Research on a Method of Constructing Sender Reputation

2.1 Introduction

Twenty years ago, in January 2004, IIJ joined MAAWG (Messaging Anti-Abuse Working Group), which works on measures to combat spam globally. I participated in the group's founding meeting in April 2004, and I have continued to attend the general meetings since. The group's name has been changed slightly to M3 AAWG[*1], and the scope of its activities has been expanded. The 60th general meeting, marking the group's 20th anniversary, took place in February 2024.

MAAWG technical discussions were initially focused on evaluating and popularizing sender authentication technologies, particularly SPF[*2] and DKIM[*3], to address the fact that it was not possible to accurately identify the sender of an email, which really should be regarded as a flaw in the email system. Those technical discussions continued over the years, with members of M3 AAWG playing a central role in creating technical specifications like DMARC[*4], ARC[*5], and BIMI[*6]. From the start, technical discussions about sender authentication were premised on the idea that, as a next step, we would need to be able to determine if we should accept email using authenticated domain names—in other words, sender reputation. Indeed, the first SPF specification, RFC4408, mentioned domain reputation, and Google and Yahoo in the US have recently been pushing heavily for email senders to get on board with sender authentication to tighten up the way incoming email is handled. Indeed, even in Japan the number of domains supporting DMARC rose sharply after the companies announced these tougher measures. As a visiting researcher at IAjapan, I study domain jp names, and as of February 2024, around a quarter of the domain names used for email had a DMARC record set, indicating a roughly three-fold rise in the proportion of domains with such a record.

The IIJ Research Laboratory conducts research on methods for building sender reputation. This article discusses a paper[*7] published in the journal of the Information Processing Society of Japan. The paper describes a sender reputation construction method and feedback loop. In this article, I focus on the sender reputation aspect of the paper. The paper was also recognized as a specially selected paper by the Information Processing Society of Japan.

2.2 Sender Reputation

DNSBL (DNS Blocklist) has long been used as a mechanism for determining whether to accept email based on sender information. It uses a DNS query to look up the source host's IP address. While the source IP address is not an appropriate way to identify the sender of an email, DNSBL has so far been used because the email address specified in the email headers and during the email delivery process to indicate who the sender is cannot be relied on. With the spread of SPF and DKIM for sender authentication, there are moves to use domain names as authenticated, trustworthy information for determining whether to accept or reject email—this is the concept of domain-name sender reputation.

In addition to domains with a negative reputation from which email should not be accepted, one can also imagine there to be legitimate domain names from which email should be accepted. When quantified, the factors behind this determination result in a reputation score. In simpler terms, reputation can be thought of as the basis for a Block List and Allow List of domain names.

Alongside the rise of sender authentication technology, we have also seen an increase in the volume of spam from actors who have registered their own domain names and properly configured SPF and DKIM. The domain names registered for this type of spam are used as throwaways, so building a Block List based on domain reputation is, unfortunately, not all that effective. It may thus be more effective to use an approach that involves building an

*1 Messaging, Malware and Mobile Anti-Abuse Working Group.
*2 Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1 (RFC7208).
*3 DomainKeys Identified Mail (DKIM) Signatures (RFC6376).
*4 Domain-based Message Authentication, Reporting, and Conformance (RFC7489).
*5 Authenticated Received Chain.
*6 Brand Indicators for Message Identification (Internet-Draft).
*7 Shuji Sakuraba et al., Sender Reputation Construction Method And Feedback Loop Using Sender Authentication, Journal of the Information Processing Society of Japan, Vol. 64, No. 1, pp. 13–23 (2023).

Allow List and combining this with email filters and the like to assess the content of emails not covered by the Allow List. Spam as a proportion of all email has declined from what it was in the past, and so if the bulk of email can be assessed using the simpler and easier methods of sender authentication and sender reputation, then more computing resources can be devoted to assessing the content of email messages.

With this background now laid out, this article describes a method for collecting the domain names of legitimate emails and building reputations on that basis.

2.3 Characteristics of Sender Authentication Technologies

Detailed descriptions of the SPF and DKIM sender authentication technologies can be found in sources such as the Sender Authentication Technology Deployment Manual[*8]. Here, therefore, I focus on parts of the paper related to the method of constructing sender reputation.

SPF authenticates the domain name of an email address representing the sender of an email over the Simple Mail Transfer Protocol (SMTP). On the sending side, an SPF record listing email sender IP addresses and the like is published on the domain's DNS server, and on the receiving end, when an email is received, the receiving server looks up the IP address to determine whether the email is from the correct sender. This mechanism means that implementing SPF on sending servers is relatively easy as all the administrator needs to do is publish an SPF DNS record, and use of SPF thus continues to spread. One problem, however, is that recipients cannot properly authenticate emails when they were sent by an entity other than the original email sender.

With DKIM, the server creates a digital signature from the email header and body for each email sent and affixes this along with other relevant information to the email header. Because it uses an authentication method that does not depend on the route by which an email was delivered, DKIM is not subject to the problems inherent in SPF, such as the inability to properly authenticate forwarded emails. But because sending mail servers need to perform the additional steps of creating the digital signature and adding the DKIM signature information to emails, DKIM has not become as widespread as SPF.

2.4 A Method of Constructing Sender Reputation

Here, I describe a method that uses sender authentication to collect SPF-authenticated domain names from which email should be accepted. In general, it is relatively easy to collect this information in the case of spam because this sort of email is itself unwanted and shouldn't be accepted. The approach has been to collect data for spam filters and block lists by extracting salient characteristics and sender information from the collected spam. When it comes to legitimate emails, however, a challenge is that the messages may contain highly confidential information, making it generally difficult to collect the desired data. And because the information is used to determine whether or not an email should be accepted, erroneously recording an email sender as a spam emitter can cause substantial damage, so accuracy is required when making such entries.

Here, we take forwarded emails to be emails that should be accepted, and describe a method for extracting the senders of forwarded emails and creating a list on that basis.

2.4.1 The Nature of Forwarded Email

Email forwarding is often used as a way of consolidating emails, such as when you use multiple email accounts and want to view them in a single place. This mechanism has long been used in email systems such as the opensource Sendmail, which can be configured to automatically redirect received emails by adding the forwarding address to the .forward file in the user's home directory. Hence, the forwarder's email forwarding settings point to the recipient of the forwarded emails, and so from the forwarded email recipient's perspective, the forwarding email sender can be regarded as an email sender from whom email should be accepted.

*8 Anti-Spam Consultation Center, Japan Data Communications Association, Sender Authentication Technology Deployment Manual (https://www.dekyo.or.jp/soudan/aspc/report.html#dam, in Japanese).

If you can collect a list of such email forwarders, you should be able to construct reputations for email senders from which email should be accepted.

2.4.2 Forwarded Email and Sender Authentication

In basic email forwarding, the email address set by the original sender is used in the envelope-from field[*9], which corresponds to the domain authenticated by SPF. Because of this mechanism, SPF authentication at the forwarded email destination fails. DKIM, meanwhile, does not use the sender's IP address for authentication, so emails to which a DKIM signature has been added can be DKIM-authenticated at the forwarding destination. The results are illustrated in Figure 1. The SDID (Signing Domain Identifier) in the figure is the domain name authenticated by DKIM.

Recently, an increasing number of email-receiving servers are refusing to accept emails that cannot be SPF-authenticated as a means of tightening defenses against email spoofing. For this reason, when forwarding email, some forwarding sources rewrite the envelope-from field to contain the domain name of the forwarding source. When this is done, emails will pass both SPF and DKIM authentication at the forwarding destination. However, the domain names authenticated in this case usually differ. The results are illustrated in Figure 2.

2.4.3 Assessing Forwarded Email Source using Sender Authentication Results

I have explained that there are two email forwarding methods and that the SPF and DKIM authentication results differ across those methods. Thus, we use the sender authentication results to determine whether an email has been forwarded, and collect this as reputational information on the forwarded email senders. First, we identify forwarding sources that do not rewrite the RFC5321.From field when

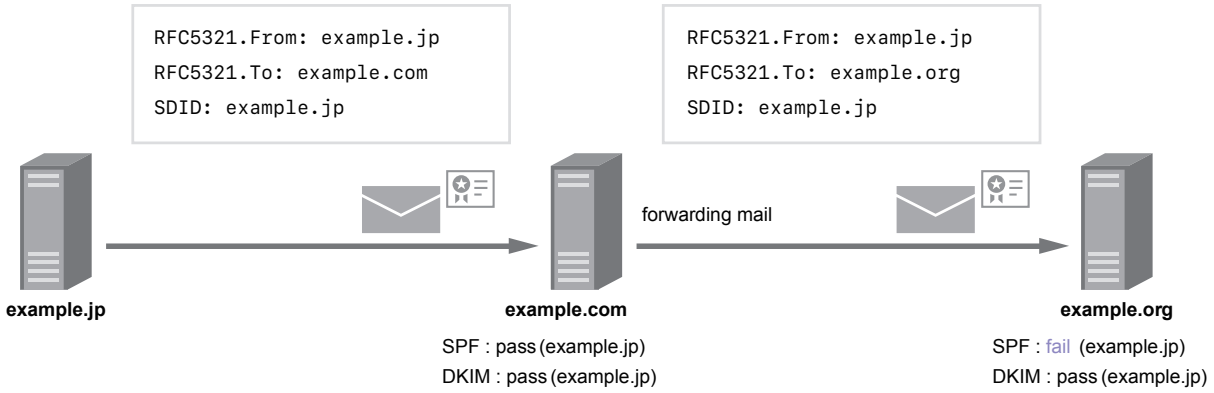

Figure 1: Results of Sender Authentication of Forwarded Email

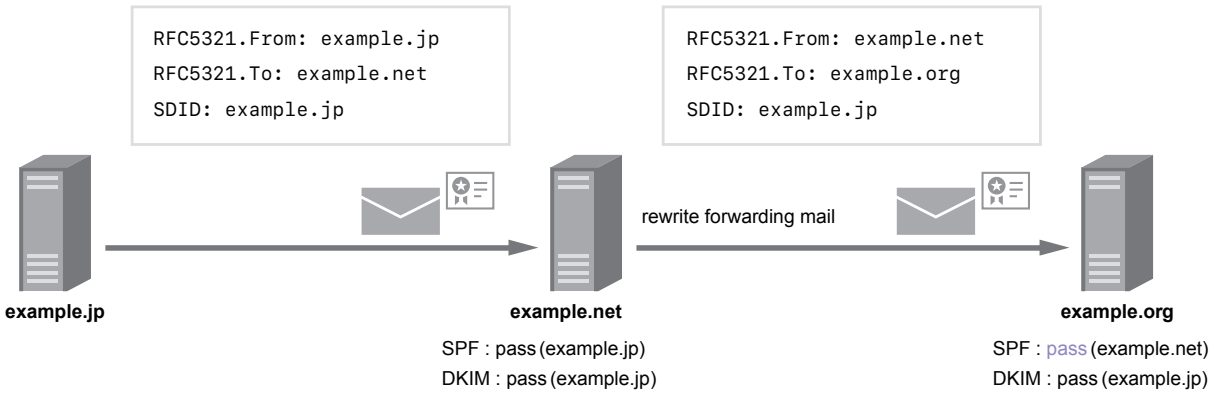

Figure 2: Sender Authentication of Forwarded Email with Sender Information Rewritten

*9 The sender email address according to the email delivery protocol (SMTP). May appear as RFC5321.From, referring to the SMTP RFC.

forwarding emails based on the following authentication criteria, and thus build a list of SPF-authenticated domains from which email should be accepted.

- Sending IP addresses for which SPF authentication fails and DKIM authentication passes
- SPF-authenticated domains that send email from the above IP address and for which SPF authentication passes

With the first criterion, we collect the source IP address of forwarded emails. There can be multiple outgoing mail servers, so to capture a broader range of legitimate email senders, we collect a list of SPF-authenticated domains for which email is sent from the forwarded email's source IP address and for which SPF authentication passes, the objective being to identify the administrative origin of sent emails. This is the second criterion. Both the forwarding IP address and the SPF-authenticated domain name sent from that address constitute sender reputation. Such IP addresses are legitimate sources from which email should be accepted, so any email sent from such addresses, including email that has not been forwarded, is considered acceptable. This makes it possible to use not only IP addresses but also SPF-authenticated domain names for sender reputation.

Next, we identify forwarding sources that rewrite the envelope-from field when forwarding emails according to the following criteria and again collect a list of SPF-authenticated domains from which emails should be accepted.

- Source IP addresses for which both SPF and DKIM authentication passes and the domain names are unrelated
- Of those above, sending IP addresses for which multiple DKIM-authenticated domain names can be obtained via DKIM authentication
- SPF-authenticated domains sent from the above IP addresses for which SPF authentication passes

Outgoing emails that have not been forwarded and that support both SPF and DKIM are expected to be closely related—for instance, usually they will have the same domain name or the same upper domain name. For example, DMARC defines an organizational domain name and assumes that the SPF- or DKIM-authenticated domain name and the sending domain name in the header are the same or have the same organizational domain name. Given this specification, the SPF- and DKIM-authenticated domain names are closely related, even for ordinary email. If the original email sender supports DKIM, and the SPF-authenticated domain name is rewritten when the email is forwarded, it is common for there to be no relation between the original DKIM-authenticated domain name and the SPF-authenticated domain name at the forwarding destination. To identify forwarding sources that rewrite the envelope-from field when forwarding email, we focus on the relationship between SPF- and DKIM-authenticated domain names. To automatically collect these sender IP addresses, we look at emails sent from the same IP address, which pass SPF authentication, and for which the email sender IP address yields multiple DKIM-authenticated domain names. These are determined to be the email forwarding sources. These email forwarding source IP addresses and the SPF-authenticated domain names that they send constitute sender reputation indicating that emails should be accepted.

2.5 Constructing and Verifying Sender Reputationn

To assess the effectiveness of these methods, we constructed sender reputations and applied this to incoming emails. We used the incoming email logs from a real-world email service. This email service performs SPF and DKIM sender authentication when receiving emails, and applies a spam filter to all emails, so the results of these operations are available in the logs. We used the results of this spam filter as the basis for evaluating the judgements made according to sender reputation.

That is, we construct sender reputation from the SPF and DKIM authentication results. Next, we check received email against the sender reputations, compare it with the results of the spam filter, and measure the volume of email classified as non-spam (ham) and as spam.

We constructed sender reputations from around 340 million incoming email log entries for the month of September 2019. At the time, spam accounted for 11.7% of email, the SPF authentication pass rate was 71.1%, and the DKIM authentication pass rate was 38.1%. From these data, we were able to extract 15,169 forwarding IP addresses, 744,660 SPF domain names sent from ordinary forwarding sources, and 11,164 domain names that rewrite the sender domain name when forwarding email.

We applied these sender reputations (indicating that email should be accepted) to the roughly 36 million emails received in the week of October 2019 immediately following the week from which we collected the reputation data (Table 1). We used the same incoming email logs when doing this. Differences between two reputation types are shown below.

(1) Ordinary forwarding sources (IPs) and SPF-authenticated domain names that do not rewrite sender information when forwarding
(2) In addition to (1), sendes (IPs) and SPF-authenticated domain names that rewrite sender information when forwarding

In Table 1, the ham column indicates the proportion of email determined not to be spam by the spam filter for which sender reputation was successfully applied. In other words, this is the true positive rate (TPR). The spam column indicates the proportion of email determined to be spam by the spam filter for which sender reputation was misapplied. That is, this is the false positive rate (FPR). This differs from the meaning of a positive result from an email filter's assessment of spam. Here, a positive result means that email should be accepted, so it is important to be aware of the relationship between true positives and false positives based on reputation.

2.6 Discussion

By assessing email forwarding source using sender authentication technology and constructing sender reputations based on this, we were able to correctly identify around 58% of acceptable email (ham). At the time, the SPF authentication rate was around 70%, so a large portion of that can be attributed to the use of sender reputation. Detecting forwarding sources that rewrite the sender information when forwarding emails and using this to add to sender reputation increased the effectiveness of our method. We were able to increase TPR by over 10 points while holding down the increase in FPR to only 0.25pt. During the period to which we applied our sender reputations, spam accounted for around 9% of received email, so the actual number of false positives was quite low. We also understand, to an extent, why these false positives occurred, so we think it will be possible to further reduce FPR.

This method of constructing sender reputation only uses the results of sender authentication and does not look at the content of email. Even though it is simple

Table 1: Results of Applying Sender Reputation

Reputation	ham(%)	spam(%)
(1)	47.45	3.01
(2)	58.01	3.26

in comparison with common email filtering methods, we still achieved high spam-detection accuracy. We do use the results of DKIM authentication, which does not have a high uptake rate, in identifying the sender of forwarded emails, but the sender does not necessarily have to support DKIM for reputation to be useful, and we only use a few DKIM-authenticated emails to determine forwarding source. So even though DKIM uptake is low, it is possible to construct adequate sender reputations. In constructing and applying sender reputations, we used SPF-authenticated domain names, which are widespread, and if SPF uptake increases further, this should enable accurate determinations about even more emails. If DKIM or DMARC uptake were to increase, we could also consider using those authenticated domain names for the purposes of sender reputation.

The fact that SPF authentication fails at email forwarding destinations has until now been considered a shortcoming of SPF authentication. However, we believe the favorable results we obtain using the method for constructing sender reputation that we describe here—using the features of SPF as a network-based method and DKIM as a digital-signing method—are actually positive for the uptake of SPF.

2.7 Conclusion

With phishing and other forms of spam becoming more sophisticated these days, such that it is hard to tell legitimate emails from malicious ones, we believe that our method, which uses information on sender trustworthiness to determine whether email should be accepted or not, is a significant contribution. The fact that our method of constructing sender reputation does not involve looking at the content of emails also makes it valuable from a privacy perspective. Further, as demonstrated when we tested this method, the fact that it makes it possible to construct sender reputation using, for example, incoming email logs means that it is possible to produce sender reputations geared to the emails that your organization receives, opening up the prospect of greater sorting accuracy. While this method may be inapplicable to a small number of incoming emails, it should make more computing resources available, which could then be used to perform deeper assessments based on the content of those email and so forth.

It has long been the case that emails would be delivered even if you had not deployed some form of sender authentication, and relatively new technologies like DMARC have thus struggled to gain traction. Yet the recent announcement of new countermeasures for incoming email from the likes of Google and Yahoo in the US has prompted more uptake of DMARC, as well as SPF and DKIM, on which it is based. This will make it possible to combat email spoofing while also increasing opportunities to apply measures of domain reputation. We will continue to pursue research relevant to achieving more accurate measures of domain reputation.

Shuji Sakuraba, Ph.D.

Senior Research Engineer, Technology Coordination Office, IIJ Research Laboratory
Dr.Sakuraba is engaged in research and development related to messaging security. He is also involved in various activities in collaboration with related external organizations aimed at bringing about safe and secure messaging environments.He has been a member of M3 AAWG (Messaging, Malware and Mobile Anti-Abuse Working Group) since its establishment. He is the chair of JPAAWG (Japan Anti-Abuse Working Group). 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. He is chairman of the Internet Association Japan's Anti-Spam Measures Committee and a visiting researcher at the association. He is a cooperating researcher at the University of Electro-Communications.