Workgroup: dkim
Internet-Draft:
draft-chuang-dkim-replay-problem-01
Published: 10 February 2023
Intended Status: Informational
Expires: 14 August 2023
Authors: W. Chuang Google, Inc.
A. Robinson Google, Inc.
B. Gondwana Fastmail Pty Ltd

DKIM Replay Problem Statement and Scenarios

Abstract

DomainKeys Identified Mail (DKIM, RFC6376) claims some responsibility for a message by associating a domain and protecting the integrity of the covered portion of a message during transit through a digital signature. DKIM survives basic email relaying. In a Replay Attack, the recipient of a DKIM-signed message sends the message further, to other recipients, while retaining the original, validating signature, thereby seeking to leverage the reputation of the original signer. This document discusses the damage this causes to email delivery and interoperability, and the associated Mail Flows. A significant challenge to mitigating this problem is that it is difficult for Receivers to differentiate between legitimate forwarding flows and DKIM Replay.

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 https://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 14 August 2023.

(https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction

DomainKeys Identified Mail (DKIM) is a well-established email protocol [RFC6376]:

DomainKeys Identified Mail (DKIM) permits a person, role, or organization to claim some responsibility for a message by associating a domain name [RFC1034] with the message [RFC5322], which they are authorized to use. This can be an author's organization, an operational relay, or one of their agents. Assertion of responsibility is validated through a cryptographic signature and by querying the Signer's domain directly to retrieve the appropriate public key.

1.1. The problem

Since many domains (including those of bad actors) list DKIM records, receiving systems track the history of messages using a DKIM-based domain name, to formulate a reputation for the name, and then to classify incoming emails. The presence of a DKIM signature

that validates the message ensures that the developed reputation was the result of activity by the domain owner, and not by other parties. Receiving filtering systems contain a rich array of rules and heuristics for assessing email. DKIM is one such identity that this system associates with sender reputation and uses to predict future sender behavior. The filter system helps protect users against spam, phishing and other abuse.

While DKIM Replay was identified as a hypothetical concern during the development of the DKIM standard, that attack has become commonplace: Attackers create, obtain access, or compromise an account at a site with a high reputation. This allows them to generate an email having content that they can (re-)broadcast widely. They send an email from that account, to an external account also under their control. This single message is delivered to the attacker’s mailbox, giving them an email with a valid DKIM signature, for a domain with high reputation. Further, Internet Mail permits sending a message to addresses that are not listed in the content To:, Cc: or Bcc: header fields. DKIM covers portions of the message content, and can cover these header fields, but it does not cover the envelope addresses, used by the email transport service, for determining actual recipient addresses. So this message can then be replayed to arbitrary thousands or millions of other recipients, none of whom were specified by the original author.

Unfortunately DKIM Replay is impossible to detect or prevent with current standards. Email authentication does not distinguish benign forwarding flows from DKIM Replay, as will be described later by itemizing the different forwarding flows and their email authentication patterns.

ARC [RFC8617] is a protocol to securely propagate authentication results seen by forwarders, such as mailing lists that re-post messages, in eMail Flow. It can be used to adjust DMARC [RFC7489] validation as described in section 7.2.1. Because ARC is heavily based on DKIM it has the same replay issue as described in section 9.5.

1.2. Glossary

This document is completely informative and omits normative language as described in [RFC2119].

This document uses delivery terminology from [RFC5598] and [RFC5321] to define the participants in a Mail Flow. In particular [RFC5598] defines mail interactions conceptually from three perspectives which are called *actors* there- users, services (Message Handling Service) or administratively (ADministrative Management Domain). This document primarily works with and expands the list of Message

Handling Services. As noted in [RFC5598] a given implementation might have multiple roles. The following are a subset of the Mail Handling Services defined in [RFC5598] to be used in this document:

- Originator - defined in Section 2.2.1. This works on behalf of the author to post the message to the relay that performs the SMTP store-and-forwarding and to validate the message. The Originator may DKIM sign the message on behalf of the author.

- Alias- as defined in Section 5.1, the Alias updates the RCPT TO envelope recipient but not the address field headers. Often used for *Auto-Forwarding*.

- Mailing Lists- defined in Section 5.3. Receives a message and reposts them to a list of members. May add list-specific header fields e.g. List-XYZ:, etc. May append text to the Subject header, and at the end of the content.

- Receiver- defined in Section 2.2.4. The Receiver works on behalf of the recipient to deliver the message to the inbox and may perform filtering.

All of these Mail Handling Services and those below may add trace and operational headers.

Modern Mail Flow has additional Mail Handling Services. These additional service definitions are:

- Email Service Provider (ESP) Bulk Sender or often abbreviated as Bulk Sender- An originating third-party service, acting as an agent of the author and sending to a list of recipients. They may DKIM sign as themselves but also sign for the author of the message.

- Outbound Filtering Service- The Originator can route some or all of their mail through an Outbound Filtering Service to provide spam or data loss protection services, instead of sending directly to the recipient's server. This service may modify the message and is administratively separate from the Originator.

- Inbound Filtering Service- The Receiver can route mail through an Inbound Filtering Service to provide spam, malware and other anti-abuse protection service. Typically this is done by publishing the Inbound Filtering Service hosts in the Receiver's MX record. This service may modify the message and is administratively separate from the Receiver. Once done filtering, the Inbound Filtering Service relays the messages to the Receiver.

It is useful to broadly identify participants in Mail Flow by functionality as defined in [RFC5598] in terms of submission, transmission and delivery. The SMTP agents are categorized under these as:

*Mail Submission Agent (MSA)

*Mail Transmission Agent (MTA)

*Mail Delivery Agent (MDA)

In addition the user interact with the MSA and MDA via:

*Mail User Agent (MUA).

The above services uses email authentication as defined in the following specifications:

*DomainKeys Identified Mail (DKIM): defined in [RFC6376].

-Note that DKIM replay is defined in [RFC6376] section 8.6.

*Sender Policy Framework (SPF)- defined in [RFC7208].

*Domain-based Message Authentication, Reporting, and Conformance (DMARC): defined in [RFC7489].

Mail Flow is an informal term for the path that messages take when they traverse between some Originator and Receiver and possibly one or more intermediary Message Handling Services. Those intermediary Message Handling Services are administratively distinct from the Originator and Receiver. Direct Mail Flow contains only the Originator and Receiver without intermediary Message Handling Services, whereas Indirect Mail Flow has Originator and Receiver with intermediary Message Handling Services.

2. DKIM Replay

A spammer finds a mail Originator with a high reputation and that signs their message with DKIM. They obtain access to an account at the Originator. This may happen via open enrollment at some Mailbox Provider or Bulk Sender, or via account hijacking for any Originator. The spammer sends a message with spam content from there to a mailbox they control. Taking advantage of the flexibility in DKIM to selectively sign headers, the spammer may intentionally leave out certain headers such as To:, and Subject: that can be added in later without damaging the existing DKIM signature. The spammer reads the signed message from the initial Receiver's inbox and potentially adds the missing headers customized by the ultimate spam victim recipient. The resulting message is then sent at scale

to the victim recipients. In addition to being DKIM authenticated via the spoofed DKIM signature, the spammer can set up SPF authentication on their servers though that will not be aligned with the DKIM. Because the signed message has high reputation, the message with spam content will tend to get through to the inbox. This is an example of a spam classification false negative – incorrectly assessing spam to not be spam.

When large amounts of spam are received by the mailbox provider, the operator’s filtering engine will eventually react by dropping the reputation of the original DKIM signer. Benign mail from the signer's domain then starts to go to the spam folder. This is an example of a spam classification false positive.

In both cases, mail that is potentially wanted by the user becomes much harder to find, reducing its value to the user. In the first case, the wanted mail is mixed with potentially large quantities of spam. In the second case, the wanted mail is put in the spam folder where the user does not expect it.

When the Receiver observes a spam campaign, operators at the Receiver may use additional signals such as SPF to reject spam. As described in the next section, SPF works well for Direct Mail Flows but is problematic for Indirect Mail Flows that are an important part of the email ecosystem.

3. Mail Flow Scenarios

The following section categorizes the different Mail Flows by a functional description, and effect on email authentication. The Mail Flow categorization in this section by email authentication is meant to demonstrate why DKIM replay is so hard to distinguish from benign Mail Flow particularly for Indirect Mail Flows. Email authentication, when present, is defined in terms of DKIM and SPF provided by some sender and validated by the Receiver. Flows involving the same service as both a forwarder and as a destination service can be thought of as two independent (though related) Mail Flows for the purposes of authentication signal propagation. Some intermediary Mail Handling Services can be composed to make even more complex Indirect Mail Flows.

3.1. Direct Mail Flow

In this case the Originator delivers mail directly to the Receiver without any intermediary Mail Handling Services.

Email authentication:

* SPF- the Originator's IPs are plainly observable by the Receiver, enabling successful authentication by the Receiver.

*DKIM- authenticates as there is no intermediary that breaks the DKIM signature.

3.2. Bulk Sender

Bulk Senders is a special case of Direct Mail Flow. The ESP acts as the Originator of messages on behalf of the Author given a message body and a list of recipients.

Email authentication:

*SPF based on a customer domain requires careful coordination with that customer, since it is their SPF record and not the ESP's.

*DKIM- the ESP may generate a DKIM signature based on the Author's domain and/or one based on the ESP 's own. Requires coordination when the customer's domain is used.

The following are examples of Indirect Mail Flows.

3.3. Transmission through an Outbound Filtering Service

In this case, the Originator relays email to Outbound Filtering Service that provides spam or data loss protection before sending the message onto the Receiver.

Email authentication:

*SPF authentication is possible, but may not be advisable. The Originator does this by publishing an SPF policy that covers the Outbound Filtering Service IPs but this IP sharing weakens authentication.

*DKIM may break if the filter performs any modifications of the message such as URL rewriting or attachment stripping. Such modifications could be supported if the filtering service has the ability to resign for DKIM on behalf of the Originator though the Originator increases risk of losing control of their private key.

3.4. Transmission through an Inbound Filtering Service

In this case an Inbound Filtering Service provides spam and abuse protections for the Receiver. The Receiver sets this up by having its MX record point to the Inbound Filtering Service and the Inbound Filtering Service relays the message to the Receiver.

Email authentication:

*SPF will be unauthenticated with the original MAIL FROM domain as the required connecting IP information is only available during

the SMTP transaction between the Originator service and the Inbound Filtering Service. The Inbound Filter Service's EHLO domain or rewritten MAIL FROM domain may authenticate.

* DKIM may break if the filter performs any modifications of the message such as URL rewriting or attachment stripping.

Because email authentication completely fails with aggressive inbound filtering, the Receiver typically will have to trust the Inbound Filtering Service to perform email authentication and DMARC policy enforcement.

3.5. Mailing List

Messages are sent to a mailing list which takes delivery, possibly append text to the Subject header, and at the end of the content, and then re-posts the message to an expanded list of recipients.

Email authentication:

* SPF will be unauthenticated with the original MAIL FROM domain as the required connecting IP information is only available during the SMTP transaction between the Originator service and the Mailing List. The Mailing List's EHLO domain or rewritten MAIL FROM domain may authenticate.

* DKIM may break if the Mailing List modifies the message. To compensate, some Mailing Lists DKIM signs the message with the identity of the Mailing List. Because this may break DMARC From header alignment, the Mailing List may also rewrite the From address.

3.5.1. Alias aka Auto-Forwarding

Automatically forwards mail to a new recipient, updating the envelope from address (MAIL FROM).

Email authentication:

* SPF will be unauthenticated with the original MAIL FROM domain as the required connecting IP information is only available during the SMTP transaction between the Originator service and the Alias. The Alias EHLO domain or rewritten MAIL FROM domain may authenticate.

* DKIM- authenticates as the Alias does not modify the messages.

In all cases of Indirect Mail Flow, SPF MAIL FROM authentication fails. To authenticate SPF, the intermediary may rewrite the MAIL FROM to provide its domain as the MAIL FROM identity or publish an

EHLO domain but this identity won't align with the DKIM domain. The Indirect Mail Flow pattern of typically passing DKIM and failing or misaligned SPF is the same DKIM Replay.

4. Proposed Solution Space

Here are some potential solutions to the problem, and their pros and cons

1. The originator includes the ENVELOPE-TO address in the signature and the Receiver verifies against the actual recipient.

   *Pros:

   -Avoids replay to destination addresses not anticipated by the DKIM signer thereby preventing DKIM replay.

   *Cons:

   -Some Indirect Mail Flows will not authenticate if they rewrite the ENVELOPE-TO. This problem is similar to SPF in being unable to support some Indirect Mail Flows.

2. Cache known DKIM signatures. Since the exact same signature is being replayed repeatedly, this allows a Receiver to detect whether a message is part of a DKIM Replay set, and suppress it.

   *Pros:

   -No changes to the DKIM standard required

   *Cons:

   -Mailing list traffic, exploder aliases, or regular BCCs will also show up as duplicates, so this is very much a heuristic guess of whether the amount of duplication is expected or not. This may lead to spam filtering false positives.

3. Strip DKIM signatures on mailbox delivery to the inbox. No DKIM signature will be available to resend by the spammer.

   *Pros:

   -Messages delivered to a mailbox can not be DKIM replayed any more.

   -No changes to the DKIM standard required.

*Cons:

-Does not help against evil or non-participanting Mailbox Provider.

-May not support MUA services that wish to independently check message integrity. A new standard could be developed to sign twice and strip only the over the wire signature used for email authentication, and leave a long term signature for message integrity. \

4. Add a per-hop signature, specifying the destination domain for the next hop

*Pros:

-Messages with this kind of signature cannot be replayed down a different pathway, since the destination won't match.

*Cons:

-Requires every site along the path to support this spec, so it will need to detect whether the next stop is making a commitment to follow the spec.

-If email goes outside of sites with this spec (without disclosure), traversing a naive forwarder remains indistinguishable from replay.

5. Privacy Considerations

6. Security Considerations

7. IANA Considerations

This document has no IANA actions yet.

8. Normative References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <https://www.rfc-editor.org/rfc/rfc1034>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/

RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC5321] Klensin, J., "Simple Mail Transfer Protocol", RFC 5321, DOI 10.17487/RFC5321, October 2008, <https://www.rfc-editor.org/rfc/rfc5321>.

[RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, DOI 10.17487/RFC5322, October 2008, <https://www.rfc-editor.org/rfc/rfc5322>.

[RFC5598] Crocker, D., "Internet Mail Architecture", RFC 5598, DOI 10.17487/RFC5598, July 2009, <https://www.rfc-editor.org/rfc/rfc5598>.

[RFC6376] Crocker, D., Ed., Hansen, T., Ed., and M. Kucherawy, Ed., "DomainKeys Identified Mail (DKIM) Signatures", STD 76, RFC 6376, DOI 10.17487/RFC6376, September 2011, <https://www.rfc-editor.org/rfc/rfc6376>.

[RFC7208] Kitterman, S., "Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1", RFC 7208, DOI 10.17487/RFC7208, April 2014, <https://www.rfc-editor.org/rfc/rfc7208>.

[RFC7489] Kucherawy, M., Ed. and E. Zwicky, Ed., "Domain-based Message Authentication, Reporting, and Conformance (DMARC)", RFC 7489, DOI 10.17487/RFC7489, March 2015, <https://www.rfc-editor.org/rfc/rfc7489>.

[RFC8617] Andersen, K., Long, B., Ed., Blank, S., Ed., and M. Kucherawy, Ed., "The Authenticated Received Chain (ARC) Protocol", RFC 8617, DOI 10.17487/RFC8617, July 2019, <https://www.rfc-editor.org/rfc/rfc8617>.

Appendix A. Acknowledgments

Many thanks goes to Dave Crocker for his patient editing and clarification. Thanks also goes to Emanuel Schorsch and Evan Burke for their advice.

Authors' Addresses

Weihaw Chuang
Google, Inc.

Email: weihaw@google.com

Allen Robinson
Google, Inc.

Email: arobins@google.com

Bron Gondwana
Fastmail Pty Ltd

Email: brong@fastmailteam.com