Network Working Group K. Ono
Internet-Draft H. Schulzrinne
Intended status: Informational Columbia University
Expires: April 15, 2010 October 12, 2009

Using Cross-Media Relations to Reduce False Positives during SPIT Filtering
draft-ono-cross-media-relations-00.txt

Status of this Memo

Copyright Notice

Abstract

Some legitimate calls are from persons or organizations connecting the callee with weak social ties, such as a restaurant the callee booked a table on-line. These legitimate calls are often mistakenly labeled as unsolicited calls at a filtering system which uses the contact list of the callee. To reduce these false positives during SPIT filtering, we propose two approaches to label incoming calls using cross-media relations from earlier communications. One approach is that a potential caller offers the callee his contact address(es) which might be used in future calls. Another is that a callee provides a potential caller with weakly secret information. In order to be identified as someone the callee contacted through other means previously, the caller should convey the information in future calls.

Table of Contents

1. Introduction

Unsolicited calls usually originate from persons or organizations, whom the callee does not know their contact addresses nor met before. Since an IP-based infrastructure is more vulnerable to unsolicited calls or SPIT (SPam over Internet Telephony) calls, as described in [RFC5039], people have recently been experiencing more SPIT calls.

Most legitimate calls, by contrast, have caller identifiers (IDs) that the callee has seen before. Some legitimate calls, however, have unknown caller IDs. Examples of these legitimate calls include confirmations of appointments, reservations, or deliveries, and recorded notifications of flight delays or school closing on a snowy day. These legitimate calls are prone to false positives during SPIT filtering. This is because their caller IDs are not found on the callee's white list even if the callers have had prior contact with the callee through transactions over the web or email exchanges [have-i-met-u-before].

This is a natural consequence of a conventional white list, which usually contains the same addresses with his contact list or address book. The contact list contains known or used contact addresses of persons or organizations with strong ties in his or her social network, such as family members, friends, and colleagues. The contact list, however, rarely includes the addresses of those with weak social ties [weak-ties], such as an operator at the customer center of an on-line shopping site, or friends of a friend in an SNS (Social Network Service) over the web.

Using a white list to label incoming calls requires caller ID authentication. For a VoIP (Voice over IP) call using the SIP (Session Initiation Protocol) [RFC3261], the SIP Identity header [RFC4474] enables a callee to authenticate the caller ID. Some legitimate calls, however, are sent with "unavailable" caller IDs. These calls without any authenticated caller IDs limit the effectiveness of labeling incoming calls based on the caller ID.

In summary, conventional whitelisting can hardly label the following types of calls:

D1: Calls from persons or organizations connecting the callee with
     weak social ties

D2: Calls from those connecting with strong social ties, but using
     new, alternative, or unknown caller IDs, e.g., from a visited
     place like a hotel

D3: Calls with unauthenticated caller IDs, e.g., through an unauthenticated domain or using the caller ID in tel-URI [RFC3966]

D4: Calls with blocked caller IDs

To cope with these difficulties of conventional whitelisting, we propose to expand filter conditions with two mechanisms; both use cross-media relations from earlier communications.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

This document defines the following term:

Cross-media relations: The information suggesting the existence of a prior contact. When the information is offered by a potential caller, it can be his contact address, which is in plain text or hash coded. When the information is offered by the callee, it needs to be a weak secret in order to be used for labeling incoming call without authentication. The weakly secret information is the value of the Message-ID [RFC5322] of an outgoing email from the callee or random components contained in the callee's customized contact address.

3. Using Cross-Media Relations

Figure 1 depicts an overview of our first mechanism. In order to cope with the difficulties of D1 and D2 described above, a potential caller offers the callee his contact addresses which he might use in future calls. If the callee agrees, these contact addresses are added to his white list. Figure 2 depicts an overview of our second mechanism. In order to cope with the difficulties of D3 and D4 mainly, a callee provides a potential caller with weakly secret information. The caller should use it in future calls in order to be identified as someone the callee has had prior contact through other means. The second mechanism allows to label incoming calls using the weakly secret information, instead of caller IDs.

```
   [Potential Caller]               [Callee]

      | HTTP Response                   |          _____________
      |------------------------------>| rcv.&   /             \
      |                               | storep |\_____________/|
      |                               |------> | White list:   |
      or                                       |               |
      | Responded Email               |        | Contact addrs.|
      |------------------------------>| rcv.&  | in plain text |
      |                               | store  | or hash coded |
      |                               |------> |               |
     ~~~                             ~~~       |               |
     ~~~                             ~~~       |               |
      | SIP INVITE Request            |query to|               |
      | From:                         |label   |               |
      |------------------------------>|------>  \_____________/
      |                               |       Filter conditions
```

Figure 1: An Overview of Mechanism I

```
                         [Callee]                  [Potential Caller]
     ______________              |                               |
    /              \  generate,|                               |
   |\______________/| send,    | HTTP Request                  |
   | Weak sercrets: | & store  |------------------------------>|
   |                |<---------|                               |
   | Random comp. in|         or
   | customized own | generate,|                               |
   | contact addrs. | send,    | Outgoing Email                |
   | & Message-IDs  | & store  |------------------------------>|
   |                |<---------|                               |
   |                |        ~~~                              ~~~
   |                |        ~~~                              ~~~
   |                |          | SIP INITE Request             |
   |                |          | To:                           |
   |                |          | or                            |
   |                | query to | New-References:               |
   |                | label    |<------------------------------|
    \______________/ <---------|                               |
   Filter conditions
```

Figure 2: An Overview of Mechanism II

4. Mechanism I: Contact Addresses of Potential Callers

Our first mechanism enables potential callers to offer their contact addresses which they might use in future calls more easily and more widely.

To make it more easily in a web transaction, we propose that an HTTP response from a potential caller conveys his or her contact addresses in an HTML META tag, HTTP-EQUIV [W3C.REC-html401-19991224] or a new HTTP header, Correspondence-URIs [I-D.shacham-http-corr-uris]. In an email exchange, the contact addresses can be contained in a vCard [RFC2426] attached to an email message sent from a potential caller. After the callee receives the contact addresses of a potential caller, the callee adds them to his or her white list. To prevent misuse, the callee should be prompted for confirmation before updating his or her white list.

To make it more widely, we propose to convey hash-coded contact addresses of potential callers. Hash-coded contact addresses are suitable if potential callers prefer concealing their routable address for privacy reasons. For example, in an SNS where subscribers prefer not to publish their routable contact addresses, subscribers should be allowed to publish their hashed contact address for the limited purpose of filtering calls.

This first mechanism is useful in a case where the contact addresses of potential callers have been determined and the number is small. In other cases, our second mechanism should be used.

5. Mechanism II: Weakly Secret Information

Our second mechanism allows a callee to provide potential callers with weakly secret information as cross-media relations. Potential callers should use this information in future calls to be identified as someone with whom the callee has had prior contact through other means.

This mechanism is useful in the following cases. One is where the previous contact was one-to-many correspondence between the callee and potential callers. For example, when joining an association, the callee is unwilling to receive all the contact addresses of potential callers in the association. Another case is where potential callers might use a different or no authenticated caller ID, due to their situation such as traveling, or due to the type of communication medium or service, such as two-stage dialing for international calls.

The information about cross-media relations depends on the

communication medium of a previous contact. A customized contact address containing a random component or a token can be used when a callee fills out contact information on a web site, or in a vCard attached to an email message. The random component or token can be automatically generated in correspondence to the URL (Uniform Resource Locator) [RFC3986], or manually specified. In the examples in Figure 3, a token, "adgs24oF", in the SIP-URI is set between the user name and the domain name preceded with "+". This is the same way as the email addressing practice called subaddressing [RFC5233]. For tel-URI, a token, "0012", follows the E.164 number like an extension. To convey this information in a later call, the caller just needs to set the destination address to the customized contact address, as the INVITE request shown in Figure 3.

```
  Web client                                                 Web server
  at callee                                         at potential caller
  |                                                                   |
  | POST /join HTTP/1.1                                               |
  | HOST: ffp.airline.com                                             |
  | Content-Length: 128                                               |
  | Conetnt-Type: application/x-www-form-urlencoded                   |
  |                                                                   |
  | phone1=sip:userA+adgs24oF@example.com&phone2=tel:+121291711110012 |
  | ...                                                               |
  |------------------------------------------------------------------>|
  | HTTP/1.1 200 OK                                                   |
  |<------------------------------------------------------------------|
  |                                                                   |
 ~~~                                                                 ~~~
 ~~~                                                                 ~~~
  |                                                                   |
  | INVITE sip:username+adgs20oF@exmale.com SIP/2.0                   |
  | To: sip:username+adgs20oF@exmale.com                              |
  |<------------------------------------------------------------------|
  |                                                                   |
 Note: To show related headers only, many mandatory headers are omitted.
```

Figure 3: Using Weak-Eecret in HTTP and SIP messages

```
  Email client                                             Email client
  at callee                                         at potential caller
  |                                                                   |
  | (Message)                                                         |
  |<------------------------------------------------------------------|
  | Message                                                           |
  | Message-ID:<56626454-8D6F-49FF-BFA0-1FF6A63E71EA@example.com>     |
  |----------------------------------------------------------------->|
  |                                                                   |
 ~~~                                                                 ~~~
 ~~~                                                                 ~~~
  |                                                                   |
  | INVITE sip:username@exmale.com SIP/2.0                            |
  | New-References:<56626454-8D6F-49FF-BFA0-1FF6A63E71EA@example.com>;|
  | type="email"                                                      |
  |<------------------------------------------------------------------|
  |                                                                   |
 Note: To show related headers only, many mandatory headers are omitted.
```

Figure 4: Using Weak Secret in Email and SIP messages

Specifically in an email exchange, as shown in Figure 4, the message identifier of an email from the callee can be used. A potential caller first sends a message to the callee requesting a real-time communication. This message is optional. If the callee accepts the request, he will respond to it by email optionally containing his contact address. As a result, the message identifier of the response email, which is set in the Message-ID header, can be used as weakly secret information to prove the acceptance from the callee. Thus, the message identifiers of outbound emails or the call identifers of SIP calls can be included by the potential caller in a later call, even if he uses a different caller ID or type of communication medium. To convey the message identifier in a SIP call, the caller should set a SIP header extension [I-D.ono-earlier-comm-references] to its value. In the example message in Figure 4, a new SIP header extension under discussion, New-References is used.

6. Enhanced Filtering

Our two mechanisms enhance a filtering process using caller IDs in the following ways. For our first mechanism, it extends white list by contain contact addresses in hash format.

> Furthermore, sharing a white list for people within an organization increases the effectiveness of the white list. A remote server maintaining common white lists is also effective, if it can be queried whether or not the caller ID is found on the

list and respond with binary, e.g., query about membership.

For our second mechanism, it adds conditionals using weakly secret information after the conditionals checking the authenticated caller ID of an incoming call on a black list nor on a white list. For incoming calls of which caller ID is not found on the white list, the expanded filtering process tests on two new conditionals. The first one is whether it contains a valid Message-ID value in the new references header under discussion [I-D.ono-earlier-comm-references]. The second is whether it contains a valid subaddress of the destination address, i.e., in the To header. That is, if the test succeeds in either condition of the message identifier or subaddress, the call request will be accepted.

Therefore, by enhancing existing filter conditions, our proposed mechanisms enable a callee to label incoming calls, not only from persons or organizations with weak ties, but also from callers who change their caller IDs. As a result, they are expected to reduce false positives that occur during filtering.

7. Security Consideration

TBD

8. IANA Consideration

This document requires no IANA Consideration.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

9.2. Informative References

[I-D.ono-earlier-comm-references]
Ono, K. and H. Schulzrinne, "Referencing Earlier Communications in SIP Requests",

draft-ono-earlier-comm-references-00 (work in progress), October 2009.

[I-D.shacham-http-corr-uris]
Shacham, R. and H. Schulzrinne, "HTTP Header for Future Correspondence Addresses", draft-shacham-http-corr-uris-00 (work in progress), May 2007.

[RFC2426] Dawson, F. and T. Howes, "vCard MIME Directory Profile", RFC 2426, September 1998.

[RFC3966] Schulzrinne, H., "The tel URI for Telephone Numbers", RFC 3966, December 2004.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.

[RFC4474] Peterson, J. and C. Jennings, "Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP)", RFC 4474, August 2006.

[RFC5039] Rosenberg, J. and C. Jennings, "The Session Initiation Protocol (SIP) and Spam", RFC 5039, January 2008.

[RFC5233] Murchison, K., "Sieve Email Filtering: Subaddress Extension", RFC 5233, January 2008.

[RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.

[W3C.REC-html401-19991224]
Jacobs, I., Hors, A., and D. Raggett, "HTML 4.01 Specification", World Wide Web Consortium Recommendation REC-html401-19991224, December 1999, <http://www.w3.org/TR/1999/REC-html401-19991224>.

[have-i-met-u-before]
Ono, K. and H. Schulzrinne, "Have I Met You Before? Using Cross-Media Relations to Reduce SPIT", IPTCOMM 09, July 2009.

[weak-ties]
Granovetter, M., "The Strength of Weak Ties", Amer.J.of Sociology 78:1360-80, May 1973.

Authors' Addresses

Kumiko Ono
Columbia University
Department of Computer Science
New York, NY 10027
USA

Email: kumiko@cs.columbia.edu

Henning Schulzrin ne
Columbia University
Department of Computer Science
New York, NY 10027
USA

Email: hgs@cs.columbia.edu