

Firewall Performance Testing Methodology

Philip Joung

with Anil Pochiraju

Application Note 03

July, 2003



Spirent Communications

26750 Agoura Road
Calabasas Hills, CA
91302 USA
e: enterprise@spirentcom.com

Sales Contacts:

North America

+1 800-927-2660

Europe, Middle East, Africa

+33-1-6137-2250

Asia Pacific

+852-2166-8382

All Other Regions

+1 818-676-2683

www.spirentcom.com/enterprise

Copyright 2003 by Spirent Communications, Inc. All rights reserved.

Table of Contents

1 Introduction.....	1
2 Why Test Your Firewall?	1
3 Methodology	2
3.1 Firewall Performance Testing Methodology	2
3.1.1 Single Protocol Tests	3
3.1.1.1 Category 1: Single Protocol Baseline Performance Tests.....	3
3.1.1.2 Category 2: Single Protocol with Inline DDoS Performance Tests.....	4
3.1.1.2.1 Protocol Testing with Inline DDoS Attacks	5
3.1.2 Mixed Protocol Tests.....	6
3.1.2.1 Mixed Protocol Baseline Testing	7
3.1.2.2 Mixed Protocol with Network Realism Performance Testing	7
3.1.2.3 Mixed Protocol with Inline DDoS Performance Testing.....	8
3.2 Other Tests	9
3.2.1 DDoS Attacks	10
3.2.2 More tests.....	10
4 Appendix	11
4.1 Things to Look for During Firewall Failures	11
5 Acknowledgements.....	11
6 References.....	12

1 Introduction

Most network deployments now use firewalls as a critical component, helping to increase the security of the networks that they protect. Indeed, firewalls have become nearly ubiquitous, especially with the recent prominence given to network security by the industry and press. To a network, firewalls look like a large fence, with only a few places left open on the fence to allow traffic to pass. As with any other part of the network, firewalls benefit from expert planning, deployment and maintenance—a “set-it-and-forget-it” style of deployment will not work for firewalls, nor will a network stay secure very long if firewalls comprise the sole protection strategy for the network.

Introducing a firewall into a network all starts with proper planning. One will need to consider things such as the protocols that the firewall will support, the applications that need access from either side of the firewall, the criticality of each application, the budget, and the importance of CRASP (compatibility, reliability, availability, scalability, performance) for the network and the firewall. The best way to ensure a high level of CRASP for the network starts with testing the candidate firewalls using the major protocols that the firewall will support in the live environment. This document describes suitable methods and techniques that can help determine the performance and reliability of the firewall.

2 Why Test Your Firewall?

Why spend the time and effort to test your firewall? For the networks that a firewall secures, often the value of the data and systems that they must help protect cost many times that of the firewall and the testing. Deploying a security infrastructure without understanding its performance and security runs the risk of doing little to protect the network, especially if it also introduces apathy from a false sense of security. Among the reasons that show the criticality of testing firewalls:

- **Security at load:** Often, security flaws do not appear until the network encounters a large load. Attacks can hide more easily within large amounts of traffic, potentially causing problems right when network downtime is most harmful.
- **Firewall behavior at load and at failure:** Firewalls often exhibit different behaviors as they encounter increasing loads. Eventually, with enough traffic, the firewall will fail, providing valuable insights. First, the conditions and behavior leading up to failure are now known, giving the security administrator things to look for in production and providing a useful pre-failure warning. Second, the failure state of the firewall will be known—firewalls that fail closed will stop all traffic from passing (essentially a successful denial of service, or DoS, attack), and firewalls that fail open permit all traffic, which is a security failure.
- **Pre-deployment capacity planning:** Deployment of a security infrastructure will most likely affect overall network performance—testing the effect on network performance will ensure that the increased security does not decrease performance beyond the levels acceptable for the business.

In the end, security is not meant to be easy, it's meant to protect important assets, and there are no shortcuts.

3 Methodology

Please note that the methodology introduced here will produce a solid picture of the CRASP of the firewall and hence its suitability for the planned deployment environment. We will touch on the testing of the security abilities of a firewall, but bear in mind that this paper focuses on performance testing rather than security testing. I welcome any further suggestions for suitable ways that this methodology could be revised, improved and extended.

Some assumptions for this methodology:

- You have a good understanding of the network and applications you're trying to secure with the firewall.
- You understand how to work with and configure firewalls (but would like to know how to test them thoroughly and methodically).
- You have access to the firewall(s) before deployment into the live environment (testing will cause many firewalls to fail—an important finding in itself).
- The firewall will be configured with the settings expected in production before testing starts, including user authentication and logging enabled.
- You have the ability to generate traffic that realistically models the major protocols and applications on your network.
- Sufficient traffic can be generated to exceed the peak traffic levels expected by your network.
- You have the ability to simulate various realistic inline DDoS attacks.
- You have read and can refer to the companion paper: *TR7: General Network System Testing Methodology*. This paper contains basic system testing methodology, applicable to the performance assessment of nearly any network system or infrastructure.

3.1 Firewall Performance Testing Methodology

Firewalls must protect the network against a variety of attacks that come from different protocols. As a result, testing only one protocol will help understand the firewall's performance and capabilities with that protocol, but will not provide insight into the differences that might occur with other protocols. Therefore, thorough testing of a firewall starts with a look the major protocols that the firewall will protect, including ones that it may not immediately start protecting. This future planning ensures that the firewall undergoes tests before going into production, when testing performance often presents a greater challenge.

We propose two main categories of tests, each of which subdivide into more tests: single protocol tests and protocol mix tests. The single protocol tests will assess the compliance and performance of the firewall with each protocol. Taken as a whole, the single protocol tests provide a good picture of firewall performance and helps determine any particular protocol anomalies that often occur with firewalls. The protocol mix tests provide further details about firewall performance with traffic mixes that more closely resemble what exists in production use. As a reminder, remember that the following

methodology requires the companion paper TR7: *General Network System Testing Methodology*, as this paper will refer to TR7 for many of the tests.



Figure 1: The most basic network testbed for testing firewall performance, with a client traffic simulator on one side and a server simulator on the other side. Firewall testbeds can become quite complicated, especially for tests that include servers and clients on both sides of the firewall and servers in the DMZ.

In general, the testbed for firewall assessments will look similar to Figure 1. The testbed includes a layer 4-7 client traffic generator (such as Spirent Avalanche 2200) to create the simulated user traffic, a firewall in the middle, and either real servers or a server simulator (such as Spirent Reflector 2200) on the other end to properly respond to the incoming user requests. While connecting the test tools to the firewall, be sure that the traffic flows in the proper direction—the clients and servers should both connect to the firewall on the same side that they will appear in production. More complicated testbeds include clients and servers on both sides of the firewall and servers in the DMZ. Even in these more complicated testbeds, the basic concepts of testing still apply—create traffic that assesses the performance, capability and behavior of the system in order to help improve CRASP.

3.1.1 Single Protocol Tests

Within the single protocol tests are two major categories: single protocol baseline performance tests and single protocol with inline DDoS performance tests. Some DDoS attacks can be especially difficult to fend off, but even the less harmful DDoS attacks will most likely affect legitimate user traffic.

3.1.1.1 Category 1: Single Protocol Baseline Performance Tests

Figure 2 shows the steps to take to conduct a thorough assessment of the single protocol performance category of tests. Besides determining the performance of the firewall with each protocol, this battery of tests also determines the protocol compliance of the firewall, locating protocol anomalies that may be harder to locate and isolate under real-world traffic. Each of the tests within the flowchart (e.g. Maximum Connections Test) comes from a methodology described in TR7.

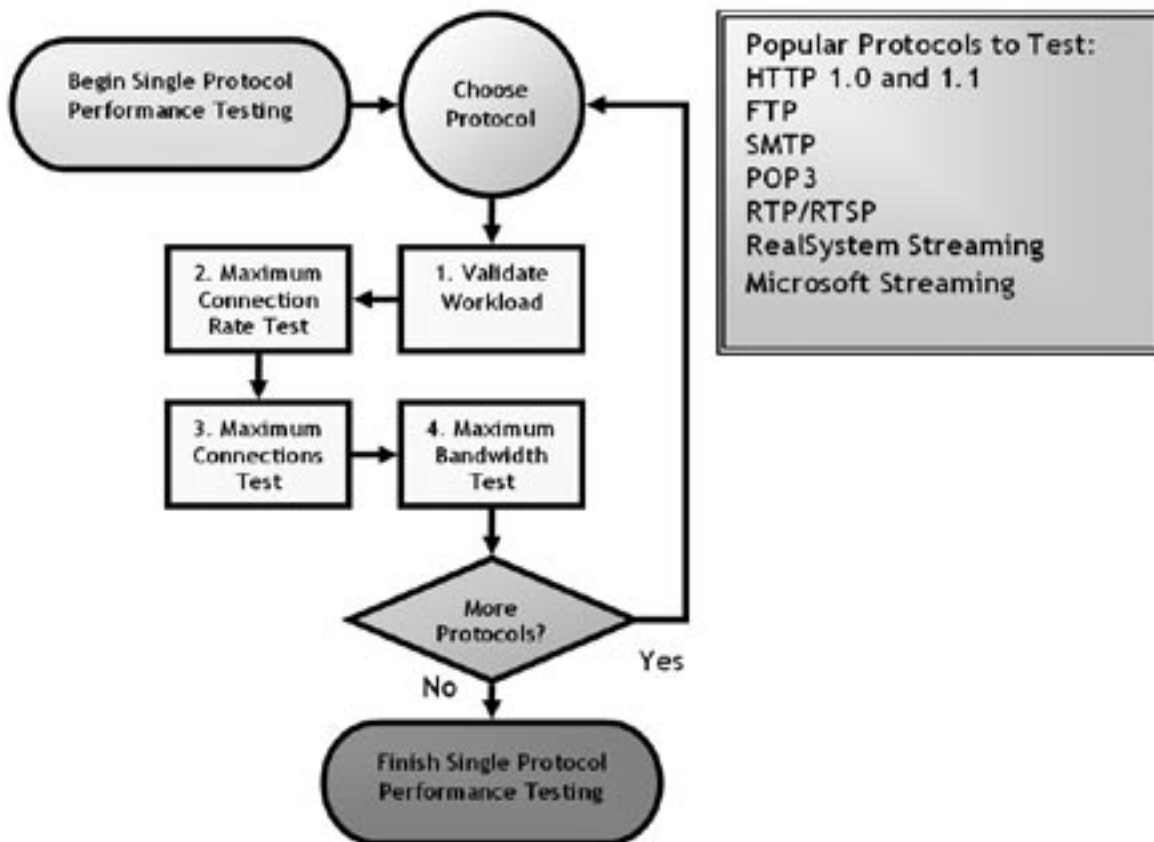


Figure 2: Testing a firewall properly requires assessment with a selection of different protocols. Because protocols have different behaviors and firewalls react to particular protocols in different ways, a full picture of firewall performance cannot rely on the information gained from just a single protocol.

Completing this first set of tests will yield a great deal of information—store the results files separately and then gather the most salient results into one place (such as a spreadsheet) in order to have a convenient location to refer to these baseline results during subsequent tests. Keep in mind that the results obtained from single protocol baseline testing will usually be the maximum performance possible for the firewall with each protocol—any other test result in this paper will likely have lower results.

3.1.1.2 Category 2: Single Protocol with Inline DDoS Performance Tests

A firewall must, of course, try to protect the network against attacks, including distributed denial of service attacks. These attacks can be quite difficult to protect against, so quantifying a DDoS attack’s effect on system performance will help in network capacity planning. The DDoS simulations can also validate the firewall’s ability to protect the network, allowing steps to be taken to improve security and performance if DDoS tests show poor results. The firewall behavior during DDoS attacks can also help during troubleshooting efforts in a live network under a real attack. This section contains methodology for testing a firewall using a single protocol in combination with an inline DDoS attack. The next section contains methodology for testing the firewall with DDoS attacks alone, which may serve as a useful baseline to compare with the results from this section.

3.1.1.2.1 Protocol Testing with Inline DDoS Attacks

Result	This category of tests will determine the firewall's performance and behavior with a particular protocol and various percentages of inline DDoS attacks.
Purpose	This test determines the performance and behavior of the firewall with various protocols while under a simultaneous attack from different DDoS attacks. Different DDoS attacks will affect the network and the firewall differently, providing useful feedback and information to improve the network's ability to respond and mitigate these attacks.
Test specification	Connections per second with inline DDoS attack.
Constraints	Be sure that the connections per second does not exceed 75% of the maximum connections per second determined in the previous section: Single Protocol Baseline Performance Tests.
Time needed	Specification: about 1 hour. Run time: about 4 hours.
Type of workload	Use the protocol workloads from the baseline tests (section 3.1.1.1) and add various inline DDoS attacks, including: floods (SYN, UDP, ICMP), LAND, Shaft, Smurf, Teardrop.
Methodology	<ol style="list-style-type: none"> 1. Choose the protocol(s) to test. Common protocols include HTTP, FTP, SMPT, POP3, and streaming media protocols. 2. Configure the test tool to ramp up this protocol to 75% of the maximum determined in the baseline tests. 3. Choose the DDoS attack to use during the test. 4. Based on the firewall's rated bandwidth capacity, configure tests to inject DDoS rates of 5%, 33%, 50% and 75% of this capacity as a start. These values may be adjusted based on results. 5. Run the test. 6. During the test, look to see how the firewall responds to this attack. If the legitimate traffic drops to zero, there is no need to run tests with higher amounts of DDoS traffic. 7. After the test, analyze the results. If steps can be taken to mitigate poor results (change settings, add patches, discuss with firewall vendor, etc), take steps to resolve the issue and retest.
What to look for	<p>Firewalls typically employ a few strategies for fending off DDoS attacks and allow legitimate network traffic to continue:</p> <p>Client connection limits: the firewall limits the number of connections from a particular client IP address, although this strategy may have little value during a DDoS attack because of the sheer number of IP addresses used.</p> <p>Server connection limits: the firewall limits the number of connections it forwards to a server to ensure that the server continues to operate.</p> <p>Capacity limits: the firewall limits the amount UDP or ICMP traffic, reserving a certain amount for TCP traffic.</p> <p>Pay attention to how the DDoS attack affects the allowed traffic. Ideally, the attack has minimal or zero effect on the allowed traffic.</p>

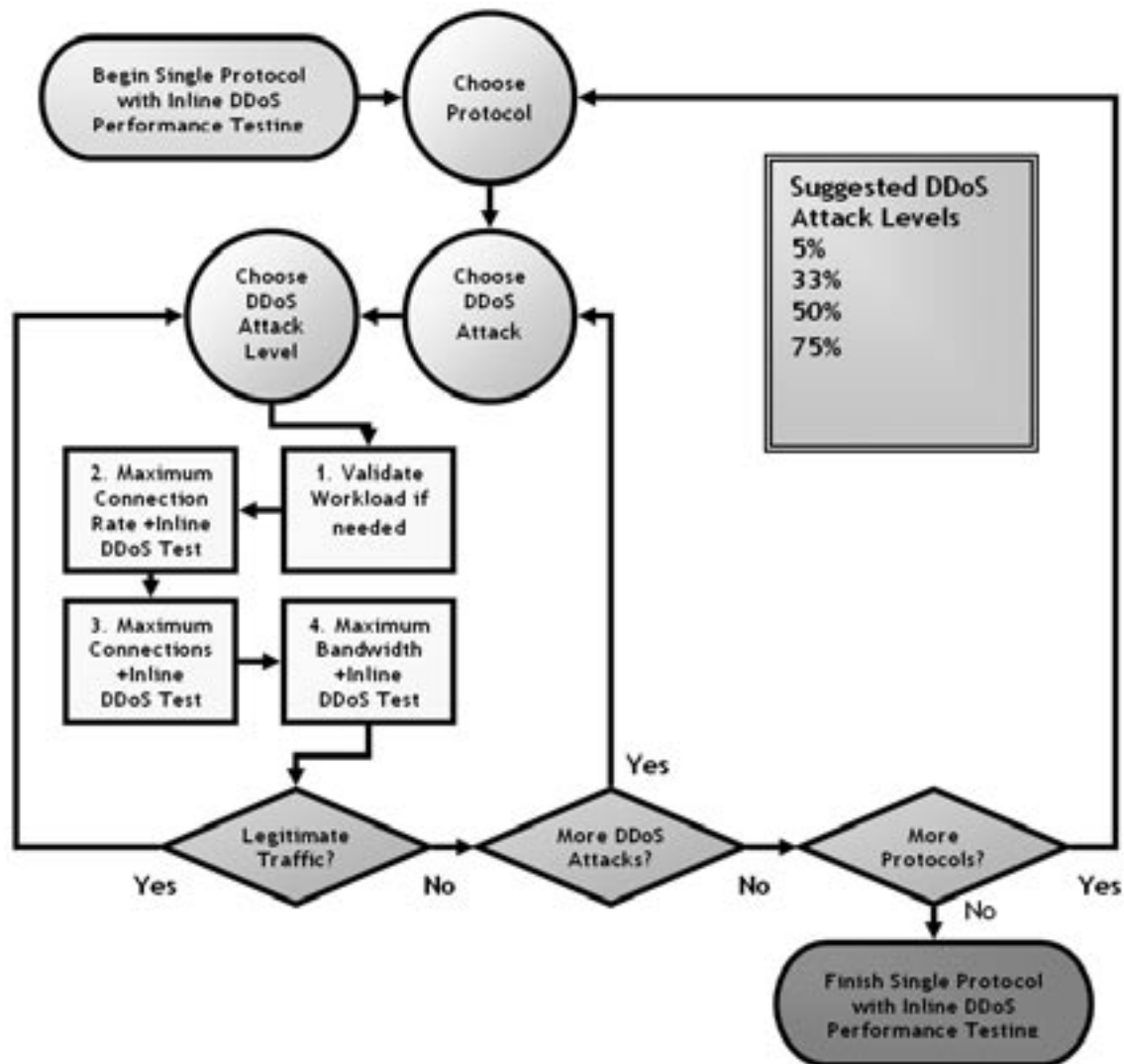


Figure 3: Comprehensively testing a firewall using single protocols and an inline DDoS attack requires several iterative steps. Because a firewall often exhibits different behaviors when encountering different protocols and DDoS attacks, obtaining behavior and results using each of these helps complete the picture of firewall performance and behavior. During testing, the DDoS percentage should be increased until no legitimate traffic passes.

3.1.2 Mixed Protocol Tests

In order to continue the characterization of the firewall, this next set of methodologies focus on using a mix of protocols. Ideally, create a mix that resembles the traffic that the firewall will eventually protect. For papers, reports and tools to help characterize the traffic on the Internet and various TCP/IP networks, see the Cooperative Association for Internet Data Analysis (CAIDA) Web site at <http://www.caida.org/analysis/workload/>.

3.1.2.1 Mixed Protocol Baseline Testing

Result	This test will determine the firewall's performance and behavior with a mix of protocols that resemble production usage.
Purpose	To protect a network, a firewall must properly handle many protocols. Each protocol will have certain behaviors that the firewall will need to properly manage and secure. Testing with a realistic mix of protocols will determine the firewall's performance and behavior under realistic conditions.
Test specification	Varies depending on goals of test, but usually connections per second or connections.
Constraints	Be sure that the connections per second does not exceed the lowest result from the previous section: Single Protocol Baseline Performance Tests.
Time needed	Specification: about 1 hour. Run time: about 2 hours.
Type of workload	Use a mix of protocols that resembles what the candidate firewall will protect. See http://www.caida.org/analysis/workload/ for tools to help characterize network traffic.
Methodology	<ol style="list-style-type: none"> 1. After determining the workload to use, validate the workload per the methodology outlined in TR7. 2. Find the lowest performance result from the single protocol baseline performance tests, which is denoted here as ProtBaseline. 3. Configure the tool to ramp up the traffic in step heights of (ProtBaseline/20) over 20 steps. The step steady time should be 240 seconds, and each step ramp time should be 20 seconds. 4. Run the test. 5. After the test, analyze the data to get a rough determination of the firewall performance, denoted here as MacroFWPerf. 6. Now, configure a new test with an initial ramp up to (MacroFWPerf-10%) over 240 seconds and hold steady for another 240 seconds. 7. Add 20 more steps of size MacroFWPerf/100 with a ramp time of 30 seconds and a steady time of 240 seconds. 8. Run the test. 9. During the test, use both the test tool's real-time reporting tool and the firewall's status reports to determine how the firewall is handling the traffic. 10. After the test, analyze the results to obtain the firewall's performance result.
What to look for	As firewalls reach their capacity, a change in behavior should occur. The appendix has a list of these changes in behavior.

3.1.2.2 Mixed Protocol with Network Realism Performance Testing

Network issues can cause significant problems for systems, often resulting in problems that the end-user can detect. Packet loss boils down to a loss of network data that must then be detected by the server and/or client so that data can be resent. This detection and resending can gobble up resources. With enough packet loss, data transmission can become impossible, resulting in timeouts. Link latencies also use resources by increasing the number of connections, requiring memory to maintain connections and processing to manage the connections.

Result	This test will determine the firewall's performance and behavior with a mix of protocols that resemble production usage and realistic amounts of packet loss and link latencies.
Purpose	Accurately determining system performance requires that the generated test traffic carefully replicate the conditions and behaviors found in real-world traffic. This includes packet loss and link speeds, both which can cause significant performance and stability problems.
Test specification	Varies depending on goals of test, but usually connections per second or connections.
Constraints	Be sure that the connections per second does not exceed the lowest result from the previous section: Single Protocol Baseline Performance Tests.
Time needed	Specification: about 1 hour. Run time: about 2 hours.
Type of workload	Use the same workload as the previous section, but add realistic amounts of packet loss and network link speeds. For information on packet loss, see http://average.matrixnetsystems.com/ . For network link speeds, use network speeds that your network is encountering or visit http://www.nielsen-netratings.com/ .
Methodology	<ol style="list-style-type: none"> 1. Adapt the workload from test 3.1.2.1 to incorporate network realism parameters such as packet loss and link speeds. 2. Use the performance result from test 3.1.2.1, denoted here as MPBaseline. 3. Configure the tool to ramp up the traffic in step heights of (MPBaseline/20) over 20 steps. The step steady time should be 240 seconds, and each step ramp time should be 20 seconds. 4. Run the test. 5. During the test, use both the test tool's real-time reporting tool and the firewall's status reports to determine how the firewall is handling the traffic. 6. After the test, analyze the results to obtain the firewall's performance result and behavior with network impairments.
What to look for	<p>As networks encounter more loss and delays, often the simultaneous connection count will increase, resulting in more resource utilization. A decrease in bandwidth and overall network utilization usually occurs as well. Users will encounter longer load times and timeouts.</p> <p>The appendix has a list of some changes in behavior to look out for. With packet loss, expect increased incidence of timeouts. With slow connection speeds, expect both the simultaneous connections and timeouts to increase.</p>

3.1.2.3 Mixed Protocol with Inline DDoS Performance Testing

Firewalls often have difficulty protecting against DDoS attacks while continuing to serve legitimate traffic. This suite of tests will help determine the performance and behavior of the firewall under various levels of DDoS attacks. The best firewalls will minimally or preferably have zero effect on legitimate traffic while under attack.

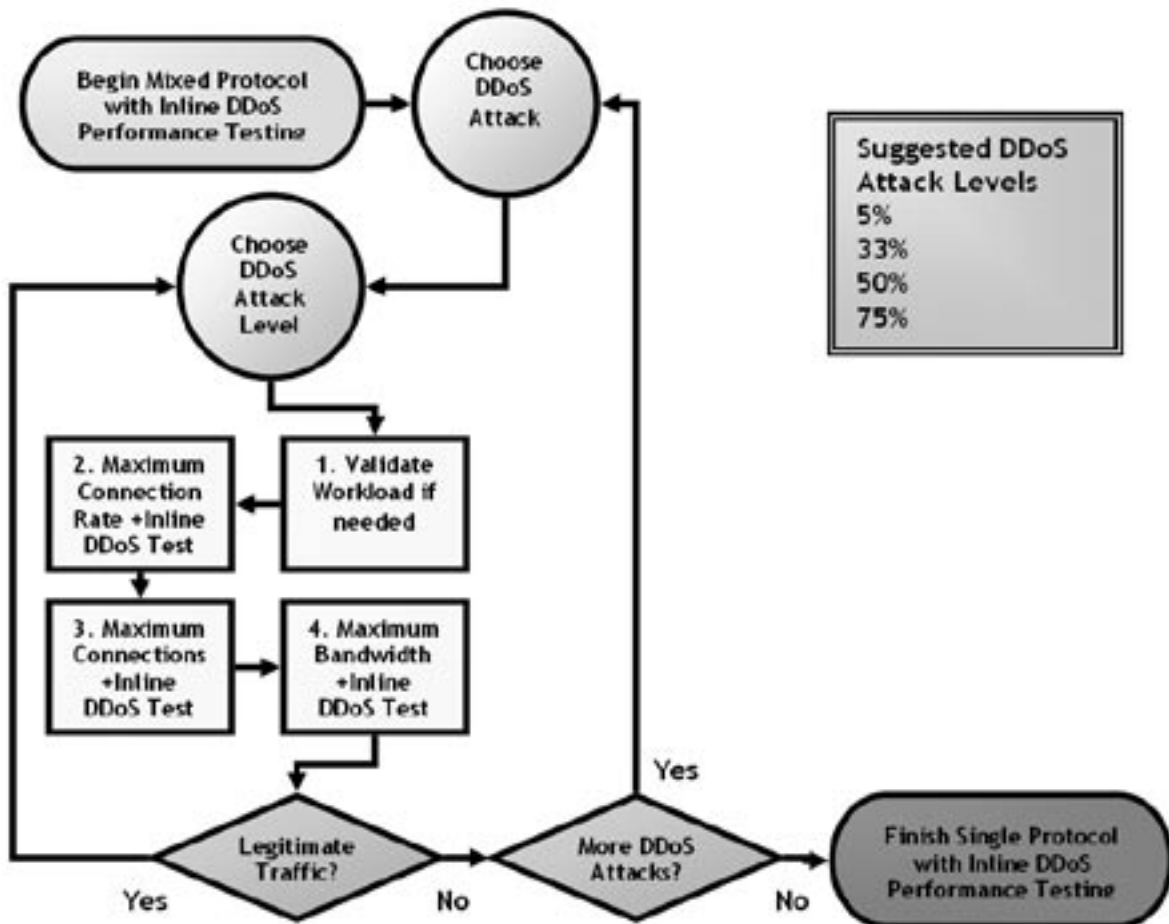


Figure 4: The steps needed to test a firewall with mixed protocols and inline DDoS attacks. Although several DDoS attack levels have been suggested, consider testing until the firewall can no longer forward legitimate traffic.

3.2 Other Tests

The above tests should help determine firewall performance and behavior under several useful scenarios. The information gathered during the tests will often be enough to apply and use in many other situations. However, many other tests exist, some of which are described and outlined below.

3.2.1 DDoS Attacks

Result	This category of tests will determine the firewall's performance and behavior with inline DDoS attacks.
Purpose	This test determines the performance and behavior of the firewall while under attack from different DDoS attacks. Different DDoS attacks will affect the network and the firewall differently, providing useful feedback and information to improve the network's ability to respond and mitigate these attacks.
Test specification	Inline DDoS attack.
Constraints	None.
Time needed	Specification: about 15 minutes. Run time: about 1 hour.
Type of workload	Various DDoS attacks, including: floods (SYN, UDP, ICMP), LAND, Shaft, Smurf, Teardrop.
Methodology	<ol style="list-style-type: none"> 1. Choose the DDoS attack to use during the test. 2. Based on the firewall's rated bandwidth capacity, configure tests with DDoS rates of 5%, 33%, 50%, 75% and 100% of this capacity. 3. Run the test. 4. During the test, look to see how the firewall responds to this attack. 5. If steps can be taken to mitigate poor results (change settings, add patches, discuss with firewall vendor, etc), take steps to resolve the issue and retest.
What to look for	Please note that not all firewalls will successfully fend off all DDoS attacks. However, the behavior of the firewall during an attack will help set expectations during an actual attack in a production network.

3.2.2 More tests

Many other tests exist:

- **Firewall Failure Mode Test:** when a firewall fails, two general scenarios will occur: fail-open or fail-close. Both scenarios are troublesome, so testing should determine which one will occur to minimize surprises. Testing this involves creating traffic well beyond the capacity of the firewall to see what happens when it fails. Among the traffic to consider are DDoS attacks.
- **Firewall Differential Latency Test:** This test helps determine the amount of latency that the firewall contributes to the overall test. Simply run two tests, the first one without a firewall and the second one with the firewall. Afterwards, compare the differences in response times.
- **VPN and Firewall Tests:** Firewalls often serve as VPN gateways to an internal company network along with their normal firewall duties. If so, the VPN functionality will consume firewall resources that can affect the other parts of the network. Testing this functionality would follow similar steps to the mixed protocol testing methodology above with VPN serving as one of the protocols. During this test, consider testing with NAT traversal turned on and turned off.

4 Appendix

4.1 Things to Look for During Firewall Failures

Understanding the performance thresholds of the candidate firewall helps in planning the overall capacity of the network. Overloaded firewalls exhibit a number of different errors in the application level and/or the TCP level. Application level errors may include incomplete transactions, application level timeouts, and HTTP errors for firewalls that happen to be application level proxies. TCP level errors include connection timeouts and TCP resets. When a firewall is overloaded, one or more of the following behaviors are usually observed:

- A failure to pass valid allowed traffic
- A leak of disallowed traffic
- Timeouts for all new incoming connections
- Resets of incoming connections and while continuing to process valid traffic
- Accepts and holds open incoming connections but refuses to forward traffic
- Accepts and holds open incoming user authentication but refuses to proxy

Often, these overloads and failures may be seen in the statistics:

- A topping out of open connections
- A drop in open connections
- Increase in time to TCP SYN/ACK
- Topping out of bandwidth
- Decrease in bandwidth
- An increase in page response time (especially to levels beyond what is acceptable by the business)

5 Acknowledgements

I would like to thank the following people for providing feedback and assistance during the creation of this document:

- Roy Chua
- Hemendra Godbole
- Alan Newman
- Andrew Foss
- John Kenney
- Sanjay Raja

6 References

- [CAIDA02] Cooperative Association for Internet Data Analysis, K. Claffy
Workload Characterization
<http://www.caida.org/analysis/workload/>, 2003
- [CERT01] Computer Emergency Response Team and Carnegie Mellon University, K. Houle, G. Weaver, N. Long and R. Thomas
Trends in Denial of Service Attack Technology
http://www.cert.org/archive/pdf/DoS_trends.pdf, October 2001
- [IETF03] The Internet Society, IETF, Spirent Communications, Network Test and GVNW Consulting, B. Hickman, D. Newman, S. Tadjudin and T. Martin
Benchmarking Methodology for Firewall Performance
Benchmark Working Group Internet Draft, January 2003
- [INTERHACK00] Interhack Research, M. Curtin and M. Ranum
Internet Firewalls: Frequently Asked Questions
<http://www.interhack.net/pubs/fwfaq/>, December 2000
- [MATRIX03] Matrix NetSystems,
Internet Average
<http://average.matrixnetsystems.com/>, 2003
- [NNR03] Nielsen//Netratings.
Nielsen//Netratings Internet Activity Reports
<http://www.nielsen-netratings.com/>, 2003
- [RFC2647] The Internet Society and Data Communications, D. Newman
Benchmarking Terminology for Firewall Performance
Request for Comments: 2647, August 1999
- [RFC2979] The Internet Society and Sun, N. Freed
Behavior and Requirements for Internet Firewalls
Request for Comments: 2979, October 2000
- [SCMAG02] SC Magazine, J. Wu, P. Joung and J. Kenney
A Practical and Realistic Approach for Testing the Performance of Firewalls: What Your Security Vendors Don't Want You to Know
<http://www.scmagazine.com/scmagazine/sc-online/2002/article/18/article.html>,
April 2002
- [UCLA02] University of California Computer Science Department, J. Mirkovic, J. Martin and P. Reihmer
A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms
Technical Report #020018
- [WEBPDA03] Internet.com Webopedia
Online Dictionary
<http://www.webopedia.com/>