To Handle, to Learn and to Manipulate the Attacker's (Uncertain) Payoffs in Security Games

(Doctoral Consortium)

Yundi Qian
Supervisor: Milind Tambe
University of Southern California
Los Angeles, CA, USA
yundi.qian@usc.edu

ABSTRACT

Stackelberg security games (SSGs) are now established as a powerful tool in security domains. In order to compute the optimal strategy for the defender in SSG model, the defender needs to know the attacker's preferences over targets so that she can predict how the attacker would react under a certain defender strategy. Uncertainty over attacker preferences may cause the defender to suffer large losses.

My thesis focuses on uncertainty in attacker preferences: such uncertainty may arise because of uncertainty over attacker's risk attitude or uncertainty over true attacker payoffs. To that end, the first part of my thesis focuses on risk-averse attackers. Extensive studies show that the attackers in some domains are in fact risk-averse rather than risk-neutral, which has never been taken into account in previous security game literatures. To handle the attacker's risk aversion attitude in viewing payoffs, I develop an algorithm to compute the defender's robust strategy against an uncertain risk-averse attacker since the defenders are also uncertain about the degree of attacker's risk aversion. The second part of my thesis focuses on learning attacker payoffs. More specifically, the concept of SSGs has also been applied to the domain of protecting natural resources, where the attacker "attacks"(illegally extracts natural resources) frequently, which reveals his preference over different targets. Based on this concept, I develop the algorithm for the defender to learn target values from the attacker's actions and then use this information to better plan her strategy.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence

Keywords

Game Theory; Stackelberg Security Games; Repeated Games

1. INTRODUCTION

Stackelberg security games (SSGs) are now established as a successful tool in the infrastructure security domain [1, 2, 7]. In this domain, the security forces (defender) deploy security resources to protect key infrastructures (targets) against potential terrorists (attackers). With limited resources available, it is usually impossible to protect all targets at all times. SSGs design the optimal strategy (resources deployment) for the defender with the use of game-theoretic approaches. In SSG model, the defender acts first and commits to a mixed strategy while the attacker learns the mixed strategy after long-time surveillance and then chooses one target to attack. The computation of the optimal strategy for the defender in SSGs requires the defender to know how the attacker views the importance of every target since it involves predicting the attacker's action under a certain mixed strategy. If the defender is unable to predict the attacker's action correctly, she may suffer significant losses.

Previous work assumes that the attacker is risk-neutral so that he would choose to attack the target with the highest expected reward for him. However, extensive studies show that the attackers in some domains are in fact risk-averse, e.g., terrorist groups in counter-terrorism domains [3, 4]. If the defender fails to take this into consideration when designing strategies, she may suffer significant losses. In addition, the defender in uncertain about the attacker's degree of risk aversion. Knowing that, the first part of my thesis takes this risk-aversion into account and develops the model and algorithm to compute the robust strategy for the defender against an uncertain risk-averse attacker [6].

Except the infrastructure security domains, the concept of SSGs has also been applied to the domain of protecting our environment and natural resources, where "defenders" (law enforcement agencies) try to protect these natural resources and "attackers" (criminals) seek to exploit them. In security games, the attacker conducts extensive surveillance on the defender and executes a one-shot attack, while in resource conservation domains, the attacker also observes the defender's strategy but carries out frequent illegal extractions. Therefore, there are frequent interactions between the defender and the attacker, which gives the defender the opportunity to learn the attacker's preference over different targets by observing the attacker's actions. Based on this concept, the second part of my thesis develops the model and algorithm for the defender to learn target values from the attacker's actions and then uses this information to better plan her strategy [5].

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

2. CONTRIBUTIONS

My thesis provides the following key contributions.

Robust strategy against risk-averse attackers in security games In the first contribution of my thesis [6], I consider the scenario where the attacker is risk-averse (including risk-neutral), and the defender is uncertain about the degree of the attacker's risk aversion. To address this issue, I develop the model and algorithm to find the robust strategy for the defender against uncertain risk-averse attackers in security games, which provides a solution quality guarantee no matter how risk-averse the attacker is. Based on previous security game research and optimization techniques, I construct a mixed integer bilinear programming problem (MIBLP) to solve the problem, which unfortunately only finds locally optimal solution and is unable to scale up. However, it provides key intuitions for our new algorithm BeRRA. The scalable BeRRA algorithm finds globally ϵ-optimal solutions by solving $\mathcal{O}(n\log^2(\frac{1}{\epsilon}))$ linear feasibility problems. The key idea of our BeRRA algorithm is to reduce the problem from maximizing the reward with a given number of resources to minimizing the number of resources needed to achieve a given reward. This transformation allows BeRRA to scale up via the removal of the bilinear terms and integer variables as well as the utilization of key theoretical properties that prove correspondence of its potential "attack sets" [1] with that of the maximin strategy. Although the BeRRA algorithm is designed for risk-averse attackers, it can also be extended to handle other risk-aware attackers, e.g., risk-seeking attackers. In addition, I find that the defender does not need to consider attacker's risk attitude if it is a zero-sum game.

Online Planning for Optimal Protector Strategies in Resource Conservation Games In the second contribution of my thesis [5], I consider the domain of protecting natural resources, e.g., protecting fish, protecting forest. Different from the infrastructure security domain, the attacker "attacks"(illegally extracts natural resources) frequently in the resource conservation domain, which gives the defender the opportunity to learn how the attacker values every target from their actions. I model the frequent interactions between the defender and the attacker as repeated games, and then recast this repeated game as a partially observable Markov decision process (POMDP). However, this POMDP formulation has an exponential number of states, making the latest general POMDP solvers infeasible in terms of computational cost. In response, I propose the GMOP algorithm, which is a dedicated algorithm that combines Gibbs sampling with Monte Carlo tree search (MCTS) for online planning in this POMDP. It uses Gibbs sampling to sample the current belief state, and then uses MCTS for online planning. Additionally, for a specific class of the game with an attacker who plays a best response against the defender's empirical distribution, and a uniform penalty of being seized across all targets, I provide an advanced sampling technique to speed up the GMOP algorithm without sacrificing solution quality, as well as provide a heuristic that trades off solution quality for lower computational cost using inference in Bayesian network.

3. FUTURE WORK

I may work on the following directions in the future.

Induce the attacker by properly setting attacker's penalties Knowing the uncertainty in how the attacker views his payoffs, my first contribution handles this uncertainty by designing robust strategies against that. My second contribution learns the attacker's preferences from their actions. In the future, I plan to work on "manipulating" the attacker's payoffs (penalties actually) to induce the attacker. In some scenario, the attacker's penalties are determined by the defender. For example, in the domain of protecting fish, some areas may be more important so that the penalties for illegal fishing in those areas may be higher. Clearly, the defender can improve her expected utility if he is able to set the penalties properly. In this sense, the defender is manipulating the attacker's actions by manipulating the penalties.

exploration vs exploitation tradeoff in resource conservation games In real-world applications, the defender does not know if the attacker has attacked a target if she does not go there. Hence there is a exploration vs exploitation tradeoff here: if the defender always goes to the targets that she believes has a lot of attackers, there might be some other targets the defender does not know that has more attackers, and the attackers at targets that are frequently visited by the defender may deviate to other targets; if the defender keeps exploiting different targets, her strategy would be less efficient. Techniques in multi-armed bandit problem and machine learning might be used to balance this exploration vs exploitation tradeoff.

4. ACKNOWLEDGMENTS

This research is supported by the United States Department of Homeland Security through the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) under award number 2010-ST-061-RE0001 and MURI grant W911NF-11-1-0332.

REFERENCES

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

[2] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordonez, and S. Kraus. Playing games with security: An efficient exact algorithm for bayesian stackelberg games. In *AAMAS*, 2008.

[3] P. Phillips. The preferred risk habitat of al-qa'ida terrorists. *European Journal of Economics, Finance and Administrative Sciences*, 2010.

[4] P. J. Phillips. The end of al-qa'ida: rationality, survivability and risk aversion. *International Journal of Economic Sciences*, 2013.

[5] Y. Qian, W. B. Haskell, A. X. Jiang, and M. Tambe. Online planning for optimal protector strategies in resource conservation games. In *AAMAS*, 2014.

[6] Y. Qian, W. B. Haskell, and M. Tambe. Robust strategy against unknown risk-averse attackers in security games. In *AAMAS*, 2015.

[7] Z. Yin and M. Tambe. A unified method for handling discrete and continuous uncertainty in bayesian stackelberg games. In *AAMAS*, 2012.