The Latest Trends in Spam

2.1 Introduction

In the first IIR Messaging Technology article in a year, we will report on technical information about email, including trends in spam and spam countermeasures. There has been a downward trend in the volume of spam over the past few years, but recently there was a temporary spike in March 2016. In this report we discuss the results of our investigation into the regions from which this increase originated. Regarding trends in email technologies, we discuss our findings regarding the adoption status of sender authentication technology that it is hoped will become more widespread in the future, with a focus on DMARC.

2.2 Spam Trends

In this section we look at changes in spam trends, based on trends in the ratios of spam detected by the spam filter provided through IIJ's email services. As we have done up until now, trends in the ratio of incoming mail determined to be spam relative to the overall volume of incoming mail collated by week are shown using graphs and other means. For some time the volume and ratio of spam had dropped to significantly lower levels than 2008 when the IIR was first published, but there was a temporary rise in March 2016.

The graph in Figure 1 that indicates spam ratio trends incorporates three years' worth of data, including the year since the last IIR report (Vol.27), covering the 53 weeks between March 30, 2015, and April 3, 2016. See IIR Vol.27 for information about trends previous to this. As you can see in the graph, the ratio of spam has generally been decreasing except for long holiday periods such as the year-end and New Year holidays. However, since around 2015 the range of reduction has narrowed. The average ratio for fiscal 2015 was 24.2%. The ratio in fiscal 2014 was 31.7%, so this represents a drop of about 7.5%. The decrease from fiscal 2013 to fiscal 2014 was 15.7%. However, in March 2016 the rate once again began to trend upward, rising to as high as 44.8% in the week of March 28, 2016. Following this, preliminary figures indicated a drop back to around 20%, so we believe this was a temporary spike. We will analyze trends for the increase in spam during this period a little later.

2.2.1 Risk Remains High

According to a report published by the National Police Agency on March 17, 2016*1, the total monetary damage caused by illegal remittances related to Internet banking in 2015 exceeded the record high of the previous year, coming to approximately 3,073,000,000 yen. There was also a record 3,823 incidents of targeted email attacks reported by affiliated business operators, so the risk associated with email remains high. Reports also indicate that in 77% of cases most sender addresses in targeted emails were spoofed, so it is clear there is an urgent need to popularize and implement sender authentication technology to protect against the spoofing of sender information.

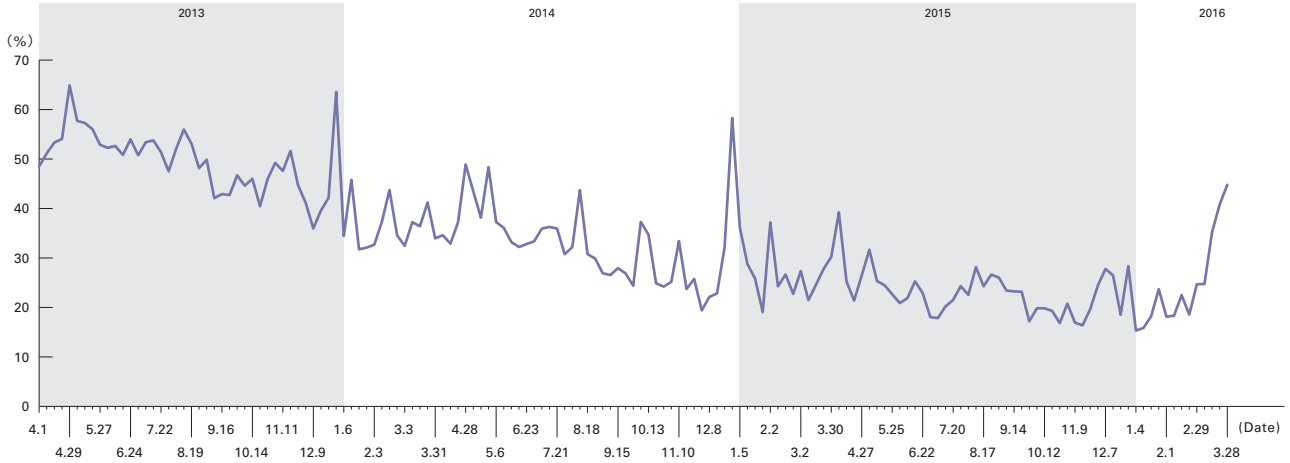

Figure 1: Spam Ratio Trends

*1 Report on Cyberspace Threats for 2015 (http://www.npa.go.jp/kanbou/cybersecurity/H27_jousei.pdf) (in Japanese).

2.2.2 Ratios for Regional Sources of Spam

The graph in Figure 2 shows sources of spam by region between January and March 2016, which corresponds to the fourth quarter of fiscal 2015.

During this period, the region from which most spam originated was the United States (US), at 16.8%. Among the surveys we have presented to date, this is the first time since IIR Vol.10 (third quarter of 2010) that the United States has been the top regional source. After holding the top position up until now, China (CN) fell to second place at 6.4%. Brazil (BR) was the third highest regional source, and also had a ratio of 6.4%. Fourth place was Japan (JP), which once again had a ratio of 6.4%. Following on from that in order was India (IN, 6.3%), Vietnam (VN, 6.1%), Mexico (MX, 5.4%), Hong Kong (HK, 3.1%), Argentina (AR, 2.5%), and Spain (ES, 2.5%). Other than Hong Kong and Vietnam, which are close to Japan, we can see that regions with large territories and high populations held the top positions. Figure 3 shows trends in the volume of spam for these top 10 countries. This time we examine trends in spam volumes rather than the ratio of spam, to analyze the spam increase in March 2016. For that reason no figures are shown on the vertical axis, but a clear comparison can be made between each region.

2.2.3 Trends in the Major Regional Sources

As you can see from Figure 3, among the top regions the United States (US), China (CN), Japan (JP), and Hong Kong (HK) were high in the rankings from the beginning, but in the March 2016 period when the volume of spam spiked they did not increase much. During this period, the peak volume of spam sent from these countries was two to three times that of the lowest volume. Meanwhile, in the other top regions of India (IN), Vietnam (VN), Mexico (MX), Brazil (BR), Argentina (AR), and Spain (ES), the peak volume was at least 10 times the minimum, and in the case of Argentina it was 85 times higher. As each of these regions saw an increase in March 2016 when the ratio of spam was high, we know they contributed to the rise in spam during this period. Because these regions are geographically dispersed, we speculate that the spike may have been caused by a botnet sending spam actively. We believe there is an ongoing need to take measures against botnets like this through international cooperation.

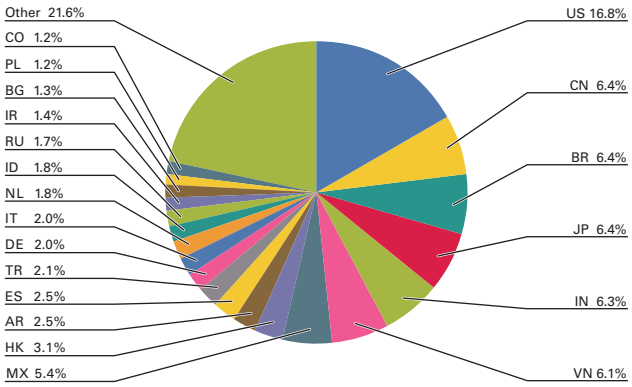

Figure 2: Ratios for Regional Sources of Spam

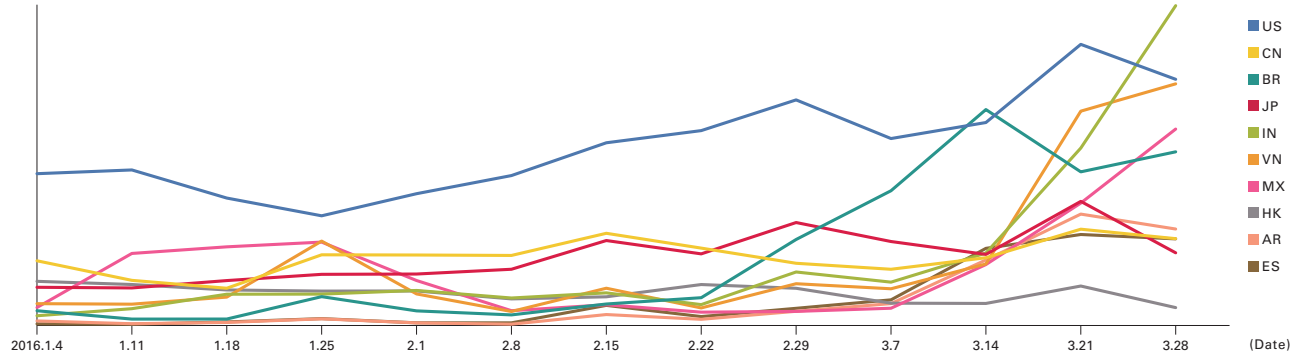

Figure 3: Trends in the Top 10 Regional Sources of Spam

2.3 Trends in Email Technologies

Here we will report on the adoption status and technological trends regarding the sender authentication technology that is also an effective spam countermeasure, with a particular focus on DMARC*2.

We have previously discussed technical details and adoption trends for the SPF*3 and DKIM*4 sender authentication technologies, but in the future we feel that the DMARC standard based on these two technologies will become the main technology used.

2.3.1 An Overview of DMARC

We have already discussed DMARC a number of times since IIR Vol.15, but here we will once again give a summary of its features. DMARC is also a type of sender authentication technology for verifying whether or not the domain used is a legitimate sender based on the sender information. Its main characteristics are as follows.

- Based on matching the domain authenticated using SPF or DKIM with From (RFC5322.From) in the email header (or verifying that the organization is the same)
- Enables recipient behavior to be indicated via policies when sender (domain management) authentication fails
- Senders can specify the report destination when authentication fails
- This information is expressed using a DNS TXT resource record

In other words, DMARC is technology that uses the authentication results of SPF and DKIM, which are either already in widespread use or becoming more prevalent. When DMARC authentication passes, it means the sender information (RFC5322.From) in the header that can be referenced by the email recipient also matches. The sender can confirm that email was sent over the correct route because they receive information in report form when authentication fails. Previously, domains that could be authenticated using SPF or DKIM did not always provide sender information that was easy for the ultimate email recipient to confirm. In a sense, using DMARC authentication has made it possible to unify the domains to be authenticated, making this clearer for the recipient as well.

2.3.2 DMARC Adoption Status

IIJ's email services have supported DMARC since 2014, and incoming mail is authenticated using DMARC. Figure 4 shows DMARC-based authentication result ratios for the three-month period from January to March 2016. Figure 5 and Figure 6 show authentication result ratios for SPF and DKIM, which DMARC authentication is based on, over the same period.

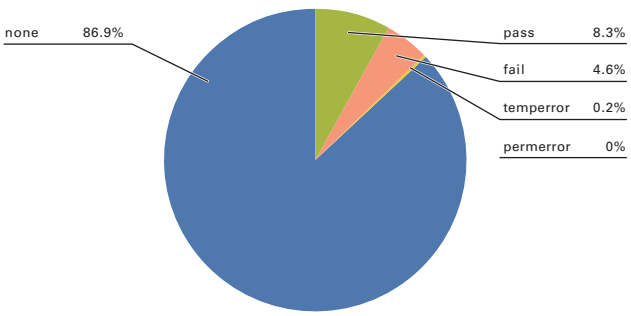

Figure 4: DMARC Authentication Result Ratios for Incoming Mail

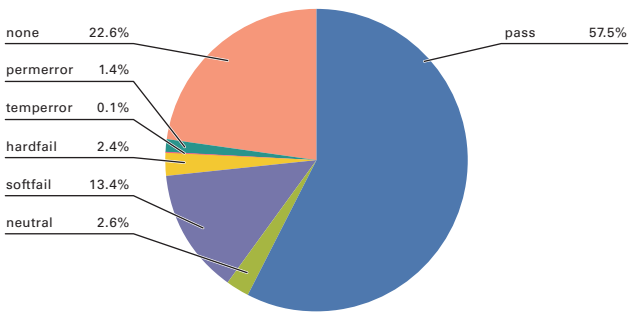

Figure 5: SPF Authentication Result Ratios for Incoming Mail

*2 Domain-based Message Authentication, Reporting, and Conformance (DMARC), RFC7489 (https://rfc-editor.org/rfc/rfc7489.txt).
*3 Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1, RFC7208 (https://www.rfc-editor.org/rfc/rfc7208.txt).
*4 DomainKeys Identified Mail (DKIM) Signatures, STD76, RFC6376 (https://www.rfc-editor.org/rfc/rfc6376.txt).

First, looking at the SPF authentication result ratios in Figure 5, by excluding "none" results that mean SPF authentication could not be performed, we can see that this time a total ratio of 77.4% of senders have implemented SPF. The last time authentication results were presented was in IIR Vol.23 in 2014, and since then this ratio has increased by 4.2%. Because it is comparatively easy for senders to implement SPF, the implementation ratio for SPF has so far tended to be high, and we can see that in this survey the ratio is still increasing slowly.

In the authentication ratios for DKIM in Figure 6, the total for implementation ratios other than "none" came to 20.1%. This was again an 8.5% increase over last time. That means even though the implementation ratio was originally low for DKIM, as it costs quite a bit for senders to deploy, we can see that these ratios are also gradually climbing. The implementation of SPF or DKIM is a prerequisite for implementing DMARC, and as shown in Figure 4, the implementation ratios for DMARC other than "none" results that indicate DMARC authentication is not possible came to 13.1%. Once SPF and DKIM are adopted, DMARC can be implemented by simply adding a DMARC record to the TXT resource record for the "_dmarc" subdomain. In light of this, we believe the implementation ratio for DMARC is lower than SPF and DKIM because recognition of DMARC is still low. We feel that in the future there will be a continued need to promote the benefits of DMARC, as well as methods for its implementation.

Another distinctive point regarding the DMARC authentication ratios in Figure 4 is that the ratio of "fail" authentication results is high at 4.6%. SPF authentication is prone to fail when mail is forwarded, and there have been moves to declare an SPF record to produce less severe "softfail" results when this phenomenon is expected. For this reason, it is possible to anticipate the ratio of "softfail" results using SPF would be the high figure of 13.4%. However, compared with the 2.4% of "hardfail" results when stronger authentication fails using SPF, and the 0.7% of "fail" results for DKIM, it could be said the 4.6% of results indicating failed DMARC authentication is very high. We will analyze the reasons for this in the next section.

2.3.3 Causes of Success or Failure in DMARC Authentication

With DMARC, when either SPF or DKIM authentication succeeds, DMARC authentication is evaluated if the domain in the "From:" email header declared a DMARC record. In other words, when the DMARC authentication result is a "pass," it means that domain (RFC5322.From) and the domain authenticated using SPF or DKIM are a match or are associated, and either the SPF or DKIM authentication result was a "pass." Consequently, we tried analyzing the factors behind "pass" results for DMARC authentication. The results are shown in Figure 7.

When a DMARC "pass" result is produced, the most prevalent pattern for SPF and DKIM authentication results was cases where both methods passed, at 69.8% of the total. In short, we found that the majority of domains that declared a DMARC record and passed DMARC authentication correctly implemented both SPF and DKIM. In cases where either SPF or DKIM authentication failed or was not implemented, and DMARC authentication passed because the other method passed, SPF pass results were most

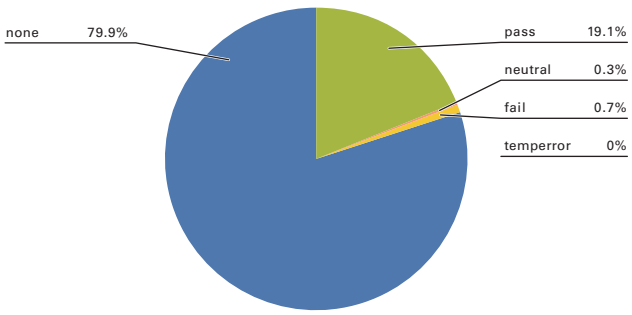

Figure 6: DKIM Authentication Result Ratios for Incoming Mail

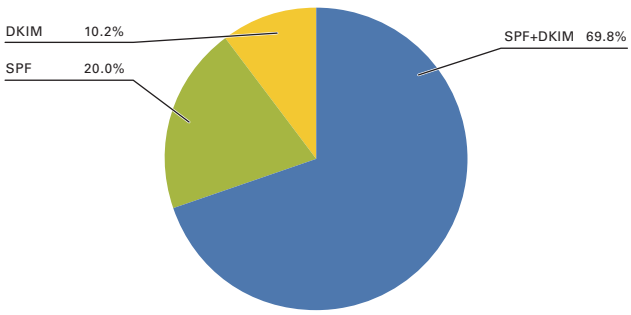

Figure 7: DMARC Pass Factors

frequent at 20.0% The ratio of cases where only DKIM authentication passed was around half that, at 10.2%. We surmise that the high deployment ratio of SPF has carried over to affect the DMARC authentication “pass” results.

Next, we will indicate the most common patterns in SPF and DKIM authentication when DMARC authentication failed in Figure 8.

The most prevalent pattern for SPF and DKIM authentication results when DMARC failed was cases in which only SPF authentication was carried out (DKIM was a “none” result), and the SPF authentication result was “fail.” Conversely, cases in which DMARC failed due to a DKIM authentication “fail” result when only a DKIM authentication result was used amounted to an extremely low ratio of just 0.6%. We believe these results demonstrate the differences in adoption rates between SPF and DKIM, as well as the robustness of DKIM authentication. Cases in which both SPF and DKIM failed made up just 0.7% of the total. Based on these factors, we found that implementing DKIM is an effective way to prevent DMARC authentication failing.

Among the DMARC failure factors, the 10.6% ratio labeled “DMARC” indicates the percentage for which DMARC authentication failed due to a mismatch between the domain authenticated using SPF or DMARC and the RFC5322.From domain authenticated using DMARC. If these were cases in which only the RFC5322.From domain is used as the misrepresented source by spoofing the sender information in SPF or DKIM, this would be a good example of correctly detecting fraudulent activity. However, if these were legitimate emails failing authentication, they could be considered unfortunate cases in which the process failed because the domains to authenticate were different, despite implementing SPF or DKIM and declaring a DMARC record. It appears that in some of these cases mail delivery has been entrusted to another provider, and the failure is caused by the SPF or DKIM authenticated domain being authenticated using the domain of the outsourcing company, resulting in a domain mismatch. Some of the mail I have received, such as email newsletters sent from major banks or other organizations, has also failed DMARC authentication for this reason. Mail sender information indicates the source of mail, so the domain used with SPF or DKIM should also be applied in a way that makes it easily and correctly identifiable as the sender’s domain.

The “none” ratio shown in Figure 8 indicates the pattern in which DMARC authentication produces a “fail” result despite the fact that the authentication result for both SPF and DKIM is “none.” This is another instance in which it would be good for DMARC if this could be detected as fraudulent activity, but it seems this is not always the case. Upon further investigation, it appears this can also be attributed to a system called Organizational Domains, in which the RFC5322.From domain that is a characteristic of DMARC and higher-level domains are treated as domains for the same organization. In other words, although the mail sent does

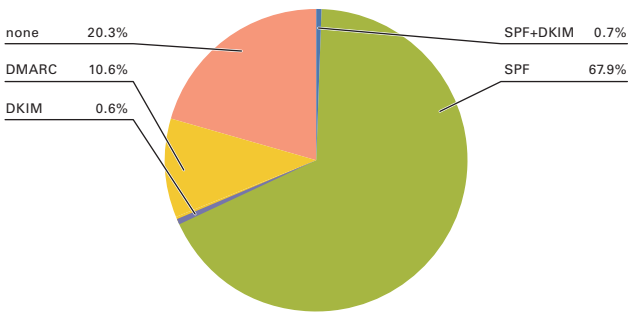

Figure 8: DMARC Failure Factors

not support either SPF or DKIM, a higher-level domain than the RFC5322.From domain in the header is declaring a DMARC record, so an attempt is made to authenticate using DMARC, producing a "fail" result. We recommend that domain administrators also configure SPF or DKIM for subdomains when declaring a DMARC record.

2.3.4 Trends in Technologies Related to DMARC

In previous IIR we have discussed that DMARC authentication can sometimes fail even for legitimate mail when resending mail, such as when sending emails to mailing lists or forwarding mail. This issue is also recognized by the organizations evaluating the specifications for DMARC, and ARC (Authenticated Received Chain)*5 has been proposed as a specification for remedying the problem. As the name suggests, this technology attempts to create an authenticated chain by linking information that is already authenticated at times such as when mail is resent. We would like to take a look at this system once the ARC specifications have been further clarified.

2.4 Conclusion

The ratio of spam began to decline gradually from 2010, but as announced in this report, there was a period in which it rose temporarily. It is said that the reason for the decrease up until now is the effectiveness of ongoing measures to prevent the activity of botnets, which are the main methods used to send spam. Hopefully this recent spike is just temporary, but we believe circumstances that enable this kind of mass-mailing capacity to continue to exist pose a threat.

Ongoing vigilance is also required with regard to the qualitative issues of spam. A high rate of incidents thought to result from spam in which monetary damage or information leaks have been caused also continue to take place in Japan. These damages are said to be related to malicious malware. There are no doubt cases in which mail is used to send this directly, or used as a trigger to infect PCs with malware. To maintain email as a fundamental communication tool, a framework for even more robust countermeasures may be necessary.

As an example of this kind of countermeasure framework, in the last IIR (Vol.27) we examined the combination of sender authentication technology centered on DMARC, domain reputations for evaluating authenticated domains, and the feedback loop for raising the accuracy of reputations. The interconnection of each of these elements improves functionality in some respects. Consequently, it would be ideal if all would become more popular, but first we hope to see DMARC become a little more widespread in Japan, as this technology can be implemented easily by senders, and has already been standardized. It is also gaining prominence in the global environment. First, we have investigated and discussed DMARC authentication result ratios in this report to better ascertain the current adoption status. We will continue to conduct a range of studies to contribute to the popularization of effective countermeasure technology.

Author:
Shuji Sakuraba
Mr. Sakuraba is a Senior Engineer in the Application Service Department of the Network Division, IIJ.
He is engaged in the research and development of communication systems. He is also involved in various activities in collaboration with external related organizations for securing a comfortable messaging environment.
He has been a member of M3AAWG since its establishment. He is acting chairperson of the Anti-Spam mail Promotion Council (ASPC) and a member of its administrative group, as well as chief examiner for the Technology Workgroup.
Additionally, he is chairman of Internet Association Japan's Anti-Spam Measures Committee.

*5 Authenticated Received Chain (ARC), draft-andersen-arc-04 (https://www.ietf.org/id/draft-andersen-arc-04.txt).