ADAM: Active Defense Algorithm and Model

Sergio Caltagirone
University of Idaho
scaltagi@acm.org

Deborah Frincke
University of Idaho
frincke@csds.uidaho.edu

***Abstract*— Defense strategies should provide more benefit than cost. However, deciding "what" a defensive action might cost is not a simple matter, particularly when considering active defense techniques that may involve external stakeholders and resources. In this paper factors are identified that are useful in assessing potential costs associated with an active defense. The model presented here begins with the development of an active defense policy based on local priorities and sensitivity to risk, as well as escalation ladder for selecting among response options. These, taken in combination, allow selection of actions to be taken as well as justifications for those actions once they have been initiated. By investing in this model, an organization frees itself to pursue active defense as a legitimate security (and protection) tool while limiting the associated risks.**

I. 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 pre-emptive 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 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.

II. 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 [1] 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 [1], 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, decision, action, analysis, escalation, and maintenance. These stages can be formally or informally defined by an organization.

A. 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.

1) 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.

2) 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

[1] 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.

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.

B. 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. 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.

D. 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.

E. 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.

F. 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).

G. 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.[2]

H. 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.

III. Goals and Assumptions

A. 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.
- *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.

B. 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.

IV. 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 [2] and the Model Penal Code §3.02 [3]).

The stages of active defense are (partially adapted from [4]):

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

[2] 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.

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

V. 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.

A. 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.

1) 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:

$$S = \{s_1, s_2, \ldots, s_n \mid \forall\, i,\ 1 \leq i \leq n, s_i, s_{i+1} \in S,$$
$$category(s_i) = category(s_{i+1}) \land$$
$$rating(s_i) < rating(s_{i+1}) \land$$
$$rating \in \Re \land -1 \leq rating \leq 1 \land$$
$$rating(s_i) \neq rating(s_{i+1})\}$$

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).

2) 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[3], 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.

3) 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.[4] 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.’ [5]

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.

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.

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

4) 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

[3] 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.

[4] 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.

[5] 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.

the threat:

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

B. 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.

1) 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.

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

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.

2) 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 [5].

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.

3) 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.

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

VI. 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 responsble active defense.

A. 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 occurance). This is expressed as:

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

VII. Algorithm

At this point, the model has been described in detail. This satisifies the first (planning) of the eight stages of active defense identified in section II. 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).

```
   $Active-Defense(t, a, P)$
1  check if $a \in P_A$, else Fail
2  check if $t \in P_T$, else Fail
3  $X \leftarrow$ADModel$(t, a, P)$
4  $n \leftarrow |X|$
5  $riskAssumed \leftarrow 0$
6  for $i \leftarrow 1$ to $n$
7      $k \leftarrow X_i$
8      while $k$ cannot be performed
9          $k \leftarrow$ get next action in $X$
10     $riskAssumed \leftarrow riskAssumed + score(k)$
11     if $riskAssumed > sum(t)$
12         break
13     execute the action $k$
14     if action $k$ achieved $goal(t)$
15         break
```

Now for a description of the algorithm. (1,2) Satisifies 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. (3) Satisfies stage 4 (decision) by retrieving from the model the decision set of actions. (4) Assigns the variable n the size of the set X. (5) Intitializes a new variable $riskAssumed$, which stores a total of the risk incurred by executing the actions. (6) Iterates over the set X. (7) Assigns a variable k the action that in the set X at index i. (8) Checks if the action k can be performed using the information available (e.g. is the IP address correct, ect.) and continues until it finds one. (9) Get the next action in the escalation ladder. (10) Adds the risk of the action k to the current risk assumed. (11) 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. (13) Satisifies stage 5 (action) by executing the action selected. (14) 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 satisified by the fact that the next iteration through the loop will escalate to the next action in the decision set.

A. 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.

VIII. 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 III-A 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 IV, 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. [6]

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 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.

IX. 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

[6] 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.

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.

X. 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.

References

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

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

[3] American Law 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.

[4] D. Dittrich, “Active defenses to cyber attacks,” September 12, 2003.

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