Advanced Security: Email

Table of Contents

Advanced Security: Email testing methodology v2.3
(16th May 2025)

1. Introduction

This methodology provides a way to test email filtering services for prolonged periods using a variety of realistic approaches and to supply results on an on-going basis.

A network of dedicated email honeypots and other sources enables us to utilise the latest threat campaigns in these tests, while in-depth knowledge of targeted attack methods allows us to emulate more direct attacks. Legitimate messages of varying types are also included to test for false positive rates.

Testing is conducted using regular endpoint clients configured to use popular email services such as Microsoft Office 365 that are, in turn, configured to use recommended settings or to use other, third-party email security services.2. Test framework

2. Test framework

The test framework collects threats and verifies that they will work against unprotected targets. It then attempts to expose protected targets to the verified threats to determine the effectiveness of the protection services.

2.1 Infrastructure

The Threat Management System is a database of attacks including live malicious URLs; malware attached to email messages; links to malware included in email messages; spear-phishing email messages; and a range of other attacks generated in the lab using a variety of tools and techniques. All attacks used are either real attacks found in the wild or otherwise highly realistic attack scenarios. Live malware threats are fed to the Threat Verification Network (TVN).

When a threat arrives at the Threat Verification Network it is sent to a Vulnerable Target System (VTS) in a realistic way to be checked that malicious behaviour is evident and active. Malicious behaviour might include executing malware, or displaying misleading content.

For example, a tester operating a VTS receives a malicious email in an email client. If that email is a phishing attack, the tester behaves as a naïve user and performs likely activities, such as clicking on links and entering information; or downloading, opening and interacting with attached files. Links to malicious websites are followed as far as possible.

This part of the process validates the threats, ensuring that they operate as the attacker intends without interference from email security services.

A Target System (TS) is identical to a Vulnerable Target Systems (VTS) used in the Threat Verification Network, except that a TS is protected by email security services.

2.1.1 Target system details

Each TS is a Windows system, deployed either as a physical or virtual PC. Each system has unrestricted internet access.

The email client used is configured to access the test's email samples via the email security service undergoing test, according to instructions provided by each email security service supplier. Configuration changes, including adding or removing policies, is permitted under advisement during the pre-test setup period.

The email client used reflects that most commonly used in the real world. For example, Microsoft Outlook; or Microsoft Outlook Web Application (OWA) via a popular browser. Consideration is given to issues around data-sharing relationships between the developers of the email clients (and browsers) and the developers of the email security services.

2.1.2 Baselining

Each email security service is permitted 10 days exposure to clean network traffic, when requested by the vendor in question.

2.1.3 Sending stages

Threats and legitimate messages are sent by email to each TS in as realistic a method as possible, at the same time. Threats are sent from different IP address ranges and email addresses to those used to send the legitimate messages.

The inboxes, logs and any other relevant elements of each tested service is monitored at the time of testing and 24 hours later to check for delayed remediation. Results are recorded as per 3. Results below.

2.2 Scope

Test cases comprise realistic attacks designed to target victims via email. Test subjects can be email platforms, optional security services, third-party security add-on services or a combination of any of the above.

2.2.1 Sample selection

2.2.1.1 Threat selection

The following categories of threats are included:

a) Social engineering (targeted)
b) Phishing (targeted)
c) Malware (targeted)
d) Business Email Compromise (targeted)

Each category comprises different scenarios representing popular attack approaches. Targeted attacks are built in the test lab and grouped into scenarios. Examples include, but are not limited to:

i.) Fake law enforcement blackmail
ii.) Emergency payment request
iii.) Fake lottery win
iv.) Fake login page to popular website

Each scenario (e.g. fake lottery win) further consists of a number of different versions of the attack. The differences between samples in each scenario may include, but are not limited to, email spoofing; attached content; linked-to content; password protection.

Targeted threats are generated in the lab according to threat intelligence gathered from a variety of sources. These threats can be considered as similar to publicly-known targeted attacks that are in common use at the time of the test run.

All threats are identified, collected and analysed independently of security vendors directly or indirectly involved in the test. Samples containing malicious code, or links to malicious code, are confirmed by the TVN as being malicious.

A replica target Company is used to test Business Email Compromise (BEC) attacks. The Company structure realistically represents a small business, with a number of identifiable employees, clients and suppliers. Separate internet domains are used for each of these organisations, with some websites being deployed where necessary. Domain ages vary but are typically older than one year.

2.2.1.2 Legitimate message selection

Legitimate samples contain popular and non-malicious website URLs and text-based messages with no harmful content. These comprise real legitimate email messages and lab-generated messages that are clearly legitimate and will not easily be confused with malicious messages by average users.

These messages are used to check for false positive detection. Candidates for legitimate sample testing include realistic email messages that may be sent directly to the target or forwarded. The email message body content is clearly legitimate and not closely resemble harmful messages.

As closely as possible, the legitimate messages contain similar but legitimate content to the malicious emails. For example, if malicious Microsoft Office macros are used in the threat set, legitimate Microsoft Office macros will be used in the legitimate set.

2.3 Configuration

Services are configured according to each vendor's recommendations, by the testing team, the vendor or both where appropriate.

3. Results

3.1 Observations

The following occurrences during the attack stage are recorded:

a) The point of detection (e.g. on arrival at the service; blocking a URL after a period of time).
b) Detection categorisation, where possible (e.g. IP address reputation, file signature, spoofing).
c) Details of the threat, as reported by the product (e.g. threat name; attack type).
d) Action on threat (e.g. deletion, quarantine, delivered with warning, delivered without warning).
e) Legitimate files allowed to pass without problems.
f) Legitimate files acted on in non-optimal ways (e.g. accusations of malicious behaviour).
g) Any anomalies (e.g. strange or inconsistent behaviour by the service).

3.2 Analysis

Each email security service is monitored to detect its ability to detect, block or warn against threats. Malware and legitimate application samples that are allowed to pass are checked to ensure that they are still valid and have not been corrupted. Corruption of malware is allowed, while corruption of legitimate content is not.

Products are scored according to their success in warning users against threats or preventing such users from downloading these threats.

3.2.1 Scoring

Services are scored according to how they handle threats and legitimate messages. Allowing a threat into the user's inbox is a bad result that brings penalties. Allowing a legitimate email into the user's inbox, conversely, is a good result that awards points. Moving or editing messages to protect the user in varying levels brings different numbers of penalties or points. The table below provides details of the scoring system.

The main criteria for all scoring are: Is the user protected? Can the user do their work? How far away is the threat kept from the user?

Slight inconveniences of having inactive threats appear in Junk folders are not penalised.

The scoring is as follows, where the Threat Scores apply when a service uses the Action against a threat, while the Legitimate Scores apply when a service uses the Action against a legitimate message:

Action	**Threat Score**	**Legitimate Score**
Inbox	-10	10
Edited (Allow)	-10	10
Junk (Allow)	-7	7
Junk Folder	5	-5
Quarantined (User)	6	-6
Quarantined (Admin)	10	-10
Edited (Deny)	10	-10
Junk (Deny)	10	-10
Notified	10	-10
Stopped	10	-10
Rejected	10	-10
Blocked	10	-10

3.2.2 Action descriptions

Inbox
Malicious messages that arrive and remain in the user's inbox have evaded the security service. All legitimate messages should appear in the inbox.

Edited (Allow)
The service may change the email messages, attempting to remove or re-direct URLs, deleting attachments or text and taking other measures to remove the threat from the attacking emails. In some cases this attempt may be fully or partially unsuccessful and the threat remains, in which case we record that the service Edited the email but Allowed the threat.

Junk (Allow)
Services that send a threat to the Junk folder but make no further attempts to remove elements of threat have Allowed the threat to remain, albeit in the Junk folder. A user could recover the email from this folder and be at risk.

Junk Folder
Platforms that classify an email as Junk but make no further attempts to remove elements of a threat have allowed the threat to remain, albeit in the Junk folder. A user could recover the email from this folder and be at risk. We only use the Junk Folder result when testing an email platform without any additional third-party security products.

Quarantined (User)
Services may intervene and move malicious messages into a user-accessible quarantine system. Points are deducted for each legitimate message that is incorrectly sent to quarantine.

Quarantined (Admin); Edited (Deny); Junk (Deny); Notified; Stopped; Rejected; Blocked
Ideally the service detects the message containing a threat and prevents any significant element of that threat from reaching the intended recipient. This includes locking the message into a 'quarantine' accessible only by an administrator.

Other effective approaches include effectively removing all harmful links, attachments or text, for an Edited (Deny) result. The service might remove the message and notify the user, or it could refuse delivery altogether, (with or without notifying the recipient or sender).

If the service miscategorises and blocks, or otherwise significantly damages or limits access to legitimate email then penalties are applied.

3.2.3 Ratings calculations

Ratings are calculated by multiplying the number of each result type by the Threat Score or Legitimate Score. These Scores, in the table above, are based on our opinion of how important these different outcomes are. Consumers of our reports may have different views on how serious it is for a legitimate email to end up in quarantine, or for a malware threat to end up in the inbox. They can use the raw data from each report to roll their own set of personalised ratings.

4. Change log

Date	Version	Change
23rd July 2021	2.2	Corrected language; specified sample size; revised scoring system; introduced baselining; clarified that services are tested simultaneously; introduced BEC framework; introduced reactive remediation.
20th January 2022	2.21	Amended specified sample size; amended baseline period.
16th May 2025	2.3	Removed commodity threats from the test set; clarified scoring when testing email platforms without third-party security services; amended legitimate sample size; amended scoring for 'Junk (allow)' result.