PREVENTING SPAM DISGUISED AS ERROR MAIL

Manabu IWANAGA

Kyushu University
Graduate School of Information Science
and Electrical Engineering
6-10-1 Hakozaki, Higashi-ku
Fukuoka Fukuoka, 812-8581 Japan
iwanaga@itslab.csce.kyushu-u.ac.jp

Toshihiro TABATA, Kouichi SAKURAI

Kyushu University
Faculty of Information Science
and Electrical Engineering
6-10-1 Hakozaki, Higashi-ku
Fukuoka Fukuoka, 812-8581 Japan
{tabata, sakurai}@csce.kyushu-u.ac.jp

ABSTRACT

Some methods against spam are based on a principle that a recipient reads only messages from senders who are registered by the recipient (challenge and response). In these schemes, some exceptions are required to show error mail (bounce message) to a sender of an original message. However, spammers can abuse this exception to send spam to users. In this paper, we propose a new method which combines a Bayesian filter and challenge-and-response using Message Authentication Code (MAC) to avoid that spam.

1. INTRODUCTION

According to popularization of email, spam is increasing because it is very inexpensive way to advertising. As users take measure to avoid spam, spammers also increase their sophistication of spamming, e.g. they fake headers of email including sender's address, and they tend to avoid using frank words which can be filtered with simple blacklist of words. According to some latest researches, approximately half of all email received by workers are spam. Massive spam filling users' mailbox not only irritates users, but makes it painful work for users to pick out messages users really need to read, especially for users who constantly send and receive messages to/from a large number of people.

To avoid spam, various schemes are proposed and used. These schemes are roughly divided between server-based and client-based. We focus on client-based anti-spam schemes. Some schemes are based on a principle that a recipient reads only messages from senders registered by the recipient. In these schemes, some exceptions are required for users to read error mail (bounce message), since it is impossible for recipients to add all possible senders of error mail, mailer daemon, to his sender-list. However, spammers can abuse this exception to send spam to users. Disguising their spam as error mail, spammers can show their spam to recipients using these schemes. In addition, if we have to consider a threat of wiretapping, the situation becomes worse.

In this paper, we propose improved scheme, using not only challenge and response but also Bayesian filtering. Our proposed scheme applies Bayesian filtering to error mail, and applies the challenge and response scheme proposed by M. Jakobsson et al. [1] to all messages except error mail. Using this scheme, we can avoid spam disguised as error mail according to words in message-body and header, assuring certain receipt of legitimate messages from registered sender.

2. RELATED WORK

2.1. Filtering by Message

Filtering by tokens in message and sender's address is well known as a simple countermeasure against spam. P. Graham introduced schemes [2][3] called Bayesian filter. This filter is according to statistical probability, not rules made by human, how often tokens in a message appeared in spam and how often in legitimate messages. Furthermore, this filter learns tokens in the message according to decision of filter itself, for later calculation.

When a user brings Bayesian filter into use, he/she imports spam and non-spam into Bayesian filter. Then the Bayesian filter calculates probability for each word, that a message including the word is spam. After that, the Bayesian filter calculates probability that an incoming message is spam, according to probability of words in the message. That is to say, if a message has many words peculiar to spam, the Bayesian filter gives high probability for the message. Usually the Bayesian filter decides that the message is spam if and only if the probability for the message is higher than some constant.

2.2. Challenge and Response

There are also other approaches against spam that a recipient registers legitimate senders, such as schemes by E. Gabber et al. [4], by R. Hall [5], and Mailblocks [6]. Here legitimate senders are considered to be people who are allowed by the recipient to send a message or who satisfied some procedure that the recipient stipulated.

In these schemes, the simplest procedure for a sender who has not exchanged by then is accessing specific URL with web browser. In this way, a recipient requires senders to have existent email address and read messages incoming to the address. To charge spammer with cost surely, some methods adopted additional process. Some schemes require a sender to enter alphabets displayed in an image, and other schemes adopts computational task. These schemes requires senders to expense cost, rather than reply to request, aiming to eliminate spam's low cost for sending to stop spamming. Even if a spammer makes his spamming program capable to recognize and reply to request for register automatically, his/her computer has to pay some cost to send a spam. If a sender wants to send a message for 10,000 recipients, he/she must pay cost 10,000 times as much as he/she want to send for one recipient. As a result, if a spammer becomes to have to expend 10 seconds for each recipient, he/she can send only 1,440 messages a day, and spamming will not pay.

Legitimate senders have to perform the procedure once for each new recipient. In schemes of challenge and response, senders have to show evidence that they are legitimate senders who are registered. However, a spammer can wiretap the evidence, because email is transferred in plaintext. Especially in the schemes based on tagged address, which is concatenation of outward address (core address) and password, an adversary only has to listen in "from" and "to" in the header of messages.

To avoid this, the scheme by M. Jakobsson et al. [1] adopts a manner that a recipient grants cryptographic key to legitimate senders. The scheme uses Message Authentication Code (MAC). In their scheme, adversaries (spammers) are defined as active. That is to say, they not only wiretap communication channel, but remove and/or inject any messages at will. Therefore, this scheme adopted MAC calculated from each message to defeat wiretapping, and only a sender and a recipient know a key of MAC. A sender makes a setup message to obtain a key. The message contains a proof that the sender has performed a certain computational task or a monetary expense. Jakobsson's scheme can also detect that a legitimate message incoming is altered to spam by spammer.

Definition

In this paper, we define "spam" as

(1) messages which are sent to large number of recipients, and whose senders ignore replies from recipients.

(2) error mail which is bounced back to a recipient according to "false" sender address in above messages.

And we call messages which are not spam "legitimate".

3. ERROR MAIL

3.1. Error Mail

When we send an email, the email is delivered by SMTP servers. Since a SMTP server that a sender posts an email is usually different from one that manages recipient's maildrop, a sender seldom has a connection with a server of recipient directly. Thus relaying occurs.

Servers in the head or middle of relaying cannot know whether a message will be delivered successfully. Therefore, if a recipient does not exist or a recipient rejects the message, it turns out failure after a sender posted. Then RFC2821 [7] says that the MTA who finds the email cannot be delivered has to make an "undeliverable mail" notification message, so-called "error mail" or "bounce mail". An error mail usually includes a reason why a delivery failed, error messages in the connection, and a message-body sender intended to send.

3.2. Spam Disguised as Error Mail

Since error mail is essential to notify failure to email users, it must not be eliminated from users' screens. However, spammers may attempt to disguise their spam as error mail. The spam is disguised as error mail and escapes from spam-protection, but a user who tries to read it encounters malicious spam.

When a user's computer receives spam in a form of error mail, following situations are possible:

(1) A spammer impersonates the user as a source of spam, and undeliverable spam was bounced back to the user's computer with error-message according the header of an email. This case is divided into two cases:

   (a) The spammer wants to hide himself/herself, so he impersonates the user. Spammers tend to hide their own address to avoid escape that their spam is filtered by their addresses, and to hinder being reported to his/her Internet Service Provider (ISP). For that reason, they often assume non-existent address, and in some case they personate other's address. As a result, a personated recipient receives large amount of error mail.

(b) It is spammer's primary purpose to bounce spam back to the user, so he sends spam to non-existent address intentionally. Spammers aim at that their spam attracts recipients' attention, or recipients' and servers' filters may behave different behave from common spam.

(2) Spammer disguises spam as error mail, and sends it to the recipient directly. It is done by the same reason as (1b) and spammers can include their advertisement in any part of messages. If a spammer eavesdrops on legitimate messages to forges his/her spam disguised as error mail from these legitimate messages, he/she takes this step.

3.3. Processing of Error Mail in Schemes of challenge and response

A user can add some tags to one's messages to recognize error mail which is really correspond to one the user sent, then ignore error mail which is not correspond to one's messages (case (1) in Section 3.2). It is safer if tags are encrypted and nobody can make the tags other than the user. A user can also achieve the same purpose, recording outbound messages to check an error mail with them.

However, a communication channel for email is not encrypted. If a spammer can wiretap users' messages, the spammer can disguise spam as error mail whose original messages is one the recipient has sent (possibility (2) in Section 3.2). In this case, it is harder for recipients' computers distinguish real error mail from disguised spam. So is the scheme [1]. Therefore additional methods against disguised spam are required to distinguish these messages.

4. OUR PROPOSED SCHEME

To prevent spam from shown to recipient as error mail, we propose improved scheme based on Jakobsson's one. In this scheme, we use a Bayesian filter to prevent disguised spam. The Bayesian filter learns non-error mail according to result of challenge and response, distinguishes error mail from disguised spam, and learns error mail according to judgment of the Bayesian filter itself. This scheme can keep using challenge and response and prevent spam that is disguised as error mail from MTAs, using a Bayesian filter.

The scheme by Jakobsson et al. [1] is based on following concepts. These concepts are realized and performed by a process on a recipient's computer, called mail proxy. Mail proxies work in the middle of a mail server and a mail client, so users can keep using their mail client without any modification.

- Messages are regarded as legitimate messages that should be shown to the recipient, if they are from a sender who performed a setup previously and consequently share a key with the recipient. That is to say, if a MAC in his/her message is correctly generated from the message and the shared key, the message is shown to the recipient.

- Messages are regarded as messages that should not be shown to the recipient, if they are from senders who are not registered. The mail proxy on the recipient's computer makes a request telling senders of the messages to install a mail proxy and perform a setup for these messages.

- A mail proxy automatically perform a setup with a mail proxy of recipient. A mail proxy also add a MAC to an outgoing message for a recipient who the mail proxy performed a setup with. So a user who has already installed mail proxy does not have to care about a setup.

We should take notice of a fact that not all messages from unregistered senders are illegitimate. All legitimate senders who begin to communicate with a recipient are not registered at first. Our proposed scheme expects that these senders perform a setup according to request and resend their messages. Therefore we added some new concepts.

- If a sender does not perform a setup within certain period after a request for a setup, the message is assumed illegitimate.

- If a sender performed a setup, the message is assumed legitimate. Note that the message is not shown to a recipient in this case, because the message without MAC may have been altered by a spammer. The sender has to resend his/her message with MAC.

In practice, a recipient's computer processes received messages with following rules.

(1) If a message has a valid MAC, then the message is regarded as legitimate one and the Bayesian filter learns the message as a legitimate message, then the message is shown to recipient.

(2) If a message is a request for a setup or a setup communication that contains evidence that the sender performed computational work stipulated, then a setup is performed automatically.

(3) If a message seems an error mail, then the Bayesian filter judges whether the message is legitimate or not. If the Bayesian filter judges that it is legitimate, the message is learned as a legitimate message, and shown to recipient. If not, the message is learned as a spam, and isolated.

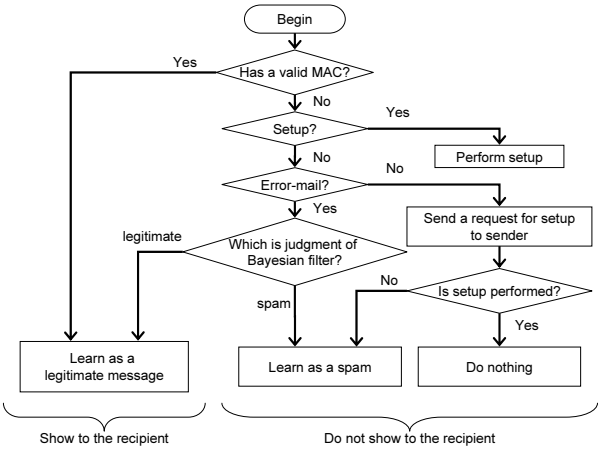

Fig. 1. How to deal with incoming messages

(4) If a message does not meet all above conditions, then the message is regarded as one from unregistered sender, and a request for a setup is sent to the sender. After that, if the sender does not perform a setup within the certain period, the Bayesian filter learns the message as a spam and the message is isolated. However, If the sender does, the Bayesian filter doesn't learn the message, because the sender will resend the message to show the message to the recipient.

Not only message-body but attached original message should be considered by the Bayesian filter, because spammers may infect their advertisement not with message-body but with attached original message. Since both legitimate error mail and spam disguised as error mail will have typical error notification sentences, so words in these sentences will be ordinary words which cannot be an evidence to judge a message to be legitimate or spam.

While our proposed scheme is bother (particularly if the procedure to a setup is complex) for senders and more complex than one using only Bayesian filtering, there is some advantage in our proposed scheme. First advantage is assurance. Filtering may make a misjudgment, and users are not likely to be aware of that. By contrast, if a recipient registers legitimate senders in a manner of [1], the recipient can read messages from them certainly. Second, a Bayesian filter needs large corpus to determine whether a message is spam or not suitably. A Bayesian filter can learn which tokens legitimate messages have and which spam messages have, according to rules above.

5. CONCLUSION AND FUTURE WORK

In this paper, we pointed out the problem that a spammer can disguise spam as error mail and avoid challenge and response anti-spam schemes, and then we proposed a scheme linking Bayesian filter with challenge and response. Our proposed scheme adopts Bayesian filtering to distinguish error mail from spam disguised as error mail in an eavesdropping regarded manner. Therefore, this scheme can protect from not only from simple sender impersonation, but also false error mail with eavesdropping, as long as a spammer attempt to advertise.

In order for challenge and response to work well, we require suitable method for challenge and response that it is not annoying for legitimate senders. And again, not only error mail processing but our proposed scheme rely on sender's address in a message, so if an adversary (not only a spammer but a DoS attacker) personates other's address, although adversary cannot send spam itself, personated person is annoyed with wrong request for a setup in our proposed method. To avoid this, we require a way to prevent personation. To prevent personation is a fundamental issue.

6. REFERENCES

[1] M. Jakobsson, J. Linn, J. Algesheimer, "How to Protect Against a Militant Spammer", Cryptology ePrint archive, report 2003/071, 2003.

[2] P. Graham, "A Plan for Spam", http://paulgraham.com/spam.html

[3] P. Graham, "Better Bayesian Filtering", In proceedings of Spam conference, January 2003, http://spamconference.org/proceedings2003.html.

[4] E. Gabber, M. Jakobsson, Y. Matias, A. Mayer, "Curbing Junk Email via secure Classification", In proceedings of Financial Cryptography '98, Volume 1465 of Lecture Notes in Computer Science, pages 198-213. Springer- Verlag, February 1998.

[5] R. J. Hall, "Channels: Avoiding unwanted electronic mail", In Proceedings of the 1996 DIMACS Symposium on Network Threats, pages 85-103.

[6] Mailblocks, http://www.mailblocks.com/

[7] J. Klensin, ed., "Simple Mail Transfer Protocol", Internet RFC-2821, April 2001.