Security at the Network Edge: A Distributed Firewall Architecture

Tom Markham
Secure Computing Corporation
tom_markham@securecomputing.com

Charlie Payne
Secure Computing Corporation
charlie_payne@securecomputing.com

Abstract

This paper introduces Network Edge Security, *a new distributed firewall architecture designed to counter the insider threat, which is undeterred by existing firewall implementations. Network Edge Security also addresses the security challenges of emerging technology trends like mobile computing, business to business computing and virtual private networks (VPNs). We describe the architecture and the rationale for its design, then we illustrate its application.*

1. Introduction

Security professionals understand that the biggest threat to an organization's assets comes not from without but from within. Compare the damage external hackers have done to the Pentagon to the recent losses at DOE labs. As reported recently Lewis Z. Koch [3]:

> "The Computer Security Institute's 1998 Computer Crime Survey, conducted jointly with the Federal Bureau of Investigation, reported that the average cost of an outside hacker penetration totaled $56,000, while the average insider attack costs a company $2.7 million."

Until recently, insider incidents were underreported to avoid embarrassment and the loss of confidence by stockholders. However, insider attacks are no longer perpetrated just by hostile users, and they are no longer invisible beyond the company's borders. Mobile code attacks (e.g., the Love Bug or the QAZ Virus) that run with the authorizations of the unsuspecting user who triggered them have turned many loyal employees into *unwitting insiders*.

Perimeter firewalls were never designed to solve the insider problem, and intrawalls, which move security enforcement closer to the user, ease the problem only slightly at the cost of significantly more complex management. The problem is that, under the current state of the practice, the difficulty of managing the network security policy is impacted directly by the complexity of the network's topology.

Without a significant change, the policy management problem will only get worse. New technology trends such as mobile computing, business to business computing and client to server VPNs are challenging security administrators to rethink the way they protect their networks. These trends strain existing security infrastructures and expose the inherent shortcomings of current firewall technology. Security industry experts discussing these trends at the High Assurance Boundary Controller workshop concluded that network security enforcement should move closer to the host. This implies that we need even more firewalls, ideally one firewall for every host. But how do we possibly manage a firewall on every host given the pain of managing far fewer firewalls today?

This paper introduces a distributed firewall architecture called Network Edge Security. Network Edge Security pushes network security policy enforcement to the edge of the network (in other words, all the way to the host) to address the insider problem. Policy management, however, remains centralized. The architecture simplifies policy management significantly – and makes centralized management of per-host firewalls practical – by decoupling policy management from the topology of the network. These characteristics make Network Edge Security a viable solution for the challenges of emerging technology trends. We describe the architecture and the rationale for its design, then we illustrate its application.

2. Losing a three-legged race

Today's network engineers and security administrators are bound at the ankles in a race against determined attackers. Network topology is the tie that binds. The two teams are attempting to work together but they actually constrain each other. Network engineers provide routing of traffic, high availability, and high bandwidth. They do this in part by deciding where within the topology to create links and install routers, firewalls, etc. Security administrators ensure that only the right users are given access to various network accessible resources. They do this by deciding where within the network topology to install firewalls/filtering

Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEXII'01)
0-7695-1212-7/01 $10.00 © 2001 IEEE

routers, and what filtering rules should be implemented in each of these intermediate network devices. Any change in one area affects the other area. Computer scientists will immediately recognize this as a coupling and cohesion problem.

The following emerging technology trends expose additional weaknesses in topology-dependent policy management:

Client to Server VPNs. VPNs are an important tool for protecting our networks. Early VPNs were often implemented between a pair of gateways. These gateways were typically routers or firewalls. However, VPN clients are being widely deployed and every copy of Windows 2000 ships with VPN capability built in. The logical result is that we will soon see client to server VPNs in widespread use. This is a double edged sword. These client to server VPNs blind adversaries that might sniff data from the network. However, they also blind intermediate network security devices such as firewalls and routers that used to be capable of filtering such traffic.

Coalition and business partners. Information is valuable if you can share it with the right folks at the right time. To achieve this both DoD and the commercial world are opening up their networks to coalition partners or business partners. Providing partners with restricted access to specific systems within your network is often the only practical way to allow them to pull the information they need in a timely manner. Security administrators must support this new way of doing business. However, it is difficult because allowing partners deep into our networks blurs the distinction between insiders and outsiders. The old perimeter firewall paradigm, "assume everyone inside the firewall is a good insider and everyone outside the firewall is suspect", no longer works.

Mobile computing. Many of our network defenses assume stationary computing. That is, a particular client or server is physically connected into the network topology at a fixed point. This allows us to place intermediate network devices (filtering routers and firewalls) at strategic points within the topology to control flows between a pair of hosts. This old paradigm is being challenged by the adoption of mobile and wireless computing. Today, a user with a laptop expects to plug into the organization's network, regardless of which building they are in, and get access to the systems they are authorized to access.

The simple example in Figure 1 illustrates the problems when mobile computing is used within an organization employing topology dependent policy management. The example organization has two groups that sometimes work together but neither group allows users outside of their group to access their group level servers. Specifically, user Joe is authorized access to the blue servers but he is not authorized access to the green servers. The organization implements this policy via packet filters in Router 1 and Router 2. Router 1 implements packet filtering rules that block any traffic to/from the green subnet. Router 2 implements packet filters that block any traffic to/from the blue subnet.

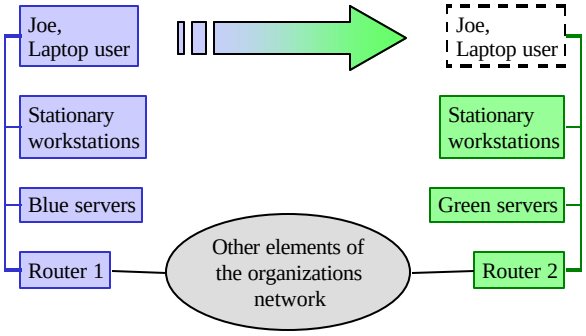

Figure 1 Mobile computing frustrates traditional network security.

Everything works fine until Joe discovers the power and efficiency of laptop computing. Joe is tasked to participate in a meeting at the green group offices. Joe takes his laptop with him and plugs into the network in the green conference room. Joe is frustrated (and ineffective) because he cannot access the blue servers, which he is authorized to access, in order to represent the blue group in the meeting. Joe must drive back to the blue building in order to answer the questions asked during the meeting.

The topology based security not only prevents mobile user Joe from accessing authorized resources. It also fails to block the threat of unauthorized access. When Joe is plugged into the green network, he could;

1. Run a network sniffer on his laptop and capture username/passwords to the green servers.
2. Use these captured passwords to gain unauthorized access to the green servers.

Security implemented in intermediate network devices and dependent upon network topology fails miserably in a mobile computing environment.

Why adopt wireless computing if users are not able to attach to the network at any point as they roam? Security technology must support mobile machines that are inside the firewall today and outside the firewall tomorrow. Or on Subnet X today and subnet Y tomorrow.

3. Topology-independent management

The key to winning the security race is to untie the network engineers from the security administrators. Since network topology is the cord that binds, we must decouple the network topology from the security policy enforcement by adopting an *end-point perspective* for security management.

Network Edge Security unties the security administrators from the network engineers by moving the PEPs (policy enforcement points) out of the network core and placing them at the edge of the network. The PEPs are then bound to the hosts, not to the network topology. This greatly simplifies security administration. For example, consider an organization with the security policy

- No machine is authorized to sniff traffic (e.g. passwords) from the network
- No machine is allowed to send packets with spoofed IP addresses
- Access rights as shown in Table 1.

Table 1 Sample security policy

Servers → --------- Clients	Human resources webserver	Engineering file server	Sales database
Human resources supervisor – Laura	FTP, HTTP		
Human resources staff - Mary	HTTP		
Engineering manager - Chris	HTTP	NFS, FTP	
Engineering consultant – Nancy		NFS, FTP	
Sales and marketing - Pat, Sam	HTTP		SQL

This trivial policy becomes non-trivial to implement using intermediate network devices within the simple topology shown in Figure 2. (This policy and network topology are grossly simplified for the sake of brevity.) The topology dependent policy implementation process generally includes the following 3 steps.

1. Specify client to server policy (e.g. Table 1).
2. The network topology including the location of clients, servers, and intermediate network devices (filtering routers and firewalls) is analyzed. Provisions are also made for mobile computers (e.g., sales team laptops) that may connect to the network at more than one point.
3. ACLs and filtering rule are created for each intermediate network device in the path between each client and each server.

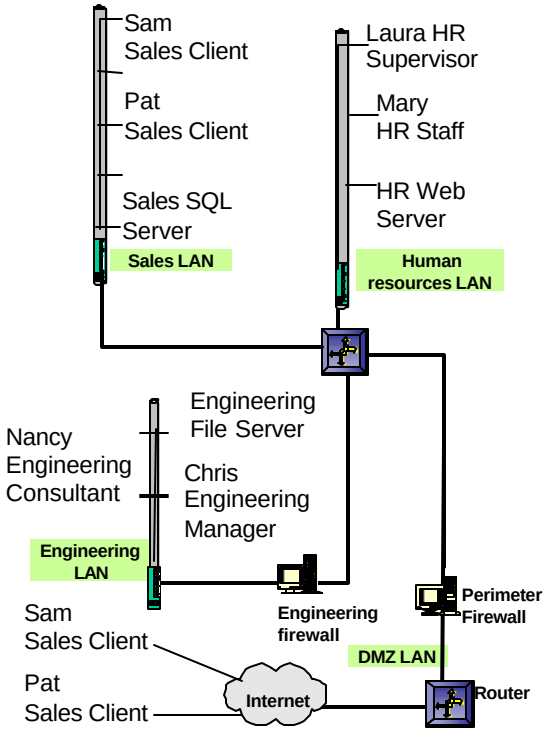

Figure 2 Simple topology for sample policy

Creating the ACLs and filter rules for the topology shown in Figure 2 is left as an exercise for the reader. Note that portions of the policy, such as no sniffing and controls within the LAN segment, cannot be enforced by the existing routers and firewalls. Now, implement the desired policy using topology independent policy management as shown in Figure 3. The implementation process generally includes the following two steps.

1. Specify client to server policy (e.g. Table 1).
2. Create filtering rules for each group of PEPs directly from the policy table. All PEPs are configured for no sniffing and no spoofing.

- **Human resources supervisor**
  - Allow FTP to {IP address of web server}
  - Allow HTTP to {IP address of web server}
- **Human resources staff**
  - Allow HTTP to {IP address of web server}
- **Engineering manager**
  - Allow HTTP to {IP address of web server}
  - Allow NFS to {IP address of file server}
  - Allow FTP to {IP address of file server}
- **Engineering consultant**
  - Allow NFS to {IP address of file server}

- Allow FTP to {IP address of file server}

- **Sales and marketing**
  - Allow HTTP to {IP address of web server}
  - Allow SQL to {IP address of SQL server}

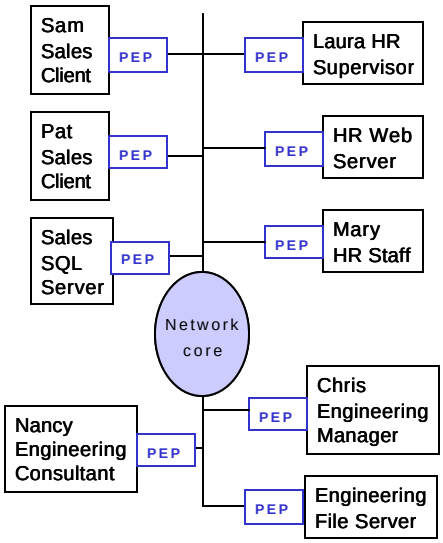

Figure 3 Topology independent security enforcement

Topology independent security enforcement unties the legs of the security administrator and network engineer. However, it opens up questions about security and scalability. Administrators now manage a relatively large number of simple policies as opposed to the complex policies managed on perimeter firewalls. It is possible to create a unique policy for each NIC. However, most organizations can significantly reduce their risk by sharing policies across groups of machines used for similar tasks. Security can be increased significantly with near zero administrative workload by simply assigning every NIC the generic "good hygiene" policy which blocks password sniffing and address spoofing.

4. The network edge security architecture

Network Edge Security is a second generation firewall architecture which changes 1st generation firewall paradigms. Imagine trying to protect your home from unauthorized access by setting up a police checkpoint on the interstate highway near your home. Of course this would be silly. However, many organizations try to protect their hosts from unauthorized access by setting up 1st generation firewalls on major network routes. Security administrators are not to be criticized for this. When the only tool you have is a hammer, everything looks like a nail. Up until now, the only tools network administrators have had are intermediate network devices.

This section will introduce the Network Edge Security architecture and show how it changes policy administration paradigms. This section also points out architectural features such as non-bypassability, tamper resistance, and scalability. The Autonomic Distributed Firewall team has implemented the PEPs using commercial Ethernet network interface cards (NICs). From this point on the paper focuses on the NIC as one instantiation of the abstract PEP.

Figure 4 shows the system architecture. The master/slave architecture provides centralized management of distributed enforcement points. The **policy server** provides all of the user interface, policy management, NIC (PEP) group management, and audit database functions. The policy server creates policy and pushes it to the NICs. The **NICs** provide packet filtering and other simple network security support functions. This approach places the complex functions at the policy server allowing the NICs to be simple, fast, and inexpensive.

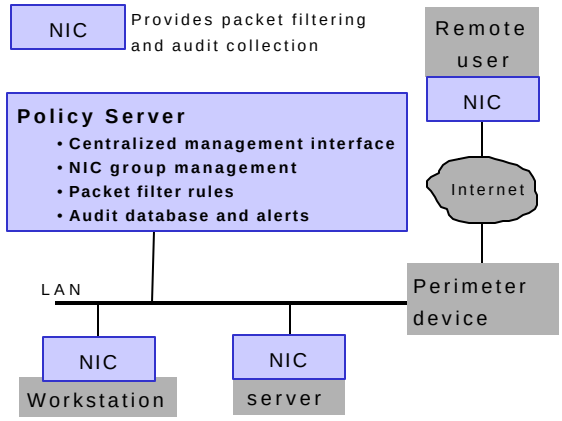

Figure 4 Network Edge Security system architecture

4.1. Network edge security is not a personal firewall

The Trusted Computer System Evaluation Criteria (A.K.A. the Orange Book) [5] spells out three fundamental truths about security mechanisms. They must be;

- Correct
- Non-bypassable
- Tamper-resistant.

"Correct" implies that the security mechanism is specified and implemented correctly. This is certainly important but it will not be addressed in this architecture level paper. "Non-bypassable" means the enforcement mechanism cannot be easily bypassed. The Network Edge Security NIC is located between the host operating system and the network. There is no path to the network except through the NIC.

"Tamper resistant" is the most interesting requirement from an architecture perspective. Some within the security community have recognized the value of moving the firewall function close to the host. The result is a growing list of commercially available software firewalls or personal firewalls. The use of these firewalls in an organizational environment raises serious questions about their tamper resistance.

Why were firewalls first built? Because commercial operating systems and applications have repeatedly shown that they are extremely difficult to secure. Applications have become more complex since firewalls were first introduced. Active-X, JavaScript, and e-mail borne viruses/worms are the current threats. Operating systems have also become more complex as more and more features are added to them. The net effect is that CERT organizations are not reducing their staffs. What happens when host based software firewalls are introduced into this environment? It is no surprise to find that they fail to meet the fundamental requirement for tamper resistance.

The DARPA Information Assurance program put one of these software firewalls to a test. The firewall was installed on a workstation and the Sandia red team began attacking it from across the network. The firewall blocked simple attacks such as port scans. The red team then crafted an e-mail message with an embedded script. This message was sent to the workstation protected by the software firewall. When the unwitting user opened the e-mail, the script executed with all of the privileges of the user. The script completely turned off the firewall but left the firewall icon on the screen. The firewall policy was completely defeated and the user was left with a false sense of security.

Ironically these software based firewalls have been disabled by other security software. Symantec's anti-virus software misidentified Network ICE's BlackICE firewall as a Trojan horse. The result: The anti-virus software disabled the firewall and left many home users' PCs vulnerable. [4] These instances of disabling software firewalls should not be a surprise. There is a circular logic underlying the architecture of host based software firewalls that cannot be corrected by the implementers writing the code. Figure 5 illustrates this circular logic. Security vendors develop firewalls to protect the host operating system. However, these host based software firewalls rely on the host operating system to protect the firewall. Does the firewall protect the operating system or does the operating system protect the firewall? The situation is even worse if a hostile user is in control of the machine.

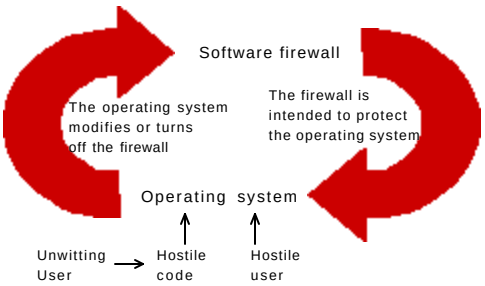

Figure 5 Host based software firewalls suffer from circular logic.

The bottom line is that while host based software firewalls offer some protection for the home user, they fall short of meeting the requirements of a security conscious organization. Personal firewalls were created largely to protect a host from the network. Network Edge Security was created to protect an organizations network from the subverted host and hostile insiders.

Network Edge Security is implemented as a remotely managed policy enforcing device that is independent (from a policy enforcement perspective) of the host operating system. The Network Edge Security NIC contains it's own processor and memory that are not accessible by the host operating system or applications. Even a hostile user at the keyboard with root/administrator access cannot force the device to change the policy established by the administrator. The NIC may be configured to send a periodic heartbeat to the policy server to reduce the threat of undetected physical tampering. The bottom line is that the Network Edge Security NIC is both non-bypassable and tamper resistant.

4.2. Scalability

Network Edge Security must scale to meet the needs of real world organizations. The management, throughput, and platforms supported must all scale up.

This is a master/slave architecture, not a client/server architecture. The key difference is that the per host firewalls are slaves to the node manager. In a client/server architecture, the clients request the server to perform a function or provide data. Within the master/slave architecture, the node manager has absolute authority and pushes policy commands to the firewall nodes (NICs) for enforcement. This

management approach supports thousands of NICs within an organization.

The performance of the Network Edge Security approach scales easily. Each time a new host is added to the network, an additional NIC processor is added to the distributed firewall. The commercial implementation allows multiple policy servers so that the policy server function also scales.

The per host firewalls devices are kept simple and host operating system independent. This allows them to be deployed across the range of hosts typical in large organizations. The NICs can protect workstations and servers on the LAN as well as remote hosts.

The following section provides additional rationale for this architecture.

4.3. Rationale for the network edge security architecture

This section focuses on the rationale for the architecture in light of the threats and issues identified in Section 1.

4.3.1. The evolving security model

First generation firewalls were developed in an environment in which the risk from the internet was considered high and the risk from insiders was considered low.

Today we recognize that the insider threat is costly even though it occurs with less frequency than attacks launched from the internet. The definition of insiders is blurred by partnerships that allow access to internal resources by users who are not members of the organization. Mobile computing also gives insiders access to networks that they could not access when their computer stayed on their desk. Finally, hostile code can turn loyal insiders into unwitting agents. These changes in the threat and risk force us to ask, where is the optimal location for firewalls.

4.3.2. Firewall location

Perimeter firewalls that separate an organization from the internet are blind to attacks launched by unwitting insiders and hostile insiders. These firewalls may provide some protection from partners that abuse their access rights unless client to server VPNs are used. These VPNs effectively blind perimeter firewalls.

It is clear that we must move the firewall functions deeper into the organization's network. Intrawalls that separate groups within an organization are a reasonable attempt to address this need. However, they suffer from the following shortcomings.

- Intrawalls, like perimeter firewalls, are blinded by client to server VPNs.
- Intrawalls do not adequately address mobile computer users that may literally carry their computer from one side of the intrawall to the other.
- Intrawalls are topology dependent. Thus, they complicate security policy management by binding security administrators to network engineers.
- Intrawalls do nothing to prevent the hostile insider from sniffing passwords.

These issues make it clear that the firewall function needs to be moved to the individual hosts. This can provide the necessary firewall functions, at the right place, as mobile/wireless computing and client to server VPNs become common.

4.3.3. Firewall implementation

How do we implement the firewall at each host? The two obvious choices are host based software firewalls and host operating system independent hardware devices.

The limitations of host based software firewalls (A.K.A. personal firewalls) have been discussed above. Software firewalls are not sufficient because they cannot provide the tamper resistance necessary to thwart hostile insiders and unwitting insiders running hostile code. In the attempt to move firewall functions from the perimeter to the host, these software implementations have gone "Over the edge". We need to pull the firewall back away from the host operating system.

The firewall implementation must be close enough to the host to be non-bypassable but independent from the host operating system. A hardware based implementation on the NIC, with its own processor and memory that cannot be accessed by the host, meets the tamper resistance requirement.

4.3.4. Management model

Given that we have concluded a hardware based firewall per host is the right architecture, how do we manage all these firewalls?

Should the firewalls be managed by individual users or by information technology professionals trained in network security? Individual users, who may be loyal to the organization, generally do not have the knowledge of network technology to properly configure a firewall. Simplified user interfaces offering "low, medium, and high" security generally

fail to address the specific functions associated with a user/host. For example;

- Should the user/host be authorized Telnet access to the servers?
- Should the user/host have unlimited access to other workstations within the organization?
- Should the user/host have access to all servers or only a specific subset of the servers?

Even the best host based user interface becomes irrelevant in the hands of a hostile insider. Naturally the trained network security staff responsible for enforcing policy should configure the firewalls.

Given that IT/IS staff should configure/enforce policy, These staffs will demand centralized/remote management. The policy server provides centralized policy specification, distribution, and auditing. Ongoing research and development is exploring ways to enhance the management capability of the system. This includes extending the system to include user authentication and other approaches to minimize the administrative workload.

5. Applying network edge security to servers shared between partners

The Network Edge Security NICs provide a low cost, high assurance tool that can be applied in innovative ways to build secure network solutions. Shared servers are discussed as an example.

Frequently partners need to share access to online data. Moving data between partners via "Sneaker net" is too slow. The pace and structure of operations today requires that partners be allowed to *interact* with our data, not simply view it. For example, vendors may need the ability to update delivery information in a logistics database. One method of providing this capability is to allow partners access into our servers. However, giving partners accounts on our servers that are connected to our internal networks introduces some risks. What if the partner abuses their access rights and gains access to other machines on our network?

Network Edge Security allows for the creation of shared servers between a pair of organizations with minimal trust between them. Figure 6 shows how to implement the set of shared servers between partners in the blue and purple organizations.

The blue organization configures their internal NIC (NIC_BI) with the following policy.

- Allow HTTP, SMTP, and SQL traffic that is initiated from the blue LAN to the server.
- Block all traffic initiated from the shared server to the blue LAN
- Block the server from sniffing any traffic flowing on the blue LAN
- The blue organization configures their external NIC (NIC_BE) with the following policy.
- Require an IPSEC VPN to NIC_PE
- Block all other connections

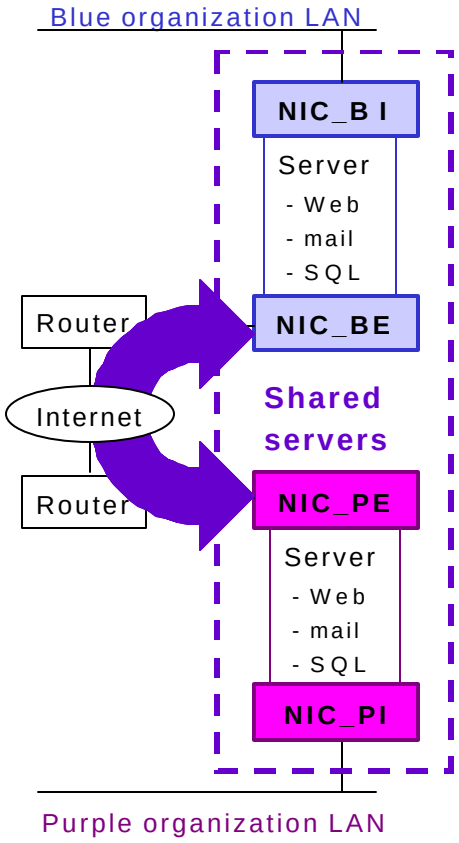

Figure 6 Shared servers between a pair of partners

The purple organization configures their NICs with the corresponding policy, replacing "blue" with "purple." This allows each organization to initiate movement of data to and from their end of the shared server. The purple organization cannot gain access to the blue LAN. Even if the purple organization is given root/administrator access to the blue server, they cannot initiate traffic to or sniff traffic from the blue LAN. The VPN (the fat arrow) between the servers ensures that no entities on the internet can view or spoof traffic between the servers. The "block all other connections" means that the external NICs connect to each other and no other machines.

The result is that each organization retains full control of their own LAN without risk that the partner can gain unauthorized access to the LAN. Each organization has a live network connection to their end

of the server. Thus, they can post data to it, manipulate it, and read it in real time. This eliminates the stale data problems associated with using sneaker-net to share information. The partner can only access data placed on the server by the respective host organization. Thus, this is much safer than actually allowing a partner into your servers, as some organizations do today.

6. Summary

Changing technologies such as IPSEC and wireless together with trends in partnering are forcing us to take a second look at the way we implement and manage firewalls. Software based firewalls are just not designed to address the insider threat. Network Edge Security represents a 2nd generation hardware firewall that addresses the existing insider threat and the emerging requirements. Network Edge Security combines topology independent policy management with strong security and scalability.

7. Acknowledgements

The authors would like to acknowledge the many useful ideas that were expressed at the Defense Advanced Research Projects Administration, Information Assurance, High Assurance Boundary Controller Workshop. This research has been supported in part through the DARPA Releasable Data Products Framework contract # F30602-99-C-0125 administered by Air Force Research Laboratory. Primary funding for this research is via the DARPA sponsored Autonomic Distributed Firewall administered by Space and Naval Warfare System Center as contract # N66001-00-C-8031.

8. References

[1] Steven M. Bellovin, smb@research.att.com, "Distributed Firewalls", *Login,* November 1999, pp. 37-39

[2] S. Bellovin, J. Smith, A. Keromytis, S. Loannidis, "Implementing a Distributed Firewall", Supported under DARPA contract F39502-99-1-0512-MOD P0001, Available at http://www.cis.upenn.edu/~angelos/Papers/df.ps

[3] Lewis Z. Koch, "Outsourcing Security", *ZDNet Interactive iWeek,* June 22, 2000, http://cgi.zdnet.com/slink?40306:2561562

[4] Robert Lemos, "Microsoft -- burned by anti-virus tools?", *ZDNet News,* Friday October 27 2000. http://www.zdnet.com/zdnn/stories/news/0,4586,2646200,00.html

[5] National Computer Security Center, 'DoD 5200.28-STD, Trusted Computer System Evaluation Criteria", Ft. Meade, MD, December, 1985