

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

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

2.1 Introduction

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

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

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

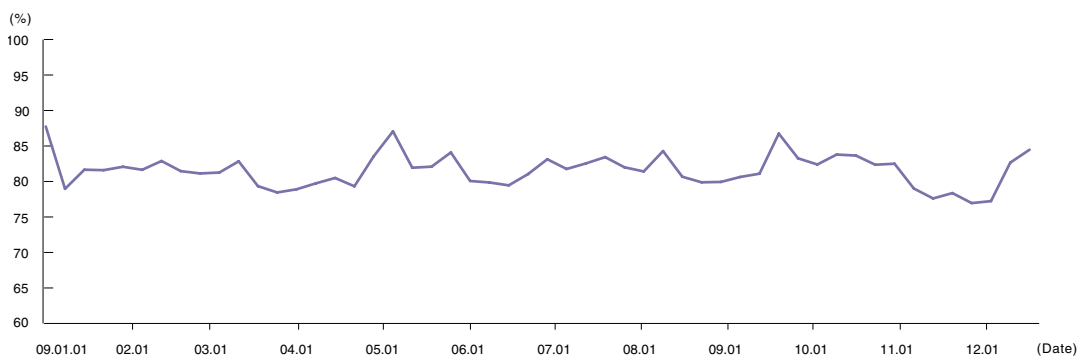


Figure 1: Spam Ratio Trends

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

2.2 Spam Trends

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

2.2.1 Spam Ratio Trends

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

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

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

2.2.2 Sources of Spam

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

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

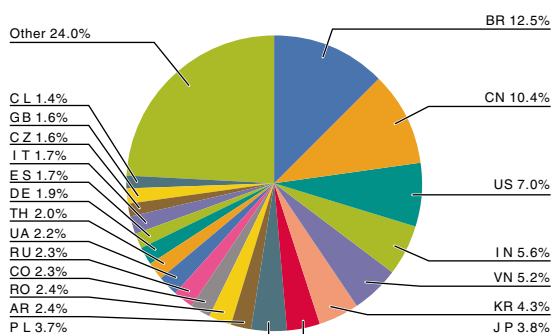


Figure 2: Regional Sources of Spam

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

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

2.2.3 Spam Sending Trends

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

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

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

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

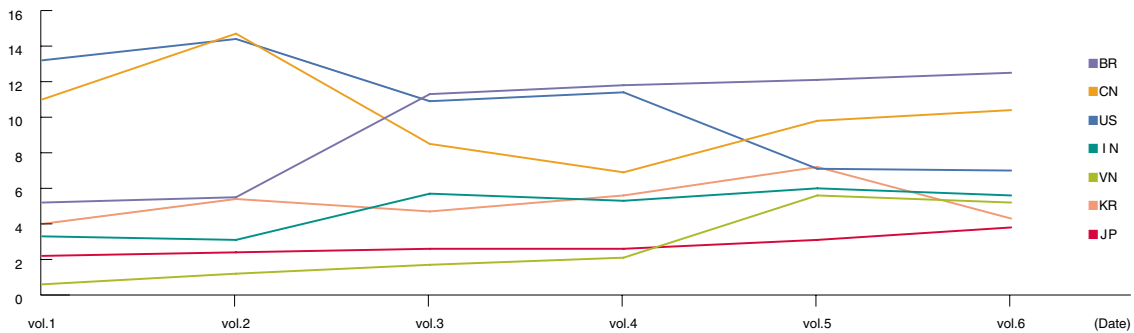


Figure 3: Trends in Sources of Spam

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

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

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

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

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

2.2.4 Sender Authentication Technology Implementation Status

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

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

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

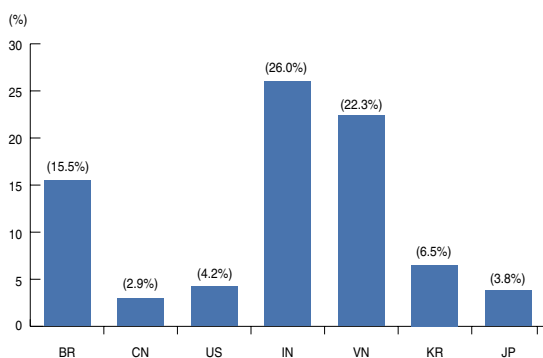


Figure 4: Ratio of Sources that Send Spam Infrequently

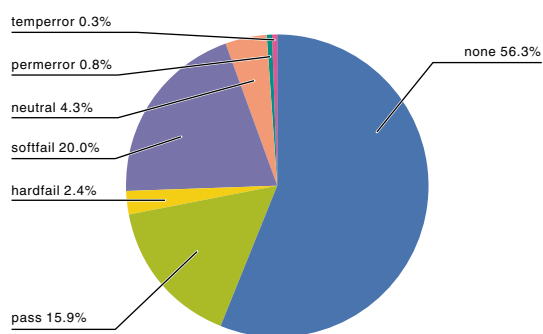


Figure 5: SPF Authentication Result Ratios

2.3 Trends in Email Technologies

2.3.1 DKIM ADSP Background

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

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

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

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

Table 1: DKIM-ADSP Naming Changes

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

2.3.2 DKIM ADPS Overview

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

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

_adsp._domainkey.example.jp

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

Table 2: DKIM-ADSP Values

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

2.3.3 DKIM Updates

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

2.4 Conclusion

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

Author:

Shuji Sakuraba

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