Keywords: anti-malware; accreditation; compliance; assessment; testing; test plan; MRG Effitas; 360 Degree Assessment

AMTSO 2018
June 5, 2018

Version 1.2

MRG Effitas Test Plan for Q2 2018 360 Degree Assessment and Certification

Sponsored and Authored by:
MRG Effitas (Zoltan Balazs, Sveta Miladinov), AMTSO (John Hawes, Scott Jeffreys)

Abstract:
This Test Plan has been prepared jointly by MRG Effitas and AMTSO as part of the AMTSO Standards V1.0. The Plan details the MRG Effitas testing activities in 360 Degree Assessment and Certification for the period May through July 2018 with reporting taking place in August 2018. This document has been developed using AMTSO Test Plan Template Version 1.6 from January 2018. Wherever conflicts might exist between this Template and the member-approved Standards, the Testing Protocol Standards will provide the prevailing rule.

www.amtso.org

Table of Contents

MRG Effitas 360 Degree Assessment and Certification Test Plan – Q2'2018

1. Introduction

A first-of-its-kind test that covers all angles, our pioneering 360 Degree Protection Test targets the key threats faced by internet users. In each test case we employ the full spectrum of Early Life Malware. We use a Time-To-Detect metric to measure how long it takes each application to detect and neutralize missed threats.

MRG Effitas has a core focus on efficacy assessments in the anti-financial fraud space, but we also publish more traditional "Real World" detection tests. Our "Time to Detect Assessment Q4 2013" measured the ability of security products to protect an endpoint from a live infection, and, in the event of a system being compromised, the time taken to detect the infection and remediate the system. The time-to-detect-and-remediate component relied on each security product being manually forced to conduct a scan every thirty minutes over a 24-hour period. For 2014, it was decided that a new approach was needed as the methodology applied in previous tests did not reflect how a security product would be used on an endpoint in the Real World.

In practice, many security applications will only detect an infection during a reboot/startup or if a scheduled scan has been set by default. For this assessment, time-to-detect will employ a methodology based on the infected endpoint being re-scanned once during a 24-hour period. The methodology employed in this test maps more closely to Real World use, and although it may not be a 100% accurate model of how an "average" system is used, it gives a more realistic assessment of a security product's ability to detect and remediate an infected endpoint.

This Programme is called a "360 Assessment" since it deals with the full spectrum of malware instead of just financial malware. In the 360 Assessments, trojans, backdoors, ransomware, financial malware and "other" malware are used.

2. Scope

This Certification Programme is designed to serve as a reflection of product efficacy based on what we have previously termed "metrics that matter". In many of our previous tests, particularly those that have focused on financial malware, we started with the assumption that the endpoint has already been compromised. Being the world's largest supplier of early-life malicious binaries and malicious URLs, and from our own simulator development, we know that all endpoints can be infected, regardless of the security solution employed.

For us, a product's ability to block initial infection (although critical in most cases) is not the only metric that matters. One also needs to measure the time taken for the security product to detect malware on a system and remediate it. When conducting these tests, we tried to simulate normal user behaviour. We are aware that a "Real World" test cannot be conducted by a team of professionals inside a lab because we understand how certain types of malware work, how malware attacks and how such attacks could be prevented. Simulating normal user behaviour means that we paid special attention to all alerts given by security applications.

A pass was given only when alerts were straightforward and clearly suggested that malicious action should be blocked. We tested a group of internet security suites and complementary security applications. With these, it is very important to note that the best choice for an average user is to keep things very simple and for the product not to present many pop-up alerts or questions.

In our last execution of this Test Plan, sixteen products were considered. We expect a similar representation set for our upcoming Q2'2018 360-Degree Test. Companies expected to be represented and their target products include the following. Specific product versions will be determined after the Public Test Notification has been issued.

- avast! Internet Security
- Avira Internet Security
- BitDefender Internet Security
- ESET Internet Security (Smart Security)
- F-Secure Computer Protection
- Kaspersky Internet Security
- Microsoft Windows Defender
- McAfee Internet Security
- Symantec Norton Security
- Trend Micro Maximum Security
- Webroot SecureAnywhere

3. Methodology and Strategy

Since its inception in 2009, MRG Effitas has strived to differentiate itself from traditional testing houses by having its primary focus on providing "efficacy assessments" and not just performing "tests". Traditionally, testing of security software has centred on measuring a product's ability to detect malware.

Testing has evolved rapidly over the last two to three years as most labs, under the guidance of AMTSO (of which MRG Effitas is a member) strived to conduct "Real World" testing. Although there is no absolute definition of this kind of testing, loosely speaking, it involves the introduction of malware to an endpoint through a realistic vector, such as a browser or USB memory stick. Real World testing mostly involves "dynamic testing" (i.e. the malware is executed and then the ability of the security product to block the malware is measured). Several testing labs also conduct "System Rescue" tests. These assess a security product's ability to remediate a preinfected endpoint.

Whilst both types of tests are useful and yield valid and meaningful data, MRG Effitas wanted to merge these tests and also go one step further by measuring the time security products take to detect infections and remediate the endpoint. To make testing more akin to Real World scenarios, no manual scanning was conducted. Instead, the system was re-scanned once a day (exactly 24 hours after the system was compromised), thereby giving security applications the opportunity to detect infections on restart. As we have stated in our previous test reports, all

malware has one primary objective, and that is to make money for the cybercriminals. Measuring initial detection rates and the time taken to detect active malware is important, particularly in today's threat landscape with the mix of malware that is prevalent. As we have repeated in our previous financial malware test reports, the longer a cybercriminal can have their malware on a system, the greater the opportunity for them to be able to capture private user information including banking passwords and social media credentials, etc.

There has been an increase in the prevalence of ransomware, such as "CryptoLocker", which, once active on the system, holds the user at ransom to decrypt system data or unlock the system in some other way (interestingly, the most common way CryptoLocker to be installed on an endpoint is via Zeus infections). For these types of malware, it is initial detection that is of the greatest importance, since the vast majority of security solutions will be unable to rescue an encrypted or locked system.

In providing these quarterly certifications, the MRG Effitas 360 Assessment & Certification Programme is the de facto standard by which security vendors, financial institutions and other corporations can attain the most rigorous and accurate determination of a product's efficacy against the full spectrum of malware that is prevalent during the period.

The detailed Methodology to be used in the assessment adheres to the following procedures.

1. Windows 10 64-bit operating system is installed on a virtual machine, all updates are applied and third party applications installed and updated according to the MRG Effitas "Average Endpoint Specification".

2. An image of the operating system will be created.

3. A clone of the imaged systems is made for each of the security applications used in the test.

4. An individual security application will be installed using default settings on each of the systems created in (3) and then, where applicable, updated.

5. A clone of the system as at the end of (4) will be created.

6. Each live URL test is conducted by the following procedure.

    a. Downloading a single malicious binary from its native URL using Microsoft Edge to the desktop, closing Microsoft Edge and then executing the binary.

    b. The security application blocked the URL where the malicious binary was located.

    c. The security application detected and blocked the malicious binary whilst it was being downloaded to the desktop.

    d. The security application detected the malicious binary when it was executed according to the following criteria: It identified the binary as being malicious and either automatically blocked it or postponed its execution and warned the user that the file was malicious and awaited user input.

7. The system under test is deemed to have been infected if the security application fails to detect or block the binary at any stage in (6) and allowed it to be executed.

8. Testing on infected systems continues for 24 hours. The system was rescanned once, 24 hours after the system was initially compromised.

9. Remediation performance of an application was determined by manual inspection of the system in contrast to its pre-infected state and not by the logs and reports of the security application itself.

10. Tests are conducted with all systems having internet access.

11. Each individual test for each security application is conducted from a unique IP address.

12. All security applications were fully-functional unregistered versions or versions registered anonymously with no connection to MRG Effitas.

13. All testing will be conducted during Q2 2018.

14. As no user-initiated scans will be involved in this test, applications relied on various technologies to detect, block and remediate threats. Some of these technologies include background scanning, startup scanning, scheduled scanning, and system monitors. A scheduled scan will be used only if enabled by default.

4. Participation

AMTSO's goal with having Voluntary Participants is that in exchange for cooperating (engaging with Testers and following disclosure requirements), Voluntary Participants have additional rights to audit their configuration and provide commentary on Test results. There must be no additional cost to Participants to be Voluntary. If a Tester charges to participate in a Public Test or any related services, and a Participant chooses to not pay the fee, that Participant must be able to choose to be a Voluntary Participant and follow this AMTSO standard.

Opt-Out Policy : Vendors can opt out if Vendor can prove that the test system or the Product was misconfigured in a way which greatly changes the test results. E.g. important modules were unnecessarily turned off compared to default configuration, or during the test the system could never reach the cloud.

Conflict of Interest Disclosure : There is no known conflict of interest.

Funding : Funding of this project is achieved by vendors subscribing to participate in this project, this gives them more in-depth information on how their product(s) performed and, if there are any issues discovered in the product during testing, our technical team provides all the data necessary to help improve the product.

Part of the funding comes from directly licensing reports so they can be used for marketing purposes.

Vendors often silently enter testing, sometimes of their newly developed product or a product

in BETA or Pre-Release phase.

Finally, part of our funding comes from third parties, they commission us to include certain product into testing, both public and private.

5. Environment

Test hardware and configuration details follow.

Physical Configuration : Our Virtual Machine hardware specification calls for 4GB RAM and a dual core processor. AES includes Adobe Flash, Reader, Java, Microsoft Office 2010, Edge & VLC Player. All Microsoft components were fully updated and all third-party components were out of date by three months. During installation of the security application, if an option to detect PUAs was given, it was selected.

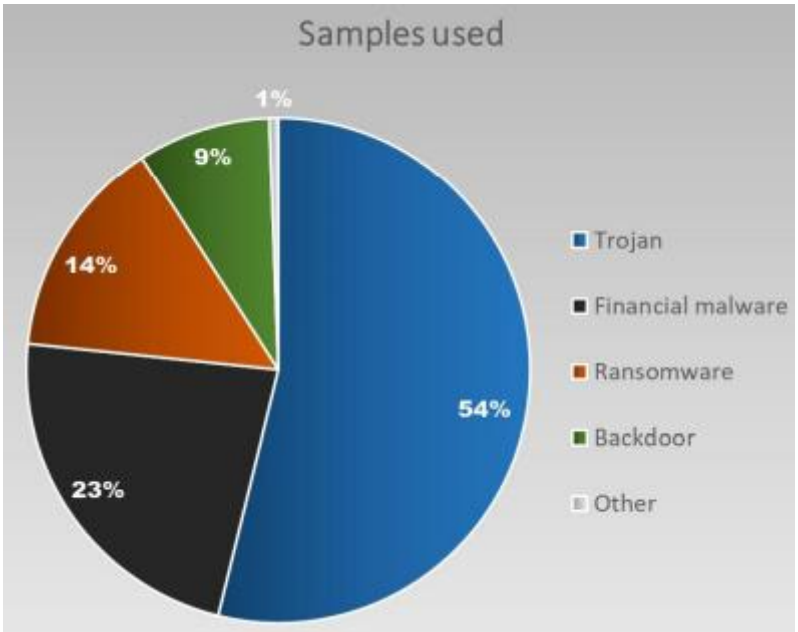
Sample Relevance : In the Wild 360 / Full Spectrum Test, approximately 50% of the malicious URLs used in this test were compromised legitimate websites which served malware. We believe that such URLs pose the greatest danger to users as this is the place where they least expect to get infected. 10% of the URLs pose as fake porn websites serving visitors with various types of malware. The remaining 40% of the URLs come from our regular honeypots or, in case of ransomware and financial malware in particular, we used URLs from newly-discovered distribution sites.

Geographic Limitations : There are no geographic limitations in terms of samples.

Curation Process : Voluntary Participants are given equal opportunities to participate in such Curation and feedback processes for all their respective Products.

Malware delivered by URLs used in this test can be considered as Zero Day in the true meaning of that phrase. It is our opinion that Ransomware currently poses the greatest threat to users, for this reason we choose to use more URLs serving this threat than before. Because of the wide spectrum of malware used in this project and the freshness of the samples, we used a smaller set than usual.

Applications that didn't protect the system from file encrypting ransomware cannot be certified because they could not remediate the threat as files usually cannot be decrypted. Our testing environment supports the use of VM aware malware, this is the reason why we were able to use more sophisticated threats which wouldn't run on Virtual Machines. 10% of the threats used in this test were introduced to the system via USB flash memory sticks. These samples came originally from live URLs, but inside archives. Testing was conducted as per the methodology detailed in Appendix 1. In total, 351 live ITW samples were used. The stimulus load comprised the following: 189 trojans, 30 backdoors, 80 financial malware samples, 50 ransomware samples, and 2 others.



Distribution of Test Data : We send all failed samples to all voluntary participants, along with detailed test logs.

6. Schedule

Start Date Range : The test commencement date is May 21, 2018.

Test Duration and Calculated End Date : The test is expected to require approximately seven weeks and is forecast to conclude on March 30, 2018.

Milestones : Delivery milestones appear in the following chart.

MRG-Effitas 360 Degree Test Project Schedule Milestones

Index	Test Activity	Start Date Range	Dependencies
1	*Test Commencement*	*May 23, 2018, Duration until August 1, 2018*	
2	*Confirm Vendor Configuration Feedback*	*May 23, 2018*	
3	*Milestone 1 – Preliminary Results*	*July 18, 2018*	*(1), (2)*
4	*Milestone 2 – Test Report First Edition – End of Testing Period*	*July 25, 2018*	*(3)*

5	*Feedback and Dispute Resolution Time – Retests as Needed*	*July 25, 2018 - August 1, 2018*	*(4)*
6	*Milestone 3 – Issue Final Report – End Date for Test*	*August 13, 2018*	*(5)*

Communications : Whenever there are significant deviations from this schedule (more than 5 days), we will notify all affected vendors within 3 business days.

Risks and Risk Management : We could not identify any risks with the test.

7. Control Procedures

Connectivity Validation : Tests are conducted with all systems having internet access. Each individual test for each security application is conducted from a unique IP address. All security applications are fully-functional unregistered versions or versions registered anonymously with no connection to MRG Effitas. Each vendor can supply a utility or another in-product feature to validate proper cloud-connectivity functionality.

Logging : In the first initial email to the vendors, we will ask if they require any special logging during the test. We expect these methods and tools to be working, otherwise we cannot guarantee the result of these logs.

Updates : Each individual security application will be installed using default settings on each of the test systems defined in the Environment section of this Test Plans and then, where applicable, updated.

8. Dependencies

Participant Actions : As this test scope are consumer products, we do not require any Vendor specific actions.

9. Scoring Process

In order to attain a quarterly MRG Effitas 360 Degree certification award, a security application must either protect the system from initial infection (behaviour protection) (a level 1 pass) or in at least 98% of all cases detect any missed malware and fully remediate the system before or on the first user initiated rescan (a level 2 pass).

Applications that meet this specification will be given certification for that quarter. Under the MRG Effitas 360 Degree Assessment & Certification, products are certified for Level 1 or Level 2 Compliance.

10. Dispute Process

After the vendors receive the test results and the logs, Vendors can dispute individual test cases if they can prove that they do not agree with the result.

11. Attestations

I understand and agree that I am submitting this Test Plan, and the following Attestations, on behalf of the entity listed below, and I represent and warrant that I have authority to bind such entity to these Attestations. All references to "I" or "me" or similar language refer to such entity. I represent and warrant that the following Attestations are true, to the best of my knowledge and belief, and each of the following commitments will be upheld to the best of my ability.

1. I will provide public notification on the AMTSO website covering my obligation for notification of a Public Test, regardless of whether a potential Participant is in actual receipt of such notification prior to the Commencement Date of a Test. (Section 1, Section 4, Section 6)

2. All products included in this Test will be analyzed fairly and equally. (Section 2, Section 3, Section 5)

3. I will disclose any anticipated or known imbalance or inequity in the Test design to all Participants in the Test. (Section 2, Section 3)

4. Although I may charge for participation in a Test, I will not charge any additional fees for a Test participant to be "Voluntary" under the Standards. (Section 4)

5. I will disclose any material conflicts of interest or other information that could materially impact the reliability of the Test. (Section 4)

6. I will disclose how the Test was funded. (Section 4)

I hereby affirm, to the best of my knowledge and belief that this Test Plan complies with the AMTSO Testing Standards, as of the date hereof.

Signature: /s/ Zoltan Balazs

Name: Zoltan Balazs

Test Lab: MRG Effitas

AMTSO Test ID: