How is E-mail Sender Authentication Used and Misused?

Tatsuya Mori
NTT Service Integration Laboratories
3-9-11 Midoricho Musashino
Tokyo, Japan 180-8585
mori.tatsuya@lab.ntt.co.jp

Yousuke Takahashi
NTT Service Integration Laboratories
3-9-11 Midoricho Musashino
Tokyo, Japan 180-8585
takahashi.yousuke@lab.ntt.co.jp

Kazumichi Sato
NTT Service Integration Laboratories
3-9-11 Midoricho Musashino
Tokyo, Japan 180-8585
sato.kazumichi@lab.ntt.co.jp

Keisuke Ishibashi
NTT Service Integration Laboratories
3-9-11 Midoricho Musashino
Tokyo, Japan 180-8585
ishibashi.keisuke@lab.ntt.co.jp

ABSTRACT

E-mail sender authentication is a promising way of verifying the sources of e-mail messages. Since today's primary e-mail sender authentication mechanisms are designed as fully decentralized architecture, it is crucial for e-mail operators to know how other organizations are using and misusing them. This paper addresses the question "*How is the DNS Sender Policy Framework (SPF), which is the most popular e-mail sender authentication mechanism, used and misused in the wild?*" To the best of our knowledge, this is the first extensive study addressing the fundamental question. This work targets both *legitimate* and *spamming* domain names and correlates them with multiple data sets, including the e-mail delivery logs collected from medium-scale enterprise networks and various IP reputation lists. We first present the adoption and usage of DNS SPF from both global and local viewpoints. Next, we present empirically why and how spammers leverage the SPF mechanism in an attempt to pass a simple SPF authentication test. We also present that non-negligible volume of legitimate messages originating from legitimate senders will be rejected or marked as potential spam with the SPF policy set by owners of legitimate domains. Our findings will help provide (1) e-mail operators with useful insights for setting adequate sender or receiver policies and (2) researchers with the detailed measurement data for understanding the feasibility, fundamental limitations, and potential extensions to e-mail sender authentication mechanisms.

1. INTRODUCTION

The high world-wide popularity of e-mail service lies in its simplicity as indicated by the name of its core protocol Simple Mail Transfer Protocol (SMTP). This simplicity makes it easier to deploy e-mail service over the world, but it also makes it easier for spammers to *abuse* the service. For instance, it is well-known that the majority of spam/phishing messages today originate from *forged* senders. In addition, forging sender identity is still an effective way of avoiding spam filtration. Thus, e-mail sender authentication mechanisms have attracted attention as a promising way of verifying sender identities. Large webmail service providers such as Google have leveraged sender authentication mechanisms to classify authenticated sending domains as either likely legit or spammy [23].

Of the several e-mail sender authentication mechanisms, we focus on Sender Policy Framework (SPF) [28], which is the most used sender authentication mechanism today [16, 26, 12, 23]. According to Refs. [16, 12, 18], roughly 60% of prominent sites publish their DNS SPF records as of July 2011. The adoption of DNS SPF has been widely spread globally as well [26, 18]. Although the adoption of SPF has been growing rapidly, there is not detailed understanding of how e-mail senders set the parameters of SPF on their DNS and how e-mail receivers should retrieve the SPF records on incoming messages. Since SPF is designed as a fully decentralized architecture, it is crucial for e-mail operators to know how other organizations are using or misusing it. In addition, as previous studies have pointed out, the simplicity and high flexibility in creating SPF records has made SPF commonly used throughout the world. However, the high flexibility also makes SPF prone to misconfiguration.

This work addresses the question "*How is DNS SPF used and misused in the wild?*" We analyse large corpus of SPF records published by working domains and e-mail delivery logs collected at an enterprise network. Answers to this question will help provide (1) e-mail operators with useful insights for setting adequate sender or receiver policies and (2) researchers with the detailed measurement data for understanding the feasibility and fundamental limitations of e-mail authentication mechanisms in order to estimate the potential needs for the extension of the mechanisms, such as Sender-Rewrite Scheme (SRS) [3].

To the best of our knowledge, this is the first extensive study to address the question. We target both *legitimate* and *spamming* domain names and correlate them with various data sets including the e-mail delivery logs collected from a medium-scale enterprise network and various IP reputation lists. On the basis of the analysis of the data sets, this work investigates how many domains properly publish SPF records, how many e-mail senders/messages pass/fail the SPF authentication test, and what types of e-mail messages are sent from the senders that pass/fail the test. We also present how current spammers leverage the SPF mechanism in an attempt to pass a simple SPF authentication test.

The remainder of this paper is structured as follows. Section 2 briefly overviews e-mail authentication mechanisms. Section 3 de-

CEAS '11 September 1-2, 2011, Perth, Western Australia, Australia

scribes the data sets we use in this work. In Section 4, we present our findings on how SPF is used and misused. We also present typical patterns of misconfiguration in publishing SPF records. In Section 5, we discuss related studies and how they compare to ours. Finally, Section 6 concludes our work.

2. OVERVIEW OF DNS SPF

This section briefly overviews the DNS SPF. We present the technical background, key technical ideas, and some known limitations. Several e-mail sender authentication mechanisms, namely SPF [28], Sender ID [17], and DomainKeys Identified Mail (DKIM) [6], have been proposed. The idea of e-mail sender authentication was first proposed back in 1997. After several years, it was developed into the stable specification of what is known as SPF today [27]. Sender ID, which adds a few modifications to SPF, was converged into a specification from SPF and other old proposals. DKIM was proposed in 2005 as a way of allowing an e-mail sender to electronically sign e-mail messages with public-key cryptography so that it could be verified by e-mail recipients.

All these sender authentication mechanisms aim to provide a way of verifying senders. As studies such as Refs. [16, 26, 18, 15] have revealed, SPF is the most widely used mechanism at the time of writing. As we show later, roughly half of the top legitimate domains have adopted SFP. Therefore, we focus on analyzing of SPF. Since Sender ID can be analyzed with the same approach, we also look at it. Note that our data set does not provide us with the DKIM-Signature field, which is essential for studying whether DKIM is used or misused. That is another reason we do not analyze DKIM in this work. Although the deployment of DKIM is much less than that of SPF, we will study its effectiveness and limitations in future work. For the global deployment of DKIM (regardless of the usage), readers may refer to the existing surveys [15, 26, 18].

We now review how SPF works. SPF works on top of the DNS. The key idea is that an administrator of a domain explicitly specifies which hosts are allowed/prohibited to send e-mail message using an sender e-mail address of the domain. Upon receiving an e-mail message, a recipient server first extracts domain name from sender e-mail address and looks up SPF record of the domain name. The recipient server determines whether the sender's IP address is authenticated to send e-mail messages using the domain by checking the SPF record.

Figure 1 shows an example of a SPF record published by an administrator of domain "`ietf.org`". The first string "`v=spf1`" specifies the version of SPF. "`ip4`", "`ip6`", and "`all`" represent "mechanisms", which specify sender's addresses. The mechanism "`all`" specifies all IP addresses *not* matched by the prior mechanisms. Finally, "−", which is placed at "all", represents a "qualifier", which, combined with each mechanism, specifies how messages should be treated by a receiver. SPF has the following four qualifiers:

- +: PASS. The mail should be accepted. (can be omitted)
- ?: NEUTRAL. No policy.
- ~: SOFTFAIL. Between NEUTRAL and FAIL.
- −: FAIL. The mail should be rejected.

In the example of `ietf.org`, senders with the IPv4 and IPv6 addresses specified with the `ip4` and `ip6` mechanisms are explicitly allowed to send e-mail messages using the domain, e.g., "foo@ietf.org" (note that a qualifier "+" is omitted for the two mechanisms), and all other IP addresses are *not* allowed to send e-mail message with the "`-all`" policy set in the SPF record. For more detailed information, refer to RFC 4408 [28].

```
"v=spf1 ip4:64.170.98.0/26
ip4:64.170.98.64/28 ip4:64.170.98.80/28
ip4:64.170.98.96/29 ip4:208.66.40.224/27
ip6:2001:1890:1112:1::0/64 -all"
```

Figure 1: SPF record of domain `ietf.org` in Apr 2011.

SPF has some known problems to be addressed. Since SPF is designed as decentralized architecture, global collaboration is required to make it really effective. As Edelman [10] indicated, incentivizing legitimate e-mail operators to deploy SPF would be one of the key success factors. The good news is the fraction of legitimate domains that adopt SPF is steadily growing [26, 18]. Another problem is that it is prone to misconfiguration because of its high degree of freedom. The final problem is that SPF is known to not work with portable e-mail addresses and forwarding services, where e-mail messages are generally delivered via e-mail servers owned by different organizations.

The key contribution of this work is the understanding of how these problems are increasingly appearing today. Such understanding is useful for defining a new research direction based on the emergence of these problems. Note that SPF is not a single solution as a sender authentication mechanism. An end-to-end authentication mechanism such as DKIM can complement the lack of an SPF mechanism in some cases (and vice versa). Our aim is to reveal the actual lack to be addressed through a pragmatic approach. Again, such information is useful in refining the sender authentication mechanisms based on observation of the existing problems.

3. DATA DESCRIPTION

This section describes the three categories of data sets used in this study: lists of domain names, SMTP delivery logs, and IP reputation lists.

3.1 Domain Names

To investigate how SPF is deployed in the Internet, it is crucial to collect representative domain names used for e-mail addresses. This work looks at two typical classes, one for legitimate organizations and the another for spammers. That is, we collect *legitimate* and *spamming* domain names.

3.1.1 Legitimate domains

To compile lists of *legitimate* domain names, we used two data sources. The first source was shown to be reasonably representative by Eggert [16]. Eggert provides a research experiment site to study the current deployment of SPF in the Internet. Alexa Top 500 Global Sites [5] was used for the experiment as a source of prominent domain names. The list contains 500 distinct domains that span 63 of distinct top-level domains (TLDs), mostly for ".com". In addition to the Alexa list, we use a commercial domain list [**?**] used for popular free e-mail service providers around the world; e.g., "gmail.com", "hotmail.co.uk", and "mail.ru". The list contains 6,202 distinct domain names that span 83 of TLDs, mostly for ".com" and ".la". Another list for legitimate domains comes from our locally collected SMTP logs. We picked a domain, i, if the number of messages from the domain, m_i, satisfied $m_i \geq m^*$ and if the number of legitimate messages from the domain, h_i, satisfied $h_i/m_i > \theta^*$. In this work, we set the thresholds $\theta^* = 0.95$ and $m^* = 100$. We note that the selection of the thresholds did not significantly affect the succeeding results. The list contains 1,675 domains. In this work, the above two lists will be referred as

ALEXA and local good domains (LGD), respectively.

3.1.2 Spamming domains

As the lists of *spamming* domain names, we used a domain blacklist and the locally compiled bad domains. For the domain blacklist, we used Spamhaus DBL [24], which is a database of domains found in spam messages. According to Spamhaus, the DBL is managed as a "zero false-positive" list, safe to use by production mail systems to reject emails that are flagged by it. The list contains 66,357 distinct domain names that span 63 TLDs, mostly for ".com", ".info", and ".ru". Similar to the legitimate domains, we compiled the list of local bad domains from our SMTP logs, i.e., we picked a domain, i, if $m_i \geq m^*$ and $s_i/m_i > \theta^*$, where s_i is the number of spam messages from the domain i. We used the same thresholds, $\theta^* = 0.95$ and $m^* = 100$. The list contains 45,123 distinct domains. In this work, the above two lists will be referred to as DBL and LBD local bad domains (LBD), respectively.

3.2 SMTP logs

We collected SMTP delivery logs from an enterprise network. The logs were collected on commercial anti-spam software, which applies content-based filtering and assigns a spam probability score to each incoming message. Each record in the logs consist of the domain name and IP address of the sender and the score of the message. We note that a large fraction of SMTP connections originating from bots are filtered at the pre-acceptance stage because a greylisting mechanism [2] is deployed on top of the e-mail delivery system. Greylisting is a mechanism that temporarily rejects e-mail messages from a sender which has not previously been seen. Greylisting is effective because if an e-mail is rejected, a spammer will likely *not* retransmit it since spammers cannot afford the time and resources to retry thousands of bounced messages. In this work, greylisted connections will *not* be used because SMTP logs of these connections do not have score information, however, we note that majority of greylisted connections are potential spam messages.

From the data set we collected, we can analyze whether a message that passed/failed the SPF test is spam or a legitimate message. Software-based filtering is error-prone and thus could affect the classification of individual e-mail messages, but we expect that the derived statistics will not be affected. For instance, let ϵ be the error probability of a message classifier. Assume that a host sent x spam messages out of y total messages. We can define a host as a spammer if $x/y \geq \theta^*$ and $y \geq \widehat{y}$, where θ^* and $\widehat{y}$ are the thresholds. Given $y = \widehat{y}$ and θ^*, the probability that a legitimate host is misclassified as a spammer is roughly $\epsilon^{\theta^* \widehat{y}}$, which quickly converges to zero for a fairly large $\widehat{y}$, e.g., given $\epsilon = 0.05$, $\widehat{y} = 10$ and $\theta^* = 0.9$, the misclassification probability becomes $0.05^9 = 1.9 \times 10^{-12}$,

The SMTP logs were collected for a month and contained 3,974,819 delivered messages originated from 142,007 distinct e-mail sender IP addresses. For each domain name, we performed DNS SPF lookup and collected the results at the time of measurement.

3.3 IP reputation lists

This work uses publicly available IP reputation lists to understand the characteristics of senders that pass/fail SPF authentication. We use one whitelist and three blacklists, namely DNSWL [9] for whitelist and Spamhaus PBL, XBL, and SBL for blacklists [4]. The three blacklists collect IP addresses for end-hosts (bot) assigned dynamic IP addresses (PBL), open proxies/relays (XBL), and spam gangs (SBL), respectively. All the lists were obtained at the time of measurement.

4. SPF IN THE WILD

Our goal in this paper is to understand how SPF is used and misused in the wild by using the data set we described in the previous section. While previous surveys on the the SPF deployment, e.g., [16, 18, 26, 12], simply check whether a domain publishes DNS TXT resource records (RRs) that contain SPF RR, this does not tell us whether the published records are valid or invalid. We do not know who uses or misuses SPF or whether the authenticated messages were legitimate or spam. To answer these fundamental questions, we first investigate the deployment of SPF from a global view. Next, we focus on the properties of messages and senders that passed or failed the SPF authentication test from a receiver view. Finally, we investigate a special usage of SPF and show how it is used.

4.1 Deployment of SPF: Global View

This section analyzes the deployment of SPF by using globally collected domain lists. We also analyze locally collected domain lists for comparison purposes.

4.1.1 Adoption of SPF

We first investigate how SPF is deployed in the two domain groups, legitimate and spamming domains. To test the validity of published SPF records, we developed a SPF analyzer that parses SPF records, processes further lookups, excludes errors, and checks the authentication. We applied the SPF analyzer to the domain names we described in the previous section.

Table 1 shows the adoption of SPF in the two domain groups. Note that we allowed invalid SPF records, e.g., `v=spf2.0`, here. We see roughly half of *legitimate* domains have adopted SPF. These numbers agree with other surveys [16, 12, 18], except the data set FREE. We note that some e-mail service providers deploy a large number of MX domains without publishing SPF records for them. Thus, the result for the list FREE could be biased and hence may not reflect the fraction of organization that adopt SPF.

A fairly large fraction of spamming domains, i.e., 20+ % of spamming domains, have adopted SPF as well. It is well known that spammers are some of the "early adopters" of SPF technology [19] and this usage by spammers is not an intended one because if all spammers adopted SPF, authenticating the senders of e-mail messages with SPF would be unnecessary. As Esquivel et al. [11] recently demonstrated, a significant volume of spam messages originates from senders associated with bad domains with SPF records. Thus, many spam messages may *pass* the sender verification test, which means using SPF authentication as spam filtering could cause false negatives. In the next section, we will analyze e-mail logs in-depth to determine how spammers leverage SPF. In particular, we will apply the sender policy, which is recorded in the SPF records, to emulate how many legitimate/spam messages can be classified / misclassified by the SPF mechanism.

Table 2 shows the adoption of SPF variants by applying strict check. We also investigate whether a SPF/Sender ID RR explicitly contains IPv6 addresses. Of the SPF-enabled domains shown in Table 1, majority of domains have valid SPF RRs and 5–10 % of domains have valid Sender ID RRs. The majority of Sender ID-enabled domains adopt both SPF and Sender ID. These observations apply to the both legitimate and spamming domains. Although the number is not large, we see that spammers are some of the "early adopters" of IPv6-enabled SPF. As measurement study done by RIPE [21] revealed, spam messages are being disseminated over IPv6. In the not-so-distant future, we may need to be ready for IPv6-aware spammers who can leverage vast amount of IP address space as a source of spam.

Table 1: Adoption of SPF: N and S are the number of total domains and SPF-enabled domains, respectively.

Domain set	N	S (%)
Legitimate domains		
ALEXA	500	289 (58%)
FREE	6,202	1,299 (21%)
LGD	1,675	754 (45%)
Spamming domains		
DBL	66,356	24,522 (37%)
LBD	45,123	9,590 (21%)

Table 2: Adoption of (1) SPF variants and (2) IPv6 in SPF RR.

Domain set	SPF	Sender ID	both	IPv6
Legitimate domains				
ALEXA	289	31	31	1
FREE	1,298	5	4	0
LGD	754	33	33	9
Spamming domains				
DBL	24,439	801	798	0
LBD	9,560	190	170	2

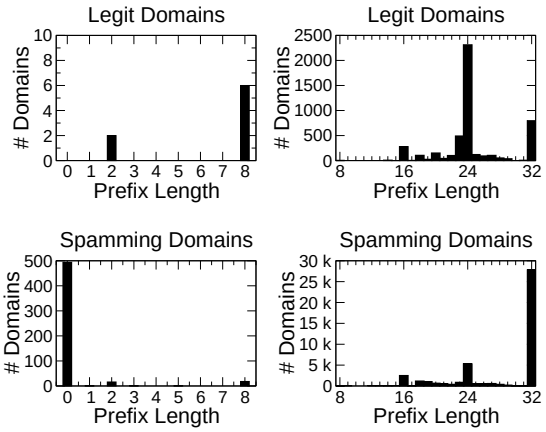

Figure 2: Distributions of prefix lengths declared in valid SPF records. Legitimate domains (top) and Spamming domains (bottom).

To summarize, the SPF authentication mechanism is now used by both legitimate (roughly 50 %) and spamming domains (roughly 20+ %). Detailed observation suggests the existence of IPv6-aware spammers who can potentially leverage a large address space as sources of spam.

4.1.2 Distributions of prefix lengths published in SPF

We then investigate the distributions of prefix lengths published in SPF. Since the size of the prefix length published in a SPF record is related the way how an organization intends to send e-mail messages from its domains, we expect to see their tactics in publishing SPF records.

Figure 2 shows the distribution of *minimum* prefix lengths published in the SPF records for each group of domains. For simplicity, we merge the two domain lists for each group. Notice that for longer prefixes, i.e., larger than "/16", we see concentration on "/16", "/24", and "/32" for both groups. This result would come from the addressing conventions used in ISPs and other organizations. Through the careful analysis, we found that the two legitimate domains with the prefix length of "/2" are likely a misconfiguration. In general, legitimate domains tend to use a fairly long prefix length, e.g., "/24", and do not use prefix lengths that are smaller than their BGP prefix lengths.

In contrast, the minimum prefix length distribution for spamming domains exhibits a weak concentration on "/0" and a strong concentration on "/32". The intrinsic characteristics of prefix lengths for spamming domains can be interpreted as follows. First, spammers may want to publish the SPF records for a very large IP address space, e.g., "/0", which is the entire IPv4 space, so that any bots with dynamic IP addresses are allowed to send spam messages using their domain. They may also want to publish the SPF record for dedicated servers as the sources of their spamming infrastructure. In the following, we will see another motivation for spammers to leverage SPF.

Let D be a set of spamming domains that publish SPF RR with minimum prefix length of "/32". The number of distinct "/32" IP addresses published by the spamming domains, D, is 10,702, which span 1,973 of distinct AS numbers. Majority of these 10,702 IP addresses are valid (non-bogon) global IP addresses (99.6%) and less than 5 % of the addresses are covered with the blacklists we used. In our SMTP logs, we observed 342 IP addresses from the 10,702 IP addresses. Of these, 142 addresses sent purely spam and 135 addresses sent purely ham. This observation indicates the former servers are used for dedicated spamming servers and the latter servers are used to make the domain look legitimate. Actually, of the 10,702 addresses, 797 addresses were listed on DNSWL and majority of the 797 addresses were owned by popular web e-mail service providers such as Google.

To summarize, the prefix lengths published in SPF concentrate on "/16", "/24", and "/32" for both legitimate and spamming domains. Spammers may have three tactics in using SPF: (1) for sending spam from entire IP space, i.e., using botnet, (2) for sending spam from particular IP address, i.e., using dedicated servers, and (3) for making them look legitimate by including legitimate IP addresses in their SPF records.

4.2 Deployment of SPF: Receiver View

In addition to the global view on the deployment of SPF, this work provides a receiver view by correlating SPF records and the SMTP logs collected from an enterprise network. Although the analysis is a case study, we believe the obtained insights are meaningful in addressing the fundamental question; "*How many legitimate/spam messages could be correctly classified or misclassified by the SPF authentication, which is widely used today?*" We analyze how messages and senders were authenticated by the SPF mechanism. We also analyze how a special but valid SPF record is used and misused.

4.2.1 Authenticated Messages

We first investigate how the messages were authenticated by SPF. Table 3 shows the breakdown of classified messages and results of SPF authentication. Of 1.6 M spam messages, 391 K (24 %) were originated from domains with valid SPF records. Of the 391 K spam messages with valid SPF RRs, 45 % were authenticated (PASS) by SPF while 50 % were either FAIL or SOFTFAIL. Thus, a large portion of spam messages from SPF-enabled spammers can pass the SPF authentication.

Similarly, of 2.3 M legitimate messages, 1.3 M (54 %) were originated from domains with valid SPF records. Of the 1.3 M legitimate messages with valid SPF RRs, 84 % were authenticated by SPF (PASS) while 14 % had sender policies of "FAIL" or "SOFTFAIL". These two policies indicate that a legitimate message from

Table 3: Breakdown of messages and results of SPF authentication.

Auth	Total	# of Spam	# of Legitimate	# of Other
PASS	1,273,994	174,380	1,073,666	25,948
NEUTRAL	48,956	22,750	24,933	1,273
SOFTFAIL	271,253	108,517	156,097	6,639
FAIL	110,867	85,152	24,574	1,141
NONE	2,269,749	1,175,843	1,065,464	28,442
TOTAL	3,974,819	1,566,642	2,344,734	63,443

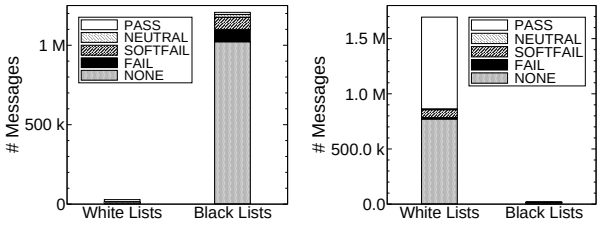

Figure 3: Results of SPF authentication for senders listed on whitelists and blacklists: # of spam messages (left) and # of legitimate messages (right).

an address of the domain can be *rejected* or *marked as a potential spam* by a receiver.

We then investigate the breakdown of classified messages for each class of senders, "whitelists" and "blacklists". We use the IP reputation lists shown in the previous section. We also use manually compiled IP reputation lists, which are derived from our SMTP logs. We set the fraction of spam/legitimate messages to $\theta = 0.95$ and the minimum number of messages to $m = 100$ as thresholds for each IP address. Figure 3 shows the results. As we have seen, while a large fraction of spam messages originated from senders listed on the blacklists do not have valid SPF records. Even though they had, majority of them failed or softfailed. However, we note some spam messages originated from servers listed on the blacklists actually pass the SPF test. For legitimate messages originated from senders listed on the whitelists, roughly half of messages passed the SPF authentication, while roughly 5+ % legitimate messages failed or softfailed.

To summarize, some of spammers not only publish SPF records, but also leverage SPF to make their messages look authenticated (legitimate) ones. 5+ % legitimate messages originated from legitimate senders can fail or softfail by the SPF mechanism. Therefore, e-mail receivers need to be careful in using sender's policies recorded in SPF records.

4.2.2 Authenticated Senders

Next, we investigate how each class of e-mail senders are authenticated by SPF. Table 4 shows the breakdown of sender IP addresses with respect to the number of distinct IP addresses. That is, if a message coming from an IP address listed in DNSWL pass SPF, we create a tuple {"PASS", "DNSWL"}. The number of distinct IP addresses associated with each tuple is shown in the table. Note that the numbers in the table include duplications. That is, an IP address could be counted as both PASS and FAIL depending on the domain name used with the IP address.

Notice that for whitelists (DNSWL and Local WL), roughly half of senders use domains with SPF record and majority of senders with SPF record pass the SPF authentication. As we have seen, this would be an expected result. We also see that roughly 15 % of SPF-enabled legitimate senders failed or softfailed the SPF authentication. As we have discussed in section 2, these senders are likely due to forwarding services or end-users using portal e-mail addresses.

For blacklists (Bot, Spam Gang, Open Proxy, and Local BL), majority of senders do not send messages from domains with SPF record, however, there are a few senders that pass the SPF authentication. We note that the fractions are bit different in the spam gang BL, where senders are likely using dedicated spamming infrastructure such as hosting servers. Of the senders listed in spam gang BL, 23 % had SPF records. Of the 23 % senders, 33 % passed the SPF authentication.

To summarize, majority of legitimate e-mail senders that use SPF correctly pass the SPF authentication. However, there are non-negligible volume of legitimate senders that fail or softfail SPF test. While majority of spam sources do not use SPF, some spammers such as those from spam gang likely adopt it.

4.3 Empty SPF

Finally, we turn our attention to a special but valid SPF record, i.e., `"v=spf1 -all"`, which mean "*no senders are allowed to send an e-mail with the domain*". Following the notation in Ref. [12], let these records be "*Empty SPF*". According to the survey on the top 1 M domains shown in [12], of the 32.5 K SPF-enabled domains with "-all" qualifiers, 1.9 K (6 %) domains have "*empty SPF*" records.

It would be reasonable to assume that "*empty SPF*" is used by a legitimate site that wants to explicitly demonstrate that it never sends mail. As of April 2011, such an example in the Alexa data set is "ibm.com"[1].

Table 5 presents how empty SPF is used in the SMTP logs. #Domains is the number of observed domains with empty SPF and #IP is the number of IP addresses that sent messages from these domains. #Ham and #Spam are the number of ham and spam messages originated from the IP addresses with the domains, respectively. Although the numbers observed are not large, we can see that empty SPF is more used for spamming. It is somewhat surprising that a non-negligible number of spam messages originate from a domain with "*empty SPF*". Of the 980 domains with "*empty SPF*", spam messages were sent from 950 domains. We further looked up the NS records of these domains and found that a majority of the 950 domains are authorized by the top 8 name servers, owned by a few hosting companies. While using an "*empty SPF*" record does not make sense from the view point of spammers, we conjecture that this observation reflects an automated process of domain management with some misconfiguration.

5. RELATED WORK

There have been several studies looking at the effectiveness of sender authentication mechanisms [23, 13, 8, 11, 20]. In [23], Taylor showed how Google's *Gmail* [1] leverages SPF and DomainKey in calculating the reputation of senders. The work revealed the effectiveness of sender's domain in compiling good reputation lists. Herzberg [13] developed an e-mail filtering system that combines IP reputation, sender authentication mechanism, and content filtering. Dalkilic et al. [8] studied empirically the effectiveness of *best-guess* SPF [25], which complements SPF records by guessing the IP address range of a domain's MTAs. Esquivel et al. [11] showed that DNS SPF records can be used to compile effective custom IP reputation lists. Qian et al. [20] proposed a way of clustering e-mail

[1]Other sub domains such as "us.ibm.com" have a SPF record with some explicit IP addresses.

Table 4: SPF authentication results and breakdown of sender IP addresses (number of distinct IP addresses).

Auth	Total	DNSWL	Local WL	Bot	Spam Gang	Open Proxy	Local BL
PASS	7.53 K	1.87 K	3.30 K	94	26	32	150
NEUTRAL	2.08 K	198	184	1.22 K	2	452	785
SOFTFAIL	4.51 K	255	451	2.79 K	39	1.09 K	1.62 K
FAIL	3.18 K	134	142	2.31 K	18	827	1.44 K
NONE	131 K	1.65 K	3.56 K	112 K	258	41.0 K	20.2 K
TOTAL	142 K	3.43 K	6.75 K	115 K	336	41.9 K	20.6 K

Table 5: Statistics of senders and messages associated with domains with empty SPF.

#Domains	#IP	#Ham	#Spam
980	1,084	327	10,110

Table 6: Number of messages for each qualifier, which is declared in the SPF records of the message that *failed* the SPF authentication.

	PASS	NEUTRAL	SOFTFAIL	FAIL
spam	245	21588	111299	74206
legit	28	24382	165224	24380
other	8	1206	6979	868

senders. They used SPF records as a part of feature vector for each sender.

While the aforementioned studies focused on the positive side of sender authentication mechanisms, to the best of our knowledge, a very few research papers have focused on the negative side. In [14], Herzberg discussed the limitations of e-mail sender authentication mechanisms in a qualitative manner. In particular, the paper discussed the risks of DNS poisoning and suggest countermeasures. In this work, we use empirical approach in addressing both positive and negative sides of sender authentication mechanisms.

6. CONCLUSIONS

We aimed to determine how SPF is used and misused in the wild. For this, we performed an extensive study based on multiple data sets. We first collected both *legitimate* and *spamming* domain names. We then correlated them with other data sources including the e-mail delivery logs collected from a medium-scale enterprise network and several IP reputation lists, i.e., whitelists and blacklists.

On the basis of the analysis of the data sets, this work presented the adoption and usage of SPF records today. We also presented how many e-mail senders/messages pass/fail the SPF authentication test, and what types of e-mail messages are sent from the senders that pass/fail the test. Our key findings are summarized as follows: (i) 50 % of legitimate domains and 20+ % of spamming domains adopt SPF today, (ii) spammers likely publish SPF with three different tactics: using entire IP address space (botnets) as spam sources, using dedicated spamming servers, and for making them look legitimate by including legitimate sender IP addresses, (iii) 10+ % of spam messages actually pass the SPF authentication check, (iv) 5+ % of legitimate messages originating from legitimate senders could be rejected or marked as potential spam by the SPF authentication, and (v) some spammers are likely misusing "Empty SPF", which does not make sense for them because it means any messages should be rejected by SPF.

We believe these findings help to provide (1) e-mail operators with useful insights for setting adequate sender or receiver policies, and (2) researchers with the detailed measurement data for understanding the feasibility, fundamental limitations, and potential extensions of the e-mail sender authentication mechanism. We will give some examples below. First, our analysis on the prefix lengths distribution can be used to make a rule-of-thumb for detecting the "self-certifying spammers", which actively leverage the SPF mechanism in an attempt to pass a simple SPF authentication test. E.g., if a domain publishes a SPF record that contains IP prefix with 0-bit mask (i.e., entire IP space), it is likely that the domain is published by spammers. Second, in order to distinguish the domains published by such "self-certifying spammers" from legitimate domains, leveraging the domain reputation systems such as [22, 7] is a promising way. That is, if a message originating from a domain passes the SPF test, and the domain is listed in the domain blacklist, we can classify the message as spam with high confidence. Moreover, the IP addresses associated with the black domain can be marked as potential sources of spam messages. Finally, our findings suggest e-mail operators to re-check their DNS SPF records carefully so that legitimate messages would not be misclassified as potential spam messages.

Our future work includes more fine-grained analyses on SPF authentication, i.e., the Return-Path test and HELO test. Analyzing such information requires access to the payload of SMTP connections. We will work on this with our spam trap server. As our study revealed, SPF is not only used as expected, but also misused in the wild. Developing effective extensions for coping with the misusage of the existing sender authentication mechanisms, on the basis of the findings in this work, is for our future work.

Acknowledgements

We thank Mitsuhiro Shigematsu, Masashi Mitsuda, and Kanako Nozue for their assistance in collecting the data sets used in this work.

7. REFERENCES

[1] Gmail. http://mail.google.com.
[2] Greylisting. `http://www.greylisting.org/`.
[3] Sender rewriting scheme. `http://www.openspf.org/SRS`.
[4] The Spamhaus Project. `http://www.spamhaus.org/`.
[5] Alexa. The top 500 sites on the web. `http://www.alexa.com/topsites`.
[6] E. Allman, J. Callas, M. Delany, M. Libbey, J. Fenton, and M. Thomas. DomainKeys Identified Mail (DKIM) Signatures. RFC 4871 (Proposed Standard), May 2007. Updated by RFC 5672.
[7] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a dynamic reputation system for dns.

In *Proceedings of the 19th USENIX conference on Security*, USENIX Security'10, Berkeley, CA, USA, 2010. USENIX Association.

[8] G. Dalkilic, D. Sipahi, and M. H. Ozcanhan. A simple yet effective spam blocking method. In *SIN '09: Proceedings of the 2nd international conference on Security of information and networks*, pages 179–185, New York, NY, USA, 2009. ACM.

[9] DNS Whitelist. `http://www.dnswl.org/`.

[10] B. Edelman. Priced and Unpriced Online Markets . *The Journal of Economic Perspectives*, 23(3):21–36, 2009.

[11] H. Esquivel, T. Mori, and A. Akella. On the Effectiveness of IP reputation for Spam Filtering. In *Proceedings of the Second International Conference on Communication Systems and Networks (COMSNETS'09)*, Jan 2010.

[12] Greg Hewgill. SPF -all Domain Survey. `https://spf-all.com`.

[13] A. Herzberg. Combining authentication, reputation and classification to make phishing unprofitable. In *Proc. Emerging Challenges for Security, Privacy and Trust, 24th IFIP TC 11 International Information Security Conference*, pages 13–24, May 2009.

[14] A. Herzberga. DNS-based email sender authentication mechanisms: A critical review . *Computer & Security*, 2010. (in press).

[15] Lars Eggert. DKIM Deployment Trends. `https://fit.nokia.com/lars/meter/dkim.html`.

[16] Lars Eggert. SPF Deployment Trends. `https://fit.nokia.com/lars/meter/spf.html`.

[17] J. Lyon and M. Wong. Sender ID: Authenticating E-Mail. RFC 4406 (Experimental), Apr. 2006.

[18] Online Trust Aliance. Email authentication resources & compliance reports. `https://otalliance.org/resources/authentication`.

[19] Paul Roberts. Spammers using sender authentication too, study says. `http://www.infoworld.com/d/security-central/spammers-using-sender-authentication-too-study-says-147`, 2004.

[20] Z. Qian, Z. M. Mao, Y. Xie, and F. Yu. On Network-level Clusters for Spam Detection. In *Proceedings of 17th Annual Network & Distributed System Security Symposium (NDSS)*, 2010.

[21] RIPE Labs. Spam over IPv6. `http://labs.ripe.net/content/spam-over-ipv6`, 2010.

[22] K. Sato, K. Ishibashi, T. Toyono, and N. Miyake. Extending black domain name list by using co-occurrence relation between dns queries. In *Proceedings of the 3rd USENIX conference on Large-scale exploits and emergent threats: botnets, spyware, worms, and more*, LEET'10, Berkeley, CA, USA, 2010. USENIX Association.

[23] B. Taylor. Sender reputation in a large webmail service. In *The Third Conference on Email and Anti-Spam (CEAS)*, July 2006.

[24] The Spamhaus Project. The Domain Block List (DBL). `http://www.spamhaus.org/dbl`.

[25] The SPF Project. FAQ/Best guess record. `http://new.openspf.org/FAQ/Best_guess_record`.

[26] WIDE antispam WG. Measurement Results on Deployment Ratio of Domain Authentications. `http://member.wide.ad.jp/wg/antispam/stats/index.html`.

[27] William Leibzon and Julian Mehnle. SPF: History/SPF-2003. `http://www.openspf.org/History/SPF-2003`.

[28] M. Wong and W. Schlitt. Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1. RFC 4408 (Experimental), Apr. 2006.