Public

Network Security Appliance Testing Methodology (Performance)

Contents

1. Test environment

The test environment is the technical infrastructure within which the tests are conducted on each Device Under Test (DUT).

1.1 Equipment used

The main types of equipment involved in this test comprise:

1. The DUT itself. This may be a hardware unit or a virtualised appliance.
2. The test network to which the DUT is attached.
3. Traffic generation devices designed to replicate a realistically busy network.
4. Network management equipment designed to control the test without interfering with the test network.

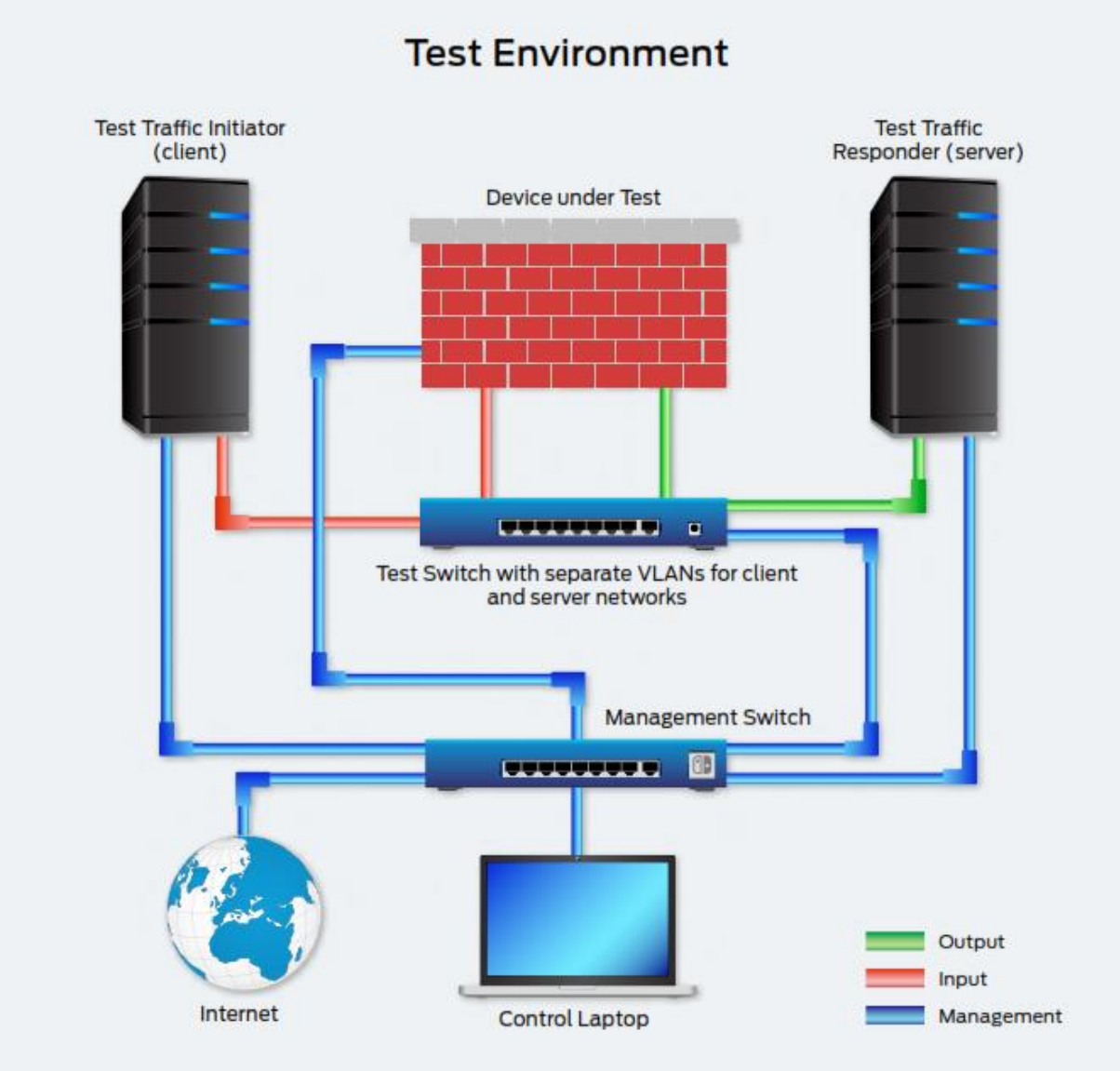
The goal is for the infrastructure to have no impact on the performance of the DUT, so that test results reflect, as closely as possible, a realistically attainable and repeatable performance. When delays are caused by the environment, such as when traffic generating equipment introduces overhead through using encryption, these will be highlighted in the report.

Specific details of the equipment used, including model names, firmware versions, configuration details and any virtualised hardware details are highlighted in each test report. Network diagrams and any further specific details are available to authorised parties engaged in the testing.

1.2 Network topology

1.2.1 Test network

The DUT is connected to a test switch configured with virtual LANs (VLANs) such that it has direct connectivity with the traffic generation systems and a management network.



1.2.2 Management network

The DUT and traffic generation systems are connected to a management network that is used to provide a means to control and monitor all systems involved in the test.

1.3 Configuration policies

The DUT will be configured according to realistic, publicly available recommendations made by the vendor. The vendor will be able to confirm the settings and may run initial tests to confirm the configuration is active or instruct SE Labs to do so.

The configuration details of the product are made available to readers of any testing report. Decisions to configure products differently to policies recommended by the DUTs' vendors are explained in all reports.

Configurations may be made available in the reports themselves or as assets, links to which will appear in the report.

2. Test cases

The test comprises different test cases that assess a DUT's abilities to transfer data quickly and reliably; and how responsive its users will find their experience of the network.

2.1 Test goals

The goal of this test is to assess the network performance capabilities of the DUT in a way that is realistic, transparent and repeatable.

2.2 Network traffic description

The type of traffic used in the test is documented in each individual report and includes details of the traffic used and the ratios of the mixture of protocols, applications and services.

In general, we expect to use the load specified by the Benchmarking Methodology Working Group of the Internet Engineering Task Force, which is supported by the NetSecOPEN standards organisation. At the time of writing we followed version 2.0 of the draft document Benchmarking Methodology for Network Security Device Performance V2. Updated references will be included in each test report.

2.3 Test case details

2.3.1 Mixed traffic capacity

This test indicates how much data can pass through the device in a real-world production environment, rather than a sterile and theoretical laboratory test.

It should answer the question, "how much throughput can I expect if I buy and use this?"

A realistic mixture of network traffic, as might be expected to pass through an enterprise network firewall, is sent through the device. Different traffic protocols are used at different times during the test.

This challenging test requires the device to check, track and respond to lots of different types of connections. It shows devices at their best or worst.

2.3.2 Application traffic capacity

This test indicates how well the device can perform with specific applications and services, rather than the mixture of protocols used in the Mixed Traffic Capacity test (see *2.3.1 Mixed traffic capacity*). The applications and services are tested in isolation to each other and not concurrently.

The results should highlight specific strengths and weaknesses in the device's ability to handle different types of network traffic. For example, a device might achieve a high throughput for FTP traffic, but SMTP traffic performance could be lower. Facebook throughput might be high, but Skype might suffer due to a lower throughput, introducing latency issues that are important to avoid when video conferencing.

We test with different loads until errors reached a threshold of over 1%. We then report the previous load, which was used before the error rate reached the threshold.

2.3.3 HTTP and HTTPS capacity

These tests indicate how much data can pass through the device when it is handling web sessions. It submits the tested device to a range of loads, starting with a low amount of network traffic and measuring its ability to transfer data as that load increases. The throughput measurements show how busy a network can be before the device starts to struggle and under-perform. At the same time, the test measures how many connections and transactions per second are possible.

The connections per second measurement shows how many basic web conversations are possible at any one time, while transactions per second measures how many groups of connections can be achieved. A basic request to a web server, and its response, is a connection. For example, requesting and receiving a single HTML page counts as one connection. A transaction comprises a single action that invokes multiple connections, such as you would experience when loading a webpage containing text, some images and other elements.

2.3.4 HTTP and HTTPS latency

This test indicates how responsive the device is when operating under normal loads.

It submits the tested device to single, 'normal' loads and measures the time it takes for web traffic to be downloaded. In line with NetSecOPEN's testing methodology and the IETF's Benchmarking Methodology for Network Security Device Performance (RFC 9411), we define 'normal' as being 50% of the maximum number of connections and transactions per second achieved by the device without significant error levels.

Latency is measured in two ways: by timing how long it takes to download the full body of a transaction (e.g. a basic web page without images, external CSS files etc.) and by timing how long it takes for the first piece of the web page to be received by the client's browser. These measurements are called 'URL response time' and 'time to first data byte' respectively.

Together the latency measurements show how smoothly users can expect to experience web browsing when the device is on the network. The URL response time shows how quickly they can expect to download full pages, while the 'time to first data byte' results shows how fast they will experience the beginning of a connection.

For example, a fast 'time to first data byte' result would mean that the user would see the web browser connect fast to the website and start downloading content. However, a relatively slow URL response time would mean that the page itself and the elements it contains might take some time to download fully. In contrast, a slow 'time to first data byte' result would mean that the user waits for the initial connection to establish but, if the URL response time was fast, the page would then quickly download.

We take an average of the 'time to first data byte' values for both HTTP and HTTPS results to create an overall result.

3. Scoring

Each report shows how close to an optimum result each DUT comes, and whether or not it achieves an acceptable level of performance. The DUT may be given an award if it achieves certain levels of performance in the six parts of the overall test. The criteria for the different award levels is as follows:

AAA: Good or Excellent in all six test elements
AA: Good or Excellent in five test elements
A: Good or Excellent in four test elements
B: Good or Excellent in three test elements
C: Good or Excellent in two test elements

To summarise SE Labs' position on acceptable and optimum ('good' to 'excellent') performance in the test cases above:

3.1 Mixed traffic capacity

Excellent throughput: Above 75% of the DUT's stated maximum.
Good throughput: Between 50% and 75% of the DUT's stated maximum.
Poor throughput: Between 25% and 50% of the DUT's stated maximum.
Very poor throughput: Below 25% of the DUT's stated maximum.

3.2 Application traffic capacity

Excellent throughput: Above 90% of the DUT's stated maximum load.
Good throughput: Between 80% and 90% of the DUT's stated maximum load.
Poor throughput: Between 70% and 80% of the DUT's stated maximum.
Very poor throughput: Below 70% of the DUT's stated maximum.

3.3 HTTP and HTTPS capacity

Excellent result: Above 90% of the traffic load.
Good result: Between 80% and 90% of the traffic load.
Poor throughput: Between 70% and 80% of the traffic load.
Very poor throughput: Below 70% of the traffic load.

3.4 HTTP and HTTPS latency

Connections Per Second

Excellent result: Under 1.5ms to first data byte.
Good result: Between 1.5ms and 2.0ms to first data byte.
Poor throughput: Between 2.0ms and 2.5ms to first data byte.
Very poor throughput: Above 2.5ms to first data byte.

Transactions Per Second

Excellent result: Under 0.5ms to first data byte.
Good result: Between 0.5ms and 1.0ms to first data byte.
Poor throughput: Between 1.0ms and 1.5ms to first data byte.
Very poor throughput: Above 1.5ms to first data byte.

4. Anomalies

Testers record any strange or inconsistent behaviour shown by the product. These will, in the first instance, be reported to the vendor for review.

5. Change log

23/01/2024 v1.1 Correction of definition of ‘normal’ for HTTP/S Latency test; Clarification on calculations of Connections and Transactions Per Second for scoring purposes.

23/03/2020 v1.0 Document created.