabilities in Adobe Reader and Acrobat before a patch was released. These vulnerabilities could be worked around even before a patch was released by prohibiting the use of JavaScript in the settings for Adobe Reader and Acrobat.

*8 Vulnerability Note VU#568372, "NTP mode 7 denial-of-service vulnerability" (http://www.kb.cert.org/vuls/id/568372). By sending specially crafted request packets to an NTP server, it is possible to create an infinite loop repeating responses and requests.

*9 Vulnerability Note VU#418861, "BIND DNS Nameserver, DNSSEC validation Vulnerability" (http://www.kb.cert.org/vuls/id/418861). There is a risk of cache poisoning when using DNSSEC.

*10 Vulnerability Note VU#120541, "SSL and TLS protocols renegotiation vulnerability" (http://www.kb.cert.org/vuls/id/120541).

*11 Details regarding this worm can be found on the F-Secure Corporation blog. "First iPhone Worm Found" (http://www.f-secure.com/weblog/archives/00001814.html).

*12 See the following official blog for details regarding the effects of this attack. Twitter blog, "Update on Last Night's DNS Disruption" (http://blog.twitter.com/2009/12/update-on-last-nights-dns-disruption.html).

*13 SEO poisoning is the act of using a search engine ranking algorithm to display links to malicious sites at the top of search results for certain phrases. For example, the Trend Micro blog below carried out a survey and analysis of words often used in searches during the Christmas season. Trend Micro Incorporated, "SEO poisoning: malicious sites also using SEO marketing?" (http://blog.trendmicro.co.jp/archives/1255) (in Japanese). For the current incidents phrases related to the earthquake that occurred near the Samoan Islands on September 30, 2009 were the target of SEO poisoning.

*14 Gumblar is also explained in Vol.4 of this whitepaper: "1.4.2 ID/Password Stealing Gumblar Malware" (http://www.iij.ad.jp/en/development/iir/pdf/iir_vol04_EN.pdf).

■ **Other**

As far as incidents not directly related to security, a popular Internet message board restricted access from multiple ISPs on a large scale, hindering consumer usage.

Additionally, because many attacks exploiting vulnerabilities in user applications such as Web browser plug-ins occurred during this period, tools for confirming the version of applications and plug-ins have been released (IPA MyJVN Version Checker*15 and Firefox PluginChecker*16, etc.). Microsoft's new Windows 7 operating system was also released, and was hailed for its improved security features.

1.3 Incident Survey

Of those incidents occurring on the Internet, IIJ focuses on those types of incidents that have infrastructure-wide effects, continually conducting research and engaging in countermeasures. In this section, we provide a summary of our survey and analysis results related to the circumstances of DDoS attacks, malware infections over networks, and SQL injections on Web servers.

1.3.1 DDoS Attacks

Today, DDoS attacks on corporate servers are almost a daily occurrence. The methods involved in DDoS attacks vary widely. Generally, however, these attacks are not the type that utilize advanced knowledge such as that of vulnerabilities, but rather cause large volumes of unnecessary traffic to overwhelm network bandwidth or server processes for the purpose of hindering services. Figure 2 shows the circumstances of DDoS attacks handled by the IIJ DDoS Defense Service between October 1 and December 31, 2009.

*15 MyJVN Version Checker enables users to confirm the version of certain software applications installed on the PC they are using (http://jvndb.jvn.jp/apis/myjvn/) (in Japanese).

*16 Visit the following URL using Mozilla Firefox (https://www-trunk.stage.mozilla.com/en-US/plugincheck/). It is also possible to confirm whether or not updates are available by clicking the "Tools (T)" menu, and then selecting "Add-ons (A)." This is a Firefox function, so different methods will be required to confirm plug-ins for other browsers, such as Microsoft Internet Explorer.

This information shows traffic anomalies judged to be attacks based on IIJ DDoS Defense Service standards. IIJ also responds to other DDoS attacks, but these incidents are excluded from the figure due to the difficulty in accurately ascertaining the facts of each situation.

There are many methods that can be used to carry out a DDoS attack. In addition, the capacity of the environment attacked (bandwidth and server performance) will largely determine the degree of impact. Figure 2 categorizes DDoS attacks into three types: attacks on bandwidth capacity*17, attacks on servers*18, and compound attacks (several types of attacks on a single target conducted at the same time).

During the three months under study, IIJ dealt with 185 DDoS attacks. This averages to 2.01 attacks per day, indicating that there was no significant change in the average daily number of attacks compared to our prior whitepaper.

Bandwidth capacity attacks accounted for 0.5% of all incidents. Server attacks accounted for 87.6% of all incidents, and compound attacks accounted for the remaining 11.9%. The largest attack observed during the period under study was a server attack that resulted in 245Mbps of bandwidth using 650,000pps packets. Of all attacks, 77% ended within 30 minutes of commencement, while 23% lasted between 30 minutes and 24 hours. The longest sustained attack lasted for approximately 12 hours.

In most cases, we observed an extremely large number of IP addresses, whether domestic or foreign. We believe this is accounted for by the use of IP spoofing*19 and botnet*20 usage as the method for conducting DDoS attacks.

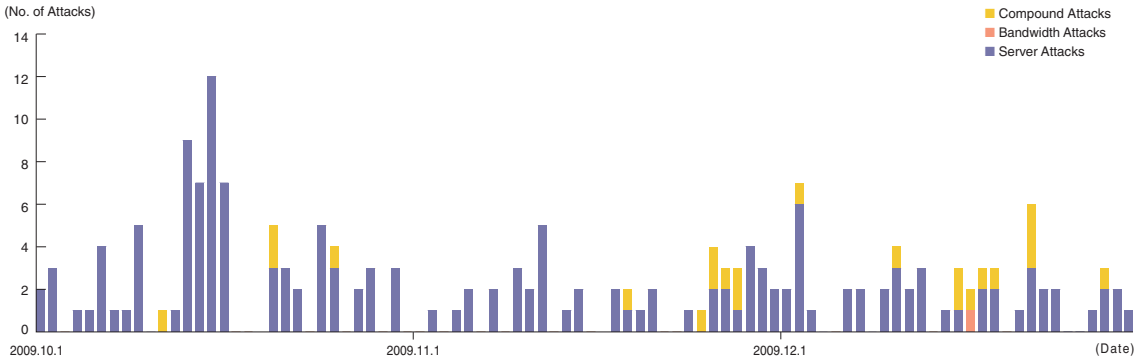


Figure 2: Trends in DDoS Attacks

*17 Attack that overwhelms the network bandwidth capacity of a target by sending massive volumes of larger-than-necessary IP packets and fragments. The use of UDP packets is called a UDP flood, while the use of ICMP packets is called an ICMP flood.

*18 TCP SYN flood, TCP connection flood, and HTTP GET flood attacks. TCP SYN flood attacks send mass volumes of SYN packets that signal the start of TCP connections, forcing the target to prepare for major incoming connections, causing the wastage of processing capacity and memory. TCP Connection flood attacks establish mass volumes of actual TCP connections. HTTP GET flood attacks establish TCP connections on a Web server, and then send mass volumes of HTTP GET protocol commands, wasting processing capacity and memory.

*19 Misrepresentation of a sender's IP address. Creates and sends an attack packet that has been given an address other than the actual IP address of the attacker in order to pretend that the attack is coming from a different location, or from a large number of individuals.

*20 A "bot" is a type of malware that institutes an attack after receiving a command from an external C&C server. A network constructed of a large number of bots acting in concert is called a "botnet."

1.3.2 Malware Activities

Here, we will discuss the results of the observations of the Malware Investigation Task Force (MITF)*21, a malware activity observation project operated by IIJ. The MITF uses honeypots*22 connected to the Internet in a manner similar to general users in order to observe communications arriving over the Internet. Most appear to be communications by malware selecting a target at random, or scans attempting to locate a target for attack.

■ Status of Random Communications

Figure 3 shows trends in the total volumes of communications coming into the honeypots (incoming packets) between October 1 and December 31, 2009. Figure 4 shows the distribution of sender's IP addresses by country. The MITF has set up numerous honeypots for the purpose of observation. We have taken the average per honeypot, showing the trends for incoming packet types (top ten) over the entire period subject to study.

Much of the communications arriving at the honeypots demonstrated scanning behavior targeting TCP ports utilized by Microsoft operating systems. As with the prior study, we observed scanning behavior for 2967/TCP used by Symantec client software and 4899/TCP used by PC remote management tools. At the same time, communications for which the goal was not clearly identifiable, such as 2582/TCP and 31138/TCP (not used by general applications), were also observed. Looking at the overall sender distribution by country, we see that attacks sourced to China and Japan, 22.6% and 20.0%, respectively, were comparatively higher than the rest.

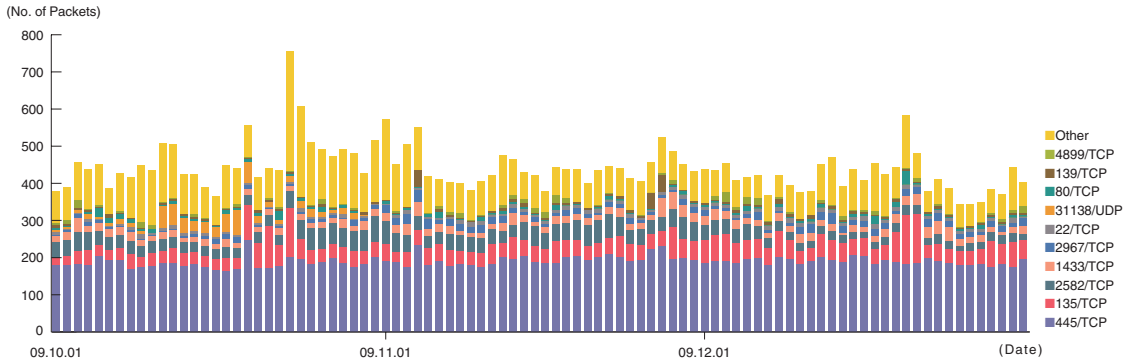


Figure 3: Communications Arriving at Honeypots (by Date, by Target Port, per Honeypot)

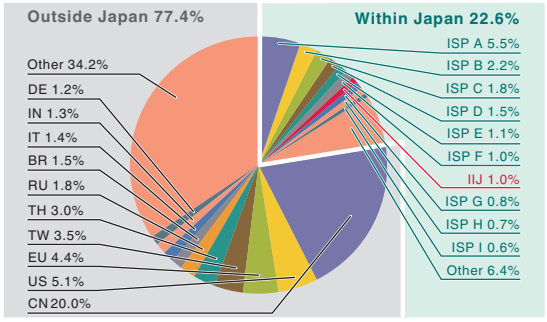


Figure 4: Sender Distribution (by Country, Entire Period under Study)

*21 Malware Investigation Task Force (MITF). The MITF began activities in May 2007, observing malware network activity through the use of honeypots to gauge trends and gather technical information, and attempting to link these findings to the creation of countermeasures.

*22 A system designed to simulate damages from attacks by emulating vulnerabilities, recording the behavior of attackers, and the activities of malware.

■ Malware Network Activity

Next, we will take a look into the malware activity observed by the MITF. Figure 5 shows trends in the total number of malware specimens acquired during the period under study. Figure 6 shows the distribution of the specimen acquisition source for malware. In Figure 5, the trends in the number of acquired specimens show the total number of specimens acquired per day*23, while the number of unique specimens is the number of specimen variants categorized according to their digest of a hash function*24.

On average, 623 specimens were acquired per day during the period under study, representing about 44 different malware variants. According to the statistics in our prior whitepaper, the average daily total for acquired specimens was 592, with 46 different variants. This indicates that both the number of specimens and the number of unique variants for this period were about the same as for the previous period.

The distribution of specimens according to source country has Japan at 60.2%, with other countries accounting for the 39.8% balance. Of the total, malware infection activity among IIJ users was 3.0%, maintaining a low value similar to the previous period.

The MITF prepares analytical environments for malware, conducting its own independent analyses of acquired specimens. The results of these analyses show that during the period under observation, 4.3% of the malware specimens were worms, 93.1% were bots, and 2.6% were downloaders. In addition, the MITF confirmed the presence of 42 botnet C&C servers*25 and 519 malware distribution sites.

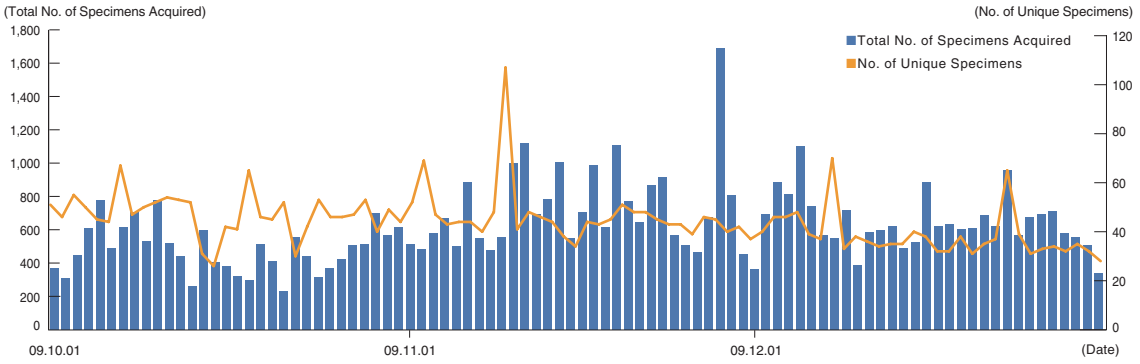


Figure 5: Trends in the Number of Malware Specimens Acquired (Total Number, Number of Unique Specimens)

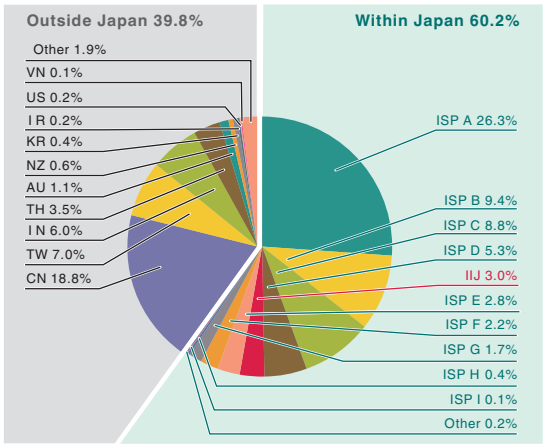


Figure 6: Distribution of Acquired Specimens by Source (by Country, Entire Period under Study)

*23 This indicates the malware acquired by honeypots.

*24 This figure is derived by utilizing a one-way function (hash function) that outputs a fixed-length value for various input. The hash function is designed to produce as many different outputs as possible for different inputs. While we cannot guarantee the uniqueness of specimens by hash value, given that obfuscation and padding may result in specimens of the same malware having different hash values, the MITF has expended its best efforts to take this fact into consideration when using this methodology as a measurement index.

*25 Abbreviation of "Command & Control." A server that provides commands to a botnet consisting of a large number of bots.

1.3.3 SQL Injection Attacks

Of the types of different Web server attacks, IIJ conducts ongoing surveys related to SQL injection attacks*26. SQL injection attacks have flared up in frequency numerous times in the past, remaining one of the major topics in the Internet security. SQL injections are known to occur in one of three attack patterns: those that attempt to steal data, those that attempt to overload database servers, and those that attempt to rewrite Web content.

Figure 7 shows trends of the numbers of SQL injection attacks against Web servers detected between October 1 and December 31, 2009. Figure 8 shows the distribution of attacks according to source. These are a summary of attacks detected by signatures on the IIJ Managed IPS Service. Japan was the source for 61.7% of attacks observed, while China and the United States accounted for 6.7% and 5.3%, respectively, with other countries following in order.

We noted the number of SQL injection attacks on Web servers similar to our prior whitepaper. Sporadic rises in attacks are those detected at multiple targets from a specific attack source.

As previously shown, attacks of various types were properly detected and dealt with in the course of service. However, attack attempts continue, requiring ongoing attention.

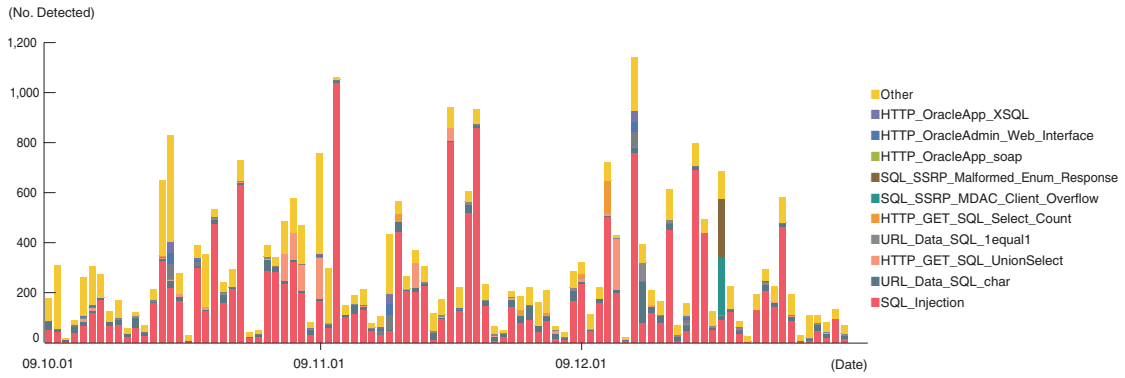


Figure 7: Trends in SQL Injection Attacks (by Day, by Attack Type)

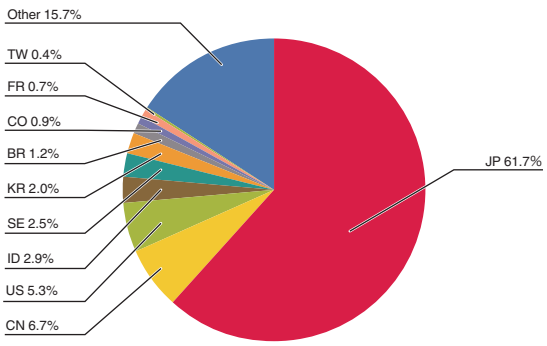


Figure 8: Distribution of SQL Injection Attacks by Source (by Country, Entire Period under Study)

*26 Attacks accessing a Web server to send SQL commands, thereby manipulating an underlying database. Attackers access or alter the database content without proper authorization, and steal sensitive information or rewrite Web content.

1.4 Focused Research

Incidents occurring over the Internet change in type and scope almost from one minute to the next. Accordingly, IIJ works toward taking countermeasures by performing independent surveys and analyses. Here we will present information from the surveys we have undertaken during this period regarding the renewed Gumblar activity, MITM attacks using a vulnerability in the renegotiation feature of SSL and TLS, and techniques for observing P2P file sharing networks.

1.4.1 Renewed Gumblar Activity

The Gumblar malware that was active from April 2009 began renewed activity in October 2009 and continued spreading through December, causing further damage. Here we will explain the situation by focusing on differences between recent incidents and the previous Gumblar*27.

■ The New Gumblar

Gumblar is a malware infection incident originating through Web content that was altered using FTP accounts that were stolen in advance. IDs and passwords on infected machines are stolen by the malware and used to perpetrate further alterations, broadening the scope of damages. This is a complex incident involving multiple websites and pieces of malware*28. The current propagation has been noticed from the alteration of a number of websites that took place around October 12*29, and infections continue to occur as of the time of writing.

As with the previous incident, the new Gumblar also induces infection via multiple websites. Previously a small number of servers were used as dedicated malware distribution sites, but for the current incident a large number of altered websites are being exploited. For this reason, it is difficult to stop the spread of the current incident by prohibiting access to or taking down malware distribution sites (Figure 9).

The total numbers of altered websites and stolen IDs and passwords are not known, but several pieces of information that have been published indicate the scale of this activity. For example, there are reports that approximately 80,000 websites have been altered to induce malware infections (over 3,000 in Japan), and that over 2,000 malware distribution sites have existed (approximately 80 in Japan)*30.

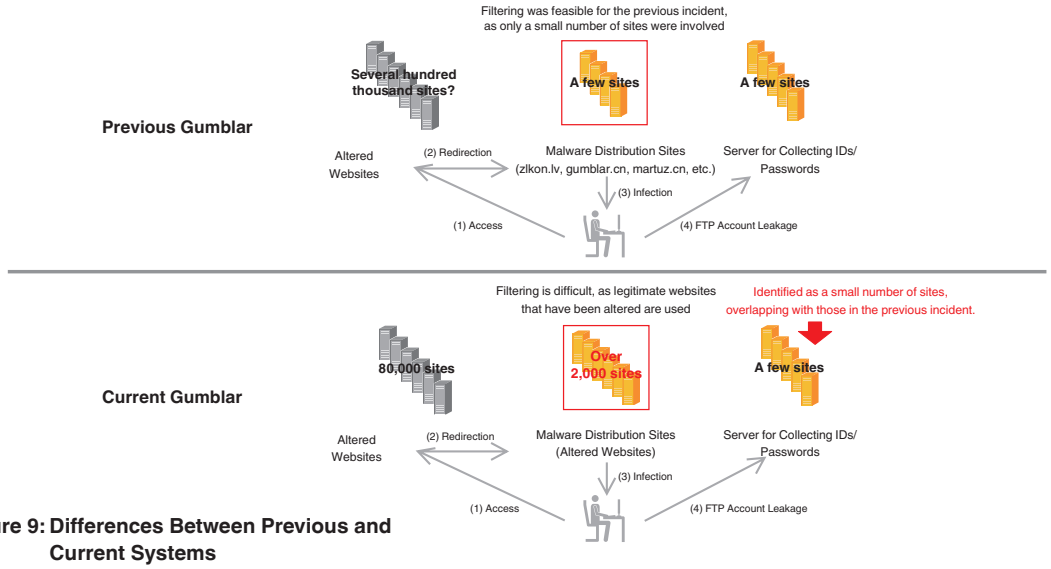


Figure 9: Differences Between Previous and Current Systems

*27 Gumblar was a portion of the name of a malware distribution website (gumblar.cn) that was used between April and May, 2009. In this whitepaper we use the name Gumblar to refer to all related websites and malware. The name Gumblar.X is also commonly used to differentiate between the current incident and the previous one.

*28 See IIR Vol.4, "ID/Password Stealing Gumblar Malware" for more information about the incident that occurred from April, such as the role of websites in Gumblar, and an analysis of the behavior of the malware (http://www.iij.ad.jp/en/development/iir/pdf/iir_vol04_EN.pdf).

*29 The cNotes reported that attacks were first observed on October 12. "The second coming of zlkon, gumblar, and martuz" (http://jvnrss.ise.chuo-u.ac.jp/csn/index.cgi?p=zlkon%A1%A2gumblar%A1%A2martuz+%BA%C6%CE%D7) (in Japanese).

*30 Site numbers in this article were sourced from the "Gumblar infection count" entry of the Analyst's Diary on the following Kaspersky Labs blog (http://www.viruslist.com/en/weblog?weblogid=208187923).

■ Currently Used Malware and Countermeasures

IIJ has analyzed multiple specimens of malware used in the current incident. As a result, we have confirmed that varieties with several additional functions compared to the previous malware are being used*31. As with the previous incident, the transmission of stolen IDs and passwords to servers is still being carried out. A distinctive header that does not follow the RFCs and is not found in normal HTTP requests is used for this communication. This means that by monitoring communications using proxy servers or Intrusion Detection Systems it is possible to identify infected users and prevent the leakage of IDs and passwords (Figure 10). Because it has been established that like the previous incident stolen IDs and passwords are only uploaded to a small number of servers, we attempted to take down*32 these servers. However, we confirmed that the malware quickly switched to other servers and resumed its activity.

■ More Recent Incidents

In parallel with this, incidents using completely different altered content and infection methods, as well as new malware and communication methods, began occurring from the beginning of December*33. These utilized more advanced malware infection methods, and exploited a vulnerability in the Java Runtime Environment*34 and a new vulnerability in Adobe Reader (including Acrobat)*35. It has been confirmed that the malware that infected users steals IDs and passwords from FTP client settings and demonstrates bot-like behavior. This incident involves the alteration of a large number of websites between late 2009 and early 2010.

As demonstrated above Gumblar is still a current incident, so care must continue to be taken with regard to managing client OS and software versions, managing passwords, and being on the lookout for Web content alterations.

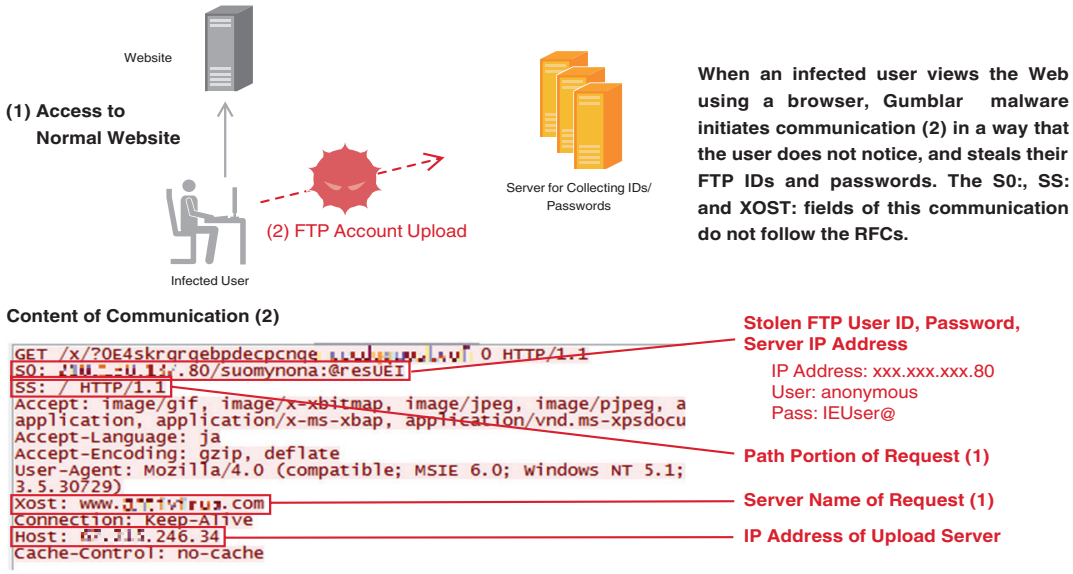


Figure 10: Communication of Gumblar Stealing FTP Account

*31 For example registry concealment, blocking access to specific websites, and preventing the startup of rootkit detection tools. Rootkits are tools that were originally used to gain root access to UNIX systems. They were often used in tandem with tools that concealed this behavior, and rootkit became a general term for tools that usurp privileges and conceal what is taking place. Because Gumblar behaves like a rootkit, using API hooking and the concealment of certain registry entries to steal IDs and passwords, it attempts to prevent itself from being detected by rootkit detection tools.

*32 IIJ issued take-down requests to JPCERT/CC for the servers it confirmed. It is possible to submit take-down requests for servers like this that are used for malicious activities by issuing an incident report notification (http://www.jpcert.or.jp/english/ir/form.html).

*33 Due to differences in the alterations that induce malware infection and the malware that is used, this incident is sometimes not referred to as Gumblar. It is known variously as GNU GPL (CODE1, LGPL), ru:8080, and 8080 due to the nature of the alterations.

*34 The fact that a vulnerability in the Java Runtime Environment (JRE) is being used has been reported in the IBM ISS Tokyo SOC Report (http://www-935.ibm.com/services/jp/index.wss/consultantpov/secpriv/b1333966?cntxt=a1010214) (in Japanese).

*35 The fact that a vulnerability in Adobe Acrobat and Adobe Reader is being used has also been reported in the IBM ISS Tokyo SOC Report (http://www-935.ibm.com/services/jp/index.wss/consultantpov/secpriv/b1333971?cntxt=a1010214) (in Japanese). A patch had not been released for this vulnerability at the time it was exploited, making this a 0-day attack. At the time of writing this issue has been addressed in "Security updates available for Adobe Reader and Acrobat" (http://www.adobe.com/support/security/bulletins/apsb10-02.html).

1.4.2 MITM Attacks Using a Vulnerability in the SSL and TLS Renegotiation

■ Background

In November 2009, Marsh Ray, Steve Dispensa and Martin Rex released details of a vulnerability*36 in the SSL and TLS protocols*37*38 that could allow Man-in-the-Middle attacks*39 to be carried out. SSL and TLS operate between the IP and application layers and ensure confidentiality and integrity for application data, authenticating the target of communications using X.509 public key certificates. As they are used together with application layer communication protocols such as HTTP, SMTP, and POP, this vulnerability affects a large number of applications and systems.

In particular, the HTTPS (HTTP over SSL) protocol*40 is implemented in a large number of Web browsers and Web servers, and Marsh Ray et al. gave an example of an attack method using HTTPS in the report. It has also been established that the vulnerability is exploitable through the release of methods for posting password information to an attacker's Twitter account by applying the vulnerability to the Twitter API*41. Thierry Zoller investigated whether or not the vulnerability could be applied to protocols other than HTTP*42. In his report, he showed that FTPS and SMTPS are vulnerable, and EAPTLS is not affected, but there are still application protocols for which the impact is not yet clear, such as POP and LDAP.

This vulnerability can be attributed to a problem in the SSL and TLS protocol specifications themselves. Fixes have been released for OpenSSL*43 and Apache*44, but most of these involve simply disabling the renegotiation feature that is causing the problem*45. More thorough measures would require an update to the current specifications and migration to implementations that follow the new specifications. The IETF*46 has been in the same line and an RFC that establishes countermeasures was published with unprecedented speed (RFC5746*47). This vulnerability affects all versions of the TLS protocol, in addition to SSL version 3.0. SSL specifications are not under IETF change control, but the RFC indicates that the TLS solution can also be applied to SSL (RFC5746 section 4.5).

■ MITM Attacks Exploiting the Renegotiation

In SSL and TLS, a client and server negotiate encryption algorithm and key information by handshake protocol before the sending and receiving of application data is carried out safely. The renegotiation feature is used to update algorithm and key information that has been accepted by both client and server. The vulnerability report indicated that when an issue in the renegotiation specifications is exploited, it is possible to intercept SSL and TLS communications using a Man-in-the-Middle attack. A specific example that was pointed out was a situation where mutual authentication (with client authentication using a public key certificate) is switched from server authentication during a session.

---

*36 Marsh Ray, Steve Dispensa, "Renegotiating TLS" (http://extendedsubset.com/Renegotiating_TLS.pdf).
This vulnerability is managed as CVE-2009-3555 (http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2009-3555) and Vulnerability Note VU#120541, "SSL and TLS Protocols renegotiation vulnerability" (http://www.kb.cert.org/vuls/id/120541).

*37 Alan O. Freier, Philip Karlton, Paul C. Kocher, Internet Draft "The SSL Protocol Version 3.0" (http://tools.ietf.org/html/draft-ietf-tls-ssl-version3-00).

*38 Dierks, T. and E. Rescorla, RFC 5246 "The Transport Layer Security (TLS) Protocol Version 1.2" (http://www.ietf.org/rfc/rfc5246.txt).

*39 Man-in-the-Middle attacks are those targeted at communications where the attacker is positioned between the two parties carrying out communications. This can result in communications being intercepted or altered (including cases where the attacker poses to each party as the other). When evaluating whether or not this attack method can succeed and investigating countermeasures, it is assumed that there is an attacker in the middle during communications. In order for a Man-in-the-Middle attack to succeed in practice, it must be used in combination with other methods for intercepting communications (for example, route hijacking over the Internet).

*40 E. Rescorla, "HTTP Over TLS" (http://www.ietf.org/rfc/rfc2818.txt).

*41 Anil Kurmus's blog identifies the Twitter API issues "TLS renegotiation vulnerability (CVE-2009-3555)" (http://www.securegoose.org/2009/11/tls-renegotiation-vulnerability-cve.html). The issues identified here were quickly fixed by Twitter.

*42 Thierry Zoller, "TLS & SSLv3 renegotiation vulnerability" (http://www.g-sec.lu/practicaltls.pdf).

*43 OpenSSL Security Advisory (http://www.openssl.org/news/secadv_20091111.txt). During proofreading, the OpenSSL project team announced the release of version 0.9.8m which implements RFC5746 (http://cvs.openssl.org/getfile?f=openssl/CHANGES&v=OpenSSL_0_9_8m).

*44 http://www.apache.org/dist/httpd/patches/apply_to_2.2.14/CVE-2009-3555-2.2.patch
During proofreading, the Apache HTTP server project team announced the release of version 2.2.15 corresponding to OpenSSL 0.9.8m which implements RFC5746 (http://www.apache.org/dist/httpd/CHANGES_2.2.15).

*45 It is possible to confirm whether or not renegotiation is enabled using the following TLS Renegotiation Test. "TLS Renegotiation Test" (http://netsekure.org/2009/11/tls-renegotiation-test/).

*46 Internet Engineering Task Force. The organization that develops Internet technical standards such as communication protocols and data formats. They issue the Request for Comments (RFC) documents that regulate standard specifications. Draft RFC documents are known as Internet drafts.

*47 E. Rescorla, M. Ray, S. Dispensa, N. Oskov, RFC5746, "Transport Layer Security (TLS) Renegotiation Indication Extension" (http://www.ietf.org/rfc/rfc5746.txt). The Internet draft this RFC is based on was proposed in November 2009, discussed quickly for the relatively short period of three months, and made an RFC in February 2010.

Figure 11 shows an example of an attack using HTTPS. This attack allows attackers as the man-in-the-middle to interrupt encrypted communications between a client and server. As a result, attackers are able to combine their HTTP requests with those of a legitimate user, and send them to the server. Note that at this point the legitimate user's application data is still encrypted, and no alteration or eavesdropping by an attacker will take place. When attackers' requests are linked with a legitimate user's existing requests using an HTTP cookie, servers will interpret the linked requests as both having come from the same user, so there is a chance that an attacker's requests will be accepted and processed.

■ RFC Modifications

Next, we will explain the modifications in the new RFC. This RFC introduces a new "renegotiation_info" TLS extension value, in addition to a "TLS_EMPTY_RENEGOTIATION_INFO_SCSV" state for the cipher suite that normally defines the encryption algorithm. Using the renegotiation_info extension, it is possible to announce to a target of communications that the current implementation can safely carry out renegotiation. More specifically, the client and server both save information shared safely by handshake protocol. When carrying out renegotiation information that cannot be known to outside parties is exchanged using the renegotiation_info extension to confirm that each party is the same as when communication began. Additionally, because some implementations detect the renegotiation_info extension as a TLS extension that cannot be processed and terminate communication, a method utilizing "TLS_EMPTY_RENEGOTIATION_INFO_SCSV" has also been made available.

■ Countermeasures and Backward Compatibility Issues

The current issue will be resolved when implementations compliant with the RFC5746 countermeasures spread, but migration is expected to take some time. From a backward compatibility perspective it is important to maintain compatibility with the current version, but when renegotiation is carried out using a previous implementation, there is no safe way of confirming that a request is coming from a legitimate party. For this reason, it is recommended that renegotiation requests from previous implementations are rejected in new implementations. This is in effect the same as the current temporary countermeasure that does not allow use of the renegotiation. In other words, when renegotiation must be carried out safely, a new implementation that supports the new specifications must be deployed for both client and server.

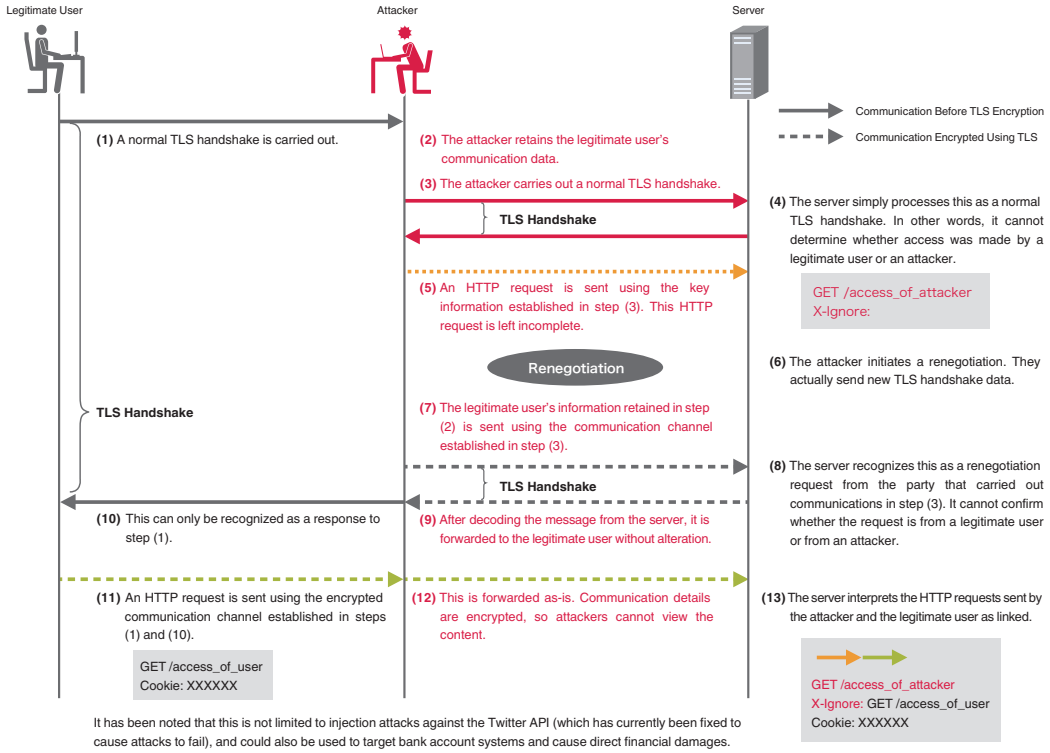


Figure 11: Attack Scenario

1.4.3 Techniques for Surveying P2P File Sharing Networks

■ Introduction

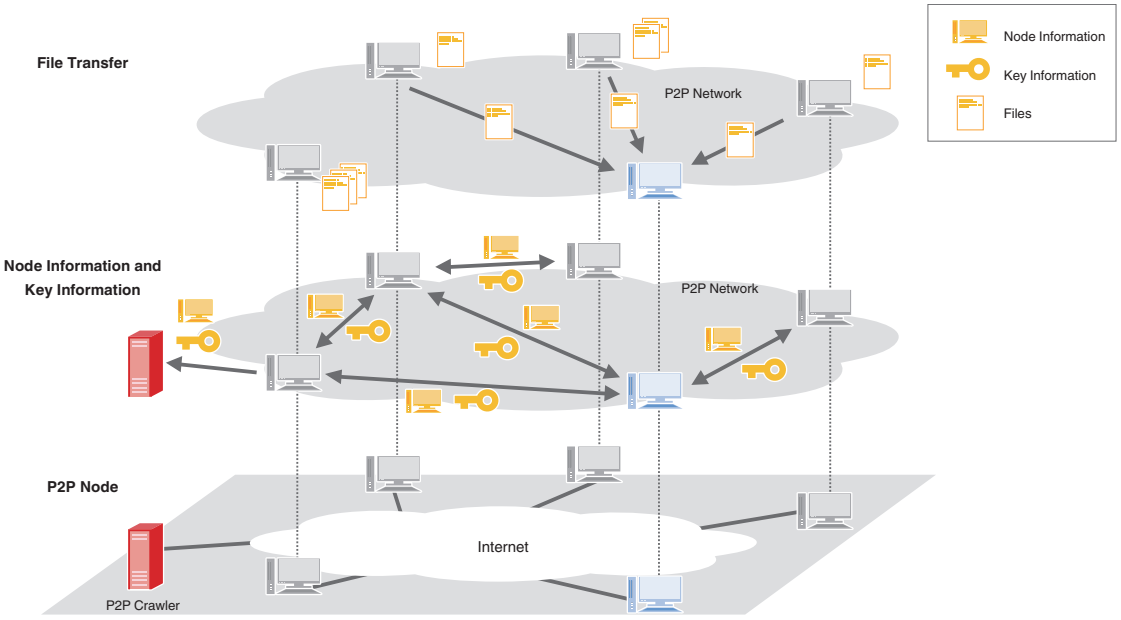
IIJ observes techniques for surveying P2P networks formed using P2P file sharing software such as Winny and Share (two of the implementations popular in Japan) from the dual perspectives of considering countermeasures to prevent information leakage and investigating the characteristics of communication volume, and has been actively participating in related research projects*48 since 2006.

P2P file sharing networks are once again drawing attention for the connection they have with copyright infringement due to the inclusion of a provision making the download of illegally copied material illegal in the amended Copyright Act*49 that came into effect on January 1, 2010. Here we take this opportunity to summarize P2P file sharing network systems such as Winny and Share, as well as techniques for surveying such networks.

■ P2P File Sharing Network Systems

P2P file sharing networks have functions for making files publicly available in order to share them, functions for searching for files, and functions for downloading files. P2P nodes exchange "key information" that indicates which nodes have which files in order to search for the desired files more efficiently. This key information exchange system is also used for notifying other nodes of the files that are made available.

One of the methods used to increase the overall download efficiency of a P2P file sharing network is the caching of files that have been downloaded, which are automatically made available to other nodes. They also feature systems where even if a user has not downloaded a file themselves, a cache is automatically created through a file transfer relay. Through this system popular files are automatically made available on large numbers of nodes. Under a pure P2P system, a P2P file sharing network is sustained through the process of nodes exchanging information on other nodes and accumulating it to allow unknown numbers of nodes to participate in and withdraw from the network freely. Figure 12 shows a summary of this explanation. Actual P2P file sharing software collects files automatically based on keywords that users specify. This means that key information collection and file transfer communications can occur constantly on P2P file sharing networks, increasing traffic volumes.



P2P nodes connected to the Internet (lower layer) form a P2P network by connecting with other nodes, exchanging and accumulating node information and key information (middle layer). Nodes that wish to download a file (the blue node) search for it using the accumulated key information, identify nodes that have that file (or a portion of it), and then connect directly to those nodes to request the transfer of the file (upper layer). On the other hand, P2P crawlers only connect to nodes and collect node information and key information (lower to middle layers), and do not transfer files.

Figure 12: P2P File Sharing Network Systems

*48 For example, the Secure Trusted Network Forum's P2P research group (http://www.scat.or.jp/stnf/) (in Japanese).

*49 The provisions of the Copyright Act can be viewed using the Japanese Law Translation (http://www.japaneselawtranslation.go.jp/?re=02).

■ Techniques for Surveying P2P File Sharing Networks

On P2P networks, there is no server which maintains centralized information on the whole network. For this reason, crawling methods are used to get an overall picture of a network*50. When carrying out a crawling survey, crawlers connect with P2P nodes and communicate using the protocol of the P2P file sharing network to acquire information about other nodes. Then, crawlers connect with the nodes that were newly discovered, and repeat the process of acquiring information about other nodes, comprehensively surveying the nodes on a P2P file sharing network (Figure 13).

By analyzing the key information that is collected along with the node information during a crawling survey, it is also possible to ascertain what kinds of files are being made public at which nodes. Surveys using this kind of crawling method have the advantage of giving an overall picture of a P2P file sharing network without side effects such as the dissemination of files.

■ The Current State of P2P Networks

We will present a portion of the results from surveys that IIJ is carrying out in cooperation with an external organization as an example of surveys of the current state of P2P file sharing networks*51. Through these surveys we have learned that approximately 2% of all Winny nodes and 3% of all Share nodes exist on the IIJ network. We are also evaluating the impact that this has on the entire network by ascertaining the amount of traffic that is generated by these nodes through communications with nodes outside the IIJ network. Figure 14 shows changes in the number of Winny and Share nodes*52 that were identified through these surveys, and Figure 15 shows the results of surveying the traffic generated by nodes on the IIJ network through communication with nodes outside the IIJ network.

The results show that the number of nodes for both Winny and Share is in a downward trend, but they still occupy approximately 6Gbps of bandwidth on a constant basis at this point in time.

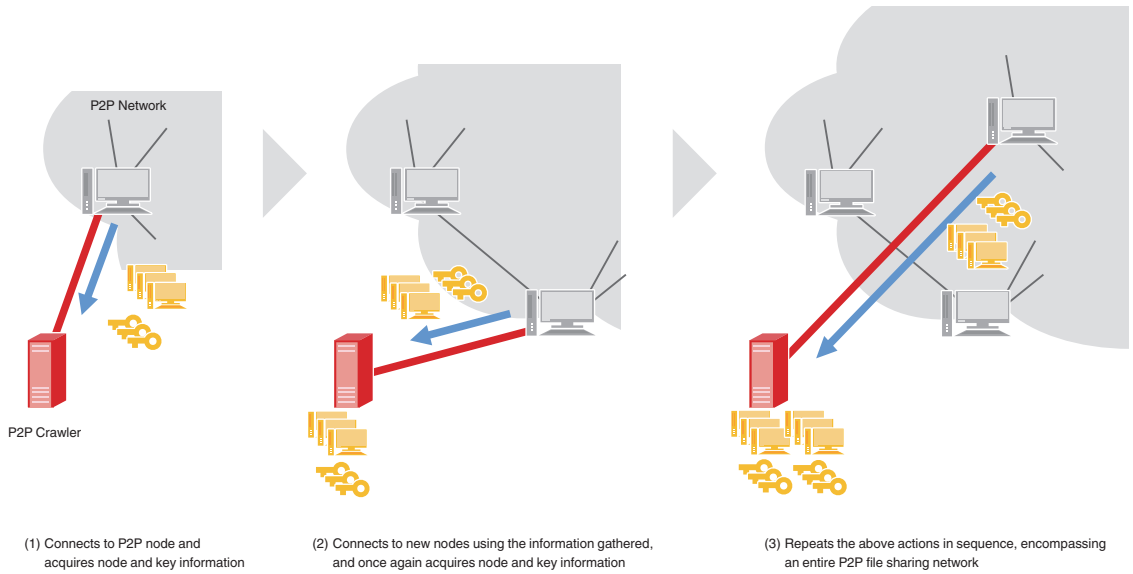


Figure 13: Crawling Method Overview

*50 Crawling survey reports are also given in the following paper. Terada et al., "P2P network observation using crawling method", Information Processing Society of Japan Computer Security Group Report Vol. 2007, No. 48, pp. 51-56 (2007) (http://jvnrss.ise.chuo-u.ac.jp/jtg/doc/CSEC07037009.pdf) (in Japanese). Winny Radar and Share Radar of Fourteenforty Research Institute, Inc. (http://www.fourteenforty.jp/) (in Japanese) are examples of products that conduct surveys like this.

*51 Surveys of the current state of P2P file sharing networks include "The Current State of P2P - Winny and Share Network Status Survey Report -" conducted by CROSSWARP Inc. (http://www.scat.or.jp/stnf/contents/p2p/p2p080910_2.pdf) (in Japanese). Several other surveys of the current state of affairs were presented at the Information Security Seminar hosted by the Secure Trusted Network Forum in September 2008, and the materials presented can be downloaded from their website (http://www.scat.or.jp/stnf/contents/p2p080910.html) (in Japanese). The next seminar was held on March 2, 2010 (http://www.scat.or.jp/stnf/contents/p2p100302/P2P.htm) (in Japanese).

*52 About the time that the amended Copyright Act came into effect on January 1, 2010, a drop in node numbers for both Winny and Share of about 20% was observed. Note that the node numbers shown in this figure are values for after the drop occurred.

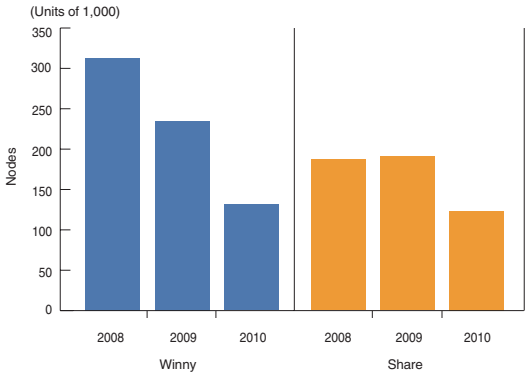
Here we presented the current state of Winny and Share, which we are surveying on a regular basis. Other P2P file sharing network implementations exist, and there is a possibility that communication characteristics may continue to evolve drastically due to changing user habits. IIJ will continue to carry out surveys like this in order to keep providing a stable network infrastructure.

1.5 Conclusion

This whitepaper has provided a summary of security incidents to which IIJ has responded.

In this volume we provided a follow-up on the continuing Gumblar incidents, and summarized vulnerabilities in the SSL and TLS communications protocols, as well as techniques for surveying P2P file sharing networks. P2P file sharing networks are not a particularly new subject, being a topic that has sparked a variety of debates, such as the volume of traffic they generate, their anonymity, information leakages, and copyright infringement, and IIJ has been surveying them over an extended period. In this volume we have only covered a few topics, such as survey methods and impact on traffic volumes, but we would like to continue surveying these networks and examine different facets of them when the opportunity arises.

By identifying and publicizing incidents and associated responses in whitepapers such as this, IIJ will continue to inform the public about the dangers of Internet usage, providing the necessary countermeasures to allow the safe and secure use of the Internet.



The survey was conducted over one week in January of each year. Nodes were counted every 24 hours after removing duplicates to gain a daily number of nodes, and the average of this figure over 7 days was used as the number of nodes for that year.

Figure 14: Comparison of Average Daily Nodes for Winny and Share

Average traffic was calculated for one week (the same week as Figure 14) in January of each year. Outgoing traffic refers to traffic from inside the IIJ network going to external networks, and incoming traffic refers to traffic from outside the IIJ network coming in.

Figure 15: Winny and Share Traffic (Temporal Average for 1 Week)

Authors:
Mamoru Saito
General Manager of the Division of Emergency Response and Clearinghouse for Security Information, IIJ Service Business Department. After working in security services development for enterprise customers, Mr. Saito became the representative of the IIJ Group emergency response team, IIJ-SECT in 2001, participating in FIRST, an international group of CSIRTs. Mr. Saito serves as a steering committee member of several industry groups, including Telecom-ISAC Japan, Nippon CSIRT Association, Information Security Operation provider Group Japan, and others. In recognition of its close activities with both domestic and international organizations, the IIJ-SECT was awarded the "commendation from Director-General, Commerce and Information Policy Bureau, Ministry of Economy, Trade and Industry (CATEGORY - Promotion of Information Security)" at the FY 2009 Informatization Month Opening Ceremony.

Hirohide Tsuchiya (1.2 Incident Summary)
Hirohide Tsuchiya, Hiroshi Suzuki (1.3 Incident Survey)
Hiroshi Suzuki (1.4.1 Renewed Gumblar Activity)
Yuji Suga (1.4.2 MITM Attacks Using a Vulnerability in the SSL and TLS Renegotiation)
Mamoru Saito, Tadaaki Nagao (1.4.3 Techniques for Surveying P2P File Sharing Networks)
Division of Emergency Response and Clearinghouse for Security Information, IIJ Service Business Department

Contributor:
Masahiko Kato, Division of Emergency Response and Clearinghouse for Security Information, IIJ Service Business Department

The Need for Anti-Spam Measures Tailored to the Regional Characteristics of the Source

In this report, we will offer our analysis of trends in the ratio of spam for the whole of 2009 including weeks 40 to 52, in addition to examining regional sources of spam for the same period. At the same time we will also investigate spam sending trends for the major regional sources of spam, explain the need for countermeasures tailored to regional characteristics, and look at technology related to DKIM sender authentication.

2.1 Introduction

In this report, we have summarized the latest developments in spam trends, information about anti-spam technologies, and other activities in which IIJ is deeply engaged. To analyze spam trends we conducted a variety of analyses based on information obtained through the Spam Filter feature of the IIJ email services. The volume of email varies depending on the day of the week according to the service under consideration. Accordingly, we have consolidated data on a weekly basis to better understand the trends revealed in our analysis. This survey covers the entire 2009 period, adding 13 weeks worth of data from the 40th week of 2009 (9/28/2009 to 10/4/2009) to the 52nd week (12/21/2009 to 12/27/2009).

Regarding spam trends, we comment on regional differences in the spam sending trends. Spam originating from Japan has decreased dramatically due to OP25B*1, but the difference between regions where countermeasures such as this are effective and specific regions that should be dealt with separately has become clear. We also report on the implementation status of sender authentication technology, which is a core technology for anti-spam measures.

Under trends in email technologies we cover DKIM sender authentication technology using digital signatures, with an explanation of the DKIM-ADSP extension that defines signing practices. Additionally, we provide an overview of the changes that have been made to the DKIM specification.

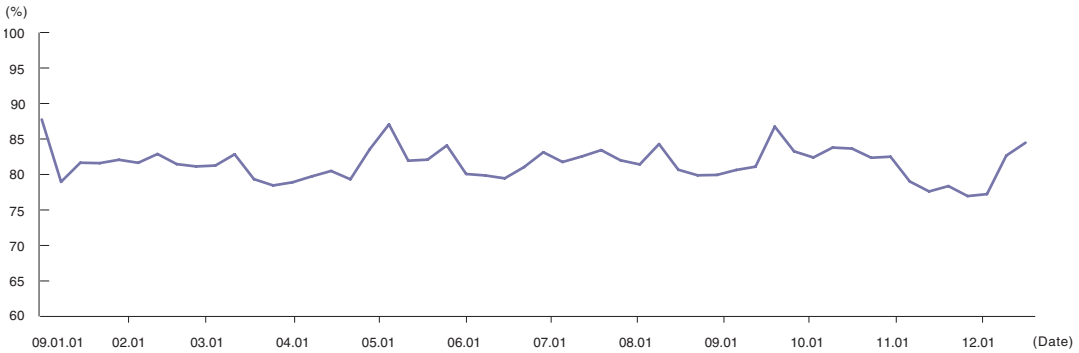


Figure 1: Spam Ratio Trends

*1 OP25B (Outbound Port 25 Blocking) is technology that suppresses the sending of spam by blocking the direct sending of mail from dynamic IP addresses assigned to consumers to external incoming mail servers.

2.2 Spam Trends

In this section, we will report on historical ratios of spam and the results of our analysis concerning spam sources based on trends detected through IIJ spam filters.

2.2.1 Spam Ratio Trends

The ratio of spam averaged 81.4% of all incoming emails over the 91-day period from week 40 to week 52, 2009. This compares to an 82.2% average in our last survey (weeks 27 through 39, 2009), indicating a slight decrease of 0.8%. The average for the same period the previous year was 81.5%, so it seems the trend is remaining constant. Figure 1 shows spam ratio trends for 2009 including the results for the current period.

Spam ratios are relative to the volume of regular emails. This means that when the volume of regular email varies due to an extended holiday or other events, it also affects the spam ratio. Seasonal differences are also observed in spam volume. For this reason, to determine upward or downward trends in spam, long-term observation is required. In light of this, we can state that spam volume has remained at a high ratio since the previous year.

Characteristics of the current period include a decrease in the spam ratio between November and early December. The volume of spam itself decreased over this period. This was not a decrease caused by the relative relationship with regular email. However, as the volume of spam shifted higher from the second half of December, the decrease is believed to have been only temporary.

2.2.2 Sources of Spam

Figure 2 shows our analysis of regional sources of spam over the period studied. Brazil (BR) remained the number one source of spam in this survey, accounting for 12.5% of total spam. Brazil has held its position as the top source of spam since it was reported in IIR Vol.3 to have taken first place in the first quarter of 2009. The 2nd to 6th top sources of spam were in descending order China (CN) at 10.4%, the United States (US) at 7.0%, India (IN) at 5.6%, Vietnam (VN) at 5.2%, and Korea (KR) at 4.3%. This order has changed since the last report, but the regions taking 1st to 6th place remain the same.

Figure 3 shows the changes in spam ratios for these six countries and Japan as reported between IIR Vol.1 and Vol.6. This graph shows that the ratio of spam from the United States (US) is in a downward trend, while Brazil (BR), India (IN), and Vietnam (VN) are trending higher. It difficult to gauge the trends for China (CN) and Korea (KR) as their ratios vary depending on the period, but they cannot be said to be decreasing, so vigilance is required.

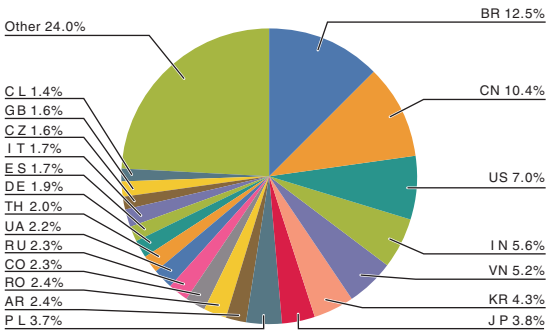


Figure 2: Regional Sources of Spam

As these figures indicate, Brazil is one of the main sources of spam sent to Japan, and the Japan Data Communication Association and JPCERT/CC have announced they will begin sharing information with Brazil regarding spam*2. Data from our IIR has been cited in the materials presented. As reported in our IIR to date, because the vast majority of spam is sent from outside Japan, coordination with regional authorities like this will be necessary in order to reduce the volume of spam.

IIJ is assisting the activities of JEAG*3, and sharing our perspective regarding the introduction of anti-spam measures such as OP25B with related organizations in countries such as Korea and China. Currently, due to the unique regional circumstances in each country, no immediate progress has been made toward effective countermeasures, but we will continue to cooperate with both domestic and international organizations to work on the creation of global anti-spam measures.

2.2.3 Spam Sending Trends

As shown in Figure 2, Japan (JP) was the source of 3.8% of spam for the current period, coming in 7th place. This ratio is a slight increase of 0.7% over the previous period. As can be seen in Figure 3, the ratio of spam sent from Japan has been increasing at a slow but steady rate since the period reported in IIR Vol.1 (June 1 to August 31, 2008).

As our analysis to date has shown, the trend for email identified as spam originating from Japan indicates that cases of mass mailing using a fixed IP address continue to be prevalent. These cases include sources thought to be data centers and hosting companies. Dynamic IP addresses that cannot be dealt with using OP25B also continue to be found among sources of spam. However, the ratio is far lower than in other regions.

For this report, we compared the ratio of sources determined to be sending spam during a specific period that sent an average of 1 or fewer spam messages per day for the countries that are the major sources of spam. In other words, this indicates the ratio of sources that sent only an extremely small number of all messages determined to be spam. Figure 4 shows the results of this comparison.

In Figure 4, while the ratios for China (CN), the United States (US), Korea (KR), and Japan (JP) are all about 5%, the ratios for Brazil (BR), India (IN), and Vietnam (VN) are high. The regions with higher ratios are all regions for which the spam source ratio in Figure 3 is increasing. As bots infected with malicious software (malware) are thought to be an increasingly common method of sending spam in recent years, we believe that bot numbers are on the rise in these regions with higher ratios.

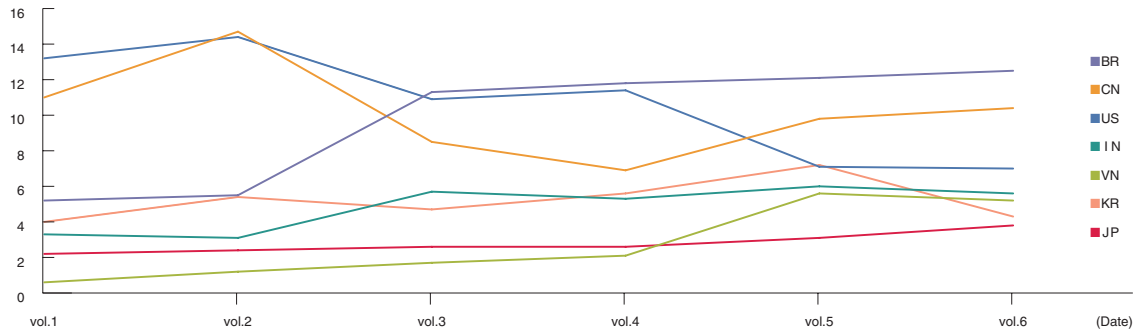


Figure 3: Trends in Sources of Spam

*2 Report materials: Regarding the start of spam information sharing with Brazil (http://www.dekyo.or.jp/soudan/image/n-image/PL_20100108.pdf).

*3 JEAG (Japan Email Anti-Abuse Group) is a working group founded by Japan's major Internet service providers (ISPs) and mobile telecommunication carriers to counter spam email abuse (http://www.iij.ad.jp/en/news/pressrelease/2005/0315.html).

PCs that are susceptible to bot infections are those used by individuals for which sufficient security measures have not been implemented, and most use a dynamic IP address that changes each connection. The results of this survey show that only a small amount of the spam sent from these regions was from the same source (IP address). This is thought to be due to the use of dynamic IP addresses. In regions like these, the introduction of network-level technology such as OP25B that prevents the direct sending of spam is effective.

On the other hand, in regions that have a high spam source ratio despite the small ratio of sources sending a low volume of spam each day, we believe that certain specific sources are sending large volumes of spam. The low ratio for Japan in Figure 4 can be explained by the fact that the volume of spam sent from dynamic IP addresses is not very large due to the introduction of OP25B. The low ratio for China and Korea is more surprising. In regions such as these that are in close proximity to Japan, we believe that specific sources are sending large volumes of spam to Japan. It was reported that a spammer arrested in 2007 was sending spam to Japan from PCs they had set up in China. In regions such as these, it should be possible to reduce the volume of spam by dealing with specific sources of mass spam.

This demonstrates that it is crucial to use countermeasures that match the circumstances and characteristics of each region to counteract spam swiftly.

2.2.4 Sender Authentication Technology Implementation Status

Figure 5 shows the authentication result ratios for SPF, a network-based sender authentication technology, during the current survey period (October 1 to December 31, 2009). Of the emails received during this period, 56.3% indicated “none” as the authentication result. This means that the domain for 43.7% of email received declared an SPF record.

This ratio of SPF implementation is almost level with the previous ratio (Vol.5), while the ratio of “pass” results climbed to 15.9%, which is 2.4% higher than the 13.5% result from the previous period. The slight reduction in the volume of spam may have had an effect on these results. Another result that stands out is the ratio of “neutral” authentication results dropping to 4.3%, which is 2.3% lower than the previous period. This means that the ratio of results for which “?all” was declared at the end of the SPF record decreased. In the SPF specification “?all” is defined as for testing purposes, so we believe that the number of domains switching from test operation to regular operation is increasing.

We will continue to survey and report the implementation status of sender authentication technologies in our IIR.

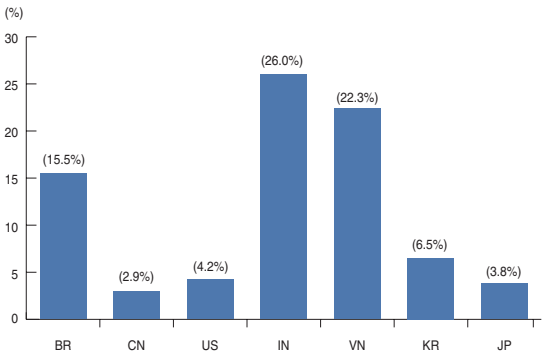


Figure 4: Ratio of Sources that Send Spam Infrequently

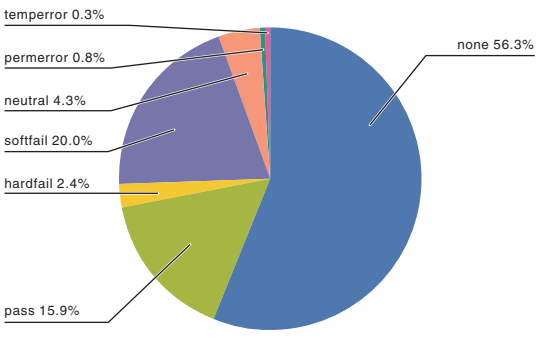


Figure 5: SPF Authentication Result Ratios

2.3 Trends in Email Technologies

2.3.1 DKIM ADSP Background

DKIM (DomainKeys Identified Mail) technology was explained in detail in IIR Vol.3. DKIM involves creating a digital signature using the header and body text of an email, and inserting this signature to the email as a DKIM-Signature header, enabling the recipient to carry out authentication. DKIM also allows authentication to be handled by the sender system without the need for a special authentication service or distribution method, as the public key that is necessary for authenticating the digital signature is published in the DNS for the sender's domain. The email recipient acquires signature information from the DKIM-Signature header of a received email, and verifies it to authenticate the sender's domain. This process is significantly different from network-based sender authentication technologies such as SPF/Sender ID.

Because all network-based technologies use existing sender information (reverse-path in SMTP and PRA information such as the From header), the location for acquiring the SPF record is predefined, and it is possible to determine whether or not a sender supports sender authentication technology by checking if an SPF record exists or not. DKIM, however, operates by simply carrying out authentication based on the DKIM-Signature header when it is present, and when this header is not inserted it is not possible to determine whether this is because the sender does not support DKIM, or because the DKIM-Signature header could not be inserted to that particular email for some reason. This is because the selector information specified in the DKIM-Signature header is necessary for acquiring the public key to be used with the digital signature, and it is not possible to determine whether or not a sender supports DKIM from their domain name alone.

Additionally, when using network-based technology it is possible for the sender of an email to specify the degree of action taken when authentication fails, depending on the type of qualifier defined before the "all" value set at the end of the SPF record. On the other hand, while the DKIM specification (RFC4871) makes it possible to verify authentication when a DKIM-Signature header is present, a sender cannot specify the action a recipient should take when authentication fails. For this reason, ADSP (Author Domain Signing Practices) were established in RFC5617 as a method for senders to declare signing practices.

In the early stages of discussing DKIM specifications, the need for a system that allows senders to declare their intentions and that differentiates between the distributor of an email according to current email usage and the actual sender of an email was pointed out. However, discussions related to determining the identity of a sender did not come to fruition, and RFC4871 was published when it was decided that the core DKIM specification should be released for the sake of early adoption. Discussions regarding sender policy continued following this, and as a result, only the basic specification was released as ADSP. For this reason, the name of the specification also changed as follows during the course of discussions.

Table 1: DKIM-ADSP Naming Changes

Published Date	Short Form	Full Name
1/10/2006	SS	Sender Signing Policy
3/3/2007	SS	Sender Signing Practices
8/26/2008	ASP	Author Signing Practices
1/3/2009	ADSP	Author Domain Signing Practices
8/2009	ADSP	RFC5617

2.3.2 DKIM ADPS Overview

The DKIM ADSP (DomainKeys Identified Mail Author Domain Signing Practices) specification is published as RFC5617. ADSP information will be published as an ADSP record in the DNS.

Specifically, the DNS TXT resource record is used. This information is acquired by querying the DNS using the domain name of the author address (author domain) indicated in the From header field of an email. This domain will be the same as the domain name indicated by the "d=" tag in the DKIM-Signature header. For example, if the author domain name was "example.jp," the ADSP record (TXT resource record) query would be sent to the following domain name.

_adsp._domainkey.example.jp

As demonstrated in this example, the domain name consists of the author domain with the "_adsp._domainkey" subdomain added. The "tag=value" format (tag format) is used to describe ADSP records, but at present only the "dkim=" tag is defined. The "dkim=" tag can be set to the following values. If any other value is set, it is treated as an "unknown" value.

Table 2: DKIM-ADSP Values

Value	Meaning
unknown	The domain might sign some or all email.
all	All mail from the domain is signed with an Author Domain Signature.
discardable	All mail from the domain is signed. Furthermore, if a message arrives without a valid Author Domain Signature, the domain encourages the recipient(s) to discard it.

2.3.3 DKIM Updates

The DKIM specification was published as RFC4871 in May 2007. It was published again as RFC5672 in August 2009, two years and three months later, with the previously ambiguous "d=" and "i=" identifiers in the DKIM-Signature header more clearly defined. However, there were no changes to the creation and verification of digital signatures that form the core of the DKIM specification, and no beneficial updates related to third party signatures, which have not been resolved to date.

2.4 Conclusion

In this volume's Messaging Technology we reported on spam and spam ratio trends, as well as information regarding the sources of spam. We also took a closer look at countries that are the main sources of spam, and the numbers of spam messages that are sent from the same source, identifying and evaluating the differences. IIJ will continue to analyze spam characteristics and trends based on emails in actual circulation, and contribute towards the development of anti-spam measures that correspond to the various needs of the global environment. With regard to trends in email technologies, we explained the DKIM sender authentication technology using digital signatures that is expected to be adopted more and more widely in the future, and gave an overview of the related ADSP specification. The SecureMX service provided by IIJ is already compatible with DKIM ADSP, and supports the latest technology for both outgoing and incoming email, with DKIP ADSP information recorded in the Authentication-Results header when email is received, in addition to sender support. IIJ will continue its efforts to stay on top of the latest trends and be the first to provide effective technologies.

Author:
Shuji Sakuraba
Shuji Sakuraba is a Senior Program Manager in the Service Promotion Section of the Messaging Service Division in IIJ Network Service Department. He is engaged in the research and development of messaging systems. He is involved in various activities in collaboration with external related organizations for securing a comfortable messaging environment. He is a MAAWG member and JEAG board member. He is a member of the Council for Promotion of Anti-Spam Measures and administrative group, as well as chief examiner for the Sender Authentication Technology Workgroup. He is also a member of Internet Association Japan's Anti-Spam Measures Committee.

Measurement Study on the Internet reachability

Internet reachability can be assessed using control-plane and data-plane measurements. However, there are biases in the results of these two measurement methods that are caused by factors such as the use of default routing. Here, we examine reasons for the biases between control-plane and data-plane measurement results, and explain a dual probing methodology that enables more accurate measurements of reachability.

3.1 Introduction

First, let us briefly review the AS (Autonomous System) and BGP (Border Gateway Protocol) terms that appear frequently in topics related to routing. An AS is a network or group of networks under a common administration and with common routing policies. As shown in Figure 1, an AS typically represents a single ISP. However, in some cases an AS may belong to more than one ISP, or conversely, a single ISP may have more than one AS. ASes are allocated a 32-bit value called an AS number, and ISPs are sometimes referred to using this AS number. For example, IIJ's AS number is 2497, so IIJ is sometimes referred to as AS 2497. A protocol called BGP (Border Gateway Protocol) is used for exchanging routing information between ASes. Each AS is allocated a block of addresses that share the same N bits from the leftmost digit of the IP address. This is called an address prefix, or simply a prefix. BGP is used to exchange information regarding the reachability of the address prefixes belonging to each AS. Additionally, the initial N bits that each AS shares are called the prefix length. When discussing the prefix length of an address it is sometimes referred to as a /N prefix.

The most fundamental service of the Internet is provision of reachability between two given points. However, we still have a great deal to learn about the basic reachability service of the Internet. Researchers and operators rely on two views of reachability to assess it: examining BGP routing information (control-plane measurements), and testing actual reachability using tools such as ping and traceroute (data-plane measurements).
Here, we show that both of these methods are insufficient for understanding reachability of the Internet as a whole, and present methods for better understanding Internet reachability through supplemental measurements. This paper is based on the results of tests carried out by IIJ senior researcher Randy Bush together with O. Maennel, M. Roughan, and S. Uhlig. For details regarding the tests covered here, see reference 1 that was presented at the ACM SIGCOMM ICM (Internet Measurement Conference) in November 2009.

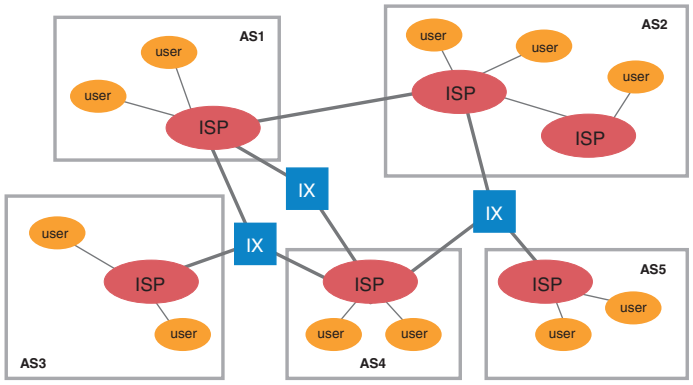


Figure 1: Overview of AS

3.2 How Far does a /25 Propagate?

Anecdotally, most providers filter prefixes more specific than /24 to bound the number of routes in the global Internet in order to reduce utilization of common resources necessary for handling announcements of overly specific prefixes, or to prevent route hijacking. As the starting point for this paper, we sought to test whether such filters are as prevalent as commonly presupposed.

We advertised a /25 prefix from AS 3130 on June 22 2008, and measured its reachability across the Internet via both control-plane and data-plane measurements. At this point, no other routing information containing this /25 existed. The results were so inconsistent that it highlighted the fact that control-plane measurements are insufficient to measure data-plane reachability. We confirmed reachability via control plane by referencing BGP monitors such as RouteViews and RIPE/RIS. As a result, we confirmed that the /25 prefix had been propagated to 11 AS locations. This matched our expectation that a /25 would be severely filtered and would not propagate far.

Our data-plane measurements were based on pinging a large set of IP addresses spread widely across the Internet, using an IP address from the /25 as the source of the ping packets. Receiving a ping-response therefore indicates that the ping target can reach our prefix. No response could mean that the IP might be down, or the pinged IP might not have a path towards the /25-address space, and so we only draw conclusions from the positive responses.

To our surprise, we found 1,024 ASes that had usable connectivity back to our /25. This represented nearly 5% of all the ASes visible in this experiment. This is not significant in proportion to the Internet as a whole, but it is an extremely large figure compared to the results of control-plane BGP routing information.

Furthermore, according to BGP monitors, all ASes containing routing information for the /25 prefix were within 2 AS-hops of AS 3130*1. In other words it did not propagate further than 2 ASes from the source. The solid line in Figure 2 shows AS number distribution for each number of hops. The prefix was announced by AS 3130, which has two tier-1 upstream providers. The /25 did not propagate further than one hop beyond those tier-1s, only reaching the “core” of the Internet.

Moreover, the results of using traceroute toward the pingable target IP addresses to measure the number of AS hops are shown as the blue dashed line in Figure 2. This was not much different from previous results of measuring reachability for the /20 prefix (the red dashed line in Figure 2). The data-plane measurement results indicate that the /25 is reachable from ASes that are further away (up to 4 hops) compared to the BGP monitor results (up to 2 hops).

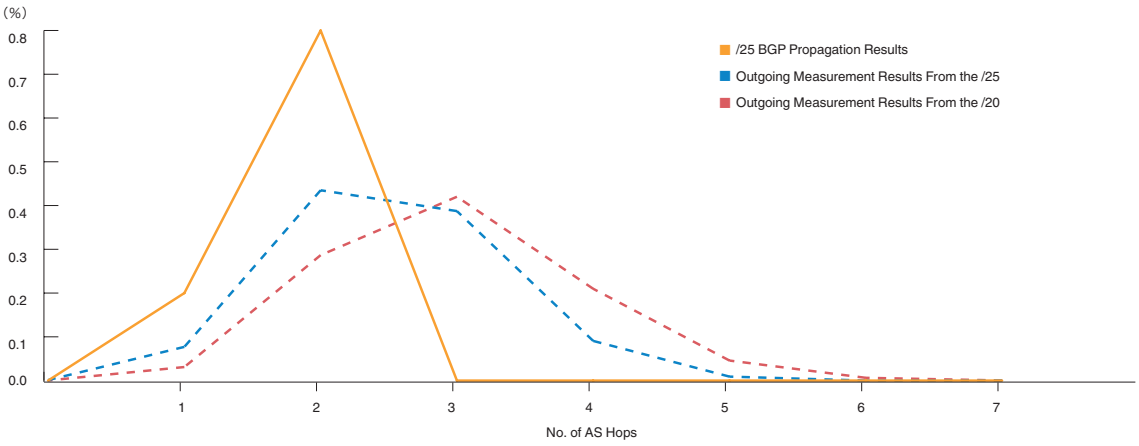


Figure 2: Distribution of the Number of AS-hops to the /25

*1 In this report, we have unified the method of measuring the number of hops with section 3.3.2, instead of using the method in reference 1.

These results show a clear difference between control- and data-plane measurements of reachability. However, data-plane measurements show real reachability, so they need to take precedence. There are two likely reasons for the difference:

- The prefix propagated further than expected on the control plane to sites which were not visible from the standard BGP monitors
- The default routes provided effective connectivity to some ASes despite the fact that they never learned of our prefix

Over 75% of those ASes with data-plane reachability were stub ASes*2. Default routing is presumed to be more common in stubs, so we sought to investigate this cause further.

3.3 Utilization of Default Routing in the Internet

Here, we use AS-path poisoning to measure the extent to which default routing is used. As illustrated in Figure 3, our test box in AS 3130 announced a set of test-prefixes to its upstream tier-1 provider. We announced these prefixes with paths containing the AS number of an AS we want to test for reachability to AS 3130, so we can be sure this AS does not have this prefix in its BGP routing tables.

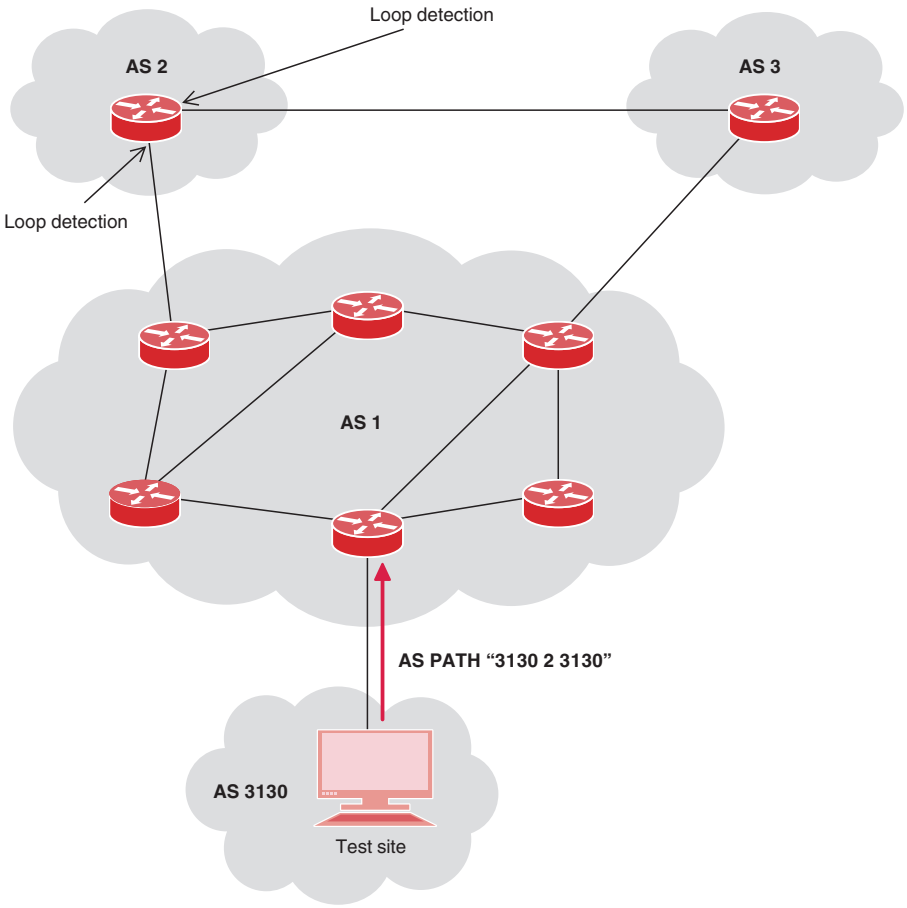


Figure 3: AS Path Poisoning

*2 ASes that do not relay communications from other ASes are called stub ASes. Meanwhile, ASes that relay communications from other ASes are called transit ASes.

For example, if AS 2 is to be measured, we announce routing information containing the “3130 2 3130” AS path. When AS 2 receives this routing information, it sees its own AS number in the AS path, and drops the announcement because of BGP loop prevention. So, as long as AS 2 has no default route, it is possible to create an environment where it cannot reach an IP address inside this test-prefix. We call this method AS path poisoning. AS path poisoning tests were conducted from April 18 2009 until May 1 2009. During that 13 day period, we tested 25,780 ASes for their use of defaults. For the experiment we used the address space 98.128.0.0/16 sliced into /24’s, and carried out concurrent tests using the following steps covering a large proportion of the Internet.

1. We check in advance that the target AS is reachable from an address with an unpoisoned prefix.
2. Next, we withdraw the test-prefix, and wait for 1.5 hours to reduce the possible influence of route flap damping.
3. We then announce the test-prefix containing the AS number of the target AS, and wait for 20 minutes to allow it to propagate.
4. We then start testing the target AS by sending pings from the test-address space to our set of target IP addresses in the relevant AS.

A typical test run takes between 2 and 3 hours. We tested a large number of ASes by conducting the tests in parallel. We also compared test probes of the target IP addresses from unpoisoned address space to check that these addresses were consistently available over the course of the whole experiment. The results showed that the vast majority of ASes (99.2%) were consistently reachable.

The results showed that 64% of all IP addresses tested were still reachable after their AS path was poisoned. We tested multiple IP addresses per AS, and found that 74.8% of ASes (19,291) answered consistently despite the poisoning. In other words, the majority of ASes have a default route configured.

Of the remaining ASes 20.9% (5,381) did not ever answer, and 4.3% (1,108) answered for some IP addresses, but not others. We noticed that some probes to the test address space failed, but this was a very small percentage (0.7%). This is thought to be due to bogon-filters.

We interpret a non-response as the AS being default-free, though this interpretation is less certain as we cannot say that no IP address in the AS ever uses default. The mixed result category reveals the complexity of network management practices for the target AS. For example, one AS apparently does not use BGP default routing, but has a default route manually configured to some routers for IP-TV and VoIP services. This illustrates that some ASes are not operated under a unified policy. Interestingly, there seems to be cultural differences in the use of default routing. In one test, results indicated that 60% of Japan ASes did not use default routing, while 36% did, and 4% had a mixed configuration.

The results of these tests were published on our website, and we took feedback from the ASes that were tested. Of the 191 ASes that replied, 94% confirmed that the results of these tests were correct. Additionally, some of the addresses in the IP address list that we pinged belonged to an address block delegated by that AS to another AS. Surprisingly, some AS administrators were not aware that they were using default routing. This can happen in cases such as when default routing from an upstream provider is accepted without applying a filter.

3.3.1 The Impact of AS Type

Intuitively, we might expect that ASes that provide transit to other networks will be less likely to use default routes than stub ASes. We tested this by breaking down our previous results by AS type. Here, we used the AS classifications provided by reference 2.

Table 1: Default Utilization Distribution by AS Category

	# Tested	Default	Default-free	Mixed
Stub	24,224	77.1%	19.3%	3.6%
Small ISP	1,307	44.5%	42.2%	13.3%
Large ISP	246	17.1%	60.6%	22.3%

As Table 1 indicates, utilization of default routing declines when going from stubs, to small ISPs, to large ISPs. The number of ASes with mixed results for default routing increases from stubs to larger ISPs. This suggests that the operation of larger ISPs is more complex. However, during our tests pings were sent to more IP addresses for larger ASes, so this must be taken into consideration when interpreting the results.
Figure 4 shows a breakdown of the test results against the number of peers with other ASes. This demonstrates that use of default routing declines until having at least 100 AS peers. Additionally, 80% of ASes with 20 or fewer peers rely on default routing, and ASes with 300 or more peers use default routing in less than 15% of the cases.

The uneven popularity of default routing in different types of ASes shown in the test results is extremely interesting. For example, when using a traceroute from a stub AS, it may be able to travel for the first few hops without explicit routing information using a default route, but when it reaches a large ISP the default route drops off, and it stops there. However, this does not necessarily mean that there is a problem with the location the traceroute has stopped at. The fact that the traceroute made it that far differs from the reachability data acquired through control-plane information, suggesting that neither data-plane nor control-plane measurements are adequate by themselves.

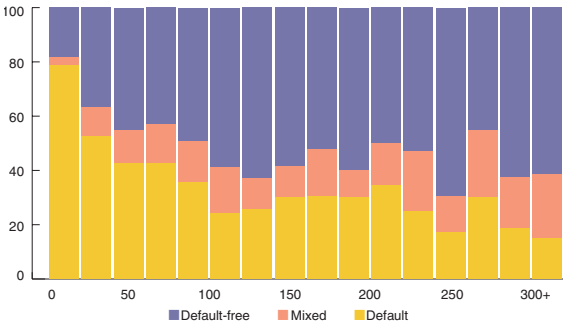


Figure 4: Default Utilization by AS Peer Numbers

3.3.2 The Impact of Defaults

In order to determine the impact that default routing has on the measurements of the Internet, we carried out simulations using the AS topology data from reference 2. In our simulations, we used the default routing utilization ratios from our test results, and assigned default routing to 77.1% of stub ASes, 44.5% of small ISPs, and 17.1% of large ISPs within the topology data. We included mixed cases with the default-free results. We adopted two methods for determining where the default route points for ASes with defaults assigned. The first was a random allocation scheme in which we chose randomly from an AS's upstream providers, and the second was a max allocation scheme in which we chose the AS's upstream provider with the maximum number of customers.

For each simulated topology we chose a random set of 1,000 sources. For each source, we then calculated how many potential destination ASes could be reached from this source using default routes.

The results showed that if we use only defaults, then we can reach only a very few ASes. The Internet hierarchy is relatively flat, so even from a small stub AS, we need only go up through a few layers of providers before we reach a large, or tier-1 provider that doesn't use default. We found that typically only 1-3 ASes could be reached in this way, and that the maximum was 5.

Here, we look at cases where routing information advertisements for a prefix are only propagated to the upstream provider one hop away, and not propagated to ASes beyond that.

Figure 5 shows the Complementary Cumulative Distribution Function (CCDF) of the number of ASes that can be reached from a random source. This graph shows that for the max default allocation we can now reach 1,000 ASes from approximately 50% of sources, and over 2,000 ASes from around 1/3 of sources. For the random allocation of default routes, we can reach somewhat fewer destinations, but the number is still substantial.

Figure 5 also shows the results of the assumption that routing information advertisements propagate two-hops. We can see that this has a dramatic impact on the number of ASes that are reachable, with 6,000 reachable about 50% of the time. There were also cases where up to 19,000 ASes were reachable.

As we observed with the propagation for the /25 prefix from 3.2, it does not actually follow a simple "hop-count" mechanism, and local filters at each AS must also be considered. However, the simulations do provide us with some valuable intuition. It is quite possible that once the /25 reaches our provider, and perhaps a few other ASes, that it will be reachable from a significant proportion of the Internet despite the limited propagation of its routing announcements. We believe these results also shed much light on the phenomena of reachability using data-plane measurements when there is no reachability using control-plane measurements.

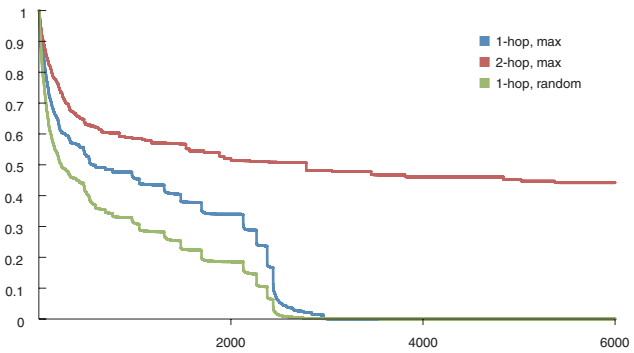


Figure 5: Distribution of Reachable ASes

3.4 Testing Reachability with Dual Probing

The existence of default routing indicates that there is a limit to the predictions that can be made through control-plane observations. Such limitations should be kept in mind before making claims based on control-plane observations alone. On the other hand, the types of data-plane measurements we have used so far are limited as well. It is easy to find situations where it is hard to interpret the results of ping probes simply because end-host (or something in the middle) behavior is so varied. Reachability is hard to measure because the following two viewpoints must be considered.

- How do I see the world?
- How does the world see me?

The first is based on the information a router receives from routing protocols. The second question is something operators would often like to know in order to debug reachability problems. In other words, how does a network appear from other parts of the Internet? Unfortunately, this information is not directly available from the network layer.
There is data available to see how the world sees us. Services such as BGP monitors, looking glasses, and traceroute servers provide public views of the Internet. However, only a sample of ASes operate these as a public service, therefore it is hard to get direct data from the world viewpoint. What we see when we combine data from the available viewpoints is actually a sampled world viewpoint.
A significant problem with this sampled view is that the operators with the sophistication and resources to operate public viewpoints tend to be larger ISPs, nearer the “core” of the Internet, so the bias in the viewpoints could mislead. For instance, we might hypothesize that these large, densely connected ISPs have fewer reachability problems than stub ISPs, so they are not useful for evaluating stub ISP reachability. There is therefore a need for methods that allow reachability to be tested from a variety of viewpoints encompassing the entire Internet.
Here, we advocate a data-plane testing method called dual probing that can be applied to a wider range of situations. Suppose a network administrator wants to check that external hosts can reach their network. A simple test would be to ping from a machine towards a large set of external IP addresses covering much of the Internet. If those IP addresses answer the probes, this indicates that the source machine is reachable from the probe’s destination. We call these “out-probes.”

Figure 6 illustrates the out-probe concept. In the “traditional” case (black solid arrow), a probe is in-bound from a public looking glass towards the test-site. In the out-probe case (green dashed arrows), a probe is sent from the network for which reachability is being tested towards many destinations in the Internet. Note that the address space under investigation must be the source address of the outgoing IP packet. In this case, it is the return traffic towards the test address space that reveals the reachability of the test IP address space.
When the ping probes are not answered, it could be due to one of the following reasons.

- The IP address simply does not answer pings.
- The ping packets are dropped by firewalls on the way towards the probed IP addresses.
- The IP addresses answer the ping probes but the answers are dropped somewhere on the path back towards the probing host.
- There is no route from the IP address in question back to the probing host or vice versa.

Only the latter two cases concern reachability of our test network. However, even the third cause may be a poor indication of unreachability, because ICMP probes are often given lower priority and may be dropped preferentially over TCP traffic. At any rate, negative responses provide little information by themselves.
If we can calibrate our expectations we will know how to interpret the responses of probes. If actual tests are carried out after first conducting tests in advance similar to the methods used in “3.3 Utilization of Default Routing in the Internet,” it is possible to gain useful information by comparing the expected results with the actual ones. In other words, using two probes separated in time allows for better interpretation of the second probe. This is also possible through using multiple IP addresses for the ping target. We call this approach dual probing. Although “dual” is used in the name, in some cases more than two probes may be involved.

Using dual probing, we can compare probe answers against probes from another prefix, called here an anchor-prefix. The anchor prefix is an old, well-established prefix known to have very good reachability. This comparison reveals far more information than a single probe from the test-prefix. Lacking a reply from the anchor probe to a particular IP address we know there is a problem probing this IP address, and so we can discount test measurements as not useful. With a reply to both, we can infer successful reachability. When we receive a ping reply to an anchor probe, but no reply to a test probe, we have evidence that there is a reachability problem somewhere between the IP address of the ping target and the test site. The evidence is not conclusive, as ICMP packets may be dropped, but over a series of such measurements we can build confidence in the results.

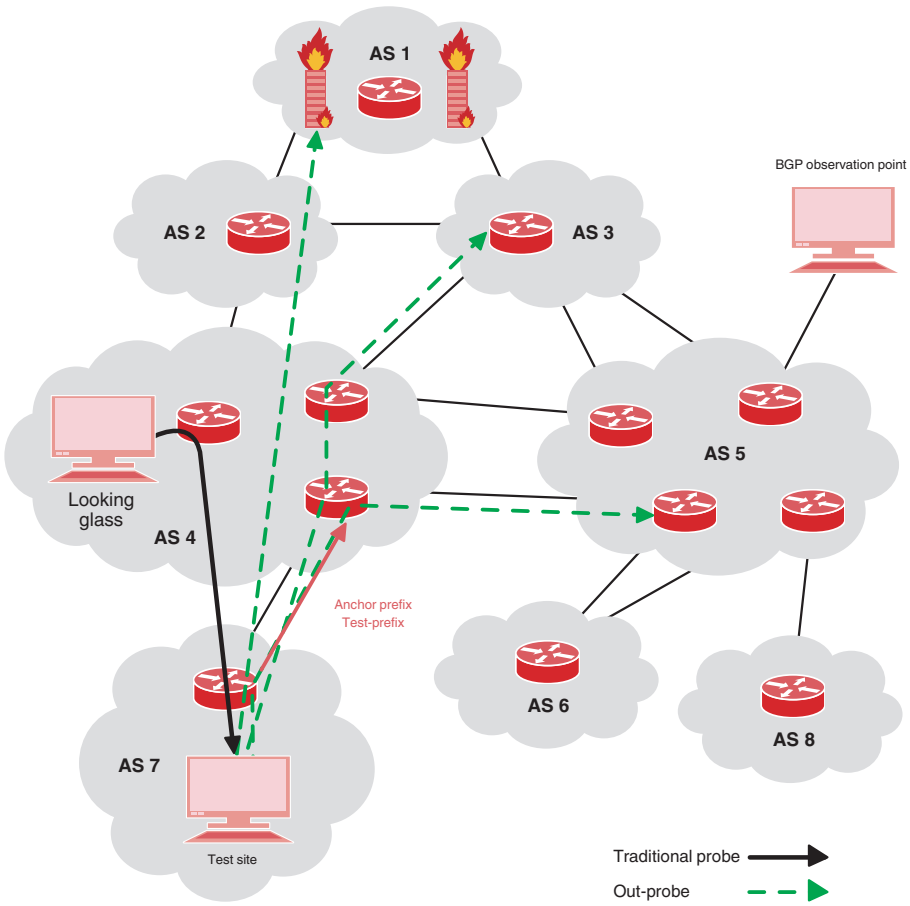


Figure 6: The Dual Probing Concept

3.4.1 Bogus Bogon Filter Detection

A bogon refers to a bogus routing announcement. These are sent either accidentally, or deliberately to hijack address space. Consequently, ISPs commonly configure either control- or data-plane filters to prevent traffic to/from obviously bogus addresses. For instance, it is common to set a filter to refuse traffic or announcements from prefixes that have not yet been allocated to ISPs by the registry. However, the configuration of these filters does not always keep up as prefixes are allocated, so reachability problems to certain prefixes can sometimes occur. In the past it was difficult to detect the location of these bogus bogon filters. We conducted tests to detect bogus bogon filters using dual probing.

ARIN allocated two large segments of new address space (173.0.0.0/16 and 174.128.0.0/16) for our experiment, and we used five smaller segments of this address space. We announced those prefixes from five different locations that volunteered to participate in our experiment: PSGNet in Seattle (USA), Verio in Ashburn (USA), SpaceNet in Munich (Germany), CityLink in Wellington (New Zealand), and IIJ in Tokyo (Japan). The anchor-IP was the normal address of a machine inside the ISP that ran the experiment. A test IP address was selected from the test-prefixes and configured as a secondary IP address on the same interface.

We ran three different measurement campaigns: the first starting in April 14 2008, the second starting on May 27 2008, and the third starting on June 12 2008. About a week was necessary to run all our probes. The first measurement campaign occurred before ARIN announced that this address space had been issued. The goal of these measurements was to understand how prevalent legitimate bogon filters were. After the first campaign, ARIN announced that the address space had been issued, and that it should be removed from bogon black-lists. In addition, we sent emails to the operators of ASes that we identified as having a filter set in the first campaign, and asked them to remove the filter. Thus, reachability problems identified in the second set of measurements are genuine problems, which needed to be fixed. The third set of measurements was used to assess how the reachability problems were changing over time.

For an AS to be identified conclusively as having a reachability problem, we required zero returns back to the test-IP, and at least five returns to the anchor-IP. If there were zero returns to the test address space, and less than five returns to the anchor IP, we considered this an indication that the AS might have a problem.

Figure 7 shows the results of these tests. The black solid line represents all ASes. The green dashed line represents ASes that had no problems. ASes for which a bogon filter was detected are indicated by the red dashed line, totaling approximately 500. The nearly 2,000 ASes represented by the yellow dashed line also potentially had a filter set.

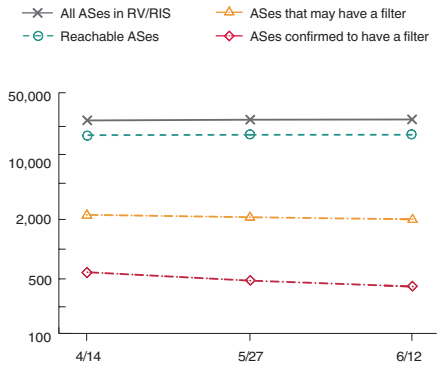


Figure 7: Bogon Filter Detection Tests

These results mean that between 2% and 7% of ASes across the whole Internet could not see the newly allocated addresses. We can also see that the problem barely changes between our second and third tests. This is a serious problem.
Figure 8 shows the distribution of ASes with a filter confirmed or with a filter potentially set by AS type. We can see that the vast majority are stub ASes. This suggests that most problems occur at the edge of the Internet. However, it is possible to incorrectly identify a problem as related to a stub AS when it is actually related to a transit AS, so this must be considered.

3.5 Impact of Methodological Issues on Measurement Confidence

So far we have shown how observations from the control plane do not indicate reachability correctly, and that the data plane can offer a different perspective. However, data-plane observations also have their own limitations. In this section we briefly discuss three methodological issues that should be taken into account for data-plane measurements.

3.5.1 Topological Coverage

The motivation for out-probes is to cover areas of the Internet where no BGP monitors or looking glasses exist — in other words, to look at reachability at the edge, rather than the core of the Internet. To achieve this, it is necessary to create an IP address list that covers the entire Internet, such as that used in "3.4 Testing Reachability with Dual Probing." This address list must have wide coverage, and be able to probe inside ASes for non-homogeneously configured parameters, in addition to limiting the number of probes that have to be sent. The quality of the IP address list that is created affects the quality of actual measurements.

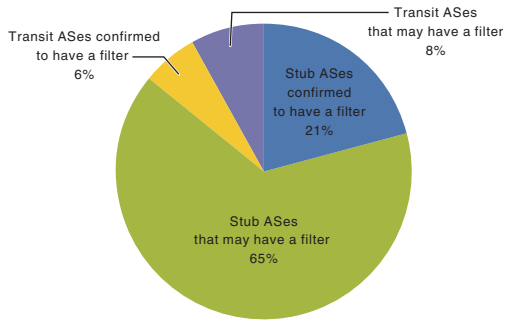


Figure 8: Distribution by AS Type

3.5.2 IP to AS Number Mapping

A general issue we encountered during all experiments is the mapping of IP to AS number. BGP routing table look-ups are used to achieve this. However, when a transit provider connects a customer, an IP address may be provided from an address block allocated to the provider. When this happens, if the router at the border of the customer AS has no reachability to certain prefixes due to a bogon filter, etc., there is a possibility that it may be detected as having no reachability from the upstream provider.

After creating correct IP address to AS number mappings, it is also important to maintain them. For example, comparing mappings we created in 2007 to those we created in 2009, only 88% of the prefixes were still mapped to the same AS. When there is an error in IP address and AS number mapping, observations may be interpreted incorrectly.

3.5.3 What Type of Probe to Use?

The decision of whether to use ping or traceroute for data-plane measurements is also crucial. Moreover, when relying on ping probes, the type of packet (ICMP, UDP, or TCP) used can have significant impact. When conducting our tests we found that for ICMP, roughly 70% of the IP addresses were reachable. For UDP, the responsiveness was only 30%. This is because it is filtered by firewalls and NATs. For TCP the responsiveness was even worse, with around 5% reachability.

3.6 Conclusion

In this paper we showed that assessing actual reachability on the Internet is more complex than it appears from publicly available BGP server data, and that data-plane reachability is different from control-plane visibility. We also demonstrated that packet reachability is provided through default routing even when routing information does not propagate. Additionally, we proposed new methods for verifying reachability using AS path poisoning and dual probing. IIJ will continue to strive for the stable operation of our Internet backbone, in addition to carrying out tests and disseminating information related to the stable operation of the Internet as a whole such as covered in this paper, so the Internet can function as a safe and stable social infrastructure.

References

1. R. Bush, O. Maennel, M. Roughan, S. Uhlig, “Internet Optometry: Assessing the Broken Glasses in Internet Reachability,” ACM SIGCOMM IMC, 2009.
2. R.Oliveria, B. Zhang, “IRL – Internet Topology Collection,” 2009.

Author:
Toshiya Asaba
President and CEO, IIJ Innovation Institute Inc. Mr. Asaba joined IIJ in its inaugural year of 1992, becoming involved in backbone construction, route control, and interconnectivity with domestic and foreign ISPs. Asaba was named IIJ director in 1999, and as executive vice president in charge of technical development in 2004. Mr. Asaba founded the IIJ Innovation Institute Inc. in June 2008, and became president and CEO of that organization.

Internet Topics: Council for Promotion of Anti-Spam Measures

One of the anti-spam measures in place in Japan is the "Act on Regulation of Transmission of Specified Electronic Mail"*1. This law was established and enacted in 2002, and was revised in both 2005 and 2008 due to the inclusion of a provision that it be revised within a three year period.
Before each revision to the act, an Unsolicited Mail Measure Committee was held to evaluate current anti-spam measures and examine possible future directions.
The most recent committee was held in July 2007, and a final report of the results of these discussions was published in August 2008*2.
This final report stated the need for a system for promoting comprehensive anti-spam measures, and a supplementary resolution covering this was made during Diet deliberations for the Act on Regulation of Transmission of Specified Electronic Mail.

As a result, the Council for Promotion of Anti-Spam Measures was established on November 27, 2008*3.
As can be gleaned from the goals of establishment and participating members that are published on the website for the Japan Data Communications Association that administers the council, a wide range of experts from academia, industry, and government take part.
At the first assembly of the council a "spam eradication declaration" was adopted, detailing resolutions towards the eradication of spam and specific measures to be taken.
An administrative group whose role is to examine practical issues and solutions regarding anti-spam measures was also formed from part of the membership when the council was inaugurated.
While engaging in a variety of discussions, members of the administrative group were principally responsible for authoring an Anti-Spam Measure Handbook as a comprehensive summary of the current state of spam and the various countermeasures that exist. This handbook was approved at the second assembly of the council on October 2, 2009, and the 2009 Anti-Spam Measure Handbook was released on October 9*4. This handbook is an all-encompassing overview that covers the current state of spam as well as activities related to systematic and technological countermeasures that are implemented by a variety of organizations.

At the second assembly the establishment of a Sender Authentication Technology Workgroup was also approved, with the goal of promoting the adoption of sender authentication technologies that serve as an infrastructure for anti-spam measures.

Sender authentication technologies have been covered in the Messaging Technology section of this IIR a number of times in the past. While it maintains compatibility with existing email distribution systems, there are a number of points regarding its utilization and effect that may be misinterpreted if they are not understood correctly.
For this reason, the council's Sender Authentication Technology Workgroup is first preparing materials to help users understand the technology before deploying it, in addition to holding information sessions for deployment beginning with members of the council.
The workgroup is examining policies for encouraging further widespread adoption based on the information gathered through these activities.

I am contributing to the activities of the Council for Promotion of Anti-Spam Measures as a member of both the council and its administrative group.
I was given the responsibility of facilitating the Sender Authentication Technology Workgroup upon its establishment. As detailed in this IIR, several varieties of sender authentication technologies exist, each with a number of differences with regard to advantages and disadvantages and cost of deployment. This means there is no one technology that can easily fulfill all requirements. We are evaluating the best uses for each technology, and examining practicalities such as the use of staggered deployment to promote adoption and methods for utilizing authentication results.
IIJ will continue to take a leadership role both in the promotion of anti-spam measures and in the industry as a whole.

Author:
Shuji Sakuraba
Shuji Sakuraba is a Senior Program Manager in the Service Promotion Section of the Messaging Service Division in IIJ Network Service Department. He is engaged in the research and development of messaging systems. He is involved in various activities in collaboration with external related organizations for securing a comfortable messaging environment. He is a MAAWG member and JEAG board member. He is a member of the Council for Promotion of Anti-Spam Measures and administrative group, as well as chief examiner for the Sender Authentication Technology Workgroup. He is also a member of Internet Association Japan's Anti-Spam Measures Committee.

*1 Act on Regulation of Transmission of Specified Electronic Mail: http://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/m_mail.html#ordinance.
*2 Unsolicited Mail Measure Committee Final Report: http://www.soumu.go.jp/menu_news/s-news/2008/080828_8.html.
*3 Council for Promotion of Anti-Spam Measures: http://www.dekyo.or.jp/soudan/anti_spam/index.html.
*4 Regarding the release of the 2009 Anti-Spam Measure Handbook: http://www.dekyo.or.jp/soudan/anti_spam/image/200910press1.pdf.

Ongoing Innovation

Vol.6 February 2010

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.

Internet Initiative Japan Inc.
Address: Jinbocho Mitsui Bldg., 1-105 Kanda Jinbo-cho, Chiyoda-ku, Tokyo, 101-0051
Email: info@iij.ad.jp URL: http://www.iij.ad.jp/

IIJ-MKTG020DA-1004CP-00001PR