

2 Email Technical Report

2.1 Introduction

The Email Technical Report summarizes the latest trends in spam, technical counter measures to spam, etc. For trends in spam, the results of a variety of analyses conducted based on various information obtained from the Spam Filter feature provided in IJ email services will be presented. Note that since the flow of email varies depending on the day of the week, in order to more easily understand the trends, the data was aggregated on a weekly basis using the week-numbering year*1 as the unit of data, and the data was analyzed focusing on the changes in the data.

This survey covers a period of 13 weeks or 91 days, from the first week of 2009 (12/29/2008 to 1/4/2009) to the 13th week (3/23/2009 to 3/29/2009).

Regarding technical counter measures to spam, continuing from the prior volume, "Sender Authentication Technologies" will be discussed. This volume will provide an overview regarding DKIM (DomainKeys Identified Mail) which uses digital signature technology.

2.2 Trends in Spam

This section provides a report focused on the trends in the ratio of spam detected by the Spam Filter feature provided by IJ and information related to sources of spam. Regarding sources of spam, there has been a significant change since McColo's network was shutdown in November 2008 (Reference: Internet Infrastructure Review Vol.2). Continuing from the prior volume, we will analyze these trends and changes.

2.2.1 Ratio of Spam

The weekly trends in the ratio of spam over a period of 91 days from the first week of 2009 to the 13th week are shown in Figure 1.

The ratio of spam averaged 81.5% of all incoming emails during this period. The ratio was highest during the first week (12/29/2008 to 1/4/2009) at 87.8%. This is because the period coincided with the year-end/new year holiday period and there was only a small amount of business email, thus the relative ratio of spam was higher. The average value for the prior period (9/1/2008 to 12/28/2008) was 82.7% and thus the ratio of spam decreased 1.2%. However, the ratio of spam is still very high and we believe it is necessary to continue to enhance spam control measures.

2.2.2 Sources of Spam

The sources of email that IJ determined to be spam are listed by country in Figure 2.

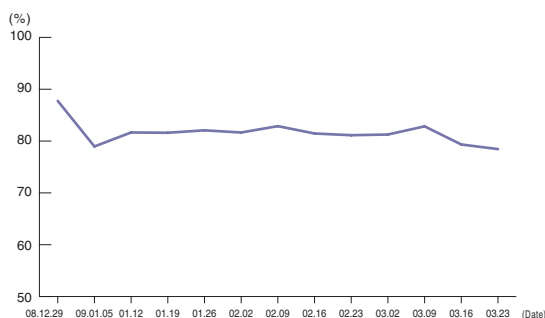


Figure 1: Ratio Proportion of Spam

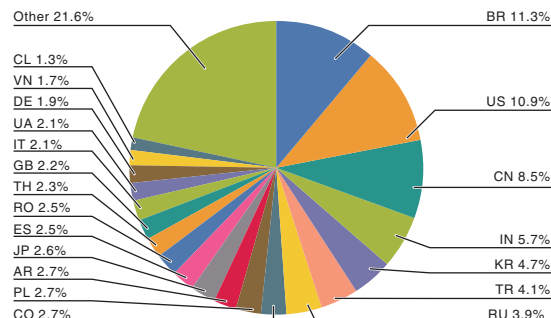


Figure 2: Sources of Spam

*1 Since the week-numbering year is determined based on ISO 8601 "Data Elements and Interchange Formats - Information Interchange - Representation of Dates and Times", a period from 2008 is also included.

In this survey, the top source of spam was Brazil (BR) with 11.3%. This is a big jump compared to our prior survey. Brazil ranked only 5th with 5.5%. In the prior survey regarding the trends in sources of spam, Brazil ranked 2nd during the final week of 2008 (52nd week), thus there was a concern regarding the future increase in spam from this region.

The United States ranked 2nd (US, 10.9%) as in the prior survey, and China which ranked 1st in the prior survey ranked 3rd (CN, 8.5%). India ranked 4th (IN, 5.7%), which is a big jump from the prior survey when it ranked 8th. Korea (KR) ranked 5th, Turkey (TR) ranked 6th, and Russia (RU) ranked 7th, with all of three continuing to rank high in the list from the prior survey. Japan (JP) ranked 11th with 2.6%. According to this survey results, spam from Brazil, the U.S.A., and China made up approximately 30%, and spam from the top 7 countries made up approximately half of all spams. In order to reduce the amount of spam received in Japan, we believe it is necessary to implement control measures on the part of the sender, such as the implementation of OP25B*2 in these top ranking countries.

The weekly trends in the ratio of spam from these 7 countries and Japan are shown in Figure 3. U.S.'s ranking continued to fall after McColo's network was shutdown in November last year, but its ranking rose sharply from the 5th week (1/26 to 2/1) and thereafter it continued to rank high in the list. From this observation, we can presume that botnets which were previously affected by the shutdown of McColo's network have gained control over new management servers (Command & Control Servers), etc. and have resumed spamming activities. In addition, Brazil, whose ranking rose starting November of last year, continued to rank high in the list and as a result ranked no.1 for the entire period under study.

China, which ranked 1st in the prior survey, experienced a downward trend, however, its ranking rose again from the 12th week to the 13th week (3/16 to 3/29), requiring ongoing attention.

In the graph in Figure 3, we can see a difference in the trends in the ratio of spam sent from Japan and the ratio of spam sent from the other top ranking countries. Whereas the ratio of spam sent from Japan stayed relatively constant at approximately 2.5%, the ratio of spam sent from the other top ranking countries fluctuated considerably. This is an interesting difference and we plan to continue our research regarding the trends in the sources of spam and the differences in situations of each country which may have caused the fluctuation.

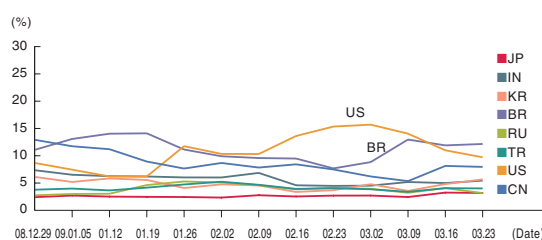


Figure 3: Trends in Sources of Spam

2.2.3 Spam Originating from Japan

As mentioned previously, in Japan, many ISP's have implemented OP25B, and thus it can be said that control measures against spam originating from Japan are relatively effective. The survey results for this report show that 97% or more of the spam was sent from outside Japan. Similar results were presented in a report created by Sophos Plc.*3, which periodically publishes a ranking of the top source countries of spam.

The results of "Communications Usage Trend Survey"*4 conducted by the Ministry of Internal Affairs and Communications in 2008 indicate that there are 90.91 million internet users in Japan, and that the proportion of broadband lines (FTTH, xDSL, CATV, etc.) used by households is 73.4%, indicating that the use of high-speed access lines is widespread. Although this condition alone would seem to suggest that in Japan there is a possibility for a massive amount of spam to be distributed over a short period of time, the results show that the actual amount of spam sent is low compared to other countries. We believe that this is due to the large number of ISP's implementing OP25B as well as the scale of implementation*5.

*2 OP25B (Outbound Port 25 Blocking) is technology that restricts access from dynamic IP addresses used by consumer users for internet connection to port number 25 which is used between mail servers in the external network, and it is known to be extremely effective in preventing spam from being sent.

*3 Sophos's URL is <http://www.sophos.com/>. In the 2008 spam source country ranking, Japan ranked 36th.

*4 Results of the survey conducted by the Ministry of Internal Affairs and Communications: http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/Releases/Telecommunications/news090407_b.html

*5 According to a study conducted by the Anti-Spam Consultation Center of the Japan Data Communication Association, 49 ISPs have implemented OP25B (http://www.dekko.or.jp/soudan/common-folder/image/About_ASCC.pdf).

Although as a result of these efforts, the amount of spam originating from Japan has been reduced considerably, it is also a fact that this ratio has never reached zero. For example, the ratio of spam originating from Japan to the total amount of spam out of the total amount of email is equivalent to 2 emails for every 100 emails received. This figure is for the enterprise email service provided by IJ, and thus in the case of mobile phone email, etc. the ratio may be even higher.

What are the sources of spam originating from Japan? One source may be spammers who have obtained a static IP address. Static IP address users are obviously not permitted to send spam, and in most cases, the sending of spam is considered improper use, and is grounds for termination of the contract with the ISP or the internet access service being used. However, it is very difficult for the administrator on the part of the sender to determine whether spam is being sent or not, and it is usually discovered after the recipient of spam makes a report, etc. In addition, it is necessary to find clear evidence that the person reporting the spam is providing correct information and that the contractee in question is actually sending spam, and it takes time to handle the problem.

Another source may unfortunately be the result of inadequate implementation of OP25B. It is known that spammers consistently target such loopholes and continue to send massive amounts of spam. One case that is recently becoming more common is where spam is sent using mobile data communications terminals. Mobile data communications are a convenient service which can be used anywhere, and the number of users has been rapidly increasing. However, at the same time, the misuse of these services is also becoming more common. There is a plan to implement OP25B, and since the effectiveness of OP25B has been confirmed, its implementation at an early stage is desirable.

In order to eliminate spam originating from Japan, it is necessary to correct these inadequacies on the part of the sender. In addition, learning from Japan's experience, to reduce the amount of outgoing spam at a global level, we believe it is important that OP25B be implemented in each country. IJ presents the advantages of OP25B at various international conferences and opportunities, and we plan to continue to actively provide information.

2.3 Trends in Email Technologies

2.3.1 Trends in Sender Authentication Technologies

In the prior volume, we introduced two technologies for sender authentication: one is a network based technology, and the other is one that uses digital signature technology. Network-based SPF/Sender ID technology was covered in a series in the prior volumes, and in this volume we will provide an overview on DKIM (DomainKeys Identified Mail) which uses digital signature technology.

According to a study conducted by WIDE Project*6, the rate of publication of SPF records for "jp" domains in March 2009 was 34.77%, which is a 1.49% increase since the time of publication of the prior volume (January 2009). This was only a small increase compared to the previous increase (8.84%), but considering the fact that the number of domains is also increasing, it can be said that the number of domains implementing SPF is steadily increasing. In addition, for "co.jp" domains which are used by incorporated companies, the rate of publication of SPF records increased from the prior volume and reached a high 41.71%. In light of the harmful effects such as bounced email described in the prior volume, this trend indicates a growing awareness regarding the protection of domain names as corporate brands.

Meanwhile, the general rate of publication of DKIM-related records is 0.38%, indicating a substantial delay in implementation. This is said to be caused by the big difference in implementation cost on the part of the sender.

2.3.2 Sender Authentication Technologies Using Digital Signature Technology

The purpose of sender authentication technologies is to enable the receiving side to determine the authenticity of the sender information claimed in the incoming email. In other words, these technologies authenticate that the email is sent from a source that is approved by the administrator (domain) of the sender identified in the sender information.

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

In network-based SPF/Sender ID technology, the sender publishes the SPF record in the DNS. In DKIM, a digital signature which can only be appended by the sender managing the private key is appended to the email header. If the private key on the part of the sender is properly managed so that it is not leaked, it is generally difficult for a third party who doesn't know the private key to create a digital signature. This mechanism enables identification of the sender.

To implement DKIM on the part of the sender, in addition to the steps required for sending email, steps to create a digital signature using the email that is to be sent, and steps to append the signature to the email header need to be carried out. Since these steps are usually performed on the outgoing mail server, there is no impact on general email senders; however, new features need to be added to the outgoing mail server. The cost of adding these features is significantly higher than the cost of implementing SPF/Sender ID in which a record is simply published in the DNS one time, resulting in the substantial difference in penetration rate.

2.4 DKIM Authentication Flow

The DKIM specifications have been published as RFC4871 by IETF. The authentication flow for DKIM is shown in Figure 4.

2.4.1 Sender Actions

The signature of the email is stored in the DKIM-Signature header field. The DKIM-Signature header

field contains the signature itself and information such as the signed header fields, a hash, an encryption algorithm, query method for retrieve the public key, the expiration date of the signature, etc. which are set as parameters in the header.

First a hash value is calculated for the email body and email header respectively. Before calculating the hash values, the email body and email header are canonicalized to prepare them for the message modifications*7 carried out when the email is transmitted. The hash value for the email body is BASE64 encoded and stored as the "bh=" tag (parameter) in the DKIM-Signature header field. The DKIM-Signature header field is always included before calculating the hash value for the email header. The other email header fields to be included in the hash calculation (i.e. header fields to be included in the signature) are selected and the selected header field names are specified in the "h=" tag of the DKIM-Signature header field. At this stage, the final signature information is unknown, thus the "b=" tag which indicates the signature information in the DKIM-Signature header field is not included in the hash value calculation of the header*8. Since the hash information for the email body is included in the hash value calculation of the header, the email body is also included in the signature.

Signature information is created using public key cryptography technology based on the hash value of the header.

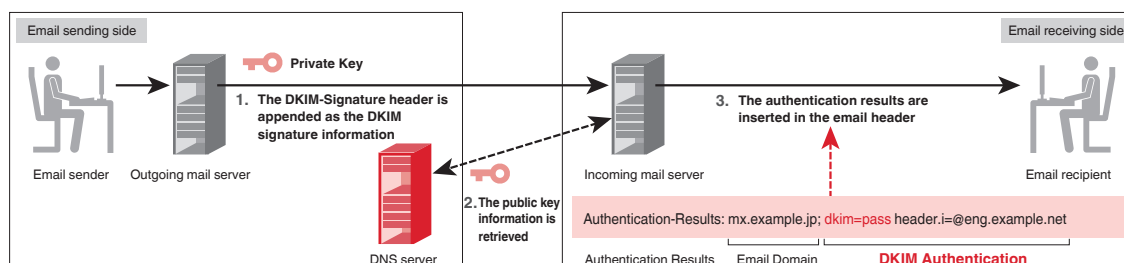


Figure 4: DKIM Authentication Flow

```
DKIM-Signature: a=rsa-sha256; d=example.net; s=brisbane;
c=simple; q=dns/txt; i=@eng.example.net;
t=1117574938; x=1118006938;
h=from:to:subject:date;
z=From:foo@eng.example.net!To:joe@example.com!
Subject:demo=20run!Date:July=205,=202005=203:44:08=20PM=20-0700;
bh=MTIzNDU2Nzg5MDEyMzQ1Njc4OTAxMjM0NTY3ODkwMTI=;
b=dzdVvYOfAKCdLXdJOc9G2q8LoXSIEniSbav+yuU4zGeeruD00!szZ
VoG4ZHRNiYzR
```

Figure 5: Example of DKIM-Signature

*7 For example, RFC5322 specifies a recommended value (98 characters or less) and a maximum value (998 characters or less) for the length of one line of an email body or header to be sent, and the mail server may automatically reformat the email by wrapping lines.

*8 The "b=" string, which indicates the parameter name, is included but the value after the "=" is set to blank before performing the calculation.

2.4.2 Process on the Receiving Side

The server receiving an email which contains DKIM signature information first retrieves the public key upon receiving the email. The retrieval method is specified in the “q=” tag of the DKIM-Signature header field but DNS is used by default. The domain name from which the public key is to be retrieved is constructed from the domain name specified in the “d=” tag of the DKIM-Signature header field and the selector specified in the “s=” tag. For example, in the header shown in Figure 5, the domain name is “example.net” and the selector is “brisbane.” Thus the location shown in Figure 6 will be referenced.

brisbane._domainkey.example.net

Figure 6: Example of Public Key Retrieval Location

The sub domain “_domainkey” following the domain of the signing entity “example.net” is a fixed name under which the DKIM key information, etc. is stored.

As described above, the retrieval location of the public key cannot be determined until the actual email is received and the DKIM-Signature header field in the email is examined. This makes it difficult to conduct a study on DKIM implementation. On the other hand, there are also benefits to using selectors. Since DNS has a caching mechanism, even if a record is rewritten, the email receiving side is not immediately updated, and the timing of the update is also variable. By storing a new public key in a domain which uses a different selector, the selector name can be changed when the matching private key is changed, allowing the receiving side to retrieve the public key that corresponds to the signature without any confusion. In addition, by creating a sub domain, the management of the key can be outsourced. This is convenient when outsourcing email operations to hosting services, etc. After the public key is retrieved, the email body is canonicalized similarly to when the email was sent. A hash value is then calculated and the value is compared with the value of the “bh=” tag. The public key is used to verify the signature using the algorithm indicated in the “a=” tag.

2.5 Conclusion

This volume’s survey results indicate that the amount of spam remains high at over 80% of all emails. As the sources of spam are rapidly changing it seems that spammers are aggressively seeking new methods for sending spam since the shutdown of McColo’s network in November last year, and we believe that it is likely that the condition may worsen in the future. IJ plans to continue to globally promote Japan’s best practices in spam control such as implementing OP25B, etc. in order to support the reduction of spam. In addition, since there is still a certain amount of spam originating from Japan, we believe it is important to work on measures for patching the remaining loopholes on the part of the sender. In this volume, an overview on DKIM, a sender authentication technology using digital signature technology, and the DKIM implementation trends were presented. DKIM not only allows you to determine authenticity of the email sender, but it also allows you to determine whether the email contents have been altered, and it can be used to secure the reliability of email which is becoming an ever more important communication tool. IJ plans to continue to promote the widespread use of these technologies and actively provide information in order to help ensure secure email connections.

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 in charge of planning and researching email systems. He is involved in various activities in collaboration with the R&D department and 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.