Advanced Threat Prevention
with
Sandbox Analysis

Lab Testing Detailed Report

DR141002G

21 November 2014

Miercom
www.miercom.com

Contents

1.0 Executive Summary

Miercom conducted an Advanced Threat Detection and Sandbox Analysis test to determine the security efficacy (catch rate) of network-based threat prevention solutions that utilize sandboxing. The objectives of this test were to evaluate the security efficacy of vendor threat prevention solutions. Vendors represented in the assessment included: Check Point, Cisco, FireEye, Fortinet and another vendor that, due to vendor EULA restrictions, we need to refer to as Vendor A.

A representative set of tests was performed to assess the capabilities of each vendor product across multiple malware sample sets comprised of attack types ranging from legacy malware to Zero-Day (unknown) malware. Detection accuracy was assessed in all attack categories. Attack propagation is comprised of web requests that are typical in a business network. The attack scenarios and test bed were created by Miercom and no vendors provided any malware for use in the test.

Check Point outperformed all other vendors in this assessment. Specifically, accuracy and performance of the Check Point 4800 Next Generation Threat Prevention appliance with Threat Emulation Cloud Service outperformed all competitors with all malware sample sets.

In addition to assessing security efficacy, several other observations were noted such as:

- Usability
- Forensic Reporting
- Vendor Specific Limitations

The main objective of this testing focused on evaluating each security offering's ability to decompose, emulate, and accurately determine whether or not unknown malware samples were in fact malicious. Any observations or findings determined by our test team to be materially significant in a security solution's overall effectiveness were also noted in the report.

We hope you find the report findings useful and meaningful to your business.

Robert Smithers

CEO

Miercom

2.0 Products Tested

Five security solutions were evaluated as part of this testing. A short description of each product is summarized below including the type and configuration of each product:

- Check Point 4800 Next Generation Threat Prevention appliance with Threat Emulation Cloud Service, vR77.20
- Cisco Web Security Virtual Appliance 8.0.5 with Sourcefire AMP subscription. This web security appliance is a stand-alone virtual solution
- FireEye NX Series 1310 Malware Detection System v7.2
- Fortinet FortiGate 100D v5.2 appliance with FortiCloud FortiGuard Sandbox Subscription
- Vendor A Gateway and Cloud sandbox subscription service – This vendor has restrictions in their product license agreement on publishing results associated with their name so their name and product details are withheld.

All products were tested using their cloud sandbox solution, except FireEye, which doesn't offer a cloud solution for web traffic.

All product signature-based malware detection was enabled.

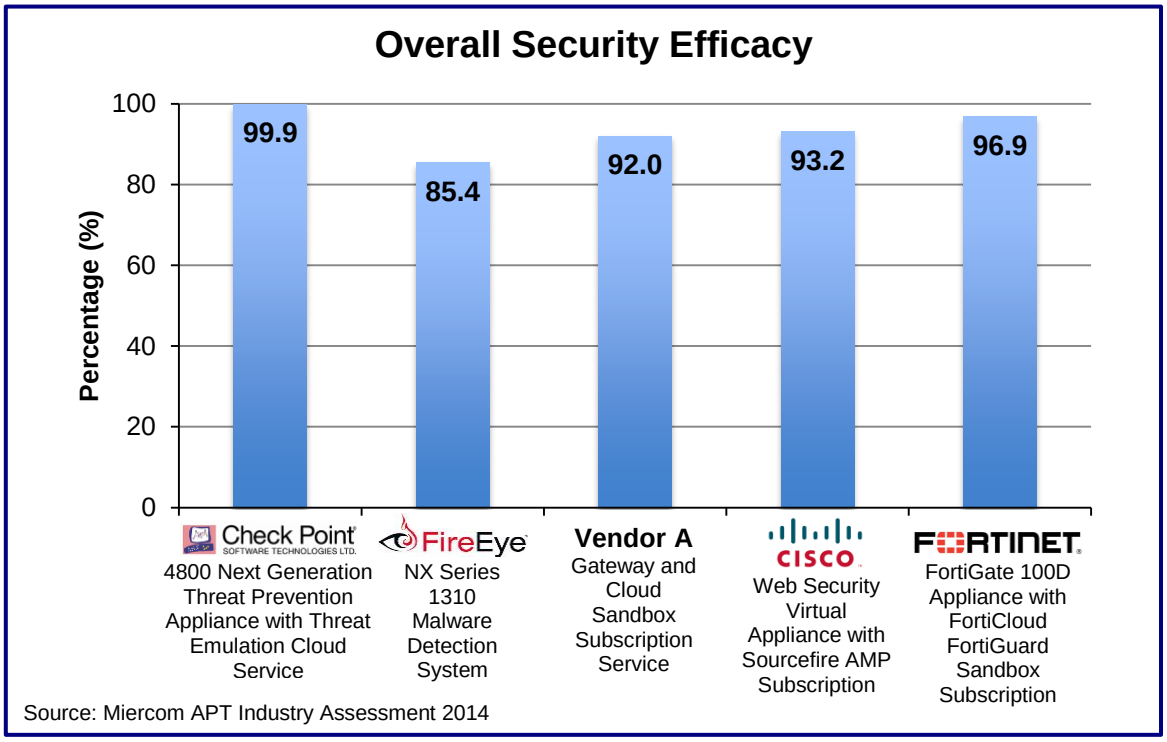
All vendor products were tested with the most current version available as of October, 2014 and were updated with the latest software updates.

See Section *6.0 About the Miercom ATP Industry Study* on *page 17* for details on Miercom Fair Test Disclosure.

3.0 Summary Results

Safeguarding enterprise networks requires detection accuracy and speed from the security solution. Further, advanced threat prevention solutions must be highly effective in their detection of both known and unknown threats. The five vendor solutions tested represent the most advanced threat prevention solutions in the industry. The chart below presents the results of the overall security efficacy of each product against the full malware sample set.

3.1 Summary Security Efficacy Results



Source: Miercom APT Industry Assessment 2014

3.2 Average Detection Times

Threat prevention solutions analyze suspicious items in the sandbox environment to determine if they are malicious. Miercom engineers observed large differences in the amount of time each sandbox solution needed to analyze a malware sample. The following table shows the average sandboxing time that both malware and benign samples were analyzed in each product's sandbox. Note, analysis of malware samples typically take longer than benign samples.

Vendor	Average Time per Sample in Sandbox
Check Point 4800	~ 3 minutes
Vendor A	~ 3 minutes
Cisco Web Security	~ 11 minutes
Fortinet FortiGate-100D	~ 14 minutes
FireEye NX Series 1310	~ 18 minutes

4.0 Threat Prevention Details and Analysis

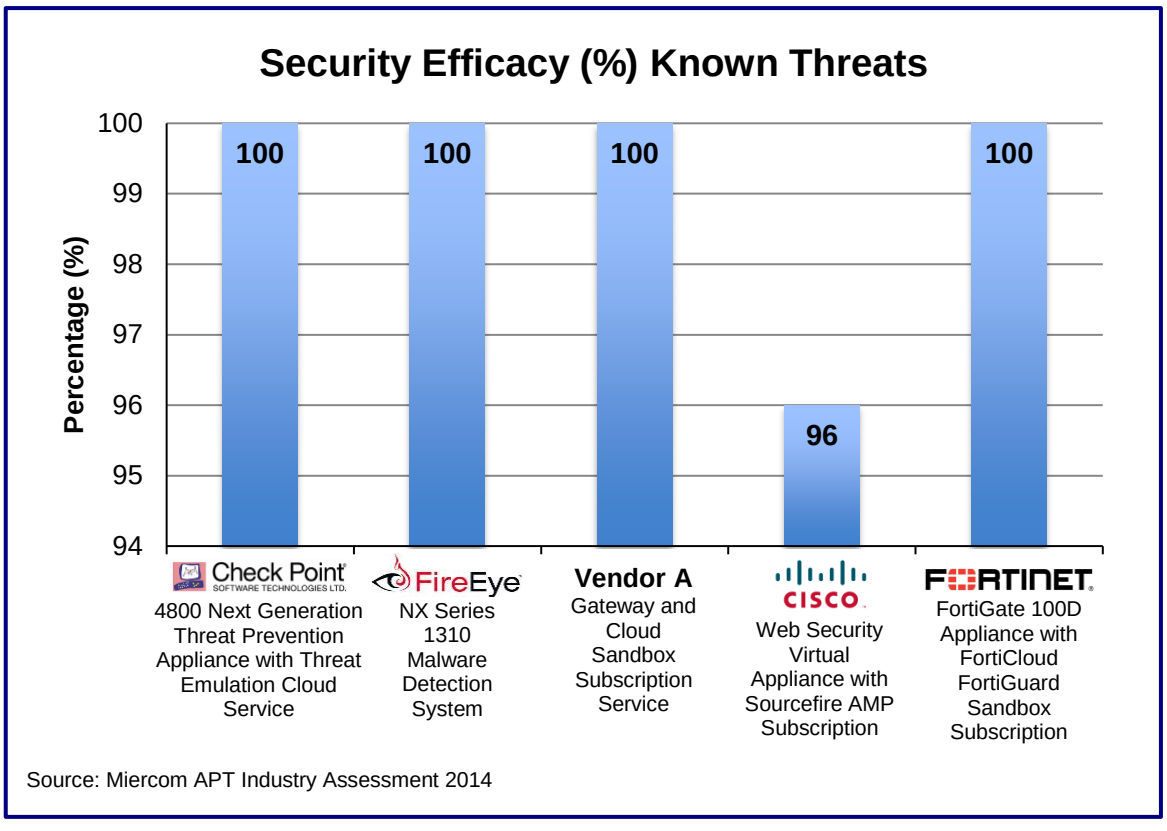
Threat prevention solutions must protect against both known and unknown threats. Typically, signature-based protection like antivirus (AV) and Intrusion Prevention Systems (IPS) detect and block known malware from infecting the organization. Unknown malware is typically detected in the sandbox where it is emulated to determine suspicious or known bad behavior. This speeds the evaluation process since the AV can typically be operated at a much faster rate than the sandbox analysis.

Each product tested has signature-based malware detection. For the purposes of testing, this protection was enabled on each appliance as a preliminary defense. If a signature triggers on a sample, the threats are immediately mitigated. If a sample is not detected based on its signature, then it is forwarded to the sandbox for further analysis.

4.1 Protection against Known Threats

Advanced threat prevention should ensure that known threats are easily identified and properly handled. This examination was considered baseline testing, evaluating the efficacy of each product's antivirus (AV) protection against known malware. The graph below represents the outcome of subjecting each product to a sample set of legacy malware.

Known Threats (Legacy Malware) = Malware samples that are at least 30 days old and have been validated by a minimum of 15 vendors confirmed on VirusTotal.

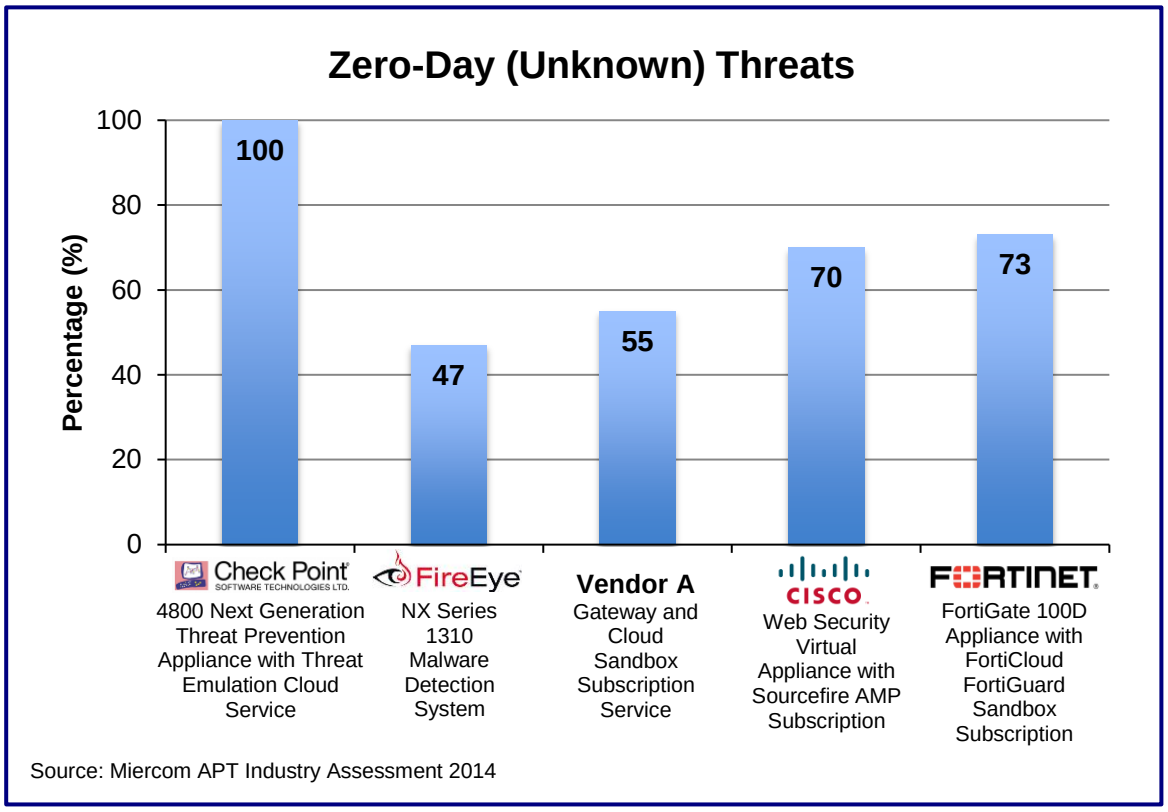


Source: Miercom APT Industry Assessment 2014

4.2 Protection against Zero-Day (Unknown) Threats

Sandbox efficacy is determined by the ability of the product to correctly analyze the sample and provide an accurate verdict as to whether that sample is malicious or benign. Zero-Day samples do not have a known signature and are therefore sent to the sandbox for analysis.

Zero-Day Threats (Unknown Threats) consist of malware samples that have been custom crafted, undetected samples acquired from external resources and private honeypots, and APTs that have undergone AV evasion techniques (encryption, black packaging, payloads that use normal and allowed egress traffic, etc.). The chart below represents the security efficacy of each product in detecting unknown threats.



Source: Miercom APT Industry Assessment 2014

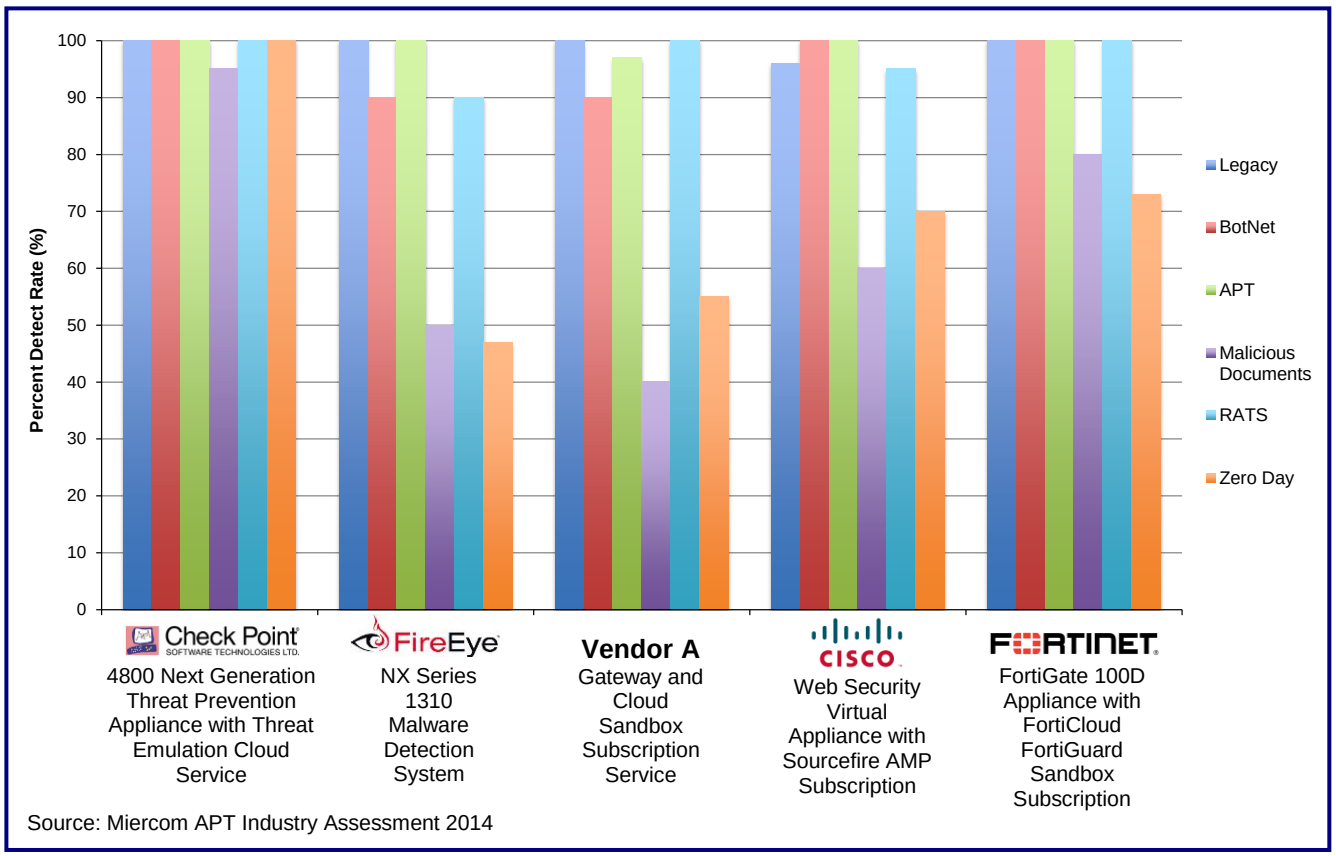
4.3 Protection by Threat Type

The chart below represents the sample classifications that each product was subjected to and the percentage of threats blocked by each product in the respective category.

To determine security efficacy, the following formula was used:

$$\text{Security Efficacy} = \frac{\text{Number of Samples Mitigated}}{\text{(Total Number of Samples)}} * 100$$

Overall Vendor Security Efficacy by Sample Set



Source: Miercom APT Industry Assessment 2014

Overall Vendor Security Efficacy Results by Percentage

	Check Point	FireEye	Vendor A	Sourcefire	Fortinet
Legacy	100	100	100	96	100
Botnet	100	90	90	100	100
APT	100	100	97	100	100
Malicious Documents	95	50	40	60	80
RATs	100	90	100	95	100
Zero-Day	100	47	55	70	73

4.4 Vendor-Specific Product Analysis

Throughout the course of the analysis, general observations regarding each product were made. These observations include attributes such as: accuracy, speed, efficacy and reporting.

Vendor	Findings
Check Point	**Summary:** Check Point 4800 had an intuitive user interface, the product was accurate, efficient, and most effective against Zero-Day threats. **Accuracy**: Highest level of accuracy of all solutions tested. Samples sent to ThreatCloud Emulation Cloud Service were properly classified and mitigated. **Efficiency**: Most efficient sandbox. First-line-of-defense (AV protection) was accurate and eliminated unnecessary overhead by reducing the number of emulated samples. **Speed**: Detection time was the fastest among all products tested at approximately 3 minutes per sample emulated. **Reporting**: Detailed forensic reporting included is highly valuable to customers for isolating incidences. The level of reporting is in line with industry standard incident response process and procedures.
Cisco	**Summary:** Cisco Web Security Virtual Appliance performed as expected for a virtual appliance. This solution is fairly accurate while identifying known threats, but performed poorly detecting Zero-Day samples. **Accuracy:** This was the only product tested that missed legacy (known malware) samples and struggled to identify threats contained within common file formats such as office documents and portable document files. **Efficiency:** Sandbox analysis was slow, and (as noted above) certain samples were commonly missed. Some known threats were sent to the sandbox even if the underlying AV definitions contained a signature for a specific sample. **Speed:** Results took an average of 11 minutes / sample. Large sample sets would bog down the performance of the product resulting in longer wait times to obtain a verdict. Such a delay could impact network production operations and thus business productivity. **Reporting**: Report was difficult to follow and lacked the level of detail expected from a product in this class.

FireEye	**Summary:** FireEye NX Series 1310 surprisingly demonstrated the worst efficacy for Zero-Day malware detection of the products tested. **Accuracy:** NX Series 1310 was the least accurate against the Zero-Day sample set and struggled with malicious documents. Legacy samples, APTs, and Botnets were accurately classified. **Efficiency:** Because FireEye NX Series 1310 sends every file to the sandbox for analysis, it was the most inefficient product tested. Results on a single sample set took days to obtain a verdict. The approach taken by this product to send every file to the sandbox for analysis causes an unnecessary amount of overhead. **Speed:** Average time to detect was ~18 minutes / sample. Some large samples sets surprisingly took several days to obtain a correct detection rating. **Reporting**: Forensic reporting was accurate and provided a good amount of detail. However, it was difficult to navigate the report and identify the source of a threat and infected nodes.
Fortinet	**Summary:** Fortinet FortiGate performed fairly well. It ranked second in overall efficacy, but with considerable limitations (see below). **Accuracy:** FortiGate struggled with the malicious document sample set and even reported several false-negatives (produced a "benign" verdict for a sample that was known to be malware). **Efficiency:** This product did a good job at reducing appliance overhead by immediately blocking known samples. However, FortiCloud suffered in performance. It would become overwhelmed with large sample sets and report false-negatives. **Speed:** FortiGate was the second slowest product tested in the test bed, which was due to FortiCloud's sandbox analysis. Known threats were rapidly identified, but it took an average of ~14 minutes per sample that were sent to the cloud for further analysis. **Reporting**: FortiGate's reporting is remedial. It provides admins with a "birds-eye" view with little forensic value.

Vendor A	**Summary:** Vendor A had an exceptionally fast product with exceptionally poor results. Typical sandbox implementation consists of a physical appliance and a cloud-based (subscription) sandbox. The subscription is based on an upload quota. Miercom identified many occurrences where instances of the same file were analyzed multiple times. This notion suggests that based on Vendor A's subscription model, customers are being charged multiple times for the same file to be analyzed. Even worse, the results varied each time. **Accuracy:** Vendor A's AV protection was accurate, however, their cloud-based sandbox was the most inaccurate of all products tested. Malware classifications were incorrect and the inaccuracy of the verdicts leaves customers with a false sense of security. **Efficiency:** Efficiency of Vendor A's product is very good, but questionable based on the inconsistent outcome of each sample analysis. **Speed:** Speed of the product was one of the fastest at an average of ~3 minutes per analysis, but consideration of scanning the same sample multiple times and the inconsistency of the analysis outcome must be taken into account. **Reporting**: Local reporting on the appliance lacked detail. However, sandbox analysis performed in the cloud provided a fairly detailed report, which was easy to read, and easy to navigate.

4.5 Summary Findings

Check Point 4800 Gateway with ThreatCloud Threat Emulation scored the highest in the following areas in all tests:

Detecting and Blocking Multiple Types of Malware Threats

Where other products fail, Check Point succeeds, and where other products succeed, Check Point outperforms the competition. No other product came close to the number of samples caught. In every sample set, Check Point was able to out-perform all competitive products tested.

Sandbox Effectiveness and Accuracy

Threat Emulation Cloud Service beat the competition on accuracy, efficiency and effectiveness. All threats that were forwarded to ThreatCloud for emulation were correctly identified, classified, and mitigated.

Forensic Reporting

Check Point has the most detailed reporting for forensic analysis and incident response. Data visualization is clear and concise, which allows for better incident

response time. The level of detail in logging is far superior to all other products tested. Sandbox analysis provided accurate timeline of malware propagation, screenshots of the events, modifications to the file system, etc.

Manageability and Effectiveness

Product deployment is straightforward. Management software is easy to use.

Performance Leader: Check Point 4800 Next Generation Threat Prevention with Threat Emulation Cloud Service

The Check Point 4800 appliance, consisting of a physical appliance and cloud-based sandbox, provided the best overall protection. Detection rate was 100 percent for legacy, botnet, advanced persistent threats (APTs), RATs and Zero-Day samples. Detection rate dropped slightly to 95 percent for Malicious Documents and that rate was better by 16 percent or greater when compared to similar products.

Check Point 4800 was the most accurate of all products tested capturing the most known and unknown malware. Check Point had the fastest sandbox analysis time for unknown malware. Data was easy to navigate on the Check Point solution and the AV and sandbox delivered fast results.

FireEye NX Series 1310 Malware Detection System

The FireEye NX Series 1310 solution consists of a physical appliance with a built-in sandbox. The product caught 100 percent of legacy and APT samples and 90 percent of botnet and RAT threats. The FireEye solution scored 50 percent or lower in identifying malicious documents and Zero-Day samples.

The dashboard was appealing but the results were difficult to navigate. The Miercom test team found FireEye NX Series to be extremely slow in that gathering the results took days. Because sandbox analysis is performed locally on the appliance, it is a good overall solution for an environment where data does not leave the network.

Cisco Web Security Virtual Appliance with Sourcefire AMP Subscription

Consisting of a virtual appliance and a cloud sandbox, Cisco Web Security was easy to deploy. Detection rate was 100 percent for botnets and APTs with 95 percent or better detection rate for legacy and RAT samples. Cisco achieved 70 percent in identifying Zero-Day samples and 60 percent detection of malicious documents.

The Cisco sandbox is fairly slow and the gathered results were very difficult to navigate.

Fortinet FortiGate-100D with FortiCloud FortiGuard Sandbox

Fortinet FortiGate-100D, consisting of a physical appliance and a cloud sandbox, was easy to deploy and navigate. The solution detected 100 percent of legacy, botnet, APTs and RAT samples. Malicious documents were detected at 80 percent and Zero-Day samples were identified at 70 percent.

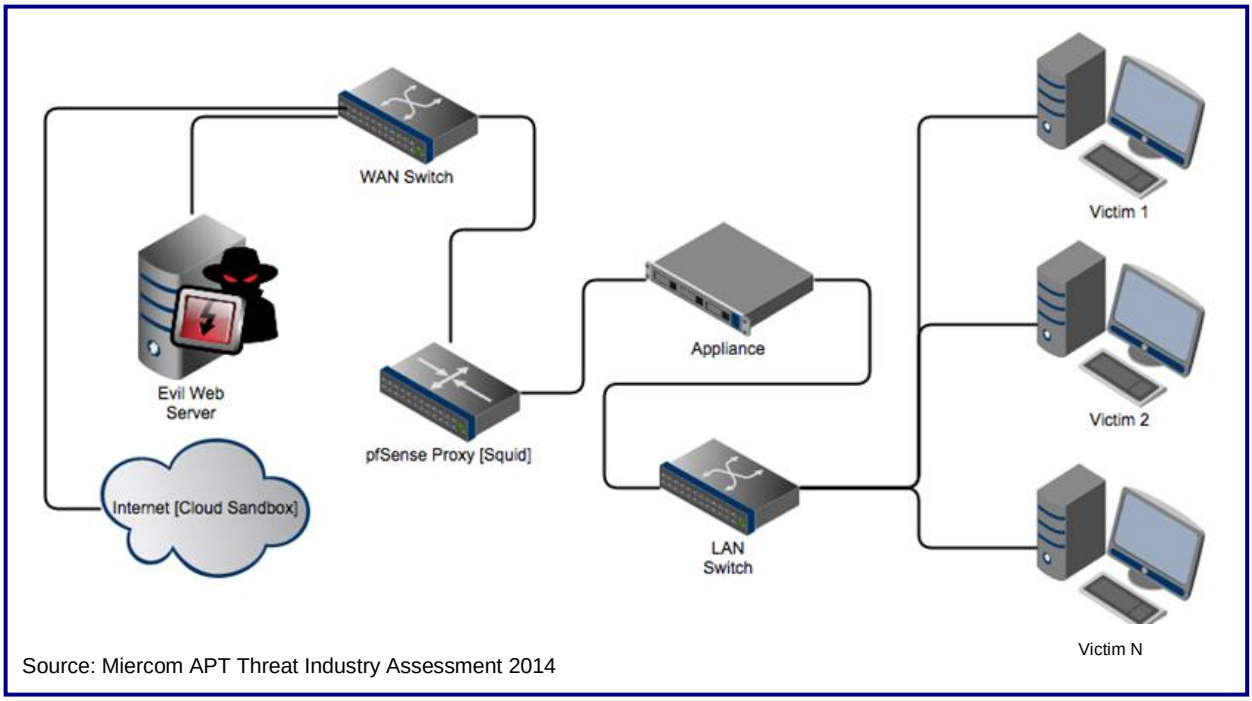
The data logs of FortiGate-100D lacked detail. Miercom engineers rated the AV with an excellent score but the solution's sandbox scored poorly. Sandbox verdicts were inconsistent.

Vendor A Gateway and Cloud Update Service

Vendor A's product, consisting of a physical appliance and a cloud sandbox, was easy to deploy. The appliance scored 97 percent or better when dealing with legacy, APT and RAT threats. Catch rate was 55 percent for Zero-Day samples and 40 percent for malicious documents.

Miercom engineers observed that not enough data was captured by the Vendor A product. Log entries were not numbered and some reverse engineering was required. The product quarantined the most false negatives, and overall results were misleading and inconsistent.

5.0 Test Bed Diagram



Source: Miercom APT Threat Industry Assessment 2014

5.1 How We Did It

A test bed was created containing each appliance, a series of victim machines for each appliance, and a malicious web server that was used to serve up the malware samples. The end-nodes (victims and malicious web server) were all virtualized, but on different hardware to ensure that the machine state was the same throughout testing.

To ensure delivery, a lightweight web application was developed to organize the sample sets. Additionally, the application performed a user agent verification on the client browser to eliminate accidental propagation of the malware.

The web server would be called upon by the victim via an HTTP GET request and the appliance would act as an intermediary. Each appliance configuration was carefully reviewed to ensure that every vendor represented in the test bed was equally deployed with comparable features.

Prior to performing the analysis, each appliance was verified working by issuing a baseline test with a select number of malware samples. This test ensured that each product was functioning and reporting on the malicious files being requested by the victim machines.

The baseline test consisted of a small legacy sample set of malware used to ensure that each appliance was working correctly. The base samples chosen were checked against VirusTotal. Each sample set contained a variety of malware classifications and each product was confirmed working before conducting tests for the record.

Evil Web Server

- Debian Linux using the LAMP Stack (Linux Apache MySQL PHP)
- The web application checked that the request originated from a custom browser agent to prevent accidental propagation of the samples
- Each sample set was organized and placed on a separate page

Victim Browser Configuration

- Windows 8.1
- Firefox v. 32.0.x
- Firefox DownThemAll plugin - To download the large sample sets automatically

The tests in this report are intended to be reproducible for customers who wish to recreate them with the appropriate test and measurement equipment. Contact Miercom Professional Services via reviews@miercom.com for assistance. Miercom recommends customers conduct their own needs analysis study and test specifically for the expected environment for product deployment before making a product selection. Miercom engineers are available to assist customers for their own custom analysis and specific product deployments on a consulting basis.

5.2 Malware Sample Sets

The malware sample sets used in this analysis were obtained from various public and private sources. Known malware (Legacy, APT, BotNet, Malicious Documents, RATs / Trojans) were obtained from VirusTotal and other public sources. Zero-Day samples were custom crafted by both internal and external resources, obtained from private honeypots that have been deployed around the globe, and APTs that have undergone AV evasion techniques, such as encryption, black packaging, payloads that use normal and allowed egress traffic, etc.

- **Legacy Malware Files**

  Legacy samples included several hundred variants of known malware that have been in circulation for 30 days or more. The malware classifications primarily consist of viruses and worms.

  The legacy sample sets contain variants of:

  Sysbot - Spyware

  AutoRun - Virus

  Danger - Trojan

  Hooker - Trojan

  Injector - Trojan

  Homepage - Spyware

  ZeroAccess Rootkit

  Hijack - Trojan

  Infector – Virus

This portion was considered baseline testing, evaluating the efficacy of each product's antivirus (AV) protection against known malware. A series of known malware tests were conducted prior to sending any new, previously unknown samples (documents, RATs, botnets, etc.). Legacy malware samples should be detected and mitigated without the need for sandbox analysis. Should any sample pass through a vendor's antivirus filter, the sandbox should then have identified it immediately due to the known heuristics of each malware sample.

- **Advanced Persistent Threat (APT)**

  Advanced Persistent Threats (APTs) are considered "back doors" into a victim network. APT malware consists of a staged payload that, when activated, allows an attacker to obtain shell access. The attacker then has command line access to the remote target at the same privilege level as the vulnerable application or service. These payloads are often masked with randomization and evasion techniques to bypass AVs. The (known) APT samples used in our testing were sourced from Mandiant's Advanced Persistent Threat sample set.

- **Botnet**

  A botnet is a collection of Internet-connected programs communicating with other similar programs in order to perform tasks. Botnets use a technique known as Command and Control, where an intermediary receives orders from an attacker and those commands are then forwarded to all infected hosts. Botnets are commonly used in spamming and DDoS operations. Variants of the Zeus and Citadel botnets were collected from high-interaction honeypots and used in this test.

- **RATs**

  RATs, or Remote Access Trojans, are malicious code disguised as something normal or desirable so they often masquerade inside other legitimate software. When activated in a victim host, they provide full remote control over that victim. The RAT sample set used in our testing consisted of a mix of MS Office documents and PDF files.

- **Malicious Documents**

  An additional sample set of malicious documents used in testing contained a mix of Microsoft Office documents (Microsoft Word, PowerPoint and Excel files) that held known macro viruses, and PDF files containing a variety of viruses, APTs and worms.

- **Zero-Day (Unknown Threats)**

  Zero-Day Threats (Unknown Threats) consisted of malware samples that have been custom crafted, undetected samples acquired from external resources and private honeypots, and APTs that have undergone AV evasion techniques (encryption, black packaging, payloads that use normal and allowed egress traffic, etc.).

6.0 About the Miercom ATP Industry Study

This report was sponsored by Check Point Software Technologies, Ltd. The data was obtained completely and independently as part of the Miercom ATP Industry Study. The study is an ongoing endeavor in which all vendors have equal opportunity to participate and contribute to the most comprehensive and challenging test program in the industry.

All vendors with products featured in this report are afforded the opportunity to represent themselves and even challenge these results in a retest if they express interest in doing so.

Among vendors who responded to notification prior to testing:

Cisco Systems declined to participate at this time in the study as a new product release is scheduled. Vendor A declined to participate and cited EULA licensing requirements prohibit competitive benchmark testing of their product.

Vendors declining to participate may still be tested without their permission.

6.1 About Miercom

Miercom has published hundreds of network-product-comparison analyses in leading trade periodicals and other publications. Miercom's reputation as the leading, independent product test center is undisputed.

Private test services available from Miercom include competitive product analyses, as well as individual product evaluations. Miercom features comprehensive certification and test programs including: Certified Interoperable, Certified Reliable, Certified Secure and Certified Green. Products may also be evaluated under the Performance Verified program, the industry's most thorough and trusted assessment for product usability and performance.

6.2 Use of This Report

Every effort was made to ensure the accuracy of the data contained in this report but errors and/or oversights can occur. The information documented in this report may also rely on various test tools, the accuracy of which is beyond our control. Furthermore, the document relies on certain representations by the vendors that were reasonably verified by Miercom but beyond our control to verify to 100 percent certainty.

This document is provided "as is," by Miercom and gives no warranty, representation or undertaking, whether express or implied, and accepts no legal responsibility, whether direct or indirect, for the accuracy, completeness, usefulness or suitability of any information contained in this report.