TEST METHODOLOGY

Software-Defined Wide Area Network (SD-WAN)

November 22, 2019

v3.0

Table of Contents

1 Introduction

1.1 The Need for the Software-Defined Wide Area Network (SD-WAN)

Modern networks are complex and difficult to manage. Software-defined networking (SDN) enables dynamic network performance monitoring and configuration by separating the network control plane from the forwarding plane. This abstraction enables the use of logical objects composed of multiple physical objects as opposed to having to manage individual physical objects with complex policies.

Wide area networks (WANs) have undergone several stages of development, from the early days of dial-up to dedicated circuits such as T1 / E1, to frame relay and ATM that then implemented multi-protocol label switching (MPLS) as an overlay technique to improve performance. Throughout their history, however, the goal of WANs has been to connect networks across long distances.

The marriage of software-defined networking (SDN) with wide area network (WAN) technology enables efficient management of complex diverse deployments through the use of common virtual private network (VPN) capabilities and the separation of data and control planes within SDN. With SD-WAN, software-managed connections can be established and managed between multiple sites over any number of link types (e.g., fixed circuit, DSL, cable, mobile, MPLS, and so on) without the operational challenges of having to manage different links. SD-WANs manage traffic according to application or service requirements (e.g., VoIP vs. Facebook), and enforce policy control capabilities (e.g., limit web-based traffic to 50% of a given link). SD-WAN options are part router, part WAN optimization and traffic shaping, part access control, and part VPN. In addition, some SD-WAN offerings provide robust security, which makes for a compelling alternative to the multiple-appliance approach that is often required at remote locations.

1.2 About This Test Methodology

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

- Management
- Impairment
- Performance
- Security effectiveness *(required for devices with built-in security)*
- Stability and reliability
- Total cost of ownership (TCO)

Since SD-WAN technology manages WAN links connecting a site with either headquarters locations or the public Internet, its stability and reliability is imperative. Therefore, regardless of any security capabilities, the main requirement of any SD-WAN is that it must be as stable, as reliable, as fast, and as flexible as the edge technology and device that it is replacing.

The following capabilities are considered essential in an SD-WAN:

- Traditional routing and policy control features, including:
  - Basic application identification and policy controls
  - Stateful networking controls

  - Virtual private network (VPN)
- Highly resilient remote office connectivity
- Prioritization of applications
- Remote configuration capabilities
- Predictable performance experience for users

Tuning: For SD-WAN technologies submitted with security and inspection abilities, tuning of security configurations will follow procedures from the NSS Labs Next Generation Firewall Test Methodology so that impairments, performance, and security are tested using the same configuration.

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

1.3 Inclusion Criteria

In order to encourage the greatest participation, and to allay any potential concerns of bias, NSS invites all vendors claiming SD-WAN capabilities to submit their products at no cost for our group test. Vendors with major market share, as well as challengers with new technology, will be included.

Where possible, the SD-WAN should be supplied as five separate appliances, with the appropriate number of physical interfaces capable of achieving the required level of connectivity for two HQ DC (Headquarters Data Center), accommodating resilience or HA, and three branch office locations. Each location will have two WAN links; comprising either an MPLS and a standard broadband connection (shown as ISP 1 and ISP 2 in Figure 1), or two standard broadband connections. Once installed in the test environment, the SD-WAN will be configured appropriately for the following use cases.

1.3.1 Multiple Use-Case Test Environment

Figure 1 presents the topology of a multiple use-case test environment that includes a redundant HA cluster at the HQ DC location and three disparate branches representing large, medium, and small enterprises. The WAN environment is provisioned with behavioral characteristics similar to those typically encountered over WAN link states. For example, the site locations are considered to be at a distance of ~1000 miles from the Headquarters (HQ DC), e.g., Denver (Colorado, USA) is HQ DC; San Francisco (California, USA) is Branch 1; Chicago (Illinois, USA) is Branch 2; Galveston (Texas, USA) is Branch 3. The test harness baseline is recorded to ensure consistent behavior, the vendor solution is deployed, and each test case is measured against the baseline.

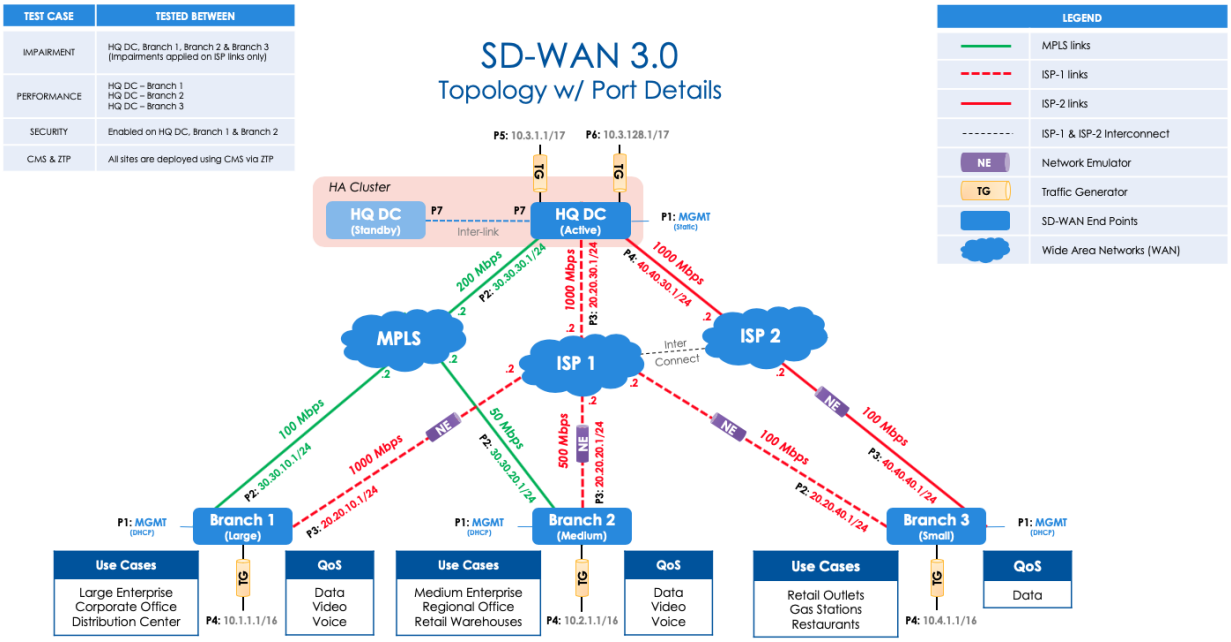

Figure 1 – Topology of Multiple Use-Case Test Environment: SD-WAN 3.0

1.3.2 Minimum Requirements for Test Inclusion

1. Minimum IPSec VPN throughput and number of devices required for HQ DC and Branches:

Location	Minimum IPSec VPN Throughput	Minimum Number of Devices
HQ DC	4 Gbps	2 (1 x Active, 1 x Standby)
Branch 1	1 Gbps	1
Branch 2	500 Mbps	1
Branch 3	200 Mbps	1

2. A central management system (CMS) for configuration and deployment of each location using zero-touch provisioning (ZTP). NSS recommends using a cloud-delivered CMS that has all of the features available in a standard enterprise deployment.
3. All applicable software licenses required for the topology (e.g., bandwidth licenses, cloud licenses, branch-to-HQ licenses, licenses required for logging, monitoring and reporting, etc.).
4. IPSec VPN encryption support for AES-256 at a minimum (If products do not support AES-256, this will be noted.)
5. Devices must be equipped with the minimum number of ports to support all LAN and WAN connections, i.e., Branches 1, 2, and 3 must each have 2 WAN ports and 1 LAN port; HQ DC must have 3 WAN ports and 2 LAN ports.
6. Devices must be provided with Ethernet (RJ-45) ports (i.e., not optical ports).
7. Devices must support routing protocols OSPF and BGP.

2 Product Guidance

NSS tests products based on evaluation criteria that are relevant to enterprise networking and security professionals. These criteria are as follows:

- **Network policy effectiveness** – The purpose of an SD-WAN is to connect multiple networks over traditional assured links and commercial broadband links and identify and manage applications and performance between sites.
- **Security effectiveness** – SD-WAN technologies with security and inspection capabilities are expected to be capable of inspecting and blocking malicious content as well be resilient against basic and complex evasion techniques.
- **Stability and reliability** – Long-term stability is particularly important for an inline device, where failure can produce network outages and business disruption.
- **Performance** – Correctly sizing an SD-WAN is essential.
- **Value** – Customers often desire low TCO, high effectiveness and predictable performance.

Products are listed in rank order according to their guidance rating.

2.1 Recommended

A *Recommended* rating from NSS indicates that a product has performed well and deserves strong consideration. Only the top technical products earn a *Recommended* rating from NSS, regardless of market share, company size, or brand recognition.

2.2 Neutral

A *Neutral* rating from NSS indicates that a product has performed reasonably well and should continue to be used if it is the incumbent within an organization. Products that earn a *Neutral* rating from NSS deserve consideration during the purchasing process.

2.3 Caution

A *Caution* rating from NSS indicates that a product has performed poorly. Organizations using these products should review their security posture and other threat mitigation factors, including possible alternative configurations and replacement. Products that earn a *Caution* rating from NSS should not be short-listed or renewed.

3 Management

3.1 Remote Initial Configuration

SD-WAN technology helps organizations achieve operational savings by enabling remote configuration of new locations rather than requiring engineers to be on-site. Many vendors offer ZTP, where onsite engineering expertise is not required other than the ability connect a device to the appropriate internal and external links and power up the device. Once online, the device will call "home" (whether that is the HQ or a cloud configuration service) to gather and download the operational configuration information. This can also be achieved via a CMS (where applicable, this will be noted in test reports). The SD-WAN is expected to be remotely configurable in order to receive a pass for this remote configuration test case. For this test, the time to configure and deploy the site will be recorded.

Time to create configuration: Time taken to create a new configuration or clone from an existing configuration, apply updates where necessary, and add a new site to the existing network topology. The metric includes creation of a template configuration, updating IP addresses for management interfaces, creating VPN tunnels for WAN links, setting up thresholds and traffic policies to allow or deny traffic, and installing the required security packages wherever required.

Time to deploy configuration: Time taken to deploy the configuration and includes connecting to the CMS, selecting the appropriate configuration to be deployed, validating for errors/issues, and provisioning the device for a desired site.

Both metrics will be components of the operational cost model used to calculate an SD-WAN's TCO.

3.2 Centralized Management System (CMS)

The CMS must provide enterprises with the ability to centrally manage, configure, and monitor devices via a graphical user interface (GUI). CMS capabilities include but are not limited to monitoring, reporting, configuration changes, and the ability to update software on the SD-WAN. These capabilities will be evaluated in the test.

4 Routing & Access Control

This section verifies that the SD-WAN is capable of consistently and effectively enforcing network configuration policies. NSS testing of SD-WAN network policy effectiveness is conducted by incrementally building upon a baseline configuration (simple routing with no policy restrictions) to create a complex, real-world, multiple-zone configuration that supports many addressing modes, policies, applications, and inspection engines. Exception events that occur must also be accompanied by detailed log information in order to enable both network and security forensics. Additionally, administrative visibility is critical. To facilitate analysis and troubleshooting, NSS requires the logging of any function or event that results in dropped traffic.

The SD-WAN must support persistent policy capabilities that ensure the prioritization of application traffic is managed securely. The SD-WAN must be able to manage policy across multiple interfaces/zones. At a minimum, the firewall must provide a "trusted" internal interface, an "untrusted" external/Internet interface, and (optionally) one or more DMZ interfaces. In addition, a dedicated management interface (virtual or otherwise) is preferred.

Every NSS test scenario employs routing. The SD-WAN must be capable of handling common enterprise routing protocol configurations (e.g., BGP and OSPF). Each of the following test cases intends to measure and record an SD-WAN's ability to perform according to its parameters. In each case, the scenario is built to mimic real-world deployments and traffic experiences in addition to common configurations in order to reveal how well the configuration policy performs.

All of the test cases possess specific industry-accepted metrics that record the performance capabilities of the tested SD-WANs. In test cases where more than one measure is taken, individual scenarios will evaluate specific metrics.

4.1 Site-to-Site VPN (IPSec, SSL, or Other)

An SD-WAN manages links between sites as VPN tunnels, thus providing secure connections over public links. While IPsec VPN is the dominant technology for securing site-to-site connections, testing will allow commonly accepted VPN configurations. NSS recommends using AES-256 encryption for the VPN tunnels.

This test determines whether the SD-WAN devices are able to dynamically establish and route traffic across multiple endpoints using VPN tunnels. Passing this test is a basic requirement for all SD-WAN devices.

4.2 Baseline Policy

A baseline policy is a routed configuration with an "allow all" policy. The SD-WAN is expected to pass traffic between all sites over VPN without incident.

4.3 Simple Policies

Simple outbound and inbound policies allow basic browsing and email access so internal clients can access untrusted, external networks without giving external clients the ability to access internal network(s). A number of combinations will be run to assess if an SD-WAN behaves according to configuration.

The SD-WAN should provide the ability to control which applications and protocols are passed based on test case criteria. A combination of policies from the following lists will be tested to simulate enterprise application prioritization:

Latency-sensitive applications and protocols (directed across service-assured link):

- VoIP
- H.323
- RTP
- RTCP
- RTSP

Latency-tolerant applications and protocols (directed across public IP link)

- HTTP
- FTP
- SCP
- IMAP
- SNMPV2
- SFTP
- POP3
- NetBIOS
- Telnet
- SMB
- NTP
- RADIUS
- LDAP
- SYSLOG
- TACACS+
- RDP
- SSH
- TFTP
- SMTP
- DNS
- Social media applications

5 WAN Impairment

A critical function of any SD-WAN is the identification and routing of traffic based on policy prioritization (autonomous or configured), which is influenced by network performance characteristics (e.g., packet loss, variability, latency, jitter, etc.). Link impairment tests subject links to a representative set of real-world conditions encountered by enterprises today. Latency, jitter, and variability are all commonly encountered on public link technologies. Products are expected to route prioritized traffic according to the quality settings for an application based on link performance. If impairment is encountered over a link where priority traffic is being routed, the SD-WAN must either throttle lower-priority traffic to assure the quality of the prioritized application, or it must reroute traffic across an alternate link if it is available, depending on network cost and expected service levels.

The various impairments described in this section are applied in order to assess the adaptability of the SD-WAN (i.e., path selection, QoS, failover, app steering, congestion avoidance). These tests go beyond packet loss, jitter, and latency to investigate misordered packets, link saturation, packet duplication, congestion, and bursty traffic. The goal is to verify how the product adapts to varied impairments.

In each test case, background traffic will be introduced to populate links with sufficient activity as to represent typical enterprise network communications. Additionally, traffic-specific flows will be introduced in order to capture accurate measurements, including RTP MOS for voice-over IP, relative MOS for video, and one-way delay for RTP. These measurements provide guidance as to how sensitive applications behave across a tested SD-WAN when the product is subjected to various impairments.

Different permutations of the following link impairments will be used to verify all SD-WAN performance features. Additionally, the impairment order and severity will be variable, representing the real-world experience of line quality, service degradation, and recovery.

5.1 Quality of Experience (QoE)

While throughput is important in SD-WAN, so is the user's QoE. A critical function of any SD-WAN is the identification and correct routing of traffic based on policy prioritization (autonomous or configured), which is influenced by network performance characteristics (e.g., congestion, packet loss, latency, packet delay variation, etc.). Link impairment tests subject connected links to testing that represents real-world conditions encountered by enterprises today. Congestion, packet loss, latency, and packet delay variation are all commonly encountered on public links.

5.1.1 Mean Opinion Score (Video & Voice)

Relative (video) MOS is an estimated perceptual quality score that considers the effects of codec; the impact of IP impairments (such as packet loss) on the group of pictures (GoP) structure and video content; and the effectiveness of loss concealment methods. The encoding specifications for video codec are used as guidelines and conformance, and vendors are free to design encoders to improve video quality and reduce the number of transmission bits. Simply put, MOS for video (relative MOS) can vary based on different advancements in the video estimation or encoding techniques. The maximum achievable MOS score for the video that will be used in the test is 4.53. VoIP (real-time protocol [RTP]) MOS, on the other hand, measures the MOS for VoIP calls based on the speech codec being used. The setup will use a G711 codec, which produces a maximum achievable MOS of 4.41 for an excellent VoIP call. Any MOS below 3.5 represents a significantly degraded voice call or video stream. NSS considers a MOS below 3.4 as failing to meet the use case.

5.1.2 Packet Loss (Video & Voice)

Packet loss for both Voice and Video plays a significant role in affecting MOS. Minimal or zero packet loss signals a reliable and effective solution for real-time communication.

5.1.3 Out-of-Order Packets (Video & Voice)

Misordered packets received due to improper application of error prevention techniques can impact MOS. Minimal or zero out-of-order packets signals proper sequencing and processing of received packets.

5.1.4 Duplicate Packets (Video & Voice)

Incorrect duplication techniques cause packets to be duplicated unnecessarily, which impacts bandwidth utilization rather than providing redundancy/reliability. Minimal or zero duplicate packets indicates effective processing of received packets.

5.1.5 RTP One-Way Delay (Voice)

The RTP one-way delay is used to assess the perceived quality of the link for Voice applications. High value indicates a possible lag in real-time applications.

5.1.6 RTP Jitter (Voice)

Variation in packet delay indicates the frequency of the received packets. High value indicates serious issues for real-time applications.

5.1.7 Connect Time (HTTP)

Time taken to establish an HTTP connection between client and server via the TCP handshake. Low value means faster connectivity.

5.1.8 Time to First Byte and Time to Last Byte (HTTP)

Time to first byte is the time it takes for a client to receive the first byte of a response to a request that it sends. Time to last byte is the time it takes for the client to receive all content in response. Low value means faster response.

5.1.9 Connection Latency (FTP)

Time taken for data to travel between source and destination. Low value means faster response.

5.1.10 Downloads Requested per Second & Downloads Successful per Second (FTP)

These values indicate the success rate of requests and responses for file downloads between the client and server. Close to equal values for both parameters indicate a very high success rate.

5.1.11 Mail Sent per Second (SMTP)

Indicates the total number of emails sent out per second. Higher value indicates better performance.

5.1.12 Sessions Requested per Second & Sessions Established per Second (SMTP)

These values indicate the success rate of requests and responses for session establishment between the client and server. Close to equal values for both parameters indicate a very high success rate.

5.2 Dynamic Path Selection with SLA Measurements

The goal of this test is to determine how long it takes for traffic to move to an available link when preconfigured impairments are applied. To limit any visible user impact, an SD-WAN should support path decisions on a per-flow basis according to available links and according to the conditions that exist on those links.

5.2.1 Packet Loss

This refers to the amount of data packets that do not reach their destination. The test will simulate various loss levels with Poisson and Gaussian distribution.

Any sustained packet loss should be identified by the SD-WAN and the links should be managed accordingly based on application or policy.

5.2.2 Packet Delay Variation (PDV)

This refers to the variation in delay of unidirectional, consecutive packets that flow between two hosts over an IP path. This impairment is most often referred to as jitter. The test will simulate a Gaussian and Internet delay with minimum and maximum values.

This test measures how an SD-WAN handles PDV impact on voice or video, both of which are delay-intolerant beyond the buffer capacity of the application.

5.3 Path Conditioning

SD-WAN technologies employ various techniques to condition WAN links in order to ensure reliability of data transmission. Some employ packet duplication, forward error correction, bonding, or load balancing.

The SD-WAN should identify the best path and guarantee priority policies (application, protocol, or other configured guidance) over known good links with other traffic transmitted as best effort.

5.3.1 Packet Reorder

This refers to the delivery of data packets out of the order in which they were originally sent. This test will simulate a Poisson and Gaussian distribution of selected packets.

Products should identify out-of-sequence packets and manage these according to the configured policy. This condition impacts voice and video applications significantly if the delay time exceeds the application buffer.

5.3.2 Packet Duplication

This refers to a packet that is duplicated on the network and is received twice at the receiving host. This test will simulate a duplication of selected packets in a Poisson and Gaussian distribution.

Products should take the next-in-sequence packet and drop the duplicates in order to preserve the whole frame sequence.

5.4 Link Saturation and Congestion

Global awareness of quality of service (QoS) can prevent congestion during the last mile of data delivery; thus, the goal of this test is to ensure reliable use of bandwidth by the controller in the SD-WAN.

5.4.1 Accumulate and Burst

This test refers to the accumulation of packets in a memory queue. Packets are burst once a configured condition is met. This test will simulate accumulation of packets until the buffer queue has (N) packets or until packets have been accumulated for a specified time (T) with a minimum interburst gap.

Burst capability testing stresses network buffering capacity. Sustained burst behaviors reveal that there is link congestion or other issues and SD-WANs should alter paths based on known good alternate paths.

5.4.2 "First-Mile" Network Behavior

The Policer limits the data rate of a network stream to ensure that it does not exceed the specified limits. This impairment emulates link saturation. Traffic policing is in accordance with the Metro Ethernet Forum (MEF) bandwidth profiles for Ethernet services. The SD-WAN is expected to utilize bandwidth appropriately.

To replicate congestion in the "first mile," impairments will be applied to the links from the HQ DC to the emulated aggregation point (ISP 1 and ISP 2).

5.4.3 "Last-Mile" Network Behavior

The Policer limits the data rate of a network stream to ensure that it does not exceed the specified limits. This impairment emulates link saturation. Traffic policing is in accordance with the MEF bandwidth profiles for Ethernet services It is expected that the product will utilize bandwidth appropriately

To replicate congestion in the "last mile," impairments will be applied to the links from the emulated aggregation point (ISP 1 and ISP 2) to the branch SD WAN sites.

5.5 Quality of Service (QoS)

Quality of service is important for business-critical applications such as voice and video. These applications must be prioritized if a link has bad performance indicators. This test measures QoS using voice traffic and video stream. The test will include MOS scores for video and call measurements for VoIP.

5.5.1 All Impairments

This test applies all of the above impairments. The SD-WAN should manage traffic according to configured QoS classification settings.

5.6 Application-Aware Traffic Steering

This test will assess how the product directs various application traffic flows for applications besides video and VoIP. Behavior will be observed and recorded to establish whether voice/video and data are sent over the same link once impairments are applied and to establish which application takes precedence.

5.6.1 Application Control Policies

These complex outbound and inbound policies consist of many rules, objects, and applications that verify whether the SD-WAN is capable of accurately determining the correct application (regardless of port/protocol used), and then taking the appropriate action.

- VoIP
- Business video (Cisco Spark, Microsoft Skype Professional, etc.)
- Popular social networking websites (web applications)
- Other basic legacy applications (e.g., FTP, Telnet)

For each application, NSS will test a product's ability to perform the following functions:

5.6.1.1 Steer

The SD-WAN should be able to accurately identify the application and direct it over the correct link according to configured policy.

5.6.1.2 Block Specific Action (Depends on Application)

For example, in the case of instant messaging, the product should allow text communications while blocking file transfers.

5.6.1.3 Drop Low-Priority Application During Congestion Event

The product should recognize when link exhaustion occurs and ensure that high-priority applications take precedence over low-priority applications.

5.7 High Availability

In a redundant network, the SD-WAN deployed in HA mode is expected to provide uninterrupted network service during any system failures, while delivering an acceptable user experience.

5.7.1 Hardware Power Fail (Master-Slave Node Negotiation)

In this test, the HQ-Master device will be subjected to a device power loss event. It is expected that the HQ-Slave device will become active and ensure that communication persists with minimal or zero loss.

5.7.1.1 Persistence of Data

The product should retain all configuration data, policy data, and locally logged data once it has been restored to operation following power failure.

5.7.2 CMS offline or Orchestrator Outage

In this test, the CMS or orchestration service will be disconnected or unreachable. During such an event, enterprises would expect that this would not impact established links. Two random impairment tests will be run with MOS scoring to validate that the performance matches that of earlier tests.

5.7.3 WAN Link Failure

In this test, an established WAN link between sites is interrupted at the HQ DC and the SD-WAN is observed to determine whether it is handling stateful session in a manner that is transparent to users. At the point of failure, the routed link traffic should be redirected without loss or interruption to the applications using the available links based on prioritization schema. The only exception to this would be where the failover links are experiencing an exhaustion event and prioritized applications are consuming all available bandwidth based on policy configuration, which could impact the non-critical applications.

An SD-WAN must be able to operate resiliently in spite of link outages. As the total traffic sent through the links will be less than the available bandwidth, it is expected that the sessions or applications should continue without interruption, and there should be no noticeable user impact.

6 Performance

This section measures the performance of a SD-WAN using various traffic conditions that provide metrics for real-world performance. Individual implementations will vary based on usage; however, these quantitative metrics provide a gauge as to whether a particular product is appropriate for a given environment and present a normative data set that is equal and comparable across all solutions.

All tests will be performed across the VPN links established according to the use case topology. Additionally, the harness baseline validation that is conducted prior to the introduction of a product will be documented in the test report.

The impairment test cases selected are the most stressful scenarios in which a WAN technology would ever be placed. The results of these tests and the application measures captured during each test case will indicate a product's ability to withstand punishing performance scenarios. In addition to these impairment scenarios, NSS will be recording standard traffic performance, the results of which will be included in test reports and scorecards.

6.1 Maximum Capacity

The use of traffic generation appliances allows NSS engineers to create "real-world" traffic at link-appropriate speeds as a background load for the tests.

The goal of these tests is to stress the policy or inspection engine and determine how it handles high volumes of TCP connections per second, application layer transactions per second, and concurrent open connections. All packets contain valid payload and address data, and these tests provide an excellent representation of a live network at various connection/transaction rates.

Note that in all tests, the following critical "breaking points"—where the final measurements are taken—are used:

- **Excessive concurrent TCP connections** – Latency within the SD-WAN is causing an unacceptable increase in open connections.
- **Excessive concurrent HTTP connections** – Latency within the SD-WAN is causing excessive delays and increased response time.
- **Unsuccessful HTTP transactions –** Normally, there should be zero unsuccessful transactions. Once these appear, it is an indication that excessive latency within the SD-WAN is causing connections to time out.

6.1.1 Theoretical Maximum Concurrent TCP Connections

This test is designed to determine the maximum concurrent TCP connections of the SD-WAN with no data passing across the connections. This type of traffic would not typically be found on a normal network, but it provides the means to determine the maximum possible concurrent connections figure.

An increasing number of Layer 4 TCP sessions are opened through the SD-WAN. Each session is opened normally and then held open for the duration of the test as additional sessions are added up to the maximum possible. Load is increased until no more connections can be established, and this number is recorded.

6.1.2 Maximum TCP Connections per Second

This test is designed to determine the maximum TCP connection rate of the SD-WAN with one byte of data passing across the connections. This type of traffic would not typically be found on a normal network, but it provides the means to determine the maximum possible TCP connection rate.

An increasing number of new sessions are established through the SD-WAN and ramped slowly to determine the exact point of failure. Each session is opened normally, one byte of data is passed to the host, and then the session is closed immediately. Load is increased until one or more of the breaking points defined earlier is reached.

6.1.3 Maximum HTTP Connections per Second

This test is designed to determine the maximum TCP connection rate of the SD-WAN with a one-byte HTTP response size. The response size defines the number of bytes contained in the body, excluding any bytes associated with the HTTP header. A one-byte response size is designed to provide a theoretical maximum HTTP connections per second rate.

Client and server are using HTTP 1.0 without keep-alive; the client will open a TCP connection, send one HTTP request, and close the connection. This ensures that all TCP connections are closed immediately upon the request being satisfied; and thus any concurrent TCP connections will be caused purely as a result of latency the SD-WAN introduces on the network. Load is increased until one or more of the breaking points defined earlier is reached.

6.1.4 Maximum HTTP Transactions per Second

This test is designed to determine the maximum HTTP transaction rate of the SD-WAN with a one-byte HTTP response size. The object size defines the number of bytes contained in the body, excluding any bytes associated with the HTTP header. A one-byte response size is designed to provide a theoretical maximum connections per second rate.

Client and server are using HTTP 1.1 with persistence, and the client will open a TCP connection, send 10 HTTP requests, and close the connection. This ensures that TCP connections remain open until all 10 HTTP transactions are complete, thus eliminating the maximum connection per second rate as a bottleneck (one TCP connection = 10 HTTP transactions). Load is increased until one or more of the breaking points defined earlier is reached.

6.2 HTTP Capacity

The aim of these tests is to stress the HTTP engine and determine how the solution 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 SD-WAN is forced to track valid TCP sessions, thus ensuring a higher workload than for simple packet-based background traffic. This provides a test environment that is as close to real-world conditions as it is possible to achieve in a lab environment, while ensuring absolute accuracy and repeatability.

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.

6.2.1 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.

6.2.2 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.

6.2.3 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.

6.2.4 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.

6.2.5 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.

6.3 Application Average Response Time: HTTP

Test traffic is passed across the infrastructure switches and through all inline port pairs of the SD-WAN simultaneously (the latency of the basic infrastructure is known and is constant throughout the tests). The results are recorded at each HTTP response size (44 KB, 21 KB, 10 KB, 4.5 KB, and 1.7 KB) at a load level of 95% of the maximum throughput with zero packet loss as previously determined in raw throughput testing.

6.4 Raw Packet Processing Performance (UDP Throughput & Latency)

This test uses UDP packets of varying sizes generated by traffic generation appliances. A constant stream of the appropriate packet size—with variable source and destination IP addresses transmitting from a fixed source port to a fixed destination port—is transmitted bi-directionally through each port pair of the SD-WAN. Each packet contains dummy data and is targeted at a valid port on a valid IP address on the target subnet. The percentage load and frames per second (fps) figures across each inline port pair are verified by network monitoring tools before each test begins. Multiple tests are run and averages are taken where necessary.

This traffic does not attempt to simulate any form of real-world network condition. No TCP sessions are created during this test, and there is very little for the flow or policy engine to do.

The goal of this test is to determine the raw packet processing capability of each inline port pair of the SD-WAN, as well as its effectiveness at forwarding packets quickly in order to provide the highest level of network performance and with the lowest latency. In addition, the latency and user response time will be recorded to determine the effect the device has on traffic passing through it under various load conditions. Test traffic is passed across the infrastructure switches and through all inline port pairs of the device simultaneously.

6.4.1 64-Byte Packets

Maximum 1,488,000 frames per second per Gigabit of traffic. This test determines the ability of a device to process packets from the wire under the most challenging packet processing conditions.

6.4.2 158-Byte Packets

Maximum 702,247 frames per second per Gigabit of traffic

6.4.3 256-Byte Packets

Maximum 452,000 frames per second per Gigabit of traffic

6.4.4 512-Byte Packets

Maximum 234,000 frames per second per Gigabit of traffic. This test provides a reasonable indication of the ability of a device to process packets from the wire on an "average" network.

6.4.5 1,035-Byte Packets

Maximum 118,483 frames per second per Gigabit of traffic. Some chipsets have difficulty with uncommon packet sizes. This test is designed to determine whether or not the SD-WAN handles uncommon packet sizes appropriately.

6.4.6 1,400-Byte Packets

Maximum 88,000 frames per second per Gigabit of traffic. This test has been included to demonstrate how easy it is to achieve good results using large packets. Readers should use caution when taking into consideration those test results that only quote performance figures using similar packet sizes.

6.5 "Real-World" Single Application Flows

Where previous tests provide a pure HTTP environment with varying connection rates and average packet sizes, the goal of this test is to simulate real-world single application traffic. Each protocol will be run to capacity and failure, at which point the maximum supported throughput per protocol will be recorded. These flows can be uni- or bi-directional to reflect both branch and HQ deployment options for the SD-WAN.

6.5.1 Single Application SIP Flow

6.5.2 Single Application SMTP Flow

6.5.3 Single Application FTP Flow

6.5.4 Single Application SMB Flow

6.5.5 Single Application RDP Flow

6.5.6 Single Application OneDrive, Dropbox

6.5.7 Single Application Office 365

6.5.8 Single Application Salesforce

7 Security Effectiveness

Required for devices that have built-in security. Cloud-delivered security not supported in this test.

7.1 False Positive Testing

The ability of the SD-WAN to identify and allow legitimate traffic while maintaining protection against threats and exploits is just as important as its ability to protect against malicious content. This test will include a varied sample of legitimate application traffic, which should be identified and allowed, or blocked, based on policy rules.

7.2 Intrusion Prevention

All products possessing protection capabilities must be tested with protection against network-delivered exploitation features enabled. These tests will accurately reflect the security capabilities of those products that possess native deep-packet inspection security capabilities. In order for an SD-WAN to be considered eligible for security effectiveness testing however, it must perform all of the test cases within the methodology with security enabled. If a product does not have security inspection capabilities, one of the modeled use cases will include a security option of the vendor's choosing (or NSS' designation in the event of no direction from the vendor.)

7.3 Evasions

Attackers can modify basic attacks to evade detection in a number of ways. If a product fails to detect a single form of evasion, any exploit can pass through the product, rendering it ineffective. NSS will verify that the product is capable of detecting and blocking basic exploits when subjected to varying common evasion techniques. Wherever possible, the product is expected to successfully decode the obfuscated traffic to provide an accurate alert relating to the original exploit, rather than alerting purely on anomalous traffic detected as a result of the evasion technique itself.

Evasions are considered highly sophisticated or advanced attacker capabilities; therefore, in order to perform well in this testing, an SD-WAN must already possess strong detection and prevention technologies prior to evasion techniques being applied.

For more details, please refer to the current NSS Labs Evasions Test Methodology.

7.4 Stability and Reliability

Long-term stability is particularly important for an inline device, where failure can produce network outages. These tests verify the stability of the SD-WAN along with its ability to maintain security effectiveness while under normal load and while passing malicious traffic. Products that are not able to sustain legitimate traffic (or that crash) while under hostile attack will not pass.

The product is required to remain operational and stable throughout these tests, and to block 100% of previously blocked traffic, raising an alert for each. If any policy-forbidden traffic passes, caused by either the volume of traffic or by the product failing open for any reason, this will result in a fail.

7.4.1 Blocking Under Extended Attack

The SD-WAN is exposed to a constant stream of policy or protocol violations over an extended period of time. The product is configured to block and alert, and thus this test provides an indication of the effectiveness of both the flow management and alert handling mechanisms.

A continuous stream of policy or protocol violations mixed with legitimate traffic is transmitted through the SD-WAN for eight hours at 10 Mbps, 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 performance with regard to policy handling.

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

7.4.2 Passing Legitimate Traffic under Extended Attack

This test is identical to the stability test run previously where the external interface of the product is exposed to a constant stream of policy or protocol violations over an extended period of time.

The product is expected to remain operational and stable throughout this test, and to pass most or all of the legitimate traffic. If an excessive amount of legitimate traffic is blocked throughout this test, caused by either the volume of traffic or by the solution failing for any reason, this will result in a fail.

7.4.3 Behavior of the State Engine Under Load

This test determines whether the product is capable of preserving state across a large number of open connections over an extended time period. At various points throughout the test (including after the maximum has been reached), it is confirmed that the product is still capable of inspecting and blocking traffic that is in violation of the currently applied network control policy, whilst confirming that legitimate traffic is not blocked (perhaps as a result of exhaustion of the resources allocated to state tables). The product must be able to apply policy decisions effectively based on inspected traffic at all load levels.

7.4.3.1 Passing Legitimate Traffic – Normal Load

This test ensures that the product continues to pass legitimate traffic as the number of open sessions reaches 75% of the maximum determined previously in performance testing.

7.4.3.2 State Preservation – Maximum Exceeded

This test determines whether the product maintains the state of pre-existing sessions as the number of open sessions exceeds the maximum determined previously in performance testing.

7.4.3.3 Drop Legitimate Traffic – Maximum Exceeded

This test ensures that the product continues to drop all traffic as the number of open sessions exceeds the maximum determined previously in performance testing.

Note: If a product allows traffic to “leak” due to the way it expires old connections, this will result in an automatic fail for the entire test.

8 Total Cost of Ownership and Value

Implementation of infrastructure and security solutions can be complex, with several factors affecting the overall cost of deployment, maintenance, and upkeep. All of the following should be considered over the course of the useful life of the SD-WAN:

- **Product Purchase** – The cost of acquisition
- **Operational Benefits –** The zero-touch provisioning concept for SD-WAN cites considerably reduced deployment requirements, specifically regarding configuration and tuning; for example, time to add a new site is measured in hours rather than days or weeks. These reduced configuration requirements contribute to operational savings for the enterprise.
- **Product Maintenance** – The fees paid to the vendor, including software and hardware support, maintenance, and other updates
- **Installation** – The time required to take a solution out of the box, configure it, deploy it into the network, apply updates and patches, and set up desired logging and reporting
- **Upkeep** – The time required to apply periodic updates and patches from vendors, including hardware, software, and other updates
- **ROI Assessment –** There are savings associated with moving from high-cost, service-assured links (e.g., MPLS) to commercial broadband. There is value both in aggregating multiple low-cost links to support demand as well as in the ease of deployment and recurring service cost reductions that are associated with moving from expensive, service-assured links to less expensive options.

The cost of the SD-WAN will be modeled based on the multiple use case scenarios depicted in the topology (Figure 1). One enterprise use-case model will be based on consumption of the entire SD-WAN value proposition; a second will be based on the enterprise expectation that all sites must support high-quality video and voice requirements in addition to basic data; and a third will be based on modest data-only requirement links.

9 Appendix A: Change Log

Version 3.0 – November 2019

- Section 1.1: Introduction updated
- Section 1.2: Updated methodology scope bullets; clarified tuning procedures
- Moved Section 1.3 Quality of Experience to Section 5.1
- Updated Figure 1 - Topology of Multiple Use-Case Test Environment: SD-WAN 3.0
- Section 1.3.2: Added additional LAN port at HQ DC site, i.e., two LAN ports at HQ DC
- Added Section 3: Management
- Section 4 (Network Policy Effectiveness) renamed to Routing & Access Control
- Section 4 subsections reordered
- Updated Section 4.2 Baseline Policy
- Updated Section 5 WAN Impairment
- Section 5.10.1: Updated test details for HA power fail
- Section 5.10.3: Updated test details for WAN link failover
- Added impairment measurement parameters in Sections 5.1.1 through 5.1.12
- Section 5: Moved WAN Impairment out of Performance section
- Section 6.4.5: Added the reason to test an uncommon packet size
- Section 7 updated with requirement for Security Effectiveness
- Sections 7.1 and 7.2 reordered
- Sections 7.3: Updated hyperlink for Evasions Test Methodology
- Section 7.4 heading edited
- Sections 7.4.1 - 7.4.3 heading levels edited

Version 2.95 – September 2019

- Section 1.5: Updated topology with multi-use case and added Minimum Requirements to Inclusion Criteria
- Section 3.3: Added details about ZTP measurement procedure
- Section 3.4.1: Removed Branch Off-ramp / Split-tunnel
- Section 3.4: Added Centralized Management System (CMS) section
- Section 4.1: Reorganized the WAN Impairment section based on various SD-WAN features
- Section 4.4.2: Changed 124-Byte Packets to 158-Byte Packets
- Section 4.4.5: Changed 1,024-Byte Packets to 1,035-Byte Packets
- Section 4.4.6: Changed 1,514-Byte Packets to 1,400-Byte Packets
- Section 4.6: Removed HTTP persistent cases
- Section 5: Removed (optional) from Security
- Section 6: Clarified TCO will be modeled based on multi-use case

Version 1.5 Proposal – September 2018

- Section 1.3: Added Branch Office “off-ramp” to topology reflecting additional test case
- Section 3.4: Moved VPN connection up from performance testing to configuration steps in order to reflect all tests are conducted via secure, established VPNs
- Section 3.4.1: Added Branch offramp / split tunnel
- Section 4.1: Introduced impairment variability and magnitude
- Section 4.1: Moved WAN link failover test case to 4.3 HA

- Section 4.3: added test cases for High Availability scenarios
- Section 4.9: Updated single app flows to reflect new enterprise apps. Added bi-direction capability of flows
- Section 5.1: Clarified that products submitted for security testing are tested with all features enabled
- Section 5.3: Evasions methodology reference updated to newer version
- Section 5.7.3: moved persistence of data requirement up to HA section 4.3
- Section 6: Clarified sites modeled for TCO will be increased in next test

Version 1.2 – 6 April 2018

- Section 1.3: Revised topology to reflect logical configuration
- Section 4.1:
    - Added saturation test case
    - Clarified test cases
    - Added section 4.1.6: First Mile Network Behavior
    - Added section 4.1.7: Last Mile Network Behavior

Version 1.1 – 1 February 2018

- Section 4 (Network Policy & Performance) moved prior to section 3 (optional security section) in order to emphasize performance nature of testing.
- Added Section 3.1.3: Remote Configuration. Clarification on what will be assessed and measured.

Contact Information

NSS Labs, Inc.
3711 South MoPac Expressway
Building 1, Suite 400
Austin, TX 78746-8022
USA
info@nsslabs.com
www.nsslabs.com

This and other related documents available at: **www.nsslabs.com**. To receive a licensed copy or report misuse, please contact NSS Labs.