TEST METHODOLOGY

Breach Detection Systems (BDS)

MARCH 5, 2018

v5.0

Table of Contents

1 Introduction

1.1 The Need for Breach Detection

Threat actors are demonstrating the capability to bypass protection offered by conventional endpoint and perimeter security solutions. Enterprises must in turn evolve their network defenses to incorporate a different kind of protection, one that NSS Labs defines as a breach detection system (BDS).

Through constant analysis of suspicious code and identification of communications with malicious hosts, breach detection systems are capable of providing enhanced detection of threats ranging from commodity malware to targeted attacks from state-sponsored threat actors that could bypass defenses such as next generation firewalls (NGFWs), intrusion prevention systems (IPS), intrusion detection systems (IDS), antivirus/endpoint protection (including host IPS), and secure web gateways (SWGs).

1.2 About This Test Methodology

NSS test reports are designed to address the challenges faced by IT professionals in selecting and managing security products. The scope of this particular methodology includes:

- Security effectiveness
- Performance
- Stability and reliability
- Total cost of ownership (TCO)

Based on the needs identified in NSS' research, the following capabilities are essential in any breach detection system:

- Centralized management of multiple devices
- Breach detection using one or more of the following methods:
  - Malware identification (signatures, heuristics, or both)
  - Network traffic analysis (flow monitoring, content analysis, or both)
  - Sandboxing that allows for modeling of internal systems
  - Emulation
- Response mechanism (alerting, session termination, etc.)
- Robust logging of conviction events
- Reporting

Security Effectiveness: At the heart of the BDS test harness is a patented technology called **BaitNET™**. This cloud-based, fully instrumented, targeted execution environment, is a unique live test harness that is used by NSS for security effectiveness testing on all leading endpoint and network security products.

Using this test harness, NSS is able to determine the detection offered by the BDS as a whole, as well as many of its subcomponents. Some basic principles:

- All products are tested in a way that does not bias the results (for example, multiple operating system and application configurations (stacks) are tested at the same time; there is verification that the attacks are delivered; extensive measures are taken to ensure that threat actors do not blacklist the test network).

- Exploits are validated for efficacy against precisely configured stacks.
- Features of the BDS products are tested as they are used in the real world.
- In order to provide the most actionable information, testing utilizes actual, live attacks from genuine cybercriminals and other threat actors.
- Scoring is based upon observed results; attack success or failure. For example:
  - Was the attacker able to execute code or manipulate existing processes (e.g., Shellcode)?
  - Was malware installed on the victim machine, or was the machine modified in other ways?
  - Was a successful outbound connection initiated?

See the current Security Stack (Network) Test Methodology for more details.

Performance: The aim of this section is to verify that network-based appliances are capable of detecting and logging breaches when subjected to increasing loads of background traffic up to the maximum bandwidth supported.

Tuning: Each vendor will be expected to be capable of monitoring the same range of operating systems and applications as in the NSS Labs Live Testing™ harness. Vendors will be informed of the specifications of all software and operating systems in the test harness and will be permitted to tune their products or create custom virtual machines (VMs) to model those environments. Vendors will be provided with a baseline sample set of malicious software ahead of testing in order to ensure their products are functioning correctly. This baseline sample set will be used to verify basic detection and performance capabilities only at the start of the test and will not count toward final security effectiveness scores.

NSS Labs test methodologies are continually evolving in response to feedback. If you would like to provide input, please contact advisors@nsslabs.com. For a list of changes, please reference the Change Log in the Appendix.

1.3 Inclusion Criteria

NSS invites all BDS vendors claiming breach detection capabilities to submit their products at no cost. Vendors with major market share, as well as challengers with new technology, will be included.

1.4 Deployment

The BDS should be supplied as a single network-based appliance at minimum. If there is an endpoint agent that complements the appliance, it can be used for endpoint visibility only. The endpoint agent must be deployed in non-blocking mode; i.e., it must be strictly in observation mode for the test. BDS typically operate out of band, in detection mode (similar to IDS), implementing multiple techniques to analyze and report on malicious traffic. BDS can also be deployed inline, in which case a tap may need to be supplied to NSS.

Network appliances should have the appropriate number of physical interfaces capable of achieving the required level of connectivity and performance. The test harness and methodology support testing of up to 10Gbit fiber interfaces. The BDS will be connected to the test network via regenerative network taps for out-of-band devices or inline at the egress with an appropriate vendor-supplied inline tap. A separate connection will be made from the management port to the management network. The BDS should be able to send all conviction events to syslog and/or have the functionality to export the logs from UI for analysis.

2 Security Effectiveness

The aim of this section is to verify that the BDS is capable of detecting and logging breaches and attempted breaches accurately, while remaining resistant to false positives. All test cases in this section are completed with background network load.

The BDS must retain all logs and event information, including information about the malicious activities that enabled systems to achieve conviction events. Event log records should be atomic, i.e., no updates are permitted after the initial logging. Any updates from additional analysis that results in changes to conviction/risk scores should be reflected in newer log records and appropriately time stamped.

This test utilizes real threats and attack methods that exist in the wild and are actually being used by cybercriminals and other threat actors, based on attacks collected from NSS' global threat intelligence network.

For detail on live testing, please refer to the latest Security Stack (Network) Test Methodology.

2.1 False Positive Testing

This test will include a varied sample of legitimate network traffic, file formats and executables. The BDS should not generate events for non-malicious traffic. In addition to known-good samples of common file formats and executables, false positive testing can include non-malicious samples that may exhibit characteristics commonly, but not uniquely, associated with malicious software.

False positive scores will be based on the propensity of the BDS to generate an event when analyzing non-malicious files or network traffic.

2.2 Detection Engine

The ability of the BDS to detect and alert on successful breaches in a timely manner is critical to maintaining the security and functionality of the monitored network. Successful infiltration or transmission of malware should be reported quickly and accurately, giving administrators opportunity to arrest the breach and minimize impact to the network.

The use of standardized logging and reporting formats, which facilitate the fast and accurate consumption of presented data, is imperative to enable administrators to make the correct decisions. The BDS should allow easy generation and exportation of reports, logs, and/or alerts into one or more of these formats:

- CSV
- XML
- JSON
- Other machine-parseable formats may be acceptable. Please coordinate with NSS to ensure compatibility and adequate capabilities.

As response time is critical in halting the damage caused by a breach, the BDS should be able to detect known malware, or analyze unknown malware, and report on them within 24 hours of initial infection, command and control (C&C) callback, or other malicious outbound traffic. Any BDS that cannot, or does not, log an attack, infection, or C&C callback within this detection window will not receive credit for the detection.

Any combination of the following tests may be conducted against each stack, and the BDS will be expected to identify the same breach as reported by the NSS Labs Live Testing™ harness:

2.2.1 Exploits

Exploits are attacks against the computer that take advantage of an underlying vulnerability in an application to perform unauthorized tasks. Exploit vectors can include, but are not limited to:

- Drive-by exploits against web browsers, plug-ins, extensions and add-ons (e.g., Java, Flash)
- Exploits delivered via social engineering—exploits that require user interaction, including but not limited to exploits embedded in documents such as PDF, .docx, HTML).

2.2.2 Malware

Traditional malware requires user interaction to download and install. Malware delivery vectors include but are not limited to:

- Web browser downloads
- Email attachments

The definition of malware includes but is not limited to so-called "grayware" samples where the overt functionality of the software:

- Has no legitimate purpose (for instance, if it is illegal in a major jurisdiction)
- Exposes the organization to undue risk in terms of regulatory compliance, financial exposure, etc.

2.2.3 Offline and Out-of-Band Infections

Mobile assets owned by the enterprise, such as laptops, may become infected when outside the security provided by enterprise security solutions. Such situations include working while on an aircraft, at coffee shops, on unprotected networks, or when using infected removable media (USB-based storage, etc). Also, non-enterprise-issued assets, such as contractor laptops, could introduce infections into the intranet when attached to it.

This test assesses the ability of the BDS to detect and report such infections when they are deployed in out-of-band or inline mode, in a solution that may or may not have endpoint agents deployed.

2.3 Evasion

NSS verifies that the BDS is capable of detecting basic exploits/drops when subjected to varying common evasion techniques. It is a *requirement* of the test that the submitted BDS has all evasion detection options enabled by default in the shipping product. Wherever possible, a component of the BDS is expected to successfully decode the evasive traffic to provide an accurate detection relating to the original exploit, rather than detecting purely on anomalous traffic seen as a result of the evasion technique itself.

Evasion testing will be conducted in accordance with the version of the NSS Labs Evasions Test Methodology published at the time BDS testing commences.

3 Performance

This section measures the performance of the BDS using traffic conditions that provide metrics for real-world performance. These quantitative metrics provide an indication of the suitability of a solution for a given environment. All tests are repeated at 25%, 50%, 75%, and 95% of the maximum rated throughput of the BDS.

The malicious sample(s) will be provided in advance of the test to ensure that every vendor can accurately catch and identify attacks. The goal of this test is to stress the device's maximum detection processing capability and find the point at which connections are overwhelmed or the table and buffer capacity is exceeded, and attacks are allowed through. At each stage, multiple instances of malicious traffic are passed, and the number detected is logged. Throughput is recorded at the highest point at which all attacks are detected.

3.1 Maximum Capacity

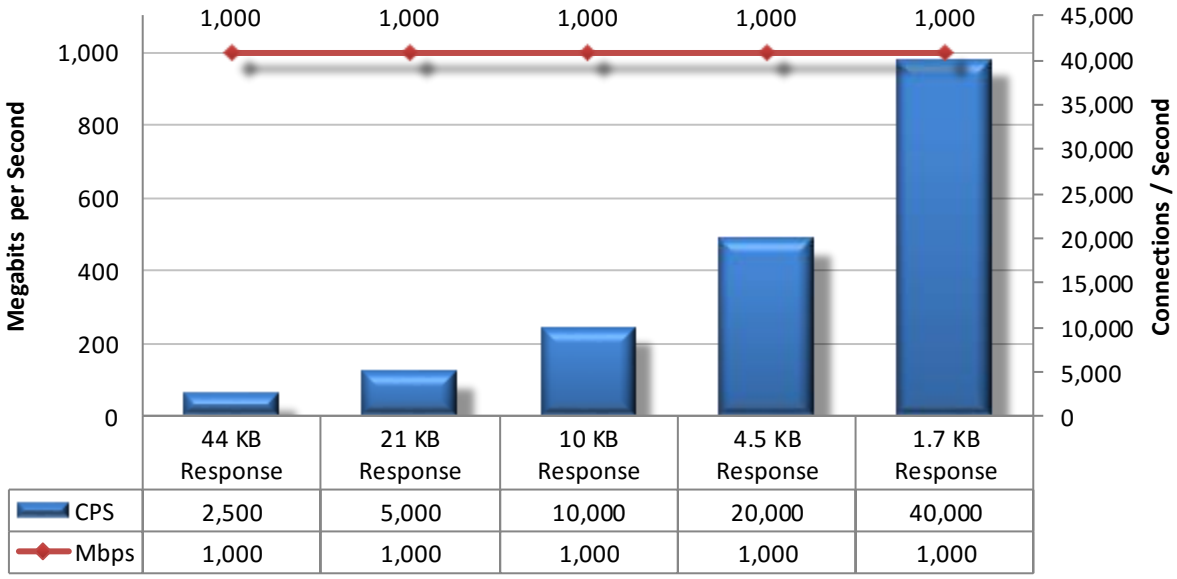
The use of automated testing and traffic generation appliances allows NSS engineers to create true "real-world" traffic at multi-Gigabit speeds as a background load for the tests.

The aim of these tests is to determine how the BDS handles increasingly challenging network loads while performing core functions. All packets contain valid payload and address data, and these tests provide an excellent representation of a live network at various connection/transaction rates.

3.1.1 HTTP Capacity

The aim of these tests is to stress the HTTP detection engine and determine how the BDS copes with network loads of varying average packet size and varying connections per second. By creating genuine session-based traffic with varying session lengths, the BDS is forced to track valid TCP sessions, thus ensuring a higher workload than for simple packet-based background traffic.

Each transaction consists of a single HTTP GET request, and there are no transaction delays (i.e., the web server responds immediately to all requests). All packets contain valid payload (a mix of binary and ASCII objects) and address data, and this test provides an excellent representation of a live network (albeit one biased towards HTTP traffic) at various network loads.



3.1.1.1 44 KB HTTP Response Size – 2,500 Connections per Second

Maximum 2,500 new connections per second per Gigabit of traffic with a 44 KB HTTP response size—maximum 140,000 packets per second per Gigabit of traffic. With relatively low connection rates and large packet sizes, all hosts should be capable of performing well throughout this test.

3.1.1.2 21 KB HTTP Response Size – 5,000 Connections per Second

Maximum 5,000 new connections per second per Gigabit of traffic with a 21 KB HTTP response size—maximum 185,000 packets per second per Gigabit of traffic. With average connection rates and average packet sizes, this is a good approximation of a real-world production network, and all hosts should be capable of performing well throughout this test.

3.1.1.3 10 KB HTTP Response Size – 10,000 Connections per Second

Maximum 10,000 new connections per second per Gigabit of traffic with a 10 KB HTTP response size—maximum 225,000 packets per second per Gigabit of traffic. With smaller packet sizes coupled with high connection rates, this represents a very heavily used production network.

3.1.1.4 4.5 KB HTTP Response Size – 20,000 Connections per Second

Maximum 20,000 new connections per second per Gigabit of traffic with a 4.5 KB HTTP response size—maximum 300,000 packets per second per Gigabit of traffic. With small packet sizes and extremely high connection rates, this is an extreme test for any host.

3.1.1.5 1.7 KB HTTP Response Size – 40,000 Connections per Second

Maximum 40,000 new connections per second per Gigabit of traffic with a 1.7 KB HTTP response size—maximum 445,000 packets per second per Gigabit of traffic. With small packet sizes and extremely high connection rates, this is an extreme test for any host.

3.2 "Real-World" Traffic

Where previous tests provide a pure HTTP environment with varying connection rates and average packet sizes, the aim of this test is to simulate a "real-world" environment by introducing additional protocols and real content while still maintaining a precisely repeatable and consistent background traffic load. The result is a background traffic load that is closer to that which may be found on a heavily utilized "normal" production network.

3.2.1 "Real-World" Protocol Mix (Enterprise Perimeter)

Traffic is generated across the BDS comprising a protocol mix typically seen in an enterprise perimeter.

3.2.2 "Real-World" Protocol Mix (Financial)

Traffic is generated across the BDS comprising a protocol mix typically seen in a large financial institution.

4 Stability and Reliability

Long-term stability is important even in passive devices such as BDS, since a failure can result in serious breaches remaining undetected.

These tests verify the stability of the BDS along with its ability to maintain security effectiveness while under normal load and while detecting malicious traffic. Products that cannot sustain logging of legitimate traffic, or that crash, while under hostile attack will not pass.

The BDS is required to remain operational and stable throughout these tests and to detect 100% of previously detected traffic, raising an alert for each. If any malicious traffic passes undetected—caused by either the volume of traffic or by the BDS failing for any reason, this will result in a FAIL.

4.1 Detection under Extended Attack

The BDS is exposed to a constant stream of security policy violations over an extended period of time. The BDS is configured to detect and alert, and thus this test provides an indication of the effectiveness of both the detection and alert handling mechanisms.

A continuous stream of malcious traffic is transmitted through the BDS for 8 hours, with no additional background traffic. This is not intended as a stress test in terms of traffic load (covered in the previous section); it is merely a reliability test in terms of consistency of detection performance.

The BDS is expected to remain operational and stable throughout this test, and it is expected to detect 100% of recognizable violations, raising an alert for each. If any recognizable policy violations are passed, caused by either the volume of traffic or by the BDS failing open for any reason, this will result in a FAIL.

4.2 Power Failure and Persistence of Data

The BDS should retain all configuration data, policy data, and locally logged data once restored to operation following power failure.

5 Total Cost of Ownership and Value

Implementation of security solutions can be complex, with several factors affecting the overall cost of deployment, maintenance, and upkeep. All of these should be considered over the course of the useful life of the solution.

- **Product Purchase –** The cost of acquisition
- **Product Maintenance –** The fees paid to the vendor (including software and hardware support, maintenance, and updates)
- **Installation –** The time required to take the device out of the box, configure it, deploy it in the network, apply updates and patches, perform initial tuning, and set up desired logging and reporting
- **Upkeep –** The time required to apply periodic updates and patches from vendors, including hardware, software, and firmware updates

Appendix A: Change Log

Version 5.0 – January 2018

- Renamed section 2.2.3 to “Offline Infections” to “Offline and Out-Of-Band Infections”
- Modified section 2.3 and added reference to NSS Labs Evasions Test Methodology
- Removed 4.4: Protocol Fuzzing and Mutation
- Changes to wording in the following sections:
    - 1.2: About This Test Methodology
    - 1.4: Deployment (Updated to include ‘inline’ solutions)
    - 2: Security Effectiveness
    - 2.1: False Positive Testing
    - 2.2 Detection Engine
    - 2.2.1: Exploits
    - 2.2.2: Malware
    - 2.2.3: Offline and Out-Of-Band Infections
    - 3: Performance
- Updated contact information with office address

Version 4.0 – November 2016

- Improved evasion scope
- Fixed grammatical errors within document

Version 3.0 –September 2015

Version 2.0 –2 May, 2014

Version 1.5 – 29 January, 2013

No Change Log available. Change Log Appendix added with version 2.0.

Contact Information

NSS Labs, Inc.
3711 South Mopac Expressway
Building 1, Suite 400
Austin, TX 78746
info@nsslabs.com
www.nsslabs.com