Internet Infrastructure Review

IIJ
Internet Initiative Japan

Vol.5

November
2009

Infrastructure Security

Large-Scale DDoS Attacks in the United States and South Korea

Messaging Technology

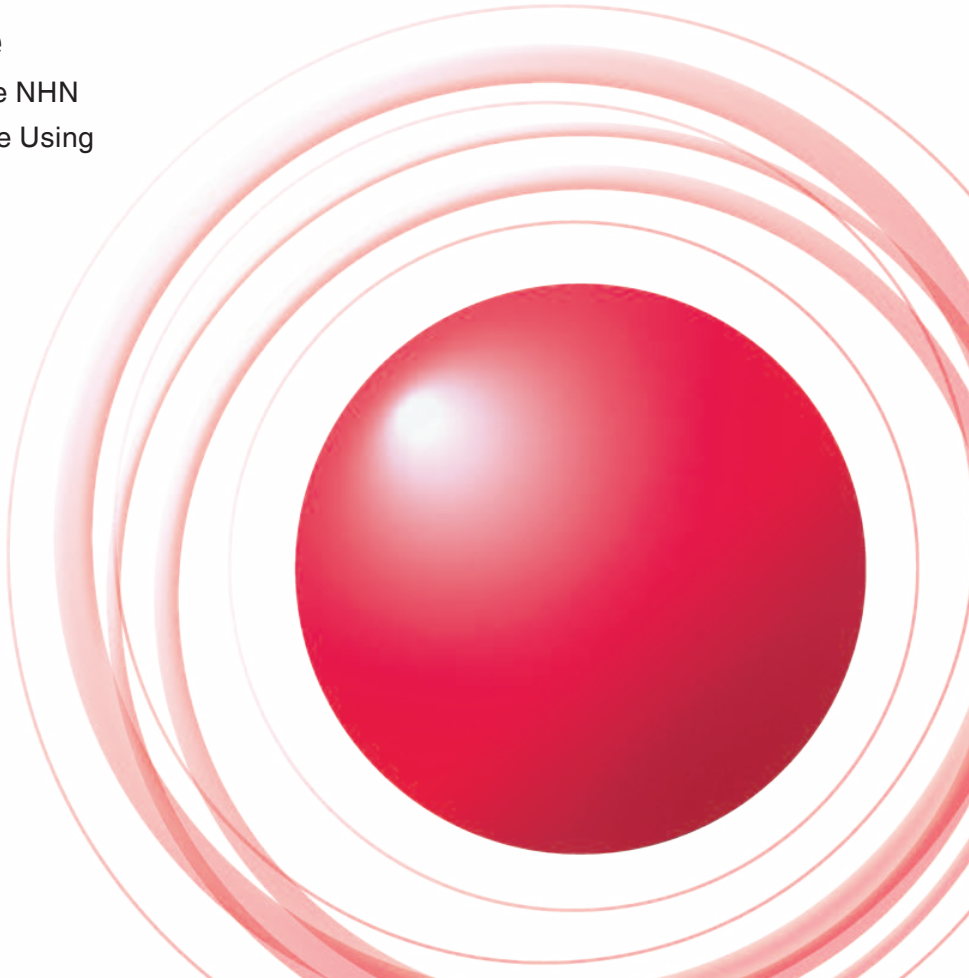
How Do We Discourage Asia from Continuing to be a Source of Spam?

Cloud Computing Technology

Live Migration of Guest Computers Using the NEMO BS Mobility Function

Service Infrastructure

The Design and Implementation of the NHN Next Generation Service Infrastructure Using Virtualization Technology and Remote Data Centers



Internet Infrastructure Review Vol.5 November 2009

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

Executive Summary

On August 17, 2009, the Ministry of Internal Affairs and Communications published the "International Comparative Evaluation Report on ICT Platform in Japan." According to this report, Japan was ranked overall as the top country out of 24 evaluated. A closer look reveals that Japan ranked at the top for broadband affordability, optical fiber installations, and broadband speeds. Indeed, Japan has developed a technological infrastructure that supports high connections speeds at low costs.

In regard to indices representing the advancement of social infrastructure, however, Japan ranked only 11th with respect to the number of Internet hosts and 13th with respect to ICT investment. This leads one to conclude that, despite having established an infrastructure for the network, Japan is lagging in server infrastructure as well as service development and related ICT investment which need well-deployed server infrastructure. Cloud computing infrastructure and related business services leveraging Japan's cheap, fast networks must be developed before one can say that Japan's ICT infrastructure and competitive ability are world-class.

As exemplified in some of the cloud computing models introduced last year and throughout this year, the Internet is an amalgamation of networking and computing. A wide variety of elements come together to comprise the Infrastructure of the Internet, which operates as a complex system built upon many different layers. Monitoring and analysis of the behavior of these various elements and layers—as well as continued technological development in pursuit of new usage models—is an indispensable part of ensuring the development of ICT infrastructure and stable operations of this complex system.

This whitepaper discusses the various monitoring and analysis activities in which IIJ engages to contribute to the perfection and growth of Internet infrastructure. We also offer information related to new and developing technologies.

Addressing Internet security and safety, we will report statistics and analyses of security incidents observed for the three months from July 1 through September 30, 2009 in the "Infrastructural Security" section. Our observations and analyses on DDoS attacks targeted multiple websites in the United States and South Korea, TCP vulnerabilities, and attacks on VoIP services exploiting SIP are also reported as focused researches on issues observed or revealed during this period. Under the heading of "messaging technology," we will report on our analysis of the state of spam, and activities related to the adoption and trends in the advancement of sender authentication technologies.

Addressing technological development, we will offer overall description about "NHN", project to migrate IIJ service infrastructure to the cloud environment. We will also address experiments and verification test results regarding our work to adapt Mobile IP technologies for migrating guest computers in a server virtualization environment on a network basis.

IIJ continues to offer solutions for the stable, secure, and innovative use of the Internet as an infrastructure supporting corporate activities. This whitepaper is one example of how we provide timely updates and information to our clients.

Author:
Toshiya Asaba
Executive vice president. Member of the WIDE Project. 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 he serves concurrently as president and CEO of that organization.

Large-Scale DDoS Attacks in the United States and South Korea

In this report, we will report on incidents to which IIJ responded between July and September 2009. At the same time, we will also cite the details behind large-scale DDoS attacks targeting Web servers in the United States and South Korea, TCP vulnerabilities announced by CERT-FI, and the mechanism behind silent phone calls caused by SIP packets.

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 July 1 through September 30, 2009. A multiple number of Web servers in the United States and South Korea were subject to large-scale DDoS attacks during this period. A series of vulnerabilities related to Web browsers were also discovered during this time, and reports cited a vulnerability in DNS servers and other servers used frequently for the Internet. Additionally, a TCP vulnerability that affects many implementations was announced. Besides these announcements and incidents, there were several incidents that resulted in direct financial damages, including cases of fake security software and extortion in connection with DDoS attacks. 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 July 1 and September 30, 2009. Figure 1 shows the distribution of incidents handled during this period.*1

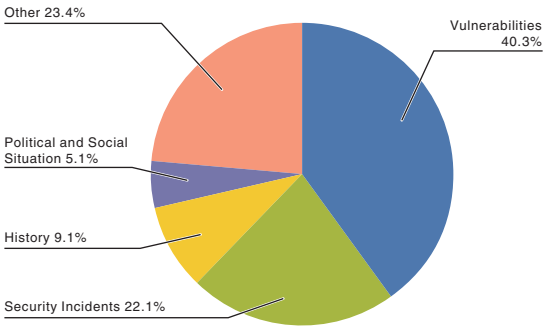


Figure 1: Incident Ratio by Category (July 1 to September 30, 2009)

*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 an attack 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, vulnerabilities were fixed in user applications such as Microsoft's Internet Explorer*2, SMB2.0*3, and Visual Studio Active Template Library*4. Many vulnerabilities were also corrected related to Web browsers, including vulnerabilities in ActiveX killbit*5, JScript*6, Adobe Flash Player and Adobe Acrobat Reader*7, and Apple QuickTime*8.

In addition to the foregoing, vulnerabilities were discovered that affect the stability of BIND9*9, Squid*10 and other software utilized frequently on servers. A Cisco router BGP vulnerability*11 was corrected, and a regular update for IOS was released to address several vulnerabilities during the period under study*12. A vulnerability related to TCP was publicly announced, which affected a large number of implementations. See "1.4.2 TCP Vulnerability (Sockstress)" for more about this TCP 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, Japan observed the 45th House of Representatives general election, the inauguration of the Consumer Affairs Agency, and other political events. However, IIJ noted no related Internet attacks.

■ History

The period in question included several historically significant days, including the observance of the end of World War II and the observance of the end of the Pacific War in Japan. In the past, historically motivated DDoS attacks and website alterations have occurred during this time of the year, and IIJ paid particular attention to political and social situations. However, no directly related attacks targeting IIJ facilities or customer networks were detected.

■ Security Incidents

Unanticipated security incidents not related to political or social situations occurred in the form of multiple large-scale DDoS attacks against web servers in the United States and Republic of Korea (South Korea) during the first part of July. See "1.4.1 DDoS Attacks in the United States and South Korea" for more related to these incidents. Additionally, attacks on P2P file sharing networks from a cloud environment*13 and a DDoS attack against Twitter*14 occurred. In August, a DDoS attack was accompanied by a demand for money, euphemistically called a cost of measures*15.

*2 Microsoft Security Bulletin MS09-034 – Critical: Cumulative Security Update for Internet Explorer (972260) (http://www.microsoft.com/ technet/security/bulletin/ms09-034.mspx).

*3 Microsoft Security Bulletin MS09-050 – Critical: Vulnerabilities in SMBv2 Could Allow Remote Code Execution (975517) (http://www.microsoft.com/japan/security/Bulletin/MS09-050.mspx).

*4 Microsoft Security Bulletin MS09-035 – Moderate: Vulnerabilities in Visual Studio Active Template Library Could Allow Remote Code Execution (969706) (http://www.microsoft.com/technet/security/bulletin/ms09-035.mspx) and Security Advisory for Adobe Flash Player APSA09-04 (http://www.adobe.com/support/security/advisories/apsa09-04.html).

*5 Microsoft Security Bulletin MS09-032 – Critical: Cumulative Security Update of ActiveX Kill Bits (973346) (http://www.microsoft.com/ technet/security/bulletin/MS09-032.mspx).

*6 Microsoft Security Bulletin MS09-045 – Critical: Vulnerability in JScript Scripting Engine Could Allow Remote Code Execution (971961) (http://www.microsoft.com/ technet/security/bulletin/ms09-045.mspx).

*7 Security advisory for Adobe Reader, Acrobat and Flash Player APSA09-03 (http://www.adobe.com/support/security/advisories/apsa09-03.html). Security updates available for Adobe Flash Player, Adobe Reader and Acrobat APSB09-10 (http://www.adobe.com/support/security/bulletins/apsb09-10.html).

*8 About the security content of QuickTime 7.6.4 (http://support.apple.com/kb/HT3859).

*9 BIND Dynamic Update DoS (https://www.isc.org/node/474). This vulnerability relates to a BIND server which holds zone information as a primary server. Even servers that provide only cache functions still frequently have zone information such as localhost, and need to be patched.

*10 Squid Proxy Cache Security Update Advisory SQUID-2009:2 (http://www.squid-cache.org/Advisories/SQUID-2009_2.txt).

*11 Cisco Security Advisory: Cisco IOS Software Border Gateway Protocol 4-Byte Autonomous System Number Vulnerabilities (http://www.cisco.com/warp/public/707/cisco-sa-20090729-bgp.shtml).

*12 Cisco Security Advisory: Summary of Cisco IOS Software Bundled Advisories, September 23, 2009 (http://www.cisco.com/warp/public/707/cisco-sa-20090923-bundle.shtml).

*13 A cNotes article reported this incident. Attacks on Share P2P networks utilizing Amazon Web Service (http://jvnrss.ise.chuo-u.ac.jp/csn/index.cgi?p=Amazon+Web+Service%A4%F2%CD%F8%CD%D1%A4%B7%A4%BFShare%A5%CD%A5%C3%A5%C8%A5%EF%A1%BC%A5%AF%A4%D8%A4%CE%B9%B6%B7%E2)(in Japanese).

*14 Twitter's tweet discussing the status of ongoing denial-of-service attacks (http://status.twitter.com/post/157191978/ongoing-denial-of-service-attack). The Arbor Networks blog detailed the decrease in traffic volume "Where Did All the Tweets Go?" (http://asert.arbornetworks.com/2009/08/where-did-all-the-tweets-go/)

*15 See the following article from Trend Micro regarding similar incidents "Botnet Extortion Attempts Extend to Japan (DDoS Attacks)" (http://blog.trendmicro.co.jp/archives/1385)(in Japanese).

■ Other

As far as incidents not directly related to security, several international undersea cables were damaged by a typhoon in Taiwan, affecting communications in and out of the area*16. A number of anti-virus software firms released 2010 updates, with concurrent releases of counterfeit software nearly indistinguishable from these programs*17. Similarly, as Microsoft released its free Microsoft Security Essentials*18 anti-virus tool during the period under study, fake security software (scareware) began to appear in search engine results, inducing unsuspecting users to click through*19.

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 July 1 and September 30, 2009.

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.

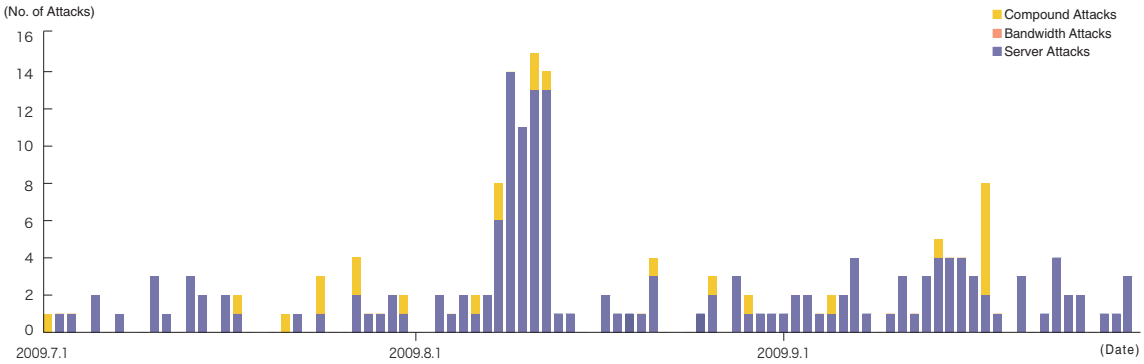


Figure 2: DDoS Attacks

*16 Reports about this incident include the following from NetworkWorld. "Asian undersea cable disruption slows Internet access" (http://www.networkworld.com/news/2009/081209-asian-undersea-cable-disruption-slows.html).

*17 Blog entry regarding counterfeit software accurately imitating Symantec products. Symantec Security Blogs: Nort "what" AV? (http://www.symantec.com/connect/blogs/nort-what-av).

*18 Microsoft Security Essentials (http://www.microsoft.com/security_essentials/).

*19 IIJ has confirmed that links leading to fake software have ranked high in English environment.

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*20, attacks on servers*21, 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 192 DDoS attacks. This averages to 2.08 attacks per day, representing an increase in the average daily number of attacks compared to our prior whitepaper. Considering the fact that a multiple number of attacks occurring between August 9 and August 12 targeted a certain website for a lengthy period, the overall trend did not otherwise vary significantly from that discussed in our prior whitepaper.

Bandwidth capacity attacks accounted for 0% of all incidents. Server attacks accounted for 87% of all incidents, and compound attacks accounted for the remaining 13%. The largest attack observed during the period under study was a compound attack that tied up 566Mbps of bandwidth using 140,000pps packets. Of all attacks, 80% ended within 30 minutes of commencement, while 19% lasted anywhere from 30 minutes to up to 24 hours. During the time period under study, IIJ noted one attack that lasted for 94 hours and 30 minutes (approximately four days).

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*22 and botnet*23 usage as the method for conducting DDoS attacks.

1.3.2 Malware Activities

Here, we will discuss the results of the observations of the Malware Investigation Task Force (MITF)*24, a malware activity observation project operated by IIJ. The MITF uses honeypots*25 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.

*20 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.

*21 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.

*22 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.

*23 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."

*24 Malware Investigation Task Force (MITF). The MITF began activities in May 2007 observing malware network activity through the use of honeypots in an attempt to understand the state of malware activities, to gather technical information for countermeasures, and to link these findings to actual countermeasures.

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

Status of Random Communications

Figure 3 shows trends in the total volumes of communications coming into the honeypots (incoming packets) between July 1 and September 30, 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 client-targeted scanning behavior using TCP ports utilized by Microsoft operating systems. As with the prior study, we observed scanning behavior attempting to exploit 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 53248/TCP and 20689/TCP (not used by general applications), were also observed. Attacks on 445/TCP, etc., targeting Microsoft vulnerabilities have continued since last October. Looking at the overall sender distribution by country, we see that attacks sourced to China and Japan, 26.6% and 24.4%, respectively, were comparatively higher than the rest.

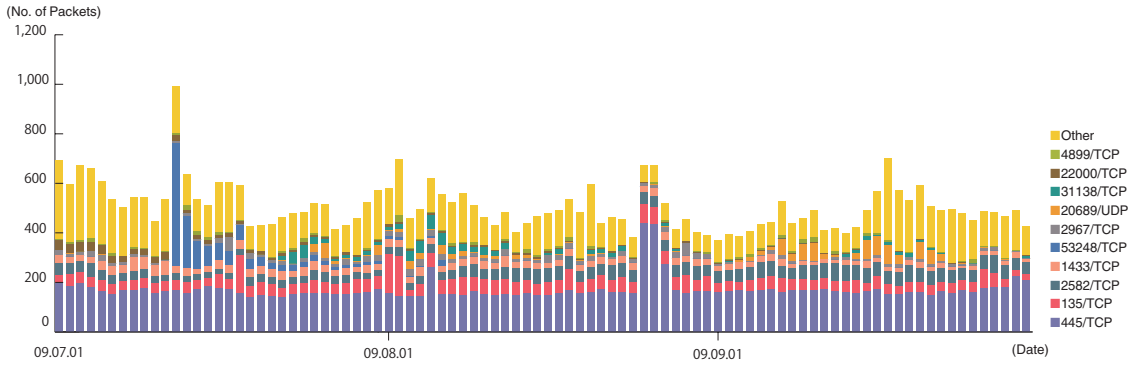


Figure 3: Communications Arriving at Honeypots (by Date, by Target Port, per Honeypot)

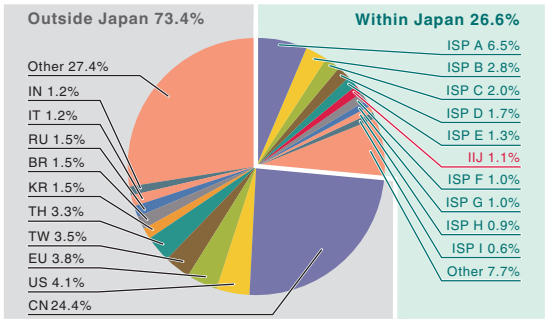


Figure 4: Sender Distribution (Entire Period under Study)

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*26, while the number of unique specimens is the number of specimen variants categorized according to their digest of a hash function*27.

On average, 592 specimens were acquired per day during the period under study, representing about 46 different malware variants. According to the statistics in our prior whitepaper, the average daily total for acquired specimens was 708, with 60 different variants, indicating a slight decline in average number of specimens and number of variants.

The distribution of specimens according to source country has Japan at 64.4%, with other countries accounting for the 35.6% balance. Of the total, malware infection activity among IIJ users was 1.5%—a significant decrease compared to the 16.8% figure reported in our prior whitepaper. Looking more closely at the malware variants, after June of this year, we see that this trend has resulted from a dramatic decline in activities attempting to infect computers with Virut*28 and its variants and activities related to Sdbot*29 and its variants on the IIJ network.

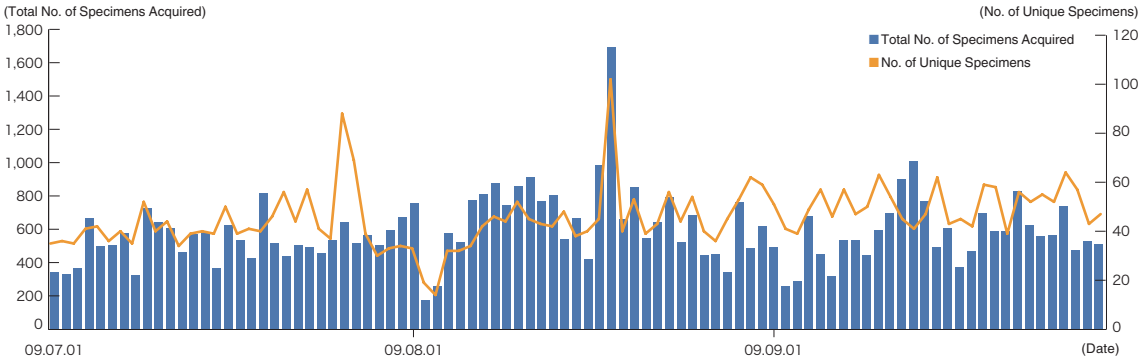


Figure 5: Trends in the Number of Malware Specimens Acquired (Total Number, Number of Unique Specimens)

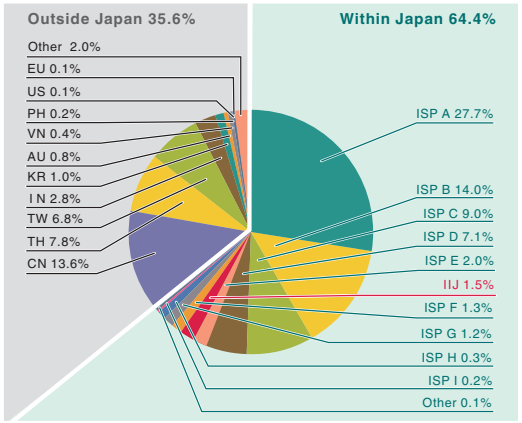


Figure 6: Distribution of Acquired Specimens by Source (Entire Period under Study)

*26 This indicates the malware acquired by honeypots.

*27 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.

*28 Virut is a virus spread through an infected file, and is not generally spread through networks. The propagation of this virus was attempted as the result of an attack exploiting a vulnerability. See Trend Micro's explanation of Virut (http://threatinfo.trendmicro.com/vinfo/virusencyclo/default5.asp?VName=MAL_VIRUT). Also see an alert from the Information-Technology Promotion Agency, Japan (IPA) regarding Virut propagation via web content (http://www.ipa.go.jp/security/txt/2009/03outline.html)(in Japanese). A similar attempt to propagate malware was detected at the Cyber Clean Center, suspected to be related to other malware infection activities (https://www.ccc.go.jp/report/200907/0907monthly.html)(in Japanese).

*29 An Sdbot is a type of bot conducting communications with the C&C server via IRC. See Trend Micro's explanation of the Sdbot (http://threatinfo.trendmicro.com/vinfo/virusencyclo/default5.asp?vname=WORM_SDBOT.GEN).

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, 4.5% of the malware specimens were worms, 89.6% were bots, and 5.9% were downloaders. In addition, the MITF confirmed the presence of 44 botnet C&C servers*30 and 548 malware distribution sites.

1.3.3 SQL Injection Attacks

Of the types of different Web server attacks, IIJ conducts ongoing surveys related to SQL injection attacks*31. 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 July 1 and September 30, 2009. Figure 8 shows the distribution of attacks according to source. These data are a summary of attacks detected by signatures on the IIJ Managed IPS Service. Japan was the source for 67.3% of attacks observed, while China and the United States accounted for 11.6% and 4.9%, respectively, with other countries following in order.

We noted a decrease in SQL injection attacks on web servers compared with our prior whitepaper. While the total number of SQL injection attacks declined, the decrease among source countries outside of Japan was particularly notable. Accordingly, the ratio of attacks sourced to Japan experienced a significant increase.

As previously shown, attacks of various types were properly detected and dealt with in the course of service. However, such attacks are constant and ongoing, calling for continued vigilance.

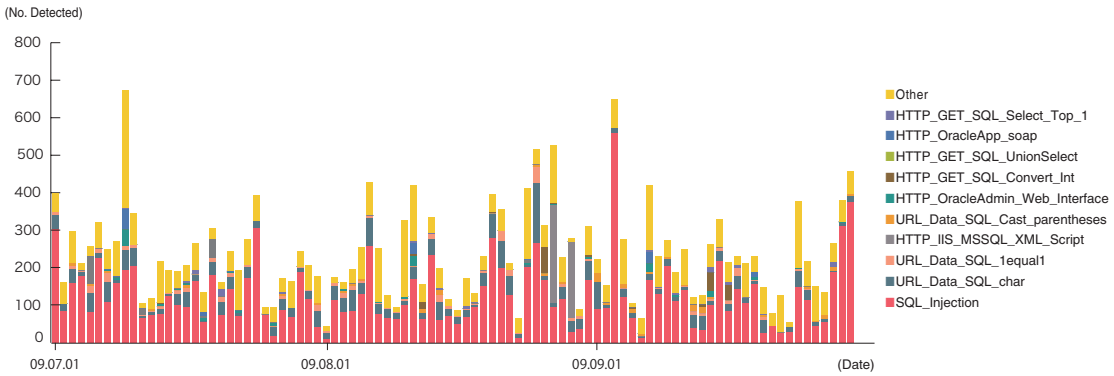


Figure 7: Trends in SQL Injection Attacks (by Day, by Attack Type)

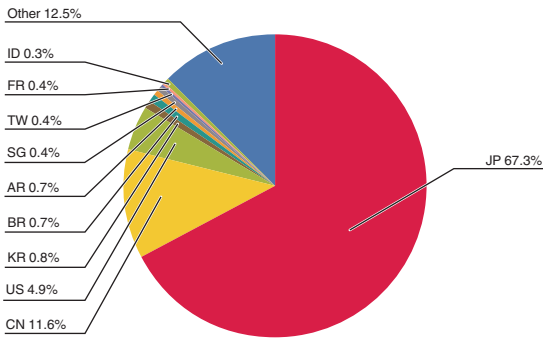


Figure 8: Distribution of SQL Injection Attacks by Source

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

*31 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 gathered from surveys performed during the period from July 1 and September 30, 2009 related to DDoS attacks in the United States and South Korea, TCP vulnerability (Sockstress), and randomly arriving SIP packets.

1.4.1 DDoS Attacks in the United States and South Korea

In early July 2009, websites in the United States and South Korea were victimized by a series of simultaneous DDoS attacks*32. In this section, we will discuss the circumstances of these attacks, based on information obtained by IIJ.

■ DDoS Attack Background

This particular series of DDoS attacks did not use botnets so widely encountered today, but rather malware designed specifically for these attacks. It is reported that this specially designed malware propagated through websites inside South Korea used for file sharing*33. Because of this, many of the IP addresses traced as the source of the attack reportedly have been identified as South Korean IP addresses*34. Malware files of a similar type were placed on similar web services outside South Korea, infecting PCs in other countries*35. While it is unclear as to what timeframe these infection activities took place, it is believed that the propagation activities took place in a concentrated period immediately before the DDoS attacks in order to evade detection and countermeasures by anti-virus software vendors*36. The total number of PCs infected by the malware is undetermined, but a subsequent announcement out of South Korea indicated that the infection affected approximately 78,000 machines within the country*37.

The DDoS attacks first occurred on July 5*38 and July 6 (Korean time), mainly targeting multiple government agencies' web servers in the United States. After July 7, the attacks moved to multiple websites in South Korea. The attacks in South Korea affected not only government agencies, but also online banking sites, webmail services, and other popular online consumer services. It has been reported that the infected PCs used for the DDoS attacks did not generate significant amounts of traffic singly. Rather than occupying communication lines using massive volumes of traffic, the attacks mainly put a direct load on the target servers*39.

This particular DDoS attack died down on July 10, eventually running its course*40. That the malware was cleaned from 95% of the approximately 78,000 machines infected within four days was mostly due to the efforts of South Korean ISPs, security organizations, and media*41. As a result of these efforts, the DDoS attack converged as of July 10. There was a report that only several hundred hard drives were destroyed as a side effect of infection with the malware.

*32 F-Secure reported on its blog that several websites in the United States became inaccessible because of this DDoS attack (http://www.f-secure.com/weblog/archives/00001720.html). South Korean news organizations also reported that several websites in that country became inaccessible due to the attack.

*33 Web-based file-sharing services (so-called "uploaders" in Japan) are used frequently in South Korea by corporations and educational institutions. It is believed that many users became infected when files containing the malware infection package were uploaded to several of these services.

*34 The following report states that more than 100,000 PCs were conscripted in this attack, between 90% and 95% of which were located in South Korea (http://www.shadowserver.org/wiki/pmwiki.php/Calendar/20090710).

*35 The following is an alert from JPCERT/CC discussing the attack from Japan, "DDoS attack against Web sites in South Korea and US" (http://www.jpcert.or.jp/at/2009/at090012.txt)(in Japanese).

*36 As one circumstantial evidence, a specimen obtained and studied by IIJ forged a file creation date of 2004, while the time stamp in the PE header of each file showed a date immediately prior to the launch of the attack, "July 4, 2009 0:38" for perfvwr.dll as an example.

*37 Information related to the number of affected machines inside South Korea is detailed in a presentation at APNIC28 by KRNIC of KISA (Korean Internet & Security Agency) (http://meetings.apnic.net/__data/assets/pdf_file/0019/14077/lee-ddos-attack.pdf).

*38 The first attack began at 2am, July 5 Korean local time. The time was 13:00 EDT July 4 (Independence Day) in the United States. There is no time zone difference between Korea and Japan.

*39 For example, the following alert from KrCERT indicates the circumstances of the DDoS attack traffic from this malware (http://www.krcert.or.kr/noticeView.do?num=340). (in Korean)

*40 With traffic returning to normal levels, the Korean NCSC (National Cyber Security Center) lowered its warning level from an "alert" to a "notice" on July 12.

*41 According to the details published by KrNIC in APNIC28 (http://meetings.apnic.net/28/program/apops/transcript#ji-young-lee). Subsequent to the DDoS attack in South Korea, information about the malware used in the attack and tools dedicated to eliminate the malware were promptly provided by several anti-virus vendors. Television news programs and popular web services issued notifications regarding the attack in conspicuous ways in efforts to alert the public and publicize information regarding corrective action. Information was also released notifying users that backing up their system clocks to an earlier time was an effective temporary measure against one of the malware variants that would destroy hard drives on a specific date (July 10).

■ Malware used in the DDoS Attacks

Once this attack occurred, IIJ first obtained several malware specimens from general malware-related information sources and organizations. These specimens included malware that first induces malware infections for DDoS attacks, malware containing functions for actual DDoS attacks and malware that continues to update attack targets.

Having analyzed and conducted demonstration tests on these specimens, we learned that the malware utilized in these attacks behaved as shown in Figure 9*42.

First, the initial malware package (msiexec*.exe, etc.) drops (creates) two types of malware (perfvwr.dll (or wmiconf.dll) and wmcfg.exe) (1). The malware perfvwr.dll (or wmiconf.dll) first turns off the personal firewall of the infected PC prior to commencing the attack. Next, the infected PC connects to three separate servers, creating a configuration file (uregvs.nls) for the attack (2). The perfvwr.dll and wmiconf.dll malware files start a DDoS attack according to the configuration file (3). The configuration file contains the length of time, target server, attack type and number of attacks for the DDoS attack. This malware-based attack is performed as shown in Figure 9 (4), in accordance with the configuration file. The results of the IIJ demonstration tests showed that the attack traffic generated per machine was 110pps for TCP SYN flood, 110pps for TCP ACK flood, and around 216pps for UDP and ICMP floods, with 107cps (commands per second) for HTTP GET flood and HTTP POST flood. We also observed behavior of intermittent increases and decreases in communications via program-embedded temporary suspension command.

Meanwhile, wmcfg.exe drops two more files—mstimer.dll and wversion.exe (5). The mstimer.dll file downloads a file called flash.gif from multiple web servers (6), extracting and updating a file called wversion.exe from flash.gif (7), while at the same time sending spam to multiple addresses (8). The wversion.exe file deletes the mstimer.dll file and itself, removing evidence (9). However, a function is inserted after the update to search and destroy files on the hard drive having certain extensions (10), writing certain character strings to the hard drive MBR*44, and preventing the PC from being turned on (11).

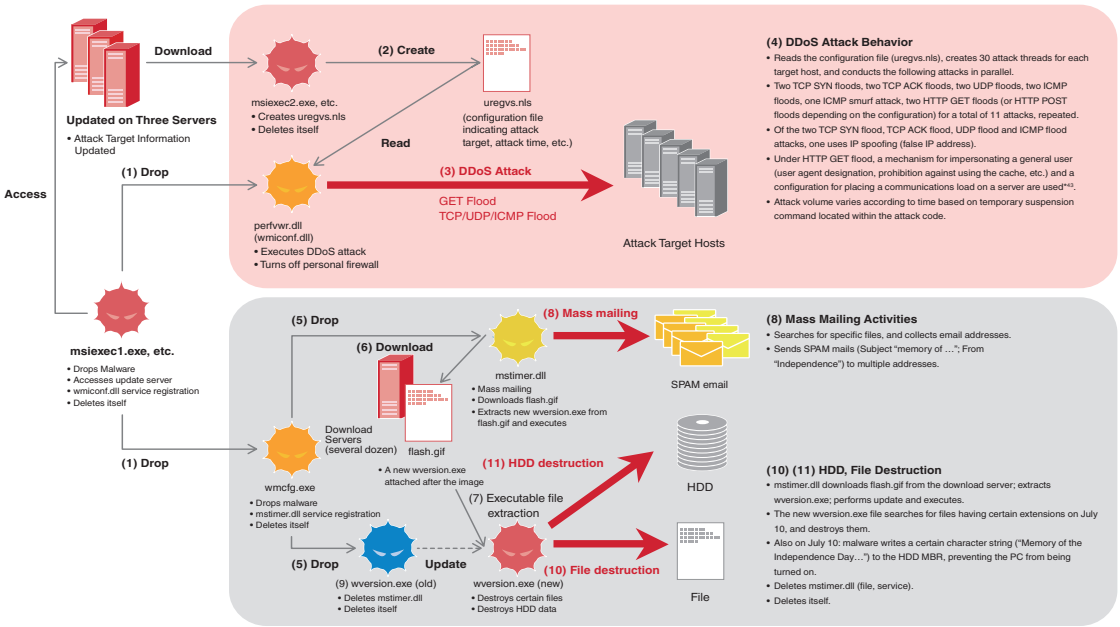


Figure 9: Behavior of Malware used in DDoS Attacks

*42 Our descriptions of this behavior is based on our best efforts at directly gathering information; however, the description includes some information that IIJ was not able to directly confirm about the roles and status of various Internet servers at the time of the DDoS attack.

*43 Various patterns were noted; character strings forged Firefox, IE7.0, or IE8.0 in the User-Agent header. "ko" (Korean language) was designated in the Accept-Language header. The Cache-Control header of some requests showed no-store, must-revalidate.

*44 MBR: Acronym for Master Boot Record. The MBR is a region at the front of the hard drive; normally programs used for launching the computer's operating system are stored in this region.

■ Summary of the Attack

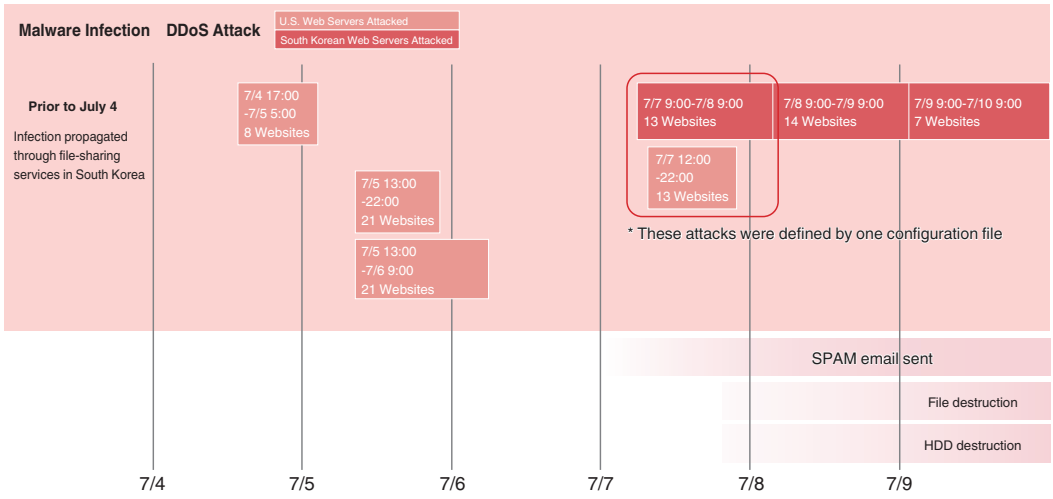
Figure 10 shows the chronological order of the progression and related events for the attack described in this section.

Normally, DDoS attacks are designed for harassment or other clear intent targeting a certain website, and are generally one type of security incident for which the purpose is comparatively easy to identify. The DDoS attack described here, however, was highly complex (malware propagates within South Korea first, attacks targets in the U.S. and then South Korea*45 and intervening malware divides into two independent malware packages, etc.) and the purpose was difficult to identify.

■ Measures against This Particular Type of DDoS Attacks

Here, we will consider the difference between the measures against the damages incurred due to this particular DDoS attack and those that would normally be taken against a DDoS attack. This attack consisted of the mass operations of numerous PCs infected with the malware, making it difficult to institute access or bandwidth controls for the individual PC IP addresses from which the attack traffic was generated. However, much of the attack traffic originated from within one country, and performing access and bandwidth controls on a network basis appeared to be an effective temporary countermeasure. A characteristic of this incident was the low volume of attack traffic from individual PCs infected with the malware. In particular, the web requests involved in the attack were disguised to imitate user behavior, making the determination between normal and attack traffic comparatively difficult. Even if anti-DDoS equipment were available, some adjustments would have been required to establish the abnormal behavior detection threshold and operating mode configurations.

At IIJ, we believe it is necessary to research measures that would deal with this type of DDoS mechanism if such an attack were to take place within Japan. Where the malware executed an attack based on a configuration file distributed beforehand, rather than dealing with the attack by taking action against a centrally controlled botnet, the only solution for stopping such an attack is to eliminate the malware on each individually infected PC. As we saw take place in South Korea, to promptly delete the malware from so many individual PCs requires the cooperation and coordination of many different organizations. To be able to accomplish something on this scale, organizations must be vigilant and ready to engage in synergistic action when called for*46.



Number of targets includes those reported in media plus targets confirmed by IIJ

Figure 10: DDoS Attacks against the United States and South Korea: Chronological Order (UTC)

*45 According to the specimen acquired, 13 U.S. websites (from July 7, 2009 18:00 to July 8 18:00) and 13 South Korean websites (from July 7, 2009 21:00 to July 8 7:00) were attacked under the configuration in the same file. We see that this is the dividing line between where the attacks transitioned to domestic South Korean servers.

*46 In Japan, cyber security exercises have been conducted such as those sponsored by "Anti-Cyber Attack Exercises in the Telecommunications Business Sector"(http://www.soumu.go.jp/menu_news/s-news/2006/061201_4.html) (in Japanese) of the Ministry of Internal Affairs and Communications, and those by Telecom-ISAC Japan (https://www.telecom-isac.jp/english/index.html). There also have been international exercises including APCERT drills (http://www.apcert.org/documents/pdf/APCERT-drill-2008.pdf). IIJ actively participates in these exercises.

1.4.2 TCP Vulnerability (Sockstress)

In September 2009, Finnish CSIRT organization CERT-FI published the state of the response to vulnerabilities related to TCP.*47 This announcement was picked up widely in the press. In this section, we will explain these TCP vulnerabilities and responses.

■ Background

The vulnerabilities themselves were first identified one year prior to the Finnish announcement. Two researchers from a security vendor Outpost24 first identified the issue. These researchers developed a tool to speed up network scanning called Unicornscan*48. During their use of this new tool, they happened to notice unexpected behavior in TCP, and created a tool called Sockstress*49 to generate traffic that exploited this behavior (this tool has not been publicly released).

Based on the information demonstrating the existence of these vulnerabilities, CERT-FI took the lead in organizing a community that encouraged product developers to take appropriate measures. In Japan, this issue is being handled by the Information Security Early Warning Partnership*50.

■ Details of the Vulnerabilities

The tool for exploiting these vulnerabilities (Sockstress) has not been released to the public, nor have the complete details about these vulnerabilities been disclosed. Here, we offer a commentary regarding “zero window size,” which has been most widely discussed in the public.

An attack using zero window size occurs as follows:

1. Establish a TCP connection from a client to a server.
2. During communications, the client specifies zero as its receive window size to declare “buffer full, cannot receive any more data.” In this state, the server will temporarily suspend data transmission via this TCP connection. The server will continue to query the current client-side receive window size in certain intervals, maintaining the connection as long as a response is received.
3. Steps 1 and 2 above from client to server are repeated over and over again.
4. Server resources are exhausted, and new TCP connections cannot be accepted.

In fact, the time for an attack to become effective depends on the server implementation, resources and capacity. It is also possible that while the load climbs high, an attack doesn’t become successful.

Maintaining a connection under a zero window size state is actually a normal operation under TCP standards RFC793*51 and RFC1122*52. Even regular clients using TCP for

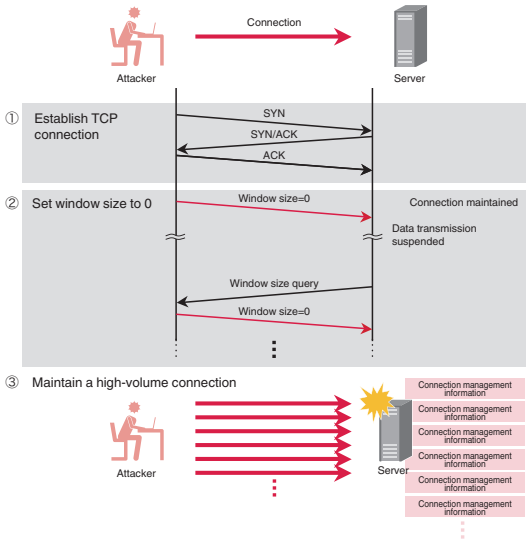


Figure 11: Server-side stack through zero window size designations

*47 Response to TCP vulnerabilities (https://www.cert.fi/haavoittuvuudet/2008/tcp-vulnerabilities.html) (http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2008-4609) (http://www.microsoft.com/technet/security/bulletin/ms09-048.mspx), etc.

*48 Unicornscan (http://www.unicornscan.org/).

*49 While Sockstress itself has not been released publicly, related information has been summarized (http://sockstress.com/).

*50 The Information Security Early Warning Partnership is a framework for vulnerability information dissemination to product developers based on the Ministry of Economy, Trade and Industry directive #235. Under this partnership, the IPA (http://www.ipa.go.jp/security/english/quarterlyrep_vuln.html#Partnership) acts as the vulnerability information reception agent, while the JPCERT Coordination Center (http://www.jpcert.or.jp/english/vh/project.html)serves as the coordinating agent to product developers. Information about vulnerabilities is published through JVN (http://jvn.jp/en/). IIJ participates in this partnership as a developer of proprietary router and other products.

*51 RFC793 Transmission Control Protocol (http://www.ietf.org/rfc/rfc793.txt).

*52 RFC1122 Requirements for Internet Hosts - Communication Layers (http://www.ietf.org/rfc/rfc1122.txt).

communications can designate zero window size as normal communications control. The issue pointed out here is that the designation of zero window size can be used to mount an attack, presenting a method that forces the server to maintain TCP connections at volumes exceeding system resources. However, this zero window size attack is not a new method. For example, the IETF*53 TCPM Working Group*54 broached this subject in July 2006*55, well before the advent of Sockstress.

■ A Protocol Issue or an Implementation Issue?

As shown above, the zero window size designation is a normal state according to protocol specifications. However, a large volume of TCP connections in this state is seen as an issue. To resolve this issue, arguments call for either making changes to the TCP protocol itself, or establishing a timeout value or other adjustment in the implementation as a workaround.

Arguments in the earlier-mentioned IETF TCPM Working Group state that this type of attack is an issue of the OS or server implementation resource management. The consensus is that the solution should be made within implementation in consideration of individual circumstances, rather than treating this as a protocol issue.

■ Effectiveness of Countermeasures

IIJ picked up the Microsoft patch as one means of dealing with this zero window size issue and performed demonstration tests of an implementation before and after applying the patch. Figure 12 shows the results of this test*56.

As shown in the results of this test, the patched implementation demonstrates a stronger resistance to this type of attack. However, this countermeasure forces the termination of existing TCP connections to make room in resources to accept new connections. We have to admit that important connections may also be unavoidably terminated. This countermeasure does not completely prevent this type of attack. Not only Microsoft, but many of the other entities publishing a patch dealing with this issue have adopted measures that similarly control the usage of limited resources.

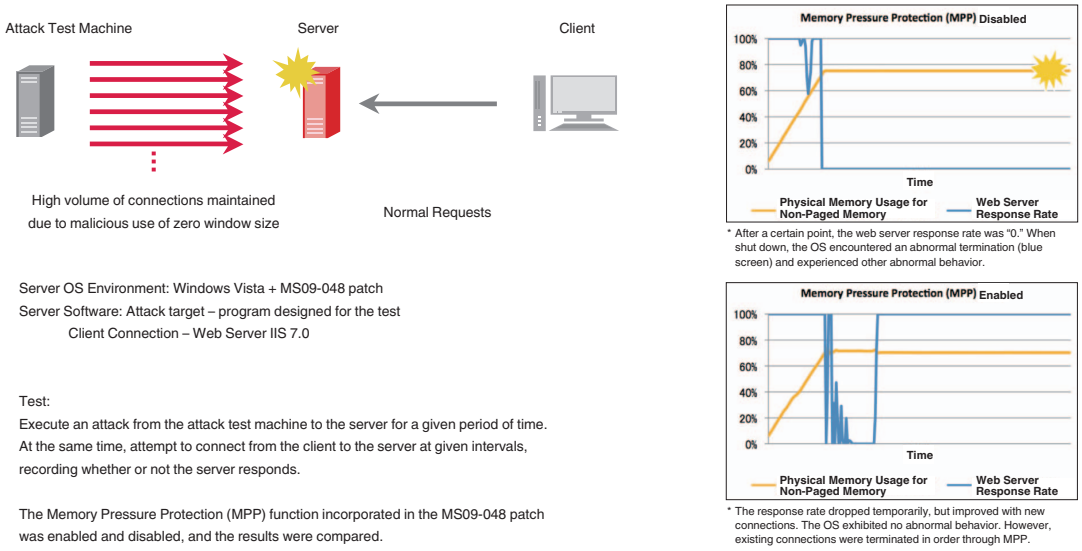


Figure 12: Demonstration Test and Results

*53 IETF is the acronym for the Internet Engineering Task Force. An organization that develops the standardization of Internet technologies. (http://www.ietf.org/)

*54 TCPM is a short name for TCP Maintenance and Minor Extensions Working Group (http://www.ietf.org/dyn/wg/charter/tcpm-charter.html).

*55 Discussion launched on the TCPM WG mailing list (http://www.ietf.org/mail-archive/web/tcpm/current/msg02189.html) and subsequently submitted Internet Draft "Clarification of sender behavior in persist condition" (https://datatracker.ietf.org/drafts/draft-ananth-tcpm-persist/), etc.

*56 In fact the Microsoft Web Server, Internet Information Service (IIS) includes functions to adjust web response according to OS load status. Simply performing the attack in question on the Web Service did not result in abnormal OS behavior. Accordingly, this experiment used this attack to place loads on other services.

As we have shown, this issue has been treated as an implementation vulnerability. However, because one cannot tell an attack connection from a normal connection from TCP standards, one has to conclude that this is an issue of system resource management for systems that receive a large volume of connections. There might be other similar issues that cannot be fundamentally resolved through implementation modifications*57, and it is likely that more will be uncovered in the future. Accordingly, servers open to the Internet must be carefully operated continuously.

1.4.3 Randomly Arriving SIP Packets

■ Fraudulent SIP Communications

From last year, IIJ has continued to observe SIP (Session Initiation Protocol)*58 packets intermittently arriving at honeypots. These SIP packets were sent to large numbers of IP addresses across the Internet, attempting to connect to terminals that could interpret SIP. Depending on the configuration, certain VoIP routers and IP telephones play a ring alert merely at the arrival of these SIP packets. This was the underlying cause behind the large number silent call reports*59.

■ SIP-Based VoIP Communication Mechanism

As indicated by the name, SIP is one protocol used to control a session, based on a request-and-response model similar to HTTP. SIP is used in IP phone services and other VoIP communications. However, while the HTTP specification defines data transmission as well, SIP only controls the initiation, modification, and termination of a session between VoIP terminals, leaving data transmission to other protocols such as Real-time Transport Protocol (RTP)*60 for audio, etc. Figure 13 shows an example of SIP communications via IP telephone.

(1) When a call is initiated via IP telephone, the origin of the call (User Agent:UA) first sends an INVITE message to the call receiver.
(2) When the call receiver gets the INVITE message, a ring tone is sounded to notify the called user. At the same time, “180Ringing” (meaning that a call is in progress) is returned to the originator of the INVITE message.
(3) When the call target picks up the receiver, the receiving terminal sends a 200 OK message to the originator of the call.
(4) Having received this message, the originator of the call sends an ACK response to the call receiver, and the session is established.

This is the basic operation. In general, SIP servers are used when connecting, rather than directly connecting the two terminals*61.

■ Attacks Targeting IP-PBX

To control adoption and maintenance costs, more corporations today are replacing existing PBX*62 systems with IP phone systems using IP-PBX*63. With the availability of low-cost IP-PBX appliances, it is likely that this trend will continue to gather momentum.

While IP-PBX does present significant adoption benefits in terms of cost savings, different than existing telephone networks, IP-PBX is connected to a network that features other connections with non-VoIP equipment, including the

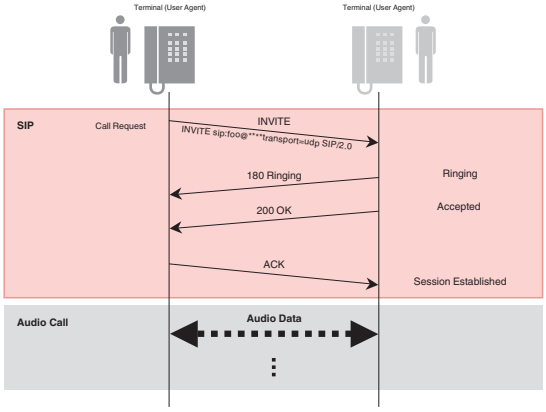


Figure 13: Initiating SIP-Based Audio Communications

*57 For example, a survey report on TCP robustness by CPNI of United Kingdom (https://www.cpni.gov.uk/Docs/tn-03-09-security-assessment-TCP.pdf) and a summary survey report of existing TCP/IP vulnerabilities by the IPA (http://www.ipa.go.jp/security/vuln/vuln_TCPIP.html)(in Japanese). These reports publicized several of these issues, providing developers with commentary on important points related to TCP/IP protocol stack implementations. The latter report also includes a guide for system operators.

*58 We also addressed random SIP packets in Vol.4 of this report (http://www.iij.ad.jp/en/development/iir/pdf/iir_vol04_EN.pdf).

*59 For example, a cNotes article reported this incident. “INVITE Flood? Fraudulent SIP Calls” (http://jvnrss.ise.chuo-u.ac.jp/csn/index.cgi?p=INVITE+Flood%3F++%C9%D4%C0%B5% A4%CASIP%C3%E5%BF%AE)(in Japanese).

*60 Real-time Transport Protocol is a data transfer protocol for transmitting a data stream in real time. Used for audio and video transmission, most VoIP equipment supports RTP.

*61 An SIP server includes proxy, redirect, and register functions; normally, communications with another party are handled via SIP servers.

*62 A PBX (Private Branch eXchange) is a telephone exchange local to a particular office. It connects internal and external (public switched telephone network) lines and controls incoming and outgoing calls.

*63 A PBX that includes VoIP functions. Asterisk (http://www.asterisk.org/) is one example.

Internet, and closed IP networks. Because of this, one must consider the fact that the IP-PBX is more susceptible to external and internal attacks than a traditional PBX. Most of today's VoIP products utilize SIP on UDP; accordingly, SIP packets with false IP addresses and caller phone numbers can be easily created. In fact, there have been cases overseas of hackers exploiting vulnerabilities to fraudulently operate an IP-PBX, capturing IP telephone service contract information in an attempt at fraudulent use*64. Other cases have involved numerous phone calls with false caller numbers, attempting to have the callees call back to premium toll numbers, thus fraudulently building up charges and stealing money*65. The random SIP packets observed by IIJ were not for the purpose of causing silent calls, but more likely attempts to find IP-PBXs having an exploitable vulnerability.

■ VoIP Security Countermeasures

It is vital to always operate equipment correctly and securely. For example, gain a correct understanding of the types of threats involved*66, receive periodic updates regarding recommended settings and product information from your VoIP equipment vendor, confirm important issues related to service usage with your ISP, etc. If possible, configure VoIP equipment with encryption functions, enable settings to only accept SIP packets from certain SIP servers, and incorporate other appropriate access controls via functions and settings to prevent SIP messages from unknown sources. Adopting VoIP-compatible firewalls, IDS and IPS, as well as a session border controller*67 is also an effective preventive measure.

We believe that the growth of VoIP will continue, along with increasingly widespread adoption of personal-use IP phones and corporate IP-PBXs. On a traditional phone, users would hesitate to answer a call coming in from a complete stranger. This same caution should apply to both traditional and newly emerging threats involved in VoIP communications.

1.5 Conclusion

This whitepaper has provided a summary of security incidents to which IIJ has responded. In this volume, we have included security incidents in which IIJ was not directly involved, mainly addressing the DDoS attacks in the United States and South Korea. We believe that our mission encompasses the collection and analysis of information stemming from incidents that occur in other countries to be better able to rapidly respond should a similar incident occur in Japan in the future.

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.

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 DDoS Attacks in the United States and South Korea)
Tadaaki Nagao, Yuji Suga (1.4.2 TCP Vulnerability (Sockstress))
Hirohide Tsuchiya (1.4.3 Randomly Arriving SIP Packets)
Division of Emergency Response and Clearinghouse for Security Information, IIJ Service Business Department

Contributors:
Shigeki Ohara System Development Section, System Infrastructure Division, IIJ Service Business Department
Masahiko Kato, Masafumi Negishi Division of Emergency Response and Clearinghouse for Security Information, IIJ Service Business Department

*64 For example, an alert issued by the Internet Crime Complaint Center (IC3: an organization involved in combating cyber crime) in cooperation with U.S. law-enforcement agencies (FBI, etc.) (http://www.ic3.gov/media/2008/081205-2.aspx).

*65 The F-Secure blog post "Beware of One-Ring Fraud," (http://www.f-secure.com/weblog/archives/00001744.html) for example.

*66 See the following report as one source of information regarding known vulnerabilities and threats. "Survey Report regarding Known SIP Vulnerabilities Version 2.0" by the IPA. (http://www.ipa.go.jp/security/vuln/vuln_SIP.html)(in Japanese).

*67 A device installed at the boundary of the VoIP network. Responsible for controlling necessary ports according to SIP packet content, and for controlling functions allowing for normal VoIP communications, even in an NAT environment.

How Do We Discourage Asia from Continuing to be a Source of Spam?

In this report, we will offer our analysis of trends in the ratio of spam and regional sources of spam detected over the period between weeks 27 and 39 of 2009. At the same time, we will also comment on the results of international activities designed to reduce or prevent spam, as well as the adoption rates of sender-authentication technologies.

2.1 Introduction

In this report, we have summarized the latest developments with respect to spam, information about anti-spam technologies, and other activities in which IIJ is deeply engaged. To analyze spam trends, we conducted an analysis of spam from a number of different angles, 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. Our survey covered the 13 weeks between week 27 (June 29 through July 5, 2009) and week 39 (September 21 through September 27, 2009)—a total of 91 days. Our prior report (Vol. 4) marked the first year since we began publishing the IIR. We will thus also provide a summary of certain data collected over that span of time.

We will also cover international activities related to the adoption of anti-spam measures. The IIJ has been active in international discussions regarding spam, mainly through the Messaging Anti-Abuse Working Group (MAAWG). We will provide an overview of other organizations and activities herein. We will also take another look at sender-authentication technologies, discussing to what extent this and other anti-spam technologies have actually been adopted.

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. Figure 1 shows the last year of data (69 weeks), including the weeks that are the focus of our survey—weeks 27 through 39, 2009.

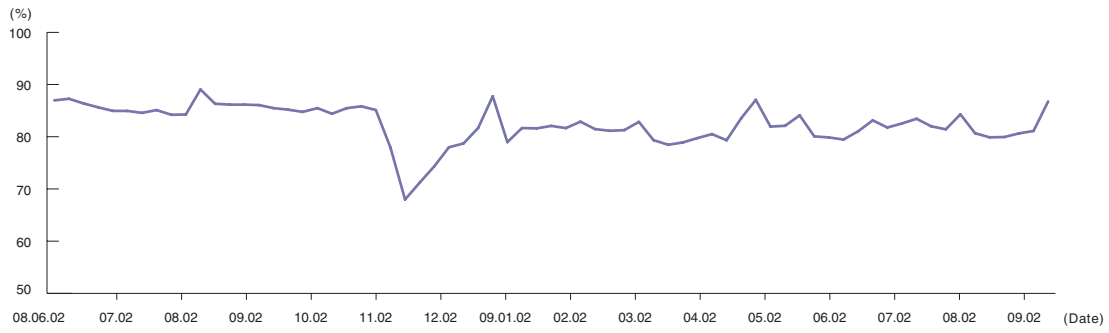


Figure 1: Spam Ratio Trends

2.2.1 Spam Ratio Trends

The ratio of spam averaged 82.2% of all incoming emails over the 91-day period from week 27 to week 39, 2009. This compares to an 81.6% average in our last survey (weeks 14 through 26, 2009), indicating a slight comparative increase. Week 39 (September 21 through September 27, 2009) had the highest ratio of spam at 86.7%. This week happened to include an extended national holiday, and the associated lower levels of overall email activity likely contributed to this bump in spam ratios. Given our experience to date, we expect spam to increase during autumn. Accordingly, we will be keeping a close watch on spam during the upcoming months.

We have noted a recent increase in spam that sends malicious programs as attachments. New variants of these programs are released at a very high rate. In some instances, anti-virus software has not been able to keep up with the changes. Users must take greater care before uncompressing or executing file attachments.

2.2.2 Sources of Spam

Figure 2 shows our analysis of regional sources of spam over the period studied.

Brazil (BR) remained at the top of the list as the number one source of spam, accounting for 12.1% of the total. Brazil actually increased its overall share of this dubious honor by 0.3%. In our opinion, there is a need to investigate the underlying causes that have made this region a major source of spam, as well as what measures can be introduced to reduce spam originating there. We noted a slight variance in the ranking of other top regions during the period investigated. The United States (US) continues to remain near the top, but fell in rank to No. 4 at 7.1%, down 4.3% since our last survey. As we reported earlier, the activities of the U.S. Federal Trade Commission (FTC) and other consumer protection agencies engaged in enforcement may be responsible for this positive trend.

Both China (CN) at 9.8% and Korea (KR) at 7.2% remain near the top of the list at No. 2 and No. 3, respectively. Combined with Vietnam (VN) at No. 6 (5.6%), these countries are the main reason that Asia has experienced growth as a source of spam. Vietnam rose rapidly up the ranks from No. 14 in our previous survey, calling for continued close monitoring. Both Vietnam and Brazil are representative of a global trend—a stronger network infrastructure due to economic growth also sets the stage to become a growing source of spam. India (IN) has also demonstrated this trend, ranking No. 5 in our survey at 6.0%, nearly unchanged since our last survey.

Japan ranked No. 9 in this survey at 3.1%, but increased in ranking and percentage as a source of spam over the previous survey. We can think of several reasons for this. One factor that stood out was the volume of so-called error emails. Among those emails originating in Japan determined to be spam, many originated from major Japanese Internet Service Providers. Looking at the transmission source (envelope From) information and transmission period of these emails, we see that the majority were error notifications in connection with mail sent to an unknown recipient. In other words, the majority of emails identified as spam coming from Japan were bounces.

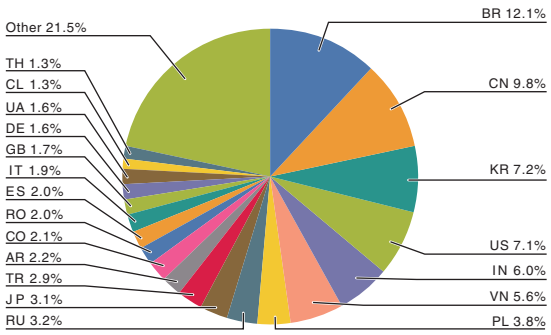


Figure 2: Regional Sources of Spam

This phenomenon occurs when IIJ customer domain names are appropriated illegally as source information for spam. In general, this type of malicious usage is difficult to prevent on the part of the domain administrator. One countermeasure is the adoption of sender- authentication technologies on the part of the sender, as offered by IIJ. Performing sender authentication on the part of the receiver as well is necessary to determine whether the domain indicated in the sender information represents the true gateway for the email being sent. In other words, this type of spam can be eliminated by introducing technology that does not return a bounce message in connection with email that uses an unauthorized domain name in the sender information.

Figure 3 shows trends in the ratio of spam from the top six countries (Brazil, United States, China, Korea, Vietnam, India) and Japan.

At the outset of the period surveyed, Vietnam was ranked lower than Japan as a source of spam. However, the Asian nation rose quickly through the ranks, reaching as high as No. 4 and No. 5 at certain points since August. These bumps pushed Vietnam higher overall, ranking No. 6 out of the countries surveyed at the end of the period studied.

From Figure 3, we see the seasonal differences in spam volume over time for each regional source. However, the variances are not necessarily consistent in each country. For example, Korea (KR) was the greatest source of spam in week 34 (the week beginning August 17, 2009), exceeding 10%. The increase in total spam volume during this week was due to a large volume of mail sent from the same source (IP address). While Brazil (BR) experienced some fluctuations in volume weekly, looking solely at the spam coming from that country, we do not see a great variance in the ratio of large-volume spammers or in those who send just one spam email in a week. Based on the patterns identified here, we see that a single source of a large volume of spam had a significant impact on the overall volume of spam sent from Korea to Japan. Brazil and other regions have a small number of high-volume spam sources, leading us to conclude that much of the spam volume is attributable to computers infected with bots or similar malicious programs.

While we can expect legal measures in different regions to have some positive effect on large-volume spam sources, there are limits to what can be accomplished legally where there are significant numbers of small-volume sources. We see here that each region and source of spam must be treated according to the individual nature of the region and source.

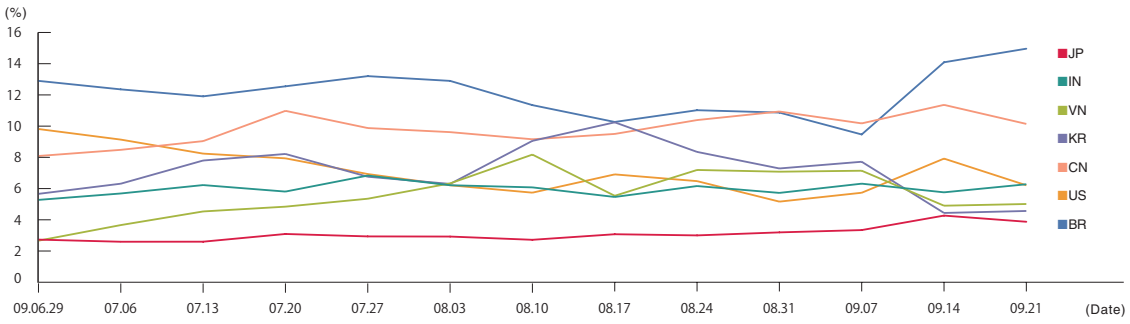


Figure 3: Trends in Sources of Spam

2.2.3 International Anti-Spam Activities

As we have clearly demonstrated in our analysis, the majority of spam sent to Japan originates outside the country. Asia, in particular, has the dishonor of being identified by security vendors as being a source of global spam.*1

The Asia Pacific Coalition against Unsolicited Commercial Email (APCAUCE) is a private organization that addresses spam in Asia. The APCAUCE is the Asian arm of the Coalition against Unsolicited Commercial Email (CAUCE), which is based in North America. This author participated in the 2006 and 2007 APCAUCE meetings, presenting anti-spam initiatives conducted in Japan. Unfortunately, the APCAUCE has not met since the 2007 symposium in India. Accordingly, several Asia-based organizations featuring the "Asia Pacific" moniker in their names have come together to sponsor the AP*Retreat*2 (an information exchange symposium), held at the same time as the APNIC 28 meeting in Beijing. The APCAUCE, however, did not participate. The issue of spam is an important topic, and the cooperation of the co-chair of the AP*Retreat allowed the opportunity for presentations regarding the state of anti-spam initiatives in Japan and China.

These types of volunteer-based organizations tend to rely heavily on the efforts of a single organizer, which is perhaps why such meetings occur only irregularly. With anti-spam measures in Asia representing such an important issue, the Internet Association Japan has stepped forward to take the lead as the Japanese representative, moving forward with plans to hold a meeting some time next year.

MAAWG is a private global organization that we have discussed in the Internet Topics section of our prior issue (Vol. 4). Presently, participants mainly come from Europe, the United States, and the author's home country of Japan. Asia does not provide many participants. With MAAWG meetings generally held in North America or Europe, it is difficult for professionals in other regions to participate. Accordingly, IIJ believes in the growing need to hold MAAWG meetings in Asia.

The London Action Plan (LAP)*3 is an example of a global organization that works in cooperation with government agencies. LAP sponsors annual meetings according to an action plan ratified in 2004 for the purpose of promoting the sharing of information, organizational coordination, and public-private interaction among enforcement authorities. LAP, MAAWG and the Contact Network of Spam Enforcement Authorities (CNSA) held a joint meeting in 2007. The fifth joint CNSA-LAP Workshop was held in Lisbon, Portugal from October 7 to October 10, 2009. The Japanese Ministry of Internal Affairs and Communications is a member of this organization, and participated in the October meeting, along with this author and the Japan Data Communications Association, presenting Japanese initiatives.*4

The LAP symposium included discussions of anti-spam laws and enforcement in different countries, as well as other anti-spam activities. Representatives from private organizations (including this author) shared information, as did the German "eco" ISP group and representatives from MAAWG, who discussed their activities in this field. Other presentations discussed how best to engage in cooperative activities with law enforcement agencies in the future. The FTC of the United States made a presentation about the circumstances surrounding the Pricewert shutdown, which we discussed in a previous issue of this report.

*1 For example, Asia is responsible for one-third of all spam according to a continent-by-continent survey conducted by Sophos. (http://www.sophos.com/pressoffice/news/articles/2009/07/dirtydozenq209.html)

*2 See http://www.apstar.org for an overview of the AP*Retreat symposium.

*3 LAP: London Action Plan (http://www.londonactionplan.org)

*4 Workshop CNSA-LAP "Spam-Fighting" (http://www.anacom.pt/render.jsp?contentId=962326)

Australia, New Zealand and several other countries in the Asia Pacific region actively participate in the LAP. Participation by government agencies throughout Asia is relatively high, with participation with from Hong Kong, Taiwan and Malaysia this year. The Seoul - Melbourne Multilateral Memorandum of Understanding on Cooperation in Countering Spam*5 was signed in 2005, representing a collection of government agencies in the Asia Pacific region. Symposia have been held in different member countries. The Ministry of Internal Affairs and Communications represents Japan, with a symposium held in Tokyo in March 2008.

These types of organizations offer government agencies opportunities to share information. While consistent enforcement of the law is of course important, technological measures to fight highly advanced spam technologies, such as botnets, are also important. The public and private sectors must work effectively in concert, cooperating in both technological and legal enforcement in the pursuit of new anti-spam measures. Sending spam from Japan has become more difficult with the introduction of technological tools, such as OP25B (Outbound Port 25 Blocking). This means that spammers are more likely to set up bases overseas, particularly farther east in Asia. Cooperation among and between government agencies in Asia is an important step in doing away with large-scale spam.

2.3 Trends in Email Technologies

2.3.1 Adoption of Sender-Authentication Technologies

We have discussed sender-authentication technologies such as DKIM which uses digital signature technology and network-based SPF/SenderID. We have also offered the results of a WIDE project survey*6 regarding adoption of such technologies, particularly the adoption of SPF on the part of the sender.

Here, we will discuss the results of our survey regarding the degree of sender-authentication technologies for emails actually released. IIJ adopted sender authentication technology*7 in 2005, additionally introducing SPF and DKIM in our continuing pursuit of email safety and security.

Figure 4 shows the ratio of authentication results when email is received in connection with some of the IIJ emails services.

The period subject to this survey was the entire month of September 2009. Of the emails received during this period, 56.2% indicated “none” as the result of SPF authentication for the domain name indicated in the sender information. This means that the domain for 43.8% of email received declared an SPF record. According to a WIDE survey, the declaration rate for “jp” domains in October 2009 was 36.8%. Our results showed a greater result under a basis of actual volume. Of course, sender

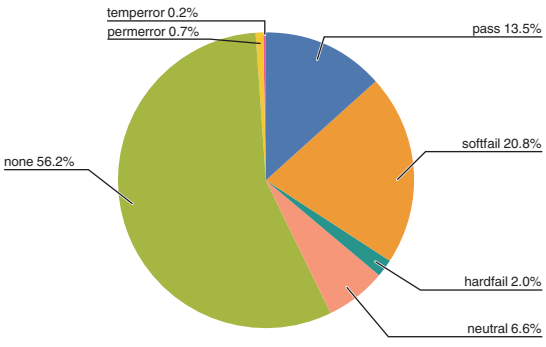


Figure 4: Authentication Results (SPF) for Email Received

*5 Seoul-Melbourne Multilateral Memorandum of Understanding on Cooperation in Countering Spam (http://www.sm-mou.org/

*6 Survey Results of the Deployment Ratio of Sender Authentication Technologies published by WIDE (http://member.wide.ad.jp/wg/antispam/stats/index.html.ja)

*7 IIJ Introduces Sender Authentication Technology (http://www.iij.ad.jp/news/pressrelease/2005/0317.html)

domains for email received include other top level domains (TLD) such as “com” and “net” in addition to “jp.” We believe that the results from analyzing emails received in Japan indicate a comparatively advanced usage of SPF.

The ratio of email receiving a “pass” as the result of SPF authentication was 13.5%, while the ratio of authentication failure (“softfail” and “hardfail”) was 29.4%. From this result, we can conclude that the volume of email that falsely represents sender information was greater than the volume of email sent from a legitimate mail server. Where the volume of spam exceeds 80% of all email received, this is a result with which we can be relatively satisfied. SPF requires care in management, as it is susceptible to forwarding problems and other issues related to false positives. But considering the declaration rates and usage of SPF, a certain level of screening through SPF authentication is one method that can be very effective.

Next, let’s look at Figure 5, which addresses DKIM authentication results.

As with our SPF survey, the period studied was the entire month of September 2009. Of all emails received, 0.8% were from senders using some type of DKIM. Given the comparative expense of DKIM for the sender compared to SPF, we see that DKIM has not been embraced on a large scale. However, DKIM offers comparatively fewer false positives than network-based SPF/SenderID, which is why we expect usage to increase in the future for important email messages. Where both SPF and DKIM require the incorporation of additional functions, the cost of authentication on the part of the receiver is the same. Accordingly, we believe that it is important to promote DKIM incorporation in parallel with SPF for authenticating incoming messages.

2.4 Conclusion

Here, we have reported on the ratio of email determined to be spam and the regional distribution of sources of spam under the heading of messaging technology. We have also discussed significant global and Asia Pacific (including Japan) international cooperatives. These activities serve as an important anti-spam measure, as private and government agencies share information through various symposia and engage in cooperative initiatives.

IIJ has presented the data and technologies published through the IIR at international conferences, promoting the adoption of activities and measures in different locales. We will continue to be active in the promotion of anti-spam measures in a variety of different settings.

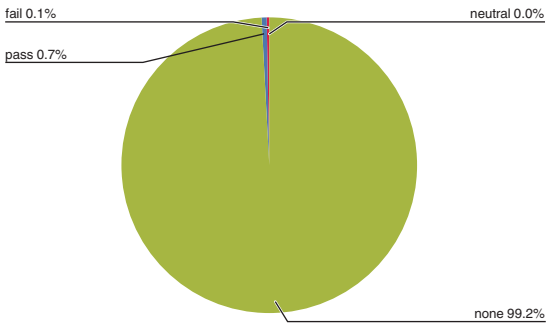


Figure 5: Authentication Results (DKIM) for Email Received

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. He is also a member of Internet Association Japan’s Anti-Spam Measures Committee.

Live Migration of Guest Computers Using the NEMO BS Mobility Function

In order to construct an efficient virtual computing environment, a system for ensuring the availability of virtual computers and managing them flexibly is required. In this whitepaper a virtual computer migration method using NEMO BS technology, which adds a mobility function to IPv6 routers, is discussed and the results of experiments using this setup are examined. NEMO BS makes it possible to migrate virtual computers beyond the network segment.

3.1 Background

The capability of individual computers is progressing rapidly as a result of the processing power of computers and technological advances in networks and data stores. Meanwhile, there is also the view that speed improvements for individual computers will slow down in the future. In recent years cloud technology that treats multiple computers as a single computer resource is gaining a lot of attention[*1][*2] as a remedy for this issue. On the other hand, the virtual computing technology that allows a single computer resource to be partitioned virtually and used as multiple different computers[*3] has been studied for many years. These technologies may appear at a glance to be targeted at different uses, but it is believed that they can supplement each other and make it possible to utilize computer resources more efficiently. For example, Amazon's EC2[*4] service provides a cloud environment that combines multiple computer resources into a single service, but virtual computer technology is also used in each individual computer resource. Using units broken down into small virtual computers instead of using large resource units consisting of individual physical computers enables the efficient and flexible use of computer resources, and consequently provides convenience in excess of the overhead created by partitioning into virtual computers.

A system for managing virtual computers flexibly is crucial for developing this kind of virtual computing environment. This is due to the fact that the ability to arrange the necessary amount of virtual computers in the necessary location in accordance with the requirements of a cloud service contributes to overall performance increases and the efficient use of resources. In this whitepaper, systems for relocating virtual computers are focused on, and a technique for migrating running virtual computers to a different segment (offlink segment) is proposed.

3.2 Live Migration Issues

Currently, a large number of virtual computer technologies are provided at a practical level. Some of these provide functions for migrating virtual computers from their parent computer (*host computer*) to another host computer. VMware's VMotion and Xen's[*5] Live Migration are examples of this kind of technology. In this whitepaper, virtual computers that are generated in the host computer are called *guest computers*, and the function for migrating running guest computers between host computers is called *live migration*. The use of a live migration function makes it possible to migrate running guest computers to another machine with very little downtime. However, the functions that are currently provided are restricted to situations in which the source host computer and destination host computer belong to the same segment.

This restriction is due to the network configuration method provided to guest computers. Host computers and guest computers do not have an equal relationship. Host computers normally have control over all resources, and allocate part of these resources to guest computers. When connecting a guest computer to a network, a configuration like the one shown in Figure 1 is used.

*1 Aaron Weiss. Computing in the clouds. netWorker, Vol. 11, No. 4, pp. 16-25, December 2007.
*2 Brian Hayes. Cloud computing. *Communications of the ACM*, Vol. 51, No. 7, pp. 9-11, July 2008.
*3 Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, et. al. Xen and the Art of Virtualization. In *SOSP '03: Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles*, pp. 164-177, ACM, 2003.
*4 Amazon. Amazon Elastic Compute Cloud (Amazon EC2), October 2009. http://aws.amazon.com/ec2/
*5 Citrix, October 2009. http://www.xen.org/

Using configuration (a) in Figure 1, the guest computers are connected to the same network as the host computer via a virtual switch provided by the host computer. Similarly, a virtual switch is also created between the host computer and the guest computers when using configuration (b). Unlike configuration (a), under this configuration the host computer also functions as an upstream router for the guest computers. As Figure 1 clearly shows, the network configuration for the guest computers is highly dependent on the host computer. When executing live migration, no changes are made to the operating environment of the guest computer itself. Consequently, it is only possible to carry out live migration using configuration (a). Under configuration (b), the address allocated to the virtual switch differs depending on the location where the host computer is attached. This means that after a guest computer is migrated, communications cannot be continued until the network environment of the guest computer is updated appropriately. The same problem also occurs under configuration (a) if the source and destination host computers for migration are connected to different segments.

Live migration makes it possible to avoid consolidating virtual computers on a single host computer, but this function is limited to migration within the same segment. Even if there is a host computer on another segment with free resources, it cannot be utilized. It is also not possible to relocate resources to improve performance, such as migrating a guest computer to a host computer that is closer to the user.

3.3 Overview of NEMO BS

NEMO BS (Network Mobility Basic Support)*6 is a protocol for adding a mobility function to IPv6 routers. In a NEMO BS environment, a *MR* (*mobile router*) compatible with NEMO BS manages the *MNP* (*mobile network prefixes*) that serve as fixed network prefixes. IPv6 hosts connected to the network provided by the MR use fixed addresses in the MNP range. The MR connects to various segments over the Internet, and maintains connectivity with the Internet according to the target network environment. During this the MNP the MR manages does not change. Hosts attached to the network controlled by the MR are always able to maintain the same network environment regardless of the location of the MR.

This function is achieved with the support of the *HA* (*home agent*) that serves as a counterpart to the MR (Figure 2). The MR acquires an address (care-of address) on the destination network depending on the network environment, and establishes a bidirectional IPv6 over IPv6 tunnel with the HA. Traffic originating from nodes within the MNP is sent to the HA using this tunnel, then transmitted from the HA to the target of communications. Meanwhile, traffic directed to nodes within the MNP is received by the HA, delivered to the MR via the tunnel, then transmitted to the final destination host.

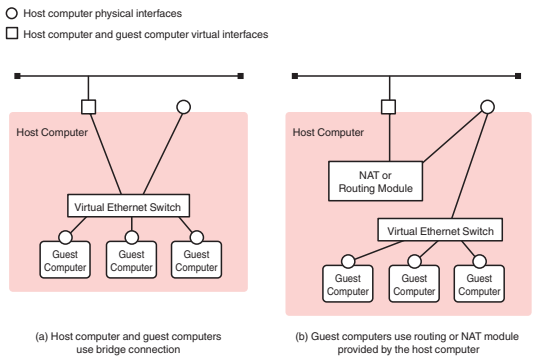


Figure 1: Guest Computer Network Configurations

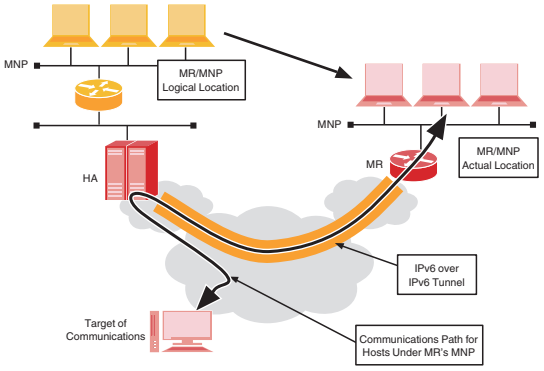


Figure 2: Operational Overview of NEMO BS

*6 Vijay Devarapalli, Ryuji Wakikawa, Alexandru Petrescu, and Pascal Thubert. Network Mobility (NEMO) Basic Support Protocol. IETF, January 2005. RFC3963

3.4 Design

Live migration is limited to migration within the same segment because the network environment for guest computers is dependent on the host computer. In other words, if the guest computer's network environment is designed so that it is fixed and does not depend on the host computer, it will be possible to migrate guest computers to host computers connected to different segments.

In this whitepaper, a method for using IP mobility technology to maintain a fixed network environment on guest computers is proposed. There are two approaches to this: equipping the guest computers with a host mobility function such as Mobile IP[*7*8], and utilizing a system such as NEMO BS on the host computer to provide a fixed network for guest computers. The former option requires the modification to guest computers (the installation of an IP mobility function), but this makes the migration of individual guest computers possible, allowing for finer control over computer resources. The latter option allows existing guest computers to be used without any modifications, but requires that the migration of guest computers and the migration of the host computer serving as the MR for NEMO BS be synchronized. In this whitepaper the latter option is focused on, assuming a scenario where the guest computers used in an existing system will continue to be utilized.

The system design may differ depending on the resource management method for host computers. In this whitepaper a design using the virtual computing environment provided by Xen is discussed, but it should be possible to apply a similar design to other systems without significant changes. Figure 3 shows an overview of this design. The configuration used is an expanded version of (b) from Figure 1. The host computer functions as the MR, providing two interfaces, one is for connecting to the Internet and the other is for providing MNP. The MNP connection interface is a virtual interface rather than a physical interface. This virtual interface is connected to the virtual switch that the host computer provides for guest computers. The host computer's virtual interface and the interfaces for guest computers connected to the virtual switch are allocated fixed addresses managed by the MR. Through the NEMO BS function, addresses within MNP do not change regardless of where the host computer is physically connected.

In the proposed environment for live migration of guest computers, multiple host computers are located on the network. These host computers act as MR and are configured with the same MNP, but host computers that have no active guest computers do not function as an MR. When migrating guest computers, regular live migration procedures are first performed to migrate guest computers to the destination host computer. At this point, guest computers are still disconnected from the network. Following this, the source host computer sends a migration complete notification to the destination host computer, and suspends its NEMO BS function. The host computer that receives the migration complete notification registers its location with the HA, and begins operation as a NEMO BS MR. Once registration with the HA is complete, the MNP for the virtual switch that the guest computers are connected to is activated, and migration is complete.

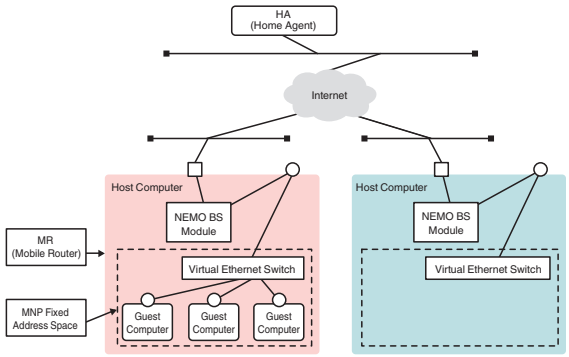


Figure 3: Guest Computer Migration Using NEMO BS

*7 Basavaraj Patil, Phil Roberts, and Charles E. Perkins. IP Mobility Support for IPv4. IETF, August 2002. RFC3344

*8 David B. Johnson, Charles E. Perkins, and Jari Arkko. Mobility Support in IPv6. IETF, June 2004. RFC3775.

3.5 Verification Tests

To confirm that the proposed design is feasible, a prototype system was implemented and operation experiments were performed. The experiment environment was configured with a total of five computers: two Xen host computers operating as MR, a computer with HA function, a computer for receiving streaming data during testing, and a control computer for initiating migration operations. A guest computer was created with the Xen host computers and configured to operate as a streaming server.

In order to carry out experiments in an environment close to real conditions, this equipment was placed into an actual Internet environment. The fixed network for the HA and MR to use was installed as a part of the network constructed for Interop Tokyo 2009*9, and the Xen host computers and stream data reception nodes were placed on the IIJ network. Each piece of equipment was connected to the others via the Internet. Figure 4 shows an overview of the experiment network.

15 MB, 520 KB/s MPEG4 stream data was continually sent from the guest computer (the streaming server) using UDP. The instructions to begin live migration of the guest computer, and the suspension and initiation of the host computer NEMO BS function were carried out over SSH using a separate control computer. The guest computer live migration command was executed remotely from the control computer, and host computer NEMO BS movement processing was executed immediately after live migration was complete. The guest computer was repeatedly migrated between the two host computers every five minutes.

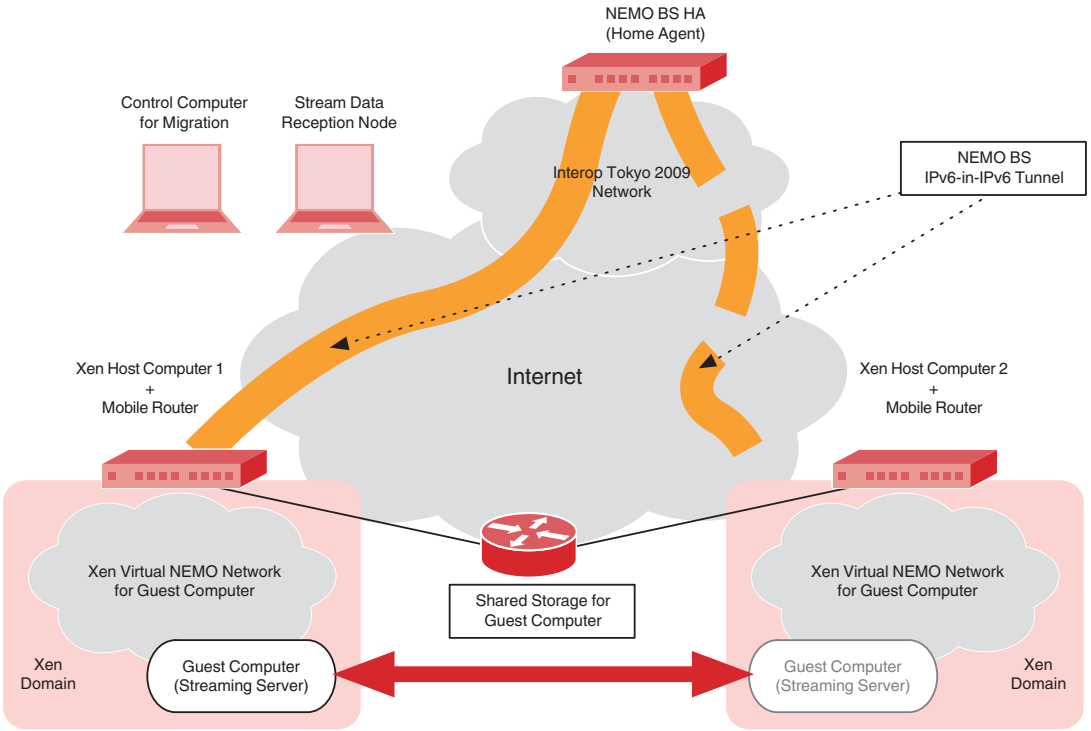


Figure 4: Experiment Network Overview

*9 Interop Tokyo 2009, June 2009. http://www.interop.jp/

3.6 Evaluation and Issues

Figure 5 shows the results of monitored stream traffic over the host computer's Internet interface. Live migration of the guest computer was carried out every five minutes, and migration between the two host computers can be seen as a movement of the traffic between two host computers.

When using live migration, the source virtual computer continues operation while virtual computer data is being copied to the destination host computer. The downtime when operation is switched over is quite small (about several hundred milliseconds). However, as network layer mobility technology was used to migrate the host computer in this proposed mechanism, node movement processing was also necessary for the host computer after the guest computer was migrated. When examining the traffic data details, it became apparent that it was taking approximately six seconds for the destination host computer to begin sending streaming data after the source host computer had suspended the distribution of streaming data.

As router advertisement messages on the network where the host computers connect were made at three to four second intervals, the host computers were in an environment where they could receive router advertisement messages in about two seconds. Even when the one second care-of address duplicate confirmation that is a part of NEMO BS processing is taken into account, movement processing should be completed in approximately three seconds in this environment. The following two points are possible reasons for processing taking longer than this in the experiment.

1. Overhead due to the execution of the remote control script from the control computer.
2. Procedures that differ from regular movement processing procedures.

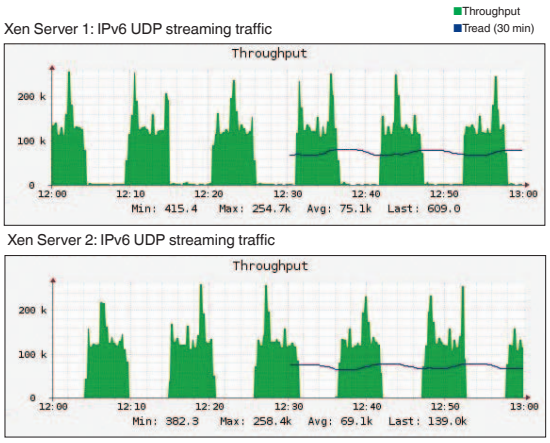


Figure 5: Stream Traffic Transition

As the suspension and initiation of the NEMO BS function were executed using a remote script via SSH, more time than usual was required for actions such as TCP connection establishment, process generation on the host computer, and controlling the NEMO BS daemon program via remote command. Additionally, movement processing that would normally be executed on a single mobile node was jointly executed on two different mobile nodes in the proposed mechanism. In other words, one MR executed registration processing after the other MR finished cancelling registration with the HA. When carrying out movement processing using the same MR, there is no need to cancel registration of old information when registering new information, so updates are made consecutively. In the proposed mechanism extra time was required for canceling registration.

3.7 Consideration

The prototype implementation and experiments aimed at verifying the combination of two technologies (NEMO BS and Xen virtualization). This section mentions some observations found during the experiments.

3.7.1 Network Storage Issues

There are two methods of providing storage for guest computers under systems that offer a virtual computing environment, one is providing part of the host computer's storage, and the other is providing network storage. Storage must be located on the network when using live migration, as the host computer that accommodates guest computers will change. As guest computers are not migrated beyond the same segment under current assumed usage patterns, there are no significant changes to the environment before or after migration even when network storage is used. However, when migrating guest computers beyond their segment as in the proposed mechanism, there is the possibility that network storage reachability and communication delay issues may occur.

One solution to this is constructing a guest computer environment that carries out all processing in memory without using external storage. However, under this method increases in the amount of data to be stored in memory will cause migration to take longer to complete. This construction also makes it difficult to handle large amounts of data.

Another method is to provide storage that can be handled efficiently and transparently from multiple locations on the Internet. For example, developing storage mirroring technology and making it possible to access the same storage data from multiple locations should ease the network quality changes to guest computers and also allow for improvements to be made to the fault-tolerance of storage.

3.7.2 Usability Issues

In the proposed mechanism the NEMO BS technology is used, but there are two issues with this method.

First, there is the fact that tunnel communications via HA is a prerequisite, although the same is true for all Mobile IP-based mobility technologies in general, and not just NEMO BS. This creates new issues to be considered, such as the HA becoming a single point of failure, and the need to introduce redundancy technology into the HA. It may be possible to resolve these issues by adopting different mobility technology that does not use tunneling, such as HIP*10, LIN6*11, or MAT*12. To achieve the offlink live migration of virtual computers, the primary requirement is to maintain the network environment of the virtual switch that virtual computers connect to. Consequently, the method used to achieve this does not need to be NEMO BS.

The second issue is that by using NEMO BS, all guest computers are managed as a cluster of nodes attached to the same fixed network. This means that when a host computer is migrated, all related guest computers must also be migrated at the same time. This issue can be dealt with using one of the following methods. One method is to have guest computers adopt a host-based mobility technology, for example Mobile IP. In this case all guest computers must support Mobile IP, increasing the requirements for deployment. NEMO BS has the large advantage of it not being necessary to modify guest computers, so standard guest computers can be used as-is. Providing a NEMO BS environment implemented as a virtual computer for each guest computer would be a compromise for this. This would operate as a kind of Mobile IP proxy. This method would make it possible to migrate individual guest computers while avoiding the need to make any changes to them.

3.8 Conclusion

Ensuring the mobility of computer resources will be crucial in the cloud computing environments of the future. Virtual computer live migration technology is a strong candidate for ensuring resource mobility. However, using current live migration technology the migration destination is restricted to other host computers within the same segment. This is because the network environment provided to guest computers is dependent on the physical network that host computers connect to. If this restriction could be removed and migration to other remote segments were made possible, it would allow for more flexible management of computer resources. In this whitepaper a technique for migrating guest computers between multiple host computers that do not share a segment using virtualization and NEMO BS technologies is proposed. The use of NEMO BS technology to provide guest computers with a fixed network environment regardless of the network a host computer connects to makes cross-segment migration possible.

In closing, while conducting this research I received a great deal of advice regarding Linux and MIPL/NEPL configurations from Masahide Nakamura, Jean Lorchat, and Martin André. I would also like to thank Takashi Miyake, Manabu Ori, and the Interop Tokyo 2009 NOC team for their assistance in preparing the test environment.

Author:
Keiichi Shima
IIJ Innovation Institute Inc. Research Laboratory
Mr. Shima is pursuing the research and development of IP mobility technology that will be necessary for transitioning continually evolving Internet terminals to wireless systems.

*10 Robert Moskowitz, Pekka Nikander, Petri Jokela, and Thomas R. Henderson. Host Identity Protocol. IETF, April 2008. RFC5201

*11 Mitsunobu Kunishi, Masahiro Ishiyama, Keisuke Uehara, Hiroshi Esaki, and Fumio Teraoka. LIN6: A New Approach to Mobility Support in IPv6. In Wireless Personal Multimedia Communication (WPMC), November 2000.

*12 Reiji Aibara, Takahiro Fujita, Kaori Maeda, and Yoshihiro Nomura. Mobile Internet Architecture with Address Translation. Special Issue on Next Generation Mobile Communication Networks and their Applications. IPSJ Journal Vol. 43, No. 12, pp. 3889-3897, December 2002.

The Design and Implementation of the NHN Next Generation Service Infrastructure Using Virtualization Technology and Remote Data Centers

NHN (Next Host Network) is IIJ's next generation service infrastructure. When NHN is adopted, service development flow and demand forecasts for equipment upgrades at IIJ are set to change dramatically. Here, we provide an overview of NHN, its technological components, architecture and implementation.

At IIJ we have implemented a service infrastructure that provides flexibility, while keeping data center costs low through the utilization of remote data centers. We designed the "NHN" (Next Host Network) in 2008, with the aim of consolidating onsite work in one place, devising systems for the efficient use of rack space and electricity, and reducing the operating costs. It was prepared as an infrastructure for the migration of both new and existing services. In this whitepaper we explain the circumstances leading up to the introduction of NHN, its technological components, architecture and implementation.

4.1 Background to Adoption of NHN

IIJ has until now operated over 200 racks and several thousand servers distributed over multiple data centers for use in our own services. As equipment was constructed with each host reserving a rack for individual services, the rack space and network devices were designed with headroom to cope with future increases in demand, resulting in lost potential for the infrastructure as a whole. Using a system configuration such as this that is vertically segmented for each service makes it difficult to expand when plans change and also reuse equipment when a service is discontinued. Service infrastructure costs piled up, and as racks used different equipment and cabling, the complexity of operation had also become an issue.

Until now we had taken factors such as support when a failure occurs into consideration, and used a data center with a favorable location in Tokyo because of the ease of access to a physical system, aiming for optimal server operation. However, with the improvement in performance and lower costs of IA servers over the past few years, and power consumption for each server continuing to rise, facility costs such as rack space, air conditioning costs for cooling, and power costs now make up a higher percentage of the total than equipment costs such as networks, servers, and storage. Our Tokyo data center receives a lot of business from customers, and even when there was space to install equipment, we were not able to install enough servers due to limitations such as a shortage in our capacity for cooling, UPS, and emergency generators.

As we saw that a configuration with systems concentrated at the Tokyo data center was reaching its limit, IIJ decided to drastically reduce costs by moving service equipment that was not dependant on location to a low-cost facility on the outskirts of Tokyo. We considered all options that would lead to reduced facility costs and power requirements, such as container-type data centers. Advances in multiplexing technology have made it possible to keep network costs down, and this also motivated our plan to move to a data center out of central Tokyo.

While considering our options, we also moved ahead with transitioning to a configuration that ensures freedom and mobility of operations even when a remote data center is used.

Moving service equipment to a data center away from the city reduces space costs, personnel expenses, and power costs, leading to cost reductions across the entire service infrastructure. We began looking into NHN with the goal of improving operational freedom at the same time.

4.2 A Design Plan Optimized Remote Data Center

For NHN, we isolated the following points that must be given priority and points that could be put on hold, based on our experience as a server operator.

- Regarding storage, foreseeable failures such as individual HDD failures are acceptable. However, there is a need to design a configuration with high reliability to prevent failures that lead to service outages.
- Server devices sometimes fail. As there are wide-ranging points of failure, and there is a limit to how much the failure rate can be curtailed, a configuration that suffers little impact when a failure occurs is ideal.
- Failures of network devices used on the edge are not very common in our experience to date. NIC redundancy settings, etc., are only implemented when required.

Next, we will detail the failure rate of devices that IIJ uses, and explain the sequence of events that lead to the plan mentioned above.

4.2.1 Storage Failure Rates at IIJ

After investigating the logs for the approximately 100 DAS*1 that IIJ uses, we confirmed that despite the number of storage units changing very little, HDD failure rates had dropped over the past few years. Another trend in the last few years is that almost no correlation was observed between storage load and failure rates. It seems that HDD failures occur regardless of load, and it is rare for failures to occur more often in HDDs in a certain storage unit.

Year	HDD Failures	HDD Failures in the Same RAID
2005	32	7 x 1 unit, 5 x 1 unit, 4 x 1 unit, *2 2 x 2 units, 1 x 12 units
2006	22	2 x 3 units, 1 x 16 units
2007	16	2 x 1 unit, 1 x 14 units
2008	7	2 x 2 units, 1 x 3 units
2009	4	1 x 4 units (as of September 2009)

As the HDD failure rate is dropping each year, storage failures other than HDD failure, and in particular those that lead directly to service outages, have become more conspicuous. Specifically, this refers to failures such as the following.

- Suspension of function due to the failure of cache memory on a RAID controller
- Unstable operation thought to be caused by a RAID controller failure
- Failure of connection interfaces on SCSI cards, etc.
- Performance degradation due to the failure of a battery backup unit on a RAID controller or it reaching its lifespan

Failure numbers amount to a few times per year overall, but when failures such as those above occur in devices without RAID controller or connection path redundancy, it is not possible to restore operation until the device is replaced onsite, leading directly to a service outage. When considering a remote data center option, we decided that for storage service outages it would take an unacceptably long time to complete the replacement of a device. For this reason, we made RAID controller and connection path redundancy a requirement when selecting devices, although this would make hardware costs higher than at present.

4.2.2 Server Failure Rates at IIJ

Regarding server failure rates, servers are often replaced on the spot as a precaution to prioritize the restoration of service, and in some cases the symptoms cannot be recreated after the device that has failed is subsequently taken back and tested, meaning that there is no statistical information regarding the parts that have failed.

We estimate that the failure rate is approximately 1-2%, based on the number of servers that were repaired in the six months between April and September 2009, and the number of servers that were switched out and are awaiting inspection. However, this figure does not include failures in redundant parts, such as the failure of one of a pair of redundant HDDs. For this reason, we estimate that a significant number of server devices fail.

*1 An abbreviation of Direct Attached Storage. At IIJ we chiefly used devices that connect externally via SCSI I/F until 2008. HDD failures for setups where a local HDD is added to a server are not included.

*2 As the devices for which 7, 5, and 4 units failed in the same RAID in 2005 were the same product type and were installed in the same period, it is highly likely they were part of a faulty batch. We removed them from service use in 2005.

Stoppages caused by memory failures and reboots stand out as reasons for failures. There were a wide range of other reasons such as thermal runaway due to fan failure, backplane failures in connectors for devices such as HDDs, power (VRM) failures, and processor failures.

When a failure to boot occurs in a server device fitted with a local HDD, it is necessary to carry out repairs onsite to relocate the HDD from the failed equipment into working server equipment. A great deal of personnel expenses are incurred in order to have operators who can perform server device maintenance such as that mentioned above on duty 24/7 at a data center. When considering a remote data center, costs will increase if operators are on duty 24/7, but if operators are not on the premises it will take a great deal of time to restore service. For this reason, we made diskless operation in which servers have no local data a prerequisite when selecting devices.

4.2.3 Network Device Failure Rates at IIJ

Failure numbers for edge L2 switches that accommodate hosts, etc. are not high in our experience to date. Excluding cases that involve specific faulty batches such as faulty condensers, the figure is less than 1% of the units in operation.

It is also possible to adopt a redundancy configuration using NIC redundancy settings, but this makes the addition of NIC and network devices necessary, increasing costs. Additionally, as adopting a redundant configuration over servers using devices such as a load balancer usually makes operation easier, we went with a design implementing NIC redundancy settings only when higher reliability was necessary, and did not add redundancy to switches accommodating the servers for the basic configuration. Also, to limit the impact of edge L2 switch failures, we designed a system that made it possible to either form a redundant configuration with devices under the control of other server switches using a load balancer, or deal with failures by migrating the contents of server devices controlled by the faulty server switch to a server under the control of another server switch using remote control.

4.3 NHN Configuration

For NHN, we used the following configuration that also supports remote data centers.

- A server pool system is used and large numbers of servers with identical specifications are installed. Onsite work such as device installation and the physical replacement of failed devices can be consolidated into monthly scheduled maintenance.
- Server configuration is standardized and no modifications are made to physical configurations such as changes to the amount of physical memory installed or the installation of a local HDD, in order to curb configuration management costs.
- Energy-saving servers are used to increase the number of servers that each rack can accommodate.
- Racks are not assigned to specific services to increase the number of servers that each rack can accommodate.
- Devices and cabling are standardized to reduce the possibility of mistakes when device installation or replacement work is outsourced on a case-by-case basis.
- Consolidation and degree of freedom is raised through combination with virtualization technologies such as Xen and OpenVZ.
- VLAN is used to make it possible to configure a logical network without changing onsite cabling.
- iSCSI is used to configure a cheap SAN. RAID controller and connection path redundancy are required for storage to prevent service outages as far as possible.
- It is assumed that diskless servers will break down. The goal is to make it possible to complete temporary restoration by switching over devices remotely when a hardware failure occurs.
- If the time it would take for switchover to a backup server to be completed when a failure occurs in a diskless server results in a longer service outage than is acceptable, two servers from separate groups are prepared in advance and a device such as a load balancer is used to create a redundant configuration, making an application-based redundancy configuration possible.
- All work that is required after installation, including OS installation, can be executed remotely.

Next, we explain each technological component.

4.3.1 Adoption of iSCSI-based IP SAN

When use in a remote data center is taken into consideration, the service downtime of storage can lead to extended failures, so we sought even higher storage reliability than we had in the past. IIJ has introduced a system using iSCSI, as it excels at IP network technology, and it makes it possible to put together cheaper systems than FC SAN. Also, by making RAID controller and connection path redundancy a requirement, we have reduced the risk of storage failure that may render our services inaccessible to a minimum.

4.3.2 Implementing a Diskless Server Combining iSCSI Storage and Energy-Saving Servers

Few current server devices support iSCSI boot, and installing iSCSI HBA on all server devices is costly. Because of this, we devised a way to make SAN boot with iSCSI possible using PXE boot, which is supported by the onboard NIC of most server devices.

Information such as the iqn*3 and IP address that is necessary for iSCSI boot is passed on using DHCP options. This makes it possible to replace a failed server device with a backup device by overwriting configuration information and booting the backup server.

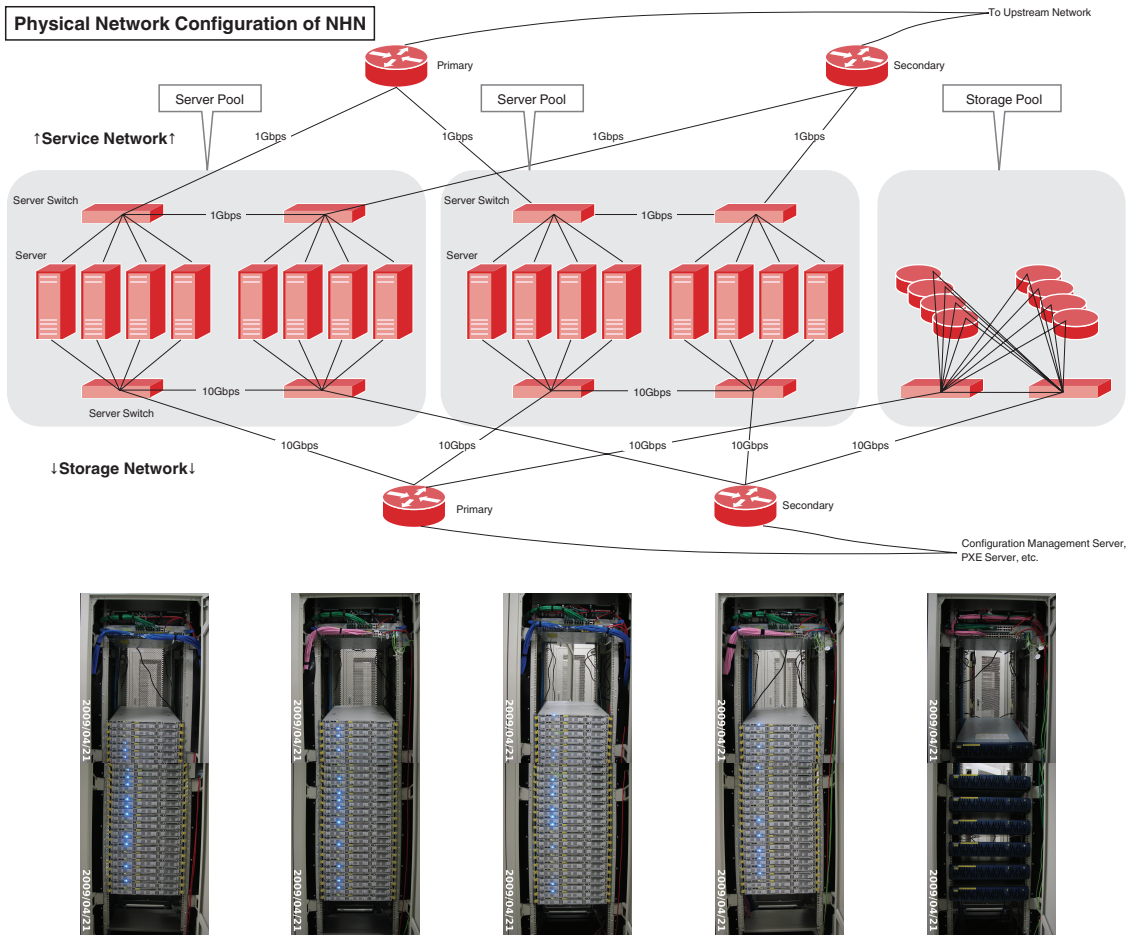
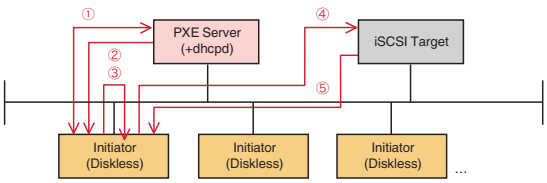


Figure 1: Physical Configuration of NHN



1. Address and parameters are acquired using DHCP.
2. Kernel and initrd images are acquired from TFTP.
3. Kernel is booted in an initrd ramfs root environment.
4. iSCSI initiator is initialized based on DHCP parameters (visible as a local HDD).
5. This is mounted as a block device, and switchroot is used to switch to the mounted location.

Figure 2: Diskless Server (Linux) Boot using PXE boot and iSCSI

*3 An abbreviation of iSCSI Qualified Name. Here this is used for the unique identification of iSCSI targets.

To avoid the need for onsite work in the event of a server failure, for NHN we used a diskless configuration for servers so they have no local data. If a failure occurs in one server device, it is possible to boot backup servers that are already in place while still maintaining the storage content of the failed device by overwriting configuration information remotely and rebooting.

4.3.3 Implementing a Virtual Network That Requires No Cabling Changes Using VLAN

For NHN, we rendered onsite cabling changes unnecessary by having multiple networks coexist over a single physical cable, making all maintenance possible remotely. Technically, each server is configured as a network using access VLAN and trunk VLAN, with multiple networks made to coexist over a single physical cable as necessary.

VLAN is not a new technology, but for NHN we implemented a system that made it possible to link it with the configuration information that IIJ manages, and automatically overwrite the VLAN settings of a server switch that accommodates a given host. This reduces VLAN setting errors in network devices, and cuts operating costs.

4.4 Results of Adoption of NHN

IIJ began looking at its future use of remote data centers in 2008, implementing drastic revisions to service host configurations under the keyword "NHN," and preparing an infrastructure for the migration of new and existing services.

We adopted new technology for NHN based on our perspective and experience as server operators, introducing virtualization, diskless servers using iSCSI, the centralized management of host information, and dynamic updates for network device VLAN settings.

When we adopted NHN for our internal system, the differences with previous service host configurations were not immediately apparent, and opinions such as the following were sometimes heard.

- I don't trust virtual servers. I'd rather use physical servers.
- This level of performance is not necessary, so I want cheaper servers.
- There is too much memory, so I'd like it reduced.
- Disk unit prices are higher than local disks.

Opinions such as these were heard at first, but even though individual servers have slightly higher prices, people understood the cost benefits when taking into account factors such as consolidation, facilities, and operating costs after the introduction of virtualization, and currently there is less resistance to in-house introduction.

Additionally, people are beginning to see the merits of servers that can be ready to use a few days after being requested, and can be returned when no longer required without requiring any physical work to be carried out at the data center. Because until now equipment was constructed using different devices for each service, it was difficult to reuse devices in other services when plans changed, and when devices were required urgently there was a delay before they could be used while waiting for their purchase and installation to be completed. Due to the introduction of NHN, service development flow and the system for making demand forecasts for equipment upgrades are gradually changing.

IIJ will continue investigating the security and I/O virtualization aspects of this technology, and polish it for eventual release to IIJ GIO and other customers.

Authors:
Yasumitsu Makino
Section Chief, System Operating Section, System Infrastructure Division, IIJ Service Business Department
Mr. Makino is engaged in the design and operation of server infrastructure for corporate and private services. Since 2008 he has consolidated equipment procurement and onsite construction work for service hosts at the System Operating Section, and worked towards optimizing the consolidation and operation of equipment through infrastructure systemization.

Shinya Hanataka
System Operating Section, System Infrastructure Division, IIJ Service Business Department
Tadashi Kobayashi
System Operating Section, System Infrastructure Division, IIJ Service Business Department

Internet Topics: About the anti-Malware engineering WorkShop 2009

Here, we introduce the anti-Malware engineering WorkShop 2009 (MWS2009), which was held at the Toyama International Conference Center over three days from October 26, 2009*1. This is a research workshop for malware analysis sponsored by the Information Processing Society of Japan and the Cyber Clean Center Steering Committee. This is the second time the workshop has been held, following on from MWS2008*2 last year, and this year over 100 researchers, students, and corporate engineers actively participated in presentations and discussions. The workshop involved creating CCC DATAset2009 for common analysis from malware activity information (malware specimens and communications data) acquired at the Cyber Clean Center*3, and sharing the achievements of research into technology for analyzing and producing countermeasures by utilizing this data set*4.

■ CCC DATAset2009

Presenters are provided with the following three types of data as CCC DATAset2009.

●Malware specimen data

Hash values for 10 malware specimen varieties (researchers must obtain malware specimens themselves). Analysis results are also provided at a later date.

●Attack communications data

Two day's worth of communications data from two honeypots (CCC DATAset2009 uses data from March 13 to 14, 2009). A packet dump showing actual communications is provided.

●Attack origin data

A year's worth of attack communications records from 94 honeypots (CCC DATAset2009 uses data from May 1, 2008 to April 30, 2009). This includes the time, the source IP address of the attack origin, the destination port number, and the name of malware.

CCC DATAset2009 content has been expanded beyond that of CCC DATAset2008*5. Presenters can verify the effectiveness of their research results using actual observed data by applying them to one of these data sets.

■ MWS2009

Last year there were 22 general oral presentations (including 8 given by students), and for MWS2009 this number rose to 30 (including 15 given by students). Presentations covered a wide range of topics, from methods for carrying out efficient malware analysis, to techniques for improving honeypots, techniques for visualizing malware activity, and attempts at predicting this activity*6. As indicated by the number of presentations given by students, another feature of this year's workshop was the more active role played by students. At IIJ we compared CCC DATAset2009 attack origin data with observed data from the MITF honeypots that we operate ourselves, and presented a big-picture look at differences in the observed results. We believe that by comparing multiple observed results for the same period we can help fuel discussions regarding the locality of events and the accuracy of observation techniques.

The MWS Cup 2009 event was also held for the first time at this workshop. In this event participants vied for the best analysis technology by analyzing malware activity record data they were given as a task within a set time limit. Participants brought tools they had created as part of their research as well as the analysis environment they use on a daily basis, and competed to analyze data handed out on CD on the day of the event as quickly and accurately as possible. Seven teams participated despite this being the first time the event was held, and once again students did remarkably well, with university teams reaching the top ranks.

The anti-Malware engineering WorkShop can be thought of as an effective venue for comparing research results in a fair way through the analysis of common data, and a place for developing human resources that have expertise in measures against malware in the future. IIJ also considers it an invaluable occasion for interacting with the academic community that we seldom have the opportunity to exchange opinions with, and we would like to continue to actively participate in this workshop in the future.

The MWS Cup 2009 event

Author:
Mamoru Saito
General Manager of the Division of Emergency Response and Clearinghouse for Security Information, IIJ Service Business Department

*1 Anti-Malware engineering WorkShop 2009 (http://www.iwsec.org/mws/2009/)(in Japanese). Held concurrently with the Computer Security Symposium 2009 (CSS2009) (http://www.iwsec.org/css/2009/english/index.html) sponsored by the Computer Security (CSEC) Group of the Information Processing Society of Japan. Photos of MWS2009 can be viewed on the MWS2009 activity archive (http://www.iwsec.org/mws/2009/photo.html)(in Japanese).

*2 IIJ also attended MWS2008 held last year. MWS2008 (http://www.iwsec.org/mws/2008/)(in Japanese). This workshop was discussed in a conversation published in IIJ.news Vol.90 (http://www.iij.ad.jp/news/iijnews/2009/vol90.html)(in Japanese).

*3 The Cyber Clean Center is an anti-bot measures project coordinated by the Ministry of Internal Affairs and Communications, the Ministry of Economy, Trade and Industry, and other related organizations (https://www.ccc.go.jp/en_index.html).

*4 Other endeavors involving the sharing of research results based on common data include the DARPA Intrusion Detection Data Sets (1998, 1999) (http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html) and the Knowledge Discovery and Data Mining Tools Competition (http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html).

*5 CCC DATAset 2008 consisted of one variety of malware specimen data, two day's worth of attack communications data, and six month's worth of attack origin data respectively.

*6 MWS2009 papers and presentations for which the author's consent has been received will be published on the following site (http://www.iwsec.org/mws/2009/)(in Japanese).

Ongoing Innovation

Vol.5 November 2009

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-MKTG020CA-0912CP-00001PR