

# Observations on the effects of defense in depth on adversary behavior in cyber warfare

Dorene L. Kewley, John Lowry

?

**Abstract--** The concept of layered defense is not new in the information assurance arena, however as technologies quickly evolve, new commercial products are put on the market, and systems become more complex, the opportunities to secure different layers within a network enclave increase. The opportunities for an adversary to exploit vulnerabilities amongst the layers also has the potential to increase. A series of experiments were conducted on the DARPA Information Assurance (IA) program to explore the concept of security layers leading to not only defense in depth, but also defense in breadth.

**Index Terms—** Defense in depth, information assurance, layered defense, strategic cyber defense.

## I. INTRODUCTION

The Defense Advanced Research Projects Agency (DARPA) Information Assurance (IA) Program continues to conduct leading-edge research into defending our nation's cyber assets through strategic cyber defense. The program's overall grand hypothesis is:

“Trustworthy systems can be built from less trustworthy components.”

In other words, trustworthy systems can be composed. It is understand that individual components in themselves can be for the most part trustworthy, however when multiple components are added to a networked system, the interactions become complex and as we have learned through experimentation, new vulnerabilities are often introduced. This occurs whether the components are commercial off the shelf (COTS) products or research prototypes. It is the security vulnerabilities that the *system* is vulnerable to that the IA program has been challenged to focus on.

To explore the assurance of these systems in a manner that can be quantified, the IA program has conducted a variety of scientific experiments on specific aspects of cyber defense. Two grand hypotheses were developed to lead the experimentation:

“Layered defenses improve system assurance;” and  
“Dynamic modification of defensive structure

improves system assurance.”

The experiments presented herein explore the layered defense hypothesis. A separate series of experiments were conducted to test the dynamic defense hypothesis and are presented under separate cover [1, 2, 3].

### A. Layered Defense

Traditional thinking lends itself to the philosophy that the more layers of protection you add, the more secure a target will be. For example, you can lock the doors and windows, put a moat around the castle, put alligators in the moat and a guard dog in the yard, a fence around the moat, and an armed guard in the watchtower to create layers of physical security. In the cyber world, however, multiple layers of defense do not necessarily add together to create a higher level of assurance.

Complexity is the bane of network security. While components and systems become more complex, they become more difficult to secure. To secure something, you must thoroughly understand how it behaves, where the vulnerable holes are, and how to plug them. As operating systems and applications add features to make themselves more “user friendly” and compatible with various devices, they become infinitely complex. When these infinitely complex components begin communicating with other components in complicated ways, securing them becomes a significant challenge. A variety of security components are then required to cover the variety of vulnerabilities in a system.

The IA Program conducted three separate experiments to further understand if and how a finite set of security layers could be used to improve the assurance of a system, thus introducing more complexities.

### B. Experimentation Approach

Experimentation on the IA Program follows the Information Design Assurance Red Team (IDART) methodology which involves using a rigorous process to analyze the vulnerabilities in a cyber system, develop attack trees to attack these vulnerabilities, and in some case actually execute the attacks [4]. The red team works in close cooperation with a blue team, which serves to defend the network and has typically designed the experiment to evaluate the effectiveness of a defense strategy, as well as a white team which oversees the experiment. While flags related to an experimental scenario are carefully defined in the experiment to motivate and direct the red team to an objective, the primary goal of the experiment is not to capture the flag. The goal is to prove or disprove the experimental hypothesis. This is a different

---

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under contract number F30602-98-C-0012.

D. L. Kewley is with BBN Technologies, Arlington, VA 22209 USA (telephone: 703-284-4623, e-mail: [dkewley@bbn.com](mailto:dkewley@bbn.com)).

J. Lowry is with BBN Technologies, Cambridge, MA 02138 USA (telephone: 617-873-2435, e-mail: [jlowry@bbn.com](mailto:jlowry@bbn.com)).

approach than more traditional red teaming where the goal is often to penetrate the security of a system without the knowledge of the defenders [5, 6].

In the case of the experiments discussed herein, the red team takes on the roll of a well-resourced adversary such as a small nation state. They have the following characteristics [4]:

- ?? They are well funded and can purchase the resources and personnel that they need.
- ?? They have aggressive programs to acquire knowledge of technologies that also may provide insider access.
- ?? They use classic intelligence methods to obtain insider information and access when necessary.
- ?? They learn all design information.
- ?? They are risk averse. They make every effort to avoid detection.
- ?? They have well-defined, specific goals.
- ?? They are creative and seek out unconventional methods to achieve their goals.

The model adversary engaged in the following experiments was a team of experts from Sandia National Labs.

## II. THE EFFECT OF ADDING LAYERS

### A. Experimental Hypothesis

An experiment titled “Multi-Layer Middleware Defense” was designed as a first attempt to quantify the cumulative effect of adding layers of protection to a networked system. The hypothesis for this experiment was:

“Adding layers has at least a cumulative impact on adversary work factor.”

The goal for the experiment was to compare adversary work factors as more layers of protection were added into the client-server database architecture designed for the experiment.

### B. Configuration

A military scenario was designed to frame the experiment into an operational context. The scenario is helpful in designing a credible network infrastructure with the associated applications, realistic data flow, and an identifiable target for the adversary to attack. In this case, the scenario involved an image server critical to the mission and determined to be a likely target of attack. The image server contained photo images, radar images, and infrared images. The blue team was challenged to apply defense layers to the CORBA-based client-server applications for controlled (secure) access to the critical imagery data on the server from clients both outside and inside of the base firewall.

While eleven possible middleware layers of defense were postulated during the experiment design phase, three were actually implemented and tested during the experiment execution. A baseline was also implemented and tested as a datum for comparison. The configurations were as follows.

#### Configuration 1 – Baseline

- ?? SSL between clients and server,
- ?? packet filtering on firewalls,

- ?? plug proxy for outside clients to server to inhibit datagram-level IP address spoofing.

#### Configuration 2 – OO-DTE

- ?? Configuration 1 plus,
- ?? Object oriented–domain type enforcement (OO-DTE) to provide fine-grained (by object) role based access control to the security aware application [7].

#### Configuration 3 – MPOG

- ?? Configuration 2 plus,
- ?? Multiprotocol object gateway (MPOG) to replace plug proxy and provide fine-grained (by server port, CORBA operation, and client certificate level) role based access control to the non-security aware application.

#### Configuration 4 – Wrapper

- ?? Configuration 3 plus,
- ?? Solaris wrapper on MPOG proxy to wrap the firewall software.

These configurations are represented in Figure 1.

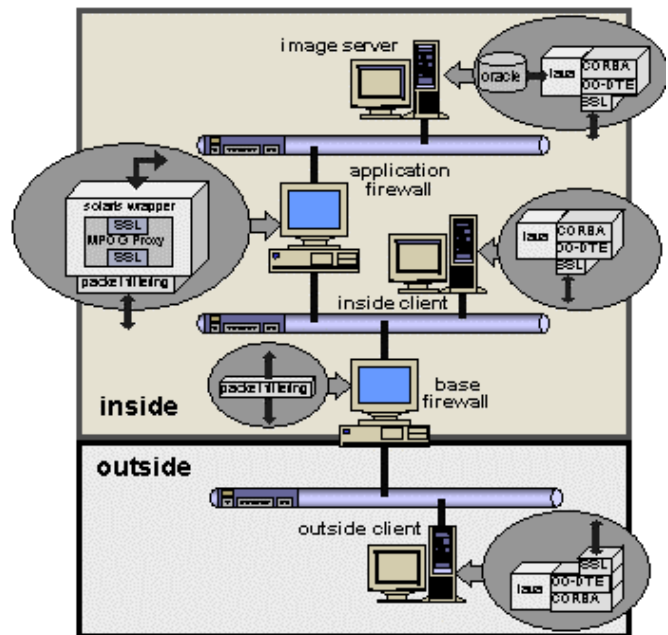


Fig. 1. Multi-layer middleware defense experiment topology and defense layers.

The red team was challenged to obtain the following confidentiality, integrity and availability flags from both outside and inside the base firewall:

**Get Image** – The adversary was challenged to exfiltrate sensitive information without being observed or detected (confidentiality).

**Modify Image** – The adversary was challenged to alter the image file that is presented to the legitimate user, without alerting the user to this fact (integrity).

**Prevent Image** – The adversary was challenged prevent legitimate users from accessing/processing images

(availability).

The adversary's work factor measured in time was the primary metric used in this experiment.

### C. Results

The expected result was that as each layer of protection was added, the adversary's work factor would increase incrementally producing a results graph that resembled a step function. The actual results were quite different. As shown in Figure 2, for the integrity flags it took the adversary the same amount of time (30 minutes) to capture the flag at configuration two as it did to capture the flag at the baseline configuration one. In other words, in this particular configuration and experiment, the OO-DTE layer did not improve the assurance of the system by increasing the adversary's work factor. In fact, the adversary was able to completely skirt the OO-DTE layer to successfully capture the flag. The results for the confidentiality flag were the same.

Adding the MPOG layer in configuration three did substantially increase the adversary's work factor from the first two configurations, however there was no additional increase in work factor when attacking configuration four with the Solaris wrappers in place. Part of the reason configuration three posed a noticeable challenge to the adversary is that the client access policy to the images was not fully enforceable until MPOG was in place on the application firewall. They were not successful when launching their attack from the LAN outside the base firewall. Insider access (still behind the application firewall but inside the base firewall) was required and granted for the adversary to be successful. In configurations three and four, it took the adversary two hours to execute the attack. Again, the results were the same for the confidentiality flag except that they spent 16 hours developing the attack as opposed to the 24 hours they spend developing the attack to capture the integrity flags.

In this particular experiment, the adversary was able to work around the security mechanisms in place at configurations two and four and successfully capture the flags. This observation lead us to conclude that defense in breadth is equally important as defense in depth.

It is interesting to note that the development time dominates the execution time. While it only took two hours to execute the attacks against the integrity flag at configurations three and four, it took 24 hours to develop the attacks. This is typical behavior for this type of risk-adverse adversary who spends a significant amount of time gathering intelligence and preparing for an attack, and a relatively small amount of time actually executing the attack [8].

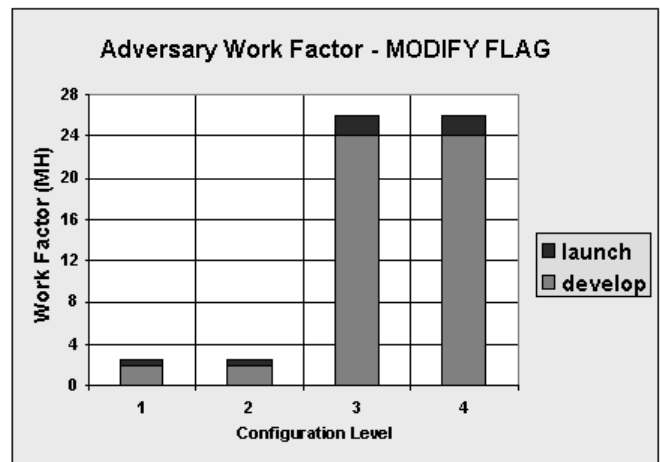


Fig. 2. Results for integrity attacks. The results for the confidentiality flags were similar.

For the availability flag, the results were quite different and even more unexpected. Denial of service (DoS) attacks are typically fairly easy for an adversary to successfully launch, however they are not stealth attacks, nor are they particularly sophisticated in most cases. The adversary was able to successfully capture the availability flag in all four configurations as shown in Figure 3. Interestingly enough, it actually took the adversary less time to capture the flag at configurations two, three, and four than it did for the baseline configuration. As the three layers of security were added, they actually introduced a new control surface for the adversary to exploit, thus making the DoS attacks trivial to them. Rather than adding to the assurance of the system, the complex interactions between the security layers were used against the blue team.

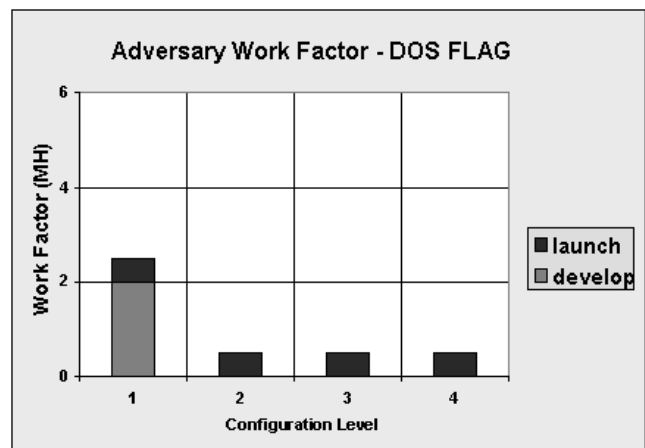


Fig. 3. Results for availability attacks.

One important conclusion of this experiment is that defense in depth without defense in breadth can be ineffective for a sophisticated adversary. This theory is demonstrated graphically in Figure 4 where if specific classes of attacks are not defended against, a hole in the brick wall occurs, leaving a



**Layer 2**

?? Firewall proxying for HTTP.

**Layer 3**

?? IPsec encapsulated security payload (ESP) in tunnel mode between the clients and the firewall [10].

**Layer 4**

?? Strong authentication at the firewall.

The red team was once again challenged to obtain confidentiality, integrity, and availability flags as follows:

**Confidentiality Flag**

?? Get image from server

?? Get image intended for client

**Integrity Flag**

?? Modify image on server

?? Modify image destined for client

**Availability Flag**

?? Attack server availability

?? Deny image to client

The red team developed extensive attack trees that were analyzed by both the red and blue teams together in a whiteboard session prior to the attack execution week. During that meeting, it was decided which attacks were expected to yield the most useful results in proving or disproving the experimental hypothesis that the security layers would increase the adversary's work factor. Again, the red team was constrained to attacks that went through the layers. Under normal circumstances, adversaries would use multiple methods and paths to attain their goals. The attacks on the prioritized list were further developed in the lab by the red team prior to execution.

**C. Results**

Execution of the attacks was limited to one week, less the set up and breakdown times. Based on the time limitations and the attack priority list, only the following layers were attacked:

?? Layer 0 – baseline;

?? Layer 1 – filtering router;

?? Layer 3 – IPsec between the clients and firewall;

?? Layers 1+3 – filtering router and firewall with IPsec.

It was observed prior to the experiment execution that the addition of layer 2, firewall proxying, would not in fact add a depth layer because IPsec encrypted packets bypass all firewall proxies when the firewall is configured in “trusted” mode as it was in this experiment. Therefore no attacks were run with layer 2 configured. Layer 4 was not engaged during the experiment as it was a lower priority.

At layer 0, the red team was able to readily capture the confidentiality and integrity flags against the server by exploiting vulnerabilities in the configuration of the remote data service (RDS) on the image server. Once exploited, they were able to both capture and relocate the images.

At layer 1, the red team was again able to readily capture the confidentiality and integrity flags. Because the filtering router

was now in place, the red team had to first spoof an authorized IP address prior to launching the same RDS exploit as used at layer 0. In this case, the addition of the filtering router did increase the adversary's work factor in both the preparation time and the execution time as they had to add a notable step to their attack.

The red team was less successful with their attacks at layer 3 with IPsec in place. They were not able to capture either the confidentiality or integrity flags in a reasonable amount of time. With layers 1+3 in place, the red team was again unsuccessful in capturing the flag, however it was not possible to measure the increase in work factor that the combination of layers presented. It can be stated that the combination of the Cisco filtering router and IPsec ESP on the firewall proved to be a sufficiently challenging set of layers for the red team [9].

The red team was successful in launching multiple denial of service attacks at both layers 0 and layer 1 using three separate exploits. They first exploited the RDS vulnerability as they did when capturing the confidentiality and integrity flags, then used a tool to kill the HTTP process on the server, and also created ICMP redirects to redirect packets destined for the server. Consistent with previous results, the addition of the filtering router did add to the adversary's work factor.

At layer 3, once again the red team was thwarted by IPsec when they launched the denial of service attack against the image server.

It can be concluded that the addition of the two layers (filtering router and IPsec) did in fact increase the adversary's work factor in capturing the flag. In particular, the IPsec layer provided the most robust boundary layer protection in this particular network configuration.

When defense in depth is concentrated in one area as it was at the boundary in this experiment, it is natural for the unconstrained adversary to search for weaker access points than the well-protected device. While it was shown that the layers did add defense in depth at the boundary, the clients remain vulnerable and would be a likely attack point given the layers of obstacles that they encountered at the boundary. Once again, the importance of depth in breadth is emphasized.

#### IV. THE EFFECT OF LAYERS ON ADVERSARY COURSE OF ACTION

An experiment to explore adversary planning and courses of action (CoA) was executed in the Summer of 2000. The experiment construct offered an opportunity to explore layered defense and how it affects the adversary [11].

**A. Experimental Hypothesis**

Traditional network security has focused the defenses at the enclave boundary through the use of firewalls. Many adversaries have found creative ways to penetrate the enclave protection to reach the vulnerable internal networks. Adversaries have also been successful at compromising client machines not protected by corporate firewalls, often connected to digital subscriber lines (DSL) at their homes. New commercial technologies have recently become available which help to secure the endpoints at reasonable costs. These

technologies include personal host based firewalls and IPsec virtual private networks (VPN). While enclave firewalls are often cost prohibitive to single users and small companies, the personal firewalls are inexpensive, and the newer operating systems such as Windows 2000™ and Solaris 8 now ship with IPsec. Protecting the endpoints with such promising technologies offers a new opportunity for adding layers of protection in both depth and breadth to the network enclave.

The hypothesis for this experiment was:

*“Distributed protection mechanisms with value-driven policies can be composed to prevent or deter unauthorized access.”*

The question remains as to what the appropriate placement of protection layers is, and how the layers can be combined to achieve increased assurance without introducing new vulnerabilities into the network.

The issue of combining layers presents two interesting observations. The first, and most obvious, observation has to do with effectiveness and cost of the layers. The extreme examples are to put a) all protection at the boundary, or b) all protection within the enclave. Examples of a) are certain classified networks that depend on very strong boundary controls but which have little or no protection within the network or at the hosts. Examples of b) do not seem to exist with the possible exception of Cheswick’s network [12] that presumably continues to operate with no significant or catastrophic security failures. The question of where the appropriate balance of layers between the two extreme examples remains.

The second observation regarding combination of layers has to do with policy and adversary behavior. It seems reasonable to believe that many classes of adversary presume that the boundary controller represents the union of all requirements from within the boundary. The configuration of the firewall is defined by meeting the needs of most (if not all) of the enclave members and consequently instantiates the most restrictive policy that is acceptable to all. Clearly this policy does not meet the needs of certain members and applications. Now that host-based firewalls and VPNs are available, there is no reason why more restrictive policies cannot be established for particular hosts, applications, or network segments within the enclave. An application domain (e.g. accounting or payroll) may wish to establish more restrictive policy domains. The first benefit from finer-grained policy to the sub-domains may be an increased resistance to attack as the sub-domain now has some control over vulnerability reduction.

A variety of internal policy enforcement mechanisms that are more restrictive than the single boundary policy could have interesting effects on the detection and behavior of internal and external attackers. The external attacker starts by trying to inventory and map the enclave based on traffic analysis of protocols seen traversing the boundary controller and mapping the boundary controller configuration. As stated earlier, it seems reasonable that some classes of adversary will leap to the conclusion that this configuration represents the most restrictive policy. This behavior would be observed as a ‘lack of caution’ once the adversary decides to look beyond or

penetrate the boundary. Mapping the interior behind the boundary may expose the existence of the adversary to more restrictive policy domains within.

When the adversary penetrates and possibly subverts a machine to further their purposes (misuse), they become a kind of insider who may be difficult to distinguish from a valid user. However, the adversary may have difficulty carrying out his mission depending on the policy enforcement mechanisms in place at sub-domains. Adversaries may have tools that expect the use of symmetric filtering by the defender. While this is often the case, a careful analysis of usage and proper configuration of tools can make detection of the adversary easier. If the adversary is detected and the defensive posture is improved, there is a possibility that the adversary can be constrained to operate only within the compromised sub-domain.

The risk of discovery and potential of being thwarted may force behavior changes on the adversary. Perhaps they will find that the benefit does not justify the cost. Perhaps the attack preparation phase will be extended. Perhaps the discovery phase will become more complex creating a greater risk of being detected.

### B. Configuration

The scenario for this experiment was revitalized from a previous experiment [2] and involved deployed peacekeeping forces that needed to order refresh and resupply items from the Defense Logistics Agency on a regular basis. These orders were placed using a standard web browser which connects to an Oracle database running on a Windows 2000™ Professional machine using a web front-end with CGI scripts. A consultant working from the Internet connects in to the database to run non-root database operations in order to improve the overall performance of the database. He connects by either using telnet over an SSH tunnel or using a host to host VPN.

Four phases of layers were developed to facilitate the exploration of the experimental hypothesis:

**Phase 1** – Enclave to enclave VPNs using IPsec in tunnel mode. Proxying at the firewalls.

**Phase 2** – In addition to the enclave to enclave VPNs and proxying firewalls, personal firewalls exist on all Windows 2000™ hosts. The personal firewalls were configured with least privileged access.

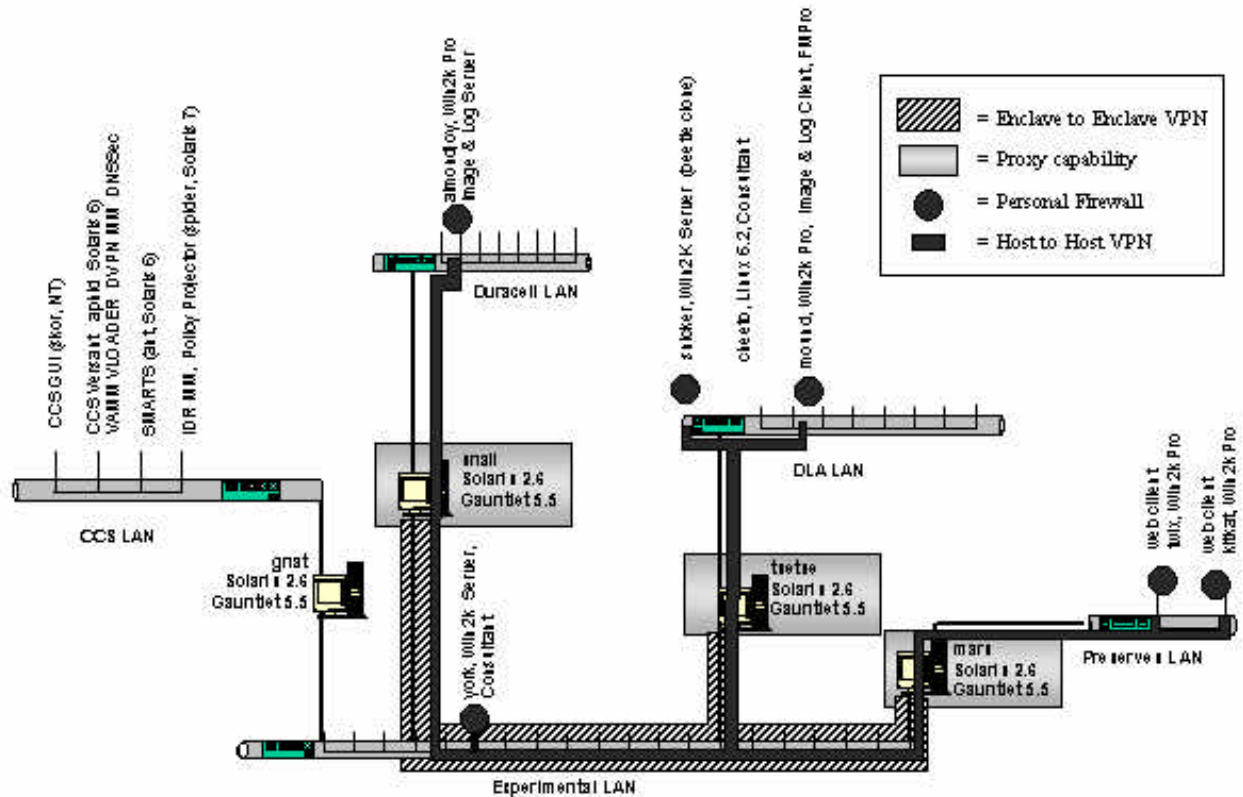


Fig. 6. Network topology and defensive layer construction.

**Phase 3** - In addition to the enclave to enclave VPNs and proxying firewalls, host to host VPNs were integrated using IPsec for all mission critical communication paths.

**Phase 4** - Enclave to enclave VPNs, proxying at the firewalls, personal firewalls on all Windows 2000™ hosts, and host to host VPNs.

The network topology and defense layers are shown in Figure 6. The network also included a variety of intrusion detection systems configured to monitor adverse activity.

The red team was challenged to achieve three flags:

- ?? modify orders in a subtle way;
- ?? delete orders in a subtle way;
- ?? add orders in a subtle way.

Prior to the experiment execution, the red team developed attack trees with weightings of “low,” “medium,” and “high” for the following attributes at each decision point along their attack paths:

- ?? perceived risk
- ?? cost to develop and implement,
- ?? likelihood of failure.

The attack paths to be executed were selected prior to the experiment execution based on lower weightings through the path. Deviations from this path based on defenses encountered were closely monitored to observe the adversary’s decision making process and course of action selection.

### C. Results

The results from the attacks executed by the red team during

the execution week are presented in Table 1.

TABLE 1. ATTACK RESULTS

	Phase 1	Phase 2	Phase 3
Add Flag	no flag	--	--
Modify Flag	conceded	captured	captured
Delete Flag	not captured, captured	--	captured

The red team had difficulty penetrating the enclave firewalls, so the consultant host became the primary target for several of the attacks. The personal firewall implemented in Phase 2 proved to be a notable obstacle for the adversary. It proved to make the host nearly invisible to scans, causing the adversary to move into noisier and riskier information gathering stages. The information gathering activities were noted by the personal firewalls, and made available to a centralized security monitoring station. The personal firewall was defeated by the use of a hostile e-mail which killed the firewall process without alerting the user. The adversary was then able to fully exploit the client host and capture a variety of modify flags. The existence of the personal firewall layer did add a challenging layer on top of the boundary firewall that the adversary had to defeat.

For the attacks on Phase 3, the existence of host to host VPNs using IPsec did not prove to be an additional challenge to the red team over the enclave VPNs. Because of the architecture of this particular network, and this particular scenario, the external consultant provided a virtual back door into the internal network. Once the consultant machine had been compromised, the red team was able to pass through the boundary firewall using IPsec as a valid user and gain access to the internal consultant machine. Once inside the network,

they did not have to continue their intelligence gathering as the critical information was now at their fingertips on the internal consultant machine.

No attacks were executed against Phase 4 because both the red team and the blue team agreed that the results would be similar to previous results based on the execution of the attacks at Phases 1, 2, and 3. The adversary was expected to have the same success rate and exploit the same vulnerabilities, thus the combination of IPsec and personal firewalls together did not create defense in depth in this case. Again, this was due to the vulnerable nature of the consultant, and the red team's inclination to take the path of least resistance (when not artificially constrained to attack directly through the layers).

For this particular network, it can be concluded that the boundary firewall and the personal firewall layers did in fact add to increase the adversary's work factor in achieving their goal. The inclusion of IPsec *in this case* provided an avenue for the adversary to exploit and thus get through the boundary firewall. It is expected that in other network configurations, the use of IPsec would in fact add a layer of security. Future experiments would need to be executed to prove this.

## V. CONCLUSIONS

Adding security layers to a system does not necessarily guarantee increased assurance. Introducing new layers of security has the potential to introduce new vulnerabilities, or control surfaces, for sophisticated adversaries to exploit. Defensive layers must be analyzed to gain a thorough understanding of how they work together before they are integrated into an operational system. As shown in previous experiments, it is sufficiently challenging to select independent defensive layers that do not interact with other layers in unusual and unexpected ways. A careful addition of layers can increase the assurance of a system by increasing the adversary's work factor if careful attention is paid to not only defense in depth, but also defense in breadth. All classes of attacks must be defended against, otherwise the adversary will likely circumvent the defenses that are integrated.

## VI. REFERENCES

- [1] D. Kewley, R. Fink, J. Lowry, and M. Dean, "Dynamic Approaches to Thwart Adversary Intelligence Gathering," accepted for publication in *Proceedings of the DARPA Information Survivability Conference & Exposition II*, June 2001.
- [2] D. Kewley, J. Bouchard, "DARPA Information Assurance Program dynamic defense experiment summary," accepted for publication in *IEEE SMC Special Journal Issue*, July 2001.
- [3] G. Schudel, B. Wood, "Adversary Work Factor as a Metric for Information Assurance," *ACM New Security Paradigm Workshop*, Cork, Ireland, September 18-22, 2000.
- [4] B. Wood and R. Duggan, "Red Teaming of Advanced Information Assurance Concepts," *Proceedings of the DARPA Information Survivability Conference & Exposition Volume II of II*, Jan. 2000.
- [5] D. Parker, *Fighting Computer Crime*, Wiley and Sons, NY, pp. 393-395, 1999.
- [6] P.D. Goldis, "Questions and Answers about Tiger Teams," *ED-PACS, The EDP Audit, Control, and Security Newsletter*, vol. 27, no. 4, pp. 1-10, 1999.
- [7] D. Sterne, G. Tally, C. McDonell, D. Sherman, D. Sames, P. Pasturel, E. Sebes, "Scalable Access Control for Distributed Object Systems," *8<sup>th</sup> Usenix Security Symposium*, Washington, D.C., August 1999.
- [8] G. Schudel, B. Wood, "DARPA Information Assurance Red Team Experiments," *Proceedings of MILCOM 2000*, Los Angeles, CA, October 2000.
- [9] D. Johnson, L. Benzinger, "Layering Boundary Protections: An Experiment in Information Assurance," *Proceedings of the 16<sup>th</sup> Annual Computer Security Applications Conference*, 2000, p. 60-66.
- [10] S. Kent, R. Atkinson, "IP Encapsulating Security Payload (ESP)" *IETF RFC 2406*, November 1998.
- [11] J. Lowry, "An Initial Foray into Understanding Adversary Planning and Courses of Action," accepted for publication in *Proceedings of the DARPA Information Survivability Conference & Exposition II*, June 2001.
- [12] S. Bellovin, W. Cheswick, "Network Firewalls," *IEEE Communications*, September 1994.