Game-Theoretic Threat Screening and Deceptive Techniques for Cyber Defense

Aaron Schlenker
University of Southern California
aschlenk@usc.edu

Abstract

My research addresses the problem faced by a defender who must screen objects for potential threats that are coming into a secure area. The particular domain of interest for my work is the protection of cyber networks from intrusions given the presence of a strategic adversary. My thesis work allows for a defender to use game-theoretical methods that randomize her protection strategy and introduces uncertainty to the adversary that makes it more difficult to attack the defender's network successfully.

Introduction

Screening for threats represents a significant security challenge, whether it is preventing an attack at on an enterprise network, a sports complex, or interdicting illicit cargo shipping. For computer network security, automated intrusion detection and prevention systems (IDPS) and security information and event management tools (SIEM) are used which generate alerts that must be investigated by human cybersecurity analysts. These analysts assess whether the alerts were generated by malicious activity, and if so, how to respond. Unfortunately, these automated systems are notorious for generating high rates of false positives. Expert analysts are in short supply, so organizations face a key challenge in managing the enormous volume of alerts they receive using the limited time of analysts. Failing to solve this problem can render the entire system insecure, e.g., in the 2013 attack on Target, IDPS raised alarms, but they were missed in the deluge of alerts.

Threat Screening Games (TSGs) (Brown et al. 2016) have been introduced to model screening domains where the screener must process a set of people or objects coming into an secure area, while an adversary tries to sneak in an attack through screening. The goal of the screener is then to find the most effective way to allocate the incoming objects to screening resources while ensuring the capacity constraints of the resources are met. Although TSGs are broadly applicable, in many real world domains TSGs fail to account for salient features in problem settings like cyber security. In my thesis work, I have addressed three of these limitations; (1) I extend TSGs to model general-sum games, (2) I introduce a time component for screening resources to model domains where screening certain objects takes varying amounts of time, and (3) I allow for attack methods to show up as different alert types with varying probabilities.

Current Results

Threat Screening Games

A *threat screening game* (TSG) is a Stackelberg game played between the screener (leader) and an adversary (follower) in the presence of a set of non-player screenees that pass through a screening checkpoint operated by the screener. The defender has a set of screening resources which are used to screen the incoming objects and must determine a randomized screening strategy to use. In security games the defender can commit to a randomized strategy by computing an optimal mixed strategy $\mathbf{q}$ to commit to. A mixed strategy $\mathbf{q}$ for the defender is a commitment to a probability distribution q_P over a set of pure strategies P, where a pure strategy is an integer assignment of incoming passengers to screening resources. Every mixed strategy $\mathbf{q}$ can be represented compactly by a marginal strategy $\mathbf{n}$, where each entry in the marginal strategy essentially denotes the expected number of passengers to be screened by a given resource. Unfortunately, computing the optimal mixed strategy $\mathbf{q}$ is an NP-hard problem in many security settings due to exponentially many pure strategies and scaling up to handle large problem sizes becomes an issue. In the security games literature, two approaches are commonly used to handle scale-up: directly computing marginal strategies (Kiekintveld et al. 2009) and column generation (Jain et al. 2010). To solve large-scale zero-sum TSGs a marginal-based approach is employed, and it is shown that such an approach significantly outperforms the use of column generation. Hence, my thesis continues to use marginal strategies.

When computing the optimal marginal strategy $\mathbf{n}$ the defender must deal with the issue of *implementable*. A marginal strategy $\mathbf{n}$ is said to be implementable if there is a corresponding mixed strategy $\mathbf{q}$ which implements the marginal strategy (which is needed for the real world implementation of a strategy). The issue of implementability in zero-sum TSGs was resolved by using a special condition on the constraints, called a "bihierarchy", which can guarantee when marginals returned for the defender are im-

plementable. The Marginal-Guided Algorithm (MGA) was then developed which converts the constraints to a bihierarchy and then computes the defender's marginal strategy.

Addressing Limitations of TSGs

The first contribution in my thesis work concentrated on solving for a defender's optimal strategy in *general-sum* TSGs (Schlenker et al. 2016). I first show this problem becomes NP-hard even when solving for the defender's optimal strategy in the relaxed marginal space because of the presence of multiple types of adversaries. To provide a solution method for general-sum TSGs, I develop the GATE algorithm which is able to solve large-scale problem instances. GATE uses hierarchical adversary type trees to break a TSG down into smaller, restricted games containing a subset of the adversary types. These restricted games are then solved using an efficient branch-and-bound search tree combined with MGA; the solution information from these 'child' nodes are then passed up to the 'parent' nodes in the hierarchical tree where the information provides (i) infeasible strategy information for pruning, (ii) tighter bounds, and (iii) branching heuristics which provide faster computation at the parent nodes. I also provide heuristics based upon the properties of TSGs that increase the computational speed of the algorithm for large scale instances.

Cyber-alert Allocation Games (CAGs)

Beyond solving TSGs with general-sum payoffs, another limitation of TSGs comes from the cyber security domain. In particular, TSGs fail to model the unique time it may take a resource to screen an object and that attacks showing up as different alert types and that attacks show up as probability distributions over alert types. To remedy these limitations, I introduce the Cyber-alert Allocation Game model (Schlenker et al. 2017) which incorporates both of these features when calculating the defender's optimal strategy. I show this problem to be NP-hard due to each individual resource having a unique screening time for the different incoming objects. To solve CAGs, I develop an algorithm which leverages the "bihierarchy" constraint structure and a branch-and-bound search to quickly find *implementable* marginal strategies for the defender. Further heuristics are developed by taking advantage of several domain features that significantly increase the computational speed of the algorithm to solve large scale problems.

Future Work

Alert Allocation

While my current work has addressed three significant limitations of the TSG model, numerous more will emerge when applying TSGs and CAGs to real world domains that I plan to address in my future work. For instance, in my work I assume that the adversary has perfect knowledge of the defender's strategy. This assumption, however, fails to capture the reality in many real world domains. Specifically, in cyber defense an adversary attempting to hack a network might only receive partial observations about how often their attacks are thwarted along with the time it takes the defender to stop the attack. Another assumption of both the TSG and CAG models is that the number of incoming passengers/alerts are known *a-priori*. However, in many domains these parameters are difficult to know exactly for the defender and not being robust to this uncertainty could lead to large losses for the defender. Even further, CAGs assume the time for a cyber analyst to resolve a specific alert type, T_a^r, is known exactly. In reality, this value represents an estimate for the time it takes an analyst to resolve an alert where the analyst may take more or less time in a specific instance. In my future work, I plan to incorporate these features into CAGs to improve it's applicability in the cyber domain and beyond.

Deception

A new thrust of my research is in developing game-theoretic methods for using deception to introduce uncertainty to an adversary attacking the defender's network. The defender is able to deceive the adversary during the adversary's reconnaissance phase when he gathers information about the network by probing all accessible machines. When an adversary is probing machines on the network, the defender is able to partially control the part of the configuration that an adversary sees for each machine on the network. By hiding part of each machine's configuration, the defender can make it more difficult for the adversary to determine the best targets for them to attack. In order to take advantage of the defender's asymmetric information advantage we propose a model which captures the belief state of an adversary and uses it to lower the expected loss to the defender. At a high-level, the adversary's perception about the expected value for exploiting a machine is based upon the observable configuration they see. This perception is then taken advantage of to divert the adversary to attack less valuable machines on the network first. The problem for the defender then is to determine how best to alter the adversary's perception of the game to minimize their expected loss from a potential attack. The work on this initial model is planned to be completed before the Doctorial Consortium.

References

Brown, M.; Sinha, A.; Schlenker, A.; and Tambe, M. 2016. One size does not fit all: A game-theoretic approach for dynamically and effectively screening for threats. In *AAAI*.

Jain, M.; Kardes, E.; Kiekintveld, C.; Ordónez, F.; and Tambe, M. 2010. Security games with arbitrary schedules: A branch and price approach. In *AAAI*.

Kiekintveld, C.; Jain, M.; Tsai, J.; Pita, J.; Ordóñez, F.; and Tambe, M. 2009. Computing optimal randomized resource allocations for massive security games. AAMAS.

Schlenker, A.; Brown, M.; Sinha, A.; Tambe, M.; and Mehta, R. 2016. Get me to my gate on time: Efficiently solving general-sum bayesian threat screening games. In *ECAI*.

Schlenker, A.; Xu, H.; Guirguis, M.; Kiekintveld, C.; Sinha, A.; Tambe, M.; Sonya, S.; Balderas, D.; and Dunstatter, N. 2017. Dont bury your head in warnings: A game-theoretic approach for intelligent allocation of cyber-security alerts. In *IJCAI*.