Defender Strategies In Domains Involving Frequent Adversary Interaction

(Extended Abstract)

Fei Fang
University of Southern California
Los Angeles, United States
feifang@usc.edu

Peter Stone
University of Texas at Austin
Austin, United States
pstone@cs.utexas.edu

Milind Tambe
University of Southern California
Los Angeles, United States
tambe@usc.edu

ABSTRACT

Recently, there has been an increase in interest in applying game theoretic approaches to domains involving frequent adversary interactions, such as wildlife and fishery protection. In these domains, the law enforcement agency faces adversaries who repeatedly and frequently carry out illegal activities, and thus, do not have time for extensive surveillance before taking actions. This makes them significantly different from counter-terrorism domains where game-theoretic approaches have been widely deployed. This paper presents a game-theoretic approach to be used by the defender in these Frequent Adversary Interaction (FAI) domains. We provide (i) a novel game model for FAI domains, describing the interaction between the defender and the attackers in a repeated game and (ii) algorithms that plan for the defender strategies to achieve high average expected utility over all rounds.

Categories and Subject Descriptors

I.2.11 [**ARTIFICIAL INTELLIGENCE**]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*

General Terms

Security, Algorithms

Keywords

Frequent Adversary Interaction, Game Theory, Repeated Games, Defender Strategy

1. INTRODUCTION

Whereas game theoretic approaches have been widely deployed in various counter-terrorism settings including protecting airports and ports [10, 15], recently there has been increasing interest in applying game theory to suppressing environmental crimes such as in protecting fisheries from over-fishing [12, 7] and protecting rhinos and tigers from illegal poaching [17]. Unfortunately, most previous work in counter-terrorism domains cannot be directly applied to these domains because they have three key differences. Firstly, in counter-terrorism domains, an attacker is assumed to conduct extensive surveillance in order to understand the defender's strategy and then executes a one-shot attack. However, in domains involving environmental crime, the law enforcement agency (the defender) is faced with multiple adversaries (the attackers) who carry out repeated and frequent illegal activities (attacks) and generally do not conduct extensive surveillance before performing an attack. Secondly, in carrying out such frequent attacks, the attackers generally spend less time and effort in each attack, and thus it becomes more important to model the attacker's bounded rationality and bounded surveillance. Thirdly, there is more attack data available — at least in comparison with the earlier counter-terrorism domains — that can be collected by the defender in such domains. We propose the term *Frequent Attacker Interaction (FAI) domains* to refer to such domains.

There are some recent efforts that have begun to address FAI domains [17, 7]. They model the problem as a repeated game and each round is a Stackelberg security game where the defender commits to a mixed strategy and the attackers respond to it; they do address the bounded rationality of attackers using the SUQR model [9]. While such advances have allowed these works to be tested in the field, whether to protect wildlife or fisheries, there are several key weaknesses in these efforts. First, the Stackelberg assumption in these works – that the defender's mixed strategy is fully observable by the attacker via extensive surveillance before each attack – is unrealistic in the context of FAI domains as mentioned above. Indeed, the attacker may experience a delay in observing how the defender strategy may be changing over time, from round to round. Second, since the attacker may lag in observing the defender's strategy, it may be valuable for the defender to plan ahead; however these previous efforts do not engage in any planning and instead rely only on designing strategies for the current round.

In this paper, we offer remedies for these limitations. First, we introduce a novel repeated game model called LEAD (LEad Attackers with Delayed observation). Generalizing the perfect Stackelberg assumption, LEAD assumes that the attackers' understanding of the defender strategy may not be up-to-date and can be inferred from the defender strategies used in recent rounds. Second, we provide algorithms that plan ahead, providing defender strategies in each round. The generalization of the Stackelberg assumption introduces a need to plan ahead and take into account the effect of defender strategy on future attacker decisions.

2. MOTIVATING EXAMPLE

In FAI domains such as protecting against wildlife poaching, the defender organizes a group of patrollers to protect a large area. The area can be divided into subareas or targets, each of different im-

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.*

portance to the defender. A population of attackers perform frequent attacks, each confined to one subarea. The defender is a set of patrollers (or rangers) who use attack data to refine their patrols. Example 1 conceptually shows that the defender can benefit from strategy change, and the focus of this paper is to design the sequence of defender strategies. For the simplicity of the example, we assume perfectly rational attackers.

Example 1. *Consider a patroller who is in charge of protecting fisheries from overfishing in an area. The area is divided into two subareas N_1 and N_2, which are of the same importance both to the defender and to illegal fishermen who place illegal fishing nets. The patroller chooses a subarea to patrol every day and she can stop any snaring in the patrolled area. The defender has been using the uniform random strategy throughout last year and is going to decide the strategy to be used this January. She can choose to continue using the uniform strategy throughout January, catching 50% of the fishing nets. However, if she always protects N_1 at the beginning of January, and then switches to always protecting N_2 in mid-January, she can catch 75% of the fishing nets as explained below. Presumably, the illegal fishermen will have no preference between the two subareas at the beginning of January due to their observation from last year. Thus, 50% of the fishing nets will be placed in N_1 and the patroller can catch these fishing nets by only protecting N_1. The illegal fishermen may realize the strategy change after a period of time (e.g., two weeks) and will then put all the fishing nets in N_2. The illegal fishermen's behavior change is expected by the defender and the patroller can catch 100% of the fishing nets by only protecting N_2 starting from mid-January.*

3. RELATED WORK

Extensive studies state and model the bounded rationality and bounded memory of human beings over the years [13, 4]. Stone et. al [16] studied the optimal strategy to lead a teammate with bounded memory given finite action set. In this paper, we are dealing with players who have conflicting interests and we consider more general cases in which the players can choose from an infinite strategy set. Bounded memory in repeated games has also been studied in some earlier work [14, 1, 5] and learning against opponents with bounded-memory is considered in some previous work [11, 3]. This paper differs from these previous work in that it aims to find *a mixed defender strategy* in every round of a repeated game and it further considers attackers' bounded rationality.

Previous work on learning in repeated Stackelberg security games [8, 2] mainly focus on learning the payoffs of the perfectly rational attackers. Qian et al. [12] model the interaction between protector and extractor in the resource conservation domains as a Partially Observable Markov Decision Process (POMDP) to learn the utility of the targets from the extractor's actions. In our problem, the payoffs are known to both players and the defender aims to maximize overall expected utility in a LEAD game. Sequential decision-making in the presence of other players is also studied in the context of computer poker [18, 6] and most work focuses on zero-sum games with a single action in every round.

Acknowledgement

This research was supported by MURI Grant W911NF-11-1-0332 and by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) under grant number 2010-ST-061-RE0001. A portion of this work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CNS-1330072, CNS-1305287), ONR (21C184-01), and AFOSR (FA8750-14-1-0070, FA9550-14-1-0087).

REFERENCES

[1] M. Barlo, G. Carmona, and H. Sabourian. Repeated games with one-memory. *Journal of Economic Theory*, 144(1):312 – 336, 2009.

[2] A. Blum, N. Haghtalab, and A. D. Procaccia. Learning optimal commitment to overcome insecurity. In *NIPS*, 2014.

[3] D. Chakraborty, N. Agmon, and P. Stone. Targeted opponent modeling of memory-bounded agents. In *Proceedings of the Adaptive Learning Agents Workshop (ALA)*, 2013.

[4] N. Cowan. *Working Memory Capacity*. Essays in cognitive psychology. Psychology Press, 2005.

[5] E. M. de Cote and N. R. Jennings. Planning against fictitious players in repeated normal form games. In *AAMAS*, pages 1073–1080, 2010.

[6] A. Gilpin and T. Sandholm. A texas hold'em poker player based on automated abstraction and real-time equilibrium computation. In *AAMAS*, pages 1453–1454, 2006.

[7] W. B. Haskell, D. Kar, F. Fang, M. Tambe, S. Cheung, and L. E. Denicola. Robust protection of fsheries with COmPASS. In *IAAI*, 2014.

[8] J. Marecki, G. Tesauro, and R. Segal. Playing repeated stackelberg games with unknown opponents. In *AAMAS*, pages 821–828, 2012.

[9] T. H. Nguyen, R. Yang, A. Azaria, S. Kraus, and M. Tambe. Analyzing the effectiveness of adversary modeling in security games. In *AAAI*, 2013.

[10] J. Pita, M. Jain, C. Western, C. Portway, M. Tambe, F. Ordonez, S. Kraus, and P. Paruchuri. Deployed ARMOR protection: The application of a game theroetic model for security at the los angeles international airport. In *AAMAS*, 2008.

[11] R. Powers and Y. Shoham. Learning against opponents with bounded memory. In *IJCAI*, pages 817–822, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.

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

[13] A. Rubinstein. *Modeling Bounded Rationality*, volume 1 of *MIT Press Books*. The MIT Press, December 1997.

[14] H. Sabourian. Repeated games with m-period bounded memory (pure strategies). *Journal of Mathematical Economics*, 30(1):1 – 35, 1998.

[15] E. Shieh, B. An, R. Yang, M. Tambe, C. Baldwin, J. DiRenzo, B. Maule, and G. Meyer. PROTECT: A deployed game theoretic system to protect the ports of the United States. In *AAMAS*, 2012.

[16] P. Stone, G. A. Kaminka, S. Kraus, J. R. Rosenschein, and N. Agmon. Teaching and leading an ad hoc teammate: Collaboration without pre-coordination. *Artificial Intelligence*, 2013.

[17] R. Yang, B. Ford, M. Tambe, and A. Lemieux. Adaptive resource allocation for wildlife protection against illegal poachers. In *AAMAS*, 2014.

[18] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret minimization in games with incomplete information. In *NIPS*, 2008.