Viewpoints

By Terry Gray

Firewalls: Friend or Foe?

We connect our computers to networks because we want to share resources. Others connect their computers to networks because they want to control our computers or data. We thus try to enable the interactions we want and to prevent the ones we don't. Hardware or software elements that enforce these distinctions are called *firewalls*. Firewalls block network traffic based on rules about the source, destination, and type of traffic involved. A firewall can operate on a specific computer ("host-based") or within the enterprise network topology ("network").

Network firewalls are important security tools, but they have significant disadvantages, and their effectiveness is often overstated. This is especially true of enterprise border firewalls, which attempt to create a moat around an entire organization. Border firewall capability is essential for dealing with cyberattack emergencies, but used routinely, it tends to be autocratic and inflexible, thereby encouraging various kinds of subversion, and it can encourage sloppy security practices within the border. Indeed, a border firewall may actually reduce overall security, just as the Maginot Line of networked forts and bunkhouses did in 1940 France.

Perimeter Protection

Network firewalls implement a *perimeter defense*, but never perfectly, so where and how to block network traffic is a contentious issue. I hold to a key premise, which I call the "Perimeter Protection Paradox": as the number of systems protected by a perimeter defense increases, the *value* of the protection mechanism increases but the *effectiveness* of the mechanism decreases. This is because the probability of at least one computer being *somehow* compromised increases with population, as do requirements for "holes" in the firewall. Every defensive perimeter creates a vulnerability zone within it. Thus, *the security of computers on a network is maximized when the network protection perimeter is minimized, that is, when it is as close to the end-system(s) as possible.*

One large-perimeter security technique that may make sense is the use of private addresses within the enterprise network. These are addresses that are not globally routed, so devices (e.g., printers) using them are invisible from the Internet. This idea can be combined with network address translation (NAT) to provide outbound connectivity for hosts, similar to that provided to home users by residential gateways. However, NAT breaks some applications and can increase support costs. Moreover, the perimeter protection paradox still applies, so insider attacks are a concern.

Host-Based Firewalls

In contrast to network perimeter firewalls, especially those placed at the enterprise border, firewalls on individual computers offer the ideal perimeter size, since there is no network vulnerability zone between firewall and host. They also allow for optimal security policy (requirements are determined by only one system) and avoid many other disadvantages inherent in network firewalls. However, they can be a support nightmare unless centrally managed. Fortunately, unlike personal intrusion-detection systems, which are notorious for false positives and corresponding support headaches, centrally managed host-based firewalls have been successfully deployed without undue pain.

It is not widely understood that host-based firewalls can provide the same perimeter protection for insecure services as do border firewalls. This is possible by using "IP access lists" to selectively block traffic originating outside a range of trusted network addresses. Thus, if it is necessary to use insecure protocols locally, one can still implement "border blocking" via host-based firewalls. Alas, many installed computers and other network devices lack such capabilities—hence the need for network, or perimeter, firewalls.

Network Firewalls

Between the end-systems and the enterprise border, there are many possible places and methods to implement traffic filtering, each with different tradeoffs between network *operations* objectives and network *security* objectives. These tensions are magnified when different constituencies have conflicting security policy goals. From a central IT operations perspective, exceptions are the enemy, because they increase support costs and reduce reliability, so policies that can be applied uniformly throughout the network win the day. For pervasive deployment, blocking policies should meet three tests: (1) provide protection that system managers cannot reasonably provide; (2) not seriously degrade performance of the network core; (3) produce widespread consensus that they are a Good Thing. Blocking traffic with forged "source" addresses meets all three tests and should be done in all

network routers. Specific subgroups within an organization may choose different tradeoffs between complexity and security—that is, they differ on the Good Thing criterion. Central IT network engineers and policymakers must help units achieve their security objectives without undermining the operational integrity of the network utility. Whether this is easy or hard to accomplish depends on where service boundaries have been drawn within the institution.

Administrative Boundaries

When the central IT group is responsible for connectivity to every Ethernet outlet, allowing departments to place conventional firewalls at subnet perimeters causes problems because the firewalls may interfere with managing the network devices beyond them. In such environments, providing for proper network management and at the same time allowing departments to have security policy autonomy requires unconventional techniques, such as "logical firewalls" or VLAN-based firewalls. These techniques preserve network manageability where there are boundary conflicts, but they also increase operational complexity.

Departmental computer labs, machine rooms, and server clusters offer a simpler problem. Often the network infrastructure within such facilities is not centrally managed, so there is no conflict in putting a conventional firewall at the point where they attach to the campus network. (However, even here, certain firewall types may conflict with network management policies.) Not all protection aims outward. In academic computer labs, firewall rules might be adjusted to protect the rest of the network from the student machines!

Firewall Disadvantages

Since firewalls must examine every packet, they often degrade network performance. Scalability becomes a major problem for firewalls within or associated with border routers.

The cost of managing firewall rules is high, especially if exceptions are permitted or change often. Even without these exceptions, it can take considerable effort to vindicate firewalls when debugging network problems. In addition, getting consensus on what to block at a college or university border is at best nontrivial. Perhaps surprisingly, blocking the most obvious ports, such as file sharing or remote control, is controversial. Ideally the "administrative distance" between policymakers and those impacted by their policies should be small. But large-perimeter firewalls are distant from those affected.

Perimeter blocking policies often restrict all but the most common types of traffic. Rather than lobby for uncommon traffic to be allowed, many users reconfigure uncommon services to masquerade as common ones—for example, implementing file sharing so that it appears to be Web traffic. "Tunneling," as this practice is called, can create security problems and make traffic engineering and capacity planning more difficult. Similarly, users often try to defeat restrictions based on source or destination by routing traffic through backdoor intermediary computers, such as proxy servers. Firewalls cannot stop these problems, since they must allow the common or diverted traffic.

Although many computer users and administrators manage their computers well, taking responsibility for their own security, not all have the time to properly manage every system. They rely on large-perimeter firewalls for protection, but such firewalls cannot protect against tunneling, routing, maintenance vulnerabilities, or attacks originating from computers inside the defensive perimeter.

Finally, many attacks exploit weaknesses in core services, especially Web servers, file sharing, domain and host name service, and remote access. These attacks cannot be blocked by a firewall without disabling the very service one is trying to provide; that is, firewalls cannot substitute for software maintenance.

Recommendations

The closer to the end-system one blocks, the more restrictive that blocking can be, and the smaller the network vulnerability zone will be. Accordingly, good security practice starts at the host. Good *network* firewall policy builds on host-based protection and includes the following:

- *In all campus routers:* block forged source addresses; enable private addresses for printers or other devices not needing external connectivity.
- *At the campus border:* be prepared to take emergency action to block attacks, recognizing that doing so may disable key services.
- *At the subnet level:* allow for departmental firewalls that do not interfere with network management.
- *For labs, machine rooms, or server clusters:* use conventional firewalls to prevent all but authorized traffic.

Comprehensive thinking about what to block and where is part of "defense in depth," a fundamental security principle. But defense in depth does not stop with firewalls. A comprehensive approach to system security involves all of the following strategies: use of secure application protocols; host hardening; proactive vulnerability probing; monitoring and intrusion detection; requirement and accountability policies; and of course, the *prudent* use of firewalls.

Many of the Internet's security problems would go away if vendors shipped computers that were "network safe" out-of-the-box. Microsoft's new multimillion-dollar initiative to improve the security of its code is a fine thing, but even a few simple changes—such as requiring passwords on privileged accounts, using IPSEC encryption whenever possible, and enabling the integral firewall by default—would dramatically improve security.

Even when vendors someday provide network-safe computers, thereby (in principle) obviating the need for network firewalls, we will still have many older systems, including some that can never be made safe. Therefore, perimeter defense strategies will remain important adjuncts to host-based security practices—and the complex tensions between network operation and network security will continue. The point, in short, is that firewalls are both friend and foe. Network security cannot be reduced to a simple problem, and whenever we try, we put ourselves at risk.

Terry Gray (gray@washington.edu) is Director of Networks & Distributed Computing at the University of Washington.