Real-World Security Games: Toward Addressing Human Decision-Making Uncertainty

(Extended Abstract)

James Pita
University of Southern California
jpita@usc.edu

ABSTRACT

Game theory is a useful tool for reasoning about interactions between agents and in turn aiding in the decisions of those agents. In fact, Stackelberg games are natural models for many important applications such as oligopolistic markets and security domains. Indeed, Stackelberg games are at the heart of three deployed systems, ARMOR; IRIS; and GUARDS, for aiding security officials in making critical resource allocation decisions. In Stackelberg games, one player, the leader, commits to a strategy and follower makes her decision with knowledge of the leader's commitment. Existing algorithms for Stackelberg games efficiently find optimal solutions (leader strategy), however, they critically assume that the follower plays optimally. Unfortunately, in many applications, agents face human followers (adversaries) who – because of their bounded rationality and possibly limited information of the leader strategy – may deviate from their expected optimal response. Not considering these likely deviations when dealing with human adversaries may cause an unacceptable degradation in the leader's reward, particularly in security applications where these algorithms have seen deployment. To that end, I explore robust algorithms for agent interactions with human adversaries in security applications. I have developed a number of robust algorithms for a class of games known as "Security Games" and am working toward enhancing these approaches for a richer models of these games that I developed known as "Security Circumvention Games".

Categories and Subject Descriptors

I.6 [**Computing Methodologies**]: SIMULATION AND MODELING

General Terms

Algorithms, Experimentation, Security, Human Factors

Keywords

Game Theory, Security, Bounded Rationality

1. INTRODUCTION

In Stackelberg games, one player, the leader, commits to a strategy publicly before the remaining players, the followers, make their decision [2]. There are many multiagent security domains, such as attacker-defender scenarios and patrolling, where these types of commitments are necessary by the security agents [1, 3] and it has been shown that Stackelberg games appropriately model these commitments [3]. Existing algorithms for Bayesian Stackelberg games are able to find optimal solutions to these attacker-defender scenarios considering an *a priori* probability distribution over possible follower types [3]. Unfortunately, to guarantee optimality, these algorithms make strict assumptions on the underlying games, namely that the players are perfectly rational and that the followers perfectly observe the leader's strategy. However, these assumptions rarely hold in real-world domains, particularly when dealing with humans.

Of specific interest in my work are a set of real-world security domains. Two domains in particular that utilize "Security Games" [8] are the security challenges faced at the Los Angeles International Airport (LAX) and by the Federal Air Marshals Service (FAMS). Here, security forces are tasked with assigning resources to protect terminals within the airports and flights leaving the airports. While Stackelberg games have been utilized to help address these problems [3], these approaches fail to take into account a human follower (adversary). In general, human adversaries may have a variety of cognitive or environmental limitations that influence their decisions. For example, such human adversaries may be governed by their bounded rationality [7] or anchoring biases due to limited observations [6]. Thus, a human adversary may not respond with the game theoretic optimal choice, causing the leader to face uncertainty over the gamut of adversary's actions. To that end, I have designed robust algorithms to address human uncertainty within "Security Games" based on bounded rationality and limited observational capabilities.

Building upon work in security domains, I have also designed a new model of security games that allow for a more complex set of security activities for the defensive resources than previous work while not turning to a general Stackelberg representation. Such a model is designed to address the decisions faced by agencies, such as the Transportation Security Administration (TSA), in protecting airports, ports, and other critical infrastructure. In these complex environments it is important that security officials are able to reason over a set of heterogeneous security activities as opposed to the homogeneous security activities previously considered in "Security Game" models. In the future it will be important

Cite as: Real-World Security Games: Toward Addressing Human Decision-Making Uncertainty (Extended Abstract), James Pita, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1359-1360.

to extend this model to also account for human uncertainty as is done in my robust approaches to "Security Games".

2. CONTRIBUTIONS

Algorithms that address human uncertainties: My thesis provides the following key contributions. First, it provides a new robust algorithm, COBRA [4], that includes two new key ideas for addressing human adversaries: (i) human anchoring biases drawn from support theory; (ii) robust approaches for MILPs to address human imprecision. To the best of my knowledge, the effectiveness of each of these key ideas against human adversaries had not been explored in the context of Stackelberg games. Furthermore, it was unclear how effective the combination of these ideas, being brought together from different fields, would be against humans. The second contribution is in providing experimental evidence that this new algorithm can perform statistically significantly better than existing algorithms and baseline algorithms when dealing with human adversaries as followers. Since this new approach considers human adversaries, traditional proofs of correctness or optimality are insufficient; instead, it is necessary to rely on empirical validation. Hence, I examined four settings based on real deployed security systems at Los Angeles International Airport [3], and compared 6 different approaches (3 based on COBRA and 3 existing approaches), in 4 different observability conditions, involving 218 human subject playing 2960 games in total to demonstrate the value of my robust algorithm. Thirdly, my detailed experiments provide a solid initial grounding and heuristics for the right parameter settings for the α parameter within the COBRA algorithm.

Compact game representations: Beyond the contributions I have made algorithmically toward addressing human followers, I have also developed a new game model known as "Security Circumvention Games" (SCGs) [5] to address a wider range of possible security applications. Specifically, previous work has addressed domains in which a single homogeneous security activity is considered such as assigning air marshals to flights. Additionally, these security activities focused on preventing a single type of threat such as a terrorist hijacking a plane. As such, "Security Games" were developed as an efficient way to represent these games. In SCGs I am able to reason about deploying resources between heterogeneous security activities where each security activity is unique in what it accomplishes. Moreover, I consider heterogeneous attacker threats that are capable of avoiding different sets of security activities and may have different impacts if successful. The benefit of SCGs are that, while they allow for a wider class of games, they still avoid turning to a general Stackelberg representation that may have too large of an action space. By taking advantage of the game structure I am able to create both a compact representation for the defender and attacker side actions. Such a model is useful in domains where security agencies such as TSA must consider the protection of a large facility such as an airport where there may be a variety of security activities considered.

3. PRACTICAL REAL-WORLD RESULTS

In developing my work I have had the opportunity to incorporate game theoretic approaches into two real-world deployed systems. First, the Assistant for Randomized Monitoring Over Routes (ARMOR) [3] has been deployed at the Los Angeles International Airport (LAX) since August 2007 to aid security officials in assigning randomized checkpoints and canine patrols. Second, Game-theoretic Unpredictable and Randomly Deployed Security (GUARDS) [5] has been delivered to the TSA and is currently under evaluation for assigning resources to heterogeneous security activities within an airport.

While ARMOR uses the traditional "Security Game" model, GUARDS is a direct application of my new security game model "Security Circumvention Games". Given that "Security Games" were not directly applicable to this specific domain, this demonstrates the benefits of exploring more robust models within the context of security games. In general, these results demonstrate the usefulness of game theoretic approaches and show that in the future game theory can be used to aid in many important multi-agent problems.

4. FUTURE RESEARCH

In the future it will be important to continue to explore alternative approaches for addressing human uncertainty. While my current results have shown the benefit of considering different forms of uncertainty that arise from human followers there may be even better strategies for addressing this uncertainty. Furthermore, I will need to explore how my current approaches transition to new and possibly more complex models such as "Security Circumvention Games". My goal is that these approaches are generally applicable and thus will work in a wide class of potential security games. Finally, as my body of work grows and we demonstrate the value of addressing human uncertainty within security games it will be crucial to begin transitioning these techniques into the real-world applications that are already utilizing game-theoretic approaches such as ARMOR and GUARDS.

5. REFERENCES

[1] N. Agmon, V. Sadov, S. Kraus, and G. Kaminka. The impact of adversarial knowledge on adversarial planning in perimeter patrol. In *AAMAS*, 2008.

[2] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, 1991.

[3] M. Jain, J. Tsai, J. Pita, C. Kiekintveld, S. Rathi, F. Ordóñez, and M. Tambe. Software assistants for randomized patrol planning for the LAX airport police and the Federal Air Marshals Service. volume 40, pages 267–290, 2010.

[4] J. Pita, M. Jain, F. Ordóñez, M. Tambe, and S. Kraus. Robust solutions to stackelberg games: Addressing bounded rationality and limited observations in human cognition. volume 174, pages 1142–1171, 2010.

[5] J. Pita, M. Tambe, C. Kiekintveld, S. Cullen, and E. Steigerwald. GUARDS - game theoretic security allocation on a national scale. In *AAMAS*, 2011.

[6] Y. Rottenstreich and A. Tversky. Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*, 104:406–415, 1997.

[7] H. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63:129–138, 1956.

[8] Z. Yin, D. Korzhyk, C. Kiekintveld, V. Conitzer, and M. Tambe. Stackelberg vs. Nash in security games: Interchangeability, equivalence, and uniqueness. In *AAMAS*, 2010.