Network Working Group
INTERNET-DRAFT
Expires in January 1999

D. Newman, editor
Data Communications
July 1998

Benchmarking Terminology for Firewall Performance

<draft-ietf-bmwg-secperf-04.txt>

Status of This Memo

Table of Contents

1. Introduction

This document defines terms used in measuring the performance of firewalls. It extends the terminology already used for benchmarking

routers and switches and adds terminology specific to firewalls. The primary metrics used in this document are bit forwarding rate and connections.

There are several reasons why firewall performance measurements are needed. First, despite the rapid rise in firewall deployment, there is no standard means of performance measurement. Second, implementations vary widely, making it difficult to do direct performance comparisons. Finally, more and more organizations are

deploying firewalls on internal networks operating at relatively high speeds, while most firewall implementations remain optimized for use over low-speed wide-area connections. As a result, users are often unsure whether the products they buy will stand up to relatively heavy loads.

2. Existing definitions

This document uses the conceptual framework established in RFCs 1242 and 1944 (for routers) and RFC 2285 (for switches). The router and switch documents contain discussions of several terms relevant to benchmarking the performance of firewalls. Readers should consult the router and switch documents before making use of this document.

This document uses the definition format described in RFC 1242, Section 2. The sections in each definition are: definition, discussion, measurement units (optional), issues (optional), and cross-references.

3. Term definitions

3.1 Allowed traffic

Definition:

Packets forwarded as a result of the rule set of the device under test/system under test (DUT/SUT).

Discussion:

Firewalls typically are configured to forward only those packets explicitly permitted in the rule set. Forwarded packets MUST be included in calculating the bit forwarding rate or maximum bit forwarding rate of the DUT/SUT. All other packets MUST NOT be included in bit forwarding rate calculations.

Measurement units:

not applicable

Issues:

See also:

policy rule set

3.2 Application proxy

Definition:

A type of proxy service that is application aware. As such these proxies can ensure that only specific types of application commands and data pass through; authenticate the user creating the connection; dynamically open and close auxiliary ports sometimes

required by applications.

Discussion:

Application proxies are aware of the type of data and application commands that are expected to be associated with a given TCP or UDP port and ensures that only such traffic is passed through those ports. For example, TCP port 21 should only allow FTP commands and responses through and not allow a non-FTP protocol to be used for potentially malicious means. An application proxy also can determine which dynamically allocated ports are required to allow the service to work properly. In the case of FTP, it requires that port 20 (ftp-data) be open for a period of time and then closed after the transfer is complete.

Measurement units:

not applicable

Issues:

circuit proxy
rule sets

See also:

allowed traffic
circuit proxy
proxy rejected

traffic rule set

3.3 Authentication

Definition:

The process of verifying that a user requesting a network resource is who he, she, or it claims to be, and vice versa.

Discussion: Trust is a critical concept in network security. Any network resource (such as a file server or printer) with restricted access MUST require authentication before granting access.

Authentication takes many forms, including but not limited to IP addresses; TCP or UDP port numbers; passwords; external token authentication cards; and biometric identification such as signature, speech, or retina recognition systems.

The entity being authenticated MAY be the client machine (for example, by proving that a given IP source address really is that address, and not a rogue machine spoofing that address) or a user (by proving that the user really is who he, she, or it claims to be). Servers SHOULD also authenticate themselves to clients.

Testers should be aware that in an increasingly mobile society,

authentication based on machine-specific criteria such as an IP address or port number is not equivalent to verifying that a given individual is making an access request. At this writing systems that verify the identity of users are typically external to the firewall, and may introduce additional latency to the overall SUT.

Measurement units:

not applicable

Issues:

See also:

user

3.4 Bit forwarding rate

Definition:

The number of bits per second of allowed traffic a DUT/SUT can be

observed to transmit to the correct destination interface(s) in response to a specified offered load.

Discussion:

This definition differs substantially from section 3.17 of RFC 1242 and section 3.6.1 of RFC 2285.

Unlike both RFCs 1242 and 2285, this definition introduces the notion of different classes of traffic: allowed, illegal, and rejected (see definitions for each term). Any bit forwarding rate measurement MUST include only allowed traffic.

Unlike RFC 1242, there is no reference to lost or retransmitted data. Forwarding rate is assumed to be a goodput measurement, in that only data successfully forwarded to the destination interface is measured. Bit forwarding rate MUST be measured in relation to the offered load. Bit forwarding rate MAY be measured with differed load levels, traffic orientation, and traffic distribution.

Unlike RFC 2285, this measurement counts bits per second rather than frames per second. Per-frame metrics are not meaningful in the context of a flow of application data between endpoints.

Units of measurement:

bits per second

Issues:

Allowed traffic vs. rejected traffic

See also:

allowed traffic
goodput
illegal traffic
rejected traffic

3.5 Circuit proxy

Definition:

A type of proxy service that copies bytes between authorized source and destinations for defined TCP ports. The data is copied

without any intelligent processing and the proxy cannot dynamically open and close application specific auxiliary ports that are sometimes required.

Discussion:

The key distinction with circuit proxies is that they are static and thus will always set up a connection if the DUT/SUT's rule set allows it. For example, if a firewall's rule set permits ftp connections, a circuit proxy will forward traffic on TCP port 20 (ftp default data) even if no control connection was first established on TCP port 21 (ftp control).

Measurement units:

not applicable

Issues:

application proxy
rule sets

See also:

allowed traffic
application proxy
proxy
rejected traffic
rule set

3.6 Concurrent connections

Definition:

The aggregate number of simultaneous connections between hosts across the DUT/SUT, or between hosts and the DUT/SUT.

Discussion:

The number of concurrent connections a firewall can support is just as important a metric for some users as maximum bit

forwarding rate.

While "connection" describes only a state and not necessarily the transfer of data, concurrency assumes that all existing connections are in fact capable of transferring data. If a data

cannot be sent over a connection, that connection should not be counted toward the number of concurrent connections.

Measurement units:

Concurrent connections
Maximum number of concurrent connections

Issues:

See also:

connections
connection establishment rate
connection overhead

3.7 Connection

Definition:

A state in which two hosts, or a host and the DUT/SUT, agree to exchange data using a known protocol.

Discussion:

A connection is an abstraction describing an agreement between two nodes: One agrees to send data and the other agrees to receive it.

Connections may be TCP sessions, but they don't have to be. Other connection-oriented protocols such as ATM also may be used, either instead of or in addition to TCP connections.

What constitutes a connection depends on the application. For a "native ATM" application like a video stream, connections and virtual circuits can be synonymous. For TCP/IP applications on ATM networks (where multiple TCP sessions may ride over a single ATM virtual circuit), the number of TCP connections is probably the most important consideration.

Additionally, in some cases firewalls may handle a mixture of native TCP and native ATM connections. In this situation, the wrappers around user data will differ. The most meaningful metric describes what an end-user will see.

Data connections describe state, not data transfer. The existence of a connection does not imply that data travels on that connection at any given time, although if data cannot be forwarded on a previously established connection that connection should not be considered in any aggregrate connection count (see concurrent

connections).

A firewall's architecture dictates where a connection is terminated. In the case of proxy-based systems, a connection terminates at the DUT/SUT. But firewalls using packet filtering or stateful packet filtering designs act only as passthrough devices, in that they reside between two connection endpoints. Regardless of firewall architecture, the number of data connections is still relevant, since all firewalls perform some form of connection maintenance; at the very least, all check connection requests against their rule sets.

Though it seems paradoxical, connectionless protocols such as UDP may also involve connections, at least for the purposes of firewall performance measurement. For example, one host may send UDP packets to another across a firewall. If the destination host is listening on the correct UDP port, it receives the UDP packets. For the purposes of firewall performance measurement, this is considered a connection. Indeed, some firewall implementations dynamically alter their rule sets to allow such connections.

Measurement units:

Connection establishment rate
Concurrent connections
Maximum number of concurrent connections

Issues:

proxy-based vs. stateful packet filtering
TCP/IP vs. ATM
connection-oriented vs. connectionless

See also:

data source
concurrent connections
connection establishment rate

3.8 Connection establishment rate

Definition:

The length of time needed for two hosts, or a host and the DUT/SUT, to agree to set up a data exchange using a known protocol.

Discussion:

Each connection-oriented protocol has its own defined mechanisms for setting up a connection. For purposes of benchmarking firewall performance, this shall be the interval between receipt of the first octet of the packet carrying a connection establishment request on a DUT/SUT interface until transmission of the last

octet of the last packet of the connection setup traffic headed in the opposite direction.

This definition applies only to connection-oriented protocols such as TCP. For connectionless protocols such as UDP, the notion of connection setup time is not meaningful.

Measurement units:

Connection establishment rate

Issues:

See also:

concurrent connections
connection connection overhead

3.9 Connection overhead

Definition:

The degradation in bit forwarding rate, if any, observed as a result of the addition of one connection between two hosts through the DUT/SUT, or the addition of one connection from a host to the DUT/SUT.

Discussion:

The memory cost of connection establishment and maintenance is highly implementation-specific. This metric is intended to describe that cost in a method visible outside the firewall.

It may also be desirable to invert this metric to show the performance improvement as a result of tearing down one connection.

Measurement units:

bit forwarding rate

Issues:

3.10 Data source

Definition:

A station capable of generating traffic to the DUT/SUT.

Discussion:

One data source MAY emulate multiple users or stations. In addition, one data source MAY offer traffic to multiple network interfaces on the DUT/SUT.

Measurement units:

not applicable

Issues:

The term "data source" is deliberately independent of any number of users. It is useful to think of data sources simply as traffic generators, without any correlation to any given number of users.

See also:

connection

3.11 Demilitarized zone (DMZ)

Definition:

A network segment or segments located between protected and unprotected networks. DMZ networks are sometimes called perimeter networks.

Discussion:

As an extra security measure, networks are often designed such that protected and unprotected segments are never directly connected. Instead, firewalls (and possibly public resources such as WWW or FTP servers) often reside on the so-called DMZ network. To connect protected, DMZ, and unprotected networks with one device, the device MUST have at least three network interfaces.

Multiple firewalls MAY bound the DMZ. In this case, the firewalls connecting the protected network with the DMZ and the DMZ with the

unprotected network MUST each have at least two network interfaces.

Measurement units:

not applicable

Issues:

Homed

See also:

unprotected network
perimeter network
protected network

3.12 Firewall

Definition:

A device or group of devices that enforces an access control policy between networks.

Discussion:

While there are many different ways to accomplish it, all firewalls do essentially the same thing: control access between networks.

The most common configuration involves a firewall connecting two segments (one protected and one unprotected), but this is not the only possible configuration. Many firewalls support tri-homing, allowing use of a DMZ network. It is possible for a firewall to accommodate more than three interfaces, each attached to a different network segment.

The criteria by which access is controlled is deliberately not specified here. Typically this has been done using network- or transport-layer criteria (such as IP subnet or TCP port number), but there is no reason this must always be so. A growing number of firewalls are controlling access at the application layer, using user identification as the criterion. And firewalls for ATM networks may control access based on data link-layer criteria.

Measurement units:

not applicable

Issues:

See also:

DMZ
tri-homed
user

3.13 Goodput

Definition:

The number of bits per unit of time forwarded to the correct destination interface of the DUT/SUT, minus any bits lost or retransmitted.

Discussion:

Firewalls are generally insensitive to packet loss in the network. As such, measurements of gross bit forwarding rates are not meaningful since (in the case of proxy-based and stateful packet filtering firewalls) a receiving endpoint directly attached to a DUT/SUT would not receive any data dropped by the DUT/SUT.

The type of traffic lost or retransmitted is protocol-dependent.

TCP and ATM, for example, request different types of retransmissions. Testers MUST observe retransmitted data for the protocol in use, and subtract this quantity from measurements of gross bit forwarding rate.

Measurement unit:

bits per second

Issues:

allowed vs. rejected traffic

See also:

allowed traffic
bit forwarding rate
rejected traffic

3.14 Homed

Definition:

The number of logical interfaces a DUT/SUT contains.

Discussion:

Firewalls MUST contain at least two logical interfaces. In network topologies where a DMZ is used, the firewall contains at least three interfaces and is said to be tri-homed. Additional interfaces would make a firewall quad-homed, quint-homed, and so on.

It is theoretically possible for a firewall to contain one physical interface and multiple logical interfaces. This configuration is strongly discouraged for testing purposes because of the difficulty in verifying that no leakage occurs between protected and unprotected segments.

Measurement units:

not applicable

Issues:

See also:

tri-homed

3.15 Illegal traffic

Definition:

Packets specified for rejection in the rule set of the DUT/SUT.

Discussion:

A buggy or misconfigured firewall may forward packets even though its rule set specifies that these packets be dropped. Illegal traffic differs from rejected traffic in that it describes all traffic specified for rejection by the rule set, while rejected traffic specifies only those packets actually dropped by the DUT/SUT.

Measurement units:

not applicable

Issues:

See also:

accepted traffic
policy
rejected traffic
rule set

3.16 Logging

Definition:

The recording of user requests made to the firewall.

Discussion:

Firewalls MUST log all requests they handle, both allowed and rejected. For many firewall designs, logging requires a significant amount of processing overhead, especially when complex rule sets are in use.

The type and amount of data logged varies by implementation. Testers SHOULD attempt to log equivalent data when comparing different DUT/SUTs.

Logging MAY take place on systems other than the DUT/SUT.

Measurement units:

not applicable

Issues:

rule sets

See also:

allowed traffic
connection

rejected traffic

3.17 Network address translation (NAT)

Definition:

A function that maps the original IP source and/or destination addresses of packets arriving at a given interface of the firewall to a different address or addresses.

Discussion:

Firewalls use NAT to ensure that the IP addresses of a protected network are not visible to systems and users on the Internet or some other untrusted network. This is typically required since many networks use RFC 1918 reserved addresses.

In the interest of conserving the IPv4 address space, RFC 1918 proposed the use of certain private (reserved) blocks of IP addresses. Connections to public networks are made by use of a device that translates one or more RFC 1918 addresses to one or more public addresses--a network address translator (NAT).

The use of private addressing also introduces a security benefit in that RFC 1918 addresses are not visible to hosts on the public Internet.

Some NAT implementations are computationally intensive, and may affect bit forwarding rate.

There are two common methods for NAT: many to one (aggregation) and one to one mapping. It should be noted that all proxy firewalls always perform NAT as a function of their architecture, while, by default, packet filtering firewalls do not. In general all application and circuit proxy firewalls, by default, perform NAT as a function of their architecture using the aggregation method, while stateful packet filtering firewalls do not perform NAT by default, but can be configured to do so.

Measurement units:

not applicable

Issues:

See also:

3.18 Packet filtering

Definition:

The process of controlling access by examining packets based on packet header content.

Discussion:

Packet-filtering firewalls forward or deny packets based on information in each packet's header, such as IP address or TCP port number. A packet- filtering firewall uses a rule set to determine which traffic should be forwarded and which should be blocked.

Measurement units:

not applicable

Issues:

static vs. stateful packet filtering

See also:

application proxy
circuit proxy
proxy
rule set
stateful packet filtering

3.19 Perimeter network

Definition:

A network segment or segments located between protected and unprotected networks. Perimeter networks are often called DMZ networks.

Discussion:

See the definition of DMZ for a discussion.

Measurement units:

not applicable

Issues:

Tri-homed

See also:
demilitarized zone (DMZ)
unprotected network
protected network

3.20 Policy

Definition:

A document defining acceptable access to protected, DMZ, and

unprotected networks.

Discussion:

Security policies generally do not spell out specific configurations for firewalls; rather, they set general guidelines for what is and is not acceptable network access.

The actual mechanism for enforcing the access policies is usually the rule set implemented in the DUT/SUT.

Measurement units:

not applicable

Issues:

See also:

rule set

3.21 Protected network

Definition:

A network segment or segments to which access is controlled by the DUT/SUT.

Discussion:

Firewalls are intended to prevent unauthorized access either to or from the protected network. Depending on the configuration specified by the policy and rule set, the DUT/SUT may allow stations on the protected segment to act as clients for servers on either the DMZ or the unprotected network, or both.

Protected networks are often called "internal networks." That term is not used here because firewalls increasingly are deployed within an organization, where all segments are by definition internal. It is also possible for a firewall to protected multiple, but different protected networks from each other.

Measurement units:

not applicable

Issues:

See also:

demilitarized zone (DMZ)
unprotected network
policy
rule set

unprotected network

3.22 Proxy

Definition:

A request for a connection made on behalf of a host.

Discussion:

Proxy-based firewalls do not allow direct connections between hosts. Instead, two connections are established: one between the client host and the DUT/SUT, and another between the DUT/SUT and server host.

As with packet-filtering firewalls, proxy-based devices use a rule set to determine which traffic should be forwarded and which should be rejected.

There are two types of proxies: application proxies and circuit proxies.

Measurement units:

not applicable

Issues:

application

See also:

application
proxy circuit
proxy

packet filtering
stateful packet filtering

3.23 Rejected traffic

Definition:

Packets dropped as a result of the rule set of the DUT/SUT.

Discussion:

Firewalls MUST reject any traffic not explicitly permitted in the rule set. Dropped packets MUST NOT be included in calculating the bit forwarding rate or maximum bit forwarding rate of the DUT/SUT.

Measurement units:

not applicable

Issues:

See also:

policy
rule set

3.24 Rule set

Definition:

The collection of access control rules that determines which packets the DUT/SUT will forward and which it will reject.

Discussion:

Rule sets control access to and from the network interfaces of the DUT/SUT. By definition, rule sets MUST NOT apply equally to all network interfaces; otherwise there would be no need for the firewall. Therefore, a specific rule set MUST be applied to each network interface in the DUT/SUT.

The order of rules within the rule set is critical. Firewalls generally scan rule sets in a "top down" fashion, which is to say that the device compares each packet received with each rule in the rule set until it finds a rule that applies to the packet. Once the device finds an applicable rule, it applies the actions

defined in that rule (such as forwarding or rejecting the packet) and ignores all subsequent rules. For testing purposes, the rule set MUST conclude with a rule denying all access.

Measurement units:

not applicable

Issues:

See also:

demilitarized zone (DMZ)
policy
protected network
rejected traffic
unprotected network

3.25 Stateful packet filtering

Definition:

The process of forwarding or rejecting traffic based on the contents of a state table maintained by a firewall.

Discussion:

Stateful packet filtering ensures that packets associated with an already established (and authorized) connection are allowed to be forwarded through the firewall. This differs from a simple packet filter firewall that would allow any packets through regardless of the state of that connection. For example, a stateful packet filtering device will reject a packet on port 20 (ftp-data) if no session has been established over the ftp control port (usually port 21).

Measurement units:

not applicable

Issues:

See also:

application proxy
circuit proxy

packet filter
proxy

3.26 Tri-homed

Definition:

A firewall with three network interfaces.

Discussion:

Tri-homed firewalls connect three network segments with different network addresses. Typically, these would be protected, DMZ, and unprotected segments.

A tri-homed firewall may offer some security advantages over firewalls with two interfaces. An attacker on an unprotected network may compromise hosts on the DMZ but still not reach any hosts on the protected network.

Measurement units:

not applicable

Issues:

Usually the differentiator between one segment and another is its IP address. However, firewalls may connect different networks of other types, such as ATM or Netware segments.

See also:

homed

3.27 Unprotected network

Definition:

A network segment or segments to which access is not controlled by the DUT/SUT.

Discussion:

Firewalls are deployed between protected and unprotected segments. The unprotected network is not protected by the DUT/SUT.

Note that a DUT/SUT's policy MAY specify hosts on an unprotected

network. For example, a user on a protected network may be permitted to access an FTP server on an unprotected network. But the DUT/SUT cannot control access between hosts on the unprotected network.

Measurement units:

not applicable

Issues:

See also:

demilitarized zone (DMZ)
policy
protected network
rule set

3.28 User

Definition:

A person or process requesting access to resources protected by the DUT/SUT.

Discussion:

"User" is a problematic term in the context of firewall performance testing, for several reasons. First, a user may in fact be a process or processes requesting services through the DUT/SUT. Second, different "user" requests may require radically different amounts of DUT/SUT resources. Third, traffic profiles vary widely from one organization to another, making it difficult to characterize the load offered by a typical user.

For these reasons, it's probably not a good idea to measure DUT/SUT performance in terms of users supported. The only exception is in cases where traffic patterns are well understood and constant--conditions that unfortunately don't exist in many networks. Instead, it's preferable to describe performance in terms of maximum bit forwarding rate and maximum number of connections sustained. It's also preferable to use the term "data

source" rather than "user" to describe the traffic generator(s) to avoid any confusion with actual user data profiles.

Measurement units:

not applicable

Issues:

See also:

data source

4. Security considerations

The primary goal of this memo is to describe terms used in benchmarking firewall performance. However, readers should be aware that there is some overlap between performance and security issues. Specifically, the optimal configuration for firewall performance may not be the most secure, and vice-versa.

Further, certain forms of attack may degrade performance. One common form of denial-of-service (DoS) attack bombards a firewall with so much rejected traffic that it cannot forward allowed traffic. DoS attacks do not always involve heavy loads; DoS describes any state in which a firewall is offered rejected traffic that prohibits it from forwarding some or all allowed traffic. Even a small amount of traffic-- such as the recent Teardrop2 attack involving a few packet fragments-- may significantly degrade firewall performance, or stop the firewall altogether. Further, the safeguards in firewalls to guard against such attacks may have a significant negative impact on performance.

Since the library of attacks is constantly expanding, no attempt is made here to define specific attacks that may affect performance. Nonetheless, any reasonable performance benchmark must take safeguards against such attacks into consideration. Specifically, the same safeguards must be in place when comparing performance of different firewall implementations.

5. References

Bradner, S., editor. "Benchmarking Terminology for Network Interconnection Devices." RFC 1242.

Bradner, S., and McQuaid, J. "Benchmarking Methodology for Network Interconnect Devices." RFC 1944.

Mandeville, R. "Benchmarking Terminology for LAN Switching Devices." RFC 2285.

Rekhter, Y., et al. "Address Allocation for Private Internets." RFC 1918.

6. Acknowledgments

The editor wishes to thank the IETF Benchmarking Working Group for agreeing to review this document. Many persons offered valuable contributions and critiques during this project: Ted Doty (Internet Security Systems), Kevin Dubray (Bay Networks), Helen Holzbaur (NSTL), Jim Hurd (NSTL), Dale Lancaster (Axent Technologies), Robert Mandeville (European Network Laboratories), Brent Melson (NSTL), Steve Platt (NSTL), Marcus Ranum (Network Flight Recorder Inc.), Greg Shannon (Ascend Communications), Christoph Schuba (Sun Microsystems), Rick Siebenaler (Cyberguard), and Greg Smith (Check Point Software Technologies).

7. Contact information

David Newman
Data Communications magazine
1221 Avenue of the Americas, 41st floor
New York, NY 10020 USA
212-512-6182 voice
212-512-6833 fax
dnewman@data.com