A Spoofed E-mail Countermeasure Method by Scoring the Reliability of DKIM Signature using Communication Logs

Kanako Konno*, Kenya Dan*, and Naoya Kitagawa†

*Department of Computer and Information Sciences, Tokyo University Agriculture and Technology, Tokyo, 184-8588, Japan
†Institute of Engineering, Tokyo University Agriculture and Technology, Tokyo, 184-8588, Japan

E-mail:*hogehoge@st.go.tuat.ac.jp, †nakit@cc.tuat.ac.jp

***Abstract*—Recently, Spoofed e-mail that cannot be determined by the visual confirmation has been increasing. DKIM is one of the most famous methods as countermeasure of spoofed e-mail. However, since DKIM allows third parties to sign, it is difficult to be determined whether or not the DKIM signature added to an e-mail is legitimate. Moreover, the method using data of DKIM signature domain passed DKIM verification is proposed, though this method do not take the reliability of that data into account. In order to solve this problems, we propose a method to measure the suspicion of pairs of e-mail sender's domain and DKIM signature domain by analyzing and scoring past communication logs focusing on the existence of e-mail delivery.**

***Keywords*—Spoofed e-mails; DKIM; Sender Domain Authentication.**

I. Introduction

E-mail is one of the most utilized services all over the world as a convenient way to communicate each other or notify various information from companies to customers. However, it has serious problem that the number of spoofed e-mail has been increasing dramatically.Spoofed e-mail is abused by attackers in order to steal personal and/or sensitive information or send malicious programs such as computer viruses.

The damage of business e-mail scams that is directly attribute to spoofed e-mails have been rapidly increasing. According to the newest statistics report of FBI, the number of victim companies is 22,314, and the total financial damage is 3.1billion US dollar from October 2013 to June 2016 [1].

Although a large number of anti-spam methods have been proposed over the years such as Bayesian Filter [2] [3], SpamAssassin [4] [5] and so on, these methods are not a direct measure to spoofed e-mail.

On the other hand, sender domain authentication is a method for directly countermeasure of spoofed e-mail. DomainKeys Identified Mail(DKIM) [6] is one of the most widely used methods among sender domain authentication. In DKIM, the sender adds the electronic signature generated from e-mail header and body to the e-mail, and the recipient verifies the reliability of the e-mail by collating it. DKIM allows to sign by third party, so it is possible for a malicious sender to lead a legitimate signature destination. This is reason why spoofed e-mail passes DKIM verification frequently.

In order to solve the problem of DKIM about third party signature, DMARC is proposed by M. Kucherawy, and E. Zwicky [8]. DMARC is the method that uses SPF method and DKIM method. In addition, there is a concept called "alignment" in DMARC. Because of this concept, the problem about third party signature can be solved. However, the penetration rate of DMARC in Japan is 21.1% [9], which means DMARC is not popular method.

At the present time, a method to verify the reliability of DKIM signature by checking past combination of e-mail sender's domain and DKIM signature domain is proposed by B. Rienthong et al. [7] as technique to solve the problem of DKIM. However, this method is not considered whether the combination of sender's domain and DKIM domain using DKIM signature verification is reliable or not.

In this paper, we propose a method to measure the suspicion of pairs of e-mail sender's domain and DKIM signature domain as a score. Our system grade the reliability about any combination of e-mail sender's domain and DKIM signature domain by analyzing past communication logs focusing on wether e-mail was delivered or not. In addition, we applied this method to the actual communication logs and scored the reliability of e-mail judged reliable in DKIM.

The paper is organized as follows. In sectionII, we describe several existing methods of countermeasure for spoofed e-mail. In sectionIII, we explain the design of our scoring system. In sectionIV, we show the part of the scoring result. Then, in sectionV, we describe the legitimacy of our method using result that is shown in the previous section. Finally, in sectionVI, we present conclusion (and future prospect?).

II. Existing methods

Sender domain authentication, including DKIM, is a method to confirm whether the e-mail sender domain is legitimate or not. In this section, we describe four existing methods of countermeasure spoofed e-mail.

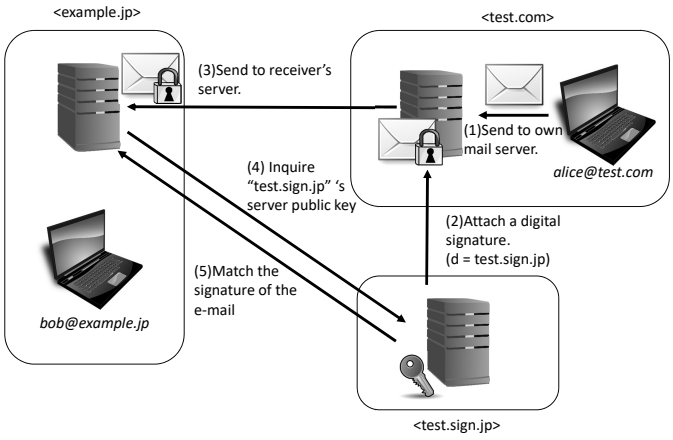

Fig. 1. Flow of DKIM verification

A. SPF

Sender Policy Framework (SPF) [10] is a method to confirm whether the e-mail transmission server's IP address is legitimate or not by checking SPF records. SPF record is a list of IP address of mail severs that the senders may use to send e-mails, which the owner of the the domain explicitly publishes on the authoritative DNS server of their domain beforehand.

B. DKIM

DomainKeys Identified Mail(DKIM) is a method to authenticate by using digital signature generated from e-mail body and e-mail header. The sender ataches the digital signature to e-mail and publish the public key that is used for generrate digital signature, and the reciever verifies the digital signature atached e-mail by using the public key which is published by the sender.

The flow of verification by DKIM is shown in the Fig.1.

C. DMARC

Domain-based Message Authentication, Reporting, and Conformance (DMARC) [8] is a policy how the reciever handles the e-mail that fails verification. DMARC uses two sender domain authentication methods, SPF and DKIM, to verify e-mails. If authentication of either or both SPF and DKIM fails, the receiver performs another processing which is declared by the addministor of the domain.

The flow of DMARC is shown in the Fig.2.

describe figure.

D. Existing DKIM signature verification method

This method is a method to verify the reliability of DKIM signature by checking past combination of e-mail sender's domain and DKIM signature domain.

The flow of this method is shown in the Fig.3.

describe figure.

However, this method has a problem with the reliability of the logs using verification of DKIM signature. Therefore, we proposed the method that scores the combinations of sender's domain and DKIM signature domain and verifies the DKIM signature's reliability.

Fig. 2. Flow of DMARC

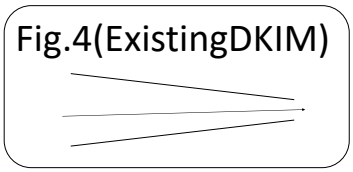

Fig. 3. Flow of DKIM signature verification method

TABLE I
EXISTS OF E-MAIL DELIVERY (EXAMPLE)

Combinations of domain		Month					
DKIM domain	From domain	Sep.	Aug.	Jul.	Jun.	May	Apr.
example.jp	example.jp	1	1	1	1	1	1
example.com	example.com	1	0	0	1	1	1

III. DESIGN OF OUR VERIFICATION SYSTEM

In this section, we describe the design of our verification system.

A. Preparation for scoring

As preparation for scoring the reliability of e-mail, we summarized the exist of e-mail delivery several months for each combinations of e-mail sender's domain and DKIM signature domain in a table like Table I.

This table shows wether e-mail was delivered or not about each combinations of e-mail sender's domain and DKIM signature domain. In the table, "1" means that an e-mail was delivered, and "0" means that an e-mail was not delivered. Our system verifies DKIM signature domain based on patterns of exists of e-mail delivery like that.

B. Scoring method

The scoring method of our system uses data about exist of e-mail delivery be applying weights and divides existence of e-mail delivery into several patterns to score the pairs of domain. First of all, we describe why we adopted such a method to our system.

At first, we considered the scoring method that the combinations of sender's domain and DKIM signature domain which e-mail were not delivered even for one month is suspicion. However, this method can only judges two-valued that the e-mail is suspicion or is not suspicion. In addition, this method cannot detect that DKIM signature moves by being updated a public key. Furthermore, in this method, older data and newer data are used in equally reliable.

Therefore, in order to these problems, we graded the combination of sender's domain and DKIM signature domain by using weight applying for each month and dividing e-mail delivery existence into several patterns. By using this method, our system can evaluate the DKIM signature domain multi-valued. In addition, by making weights heavier from old data to new data, our system can focus on newer data than older data.

Then, we explain about the patterns of existence of e-mail delivery, for example, DKIM signature domain may have moved, not have moved, and so on. In our system, we scored by dividing into the following four patterns.

1) The domain that there is no e-mail delivery for the past N months.
2) The domain that is newly created domain.
3) The domain that is not currently used after the move is completed.
4) Other than those above.

Pattern 1 means that DKIM signature domain may be not currently used, so if an e-mail signed by that DKIM domain is received, the e-mail should be judged as malicious e-mail. In the case of pattern 2, that e-mail is spoofed e-mail at that time, but there is a possibility that the domain has moved during a period when we had recorded the communication logs, so the e-mail of pattern 2 may be reliable. Pattern 3 is opposite to pattern 2. The DKIM domain had been used in the past, but there is no e-mail delivery using that DKIM signature, so the e-mail of pattern 3 may be suspiciousness. Pattern 4 is other than those above, which is included that e-mails were delivered using that DKIM signature every month, an e-mail is delivered or not delivered alternately, and DKIM signature domain had moved soon when system start analyze the communication logs.

Next, based on above ideas, we considered the flow of scoring. The flow of scoring is shown in Fig.4.

In this flow, first of all, we confirm whether e-mail was delivered even once for each combinations of domains. If there was no e-mail delivery, that DKIM signature domain's reliability is 0 points. Next, we check wether the DKIM signature domain is newly created. In that case, we graded using weight 1 with 40 points as the reference point. weight 1 is calculated from (1).

$$a = b + c \tag{1}$$

The reason why 40 points as the reference point, ... Then, we confirm wether the DKIM signature domain had moved completed and that DKIM domain is not currently used. In this case, we graded using weight 2 with 60 points as the reference point. weight 2 is calculated from (2).

$$a = b + c \tag{2}$$

The reason why 40 points as the reference point, ... Finally, we graded the remaining patterns using weight 3 with 0 points as the reference point. weight 3 is calculated from (3).

$$a = b + c \tag{3}$$

The reason why 0 points as the reference point, ...

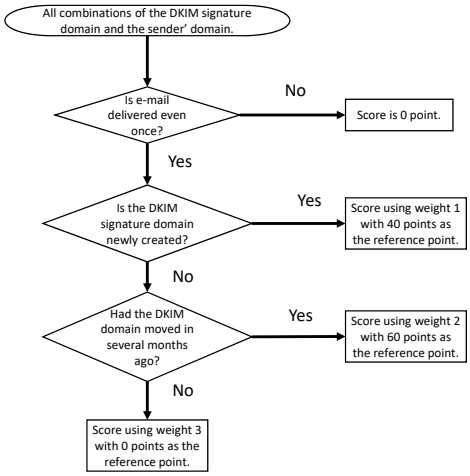

Fig. 4. flow of scoring

TABLE II
WEIGHT1

Month	Sep.	Aug.	Jul.	Jun.	May	Apr.
Weight	13	12	11	9	8	7

TABLE III
WEIGHT2

Month	Sep.	Aug.	Jul.	Jun.	May	Apr.
Weight	7	8	9	11	12	13

IV. EXPERIMENT

We use our verification system with actual communication logs for 6 months and graded the e-mail suspiciousness. The communication logs that we used in this experiment has about 200,000 combinations of sender's domain and DKIM signature domain per one month. In other word, we made the experiment using about 1,200,000 combinations of domains. In the situation that use communication logs for 6 months to score, weights are as Table II, Table III, and Table IV .

The score example of conceivable all e-mail delivery patterns in this experiment is as below Table V.

V. EVALUATION

In order to confirm that the score of reliability scoring by our system, we check the result of the domain for load

TABLE IV
WEIGHT3

Month	Sep.	Aug.	Jul.	Jun.	May	Apr.
Weight	25	21	18	15	12	9

balancing. The results of scoring about one major company's load distribution domain are Table VI. From this result, it can be said that we can properly measure the reliability of e-mail by using our verification method.

VI. CONCLUSION

In this paper, we proposed a method that measure the suspicion of combinations of e-mail sender's domain and DKIM signature domain by analyzing and scoring communication logs that foucuses whether e-mail was delivered or not.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

[1] FBI (Federal Bureau of Investigation), "Public Service Announcement, Business E-mail Compromise: The 3.1 billion dollar scam" [online] Available: https://www.ic3.gov/media/2016/160614.aspx, June 2016.

[2] I.Androutsopoulos, J.Koutsias, K.V.Chandrinos, G.Paliouras, C.D.Spyropoulos, "An evaluation of Naive Bayesian anti-spam filtering," Proceedings of the work-shop on Machine Learning in the New Information, pp.9-17, 2000.

[3] I.Androutsopoulos, J.Koutsias, K.V.Chandrinos, C.D.Spyropoulos, "An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages," Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, pp.160- 167, 2000.

[4] "The Apache SpamAssassin Project," [online] Available: http://spamassassin.apache.org/

[5] J.Mason, "Filtering spam with spamassassin,"In HEANet Annual Conference, 2002.

[6] D. Crocker, T. Hansen, and M. Kucherawy, "DomainKeys Identified Mail (DKIM) signatures", (STD 76). Sep. 2011.

[7] B. Rienthong, N. Kitagawa, and N. Yamai, "Countermeasure of spoofed emails by checking the reliability of DKIM signature". Nov. 2016.

[8] M. Kucherawy, and E. Zwicky, "Domain-based message authentication, reporting, and conformance (DMARC)", (RFC 7489). Mar. 2015.

[9] L. Eggert, "DMARC Deployment "[online] Available: https://eggert.org/meter/dmarc, Mar. 2017.

[10] M. Wong, and W. Schlitt, "Sender Policy Framework (SPF) for authorizing use of domains in e-mail", (RFC 4408). Apr. 2006.

TABLE V
EXAMPLE OF CONCEIVABLE ALL PATTERNS

Sep.	Aug.	Jul.	Jun.	May.	Apr.	Poitns
0	0	0	0	0	0	0
0	0	0	0	0	1	13
0	0	0	0	1	0	12
0	0	0	0	1	1	25
0	0	0	1	0	0	15
0	0	0	1	0	1	24
0	0	0	1	1	0	27
0	0	0	1	1	1	36
0	0	1	0	0	0	18
0	0	1	0	0	1	27
0	0	1	0	1	0	30
0	0	1	0	1	1	39
0	0	1	1	0	0	33
0	0	1	1	0	1	42
0	0	1	1	1	0	45
0	0	1	1	1	1	54
0	1	0	0	0	0	21
0	1	0	0	0	1	30
0	1	0	0	1	0	33
0	1	0	0	1	1	42
0	1	0	0	0	0	36
0	1	0	1	0	1	45
0	1	0	1	1	0	48
0	1	0	1	1	1	57
0	1	0	1	0	0	39
0	1	1	0	0	1	48
0	1	1	0	1	0	51
0	1	1	0	1	1	60
0	1	1	0	0	0	54
0	1	1	1	0	1	63
0	1	1	1	1	0	66
0	1	1	1	1	1	75
1	0	0	0	0	0	53
1	0	0	0	0	1	34
1	0	0	0	1	0	37
1	0	0	0	1	1	46
1	0	0	1	0	0	40
1	0	0	1	0	1	49
1	0	0	1	1	0	52
1	0	0	1	1	1	61
1	0	1	0	0	0	43
1	0	1	0	0	1	52
1	0	1	0	1	0	55
1	0	1	0	1	1	64
1	0	1	1	0	0	58
1	0	1	1	0	1	67
1	0	1	1	1	0	70
1	0	1	1	1	1	79
1	1	0	0	0	0	65
1	1	0	0	0	1	55
1	1	0	0	1	0	58
1	1	0	0	1	1	67
1	1	0	0	0	0	61
1	1	0	1	0	1	70
1	1	0	1	1	0	73
1	1	0	1	1	1	82
1	1	0	1	0	0	76
1	1	1	0	0	1	73
1	1	1	0	1	0	76
1	1	1	0	1	1	85
1	1	1	0	0	0	79
1	1	1	1	0	1	88
1	1	1	1	1	0	91
1	1	1	1	1	1	100

TABLE VI
RESULT OF ONE MAJOR COMPANY'S DKIM DOMAIN SCORE

DKIM signature domain	Sebder's from domain	Score
yahoo.co.jp	arch.t.u-tokyo.ac.jp	0
yahoo.co.jp	gesurg.med.osaka-u.ac.jp	0
yahoo.co.jp	hotmail.co.jp	0
yahoo.co.jp	hotmail.com	0
yahoo.co.jp	kaikou.or.jp	0
yahoo.co.jp	m.titech.ac.jp	0
yahoo.co.jp	eng.u-hyogo.ac.jp	0
yahoo.co.jp	osaka-med.ac.jp	0
yahoo.co.jp	shirayama.name	0
yahoo.co.jp	nf-planning.com	12
yahoo.co.jp	vip.163.com	12
yahoo.co.jp	yuuten.mail.tokimeki-land.com	12
yahoo.co.jp	goleiro-pro.com	13
yahoo.co.jp	hightec-sys.com	13
yahoo.co.jp	i.softbank.jp	13
yahoo.co.jp	nifty.com	13
yahoo.co.jp	elect.chuo-u.ac.jp	15
yahoo.co.jp	faculty.chiba-u.jp	15
yahoo.co.jp	jfe-steel.co.jp	15
yahoo.co.jp	j-monkey.jp	15
yahoo.co.jp	ya2.so-net.ne.jp	15
yahoo.co.jp	icloud.com	18
yahoo.co.jp	takahashi.name	18
yahoo.co.jp	herb.ocn.ne.jp	21
yahoo.co.jp	pu-hiroshima.ac.jp	24
yahoo.co.jp	marianna-u.ac.jp	27
yahoo.co.jp	aitech.ac.jp	33
yahoo.co.jp	atlastanker.com	36
yahoo.co.jp	oregano.ocn.ne.jp	43
yahoo.co.jp	iog.u-tokyo.ac.jp	46
yahoo.co.jp	ybb.ne.jp	48
yahoo.co.jp	affrc.go.jp	53
yahoo.co.jp	cd5.so-net.ne.jp	53
yahoo.co.jp	zeus.eonet.ne.jp	53
yahoo.co.jp	naturecenter-risen.com	65
yahoo.co.jp	umin.ac.jp	73
yahoo.co.jp	cea.jp	75
yahoo.co.jp	gmail.com	100
yahoo.co.jp	n04.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n05.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n06.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n07.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n08.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n09.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n10.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n11.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n12.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n13.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n14.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n24.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n25.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n26.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n27.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n28.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n29.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n30.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	n31.bulk.ogk.yahoo.co.jp	100
yahoo.co.jp	ohkubo-kurume.com	100
yahoo.co.jp	susineta.com	100
yahoo.co.jp	yahoo.co.jp	100