SPECTRE: A Game Theoretic Framework for Preventing Collusion in Security Games (Demonstration)

Shahrzad Gholami, Bryan Wilder, Matthew Brown, Arunesh Sinha, Nicole Sintov, Milind Tambe
University of Southern California, USA,
{sgholami,bwilder,matthew.a.brown,aruneshs,sintov,tambe}@usc.edu

ABSTRACT

Several models have been proposed for Stackelberg security games (SSGs) and protection against perfectly rational and bounded rational adversaries; however, none of these existing models addressed the destructive cooperation mechanism between adversaries. SPECTRE (**S**trategic **P**atrol planner to **E**xtinguish **C**ollusive **Th**R**E**ats) takes into account the synergistic destructive collusion among two groups of adversaries in security games. This framework is designed for the purpose of efficient patrol scheduling for security agents in security games in presence of collusion and is mainly build up on game theoretic approaches, optimization techniques, machine learning methods and theories for human decision making under risk. The major advantage of SPECTRE is involving real world data from human subject experiments with participants on Amazon Mechanical Turk (AMT).

Categories and Subject Descriptors

H.4 [**Security and Multi-agent Systems**]:

Keywords

Game Theory, Stackelberg Security Games, Human Behavior Models, Cooperation Mechanism, Collusion

1. INTRODUCTION

Security agencies including the US Coast Guard (USCG), the Federal Air Marshal Service (FAMS) and the Los Angeles Airport (LAX) police are several major domains that have been deploying Stackelberg security games (SSGs) and related algorithms to protect against adversaries strategically [3]. The security games introduced in these domains, mostly, include two players: a defender and an adversary. The interaction between the defender and the attacker was modeled as a single-shot game and the attacker was defined as a perfectly rational player. This well known class of SSGs is sequential, i.e. one player (the leader or the defender) commits to a strategy which can be observed by the other player (the follower or adversary) before choosing his own strategy. On the other hand, SPECTRE deals with the sequential single-shot security games but in which the defender commits to a strategy and the two attackers can either attack individually or cooperatively in terms of sharing the pay-offs. One of the applications of this type of game is in wildlife protection domain which is an active area of research [1], [2].

According to the references, international illegal animal trafficking is increasing incredibly and based on the estimations, it is worth at least $5 billion, annually. The main types of wildlife commodities that are subject to these illegal trades include elephant ivory, rhino horn, tiger parts and caviar, to name a few. These activities have the potential to introduce several threats to the national security and evironment arround the world. Biodiversity loss, potential extinctions, introduction of invasive species and desease transmission into healthy ecosystems, all can impact the environment adveresly. In addition to that, some connections have been observed among wildlife trafficking, organized crime and drug trafficking which means that poor law enforcement, poor patrol scheduling or corrupt rangers at wildlife sources, corrupt governments at transit countries and porous borders can all threaten the national security [4]. Despite the evidence of this illegal exchange between different groups of criminals, the destructive synergistic effect of cooperation among adversaries is unexplored in related literature in security game domain.

To combat this illegal wildlife trade, exploitation and collaboration among criminals and adversaries, SPECTRE is designed, employed within a simulation game in wildlife domain and deployed on Amazon Mechanical Turk (AMT) to demonstrate the concept of collusion in security games. It is worth noting that each of the two adversaries in this type of game can be a representative for either a poacher who is directly hunting in the field or a trader who is illegally exchanging the animals or financing other illegal commodities via animal trafficking.

2. HOW TO EXTINGUISH COLLUSION?

SPECTRE system has two main compartments: i) human subject experiment interface (HSI) and ii) the core software for learning the governing model and designing patrol plans. (Demo Link: `https://youtu.be/zRzg2g4PxWA`)

2.1 Human Subject Interface

To see how human adversaries make decisions about cooperating with each other, SPECTRE Human Subject Interface (HSI) was designed in wildlife protection domain. In this software real human subjects are asked to play a game

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), John Thangarajah, Karl Tuyls, Stacy Marsella, Catholijn Jonker (eds.), May 9–13, 2016, Singapore.*

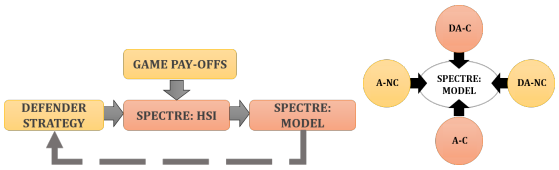
in which two players are involved. In this game, they have the role of a poacher in a national park in Africa. There are different number of hippopotamus distributed over the park which indicates animal density distribution over the area and are considered to visually represent the distribution of wildlife to the players. The entire park area is divided into two sections (right and left) and each human subject can only attack in one section (either right or left); however, they can explore the whole park to find out about reward, penalty and coverage at each sub-region. The other section of the park is only available to another player who is playing the same game. Two players have access to each others' information to make wise decisions about cooperation with each other. Each section of the park is divided into 3×3 grid, i.e. each player has 9 cells (sub-regions) accessible to him to attack. Players are able to choose different sub-regions and all of the information about success and failure likelihood, reward for the attacker (which is animal density in each sub-region) and penalty at each sub-region (either on left or side of park) will be shown to them. To help the human subjects to have a better view of the success/failure percentage (which is defender coverage) over all the sub-regions, we put a heat-map of that overlaid on Google Map view of the park. Also, to help the players to have a better understanding of the cooperation mechanism in this game, we provided a table that summarizes all possible pay-offs for cooperative attacks based on the cooperation bonus considered for each game. The human subjects need to make

Figure 1: Hunters vs Rangers game interface

decisions about: i)whether they are inclined to cooperate with the other player or not and ii)which region of the park to put their snare (trap) where there is less chance of getting caught and also a high chance of capturing a hippopotamus. So the human subjects may decide to attack "individually and independently" or attack "cooperatively" with the other player. In both situations, they will attack different sections separately but if both of them agree to attack cooperatively, they will share all of their pay-offs with each other, equally (fifty-fifty). To enhance understanding of the game, participants were asked to play one trial game to become familiar with the game interface and procedures. Then we provided a validation game to make sure that the players have read the instructions of the game and are fully aware of the rules and options of the game. Finally, the third game which is the main game is shown to the human subjects and their decisions are recorded. Then we analyze the human subject decisions to derive a more accurate model to describe the human adversary behavior in security games in presence of cooperation mechanism.

2.2 SPECTRE: A Patrol scheduler software

To study how colluding adversaries behave, different pay-off structures and defender strategies can be applied to the software. The main output of this software is human decisions about cooperation and the targets they have attacked either in cooperation or individual attack. Figure 2(a) illustrates how SPECTRE functions in human subject experiments analysis.



(a) General diagram (b) SPECTRE Components

Figure 2: SPECTRE framework for human adversary analysis

For the initial experiments, SPECTRE HSI starts with rational adversary assumption and generates defender strategies on that basis. After collecting data based on this initial input, new strategies based on bounded rational adversaries are generated and re-applied to the software. The main idea for breaking the cooperation is to put one adversary in a better condition in terms of defender coverage and the other one in a worse condition, then cooperation will not be preferred by one of adversaries and cooperation breaks. So for human behavior analysis, SPECTRE considers four groups of adversaries shown in 2(b): i) a disadvantaged attacker who is inclined to cooperate, DA-C , ii) a disadvantaged attacker who is not inclined to cooperate, DA-NC, iii) an advantaged attacker who is inclined to cooperate, A-C, and iv) an advantaged attacker who is not inclined to cooperate, A-NC. Based on the decisions made by these four groups and human behavior models under risk such as Prospect Theory and SUQR, SPECTRE predicts adversaries behavior and generate an optimal defender strategy which breaks the collusion and maximize the defender utility.

Acknowledgments

This research was supported by MURI Grant W911NF-11-1-0332.

REFERENCES

[1] F. Fang, T. H. Nguyen, R. Pickles, W. Y. Lam, G. R. Clements, B. An, A. Singh, M. Tambe, and A. Lemieux. Deploying paws: Field optimization of the protection assistant for wildlife security. 2016.

[2] F. Fang, P. Stone, and M. Tambe. When security games go green: Designing defender strategies to prevent poaching and illegal fishing. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[3] M. Tambe. *Security and game theory: Algorithms, deployed systems, lessons learned*. Cambridge University Press, 2011.

[4] L. S. Wyler and P. A. Sheikh. International illegal trade in wildlife: Threats and us policy. DTIC Document, 2008.