pre-workgroup
Internet-Draft
Expires: April 27, 2006

D. Otis
Trend Micro, NSSG

October 24, 2005

Review of Threats Associated with Email and DomainKeys Identified Mail (DKIM)
draft-otis-dkim-threats-01

Status of this Memo

Copyright Notice

Abstract

This document is intended to provide an alternative perspective to the document prepared by Jim Fenton for DomainKeys Identified Mail (DKIM), although it borrows from his substantial effort. This review removes emphasis on email-address domains, as DKIM allows signatures to be independent of the email-address. This document also considers the impact of adding an opaque-identifier and implementing abusive message replay abatement. This document considers threats against Internet mail and threats created when employing a signature-based

method for establishing an accountable domain for a message, in particular DKIM. This document also ranks threat levels, modes of access, Bad Actors and their capabilities, and possible motivations for various attack scenarios.

Table of Contents

1. Introduction

Message signing, as exemplified by DomainKeys Identified Mail (DKIM) [I-D.allman-dkim-base], is a mechanism to allow an assertion of an accountable domain for an email message in transit. The assertion is made by means of a digital signature included within a header. This signature also validates the integrity of selected headers and message body content subsequent to signing the message. This review is based upon the work of Jim Fenton, but differs from the perspectives regarding the role of an email-address and how replay, intra-domain, and DoS attacks may be handled.

For example, on the DKIM list Jim suggested Sender-ID's path-registration and authorization mechanism could be a possible solution for a message replay abuse problem. Unfortunately path-registration imposes upon DKIM problematic limitations that would also be very disruptive. Path-registration would also essentially constrain mailbox-addresses to specific providers. In addition, path-registration already exposes email-address domain owners to risks where path authorization may form the basis for accrual of unfair reputations. Within shared environments, requisite checks to ensure email-address domain exclusivity may not have been made. Such checks may be beyond the control of the email-address domain owner, while those who are in control remain unaccounted.

With DKIM, once an accountable domain has been established for a message and where the reputation of this domain can be defended, the recipient may assess the reputation accrued by the signing-domain when deciding whether to accept the message. This assessment can be made using locally-maintained white-lists, and reputation/ accreditation services. By applying a signature, the conduct permitted by the signing-domain may be accurately accrued to establish a valid reputation. Good conduct is generally maintained when there are expectations that future messages will be accepted by the recipient from the signing-domain.

2. Scope of DKIM

DKIM verifies the association of a specific domain with a message using a signature and a public key. This mechanism also ensures selected headers and body content have not been altered since the domain's association. The DKIM effort is not intended to address threats associated with message confidentiality nor provide a signature suitable for long-term archival. The scope of DKIM does not include semantics for reputation or accreditation services or white-listing practices. DKIM does not provide a direct method to verify the identity of a message's author. DKIM does not provide safe mechanisms for authorizing messages associated with different

domain signatures.

3. Vulnerabilities

Email is exposed to several security related threats where exploitation of a vulnerability often results in substantial damages. The following table provides a general overview of vulnerabilities and side-effects created by defensive strategies.

Vulnerability	Damage	Prevalence	Severity
Unfair Reputations	Loss of Service	Low	Medium
Collateral IP Blocking	Loss of Service	Low	Medium
Filtering Errors	Loss of Service	Low	Medium
Junk Folder	Loss of Service	Low	Medium
Delete w/o DSN	Loss of Service	Low	Medium
DoS	Loss of Service	Low	Medium
Name Blocking	Loss of Service	Low	High
Message Spoofing	Defrauded User	Medium	High
OS Flaws	Compromised System	Medium	High
Stolen Passwords	Compromised Account	Medium	High
Malware Payloads	Compromised System	High	High
Timing Analysis	Compromised Key	Low	High
Network Exploits	Compromised Privacy	Low	High

An Unfair Reputation regards assessments made against the email-address. These unfair assessments may occur when the email-address domain owner is assumed to control the use of their email-address domain. The email-address domain owner may have been obliged to authorize the shared systems of a service provider when registering prescribed email paths. Checks are often not performed by an unaccounted provider that should ensure only the owner is able to utilize their email-address domain. The public nature of system authorizations and prevalence of compromised systems place the email-address domain owners at great risk when reputation is based upon the email-address.

Collateral IP Blocking occurs when email systems are being shared and many email-address domains utilize common IP addresses. When one of these email-address domains displays abusive conduct, an IP address based reputation service may list the IP address and consequently block all other email-address domains sharing the IP address.

Filtering Errors may occur when erroneously relying on a phrase or name. These errors cause the normal processing of the message to be altered. Abusers often spoof many elements within their message

largely to avoid being filtered and this may increase the filtering program's sensitivity to otherwise innocent domains. This type of erroneous sensitivity could be viewed as another type of unfair reputation. The message may be "bounced", placed into the "Junk Folder", or deleted without the issuing of a Delivery Status Notification (DSN). With the breakdown of email policy enforcement, email filtering has become a common alternative to manual, and also highly error-prone, deletion of unwanted email.

The Junk Folder has become the catch-all repository of questionable email. The increased use of multi-level acceptance criteria also increases the portion of email placed into the Junk Folder. Ideally, acceptance criteria would provide a binary go/no-go result. Stronger methods of authentication that singularly identify an accountable entity may assist in achieving such a goal.

Deletion without Delivery Status Notification may occur when messages have been accepted but then, perhaps through the use of analytical heuristics, the message is subsequently identified as unwanted. In the case of message deletion, the provider does not notify the sender since even the bounce-address is typically considered invalid. This mode of operation represents a significant reduction in the integrity of email delivery.

Denial of Service attacks may not be intentional, and could be the result of unsolicited bulk email. An enterprise attempting to run their own email service may find their networks unable to deal with the vast amount of unwanted email. Being able to reject unwanted email early in the exchange is critical when network resources are limited. Signatures carried within the message will not allow for early rejection and thus not offer any DoS protection. Simply dropping the connection may result in a storm of retries. DoS protections based upon a domain, rather than an IP address, can be achieved by verifying the EHLO name, as it still permits early rejection while also avoiding IP address collateral blocking, see [I-D.crocker-csv-intro] and [I-D.crocker-csv-csa]. Domain-based assessments derived from the EHLO or a domain signature could utilize a name based reputation service, commonly referred to as a Right-Hand-Side Black-hole List (RHSBL). Attacks from unlisted domains would be retained together with the IP addresses within a local list.

Name Blocking may result from unfair reputation accrual. Such accrual could occur when feedback is based upon email-address domains held accountable for having authorized a system sending abusive email. The damage would be much worse than that caused by collateral IP address blocking. In the case where the IP address is blocked, the email-address domain owner would be able to obtain services elsewhere as a remedy. In the case of Name Blocking, there may not

be any remedy, which represents a serious problem. This problem may become endemic with email-address authorization mechanisms.

Message Spoofing uses many techniques to mislead either the recipient or a message filter. Often the email-address domain appears random, or contains several randomly generated domain labels. Such randomly generated labels may still be valid when a DNS wildcard resource record is used. In the case where some level of authentication is asserted by the MUA, the domain used to achieve the authentication may rely upon the recipient seeing the "pretty-name" rather than the actual email-address. For various reasons, methods that attempt to select a visible header to authorize an email-address domain, still may permit hidden headers. The complexity of the selection algorithms may also confuse the average recipient, where any indication of the message having been authorized may be exploited.

Operating System flaws will always exist. Some operating systems offer better secured internal communication and program execution. Minimizing the exposure of system services to external access reduces the number of exploits possible. Automated system monitoring is often needed to react to attack attempts. Otherwise, the scale of deployment may require overwhelming manual monitoring.

Stolen Passwords are a common problem. There are many enterprises that use protocols that send passwords in the clear. With the prevalence of wireless networks with weak protections, passwords sent in the clear should be considered the same as publishing them. In addition, many of the programs that compromise systems also do keyboard and paste buffer logging sent covertly to the criminals stealing sensitive information. Even access to a microphone near the keyboard may reveal passwords. Once access is gained, even with a low privilege, many applications automatically assert the password when sending or receiving messages.

Malware Payloads may be carried in items that receive special handling. Examples could be pictures or scripts embedded within a message or referenced via URLs. It could also be attachments added to the message, often where the name of the attachment has been obfuscated to convince the recipient they can safely invoke the operating system's default handling routines.

Timing Analysis is an attack method that takes advantage of knowing the algorithm used for processing the signature. When the private portion of the key is involved in the process, care must be taken to ensure the time performing the operation is not revealed. This timing enables techniques that deduce the private key. Even with random latency variations introduced by the network, this variation can be significantly reduced by finding the median from additional

measurements.

Network Exploits are also becoming more common. These attacks exploit various elements of the network infrastructure, often misdirecting network traffic. This may involve falsified configuration information, falsified connection redirects or hijacks, poisoned domain name servers, and even corrupted routing elements.

4. Security Requirements

The use of private keys within the signing MTA server increases the level of security required to safely provide the DKIM signing services. External access to non-essential services should be prevented using appropriate measures. The Operating System should also be monitored for execution of unauthorized programs and access attempts to otherwise protected services. When the server is running within a virtual machine, special care must be exercised with respect to timing attacks on private keys where even processing loads may reveal sensitive information. The generation of a signature should be staged to mask the actual time expended as a means to protect the private key. As email is often held in queues during processing, results may be held for an assured duration to conceal the time related to signing.

5. Ranking of Bad Actors

The problems confronted by DKIM can be generalized as the result of abusive acts by individuals taking advantage of the open nature of email. For the purposes of this document, these abusive individuals will be referred to as Bad Actors. Bad Actors have become more sophisticated, and their motivations have become increasingly criminal in nature.

At the low end of severity, Bad Actors may represent unsophisticated individuals taking advantage of many commercially available tools. These tools may facilitate messages being sent in bulk, or may employ criminal strategies that avoid being identified and subsequently blocked. Unsolicited messages sent in bulk often becomes the basis used for blocking future exchanges. As newer methods of identification are attempted, often these bulk distribution tools represent a significant portion of applications adopting the newer protocol extensions. The adoption of the extensions is often based upon another strategy where changing identities becomes a small cost for sustaining their nefarious enterprise.

The next level of severity could be considered a surreptitious service provider that specializes in sending unsolicited email. These Bad Actors often employ infrastructure specifically designed to

obfuscate their identity. Their infrastructure may include open-proxies, open-relays, and the use of multiple points of access which take advantage of asymmetric routing as a means to disguise source IP addresses. Their infrastructure may also be established using thousands of compromised systems that may have been leased.

Another technique used by Bad Actors is to send to low priority backup MTAs with the expectation messages will be bounced and the intended recipient is contained within the bounce-address. This bounced message technique is another method used to obfuscate the source IP address of the message. Although this IP address is typically found in the Received headers, the reputation of this Received header IP address is not always checked. Bad Actors offering such services often provide email address lists to their clientele that may have been obtained from compromised systems, harvested, purchased, or discovered by means of dictionary attacks.

The highest level of severity would be represented by criminals who intend to commit fraud and who are financially-motivated. These Bad Actors are becoming organized and specialized with rapidly growing sophistication and threaten not only email, but the network itself. These Bad Actors should be expected to employ all of the above mechanisms, in addition to attacks on Internet infrastructure, such as DNS cache-poisoning attacks, and IP routing attacks via compromised network routing elements, and more.

6. Capabilities of Bad Actors

6.1 General capabilities

In general, Bad Actors described above should be expected to have access to the following:

1. An extensive corpus of messages from targeted domains

2. Knowledge of the practices used by targeted domains

3. Access to public keys and associated authorization records published by the targeted domains

and the ability to do at least some of the following:

1. Generate substantial numbers of unsigned messages that might represent either an intentional or unintentional denial of service attack

2. Transmit messages using any envelope information desired

3. Transmit messages using any message headers desired

4. Submit messages to MTAs from multiple locations within the Internet

5. Sign messages on behalf of domains from registrars that protect the domain owner's privacy or that have poor vetting practices

6. Generate substantial numbers of messages with invalid signatures which may be an attempt to create a denial of service attack by overwhelming DKIM verification

7. Resend messages which may have been previously signed by other domains

6.2 Advanced capabilities

Certain classes of Bad Actors may have substantial financial motivation, and therefore could be expected to have more capabilities at their disposal. These include:

1. Access to significant computing resources, perhaps through the conscription of many compromised systems. This could allow Bad Actors to perform various types of brute-force attacks.

2. Manipulation of IP routing. This could be used to submit messages from specific IP addresses or difficult-to-trace addresses, or to cause diversion of messages to a specific domain.

3. Influence over portions of DNS using mechanisms, such as cache poisoning. This might be used to influence message routing, or to falsify DNS-based key or policy advertisements.

4. Ability to eavesdrop on some existing traffic, perhaps from a wireless network, a cable or DSL modem, or a compromised server or router.

The last three of these mechanisms could permit Bad Actors to redirect traffic and masquerade as a desired destination. Such redirection provides a means to deceive the recipient or those attempting to send messages, and may be used to obtain access-related information among other sensitive data.

7. Bad Actor's Points of Access

In the following discussion, the term "Administrative Unit", taken

from [I-D.crocker-email-arch], is used to refer to a portion of the email path under common administration. Recipients usually establish mutual authentications with Administrative Units receiving and verifying their email using shared secrets and server certificates. Administrative Units that perform message signing also usually establish mutual authentication using shared secrets and server certificates.

7.1 Bad Actors with Access in the Administrative Unit

Bad Actors can obtain access anywhere in the Internet. Bad Actors that have privileged access within the Administrative Unit of the signing-domain or the recipient domain have capabilities beyond those elsewhere, as described in following sections.

7.1.1 Access in the Signing-Domain's Administrative Unit

Bad Actors may gain access using compromised accounts or systems within the Administrative Unit corresponding to the signing-domain. Although submission of messages generally occurs prior to the application of a message signature, DKIM could still be effective at isolating compromised accounts, and should be effective at isolating compromised message signing systems when each system utilizes specific keys. Defense in such cases would be improved by monitoring for account abuse and system integrity, in addition to limiting access to local system services.

DKIM can be effective at improving email security within the Administrative Unit, especially in the case where an Administrative Unit has systems coupled by the Internet. DKIM signatures can validate legitimate externally-originated messages considered within the Administrative Unit.

7.1.2 Access in the Recipient's Administrative Unit

Bad Actors may gain access using compromised systems within the Administrative Unit of the message recipient. Since messages typically undergo DKIM verification one time within a possible successions of systems carrying messages to their destination within the Administrative Unit, DKIM may not detect invalid messages from compromised systems that are subsequent to the DKIM verification. Bad Actors may also masquerade as a system within the succession as another means of introducing invalid messages. Defense in such cases would be improved by verifying DKIM at the last system in the succession, using authentication mechanisms like SMTP AUTH, by monitoring system integrity, in addition to limiting access to local system services.

7.2 Bad Actors with External Access

DKIM focuses primarily on Bad Actors that do not have privileged access within the Administrative Units of the signer or the recipient. Outside these Administrative Units, the trust relationships required for authenticated message submission do not exist and do not scale adequately to be practical.

Bad Actors with only external access will usually attempt to exploit the open nature of email. Most Administrative Units will accept messages with a valid email address for a local domain, often after investigating the reputation of the source IP address. Bad Actors may generate messages without signatures and rely upon techniques that obfuscate their source IP address.

When signing-domains accrue reputations serving as a basis for acceptance, Bad Actors may take advantage of registrars that protect the privacy of domain owners, or that do not verify the domain owner's identity in a reliable manner. Bad Actors may then use these domains without an initial negative reputation to generate messages with valid signatures. Bad Actors may send messages containing a diversity of email addresses to avoid filtering techniques. Often this ploy by Bad Actors is attempted while posing as mailing lists, greeting cards, or other agents which legitimately send or re-send messages on behalf of others.

8. Representative Bad Acts

One of the most common bad acts attempted is the delivery of messages which obfuscate their true source within the network or associated domain. The purpose of obfuscation might be to gain acceptance or to defraud the recipient. The severity of this problem ranges from messages merely being unwanted, to defrauding the recipient or compromising their system with a payload of malware.

8.1 Use of Arbitrary Identities

Arbitrary Identifiers typify those bad acts aimed at obfuscation to gain acceptance of messages. Such methods use a wide range of techniques as previously described.

DKIM may be effective for abating the misuse of a domain which asserts all messages are signed by the domain. The effectiveness of DKIM at preventing domain misuse would depend upon the prevalence of recipients validating DKIM signatures and obtaining domain-wide assertions. For cases where a domain is not being misused, or when signed by a domain controlled by Bad Actors, the recipient would then be reliant upon reputation or accreditation services for protection.

8.2 Use of Specific Identities

A second major class of bad acts involves the assertion of specific identities in email.

8.2.1 Exploitation of Social Relationships

One reason for falsifying an associated domain is to encourage a recipient to act on a particular email message that appears to be from an acquaintance or previous correspondent trusted by the recipient. This tactic has been used by email-propagated malware which emails to addresses in the compromised system's address book. In this case, the sender's email address and related signatures may not have been falsified, so DKIM would not be directly effective in preventing this act, but could facilitate the isolation of the compromised system when an opaque-identifier has been included, see [I-D.otis-mass-reputation]. Using opaque-identifiers would allow rapid correlations of malware sources by third-parties monitoring for this type of threat.

It is also possible for address books to be harvested and used by an attacker to send messages from elsewhere. DKIM may be effective at mitigating these acts when the recipient verifies the DKIM signature and when the sending domains assert all messages are signed by the domain. It is also possible for the recipient to retain a binding of the opaque-identifier with the signing-domain associated with the sender's email-address, see [I-D.otis-mass-reputation].

8.2.2 Identity-Related Fraud

Bad acts related to email-based fraud often involve the transmission of messages misusing visible associated domains as part of the fraud scheme. The misuse of a specific associated domain sometimes contributes to the success of the fraud by convincing the recipient that the message is valid.

To the extent the success of the fraud is enhanced by the misuse of a specific visible associated domain, Bad Actors may have significant motivation and resources to circumvent measures that target specific headers for unauthorized use. There could be exigent cases where a domain-wide assertion becomes beneficial if it prohibits the appearance of the domain in any originating header unless also signed by the domain. This would be accomplished with a domain-wide assertion made by the signing-domain, similar to the assertion that all messages are signed by the domain, see [I-D.otis-mass-reputation].

8.2.3 Reputation Attacks

Another motivation for using a specific associated domain in a message is to harm the reputation of the domain. For example, a commercial entity might wish to harm the reputation of a competitor, perhaps by sending a copy of their competitor's signed promotion as unsolicited bulk email. A reputation service would categorize abuse primarily by the recipient, and not the message content. A reputation service would not be able to differentiate between valid and invalid uses of a signed message.

While reputation services must accrue behaviors based upon verified identifiers, there must also be a means to mitigate ongoing abuse. Without a means to abate ongoing abuse, the reputation service, which must be responsive to the needs of their subscribers, would have little choice but to list the domain being attacked and expect them to undergo a rehabilitation process. Rehabilitation may involve a demonstration of having a means to respond to abuse. See the following section on Message Replay Attacks.

9. Attacks on Message Signing

Bad Actors can be expected to exploit all of the limitations of message authentication systems. They are also likely to be motivated to degrade the usefulness of message authentication systems in order to hinder their deployment, when the systems prove effective. Some categories of bad acts are described below. Additional postulated attacks are described in the Security Considerations section of [I-D.allman-dkim-base].

9.1 DoS Attack

Checking either the authorizations associated with a message signature or the verification of the signature will not afford any Denial of Service protections. There are only two choices available where DoS protections are possible. The DoS protections could be based upon the remote IP address or upon the verification of the EHLO name. In the case of the EHLO name, the problem associated collateral blocking is overcome, and a subsequent reputation check on the signing-domain may be skipped when the EHLO name is within the signing-domain.

If the strength of the EHLO is not considered adequate, it may be possible to added a signature, the last digit of the year, and the week number prefixed to the EHLO name delineated with an '_'. The hash of the EHLO would could include a string that represents all but the lower three bits of the remote IP address. After all, the EHLO is currently permitted to fail verification.

9.2 Invalid Signatures

Messages with invalid signatures would be handled as messages without signatures. Such messages would be handled in the normal fashion where the reputation of the remote IP address would be assessed. Rejections based upon the remote IP address often creates a problem when the address is being shared. When a Bad Actor sharing the IP address sends abusive email, other entities may be collaterally blocked when a negative reputation is then applied. Such blocking may encourage those that find themselves blocked to adopt message signing as an alternative basis for reputation assessment.

9.2.1 Canonicalization and Message Normalization

Until signatures are known to be reliably valid throughout the email infrastructure, invalid signatures should be treated in the same manner as a message without a signature. As with any message, these messages may be introduced by Bad Actors. The intent may be to have the message appear as though it was legitimately sent, but "broken" in transit. This should not affect how the message is handled, and the reputation of the remote IP address should still be assessed.

To minimize the number of broken messages and thus improve the reliability of the message signature, message normalization is required to ensure line-lengths are compliant prior to signing. The current simple canonization is rather fragile, whereas the relaxed canonicalization allows for several exploits. On the DKIM list, Earl Hood has recommended what appears to be a suitable replacement for the no-white-space method. This technique removes white-spaces preceding end-of-lines and streaming white-space at the beginning and end of the message body.

9.3 Body Length Parameter

The Body Length parameter available as an option within DKIM must be used with great caution. Recipients that find the message length has increased, should discard the portion of the message that prevents signature validation when included as signed content. The concept behind this option was to accommodate the appending of information by providers or some list servers. As this option can be easily exploited, especially through a list server, mandating the deletion of added information would be the only solution that prevents this option from inviting an abusive message replay problem.

9.4 Multiple Signatures

The DKIM draft offers little guidance with respect to multiple signatures. Multiple signatures, rather than offering greater

information, may obfuscate the role played by the signer. There have been suggestions that signatures be sequenced by way of a count added, or by their header order within the message. Unfortunately these techniques do not clarify which signature is significant with respect to the initial signer. Any miscreant could either rearrange signature headers, add sequence numbers that appear as if signatures have been dropped, or add several "throw-away" signatures where reputation accrual may be diffused. Multiple signatures also provide plausible deny ability when a reputation service attempts to locate the accountable domain. An association with an email-address may not clarify the role of the signer, as the signing may be done by domains that are independent of any message headers. There should also be a limit with respect to the number of signatures allowed within a message, as each signature represents added message overhead.

It may be advisable to have signature headers that clarify the role of the signer. The current signature header could be considered the "Originating" signer. A list server may add their signature as a "Remailing" signer. The receiving domain within the recipient's Administrative Unit may wish to add their "Verifying" signature as verification that various checks have been completed. Any previous signatures of the same role would be deleted and perhaps logged within the Received headers. The "Verifying" signature should not be considered valid beyond the Administrative Unit.

For additional protection against message replay attacks, a practice could be established that restricts these three roles to a single signature per message. A Remailing signature may encompass the Originating signature. A Verifying signature may encompass both the Originating signature and the Remailing signature. When adding a signature that encompass previous signatures, the signature information associated the 'b=' tag with could be overwritten with "remailing-checked", "verifying-checked" or "invalid." The typical recipient could rely upon the Verifying signature provided by the Administrative Unit at the MUA, but would not have access to signatures that could be used to stage a message replay attack.

In a situation where message replay attacks become problematic for messages sent to a specific domain, the signing domain may wish to preclude sending signed messages to these domains. The receiving domain may wish to prevent the possibility that any of their users could be capable of such an attack by obfuscating Originating and Remailing signatures and adding the Verifying signature.

9.5 Use of "throw-away" domains

Bad Actors may also introduce messages with valid signatures from domains they control, perhaps "throw-away" domains registered under

false pretenses or using registrars that protect the privacy of the domain owner. In other words, the existence of a message signature does not imply the conduct of a signing-domains is trustworthy. The already common use of such domains require domain-based accreditation or reputation systems. A reputation service may not be able to differentiate between a new and a throw-away domain. A Bad Actor could also acquire a domain previously used for legitimate purposes. Reputation services may extend the weighing of behaviors to include that of the registrar. Current name based reputation systems are known as Right-Hand-Side reputation services. Even without the use of a name-based reputation service, local reputation, mostly in the form of white-lists, can be maintained by domains to improve the deliverability of email from domains where existing relationships have been established.

Accreditation and reputation, or even local white-lists, require a verifiable identity upon which to base the accrual of behaviors and possible feedback reports. Message signing provides an identity which is intended to be sufficiently reliable for this purpose. A verifiable identity is necessary for accreditation and reputation systems to operate, provided there is a means established to either prevent or abate message replay attacks.

Providers operating shared email services, mailing lists, and other legitimate agents may commonly sign messages with headers containing differing domains. This common practice offers valuable freedoms for typical users. This practice may allow a Bad Actor to sign messages containing divergent domains and to also appear legitimate. Nevertheless, the assessments of the signing-domain should remain the primary factor when deciding whether to accept the messages. When the signing-domain is provided by a registrar that ensures domain owner privacy and has not obtained a recent reputation, little confidence in the signing-domain can be obtained. As with message replay abatement, such mysterious signing-domains may be given a Transient Negative Completion reply. Over time, the signing-domain will become known, or the administrators of the signing-domain may contact the relevant reputation and accreditation services.

9.6 Message Replay Attack

Attacks based upon abusive retransmission of an otherwise valid message is referred to in this document as a "message replay attack". DKIM is able to provide an option that offers a means to promptly abate this type of replay, while also identifying all culpable sources, see [I-D.otis-mass-reputation]. This could be accomplished using an opaque-identifier revocation mechanism that is checked when the message appears to be coming from a different domain, or when the recipient has changed. Abusive replay abatement is essential for the

protection of the signing-domain's reputation.

Message replay must be allowed to occur, even at high levels. Legitimate replays may result when a message is distributed through a list server, for example. As valid message replay is performed at systems unrelated to the signing-domain, this replay process must be monitored for abuse, and strategies will be needed to deter efforts attempting to elude an abatement process. There are methods that can be used to mitigate the precautions needed to deal with a potential replay problem. The EHLO verification, essential for name based DoS protections, could also mitigate replay abatement. A signed hash of a salted [RFC2821] RCPT TO header could be another replay abatement mitigation strategy.

Part of this mitigation strategy may involve delaying the complete processing of messages identified as having potential replay risk. This delay is needed to allow abuse-monitoring the requisite time to react, which could be done within minutes. This processing delay may occur at the transmitter, or the receiver, or at a combination of both. Transient Negative Completion replies indicating the request was aborted with an error in processing should cause the message to be held at the transmitter, see [RFC2821]. The receiver may opt to queue a limited number of messages identified as having a replay risk, and once that limit has been reached, a Transient Negative Completion reply is issued. This mechanism could be further mitigated by white-lists of trusted message resenders, such as list servers.

When a message appears to be coming from a different domain and has an invalid hash of the [RFC2821] RCPT TO, as yet another option, messages may be held within a queue for a brief period to allow time for an opaque-identifier revocation to occur, see [I-D.otis-mass-reputation]. While opaque-identifiers may normally reflect the account used to gain access, serializing the opaque-identifier per a recipient of bulk email may also isolate the culpable recipient.

Revocation of keys would not be a practical solution, as this will likely impact the delivery of many unrelated messages and will not likely be as prompt. However, revocation of the opaque-identifier could also act as acknowledgement to a reputation or accreditation service that the signing-domain has responded to the reported abuse. Even the listing of revocations could become a service by delegating the revocation zone. Lengthy delays in responding would provide little protection against such acts and likely precipitate a negative reputation as well as increased abuse.

Bad Actors may obtain a reply from an individual within a signing-domain that carries a copy of their desired content. The reply may

then be used to distribute the desired content in bulk where no account has been obtained from the signing domain. The person replying may be unaware of the risk. When Bad Actors obtain an account from a provider that offers services to the public, and they send a small number of messages with desired content to addresses controlled by Bad Actors, the activity of the account will appear normal, but messages obtained can be used for abusive replay.

Some other suggestions to abate message replay abuse burden the recipient. These suggestions are usually based upon content filtering of messages that have been signed. Once an unwanted message is discovered through filtering, signature finger-prints may then be used to identify any replicate message. There is no assurance Bad Actors will not limit the number of replicate messages sent to a specific domain. In such a case, filters would not be greatly advantaged by the signature. There is also a suggestion that the signature finger-print itself accrues a reputation. It is then expected the recipient would obtain the services of a signature finger-print reputation service. The amount of centralized data exchanged to support a signature finger-print reputation scheme would represent a significant increased burden for the recipient that would be difficult to scale for either the service or the recipient.

Other suggestions attempt to burden the sender. Some suggestions have the signing-domain employ outbound content filtering. Although outbound filtering could be a reasonable practice, the effectiveness of filtering is never 100%. Bad Actors attempting to accumulate signed messages sent to themselves would be advantaged by outbound filters. Messages that manage to evade outbound filters are also likely to evade inbound filters employed in other domains. Another strategy suggests to enforce greater accountability for accounts whose messages are to be signed by DKIM. Even with exorbitant fines exacted and with extensive vetting processes, there remains the problem created by the millions of compromised systems. The stricter policies would increase the harm created by the compromised systems.

Perhaps the greatest suggested burden placed upon the sender is to register the paths for all their valid email. An onerous enough task is made more difficult with the path registration being predicated upon an email-address domain within a specific header selected in a non-compliant fashion, and not the signature itself. See the section on Sender Signing Policy below. This mingling of mail-address domain paths with a signing-domain path is actually attempting to solve two different problems simultaneously. Neither message replay abuse, nor the misuse of an email-address is effectively prevented. For many years from now, there will be recipients that use forwarded accounts, or users wishing to send messages to simple exploding list servers. These uses, as well as hundreds of others, will create a multitude of

exceptions that must be accommodated. Each method of accommodation represents a means for exploitation.

10. Threats to Delivery

DKIM relies upon more of the network infrastructure. Normal email exchanges depend upon the recipient's DNS to locate MTAs that accept delivery for the recipient. DKIM adds reliance upon the signing-domain's DNS for distributing public-keys. With DKIM's Sender Signing Policy (SSP) [I-D.allman-dkim-ssp], as currently defined, delivery also depends upon allowances made for "third-party" signers when the [RFC2822] Sender, Resent-From, or Resent-Sender headers are used by the signing-domain. Such increased dependency, especially with respect to policy assertions, represents additional avenues for attack and will negatively impact many commonly used email services.

While DNSSEC [RFC4033][RFC4034][RFC4035] offers protection from various attacks on DNS, its greater overhead may increase DKIM DoS vulnerabilities. There could be increased susceptibility at the recipient's DNS resolver, especially when wildcard public-keys or policies have been published. Synthetic labels may allow an attack with increased requisite interaction and caching prior to verifying the signature.

An SSP policy which excludes third-party signing, and is the only mode that can repudiate Bad Actors, may cause messages to become lost when remailed, or mailed. For example, such a policy could prevent messages containing an email-address in the [RFC2822] From header from surviving mailing lists that sign messages, or when sent as signed news articles or invitations. Such a consequence will discourage the vital adoption of the only policy that affords protection from Bad Actors.

11. Urgent Response to Threats for Consumer Protections

11.1 Restricted Two-Party Communication

While "phishing" attacks may be creating an exigent situation requiring an urgent response, parsing a complex policy record for qualifiers and lists of authorized "third-party" signers is not something that can be incorporated quickly. A simpler record is required to allow immediate incorporation into MTAs as a means to offer the most expedient response to this type of threat. This lookup should be offered as a service by the industry to combat this specific threat.

A Two-Party listing service should simply return a record to indicate the queried domain prohibits the use of this domain within any email

parameter or header unless also signed by this domain. This would limit the use of the listed domain to two-party exchanges. Exceptions for messages containing the prohibited domain would be made when the only recipient is the prohibited domain.

11.2 Unrestricted Multiple-Party Communication

When multiple-party email exchanges are to be protected, an assertion that all emails are exclusively signed by the domain should be based upon the header specified by [RFC2822] as having introduced the message into the email transport system. Compliance with [RFC2822] email transport introduction allows common services to be retained without specific white-listing. No information has been lost with respect to the signing domain and the email-address contained within the From header. The receiving domain may assess such messages according to the relationships indicated. However, acceptance should be primarily based upon the reputation of the signing-domain. Where it is absolutely critical that the From header for a specific domain be protected, the Two-Party listing service should be used. By removing any disruption in services resulting from a domain signing all their messages and making assertions on that basis would allow immediate repudiation of Bad Actors. The current SSP policy necessitates a disruption in some services in order to repudiate messages from Bad Actors.

Authorizing third-party signing defeats protection from Bad Actors. SSP's use of the From header to designate which domain establishes signing policy is disruptive and limits the scope of messages that can be afforded protection. Messages being remailed where they are signed and reintroduced into the email transport system following [RFC2822] conventions, may be designated by SSP policies as being signed by "third-party" domains. Rather an adhering to [RFC2822] conventions for establishing the header having introduced the message into the transport, the often visible From header was used as a basis instead, while risking the loss of a substantial number of otherwise valid messages and limiting the adoption of a protective policy.

11.3 Opportunistic Protection without Domain-wide Policy Assertions

SSP designating the From, or when multiple-addresses exist, the Sender header for establishing signing policy is primarily to abate attempts at falsifying the author's email-address when third-party signing is not permitted. Such falsification is often the case in "phishing" attacks. The success of such an effort remains doubtful as many MUAs display just the "pretty-name." Policy based upon the visible header also does not deal with threats associated with domains that have a similar appearance and with MUAs susceptible to character-set attacks.

Attempts to base policy on the [RFC2822] From header significantly decreases protections afforded by DKIM by imposing dire consequences when indicating emails are signed exclusively by the domain. While white-lists of authorized third-party signers may mitigate some unintended message loss, such an authorization strategy burdens the recipient, is expensive to maintain, and is not practical.

DKIM can offer significant protection for multiple-party email commutations without the separate publication of signing policy. There is an alternative method that can be used to extend protection to recipient. This strategy would take advantage of the situation that protection is often desired for prior correspondents. With the messages being signed, the message itself offers a secure means to communicate what elements within the message can be relied upon to uniquely identify the message's author.

For domains that assure those granted access are limited to specific email-addresses, any email-address from this domain can then be used to uniquely identify the author. Domains that do not limit the email-address can alternatively offer an opaque-identifier within the signature to uniquely identify the account used to gain access. The recipient could request that bindings of the requisite identifiers be saved. With these bindings saved, the recipient can then be alerted when these identifiers have changed in subsequent messages. In some cases, it would be possible to automatically retain the bindings at the MTA rather than just the MUA.

By retaining binding for those specific email-addresses considered important to the recipient, greater analysis can occur to defeat "pretty-name", character-set, and domain look-alike attacks. When an identifier appeared to have changed, the recipient should be shown the email-address and other identifiers using a consistent character-set and asked if the information should replace or be merged with current bindings, or perhaps be ignored, or flagged as a spoofing attempt.

12. Sender Signing Policy

DKIM Sender Signing Policy (SSP) [I-D.allman-dkim-ssp] attempts to introduce several constraints on an email-addresses found in one of two headers in conjunction with DKIM signatures. These constraints may indicate that a signature is required either of some third-party domain, or of a first-party domain, or no signature is required at all. The SSP also introduces a constraint placed upon the [RFC2822] From header that may be conditionally relegated to the Sender header when there are multiple email-addresses within the From header. Conditionally relegating the From to the Sender header may confuse recipients for two reasons. The Sender header may be indicated by

some MUAs as being the significant email addresses. It also may not be obvious to the recipient when there are multiple email-addresses within the From header.

It is also not clear how a third-party signature should be handled in the case where there are multiple signatures added to the message. Should a message that has a first-party signature and a policy that third-party signatures are prohibited, then cause a message to be rejected when a signature is added by a third-party? What happens when the first-party signature has expired, but the third-party signature is still valid?

Unless the domain-wide assertion that all emails are signed follows normal email conventions with respect to which header introduced the message into the email transport system, there will be messages that are not be protected by the SSP From/Sender assertion. This failure to provide full protection in such cases was created by a desire to ensure the significant header related to policy is always visible. Nevertheless, to establish protections without disrupting email services, it would be beneficial to have an Introduction assertion. With an Introduction assertion, all messages that would normally be considered to have been introduced by a domain using [RFC2822] definitions would be assured protection. This approach would suffer from far fewer policy conflicts by other domains and provide much greater delivery integrity.

The SSP strategy also fails to consider there are other methods that may be used to qualify email and assumes the From/Sender headers are always significant. An MUA that uses SPF Classic may indicate a message has been qualified based upon the [RFC2821] MAILFROM, but this parameter is not checked against the DKIM signing-domain, and yet the recipient may see a message as verified as a result of the MAILFROM.

Once normal email conventions are allowed, a new assertion is required to protect those domains that have become "phishing" targets. This new assertion would prohibit other domains the use parameters and headers that could contain an email-address considered to have introduced the message. It would cover the MAILFROM, From, Sender, Reply-To, and corresponding Resent-* headers. Such a domain-wide assertion would curtail exploits still possible with an SSP policy that erroneously assumes what the recipient considers significant.

Using non-compliant conventions with respect to the significant header also disrupts normal email practices. A domain making a domain-wide assertion that all messages are signed may be unable receive protection for some of their messages, and may find that some

of their messages have been subsequently prohibited by other domains. Even appending a Sender or Resent-* header will not offer a solution, as the From header may retrain significance.

Attempting to bind specific headers to a signing-domain has an unfortunate consequence that some mechanisms may unfairly extend accountability to the email-address. Any authorization of third-party signers with respect to a specific header's email-address is highly problematic. This authorization is likely to be assumed as having authenticated the email-address causing unfair reputation accrual. The SSP mechanism may also require an extensive number of lookups. This mechanism will require lookups walking up the label tree, to qualify the local-part of an email address, and, with a pending proposal, to also authorize specific third-party signers.

13. IANA Considerations

This document defines no items requiring IANA assignment.

14. Security Considerations

This document describes the security threat environment in which DomainKeys Identified Mail (DKIM) is expected to provide some benefit.

15. References

15.1 Normative References

[RFC2821] Klensin, J., "Simple Mail Transfer Protocol", RFC 2821, April 2001.

[RFC2822] Resnick, P., "Internet Message Format", RFC 2822, April 2001.

[RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.

[RFC4034] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Resource Records for the DNS Security Extensions", RFC 4034, March 2005.

[RFC4035] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions", RFC 4035, March 2005.

15.2 Informative References

[I-D.allman-dkim-base]
Allman, E., "DomainKeys Identified Mail (DKIM)", draft-allman-dkim-base-01 (work in progress), October 2005.

[I-D.allman-dkim-ssp]
Allman, E., "DKIM Sender Signing Policy", draft-allman-dkim-ssp-01 (work in progress), October 2005.

[I-D.crocker-csv-csa]
Crocker, D., "Client SMTP Authorization (CSA)", draft-crocker-csv-csa-00 (work in progress), October 2005.

[I-D.crocker-csv-intro]
Crocker, D., Otis, D., and J. Leslie, "Certified Server Validation (CSV)", October 2005.

[I-D.crocker-email-arch]
Crocker, D., "Internet Mail Architecture", draft-crocker-email-arch-04 (work in progress), March 2005.

[I-D.otis-mass-reputation]
Otis, D., "MASS impacts upon reputation", draft-otis-mass-reputation-03 (work in progress), September 2005.

Author's Address

Douglas Otis
Trend Micro, NSSG
1737 North First Street, Suite 680
San Jose, CA 95112
USA

Phone: +1.408.453.6277
Email: doug_otis@trendmicro.com

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.