December 2023

EMAIL SCANNING OPTIMIZATION TECHNIQUES

Suresh Gopathy

Gajendar Pandey

EMAIL SCANNING OPTIMIZATION TECHNIQUES

AUTHORS:
Suresh Gopathy
Gajendar Pandey

ABSTRACT

Typically, a Secure Mailbox (SM) or Secure Email Gateway (SEG) that receives emails for an organization scans emails and performs threat detection/handling for the emails (e.g., allowing or dropping emails) based on various handling policies configured for the organization. Currently, an SM/SEG completely scans all email each time an email traverses the SM/SEG (e.g., for an initial email and any subsequent replies/responses for a conversation/email thread), which can result in high usage of compute resources for an organization, potentially increasing the cost of email services, as well as increasing the latency of email delivery. In order to address such issues, techniques presented herein provide for the ability to prevent repeat scanning of email content that has already been scanned for a given conversation thread, which can be identified using a message identifier (M-ID), and to correlate results of previous scans of the conversation thread with a current scan of the thread (having a same message ID) in order to perform threat detection for emails.

DETAILED DESCRIPTION

Enterprise organizations typically utilize a Secure Mailbox (SM) or Secure Email Gateway (SEG) in order to scan emails and performs threat detection/handling for the emails (e.g., allowing or dropping emails) based on various handling policies configured for the organization. In current implementations, an SM/SEG typically scans all email that traverses the SM/SEG (e.g., for an initial email and any subsequent replies/responses for a conversation/email thread), which can result in high usage of compute resources for an organization, potentially increasing the cost of email services, as well as increasing the latency of email delivery.

Consider an illustrative example through which multiple scans may be performed for each of multiple emails exchanged for a conversation thread. For example, consider

that an SM/SEG receives an initial/original message from a given sender and scans content of the email via a threat detection scanning process. Next, consider that a given recipient of the email sends an initial reply email to the original message sender (Reply-1). When the SM/SEG receives the Reply-1 email, the reply email also has the original message embedded in it. However, the SM/SEG interprets the Reply-1 email as a new email and, thus, performs content scanning on the Reply-1 content and also performs a repetitive scan of the original email content. Next, consider that the sender then sends another reply email (Reply-2) such that when the SM/SEG receives the second reply email, it will perform scanning on the Reply-2 content and will repeat the Reply-1 content scanning and the original content scanning again.

In order to address such issues, techniques presented herein provide for the ability to prevent repeat scanning of email content that has already been scanned for a given conversation thread, which can be identified using a conversation or message identifier (M-ID), and to correlate results of previous scans of the conversation thread with a current scan of the thread (having a same message ID) in order to perform threat detection for emails.

Figure 1, below, illustrates example details for an operational flow involving an SM/SEG that may perform techniques in accordance with this proposal in order to prevent repetitive scanning of email content that has previously been scanned through the use of a conversation ID that can be used to track email exchanges for a given conversation thread.

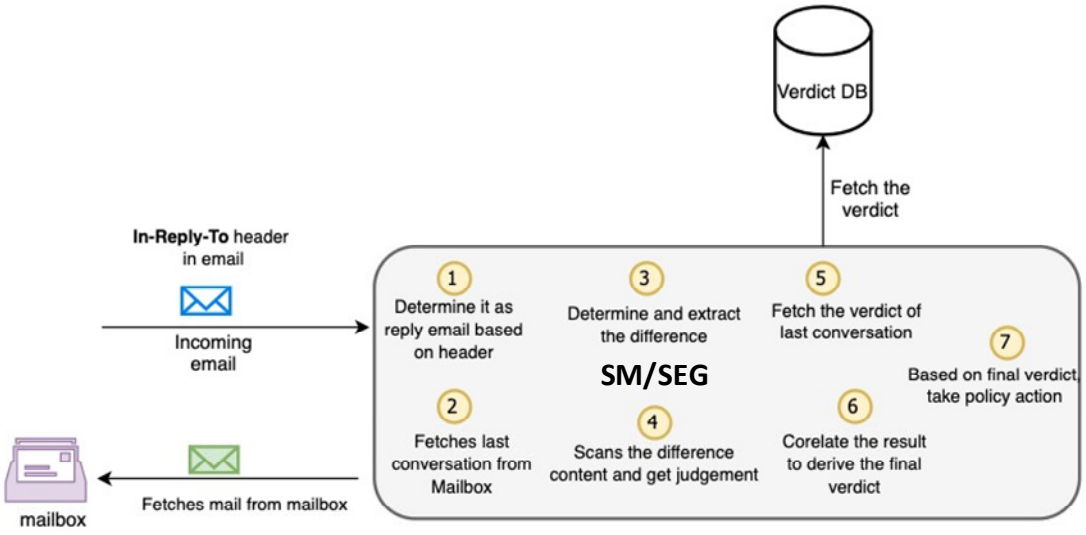

Figure 1: Example Operational Flow to Prevent Repetitive Email Scanning

As illustrated in Figure 1, consider at (1) that the SM/SEG, on receipt of an email determines whether the email is a new email or a reply to an earlier email. In at least one instance, the SM/SEG can determine whether the email is a new email or is a reply email through analysis of the message to determine whether an 'In-Reply-To' header is contained in the message such that, if the header is present, it can include a message ID (M-ID) through which to identify the conversation/earlier messages for the thread, as generally prescribed by Internet Engineering Task Force (IETF) Request for Comment (RFC) 2822.

As shown at (2), upon determining that the email is a reply email, the SM/SEG can fetch the original conversation from storage using the M-ID contained in the 'In-Reply-To' header. For example, using the M-ID, the SM/SEG can fetch the last message of the conversation thread from a user's mailbox (sender or receiver) using a mailbox Application Programming Interface (API) or other function call.

Upon obtaining the last message in the thread, as shown at (3), the SM/SEG can determine and extract/identify any differences between the original email and previous email(s) for the conversation thread that are fetched from storage. For example, the SM/SEG can extract the content of the current email as well as previous email(s) for the thread and can compare the content in order to identify any differences between the content of the current email and previous email(s). Any new content that may be identified for the current email can be scanned using the threat detection scanning process (e.g., performed using scanning servers/blades), as shown at (4), and determine any appropriate handling/judgement for the email content based on the scanning.

As shown at (5), the SM/SEG can also fetch previous scan results from storage (e.g., a 'verdict' database or the like) and can compare/correlate a result (e.g., a verdict) from the current scan and any previous verdicts performed for the conversation thread in order to derive a final (current) verdict for the received email, as generally shown at (6). Thereafter, based on the correlated/overall verdict, the SM/SEG can take an appropriate action based on policy for the received email (e.g., drop, delete, quarantine, remediate, deliver, etc.), as generally illustrated at (7).

Accordingly, techniques of this proposal provide a process that can be utilized by an SM/SEG in order to prevent repeat scanning of email content that has already been scanned for a given conversation thread and to also facilitate the correlation of results of

previous scans of the conversation thread with a current scan of the thread in order to perform threat detection for emails.