MARF Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 30, 2012

J. Falk, Ed.
Return Path
M. Kucherawy, Ed.
Cloudmark

January 27, 2012

Redaction of Potentially Sensitive Data from Mail Abuse Reports
draft-ietf-marf-redaction-08

Abstract

Email messages often contain information that might be considered private or sensitive, per either regulation or social norms. When such a message becomes the subject of a report intended to be shared with other entities, the report generator may wish to redact or elide the sensitive portions of the message. This memo suggests one method for doing so effectively.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

[ARF] defines a message format for sending reports of abuse in the messaging infrastructure, with an eye toward automating both the generating and consumption of those reports.

For privacy considerations it might be the policy of a report generator to anonymize, or obscure, portions of the report that might identify an end user who caused the report to be generated. This has come to be known in feedback loop parlance as "redaction". Precisely how this is done is unspecified in [ARF] as it will generally be a matter of local policy. That specification does admonish generators against being too over-zealous with this practice, as obscuring too much data makes the report non-actionable.

Previous redaction practices, such as replacing local-parts of addresses with a uniform string like "xxxxxxxx", frustrated any kind of prioritizing or grouping of reports. This memo presents a practice for conducting redaction in a manner that allows a report receiver to detect that two reports were caused by the same end user without revealing the identify of that user. That is, the report receiver can use the redacted string, such as an obscured email address, to determine that two such unredacted strings were identical; the reports originally contained the same address.

Generally, it is assumed that the recipient-identifying fields of a message, when copied into a report, are to be obscured to protect the identity of the end user who submitted the complaint about the message. However, it is also presumed that other data will be left intact, and those data could be correlated against log files or other resources to determine the intended recipient of the original message.

2. Keywords

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 [KEYWORDS].

3. Recommended Practice

When redacting of reports is desired, in order to enable a report receiver to correlate reports that might refer to a common but anonymous source, the report generator SHOULD use the following practice:

1. Select a transformation mechanism (see Section 4) that is consistent (i.e., the same input string produces the same output each time) and reasonably collision-resistant (i.e., two different inputs are unlikely to produce the same output).
2. Identify string(s) (such as local-parts of email addresses) in a message that need to be redacted. Call these strings the "private data".
3. For each piece of private data, apply the selected transformation mechanism.
4. If the output of the transformation can contain bytes that are not printable ASCII, or if the output can include characters not appropriate to replace the private data directly, encode the output with the base64 algorithm as defined in Section 4 of [BASE64], or some similar translation to a form valid replacement in the original context. For example, replacing a local-part in an email address with transformation output containing an "@" character (ASCII 0x40) or a space character (ASCII 0x20) is not permitted by the specification for local-part ([SMTP]), so the transformation output needs to be encoded as described.
5. Replace each instance of private data with the corresponding (possibly encoded) transformation when generating the report. Note that the replaced text could also be in a context that has constraints such as length limits that need to be observed.

This has the effect of obscuring the data (in a potentially irreversible way) while still allowing the report recipient to observe that numerous reports are about one particular end user. Such detection enables the receiver to prioritize its reactions based on problems that appear to be focused on specific end users that may be under attack.

4. Transformation Mechanisms

This memo does not specify a particular transformation mechanism as a requirement. The interoperability that this memo seeks to provide is enabled by the consistency of the transformation.

The issue of the security of the transformation, frustrating attempts to reverse the transformation, is a matter of local policy. A continuum of possible transformations exists, from trivial ones such as rot13, CRC32 and base64, through strong cryptographic encodings such as [HMAC] and even full encryption, or private transformations such as mapping an email address to an internal customer number. An operator wishing to perform report redaction needs to select a consistent transformation that obscures the private data and is resilient to attempts to extract the original data to the extent required by local policy, keeping in mind that the environment in

which the transformation is operating is not a highly secure one. See Section 5.3 for further details of this issue.

An implementation MAY choose any transformation that has a reasonably low likelihood of collision.

5. Security Considerations

5.1. General

General security issues with respect to these reports are found in [ARF].

5.2. Digest Collisions

Message digest collisions are a well-understood issue. Their application here involves a report receiver improperly concluding that two pieces of redacted information were originally the same when in fact they are not. This can lead to a denial of service, where the inadvertently improper application of complaint data causes unjustified corrective action. Such cases are sufficiently unlikely as to be of little concern.

5.3. Information Not Redacted

Although the identity of the user causing a report to be generated can be obscured using this mechanism, other properties of a message (such as the Message-ID field) that are not redacted could be used to recover the original data by locating them in the message logs of the originating system or via other data correlation techniques. It is incumbent on the report generator to anticipate and redact or otherwise obscure such data, or accept that such recovery is possible even from the very simplest kinds of feedback.

It is for this reason that the normative portions of this memo do not include stronger assertions about cryptography used in the transformation. Given the ultimate recoverability of the redacted information, the cryptographic strength of the transformation is not a critical security measure.

The process of redacting a feedback report satisfies a privacy requirement established by local policy, and is not meant to provide strong security properties.

[FBL-BCP] and Section 8 of [ARF] discuss topics related to establishment of bilateral agreements between report producers and consumers. The issues raised here are also things to be considered

when establishing such agreements.

6. Privacy Considerations

While the method of redaction described in this document may reduce the likelihood of some types of private data from leaking between ADMDs, it is extremely unlikely that report generation software could ever be created to recognize all of the different ways that private information could be expressed through human written language. If further protections are required, implementers may wish to consider establishing some sort of out-of-band arrangements between the relevant entities to contain private data as much as possible.

7. IANA Considerations

This memo includes no request to IANA.

[RFC Editor note: This section may be removed prior to publication.]

8. References

8.1. Normative References

[ARF] Shafranovich, Y., Levine, J., and M. Kucherawy, "An Extensible Format for Email Feedback Reports", RFC 5965, August 2010.

[BASE64] Josefsson, S., "The Base16, Base32, and Base64 Data Encodings", RFC 4648, October 2006.

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

8.2. Informative References

[FBL-BCP] Falk, J., "Complaint Feedback Loop Operational Recommendations", RFC 6449, November 2011.

[HMAC] Krawczyk, H., Bellare, M., and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, February 1997.

[SMTP] Klensin, J., "Simple Mail Transfer Protocol", RFC 5321, October 2008.

Appendix A. Example

Assume the following input message:

```
From: alice@example.com
To: bob@example.net
Subject: Make money fast!
Message-ID: <123456789@mailer.example.com>
Date: Thu, 17 Nov 2011 22:19:40 -0500

Want to make a lot of money really fast?  Check it out!
http://www.example.com/scam/0xd0d0cafe
```

On receipt, bob@example.net reports this message as abusive through whatever mechanism his mailbox provider has established. This causes an [ARF] message to be generated. However, example.net wishes to obscure Bob's email address lest it be relayed to the offending agent, which could lead to more trouble for Bob.

Thus, example.net plans to redact the local-part of the recipient address in the To: field. Local policy and security requirements suggest the algorithm known as "H" (a hash of a key concatenated with the data to be obscured) using SHA1 is adequeate. It has thus selected a redaction key of "potatoes", and the private data in this case is the string "bob". The concatenation of "potatoesbob" is digested with SHA1 and then base64-encoded to the string "rZ8cqXWGiKHzhz1MsFRGTysHia4=".

Therefore, when constructing the ARF message in response to Bob's complaint, the following form of the received message is used in the third part of the ARF report:

```
From: alice@example.com
To: rZ8cqXWGiKHzhz1MsFRGTysHia4=@example.net
Subject: Make money fast!
Message-ID: <123456789@mailer.example.com>
Date: Thu, 17 Nov 2011 22:19:40 -0500

Want to make a lot of money really fast?  Check it out!
http://www.example.com/scam/0xd0d0cafe
```

Note, however, that it is possible the redacted information can be recovered by agents at example.com searching their logs for the original envelope associated with the message, by correlating with the Message-ID contents which were not redacted here. It is expected that feedback loops generating such reports involve senders that have been vetted against such information leakage.

Appendix B. Acknowledgements

Much of the text in this document was initially moved from other MARF working group documents, with contributions from Monica Chew, Tim Draegen, Michael Adkins, and other members of the Messaging Anti-Abuse Working Group. Additional feedback was provided by John Levine, S. Moonesamy, Alessandro Vesely, and Mykyta Yevstifeyev.

Authors' Addresses

J.D. Falk (editor)
Return Path
100 Mathilda Place, Suite 100
Sunnyvale, CA 94086
US

Email: ietf@cybernothing.org
URI: http://www.returnpath.net/

M. Kucherawy (editor)
Cloudmark
128 King St., 2nd Floor
San Francisco, CA 94107
US

Email: msk@cloudmark.com