Operations Area Working Group
Internet-Draft
Intended status: BCP
Expires: April 21, 2013

F. Baker
Cisco Systems
P. Hoffman
VPN Consortium

October 18, 2012

On Firewalls in Internet Security
draft-ietf-opsawg-firewalls-01

Abstract

This document discusses the most important operational and security implications of using modern firewalls in networks. It makes recommendations for operators of firewalls, as well as for firewall vendors.

Table of Contents

1. Introduction

In this document, a firewall is defined as a device or software that imposes a policy whose effect is "a stated type of packets may or may not pass from A to B". All modern firewalls allow an administrator to change the policies in the firewall, although the ease of administration for making those changes, and the granularity of the policies, vary widely between firewalls and vendors.

Given this definition, it is easy to see that there is a perimeter (the position between A and B) in which the specific security policy applies. In typical deployed networks, there are usually some easy-to-define perimeters. If two or more networks that are connected by a single device, the perimeter is inside the device. If that device is a firewall, it can impose a security policy at the shared perimeters of those networks.

Many firewalls also employ some perimeters that are not as easy to define. Some of these perimeters in modern firewalls include:

- An application-layer gateway (ALG) in front of a server creates a perimeter between that server and the network it is connected to. The ALG blocks some of the flows in the application protocol based on policies such as "do not all traffic from this network" and "do not allow the client to send a message of this type".

- Routing domains that are controlled with role-based administration create perimeters in a routed network. Role-based administration makes rules such as "Domain X cannot see Domain Y in its routing table"; this prevents any host in Domain X from sending traffic to any host in Domain Y.

- [[[MORE HERE with other interesting perimeters]]]

Modern firewalls apply perimeters at three layers:

Layer 3: Most firewalls can filter based on source and destination IPv4 addresses. Many (but, frustratingly, not all) firewalls can filter based on IPv6 addresses.

Layer 4: Most firewalls can filter based on TCP and UDP ports. Many (but, frustratingly, not all) firewalls can also filter based on transports other than TCP and UDP.

Layer 7: Modern firewalls can filter based on the application protocol contents, such as to allow or block certain types of protocol-defined messages, or based on the contents of those messages.

Note that many firewall devices can only create policies at one or two of the layers.

Hardware-based firewalls by their nature inspect traffic flowing through them, sometimes using proprietary mechanisms to make traffic analysis as fast as possible on the given hardware. Some firewalls use network visibility protocols such as NetFlow and sFlow to help capture and analyze traffic. [[References needed]]

1.1. Modern Firewall Features That Should Not Be Confused with Firewalling

There are a few features that appear in any firewall devices that have become associated with firewalls but in fact are not used for firewalling. Those non-firewalling features include:

Network Address Translation (NAT) [RFC2993], which is not used for security policy

IPsec [RFC4301], which is used for virtual private networks (VPNs). Although the core IPsec protocol has firewalling in it, when IPsec appears in a firewall device, it is normally only associated with the application of authenticated encryption and integrity protection of traffic.

"SSL VPN" is a set of technologies that rely on tunneling traffic through the TLS [RFC5246] protocol running on port 443. Some firewalls offer SSL VPNs as an alternative to IPsec.

Traffic prioritization is a feature common in firewalls, but does not meet the definition of firewalling at all.

1.2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Some terms which have specific meanings in this document (such as "firewall") are defined earlier in this section.

2. High-Level Firewall Concepts

2.1. The End-to-End Principle

One common complaint about firewalls in general is that they violate the End-to-End Principle [EndToEnd]. The End-to-End Principle is

often incorrectly stated as requiring that "application specific functions ought to reside in the end hosts of a network rather than in intermediary nodes, provided they can be implemented 'completely and correctly' in the end hosts" or that "there should be no state in the network."

What it actually says is heavily nuanced, and is a line of reasoning applicable when considering any two communication layers. The document says that it "presents a design principle that helps guide placement of functions among the modules of a distributed computer system. The principle, called the end-to-end argument, suggests that functions placed at low levels of a system may be redundant or of little value when compared with the cost of providing them at that low level."

In other words, the End-to-End Argument is not a prohibition against lower layer retries of transmissions, which can be important in certain LAN technologies, nor of the maintenance of state, nor of consistent policies imposed for security reasons. It is, however, a plea for simplicity. Any behavior of a lower communication layer, whether found in the same system as the higher layer (and especially application) functionality or in a different one, that from the perspective of a higher layer introduces inconsistency, complexity, or coupling extracts a cost. That cost may be in user satisfaction, difficulty of management or fault diagnosis, difficulty of future innovation, reduced performance, or other forms. Such costs need to be clearly and honestly weighed against the benefits expected, and used only if the benefit outweighs the cost.

From that perspective, introduction of a policy that prevents communication under an understood set of circumstances, whether it is to prevent access to pornographic sites or prevents traffic that can be characterized as an attack, does not fail the end to end argument; there are any number of possible sites on the network that are inaccessible at any given time, and the presence of such a policy is easily explained and understood.

What does fail the end-to-end argument is behavior that is intermittent, difficult to explain, or unpredictable. If I can sometimes reach a site and not at other times, or reach it using this host or application but not another, I wonder why that is true, and may not even know where to look for the issue.

2.2. Building a Communication

Any communication requires at least three components:

- a sender, someone or some thing that sends a message,
- a receiver, someone or some thing that receives the message, and
- a channel, which is a medium by which the message is communicated.

In the Internet, the IP network is the channel; it may traverse something as simple as a directly connected cable or as complex as a sequence of ISPs, but it is the means of communication. In normal communications, a sender sends a message via the channel to the receiver, who is willing to receive and operate on it. In contrast, attacks are a form of harassment. A receiver exists, but is unwilling to receive the message, has no application to operate on it, or is by policy unwilling to. Attacks on infrastructure occur when message volume overwhelms infrastructure or uses infrastructure but has no obvious receiver.

By that line of reasoning, a firewall primarily protects infrastructure, by preventing traffic that would attack it from it. The best prophylactic might use a procedure for the dissemination of flow specification rules from [RFC5575] to drop traffic sent by an unauthorized or inappropriate sender or which has no host or application willing to receive it as close as possible to the sender.

In other words, as discussed in Section 1, a firewall compares to the human skin, and has as its primary purpose the prophylactic defense of a network. By extension, the firewall also protects a set of hosts and applications, and the bandwidth that serves them, as part of a strategy of defense in depth. A firewall is not itself a security strategy; the analogy to the skin would say that a body protected only by the skin has an immune system deficiency and cannot be expected to long survive. That said, every security solution has a set of vulnerabilities; the vulnerabilities of a layered defense is the intersection of the vulnerabilities of the various layers (e.g., a successful attack has to thread each layer of defense).

3. Firewalling Strategies

There is a great deal of tension in firewall policies between two primary goals of networking: the security goal of "block traffic unless it is explicitly allowed" and the networking goal of "trust hosts with new protocols". The two inherently cannot coexist easily in a set of policies for a firewall.

3.1. Blocking Traffic Unless It Is Explicitly Allowed

The security goal of "block traffic unless it is explicitly allowed" prevents useful new applications. This problem has been seen repeatedly over the past decade: a new and useful application protocol is deployed, but it cannot get wide adoption because it is blocked by firewalls. The result has been a tendency to try to run new protocols over established applications, particularly over HTTP [RFC3205]. The result is protocols that do not work as well they might if they were designed from scratch.

Worse, the same goal prevents the deployment of useful transports other than TCP, UDP, and ICMP. A conservative firewall that only knows those three transports will block new transports such as SCTP [RFC4960]; this in turn causes the Internet to not be able to grow in a healthy fashion. Many firewalls will also block TCP and UDP options they don't understand, and this has the same unfortunate result.

[[[MORE HERE about forcing more costly and error-prone layer 7 inspection]]]

3.2. Typical Firewall Categories

Most IPv4 firewalls have pre-configured security policies that fall into one of the following categories:

I: Block all outside-initiated traffic, allow all inside-initiated traffic

II: Same as I, but allow outside-initiated traffic to some specific inside hosts. The specified hosts are often added by IP address (or sometimes by DNS host name), and the host may be limited to particular transport and application protocols. For example, a rule might allow traffic destined to 203.0.113.226 on TCP ports 80 and 443.

III: Same as I or II, but allow some outside-initiated traffic over some protocols to all hosts. For example, a firewall protecting a farm of web servers might want to allow traffic using TCP ports 80 and 443 to all addresses protected by the firewall so that new servers can be deployed without having to update the firewall rules.

Firewalls that understand IPv6 may have a fourth category:

IV: Allow nearly all outside-initiated traffic. [[[MORE HERE about why this is considered a good idea by some and a bad idea by

others]]]]

3.3. Newer categories of firewalling

[[[MORE HERE on blocking traffic based on dynamic origin reputation such as the long-expired vyncke-advanced-ipv6-security]]]

4. Recommendations for Operators

[[[MORE HERE with the following outline]]]

Firewalling strategies
- None. This is really the operator's choice.
- Be aware that deep packet inspection causes varying amounts of delay in firewalls, particularly for long-lived flows

Don't enforce protocol semantics in the firewall
- Applications are easier to change than firewalls

Avoid using application-layer gateways for firewalling
- Use the security in the applications servers instead
- Servers are easier to change than firewalls
- However, ALGs are useful for IPv4-IPv6 conversion and proxying in some protocols

Allow fragments
- Except in specific protocols where layer 7 content filtering is deemed crucial

Document your intended firewall strategy and settings
- Be sure that other operators of the firewall are able to see it

Don't rely on a NAT for security (see Appendix A)

If using IPsec or SSL VPN, test whether the filtering rules for the rest of the firewall apply

5. Recommendations for Firewall Vendors

[[[MORE HERE with the following outline]]]

Make a set of NAT-like rules for IPv6 easily choosable

Interface for pinholing of IPv4 NATs needs clearly identify security issues

Follow the BEHAVE RFC rules for binding timeouts on NATs

Keep a summary log of non-normal events to aid reviewing

Make leaving notes about the firewalling rules easy and useful

Implement draft-ietf-pcp-base and probably the follow-on protocols from that WG

6. IANA Considerations

None.

7. Security Considerations

This document is all about security considerations. It introduces no new ones.

8. Acknowledgements

Warren Kumari commented on this document.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

9.2. Informative References

[EndToEnd] Saltzer, JH., Reed, DP., and DD. Clark, "End-to-end arguments in system design", ACM Transactions on Computer Systems (TOCS) v.2 n.4, p277-288, Nov 1984.

[RFC2993] Hain, T., "Architectural Implications of NAT", RFC 2993, November 2000.

[RFC3205] Moore, K., "On the use of HTTP as a Substrate", BCP 56, RFC 3205, February 2002.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC4960] Stewart, R., "Stream Control Transmission Protocol", RFC 4960, September 2007.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, August 2009.

Appendix A. IPv4 NATs Are Not Security Devices

Their security is a side-effect of their design. [[[MORE HERE about the history and why some operators mistake the security policy of NATs with firewalls.]]]

[[[MORE HERE about how pinholes mess badly that security policy.]]]

[[[MORE HERE about PCP and how to integrate it with a firewall security policy.]]]

Recommendations for deploying NATs in firewalls include:

- NATs should only be used when more IPv4 addresses are needed
- Operators should not pinhole to addresses that are unpredictably assigned by DHCP

Appendix B. Origin Reputation and Firewalls

[[[MORE HERE with the following outline]]]

Letting someone else curate your security policy
Different types of reputation for different layers
draft-ietf-repute-model
draft-vyncke-advanced-ipv6-security
draft-hallambaker-omnibroker
Recommendations
Check logs to be sure updates are happening
Check vendors' policies

Authors' Addresses

Fred Baker
Cisco Systems

Email: fred@cisco.com

Paul Hoffman
VPN Consortium

Email: paul.hoffman@vpnc.org