Thinking About Firewalls

Marcus J. Ranum

mjr@tis.com

Trusted Information Systems, Inc.

Glenwood, Maryland

Abstract

Generally, he who occupies the field of battle first and awaits his enemy is at ease. —— Sun Tzu

Many companies connect to the Internet, guarded by "firewalls" designed to prevent unauthorized access to their private networks. Despite this general goal, firewalls span a continuum between ease of use and security. This paper describes some of the considerations and tradeoffs in designing firewalls. A vocabulary for firewalls and their components is offered, to provide a common ground for discussion.

Why a Firewall?

Against those skilled in the attack, an enemy does not know where to defend. Against the experts in defence, the enemy does not know where to attack. —— Sun Tzu

The rationale for installing a firewall is almost always to protect a private network against intrusion. In most cases, the purpose of the firewall is to prevent unauthorized users from accessing computing resources on a private network, and often to prevent unnoticed and unauthorized export of proprietary information. In some cases export of information is not considered important, but for many corporations that are connecting this is a major though possibly unreasoning concern. Many organizations will want simply to address the problem by not connecting to the Internet at all. This solution can be difficult to implement. If the private network is loosely administered or decentralized, a single enterprising individual with a high speed dialup modem can quickly effect an Internet SLIP connection which can compromise the security of an entire network.

Often it is safe to say that a firewall needs to be put in place for the "CYA"[1] factor. Even though an employee could compromise proprietary information by carrying it offsite on a DAT or floppy disk, the Internet represents a tangible threat, populated with dangerous "vandals."[2] It could very easily cost a network manager his job if a break-in occurs via this route, even if the damage is no more extensive than could have been inflicted over a dialup line or by a disgruntled employee. Generally, for a would-be Internet site, the technical difficulties of implementing a firewall are greatly outweighed by the public relations problems of "selling" upper management on the idea. In summary, because Internet services are so highly visible, they are much more likely to require official oversight and justification.

---

[1] "*Cover Your Assets*" - this is a family publication.

[2] The Vandals were a collection of tribes of roughneck barbarians who sacked Rome in 455 and looted it of all its portable wealth. Some use the term "hackers" to describe Internet snoopers, but "vandals," "crackers," or "jerks" is more appropriate.

Design Decisions

Examine your environment* — *Miyamoto Musashi

In configuring a firewall, the major design decisions with respect to security are often already dictated by corporate or organizational policy; specifically, a decision must be made as to whether security is more important than ease-of-use, or vice versa. There are two basic approaches that summarize the conflict:

- That which is not expressly permitted is prohibited.
- That which is not expressly prohibited is permitted.

The importance of this distinction cannot be overemphasized. In the former case, the firewall must be designed to block everything, and services must be enabled on a case-by-case basis only after a careful assessment of need and risk. This tends to impact users directly, and they may see the firewall as a hindrance. In the second case, the systems administrator is placed in a reactive mode, having to predict what kinds of actions the user population might take that would weaken the security of the firewall, and preparing defenses against them. This essentially pits the firewall administrator against the users in an endless arms race that can become quite fierce. A user can generally compromise the security of their login if they try or aren't aware of reasonable security precautions. If the user has an open access login on the firewall system itself, a serious security breach can result. The presence of user logins on the firewall system tends to magnify the problem of maintaining the system's integrity. A second important statement of policy is implicit in the "that which is not expressly permitted is prohibited" stance. This stance is more "fail safe," since it accepts that the administrator is ignorant of what TCP ports are safe, or what holes may exist in the manufacturer's kernel or applications. Since many vendors are slow to publicise security holes, this is clearly a more conservative approach. It is an admission of the fact that what you don't know can hurt you.

Levels of Threat

If ignorant both of your enemy and yourself, you are certain in every battle to be in peril* — *Sun Tzu

There are several ways in which a firewall can fail or be compromised. While none of them are good, some are decidedly worse than others. Since the purpose of many firewalls is to block access, it's a clear failure if someone finds a loophole through it which permits them to probe systems in the private network. An even more severe situation would result if someone managed to break into the firewall and reconfigure it such that the entire private network is reachable by anyone. For the sake of terminology, this type of attack will be referred to as "destroying" a firewall, as opposed to a mere "break-in." It is extremely difficult to quantify the damage that might result from a firewall's destruction. An important measure of how well a firewall resists threat is the information it gathers to help determine the course of an attack. The absolute worst thing that could happen is for a firewall to be completely compromised without any trace of how the attack took place. The best thing that can happen is for a firewall to detect an attack, and inform the administrator politely that it is undergoing attack, but that the attack is going to fail.

One way to view the result of a firewall being compromised is to look at things in terms of what can be roughly termed as "zones of risk." In the case of a network that is directly connected to the Internet without any firewall, the entire network is subject to attack. This does not imply that the network is vulnerable to attack, but in a situation where an entire network is within reach of an untrusted network, it is necessary to ensure the security of every single host on that network. Practical experience shows that this is difficult, since tools like *rlogin* that permit user-customizable access control are often exploited by vandals to gain access to multiple hosts, in a form of "island hopping" attack. In the case of any typical firewall, the zone of risk is often reduced to the firewall itself, or a selected subset of hosts on the network, significantly reducing the network manager's concerns with respect to direct attack. If a firewall is broken into, the zone of risk often expands again, to include the entire protected network. A vandal gaining access to a login on the firewall can begin an island hopping attack into the private network, using it as a base. In this situation, there is still some hope, since the vandal may leave traces on the firewall, and may be detected. If the firewall is completely destroyed the private network can undergo attack from any external system and reconstructing the course of an attack becomes nearly impossible.

In general, firewalls can be viewed in terms of reducing the zone of risk to a single point of failure. In a sense, this seems like a bad idea, since it amounts to putting all of one's eggs in a single basket, but practical experience implies that at any given time, for a network of non-trivial size, there are at least a few hosts that are vulnerable to break-in by even an unskilled attacker. Many corporations have formal host security policies that are designed to address these weaknesses, but it is sheer foolishness to assume that publishing policies will suffice. A firewall enhances host security by funneling attackers through a narrow gap where there's a chance of catching or detecting them first. The well-constructed medieval castle had multiple walls and interlocking defense points for exactly the same reason.

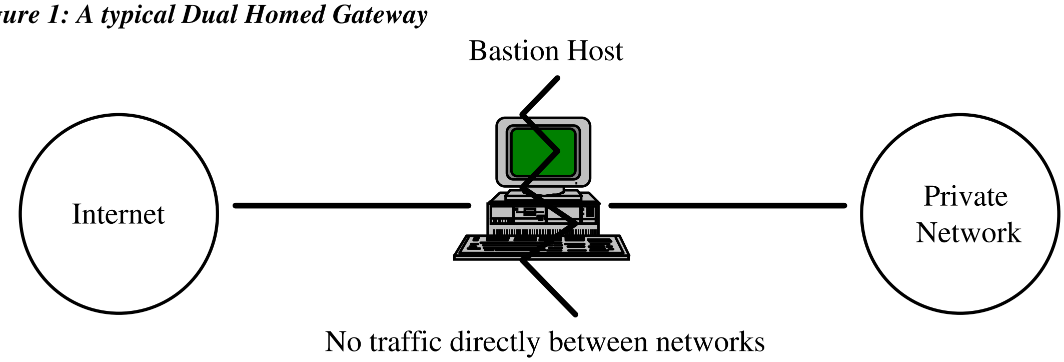
Firewalls and Their Components

There may be a hundred combat postures, but there is only one purpose: to win.* — *Heiho Kaden Sho

In discussing firewalls there is often confusion of terminology since firewalls all differ slightly in implementation if not in purpose. Various discussions on USENET indicate that the term "firewall" is used to describe just about any inter-network security scheme. For the sake of simplifying discussion, some terminology is proposed, to provide a common ground:

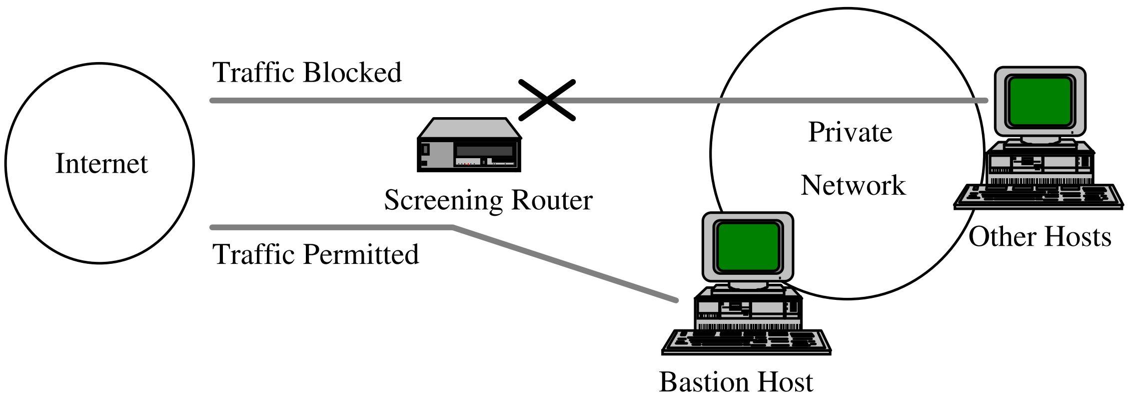
> **Screening Router** — A screening router is a basic component of most firewalls. A screening router can be a commercial router or a host-based router with some kind of packet filtering capability. Typical screening routers have the ability to block traffic between networks or specific hosts, on an IP port level. Some firewalls consist of nothing more than a screening router between a private network and the Internet.
>
> **Bastion host** — Bastions are the highly fortified parts of a medieval castle; points that overlook critical areas of defense, usually having stronger walls, room for extra troops, and the occasional useful tub of boiling hot oil for discouraging attackers. A bastion host is a system identified by the firewall administrator as a critical strong point in the network's security. Generally, bastion hosts will have some degree of extra attention paid to their security, may undergo regular audits, and may have modified software.
>
> **Dual Homed Gateway** — Some firewalls are implemented without a screening router, by placing a system on both the private network and the Internet, and disabling TCP/IP forwarding. Hosts on the private network can communicate with the gateway, as can hosts on the Internet, but direct traffic between the networks is blocked. A dual homed gateway is, by definition, a bastion host.

Figure 1: A typical Dual Homed Gateway



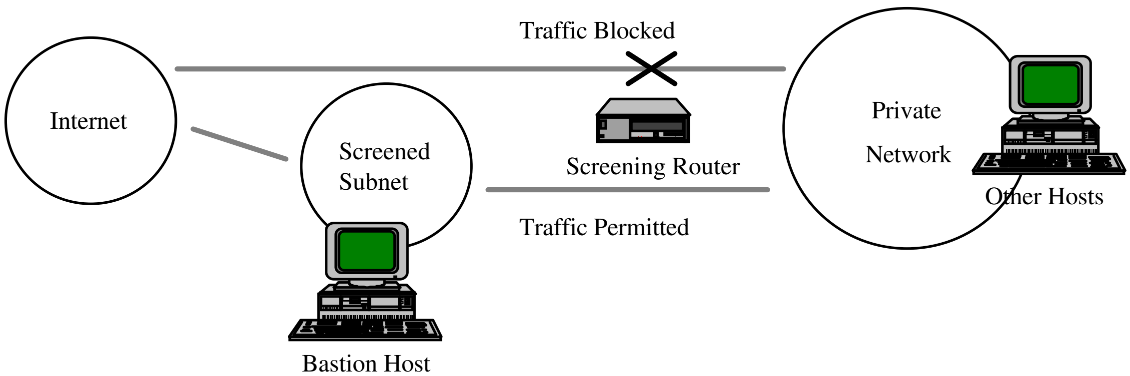
Screened Host Gateway — Possibly the most common firewall configuration is a screened host gateway. This is implemented using a screening router and a bastion host. Usually, the bastion host is on the private network, and the screening router is configured such that the bastion host is the only system on the private network that is reachable from the Internet. Often the screening router is configured to block traffic to the bastion host on specific ports, permitting only a small number of services to communicate with it.

Figure 2: A typical Screened Host Gateway



Screened Subnet — In some firewall configurations, an isolated subnet is created, situated between the Internet and the private network. Typically, this network is isolated using screening routers, which may implement varying levels of filtering. Generally, a screened subnet is configured such that both the Internet and the private network have access to hosts on the screened subnet, but traffic across the screened subnet is blocked. Some configurations of screened subnets will have a bastion host on the screened network, either to support interactive terminal sessions or application level gateways.

Figure 3: A typical Screened Subnet



Application Level Gateway (or "**proxy gateway**") — Much of the software on the Internet works in a store-and-forward mode; mailers and USENET news collect input, examine it, and forward it. Application level gateways are service-specific forwarders or reflectors, which usually operate in user mode rather than at a protocol level. Generally, these forwarding services, when running on a firewall, are important to the security of the whole. The famous *sendmail* hole that was exploited by the Morris Internet worm is one example of the kinds of security problems an application level gateway can present. Other application level gateways are interactive, such as the *FTP* and *telnet* gateways run on the Digital Equipment Corporation firewalls. In general, the term "application level gateway" will be used to describe some kind of forwarding service that runs across a firewall, and is a potential security concern. In general, crucial application level gateways are run on some kind of bastion host.

Hybrid Gateways — Hybrid gateways are the "something else" category in this list. Examples of such systems might be hosts connected to the Internet, but accessible only through serial lines connected to an ethernet terminal server on the private network. Such gateways might take advantage of multiple protocols, or tunneling one protocol over another. Routers might maintain and monitor the complete state of all TCP/IP connections, or somehow examine traffic to try to detect and prevent an attack. The AT&T corporate firewall*[1]* is a hybrid gateway combined with a bastion host.

Taking the components described above, we can accurately describe most of the forms that firewalls take, and can make some general statements about the kinds of security problems each approach presents. Assuming that a firewall fulfills its basic purpose of helping protect the network, it is still important to examine each type of firewall with respect to:

Damage control — If the firewall is compromised, to what kinds of threats does it leave the private network open? If destroyed, to what kinds of threats does it leave the private network open?

Zones of risk — How large is the zone of risk during normal operation? A measure of this is the number of hosts or routers that can be probed from the outside network.

Failure mode — If the firewall is broken into, how easy is this to detect? If the firewall is destroyed, how easy is this to detect? In a post mortem, how much information is retained that can be used to diagnose the attack?

Ease of use — How much of an inconvenience is the firewall?

Stance — Is the basic design philosophy of the firewall "That which is not expressly permitted is prohibited" or is it "That which is not expressly prohibited is permitted?"

Firewalls using Screening Routers

In large scale strategy, when the enemy embarks on an attack, if you make a show of strongly suppressing his technique, he will change his mind. — Miyamoto Musashi

Many networks are firewalled using only a screening router between the private network and the Internet. This type of firewall is different from a screened host gateway in that usually there is direct communication permitted between multiple hosts on the private network, and multiple hosts on the Internet. The zone of risk is equal to the number of hosts on the private networks, and the number and type of services to which the screening router permits traffic. For each service provided via peer-to-peer connection the size of the zone of risk increases sharply. Eventually it is impossible to quantify. Damage control is difficult as well since the network administrator would need to regularly examine every host for traces of a break-in. If there is no regular audit one must hope to stumble on a clue such as a mismatched system accounting record*[2]*.

In the case of total destruction of the firewall, it tends to be very hard to trace or even to discover. If a commercial router (which does not maintain logging records) is used, and the router's administrative password is compromised, the entire private network can be laid open to attack very easily. Cases are known where commercial routers have been configured with erroneous screening rules, or have come up in some pass-through mode because of hardware or operator error. Generally, this configuration is a case of "That which is not expressly prohibited is permitted" as the ingenious user can fairly easily piggyback protocols to achieve a higher level of access than the administrator expects or wants. Given a collaborator on an external host, it is left as an exercise to the reader to implement a remote login stream protocol over Domain Name Service packets[3].

Screening routers are not the most secure solution, but they are popular since they permit fairly free Internet access from any point within the private network. Many consultants and network service providers offer screening routers in a "firewall" configuration. It is uncertain if the various trade-offs involved are clear to the customer; the author would not recommend use of a screening router to protect sensitive information or trade secrets, since screening routers are very permeable from the inside.

Dual Homed Gateways

It was Hannibal's principle to keep his forces united and to garrison but one single fortress — Emperor Napoleon I

An often used and easy to implement firewall is the dual homed gateway. Since it doesn't forward TCP/IP traffic, it acts as a complete block between the Internet and the private network. Its ease of use is determined by how the systems manager chooses to set up access; either by providing application gateways such as *telnet* forwarders or by giving users logins on the gateway host. If the former approach is taken, the stance of the firewall is clearly "That which is not expressly permitted is prohibited"; users can only access Internet services for which there is an application gateway. If users are permitted logins, then, in the opinion of the author, the firewall's security is seriously weakened. During normal operation, the only zone of risk is the gateway host itself, since it is the only host that is reachable from the Internet. If there are user logins on the gateway host, and one of the users chooses a weak password or has their account otherwise compromised, the zone of risk expands to encompass the entire private network. From a standpoint of damage control, the administrator may be able to track the progress of an intruder, based on the access patterns of the compromised login, but a skillful vandal can make this quite difficult. If a dual homed gateway is configured without direct user access, damage control can be somewhat easier, since the very fact that someone has logged in to the gateway host becomes a noteworthy security event. Dual homed gateways have an advantage over screening routers from the standpoint that their system software is often easier to adapt to maintain system logs, hard copy logs, or remote logs. This can make a post-mortem easier for the gateway host itself, but may or may not help the network administrator identify what other hosts on the private network may have been compromised in an island-hopping attack.

Attacking a dual homed gateway leaves the attacker a fairly large array of options. Since the attacker has what amounts to local network access if a login can be obtained, all the usual attacks that can be made over a local network are available. NFS-mounted file systems, weaknesses in *.rhosts* files, automatic software distribution systems, network backup programs and administrative shell scripts — all may provide a toehold on systems on the internal network. Once a toehold is secured, it then provides a base from which to launch attacks back at the gateway itself. The weakest aspect of the dual homed gateway is its failure mode. If the firewalll is destroyed it is possible that a skillful attacker might re-enable routing and throw the entire private network open to attack. In the usual UNIX-based dual homed gateway, TCP/IP routing is often disabled by modifying a kernel variable named *ipforwarding*; if systems privileges can be obtained or stolen on the gateway, this variable can be changed. Perhaps this seems far-fetched, but unless great care is paid to monitoring the software revision levels and configuration on the gateway host, it is not improbable that a vandal with a copy of the release notes for the operating system version and a login can compromise the system.

---

[3]The author has implemented just such a scheme, tunnelling TCP/IP traffic through a firewall. Running over T1 lines interactive session response is quite good and users have complete access to *telnet*, *rlogin*, *FTP*, etc. As PPP (Point to Point Protocol) becomes more widespread, so will this threat.

Screened Host Gateways

***Ground to which access is constricted, where the way out is tortuous, and where a small enemy force can strike my larger one is called "encircled."* — Sun Tzu**

Several articles have described screened host gateways, and how to construct them*[3,4]*. Generally, the screened host gateway is very secure, while remaining fairly easy to implement. Typically, a bastion host is configured on the private network, with a screening router between the Internet and the private network, which only permits Internet access to the bastion host. Since the bastion host is on the private network, connectivity for local users is very good, and problems presented by exotic routing configurations do not present themselves. If the private network is, as many are, a virtual extended local area network (e.g.: no subnets or routing) the screened host gateway will work without requiring any changes to the local network, as long as the local network is using a legitimately assigned set of network addresses. The zone of risk of a screened host gateway is restricted to the bastion host, and the screening router, and the security stance of the screened host gateway is determined by the software running on that system. If an attacker gains login access to the bastion host, there is a fairly wide range of options for attacking the rest of the private network. In many ways, this approach is similar to the dual homed gateway, sharing similar failure modes and design considerations with respect to the software running on the bastion host.

Screened Subnets

When the enemy gets in an inconvenient position, do not let him look around. — Miyamoto Musashi

A screened subnet is usually configured with a bastion host as the sole point of access on the subnet. The zone of risk is small, consisting of that bastion host or hosts, and any screening routers that make up the connections between the screened subnet, the Internet, and the private network. The ease of use and basic stance of the screened subnet will vary, but generally a screened subnet is appealing only for firewalls that are taking advantage of routing to reinforce the existing screening. This approach forces all services through the firewall to be provided by application gateways, and places the stance strongly in the "That which is not expressly permitted is prohibited" category.

If a screened subnet based firewall with inter-network routing blocked is attacked with an intent to destroy it, the attacker must reconfigure the routing on three networks, without disconnecting or locking himself out, and without the routing changes being noticed. No doubt this is possible, but it can be made very difficult by disabling network access to the screening routers, or by configuring the screening routers to only permit access from specific hosts on the private network. In this case, an attacker would need to break into the bastion host, then into one of the hosts on the private network, and then back out to the screening router — and would have to do it without setting off any alarms.

Another advantage of screened subnets is that they can be put in place in such a way that they hide any accidents of history that may linger on the private network. Many sites that would like to connect to the Internet are daunted by the prospect of re-addressing and re-subnetting existing networks. With a screened subnet with blocked inter-network routing, a private network can be connected to the Internet and changed gradually to new subnet and network addresses. In fact, this approach has been observed to significantly accelerate the adoption of new network addresses on loosely controlled private networks. Users will be more receptive to changing their host addresses if they can realize the benefits of Internet connectivity thereby, since hosts that are not correctly addressed cannot use the firewall properly. In most other respects, the screened subnet is very much dependent on the suite of software running on the bastion host. Screening a whole subnet provides functionality similar to the dual homed gateway or screened host gateway; it differs primarily in the extra level of complexity in routing and configuration of the screening routers.

Hybrid Gateways

There is in guerrilla warfare no such thing as a decisive battle; there is nothing comparable to the fixed, passive defense that is a feature of orthodox war. —— Mao on guerrilla warfare

"Security through obscurity" is not sufficient in and of itself, but there is no question that an unusual configuration, or one that is hard to understand, is likely to give an attacker pause, or to make them more likely to reveal themselves in the process of trying to figure out what they are facing. On the other hand there is a real advantage to having a security configuration that is easy to understand, and therefore easier to evaluate and maintain. Since the hybrid gateway is mentioned here in the category of "something else" no attempt will be made to describe the indescribable. Some hypothetical hybrids may serve to show how hybrid gateways might differ from and be similar to the other types.

Let us postulate a hybrid gateway that consists of a box sitting on the Internet, which is capable of routing traffic, but also maintains a complete notion of the state of every TCP connection, how much data has gone across it, where it originated, and its destination. Presumably, connections can be filtered based on arbitrarily precise rules, such as: "permit traffic between host A on the private network and all hosts on network B on the Internet via the *telnet* service if and only if the connection originated from host A between the hours of 9:00 am and 5:00 pm and log the traffic." This sounds terrific, providing arbitrary control with great ease of use, but some problems simply refuse to go away. Consider that someone wishing to circumvent the firewall, who broke into the private network via an unguarded modem, might very easily set up a service engine that was piggybacked over the *telnet* port. This is actually a fairly easy firewall to destroy.

Another hybrid gateway might take advantage of various forms of protocol tunneling. Suppose the requirement is to connect to the Internet with very tight restrictions, but that a high degree of connectivity is required between the private network and an external network that is somewhat trusted (for example, a corporate R&D department needs to be able to run X-windows applications on a supercomputer at another facility). The usual archetypal gateways discussed here could provide general purpose e-mail connectivity, but for secure point-to-point communications, an encrypted point-to-point virtual TCP/IP connection might be set up with the remote system, after users had authenticated themselves with a cryptographic smart card. This would be extremely secure, and might be made fairly easy to use, but has the disadvantage that the protocol driver needs to be added to every system that wants to share communication. It is hard to make any guesses about the failure mode of such a system, but the zone of risk is neatly limited to all the hosts which are running the tunneling protocol driver, and to which the individual user has smart card access. Some of this might be implemented in hardware or in the routers themselves[4]. In the future, it is likely that the rapid growth of the Internet will fuel more development in this area, and we will see various hybrid gateways arise. The basic issues surrounding configuring a firewall will probably remain the same as the ones discussed here.

Other firewall-related tools

Subtle and insubstantial, the expert leaves no trace; divinely mysterious, he is inaudible. —— Sun Tzu

There is active research and development on tools to aggressively seek out and identify weaknesses in an entire network, or to detect the patterns that might indicate when an attack is in progress. These tools range from the simple checklist*[5]* to complex "expert systems" with inference engines and elaborate rule bases. Many firewalls today run software that is designed to go forth and gather information relating to possible attacks and their origins, often using and abusing tools like *finger* and *SNMP[6,7].* Unless true artificial intelligence is developed, however, these tools cannot guard against an unknown form of attack, since they cannot possibly match the creativity of a network vandal. While often

---

[4]Routers performing point-to-point encryption for specified networks are beginning to appear on the market. Until recently, the only demand for such hardware was for government applications, and most solutions employed Type I cryptography.

billed as being "proactive" they are in fact reactive, and generally will serve only to catch systems crackers armed with last year's bag of tricks. Catching the small fry is still worth doing, but it is likely that they are less of a threat than the fellow who is so eager to break into your network that he is doing research and development in new system cracking techniques.

No Conclusions, but Observations

Those skilled in war cultivate the Tao and preserve the laws and are therefore able to formulate victorious policies. —— Sun Tzu

It is the privilege of a writer to use the last section of a publication to state his opinions and call them "conclusions." In dealing with firewalls, it is simply not reasonable to say that any particular approach is best, since there are so many factors that determine what the best firewall for a given situation may be. Cost, corporate policy, existing network technology, staffing, and intra-organizational politics may all easily outweigh the technical considerations presented here.

There are a few observations worth making about firewalls at a very general level. Firstly, a firewall is a leverage-increasing device from a network management point of view. Rather than looking at it as "all eggs in one basket," it can also be viewed as a trustworthy basket, and a single point from which a very important security system can be controlled. The size of the zone of risk is crucial to the design; if it is small, security can be maintained and controlled easily but if security is compromised, the damage can be more severe. The ideal would be to have such strong host-based security that a firewall would be redundant. Systems administration costs and a hard dose of reality prevents this ideal from being obtainable.

A second important aspect of firewall building is that it is not something to undertake in a vacuum. Many sites are connected with a simple firewall consisting of a screening router and nothing more because someone told them that it was "secure enough." In setting up a firewall one must trade off time and money, security, and risk. "Secure enough" depends largely on what you risk losing. Finally, it is important when approaching implementing a firewall to avoid the urge to start from scratch. System security is a lot like pregnancy; one is seldom only broken into a little bit, and it only takes a little mistake or a moment of inattention to find oneself in a delicate position. Leaning on the experiences of others, and learning from their mistakes and successes is very important. Setting up a firewall is definitely an area where having a wide background in experience to draw upon is important. The vandals on the network have a wide background in experience to draw upon as well, and a firewall administrator must communicate with others, and must keep up to date on other firewall-related happenings on the network. Static defenses do not work unless they keep up with emerging tricks of the trade, or one's firewall may be the next Maginot Line.

The purpose of this paper is not to discourage companies from connecting to the Internet. The Internet is an incredibly valuable resource, one which will in the coming years completely change the way people work and communicate on a global level. The benefits of connection far outweigh the costs, but it is wise to reduce the costs and potential costs as much as possible, by being aware of the dangers and being as protected as is necessary. The purpose of this paper is not to sow "fear, uncertainty and doubt" in order to sell any particular firewall or software. If the author has succeeded to any degree in describing some of the considerations and tradeoffs in setting up a firewall, this paper has been a success, and the reader is left to their own devices to determine how best to proceed.

References

[1] Bill Cheswick, "The Design of a Secure Internet Gateway," USENIX proceedings. Available for FTP from research.att.com: /dist/secure_internet_gateway.ps

[2] Cliff Stoll, "The Cuckoo's Egg"

[3] Smoot Carl-Mitchell and John Quarterman, "Building Internet Firewalls," UNIX World, February 1992

[4] Simson Garfinkel and Gene Spafford, "Practical UNIX Security," O'Reilly and Associates, June 1991

[5] Dan Farmer, "COPS and Robbers, UN*X System Security," Internet software. Available for FTP from cert.sei.cmu.edu: /pub/cops

[6] Bill Cheswick, "An Evening with Berferd in which a cracker is Lured, Endured, and Studied," USENIX proceedings, Jan 20, 1990 Available for FTP from research.att.com: /dist/internet_security/berferd.ps

[7] Marcus J. Ranum, "An Internet Firewall," proceedings of World Conference on Systems Management and Security, 1992 Available for FTP from decuac.dec.com: /pub/docs/firewall/firewall.ps