



NSS Labs

Independent Security Testing & Certification

AVOIDING PITFALLS IN EFFECTIVE NETWORK PERFORMANCE MEASUREMENT

By Christian Stankevitz, CTO

SUMMARY

Effective performance of network products must be measured in the context for which the product will be implemented. Typical packet blasting and simple capture/replay technologies are insufficient for realistic determination of network performance for sophisticated state-based network products.

The history of enterprise and carrier network implementations has vividly demonstrated that network products perform significantly different depending on product configurations and network traffic conditions. Unfortunately, it is not uncommon to see product manufacturers' data sheets and even test lab reports that cite performance levels measured under ideal network conditions with little or no product configuration. These misleading performance measurements can then impact production networks when the implemented product fails to perform as advertised. No network administrator wants to be in the precarious position of discussing why "we just spent how much to find that it can't handle our network?" Thus, obtaining realistic independently validated performance metrics for network products before purchase has become increasingly important.

Measuring the performance of modern, stateful network products through simple network packet blasting (i.e. RFC 2544) is like measuring the speed of an M1 Abrams Tank on a smooth paved road. Measurements can be taken quite accurately but if military tanks were designed to run only on paved roads, they would have been equipped with wheels instead of tracks.

Similarly, if modern deep inspection firewalls, intrusion prevention systems, policy based routers, and other content-aware network products were designed to only retransmit empty network packets from one interface to another, the simple use of packet blasting and replay-based test tools would be sufficient. However, these sophisticated products are designed to track application layer dynamics such as unique key exchange, individual client session establishment, and rate-based anomalies to indicate possible denial of service attacks.

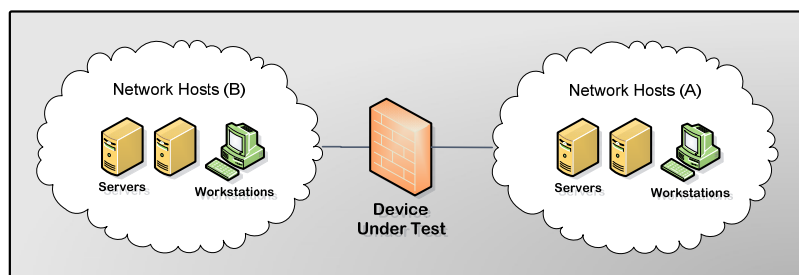


Figure 1 – Typical DUT configuration in a lab environment.

In many cases, the default functionality of the device under test (DUT) causes the test equipment to 'break' leading lab personnel to 'tweak' the DUT until the capture/replay tools can work. This type of scenario, of



course, leads to results that reflect neither real-world deployment configurations nor real-world traffic conditions.

Performance testing of network systems must be scientifically measured using product configurations and traffic conditions that are indicative of real-world deployments. Products that are designed to track TCP session state must be tested with realistic TCP traffic from IP addressing topographies that closely resemble the connection dynamics of the target implementations. Static UDP traffic flows from fixed quantities of IP addresses provide insufficient connection dynamics that directly impact the effective performance of modern stateful systems. Stateful benchmarks must include proper network dynamics such as TCP session establishment, windowing, retransmissions, realistic IP address lifetimes, and randomization.

To address complex multiprotocol network dynamics some manufacturers and test environments have resorted to capture/replay technology in which network captures from production IT implementations are replayed at high speed to emulate larger, more bandwidth intensive networks. Using accelerated replays of network capture files to emulate larger more complex bandwidth intensive networks is like audio recording musicians in a quartet and playing it back with multiple MP3 players to recreate the dynamics of a full philharmonic orchestra. The resulting “music” not only sounds bad but also fails to recreate the fundamental complexity of overtones and rhythm of the full orchestra. Similarly, network capture/replay technology fails to generate traffic with proper TCP connection dynamics and application layer state management. The replay of common protocols, such as HTTP and Microsoft SMB, fails to account for basic content dynamics such as HTML cookies, individual client authentication, and dynamic port establishment. Effective performance measurement of complex network products, such as modern intrusion prevention systems, deep inspection firewalls, application gateways / proxies, and anti-spam systems, requires valid application layer state and content dynamics. Replaying the same authentication tokens, client cookies, and transaction parameters across multiple IP addresses fundamentally breaks application layer state resulting in invalid traffic.

“Unrealistic traffic generation can skew performance numbers by 20% to 50%, depending on the content.”

Effective network performance benchmarks must be measured in the context of meaningful traffic conditions that address the complexity of application layer state management. Published performance levels that do not address application state management result in wide gaps between manufacturer stated specifications and effective performance. After measuring 100’s of products, NSS Labs has commonly found gaps ranging from 20% to 50% of the stated performance levels.



CONCLUSION

Meaningful performance specifications must disclose not only the conditions of the benchmark but also the applicability of the benchmark traffic to the network product. Stateless packet blasting and replay traffic is meaningless for benchmarking products that properly track network and application layer state. Modern network products require new benchmarking methodologies that reflect the application layer content complexity and dynamics of real world implementations.

REFERENCES

RFC 2544 - Benchmarking Methodology for Network Interconnect Devices

Tomahawk - <http://tomahawk.sourceforge.net/>

Flowreplay - <http://tcpreplay.synfin.net/trac/wiki/Documentation>



ABOUT THE AUTHOR

Christian Stankevitz (cstankevitz@nsslabs.com) leads the NSS Labs testing group as the company's CTO. A Certified Six Sigma Black Belt, Mr. Stankevitz has over 13 years of telecommunications testing experience. He developed the first latency generator used for application impact analysis across variable latency telecommunications networks. He has been a supercomputer researcher, a telecommunications architect for global networks, and his network designs and methodologies have been implemented by several F500 enterprises, carriers, and leading manufacturers for the past decade.

NSS LABS

Founded in 1991, NSS Labs is the globally recognized leader in independent security performance testing and certification. With operations in Chicago, IL and San Diego, CA, NSS Labs offers a range of specialized networking and security testing services to vendors and end-user organizations world-wide. NSS Labs' comprehensive certification schemes address a wide range of security products, including Intrusion Prevention Systems (IPS), Intrusion Detection Systems (IDS), Secure Content Appliances (SCA), Content Monitoring and Filtering (CMF) devices, Unified Threat Management (UTM) systems, firewalls, VPNs, Web Application firewalls, vulnerability scanning, cryptographic devices and PKI products. More info: www.nsslabs.com.