# Adversary Modeling to Develop Forensic Observables

John Lowry, Rico Valdez, Brad Wood
jlowry@bbn.com, rvaldez@bbn.com, bwood@bbn.com

Keywords: Computer forensics, Adversary modeling. Observables, Theory of Observables

*Abstract: Observables of malicious behavior in the cyber realm are derived from intuition or analysis of previous (a-posteriori) events. This creates an untenable situation where cyber defenders are unprepared for novel attacks or malicious behaviors – particularly those expected to be used by sophisticated adversaries. Development of a complete theory of observables with a particular focus on development of a-priori observables is critical to defend against computer network attack and computer network exploitation. Monitoring of a-priori observables will greatly assist in the areas of indications and warnings and attack sensing and warning. Forensic development and analysis of a-priori observables is critical to determine the type of adversary, adversary mission, and ultimately attribution.*

## 1 Introduction

Practitioners of computer network defense (CND) are fighting an asymmetrical battle against a wide variety of attackers and attacks. The most public of these are worms and virii released by 'hackers'. In less well-known cases, attacks are conducted by a variety of named actors that include organized crime, 'insiders', and nation states. This asymmetry is the result of many factors that include policies against active defense; poorly designed, configured and operated systems; and lack of adequate tools, techniques, and procedures. In the last category, there exist few if any observables that would permit monitoring for indications and warnings (I&W) and attack sensing and warning (AS&W). While forensic and counter intelligence (CI) experts are often called in during or after an event, they too lack observables for novel attacks or variations of attack.

The current sets of cyber observables are developed after an attack or event takes place. These are termed *a-posteriori* observables because they follow the pattern of *event—analysis—observables*. Properly specified, these observables will catch most or all repeat events or new events that use the same techniques.[i] These observables have no value in identifying new types of events or novel variations of known events. Since the vulnerability space is huge, defenders are forced into a responsive mode of operation. What is needed is an additional set of observables that will permit the detection and analysis of novel events and attacks. These must be developed *a-priori* and follow the pattern of *threat—analysis—observables*.

## 2 Threat Modeling

Any threat model must start with analysis of adversary behavior and incorporate sufficient knowledge of the defended system. For development of *a-posteriori* observables, real behaviors and real systems are used. For development of *a-priori* observables, hypothetical or potential adversarial behavior is modeled. While this may seem to be an open ended problem, it is important to remember that adversaries typically have a mission, have means (attacks) that have to be used or

have a high probability of being used, and have to operate within the constraints of the environment provided by the target environment. Consequently, the steps in adversary modeling to extract *a-priori* observables are:

- *Hypothesize* potential adversaries or malicious acts.
- Identify *threats* and adversary *missions.*
- Identify the *means* that would *have to be used* or have a *high probability* of being used.
- Develop *observables* for those *means.*

### *2.1 Cyber Adversaries*

The cyber defense community frequently makes the mistake of assuming a kind of omnipotence on the part of the cyber-adversary, particularly when sophisticated cyber-adversaries are discussed. There are several good reasons for this. First, defenders are often quite aware of their own inadequacies and the asymmetric advantage held by adversaries today. This leads them to "game themselves." Knowing all of the problems they face, they tend to assume that a cyber-adversary will know all of their inadequacies too. Second, experience with red teams at various levels has demonstrated that it is "impossible" to pass a security evaluation. This is misleading on several counts since a security evaluation is not representative of an operational scenario. It has none of the characteristics of an operation (or game used in modeling) where moves are met with counter-moves and, most importantly, the security evaluator has no mission to execute (besides demonstrating that they can penetrate) and the security evaluator does not encounter or consider risk. Finally, the cyber-defender has no appreciation that the cyber-adversary has very real constraints in the form of resource availability and consumption. In spite of the above, the cyber-adversary still poses a very significant and frequently overwhelming threat.

It is important to note that just as in every other instance of human conflict, the cyber-adversary is composed of human beings and has properties in common with all other types of adversary.

1. Cyber-adversaries have goals and objectives. There is a reason why the defender's system is under attack. An understanding of a cyber-adversary's goals and objectives, or of the possible goals and objectives any unknown cyber-adversary might have for targeting a particular system, constrains the space of possibilities and identifies opportunities for defense.

2. Cyber-adversaries have resource limitations. Even the most wealthy and capable cyber-adversary faces some kind of resource limitation including time (to carry out the mission), opportunity, qualified and educated personnel, and risk. In short, there are variables that are not under the cyber-adversary's control. Risk is a complicated resource to consider and can change dynamically with time and relative to other variables but is best approximated as a threshold above which the mission or mission element cannot be prosecuted.

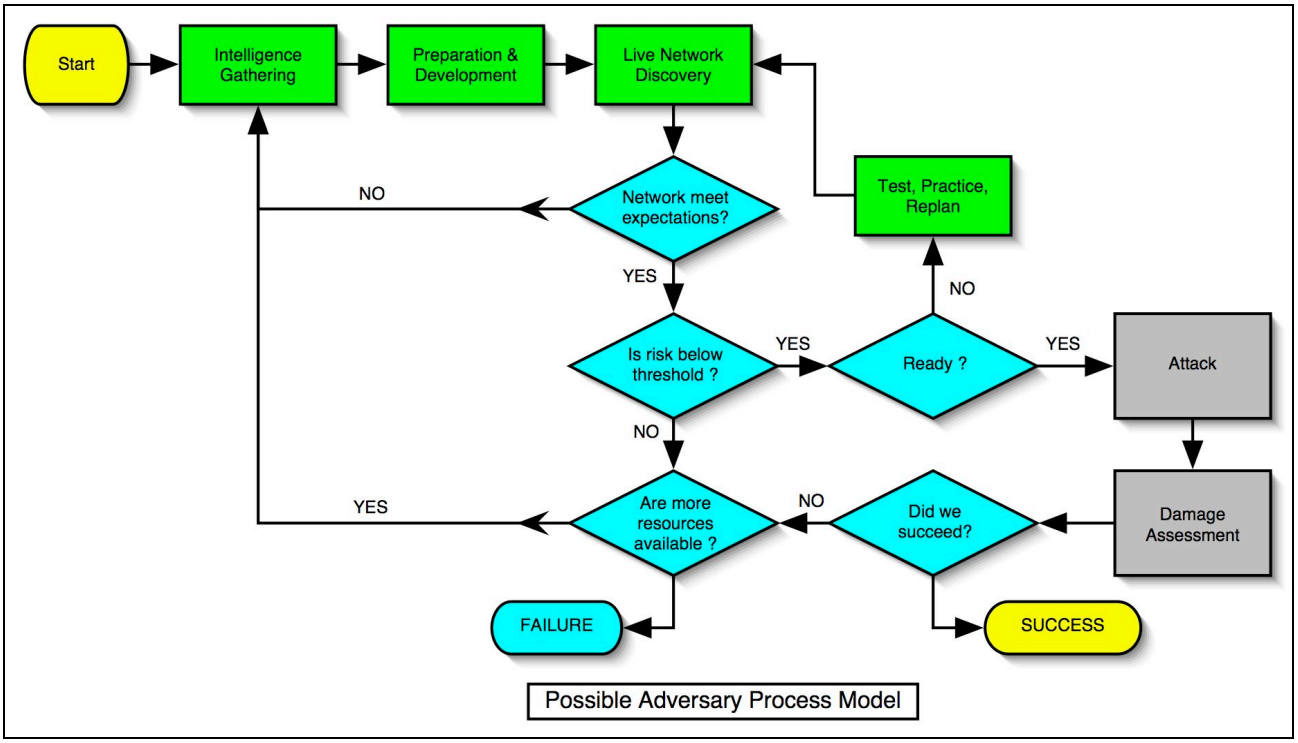


**Figure 1: High-level adversary process model.**

3. Cyber-adversaries engage in mission planning, practice, development and testing – in short all the things done in the kinetic world. This is outlined in Figure 1 above. For certain low-level threats, this kind of activity may be ad-hoc, but it is safe to presume that the amount of energy a cyber-adversary devotes to this process goes up with sophistication and degree of perceived risk. Observations of model adversary behavior over several mock attacks suggest the fraction of time devoted to different portions of the process. This is illustrated in Figure 2 below. These two charts suggest that identifying observables for, or identifying means to frustrate the Intelligence/Logistics, Detailed Preparation, and Test & Practice phases could be highly effective.

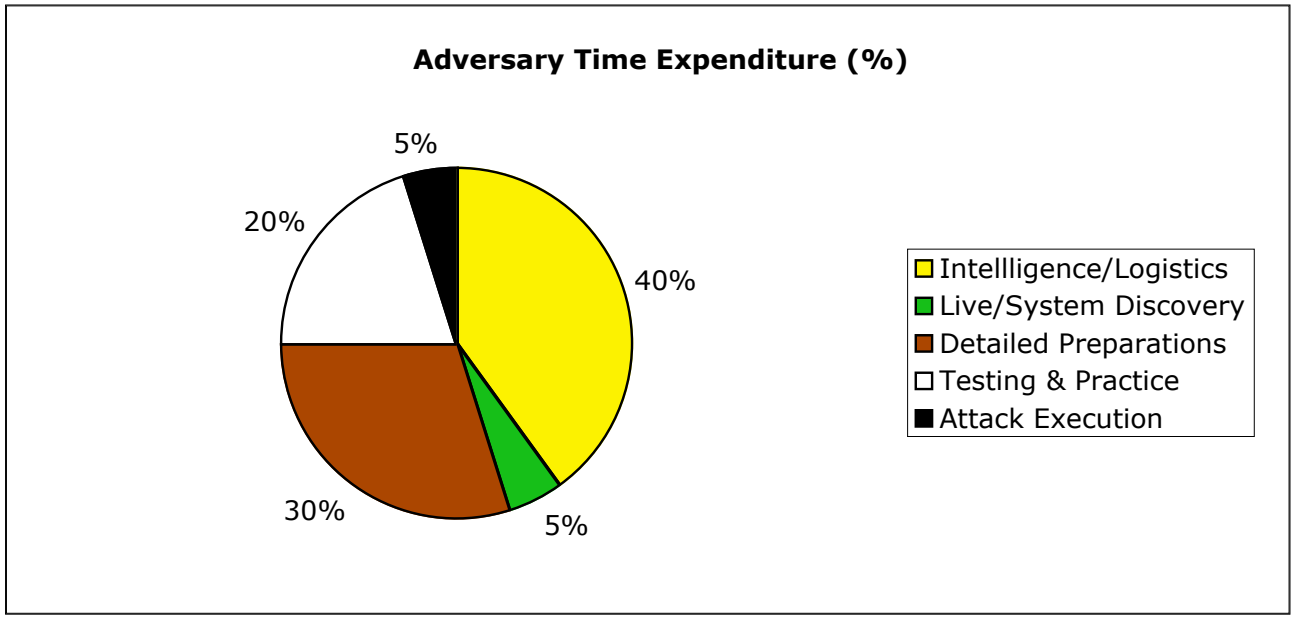


**Figure 2: Adversary Time Expenditure**[ii]

4. Cyber-adversaries translate their behavior into the world of computers and networks. This is a constrained physical and virtual space with "rules" and limitations. Kinetic adversaries translate their behavior into the "real" world of physical space: gravity, mountains, deserts, numbers of warriors, and so on. Adversary operations in the real world not only impose limitations but also yield observables. The same is true for cyber-adversaries.

### *2.2 Schema for Cyber Adversaries*

The choice of an adversary[iii] schema[iv] is critical for modeling because a schema guides and focuses thinking about the adversary. The most commonly used schema for adversaries is the named actor or naming schema. As will be described below, its best use is to focus on particular and well-identified adversaries. Properly used, this has the advantage of focusing analysis and resources on particular threats such as *China* or *al-Qaeda*. However, the named actor schema is frequently misused to generate what are apparently named actors but are actually inappropriate generalizations. Two of the frequently used ones are *nation-state* and *terrorist.* A cursory examination of these terms will reveal that both *Russia* and *Botswana* are nation-states but represent very different threats and that classifying them together as nation-states hides important differences. The same is true for *al-Qaeda* and the *IRA*.

#### 2.2.1 The Named Actor Schema

As noted above, the named actor schema works well when talking about precisely identified adversaries. *Russia, China, France, North Korea, Great Britain,* and *Jamaica* are all pose distinct threats and need to be treated separately. All general named actor schemas include "Nation/State" to imprecisely represent 'a well resourced, generally hostile, internationally recognized nation that

poses a significant threat'. Table 1[v] below is typical of named actor schemas and shows the difficulty of attempting to discuss the nature of threat and risk at this level of generality.

| Adversary | Description |
|---|---|
| **Malicious** | |
| Nation States | Well-organized and financed. Use foreign service agents to gather classified or critical information from countries viewed as hostile or as having economic, military, or political advantage. |
| Hackers | A group or individuals (e.g., hackers, phreakers, crackers, trashers, and pirates) who attack networks and systems seeking to exploit the vulnerabilities in operating systems or other flaws. |
| Terrorists/ Cyberterrorists | Individuals or groups operating domestically or internationally who represent various terrorist or extremist groups that use violence or the threat of violence to incite fear with the intention of coercing or intimidating governments or societies into succumbing to their demands. |
| Organized Crime | Coordinated criminal activities, including gambling, racketeering, narcotics trafficking, and many others. An organized and well-financed criminal organization. |
| Other Criminal Elements | Another facet of the criminal community, but one that is normally not very well organized or financed. Usually consists of very few individuals or of one individual acting alone. |
| International Press | Organizations that gather and distribute news, at times illegally, selling their services to both print and entertainment media. Involved in gathering information on everything and anyone at any given time. |
| Industrial Competitors | Foreign and domestic corporations operating in a competitive market and often engaged in the illegal gathering of information from competitors or foreign governments through corporate espionage. |
| Disgruntled Employees | Angry, dissatisfied individuals who can inflict harm on the local network or system. Can represent an insider threat depending on the current state of the individual's employment and access to the system. |
| **Nonmalicious** | |
| Careless or Poorly Trained Employees | Users who, through lack of training, lack of concern, or lack of attentiveness, pose a threat to information and information systems. This is another example of an insider threat or adversary. |

**Table 1: Example Named Actor Schema**

Once this level of naming schema is established, many analyses proceed to try and develop an understanding of *resources, opportunity,* and *motivation.* Unfortunately, these are often divorced from the just-developed naming schema and are themselves developed independently, at a high level of abstraction, and with no linkages to the named actor. Some analyses do try to develop an understanding of resources, opportunity and methods that are linked but quickly discover that the range and breadth of each of these categories prevents a solid characterization. For example, one analysis discusses the motivation of "embarrass the target" but it is obvious that this can apply to any of the adversaries in Table 1. There is no understanding of whether "embarrassment" is a

primary or secondary goal, the degree of embarrassment desired, in what context, etc. Understanding of the threat posed by this motivation is impossible at a general level and must be developed once specifics are available about the real actor and real target.

### 2.2.2 Class Schema

The purpose of the class schema is to provide an appropriate level of abstraction that can be combined with specific named actors as they become known. It will permit analysis at a high level of abstraction – obviously with high-level results, but also at finer degrees of granularity.[vi] The schema is defined in terms of resources, opportunity, and motivation, which are elements of capability and intent. It is open ended, but as a starting point, identifies four classes of adversary with Class I as the least capable and Class IV as the most capable.

The class schema allows the characterization of named actors as part of the next level of granularity. At a very abstract level, the traditional named actors can be grouped by class as shown in Table 2 below. Even this isn't very informative – it requires additional specificity and intelligence to group real-world actors into a particular class.[vii] Note that any first-world country would automatically go into the Class IV category because of the capability (resources and opportunity) they possess, however, the classification requires evaluation of motive and intent. Consequently, strong allies would not be considered as adversaries simply because they are not adversaries.[viii] Certain countries belong in the Class II category because it would be extremely difficult for them to develop a capability. Similar things can be said about other named actors. For example, we have not seen development of a Class IV terrorist organization in the *cyber* arena mostly because of the excessive resources required for them to become a Class IV adversary. The kind of resources and opportunity required simply haven't been available.

| **Class** | **Named Actor** |
|---|---|
| Class IV | First-world and certain second-world countries, including military and intelligence agencies. Future terrorist organizations. Future organized criminal groups. Some types of insider. |
| Class III | Almost every country not in the Class IV category. Some terrorist organizations. Some organized criminal groups. Some types of insider. Some types of radical organizations. |
| Class II | A very few countries. Many terrorist organizations. Many organized criminal groups. Many types of insider. Many types of radical groups. Very expert hackers and hacker coalitions. |
| Class I | Some terrorist organizations. Some organized criminal groups. Many types of insider. Many types of radical groups. Beginner to journeyman hackers. |

**Table 2: Class to Named Actor Type Mapping**

### *2.3 Class IV Process Model*

The process model above shows a high-level process model of adversary behavior. However, it can be expected that a Class IV adversary will engage in a much more detailed set of behaviors. There is a strategic set of goals followed by assignment of missions and mission objectives.

The adversary's strategic planning can be represented in a Warnier/Orr diagram. The goal is to identify effects that can be achieved, i.e., to identify the top-level opportunities and resources available to carry out the strategic mission. The figure below suggests the following high-level strategic planning process:

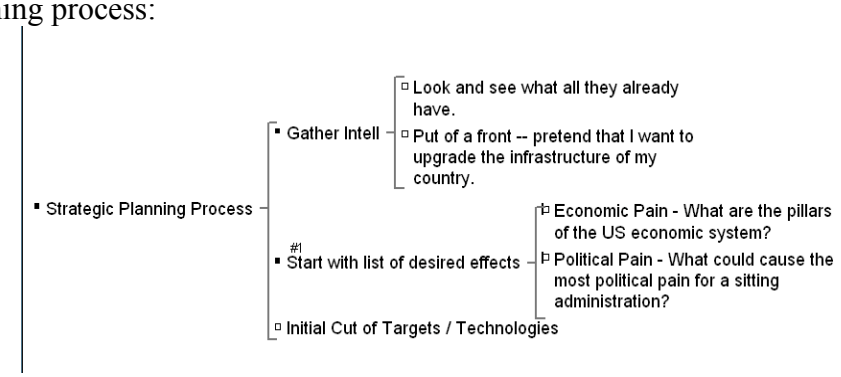


**Figure 3: High Level Warnier/Orr Diagram**

1) The adversary will study their enemy to determine what they have in place and how they operate.
2) The adversary will develop a list of desired effects that the adversary wishes to have on their enemy.
3) The adversary also takes an initial, high-level cut at the targets of interest.

What is further implied is that this process is iterated and refined, so that as this process is repeated, the intelligence improves, the list of desired effects becomes more precise, and the list of targets becomes more precise. In fact, one can apply spiral development process to this paradigm to obtain an optimal result as suggested by Figure 4: Adversary Target Management Process below.

Several areas in this process appear that may yield observables depending on the specific details of how an adversary operates. Among these are typical ones like resource planning and allocation. It may not be possible to determine when resources are being allocated through strictly cyber means but it would be highly suggestive if an adversary who previously had little or no interest in particular types of computer and network technology suddenly increased their ability in these areas – particularly if there were no obvious need to do so.[ix]

Another area that stands out is the need to gather target intelligence. The adversary typically needs detailed intelligence to successfully attack a computer or network system. Detail of configuration and deployment are critical in order to reduce the risk of discovery and failure. While many high-level details can be gathered through published material and training courses, the particulars of any

installation are critical and one can expect an adversary to attempt to obtain these. Frequently this is done through intrusion, which can be cyber (network based) or by an insider although other means exist.

Finally, for this class of adversary, there is frequently a desire to determine effects and progress. In a complicated multi-stage attack, monitoring of effects and progress is critical. This suggests active monitoring[x] for effects and those monitors will have some element of observability.

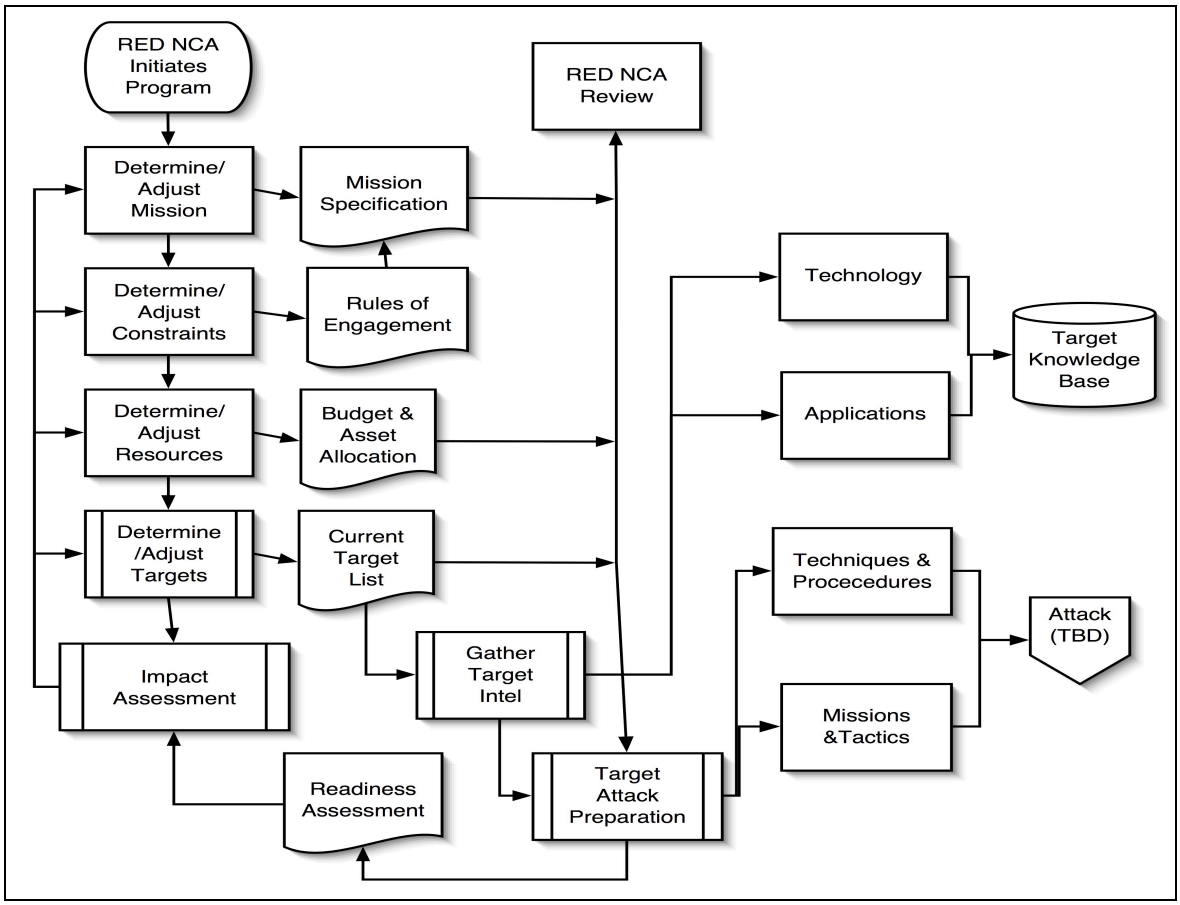


**Figure 4: Adversary Target Management Process**

## 3 Class IV Adversaries and Forensics

The discipline of computer forensics has been largely focused on the development of a set of excellent tools and procedures. However, the majority of efforts have remained at this level and not progressed to meet the challenge of Class III and Class IV adversaries. With the resources available to these adversaries, it is not apparent that analysis of single exploits or events will help to identify and analyze the presence of these adversaries. For example, it is understood that an adversary will not use his most valuable or sophisticated techniques or methods unless there is sufficient payoff.[xi]

Consequently, identification of Class IV adversaries must look for supporting evidence. Fortunately, the kinds of process and control exercised by this type of adversary is likely to leave such evidence.

### *3.1 Forensics, Observables, and Class IV Adversary Time Expenditure*

Figure 2: Adversary Time Expenditure above contains a hidden lesson and is very suggestive as to where additional observables and lines of evidence need to be developed. By far, the majority of time spent by an adversary is in intelligence and logistics. Detailed preparations, and testing and practice follow in terms of total expenditure. Tellingly, a minimal amount of time is spent on live discovery and system discovery and yet that is one of the primary focuses for CND and forensics practitioners. Significant effort needs to be made to develop observables for areas where the adversary spends the majority of time.

Actual attack execution takes the least amount of time. The model of adversary time expenditure was developed over multiple engagements and experiments designed to understand adversary processes using national-level red teams tasked to model Class III and Class IV adversaries. The hidden lesson is that *the defenders lost every engagement if the adversary was allowed to proceed to attack execution without being discovered and challenged*. Further, unless the system was completely instrumented, there was little or no evidence that could identify the class of adversary or to provide attribution. Adversary mission could frequently be inferred but not supported through overall system effects.

### *3.2 Forensics, Observables, and Class IV Adversary Target Management Process*

Figure 4: Adversary Target Management Process above also suggests the existence of a set of observables and forensic evidence. While many of the elements may require all-source intelligence or may never be observable, several phases of the process will be observable. Development of targets may very well be visible in certain environments. For example, there is surprising diversity in air traffic control systems in the United States today. This is due primarily to the way systems for airports and regional control centers were acquired and are managed. Significant effort by an adversary would have to be made to target these systems and to gather target intelligence. Any forensics practitioner called in to investigate one of these systems should be looking for evidence as to whether the incident could be the result of intelligence gathering and not assume that an adversary mission has been attempted.

Adversary impact assessment was mentioned earlier and technical means to perform impact assessment should also be looked for. Trojans and backdoors are highly likely to be used by a Class III or Class IV adversary. This was universally observed during engagements with model adversaries and frequently described as “consolidating the position.” However, other techniques representing “heartbeats”, whether regular or stimulated, have been used.

Evidence for multiple phases of target management strongly suggests the class of adversary. Observables for these activities may be found using current technologies and methods. However, an

effort must be made to develop new ones based on understanding adversary behavior, possible missions, and procedures as well as techniques.

Explicitly stated, a Class III or IV adversary will provide multiple opportunities for observation. These observables will not only represent the different phases of target management but it is highly likely that multiple observables will exist within a particular phase. In short, multiple "exploits" and "techniques" can be expected to be used, each with observables.

### *3.3 Types of Observables*

In general, there are two different classes of observables. The first class of observables is based on the *state* the attacker has achieved or wishes to achieve. State observables are the most important due to the fact that they are based on *necessary* states the adversary must attain in order to attain his mission.

The second class of observables is based on the *transitions* or attempted transitions to and from *states*. These are important because they may be easier to detect and they may indicate the state an adversary is attempting to attain and possibly the state that an adversary is currently in. States and transitions map easily onto graphs with nodes and edges where the state (or node) represents a kind of access to the system and the transition (or edge) represents techniques (exploits) used to move from one state to another.

This kind of graph is universally used – whether drawn as a diagram or described in text. It is an overview of an attack and can be recaptured through forensic analysis. The attack graph is a representation of some or all of the adversary's plan and as such can be used to determine adversary's mission and to assist in attribution.

States and transitions have properties that, if determined, can also assist in classifying adversaries. This is intuitively understood when considering properties commonly lumped into the category of "sophistication". Resistance to detection, analysis, and detection are the primary categories of concern.[xii]

In some cases, the observable is composed – individual events that together fail to cross a threshold or may even represent normal behavior. Typically, composed observables are made up of information from various sources or are arrived at by the correlation of several detectable events. For example, if event *x* followed by event *y* followed by event *z*, this "chaining" of sequential events together can provide a single observable that is unlikely to be the result of legitimate activity on the system. This can also be used for combinations of states.

Temporal information can also be incorporated in these observables. It may be the case that some activities must happen within a set amount of time of another activity. If temporal information is included in the observable, it may cross a threshold of significance. Again, this temporal information can be incorporated into observables on transitions or on states.

### 3.3.1 State Analysis

From studying attacks, it if apparent that an attacker typically has multiple steps that must be completed in order to successfully achieve the goals of the attack. This is more likely to be true in more secure networks or hosts that are protected by layers of defenses. These steps can be thought of as sub-goals of the attack, as they themselves are goals and may require other such steps and sub-goals. This composition of states is also found in attack trees. The decomposition can go very deep depending on the level of detail that is desired.

The general concept of attack graphs is critical to develop observables. Adversary goals can be translated to states with little difficulty, as the concept of being in a particular state is typically intrinsic in the purpose of the goal. In addition, the graph can be abstracted to a level that is less detailed, but much easier to manage and work with. Attacks are represented in the state transitions and the pre- and post-conditions of the attacks, represented as states, are what are listed in each node.

Using this model, the states that are necessary in order to achieve a specific goal become clear. Necessary states and frequently the order in which they need to be achieved can be identified. The transitions between these states are represented by different attacks, exploits, methods, and tactics. If the adversary can be detected in a specific state, the method that was used to reach that state becomes significantly less important. This is a critical point. In most cases, we don't need to know *how* the adversary achieved a certain state as much as the fact that they did. Some ramifications of this are –

- Detection now becomes a matter of engineering since we can derive technical means to determine presence at a state
- Sophistication is no longer an independent variable capable of masking presence, as all adversaries are required to pass through the same states
- New and novel attacks never before seen or thought of can be detected.

Determining what these states are and the sequence in which they must be attained will be tailored to the system of interest. Although there will be some set of states that can be broadly applied to many different systems, the states necessary to achieve different adversary goals will be dependant upon the system. While this introduces additional work into the application of this methodology, it is necessary to develop observables that have a high degree of fidelity and utility. It should be noted that this is considerably less resource intensive than developing detailed attack trees, and if attack trees already exist, they can be translated into this new format without a heavy resource commitment. It is highly recommended that this type of activity take place for high-value targets, as it is important to understand the threats to these information systems.

### 3.3.2 State Transitions

The transitions between states are frequently more easily detected than the existence of being in the state itself. The transition is typically where the adversary is interacting with the system and creating the most observables. Hence, these transitions should not be discarded. At a minimum, they can give an indication of the state that the attacker is attempting to transition into or the state that the adversary has come *from*. It may also be the case that there is no good way to detect that the adversary is in a certain state and only state transitions can be detected. This may be sufficient if all the possible ways to transition into a specific state can be identified, which may be the case for well-defined states with limited transition vectors. It is important to select well-defined states to use in attack graphs.

## *3.4 Future Development*

It is important to note that much of this work comes from observation and experience with real and model adversaries. Technical work on states and state transitions is ongoing and subject to refinement. However, these initial findings and results appear to be significant and will be the focus of additional work.

# 4 Support

While the development of new models and characterizations of cyber-adversaries has been informally pursued for several years and within multiple government-supported programs, the full development and presentation is made under an effort called *Theory of Observables* within the *Proactive and Predictive Cyber Indications and Warnings* contract from the Advanced Research and Development Activity (ARDA). ARDA's web site is located at www.ic-arda.org.

This material is based upon work supported by the Air Force Research Laboratory and Advanced Research and Development Activity (ARDA) under Contract No. F30602-03-C-0232. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Air Force Research Laboratory or the Advanced Research and Development Activity (ARDA).

---

[i] However, there is a distressing tendency to over-specify the resulting observables as signatures, which significantly reduces the value of the observable for that kind of attack.

[ii] "An Initial Foray into Understanding Adversary Planning and Courses of Action," in the proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX II), Anaheim, CA (12-14 June 2001), vol. 1, pp. 123-133.

[iii] The term "adversary" will be used throughout the rest of this document to mean a cyber-adversary exclusively unless otherwise noted.

[iv] Schema: 1. A diagrammatic representation; an outline or model. 2. A pattern imposed on complex reality or experience to assist in explaining it, mediate perception, or guide response. *The American Heritage® Dictionary of the English Language, Fourth Edition*

[v] This table is taken from the IATF. All the adversaries have been over-generalized with the possible exception of "Disgruntled Employee". Other named actor schemas usually roll up this adversary into the general term "Insider". IATF Release 3.1, Appendix I, p. 8, http://www.iatf.net/framework_docs/version-3_1/index.cfm.

[vi] At this stage of the program, the development of finer granularity is a work in progress but early indications strongly suggest that this is possible and useful.

[vii] Specificity is not appropriate for this kind of paper. The author does not have access to suitable information to make such a classification.

[viii] There are very few countries with which the U.S. shares this kind of relationship. Even then, the best of allies are always a bit suspicious of each other. Consequently, it would be a mistake to say that ally "A" posed no threat, but it would be acceptable to state that ally "A" sits below a threat threshold and exclude them from the model based on judgment.

[ix] Just as al-Qaeda needed to develop persons with the ability to maintain and guide aircraft that were already aloft, certain adversaries will need to acquire knowledge, training, or components of advanced telecommunications switching and SCADA systems in order to be effective.

[x] Passive monitoring, such as monitoring news reports, will also be used but will not generate specific observables – or perhaps any observables at all.

[xi] Leonhard, Robert R., The Principles of War for the Information Age, Presidio Press, ISBN 0-89141, 1998. "The law of economy states that man is weak and lacks resources sufficient to serve his goals in conflict. Further, the supreme danger of armed conflict causes it to be an exceedingly wasteful enterprise. Therefore, to prevail in conflict, one must economize as much as possible." The law of economy in this context deals primarily with resources. Most importantly, it is focused on the best use of resources to achieve a goal.

[xii] The concept of *rate* as in tempo of attack is frequently cited as an important property in determining sophistication. This is a mistake as rate is more likely to be an indicator of sub-type of an adversary class. It seems reasonable to assume that "slow and cautious" is a trait of an adversary engaged in computer network exploitation (CNE) and may represent foreign intelligence. A fast attack might be considered a trait of computer network attack (CNA) as represented by a foreign military. While it seems reasonable, it is important to note that *there is no evidence to support assignment of one behavior or the other to a particular sub-type and that a Class IV adversary will select the tempo based solely on mission needs.*