IST-128 Workshop on Cyber Attack Detection, Forensics and Attribution for
Assessment of Mission Impact

Position Paper: Estimating Attack Intent and Mission Impact From Detection Signals

April 1, 2015

Patrick McDaniel and Robert J. Walls
School of Electrical Engineering and Computer Science
Penn State University
360A IST Building
University Park, PA 16802
{mcdaniel, rjwalls}@cse.psu.edu

1. Introduction

While measuring *security* is an unsolved and important area, measuring *system behaviors* in terms of performance and capability is a well-established science. We argue that measuring security—and hence understanding environmental threats—relies on the projection of system measurements (detection signals) onto mission needs and adversarial objectives. Put succinctly, the best security metric identifies how well the observed system can achieve its mission objectives. The best attack metric identifies how well the adversary is achieving its adversarial goals.

Historically, defensive cyber-systems have focused at identifying attacks based on observable system behaviors; this is the basis for modern anomaly and intrusion detection systems. Such measurements attempt to identify adversarial behavior based on models of normal or aberrant behavior (e.g., signatures). The goal is to identify what attack is occurring and specifically *not* what impact that attack has on the system or environmental goals. However, simply identifying attack type does not often provide a clear view of what the goals of the adversary are, how the attacks impacts ongoing mission objectives, or how its effects can (or should) be mitigated.

This paper introduces a vision for security that attempts to infer attack intention and the impacts of an attack on the missions in progress, rather than diagnosing the identity of the attack itself. Presented below, we see this effort as breaking down into two interrelated phases of analysis. The first phase discussed Section 1.1 posits how detection signals can be used to identify resource or performance related impacts that impact an active cyber-mission. The second focus discussed in Section 1.2 attempts to project those state changes on a mission plan described by an operational model. We conclude by exploring a range of challenges introduced by this research agenda.

The effort highlighted throughout is begin carried out within the Cyber-Security Collaborative Research Alliance (CSec CRA, or just CRA) [CRA15]. The CRA is a consortium of academic, military and industrial researchers been investigating the techniques for ensuring mission progress in the presence of adversarial action. The goal of the CRA program is to understand and model the risks, human behaviors and motivations, and attacks within military cyber-maneuvers. The overarching scientific goal of this effort is to develop a rigorous science of cyber-decision making that enables military environments to a) detect the risks and attacks present in an environment, b) understand and predict the motivations and actions of users, defenders, and attackers, c) alter the environment to securely achieve maximal maneuver success rates at the lowest resource cost.

2. Overview

Figure 1 describes a preliminary analysis framework. At a high level, we map attacks onto the adversarial goals and impacts on a system. This requires us to manually or automatically identify

how an attack manifests on the victim, as well as the local impacts on its resources. Once identified, the impacts are mapped onto the mission objectives and plans to determine when a mission outcome may be in jeopardy. This analysis is used in the context of a mission plan to determine when an attack is impacting a mission, identify where the impacts of an attack will present problems (now and possibly later), and to enable alteration of mission strategies to increase the likelihood of a positive mission outcome.

Figure 1 – Intent and Impact Analysis

1.1. Attack Identification and Intent Analysis

The first stage in this approach is the identification of system measurements that can indicate the presence of an attack. This is the widely studied detection problem, and we defer to the vast literature and systems for solutions that address them. Here, it is sufficient to assume the identification of attacks. Note that system performance measurements may also be used to identify system state.

The second stage is to relate those known attacks to impacts onto intents. Here, we define an *intent* of an attack as a set of one or more impacts (e.g., availability, integrity, confidentiality, or performance) on *resources* (targets). Note that an attack can have multiple intents. Initially, we will hand-label intents based on the documented behaviors of the known attacks as well as our experimental observations. In the longer term, we seek to infer intents based on system measurements. Such inference can be difficult because causality in complex systems is inherently vague and often unknowable from simple measurements.

This investigation of intent is similar the investigation of attack strategies. For example, attack trees are a means of creating structured models enumerating the ways that attacks can be used in concert to achieve a particular adversarial goal [Sch99]. Other methods of modeling attackers used attack patterns [HM04, GW05] which was developed from fault analysis techniques in aviation and nuclear power systems.

One interesting question that comes about from this effort is what exactly are the scope and semantics of resources and impacts. One approach is to develop ontologies [Gru95, Gua98, OCWD14] for resources and impacts. Such ontologies provide a way of articulating these features at different levels of abstraction and granularity. To see why this is necessary, consider two kinds of network-based denial of service attacks. Attack *A* floods the network interface of a victim machine with large packets, while attack *B* sends many TCP syn-requests that consume entries in the operating system connection table. The intents of these attacks are similar (reduce the network performance), but have vastly different vectors and consequences (consume bandwidth vs. preventing successful incoming connections).

1.2. Mission Impact Analysis

One of the areas of concentration within the CRA is the development of formalism for describing mission plans and strategies called an operational model. To simplify, an operational model is an annotated finite state machine that describes transitions (maneuvers) that can be undertaken to move a cyber-mission from an initial state (start) to an end state. Figure 2 shows a partial example of a mission as represented in the substantially simplified operational model. This example mission implements a generic request/response exchange relating to the acquisition of data through a series of discrete steps. Each state in the model is annotated with a set of preconditions that represent requirements for a state to be reached. Importantly, the preconditions are formulated as an expression over the resource states that are affected by attacks. This allows us to track the system state changes over time, and importantly how an attack impacts the mission.

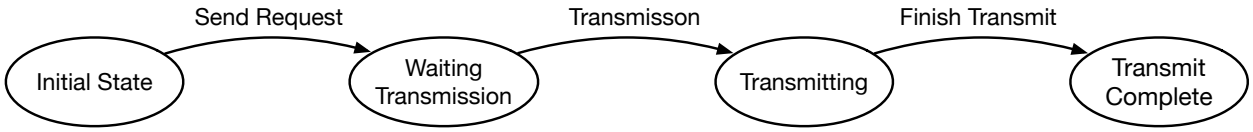

Figure 2 - Simplified Operational Model Example

Focusing on the example, the states "waiting transmission" and "transmit complete" have practical preconditions for its operation. The "waiting transmission" state can only be reached if the source from which the data is acquired is reachable and is receiving requests. Further, "transmit complete" state can only be reached if connectivity is maintained and there is sufficient bandwidth to support the entire transmission.

Attack intents allow us to reason about progress of an environment executing a mission using the operational model. Once detected, we can formally reason about the effects of the impacts on the preconditions of the operational model states by evaluating the precondition expressions over the resource states. That is, the impacts restrict the set of reachable states by making the preconditions unsatisfiable. To see why, consider again the execution of attacks A (packet flood) and B (syn floods) in executing the sample data acquisition mission. Attack A does not prevent the system from entering into a wait state because it restricts the bandwidth but allows the request to connect (with some probability). However, such an attack would prevent the process from reaching the desired end state (transmit complete) because there is not sufficient bandwidth to complete the transfer. Conversely, attack B would prevent the wait state from ever being reached and therefore the mission would fail.

There are several advantages to this approach intent-impact analysis. First, an observer can determine whether a mission can be completed successfully in the presence of an attack before an impact is realized. In the case of the above example, a system under attack A would know that bandwidth needed later would is not available and would never send a request in the first place.

Second, this analysis provides for missions to alter their mission strategies when it is determined that a mission end-state is not achievable. In this case, the analysis could identify alternate paths through the state machine that would arrive at the end state. For example, the operation could employ countermeasures to mitigate the effects of the attack. In the case a new state could be introduced that enables syn puzzle countermeasures as a precondition to the "protected" wait state. In this way, the model can codify responses to adversarial action and predict future progress.

3. Research Challenges

Reasoning about attack intent and mission impacts introduces a number of intriguing research issues. These include:

- Understanding how to represent intent, at what level of granularity, and how large is an open issue. While ontology development will help, a clear understanding of these issues can only come about through experimental and operational experience.
- Determining causality and intent of an attack is difficult. For example, it is often difficult to determine the difference between intended system behavior (e.g., excess CPU load based on local workloads) and adversarial actions.
- New attacks will exploit new systemic features. It is our expectation that intents will remain largely the same (once we have evaluated a sufficiently large sample of attacks). Yet, this hypothesis needs to be confirmed.

The answers to all of these questions will be the substance of the CRA research efforts in the coming years.

Bibliography

[CRA15] Cyber Security Collaborative Research Alliance (CSec CRA), 2015, http://cra.psu.edu/

[Gru95] Gruber, Thomas R. "Toward principles for the design of ontologies used for knowledge sharing?" *International journal of human-computer studies* 43.5 (1995): 907-928, 1995.

[Gua98] N. Guarino (ed.), Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

[GW05] Michael Gegick and Laurie Williams. Matching attack patterns to security vulnerabilities in software intensive system designs. In SESS '05: Proceedings of the 2005 workshop on Software engineering for secure systems—building trustworthy applications, pages 1–7, New York, NY, USA, 2005. ACM.

[HM04] Greg Hoglund and Gary McGraw. Exploiting Software: How to Break Code. Addison Wesley, 2004.

[OCWM14] Alessandro Oltramari, Lorrie Cranor, Robert J. Walls, and Patrick McDaniel. "Building an Ontology of Cyber Security." *International Conference on Semantic Technologies for Intelligence, Defense, and Security* (STIDS), 2014.

[Sch99] Bruce Schneier. Attack Trees. Dr. Dobb's Journal, December 1999.