Email security services testing methodology

Methodology version 2.1; Created: 15/08/2019

Contents

1.0 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 enables us to utilise the latest 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 the email security services under test.

2.0 Test framework

The test framework collects threats, verifies that they will work against unprotected targets and exposes protected targets to the verified threats to determine the effectiveness of the protection mechanisms.

2.1 Threat Management System (TMS)

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).

2.2 Threat Verification Network (TVN)

When threats arrive at the Threat Verification Network they are sent to Vulnerable Target Systems in a realistic way. For example, a target would load the URL for an exploit-based web threat into a web browser and visit the page; while its email client would download, process and open email messages with malicious attachments, downloading and handling the attachment as if a naïve user was in control. Links to malicious websites will be followed as far as possible.

Replay systems may be used to ensure consistency when using threats that are likely to exhibit random behaviours and to make it simpler for other labs to replicate the attacks.

2.3 Target Systems (TS)

Target Systems (TS) are identical to the Vulnerable Target Systems used on the Threat Verification Network, except that they are protected by email security services.

2.4 Service configuration

Services will be configured according to each vendor's recommendations, by the vendor where appropriate. Additional changes may be made to take into account the issues surrounding IP address reputation, which allows for testing that is as close to real life as possible, while also allowing for tests to be replicated across all services. These may include, but are not limited to:

a) Adding header metadata to provide original source IP address values
b) Adding source IP address values into the SMTP negotiation (e.g. XCLIENT)
c) Whitelisting the attacking systems' IP address(es) to allow reputation systems to accept the message for analysis, after which it may block the message based on the source IP address (see a) and b) above)

2.5 Threat selection

The following categories of threats are included:

a) Public (commodity)
b) Social engineering
c) Phishing
d) Targeted

Each category (other than the 'public' one) comprises different scenarios that represent popular approaches to each attack category. 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 10 different versions of the attack, varying in sophistication from very basic to sophisticated. These differing levels of sophistication might include changes in grammar; spelling; more or less convincing matches (or mismatches) of name and email address; email headers; and the inclusion of HTML and other rich content.

Public threats are sourced directly from attacking systems on the internet at the time of the test and can be considered 'live' attacks that were attacking members of the public at the time of the test run. Multiple versions of the same prevalent threats may be used in a single test run, but every version will be verified unique through its cryptographic signature.

Private threats are generated in the lab according to threat intelligence gathered from a variety of sources and can be considered as similar to more targeted attacks that are in common use at the test of the test run.

All threats are identified, collected and analysed independently of security vendors directly or indirectly involved in the test. The full threat sample selection will be confirmed by the Threat Verification Network as being malicious.

False positive samples will be messages containing popular and non-malicious website URLs, text-based messages with no harmful content and attached legitimate applications. These will comprise real legitimate email messages and generated messages that are clearly legitimate and will not easily be confused with malicious messages by users.

2.6 Target System details

The Target Systems are Windows PCs, deployed either as physical or virtual systems. Each system has unrestricted internet access and it isolated from other Target Systems using Virtual Local Area Networks (VLANs).

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

3.0 Threat selection and management

3.1 Sample numbers and sources

The Target Systems will be exposed to a selection of undesirable email messages. The range includes commodity 'public' threats and a wide range of targeted attacks of varying sophistication. Phishing

attacks, for example, will range from the obvious to the highly convincing. Ten email messages will be used for each scenario, as described in 2.5 Threat selection, providing a subtotal of 60 in each category and 240 in total for all four categories of threat.

3.2 Sample verification

Threats will be verified using Vulnerable Target Systems, as outlined above (see 1.0 Test framework).

Threat verification occurs throughout the test period, with live public threats being used within minutes or hours after they are verified as being effective against the Vulnerable Target Systems on the Threat Verification Network. In some cases, such as when significant attacks are discovered over a weekend, attacks may be deployed in the test a day or two after initial discovery to reflect the reality of users checking their email after a day or two without access.

In cases where a threat is initially verified to be effective, but which is found not to be effective during testing (e.g. its C&C server becomes unavailable) the threat sample will be excluded from the test results of each service.

3.3 Attack stage

Threats are introduced to the system in as realistic a method as possible. This means that threats found as email attachments are sent to target systems in the same way – as attachments to email messages. Links to web-based threats are downloaded directly from their original sources, via clicking through in the email.

4.0 Legitimate sample selection

Non-malicious email test messages are used to check for false positive detection. The number of these messages will match the number in one category of threats (currently 60). Candidates for legitimate sample testing include realistic email messages with a varying range of minor header and MIME misconfigurations and other corruptions. They may be sent directly to the target or forwarded. The email message body content will be clearly legitimate and not closely resemble harmful messages.

Attachments include a range of popular office documents; plain text and HTML; and files compressed with a range of popular compression utilities.

5.0 Measuring success

The following occurrences during the attack stage will be recorded.

5.1 The point of detection

(e.g. on arrival at the service; blocking a URL after a period of time).

5.2 Detection categorisation, where possible

(e.g. URL reputation, signature or heuristics).

5.3 Details of the threat, as reported by the product

(e.g. threat name; attack type).

5.4 Action on threat

(e.g. deletion, quarantine, delivered with warning, delivered without warning)

5.5 Legitimate files allowed to pass without problems

5.6 Legitimate files acted on in non-optimal ways

(e.g. accusations of malicious behaviour; blocking of installation)

5.7 Any anomalies

(e.g. strange or inconsistent behaviour by the product.)

6.0 Measuring product effectiveness

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 applications is not.

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

SE LABS LTD
ONE Croydon, CR0 0XT, United Kingdom.
Registered in England: 9688006.
Tel: +44(0)203 875 5000; Email: info@selabs.uk