ACTIVE RESPONSE

A Thesis
Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a
Major in Computer Science
in the
College of Graduate Studies
University of Idaho

by
Sergio Paul Caltagirone

May 2005

Major Professor: Deborah Frincke, Ph.D.

AUTHORIZATION TO SUBMIT THESIS

This Thesis of Sergio Paul Caltagirone, submitted for the degree of Master of Science with a major in Computer Science and titled " Active Response," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor ______________________________ Date __________
Dr. Deborah Frincke

Committee Member ______________________________ Date __________
Dr. Paul Oman

Committee Member ______________________________ Date __________
Dr. Steffen Werner

Department Administrator ______________________________ Date __________
Dr. Robert Hiromoto

Discipline's College Dean ______________________________ Date __________
Dr. Charles Peterson

Final Approval and Acceptance by the College of Graduate Studies

______________________________ Date __________
Dr. Margrit Von Braun

ABSTRACT

Active response has been long misunderstood and misclassified as a strike-back methodology. This work greatly expands the scope of active response in order to facilitate an open discussion of response methodologies. This discussion must include the multitude of problems facing response such as the ethical and legal concerns, and the technical feasibility of a response. This work attempts to frame this discussion by providing a taxonomy of response actions, a model illustrating the process of response, and a decision model to determine an appropriate response. These enhancements to the topic of response, as well as an implementation of response using evolutionary techniques, illustrate the potential of active response as a legitimate security tool in the protection of vulnerable assets.

VITA

Sergio Caltagirone is a graduate of the University of Portland in Portland, Oregon with degrees in computer science and theology. His interests are wide and varied, including human-computer interaction, comparative theology, ethics, and law. He worked for four years as a developer on large-scale web applications in the Portland/Vancouver area. He enjoys bicycle racing, computer gaming and is an avid reader of social science fiction.

ACKNOWLEDGMENTS

I would like to thank many people for their support during the development of this work. My parents, Paul and Jean, and my brother, Dante, for their support throughout my life and educational career; and Sherrie for her great patience during my late night writing binges. I would also like to thank Dr. Deborah Frincke for her invaluable guidance and judgement throughout this process. Stan Gotshall, Ryan Blue, Sean Melton, Doug Shikashio, and Paul Ortman also provided great discussion through the years on active response and its implications — thank you all very much.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

Chapter 1

INTRODUCTION

A computer security threat to most organizations is not a small event. While most organizations may experience the loss as financial, with the loss of productivity and data (and potentially the trust of investors), the threat to some organizations go far beyond the realm of financial risk. Take for example a stock exchange, although its primary function is financial transactions, there are greater diplomatic and national security interests at stake. Additionally, hospitals, national security organizations, Supervisory Control And Data Acquisition (SCADA) systems (e.g. infrastructure control systems), air traffic control, and other vital systems also measure risk in units other than financial (e.g. ethical, legal, national security, etc.) and exact great cost if successfully attacked.

The threats to these systems are well documented [1–3] and sometimes successful, as in the cases of a regional air traffic control system [4], a nuclear power safety control system [5] and a hospital records system [6]. There is no argument that these systems should be strongly protected with traditional defenses such as Intrusion Detection Systems (IDS), firewalls, and better designs. But what happens if a threat bypasses these defenses? Do we unplug? What if the system provides services that are more valuable than the system itself (e.g. life/safety/national security critical systems) and cannot be disconnected? Do we do nothing and allow the attacker continued access? What if the attacker ends up shutting down the system in the end?

Some would consider the idea of allowing the attacker to continue in any system abhorrent and initialize a strike-back [7,8]. Others vehemently reject strike-back in any scenario [9]. The problem with these positions is that they are not supported by anything other than emotion or retributive justice. These types of arguments have caused response to be viewed as a strike-back only methodology. There must be another way; response must be more than 'to strike-back or not to strike-back'. This thesis explores the breadth of response and proposes a definition, taxonomy, and decision model of response among other contributions.

1.1 Problem Statement

A defender cannot choose whether to respond to a known security event as deciding not to take action is a response in every sense of the term. However, since any action or inaction is a response, what is an appropriate set of actions to take during a security event in order to mitigate the threat given the immense social and technical considerations of response?

Notice that we are not searching for the *best* set of actions, but merely an appropriate set. This distinction is important because determining the *best* set of actions requires an analysis that is not feasible at this point in time given the state of the art in computer security, ethics, and law.

This problem, while discussed in a few settings, such as the Agora Workshop in Seattle and the Active Response Continuum Workshops, has never been pursued in formal venues. As described in Chapter 2, the amount of research directly on the topic has discussed a very limited aspect of response, strike-back, or discussed it in terms of only one viewpoint, such as international law. While this type of research is very valuable for the discussion of response, there has not yet been a holistic view taken of the topic, consolidating all of the the view points and resources and presenting an overall model for response and its many issues. This is what this work attempts to accomplish.

1.2 Motivation

There is a very pragmatic motivation to solve this problem: whether we agree with it or not, system administrators have been utilizing active response for over 20 years and will continue to do so — especially given the greater variety of active response tools becoming available by companies such as Lycos [10] and IBM [11]. It is therefore imperative that research on the topic begin to inform these users of the implications of assuming an active response position — and provide a positive framework with which to develop a more effective dialog on the topic.

Additionally, a greater number of life/safety/national security critical systems are being networked and becoming vulnerable to threats. While providing the standard defenses (e.g. firewalls and intrusion detection systems) for these systems is both prudent and necessary, an additional tool is needed to protect these systems in the event that an attacker is able to circumvent these defenses. Currently, if an intruder is able to circumvent defenses and enter a system, there is no policy governing the actions to take in order to mitigate the threat. The best advise at the current time is from computer forensics, and that advise is to disconnect the system, either by removing network connectivity or shutting the system down. However, some systems cannot do either (e.g. medical databases, air traffic control, etc.), and in that case, a policy is needed to determine the best actions to take. These are the motivations for attempting to solve the problem of active response.

1.3 Methodology

To accomplish the task of determining appropriate active response actions, several methods are employed. The first is an implementation of a competitive co-evolutionary genetic algorithm which attempts to develop reasonable active response strategies to a multitude of threats. The genetic algorithm provides a 'proof-of-concept' that it is possible to model active response using a genetic algorithm and that it is possible to develop reasonable active response strategies. Additionally, a model of the response process and a formal decision model is developed and then evaluated against a set of criteria. With this methodology, a general framework for active response is produced.

1.4 Organization

This thesis is divided into four major chapters: the introduction, the background, the approach and results, and the conclusion. The introduction provides the reader with a clear understanding of the motivations and goals of the work and the background places the work in a historical and research context while providing the reader with an understanding of necessary concepts. The third chapter discusses the research presented in the three papers and their results.

Unlike a traditional monolithic thesis, this work revolves around three separate papers which address different aspects of response. While segregated, these papers share a similar theme and overarching motivation: the understanding and analysis of response in a computer security scenario. The last chapter, the conclusion, provides a summary of this work and a summary of the contributions made in the three central papers.

Chapter 2

BACKGROUND AND PREVIOUS RESEARCH

Active response [1] does not have the research foundation enjoyed by other topics. In fact, very little research has directly referenced active response itself. Several reasons exist to explain the lack of extensive research in this topic. First and foremost, it is a difficult topic to address because of the difficulty in calculating risks and categorizing threats. Second, active response can be a potentially dangerous topic; unintended consequences can occur if the wrong system is targeted for defense leading to additional risk or international incidents. Third, some active response actions are morally ambiguous, which some researchers find uncomfortable. Fourth, there are fundamental questions still to research. These, and other reasons, have kept researchers from the topic of active response, leaving a lack of research.

And although there is a lack of active response research, this does not mean there is no background research to discuss. Active response is a topic that rests at the intersection of multiple diverse disciplines, such as ethics, law, and computer security. All of these academic fields, in addition to others, provide fundamental research which active response draws upon. These fields, although not directly referencing active response, contribute indirectly and importantly.

This section will focus on both the limited research directly addressing active response and the extensive research that is indirectly connected to the topic. There are seven fundamental research areas that will be discussed: computer security,

[1] 'Active response' and 'active defense' will be treated as synonymous terms within this work, although the discrepancy between these terms is properly noted in Section B.4.2.

active response, intrusion detection, information warfare, ethics, law (both domestic and international), and evolutionary computation. Several pieces of important research will be presented in each area. The research presented in the following subsections will then be discussed with regard to its contribution to the topic of active response.

Some of these areas have extensive research histories. Therefore, not all nuances of the research can be presented or discussed, but a genuine effort will be made to introduce the reader to important topics and theories when necessary for the purpose of a complete understanding — especially when it has direct relevance to active response.

2.1 Computer Security

2.1.1 A Definition of Computer Security

"Broadly speaking, security is keeping anyone from doing things you do not want them to do to, with, or from your computers or any peripherals."

— William Cheswick and Steven Bellovin [12]

"Computer security is the protection of the integrity, availability, and, if needed, confidentiality of automated information and the resources used to enter, store, process, and communicate it."

— An Introduction to Computer Security: The NIST Handbook [13]

2.1.2 The Computer Security Background

In the 1950s, with the coming of age of the punchcard systems and large scale use of computers, the primary concern with regard to security was the physical protection of the systems. This was due to the limited access and single user aspect of the systems. In the mid-1960s, time-sharing systems, which allowed multiple users and absent-user processes, were developed and introduced [14].

Because of the nature of these early systems, security to protect users from other users and provide administrator access was necessary. However, even in this environment, the designers of these systems were still well aware of the potential future social implications that would arise due to this technology [15].

In the 1970s, techniques of engineering and guaranteeing security through the use of security requirements and categories were researched [16]. A major outcome of this push was the so called 'Orange Book' or the *Trusted Computer Systems Evaluation Criteria* (TCSEC) [17,18]. These, and other methods such as provably secure systems [19], were developed in an attempt to prevent users from escalating privileges (i.e. executing or accessing programs they have not been given permission to) and establishing covert channels (i.e. unallowed communication between entities).

Active research has continued in these avenues, however in the 1980s and continuing presently, the computer security community has been increasingly focused on the vulnerabilities and external threats to computer systems. Although these issues were not completely disregarded previous to the 80s, the immediacy of the issues were not felt until network connectivity expanded and computers were threatened by hackers (then called crackers), worms, and viruses. [2]

In the 1980s, there were several works that foreshadowed the impending crisis that would envelop networked computers. First, a thesis was written by Lt. Meyers describing a trojan (i.e. a malicious program masquerading as a legitimate program) being launched by a remote attacker [20]. Second, in 1983, a graduate student named Fred Cohen demonstrated the first computer virus (i.e. "a program that can 'infect' other programs by modifying them to include a possible evolved copy of itself") [21]. The growing understanding of threats to networked

[2] Those who maliciously trespass into computer systems, or escalate existing privilege will be herein referred to as 'hackers' rather than the classical term 'crackers' — however these terms are now synonymous and should be read as such.

computer systems culminated with the release of the first worm in 1988.

The 'Morris' worm , named after its author Robert Morris, son of the director of the National Security Agency (NSA) and doctoral student at Cornell University, took the computer security community by suprise [22]. The worm seriously disrupted computer service throughout the then-young 'Internet' by utilizing enormous (by 1988 standards) amounts of bandwidth in order to propagate. This event was so significant in the increased awareness of computer security that the Defense Advanced Research Projects (DARPA) created the Computer Emergency Response Team (CERT) [3] as "...a central switchboard and coordinator for computer security emergencies...", which is its continuing mission in 2005 [22].

Closely following the worms and viruses of the 1980s, the hackers soon took center stage. A number of cases brought hackers into the light in both the computer security research community and the media. In 1988, Clifford Stoll published his account of tracking and finally capturing German hackers attempting to access restricted United States government computers [23]. Following Stoll was Cheswick's similar account of watching and studying the activity of hackers in a system [24]. These reports were followed up in a paper by Bellovin, which described the tools and techniques that computer criminals employ in an attempt to compromise systems [25]. However, these scholarly and very exciting papers took a backseat to the media attention paid to one particular hacker, Kevin Mitnick [26]. At that point, 'hacker' entered the popular lexicon, and the threat that computer criminals posed became evident to the general public.

Computer security researchers have not just sat idly by watching these threats take control of networks. They have gone out and created a multitude of tools with which to protect systems and networks from these threats (with mixed

[3] http://www.cert.org

results). There has been serious research in intrusion detection systems (as described in a following section), as well as in firewalls (i.e. systems that filter network traffic based on rules) [12, 27], anti-virus protection (i.e. programs that detect malicious programs based on unusual behavior or a known signature) [28], and security requirements and design [16] . These preventative measures have become a large industry because of their necessity in the current environment.

Despite all of these advances in computer security research during the past 50 years, and the tools developed therein, it can be argued that, if using economic damage as an indicator, worms, viruses and hackers have far outpaced the defensive tools in their ability to compromise and damage systems [29]. When one enters the field of computer security, it seems like the Internet is a large, anonymous, free-for-all where some get caught but most do not, in a modern version of the 'Wild West'. So it is in this environment, plagued by infestations of worms, pandemics of viruses, and vandals waiting in every dark alley that active response enters. And although it may not be the saving grace, or the 'silver bullet,' of security, it may help stem the growing tide of lawless destruction currently overcoming our systems and threatening the up-and-coming information infrastructure.

2.2 Active Response

As stated before, there is a very limited amount of literature directly referencing active response. Most of the literature discusses the ramifications of hack-back [4] or other offensive techniques. In 2002, Jayawal, Yurcik and Doss called for more effective ways of protecting networked systems from attack and examined the possibility of hack-back [30]. In the same year, Mullen presented justifications for

[4] Hack-back: retaliating against the attacker using techniques that share many attack characteristics (a.k.a. strike-back or counter-hack).

strike-back at defcon [31], wrote a corresponding article at SecurityFocus [8] and published a whitepaper on strike-back [7].

Mullen's work inspired a Reuters news article about strike-back, especially the ethical and legal implications [32]. Researchers such as Schneier criticized Mullen's position, using an analogy to the Recording Industry Association of America's (RIAA) attempt to attack copyright infringer's computers [9]. Mullen responded in [33]. However, active response considers a greater range of actions than only hack-back.

Some researchers emphasize the response decision-making process. Loomis, in [34], while implying hack-back, presents an objective discussion of the ethical and legal aspects of response decisions. Grove, et al., distinguishes 'active defense' from 'passive defense' and undertakes a thorough discussion of international law implications of active response [35]. Bruschi and Rosti discuss a response strategy for denial of service (DoS) attacks in which they limit the capabilities of the attacker rather than strengthening defenses [36]. Additionally, they provide AngeL, an implementation of their strategy [37,38].

2.2.1 The Technical Aspects of Active Response

Works such as ADAM [39], Loomis in [34], and Grove et al. in [35] discuss the nebulous question of: which situations, if ever, it is proper to respond? Some, like ADAM [39], go even further and attempt to suggest actions to take in a response scenario. However, these works abstract the larger question of what response actions are technically feasible. It may be that an organization would like to respond, knows the actions they should take, but cannot take those actions because they are not technically able to. There are two primary aspects to a response technology: the identification of the threat and the ability to respond (there are other questions, but those are discussed further with Intrusion

Detection in Section 2.3).

The Identification of the Threat

Active response is about mitigating a threat. A threat must have materialized from some point, so identifying the source of the threat can make some types of actions available. However, finding the source of a threat is sometimes difficult, if not close to impossible, because of network layer abstractions and the agent/handler model of attack [40]. The Internet, and modern networks in general, are designed in a distributed fashion; that is, that information and processing occurs in more than one place. The agent/handler model takes advantage of this distributed system. The agent/handler model is when an attacker compromises a machine and then uses that machine to launch an attack against another machine.

Take for example one of the most difficult attacks on networked systems, the Distributed Denial of Service (DDoS) attack. One popular model utilized by attackers is agent/handler, where an attacker compromises multiple machines (sometimes on the order of thousands). Some of those machines are designated *handlers*, others are *agents* [40]. From a single point, the attacker can control the handlers, which in turn, control their agents. In a DDoS attack, the purpose of the agents are to send network data to a victim. In general, a server can handle normal network traffic behavior, but because there are so many agents, all of the network data overwhelms the victim, causing their network queue to fill and to stop processing legitimate requests. The difficulty in mitigating a DDoS attack is that there will be too many agents to stop; to catch the actual attacker requires going through at least two layers of abstraction (through the agents, through at least one level of handlers, and then on to the attacker).

If the handlers and agents used in the attack are controlled by the victim, then all the victim has to do is to look through the host logs of the agents and handlers to determine the source of their compromise and trace it back (still not an easy task). But this is rarely the case, and attackers spread the handlers and agents between multiple networks (sometimes the sheer number of agents necessary to overcome a server requires the use of machines outside one network). Therefore, in the case of the DDoS attack, there are two primary network-centric methods of searching for the true source of the attack: one is to utilize the functionality of the network equipment comprising modern networks, the other is to elicit cooperation from other network providers so that they can trace the attacker through their resources, but will not be discussed here.

Some network equipment providers are beginning to develop technology that traces the source of an attack through a network (including the Internet). Cisco Systems, a manufacturer of networking equipment, has implemented an Internet Protocol (IP) tracker, which allows a victim to trace traffic from a Denial of Service (DoS) back to its source through the network equipment [41]. Additionally, research in the area of tracing and tracking has led to some significant improvements in methodology. One method that received particular attention is Kohno, Brodio and Claffy's work in producing a unique identifier for each machine based on clock-skew [42]. This means that because each machine has a unique amount that the clock is off from official time, the packets can be traced back to the machine with a particular clock-skew. Additionally, [43] and [44] have identified methods of using network equipment (e.g. routers and switches) to mark packets (i.e. a watermark) so that the path of the packet can be traced back to the source. These are promising steps forward, but there is still no 'silver bullet' in the hunt for the attackers.

The Ability to Respond

Assuming away the need to obtain the identity of the attacker (or assuming that the identity is available) allows for a wide variety of actions to take place in defense of a system. Active response actions can be placed into two broad categories: those having effect within the organization's boundaries and those having effect outside it's boundaries.[5] We will discuss each category and the actions associated with the category separately.

There are a large number of actions that can be taken within the boundaries of an organization. The first, and most obvious, is to simply remove the asset from being a target. This can be done by either turning off the power or segregating it on the network. Additionally, the source IP, source port, or other traffic characteristics can be used to block the offending traffic at the firewall. If the attack is known and has an Intrusion Detection System (IDS) signature, and if an IDS is running in inline mode [6] the IDS can deny the malicious packets.

The use of a honeypot [7] is also an internal measure to collect intelligence on an attacker. One internal method to help prevent DoS and DDoS attacks is to use the zombie control process against the attacker. Because a zombie (i.e. a DoS or DDoS agent) does not care about the origin of its control signal, a system administrator can send the 'kill' or 'stop' signal to all of the machines within a network (in case they may be infected zombies) in order to stop the zombies from launching an attack. A novel method of internal response is IBAN, presented in [46]. IBAN is a network agent that contains IDS rules. When a traffic pattern matches one of its

[5] This categorization is broad and only for discussion purposes. Additionally, one cannot make judgements regarding the actions based solely on category because contacting the FBI would be considered an outside action and is considered by most to be very legitimate.

[6] Inline mode, is when an IDS controls traffic to the network based on its internal rule-set.

[7] A honeypot is a non-production system attached to the network with the purpose of being attacked or compromised so that an organization can gain intelligence on those attacking them [45].

rules, it blocks traffic from the supposedly compromised host to other hosts in order to prevent further compromise. This only comprises some of the actions that can be taken with an organization's boundaries, now actions outside of those boundaries will be examined.

Compared to the number of actions available to a defender inside their own boundaries, the actions which reach outside of an organization's boundaries are far fewer. One external action which should be stressed to a greater extent is the notification of the FBI, CERT, Secret Service, or other organizations that handle computer crime and computer security related events. These organizations, while the response may be delayed, can provide additional resources external to an organization that they would not have access to otherwise. Some low-level and low-risk external response actions would also include sending TCP RST packets that close an attacker's TCP connection, stopping communication for a short time.

Other external actions can be geared towards intelligence gathering. Utilities such as ping, finger, trace-route, nmap, and other tools can be used to determine the true location of the attacker and their system characteristics and specifications. This intelligence can be used in even stronger defensive actions, for example a defender could write a worm that searches for a specific IP or Media Access Control (MAC) address in order to exploit and disable the machine. A defender could also conduct a hack-back, where the defender compromises the attacker's machine in order to disable it from further attacks.

Meer, Temmingh, and Walt in [47] propose passive strike-back as an option. Passive strike-back involves "camouflage, disinformation, misdirection, obfuscation, and proportional response" [47]. They provide some examples that are helpful and useful active response measures such as sending false terminal signals so that the attacker will execute any code (on their machine) the defender wishes. These are some of the many examples of active response actions that are possible given the state-of-the-art of computer science.

2.3 Intrusion Detection

In order to respond to an attack, there must be a method of gaining knowledge of the attack. While many attacks are still caught by watchful system operators, Intrusion Detection Systems (IDS) have become a popular technology to assist in the detection process. An intrusion detection system is a technology whose purpose is "to identify, preferably in real time, unauthorized use, misuse, and abuse of computer systems" [48].

In 1980, James Anderson in [49] was the first to identify the use of audit trails in tracking system misuse. Later, Denning in [50] developed the first complete IDS model that is still being used as the basis of current research. Contemporary IDSs fall into two categories: misuse (i.e. signature based) and anomaly. Misuse-based IDSs uses signatures of known threats and match patterns based on those signatures to determine whether an action is misuse [51]. Anomaly-based IDSs attempt to distinguish normal user behavior from non-normal behavior in order to determine the classification of an event — this can be done using a number of statistical models (e.g. threshold, mean/standard deviation, Markov Processes, etc.) [51].

Additionally, an IDS must receive data from one or more sources in order to evaluate user actions and events. Anderson's paper [49] describes analyzing audit trails from host systems. While in 1990, Heberlein et al. presented the Network Security Monitor (NSM) which detects intrusions from network traffic [52]. This demonstrates the two categories of intrusion detection systems by data source: host-based and network-based IDSs. However, some modern systems are hybrids and collect data from both sources in order to make a determination regarding a suspicious event [48].

Regardless of the type of IDS used, response requires the determination that a threat has materialized, and more importantly, the source, type and victim of the attack. Intrusion detection systems, apart from detecting potential illegitimate behavior, also identify evidence of an attack, providing a greater range of data regarding the threat so that a more accurate assessment can be completed [51]. Additionally, an IDS can assign confidence values to the evidence collected, allowing an assessment of the threat to take into account statistical measures of risk [50]. Overall, intrusion detection systems cannot be completely decoupled from response models because a response would probably depend on the data being identified and gathered by an IDS.[8]

2.4 Information Warfare

Information warfare and active response are sometimes confused. It is true that they both share the potential of using information systems in an offensive manner, but that is where the similarities end. Information warfare is defined as using information technology to gain a “military advantage using tactics of destruction, denial, exploitation, and/or deception” [53]. This differs from active response in that active response is concerned only with actions which might successfully mitigate a threat. Active response does attempt to gain an advantage, only to return to a previous security state.

This significant difference, however, does not mean that information warfare research is inapplicable to active response. Information warfare is one of the few topics which actually discuss a method of response in a threat-filled environment. And in some situations, active response can be argued to subsume information warfare. If anything, information warfare can inform active response of the abundance of threats and corresponding responses.

An example of a threat is the North Korean ‘hacking army,’ which according to

[8] In [51], Kemmerer and Vigna suggest that response is an integral part of intrusion detection systems, but do not go any further than to acknowledge the potential of response and some of the problems.

CNET [54] is intimidating Australian businesses as well as businesses elsewhere with the threat of economic sabotage and espionage. In [55], Paul Joyal describes an intense history of industrial espionage beginning with the fourteenth century and dating to contemporary industrial espionage tactics using intelligence agencies and information technology.

Industrial and economic infrastructure is only one area where information warfare can be utilized. There are also a growing number of military targets which are susceptible to information warfare because of the attempt to network these systems and share information. Weapons and defense systems could be potentially compromised by worms, viruses, and hackers rendering them useless — or even more dangerously, turn against their own military. Of course this is only hypothetical, but the fundamental and underlying principles about such an attack are sound, and simply because such an attack has not been documented does not mean that the threat does not exist. For example, a worm is generally unknown to exist until it is released in the wild and computer users begin to become affected — an attack for which there is no defense is called a zero-day exploit.

More important than defining the threats in an information warfare scenario, is what information warfare research says on the potential response to a threat — in other words, defensive operations. Denning, in [56] identifies six areas of defense: prevention, deterrence, indications and warnings, detection, emergency preparedness, and response. According to Denning, the indications and warning category provides for the ability to respond in the early stages of an attack. However, this ability to thwart the attack is diminished, as Denning states, by the fact that “attacks over computer and telecommunications systems can take place at lightning speeds” [56]. The response category provided is more accurately described as forensics/repair, where the damage is assessed and action is taken to prevent future attacks. Denning places more emphasis on the ability to protect systems through the improvement of design and requirements, and the use of cryptography, rather than other types of defensive actions. Erbschloe echoes Denning in [57], in that preventative defenses are to be relied on almost exclusively for protection from information warfare threats.

Yurcik in [53, 58] provides the clearest indication of which actions are deemed appropriate responses to an information warfare scenario. He goes as far as saying that if threatened, or attacked, a nation-state, such as the United States may be able to employ force (either kinetic or technological) to protect itself within the given boundaries of the UN Charter. Barkham agrees, but goes even further in his analysis of the *jus ad bellum* (just war) doctrine in [59] to provide greater support for the use of force in the event of a cyber attack. Importantly, Col. Cabana in [60] disagrees and argues that the military should have *only* a support role in the protection of civilian infrastructure during a cyber attack. The Colonel argues civilian law enforcement is better equipped to handle such an event and that civilian authority takes precedence over military power — drawing on the analogy of border patrol and protection.

Even with this short discussion on information warfare and its application to active response, there is clearly an ongoing debate on the proper response to a large-scale or devastating cyber attack on the United States or other nation-states. Active response will utilize the debate of these researchers to further the discussion of the boundaries of response, particularly with regards to international laws and treaties, and apply that to an active response strategy.

2.5 Ethics

One of the largest debates regarding active response is its ethicalness and whether at any time it is ethical to engage in offensive action for the protection of information systems. This section attempts to introduce the reader to the very old and distinguished research topic of ethics so that a discussion of its implications to active response actions can be engaged.

2.5.1 An Introduction to Ethics

The general idea of ethics is simple: it is to find general principles or explanations of morality or moral systems — behavior or action that is "right" or "good" [61]. Additionally, not all situations require ethics, such as whether one should spread their toast with grape jelly or strawberry jelly. This situation does not involve an ethical question because all choices are neither right nor wrong, they are simply preferences.

However, when situations of preference are discarded, ethical questions remain. One of the most important principles in ethics is universality [62]. This states that if an ethic declares an action right or wrong given a situation, it must also declare the action right or wrong given another situation with similar circumstances. In other words, it must be consistent when faced with similar circumstances — one action cannot be right for one person, but wrong for another in the same situation.

Another important principle is ethical justification. When one is asked why they made such an ethical decision, the person cannot respond with a statement like, "because it felt right." Rather, the decision must be justified using reason that refers to an underlying moral principle, such as "in general, lying is wrong."

The reason there are so many ethical theories is because philosophers "have varied considerably in their opinions regarding the basic ethical concepts" [63]. This is not a disagreement on what is good or bad, but rather what is the highest good. Some philosophers have determined that the consequences of an action are more important than the action itself, while other have argued the opposite.

This short introduction should be sound enough to provide the reader with a context to consider the following theories and their application to active response. For a more thorough introduction to ethics, see [63] and [64].

The theories presented in this section present a representative cross-section of the variety of ethical theories available. These particular theories were chosen for their popularity or direct application to the study of active response. Secondly, John Rawls in [65] and Nancy Davis in [66] both argue that teleological and deontological theories “exhaust the possibilities regarding theories of right action” [65]. However, this is not a complete discussion of all applicable theories; it will be left to other philosophers and scientists to discuss additional theories in the context of active response.

2.5.2 Teleological Ethics (Consequentialism)

Teleological ethics (e.g. consequentialism) is the category of ethical theories that are primarily concerned with the consequences of an action rather than the action itself. More specifically, it “assesses the rightness or wrongness of actions in terms of the value of their consequences” [67]. The value of a consequence is based on the amount of ‘goodness’ that is produced. Many theories have alternating views of what constitutes ‘goodness.’ Utilitarianism, a popular teleological theory, maintains that goodness is measured in happiness — therefore the act (or rule) is right if it maximizes happiness.

Consequentialism, as applied to active response, produces an easy method of determining the ethicalness of a case of active response. The only question needed to be asked is whether the active response action produced results of greater good than would have been produced by any other action. Given this criterion, the only missing component is a definition of ‘good.’ However, this criterion shows the critical flaw of consequentialism — calculating ethicalness [68].

To successfully apply pure consequentialism, it is necessary to examine all consequences of an action. Additionally, it is necessary to assign costs to the consequences so that they can be weighed against other consequences. However, it can be easily argued that one can never enumerate all of the consequences of the action (including ones far in the future) — this is made more difficult with the requirement that these must be known before the action is taken.

Of course, on the surface, consequentialism seems simple enough to apply to active response — simply enumerate the potential consequences of the active response action and compare that to other actions. But active response faces a problem: the interconnectedness of information systems. If the active response action chosen is to launch a denial of service (DoS) attack against a target, but the DoS attack also affects a router that handles pharmaceutical transactions for a hospital, which causes one or more people to die, then the action was probably not ethically justified. This example should logically lead to the conclusion that although enumerating consequences in the physical world are difficult enough, enumerating them in the virtual world are even more difficult because of the interconnectedness of the systems.

With this problem, consequentialism can be difficult to implement successfully, especially in the virtual environment. But, although this is a difficult problem, consequentialism does offer active response at least one benefit. As opposed to other ethical theories, consequentialism is not primarily concerned with honoring values, but rather promoting them. Therefore, an organization could use consequentialism as a justification for promoting certain values — although the benefits of the actions must still outweigh the costs [68]. However, to use consequentialism in active response would still require a tremendous leap by calculating many potential consequences that may not be able to be detected before the action is taken. Therefore, if consequentialism is used to justify an active response action, the organization is burdened with a tremendous, if not impossible, task.

2.5.3 Deontological

Deontology is an ethical theory popularized in the form of Kantian ethics [69]. Deontology is concerned with the actions themselves, as opposed to the consequences. To a deontologist, there are several ethical duties that we must abide by and there are several kinds of actions that are intrinsically right or wrong. Therefore, by placing a particular action in one of these categories, it is possible to ascertain the rightness or wrongness of that action based on the value of the category.

One of the most important aspects of deontology is how an action is actually determined to be right or wrong. To a deontologist, an action is right if everyone can do the same action without terrible consequences. Therefore, deontology defines lying to be wrong because if everybody lied, then there would be no difference between fact and fiction, which would lead to a logical contradiction.

This is an important theory to active response, because many ethical judgments about active response are argued from this theory. Most opponents of active response argue that all cases of hack-back and preemptive strikes (which compose a subset of potential active defense actions) are wrong [9,70]. Therefore, any action that constitutes an unauthorized use or disruption in a computer service is wrong. Deontologists would say that this is a correct rule because if everyone 'actively responded,' then it would cause a detriment to society.

In particular, Eugene Spafford argues in [70] from a deontological approach that all hacking is wrong (except in extreme or rare cases). He first dismisses any consequentialist forms by claiming it is impossible to measure the ′goodness′ and that consequentialism can justify, or even require, monstrous acts such as murder. Spafford justifies a deontological position by providing counter-arguments to hacker claims such as freedom of information and social protection, followed by appealing to the reader′s sense of vulnerability that computers systems operate in many life critical environments. He argues that ethically justifying hackers would be a mistake.

Smith, Yurcik and Doss argue against Spafford's thesis in [71] and propose a "security justification." They state that because networked systems have become so riddled with security holes, and that some of these systems operate business and life critical applications, there is room in an ethical understanding for so-called ethical hacking. They propose a definition: "Ethical hacking is fixing a system by compromising it — destructive testing in other domains — which has a long history of achievement. . . " [71]. They argue that to improve system security, since it has not improved by itself, we must turn to 'penetration testing.' Furthermore, because of the increase in security that the use of this type of hacking brings, it is ethically justified.

Spafford and deontology have an important point when it comes to hacking in general. If we followed the deontologist's view of how to classify kinds of acts, then yes, hacking would be unethical because it would be dangerous to society if everybody did it. However, even Spafford makes an important admission that in some cases, although unethical, hacking may be required to prevent a greater wrong — such as in loss of life. But he contends that these situations have not occurred, or are too rare to worry about.

However, the argument between Yurcik and Spafford is really an age old dilemma, that between the teleological and deontological frameworks. Yurcik is arguing (from the teleological) that penetration testing, while ethically questionable, can be used to improve the greater security of our systems — and therefore the greater security outweighs the short term hack. Spafford would deny any type of penetration testing because it falls within the category of hacking. These two positions cannot be reconciled and the debate will continue from both camps in the future. However, it is important that in any ethical discussion, it is clear which framework is being chosen, teleological or deontological, and that both sides provide adequate justifications.

2.6 Law

Most of the literature produced regarding topics applicable to active response is on the subject of law and concerns either the legality of an active response policy or the justification of an active response action.

2.6.1 An Introduction to Law

To understand the legal discussion, it is necessary to present a short introduction of a few legal concepts. This section has been partially adapted from [72].

First of all, it is important to note that the discussion will use concepts primarily from the Model Penal Code (MPC). The Model Penal Code was developed by The American Law Institute in 1962 in an effort to assist legislators reconcile the criminal statute with contemporary legal understanding [73]. Since its inception, the MPC has been a strong influence in the redevelopment of both federal and state statutes. In fact, many statutes simply mirror the text of the MPC. Therefore, for our purposes, the MPC will serve as our penal code, but when serious inconsistencies exist (or a particular statute does not exist), federal and particular state statutes will be examined as well.

An offense is made up of three parts (which can contain zero or more elements), the actus reus, the attendant circumstances, and the result. Additionally, mens rea is attributed to each element of the offense. The actus reus is the voluntary action (or omission) of a person, which causes a result. An attendant circumstance is simply a fact surrounding an event. The result of an offense is simply that, result of conduct of an actor.

More abstract than the actual elements of the offense is the concept of mens rea. Mens rea is the actor's mental state with respect to each element. Important in the concept of mens rea is that each element of the offense may have a different level of mens rea. Therefore each element's mens rea requirement must be inspected. The MPC makes it clear that for an actor to be guilty of an offense, they must have acted either purposefully, knowingly, recklessly, or negligently with respect to each element of the offense. Where the MPC defines these as:

Purposefully *(MPC 2.02(2)(a))*: "A person acts purposefully with respect to each element of the offense when . . . (i) it is his conscious object to engage in conduct of that nature or to cause such a result; and (ii) if the element involves the attendant circumstances, he is aware of the existence of such circumstances or he believes or hopes that they exist" [73].

Knowingly *(MPC 2.02(3)(b))*: "A person acts knowingly with respect to a material element of an offense when . . . (i) he is aware that his conduct is of the nature that such circumstances exist; and (ii) he is aware that it is practically certain that his conduct will cause such a result" [73].

Recklessly *(MPC 2.02(3)(c))*: "A person acts recklessly with respect to a material element of an offense when he consciously disregards a substantial and unjustifiable risk that the material element exists or will result from his conduct" [73].

Negligently *(MPC 2.02(3)(d))*: "A person acts negligently with respect to a material element of an offense when he should be aware of a substantial and unjustifiable risk that the material element exists or will result from his conduct" [73].

Additionally, the mens rea definitions have an explicit hierarchy, such that Purpose > Knowledge > Reckless > Negligent. If negligent is the requirement, then it is sufficient if recklessness, knowledge, or purpose can be shown, and if knowledge is required, then purpose or knowledge suffices, etc.

2.6.2 United States Domestic Law

An active response occurring within the boundaries of the United States, whether it is legal or not, is covered under United States domestic case law and statutes. It is for that reason that statutes and case law must be discussed. This section should provide a description of the legal boundaries that active response must operate within.

The Computer Fraud and Misuse Act (18 USC 1030)

In 1986, federal legislators realized the need to formally criminalize computer misuse. Prior to 1986, any computer misuse was prosecuted, if at all, under theft and fraud offenses. However, the theft and fraud statutes were not designed to prosecute offenders who were not stealing data or entering systems fraudulently — although some argued otherwise. To remedy the situation, the US Congress formalized a statute that would make computer misuse its own offense: the Computer Fraud and Abuse Act (18 USC 1030). It has gone through several revisions since its inception in 1986 and its current form was adopted in 1999.

For our purposes, the important parts of the code are:

(a) Whoever . . . (2) intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains

(A) information contained in a financial record of a financial institution, or of a card issuer as defined in section 1602(n) of title 15, or contained in a file of a consumer reporting agency on a consumer, as such terms are defined in the Fair Credit Reporting Act (15 U.S.C. 1681 et seq.);

(B) information from any department or agency of the United States; or

(C) information from any protected computer if the conduct involved an interstate or foreign communication;

(a)(5)

(A)(i) Knowingly causes the transmission of a program, information, code, or command, and as a result of such conduct, intentionally causes damage without authorization, to a protected computer;

(B) by conduct described in clause (i), (ii), or (iii) of subparagraph (A), caused (or, in the case of an attempted offense, would, if completed, have caused) - (i) loss to 1 or more persons during any 1-year period (and, for purposes of an investigation, prosecution, or other proceeding brought by the United States only, loss resulting from a related course of conduct affecting 1 or more other protected computers) aggregating at least $5,000 in value; (ii) the modification or impairment, or potential modification or impairment, of the medical examination, diagnosis, treatment, or care of 1 or more individuals;

(e)(2) A "protected computer" means a computer (A) exclusively for the use of a financial institution or the United States Government, or, in the case of a computer not exclusively for such use, used by or for a financial institution or the United States Government and the conduct constituting the offense affects that use by or for the financial institution or the Government; or

(e)(4) A "financial institution" means (A) an institution, with deposits insured by the Federal Deposit Insurance Corporation;

(e)(6) the term "exceeds authorized access" means to access a computer with authorization and to use such access to obtain or alter information in the computer that the accessor is not entitled so to obtain or alter;

(e)(8) The term "damage" means any impairment to the integrity or availability of data, a program, a system, or information;

As can be seen, 18 USC 1030 is very limited in its effect on computer crime. It limits itself to primarily government and financial systems (with some protection of medical records). Because of this, most state and federal prosecutors have used the Wire Fraud statutes when other systems are attacked. However, the application of these statutes become difficult when the hacker acquires property of little financial value — or simply enters a system without damage.

Significant Case Law Regarding 18 USC 1030

United States v. Sullivan, *40 Fed. Appx. 740 (2002)*: In (e)(8), damage need not be "actual destruction of a computer system" [74].

United States v. Czubinski, *106 F.3d 1069 (1997)*: Simply browsing taxpayer's information without intent to commit fraud (nothing "more than to satisfy curiosity") is not criminal under 18 USC 1030 [75].

United States v. Middleton, *231 F.3d 1207 (2000)*: The court found that the word "person" (5)(B)(i) refers to corporations as well as individuals. Also, the court found that the $5000 damage threshold necessary to find criminal liability, can be met using "any loss that you find was a natural or foreseeable result of any damage that occurred," including costs "to restore the data, program, system, or information that you find was damaged or what measures were reasonably necessary to re-secure the data, program, system, or information from further damage" [76].

United States v. Morris, *928 F.2d 504 (1991)*: Intention is only necessary with regard to unauthorized access and not damages caused, pursuant to (a)(5)(A) [77].

United States v. Sablan, *92 F.3d 865 (1996)*: Upheld the Morris decision in that the mens rea requirement of intention is only applied to unauthorized access, which meets constitutional standards for the statute [78].

Shurgard Storage Centers, Inc. v. Safeguard Self Storage, Inc., *199 F.Supp.2d 1121, 1128 (W.D. Wash. 2000)*: The court greatly expanded the understanding of "protected computers" to mean any computer connected to the Internet. It also broadens the meaning of "exceeding authorized access" to using pre-existing authorization to "obtain or alter information in the computer that the accessor is not entitled so to obtain or alter" [79].

Computer Trespass Statutes

As the reader can see, the current federal law does not protect systems that are not governmental, financial or medical. As the court in *Shurgard* noted, Congress purposefully limited itself to computers where "there is compelling federal interest," and left the states to enact their own laws [79]. As such, several states have created separate computer crime laws that are much more generally applicable to most computer hacking. These state statutes make it a crime to intentionally access a computer in an unauthorized manner. A few of the state statutes are given here for comparison.

Rhode Island (§11-52-3): *Whoever, intentionally and without authorization, directly or indirectly, accesses, alters, damages, or destroys any computer, computer system, computer network, computer software, computer program, or data contained in a computer, computer system, computer program, or computer network shall be guilty of a felony…*

Virginia Computer Crime Act (§18.2-152.5): *A person is guilty of the crime of computer invasion of privacy when he uses a computer or computer network and intentionally examines without authority any employment, salary, credit or any other financial or personal information relating to any other person. "Examination" under this section requires the offender to review the information relating to any other person after the time at which the offender knows or should know that he is without authority to view the information displayed.*

The University of Dayton School of Law has drafted a Model State Computer Crimes Code for the purpose of assisting state legislators with the design of their own computer crime statutes [80].

Model State Computer Crime Code (§4.01.1): *(A) No person shall purposely, knowingly or recklessly gain access to or cause access to be gained to any computer, computer system, computer network, computer program, computer data base, or computer* *material without the express or implied authorization of the owner or an agent of the owner empowered to authorize access to the computer, computer system, computer network, computer program, computer data base, or computer material. Any person who violates this section is guilty of the crime of computer trespass* [80].

Necessity Defense

There are some who have posited informally that an active response action may be justified using a necessity defense. These claims are usually supported by a belief that active response is used in situations where there is no alternative to breaking the law. The law anticipates these circumstances and so devised the necessity defense: when faced with a number of horrible choices, a person may choose the best one regardless of legality. However, the necessity defense does not remove the legal liability of the actor, they are still guilty of the offense. Rather, the law has decided that it serves no purpose to punish in these cases.

The Model Penal Code developed the Choice of Evils justification for the purpose of providing a justification in a case of necessity.

MPC §3.02: (1) Conduct which the actor believes to be necessary to avoid a harm or evil to himself of another is justifiable, provided that: a. The harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged; and b. Neither the Code nor other law defining the offense provides exceptions or defenses dealing with the specific situation involved; and c. A legislative purpose to exclude the justification does not otherwise plainly appear. (2) When the actor was reckless or negligent in bringing about the situation requiring a choice of harms or evils or in appraising the necessity for his conduct, the justification afforded by this section is unavailable in a prosecution for any offense for which recklessness or negligence, as the case may be, suffices to establish culpability [73].

There are a few of interesting points in the statute provided by the American

Law Institute. First, there must be a quantifiable benefit to breaking the law (§(1)(a)). Upon prosecution, it must be quantifiably proven that the choice was the best one. Second, the actor cannot have been reckless or negligent in bringing about the situation — they cannot be the instigators (§(2)). Additionally, the justification is not available to offenses where recklessness or negligent is the mens rea requirement. These specifications are provided to guarantee that the necessity defense will only be utilized in situations where society does not wish to punish a person for their choice.

There are two approaches to a necessity defense statute. The first is that of the Model Penal Code, the second is illustrated by New York State. The New York statute requires that the conduct must be such that it qualifies as "an emergency measure to avoid an imminent . . . injury" (N.Y. Penal Law §35.05 (2002)). This requirement forces the defendant to prove the imminence of the harm to be avoided and removes the defense from those who act to stop great future harms.

Significant Case Law Regarding The Necessity Defense

United States v. Schoon, *971 F.2d 193 (1992)*: To invoke the defense, the defendant must show that: (1) they were faced with a choice of evils and chose the lesser evil; (2) they acted to prevent imminent harm; (3) they reasonably anticipated a direct causal relationship between their conduct and the harm to be averted; and (4) they had no legal alternatives to violating the law [81].

Use of Force in Defense of Property

A question that is often asked in regards to active response is whether force can be used in defense of computer systems. Although this specific question has not been adjudicated or decided by legislatures, we can examine already existing statutes in search of some instruction with regards to this issue.

While each state will have their own unique language, there will be many concepts which are common among them all. In Utah (§76-2-406 (2002)), a person may use non-deadly force to prevent or terminate criminal interference with real property or personal property. In North Dakota (§12.1-05-06 (2001)), "force is justified if it used to prevent or terminate an unlawful entry or other trespass in or upon premises, or to prevent . . . damaging of property." However, the statute requires that the person using force must first request the perpetrator to desist, but the request is not necessary if it would be useless, dangerous, or "damage would occur before the request could be effectively made."

These statutes may be able to be coupled with the computer trespass statutes, whereby 'criminal interference' may be satisfied if the computer trespass statute was satisfied. Additionally, pursuant to US v. Middleton [76], the term damage is very widely defined. Barnes notes that some 'defense of property' statutes exclude the use of force if there is a "substantial risk of serious bodily harm," which in the case of active response does not exist [82].

2.6.3 International Law

While most will have a clear understanding of the connection between active response and United States domestic law, there is a slightly more abstract relationship between active response and international law.[9] Because standard international boundary delimiters are not applicable on the Internet, an active response scenario may include the attackers (which may or may not be known to the defenders) existing in another country. If the attacker's systems are harmed during the defense of a system, then that may constitute hostile action or a 'use of force' against the state in which the attacker resides. This could lead to a

[9] The term international law in this context does not refer to the domestic laws of other nations, but rather the treaties, councils, and agreements between nations which govern their actions with other countries. It also includes widely accepted international relations concepts that may or may not be codified in a treaty.

diplomatic problem between the states, or even worse, the outbreak of war.

However, for a state to legitimately retaliate against another state due to a cyber threat/attack, a 'use of force' and an 'armed attack' must have occurred. The definition of 'use of force' and 'armed attack' is determined by the United Nations (UN) Charter [83] and its interpretation. The problem is that the UN Charter was created to protect against kinetic threats and extending these doctrines into a virtual environment such as the Internet is difficult.

There are three parts of the UN Charter that are particularly interesting in this regard: Articles 2(4) [10], 39 [11], and 51 [12]. Article 2(4) prohibits the threat or use of force by states. Article 39 reserves the exclusive right of the Security Council to authorize the use of force at a threshold lower than a state's right to utilize self-defense. Article 51 preserves the rights of the member states to use self-defense in the event of an 'armed attack'. But what is a 'threat or use of force' and an 'armed attack' on the Internet?

In all cases, the four principles of armed conflict must be adhered to:

- **Necessity**: There must be no other alternative but to enter into armed conflict.

- **Proportionality**: Civilian losses must be proportional to the military

[10] "All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations" [83].

[11] "The Security Council shall determine the existence of any threat to the peace, breach of the peace, or act of aggression and shall make recommendations, or decide what measures shall be taken in accordance with Articles 41 and 42, to maintain or restore international peace and security" [83].

[12] "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security. Measures taken by Members in the exercise of this right of self-defence shall be immediately reported to the Security Council and shall not in any way affect the authority and responsibility of the Security Council under the present Charter to take at any time such action as it deems necessary in order to maintain or restore international peace and security" [83].

advantage gained.

- **Discrimination**: The best effort must be made to discriminate between civilian and militant actors.

- **Perfidy**: "Acts inviting the confidence of an adversary to lead him to believe that he is entitled to, or is obliged to accord, protection under the rules of international law applicable in armed conflict, with intent to betray that confidence" (Geneva Convention, Article 37 of Protocol I).

Several researchers have commented on this question and the application of these articles to active response, information warfare, and similar activities. Yurcik in [58] and Yurcik and Doss in [53] suggest that pursuing the "ability to use force on the Internet is lawful and a fundamentally important aspect of deterrence and international peace and security." They provide a framework from which to begin determining the appropriate policy to purse these activities, and question whether offensive counter measures (e.g. hack-back) are appropriate or reasonable.

Grove et al. in [35] provides a detailed and thorough analysis of information warfare applications. Grove et al. suggests three possible approaches to a cyber attack: attempt to have the source state locate the attacker, take covert action, or take overt action to eliminate threat. The problem with these approaches, as Grove points out, is that the source state may be unwilling, espionage (covert operations) is prohibited in the source state (but not prohibited by international law), and overt action may constitute an act of war. Grove suggests that bilateral or multilateral treaties (e.g. The Council of Europe Convention on Cybercrime) may help improve the cooperation between states on this issue and reduce the amount of covert or overt action necessary.

However, if overt action is necessary, Grove states that there is a "sizeable grey area between actions that clearly constitute a use of force, and those that clearly do not" [35]. Grove goes on to provide six alternatives in responding to an information attack: in-kind, real-time disabling, and physical response, technical sanctions, preemptive response, and considered response.

The first is an in-kind response, or the use of non-force to response to non-force. The problem with this is that the response may go through neutral countries, thus violating their neutrality. Additionally, under Article 51 members are required to report actions of self-defense to the Security Council. A real-time disabling response is one where the non-force threat is disabled. However, care needs to be taken that the response is not a use of force as the use of force is never allowed in response to a non-force threat. A physical response is the use of kinetic force in response to a threat in cyberspace. The legality of this is in question and the principle of proportionality needs to be taken into consideration. Technical sanctions are the use of electronic boycotts or blockades preventing communication from one area to another. A preemptive response is one which occurs before the damage of the attack has occurred; this theory has not received wide spread support. The last response is a considered response, where a counter-attack is launched. Barkham warns against an automated counter-attack as this is a matter of state, and humans need to be kept in the loop throughout the process.

Barkham in [59] provides a very thorough analysis of the 'use of force' doctrine in international law as applied to information warfare, particularly Article 2(4) of the UN Charter. Barkham concludes that Article 2(4) does not prevent states from using information warfare techniques in a manner that does not destroy life or property and techniques such as subversion, electronic blockades, and incursions are permitted under Article 2(4). Barkham sees this as a very large loophole that allows states too much flexibility in international law and he argues that multilateral treaties would be a better solution than to reinterpret Article 2(4) to include these. Barkham points out that a large, and continuing, problem with information warfare is that the technology is so easy to utilize that differentiating between state and non-state actors is difficult, if impossible, in some situations.

Michael Schmitt in [84] provides a very complete framework from which to derive decisions as to whether an information warfare action falls under Article 2(4). The framework utilizes six criteria in a holistic manner: severity, immediacy, directness, invasiveness, measurability, and presumptive legitimacy. The Schmitt analysis was successfully performed on a case study in [85] to demonstrate its effectiveness. However, Barkham in [59] argues that the Schmitt analysis is incomplete and lacking in several places. First, Barkham argues that there can be no presumptive legitimacy because this requires determining legitimacy by "asking whether the attack is legitimate." Additionally, Barkham argues that no information warfare framework can require an *ex post* analysis because the attack cannot be readily assessed at the time of the attack to determine the appropriate response.

The Council of Europe Convention on Cybercrime [86] will become an important document in the near future and has some implications in the use of active response. The treaty requires that member states adopt specific pieces of legislation domestically to prevent computer crime. Additionally, the treaty allows for mutual assistance in the investigation and prosecution of computer criminals. This is an important tool because it means that there will be greater cooperation in cases where computer criminals launch an attack from one country into another. This treaty could provide an important diplomatic alternative to other active response measures because of increased cooperation with other states. We will now discuss genetic algorithms and co-evolutionary techniques which are used to model an active response scenario later in this work.

2.7 Genetic Algorithms and Co-Evolutionary Techniques

Research in evolutionary computing techniques has been popular in the recent decades. Researchers in the field have created many types of evolutionary paradigms, artificial life, genetic algorithms, genetic programming, neural networks, and particle swarm among others. For our purposes, genetic algorithms and co-evolution will be focused on. The co-evolutionary model is very useful in modeling security scenarios that require an adversarial relationship — such as attacker/defender in active response. It will be shown later in Section 3.8 that this model is successful in generating active response strategies. However, a short background into the topic and previous research is required. The rest of this section has been adopted from [87].

A genetic algorithm is a computation paradigm that utilizes a population of encoded chromosomes and operations upon those chromosomes for the purpose of traversing search spaces by increasing population fitness as individuals near a goal. This paradigm was first introduced by John Holland in [88]. However, although his original work describes natural competitive co-evolving populations, his theories and experiments are only subject to fixed environments.

Since Holland's pioneering work, John Koza has been a leader in the field, developing and championing the technique of genetic programming. See [89] for more details about Koza's work in genetic programming. With regard to co-evolution, he put forth two papers discussing his experiments in the subject [90, 91]. In these works, Koza describes a "hierarchical co-evolution," which is where the environment for the first population consists of a second population and vice versa. He also describes "relative fitness," which is where an individual's fitness is determined by its performance against all of the individuals of the opposing population. He puts these into practice by attempting to evolve a game strategy using a tree structure and succeeds in evolving the optimal strategy for each player without any direct knowledge of such strategy.

Robert Axelrod is generally regarded as the first to apply evolutionary techniques to game theory using the Prisoner's Dilemma in [92]. Prisoner's Dilemma is where two players must decide whether to 'rat out' the other person or to not talk at all. If both players rat the other person out, they receive no reward. If player one talks and player two does not, then player one gets a reward and the other player two does not. Axelrod utilized genetic algorithms to develop a strategy for the game. Using several fixed programs submitted in a competition by others, his algorithm was able to evolve the optimal strategy of TIT-FOR-TAT.

Miller followed Axelrod's work in [93] by applying a co-evolutionary technique to the search for a Prisoner's Dilemma strategy. Instead of the tree structure used by Koza or the linear gene structure of Axelrod, Miller utilized finite-state automata for the problem representation. His experiments showed widely varied convergence after 10 generations depending on the information about the other player available. The results of the co-evolutionary experiment showed that the top performer was just slightly less optimal than Axelrod's, but was more tolerant to short-term defections by the other player. This leads to a potential conclusion that while a co-evolutionary technique may not yield optimal strategies, the results may be very good if the game is modeled imperfectly or contains many dynamic components.

Rosin and Belew produced two papers with regard to competitive co-evolution in 1995 [94] and 1996 [95]. They used Tic-Tac-Toe, Nim, and Go as the games with which to evolve two competitive players. They developed two novel techniques, "competitive fitness sharing" and "shared sampling," which improves performance. The primary purpose of the work is to improve the "parasite" (population being tested against) population so that stronger "hosts" will be evolved. The usual method of fitness evaluation involves summing the scores during the interactions. However, competitive fitness sharing works such that each parasite's score is the number of hosts that defeat it and a host's fitness is the sum of the scores of the parasites defeated by it. In this way a host is rewarded for defeating parasites few other hosts can.

The other technique developed by Rosin and Belew is "shared sampling," where a host is tested not against the entire parasite population as it normally would be, but instead against a mixed set of parasites that tend to both defeat many hosts and are defeated themselves. Both of these novel techniques greatly improved the players evolved for these games.

Potter, De Jong, and Grefensttete presented [96] in 1995, which developed a solution for evolving agents with many subtasks. Their solution was to use multiple genetic algorithms, each evolving a single subtask. When their populations had converged, they took the best of each population combining them into a single agent, which was then able to effectively complete a more complex task composed of the trained subtasks. Each population was guided towards its subtask not by direct instruction, but rather by initial seeding. This is not an example of competitive co-evolution, but of a novel use of cooperative co-evolution. This work shows that complex rule-based behavior, such as active response actions, can be successfully evolved from simpler elements.

Haynes and Sen in [97] described their (failed) attempt at co-evolving predator and prey populations. In their experiment, they used a grid with multiple predator and prey agents where the predators could communicate together and the prey could not. They expected an arms race, where one evolutionary jump by one population is quickly matched by a counter-evolutionary jump in the other. However, the prey evolved a very simple, yet effective strategy: they all moved quickly in a straight line so that the predators were always chasing and could not surround them. This strategy fails when pitted against a greedy algorithm, and hence did not produce a novel strategy that performs better than current strategies. Possible improvements would be if the predators could predict prey action (n-look ahead) or if the predators could move quicker instead of only smarter. These improvements may have produced more complex prey strategies. The lesson derived from this research is that prey will always tend towards simple, effective strategies which exploit the actions, or lack thereof, available to the predators.

Chapter 3

RESEARCH APPROACH AND RESULTS

Now that the reader is well-exposed to the background material relating to active response, the research and results of the research will be presented. The research presented in this chapter is the combined knowledge of the three papers presented in the appendices. It is suggested that while this chapter will present the reader with the breadth of research accomplished on the topic, the depth of the research can only be discovered by examining the papers from which this research is based [39, 87, 98].

Active response is many things to many people: some consider response strike-back [7, 8, 31], some consider it deception and obfuscation [47], and others consider it an extension of intrusion detection [51]. Active response is all of these and more. However, there is no overarching and unifying framework for the concept of active response. The essence of response must be distilled and captured so that a well-founded and organized dialog on active response can be productive on informing the topic. Hopefully, this dialog will spring a multitude of approaches to response, but until that point, the following framework is proposed with which to discuss active response.

3.1 Definition

In order to unify the concepts and preserve the purpose of active response, the following criteria for a definition are proposed [98]:

- Active response is *time-bound* and takes place exclusively during a period when an attack is known to be in progress.[1]

- Active response is *purposeful*. Actions would only be considered 'response' when used to mitigate the threat for the purpose of returning to a more secure state and no further.

- Active response, being purposeful, has *limitations.* Threat mitigation need not mean threat elimination, but instead indicates that the actions seek to diminish or contain the threat with respect to the resource considered to be threatened. Neither punitive responses nor warning responses, which might be intended to have the general effect of reducing threats from a broad perspective, are included in our definition of active response.

- Active response is comprised of a *sequence* of actions.

- The decision to apply specific active response actions is *controllable* and *deliberate*,[2] though an active response action sequence's consequences might be neither.

- Active response is *technologically independent* and can be executed by an operator or automatically by an intrusion detection system using a pre-defined ruleset.

- The timeline of active response is to be considered considered *subjectively*, from the perspective of the decision-making entity and not from any other body.

[1] Accuracy of detection clearly has a role to play. The degree of certainty to which an attack is known to be in progress certainly has a role to play. We have chosen not to require that aspect of the issue in our initial identification of required characteristics of active response; it is an area for future work.

[2] The decision to apply certain responses when certain threats are identified could still be made in advance of the particular threat, as with an automated active response.

This leads us to the following definition of active response.

Definition 1. **Active Response**: *Any action sequence deliberately performed by an individual or organization between the time an attack is detected and the time it is determined to be finished, in an automated or non-automated fashion, in order to mitigate the identified threat's negative effects upon a particular asset set* [98].

This definition satisfies all of our criteria and is successful in representing all of the ideas of active response, while preserving its primary purpose of defense and threat mitigation — rather than allowing it to be defined in terms of retaliation and retribution.

3.2 Response as Integral to Security

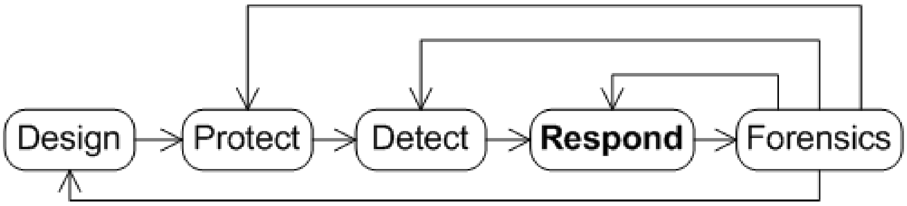


Figure 3.1: Response as Part of the Security Timeline

Figure 3.1 represents a simplified security timeline. A system is generally designed/specified and then built, protection is then designed based on a threat model, and a system is then put in place to audit the system based on threats in order to detect whether the system has been compromised. When the system is compromised, organizations quickly jump to forensics, analyzing how the system was compromised and how to protect it in the future. However, this neglects any opportunity for a response and the ability to mitigate the threat before the entire value of the system is lost — allowing the recovery of partial system value.

The mere existence of response does not imply an active response action. The only implication is that response is considered and either action or no-action is decided. In this way, response is preserved as a legitimate security tool when it is necessary, but its use is still strongly considered and not an automatic 'knee-jerk' reaction to any security event.

I propose that response is integral to the computer security timeline. It occurs between the time the threat is detected and is finished (when forensic analysis begins), and allows for the mitigation of the threat before complete system loss or compromise.

3.3 A Response Taxonomy

Based on the definition given in Section 3.1, active response is a broad concept that encompasses many actions. These actions range from no action all the way to a strike-back methodology with maximal damage inflicted on the attacker. In order to illustrate the great range of action available to a defender, the following taxonomy of active response action is proposed. This taxonomy is adapted from [39, 98, 99].

1. *No Action*: A threat is detected, but no action is taken.

2. *Internal Notification*: Using the organizational structure to notify the designated responder(s) of an active response situation.

3. *Internal Response*: Applying active response actions within the domain over which the responder has authority (e.g. close a threat vector's associated port).

4. *External Cooperative Response*: Employing entities external to the responding organization to mitigate a threat.

5. *Non-cooperative Intelligence Gathering*: Using external services (e.g. finger, nmap, netstat, etc.) to gather intelligence on the threat. Sometimes referred to as "look but don't touch."

6. *Non-cooperative 'Cease and Desist'*: Stopping harmful and unauthorized services (e.g. zombie control processes) without compromising legitimate usability.

7. *Counter-strike*: An external action to reduce or deny the capabilities of an attacker to continue the attack.

8. *Preemptive Defense*: With knowledge of a forthcoming attack, execute active defense actions to preempt (and disable) the upcoming attack.

3.4 Evaluating a Response

The previous sections propose active response as a legitimate security tool and as integral to the security timeline that should be considered during a security event. However, there are several reasons why active response has not been widely adopted or embraced. One of those reasons is that there is a great number of issues with an active response methodology — many of which have no solution in the foreseeable future. These issues are part of the necessary dialog regarding active response and must be discussed for a consensus to arise on the weight and application of these issues.

There are five issues: ethical, legal, risk analysis, technical, and unintended consequences [98]. Most of these issues have been thoroughly discussed in the background provided in Chapter 2, therefore only the outline of the issues are presented.

3.4.1 *Ethical*

The ethical issues surrounding active defense compose a large segment of discussion. The question is "whether certain response actions are ethical, particularly those with the potential to overflow the boundaries of the responder's domain of responsibility" [98]. This is difficult question to address, and is made even more difficult by the fact that the decision of ethical action must be made quickly in an active response scenario. It is therefore imperative that because of the nature of the ethical question, these decisions be made *a priori* and integrate them within the decision making process [98].

The two frameworks from which to work are the teleological (only consequences are determinative) and deontological (only actions themselves are determinative). These ethical frameworks have been covered in depth in Section 2.5, no additional discussion is necessary.

3.4.2 *Legal*

The legal question is rather strongly related to the ethical question in Section 3.4.1. The legal question in active response is regarding the applicability of existing law to a cyber domain and "which actions are permitted under the environment governing the decision-maker" [98]. The laws implied in this question are not only criminal domestic law, but also civil, international, and foreign domestic. The legal background necessary for these questions has been previously discussed in Section 2.6 and no further discussion is necessary.

3.4.3 *Risk Analysis*

In order to make a rational, informed, thorough, and complete decision regarding active response actions, a risk analysis is necessary. However, can a satisfactory risk analysis be accomplished? Can the ethical and legal risks be realistically evaluated in the face of enormous unknowns? Can ethical risks be realistically valued? Can legal risks? These are the questions that should frame ongoing research.

[98] proposes two limiting factors to risk analysis that may help in the exposure of these issues: (1) risk analysis should be treated as a due diligence requirement, and (2) risk analysis need not include all risks. These two factors allow a realistic and manageable, but considerable, risk analysis.

3.4.4 Technical

[98] described three technical issues with active response:

1. Do contemporary response-triggering systems provide enough confidence in their alerts to base particular responses upon? Is there enough information incorporated in the alert to support response?

2. Can response be quick enough to be effective?

3. Is identification and authentication of the attacker/attacker's resource via trace-back [3] and other means viable in this environment, particularly given the prevalence of anonymity techniques and utilization of "innocent bystander" resources?

These questions have been thoroughly addressed in Section 2.2.1 and need no further discussion.

3.4.5 Unintended Consequences

Unintended consequences are not an issue unto themselves, but rather reflect the unintended consequences of the legal, ethical, and technical issues. However,

[3] Trace-back is the attempt to find the attacker by tracing their network traffic through the network signaling equipment that composes contemporary networks.

because the question of unintended consequences is of such magnitude, it is treated as a separate issue to emphasize the importance of analysis in this area.

[98] states that the costs of unintended consequences derive from three sources:

1. An unintended (counter) response elicited from the attacker (i.e. you want them to stop, but an unexpected result occurs — perhaps behavior escalates or is diverted).

2. Damage to the perceived source of the threat excluding the attacker (i.e. damage to a zombie or co-opted system).

3. Unplanned damage to the responder′s domain.

These sources will now be discussed.

Unanticipated Attacker Response

One of the risks with executing an active defense action is that the attacker may detect it and alter their behavior. As with any form of self-defense, there is always the possibility that resistance will lead to escalation (though it may also lead to cessation of the attack). Some examples of attacker tactic change in the cyber domain are diverting their attack to another resource, becoming angry and launching a more vicious attack, or even alerting other attackers to join the assault.

However, deception, diversion or tactical change may be a useful alternative to direct frontal assaults (i.e. more detectable forms of response). The attacker could be diverted to a honeypot/honeynet or other disposable resources — which may then provide additional data as to the source of the threat and better inform any future actions against this attacker.

Damage to Non-Attackers

Accurate tracking, an identification/authentication aspect mentioned in the Section 2.2.1 is difficult to ensure. As described earlier, there are multiple methods that attackers may utilize in their quest for anonymity. And because methods of detecting, tracking, and tracing attackers are highly unreliable, there is a high likelihood the attacker has been misidentified. Therefore, any action taken based on that information includes an uncertainty that the target is innocent or unaware of the attack.

Even worse is the real risk that the target for the active response action may be a life, safety, or national security critical system, possibly of more value than the one under attack. A byproduct of active response could be an increase of threat to critical systems, that may be used as shields from active responses.[4] This unintended consequence has both ethical and technological effects, with the potential for legal effects as well.

Damage to Own Resources

The third source of unintended consequence costs come from effects upon one's own resources. This risk is common whenever an organization makes a change in policy or design. The risk is that by changing a policy or design, the responder may unintentionally block legitimate users from a resource or harm internal assets, causing even more damage than the original attack.

[4] This concept is known as 'perfidious action' and is considered a war crime by the Geneva Convention (Protocol 1, Part III, Article 37) because it causes defenders to begin targeting innocents in order to protect themselves. See Section 2.6.3 for a discussion of this and other international law issues.

3.5 The Eight Stages of Response

Caltagirone and Frincke in [98] proposed an eight-stage response model consisting of: planning, detection, evaluation, decision, action, analysis, escalation, and maintenance.

Of the eight stages, three are not within the scope of the definition of active response: planning, detection and maintenance. However, these stages are necessary for a thorough analysis of active response and cannot be decoupled from the process of response. The planning stage is necessary because active response requires a complex and thorough risk analysis that cannot be accomplished during an attack, but must rather be done before active response is considered. Therefore, planning for active response is part of the active response process. Additionally, detection and response cannot be completely decoupled since the detection stage is the source of all information regarding the attack. Finally, maintenance is an important component of any policy-driven endeavor in order to keep the policy current and relevant. Therefore, like planning, maintenance is part of the response process although not included in the definition of active response.

The eight-stage process model will now be described and discussed.

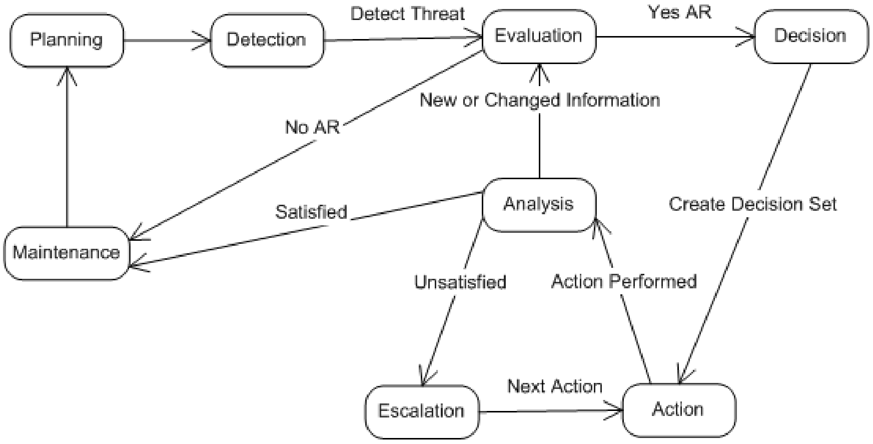


Figure 3.2: The Eight Stage Response Cycle

3.5.1 Planning

The model for response that has been chosen is planning-centric. This means that the ***active response policy*** developed in this stage will be the standard by which decisions are made. A planning-centric model requires the largest amount of resources to be spent on the development of the plan because of the issues discussed in Section 3.4. With proper planning as incorporated with risk analysis, the risks can be reduced and active response made a more viable option.

An ***active response policy*** is developed in the planning stage. The policy is an unambiguous analysis of the risks and costs (in every category) of each threat and potential mitigating active response action. For the development of the policy, a number of inputs are necessary: the assets to be protected, the threats to those assets, the risks/costs if the threats were successful (or partially successful), and the potential mitigation active response actions and corresponding risks/costs. With this information, the information is scaled and the actions are ordered based on relative probability of success and risk. The output of this process should be a clear analysis of all of the risks and costs involved, allowing for the most informed decision.

3.5.2 Detection

Detection is the discovery of a threat. The discovery itself is technologically independent, meaning that the threat could have been discovered by a human operator or by an automated IDS. The purpose of detection is two-fold: to determine the start and end of an attack and to identify important evidence in the attack in order to make the most informed active response decision with the highest confidence. This evidence is then fed to the evaluation stage.

3.5.3 Evaluation

Evaluation takes the threat data provided by the detection stage and compares that to the policy developed in the planning stage. The calculus of response decisions is used to identify whether active response is appropriate. If active response is found to be appropriate, data is passed to the decision stage where a response is chosen or data is passed elsewhere and the cycle moves to the maintenance stage.

3.5.4 Decision

The decision stage determines exactly which actions, identified in the active response policy as potential mitigation techniques, are selected for execution. The actions selected are placed into the ***decision set*** $(a_1, a_2, \ldots, a_k)$, an ordered set of actions that will execute in sequence until the threat is mitigated to satisfaction. The decision stage is examined in greater detail in Section 3.6.

3.5.5 Action

In this stage, the active response action determined by the decision set is executed. After execution of an action, response moves to the analysis stage.

3.5.6 Analysis

The analysis stage is the watchdog. It determines whether the active response should continue based on a number of factors:

- Has the threat been successfully mitigated? (go to maintenance)
- Was the action unsuccessful? (go to escalation)
- Has the environment changed enough to require re-evaluation (go to evaluation)?

3.5.7 Escalation

Escalation selects the next action in the decision set to be executed and feeds that into the action stage for execution. Escalation is necessary if an action is unsuccessful in providing sufficient mitigation.

3.5.8 Maintenance

Maintenance is the final stage of the response cycle. Although not explicitly allowed as an element of active response by our definition, maintaining an effective active response policy is essential when implementing policy-based response. The maintenance stage takes any forensic and post-mortem data and applies it to the policy so that past knowledge is included in future decisions.

3.6 Formal Decision Model

In [39], Caltagirone and Frincke develop a formal decision model for active response. The decision model builds on the eight-stage response model (previously discussed in Section 3.5) as its foundation. It attempts to produce a pragmatic, implementable model, whose purpose is to inform an organization as to an appropriate set of active response actions to execute given the defined policy. The model is separated into four distinct stages: scoring chart, asset evaluation, action evaluation, and escalation ladder creation. When these stages have been completed, an organization will have an ***active response policy*** to inform their decisions.

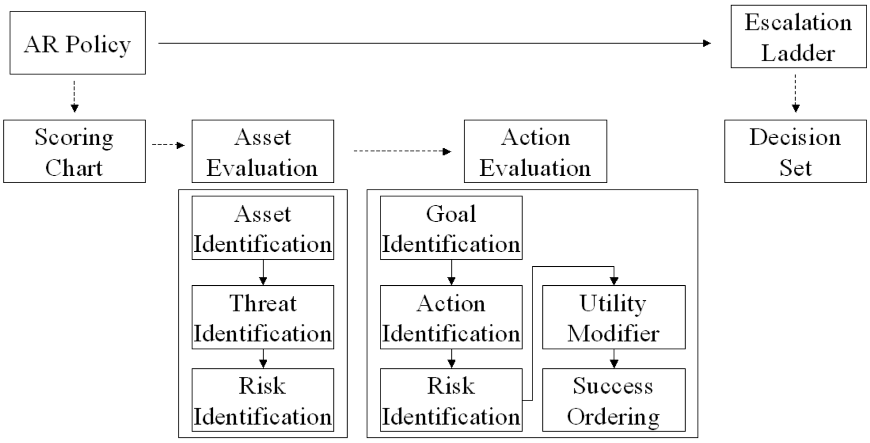


Figure 3.3: Decision Model Overview

3.6.1 Assumptions

The decision model makes several assumptions before attempting to evaluate a potential active response.

- *Assets can be estimated*: The model assumes that the assets and risks of an organization can be accurately estimated with respect to the given categories.
- *Responses can be evaluated*: The model assumes that all of the active defense actions to a given threat have been included and that the model will not be used to evaluate actions that have not been included.

- *Consequences are enumerable*: The model assumes that all the consequences of an action are known and have been included in the active defense policy.

- *Ethical considerations can be evaluated*: The model assumes that all ethical considerations have been evaluated correctly to provide their accurate weight.

- *Legal consequences are known*: The model assumes that all legal consequences are known, the laws have been tested, and interpretations will be static.

These assumptions are tempered with the fact that an organization has the freedom to choose only assets or actions on which they can perform an acceptable risk evaluation.

3.6.2 Scoring Chart

Before the active response policy can be developed, there must be a method which allows the diverse categories of risks to be equated. Therefore, an organization must have a reasonable method of scoring the risks. ADAM includes five threat-risk categories, which can be modified to fit an organization's strategic goals. The categories are: legal, national security, financial, ethical consequences, and ethical actions.

The first three are traditional risk areas. However, when active response activities are contemplated, it is important to include ethical considerations. Clearly, performing an active response action places ethical risks on an organization. While some organizations may minimize the weight of this category, others may place a high value upon it. Goals important for maintaining an ethical organization should be supported by any active response model.

Ethical risks are further subdivided into two parts: ethical consequences and ethical actions. Choosing between a teleological (only the consequences of an action are deemed necessary for ethical consideration) and a deontological (only the act in and of itself is considered) ethical theory is an overly burdensome way to approach the issue. Therefore, both are represented. The Ethical Consequences category represents the teleological perspective, which defines the 'ethicalness' of the potential consequences of an active response action. The Ethical Action category represents the deontological perspective, which describes the 'ethicalness' of the action an organization takes in and of itself.

Scoring in any of these categories is difficult — as even financial and legal risks cannot be assessed with full accuracy since they draw on qualitative determinations and changeable environments. It is also correct that ethical scoring in particular is highly subjective and difficult for an organization to perform. On the other hand, an organization that cannot answer these questions without the pressure of a live attack damaging key assets will certainly not be better positioned to do so once the attack occurs.

To simplify the scoring task in our preliminary model, we have initially required only potential active response actions (because consequences are not considered in a deontological framework) in the Ethical Actions category. In the Ethical Consequences category, all potential consequences need to be considered.

The actual construction of the chart is not overly complicated: for each risk category (e.g. ethical action, ethical consequence, legal, national security, financial, etc.) there is a real-number scale ranging from -1 to 1; -1 being the greatest benefit in the category, 1 being the greatest loss in the category, and 0 being no loss or benefit.[5] For each number in the scale, a risk is equated. For example, in financial risks, an organization may determine that the worst risk is $1,000,000, that would then be equated to 1, while a loss of $500,000 may be equated to .5.[6] This scale (from -1 to 1) provides advantages to the model, including allowing for costs and

[5] It is forseeable that some categories may not have benefits, and so the range would be 0 to 1.

[6] The range and scale are not required to be symmetric.

benefits to be modeled as well as probabilities of success to be equated with risk.

3.6.3 *Asset Evaluation*

In asset evaluation, the organization looks to their standard security practices and security policy (which should be in place) in order to produce a set of assets that are going to be protected using active response. This also includes a listing of the specific threats to be included in the active response policy.

Once a complete set of assets and threats are compiled, all of the risks to the assets from those threats must be identified. Then the costs of those risks must be located in the scoring chart in order to determine their score. This allows a threat/asset combination to receive a score that is the sum of the scores of all of the risks, allowing the threat/asset to be compared to potential active response actions and their risks.

3.6.4 *Action Evaluation*

Action evaluation is very similar to the asset evaluation described in the previous section. In this step, the organization lists all of the potential active response actions that can be executed given the set of threats in the Asset Evaluation. The risks of these actions to the organization are then determined and scored using the scoring chart. Additionally, for each action, its relative probability of success is determined. This provides a way to compare the potential successfulness of actions.

Once an action and its risks have been identified, an organization applies a weight to the risks of the action in each category depending on the organization's unique mission and goals. For example, a financial institution may determine that financial and legal risks are its primary concerns and would weigh those risks higher than other categories. Similarly, a medical facility may decide that financial and ethical risks are more important than other categories and weigh those higher. Once the identification, scoring, and weighing has been accomplished, the ***active response policy*** has been created and an escalation ladder can be produced.

3.6.5 Escalation Ladder Creation

At this point, the ***active response policy*** has been created and a set of actions can be determined for each threat/asset relation. The purpose of the ladder is to determine the most appropriate set of actions an organization can take (given the data provided in the policy) when a threat is detected. Formally, the escalation ladder is an ordered set of actions that are progressively executed until the threat is successfully mitigated or another condition, like a risk threshold, is met.

The set of actions is created by ordering the actions based on a simple formula equally balancing risk and success. The formula takes the total score of a threat/asset, subtracts the score of the action, and then further subtracts the probability of success for the action. Once this formula has been applied to each action, the actions are ordered from lowest to highest based on the function just described. This ordering creates a set where the initial actions have the highest success to risk ratio. At this point, the escalation ladder is complete and can be used in the active defense algorithm described next.

3.7 Active Defense Algorithm

The active defense algorithm presented in this section is provided as an implementable form of the eight-stage response cycle (Section 3.5) and formal decision model (Section 3.6). It provides an additional method of analyzing and validating the model.

The algorithm takes as input, the threat, t, the asset being threatened, a, and the active response policy, P, developed in Section 3.6. The algorithm follows.

Algorithm 1 $Active - Defense(t, a, P)$

```
1: if $a$ not $\in P_A$ then
2:     fail
3: end if
4: if $t$ not $\in P_T$ then
5:     fail
6: end if
7: $X \leftarrow$ADModel$(t, a, P)$
8: $n \leftarrow |X|$
9: $riskAssumed \leftarrow 0$
10: for $i \leftarrow 1$ to $n$ do
11:     $k \leftarrow X_i$
12:     while $k$ cannot be performed do
13:         $k \leftarrow$ get next action in $X$
14:     end while
15:     $riskAssumed \leftarrow riskAssumed + score(k)$
16:     if $riskAssumed > sum(t)$ then
17:         break
18:     end if
19:     execute the action $k$
20:     if action $k$ achieved $goal(t)$ then
21:         break
22:     end if
23: end for
```

Now for a description of the algorithm. Steps **1-6** satisfy stage 2 (evaluation) by deciding whether the asset and threat are covered in the active defense policy — if it is not in the policy, then fail and do not execute an action. Step **7** satisfies stage 4 (decision) by retrieving from the model the decision set of actions. Step **8** assigns the variable n the size of the set X. Step **9** initializes a new variable $riskAssumed$, which stores a total of the risk incurred by executing the actions. Step **10** iterates over the set X. Step **11** assigns a variable k the action that in the set X at index i. Steps **12-14** search for the next action k in the escalation ladder that can be performed using the information available (e.g. is the IP address correct, etc.). Step **15** adds the risk of the action k to the current risk assumed. Steps **16-18** check if the current amount of risk (total risk) has exceeded the risk of the threat, if it has then get out of the loop. Step **19** satisfies stage 5 (action) by executing the action selected. Steps **20-23** satisfy stage 6 (analysis) by checking if the action has achieved its stated goal in $goal(t)$, if it has then no need to continue. Stage 7 (escalation) is satisfied by the fact that the next iteration through the loop will escalate to the next action in the decision set.

One may notice that if the algorithm determines that the threat/asset combination is not found in the policy (steps 1-6), the algorithm fails and does not continue executing the active response. This is an important step as the algorithm defaults to no-action if the policy has not sufficiently described the current security state. The reason for this is straightforward: if the policy is not thoroughly described, then no risk analysis can be performed in the decision stage. Thus, no decision set can be created, preventing any action from being executed. Additionally, if the algorithm were to begin executing actions for which a thorough risk analysis had not been completed, an organization *could* assume enormous or very slight risk. The lack of knowledge of assumed risk provides no justification for active response.

3.8 Evolutionary Active Response Model

There are several approaches one could take in modeling active response. Competitive co-evolution may not seem like the first choice, but its ability to model adversarial relationships, the freedom it provides to search a solution space, and past success on other problems make it an excellent choice. In this case, an active response game was created and the individuals in the populations evolved strategies to beat their opponents — allowing the individuals with the best strategies to procreate and move on to the next generation. This technique was successful and met its goal of producing realistic and reasonable strategies given the scenario [87]. In the end, a baseline is created, determining that evolutionary techniques can be successfully applied to active response and possibly other computer security problems.

This section will provide a general overview of the experiment, however all of the details are provided in Appendix D and [87].

3.8.1 The Scenario

Because active response is not a general purpose security tool, it must be tailored specifically for each threat and each asset. Therefore, any experiment in active response must provide a scenario so that the threat and asset can be identified.

For this experiment, a realistic scenario was chosen. The scenario is based on a medical patient database. This database is hosted by a medical facility that provides access via the Internet to other facilities. The data stored in the database is necessary for patient care at the facilities that use it. If the data's integrity or availability were threatened, then patient care would also be threatened. This scenario implies that the worst threats are those that compromise availability and integrity. Likewise, the riskiest active defense actions would be those that do the same.

3.8.2 Active Response Model

For this experiment, 12 defensive, and 11 offensive actions were chosen, with each set sharing only one action: the null action (signifying no action). These actions were then given a relative risk value. The risk value of the offensive actions signified the cost to the defender if the attack is successful, while the defensive action risk value signified the cost of executing the action.

Defensive actions were also flagged with whether the attacker's IP address is necessary or whether the successful execution of the defensive action meant that the attacker was permanently stopped (e.g. FBI). An additional dimension was added by allowing the attacker to spoof their IP address and allowing the defender the ability to use traceback to discover their true IP address in order to execute additional actions.

3.8.3 The Game

The game that the attackers and defenders played is basic. An attacker confronted a defender. The attacker executed their first action and the defender was allowed to cycle through their entire action set until either (1) they had no more actions or (2) the defender executed an action to stop the attack action. Each action the defender executed accumulated risk, while only if the attacker was successful was the attacker action risk added to the score.

The attacker's goal was to accumulate the greatest amount of risk to the defenders. On the other hand, the defender's goal was to have the least amount of risk accumulated. Therefore, it was in the defender's best interest to discover the strategy that stopped the most attacks and limited the cost of their own actions and it was in the attacker's best interest to discover the strategy that circumvented the most defensive actions or caused the defenders to execute expensive actions in an attempt to mitigate.

3.8.4 The Experiment

The goal of this experiment was to discover whether it is possible to evolve an active defense strategy that is reasonable and could be considered viable enough to utilize against a threat.

To accomplish this, two populations were defined, one of attackers and one of defenders. These populations were evolved simultaneously and their fitness value was derived by their performance in the game as noted in Section 3.8.3. The attackers searched for the strategy incurring the highest risk on the defenders and the defenders searched for a strategy to minimize their risk. These populations were then placed in competition with each other in order to obtain a fitness value for the genetic operation of selection. One hundred trials of the experiment were run to produce data on the fitness of the populations and frequency data on actions chosen by the top individuals.

3.8.5 The Results

The experiment was a success. First, the implementation of the genetic model is validated by the fact that the two populations followed the standard behavior of co-evolutionary adversarial populations. As shown by [94, 95, 97], competitive co-evolutionary populations should mirror each other's fitness (with a slight time difference) — so as one population derives a new stronger strategy, thereby increasing its fitness, its competitive population will similarly evolve a new strategy increasing its fitness as well.

Second, both the attackers and defenders evolved reasonable strategies. The attackers chose a strategy that inflicted a great amount of risk while minimizing the defensive actions available to mitigate the threat. The general attacker strategy was to spoof their IP address and then to 'Poison DNS' and hack the server to change the records. The defender's strategy, on the other hand, was to rely heavily on external entities (e.g. ISP and FBI), while utilizing low-risk internal responses such as sending TCP RST packets.

An example attacker strategy: (Spoof IP Address) (Poison DNS) (Port Scan the Server) (Port Scan the Server) (Poison DNS) (Port Scan the Server) (Hack Server, Install Backdoor) (Poison DNS).

An example defender strategy: (Ask ISP to Shut-off Attack) (Use Traceback) (Use Traceback) (Contact Administrator) (Contact FBI) (Filter IP at firewall) (Contact Chief Technology Officer) (Null Action)

3.9 Evolutionary Model Critique

The evolutionary model is one which deserves much more attention in regards to potential enhancements and improvements. There has been no study on how modifying the model's parameters effects changes in strategy development. Because of this, the conclusions drawn from this experiment are limited to only the parameters stated in the experiment description. In order to rectify this, the genetic algorithm parameters and game parameters would need to be modified independently and the effects studied. This would allow significant conclusions as to the robustness of the co-evolutionary approach and potentially provide additional insights as to new attacker or defender strategies.

Additionally, modifying the game, environment, and scenario slightly to include a greater range of 'real world' behaviors would provide interesting conclusions. For example, the effects of allowing the attackers to implement a 'low and slow' attack and conversely giving the defenders the ability to detect or potentially defend against these attacks without the ability to detect them. Another enhancement would be to model zero-day or devastating attacks for which a defense would need to be evolved and would require a time-delay (e.g. the effects of the blaster worm on an unprepared population). These enhancements would significantly improve the model's ability to develop strategies in a more realistic environment.

Chapter 4

CONCLUSION

4.1 Conclusions

Active response is not, as some view it, a strike-back methodology, but rather a topic which, by our definition, incorporates a large number of actions for the purpose of mitigating a threat. Already, a large number of public and private organizations, as well as individuals, are utilizing response as more applications incorporate response technology. Without any framework, these organizations and individuals have no manner to judge the various risks and costs of their response. This thesis has presented such a framework, consisting of a taxonomy, a summary of issues, a formal decision model, an implementable algorithm, and an evolutionary technique for generating active response strategies.

While this framework makes the multitude of problems and issues of active response clearer, it does not remove them as an obstacle to active response. There is still much more yet to be done in the topic, especially input from lawyers and ethicists. However, the framework allows organizations and individuals contemplating response to successfully analyze and model an active response scenario (given the appropriately allocated resources) to the point where active response can be utilized as an additional layer of defense and a legitimate security tool.

4.2 Summary of Contributions

There are several novel contributions that this work provides to the computer security community:

- **A Definition of Active Response**: A set of requirements for any definition of active response is presented as well as a definition of active response which satisfies those requirements.

- **A Taxonomy of Active Response Actions**: A taxonomy of active response actions is developed and provides a greater discussion of the definition of an active response action.

- **A Summary of Issues**: A list and summary of issues regarding active response is presented as well as a short discussion of these issues to further frame the problem of active response.

- **Response Model**: A eight-stage model is developed that describes the stages of active response and the responsibilities in each stage to produce a proper response.

- **Formal Decision Model**: A formal decision model is developed for the purposes of developing an active response policy which can determine appropriate active response actions to take during an attack.

- **Active Defense Algorithm**: An algorithm that combines the formal decision model and response model in order to produce an implementable version of the models presented.

- **Evolutionary Active Response Model**: A genetic co-evolutionary implementation of active response is presented in order to illustrate that it is possible to evolve reasonable active response strategies.

4.3 Future Research

There are several avenues available for future research on this topic, many that are directly referenced in this work. First and foremost is the need for additional research on the numerous social challenges to response, including the ethical and legal implications of active response. This discussion will need to include lawyers and ethicists as well as computer scientists in order to overcome these challenges. Specifically, the concept of evaluating and comparing legal consequences (from a civil, criminal, domestic, and international viewpoint) needs to be defined. Also, an ethical framework for response must be developed, providing ethical guidelines for response decisions.

From a technology standpoint, the taxonomy presented in this work needs to be developed further. This refinement process must include a thorough study as to the current state of response tools and techniques and their classification. Additionally, new tools and techniques themselves must be developed. Specifically, more research needs to be undertaken as to the opportunities provided by traceback and the limits of network infrastructure equipment to provide additional response capabilities.

In addition to social and technology requirements, a greater diversity of decision models and methods of calculating and equating risks in response is needed. The largest question looming over the decision models is which one produces the *best* decision set. The model provided here only attempts to provide an appropriate set of actions, relying on subjective qualifications to determine appropriateness. However, it would be worthwhile to research a set of rigid criteria to evaluate a response decision set.

Another question left unanswered is: what is the true relationship between the risk and success of an active response action. The model provided here gives them equal standing, but is that the true relationship? Or is the relationship one of the parameters to be tuned by an organization's own needs? What would determine that relationship? These questions need to be seriously addressed.

Lastly, work is required on applying competitive co-evolutionary techniques to active response. The implementation and experiment provided here is only a cursory look a potential application. However, it needs a much more thorough understanding and experiment that bring in more components of 'real world' security environments.

4.4 Looking to the Future of Response

There is no doubt that active response is an emerging topic in the field of computer security: the increased attention of the news media to response situations, a greater discussion in the academic community, more products including the ability to respond to the ability to detect, and the increased deployment of stand-alone response tools (e.g. Lycos Make-Love-Not-Spam screensaver). But with the increased attention and greater range of products, what is the future for active response — where will it end up?

The 4th Workshop on the Active Response Continuum to Cyber Attacks concluded in 2005 and the results of the workshop are clear: the state of the topic is unclear. There is quite a bit of disconnect within the community regarding the direction response should head. There are some who advocate a slow transition into response while others are desperately searching for a solution in order to solve their immense security problems. The greatest outcome of the increasing academic discussion is the finding that there are serious questions that still need to be answered, especially in the areas of concern detailed in this work.

On this topic, a strong pragmatic value is associated with a potential solution to these problems. Academics and industry experts tend to agree that response is a necessary tool since other current tools are ineffective in completely protecting assets. In this vein, industry, academia, and government will slowly come together in the upcoming years to further discuss this problem, solutions, and policy. There is no doubt that industry, and some in academia, will continue to develop response tools and 'proof-of-concept' applications. From that basis, there will be a greater demand on academia and government to solve the problems of response and develop policy with regard to these tools and their use.

Currently, governments are in the midst of developing policy to deal with both the domestic and international consequences of cyber criminals. Similarly, the foreign relations and international law communities will develop a framework for electronic response and self-defense over time, which will transfer slowly to governmental policy. However, this cycle is long and complicated, so the road to a consistent and applicable policy will include many false starts.

Industry will continue on its current path for the next few years, developing products which include the ability to respond within an organization's boundaries with few products reaching outside of the bounds. However, most organizations are risk-adverse and the number that are actually interested in implementing response technology that reaches outside of an organization is low. Therefore, over the next several years there will be a greater advancement and understanding of 'internal response' (especially from IDS and networking components).

In academia, response will slowly be understood and develop into its own sub-topic of computer security, much the same way that intrusion detection systems did in the 1990s. However, there is a necessity to include the greater academic community beyond computer science as well — and so academics must reach out to the lawyers and ethicists to help address the concerns of response from these viewpoints. Therefore, academia will be developing the discussion of this topic, as well as slowly producing new tools and techniques for response — probably focusing more on ‘internal response’ in the beginning and then slowly moving outward.

In the end, it cannot be forgotten that response is a study of social and ethical computing. One is ultimately responding to a human on the other side of the technology.

BIBLIOGRAPHY

[1] C. Taylor, A. Krings, and J. Alves-Foss, "Risk analysis and probabilistic survivability assessment (RAPSA): An assessment approach for power substation hardening," in *ACM Workshop on Sci Aspects of Cyber Terrorism*, Washington, DC, 2002.

[2] P. Oman, E. Schweitzer, and D. Frincke, "Concerns about intrusions into remotely accessible substation controler and SCADA systems," in *27th Annual Western Protective Relay Conference*, Spokane, WA, 2000.

[3] P. Oman, E. Schweitzer, and J. Roberts, "Safeguarding IEDs, substations, and SCADA systems against electronic intrusions," in *2001 Western Power Delivery Automation Conference*, Spokane, WA, 2001.

[4] CNN, "Teen hacker faces federal charges," *CNN.com*, 1998. [Online]. Available: http://www.cnn.com/TECH/computing/9803/18/juvenile.hacker/

[5] K. Poulsen, "Slammer worm crashed Ohio nuke plant network," *Security Focus*, 2003. [Online]. Available: http://www.securityfocus.com/news/6767

[6] J. Krim, "Dozens of privacy bills introduced after spate of security breaches," *Washington Post*, p. E01, April 9 2005.

[7] T. Mullen, "Defending your right to defend: Considerations of an

automated strike-back technology," 2002. [Online]. Available: http://www.hammerofgod.com/strikeback.txt

[8] ——, "The right to defend," *Security Focus*, 2002. [Online]. Available: http://www.securityfocus.com/columnists/98

[9] B. Schneier, *Counterattack*. Cryto-Gram Newsletter: Dec 15, 2002. [Online]. Available: http://www.schneier.com/crypto-gram-0212.html

[10] P. Roberts, "Lycos, spammers trade blows," *PCWorld.com*, 2004. [Online]. Available: http://www.pcworld.com/news/article/0,aid,118816,00.asp

[11] CNN, "Spamming spammers?" *CNN.com*, 2005. [Online]. Available: http://money.cnn.com/2005/03/22/technology/ibm_spam/index.htm

[12] W. R. Cheswick and S. M. Bellovin, *Firewalls and Internet Security: Repelling the Wily Hacker*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc, 1994.

[13] B. Guttman, E. Roback, United States Dept of Commerce Technology Administration, and National Institute of Standards and Technology, *An introduction to computer security: the NIST handbook*. Gaithersburg, MD: U.S. Dept. of Commerce Technology Administration, National Institute of Standards and Technology, 1995.

[14] F. J. Corbato and V. Vyssotsky, "Introduction and overview of the Multics system," in *1965 Fall Joint Computer Conference*, 1965, pp. 185–196.

[15] E. E. David and R. M. Fano, "Some thoughts about the social implications of accessible computing," in *Fall Joint Computer Conference*, vol. 27. Washington D.C.: Spartan Books, 1965, pp. 243–247.

[16] J. P. Anderson, "Computer security technology planning study," ESD/AFSC, Tech. Rep. ESD-TR-73-51, 1972.

[17] B. DeWolf and P. Szulewski, "Final report of the 1979 summer study on air force computer security (draper report)," 1979.

[18] *Trusted Computer System Evaluation Criteria*. Department of Defense, 1983.

[19] P. G. Neumann, L. Robinson, K. N. Levitt, R. S. Boyer, and A. R. Saxena, "A provably secure operating system," Stanford Research Institute, Report M79-225, 1975.

[20] P. Meyers, "Subversion: The neglected aspect of computer security," Masters of Science, Naval Postgraduate School, 1980.

[21] F. Cohen, "Computer viruses," Thesis Ph D, University of Southern California, 1986.

[22] E. Spafford, "Crisis and aftermath," *Communications of the ACM*, vol. 32, no. 6, pp. 678–687, 1989.

[23] C. Stoll, "Stalking the wily hacker," *Communications of the ACM*, vol. 31, no. 5, pp. 484–497, May 1988.

[24] B. Cheswick, "An evening with Berferd in which a cracker is lured, endured, and studied," in *USENIX*, San Francisco, 1992. [Online]. Available: http://www.securityfocus.com/data/library/berferd.ps

[25] S. M. Bellovin, "There be dragons," in *Third Usenix UNIX Security Symposium*, Baltimore, MD, 1992, pp. 1–16.

[26] T. Shimomura and J. Markoff, *Takedown*. New York, NY: Hyperion Books, 1996.

[27] M. J. Ranum, "Thinking about firewalls," in *Second International Conference on Systems and Network Security and Management (SANS-II)*, 1993.

[28] J. O. Kephart, "A biologically inspired immune system for computers," in *Fourth International Workshop on the Synthesis and Simulation of Living Systems (Artificial Life IV)*. Cambridge, MA: MIT Press, 1994, pp. 130–139.

[29] Reuters, "PC viruses spawn $55 billion loss in 2003," 2004. [Online]. Available: http://news.zdnet.com/2100-1009_22-5142144.html

[30] V. Jayawal, W. Yurcik, and D. Doss, "Internet hack back: Counter attacks as self-defense or vigilantism?" in *International Symposium on Technology and Society*, Raleigh, North Carolina, 2002, pp. 380–386.

[31] T. Mullen, "Neutralizing Nimda: Technical, moral and legal discussions of an automated strike-back," in *Defcon*, 2002. [Online]. Available: http://www.blackhat.com/presentations/bh-usa-02/bh-us-02-mullen-nimda.ppt

[32] Reuters, "Computer under attack can hack back, expert says," *USAToday.com*, 2002. [Online]. Available: http://www.usatoday.com/tech/news/computersecurity/2002-08-05-hack-back_x.htm

[33] T. Mullen, "Strikeback, part deux," *Security Focus*, 2003. [Online]. Available: http://www.securityfocus.com/columnists/134

[34] C. Loomis, "Appropriate response: More questions than answers," in *SecurityFocus*, 2001. [Online]. Available: http://www.securityfocus.com/infocus/1516

[35] G. D. Grove, S. E. Goodman, and S. J. Lukasik, "Cyber-attacks and international law," *Survival*, vol. 42, no. 3, pp. 89–104, 2000.

[36] D. Bruschi and E. Rosti, "Disarming offense to facilitate defense," in *2000 Workshop on New Security Paradigms*. Ballycotton, County Cork, Ireland: ACM Press, 2000, pp. 69–75.

[37] ——, "AngeL: A tool to disarm computer systems," in *2001 Workshop on New Security Paradigms*. Cloudcroft, New Mexico: ACM Press, 2001, pp. 63–69.

[38] D. Bruschi, C. L., and E. Rosti, "Less harm, less worry or how to improve network security by bounding system offensiveness," in *16th Annual Computer Security Applications Conference*. New Orleans, Louisiana: IEEE Computer Society, 2000, pp. 188–195.

[39] S. Caltagirone and D. Frincke, "ADAM: Active defense algorithm and model," in *Aggressive Network Self-Defense*, N. R. Wyler, Ed. Rockland, MD, USA: Syngress Publishing, 2005, pp. 287–311.

[40] S. M. Specht and R. B. Lee, "Distributed denial of service: Taxonomies of attacks, tools, and countermeasures," in *17th International Conference on Parallel and Distributed Computing Systems*, 2004, pp. 543–550.

[41] *IP Source Tracker*. San Jose, CA: Cisco Systems, 2004. [Online]. Available: http://www.cisco.com/

[42] T. Kohno, A. Broido, and K. Claffy, “Remote physical device fingerprinting,” in *IEEE Symposium on Security and Privacy*, 2005.

[43] T. W. Doeppner, P. N. Klein, and A. Koyfman, “Using router stamping to identify the source of IP packets,” in *7th ACM Conference on Computer and Communications Security*. Athens, Greece: ACM Press, 2000, pp. 184–189.

[44] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, “Practical network support for IP traceback,” *ACM SIGCOMM*, vol. 30, no. 4, pp. 295–306, 2000.

[45] A. Hess, J. M., and G. Schafer, “Combining multiple intrusion detection and response technologies in an active networking based architecture,” in *17th DFN-Arbeitstagung uber Kommunikationsnetze*, Dusseldorf, Germany, 2003.

[46] W. L. Cholter, P. Narasimhan, D. Sterne, R. Balupari, K. Djahandari, A. Mani, and S. Murphy, “IBAN: Intrusion blocker based on active networks,” in *2002 DARPA Active Networks Conference and Exposition*, San Francisco, CA, 2002, pp. 182–192.

[47] H. Meer, R. Temmingh, and C. van der Walt, “When the tables turn: Passive strike-back,” in *Aggressive Network Self-Defense*, N. R. Wyler, Ed. Rockland, MA: Syngress Publishing, 2005, pp. 339–372.

[48] M. Crosbie and G. Spafford, “Active defense of a computer system using autonomous agents,” Purdue University, Technical Report 95-008, February 1995.

[49] J. P. Anderson, “Computer security threat monitoring and surveillance,” James P. Anderson Co.,” Technical Report, April 1980.

[50] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, 1987.

[51] R. A. Kemmerer and G. Vigna, "Intrusion detection: A brief history and overview," *IEEE Computer*, vol. 34, no. 4, pp. 27–30, 2002.

[52] L. T. Heberlein, G. V. Dias, K. N. Levitt, B. Mukherjee, J. Wood, and D. Wolber, "A network security monitor," in *IEEE Symposium on Research in Security and Privacy*, Oakland, CA, 1990, pp. 296–304.

[53] W. Yurcik and D. Doss, "Internet attacks: A policy framework for rules of engagement," in *29th Research Conference on Communcation, Information, and Internet Policy*, L. F. Cranor, Ed. Alexandria, VA: MIT Press, 2001.

[54] M. Kotadia, "Australia vulnerable to Korean hacking army." CNET.com, 2004. [Online]. Available: http://asia.cnet.com/news/security/0,39037064,39197226,00.htm

[55] P. M. Joyal, "Industrial espionage today and information wars of tomarrow," in *19th National Information Systems Security Conference*, 1996, pp. 139–151.

[56] D. E. Denning, *Information Warfare and Security*. Boston: Addison-Wesley, 1999.

[57] M. Erbschloe, *Information Warfare: How to Survive Cyber Attacks*. New York: Osborne/McGraw-Hill, 2001.

[58] W. Yurcik, "Information warfare survivability: Is the best defense a good offense?" in *5th Annual Ethics and Technology Conference*, Loyola University, Chicago, IL, 2000.

[59] J. Barkham, "Information warfare and international law on the use of force," *New York University Journal of International Law and Politics*, vol. 34, pp. 57–113, 2001.

[60] Colonel N. C. Cabana, "Cyber attack response: The military in a support role," 2000. [Online]. Available: http://www.airpower.maxwell.af.mil/airchronicles/cc/cabana.html

[61] R. Crisp, "Ethics," in *Routledge Encyclopedia of Philosophy*, E. Craig, Ed. London: Routledge, 1998, vol. 3, pp. 435–437.

[62] R. M. Hare, "Universal prescriptivism," in *A Companion to Ethics*, P. Singer, Ed. Oxford, England: Basil Blackwell, 1991, pp. 451–463.

[63] P. E. Davis, Ed., *Introduction to Moral Philosophy*. Columbus, Ohio: Bell and Howell Company, 1973.

[64] P. Singer, Ed., *A Companion to Ethics*. Oxford, England: Blackwell, 1991.

[65] J. Rawls, *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press, 1999.

[66] N. Davis, "Contemporary deontology," in *A Companion to Ethics*, P. Singer, Ed. Cambridge, Massachusetts: Basil Blackwell, 1991, pp. 205–218.

[67] D. McNaughton, "Consequentialism," in *Routledge Encyclopedia of Philosophy*, E. Craig, Ed. Cornwall, England: Routledge, 1998, vol. 2, pp. 603–606.

[68] P. Pettit, "Consequentialism," in *A Companion to Ethics*, P. Singer, Ed. Oxford, England: Basil Blackwell, 1991, pp. 230–240.

[69] I. Kant, *Groundwork of the Metaphysics of Morals*. New York: Harper Torchbooks, 1958.

[70] E. Spafford, “Are computer hacker break-ins ethical?” in *Computers and Ethics in the Cyberage*, D. M. Hester and P. J. Ford, Eds. Upper Saddle River, New Jersey: Prentice Hall, 2001, pp. 332–344.

[71] B. Smith, W. Yurcik, and D. Doss, “Ethical hacking: The security justification,” in *5th Annual Ethics of Electronics Information in the 21st Century Symposium*, University of Memphis, Memphis, TN, 2001.

[72] S. Caltagirone, “Criminal law perspectives of contemporary issues in computer security,” University of Idaho, Technical Report CSDS-DF-TR-05-28, 2005.

[73] A. L. Institute., *Model penal code: official draft and explanatory notes: complete text of model penal code as adopted at the 1962 annual meeting of the American Law Institute at Washington, D.C., May 24, 1962*. Philadelphia, Pa.: American Law Institute, 1985.

[74] *United States v. John Michael Sullivan*, ser. Fed. Appx. United States Court of Appeals for the Fourth Circuit, 2002, vol. 40.

[75] *United States v. Richard W. Czubinski*, ser. Federal Reporter. United States Court of Appeals for the First Circuit, 1997, vol. 106.

[76] *United States v. Nicholas Middleton*, ser. Federal Reporter. United States Court of Appeals for Ninth Circuit, 2000, vol. 231.

[77] *United States v. Robert Morris*, ser. Federal Reporter. United States Court of Appeals for the Second Circuit, 1991, vol. 928.

[78] *United States v. Bernadette Sablan*, ser. Federal Reporter. United States Court of Appeals for the Ninth Circuit, 1996, vol. 92.

[79] *Shurgard Storage Centers, Inc. v. Safeguard Self Storage, Inc.*, ser. Federal Supplement. United States District Court for the Western District of Washington, Seattle Division, 2000, vol. 199.

[80] S. Brenner, *Model State Computer Crimes Code*. University of Dayton School of Law, 2001. [Online]. Available: http://cybercrimes.net/99MSCCC/99MSCCCMain.html

[81] *United States v. Schoon*, ser. Federal Reporter. United States Court of Appeals for the 9th Circuit, 1992, vol. 971.

[82] D. W. Barnes, "Judges and legislatures in 21st century torts: Integrating cases and statutes," in *Association of American Law Schools Conference on Torts*, New York, New York, 2003.

[83] *Charter of the United Nations*. The United Nations, 1945.

[84] M. N. Schmitt, "Bellum americanum: The US view of twenty-first century war and its possible implications for the law of armed conflict," *Michigan Journal of International Law*, vol. 19, no. 4, pp. 1050–1090, 1998.

[85] J. B. Michael, T. C. Wingfield, and D. Wijesekera, "Measured responses to cyber attacks using Schmitt analysis: A case study of attack scenarios for a software-intensive system," in *27th Annual International Computer Software and Applications Conference*, Dallas, Texas, 2003, pp. 621–626.

[86] *Convention on Cybercrime*. The Council of Europe, 2001.

[87] S. Caltagirone, "Evolving active defense strategies," University of Idaho, Technical Report CSDS-DF-TR-05-27, 2005.

[88] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: University of Michigan Press, 1975.

[89] J. R. Koza, "Genetic programming," in *Encyclopedia of Computer Science and Technology*, J. G. Williams and A. Kent, Eds. Marcel-Dekker, 1998, vol. 39, pp. 29–43.

[90] ——, "Genetic evolution and co-evolution of game strategies," in *The International Conference on Game Theory and Its Applications*, Stony Brook, New York, 1992.

[91] ——, "Genetic evolution and co-evolution of computer programs," in *Artificial Life*, C. Taylor, C. Langston, J. D. Farmer, and S. Rasmussen, Eds. Santa Fe, New Mexico: Addison-Wesley, 1991, vol. X, pp. 603–629.

[92] R. Axelrod, "The evolution of strategies in the iterated Prisoner's Dilemma," in *Genetic Algorithms and Simulated Annealing*, L. Davis, Ed. London: Morgan Kaufman, 1987, pp. 32–41.

[93] J. H. Miller, "The coevolution of automata in the repeated Prisoner's Dilemma," *Journal of Economic Behavior and Organization*, vol. 29, no. 1, pp. 87–112, 1996.

[94] C. Rosin and R. Belew, "Methods for competitive co-evolution: Finding opponents worth beating," in *Sixth International Conference on Genetic Algorithms*, L. Eshelman, Ed. San Francisco, CA: Morgan Kaufmann, 1995, pp. 373–380.

[95] ——, “New methods for competitive co-evolution,” *Evolutionary Computation*, vol. 5, no. 1, pp. 1–29, 1996.

[96] M. A. Potter, K. A. De Jong, and J. J. Grefenstette, “A coevolutionary approach to learning sequential decision rules,” in *Sixth International Conference on Genetic Algorithms*, L. Eshelman, Ed. San Francisco, CA: Morgan Kaufmann, 1995, pp. 366–372.

[97] T. Haynes and S. Sen, “Evolving behavioral strategies in predators and prey,” in *IJCAI-95 Workshop on Adaptation and Learning in Multiagent Systems*. Montreal, Quebec, Canada: Morgan Kaufmann, 1996, pp. 32–37.

[98] S. Caltagirone and D. Frincke, “The response continuum,” in *6th IEEE Information Assurance Workshop*. West Point, NY, USA: IEEE, 2005.

[99] D. Dittrich, “Active defenses to cyber attacks.” University of Washington Information School: Agora Workshop, September 12, 2003.

[100] W. Schwartau, “Can you counter-attack hackers?” *CNN.com*, 2000. [Online]. Available: http://archives.cnn.com/2000/TECH/computing/04/07/self-defense.idg/

[101] M. Delio, “‘Stung’ Russian hacker guilty,” *Wired News*, 2001. [Online]. Available: http://www.wired.com/news/politics/0,1283,47650,00.html

[102] M. Bishop and D. Frincke, “Who watches the security educators?” *IEEE Security and Privacy*, vol. 1, no. 3, pp. 56–58, 2003.

[103] W. Schwartau, “Striking back,” *Network World*, 1999. [Online]. Available: http://www.nwfusion.com/news/0111vigilante.html

[104] N. R. Wyler, Ed., *Aggressive Network Self-Defense*. Rockland, Maryland, USA: Syngress, 2005.

[105] V. D. Sokolovskii, *Soviet Military Strategy*. Englewood Cliffs, New Jersey: Prentice-Hall, 1963.

[106] *Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms*. Department of Defense, 2001.

[107] *Joint Publication 3-40: Joint Doctrine for Combating Weapons of Mass Destruction*. Department of Defense, 2004.

[108] D. J. Welch, N. Buchheit, and A. Ruocco, "Strike back: Offensive actions in information warfare," in *1999 Workshop on New Security Paradigms*. Caledon Hills, Ontario, Canada: ACM Press, 1999, pp. 47–52.

[109] J. Nathan, *Snort flexresp2 README*, 2004. [Online]. Available: http://cerberus.sourcefire.com/ jeff/archives/snort/

[110] D. Radcliff, "Hack back," *Network World*, May 29 2000.

[111] S. Gaudin, "Plan to counterattack hackers draws more fire," 2004. [Online]. Available: http://www.internetnews.com/article.php/3335811

[112] R. Tadjer, "Detect, deflect, destroy," *Internet Week*, 2000. [Online]. Available: http://www.internetweek.com/indepth/indepth111300.htm

[113] X. Wang, D. S. Reeves, and S. F. Wu, "Tracing based active intrusion response," *Journal of Information Warfare*, vol. 1, no. 1, 2001.

[114] D. Yu and D. Frincke, "Alert confidence fusion in intrusion detection systems with extended dempster-shafer theory," in *43rd ACM Annual Southeast Conference*, Kennesaw, GA, 2005.

[115] D. Frincke and E. Wilhite, "Distributed network defense," in *IEEE Workshop on Information Assurance and Security*, West Point, NY, 2001, pp. 236–238.

[116] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives*. Cambridge, Massachusetts: Cambridge University Press, 1976.

[117] Symantec, "Symantec Internet security threat report," February 2003. [Online]. Available: http://www.securitystats.com/reports/Symantec-Internet_Security_Threat_Report_vIII.20030201.pdf

[118] CERT, "Testimony of Richard D. Pethia before the House Select Committee on Homeland Security: Cyber security - growing risk from growing vulnerability," July 2003. [Online]. Available: http://www.cert.org/congressional_testimony/Pethia_testimony_06-25-03.html

[119] B. Stalbaum, "The zapatista tactical floodnet," 2003. [Online]. Available: http://www.thing.net/ rdom/ecd/ZapTact.html

[120] J. Larson and J. Haile, "Understanding IDS active response mechanisms," 2002. [Online]. Available: http://www.securityfocus.com/infocus/1540

[121] M. Crosbie and E. Spafford, "Applying genetic programming to intrusion detection," in *1995 AAAI Fall Symposium on Genetic Programming*, Cambridge, Massachusetts, 1995, pp. 1–8.

[122] W. M. Spears and D. F. Gordon, “Evolving finite-state machine strategies for protecting resources,” in *the 12th International Symposium on Foundations of Intelligent Systems*. Springer-Verlag, 2000, pp. 166–175.

Appendix A

PAPERS IN RESPONSE AND DEFENSE

The following appendices contain three papers, each discussing a different aspect of active response and providing a unique perspective on the problem. However, while each paper has a slightly different emphasis, each revolves around the same theme, active response, and provides the necessary input in order to develop a cohesive and thorough framework. Each paper will be summarized and discussed in turn[1], while the papers in their entirety are available in the following chapters.

A.1 The Response Continuum

One of the primary problems with the study of active response is that the discussions are very disorganized and do not have a central framework with which to work from. This paper attempts to solve this problem by presenting such a framework which enhances the discussion of active response. The framework develops six of core elements of active response: the need for active response and the previous work on the topic, a common definition of active response, a response taxonomy, a model of response and various representations, issues in evaluating a response, and finally, the eight stages of a response scenario.

Before a legitimate discussion of active response can be mounted, the need for active response must first be expressed. However, with the United States

[1] The papers are not presented in chronological order, but rather in an order that best suits discussion.

Department of Defense (DOD) [100] and Federal Bureau of Investigation (FBI) [101], along with private individuals engaging in active response actions, the question no longer becomes whether active response is a legitimate tool but rather how should active response be used. From the point that the case for active response has been made, the framework moves to a definition of active response.

The definition given in Section B.4.2 is: *Any action sequence deliberately performed by an individual or organization between the time an attack is detected and the time it is determined to be finished, in an automated or non-automated fashion, in order to mitigate the identified threat's negative effects upon a particular asset set*. This definition is purposefully broad due to fact that a response encompasses many actions in many situations.

In order to help with the breadth of the definition, a taxonomy of active response actions is proposed. The taxonomy is developed from the attack target's point-of-view and action categorization is based on the action's scope and effect. The taxonomy contains eight (8) categories: no action, internal notification, internal response, external cooperative response, non-cooperative intelligence gathering, non-cooperative 'cease-and-desist', counter-strike, and preemptive defense. These categories encompass many actions, and similarly an action may fit into more than one category. Further order and subdivisions of this taxonomy is expected in future work.

The taxonomy of actions is an important step because, as a general rule, the greater the interaction with external networks the greater the cost and/or risk.[2] To illustrate the cost of an active response action, two models are utilized, a vector and graphical representation. The vector representation allows an action to be represented as a vector of risks. The vector representation allows certain

[2] Cost and risk in this context denote different elements, as cost is a fixed or known value while risk has an associated probability that all or part of the cost will be emerge. However, the word risk will be used exclusively for readability when the two do not need to be differentiated.

functions, such as $<$, to be defined so that actions can be compared and manipulated in a decision model for the purpose of choosing actions based on risk and other factors.

Additionally, a graphical representation is presented (Section B.4.3), providing an example of how an active response policy determines the action and risk to an organization. It also illustrates the idea of escalating a response based on increasing attacker activities. The graphical representation, while illustrated in two (2) dimensions, can be extended to n-dimensions, where n is the number of risks being associated with an active response action. This allows each of the risks to be modeled independently (instead of combined as illustrated) and the graph can be analyzed based on the influence of specific risk.

However, the vector and graphical representation illustrate potential models for the risk/cost of actions — but what are those risks? The framework answers this by providing a discussion of the numerous issues associated with active response. The framework provides discussion on five primary risk categories: ethical, legal, risk analysis, technical, and unintended consequences. This discussion is meant to frame the discussion of the issues rather than to answer the questions associated with the issues themselves.

After presenting the issues surrounding active response, the framework turns to presenting a planning-centric eight stage response model. These eight stages represent the various stages undertaken by an operator or system in an active response scenario. Presented in Section B.6, the eight stages are: planning, detection, evaluation, decision, action, analysis, escalation, and maintenance. This eight-stage model is further developed in ADAM.

A.2 ADAM: Active Defense Algorithm and Model

Taking advantage of the framework developed in “The Response Continuum” [98], ADAM presents a formal decision model and algorithm of active response.[3] The purpose of such a model and algorithm is two-fold, it (1) further refines and defines the concept of active response, and (2) illustrates an implementable model of active response for organizations to consider. With these motivations, we will now discuss ADAM and its ramifications for active response.

As described in Section A.1, the eight-stage model of response is planning-centric. In the case of this model, planning-centric means that the decision making process relies primarily on on the planning stage and the development of an active response policy. There are two reasons for a planning-centric model, rather than an *ex post facto* (i.e. one that decides during or after an attack) model: (1) the amount of risk analysis necessary to make an adequate decision precludes the ‘real-time’ or ‘near real-time’ response required to mitigate most attacks, and (2) provides some justification after the response that legal due diligence was performed and that the decision was not ad hoc or retributive.

ADAM produces three work products, an **active response policy**, an **escalation ladder**, and an **algorithm** (defined in Algorithm 2) illustrating the use of the policy and escalation ladder. Due to the planning-centric nature of the model, it should be no suprise that ADAM allocates most of its resources to the development of the active response policy.

Before the details of the model are discussed, an important aspect of the model needs to be presented: the scoring chart. The scoring chart is a method of scoring risks, allowing one to compare disparate risks. The scoring chart works by

[3] ADAM refers to active response as active defense, these terms in this context are synonymous. See Section B.4.2 for a short discussion on the use of these two terms.

creating risk categories (five are presented for example: ethical consequence, ethical action, legal, financial, and national security) within which are detailed costs placed on a scale of real numbers from -1 to 1, where $cost(1) > cost(-1)$.

The **active response policy** is divided into two parts, the asset evaluation, and the action evaluation. In the asset evaluation, an organization lists all of its assets that they wish to protect using active response, the threats to those assets, the goal of mitigation, and the risks of those threats to those assets. The risks are then scored using the scoring chart allowing an easy method of comparison. The action evaluation follows the same pattern. For each threat, a list is created that includes potential active response actions, the probability of successfully mitigating the threat, and associated risks. Those risks are then also scored using the scoring chart.

The organization then assigns a utility modifier to each risk category. A utility modifier is a real number (greater than 1), that represents the organization's interest in each risk category. For example, a hospital may weight ethical consequence above national security, so it may assign a modifier of 1.2 to the ethical consequences category while the national security modifier remains at 1. Whereas a financial institution may weight financial risks a 1.5 and ethical action 1.1. The organization then multiplies each risk score by the modifier assigned to that category, thereby producing an active response policy tailored to the unique mission of the organization.

Once the active response policy has been created, then the **escalation ladder** can be determined. The escalation ladder's purpose is to provide a set of appropriate actions that will be used to mitigate the threat (if and when the threat materializes). The escalation ladder, more specifically, is an ordered set of actions that are progressively executed until the threat is successfully mitigated. The set is created by ordering the actions based on a simple formula equally balancing risk and success.

The algorithm described in Algorithm 2 illustrates the use of the active response policy and escalation ladder just created within the framework of the eight-stage response model. The algorithm is instantiated in the evaluation stage after evidence of an attack is detected and follows the model description in B.6. It first checks if the threat is in the policy, and then executes actions in the order described in escalation ladder (checking back to see if the situation has changed) until the threat is mitigated or the accumulated risk is too great.

ADAM provides a pragmatic approach to active response because it is geared towards organizations exploring active response as a legitimate security tool. Additionally, ADAM provides a formal decision model and algorithm for greater discussion on the topic of active response. Overall, ADAM is the first attempt at a formal model of active response implementing the framework described in [98].

A.3 Evolving Active Defense Strategies

Competitive co-evolution is a powerful tool with which to model adversarial strategies. It has been successful in the field of evolutionary computation for some time and continues to be researched. The model of competitive co-evolution is easy to understand. In the model there are two populations of individuals, where each individual attempts to maximize their fitness and procreate in order to pass their strategy on to the next generation. The fitness of an individual is determined by the strength of their strategy when played against each of the individuals in the opposing population. The individuals evolve their own strategy by utilizing the genetic operators of mutation, selection, and crossover.

In computer security, there is a very strong adversarial relationship between the defenders of electronic assets, and the attackers who wish to exploit them. If this relationship can be successfully modeled, then competitive co-evolution can be utilized in order to develop strategies for the protection of the assets. Because competitive co-evolution has been successfully used to develop game strategies (e.g. Prisoners Dilemma [93], TIC-TAC-TOE and Go [95], etc.), the attacker/defender relationship was also modeled as a game in order to take advantage of the previous research more directly. The details of the game algorithm can be examined in Algorithm 3.

In the game, there are two set of actions, one set for attackers and one set for defenders — both sets share the null action, or an action that symbolizes no action taken. The game is initialized by populating each individual with a random set of actions (each individual begins with the same number of actions, but the actions and placement of the actions are random). Each defender is then paired with each attacker in order to calculate the fitness of the individuals.

Once the individuals are initialized and an attacker and defender are paired, then the game begins. The attacker executes their first action and the defender then executes each of their actions in turn in an attempt to find an action that mitigates the attacker's action. If an action is found that permanently stops the attacker, then the attacker is done; otherwise, the defender only successfully stops the current attack action or does not stop the action — in either case, the attacker then executes the next action and the defender goes through their actions again. In any case, the cost of the defender's executed actions are combined, and if the attacker's action is successful its cost is also combined. These totals are then summed to provide the individual's fitness — the attacker's goal is to maximize the cost to the defenders, and the defender's goal is to minimize the cost to themselves with the better individuals having a greater probability of being selected to reproduce.

Using this game and a competitive co-evolutionary model some valuable conclusions were derived. First, the fitness of the populations successfully model a pattern found in most competitive co-evolutionary models validating the implementation. Second, the model produced reasonable and effective strategies for both the attackers and defenders — providing validation that this technique is successful in developing active response strategies. Overall, this model provides a primitive baseline with which to launch a more formal and thorough analysis of competitive co-evolution as applied to computer security strategies.

Appendix B

THE RESPONSE CONTINUUM

S. Caltagirone and D. Frincke, "The Response Continuum," to appear in the proceedings of the 6th IEEE Information Assurance Workshop, West Point, NY, USA, 2005.[1]

B.1 Introduction

Response is implicitly present in every security defense system, whether that response is to inform a system administrator, to close off access, to involve law enforcement, or even to ignore the misuse. A reasoned debate as to how to determine an acceptable level of response seems appropriate. However, only recently has "response" [2] been addressed with academic rigor by the mainstream research community. We propose a framework for evaluating response possibilities both qualitatively and quantitatively. This framework provides a common starting point for response discussions with an emphasis on active response as an integral part of contemporary computer security.

Many factors may have influenced the hesitation in addressing response. First, researchers have logically argued that it is of primary importance to reduce system design vulnerabilities and/or accurately and rapidly detect misuse, as a precursor to formal response. Second, experimenting with the more extreme forms of response is difficult to do safely within most university environments,

[1] Republished with permission from IEEE.

[2] Alternately called "active defense."

and has associated costs: increased supervision of students, separation of equipment from the mainstream, potential for bad publicity, harm caused if experiments overflow the university testbed, to name a few. Too, some aspects of the discussion of response techniques are akin to those about whether students learn more from a defensive, or an offensive, philosophical approach [102]. For researchers teaching from a defensive posture, it would be somewhat foreign to study response in detail and correspondingly natural to incorporate primarily protective measures into their research and classrooms. Further, response has tended to be "folded in" as a sideline to detection, rather than designed in as a key characteristic of a protective system, and this may also have led to the lack of focused attention upon active response. Finally, some equate response research with *advocacy* of extreme forms of response. These are all valid concerns.

However, just because a discussion is awkward does not mean it should be avoided. There is a growing frustration among some security practitioners, many of whom are increasingly dissatisfied with the effectiveness of current remedies. This frustration has led a few to take more aggressive measures [8, 103, 104]. Therefore, we believe it is incumbent now more than ever for the academic community to seriously re-examine the question of response, and to lead the public debate regarding where/when/why various forms of response are appropriate.

To support this debate, we offer three things:

- a common definition of response,
- an ordered continuum of response actions, and
- a straw-man schema for evaluating choices among possible response actions.

B.2 The Need for Response

Response has long been incorporated as part of most good defense strategies, and computer defense is no exception. We note some uses of “response” to computerized threats. In 1998, the US Department of Defense, while responding to an attempted denial-of-service (DoS), launched an applet that shut down the browsers of attackers preventing them from attacking further [100]. In 2001, a U.S. District court judge allowed the FBI to compromise a Russian hacker's computers and install a keylogger to gather evidence on his illegal activities regarding computers inside the U.S. [101].

In the context of decisions such as these, it is legitimate to ask: why were these particular response actions chosen? Were they at the appropriate level of force given the context of the perceived threat and the degree of certainty of the defenders about their circumstance? What was the external cost of these response decisions and did it match what was predicted? The problem is that there is no agreed-upon tool, technique, method, standard or policy, upon which one can rely for making good response choices when a threat is identified, perceived, or predicted. Currently, those responsible for defending the front lines of most computing systems rely heavily on relatively ad hoc policies or instincts guided by experience and intuition, rather than a well-defined scientific model.

Any ad hoc method of making response decisions is most unsatisfactory, and is particularly unacceptable when system value is high and/or availability cannot be interrupted, as in life/safety/national security critical systems. Ideally, it would be replaced with a rigorous and scientific model that can support appropriate response selection. With the growing number of exploitable vulnerabilities in critical systems (e.g. air traffic control [4], nuclear power safety systems [5], etc.), and corresponding pressure to protect such systems, the immediate need for such a model is evident.

B.3 Related Work and Previous Discussion

Much discussion of response has emphasized hack-back,[3] and so we begin by outlining the essence of that discussion. In 2002, Jayawal, Yurcik and Doss called for more effective ways of protecting networked systems from attack and examined the possibility of hack-back [30]. In the same year, Mullen presented justifications for strike-back at defcon [31], wrote a corresponding article at SecurityFocus [8] and published a whitepaper on strike-back [7].

Mullen's work inspired a Reuter's news article about strike-back, especially the ethical and legal implications [32]. As well, researchers such as Schneier criticized Mullen's position, using an analogy to the Recording Industry Association of America's (RIAA) attempt to attack copyright infringer's computers [9]. Mullen responded in [33]. However, active response considers a greater range of actions than only hack-back.

Some researchers emphasize the response decision-making process. Loomis, in [34], while implying hack-back, presents an objective discussion of the ethical and legal aspects of response decisions. Grove, et al., distinguishes 'active defense' from passive defense and undertakes a thorough discussion of international law implications of active defense [35]. Bruschi and Rosti [36] discuss a response strategy for DoS attacks in which they limit the capabilities of the attacker rather than strengthening defenses. Additionally, they provide AngeL, an implementation of their strategy [37,38]. [39] provides an organizational model and structure for codifying how decisions about response could be made.

[3] Hack-back: retaliating against the attacker using techniques that share many attack characteristics (a.k.a. strike-back)

B.4 Active Response: Definitions and Models

B.4.1 Characteristics of a Definition

We propose the following requirements for any definition of active response:

- Active response is *time-bound*, and takes place exclusively during a period when an attack is known[4] to be in progress.

- Active response is *purposeful*. Actions would only be considered 'response' when used to mitigate the threat for the purpose of returning to a more secure state and no further.

- Active response, being purposeful, has *limitations.* Threat mitigation need not mean threat elimination, but instead indicate that the actions seek to diminish or contain the threat with respect to the resource considered to be threatened. Neither punitive responses nor warning responses, which might be intended to have the general effect of reducing threats from a broad perspective, are not included in our definition of active response.

- The decision to apply specific active response actions is *controllable* and *deliberate*,[5] though an active response action sequence's consequences might be neither.

- Active response is comprised of a *sequence* of actions.

[4] Accuracy of detection clearly has a role to play. The degree of certainty to which an attack is known to be in progress certainly has a role to play. We have chosen not to require that aspect of the issue in our initial identification of required characteristics of active response: it is an area for future work.

[5] The decision to apply certain responses when certain threats are identified could still be made in advance of the particular threat, as with an automated active response.

- Active response is *technologically independent* and can be executed by an operator or automatically by an intrusion detection system using a pre-defined ruleset.

- The timeline of active response is to be considered considered *subjectively*, from the perspective of the decision-making entity, and not from any other body.

We have been deliberately general in some of these requirements. For example, when precisely an attack should be considered "detected" or to have "finished" depends on several factors and in our opinion is again a matter for the academic, legal, and other stakeholder communities to discuss. We identify some of the relevant factors in [39].

B.4.2 Definition

This leads us to the following definition of active response.

Definition 1. **Active Response**: *Any action sequence deliberately performed by an individual or organization between the time an attack is detected and the time it is determined to be finished, in an automated or non-automated fashion, in order to mitigate the identified threat's negative effects upon a particular asset set.*

How well does this match our requirements?

- **Time bound** and considered **subjectively**. Active response takes place only when the organization believes the attack to be in progress.

- **Purposeful.** The reason for the action sequence must be to address a specific problem, and the organization must choose to take these particular actions.

- **Limited.** It is sufficient for the response to be intended to improve the situation.

This definition disallows reactions to threats such as retaliation and retribution, even though these might be considered by some to have deterrent value and thus be "response." We believe that it would be better to consider policies about actions in those categories separately, as the issues involved are significantly different. Also note that the term 'attack' is used, but no implications as to the motivations behind the security event are made.

We end this section by noting the prevalence of alternate terminology to active response: active defense. While *active defense* may well be more descriptive of the rationale behind the *actions* being taken, the phrase active defense has a difficult history of prior use in the military [105–107]. Militaries use the phrase active defense as meaning limited offensive action to deny a contested resource or position, implying target destruction. Since we are allowing a continuum of potential responses in our definition, the phrase active response seems more suitable.

B.4.3 A Model of Response

When a system administrator chooses to manually respond to a threat, normally this is not done in a single "response", but rather in an almost interactive collection of actions. More formally, response is usually realized as a temporally ordered sequence of actions executed in a manner that supports feedback regarding the actions' effectiveness through analysis and continued detection.

For example, an attack may be suspected and actions $a_1, a_2, \ldots, a_k$ are chosen for mitigation. If the confidence level in the detection is low, then the $a_1, a_2, \ldots, a_k$ would, we anticipate, be relatively benign. Suppose, though, that ongoing analysis of the suspected attack yields new information, which leads to a greater certainty. The defender might then choose to perform additional actions $a_{k+1}, \ldots, a_{k+n}$. If the results are unsatisfactory, and/or the risk increases, a stronger action a_{k+n+1} may be taken next.

How actions $a_1, a_2, \ldots, a_k$ are chosen is dependent on the decision model. One possible decision model is our prior work on ADAM [39] summarized here. ADAM supplies a framework an organization can use to inform their decisions as to the active response actions to execute. ADAM weighs multiple categories of organizational 'cost'[6] with the relative probability of success of the action to mitigate the threat. This model allows organizations to balance cost with success to obtain a sequence of graduated responses — and to indicate reasonable cutoff points for when those responses are useful. These principles of cost and success should be a guiding factor in action choice, although their balance is arguable (see [108] for a thorough discussion of military theories to guide offensive action, for example).

In the following subsections, we provide both a graphical and a vector representation to illustrate active response costs. We believe that this representation can be helpful to clarify the relationship between a response policy and the active response measures taken to mitigate a threat.

Vector Representation

We propose a vector representation of the cost categories, which can be mapped onto our model, to permit direct comparison of the costs. Let $\overrightarrow{C}$ = $< C_{ethical}, C_{legal}, C_{risk}, C_{technical}, C_{unintendedconsequences} >$. Then $\overrightarrow{C}$ is the cost vector of all of the components making up the "costs" of the risks of a particular active response action.

We can use this in several useful ways, which we will only enumerate here:

1. For a given entity wishing to make a judgement about a particular action,

[6] Cost here has to do with predicted loss to the organization for employing active defense — whether financial, legal sanction, or violation of corporate ethics.

we can establish a $\overrightarrow{CMAX}$, where $\overrightarrow{CMAX}$ is the vector of the maximum cost we're willing to accept in each category: $< C_{ethical}, \ldots >$.

2. We can define an ordering function $\leq$, where $\overrightarrow{C} \leq \overrightarrow{T}$ (where both are cost vectors for active defense actions) iff each of $C_{ethical} \leq T_{ethical}$, $C_{legal} \leq T_{legal}$, $\ldots$

3. We can establish a time-sequence or environment-dependent version of the C_{MAX}, so that at a given time or in a given environment we change the thresholds allowed for the costs. This allows any changes in tolerance for active defense results to be recorded over time, or over environmental change.

4. We can establish a weighting vector W, where $W_{ethical}$ has the weight associated with the value placed on the ethical components, likewise for legal, financial, etc. This allows different organizations to weight costs based on their own particular goals and mission. The costs which an organization wishes to include are dependent on the organization's own situation and environment.

Graph Representation

On the Y axis we represent the cost of an active response action; where as Y increases, the cost for the responder increases. The X axis represents the cost of the attacker's action to the asset; where $X < 0$ are attacker activities before the initial compromise (e.g. port scanning, vulnerability scanning, and other intelligence gathering techniques), $X = 0$ is the compromise event, and $X > 0$ are attacker activities after the initial compromise (e.g. adding users, copying data, installing back-doors, etc). The line that connects response actions is determined by the policy (as described in Section B.6) and is referred to as the ***policy***

determinate because the policy is determining the actions, which by proxy determines the cost to whomever is responding.

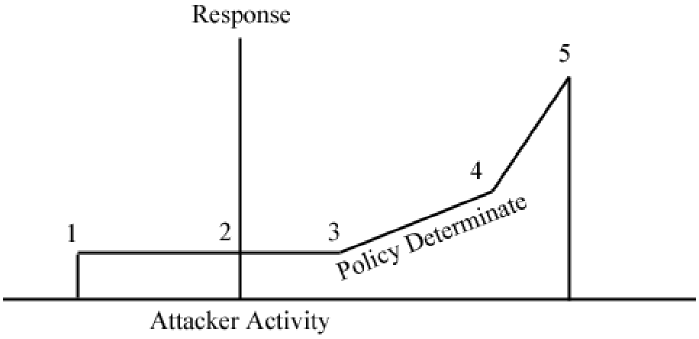


Figure B.1: An Example Response Continuum: (1) Attacker found portscanning, IDS utilizing more resources to watch for threat (2) attacker compromises system, (3) IDS detects intrusion, alerts internal and external authorities (4) files being copied, network tools used to determine source of threat (5) firewall rule changed, counter measures stop intrusion

B.4.4 Towards a Response Taxonomy

This section provides a top-level view of a taxonomy for organizing active response actions, loosely based on the degree of control and the scope of the possible effects of what the responder does. This taxonomy has been adapted from [39] and [99]. We anticipate future work that will further subdivide this taxonomy or ordering.

1. *No Action*: A threat is detected, but no action is taken.

2. *Internal Notification*: Using the organizational structure to notify the designated responder(s) of an active response situation.

3. *Internal Response*: Applying active response actions within the domain over

which the responder has authority (e.g. close a threat vector's associated port).

4. *External Cooperative Response*: Employing entities external to the responding organization to mitigate a threat.

5. *Non-cooperative Intelligence Gathering*: Using external services (e.g. finger, nmap, netstat, etc.) to gather intelligence on the threat. Sometimes referred to as "look but don't touch."

6. *Non-cooperative 'Cease and Desist'*: Stopping harmful and unauthorized services (e.g. zombie control processes) without compromising legitimate usability.

7. *Counter-strike*: An external action to reduce or deny the capabilities of an attacker to continue the attack.

8. *Preemptive Defense*: With knowledge of a forthcoming attack, execute active defense actions to preempt (and disable) the upcoming attack.

Examples of the Response Continuum

There are several publicly available systems which implement response all along the response continuum outlined in the previous section, ranging from rudimentary support for notification through relatively sophisticated and/or powerful responses. To illustrate our response continuum, we outline a few of them here.

Internal Notification Snort,[7] the open source intrusion detection system (IDS), is a very popular notification-based system. Most of other IDSs include a

[7] http://www.snort.org

notification feature.

Internal Response Internal response can come in the form of firewall rules, detaching a machine from the network, or simply destroying TCP connections, among other actions. An extension of Snort is flexresp2, which is an active response tool that terminates TCP connection attempts [109]. Snort can also be run in 'inline' mode allowing it to drop or modify packets that are flagged as malicious.

Automatic or Manual External Cooperative Response Relatively few systems have automated external cooperative response. However, they do exist. DShield,[8] a distributed IDS which collects and analyzes firewall logs from several commercial products, has implemented a 'Fightback' feature that allows an ISP to be notified if firewall logs show attacks emanating from their network.

Non-cooperative 'Cease and Desist' These methods are difficult to implement properly because they must stop harmful services without impinging upon the usability of a network or host. However, there are tools which are reasonably successful within limited domains. Bindview devised ZombieZapper, which acts as a zombie master in to seek shutdown all zombie bots in a network [110].

Counter Strike Use of counter strike is highly controversial. Lycos recently released, then retracted, the 'makelovenotspam' screensaver with some counter strike characteristics. The Lycos screensaver would continually request information from a list of websites known to send spam — effectively creating a pseudo-zombie army to launch a DDoS attack against spam sites [10]. However, ISPs began blocking this traffic as they would any DDoS attack, and public

[8] http://www.dshield.org

pressure mounted until Lycos ultimately removed the product from service. Additionally, Symbiot Inc. created iSMS, which can run the gamut of responses, from blocking traffic, to denial of service attacks, to gaining administrator access on an attacker's machine [111].

Preemptive Defense One example of preemptive defense was designed cooperatively by Mazu Networks and Asta Networks. Their tool, from an ISP's edge router, can detect and block a denial of service (DoS) attack [112] before it leaves their network. Although the attack has already been launched, from the perspective of the potential victim the attack was preempted.

B.5 Evaluating a Response Sequence

We define response evaluation using five components: ethical, legal, risk analysis, technical, and unintended consequences. Each of these can be assessed as a "cost", though "cost" in the case of ethics, etc, should be read as "relative value" rather than "money." This section briefly outlines these evaluation areas.

B.5.1 Ethical

One hotly contested issue surrounding active response decisions is whether certain response actions are ethical, particularly those with the potential to overflow the boundaries of the responder's domain of responsibility. As an example of the distinctions drawn, many who would consider launching a DoS attack against an attacker's firewall unethical would consider modifying one's own firewall rules ethical. The key is that the calculus of response decisions should include components that allow such distinctions to be drawn clearly and rationally, in a way that can be supported within a logical framework. This is particularly important because the speed with which most responses would have to be launched will necessarily lead to automation, so codification of the ethics involved should be done in advance.

There are two frameworks we might utilize in the ethical debate: the teleological (only consequences matter) and deontological (only the actions and types of actions matter). Rawls in [65], and Davis in [66], both argue that teleological and deontological theories "exhaust the possibilities regarding theories of right action." There are strong arguments made using both approaches. Spafford [70], argues from deontology that offensive action would be unethical. Welch et al [108] argue from the teleological that if life, safety, or national security critical systems were significantly threatened, offensive action could be supported.

The key, however, is to have these discussions and make these determinations *a priori*, and integrate them within the decision making process.

B.5.2 Legal

The legal risks to active response are significant. Not only are there questions of the applicability of existing law to a cyber domain, but also the question of which actions are permitted under the environment governing the decision-maker. This component of the calculus incorporates assessment of all facets of law, criminal, civil, domestic, international and foreign domestic. The international question is particularly difficult because the nature of networks do not (generally) limit themselves to national boundaries; and if an action is taken in another nation, then the question is if the action constitutes a 'use of force' and what diplomatic repercussions there may be.

Information warfare best informs the international issue. Grove et al [35], Barkham [59], and Yurcik [53, 58] have all contemplated and analyzed the international question. Additionally, Schmitt proposed the 'Schmitt Analysis,' which is a useful tool in determining whether a cyber attack is a 'use of force' in international law [84].

With regards to domestic law in the United States, the Computer Fraud and Misuse Act (§18 USC 1030), the Wiretap Act (§18 USC 2511), and the Electronic Communications Privacy Act (§18 USC 2510), with corresponding case law, are the primary guides. There are also state statutes regarding computer trespass (see Rhode Island §11-52-3, Virginia §18.2-152.5, and the University of Dayton School of Law Model State Computer Crime Code §4.01.1 [80] for examples).

Some additional areas for consideration include the 'necessity defense' (a.k.a. choice of evils justification) (see Model Penal Code §3.02 [73]) and the 'use of force in defense of property' state statutes (see Utah §76-2-406 and North Dakota §12.1-05-06 for examples). Extending the use of force in self defense where 'self' encompasses electronic assets, is difficult. However, there are certain legal theories that may be helpful: minimal force, proportional force, and immediate (or immanent) threat. These theories are supported in both United States domestic law and International Law (Article 51 of the UN Charter [35] and the Model Penal Code §3.02 [73]). Further identification of risks and identification of the appropriateness of categories of active response can be expected from the legal community.

B.5.3 Risk Analysis

Risk analysis must be performed properly if it is going to serve as the basis of decisions about response. However, can a satisfactory risk analysis be accomplished? Can the ethical and legal risks be realistically evaluated in the face of enormous unknowns? This question should frame ongoing research — much information would be gained if a few organizations attempt the task and report on the outcome via case study, or researchers propose models to examine in abstract, or a combination.

Given the challenges of risk analysis we propose two limiting factors. One, that the extent of risk analysis performed be treated as a due diligence requirement. Second, the risk analysis may not include ALL risks. Accuracy of prediction normally is reduced the further forward one looks, and the ripple effect of an event may take it well out of the range of what an organization is capable of assessing. In practice, risk analysis will be limited to a finite time-frame. This pragmatic consideration has implications on the ethical side (is it ethical to take external actions for which risk analysis cannot be performed properly?) and the legal side (what is due diligence in this context?).

B.5.4 Technical

There are a number of technical issues when it comes to active response. Some of the more significant questions are:

1. Do contemporary response-triggering systems provide enough confidence in their alerts to base particular responses upon? Is there enough information incorporated in the alert to support response?

2. Can response be quick enough to be effective?

3. Is identification and authentication of the attacker/attacker's resource via trace-back[9] and other means viable in this environment, particularly given the prevalence of anonymity techniques and utilization of "innocent bystander" resources?

In essence, technical questions of active response can be summarized by asking whether our technology is reliable and accurate enough to execute

[9] Trace-back is the attempt to find the attacker by tracing their network traffic through the network signaling equipment that composes contemporary networks.

response, particularly the more extreme forms of response on our continuum, or if the modern network environment precludes use of certain forms of response. We note that changes in technology make evaluation of this component of the response decision calculus one that changes rapidly; there are advances in intrusion detection systems (see [113, 114]) and network tracking (see [42–44]) and evaluation of confidence level that may cause evaluations in this category to change significantly.

B.5.5 Unintended Consequences

Unintended consequences are placed in a separate category in our calculus of response, primarily because concern about these is a key issue for response. Discussions of unintended consequences could be placed under legal, ethical, or technological - but because of the magnitude of the issue, we prefer to treat it separately as well, to emphasize the importance of analysis in this area.

The costs of unintended consequences derive from:

1. An unintended (counter) response elicited from the attacker (i.e. you want them to stop, but an unexpected result occurs — perhaps behavior escalates or is diverted)

2. Damage to the perceived source of the threat excluding the attacker (i.e. damage to a zombie or co-opted system)

3. Unplanned damage to the responder's domain.

These sources will now be discussed.

Unanticipated Attacker Response

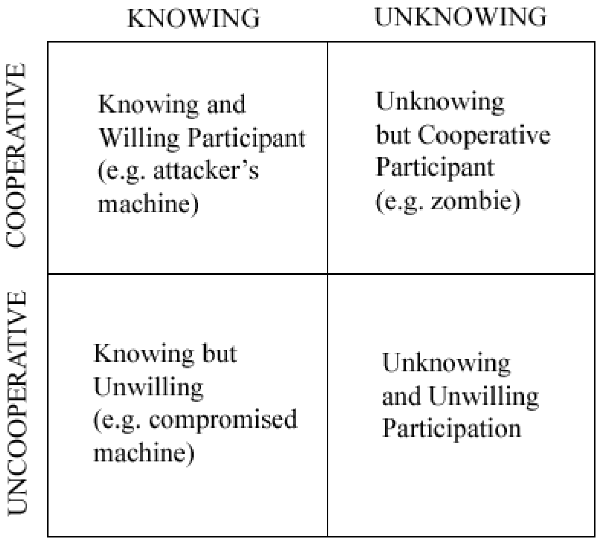
It is possible that actions executed to mitigate an attack will alter the attacker's behavior. As with any form of self defense, there is always the possibility that resistance will lead to escalation (though it may also lead to cessation of the attack). Some examples of attacker tactic change in the cyber domain: they divert their attack to another resource, they become angry and launch a more vicious attack, or even alert other attackers to join the assault. Without sufficient knowledge of the attacker, such as their abilities, their contacts, and resources at their disposal, this variable is outside the responder's control, and is an area where additional external information would be useful in informing a decision.

However, deception, diversion or tactical change may be useful also. The attacker could be diverted to a honeypot/honeynet or other disposable resources — which then may provide additional data as to the source of the threat and better inform any future actions against this attacker.

Damage to Non-Attackers

Accurate tracking, an identification/authentication aspect mentioned in the technology section is difficult to ensure. Attackers may utilize Internet Protocol (IP) spoofing, Media Access Control (MAC) spoofing, and the use of zombies, handlers, or other proxies, for instance. Any action taken outside of resources controlled by an organization or cooperating entities necessarily involve uncertainty with regards to whether the attacker has been correctly identified, and also which of the resources associated with the attack belong to the attacker in the sense of ownership and not simply controlled.

For example, it is possible that the resource targeted for response was either incorrectly identified, or was engaged without the knowledge/support of the true owner of the resource. There are real risks that the target of response might be a life, safety, or national security critical system, possibly of more value than the one under attack. A byproduct of active response could be an increase of threat to critical systems, to be used as shields from active responses. This unintended



	KNOWING	UNKNOWING
COOPERATIVE	Knowing and Willing Participant (e.g. attacker's machine)	Unknowing but Cooperative Participant (e.g. zombie)
UNCOOPERATIVE	Knowing but Unwilling (e.g. compromised machine)	Unknowing and Unwilling Participation

Figure B.2: Attack Participation Taxonomy

consequence has both ethical and technological effects, with the potential for legal as well.

Damage to Own Resources

The third source of unintended consequence costs come from effects upon one's own resources. This risk is common whenever an organization makes a change in policy or design. The risk is that by changing a policy or design, the responder may unintentionally block legitimate users from a resource or harm internal assets, causing even more damage than the original attack.

B.6 The Eight Stages of Response

We propose study of the response process in eight stages: planning, detection, evaluation, decision, action, analysis, escalation, maintenance. These stages also meet our goals for the response definition in Section B.4.1.

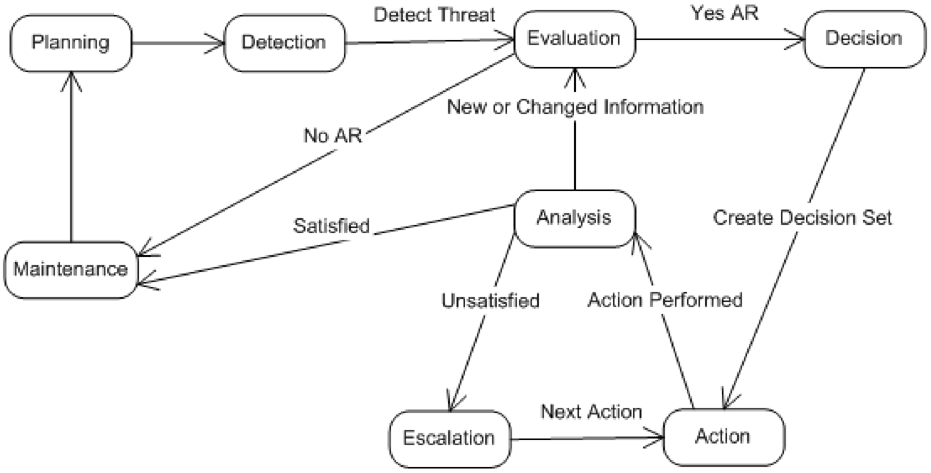


Figure B.3: The Eight Stage Response Cycle

B.6.1 Planning

Every stage of the response process is informed by the policy developed in this first stage, giving a planning-centric model which we believe best serves response. Because of the risks and unknowns involved in assuming an active response position, this stage requires the greatest investment. However, with proper planning as incorporated with risk analysis, the risks can be reduced and active response more likely to be a viable option, or else a reasoned decision to limit response to low risk categories can be made.

In the planning stage, an ***active response policy*** is developed. The policy is an unambiguous and complete analysis of the risks and costs (in every category) of each threat and potential mitigating active response action allowing the risks and costs to be compared. The policy takes as input: the assets to be protected, the threats to those assets, the risks/costs if the threats were successful (or partially successful), and the potential mitigation active response action and corresponding risks/costs. After the inputs have been provided, then, using a scaling method, such as ADAM [39], the actions are ordered based on their relative probability of success and risk. An important step is for each threat to be assigned an unambiguous goal that defines the state at which a threat is successfully mitigated.

To develop the policy, we would anticipate involvement of an array of stakeholders, internal and external. The output of this process should be a clear analysis of all of the risks and costs involved with each asset and potential mitigating action allowing for the most informed decision.

B.6.2 Detection

Detection is the discovery of a threat, whether automated or unautomated. In most instances it is preferable that detection occur early in the attack (e.g., during portscanning) or at least prior to damage (e.g., when an attacker moves past target identification phase). Detection ideally will provide sufficient high-confidence data with regards to the origin, method, and target of the threat to enable response — degree of confidence required for individual response options would be a matter of policy, and considered during the next phase, Evaluation.

B.6.3 Evaluation

Evaluation takes the threat data provided by the detection stage, and compares that to the policy developed in the planning stage. The calculus of response decisions is used to identify whether active response is appropriate, and then the next step is either the decision stage (choose a response), or else the data is passed elsewhere and the cycle moves to the maintenance stage.

B.6.4 Decision

The decision stage determines exactly which actions, identified in the active response policy as potential mitigation techniques, are selected for execution. The actions selected are placed into the ***decision set*** $(a_1, a_2, \ldots, a_k)$, an ordered set of actions that will execute in sequence until the threat is mitigated to satisfaction.

B.6.5 Action

It is this stage that the active response action determined by the decision set is executed. After execution of an action, response moves to the analysis stage.

B.6.6 Analysis

The purpose of the analysis is to determine whether the threat was successfully mitigated (which will invoke maintenance), or the action was unsuccessful (which will invoke escalation) according to the policy set in stage one. Analysis includes considering whether environmental changes require (re)evaluation.

B.6.7 Escalation

Escalation refers not necessarily to increased force,[10] but rather to executing the next action described in the decision set. Escalation comes when action is unsuccessful in providing sufficient mitigation. In the escalation stage, the next action in the decision set is selected and fed into the action stage for execution.

B.6.8 Maintenance

Maintenance is the final stage of the response cycle. Maintaining an effective active response policy is essential when implementing policy-based response. The

[10] Whether a gradual increase in force is utilized, is determined by the decision model chosen.

primary goal of the maintenance stage is to take as input the results of the evaluation or analysis stage and review the policy in view of any forensic or post mortem analysis that occurs after an active response (or failed evaluation) and update the policy accordingly.

B.7 Conclusions and Future Research

There are several opportunities for future work indicated by this work. The authors would like to first identify the need for greater discussion forums for this topic. Each of the risk categories identified in Section B.5 needs more attention and potential solutions identified. Intrusion detection systems need to support response through increased information required for response decisions, including confidence levels associated with alerts. Lastly, the authors acknowledge that the taxonomy and model presented here are only a starting point and more discussion is needed regarding the sufficiency or deficiency of these elements.

Appendix C

ADAM: THE ACTIVE DEFENSE ALGORITHM AND MODEL

S. Caltagirone and D. Frincke, "ADAM: Active Defense Algorithm and Model," in Aggressive Network Self-Defense, N.R. Wyler and G. Byrne, Eds. Rockland, MD, USA: Syngress Publishing, 2005, pp. 287-311.[1]

C.1 Introduction

> *"Security against defeat implies defensive tactics; ability to defeat the enemy means taking the offensive."* [4:5] - Sun Tzu, The Art of War

Active defense is, as Sun Tzu so eloquently phrased it, the "ability to defeat the enemy [by] taking the offensive." There are many possible interpretations of this remark, from a belief that it advocates disabling the enemy through attack, to a milder approach that emphasizes preemptive reduction of the enemy's ability to perform attacks. While neither may be an accurate reflection of Sun Tzu's meaning, modern system defenses are increasingly adding the capacity to *defend* to the capacity to *detect*. Some possible defensive actions have similarities to traditional "attacks" on a computer system, including the potential to have effects outside the defending system's boundaries. When, if ever, should such methods be employed?

In this paper, we propose a model for making decisions about the selection of defensive actions. We informally define active defense as *any action sequence performed by an individual or organization between the time an attack is detected and the*

[1] Republished with permission from Syngress Publishing

time it is known to be finished, in an automated or non-automated fashion, to mitigate a threat against a particular asset. We incorporate assessment strategies for "cost" and "benefit". Our intent is to be general enough to support decisions about defense for organizations and individuals who seek an additional security mechanism with which to protect themselves from constant attack, or to protect a significant asset; for example, a medical facility whose patient databases are constantly being probed for vulnerabilities.

There are substantial risks when an organization takes an active role to stop attackers, and decisions regarding whether those risks are acceptable require planning as well as an understanding of the implications of the actions being considered. Without a model, organizations that assume a position of active defense may unknowingly take on unacceptable risks. We propose the following criteria for any decision model selected:

1. It should allow an organization to create an active defense policy and escalation ladder tailored to internal priorities as well as required and/or desirable external criteria.

2. It should provide organizations with a sense of confidence that they have properly assessed the acceptability of the risks involved in engaging in active defense activities.

3. It should lend itself to automated response.

This paper proposed for consideration a preliminary model, herein called ADAM (Active Defense Algorithm and Model). Note that it is presented for study, and not for employment in an organization at this stage of its development.

The remainder of the paper is divided as follows. First, we provide an outline of the five essential aspects of active defense. Second, the paper will define the goals and assumptions of ADAM. Third, the model itself will be defined and described. Fourth, an algorithm will be presented, which utilizes the model and provides an example of the application of active defense. Lastly, the ADAM model will be analyzed with respect to the stated goals and assumptions.

C.2 Active Defense

Before describing any model of active defense, a definition must be agreed upon:

Definition 1. **Active Defense** *any action sequence performed by an individual or organization between the time an attack is detected and the time it is known to be finished, in an automated or non-automated fashion, to mitigate a threat against a particular asset.*

We emphasize the following aspects of this definition.

1. *Active defense is time-bound.* The active defense action sequence includes only those actions that take place during the time that a specific attack is believed to pose a threat. Both preparatory defensive activities as well as post-mortem forensic analysis are specifically excluded from the active defense action sequence.

2. *Active defense is purposeful.* The active defense action sequence is performed to preserve something the defender considers to be an asset. Note here that an *asset* is anything perceived as providing a positive benefit to the defender and is not limited to those benefits under the defender's direct control. In particular, the asset is not necessarily owned by the defender.

3. *Mitigation does not require elimination.* The use of the word 'mitigate' in the definition does not require that the source of the threat be eliminated. It would be sufficient if the threat is diminished or contained, for example, an active defense strategy might instantiate temporary protections that would preserve the asset in question during the attack (e.g. changing an IP address or DNS entry during a Denial of Service attack).

Under this definition, then, the goal of the organization is to act only until the protections around the threatened assets have reached a predetermined protection threshold — in other words, the threat [2] to assets has been sufficiently mitigated. This predetermined protection threshold is organizationally determined and can encompass a wide tolerance for risk, each associated with a different protection goal. Protection goals might range from 'remove the threat to the asset until a more permanent solution can be found' to 'permanently remove the threat'. To achieve the first of these, the active defense sequence used to protect a service might include changing the port that this service is running on, or blocking traffic to the port at the firewall. To achieve the second might involve stronger measures aimed at disabling the attacker. The key point is that the goals of an active defense action, and by proxy the protection threshold, must be established in light of an organization's own context, capabilities, and values; which should be tempered by external variables such as time, resources, perception of consequences, regulations, and cooperation of upstream providers.

Additionally, numerous actors can participate in a given active defense action, particularly in the case where the threatened asset is perceived as beneficial to many organizations. These actors could be autonomous agents in an intrusion detection system [115], or system administrators working together, using phones and email to synchronize. The decision how to divide the active defense sequence between actors, and in which circumstances they would coordinate, is for each organization to define based on their own security policy and organizational structure. Regardless of whether participation in active defense is autonomous or live, individual or multiple, the core stages of active defense remain the same.

The eight core stages of active defense are: planning, detection, evaluation,

[2] Threat is utilized to distinguish between the actions executed in defense of an asset by an organization, and the actions of an attacker against an asset; and can be meant to encompass all actions of an attacker to reach a goal.

decision, action, analysis, escalation, and maintenance. These stages can be formally or informally defined by an organization.

C.2.1 Planning

Planning should be done well before any attempt at implementing active defense is made by an organization. An active defense plan includes two components: an active defense policy and an escalation ladder. Both components support an approach that balances the risks acquired by assuming a position of active defense against the benefits of so doing. Unfortunately, at present, planning is not always employed in active defense — anectdotally it appears often when the 'stronger' techniques that have been employed were the result of an angry or frustrated operator and not the result of corporate strategy.

Active Defense Policy

The *active defense policy* describes the assets to be protected by active defense, includes an evaluation of the threats (or classes of threats) that exist against those assets, and identifies the value of the asset with respect to the consequences of a successful attack. Additionally, the policy describes the potential actions that can be taken to mitigate the risk of the threat, as well as the risks assumed by conducting an action. Given the definition of an asset as anything of benefit to the organization, this can be a cumbersome step. However, most owned assets should already have been identified during the course of the organization's development of an internal preparation and risk strategy. Organizations that do not already formalize the value of their assets and the methods they use to protect them are likely not good candidates for successful adopters of active defense practices.

Note also that most assets, and most threats, will be excluded from the active defense policy. The selection criteria of candidates for active defense must balance the likelihood that a given threat will cause significant harm against the likelihood of negative consequences (intended or unintended) from adopting an active defense posture, and the potential that a better (or safer) method exists. For example a ’typical’ scan of an HTTP server with no other factors probably does not warrant active defense, and an organization’s spare workstation is probably not an asset worth accepting the potential negative consequences to protect. Too, an organization may rely on its local ISP to maintain connection with its customers, but rather than actively defending that ISP itself, it is probably preferable to find a backup provider or obtain insurance against business loss due to network failure. An active defense policy should be consistent with an organization’s formal security policy that describes valuable owned and utilized assets and the risks associated with damage or loss of those assets, and also tempered with knowledge of an organization’s context and willingness to engage in active defense. It is here that the acceptable protection threshold will be set for each selected asset.

Escalation Ladder

The escalation ladder is an ordered sequence of active defense actions that an organization may consider utilizing with regard to each threat and asset. Each step is ordered according to a predetermined organizationally defined criteria. For instance, a company might identify risk of damaging external resources as the primary element of these criteria, with risk of business loss being the secondary element. As an example of an escalation ladder, consider the case where an intruder is detected using organizational resources to launch a denial of service attack. The lowest rung of the escalation ladder — the first sequence in the order — might be to notify the chief information security officer of the systems under attack. The second rung might be to seek out and block the intruder’s incoming port via the firewall; the third might be to cut off all outgoing packets.

The escalation ladder for a given threat and asset will normally contain one or more rungs, where each rung will be associated with a risk level as defined by the active defense policy and developed through the use of the model. We deliberately use the format of a ladder, rather than a lattice, to simplify our model. In practice the more complex form may become necessary.

C.2.2 Detection

Detection is the automated or non-automated discovery of a past, ongoing, or future threat against an asset. This is the first of the active defense stages involving 'real time' activity. Note that only the ongoing and possibly anticipated threats fall within the active defense definition as we have provided it here — past threat management has its own sequence of actions, such as forensics and repair.

C.2.3 Evaluation

After detection of a threat, it must be decided whether the threat is included in the active defense policy (since some threats may not be covered), which assets are at risk, and estimate how great those risks might be. Evaluation places a detected threat in the context of the active defense policy and drives the decisions (and actions) that follow. Evaluation may be as simple as a table lookup, or as complex as launching an extensive investigation. In the latter case, evaluation may include activities that include active components — such as intelligence gathering — and hence have their own associated escalation ladder. The result of evaluation is to properly identify the real time situation for purposes of decision making.

C.2.4 Decision

The decision stage is when a decision set is created. The decision stage utilizes the output of evaluation to place the asset threats properly in the context of the active defense policy and the predefined escalation ladder. A decision set is the combination of active defense actions selected to be performed to mitigate the threat. Rules for selecting the decision set can be complex. One possibility is to estimate the sum of the projected consequences of the elements of each rung of the escalation ladder and choose lowest that is less than or equal to the protection threshold.

C.2.5 Action

An active defense action is any automated or non-automated activity performed for the specific purpose of mitigating the threat against an asset. Actions can range from notifying the chief information security officer of a detected threat, to shutting down a port, to the use of a denial of service (DoS) attack against the attacker, to the initiation of a virus against the attacker. A successful action does not imply a decreased risk from the threat.

There are two types of active defense actions: atomic and composite. An atomic action is one that cannot be divided into sub-actions, while a composite action is an action consisting of two or more other actions (atomic or composite). An atomic action may be something like 'shutting down a port at a firewall', while a composite action may be 'disabling all communication to the server'. This distinction allows greater flexibility for an organization when developing their active defense response. However, when using composite actions, the cost of the action will be the sum of the costs of the actions of which it is comprised.

C.2.6 Analysis

After an action is performed, an analysis must be made of whether the action has successfully mitigated the threat to the satisfaction of the threshold stated in the active defense policy. If the action has not satisfied the threshold, then escalation is necessary. If the action has satisfied the protection threshold, then determination must be made if the attack is ongoing — and whether actions taken need to be kept in place or whether an organization can revert to a state of less risk further down the escalation ladder.

For an active defense model to be successfully implemented, the organization must be confident in their ability to assess whether an action was effective in meeting the protection threshold. It is therefore prudent and necessary that an organization limit their use of active defense to areas where such assessment is possible — either directly or indirectly (such as determining whether an appropriate level of service is restored).

C.2.7 Escalation

Escalation here refers to the change in state by performing the next action described in the escalation ladder. Escalation may be tied to an increased cost of some type, perhaps risk or financial, assumed by the organization, or possibly tied to some estimate of 'increasing use of force'. This 'cost' is something that should be established when the ladder was devised — and, as noted earlier, it may ultimately become useful to model the ladder as a lattice involving a variety of cost hierarchies instead of combining these into one. Escalation may be repeatedly performed — this is anticipated to occur when the action sequence performed is unsuccessful in mitigating the threat to the satisfaction of the protection threshold described in the active defense policy.[3]

C.2.8 Maintenance

Maintenance is important to the security of any organization. Maintaining an effective active defense policy includes adding or removing assets, threats and risks. Additionally, after the analysis and escalation stages of an active defense, the policy should be reviewed to reflect any lessens learned during the post-mortem of the active defense action. It is also necessary to update the escalation ladder if the active defense policy changes. By our definition, maintenance activities are not considered part of the active defense action sequence per se — since they may occur during the course of the attack, or may occur only after it is over — but they are part of our model for managing active defense.

C.3 Goals and Assumptions

C.3.1 Goals

We have identified several goals for our model.

- *Generalizable*: The model should allow any organization or individual the ability to create an active defense policy and escalation ladder.

- *Useful*: The model should be practical and useful to any organization contemplating active defense.

- *Expandable*: The model should allow organizations to include elements that are not included in the model with no changes to the model in general.

[3] Our model does not presently explicitly specify whether an escalation requires a repeat of the full Detection/Decision/Analysis/Escalation phases, or if Escalation has an Assessment of Outcomes loop built in.

- *Mitigates Legal Risk*: Allows an organization to 'prove' that they have applied proportional and minimal force necessary to repel an attack in the face of a legal challenge.

- *Mitigates Ethical Risk*: The model should allow an organization to include their own deontological or teleological ethical considerations and be confident that the actions suggested by the model are consistent with those considerations.

- *Minimizes Unintended Consequences*: The model should attempt to minimize the unintended consequences of an active defense action. This is a key area of active defense that warrants further study — in particular, how might unintended consequences be identified? How much cost might be associated and how should this be assessed when the cost is to another organization?

- *Consistent*: Every element in the model should be consistent with every other element in the model.

- *Thorough*: The model should allow any organization the ability, with the proper time investment, to create a complete assessment of risk and benefit for each potential active defense action.

- *Automated*: The model should allow explicit analysis and action by automated methods.

C.3.2 Assumptions

- *Assets can be estimated*: The model assumes that the assets and risks of an organization can be accurately estimated with respect to the given categories.

- *Responses can be evaluated*: The model assumes that all of the active defense actions to a given threat have been included, and that the model will not be used to evaluate actions that have not been included.

- *Consequences are enumerable*: The model assumes that all the consequences of an action are known and have been included in the active defense policy.

- *Ethical considerations can be evaluated*: The model assumes that all ethical considerations have been evaluated correctly to provide their accurate weight.

- *Legal consequences are known*: The model assumes that all legal consequences are known, and that the laws have been tested and interpretations will be static.

These assumptions are tempered with the fact that an organization has the freedom to choose only assets or actions on which they can perform an acceptable evaluation.

C.4 Escalation Stages

Active defense actions vary considerably in many aspects — from effectiveness and risk, to legality and ethicalness. Tracking down an attacker with common tools such as ping and finger is not the same as sending them a virus. It is important to identify the stages of active defense actions because as the model is concerned with an organization assuming liability, taking action should begin at the lowest stages and progress upward until the protection goal is met.

Additionally, a logical and measured progression through the stages can, to some extent, defend an organization legally by showing due diligence was practiced and the defense was not ad hoc. Although legal precedent with regard to the use of force in self-defense of electronic assets has not been established, there will likely be elements of traditional legal theory involved. Most importantly, that the minimal force necessary to repel the attack was used, that the force was proportional to the threat, and that the threat was immediate (some choose to also impose an imminence standard). These theories are supported by both United States and International law (Article 51 of the UN Charter [35] and the Model Penal Code §3.02 [73]).

The stages of active defense are (partially adapted from [99]):

1. *Internal Notification*: Using the organizational structure to notify the appropriate persons of an active defense situation

2. *Internal Response*: Applying active defense actions within an organization's boundaries (e.g. shutting down the port on a firewall)

3. *External Cooperative Response*: Employing the assistance of other entities outside of an organization to mitigate a threat

4. *Non-cooperative Intelligence Gathering*: Using external services (finger, nmap, netstat) to gather intelligence on the attacker

5. *Non-cooperative 'Cease and Desist'*: Shutting down harmful services that do not affect usability on a network or host (e.g. Zombie Zapper™from BindView).

6. *Counter-strike*: An offensive action designed to deny an attacker the ability to continue an attack.

7. *Preemptive Defense*: With knowledge of a forthcoming attack, execute active defense actions to preempt (and disable) the upcoming attack

These stages are not argued to be complete or sufficient, but merely a starting point and an example of categorizing active defense actions based on their perceived risk. Because of the generalized nature of this model, we would anticipate extensive tailoring by a given organization.

C.5 An Active Defense Algorithm and Model (ADAM)

We propose a preliminary model ADAM (Active Defense Algorithm and Model), intended to illustrate an algorithmic method of how an organization might go about devising an active defense policy and an escalation ladder. The model is separated into three stages, asset evaluation, action evaluation, and the escalation ladder. Asset and action evaluation stages are used to formulate the active defense policy, while the escalation ladder decides which actions in the policy are best suited to mitigate the threat and in what order they should be executed.

C.5.1 Asset Evaluation

The first stage in the creation of an active defense policy is asset evaluation. In this stage, an organization identifies which assets, if threatened, are candidates for an active defense action. Ideally these will be drawn from an existing plan that the organization has in place for risk management. Additionally, the threats against each identified asset that are considered a potential trigger for an active defense action are enumerated, and in most cases these also can be drawn from existing planning documents. More importantly in this stage, is that the risks to an organization are properly listed for each threat, and each risk is valuated. This helps to quantify an organization's exposure to risk if the threat materializes and is successful. Later this will be used to decide if the risks of an active defense action outweigh the loss of the asset.

Scoring Chart

The scoring chart is used to compare the risk of a threat materializing with the risk of an active defense action. Therefore, an organization must have a reasonable method of scoring the risks. In our preliminary model we include five threat-risk categories, which can be modified to fit an organization's strategic goals. Our categories are: legal, national security, financial, ethical consequences, and ethical actions.

The first three are traditional risk areas. However, when active defense activities are contemplated, it is important to include ethical considerations as well as the others. Clearly, performing an active defense action places ethical risks on an organization. While some organizations may minimize the weight of this category, others may place a high value upon it. Goals important for maintaining an ethical organization should, in our opinion, be supported by any active defense model.

We have further subdivided ethical risks into two parts: ethical consequences and ethical actions. It is our belief that choosing between a teleological (only the consequences of an action are deemed necessary for ethical consideration) and a deontological (only the act in and of itself is considered) ethical theory is an overly burdensome way to approach the issue. Therefore, we have chosen to represent both, with the teleological perspective represented by the Ethical Consequences category, which defines the 'ethicalness' of the potential consequences of an active defense action; and the deontological represented by the Ethical Action category, which describes the 'ethicalness' of the action an organization takes in and of itself.

Scoring in any of these categories is difficult — as even financial and legal risks cannot be assessed with full accuracy, drawing as they do on qualitative determinations and changeable environments. It is also correct that ethical scoring in particular is highly subjective and difficult for an organization to perform. On the other hand, an organization that cannot answer these questions without the pressure of a live attack damaging key assets will certainly not be better positioned to do so once the attack occurs.

To simplify the scoring task in our preliminary model, in the Ethical Actions category we have initially required only potential active defense actions (because consequences are not considered in a deontological framework). In the Ethical Consequences category, all potential consequences need to be considered.

We proceed to explaining the scoring system. A score s is identified as a three-tuple $(category, rating, risk)$ in a set designated as S, defined by:

$S = \{s_1, s_2, \ldots, s_n \mid \forall\, i,\ 1 \leq i \leq n, s_i, s_{i+1} \in S$,

$$category(s_i) = category(s_{i+1}) \wedge$$

$$rating(s_i) < rating(s_{i+1}) \wedge$$

$$rating \in \Re \wedge -1 \leq rating \leq 1 \wedge$$

$$rating(s_i) \neq rating(s_{i+1})\}$$

This set of tuples is used to score the risks of a threat. Each category is used to denote a particular type of threat-risk (e.g. legal, national security, etc.). Within each category there exists ratings along a scale from -1 to 1, where each rating is a real number and unique. For each rating there is an associated risk, which increases as the rating increases (e.g. rating(1)=$10,000 and rating(.6)=$2,000). The risks do not have to be symmetric(e.g. if rating(1)=$1,000, then rating(-1) does not have to be -$1,000).

Asset Identification

As usual, the key to a good security policy is proper identification of assets and their value. As noted earlier, for our purposes, an asset need not be something that the organization owns, but can include anything that benefits the organization (including external resources and services). Again as noted earlier, not all assets need to be explicitly identified in the active defense policy; because of the nature of active defense, an organization my only choose certain assets to protect with an active defense policy. Those not explicitly included are assumed to be excluded from active defense protections. Further note that in the case of a real organization, asset values can fluctuate significantly[4], and this changeability would need to be reflected in any implementation of ADAM.

In terms of asset identification in the context of asset defense, note that it is a risk of this technique that an asset may be *overvalued*, and less of a risk if an asset is missed or *undervalued*. The purpose of the active defense techniques described here are to assist an organization in managing risk in those specific cases when active defense will be used. It is distinctly *not* the purpose to employ active defense as widely as possible. Thus, if an asset is left out, it is acceptable but if an asset is overvalued, it may be protected with unnecessary force.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be the set of assets of an organization to be considered for active defense measures.

[4] Consider the value to an e-commerce business of having multiple servers present to take orders. During the holiday rush, all servers may be needed (and hence all valuable). During a slow time or when inventory is being taken, not all servers are needed (and hence some are not as valuable). Hence the value of the individual server asset changes over time. Also consider a computerized life support system. When a patient is present and dependent upon it, the value is high! If there is no patient, the value is reduced. Active defense of this asset may be warranted only in former case.

Threat Identification

After identification, the threats to each asset are enumerated under the classical categories of confidentiality, integrity, and availability. The threats identified can be as general or as specific as necessary to satisfy the organization.[5] As before, for active defense purposes we include only those threats for which it is reasonable to consider employment of active defense techniques. The observations linked to threats can be as specific as 'an attacker probes port 25 and 26 in order during non-operational hours', or can be as general as 'a probe of network ports is detected.' [6]

Additionally, the organization must also determine protection goals for each threat. The protection goal is the state at which a threat is deemed to be sufficiently mitigated — this is how the protection threshold, mentioned earlier, is set. The existance of a protection goal provides three benefits: it prevents an organization from accidentally assuming more risk than necessary, supports any later need the organization may have to prove in court that they did only what was necessary to achieve an appropriate protection goal, and (3) helps guide the development of a response to a threat by providing a threshold.

The goals for each threat are going to be different depending on the organization and their needs. For example, a national security organization may have a goal to prevent any future threat from that particular assailant, while a business may only be concerned with halting the current threat. The level of goal will be dependent on an organization's available resources and their protection needs.

[5] Here we use 'threat' interchangeably between those identifiable activities or threat symptoms that might indicate some specific threat to the organization is in play, and the actual goal/threat that is the purpose of the opponent.

[6] It is beyond the scope of this paper to define a taxonomy of threats and threat symptoms, though such would clearly be of benefit. At this preliminary stage, it suffices to recommend that the threats be specific enough to detect and analyze easily, and general enough that new attacks could be placed into a categorization/hierarchy of threats.

Therefore, for each threat, a clear and unambiguous goal must be declared which will guide the responses to the threat. These goals must also be approved by the management in the organization responsible for assuming the risk if anything goes wrong while executing the active defense actions.

A threat t is identified as a three-tuple $(threat, goal, sum)$ in a set designated T.

$$T = \{t_1, t_2, \ldots, t_n\}$$

Threats are associated with assets in the formal notation by the use of the relation AT.

$$AT = \{(a, T') \in Ax2^T |\ \forall\ t \in T', t \text{ threatens } a\}$$

Risk Identification

After each threat has been identified, then it is necessary to calculate potential risks. For each threat, the organization should list all possible risks (in the aforementioned categories). Each risk must then be scored.

To calculate the score of each risk requires two steps. The first step is to assign a probability, between 0 and 1, that the risk will manifest itself. The second is to locate a score on the scoring chart that represents the total cost to the organization.

An important requirement in determining the total cost of the risk is the time interval which an organization calculates risk. It is not possible to calculate the total risk cost over all time because of the number of unknown variables, however one can calculate risk given a specific time interval. Therefore, the score assigned to a risk will be that within the time interval.

A risk r is identified as a four-tuple $(risk, category, probability, score)$ in a set designated as R.

$$R = \{r_1, r_2, \ldots, r_n |\ score \in S \wedge\ 0 \leq probability \leq 1\}$$

For example, a university may anticipate that if the threat is successful, it will result in the loss of some enrollment (financial risk). They would further estimate that the probability of this risk manifesting itself is 0.3. The university then computes the total cost (of lost enrollment dollars) as approximately \$100,000 — which corresponds to a score of .2 (in the financial category of the scoring chart).

After all of the risks have been determined, then it is possible to assign a total risk cost to a threat by calculating the sum of all of the risks associated with a threat. This is accomplished by first defining a relation TR between the threat and risk sets, which allows the reference of only risks that apply to a specific threat.

$$TR = \{(t, R') \in Tx2^R \mid \forall r \in R', r \text{ is a risk of } t\}$$

The $sum(t)$ of a threat is then defined as the sum of the products ($probability$, $score$) for each risk associated with the threat:

$$\sum_{\forall r \in R'} probability(r) * score(r)$$

C.5.2 Action Evaluation

Action evaluation is the next, and final, step in the development of an active defense policy. In this step, an organization identifies all of the potential actions it can perform to mitigate threats and the risks associated with those actions. At the end of this step, an organization should have created an active defense action chart, which will be used to develop the escalation ladder.

After this step, it will be important that the consequences of actions are known. When considering active defense, it is a greater error to *underestimate* the negative consequences of an action than it is to underestimate the benefits. In the former case, an action may be selected without full understanding of the risk involved - in other words, riskier actions might be performed more often, putting the organization at greater risk. In the second case, an active defense action may be selected less often. This still leaves the 'regular' (less risky, non-active defense techniques) in place to protect assets.

Action Identification and Classification

An organization must identify the possible actions that can be performed to mitigate a threat against a particular asset to obtain the goal (within their available resources). Additionally, actions must include organizational requirements, such as notifying the proper higher-ups, filing a report, etc. As described before, an action can be of two types, atomic and composite — where a composite action is made of other atomic or composite actions.

An action k is identified as a four-tuple $(action, acts, success, score)$ in a set designated as K.

$$K = \{k_1, k_2, \ldots, k_n \mid success \in \Re \wedge 0 \leq success \leq 1\}$$

The success of a composite action is defined as the product of the success of its sub-actions.

$$success(k) = \prod_{\forall k_i \in acts(k)} success(k_i)$$

Actions and threats are associated using the relation TK as defined by:

$$TK = \{(t, K') \in Tx2^T \mid \forall\ k \in K', k \text{ can mitigate } t\}$$

The four aspects of an active defense action that the model incorporates are: action, acts, success, and score. The first, $action$, is a unique identifier. The second, $acts$, provides a sequence of actions, of which the action is comprised. This is the empty set, $\{\}$, if the action is atomic. The third, $success$, is the probability (between 0 and 1) of the action (by itself) mitigating the threat to the satisfaction of the goal. Normally it will not be possible to assign an accurate probability to the success of an action, so probabilities can be assigned relative to the other actions. In such a case, though, the prediction will be of relative likelihood of success rather than actual success. The fourth parameter, $score$ will be discussed in detail in the next section; simply put, it is used to quantify a combination of factors that are useful in determining whether or not to choose a given action.

Utility Modifiers

Because each organization has its own unique goals, categories should not be weighted equally. A utility modifier is associated with each specific category to provide relative weighting based on the utility of that goal to the organization. This comes from the idea of a utility function developed by many other authors, including for instance Keeney and Raiffa in [116].

If, for example, a national security organization was concerned with the national security implications of an action above financial considerations, then it could place a higher utility modifier on the national security category to give it more weight in the escalation ladder.

To use the modifier, an organization multiplies each risk's $score$ in that category with the corresponding modifier. For example, we may multiply every National Security risk $score$ by 1.2 while we multiply every Ethical Action $score$ by 1.3. This would place a 10% greater weight on Ethical Action than on National Security, and a 20% greater weight on National Security over all other categories. Note that this implies that the values in each score category have already been normalized.

Risk Identification

The method of identifying the potential risks of an active defense action is identical to identifying risks of threats as previously defined. For each action, all of the risks must be identified in the suggested categories of Legal, National Security, Financial, Ethical Consequences, and Ethical Actions. Additional categories may be added by an organization if necessary. As risks of actions are identified, they are placed in the already defined set of risks designated as R. A relation between actions and risks is then identified as KR.

$$KR = \{(k, R') \in Kx2^R \mid \forall\ r \in R', r \text{ is a risk of } k\}$$

The $score$ of the action four-tuple is then defined as the sum of the products ($probability$, $score$) for each risk associated with the action, plus the total risk of any sub-actions (if a composite action). $umod$ is the utility modifier for the category of the risk.

$$\begin{aligned}\forall (k, R') \in KR, k_{score} &= \sum_{\forall r \in R'} umod * (r_{prob} * r_{score}) \\ &+ \sum_{\forall k \in acts} k_{score}\end{aligned}$$

C.6 Escalation Ladder

So far, this paper has presented the first two stages of the model, asset evaluation, and action evaluation. Once completing these two stages, an organization now has two yardsticks with which to analyze their risks with respect to active defense. This has answered the question: what risks are involved for an organization if an active defense policy is initiated. The question still left to answer is: if faced with a threat against an asset, how does a particular active defense policy describe what an organization should do?

The escalation ladder answers these questions of how to proceed and what actions to perform. An escalation ladder is an ordered set of actions that are progressively executed (i.e. the ladder is ′climbed′) until a threat is successfully mitigated. A ladder is created by ordering the actions based on a simple formula to balance risk and potential success. By iterating through the ordered actions, an organization can be assured that the defense is escalated responsibly and following the legal theory that defense should use minimal and proportional force. In the end, the escalation ladder and the algorithm will provide the defender a method of executing a responsible active defense.

C.6.1 Ladder Creation

The escalation ladder for a given threat t is created by ordering the actions in the relation TK using the formula $Score(Action) - Sum(Threat) - Success(Action)$ and not including any actions that have greater risk (designated as $score$) than the threat. Formally, an escalation 'rung' x is identified as a three-tuple $(t, k, order)$ (where t is the threat and k is an action) in a set designated as X, defined by:

Let $order(x) = score(k) - sum(t) - success(k)$ in

$$X = \{e_1, e_2, \ldots, e_n \mid \forall i \mid 1 \leq i \leq n \wedge < t, k > \in TK$$

$$\mid sum(t) \leq score(k) \wedge order(e_i) \leq order(e_{i+1})\}$$

Given that an organization provided reasonable probabilities as per the success of the actions, the estimated probability that escalation ladder will successfully mitigate the threat is the probability that at least one of the actions in the set is successful (i.e. alternative occurrence). This is expressed as:

$$\sum_{\forall x \in X} success(k) - \prod_{\forall x \in X} success(k)$$

C.7 Algorithm

At this point, the model has been described in detail. This satisfies the first (planning) of the eight stages of active defense identified in section C.2. The algorithm presented here satisfies the next five stages (minus detection and maintenance). The algorithm takes as parameters, the threat t, the asset a being threatened, and an active defense policy P. The first two parameters are most likely from an intrusion detection system from the second stage (detection).

Algorithm 2 $Active - Defense(t, a, P)$

```
 1: if $a$ not $\in P_A$ then
 2:    fail
 3: end if
 4: if $t$ not $\in P_T$ then
 5:    fail
 6: end if
 7: $X \leftarrow$ADModel$(t, a, P)$
 8: $n \leftarrow |X|$
 9: $riskAssumed \leftarrow 0$
10: for $i \leftarrow 1$ to $n$ do
11:    $k \leftarrow X_i$
12:    while $k$ cannot be performed do
13:       $k \leftarrow$ get next action in $X$
14:    end while
15:    $riskAssumed \leftarrow riskAssumed + score(k)$
16:    if $riskAssumed > sum(t)$ then
17:       break
18:    end if
19:    execute the action $k$
20:    if action $k$ achieved $goal(t)$ then
21:       break
22:    end if
23: end for
```

Now for a description of the algorithm. (1-6) Satisfies stage 2 (evaluation) by deciding whether the asset and threat are covered in the active defense policy — if it is not in the policy, then fail and do not execute an action. (7) Satisfies stage 4 (decision) by retrieving from the model the decision set of actions. (8) Assigns the variable n the size of the set X. (9) Initializes a new variable $riskAssumed$, which stores a total of the risk incurred by executing the actions. (10) Iterates over the set X. (11) Assigns a variable k the action that in the set X at index i. (12) Checks if the action k can be performed using the information available (e.g. is the IP address correct, etc.) and continues until it finds one. (13) Get the next action in the escalation ladder. (15) Adds the risk of the action k to the current risk assumed. (16-18) Checks if the current amount of risk (total risk) has exceeded the risk of the threat, if it has then get out of the loop. (19) Satisfies stage 5 (action) by executing the action selected. (20-22) Satisfies stage 6 (analysis) by checking if the action has achieved its stated goal in $goal(t)$, if it has then no need to continue. Stage 7 (escalation) is satisfied by the fact that the next iteration through the loop will escalate to the next action in the decision set.

C.7.1 Contingency Plan

Step 8 in the algorithm is considered the contingency plan. It allows active defense to continue although an action could not be completed. A major concern with active defense is that the information available to network tools about a threat or attacker can be incorrect or unavailable. More dangerous is the fact that the situation can change between actions (the attack can change, the attack is using a new source, etc.) In these cases, the algorithm skips that action and moves onto the next 'rung' of the escalation ladder. As an additional measure, confidence values can be added to network data such that an action will not be taken using that data until a specific threshold (confidence) is met. More can be added to this test as necessary by an organization to guarantee that actions are only being executed under certain conditions.

C.8 Analysis

At this point it is necessary to look at the model objectively and to determine whether it has satisfied the goals stated in section C.3.1 To accomplish this, each goal will be examined in turn.

1. **Generalizable**
   The model is generalizable because it does not discriminate towards any particular organization and can also be used by individuals. An organization can add or remove threats, assets, risk, categories, and escalation stages as necessary to fit the model to existing security policies and threat models; an organization can also use the utility modifier to match the model to the organization's risk focus. The flexibility of the model allows any organization or individual to modify the model to meet their needs and to address their particular concerns.

2. **Useful**
   This goal can only be shown to be met when organizations actually attempt to adopt the model. However, every effort has been made to develop the model in a pragmatic direction; and address the concerns that both public and private organizations would have with active defense — namely legal, ethical, and unintended consequences.

3. **Expandable**
   Since the organization that is developing the active defense policy can determine the categories, assets, threats, risk charts, and all other aspects of the model, the model can be expanded as large as necessary to accommodate any organization.

4. **Mitigate Legal Risk**

As discussed earlier in section C.4, understanding of the legal issues involved in protection of electronic property is a highly volatile area. Also note that neither author is a lawyer, and is not offering legal advice here. However, it is useful to recount here three of the legal theories often cited with regard to the use of self-defense for consideration by the readers. The three theories are that the minimum amount of force is used to mitigate the threat, the force was proportional, and that the threat was immediate. The model we present here incorporates these through the use of stages to escalate a defense so that the least amount of force was used. [7]

5. **Mitigate Ethical Risk**

   A major issue with active defense is the question of whether active defense actions are ethical. The model addresses this question by incorporating both teleological and deontological ethics into the risks of an action. In this way, the model only suggests actions that an organization has deemed ethical in certain circumstances.

6. **Minimize Unintended Consequence**

   Unintended consequences are difficult to protect against, and in particular it is a trait of them that they may not even be knowable in advance, or repairable once they occur. The model provides two methods to address this concern. The first is that confidence values can be added as input, providing additional information as to the validity of the threat, and source of the threat (so that actions are not executed against innocent targets). The second method is that each action is assigned a probability that it will be successful, if an action is not successful (the inverse of the assigned probability) then it must be assumed that an unintended consequence did occur; and by this

[7] Note that in the technical realm, it is not at all clear what 'least force' means, and so any organization using these strategies would need to seek legal advice before setting these values.

method, an estimate of the probability that unintended consequences will occur with a specific action is produced. Although these are not foolproof methods, unintended consequences, by their nature are difficult to predict and mitigate and these provide at least a level of planning. In general, the more 'active' the defense, the more likely that there will be unintended consequences and hence some loss to the organization (and others)in employing the technique.

7. **Consistent**

   A consistency proof is beyond the scope of this article.

8. **Thorough**

   A proof of thoroughness is beyond the scope of this article — and not something that a model alone can enforce. We note, however, the following. Since the model requires that the organization fully enumerate all of their assets and risks that will be covered by the active defense policy, the thoroughness is in the hands of the implementor. The primary issues from an active defense perspective are the *undervaluing* of risks assumed as a consequence of employing active defense, and the *overestimating* of the value of the asset. Leaving out an asset reduces those things protected by active defense — leaving protection to the remainder of the security methods in place.

9. **Automated**

   The model was designed with this goal in mind. It can easily be implemented in a contemporary intrusion detection system because it is only a series of sets used to create a graph, which autonomous agents can analyze easily using well-known algorithms. Also, the algorithm presented is obviously designed to be implemented in an automated system.

C.9 Conclusion

This paper has used a preliminary model, ADAM, to bring out a discussion of the factors that should influence an organization that is considering the use of active defense techniques. The four primary considerations are ethical, legal, unintended consequences, and risk valuation. ADAM illustrates one method of addressing these considerations in a form that is pragmatic in nature. ADAM itself is divided into two parts: the active defense policy, which describes an organization's assets, threats, risks, and potential mitigating actions; and the escalation ladder, which is an ordered set of actions to execute based on the information provided in the active defense policy.

The creation of the active defense policy and escalation ladder require a tremendous resource commitment on the part of any organization. However, the questions regarding what one should do in an active defense situation are astounding and require such a commitment to explore the real ramifications of an active defensive position.

C.10 Acknowledgements

The authors would like to thank others for their help with this paper. Ryan Blue for his help correcting the formal notation, Stan Gotshall for his lengthy discussions with us and insights on the issue of active defense; and Dave Dittrich and Barbara Endicott-Popovsky for providing the impetus and inspiration to begin formal research on this topic.

Appendix D

EVOLVING ACTIVE DEFENSE STRATEGIES

S. Caltagirone, "Evolving Active Defense Strategies," University of Idaho, Moscow, ID, Technical Report CSDS-DF-TR-05-27, 2005.

D.1 Introduction

It's been no surprise to the security community that the number of attacks on computer systems has been steadily increasing over several years [117]. This pattern is more disturbing when the number of national infrastructure systems accessible over the Internet is also growing [118]. Together, this is a dangerous combination that puts the US national infrastructure at risk and threatens national security. In addition to infrastructure, medical and financial systems are equally vulnerable and accessible, placing life-critical systems in danger.

There is no question that these systems are being protected by extensive security measures such as firewalls and intrusion detection systems. However, what happens if an intruder were to bypass those defenses? It is like having a castle with very thick and tall walls, but having no security force inside — if an attacker were to tunnel under the walls, they would have full and complete access until they choose to leave. The defenses on these critical systems are good, but not sufficient to protect against the level of sophistication that intruders are reaching. Additionally, these defenses can not be maintained fast enough to keep up with the number of vulnerabilities discovered weekly; thereby providing attackers with an opportune time between when a vulnerability is known, and when a vulnerability can be protected against. Preventative defenses are no longer sufficient, especially for national infrastructure and life-critical systems.

This leaves only one choice: to implement a defense that can mitigate a threat between the time it is detected and the time that it has achieved its goal. This is the essence of active defense.

Active defense is a topic that most security researchers have shied away from. It has been portrayed as a rash tool of the vigilante [32,34]. But in reality, active defense actions are varied in scope; from the notification of appropriate personnel, to the notification of authorities, to rewriting firewall rules, to initiating a denial of service attack against the attacker. An active defense is any set of actions taken to mitigate a threat against an asset between the time the threat is detected until the time it has completed its objectives [39]. There have been very few published instances of active defense actually being utilized. Some famous cases involve Cliff Stoll tracking German hackers in the early 80's [23] and the US Department of Defense initiating an attack against a group attempting to use a distributed denial of service attack (DDoS) [119].

However, certain companies, such as Cisco Systems, have recently begun to include 'response' technologies in their firewall and intrusion detection systems. These 'responses' limit themselves to rewriting firewall or routing rules in an attempt to block an emerging threat [120]. Yet although the technology now allows us to undertake active defense action, there is still no clear method of creating and evaluating the effectiveness of an active defense strategy. An active defense strategy is the ordered set of actions that will be taken in response to the detection of a threat.

In this paper, we will describe a novel method to derive active defense strategies using evolutionary techniques and genetic algorithms. While active defense strategies can be derived in other ways [39], an evolutionary environment can provide a unique setting where the strategies are not determined solely by calculated risk and success, but by which strategies perform best against an attacker's strategy.

More specifically, competitive evolution will be utilized so that both an attack and defend strategy are simultaneously co-evolved. The strategies are then evaluated based on their performance against their evolved counterparts. In this way, we hypothesize, that using evolutionary strategies to derive active defense strategies will yield results that are reasonable based on a commonsense understanding of security and the use of active defense.

D.2 Background

D.2.1 Active Defense

There has been very little previous research directly relating to the topic of active defense. Most previous work has been applied to international policy and developing a doctrine of information warfare. In 2000, Grove, et al. in [35] attempted to determine the international legal implications of an active defense with particular focus on the UN charter and established laws of armed conflict. The paper concluded that active defense, when utilized by a nation-state, as a response to an attack by (or sponsored by) another nation-state, was an appropriate and acceptable defense with regard to established international laws and policies.

William Yurcik has also published a series of papers regarding whether the United States should pursue a policy of information warfare, and whether that policy would be (1) in the best interests of the United States, and (2) whether it would be consistent with international law. In his 2000 paper [58], Yurcik concludes that international law is vague with regard to the issue and leaves room for the United States to act if attacked via the Internet. In his 2001 paper [53], Yurcik presents a framework with which to develop an information warfare policy. This is primarily based in current rules of engagement of the United States military and the potential harm to civilian infrastructure during an attack.

Although information warfare and active defense are regularly confused because of potential offensive action, there is a significant difference between the two. Information warfare is concerned with achieving a "military advantage using tactics of destruction, denial, exploitation, and/or deception." [53] Active defense, on the other hand, is not concerned with military advantage, and only attempts to mitigate a threat until a previous security state has been reached. This difference does not mean that information warfare research is not valuable; on the contrary, information warfare research is very valuable because of the stress on offensive action - the most questionable element of active defense.

In this experiment two topics are being combined in the hope of developing a new technique for information assurance policy and strategy creation - active defense and evolutionary computation. This is not the first time that evolutionary techniques have been brought to bear on the problems of information assurance. Most of the work combining evolution and security has been completed in the field of intrusion detection systems.

In a seminal paper on the topic [48], Crosbie and Spafford developed a prototype system, where the agents on the system were to taught to detect intrusive behavior using genetic programming techniques. To accomplish this, they developed a meta-language to examine specific aspects of the system such as network data and disk access. This language was used in the parse trees developed through the use of genetic programming, which were used as rules in the agents. If a rule was broken on the system, an agent raised the suspicion level and other agents began to look more closely at their own data by incorporating more strict rules. When a sufficient number of agents have raised the suspicion level, the level goes above a threshold and the security officer is notified of a potential system intrusion.

While Crosbie and Spafford's work remains one of the few in the topic, some other work has continued. In [121], the authors use genetic algorithms to design an intelligent decision system for intrusion detection. They want to find a new method of limiting the number of false positives in an anomaly-based intrusion detection system (one that looks for a pattern of intrusion rather than specific behavior). Their primary objective is a classifier, which will classify and then execute an action based on the classification. They do this by evolving a population of classifiers and testing against a previously developed knowledge base. Their experiments were relatively successful in classifying the data and suggest that a system to evolve defensive security actions is possible.

Working in a similar vein was Spears and Gordon regarding the evolution of finite-state machine strategies for a defender-adversary game [122]. This work attempted to evolve strategies for a game involving two players competing for limited resources; the adversary's strategy was fixed, while the defender evolved to beat it. Their results were promising, however limited by the existence of cycles in the strategies. Their work has potential applicability in the survivability and defense of networks.

D.2.2 Genetic Algorithms and Co-Evolution

Research in evolutionary computing techniques has been popular in the recent decades. Researchers in the field have created many types of evolutionary paradigms, artificial life, genetic algorithms, genetic programming, neural networks, and particle swarm among others. For our purposes, genetic algorithms and co-evolution will be focused on.

A genetic algorithm is a computation paradigm that utilizes a population of encoded chromosomes, and operations upon those chromosomes for the purpose of searching search spaces by increasing population fitness as individuals near a goal. This paradigm was first introduced by John Holland in [88]. However, although his original work describes natural competitive co-evolving populations, his theories and experiments are only subject to fixed environments.

Since Holland's pioneering work, John Koza has been a leader in the field, developing and championing the technique of genetic programming. See [89] for more details about Koza's work in genetic programming. With regard to co-evolution, he put forth two papers discussing his experiments in the subject [90, 91]. In these works, Koza describes a "hierarchical co-evolution," which is where the environment for the first population consists of a second population and vice versa. He also describes "relative fitness," which is where an individual's fitness is determined by its performance against all of the individuals of the opposing population. He puts these into practice by attempting to evolve a game strategy using a tree structure; and succeeds in evolving the optimal strategy for each player without any direct knowledge of such strategy.

Robert Axelrod is generally regarded as the first to apply evolutionary techniques to game theory using the Prisoner's Dilemma in [92]. Prisoner's dilemma is where two players must decide whether to rat out the other person or to not talk at all. If both players rat the other person out, they receive no reward. If player one talks and player two does not, then player one gets a reward and the other player two does not. If both talk, neither gets a reward. Axelrod utilized genetic algorithms to develop a strategy for the game. Using several fixed programs submitted in a competition by others, his algorithm was able to evolve the optimal strategy of TIT-FOR-TAT.

Miller followed Axelrod's work in [93] by applying a co-evolutionary technique to the search for a Prisoner's Dilemma strategy. Instead of the tree structure used by Koza, or the linear gene structure of Axelrod, Miller utilized finite-state automata for the problem representation. His experiments showed widely varied convergence after 10 generations depending on the information about the other player available. The results of the co-evolutionary experiment showed that the top performer was just slightly less optimal than Axelrod's, but was more tolerant to short-term defections by the other player. This leads to a potential conclusion that while a co-evolutionary technique may not yield optimal strategies, the results may be very good if the game is modeled imperfectly or contains many dynamic components.

Rosin and Belew produced two papers with regard to competitive co-evolution in 1995 [94] and 1996 [95]. They used Tic-Tac-Toe, Nim, and Go as the games with which to evolve two competitive players. They developed two novel techniques, "competitive fitness sharing" and "shared sampling," which improved performance. The primary purpose of the work is to improve the "parasite" (population being tested against) population so that stronger "hosts" will be evolved. The usual method of fitness evaluation involves summing the scores during the interactions. However, "Competitive fitness sharing" is that each "parasite" is scored with the number of "hosts" that defeat it, and a host gets a fitness that is the sum of the scores of the "parasites" defeated by it. In this way a "host" is rewarded for defeating "parasites" few other "hosts" can.

The other technique developed by Rosin and Belew is "shared sampling" where a "host" is tested not against the entire "parasite" population as normally would be, but instead against a mixed set of "parasites" that tend to both defeat many "hosts," and get defeated themselves. Both of these novel techniques greatly improved the players evolved for these games.

Potter, De Jong, and Grefensttete presented [96] in 1995, which presented a solution for evolving agents with many subtasks. Their solution was to use multiple genetic algorithms, each evolving to a single subtask. When these populations had converged, they took the best of each population combining into a single agent, which was then able to effectively complete a more complex task composed of the trained subtasks. Each population was guided towards its subtask not by direct instruction, but rather by initial seeding. This is not an example of competitive co-evolution, but of a novel use of cooperative co-evolution. This work shows that complex rule-based behavior, such as active defense actions, can be successfully evolved from simpler elements.

Haynes and Sen in [97] described their (failed) attempt at co-evolving predator and prey populations. In their experiment, they used a grid with multiple predator and prey agents; where the predators could communicate together and the prey could not. They expected an arms race, where one evolutionary jump by one population is quickly matched by a counter-evolutionary jump in the other. However, the prey evolved a very simple, yet effective strategy, they all moved quickly in a straight line so that the predators were always chasing and could not surround them. This strategy fails when pitted against a greedy algorithm, and hence did not produce a novel strategy that performs better than current strategies. Possible improvements would be if the predators could predict prey action (n-look ahead), or if the predators could move quicker instead of only smarter; these improvements may have produced more complex prey strategies. The lessons from this research are that prey will always tend towards simple, effective strategies which exploit the actions, or lack thereof, available to the predators.

D.3 Experiment

The goal of this experiment is to discover whether it is possible to evolve an active defense strategy that is reasonable and could be considered viable enough to utilize against a threat.

D.3.1 Scenario

Active defense is a security tool. However, it is unlike other security tools in that it is not for general use. Active defense is a tool that must be tailored for each threat and each asset. For that reason, any attempt to experiment with active defense must provide a scenario for which the actions and risks are tailored.

For this experiment, a realistic scenario was chosen. The scenario is based on a medical patient database. This database is hosted by a medical facility that provides access via the Internet to other facilities. The data stored in the database is necessary for patient care at the facilities that use it. If the data's integrity or availability were threatened, then patient care would also be threatened. This scenario implies that the worst threats are those that compromise availability and integrity; and likewise, the riskiest active defense actions would be those that do the same.

D.3.2 Active Defense Modeling

The first step in developing an active defense strategy is the identification of potential actions and associated risk. On the other side, a set of attacking actions must also be defined — as well as the risks of those attacks to the asset. An important element is that both the attacker and defender also have a 'null' action available to them - which is equivalent to no action (and no risk). The null action is included because sometimes the best action is no action.

In this experiment, 12 defensive actions were defined. Those actions were given a risk amount, which is only relative to the other actions (defensive and offensive). Additionally, a flag was set if the attacker's IP address was a necessary piece of information to carry out the action; and whether the action, if successful, would permanently stop the attacker. These defensive actions and parameters are described in Table D.1.

Table D.1: Defensive Actions

Action	**Risk**	**IP Necessary**	**Permanent Stop**
Contact Administrator	2		
Contact CTO	5		
Shutdown Port at Firewall	1700		
Filter IP at Firewall	200	X	
Shutdown Server	2000		
Send TCP RST Packet	6	X	
Ask ISP to Shutoff Attack	10		
Contact FBI	5		X
Use Traceback	6		
Send Virus Against Attacker	1000	X	X
DoS Attacker	1000	X	X
Hack Attacker	1500	X	X

Eleven offensive actions were also defined. The risks of the offensive actions are not determined by any formula, but as relative to the other actions and the value of the asset (mainly availability and integrity). These are described in Table D.2.

Table D.2: Offensive Actions

Action	**Risk**
Spoof IP Address	0
Port Scan Server	0
Ping Server	0
DoS Server	800
DDoS Server	800
Poison DNS	200
Install Backdoor	900
Download Records	1000
Change Records	1500
Send Virus Against Server	400

A table was then created that specified which defensive actions stop which offensive actions — Table D.3 is a reproduction of that table.

Table D.3: Defense Action Mitigation Matrix

Defensive Action	**Stops These Actions**
Contact Administrator	
Contact CTO	
Filter IP at Firewall	Port Scan, Ping, DoS, DDoS, Backdoor, Download Records, Change Records, Virus
Shutdown Server	All Actions
Send TCP RST Packet	Port Scan, Ping, Backdoor, Download Records, Change Records
ISP Shut-Off	Port scan, Ping, DoS, Backdoor, Download, Change Records
Contact FBI	Download Records
Use Traceback	
Send Virus	Port Scan, Ping, DoS, Backdoor, Download, Change Records, Send Virus against Server
DoS Attacker	Port Scan, Ping, DoS, Backdoor, Download, Change Records
Hack Attacker	Port Scan, Ping, DoS, DNS, Backdoor, Download, Change Records

The 'Use Traceback' defensive action plays an important role. It itself does not stop any actions, however if an attacker has used the 'Spoof IP' action, then traceback will find the IP of the attacker so that defensive actions that require a valid IP will then be effective. Additionally, the Contact CTO and Contact Administrator actions do not do anything very valuable, but are there because in a real active defense scenario those would are required before an active defense can be initiated (although that is not modeled here).

D.3.3 The Game

Active defense in this experiment is played like a game. The attacker plays one action, if that action is neither null nor ′IP Spoof′, then the defender has the opportunity to counter the attack. If the defender executes the 'Use Traceback' action, then the attacker′s IP address is known until the attacker executes the ′IP Spoof′ action again. All of the defender′s actions are iterated through until (1) there are no more defensive actions, or (2) the attack is successfully stopped. Each action that the defender takes accumulates risk, however only if the attack is not stopped is the attack risk also added. This is illustrated in Table D.4.

Table D.4: Risk Assignment

	Attacker	
Defender	*Success*	*Stopped*
Success	Defender Risk	Defender Risk
Unsuccessful	Attacker + Defender Risk	Defender Risk

The algorithm for the game is described in Algorithm 3.

Algorithm 3 Co-Evolutionary Game Algorithm

```
$totalRisk \leftarrow 0$
for all $aAction \in AttackerActions$ do
    $stopped \leftarrow false$
    for all $dAction \in DefenderActions$ do
        $totalRisk$ += risk($dAction$)
        if permanentStop($dAction$,$aAction$) then
            return $totalRisk$
        else if stops($dAction$,$aAction$) then
            $stopped \leftarrow true$
        end if
    end for
    if !$stopped$ then
        $totalRisk$ += risk($aAction$)
    end if
end for
return $totalRisk$
```

D.3.4 Genetic Algorithm

To accomplish this experiment, two populations will need to be evolved in parallel - one for attackers and one for defenders. These populations will then be placed in competition to obtain the fitness of the individuals. The parameters of the algorithm are described in Table D.5.

Table D.5: Genetic Algorithm Parameters

Parameter	**Value**
Paradigm	Generational
No. of Populations	2
Population Size	60
No. of Trials	100
Fitness Evaluations	132,000
Parent Selection	Tournament Selection
Elitism	Top 2 are kept
Mutation Type	Uniform Random Replacement
Mutation Rate	$1/n$
Crossover Type	2-pt Crossover
Crossover Probability	100
No. of Actions in Chromosome	8
No. of Initial Actions	4

Initialization

The chromosomes were randomly initialized with actions. However, they were not filled. The chromosome was first filled with null actions, then, a certain number (see Table D.5) of random actions were chosen to be placed in random locations within the chromosome array.

Mutation

The mutation algorithm chosen is uniform random replacement. Each allele has probability 1/n (where n is the number of actions in the chromosome) to be randomly changed with another action. The action randomly chosen can be null;

this means that the action can be ′removed.′ Because no-action is a legitimate active defense strategy, this does not affect the evolution of a strategy.

Crossover

Two-point crossover was chosen as the method of crossover. In two-point crossover, two points are randomly chosen and the mid-section is swapped between the chromosomes.

Selection

Selection for crossover was done with 2-element tournament selection. In this method, two chromosomes are randomly selected, and the one with the best fitness becomes a parent in crossover.

Fitness Evaluation

In this experiment, the fitness evaluation follows what is normal procedure for competitive co-evolution. Each member of a population is independently tested against each member of the opposing population. The fitness of the chromosome is the sum of the risk produced in the competitions with all the elements of the opposing population.

D.4 Results

Using a competitive co-evolution technique, the experiment produced successful results. The results fall into two sections, population fitness and the development of a strategy. The population's fitness shows very clear competitive co-evolutionary behavior as well as an insight into the choices made by individuals in the population. The strategy developed in the experiment also shows a clear trend and supports the hypothesis that a strategy can be successfully evolved.

D.4.1 Example Strategies

Here is an example attack and defend strategy developed through competitive co-evolution.

Attacker: (Spoof IP Address) (Poison DNS) (Port Scan the Server) (Port Scan the Server) (Poison DNS) (Port Scan the Server) (Hack Server, Install Backdoor) (Poison DNS)

Defender: (Ask ISP to Shut-off Attack) (Use Traceback) (Use Traceback) (Contact Administrator) (Contact FBI) (Filter IP at firewall) (Contact Chief Technology Officer) (Null Action)

D.4.2 Population Fitness

In most genetic algorithms, the average fitness of the population over time rises or falls as it reaches a solution. However, with competitive co-evolution, this is not the case. As shown by Figure D.1 and Figure D.2, the fitness of the populations quickly rise to a level then fluctuates near that level over time. More importantly is the fact that the change of the population fitness should mirror that of the opposing population.

As one population makes a change, the opposing population will make a change to counter the change. This effect has been previously shown in [94, 97]. If a close examination of the two graphs is made, the attacker population, while making more significant changes, and the defender population closely mirror each other. The reason for the larger changes in the attacker is due to the fact that the attacker is free to experiment with riskier actions, while the defender is going to usually choose the same safe actions. This conclusion about the behavior of the populations is supported by the individual choices of the chromosomes (as described in the next section).

D.4.3 Strategy Results

The purpose of this experiment was to determine whether competitive co-evolution was a useful technique in developing active defense strategies. After running the genetic algorithm for 1000 generations over 100 trials, a general strategy emerged. The strategy produced is reasonable and effective.

The attackers choose a strategy that inflicts the most amount of risk while minimizing the defensive actions available to mitigate the threat. As shown by Table D.7, the attackers normally choose to spoof their IP for the first 1-3 actions, then to poison the DNS next, and then finally either poison the DNS or change the records the last four actions. This strategy was developed consistently over many iterations of the experiment.

This evolved attack strategy is successful because spoofing the IP address early in the actions limits the defensive actions possible to only those not requiring the attackers IP address, or the 'Use Traceback' action. Additionally, 'Poison DNS' is an attack that is difficult to defend and therefore has the highest likelihood of succeeding. Lastly, the attacker choose to either change the records, or 'Poison DNS' because the actions necessary to mitigate the 'Poison DNS' attack do not mitigate the changing the records — and the ones that do mitigate are very risky for the defender. Therefore, the defender is likely not to take an action but rather absorb the attack.

From the defender's viewpoint, as illustrated by Table D.6, the most telling aspect of their strategy comes from the actions that no top-rated defender choose: shutting down the port at the firewall, shutting down the server, sending a virus, initiating a DoS, and attempting to hack the attacker. Since these actions are very risky (as compared to the other actions), the top defenders found that the best strategies did not include them — but rather relied primarily on five other actions: using traceback, contacting ISP, contacting the administrator, contacting the CTO, and using TCP RST packets.

From both the actions chosen and avoided, some general observations of the defender's strategy can be made. Obviously, the defender needs to execute 'Use Traceback' to gather the requisite IP data for other actions; and since the attackers generally use IP Spoofing early in the attack, the defenders also show heavier use of 'Use Traceback' in the early actions rather than the later.

The defender also heavily favored contacting the ISP to ask for assistance. If the frequency of the ISP and FBI actions are combined, they form the most frequent action chosen (aside from the null action) — illustrating why organizations today rely so heavily on external entities to assist during an attack, because the risks are low and the benefits are high. This is also why the attackers choose to heavily use the DNS attack, because the ISP and FBI could not counteract the DNS attack.

In a similar vein, contacting the administrator and CTO were also high frequency actions. Since these actions do not stop an attack, the only justification for them being so frequent is that because they added no risk to the defender, they did not harm the fitness of individuals and were therefore not biased against during crossover (allowing them to reproduce). In a real scenario, these actions would also be quite frequent as most organizations would probably insist on some human interaction with a potential response — at least a notification as these actions provide.

Compared to the other actions chosen by the defenders, the choice to send TCP RST packets was the only 'offensively' minded action chosen with any regularity. However, the same rationale of the other actions apply because sending TCP RST packets provides the defenders with a reasonable response while limiting the risk from their own actions.

Given these choices and non-choices by the defenders, the defensive strategy comes to rely heavily on external entities (e.g. ISP and FBI), while utilizing low-risk internal responses such as sending TCP RST packets. This is a strategy also used by people who informally execute active defense: get the attacker's data (if possible), then hand the defense to another entity for them to stop the attack and assume the risk — although more are moving towards utilizing TCP RST and IP filtering as an option. This defense and attack show that competitive co-evolution can indeed derive security strategies if used with the game model.

D.5 Conclusions

Active defense is a difficult topic because uncertain risks must be weighed with benefits. However, if a method could be introduced to incorporate these issues and produce a strategy, then active defense could be used as a legitimate security tool. In this paper a new technique for developing active defense strategies was illustrated. The technique involved modeling security as a game and utilizing competitive co-evolution as the tool determining the best strategy.

The rules of the game are the actions (of both attacker and defender), the risks of the actions, the relationship of the actions, and the accumulation of risk. Then a population of defenders and attackers is initialized, and standard evolutionary techniques of mutation and crossover are applied in order to produce better strategies that will survive encounters with the opposing population.

The technique was shown to be successful. The fitness of the populations over time was shown to mirror each other, and general conclusions about the nature of the strategies emerged from the experiment. It was shown that in general, the attacking population was much more diverse in its use of actions, while (overall) the defensive population took much more conservative actions and did not overly diversify.

Additionally, the strategies that emerged were realistic and followed current active defense understandings. These strategies were that attackers spoofed their IP, then poisoned the DNS (since there are few defenses for it), and then attempted to change the records in the server. On the other hand, the defenders took a conservative route, using traceroute, contacting ISP and FBI to mitigate the attack, and finally attempting to use TCP RST packets. This shows a strongly risk adverse defender because these actions assume little risk, but have the potential for mitigation.

These results support the hypothesis that using evolutionary techniques, especially competitive co-evolution is successful in producing active defense solutions when coupled with a game-like paradigm. This conclusion can support extending competitive co-evolution to other security strategy production and further work into evolutionary computation as applied to computer security.

Table D.6: Defensive Action Frequency

DEFENSE ACTION	**DEFENSE POSITION**							
	0	1	2	3	4	5	6	7
Null Action	58	58	57	48	57	53	50	52
Contact Administrator	8	2	5	6	6	10	5	5
Contact Chief Technology Officer	3	2	2	6	9	5	7	9
Shutdown Port at Firewall	0	0	0	0	0	0	0	0
Filter IP at Firewall	0	1	1	2	2	1	0	2
Shutdown Server	0	0	0	0	0	0	0	0
Send TCP RST Packet	3	4	6	5	6	5	7	5
Ask ISP to Stop Attack	7	15	7	10	9	7	18	11
Contact FBI	4	2	5	4	1	5	3	7
Use Traceback	17	16	17	19	10	14	10	9
Send Virus on IP	0	0	0	0	0	0	0	0
Initiate DoS Against IP	0	0	0	0	0	0	0	0
Attempt to Hack Attacker	0	0	0	0	0	0	0	0

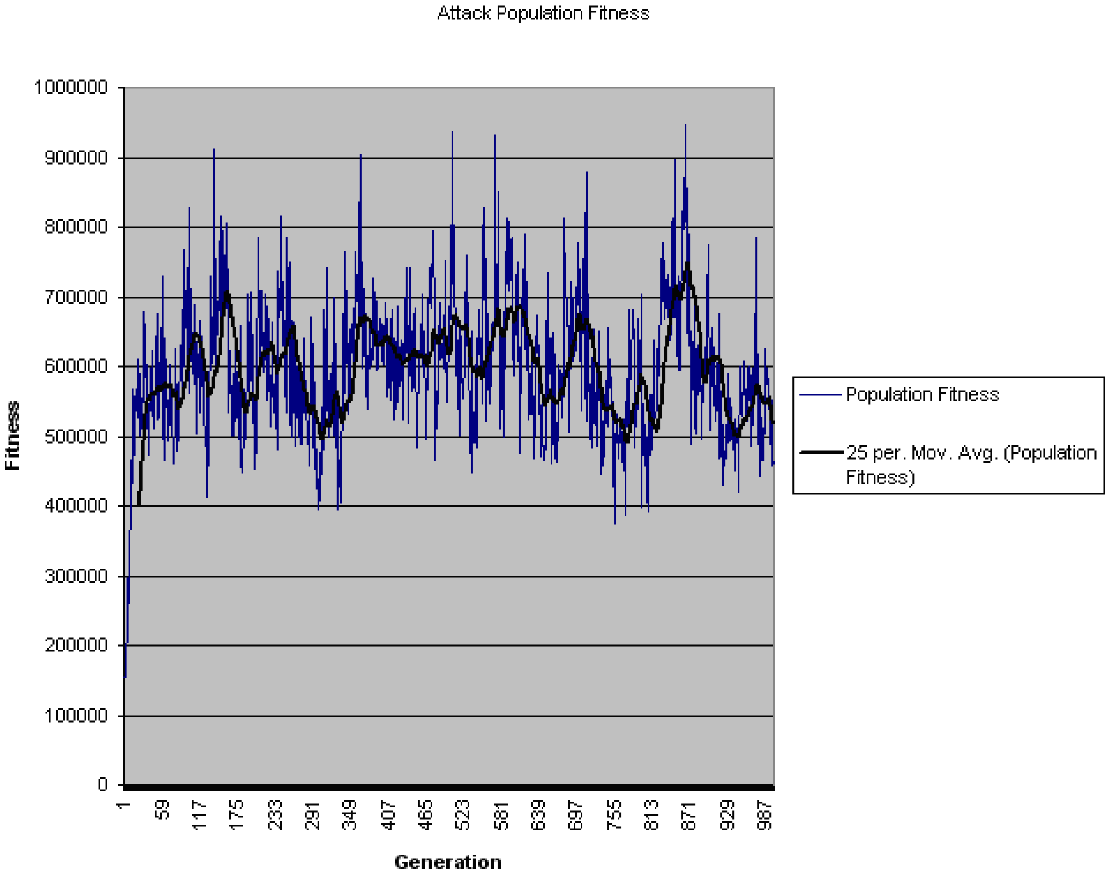


Figure D.1: Attack Population Fitness

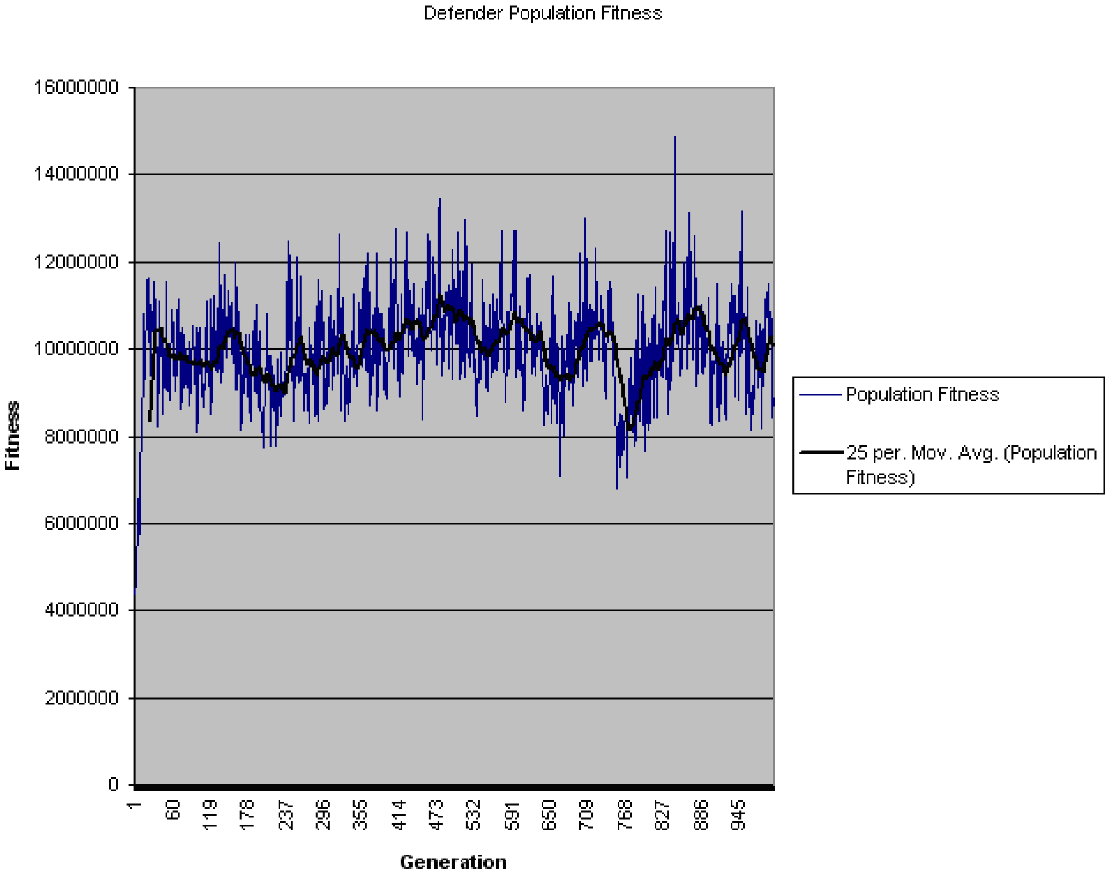


Figure D.2: Defender Population Fitness

Table D.7: Attacker Action Frequency

ATTACK ACTION	**ATTACK POSITION**							
	0	1	2	3	4	5	6	7
Null Action	54	51	56	48	56	43	46	49
Spoof IP Address	39	24	19	7	4	2	0	3
Port-Scan the Server	0	4	6	7	6	5	6	1
Ping the Server	0	1	0	2	3	2	5	1
DoS the Server	0	0	0	0	0	2	2	4
DDoS the Server w/ Zombies	0	1	0	2	2	6	6	5
Poison DNS	7	12	8	17	10	12	8	11
Hack Server, Install Backdoor	0	1	2	2	1	7	4	3
Hack Server, Download Records	0	0	1	0	2	4	2	4
Hack Server, Change Records	0	2	7	8	10	10	13	12
Send Virus Against Server	0	4	1	7	6	7	8	7