DKIM Working Group
Internet-Draft
Intended status: Informational
Expires: September 29, 2011

M. Kucherawy
Cloudmark

March 28, 2011

RFC4871 Implementation Report
draft-ietf-dkim-implementation-report-06

Abstract

This document contains an implementation report for the IESG covering DomainKeys Identified Mail (DKIM) in support of the advancement of that specification along the Standards Track.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 29, 2011.

Table of Contents

1. Introduction

DomainKeys Identified Mail (DKIM), published in May 2007, has reached a level of maturity sufficient to consider its advancement along the standards track. Enclosed is a summary of collected interoperability data provided from sources that are aggregating such information as well as from a more formal DKIM interoperability event that took place in October 2007.

2. Definitions

DomainKeys Identified Mail is defined in [DKIM].

Various terms specific to email are used in this document. Their definitions and further discussion can be found in [EMAIL-ARCH].

3. DKIM Interoperability Event

In October 2007, Alt-N Technologies of Dallas, Texas hosted an interoperability and testing event at their headquarters. Twenty organizations sent engineers and their various DKIM implementations to connect to a private internal network and exchange test messages and tabulate observed results.

3.1. Participants

The interoperability event included participants from all of the following organizations: Alt-N Technologies, AOL Inc., AT&T Laboratories, Bizanga Ltd., Brandenburg InternetWorking, Brandmail Solutions, ColdSpark, Constant Contact, Inc., DKIMproxy, Domain Assurance Council, Google Inc., ICONIX Inc., Internet Initiative Japan (IIJ), Ironport Systems, Message Systems, Port25 Solutions, Postfix, Sendmail, Inc., StrongMail Systems, and Yahoo! Inc. Most of the participants traveled to Dallas and participated in person, but a few operated remotely.

Nearly all of the implementations were based on disjoint code development projects. A few were based on a common open source base project.

3.2. Testing Methodology

Participants were encouraged before the event to craft a set of test messages meant to exercise their own implementations as well as those of the other participants, both in terms of successful verifications as well as some expected to fail. Test cases were developed with the intent of confounding verifiers that may not have implemented the [ABNF] of [DKIM] correctly.

The participants set up Mail Transfer Agents (MTAs) equipped with their own DKIM signing and verifying modules, and their own tools to generate mail to be signed along with tools to analyze the results post-verification. They then sent their own batteries of test messages, looking for both expected and unexpected failures in response. Some implementations included "auto-responders" that would reply with verification results, while others simply collected the results that would then be shared manually.

3.3. Observations

All of the packages implemented all of the required portions of [DKIM] in terms of both signature and key features. Most of the packages implemented all of the optional features of both signatures and keys. There were at least two implementations of each optional

feature.

The interoperability testing was largely successful. As might be expected, there were many verification false negatives or false positives that were the result of bugs in corner cases of some of the implementations presented for testing. In such cases the developers were able to identify the issue as resulting from their own mis-reading of the specification and not an error in the specification itself.

Several of the failures did occur as a result of specification ambiguities. The participants discussed each of these in turn and were able to come to consensus on how they believed the specification should be changed to resolve them.

The participants agreed to keep the results about the specific tests private. Accordingly, those data are not presented here.

3.4. Results

The handful of interoperability issues described above that referred to weaknesses or ambiguities in [DKIM] resulted in several errata being opened via the RFC Editor web site. These are being addressed in an RFC4871bis draft effort that is now starting from within the DKIM working group.

The errata items, in summary:

- explicit canonicalized forms of empty bodies for each canonicalization method, along with their SHA1 and SHA256 hash values (errata #1376 and #1377)
- clarification about normative text regarding the "a=" tag (errata #1378)
- ABNF corrections regarding the "z=" tag (errata #1379)
- informative discussion regarding the "x=" tag (errata #1380)
- normative clarifications about "q=", "h=", "k=", "s=" and "t=" tags (errata #1381 and #1382)
- correction of "g=" description to match its ABNF (errata #1383)
- clarifications about "relaxed" body canonicalization (errata #1384)

- correction to the signature example (errata #1386)
- ABNF corrections regarding the "h=" tag (errata #1461)
- ABNF corrections regarding the "v=" tag (errata #1487)
- discussion of DomainKeys compatibility (errata #1532)
- discussion about what constitutes the actual value of the "b=" tag (errata #1596)

4. Collected DKIM Interoperability and Use Data

Several implementations are collecting private data about DKIM use, signature survivability, which properties of the base specification are observed in public use, etc. This section includes collection methods and summary reports provided by those implementations.

4.1. The OpenDKIM Project

The OpenDKIM Project (http://www.opendkim.org) is an open source project providing a DKIM support library, an email filter for use with Mail Transfer Agents (MTAs), and a set of tools to assist with deployment of DKIM.

4.1.1. Details

Recent releases have included an optional feature to record statistics about messages with and without DKIM signatures. Sites enabling this feature can choose to share the data with the project's development team as part of this interoperability report work. The data can be anonymized to conceal the sending domain and client IP addresses, though these data are passed through a one-way hash to enable collation of data from common sources.

4.1.2. Results

At the time of writing of this document, seven months of data had been collected. The results of this effort are as follows:

Reporting Hosts: 21 individual MTAs representing nine distinct ADMDs

Total Messages: 11367042

Signatures: 8146244 messages (71.7%) were not signed; 3186313 (28.0%) had one signature; 34156 (0.3%) had two signatures; the remainder (less than 0.01%) had more.

Signing Algorithms: 53.4% of signatures used "rsa-sha1", while the balance used "rsa-sha256".

Header Canonicalization Algorithms: 14.8% of signatures used "simple", while the balance used "relaxed"; when grouped by domains, 11.8% of domains used "simple" while the balance used "relaxed".

Body Canonicalization Algorithms: 26.2% of signatures used "simple", while the balance used "relaxed"; 19.1% of observed signing domains used "simple" while the balance used "relaxed".

Keys in Test Mode: 54.5% of keys retrieved from the DNS were tagged as being in test mode.

Keys with Specific Granularity: 434 keys were retrieved that had specific names in their "g=" tags.

Keys with Syntax Errors: Less than 0.1% of keys retrieved from the DNS had syntax errors.

DomainKeys Compatibility: 1% of the retrieved keys appeared to be intended for use with the older DomainKeys proposal rather than DKIM

AUID use: 49.8% of signatures did not contain an explicit AUID ("i=" value). Of those that did, 89.4% used a domain matching the SDID ("d=" value). Across all "i=" tags present, 43.6% provided no local-part, 52.8% included a local-part matching the one found in the From: field, and the remainder had a different local-part.

Missing Keys: 2.1% of signatures received referenced keys that were not found in the DNS

Optional Signature Tags: Of the optional signature tags supported by the base specification, "t=" was seen 45.3% of the time, "x=" was seen 3.4% of the time; "l=" was seen 3.7 of the time; "z=" was seen 5.9% of the time. (The "z=" statistic is likely inflated due to the nature of the sampling.)

Body Length Limits: Of the signatures for which "l=" was used, 11.9% of them signed none of the body, and 92.5% of the rest had the body extended after signing.

Signature Pass Rates: Overall, 92.3% of observed signatures were successfully verified.

Pass Rates for Non-List Mail: Where "list mail" is defined as any mail bearing one of the header fields defined in [LIST-ID] or in [LIST-URLS], or a "Precedence: list" field, selecting only for mail that is not list mail revealed a successful verification rate of 93.5%; selecting only for list mail produced a 90.5% success rate.

DNSSEC: There has been scant but detectable uptake in the checking of whether or not DKIM or ADSP records in the DNS are protected by DNSSEC.

Common errors: The top five verification errors observed: Key not found in DNS (28.3%), key granularity mismatch (13.8%), DNS retrieval failure such as timeouts (10.6%), key type unknown (2.3%), key timestamp is in in the future (2.2%).

Detected Header Field Changes: A subset of the reporting sites are capable of reporting header fields known to have been changed in transit (e.g. when "z=" tags were used by the signer). In such cases, changes to the "To:" field were more common than any other by more than an order of magnitude.

Most Commonly Signed Fields: From: (100%), Subject: (95.8%), Date: (95.6%), To: (95.4%), MIME-Version: (92.9%), Content-Type: (85.0%), Message-Id: (75.6%), Reply-To: (43.8%), Received: (35.2%), List-Unsubscribe: (31.3%). All others are below 30%.

Identities: 76% of the signatures observed included a "d=" value matching the domain in the From: field.

Low-use Signing Domains: 50942 unique signing domains were observed. Of these, 26% of them sent a single signed message in the sample period, 14.4% sent two and 8.3% sent three.

4.1.3. Conclusions

The results of the OpenDKIM work are updated constantly as more data feeds come online and more data are reported. Based on the data available at the time of writing, some conclusions are possible.

At least some implementations of all of the optional signature features, all of the canonicalization combinations and all of the signing algorithms are in general use. None of the features had zero use counts.

Overall signature pass rates are generally quite high. The impact of signature survivability when correlated against Mailing List Manager (MLM) activity is detectable based on observed data. More research into this is recommended. The DKIM Working Group is already working on an Informational draft to discuss those issues.

That the "To" field is the one most often associated with verification failures suggests some MTAs handling the message are correcting cases where the field is improperly formed. A common case is failing to quote the comment portion of that field when required

to do so by [MAIL]. Such corrections cause signatures to become invalid. The working group may want to consider revising Section 5.5 of [DKIM] to either reduce the list of recommended header fields for signature content or provide some additional text warning of this potential issue.

The high counts (i.e., nearly half) of low-use signing domains suggest that spammers, who typically rotate domain names with high frequency, have adopted DKIM as a tool to try to get through message filters.

4.2. AOL, Inc. Data

A one-day summary of observed traffic from AOL, Inc. reports the following:

Ratio of DKIM-signed mail: 42%

Properly formed signatures: 1.4 billion

Malformed signatures: 3000

Unique signing domains: 50,000-90,000

Key retrieval errors: 14 million (1%)

Signature refers to nonexistent domain: 10 million (0.7%)

Signature refers to nonexistent key: 36 million (2.5%)

Signature refers to revoked key: 138,000 (~0%)

Verified signatures: 1.2 billion (85.7%)

AUID matches From: domain: 1.2 billion (85.7%)

Failed signatures (body changed): 78 million (5.6%)

Failed signatures (other): 34 million (2.4%)

Expired signatures: less than 1 million (~0%)

4.3. Google Mail Data

Google Mail reports the following:

Unsigned mail: 72.1%

AUID matches From: domain: 68.7%

Signed mail that verified: 14.7%

Signed mail that verified in test mode: 11.7%

Signed mail that failed: 0.2%

Signed mail that failed in test mode: 0.2%

Body hash mismatch: 0.5%

Signature missing required parameters: 0.3%

Granularity mismatch: 0.2%

These data are reported based on an implementation that only evaluates one signature per message.

All other reportable anomalies occurred in vanishingly small percentages.

5. IANA Considerations

This memo contains no actions for IANA.

6. Security Considerations

This document is an implementation report and thus has no security considerations.

7. References

7.1. Normative References

[DKIM] Allman, E., Callas, J., Delany, M., Libbey, M., Fenton, J., and M. Thomas, "DomainKeys Identified Mail (DKIM) Signatures", RFC 4871, May 2007.

[EMAIL-ARCH] Crocker, D., "Internet Mail Architecture", RFC 5598, July 2009.

7.2. Informative References

[ABNF] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC 5234, January 2008.

[LIST-ID] Chandhok, R. and G. Wenger, "List-Id: A Structured Field and Namespace for the Identification of Mailing Lists", RFC 2919, March 2001.

[LIST-URLS] Neufeld, G. and J. Baer, "The Use of URLs as Meta-Syntax for Core Mail List Commands and their Transport through Message Header Fields", RFC 2369, July 1998.

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

Appendix A. Acknowledgements

The author wishes to acknowledge the following for their review and constructive criticism of this document: Dave Crocker, Tony Hansen, Jeff Macdonald, S. Moonesamy and Rolf Sonneveld.

The author also wishes to acknowledge Margot Koschier of AOL, Inc., Monica Chew of Google, and the members of The OpenDKIM Project for providing data used in this report.

The working group expresses its profound thanks to Alt-N Technologies for graciously hosting the 2007 DKIM interoperability event.

Author's Address

Murray S. Kucherawy
Cloudmark
128 King St., 2nd Floor
San Francisco, CA 94107
US

Phone: +1 415 946 3800
Email: msk@cloudmark.com