Predicting and Planning against Real-world Adversaries:
An End-to-end Pipeline to Combat Illegal Wildlife Poachers on a Global Scale

by

Shahrzad Gholami

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(Computer Science)

Spring 2019

Acknowledgments

I would like to take this opportunity to thank my advisor, Prof. Milind Tambe, for his incredible mentorship and guidance through the challenging journey of the Ph.D. program. Milind, joining the Teamcore research lab and the USC Center for Artificial Intelligence in Society was a remarkable experience for me. Besides all of your encouragement and support for scientific achievements, your direction also helped make this experience an incredible one for me in various other ways. I was fortunate to have the chance to work on several real-world problems, which received attention from decision-makers and stakeholders. I believe that it is the greatest reward for a researcher if their work and efforts can directly help solve real-world problems and positively influence the tough situations. I was lucky that I was given the chance to work with so many people with a variety of background that I could never imagine. I traveled a lot and I learned that hearing about other people's ideas and concerns can help one grow out of their own boundaries. Also, being in a large team with a diversity of nationalities led to a lot of precious friendship for me while I was away from my family and my home town. It was marvelous to get to know so many amazing people from all around the world.

Next, I would like to thank my mentor, Prof. Bistra Dilkina who supported me generously for several of my research papers. Our frequent interactions and your invaluable feedback and insights enriched my research accomplishments. I am very grateful for all of your guidance. Also,

I would like to thank Dr. Andrew Plumptre, a conservation biology scientist, without whom this research would have not been possible. Your insights and feedback to do this research were precious to me and I am very grateful for your help with the deployment of my research outcomes. I am also very thankful for my great collaborators who were also my coauthors in the research publications: Fei Fang, Nicole Sintov, Long Tran-Thanh, Bryan Wilder, Matthew Brown, Debarun Kar, Benjamin Ford, Amulya Yadav, Sara Mc Carthy, Thanh H. Nguyen, Arunesh Sinha, and Lily Xu. It was undoubtedly wonderful to have you beside me.

In the last five years, I had the chance to meet and work with so many great people at the USC Center for Artificial Intelligence in Society: Haifeng Xu, Aaron Schlenker, Leandro Marcolino, Yasaman Abbasi, Aida Rahmattalabi, Elizabeth Bondi, Elizabeth Orrico, Omkar Thakoor, Han Ching Ou, Kai Wang, Sarah Cooney, Donnabell Dmello, Laksh Matai, Amit Plaha, Dana Thomas, Lakshmi Chirumamilla, Samarth Kulkarni, Andrew Perrault, Jackson Killian, Eric Ewing, Aditya Mate, and Elizabeth Ondula, Caleb Robinson, Amrita Gupta, Aaron Ferber, Hailey Winetrobe, and Tye Hines. It was absolutely fabulous to get to know each and every one of you.

Finally, I would like to thank my parents who have perpetually provided love and support to me. They encouraged me to be myself, believe in my decisions and follow my heart in every single step of my life. I have always been blessed to have you with me and I cannot imagine how much you sacrificed your life for me. Also, I would like to thank my very own younger sister, Yasi, who brightened up our lives while we used to be a single-child family for 9 years! I am very grateful to have such a compassionate and caring person like you in my life. It has definitely not been easy to be 8000 miles away from home and my beloved family. During all these years, I have been very fortunate to share most of my life in the US with my dearest Iman, the rock of my life. Iman, thanks for supporting me unconditionally through all these years. You always encouraged me to look forward and be faithful and confident in my path. You always make me smile and you are very precious to me.

Table of Contents

List Of Figures

List Of Tables

Abstract

Security is a global concern and a unifying theme in various security projects is strategic reasoning where the mathematical framework of machine learning and game theory can be integrated and applied. For example, in the environmental sustainability domain, the problem of protecting endangered wildlife from attacks (i.e., poachers' strikes) can be abstracted as a game between defender(s) and attacker(s). Applying previous research on security games to sustainability domains (denoted as Green Security Games) introduce several novel challenges that I address in my thesis to create computationally feasible and accurate algorithms in order to model complex adversarial behavior based on real-world data and to generate optimal defender strategy.

My thesis provides four main contributions to the emerging body of research in using machine learning and game theory framework for the fundamental challenges existing in the environmental sustainability domain, namely (i) novel spatio-temporal and uncertainty-aware machine learning models for complex adversarial behavior based on the imperfect real-world data, (ii) the first large-scale field test evaluation of the machine learning models in the adversarial settings concerning the environmental sustainability, (iii) a novel multi-expert online learning model for constrained patrol planning, and (iv) the first game theoretical model to generate optimal defender strategy against collusive adversaries.

In regard to the first contribution, I developed bounded rationality models for adversaries based on the real-world data that account for the naturally occurring uncertainty in past attack evidence collected by defenders. To that end, I proposed two novel predictive behavioral models, which I improved progressively. The second major contribution of my thesis is a large-scale field test evaluation of the proposed adversarial behavior model beyond the laboratory. Particularly, my thesis is motivated by the challenges in wildlife poaching, where I directed the defenders (i.e., rangers) to the hotspots of adversaries that they would have missed. During these experiments across multiple vast national parks, several snares and snared animals were detected, and poachers were arrested, potentially more wildlife saved. The algorithm I proposed, that combines machine learning and game-theoretic patrol planning is planned to be deployed at 600 national parks around the world in the near future to combat illegal poaching.

The third contribution in my thesis introduces a novel multi-expert online learning model for constrained and randomized patrol planning, which benefits from several expert planners where insufficient or imperfect historical records of past attacks are available to learn adversarial behavior. The final contribution of my thesis is developing an optimal solution against collusive adversaries in security games assuming both rational and boundedly rational adversaries. I conducted human subject experiments on Amazon Mechanical Turk involving 700 human subjects using a web-based game that simulates collusive security games.

Chapter 1

Introduction

Security remains a global concern, whether it is the challenge of protecting airports, ports and other critical infrastructure; protecting wildlife, forests and fisheries; suppressing urban crimes; or preventing cyber intrusions. A unifying theme in various security projects is strategic reasoning where the mathematical framework of machine learning and game theory can be integrated and applied. In general, solutions to security problems require us to develop a data-to-decision pipeline which consists of two critical elements: (i) a learned model of human behavior based on the real-world data, and (ii) strategic reasoning about such behavior to develop decision solutions. Green security games were introduced to address the challenges specific to wildlife protection (and in general environmental sustainability problems). This line of research is motivated by the endangered Wildlife crisis in Africa where animals such as elephants and rhinos are threatened by extreme poaching and habitat loss [59,65]. Even though 84% of elephants currently reside in protected areas, they are still observed to have an extremely high rate of mortality [18] which serves to highlight the great need to take intelligent action towards thwarting poachers and reversing the downward trend in biodiversity loss.

1.1 Problem Addressed

In the sustainability domain, the problem of protecting endangered wildlife from attacks (i.e., poachers' strikes) can be abstracted as a game between defender(s) and attacker(s) where the complexity of the attacker's behavior can be modeled and represented by machine learning models built based on historical attack records. To develop solutions to such problems, I have explored two significant research areas in my thesis:

- **Machine learning and human behavior modeling with real-world data**: Understanding and modeling the behavior of the human players is fundamental to strategy design in games. Also, the evaluation of the models, prior to long-term deployments of game theoretical strategies is critical. Two significant research problems are (i) how to capitalize on the availability of imperfect data to learn behavior models of humans? (ii) how do the behavior models learned based on imperfect historical data perform in real-world settings?

- **Game theoretical frameworks for strategy design in security settings**: While the learned machine learning models for human behavior in games are beneficial, they might suffer from inaccuracies due to imperfect data. A significant research problem is how to develop efficient algorithms to compute optimal security solutions in the presence of an imperfect human behavior model. Furthermore, in many complex domains where several players are involved, adversaries collude in their attacks and impose more damage to the defenders. Thus it is important to develop game-theoretic defending solutions against such collusive and collaborative adversaries.

(a) Large-scale evaluation across multiple national parks in Africa.

(b) Park rangers detecting well-hidden snares in Murchison Fall park via the field test: Photo taken by Uganda Wildlife Authority.

(c) An elephant spotted with a snare caught on its foot: Photo taken by Singita.

Figure 1.1: The example domain motivating my research. My algorithms have been deployed in the wildlife protection domain.

1.2 Main Contributions

My thesis is focused on providing computationally feasible and accurate approaches to address the challenges in these areas for problems with adversarial interactions. The models and algorithms developed in my thesis advance the state of the art to a new generation of security games where adversarial behavior is presented by complex machine learning models, which are aware of the uncertainty in past attack data. I conducted a large-scale field test evaluation of adversarial behavior model in the environmental sustainability domain, in particular, wildlife poaching, where I directed the defenders (i.e., rangers) to the hotspots of adversaries that they would have missed. During these experiments across multiple vast national parks, several snares and snared animals were detected, and poachers were arrested, potentially more wildlife saved (Figure 1.1). The data-to-decision solution proposed in this thesis (Figure 1.2), that combines machine learning and game-theoretic patrol planning [27] is planned to be deployed at 600 national parks around the world in the near future to combat poaching.

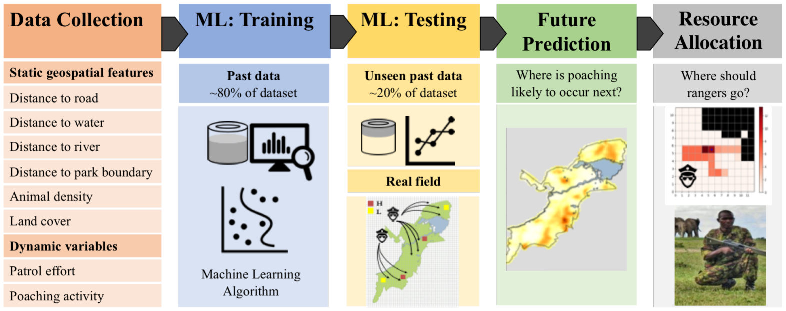


Figure 1.2: The proposed data-to-decision pipeline to combat illegal wildlife poachers.

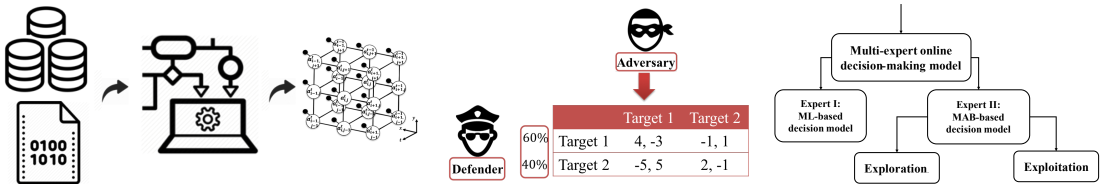
1.2.1 Machine learning and adversarial human behavior modeling with real-world data

Game theoretical frameworks have been extensively used for optimal resource allocation and scheduling problems [1, 6, 25, 87]. More specifically, Stackelberg security game (denoted as SSG) models have received significant attention for security resource allocation problems where two players are involved [17,46,75,81,89]. The underlying problem in all of the domains with security challenges is how to make the best use of the limited security resources. This problem can be abstracted as how to design the defender's strategy in a security game, where the defender tries to protect a set of targets (e.g., terminals at an airport, regions in national parks, etc.) from potential attacks. Most of the past work assumes a perfectly rational attacker while designing defender strategies, which is not true for domains like wildlife protection, where poachers are boundedly rational [23].

Green Security Games were introduced to address the challenges specific to wildlife protection (and in general environmental sustainability problems), such as boundedly rational adversaries. While there has been work on learning these adversary models, this has been mostly done based on simulated games, where data is collected from human subject experiments in the laboratory [42,72,99] rather than from real-world poachers. These methods are additionally unable to scale to real-world setups, which typically have an enormous number of targets (e.g., 3900 targets of 1x1 sq. km in Murchison Fall park) and diverse geo-spatial characteristics which drive human adversaries in the real world.

The first major contribution of my thesis is to develop bounded rationality models for adversaries based on the real-world data, which takes into account the major challenges of adversarial behavior modeling in Green Security Games [44, 71] and can be integrated into the game theoretical decision models to generate optimal strategies for the defenders. In particular, I have proposed bounded rationality models which account for uncertainty in past attack evidence collected by defenders and spatio-temporal dimensions in illegal activities. I proposed two novel predictive behavioral models, a hybrid spatio-temporal model, and an imperfect-observation-aWare Ensemble (iWare-E) method, which I improved progressively [26, 27]. More particularly, the latest iWare-E model I proposed considers the major challenge of adversarial behavior modeling in the wildlife protection domain, which is the imbalanced non-uniform uncertainty on the evidence of attacks collected by the defenders. This approach improves the accuracy and runtime of the algorithm compared to the state-of-the-art by using multiple fast running weak-learners involved in a structured ensemble model compatible with the data collection scheme in Green Security Games (Figure 1.3(a)).

To evaluate the predictive power of the adversarial models proposed in my thesis, I conducted a large-scale field test based on the machine learning techniques beyond the laboratory, in the real world, across multiple national parks for more than a year. The adversarial model that I developed was used to detect the most high-risk regions within infrequently patrolled areas in two national parks in Uganda (Queen Elizabeth and Murchison Fall National Parks, covering 5000 sq. km and 2500 sq. km, respectively). When park rangers visited the places my model recommended, they were able to find many active and inactive snares which imply that animals were saved before being killed by the well-hidden snares (Figure 1.1(a) –1.1(b)).


0100
1010
Adversary
Defender
Multi-expert online decision-making model
Expert I: ML-based decision model
Expert II: MAB-based decision model
Exploration
Exploitation

		Target 1	Target 2
60%	Target 1	4, -3	-1, 1
40%	Target 2	-5, 5	2, -1

(a) Behavior Modeling via Machine Learning and Data Science.

(b) Game Theoretical Reasoning and Randomization for Decision Making.

(c) Exploration vs. Exploitation for fine-tuned Decision Making.

Figure 1.3: The research areas in artificial intelligence that my thesis addresses.

1.2.2 Novel challenges for defender strategy design in Green Security Games

Previous work in the field of Green Security Games has led to the development of several algorithms which serve as game-theoretic decision aids to optimize the use of limited human patrol resources to combat poaching [42, 72]. The basic premise behind most of this work is that repeated interactions between patrollers and poachers provide the opportunity to gather data which can be used to learn models of poacher behavior [22]. Thus, most previous algorithms design patrol routes assuming poachers attack according to a fixed ”*learnable*” model (which could either have a functional form [22, 72], or it could be a black-box model [27, 95]). Most of these algorithms then try to solve a repeated Stackelberg game, where the patrollers (defenders) conduct randomized patrols against poachers (attackers) while balancing the priorities of different locations in the park. Unfortunately, this approach suffers from serious shortcomings, which impedes usability in the real world.

The Green Security Game approach can be expected to provide good results only if the collected historical data is a good representation of the actual poaching activities that occurred in the past (and those that will occur in the future), which would allow us to learn an accurate adversarial model for attacker behavior. Unfortunately, in the Green Security Games domain, while we can conduct a regional evaluation of the predictive models for patrolled areas, it is extremely difficult to know ahead of time whether the learned model of attacker behavior is accurate or not over the entire protected area. Due to logistical issues, many defender resources (i.e., rangers) only conduct patrols either close to their sparsely spread patrol posts, or in areas that are easily accessible to them. This issue is so prevalent that it has a special name in ecological research: the silent victim problem [56]. As a result, the attack data collected in these domains may be highly biased (in a spatial sense). Due to such biased data collection, the data sample might not fairly represent the entire space of the problem [53], and the learned model of the attacker behavior might have different prediction accuracy in the park areas that have high vs. low patrol densities. Thus, it may or may not be optimal to rely on learned models of attacker behavior in patrol planning, and there is no straightforward method to determine the optimal course of action prior to deployment, i.e., whether to use the learned model (or not) in patrol planning. Moreover, a sub-optimal choice may lead to arbitrary losses for the defender.

To tackle the blind-spots of machine learning models in adversarial behavior modeling based on the real-world data, I have proposed a novel multi-expert online learning model for constrained patrol planning, which benefits from several expert planners where insufficient or imperfect historical records of past attacks are available to learn adversarial behavior. I introduced a model to integrate complex machine learning models of adversarial behavior along with an online learner to design efficient and feasible randomized defender strategies in Green Security Games [31] (Figure 1.3(c)).

The final contribution of my thesis is to develop an optimal solution against collusive adversaries in security games assuming both rational and boundedly rational adversaries. Research on security games has focused on settings where the defender must protect against either a single adversary or multiple, independent adversaries. However, there are a variety of real-world security domains where adversaries may benefit from colluding in their actions against the defender, e.g., wildlife poaching, urban crime, and drug trafficking. Given such adversary collusion may be more detrimental for the defender, she has an incentive to break up collusion by playing off the self-interest of individual adversaries.

Most previous work on security games assumes that different adversaries can be modeled independently [43, 47, 70]. However, there are many real-world security domains in which adversaries may collude in order to more effectively evade the defender. One example domain is wildlife protection. Trade in illicit wildlife products is growing rapidly, and poachers often collude both with fellow poachers and with middlemen who help move the product to customers [92]. These groups may coordinate to gain better access to information, reduce transportation costs, or reach new markets. This coordination can result in higher levels of poaching and damage to the environment. Additionally, connections have been observed between illicit wildlife trade and organized crime as well as terrorist organizations, and thus activities such as poaching can serve to indirectly threaten national security [94].

Despite mounting evidence of the destructive influence of collusive behavior, strategies for preventing collusion have not been explored in the security games literature (there are some recent exceptions, which we discuss in Section 3). Furthermore, analysis of collusive adversary behaviors is complicated by the bounded rationality of human adversaries; such analysis with data from human players is also missing in the security games literature.

I proposed COllusive Security Game (COSG) model with three players: one defender and two adversaries with the potential to collude against the defender. To generate optimal defender strategy against collusive adversaries, I provided two algorithms, (i) SPECTRE-R, which optimizes against collusive adversaries assuming them to be perfectly rational, and (ii) SPECTRE-BR, which optimizes against the learned behavior model to better prevent collusion between bounded rational adversaries. I conducted human subject experiments on Amazon Mechanical Turk (AMT) involving 700 human subjects using a web-based game that simulates collusive security games. Through the first round of the experiments, I collected data from human players responding to the strategies designed based on rational assumptions about adversarial behavior and I learned the parameters of the behavioral models from the experimental data. Then I designed defender strategies based on the learned adversarial behavior model, and I tested the defender strategies against human subjects again. I showed that when playing against human adversaries in experiments, my algorithms were able to learn from past human behavior and exploit their biases to more successfully prevent collusion among human adversaries than standard game theoretic approaches [29, 30] (Figure 1.3(b)).

1.3 Thesis Outline

In Section 2, I discuss the adversarial behavior models and the game-theoretic models which inspire the development of the predictive models and game theoretic decision solutions in this thesis. In Section 3, related work is discussed for background in the Green Security Game research. Next, in Section 4, I introduce the hybrid spatio-temporal model that I proposed for adversarial behavior in GSGs. Section 5 introduces the ensemble technique for adversarial behavior modeling and the real field tests conducted in multiple national parks to evaluate the model. In Section 5.5, I discuss how we can quantify uncertainty in predictions of our behavioral model by using Gaussian Process models and how we can incorporate such predictive models in the patrol planning models. In Section 6, I present the multi-expert online learning algorithm which recommends the best defender strategy based on several expert planner models using predictive adversarial models and online learning framework. In Section 7, I discuss my game theoretical model for collusive rational and bounded rational adversaries. Finally, in Section 8, I discuss the relevant future work for the use of machine learning and game theoretical frameworks to address key challenges in real-world problems and conclude my thesis.

Chapter 2

Background

2.1 Security Games

Stackelberg Security Game: The Stackelberg Security Game model, introduced almost a decade ago, has led to a large number of applications and has been discussed widely in the literature [50, 69, 87]. All of these works consider adversaries as independent entities and the goal is for a defender (leader) to protect a set of targets with a limited set of resources from a set of adversaries (followers)[1]. The defender commits to a strategy and the adversaries observe this strategy and each select a target to attack.

The defender's pure strategy is an assignment of her limited resources to a subset of targets and her mixed strategy refers to a probability distribution over all possible pure strategies. This mixed strategy is equivalently expressed as a set of coverage probabilities, $0 \leq c_t \leq 1$, that defender will protect each target, t [50]. Defender's utility is denoted by $U^u_\Theta(t)$ when target t is uncovered and attacked by the adversary and by $U^c_\Theta(t)$ if t is covered and attacked by the adversary. The payoffs for the attacker are analogously written by $U^u_\Psi(t)$ and $U^c_\Psi(t)$. The expected

[1] We use the convention in the security game literature where the defender is referred as ”she” and an adversary is referred to as ”he”.

utilities of the defender, $U_{\Theta}(t,C)$, and attacker, $U_{\Theta}(t,C)$ for the defender coverage vector C, are then computed as follows:

$$
\begin{aligned}
U_{\Theta}(t,C) = &\quad c_t \cdot U^{c}_{\Theta}(t) + (1 - c_t)U^{u}_{\Theta}(t) && (2.1) \\
U_{\Psi}(t,C) = &\quad c_t \cdot U^{c}_{\Psi}(t) + (1 - c_t)U^{u}_{\Psi}(t) && (2.2)
\end{aligned}
$$

The solution concept for security games involves computing a strong Stackelberg equilibrium (SSE) which assumes that the adversaries maximize their own expected utility and break ties in favor of the defender [47, 87].

2.2 Quantal Response Models

Another important area within game theory that provides concepts that we will use in this thesis for modeling and analyzing adversary behaviors is that of behavioral models [14]. This area is particularly relevant given our focus on modeling human adversaries in this study. In real-world settings, it is useful to model human adversaries as not strictly maximizing their expected utility, but rather, as their choosing strategies stochastically [61]. Quantal response equilibrium (QRE) is a solution concept based on the assumption of bounded rationality [63]. The intuition behind the QR model is that the higher the expected utility for an action, the higher the probability of the adversary selecting that action. SUQR [72] has been proposed as an extension of QR and seen to outperform QR in modeling human adversaries [97]. This model is used frequently in security games literature to predict the probability of attack at each target. The logit function shown in Equation 2.3 is the most common specification for QR and SUQR functional form where q_t is the probability of choosing strategy t among all possible strategies in set of T.

$$q_t = \frac{e^{\hat{U}_\Psi(t,C)}}{\sum\limits_{t \in T} e^{\hat{U}_\Psi(t,C)}} \tag{2.3}$$

In SUQR model, $\hat{U}_\Psi(t,C)$ refers to subjective utility, and it replaces expected utility. Subjective utility in SUQR is defined as a linear combination of key domain features including the defender's coverage probability and the adversary's reward and penalty at each target which are respectively weighted by ω_1, ω_2 and ω_3.These are assumed to be the most salient features in the adversary's decision-making process.

$$\hat{U}_\Psi(t,C) = \omega_1 \cdot c_t + \omega_2 \cdot U^u_\Psi(t) + \omega_3 \cdot U^c_\Psi(t) \tag{2.4}$$

2.3 Markov Random Field and EM algorithm

A Markov Random Field (MRF) is a graphical model that represents a joint probability distribution via an undirected graph. This graph consists of a set of nodes and links. Each node corresponds to a random variable or group of random variables and each link builds a connection between a pair of nodes. Since this model has the capability to represent complex behavior in real-world settings, we use this framework as one of the approaches to model illegal poaching activity in this study. To use this framework in the wildlife protection domain, we slice the forest area and the entire time period of study into smaller pieces called targets. Each target is represented by two random variables, $o^t_{i,j}$ and $a^t_{i,j}$, which denote the records of the detected poaching activities and the actual poaching attacks occurred at time step t and location i, j. $o^t_{i,j}$ denotes the known observed data and $a^t_{i,j}$ denotes the hidden variable for actual illegal activity that is

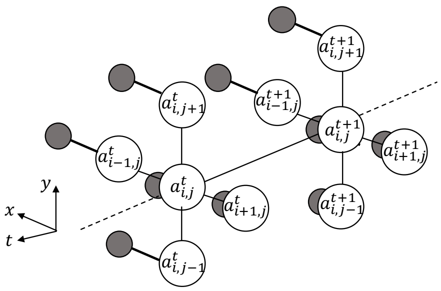


Figure 2.1: Spatio-temporal Markov Random Field graphical models

unknown to us. This Spatio-temporal MRF model is shown in Figure 2.1 where each node is connected to 6 other nodes of hidden variables, i.e., $a_{i,j}^{t}$ connects to $a_{i\pm1,j}^{t}$, $a_{i,j\pm1}^{t}$ and $a_{i,j}^{t\pm1}$ along with the observed poaching sign, $o_{i,j}^{t}$, shown as gray nodes. This model contains only pairwise cliques and the joint probability over the MRF network is:

$$P(a_1,...,a_N,o_1,...,o_N) = \prod_{i\neq j}\psi_{i,j}(a_i,a_j)\prod_{k}\phi_k(a_k,o_k) \tag{2.5}$$

To avoid index overload, nodes are labelled with serial numbers and $\psi_{i,j}(a_i,a_j)$ and $\phi_k(a_k,o_k)$ are potentials or compatibility functions between each pair of nodes in the graph.

Compatibility functions for this model could be defined in different ways, if we assume two possible labels of $L=\{0,1\}$ for hidden and observation variables of a_i and o_i, then the potential functions of $\psi_{i,j}(a_i,a_j)$ and $\phi_k(a_k,o_k)$ are represented in the form of Equation 2.6.

$$\psi_{i,j}(a_i,a_j) = \begin{bmatrix} \psi_{i,j}^{00} & \psi_{i,j}^{01} \\ \psi_{i,j}^{10} & \psi_{i,j}^{11} \end{bmatrix} \tag{2.6}$$

$$= \begin{bmatrix} P(a_i = 0|a_j = 0) & P(a_i = 0|a_j = 1) \\ P(a_i = 1|a_j = 0) & P(a_i = 1|a_j = 1) \end{bmatrix} \tag{2.7}$$

$$\phi_k(a_k,o_k) = \begin{bmatrix} \phi_k^{00} & \phi_k^{01} \\ \phi_k^{10} & \phi_k^{11} \end{bmatrix} \tag{2.8}$$

$$= \begin{bmatrix} P(o_k = 0|a_k = 0) & P(o_k = 0|a_k = 1) \\ P(o_k = 1|a_k = 0) & P(o_k = 1|a_k = 1) \end{bmatrix} \tag{2.9}$$

- **Hard pairwise potential functions:** This case is represented by 2×2 matrix of constant elements as shown below when parameters α_1, α_2 and β are known:

$$\psi_{i,j}(a_i,a_j) = \begin{bmatrix} \alpha_1 & 1-\alpha_2 \\ 1-\alpha_1 & \alpha_2 \end{bmatrix} \tag{2.10}$$

$$\phi_k(a_k,o_k) = \begin{bmatrix} 1 & 1-\beta \\ 0 & \beta \end{bmatrix} \tag{2.11}$$

For detection compatibility function $\phi_k^{10} = 0$ since it is not possible that rangers observe a poaching sign when an actual attack has not happened. β is the detectability parameters which models the imperfect observations of rangers.

- **Smooth pairwise potential functions with static and dynamic features:** The hard pairwise potential functions discussed earlier does not take into account the environmental factors that can influence the decision-making process by the poachers and thus simplifies the modeling approach. For instance, the probability that two neighbouring targets are attacked simultaneously could be dependent on the similarity between the features of those two targets including animal density ρ_i, slopes_i, forest cover h_i, NPP npp_i, distance from patrol post dp_i, distance from town dt_i distance from rivers dr_i, patrol coverage c_i, etc. In Equation 2.12 this dependency is shown as a linear combination of difference between feature vectors $\mathbf{X}_i$ and $\mathbf{X}_j$ of targets i and j, parametrized by α.

$$\psi_{i,j}(a_i,a_j) = \begin{bmatrix} \frac{e^{-\alpha_1(\mathbf{X}_i-\mathbf{X}_j)}}{1+e^{-\alpha_1(\mathbf{X}_i-\mathbf{X}_j)}} & \frac{1}{1+e^{-\alpha_2(\mathbf{X}_i-\mathbf{X}_j)}} \\ \frac{1}{1+e^{-\alpha_1(\mathbf{X}_i-\mathbf{X}_j)}} & \frac{e^{-\alpha_2(\mathbf{X}_i-\mathbf{X}_j)}}{1+e^{-\alpha_2(\mathbf{X}_i-\mathbf{X}_j)}} \end{bmatrix} \tag{2.12}$$

Similarly, the probability that attack has happened but the park rangers have not detected that attack, could be a function of the patrolling effort devoted to that target and the passability of the area or in particular forest land cover and slope. In Equation 2.13 this dependency is shown as a linear combination of patrol coverage c_i and forest cover h_i, weighted by β.

$$\phi_k(a_k,o_k) = \begin{bmatrix} 1 & e^{-\beta[c_k,h_k]} \\ 0 & 1-e^{-\beta[c_k,h_k]} \end{bmatrix} \tag{2.13}$$

In section 4, we will discuss our approach for modeling the smooth compatibility functions.

Model fitting: Estimating unknown parameters with EM algorithm The parameters of potential functions, discussed in the previous sections, are not known to us and in the real-world applications, we need to estimate them from the existing data. We use the EM algorithm to estimate the unknown parameters introduced in an MRF model.

In the EM model, the E-step calculates the conditional expectation:

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \mathbb{E}\left[logP(\mathbf{a},\mathbf{o}|\theta) \middle| \mathbf{o},\theta^{(t)} \right] \\ &= \sum_{\mathbf{a}\in\mathscr{A}} P(\mathbf{a}|\mathbf{o},\theta^{(t)}).logP(\mathbf{a},\mathbf{o}|\theta) \end{aligned} \tag{2.14}$$

$\mathscr{A}$ is the set of all possible configuration for **a** which is the actual attack. $\theta = \{\alpha, \beta\}$

The M-step maximize $Q(\theta|\theta^{(t)})$ to obtain the next estimate:

$$\theta^{(t+1)} = arg\max_{\theta} Q(\theta|\theta^{(t)}) \tag{2.15}$$

And then put $\theta^{(t)} \leftarrow \theta^{(t+1)}$ and repeat.

The conditional independence gives us the following equation:

$$P(\mathbf{o}|\mathbf{a}) = \prod_{i\in\mathscr{N}} P(o_i|a_i) \tag{2.16}$$

For the MRF model, the joint probability of o_i and a_i at each node could be written as:

$$P(a_i,o_i|a_{N(i)}) = P(o_i|a_i)P(a_i|a_{N(i)}) \tag{2.17}$$

$$\begin{aligned} logP(\mathbf{a},\mathbf{o}|\theta) &= \sum_{i\in\mathscr{N}} logP(a_i,o_i|a_{N(i)};\theta) \\ &= \sum_{i\in\mathscr{N}} logP(o_i|a_i;\beta)+logP(a_i|a_{N(i)};\alpha) \\ &= \sum_{i\in\mathscr{N}}\sum_{L(a_i)\in\mathscr{L}} logP(o_i|L(a_i);\beta)+logP(L(a_i)|a_{N(i)};\alpha) \end{aligned} \tag{2.18}$$

By substituting the above values in E-step, we will get:

$$Q(\theta|\theta^{(t)}) = \sum_{i\in\mathscr{N}}\sum_{L(a_i)\in\mathscr{L}} P^{(t)}(L(a_i)|o_i)\Big(logP(o_i|L(a_i);\beta)+logP(L(a_i)|a_{N(i)};\alpha)\Big) \tag{2.19}$$

Where:

$$P^{(t)}(L(a_i)|o_i) = \frac{P^{(t)}(o_i|(L(a_i);\beta).P^{(t)}(L(a_i)|a_{N(i)};\alpha)}{P(o_i)} \tag{2.20}$$

By replacing ψ and ϕ function as $P(o_i|L(a_i);\beta)$ and $P(L(a_i)|a_{N(i)};\alpha)$ in Equation 2.19 and finding the derivatives and putting them equal to zero to maximize Q, we can find the update rules for the EM algorithm. After enough number of iteration, we can find the final estimated parameters and replace them in the MRF model to predict the poaching activity and detection probability at each node.

Chapter 3

Related Work

3.1 Green Security Games

Game theoretic models, in particular security games are well known to be effective models of protecting valuable targets against an adversary, and have been explored extensively [5,42,45,52,64] and the problem of patrol planning has been well studied in this context [6,73]. However, much of this work assumes a perfectly rational adversary, which is not true for the wildlife protection domain, where poachers are boundedly rational. Green Security Games [23] were introduced to address the challenges specific to this domain, such as boundedly rational adversaries. While there has been work on learning these adversary models, this has been mostly done based on simulated games where data is collected by human subject experiments in the laboratory [28,29,42,72,99] rather than real world poachers. These methods are additionally unable to scale to real-world setups which typically have an enormous number of targets (e.g., 3900 targets of 1x1 sq. km in Murchison Fall park) and diverse geo-spatial characteristic.

In patrol planning for wildlife protection, PAWS was introduced as a risk-based randomized patrol generation algorithm which has been tested in the real world [22, 23, 98]. However, it relies on a specific type of explicit attacker behavior model such as Quantal Response and Subjective Utility Quantal Response [71]. Therefore, a framework for patrol planning to generate implementable patrolling routes against a black-box attacker was proposed in [95]. Although this framework can handle complex data-driven predictive model, it was not able to scale up for continuous patrol effort values. Furthermore, none of the aforementioned studies account for naturally occurring uncertainty in crime evidence collected by defenders and its consequent effects on planning, which I have directly addressed in my thesis.

3.2 Modeling approaches for adversarial behavior based on the real-world data

In data-driven wildlife protection literature, Critchlow et al. [19] analyzed spatio-temporal patterns in illegal activity in Uganda's Queen Elizabeth Protected Area (QEPA) using Bayesian hierarchical models. With real-world data, they demonstrated the importance of considering the spatial and temporal changes that occur in illegal activities. However, in this work and other similar works with spatio-temporal models [77, 78], no standard metrics were provided to evaluate the models' predictive performance (e.g., precision, recall). As such, it is impossible to compare our predictive models' performance to theirs. While [20] was a field test of [19]'s work, [77, 78] do not conduct field tests to validate their predictions in the real-world.

Also, [71] introduced a two-layered temporal Bayesian Network with hidden variables, CAPTURE [71]. CAPTURE assumes one global set of parameters for all of QEPA which ignores local differences in poachers' behavior. Additionally, the first layer, which predicts poaching attacks, relies on the current year's patrolling effort which makes it impossible to predict future attacks (since patrols haven't happened yet). While CAPTURE includes temporal elements in its model, it does not include spatial components and thus cannot capture neighborhood specific phenomena. In contrast to CAPTURE, [44] presented a behavior model, INTERCEPT, based on an ensemble of decision trees and was demonstrated to outperform CAPTURE. However, INTERCEPT assumes perfect detection of poaching activity by park rangers, leading to biases in final predictions. While this model accounted for spatial correlations, it did not include a temporal component. In contrast to these predictive models, my hybrid spatio-temporal model addresses both spatial and temporal components.

In the Machine Learning literature, spatio-temporal models have been used for prediction tasks in image and video processing. Markov Random Fields (MRF) were used by [86, 100] to capture spatio-temporal dependencies in remotely sensed data and moving object detection, respectively. Although, these models have not been used in adversarial settings, their expressiveness for spatio-temporal events make them suitable for adversarial behavior modeling that we will discuss in chapter 4. [11] proposed using the imagery from longwave thermal infrared cameras mounted on unmanned aerial vehicles (UAVs or drones) to spot poachers at night real-time and report them to park rangers before they are able to harm animals. They build upon deep learning techniques like Faster RCNN to detect moving objects in videos. To facilitate the image data labeling task for the object (e.g., animals and poachers) detection in green security games, [12] introduced VIOLA, a video labeling application for security domains. Unfortunately, in the national parks under study in this thesis, drones are not flown for patrolling purposes and we do not have access to that type of data to augment our original data sets by. The real-world data sets available to us are collected based on foot patrols conducted by the park rangers.

In the green security games, the wildlife crime data set suffers from a severe imbalance across the observation labels and an asymmetric uncertainty on the illegal activity records where positive instances are certain and negative instances (patrollers' records when they visited a place and did not find any illegal activities) are uncertain with different levels of uncertainty depending on the amount of the patrol effort spent for the collection of those instances. A closely related line of research to address these challenges in machine learning literature is the learning from positive and unlabeled examples approaches, which are well-explored the in text mining domain, in particular. The term PU learning (learning from positive and unlabeled examples) was first coined in [57] as a two-class (positive and negative) classification problem, where there are only labeled positive training data, but no labeled negative training data is available. They proposed a technique names A-EM to address this issue in the document classification domain. The main idea behind their novel model is that they add a large set of irrelevant documents (negative classes), which contains almost no positive document to the unlabeled class data and then the EM algorithm generates a sequence of classifiers. They also propose a classifier selection (or catch) criterion to select a good classifier from the set of classifiers produced by EM. The other existing techniques to handle this challenge include (i) a two-step strategy where some reliable negative instances are detected and used for learning [58], (ii) methods based on weighted positive and unlabeled data where positive and negative instances are given different weights [21], and (iii) methods that can handle noisy negative data including the SVM techniques.

In ensemble modeling literature, ensemble-based techniques are well-known to improve performance of single models (i.e., weak learner) and they have been widely used to address imbalance in positive and negative instances of observations in a variety of domains from chaotic behavior modeling for stock market prediction [16], knowledge base population in text analysis [76] and vowel discrimination tasks [38]. Ensemble techniques can be categorized as iterative based ensembles or parallel ensembles [36]. [44] leverages iterative based ensembles for adversary behavior modeling. However, parallel ensembles which are based on parallel re-sampling and bagging of weak learners have also been shown to be very time saving and easy to develop in many practical problems to learn human behavior, e.g., in online banking fraud detection [93]. In this thesis, in section 5, I propose a parallel ensemble for which we re-sample via filtering of negative instances of crime depending on the amount of the defenders' effort to collect those instances. By this we are able to minimize the adverse effects of uncertainty in negative instances of crime and boost the prediction accuracy by generating more specialized weak learners based on more confident subsets of data.

3.3 Field test evaluation of adversarial behavioral models

It is vital to validate predictive models in the real world, and both [20] and [44] have conducted field tests in QEPA. [44] conducted a one month field test in QEPA and demonstrated promising results for predictive analytics in this domain. Unlike the field test we conducted, however, that was a preliminary field test and was not a controlled experiment. On the other hand, [20] conducted a controlled experiment where their goal, by selecting three areas for rangers to patrol, was to maximize the number of observations sighted per kilometer walked by the rangers. Their test successfully demonstrated a significant increase in illegal activity detection at two of the areas, but they did not provide comparable evaluation metrics for their predictive model. Also, our field test was much larger in scale, involving 27 patrol posts compared to their 9 posts.

3.4 Game theoretical frameworks for patrol planning

There has been a lot of effort in GSGs at learning models of attacker behavior from historical patrolling data, which has then been used inside Stackelberg game solvers to generate patrol plans [87]. A lot of initial effort in this direction assumed attackers behaved according to parametric models, e.g., Quantal Response [71], Subjective Utility Quantal Response [22, 23, 99], SHARP model [42], etc., and tried to learn model parameters which best fit the historical data. Unfortunately, the assumption of having a fixed model of attacker behavior is quite restrictive and is not robust to any errors in our knowledge about the model type. As a result, there has also been recent effort at learning black-box machine learning models of attacker behavior from past patrolling data which can be used to plan patrols [27, 35, 95]. Sinha et al. [84] proved sample complexity results for learning in Stackelberg Security Games which showed that a huge amount of prior knowledge (historical data) is required to achieve good performance in the GSG approach. Moreover, as mentioned in the introduction, the poaching data collected in these domains is highly biased (in a spatial sense) and as a result, planning patrols based on this data may lead to arbitrary losses. Moreover, In our work, we propose online learning approaches which do not rely on past data to learn attacker models (or at least trade off between (i) relying on past data; and (ii) online learning approaches), and as we show in our evaluation section, this may lead to significant improvements in solution quality.

In the field of repeated Stackelberg Security Games, Klima et al. [48, 49] solved the problem of patrol planning for repeated border patrols with online learning algorithms. They provided an experimental analysis of the performance of several well-known online learning algorithms. However, they emphasized empirical results and they do not provide theoretical analysis. Balcan et al. [3] solved repeated Stackelberg Security Games with varying attacker types captured with different payoff matrices and proposed an online learning approach, but they assumed perfect rationality of attackers and complete knowledge of the payoff matrices, which is unrealistic to expect in the wildlife poaching domain. Blum et al. [9] optimizes defender strategy with no prior knowledge in repeated Stackelberg Security Games but they consider a query based model, where they try to learn good approximations of the payoff matrices with the least amount of queries, which is an orthogonal setting compared to our work.

In another closely related work, Xu et al. [96] proposed an online learning approach to solving repeated Stackelberg Security Games under no assumptions on the adversary's behavior. While the problem that we are solving in chapter 6 is similar to the one considered in [96], their work do not take into account spatio-temporal scheduling constraints while planning patrols. As a result, the generated patrols are un-implementable in the real-world, and thus, their approach is not easily usable in the real-world. In our work, we ensure that our proposed algorithms generates patrols which take into account several important scheduling constraints. Moreover, Xu et al. [96] do not take into consideration any prior knowledge and learn models from scratch, whereas our approach learns whether models based on prior knowledge are better (or worse) than models learned on-the-fly and takes decisions accordingly.

[91] proposed a deep reinforcement learning framework to address green security games with real-time information such as attackers' footprints. They designed a deep reinforcement learning-based algorithm to compute the defenders' best strategy against a best-responding attacker. Although this technique takes into account the real-time information, it does not use the historical real-world data collected by the park rangers in the past. Additionally, they constrain each poacher to a path and the defender agent in their system explores to find any footprint from

the attacker and follows the footprints. If there are multiple footprints at the same cell, the defender will randomly choose one to follow. However, in our study, we do not restrict attackers' actions to a path and the defender's strategy in this thesis is computed against adversaries with no restrictions in attacks. [41] also proposed a policy learning approach for continuous space security games using neural networks. Due to the sparsity of the historical data, in our study, we focus on the discretized action space (1 km $\times$ km cells) rather than continuous space. [13] proposed a multi-stage game theoretical model to address the broken signals in security games when sensors are involved. Their work is motivated by the mobile sensors, e.g., unmanned aerial vehicles (UAVs), which are used in security domains and can help with the tasks such as searching for poachers in conservation areas. Their game theoretical model aims to address real-world uncertainty in the sensor's detection of adversaries and adversaries' detection of the sensor's signals.

3.5 Game theoretical frameworks to handle collusive adversaries

Security game models where an adversary is capable of attacking multiple targets simultaneously have been explored in [51, 101]. Yin et al. [101] studied the Stackelberg vs. Nash in security games and extended their analysis to the case that the follower can attack multiple targets. However, a generalized form of SSG where the attacker attacks multiple targets simultaneously was proposed in [51]. To address cooperation between adversaries, [34] introduced a communication network based approach for adversaries to share their skills and form coalitions in order to execute more attacks. However, no previous work on security games has conducted behavioral analysis or considered the bounded rationality of human adversaries in deciding whether to collude in the first place.

Another area of related work, as well as one that provides concepts that we will use in this thesis for modeling and analyzing adversary behaviors in COSGs is that of behavioral models in game theory [14]. This area is particularly relevant given our focus on modeling human adversaries in this study. In real-world settings, it is useful to model human adversaries as not strictly maximizing their expected utility, but rather, as their choosing strategies stochastically [61]. Quantal response equilibrium (QRE) is a solution concept based on the assumption of bounded rationality [63]. The intuition behind the QR model is that the higher the expected utility for an action, the higher the probability of the adversary selecting that action. SUQR [72] has been proposed as an extension of QR and seen to outperform QR in modeling human adversaries [97]. This model is used in chapter 7 to predict the probability of attack at each target.

Another relevant aspect of bounded rationality is how humans weight probabilities. Prospect Theory (PT) proposes that individuals overweight low probabilities and underweight high probabilities; essentially, probabilities are transformed by an inverse S-shaped function [40, 90]. Various functional forms have been proposed to capture this relationship [40, 90]. Later work, specific to security games, has found the opposite of what Prospect Theory suggests: human players underweight low probabilities and overweight high probabilities [43]. This corresponds to an S-shaped weighting function. In either case, incorporating a model of probability perception allows the defender to exploit inaccuracies in the adversary's reasoning. Human subject experiments have been conducted for security games to test both bounded rationality and probability weighting [43], but have never included the collusive actions investigated in chapter 7.

Additionally, humans' decisions in strategic settings can be influenced by the relative advantage of participants. According to Inequity Aversion (IA) theory humans are sensitive to inequity of outcome regardless of whether they are in the advantaged or disadvantaged situation and they make decisions in a way that minimizes inequity [24]. Inequity aversion has been widely studied in economics and psychology and is consistent with observations of human behavior in standard economic experiments such as the dictator game and ultimatum game in which the most common choice is to split the reward 50-50 [7]. Along these lines and contrary to the theoretical predictions, IA theory also supports our analyses in the security game domain.

Chapter 4

A Hybrid Spatio-temporal Approach for Adversarial Behavior Modeling

This chapter discusses an approach for adversarial behavior modeling in Green Security Games via a hybrid spatio-temporal model that consists of two components including an ensemble model and a spatio-temporal Markov Random Field.

4.1 Problem Domain

At many sites now, rangers patrol and collect data related to snares they confiscate, poachers they arrest, and other observations. Given rangers' resource constraints, patrol managers could benefit from tools that analyze these data and provide future poaching predictions. However, this domain presents unique challenges. First, this domain's real-world data are few, extremely noisy, and incomplete. To illustrate, one of rangers' primary patrol goals is to find wire snares, which are deployed by poachers to catch animals. However, these snares are usually well-hidden (e.g., in dense grass), and thus rangers may not find these snares and (incorrectly) label an area as not having any snares. Second, poaching activity changes over time, and predictive models must account for this temporal component. Third, because poaching happens in the real world, there are mutual spatial and neighborhood effects that influence poaching activity. Finally, while field tests are crucial in determining a model's efficacy in the world, the difficulties involved in organizing and executing field tests often precludes them.

Previous works in this domain have modeled poaching behavior with real-world data. Based on data from a Queen Elizabeth Protected Area (QEPA) dataset, [71] introduced a two-layered temporal graphical model, CAPTURE, while [44] constructed an ensemble of decision trees, INTERCEPT, that accounted for spatial relationships. However, these works did not (1) account for both spatial and temporal components nor (2) validate their models via extensive field testing.

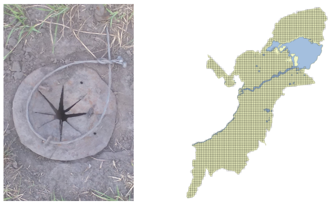
In this chapter, I provide the following contributions. (1) I introduce a new hybrid model that enhances an ensemble's broad predictive power with a spatio-temporal model's adaptive capabilities. Because spatio-temporal models require a lot of data, this model works in two stages. First, predictions are made with an ensemble of decision trees. Second, in areas where there are sufficient data, the ensemble's prediction is boosted via a spatio-temporal model. (2) In collaboration with the Wildlife Conservation Society and the Uganda Wildlife Authority, I designed and deployed a large, controlled experiment to QEPA. Across 27 areas I designated across QEPA, rangers patrolled approximately 452 kilometers over the course of five months; to our knowledge, this is the largest controlled experiment and field test of Machine Learning-based predictive models in this domain. In this experiment, I tested our model's selectiveness: is our model able to differentiate between areas of high and low poaching activity?

In experimental results, (1) I demonstrate our model's superior performance over the state-of-the-art [44] and thus the importance of spatio-temporal modeling. (2) During our field test, rangers found over three times more snaring activity in areas where I predicted higher poaching activity. When accounting for differences in ranger coverage, rangers found twelve times the number of findings per kilometer walked in those areas. These results demonstrate that (i) our model is selective in its predictions and (ii) our model's superior predictive performance in the laboratory extends to the real world.

4.2 Wildlife Crime Dataset: Features and Challenges

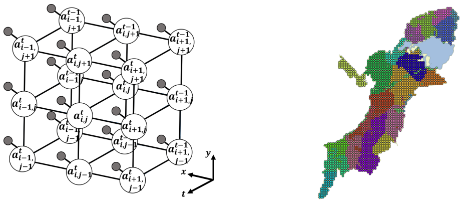
This study's wildlife crime dataset is from Uganda's Queen Elizabeth Protected Area (QEPA), an area containing a wildlife conservation park and two wildlife reserves, which spans about 2,520 square kilometers. There are 37 patrol posts situated across QEPA from which Uganda Wildlife Authority (UWA) rangers conduct patrols to apprehend poachers, remove any snares or traps, monitor wildlife, and record signs of illegal activity. Along with the amount of patrolling effort in each area, the dataset contains 14 years (2003-2016) of the type, location, and date of wildlife crime activities.

Rangers lack the manpower to patrol everywhere all the time, and thus illegal activity may be undetected in unpatrolled areas. Patrolling is an imperfect process, and there is considerable uncertainty in the dataset's negative data points (i.e., areas being labeled as having no illegal activity); rangers may patrol an area and label it as having no snares when, in fact, a snare was well-hidden and undetected. These factors contribute to the dataset's already large class imbalance; there are many more negative data points than there are positive points (crime detected). It is thus necessary to consider models that estimate hidden variables (e.g., whether an area has been attacked) and also to evaluate predictive models with metrics that account for this uncertainty, such as those in the Positive and Unlabeled Learning (PU Learning) literature [54]. We



(a) Snare (b) QEPA grid

Figure 4.1: Photo credit: UWA ranger



(a) Spatio-temporal model (b) Geo-Clusters

Figure 4.2: Geo-clusters and graphical model

divide QEPA into 1 square kilometer grid cells (a total of 2,522 cells), and we refer to these cells as targets. Each target is associated with several static geospatial features such as terrain (e.g., slope), distance values (e.g., distance to border), and animal density. Each target is also associated with dynamic features such as how often an area has been patrolled (i.e., coverage) and observed illegal activities (e.g., snares).

4.3 Models and algorithms

4.3.1 Prediction by Graphical models

4.3.1.1 Markov Random Field (MRF)

To predict poaching activity, each target, at time step $t \in \{t_1, ..., t_m\}$, is represented by coordinates i and j within the boundary of QEPA. In Figure 4.2(a), we demonstrate a three-dimensional network for spatio-temporal modeling of poaching events over all targets. Connections between nodes represent the mutual spatial influence of neighboring targets and also the temporal dependence between recurring poaching incidents at a target. $a^t_{i,j}$ represents poaching incidents at time step t and target i, j. Mutual spatial influences are modeled through first-order neighbors (i.e., $a^t_{i,j}$ connects to $a^t_{i\pm1,j}$, $a^t_{i,j\pm1}$ and $a^{t-1}_{i,j}$) and second-order neighbors (i.e., $a^t_{i,j}$ connects to $a^t_{i\pm1,j\pm1}$); for simplicity, the latter is not shown on the model's lattice. Each random variable takes a value in its state space, in this study, $\mathscr{L} = \{0,1\}$.

To avoid index overload, henceforth, nodes are indexed by serial numbers, $\mathscr{S} = \{1,2,...,N\}$ when we refer to the three-dimensional network. We introduce two random fields, indexed by $\mathscr{S}$, with their configurations: $\mathscr{A} = \{a = (a_1,...,a_N)|a_i \in \mathscr{L}, i \in \mathscr{S}\}$, which indicates an *actual* poaching attack occurred at targets over the period of study, and $\mathscr{O} = \{o = (o_1,...,o_N)|o_i \in \mathscr{L}, i \in \mathscr{S}\}$ indicates a *detected* poaching attack at targets over the period of study. Due to the imperfect detection of poaching activities, the former represents the hidden variables, and the latter is the known observed data collected by rangers, shown by the gray-filled nodes in Figure 4.2(a). Targets are related to one another via a neighborhood system, $\mathscr{N}_n$, which is the set of nodes neighboring n and $n \notin \mathscr{N}_n$. This neighborhood system considers all spatial and temporal neighbors. We define neighborhood attackability as the fraction of neighbors that the model predicts to be attacked: $u_{\mathscr{N}_n} = \sum_{n \in \mathscr{N}_n} a_n / |\mathscr{N}_n|$.

The probability, $P(a_i|u_{\mathscr{N}_n}, \alpha)$, of a poaching incident at each target n at time step t is represented in Equation 4.1, where α is a vector of parameters weighting the most important variables that influence poaching; Z represents the vector of time-invariant ecological covariates associated with each target (e.g., animal density, slope, forest cover, net primary productivity, distance from patrol post, town and rivers [19, 74]). The model's temporal dimension is reflected through not only the backward dependence of each a_n, which influences the computation of $u_{\mathscr{N}_n}$, but also in the past patrol coverage at target n, denoted by c_n^{t-1}, which models the delayed deterrence effect of patrolling efforts.

$$p(a_n = 1|u_{\mathscr{N}_n}, \alpha) = \frac{e^{-\alpha[Z, u_{\mathscr{N}_n}, c_n^{t-1}, 1]^\intercal}}{1 + e^{-\alpha[Z, u_{\mathscr{N}_n}, c_n^{t-1}, 1]^\intercal}} \tag{4.1}$$

Given a_n, o_n follows a conditional probability distribution proposed in Equation 4.2, which represents the probability of rangers detecting a poaching attack at target n. The first column of the matrix denotes the probability of not detecting or detecting attacks if an attack has not happened, which is constrained to 1 or 0 respectively. In other words, it is impossible to detect an attack when an attack has not happened. The second column of the matrix represents the probability of not detecting or detecting attacks in the form of a logistic function if an attack has happened. Since it is less rational for poachers to place snares close to patrol posts and more convenient for rangers to detect poaching signs near the patrol posts, we assumed dp_n (distance from patrol post) and c_n^t (patrol coverage devoted to target n at time t) are the major variables influencing rangers' detection capabilities. Detectability at each target is represented in Equation 4.2, where β is a vector of parameters that weight these variables.

$$p(o_n|a_n) = \begin{bmatrix} p(o_n = 0|a_n = 0) & p(o_n = 0|a_n = 1, \beta) \\ p(o_n = 1|a_n = 0) & p(o_n = 1|a_n = 1, \beta) \end{bmatrix} = \begin{bmatrix} 1, & \frac{1}{1+e^{-\beta[dp_n, c_n^t, 1]^\intercal}} \\ 0, & \frac{e^{-\beta[dp_n, c_n^t, 1]^\intercal}}{1+e^{-\beta[dp_n, c_n^t, 1]^\intercal}} \end{bmatrix} \tag{4.2}$$

We assume that (o,a) is pairwise independent, meaning $p(o,a) = \prod_{n \in \mathscr{S}} p(o_n, a_n)$.

4.3.1.2 EM algorithm to infer on MRF

We use the Expectation-Maximization (EM) algorithm [8] to estimate the MRF model's parameters $\theta = \{\alpha, \beta\}$. For completeness, we provide details about how we apply the EM algorithm to our model. Given a joint distribution $p(o, a|\theta)$ over observed variables o and hidden variables a, governed by parameters θ, EM aims to maximize the likelihood function $p(o|\theta)$ with respect to θ. To start the algorithm, an initial setting for the parameters θ^{old} is chosen. At E-step, $p(a|o,\theta^{old})$ is evaluated, particularly, for each node in MRF model:

$$p(a_n|o_n,\theta^{old}) = \frac{p(o_n|a_n,\beta^{old}).p(a_n|u^{old}_{\mathscr{N}_n},\alpha^{old})}{p(o_n)} \tag{4.3}$$

M-step calculates θ^{new}, according to the expectation of the complete log likelihood, $\log p(o,a|\theta)$, given in Equation 4.4.

$$\theta^{new} = \arg\max_{\theta} \sum_{a_n \in \mathscr{L}} p(a|o,\theta^{old}).\log p(o,a|\theta) \tag{4.4}$$

To facilitate calculation of the log of the joint probability distribution, $\log p(o,a|\theta)$, we introduce an approximation that makes use of $u^{old}_{\mathscr{N}_n}$, represented in Equation 4.5.

$$\log p(o,a|\theta) = \sum_{n \in \mathscr{S}} \sum_{a_n \in \mathscr{L}} \log p(o_n|a_n,\beta) + \log p(a_n|u^{old}_{\mathscr{N}_n},\alpha) \tag{4.5}$$

Then, if convergence of the log likelihood is not satisfied, $\theta^{old} \leftarrow \theta^{new}$, and repeat.

4.3.1.3 Dataset preparation for MRF

To split the data into training and test sets, we divided the real-world dataset into year-long time steps. We trained the model's parameters $\theta = \{\alpha,\beta\}$ on historical data sampled through time steps $(t_1,...,t_m)$ for all targets within the boundary. These parameters were used to predict poaching activity at time step t_{m+1}, which represents the test set for evaluation purposes. The trade-off between adding years' data (performance) vs. computational costs led us to use three years ($m = 3$). The model was thus trained over targets that were patrolled throughout the training time period (t_1,t_2,t_3). We examined three training sets: 2011-2013, 2012-2014, and 2013-2015 for which the test sets are from 2014, 2015, and 2016, respectively.

Capturing temporal trends requires a sufficient amount of data to be collected regularly across time steps for each target. Due to the large amount of missing inspections and uncertainty in the collected data, this model focuses on learning poaching activity only over regions that have been continually monitored in the past, according to Definition 1. We denote this subset of targets as $\mathscr{S}_c$.

Definition 1 ***Continually vs. occasionally monitoring:*** *A target i,j is continually monitored if all elements of the coverage sequence are positive; $c_{i,j}^{t_k} > 0, \forall k = 1,...,m$ where m is the number of time steps. Otherwise, it is occasionally monitored.*

Experiments with MRF were conducted in various ways on each data set. We refer to a) a *global* model with spatial effects as **GLB-SP**, which consists of a single set of parameters θ for the whole QEPA, and b) a *global* model without spatial effects (i.e., the parameter that corresponds to $u_{\mathscr{N}_n}$ is set to 0) as **GLB**. The spatio-temporal model is designed to account for temporal and spatial trends in poaching activities. However, since learning those trends and capturing spatial effects are impacted by the variance in local poachers' behaviors, we also examined c) a *geo-clustered* model which consists of multiple sets of local parameters throughout QEPA with spatial effects, referred to as **GCL-SP**, and also d) a *geo-clustered* model without spatial effects (i.e., the parameter that corresponds to $u_{\mathscr{N}_n}$ is set to 0) referred to as **GCL**.

Figure 4.2(b) shows the geo-clusters generated by Gaussian Mixture Models (GMM), which classifies the targets based on the geo-spatial features, Z, along with the targets' coordinates, $(x_{i,j}, y_{i,j})$, into 22 clusters. The number of geo-clusters, 22, are intended to be close to the number of patrol posts in QEPA such that each cluster contains one or two nearby patrol posts. With that being considered, not only are local poachers' behaviors described by a distinct set of parameters, but also the data collection conditions, over the targets within each cluster, are maintained to be nearly uniform.

4.3.2 Prediction by Ensemble models

A **Bagging ensemble model** or **B**ootstrap **agg**regation technique, called Bagging, is a type of ensemble learning which bags some weak learners, such as decision trees, on a dataset by generating many bootstrap duplicates of the dataset and learning decision trees on them. Each of the bootstrap duplicates are obtained by randomly choosing M observations out of M with replacement, where M denotes the training dataset size. Finally, the predicted response of the ensemble is computed by taking an average over predictions from its individual decision trees. To learn a Bagging ensemble, we used the *fitensemble* function of MATLAB 2017a. **Dataset preparation** for the Bagging ensemble model is designed to find the targets that are liable to be attacked [44]. A target is assumed to be attackable if it has ever been attacked; if any observations occurred in the entire training period for a given target, that target is labeled as attackable. For this model, the best training period contained 5 years of data.

4.3.3 Hybrid of MRF and Bagging ensemble

Since the amount and regularity of data collected by rangers varies across regions of QEPA, predictive models perform differently in different regions. As such, we propose using different models to predict over them; first, we used a Bagging ensemble model, and then improved the predictions in some regions using the spatio-temporal model. For global models, we used MRF for all continually monitored targets. However, for geo-clustered models, for targets in the continually monitored subset, $\mathscr{S}_c^q$, (where temporally-aware models can be used practically), the MRF model's performance varied widely across geo-clusters according to our experiments. q indicates clusters and $1 \leq q \leq 22$. Thus, for each q, if the average Catch Per Unit Effort (CPUE), outlined by Definition 2, is relatively large, we use the MRF model for $\mathscr{S}_c^q$. In Conservation Biology, CPUE is an indirect measure of poaching activity abundance. A larger average CPUE for each cluster corresponds to more frequent poaching activity and thus more data for that cluster. Consequently, using more complex spatio-temporal models in those clusters becomes more reasonable.

Definition 2 ***Average CPUE*** *is* $\sum_{n \in \mathscr{S}_c^q} o_n / \sum_{n \in \mathscr{S}_c^q} c_n^t$ *in cluster q.*

To compute CPUE, effort corresponds to the amount of coverage (i.e., 1 unit = 1 km walked) in a given target, and catch corresponds to the number of observations. Hence, for $1 \leq q \leq 22$, we will boost selectively according to the average CPUE value; some clusters may not be boosted by MRF, and we would only use Bagging ensemble model for making predictions on them. Experiments on historical data show that selecting 15% of the geo-clusters with the highest average CPUE results in the best performance for the entire hybrid model (discussed in the Evaluation Section).

4.4 Evaluations and Discussions

4.4.1 Evaluation metrics

The imperfect detection of poaching activities in wildlife conservation areas leads to uncertainty in the negative class labels of data samples [44]. It is thus vital to evaluate prediction results based on metrics which account for this inherent uncertainty. In addition to standard metrics in Machine Learning (e.g., precision, recall, F1) which are used to evaluate models on datasets with no uncertainty in the underlying ground truth, we also use the L&L metric introduced in [54], which is a metric specifically designed for models learned on Positive and Unlabeled datasets. L&L is defined as $L\&L = \frac{r^2}{Pr[f(Te)=1]}$, where r denotes the recall and $Pr[f(Te) = 1]$ denotes the probability of a classifier f making a positive class label prediction and is estimated by the percentage of positive predictions made by the model on a given test set.

4.4.2 Experiments with real-world data

Evaluation of models' attack predictions are demonstrated in Tables 4.1 and 4.2. To compare models' performances, we used several baseline methods, i) Positive Baseline, **PB**; a model that predicts poaching attacks to occur in all targets, ii) Random Baseline, **RB**; a model which flips a coin to decide its prediction, iii) Training Label Baseline, **TL**; a model which predicts a target as attacked if it has been ever attacked in the training data. We also present the results for Support Vector Machines, **SVM**, and AdaBoost methods, **AD**, which are well-known Machine Learning techniques, along with results for the best performing predictive model on the QEPA dataset, INTERCEPT, **INT**, [44]. Results for the Bagging ensemble technique, **BG**, and RUSBoost, **RUS**, a hybrid sampling/boosting algorithm for learning from datasets with class imbalance [82], are also presented. In all tables, **BG-G*** stands for the best performing model among all variations of the hybrid model, which will be discussed in detail later. Table 4.1 demonstrates that **BG-G*** outperformed all other existing models in terms of L&L and also F1.

Table 4.2 provides a detailed comparison of all variations of our hybrid models, **BG-G** (i.e., when different MRF models are used). When **GCL-SP** is used, we get the best performing model

Test set	2014					2015					2016				
Models	**PB**	**RB**	**TL**	**SVM**	**BG-G***	**PB**	**RB**	**TL**	**SVM**	**BG-G***	**PB**	**RB**	**TL**	**SVM**	**BG-G***
Precision	0.06	0.05	0.26	0.24	0.65	0.10	0.08	0.39	0.4	0.69	0.10	0.09	0.45	0.45	0.74
Recall	1.00	0.46	0.86	0.3	0.54	1.00	0.43	0.78	0.15	0.62	1.00	0.44	0.75	0.23	0.66
F1	0.10	0.09	0.4	0.27	0.59	0.18	0.14	0.52	0.22	0.65	0.18	0.14	0.56	0.30	0.69
L&L	1.00	0.43	4.09	1.33	6.44	1.00	0.37	3.05	0.62	4.32	1.00	0.38	3.4	1.03	4.88
Models	**RUS**	**AD**	**BG**	**INT**	**BG-G***	**RUS**	**AD**	**BG**	**INT**	**BG-G***	**RUS**	**AD**	**BG**	**INT**	**BG-G***
Precision	0.12	0.33	0.62	0.37	0.65	0.2	0.52	0.71	0.63	0.69	0.19	0.53	0.76	0.40	0.74
Recall	0.51	0.47	0.54	0.45	0.54	0.51	0.5	0.53	0.41	0.62	0.65	0.54	0.62	0.66	0.66
F1	0.19	0.39	0.58	0.41	0.59	0.29	0.51	0.61	0.49	0.65	0.29	0.53	0.68	0.51	0.69
L&L	1.12	2.86	6.18	5.83	6.44	1.03	2.61	3.83	3.46	4.32	1.25	2.84	4.75	2.23	4.88

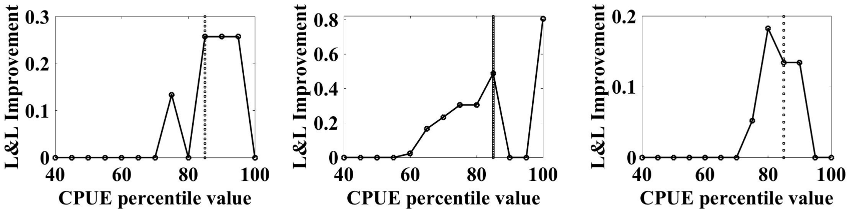
Table 4.1: Comparing all models' performances with the best performing BG-G model

Test set	2014				2015				2016			
MRF models	**GLB**	**GLB-SP**	**GCL**	**GCL-SP**	**GLB**	**GLB-SP**	**GCL**	**GCL-SP**	**GLB**	**GLB-SP**	**GCL**	**GCL-SP**
Precision	0.12	0.12	0.63	0.65	0.19	0.19	0.69	0.69	0.18	0.19	0.72	0.74
Recall	0.58	0.65	0.54	0.54	0.52	0.58	0.65	0.62	0.50	0.46	0.66	0.66
F1	0.20	0.20	0.58	0.59	0.28	0.29	0.65	0.65	0.27	0.27	0.69	0.69
L&L	1.28	1.44	6.31	6.44	0.99	1.14	4.32	4.32	0.91	0.91	4.79	4.88

Table 4.2: Performances of hybrid models with variations of MRF (BG-G models)

in terms of L&L score, which is denoted as **BG-G***. The poor results of learning a global set of parameters emphasize the fact that poachers' behavior and patterns are not identical throughout QEPA and should be modeled accordingly.

Our experiments demonstrated that the performance of the MRF model within $\mathscr{S}_c^q$ varies across different geo-clusters and is related to the CPUE value for each cluster, q. Figure 4.3(a) displays an improvement in L&L score for the **BG-G*** model compared to **BG** vs. varying the percentile of geo-clusters used for boosting. Experiments with the 2014 test set show that choosing the 85th percentile of geo-clusters for boosting with MRF, according to CPUE, (i.e., selecting 15% of the geo-clusters, with highest CPUE), results in the best prediction performance. The 85th percentile is shown by vertical lines in Figures where the **BG-G*** model outperformed the **BG** model. We used a similar percentile value for conducting experiments with the MRF model on test sets of 2015 and 2016. Figure 4.3(b) and 4.3(c) confirm the efficiency of choosing an 85th percentile value for those test sets, as well. Also, Table 4.1 demonstrates that for **BG-G***



(a) Test set 2014 (b) Test set 2015 (c) Test set 2016

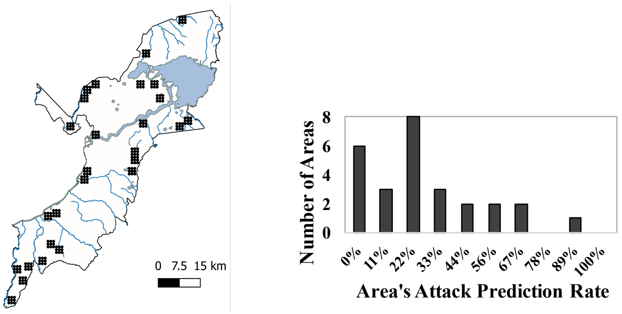
Figure 4.3: L&L improvement vs. CPUE percentile value; BG-G* compared to BG

recall increased up to almost 10% for the 2015 test set which would result in marking roughly 10% more vulnerable targets as attacked and thus protecting more endangered animals.

4.5 QEPA Field Test

While our model demonstrated superior predictive performance on historical data, it is important to test these models in the field.

The initial field test we conducted in [44], in collaboration with the Wildlife Conservation Society (WCS) and the Uganda Wildlife Authority (UWA), was the first of its kind in the Machine Learning (ML) community and showed promising improvements over previous patrolling regimes. Due to the difficulty of organizing such a field test, its implications were limited: only two 9-sq km areas (18 sq km) of QEPA were patrolled by rangers over a month. Because of its success, however, WCS and UWA graciously agreed to a larger scale, controlled experiment: also in 9 sq km areas, but rangers patrolled 27 of these areas (243 sq km, spread across QEPA) over five months; this is the largest to-date field test of ML-based predictive models in this domain. We show the areas in Figure 4.4(a). Note that rangers patrolled these areas in addition to other areas of QEPA as part of their normal duties.



(a) Patrolled areas (b) Prediction rates

Figure 4.4: Patrol Area Statistics

This experiment's goal was to determine the selectiveness of our model's snare attack predictions: does our model correctly predict both where there are and are not snare attacks? We define attack prediction rate as the proportion of targets (a 1 km by 1 km cell) in a patrol area (3 by 3 cells) that are predicted to be attacked. We considered two experiment groups that corresponded to our model's attack prediction rates from November 2016 - March 2017: High (group 1) and Low (group 2). Areas that had an attack prediction rate of 50% or greater were considered to be in a high area (group 1); areas with less than a 50% rate were in group 2. For example, if the model predicted five out of nine targets to be attacked in an area, that area was in group 1. Due to the importance of QEPA for elephant conservation, we do not show which areas belong to which experiment group in Figure 4.4(a) so that we do not provide data to ivory poachers.

To start, we exhaustively generated all patrol areas such that (1) each patrol area was 3x3 sq km, (2) no point in the patrol area was more than 5 km away from the nearest ranger patrol post, and (3) no patrol area was patrolled too frequently or infrequently in past years (to ensure that the training data associated with all areas was of similar quality); in all, 544 areas were generated across QEPA. Then, using the model's attack predictions, each area was assigned to an experiment group. Because we were not able to test all 544 areas, we selected a subset such that no two areas

Experiment Group	Exhaustive Patrol Area Group Memberships	Final Patrol Area Group Memberships
High (1)	50 (9%)	5 (19%)
Low (2)	494 (91%)	22 (81%)

Table 4.3: Patrol Area Group Memberships

overlapped with each other and no more than two areas were selected for each patrol post (due to manpower constraints). In total, 5 areas in group 1 and 22 areas in group 2 were chosen. Note that this composition arose due to the preponderance of group 2 areas (see Table 4.3). We provide a breakdown of the areas' exact attack prediction rates in Figure 4.4(b); areas with rates below 56% (5/9) were in group 2, and for example, there were 8 areas in group 2 with a rate of 22% (2/9). Finally, when we provided patrols to the rangers, *experiment group memberships were hidden to prevent effects where knowledge of predicted poaching activity would influence their patrolling patterns and detection rates*.

4.5.1 Field Test Results and Discussion

The field test data we received was in the same format as the historical data. However, because rangers needed to physically walk to these patrol areas, we received additional data that we have omitted from this analysis; observations made outside of a designated patrol area were not counted. Because we only predicted where snaring activity would occur, we have also omitted other observation types made during the experiment (e.g., illegal cattle grazing). We present results from this five-month field test in Table 4.4. To provide additional context for these results, we also computed QEPA's park-wide historical CPUE (from November 2015 to March 2016): 0.04.

Areas with a high attack prediction rate (group 1) had significantly more snare sightings than areas with low attack prediction rates (15 vs 4). This is despite there being far fewer group 1

Experiment Group	**Observation Count(%)**	**Mean Count(std)**	**Effort(%)**	**CPUE**
High (1)	15 (79%)	3 (5.20)	129.54 (29%)	0.12
Low (2)	4 (21%)	0.18 (0.50)	322.33 (71%)	0.01

Table 4.4: Field Test Results: Observations

areas than group 2 areas (5 vs 22); on average, group 1 areas had 3 snare observations whereas group 2 areas had 0.18 observations. It is worth noting the large standard deviation for the mean observation counts; the standard deviation of 5.2, for the mean of 3, signifies that not all areas had snare observations. Indeed, two out of five areas in group 1 had snare observations. However, this also applies to group 2's areas: only 3 out of 22 areas had snare observations.

We present Catch per Unit Effort (CPUE) results in Table 4.4. When accounting for differences in areas' effort, group 1 areas had a CPUE that was over ten times that of group 2 areas. Moreover, when compared to QEPA's park-wide historical CPUE of 0.04, it is clear that our model successfully differentiated between areas of high and low snaring activity. The results of this large-scale field test, the first of its kind for ML models in this domain, demonstrated that our model's superior predictive performance in the laboratory extends to the real world.

4.6 Conclusion

In this chapter, a hybrid spatio-temporal model to predict wildlife poaching threat levels was presented. We validated our model via an extensive five-month field test in Queen Elizabeth Protected Area (QEPA) where rangers patrolled over 450 sq km across QEPA — the largest field-test to-date of Machine Learning-based models in this domain. On real-world historical data from QEPA, our hybrid model achieves significantly better performance than prior work. On the data collected from our field test, we demonstrated that our model successfully differentiated between areas of high and low snaring activity. These findings demonstrated that our model's predictions are selective and also that its superior laboratory performance extends to the real world. Based on these promising results, future work will focus on deploying these models as part of a software package to UWA to aid in planning future anti-poaching patrols.

Chapter 5

Imperfect-observation-aware Ensemble Approach for Adversarial Behavior Modeling

This chapter discusses an approach for adversarial behavior modeling in Green Security Games via an ensemble model that consists of multiple fast running weak learners involved in a structured ensemble model compatible with the data collection scheme in Green Security Games. The evaluation results based on the historical data and the real field tests in Uganda are presented.

5.1 Problem Domain

Park rangers play a key role as the defenders of these protected areas, and are responsible for removing snares and traps placed by the poachers. Furthermore, they regularly collect records of illegal activities detected. While this data can provide significant insight and allow us to better model poachers' adversarial behavior, these records of attacks are unfortunately limited to the regions that the park rangers choose to visit (e.g., only about 60% of the protected areas are patrolled in each year). Moreover, the certainty about the absence of attacks largely depends on the amount of the patrol effort devoted to each area. Due to the vastness of the protected areas (e.g., Murchison Fall covers about 5000 sq. km shown in Figure 5.1), the limited number of outposts and rangers across the protected areas (e.g., about 30 outposts) and well-hidden placement of snares in the ground by poachers (Figure 5.2), it is not possible to conduct foot patrolling thoroughly throughout the area. Thus, it becomes necessary to consider this inherent uncertainty in real crime data in order to be able to use real data collected from the rangers we need to correctly model poachers' behavior.

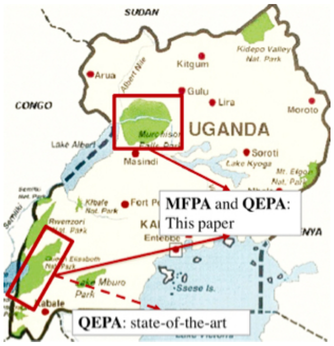


Figure 5.1: Protected areas in Uganda: we present seasonal poachers' behavior analysis across two different protected areas (7500 sq. km in total). State-of-the-art focused only on a single area of 2500 sq. km with annual coarse-grained crime analysis.

Previous work on data-driven modeling of wildlife poachers' behavior suffers from the following limitations: (i) they learn poachers' behavior without reasoning about the corresponding uncertainty in labels(due to insufficient amount of patrol effort) [44]. This results in unreliable predictions and consequently misleads the park rangers. Furthermore, (ii) they consider an annual basis for the temporal trend in crime predictions which results in missing short-term patterns in poachers' behavior [26, 44]. From a practical point of view, (iii) the computationally expensive techniques, including Markov Random Field and Dynamic Bayesian Networks [26, 71] proposed by many of these studies suffer from long runtimes and cannot be integrated into low resource outposts within the African protected areas. Last, to prove the reliability of the results to the law

Figure 5.2: Well-hidden snares detected by rangers, Photo credit: Uganda Wildlife Authority

enforcement agencies in Uganda, models have to be evaluated in different sites. However, (iv) none of the previous studies showed their models' performance across multiple protected areas.

In this chapter, we propose a new **i**mperfect-observation a**Ware E**nsemble (iWare-E[1]) method which takes into account the major challenge of adversarial behavior modeling in the wildlife protection domain, i.e., imbalanced non-uniform uncertainty on evidence of crime collected by defenders. (I) This approach significantly improves accuracy and runtime of the algorithm compared to state-of-the-art by using multiple fast running weak learners involved in a structured ensemble model compatible with the data collection scheme in protected areas. (II) we propose a scalable planning algorithm to design patrols, which utilizes the behavior prediction model (as a black box) and applies a piecewise linear approximation to reason about continuous values of patrol effort, which allows us to generate fine-grained patrols. we show that this approach results in up to 150% improvement in solution quality compared to the state-of-the-art. (III) Moreover, we evaluate all models on fine-grained temporal resolutions, i.e., seasonally, and for the first time, (IV) we evaluated all of our models on a larger scale based on real-world data across multiple protected areas including Murchison Fall and Queen Elizabeth in Uganda, covering 5000 sq. km and 2500 sq. km, respectively.

[1] To be pronounced similar to ivory

5.2 Predictive Model and Algorithm

5.2.1 Domain Features

The wildlife crime datasets in this study are from Uganda. We study Murchison Fall National Park jointly with Bugungu and Karuma wildlife reserves, and Queen Elizabeth National Park with Kigezi and Kyambura wildlife reserves. We refer to these protected areas as MFPA and QEPA, which span about 5000 sq. km and 2500 sq. km, respectively. There are 30 and 20 patrol posts situated across these protected areas from which Uganda Wildlife Authority (UWA) rangers conduct patrols. Along with the amount of patrolling effort in each area, the datasets contain 14 years (2003-2016) of the type, location, and date of wildlife crime activities. To study wildlife crime, we divide the protected areas into 1 sq. km grid cells. Each of these cells is associated with several static geo-spatial features such as terrain (e.g., slope), distance values (e.g., distance to border, roads, and towns), and animal density. Additionally, each cell is associated with dynamic features such as patrol effort (coverage) across time and observed illegal activities (e.g., snares). Patrol effort is the amount of distance walked by park rangers across a cell at a specific time step. Since park rangers do not have unlimited manpower to patrol each cell thoroughly, it is possible that the amount of distance walked by them is not sufficient and consequently, some of the well-hidden snares are not detected by them. This fact is the source of uncertainty over the negative instances of crime and has to be considered in the adversarial reasoning.

5.2.2 Dataset Preparation

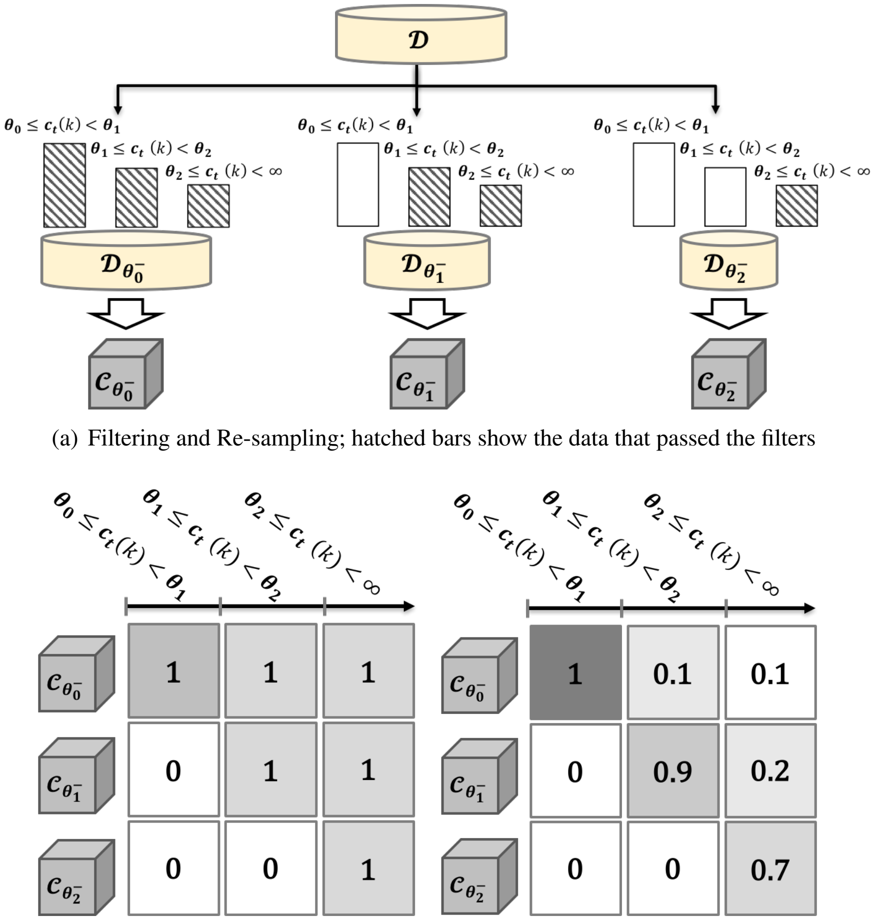
We create the wildlife crime datasets, $\mathscr{D} = (\mathbf{X}, \mathbf{y}, \mathbf{w})$, studied in this study from a dataset of recorded illegal activity by discretizing the records by time and by location so that we have a set of T time steps and N locations. $\mathbf{X} \in \mathbb{R}^{TN \times f}$ is a matrix of f predictor features recorded at each of these T discrete time steps and N locations. Each row of predictor features $X(k)$ includes several time-invariant geo-spatial features (discussed earlier) associated with each location (e.g., average animal density, slope, forest cover, net primary productivity, distance from patrol post, town, rivers, park boundaries, salt licks and water holes) and a set of time-variant covariates, patrol effort $c_{t-1}(k)$, that is the amount of patrol coverage during the previous time step $t-1$, which models the potential deterrence effect of patrols and $\mathbf{c_t}(\mathbf{k})$ the amount of patrol effort in the current time step, which models the effort devoted to each data point at the data collection time. $\mathbf{y} \in \{0,1\}^{TN}$ denotes the observation vector associated with all data points. Additionally, each data point in the dataset is associated with a weight $\mathbf{w} \in \{0,1\}^{TN}$. In the original dataset all weights are 1, however, if data point k is recognized as a sufficiently uncertain data point by the algorithm, $\mathbf{w}(k)$ will be changed to 0 and k is disregarded from the training set. To train any predictive model in this study, we divide this data into two sets for training, $\mathscr{D}^{tr}$, and testing, $\mathscr{D}^{ts}$. For our study, we used a training set which includes the first $T-1$ years of crime data (corresponding to 6 years) and tested on the data in next successive year.

5.2.3 Uncertainty in Poaching Activity Detection

While park rangers attempt to remove and record any illegal activity signs (e.g., snares and traps), it is often the case that they do not detect such signs, particularly if the snares are well-hidden. The success with which they detect these signs is linked to the amount of effort exerted in patrolling these regions. While positive records of poaching are assumed to be reliable in this study regardless of the amount of patrol effort, there is an intrinsic uncertainty associated with negative labels in the dataset, which depends on the patrol effort amount c_t (i.e., distance walked) devoted to each region during the data collection period, t. In particular given a threshold for patrolling effort θ, negative data samples recorded based on a patrol effort of $c_t \geq \theta$ are relatively more reliable (i.e., more probable to be actual negative samples) compared to the ones that were recorded based on a patrol effort of $c_t \leq \theta$. We use the notation subscript of θ_i^- to represent an instantiation of weight vectors in our dataset where negative samples recorded by a patrol effort of $c_t \leq \theta_i$ are ignored. In other words, for each data point k in $\mathscr{D}$, if $\mathbf{y}(k) = 1$, then $\mathbf{w}_{\theta_i^-}(k) = 1$. If $\mathbf{y}(k) = 0$ then $\mathbf{w}_{\theta_i^-}(k) = 1$ when $\mathbf{c}_t(k) \geq \theta_i$ and $\mathbf{w}_{\theta_i^-}(k) = 0$ when $\mathbf{c}_t(k) \leq \theta_i$.

5.2.4 Imperfect Observation-aware Ensemble

Due to diversified and robust characteristics of ensemble techniques, we propose a new **i**mperfect observation-a**ware** **E**nsemble model (iWare-E), which is able to handle the intrinsic uncertainty in the poaching activity data collection scheme by park rangers mentioned earlier. This ensemble technique outlined in Algorithm 1 involves multiple weak learners (also known as experts or ensemble members) which vote on the final predictions. Each weak leaner is trained based on a subset of the dataset, $\mathscr{D}_{\theta_i^-}$, filtered by a threshold θ_i where i is in $\{0, 1, ..., I-1\}$ and $\theta_i \leq \theta_{i+1}$. Line 2 in Algorithm 1 indicates that for any choice of $[\theta_{min}, \theta_{max}]$, I number of equally or unequally distanced intermediate thresholds θ_i can be obtained such that $\theta_{min} \leq \theta_i \leq \theta_{max}$ and consequently I weak learners, $\mathscr{C}_{\theta_i^-}$ can be trained on the corresponding $\mathscr{D}_{\theta_i^-}$ (line 6 in Algorithm 1). Figure 5.3(a) shows how patrol effort is filtered by different thresholds to generate a different sub-dataset and a corresponding expert in ensemble. The leftmost branch in the Figure 5.3(a) represents the case that $\theta_0 = 0$, i.e., the entire dataset and the rightmost branch represents the the case where negative instances of crime associated with $c_t \leq 2$ are disregarded.


$\mathcal{D}$
$\theta_0 \leq c_t(k) < \theta_1$
$\theta_1 \leq c_t\ (k) < \theta_2$
$\theta_2 \leq c_t\ (k) < \infty$
$\mathcal{D}_{\theta_0^-}$
$\mathcal{D}_{\theta_1^-}$
$\mathcal{D}_{\theta_2^-}$
$\mathcal{C}_{\theta_0^-}$
$\mathcal{C}_{\theta_1^-}$
$\mathcal{C}_{\theta_2^-}$

(a) Filtering and Re-sampling; hatched bars show the data that passed the filters

	$\theta_0 \leq c_t(k) < \theta_1$	$\theta_1 \leq c_t\ (k) < \theta_2$	$\theta_2 \leq c_t\ (k) < \infty$
$\mathcal{C}_{\theta_0^-}$	1	1	1
$\mathcal{C}_{\theta_1^-}$	0	1	1
$\mathcal{C}_{\theta_2^-}$	0	0	1

(b) Qualification Matrix Example

	$\theta_0 \leq c_t(k) < \theta_1$	$\theta_1 \leq c_t\ (k) < \theta_2$	$\theta_2 \leq c_t\ (k) < \infty$
$\mathcal{C}_{\theta_0^-}$	1	0.1	0.1
$\mathcal{C}_{\theta_1^-}$	0	0.9	0.2
$\mathcal{C}_{\theta_2^-}$	0	0	0.7

(c) Votes Power Matrix Example

Figure 5.3: Schema of iWare-E model

To address the voting scheme among the ensemble members, $\mathscr{C}_{\theta_i^-}$, we propose a binary vote qualification matrix, $\mathbf{V}^q$ which determines the qualification, 1, or disqualification, 0, of weak learners (each represented by a row), across ranges of c_t indicated by $[\theta_i, \theta_{i+1})$ (each represented by a column). Since each of these models are qualified to make predictions on data points which fulfill the condition $c_t \geq \theta_i$, the vote qualification matrix is a triangular matrix with size $I \times I$ (lines 7 through 11 in Algorithm 1). An example of this matrix is illustrated via the table in Figure 5.3(b), where each column represents an interval on $c(t)$ and each row represents a trained expert in the ensemble. It is worth noting that number intervals and number of experts are always equal (denoted by I here). If an expert is qualified to make predictions on an interval, the corresponding $\mathbf{V}^q$ element is 1. Furthermore, we also introduce a vote power matrix $\mathbf{V}^p$ of size $I \times I$ which contains the weights or vote power of each of the weak learners (each represented by a row), across ranges of c_t indicated by $[\theta_i, \theta_{i+1})$ (each represented by a column). An example of a vote power matrix is shown with different shade of gray rectangles and numbers associated with them in Figure 5.3(c).

The actual weights on the weak learners are a combination of qualification and vote power matrices, $\mathbf{V}^{qp} = \mathbf{V}^q \circ \mathbf{V}^p$. To ensure proper weighing of qualified weak learners within each range of $[\theta_i, \theta_{i+1})$, $\mathbf{V}^{qp}$ is normalized such that each column sums up to one (lines 13 and 14 in Algorithm 1). While $\mathbf{V}^q$ depends on the structure of the ensemble method, $\mathbf{V}^p$ is a hyper parameter. To tune this hyper parameter, we choose an initial $\mathbf{V}_o^p$ and a validation set, and then we use Algorithm 1 to minimize the error between actual observations and estimations by the model. This tuned $\mathbf{V}^p$ is used for training the ensemble on other sets via Algorithm 1. To make prediction on the test set and evaluate the model, the appropriate interval (which depends on the value of c_t) is obtained (line 3 in Algorithm 2) and then, the weighted average of all experts' predictions is computed using $\overline{\mathbf{V}}^{qp}$ (line 3 and 5 in Algorithm 2).

```
Algorithm 1: Train iWare-E
    input : Train dataset ($\mathscr{D}^{tr}$, $\mathbf{w^{tr}}$, $\mathbf{c_t^{tr}}$);
            Threshold parameters ($\theta_{min}$, $\theta_{max}$, $I$);
            Vote power matrix, ($\mathbf{V}^p$, size $I \times I$)
    output: Classifiers and weights matrix ($\mathscr{C}_{\theta_i^-}$ and $\overline{\mathbf{V}}^{qp}$)
    find threshold values for I intervals on $c_t$;
    $\theta \leftarrow$ FindThresholdVector($\theta_{min}$, $\theta_{max}$, $I$);
    train the classifiers;
    for $i \leftarrow 0$ to $I-1$ do
        $\mathscr{D}^{tr}$,$\mathbf{w}^{tr}_{\theta_i^-}$ $\leftarrow$FilterData($\mathscr{D}^{tr}$,$\mathbf{w^{tr}}$,$\mathbf{c_t^{tr}}$);
        $\mathscr{C}_{\theta_i^-}$ $\leftarrow$TrainABaggingEnsemble($\mathscr{D}^{tr}$,$\mathbf{w}^{tr}_{\theta_i^-}$);
        build vote qualification matrix, row is a member and column is an interval on $c_t$;
        for $j \leftarrow i$ to $I-1$ do
            $\mathbf{V}^q\ (j,i) \leftarrow 1$;
        for $k \leftarrow 0$ to $i-1$ do
            $\mathbf{V}^q\ (k,i) \leftarrow 0$;
    find total weights for member;
    $\mathbf{V}^{qp}$ $\leftarrow$ MultiplyElementWise($\mathbf{V}^q$,$\mathbf{V}^p$);
    $\overline{\mathbf{V}}^{qp}$ $\leftarrow$ ColumnWiseNormalizeToSumOne($\mathbf{V}^{qp}$);
```

5.3 Predictive Model Evaluation

5.3.1 Evaluation on Historical Data

For the illegal activity datasets we use, each protected area is divided into small 1×1 km regions and time steps of 3 months long are considered as opposed to the state-of-the-art [26] that considered coarse time steps of one year long, which makes it vulnerable to missing fine-grained temporal trends in poaching. To convince law enforcement agencies, it was essential to evaluate

Algorithm 2: Predict by iWareE

```
input  : Test dataset ($\mathscr{D}^{ts}$, $\mathbf{w^{ts}}$, $\mathbf{c_t^{ts}}$);
Threshold parameters ($\theta_{min}$, $\theta_{max}$, $I$);
Classifiers and weights matrix ($\mathscr{C}_{\theta_i^-}$ and $\overline{\mathbf{V}}^{qp}$)
output: Predicted probability of crime observation ($\mathbf{p}$)
test the classifiers;
for $\mathscr{D}^{ts}(k) \in \mathscr{D}^{ts}$ do
    $i^* \leftarrow$FindRelatedInterval($\mathbf{c_t^{ts}}(k)$);
    for $i \leftarrow 0$ to $I-1$ do
        $\mathbf{p}(k) \leftarrow \mathbf{p}(k) + \mathscr{C}_{\theta_i^-}(k) \cdot \overline{\mathbf{V}}^{qp}(i,i^*)$;
```

the predictive model across different protected areas and demonstrate superior performance of the model for smaller temporal resolutions.

For these datasets, the patrol effort is the amount of distance that park rangers walk through a 1 × 1 km region during a single time step of study. We tune hyper parameter based on training from 2007-2012 and validating on the 2013 dataset. Three different sets are used to evaluate our model, trained on the data from the years 2008-2013, 2009-2014 and 2010-2015 and tested on 2014, 2015 and 2016 respectively. Due to space consideration, detailed comparison of the proposed model with all possible baselines (e.g., Positive, Random, Training Label baselines) are presented in the supplementary material in the online Appendix[2]. We selected $\theta_0 = 0$ and $\theta_{I-1} = 7.5$ with 16 equally-distanced intermediate values of θ_i. Since the number of the data points with $c_t > 7.5$ was significantly lower compared to the ones with $c_t \leq 7.5$, we chose $\theta_{max} = 7.5$ to guarantee reasonable training datasets for all weak learners.

We compare the performance of the proposed model with the latest best performing existing models examined on the QEPA dataset in [26] in terms of standard machine learning metrics including AUC, Precision, Recall, F1. Since the metrics are used to evaluate models on datasets

[2] https://www.dropbox.com/s/cu08xr0txd8ur41/Appendix.pdf?dl=0

Test	**2016**					
	state-of-the-art				**iWare-E**	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
AUC	0.59	0.60	0.63	0.69	0.64	0.68
Prec.	0.20	0.17	0.19	0.22	0.20	0.21
Recall	0.40	0.57	0.59	0.62	0.57	0.58
F1	0.27	0.26	0.28	0.32	0.29	0.30
L&L	0.61	0.73	0.82	1.01	0.89	0.89
L&L %	8.73	10.41	11.02	14.4	11.0	12.0
Test	**2015**					
	state-of-the-art				**iWare-E**	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
AUC	0.58	0.63	0.67	0.69	0.68	0.69
Prec.	0.18	0.17	0.19	0.21	0.20	0.19
Recall	0.41	0.59	0.62	0.61	0.62	0.65
F1	0.25	0.26	0.29	0.31	0.30	0.30
L&L	0.59	0.80	0.95	1.02	0.99	1.01
L&L %	8.43	11.37	11.95	14.52	12.0	13.0
Test	**2014**					
	state-of-the-art				**iWare-E**	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
AUC	0.56	0.57	0.68	0.66	0.68	0.69
Prec.	0.23	0.19	0.26	0.26	0.25	0.26
Recall	0.33	0.54	0.64	0.61	0.62	0.64
F1	0.27	0.28	0.37	0.36	0.36	0.37
L&L	0.46	0.62	1	0.96	0.96	1.02
L&L %	7.68	10.36	16.3	15.97	16	17

Table 5.1: Comparing all models' performances for MFPA

with no uncertainty in the underlying ground truth, we also use the L&L metric [54], which is a metric specifically designed for models learned on Positive and Unlabeled datasets. L&L is defined as $L\&L = \frac{r^2}{Pr[f(Te)=1]}$, where r denotes the recall and $Pr[f(Te)=1]$ denotes the probability of a classifier f making a positive class label prediction and is estimated by the percentage of positive predictions made by the model on a given test set. We also discuss our algorithm runtime compared to the state-of-the-arts.

Test	**2016**					
	state-of-the-art				**iWare-E**	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
AUC	0.53	0.61	0.58	0.68	0.69	0.71
Prec.	0.13	0.08	0.08	0.11	0.13	0.12
Recall	0.13	0.59	0.58	0.62	0.57	0.62
F1	0.13	0.14	0.14	0.18	0.21	0.19
L&L	0.31	0.82	0.79	1.19	1.28	1.27
L&L %	1.85	4.84	4.43	7	7.0	7.0
Test	**2015**					
	state-of-the-art				**iWare-E**	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
AUC	0.53	0.63	0.62	0.7	0.66	0.71
Prec.	0.12	0.09	0.09	0.1	0.12	0.12
Recall	0.09	0.60	0.59	0.62	0.47	0.61
F1	0.10	0.16	0.15	0.18	0.19	0.2
L&L	0.18	0.90	0.81	1.06	0.94	1.21
L&L %	1.12	5.62	4.98	6.6	6.0	7.0
Test	**2014**					
	state-of-the-art				**iWare-E**	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
AUC	0.58	0.70	0.68	0.74	0.68	0.76
Prec.	0.12	0.07	0.06	0.08	0.09	0.09
Recall	0.17	0.65	0.62	0.62	0.52	0.72
F1	0.14	0.12	0.11	0.14	0.15	0.16
L&L	0.53	1.17	1.02	1.34	1.22	1.75
L&L %	1.95	4.33	3.76	4.95	5.0	6.0

Table 5.2: Comparing all models' performances for QEPA

	state-of-the-art				iWare-E	
Mdl.	**SVB**	**DTB**	**MRF**	**HY**	**SVB-iW**	**DTB-iW**
MFPA	80	65	45850	15328	1250	183
QEPA	78	71	31115	10348	1309	175

Table 5.3: Average runtime over all years (seconds)

Table 5.1 summarizes the performance of different models including bagging ensemble of SVM (denoted as SVB), bagging ensemble of decision trees (denoted as DTB), Markov Random Field (denoted as MRF), hybrid of last two ones (HY) (presented in [26]) as the existing models in literature against iWare-E model with two different weak learners including SVB and DTB, which are denoted as SVB-iW and DTB-iW, respectively. For all cases, applying iWare-E ensemble on SVB or DTB improves the AUC and L&L scores up to 10% compared to SVB or DTB. Furthermore, in comparison with the HY model, which is the state-of-the-art model, our results are improved up to 10% (see recall for QEPA 2014 test set) while the runtime is significantly reduced as we will discuss later.

Currently, the SMART software (a Spatial Monitoring and Reporting Tool) is used worldwide by park rangers to collect data and make decisions about patrolling routes. However, this software does not exploit historical data to predict poachers' behavior. In order to make possible the integration of our software in platforms like SMART, we have to develop fast models able to be run on non-high-performance machines. To that end, we present runtime analysis of all models for both parks in Table 5.3. Notably, DTB-iW completes in less than 200 seconds and SWD-iW completes in about 1300 seconds, which are significantly lower compared to MRF and HY models that suffer from the slow speed of EM algorithm. Although, SVB and DTB are fastest, they do not perform well in terms of accuracy as discussed earlier. All the experiments were performed on a machine with 2.6GHz and 8GB RAM.

5.3.2 Evaluation in the Field: Queen Elizabeth Protected Area

In this section, we present the field test results for the Queen Elizabeth Protected Area and we compare our results with the existing patrolling practices.

5.3.2.1 Experiment results

Fortunately, for QEPA, we have access to the field test data studied in [26], where attack prediction labels were defined as the proportion of 1×1 sq. km cells inside a patrol area of 3×3 sq. km, that were predicted to be attacked by the model. When attack prediction rate was more than 50%, the area was classified as high (H) and when it was less than 50%, the area was classified as low (L). So two experiment groups of high and low were generated according to model's attack prediction rates from November 2016 - June 2017. Figure 5.4(a) and 5.4(b) shows two pictures adopted from a footage provided by the park ranger while conducting our field tests. Table 5.4 summarizes the field test results for the QEPA dataset across time and prediction sources. The first row indicates three time steps of three months long when field tests were executed by park rangers. For each of these time steps and for each source of prediction, three different values are reported for high (H) and low (L) regions, i.e., Ar, C, E denote the number of 3×3 sq. km regions, counts of observations and amount of patrol effort, respectively. Predictive model names are outlined in the first column. Hybrid model and iWare-E models are indicated by HY and iW, respectively. iW denotes the best performing model for QEPA which is DTB-iW.

Predictions of the iWare-E model depend on the amount of patrol effort devoted to each cell which is designed to take into account the detectability challenge. We assumed a fixed amount of 4.5 km for patrol effort during 3 months for each individual cell. This value was selected based on the historical distribution of patrol effort over different months. Table 5.4 demonstrates

(a) Detecting a well-hidden snare by the park rangers (b) Recording the detected snares by the rangers via their smart phones

Figure 5.4: Pictures adopted from the footage of the our field tests provided by the park rangers and Uganda Wildlife Authority

Months		**11/1/2016-1/31/2017**			**2/1/2017-4/30/2017**			**5/1/2017-6/30/2017**		
Counts		**Ar**	**C**	**E**	**Ar**	**C**	**E**	**Ar**	**C**	**E**
HY	H	5	1	10.1	5	9	8.9	5	7	10.51
	L	22	0	11	22	3	8.87	22	2	10.3
iW	H	4	1	9.95	9	10	9.1	10	9	9.45
	L	23	0	10	18	2	9.83	17	0	10.36

Table 5.4: Comparison across sources of prediction, QEPA

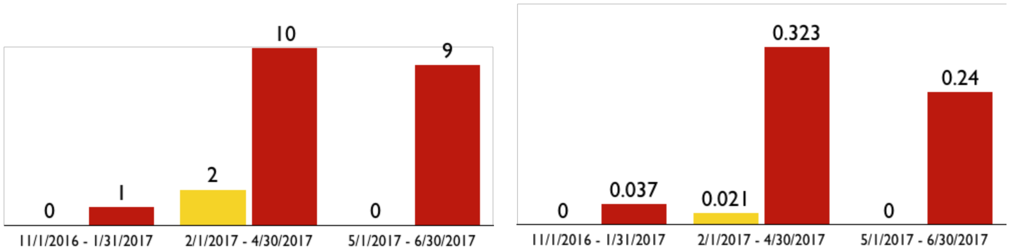
that iWare-E model is more selective between high and low groups and outperforms state-of-the-art predictions (HY). Furthermore, average monthly values of patrol effort, column named as E, computed for each high and low groups shows that park rangers are covering areas nearly uniformly. This indicates the shortcoming of their current method for planning as poachers are not attracted to all regions equally. Table 5.5 summarizes the statistical significance test based on occurrence of attack for all sources of prediction across time. Chi-Square test results for binary outcomes show that for the proposed iW model, p-value is below the significance level of 0.05 for the last two time steps, which determines that it is not likely that observations are due to chance for those time steps. To give an example of how we used the Chi-Square test for our problem, we review the steps of the test in the first time step for iW model. In that case, we recommended 27 areas of 9 sq. km (or 243 cells) that some of them are predicted as high-risk and some of them are predicted as low-risk. However, not all of the cells were patrolled by the park rangers. So to adjust the numbers for the statistical test, we looked at the total number of the patrolled cells in each group. So in each risk group, some of these patrolled cells had observation of 1 and some have observation of 0. Our null hypothesis is that whether the higher number of observations across the high-risk cells compared to the low-risk cells is due to chance. So we created the contingency table according to section 11-2 of [10] and conducted the Chi-Square test for the binary outcomes.

Months	**11/1/2016-1/31/2017**	**2/1/2017-4/30/2017**	**5/1/2017-6/30/2017**
HY	2.74×10^{-4}	5.76×10^{-4}	3.22×10^{-3}
iW	6.38×10^{-2}	5.92×10^{-7}	4.26×10^{-3}

Table 5.5: Statistical Significance Test: p values (≤ 0.05)

5.3.2.2 Comparison with historical patrolling data

The regions that we have recommended for patrolling are infrequently patrolled areas in the past and due to the spatio-temporal nature of the poaching activity, we do not have access to the identical historical data for the patrollers' existing practices. As such, we compare the number of findings in our recommended regions with the average number of snare observations across all locations and all times. To that end, we computed the historical average value for all patrolled 1×1 sq. km cells in QEPA (i.e., if snare has ever been found during a three-month time period in a cell, the snare observation is labeled as one, otherwise zero). This average number is equal to 0.0532 for QEPA. In Figure 5.5(a) and 5.5(b), we visualized the actual data and normalized data across all different time steps. As illustrated in Figure 5.5(b), our average hit rates in the patrolled high-risk regions in the field tests are more than the historical average value for the second and third time periods —0.323 and 0.24, respectively. This implies that our predictive



(a) bar chart for binary trap observations across the time steps

(b) bar chart for normalized binary trap observations based on patrolled cells across the time steps

Figure 5.5: Visualization of the field test results for iW model

model is capable of recognizing high-risk regions and if it is deployed for a long period of time, it has the potential to direct the park rangers to the right places to increase the number of findings compared to their current practices.

5.3.3 Evaluation in the Field: Murchison Falls Protected Area

In this section, we present the field test results for the Murchison Falls Protected Area and we compare our results with the existing patrolling practices.

5.3.3.1 Experiment results

The i-WareE model's performance on the historical data and also in the real field, i.e., Queen Elizabeth Protected Area, was evaluated in [27]. However, since protected areas have different ecological and topographical characteristics, it is vital to extend the field tests to other protected areas as well and evaluate the model's performance in a more comprehensive manner. For instance, Queen Elizabeth Protected Area (QEPA) and Murchison Falls Protected Area (MFPA) are both in Uganda but about 500 km apart. They are both large savanna parks but QEPA contains more scrub and woodland while MFPA has large grasslands. The shape of QEPA is long and access is easy from the edges to the center by poachers while MFPA is more circular with a more protected core and most poaching exist at the edges of the park. Furthermore, from the computational perspective, the amount of labels imbalance between positive and negative instances of poaching observations is different.

In this study, we demonstrate the field test results of the experiments that are conducted in Murchison Falls Protected Area, Uganda, for the first time. Previous field test study in QEPA classified the recommended areas into low vs. high-risk groups according to the model's prediction scores and requested the park ranger teams to conduct patrol over those areas for several months. Although, two risk categories are already providing insightful information, presenting higher levels of granularity is more beneficial from the conservation area management perspective. To that end, we classified our recommended areas into three risk groups of low, medium and high and requested the park ranger teams to patrol those areas. These areas were infrequently patrolled in the past. It is worth noting that to prevent any biases in data collection, we did not reveal the risk groups to the park rangers before conducting the experiments. Additionally, previous work selected multiple areas each consists of 9 contiguous 1×1 sq. km cells. However, we selected multiple areas consists of 4 contiguous 1×1 sq. km cells. With this fine-grained area selection scheme, we can ensure that the areas are patrolled sufficiently and thoroughly over the course of experiments.

Table 5.6 summarizes the field test results for the MFPA across time and risk groups. The first row indicates the time steps when the field tests were executed by the park rangers. The MFPA experiments were executed for 5 months. In total, 38 attacked cells were detected and snares were removed from there by the park rangers to save the wildlife. Since the predictive models are trained based on the data labels with three months time resolution, we demonstrate the field test observation with a similar approach. The data collected from November to December are aggregated and shown in the first column. Similarly, the data from January to March are grouped together and demonstrated in the second column in Table 5.6. Each row shows the data for the corresponding risk group. H, M, and L denotes the high-risk group, the medium-risk group, and the low-risk group, respectively. For each of the aforementioned time steps and for each risk group, four different values are reported, i.e., a count of binary observations or snares in this study (Obs.#), number of 1×1 sq. km cells (Cell#), amount of patrol effort in km (Patrol E.) and the normalized number of observations by the number of patrolled cells (Nrml.#).

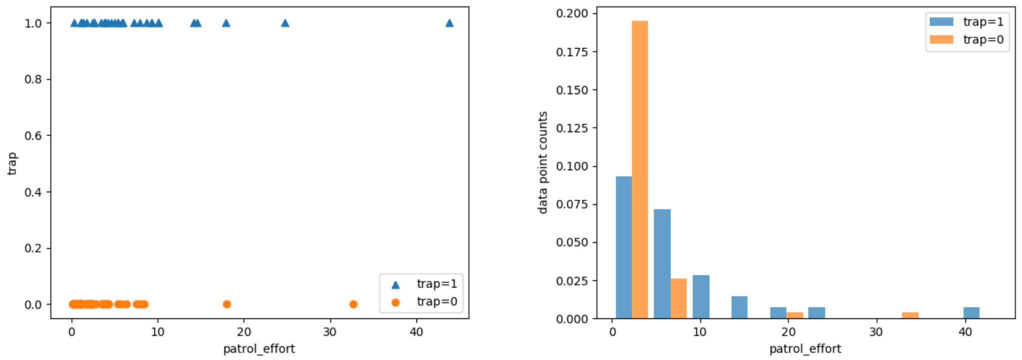
Table 5.6 implies that the number of snare activities observed in the high-risk regions is higher than the medium-risk regions and the number of snare activities observed in the medium-risk regions is higher than the low-risk regions for both time steps. Since the number of the park ranger resources is limited, not all of the recommended cells are patrolled at each time step. Hence the corresponding actual numbers of patrolled cells are also reported and based on that the normalized number of observations are computed. We used a Pearson's chi-squared statistical test of independence to assess whether the observations on two variables (i.e., Obs. and Risk Gr. in Table 5.6) are independent of each other. The contingency table for our analysis consists of two columns for Obs. in $\{0,1\}$ and three rows for high, medium, low-risk groups predicted by the adversary model. Each element in the contingency table represents the number of records made via patrolling in the field test experiments. The Chi-Square test applied on the entire experiments records show that the p-value is 1.05×10^{-2} and is below the significance level (0.05), which determines that it is not likely that observations are due to chance. In other words, this rejects the null hypothesis that the column variable (attack observations) is independent of the row variable (risk level predictions) in our contingency table.

Months	**11/1/2017-12/31/2017**				**1/1/2018-3/30/2018**			
Risk Gr.	**Obs.#**	**Cell#**	**Patrol E.**	**Nrml.#**	**Obs.#**	**Cell#**	**Patrol E.**	**Nrml.#**
H	6	18	71.62	0.33	17	36	197.4	0.47
M	5	21	31.86	0.24	7	34	83.36	0.21
L	2	10	12.62	0.20	1	13	45.1	0.08

Table 5.6: Field test results in MFPA

Since the aggregated patrol effort values for each risk group is not equal for all categories and more efforts are recorded for high-risk and medium-risk categories, we conducted an extra analysis to show that there is not a linear relationship between binary trap observations and patrol effort values. This can explain that higher number of observations in higher risk groups cannot be solely associated with the higher amount of patrol effort. Thus there are other factors (i.e., other predictor features), which also influence the prediction of the risk categories and our model is able to recognize them with reasonable accuracy. To that end, we used the point biserial correlation coefficient, which is a correlation coefficient used when one variable (e.g. trap observation) is dichotomous. The point-biserial correlation is mathematically equivalent to the Pearson correlation. Based on this test, correlation is 0.278 and the p-value is 0.0069, which indicates that the linear relationship is weak and the p-value is below the reference significance level (0.05).

Figure 5.6(a) confirms that even for low levels of the patrol effort, there is a chance to find traps and also for high levels of patrol effort, there is a chance to not to find any trap. Our observations indicate that this relationship between observations and patrol effort is not a strictly linear one. We can observe that his relationship is nonlinear and complex and in fact, we are trying to capture this relationship via our machine learning approach. Also, we can observe from Figure 5.6(b) that for both types of observations, the peak (mode) of the distribution happens at the small levels of patrol effort.



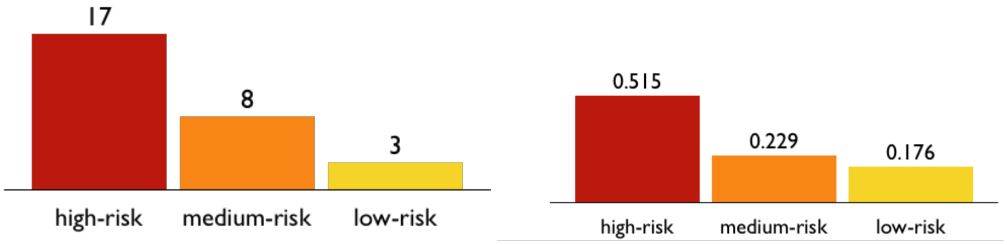
(a) Binary Trap Observations vs. Patrol Effort

(b) Histogram for Observations Frequency vs. Patrol Effort

Figure 5.6: Additional Analysis on the Field Test Results

5.3.3.2 Comparison with historical patrolling data

Similar to the Queen Elizabeth Protected Areas, the regions that we have recommended for patrolling are infrequently patrolled areas in the past and due to the spatio-temporal nature of the poaching activity, we do not have access to the identical historical data for the patrollers' existing practices in order to contrast absolutely with our field tests. As such, we compare the number of findings in our recommended regions with the average number of snare observations across all locations and all times. To that end, we computed the historical average value for all patrolled 1×1 sq. km cells in MFPA (i.e., if snare has ever been found during a three-month time period in a cell, the snare observation is labeled as one, otherwise zero). This average number is equal to 0.104 for MFPA. In Figure 5.7(a) and 5.7(b), we also visualize the aggregated actual data and normalize data across all 5 months. Our hit rates in the patrolled high-risk and medium-risk regions in the field tests —0.515 and 0.229, respectively —are more than the historical average value. This implies that our predictive model is capable of recognizing high-risk regions and if it is deployed for a long period of time, it can direct the park rangers to the right places to increase the number of findings compared to their current practices.



(a) bar chart for binary trap observations across risk groups

(b) bar chart for normalized binary trap observations based on patrolled cells across risk groups

Figure 5.7: Visualization of the aggregated field test results for iW model

5.4 Patrol Planning Model

The goal of developing these predictive models is to allow the rangers to leverage this additional information in order to better detect and reduce the number of attacks in protected areas. While there has been much work in Green Security Games (GSG) doing patrol planning in these domains [23, 60], much of this work has assumed explicit models for how poachers behave. These models, ranging from perfect rationality to bounded rationality models like Quantal Response (QR) and Subjective Utility Quantal Response (SUQR) [62, 68] can make planning much simpler due to their explicit nature. However when there is access to data on poaching activity, we can achieve much more accurate representations poachers behavior with machine learning models. These models are much more difficult to optimize from a planning perspective since we only have black box access to the predictions given a desired input. While there has been some prior work in GSG planning patrols which optimize black box functions [95], which we build off of, there are several key differences which make it so that these solution methods are not appropriate for our problem. The most important is that previous work is limited to optimizing over discrete levels of patrol effort. For more general machine learning models such as iWare-E, which can make predictions based off of continuous values of patrol effort, this can result in either large losses in solution quality when discretization levels are too coarse, or large runtimes when discretization is too fine (which we show in Figures 5.9 and 5.10). To address these issues we propose to instead approximate the machine learning model through the use of piecewise linear (PWL) functions. This allows us to reason about continuous values of patrol effort and achieve significant improvements in solution quality (up to 150% improvement) while remaining scalable (up to 400× increase in speed).

Following standard practice in Green Security Games [23, 43] we model the wildlife conservation patrolling problem as a game played on a graph $G(N,E)$ of nodes and edges, over a period of time T. We discretize the conservation area into a set of N grid cells, corresponding to the 1×1 km regions of the dataset. In order to protect the conservation area, rangers conduct patrols over these N grid cells. Patrolling a grid cell takes a certain amount of time and effort, and we assume that the ranger may only spend T time steps patrolling in any given day. Note that this time discretization is distinct from the 3 month long time steps considered in the dataset. Here we define a time step as the minimum amount of time it would take to cross a single grid cell (so that the ranger must spend at least 1 time step in each grid cell they choose to visit). A single patrol corresponds to a 1 unit flow on the time unrolled graph $G(N',E')$, with a set of nodes and directed edges given by:

$$N' \;:=\; \left\{ v' = (v,t) \;:\; v \in N \, t \in \{1,T\} \right\}.$$

$$E' \;:=\; \left\{ ((u,t_1),(v,t_2)) \;:\; \begin{array}{l} (u,v) \in E \cup \{(w,w)\} \;\; u,v,w \in N \\ t_2 = t_1 + 1 \\ t_1,t_2 \in \{1,T\} \end{array} \right\}.$$

One of the grid cells is a designated patrol post. All patrols must begin and end at this grid cell, and so we designate this grid cell as the source $s \in N$. For notational convenience, let $s_1 = (s,1) \in N'$ and $s_T = (s,T) \in N'$ the source node in the time unrolled graph at the first and last time steps respectively. The goal of these patrols is to detect signs of poaching activity, and rangers may conduct multiple rounds of successive patrols. The ability of a patrol to detect illegal activity at any grid cell $v \in N$ will depend on the level of patrol effort at that cell c_v^r (where r indicates the r^{th} round of patrols). Each patrol may expend a total of T units of patrol effort on any single patrol. Since a single unit of patrol effort is necessary to cover any cell for a single time step, a feasible patrol corresponds to a single unit flow on G' originating at the source s_1 and where the sum of the total flow on all edges in E' is equal to T. We then denote the set of feasible patrol efforts as $\mathscr{F}$ given by:

$$\mathscr{F} := \left\{ f_{u',v'} : \begin{array}{l} \sum_{u:(u',v')\in E'} f_{u',v'} = \sum_{u':(v',u')\in E'} f_{v',u'} \;\; \forall v' \in N' \\ \sum_{u':(s_1,u')\in E'} f_{s_1,u'} = \sum_{u':(u',s_T)\in E'} f_{u',s_T} = 1 \\ \sum_{(v',u')\in E'} f_{u',v'} = T \end{array} \right\}$$

Optimizing Detection Probability: We assume that for each grid cell there exists function $g_v : C_v^r \times C_v^{r-1} \to P_v^r$ which maps the current and past total defender patrol effort at a particular grid

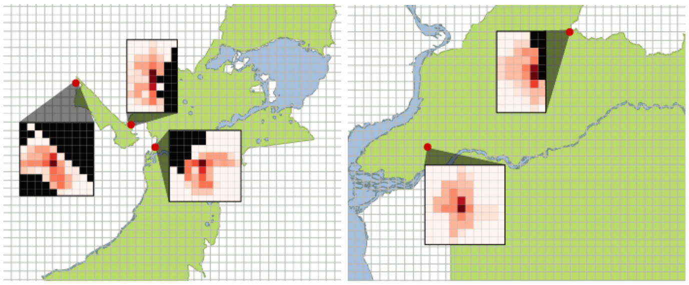


Figure 5.8: Sample heat map of patrol efforts computed for the QEPA (left) and MFPA(right)

cell $v \in N$ to a corresponding likelihood that there will be a detected attack P_v^r at that grid cell in round r. What we would like to do is solve for a series of patrols which maximizes the probability of detecting attacks over the entire area. The rangers conduct a total of K patrols within a single round r; since each patrol has a probability $\sum_{u':(u',v')\in E'} f_{u',v'}$ of visiting each cell v (ie. the sum of the total flow visiting that cell across all time steps), the expected aggregate patrol effort c_v at v is this probability times the total number of patrols K. The following Mathematical Program (MP) computes the optimal patrol effort which maximizes the predicted total detected attacks:

$$\max_{c,f} \quad \sum_{v\in N} g_v(c_v^r, c_v^{r-1})$$

$$\begin{aligned} & f_{u',v'} \in \mathscr{F} && \forall (u',v') \in E' \\ & K\textstyle\sum_{u':(v',u')\in E'} f_{u',v'} = c_v && \forall v \in N, v' = (v,t) \\ & \textstyle\sum_{v\in N} c_v = T \times K \end{aligned} \tag{$\mathscr{P}$}$$

Using the iWare-E model to generate these predictions we only have black box access to these functions g_v, so we instead use piecewise linear (PWL) approximations of the g_v in our objective functions. In order to construct these functions, we build datasets D_g of $m_r \times m_{r-1} \times N$ sample

Segments		Runtime (s)		Detections		Error (%)	
		QEPA	MFPA	QEPA	MFPA	QEPA	MFPA
($\mathscr{P}_1$)	5	0.25	0.26	8.8	18.8	17	2.7
	10	1.9	1.8	14.5	22.7	9.4	2.9
	20	1.2	12.2	17.4	23.5	0.2	8.2
	40	20.2	57.1	19.4	25.5	4.3	2.7
	80	45.5	97.9	19.7	26.7	4.2	1.8
($\mathscr{P}_2$)	25	21.4	4.3	11.1	22.5	37.7	11.1
	36	131.7	104	13.4	23.8	15.4	7.7

Table 5.7: Performance of the PWL approx. MILP (($\mathscr{P}_2$)) (top) and 2D-PWL approximation MILP ($\mathscr{P}_2$) (bottom) with increasing segments.

points p from the N functions g_v, giving the probability of detection P_v for m_r possible effort values for the current round and m_{r-1} effort values for the previous rounds of patrol:

$$D_g := \left\{ p = \left\langle C^r_{v,i}, C^{r-1}_{v,j}, g_v(C^r_{v,i}, C^{r-1}_{v,j}) \right\rangle : \begin{array}{l} \forall v \in N \\ i \in \{1, m_r\} \\ j \in \{1, m_{r-1}\} \end{array} \right\}$$

Using this dataset D_g we can construct our PWL approximation by representing any set of feasible patrol efforts (c^r, c^{r-1}) and corresponding predicted detection of attack $g(c^r, c^{r-1})$ as a convex combination of their nearest neighbors in the dataset D_g.

At round r we already have data on the previous $r-1$ round's patrolling effort at each cell which we denote $\tilde{c}^{r-1}_v$. We use the notation capital $C \subset p \in D_g$ to denote patrol effort data used to construct the piecewise linear objective and lowercase $\tilde{c}$ to denote known past patrol effort data. Because the $\tilde{c}^{r-1}_v$ are known we can directly express them as a convex combinations of the closest two data points (C^+, C^-) so that $\tilde{c}^{r-1}_v = \lambda^{r-1}_v C^+_v + (1 - \lambda^{r-1}_v) C^-_v \ \ \forall v \in N$. We want to plan patrols for the current round r, meaning that for the patrol effort c^r_v, we do not know beforehand what the two closest data points in D_f will be. Instead we express c^r_v as a convex combination of *all* points p_c, and constrain the weights λ^r on the points to belong to a *Specially Ordered Set of Type*

2 (SOS2) which are an ordered set of variables where at most two consecutive variables may be non-negative. The objective function of MP ($\mathscr{P}$) can then be expressed as:

$$\begin{aligned} g_v(c_v^r, \bar{c}_v^{r-1}) &= \sum_i \lambda_{v,i}^r \Big(\lambda_v^{r-1} g_v(C_{v,i}^r, C_v^+) \\ &\qquad + (1 - \lambda_v^{r-1}) g_v(C_{v,i}^r, C_v^-) \Big) \\ &= \sum_i \lambda_{v,i}^r \tilde{g}_v(C_v^r) \end{aligned} \tag{5.1}$$

Where we add additional constraints:

$$\begin{aligned} &\sum_{i \in [m_r]} \lambda_{v,i}^r C_{v,i}^r = c_v^r && \forall v \in N \\ &\lambda_{v,i}^r \in \text{SOS2} && \forall v \in N, i \in [m_r] \\ &\lambda_{v,i}^r \in [0,1] && \forall v \in N, i \in [m_r] \\ &\sum_{i \in [m_r]} \lambda_{v,i}^r = 1 && \forall v \in N \end{aligned} \tag{5.2}$$

So that MP ($\mathscr{P}$) is now expressible as a Mixed Integer Linear Program which we refer to as MILP ($\mathscr{P}_1$).

Two Stage Planning: Given that the predictions are functions of past and current patrol, we have the ability to plan for multiple rounds of patrolling. We want to maximize the probability of detecting an attack in two rounds, r and $r+1$, ie. $\sum_{v \in N} g_v(c_{v,i}^r, \tilde{c}_v^{r-1}) + \sum_{v \in N} g_v(c_{v,i}^{r+1}, c_v^r)$. We already know how to construct the PWL approximation of $g_v(c_{v,i}^r, c_v^{r-1})$ since we have $\tilde{c}_v^{r-1}$ as data; however both c_v^r and c_v^{r+1} are variables and must be expressed as convex combinations of points in $\mathscr{D}$ using the same type constraints as (5.2). We can then express the tuple (c_v^r, c_v^{r+1}) as a convex combination of the 4 closest points in $\mathscr{D}$ with weights $\Lambda_{i,j}$ using from the weights $\lambda_{v,i}^r$ and $\lambda_{v,j}^{r+1}$ with the following constraints $\sum_i \Lambda_{i,j}^v = \lambda_{v,j}^{r+1}$ $\forall v \in N, j \in [m_{r+1}]$ and $\sum_j \Lambda_{i,j}^v = \lambda_{v,i}^r$ $\forall v \in N, i \in [m_r]$. With these we are guaranteed to have only 4 non-zero $\Lambda_{i,j}^v$ since there are only 2 non-zero $\lambda^r_{v,i}$ and $\lambda^{r+1}_{v,j}$. The two stage optimization problem, MILP ($\mathscr{P}_2$) can then be expressed as:

$$\max_{\lambda,\Lambda} \quad \sum_{v\in N}\sum_{i,j}\Lambda^v_{i,j}g(C^{r+1}_{v,i},C^r_{v,j})+\sum_i \lambda^r_{v,i}\tilde{g}_v(C^r_{v,i})$$

$$\begin{aligned}
& f^r_{u',v'}, f^{r+1}_{u',v'} \in \mathscr{F} && \forall (u',v') \in E' \\
& K\textstyle\sum_{u':(v',u')\in E'} f^r_{u',v'} = c^r_v && \forall v \in N, v'=(v,t) \\
& \textstyle\sum_{v\in N} c^r_v = T\times K && \\
& && \\
& \textstyle\sum_i \lambda^r_{v,i}C^r_{v,i} = c^r_v && \forall v\in N \\
& \textstyle\sum_i \lambda^{r+1}_{v,i}C^{r+1}_{v,i} = c^{r+1}_v && \forall v\in N \\
& \textstyle\sum_i \Lambda^v_{i,j} = \lambda^{r+1}_{v,j} && \forall v\in N, j\in[m_{r+1}] \\
& \textstyle\sum_j \Lambda^v_{i,j} = \lambda^r_{v,i} && \forall v\in N, i\in[m_r] \\
& && \\
& \lambda^r_{v,i}, \lambda^{r+1}_{v,j} \in \text{SOS2} && \forall v\in N, i\in[m_{r+1}], j\in[m_{r+1}] \\
& \lambda^r_{v,i}, \lambda^{r+1}_{v,j}, \Lambda^v_{i,j} \in [0,1] && \forall\forall v\in N, i\in[m_r], j\in[m_{r+1}] \\
& \textstyle\sum_i \lambda^r_{v,i} = \sum_i \lambda^{r+1}_{v,i} = 1 && \forall v\in N
\end{aligned} \tag{$\mathscr{P}_2$}$$

5.4.1 Evaluation of the Prescriptive Model

Using the predictions made on the QEPA and MFPA datasets we generated patrols for each of the patrol posts in both national parks. Samples of these can be see in Figure 5.8 where we show a heat map of the patrol effort corresponding to the distribution over computed patrols around posts for both the protected areas. These are currently being evaluated for real world deployment in both QEPA an MFPA. To evaluate the piecewise linear approximation with iWare-E prediction model, we look at the expected total predicted detections of illegal activity of the patrol schedules generated by the MILP. Given an optimal solution we can compute the actual predicted number of detected attacks using the iWare-E model. We then compare this prediction to the optimal objective value of the MILP used to compute c. These results are shown in Table 5.7 under the error column, where we measure the percent difference in these two values, averaged over all posts in the protected area. We see that we can get low approximation error when using the piecewise linear objective.

We also show the importance of being able to reason about continuous levels of patrol effort, where in Figure 5.9 we show the significant improvement in utility of the patrols computed, measured in terms of number of predicted detected attacks. For this comparison we let each breakpoint in the PWL approximation correspond to a discrete level of patrol effort and compared the number of predicted detections of both solutions. We see that even as we increase the number of levels of patrol effort to 80 levels, we still outperform the previous state-of-the-art by approximately 130% for the QEPA dataset and 150% for the MFPA dataset. Additionally the PWL objective allows us to be much more scalable; as an example, the previous state-of-the-art method requires 80 levels of discretization to achieve the same average utility (in terms of predicted detections) as 10 levels for QEPA and 5 for MFPA. This difference in discretization results in 400×, and 140× decrease in runtime for QEPA and MFPA respectively when using our method. We show similar improvements in runtime for more of these fixed utility comparison points in Figure 5.10 where it can be see that it takes significantly more computational power for the previous state-of-the-art to match our results.

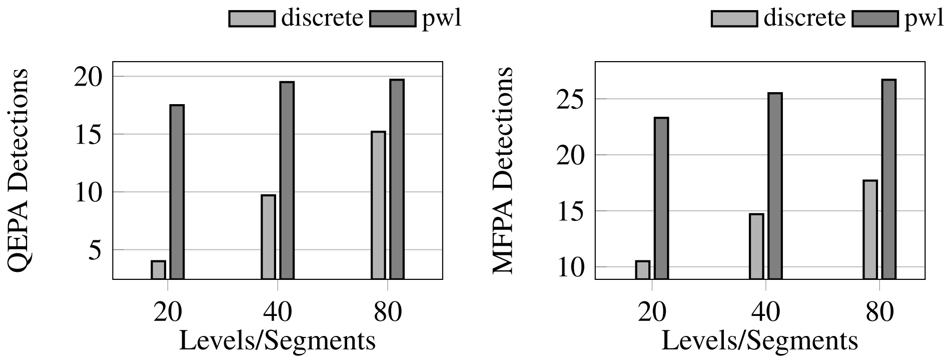


Figure 5.9: Improvement in solution quality of patrols planned for the QEPA and MFPA, using MILP ($\mathscr{P}_2$) compared to previous work using discrete levels of patrol effort. The utility is measured in number of predicted detected attacks.

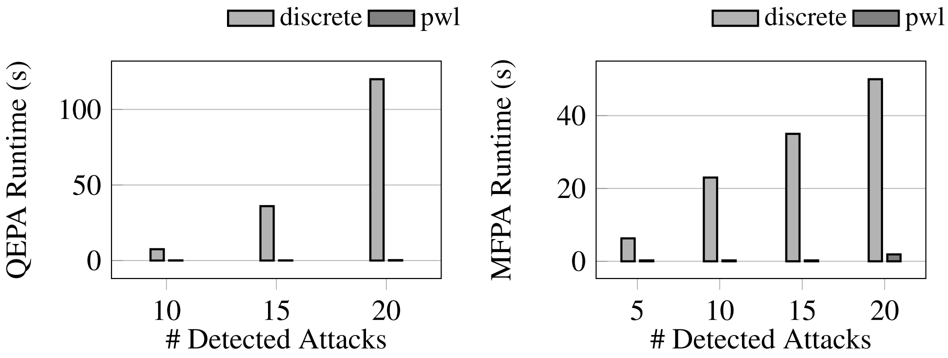


Figure 5.10: Improvement in runtime for computing patrols for QEPA and MFPA, using MILP($\mathscr{P}_2$) compared to previous work using discrete levels of patrol effort for comparable solution quality, measured in predicted number of detected attacks. Results are averaged over 20 trials.

Figure 5.11: Park rangers in SWS with snares they removed during field tests in December 2018. Photo: WWF Cambodia.

5.5 Uncertainty Quantification for Adversarial Behavior Predictions

Prescribing patrol plans to park rangers requires in-depth knowledge of the poachers' behavior. Learning the poachers' behavior is a challenging machine learning problem since (a) the wildlife crime datasets are usually extremely imbalanced, with up to 99.5% negative labels; (b) negative labels indicating absence of illegal activity are not reliable due to the inherent difficulty of detecting well-hidden poaching signs in the forest, as shown in Figure 5.12; (c) historical crime observations are not collected thoroughly and uniformly in many protected areas, so datasets suffer from potential biases; and (d) poaching patterns and landscape features vary from one protected area to another, so a universal predictive model cannot be recommended.

In the following subsections, We improve upon the previously proposed methodology by addressing these four challenges. To do so, we use Gaussian processes to quantify uncertainty in predictions of poaching risk and exploit these uncertainty metrics to increase the robustness of our prescribed patrols. We evaluate our approaches on historical crime data from three real-world datasets from Uganda and Cambodia, which have different characteristics from both ecological and data quality perspectives.

The historic patrolling and poaching data we used for both MFNP and SWS was collected over the years using SMART, a monitoring and reporting tool for protected area management [85].

Figure 5.12: Well-hidden snares detected and removed by park rangers during our field tests in MFNP. Photo: Uganda Wildlife Authority.

SMART is developed by a consortium of leading conservation organizations and used in more than 600 protected areas across 55 countries. Although SMART records significant amounts of historical data, machine learning techniques have not been systematically applied to predict poacher behavior. In the coming year, PAWS will be integrated into the SMART software and become available to park managers around the world. The enhancements and field tests in this paper outline significant steps to expanding PAWS on a global scale.

5.5.1 Enhancements to Address Uncertainty in Predictions

A key challenge we address in this study is the need to explicitly reason about the uncertainty in the model predictions. Hence, we augment the iWare-E method with Gaussian process (GP) classifiers [79] as the weak learners:

$$f(\mathbf{x}_i) \sim \mathcal{GP}(\mu(\mathbf{X}), \Sigma(\mathbf{X})) \ , \tag{5.3}$$

with mean $\mu(\mathbf{X})$ and covariance matrix $\Sigma(\mathbf{X})$. Due to the formal definition of the covariance functions, GPs enable us to explicitly quantify the uncertainty of predictions as a percentage. We then use these uncertainty values to make more informed decisions in patrol planning.

In the iWare-E ensemble method, a classifier trained on data at threshold θ_i can make predictions on a new observation with effort at threshold θ_j if there is a value in the corresponding vote matrix $\mathbf{V}^{qp}$, that is, if $\mathbf{V}^{qp}[i, j] > 0$. [27] proposed a voting matrix with values only in the upper triangular matrix, so classifiers trained on data with high patrol effort would only predict on data with equal or higher patrol effort. However, our experiments showed that we could improve the approach. We achieved the best results by setting the voting matrix as a lower triangular matrix, which means that each classifier predicts on data points with patrol effort equal or less than data it was trained on. For example, the final classifier with threshold θ_I will be used in all predictions. This finding aligns with the intuition that higher classifiers provide better predictions, as they have been trained on higher quality data: negative labels at high patrol effort are more reliable to be true negatives.

In picking thresholds θ_i for patrol effort, [27] used 16 equally-distanced values from $\theta_1 = 0$ to $\theta_I = 7.5$. However, we found that the best approach was to select these thresholds based on patrol effort percentiles, to produce a consistent amount of training data for each classifier. This approach then simplifies the model such that the number of classifiers becomes a single hyperparameter, rather than having to pick $\theta_{\min}$, $\theta_{\max}$, and $\Delta\theta$ separately. The number of classifiers should be selected based on the characteristics of the data; we used more classifiers (20) for QENP and MFNP, which have well-behaved label imbalance (see Figure 5.14), than for SWS (10). In addition, selecting thresholds based on percentile better accounts for sparsity in the data: there may be very few cells patrolled with effort between 5 and 6, and those points may all be negative labels.

5.5.2 Evaluation on Historical Data

We study Murchison Falls National Park (MFNP) and Queen Elizabeth National Park (QENP) in Uganda, and Srepok Wildlife Sanctuary (SWS) in Cambodia. These protected areas cover 5000 sq. km, 2500 sq. km and 4300 sq. km, respectively. MFNP and QENP are critically important for ecotourism and conservation in Uganda, and provide habitat to elephants, giraffes, hippos, and lions [19]. SWS is the largest protected area in Southeast Asia and is home to leopards, bears, and banteng. SWS once housed a native population of tigers, but they fell prey to poaching; the last tiger was observed in 2007. In the intervening decade, SWS has been identified as a promising site for tiger reintroduction [33]. Effectively managing the landscape by reducing poaching will be critical to successful reintroduction.

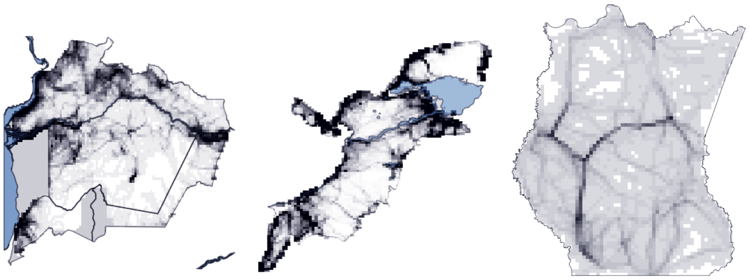

(a) MFNP, Uganda (b) QENP, Uganda (c) SWS, Cambodia

Figure 5.13: The protected areas used in this study. Visualized are the aggregate patrol effort for each protected area.

To combat poaching, park rangers conduct patrols through protected areas and use GPS trackers to mark their observations. They confiscate animal traps, rescue live animals caught in snares, and arrest any poacher they encounter [20]. Their GPS trackers are connected to the SMART system, recording many years of wildlife crime data [85]. However, the data are biased due to the inability of park rangers to detect all instances of poaching. There are additional data collection issues due to the nature of these patrols: rangers may have to address an emergency in the field, such as hearing a poacher in the distance, and lose the opportunity to record snares or bullet cartridges they found.

We study SWS in Cambodia for the first time and conduct a month-long field test. This park introduces many challenges. (a) It is particularly under-resourced, with only a few dozen rangers enforcing an area the size of Rhode Island. (b) Unlike MFNP or QENP where rangers conduct foot patrols, park rangers in SWS travel by motorbike. This faster form of transport means that waypoints in the dataset (typically recorded once every 30 minutes) are even more sparse, making it difficult to interpolate their trajectory between sequential points. Additionally, negative labels in the dataset are likely less reliable because patrollers are less able to carefully observe their surroundings when traveling quickly. (c) SWS experiences strong seasonality: many rivers are impossible to cross during the wet season, but dry up during the dry season.

Patrol observations come from SMART conservation software, which records the GPS location of each observation along with date and time, patrol leader, and method of transport. Park rangers enter their observations: animals or humans spotted; signs of illegal activity such as campsites or cut trees; and signs of poaching activity such as firearms, bullet cartridges, snares, or slain animals. We categorize these observations into poaching and non-poaching. Additionally, we rebuild historic patrol effort from these observations by using sequential waypoints to calculate patrol trajectories.

We augment these patrol observations with geospatial features about each park, provided as GIS shapefiles from the data specialists at WWF, WCS, and UWA. The features differ between parks, but typically include terrain features such as rivers, elevation maps, and forest cover; landscape features such as roads, park boundary, local villages, and patrol posts; and ecological features such as animal density and net primary productivity. We use these static features to build data points in our predictive model, either as direct values (such as slope) or as distance values (such as distance to nearest river). We do not explicitly encode longitude or latitude as features, which could be extrapolated through the various distance values.

To study the data, we discretize the protected areas into 1×1 km grid cells. Each cell is associated with the geospatial features described above. We partition time into three-month time intervals, which allows us to capture seasonal trends and corresponds to approximately often park rangers plan new patrol strategies. The dynamic features in our dataset are patrol effort and poaching activity. For each time interval, we aggregate the patrol effort at each cell and assign a positive label $(y = 1)$ if park rangers observed poaching-related activity during that time period. We compute patrol effort as the distance walked by park rangers across a cell. Ideally, we would be able to encode the time spent in each cell to measure patrol effort, as walking slowly through an area likely increases chance of detecting illegal activity. However, rangers typically record waypoints only once every 30 minutes and activities such as lunch breaks or setting up camp overnight are not identified.

We build the datasets $\mathscr{D} = (\mathbf{X}, \mathbf{y})$ based on these historical patrol observations. The records are discretized into a set of T time steps and N locations to create a matrix $\mathbf{X} \in \mathbb{R}^{T \times N \times k}$, where k is the number of features. Each feature vector $\mathbf{x}_{t,n}$ contains multiple time-invariant geospatial features associated with each location and two time-variant covariates: $c_{t-1,n}$, the amount of patrol coverage during the previous time step $t-1$, which models the potential deterrence effect of past patrols; and $c_{t,n}$, the patrol effort in the current time step, which models the effort devoted to the cell during time t. Additionally, we have the observation vector $\mathbf{y} \in \{0,1\}^{T \times N}$ as a binary encoding of whether any illegal activity was detected at each data point (t,n).

We generate predictive poaching models with four years of data for each park, training on the first three years and testing on the fourth. This setup simulates the ability of each model to

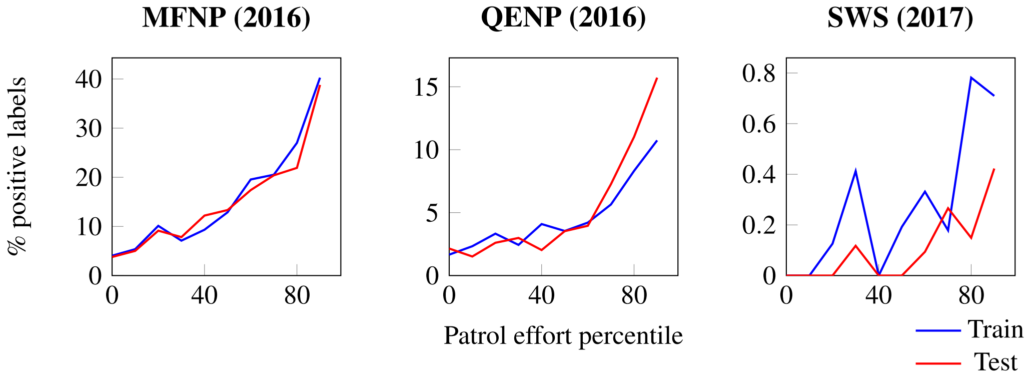


Figure 5.14: Percentage of positive labels at different thresholds of patrol effort. The year in parentheses is used for the test set; the previous three years of data comprise the training set. Note that the *y*-axis varies drastically between datasets.

predict future incidences of poaching. Although we have up to 18 years of data for each park, earlier years are increasingly less reliable: a park may hire more rangers; the sale price of illegal wildlife goods may increase, making poaching more attractive; or logging may increase in a region, thus changing the landscape.

The dataset from SWS created new challenges of class imbalance. Observe in Table 5.8 the extreme label imbalance in SWS for test year 2017, with only 0.3% positive labels in the training set and 0.1% in testing. Given these new challenges of significantly imbalanced data, we used a balanced bagging classifier to undersample negative labels [37,55]. This undersampling approach improved our AUC by 10% on average. With datasets in the wildlife crime domain, undersampling is preferred to oversampling the minority class because the positive labels are inherently more noisy due to random factors that influence whether rangers detect poaching activity.

We implement the iWare-E ensemble method with three base classifiers: bagging ensembles of SVMs (SVB-iW), bagging ensembles of decision trees (DTB-iW), and bagging ensembles of Gaussian process classifiers (GPB-iW). We compare these models to baseline models, using those same weak learners but without iWare-E (referred to as SVB, DTB, and GPB).

		number of points		percent positive	
		train	**test**	**train**	**test**
MFNP	**2014**	9,254	4,285	14.0%	16.3%
	2015	11,657	2,661	15.4%	12.6%
	2016	11,150	2,055	15.6%	13.4%
QENP	**2014**	10,541	2,755	4.3%	3.7%
	2015	9,436	3,335	4.2%	6.1%
	2016	9,025	3,233	4.6%	5.6%
SWS	**2016**	18,534	7,636	0.2%	0.4%
	2017	20,664	8,491	0.3%	0.1%
	2018	23,805	11,166	0.3%	0.7%

Table 5.8: About the datasets: percentage of positive labels

		without iWare-E			with iWare-E		
		SVB	**DTB**	**GPB**	**SVB**	**DTB**	**GPB**
MFNP	**2014**	0.52	0.59	0.63	0.69	0.72	0.72
	2015	0.51	0.61	0.66	0.68	0.71	0.71
	2016	0.52	0.60	0.62	0.66	0.71	0.71
QENP	**2014**	0.50	0.68	0.69	0.60	0.72	0.64
	2015	0.50	0.59	0.60	0.62	0.70	0.71
	2016	0.50	0.64	0.60	0.64	0.74	0.74
SWS	**2016**	0.81	0.80	0.78	0.76	0.72	0.68
	2017	0.68	0.71	0.73	0.86	0.83	0.82
	2018	0.51	0.53	0.55	0.67	0.69	0.71

Table 5.9: Comparing performance (AUC) of each model across all three datasets

The performance of the different predictive models evaluated on MFNP, QENP, and SWS for various choices of weak learners used in iWare-E is presented in Table 5.9[2]. The year listed is the test set; for example, MFNP (2016) indicates that 2013–2015 were used for training and 2016 for testing. We conducted three experiments on each of the parks: with test sets 2014, 2015, and 2016 for MFNP and QENP, and 2016, 2017, and 2018 for SWS. The iWare-E approach consistently improves AUC across all models. In general, SVMs are suboptimal weak learners in this domain, and decision trees and GPs have comparable performance. Thus, introducing GPs to the iWare-E method does not inhibit performance. As we will show, the ability of GPs to estimate uncertainty provides important advantages in patrol planning that make it the preferred weak learner.

Algorithm runtime for all models is listed in the online appendix[2]. Overall, DTB-iW is the most efficient and runs nearly instantly, and SVB-iW completes in under two minutes. GPB-iW is significantly more computationally expensive, requiring up to 7 hours. To improve runtime with GPB and GPB-iW, we used bagging classifiers on small subsets of the data with minimal trade-off in performance.

5.5.3 Uncertainty in Crime Prediction

We analyze the predicted probabilities and the uncertainties associated with the predictions of the GPB-iW model. By construct, in the Gaussian process model, each probability score is associated with a variance value which indicates the amount of uncertainty in the prediction. Figure 5.15 depicts the joint probability of detection of attack and attack $\Pr[o = 1, a = 1]$, generated by GPB-iW, for different levels of patrol effort in MFNP. For instance, the middle plot shows the predicted probability of detecting an attack if each 1 sq. km area in the park is patrolled with 1 km of patrol

[2] Additional performance metrics andavailable in the online appendix: `https://www.dropbox.com/s/s46pz94sxsyoz8h/KDD19_PAWS_appendix.pdf`

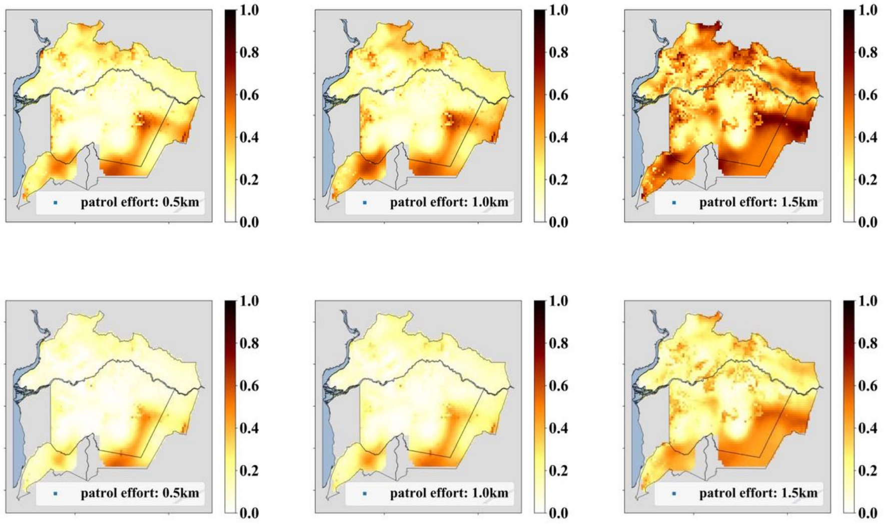


Figure 5.17: Prediction values and uncertainties for different levels of patrol effort in MFNP.

effort by park rangers during three months of patrolling. Although the predicted probability for many cells increases as patrol effort increases, several cells show a zero or near-zero change in predicted probability. Such observations indicate that there is either no potential attack ($\Pr[a = 1] \simeq 0$) or no high-quality patrol coverage in those regions.

Figure 5.16 presents the corresponding uncertainties in the predictions from Figure 5.15. These uncertainty heatmaps show the model's confidence about each prediction. For example, the southeast region of the park suffers from the highest amount of uncertainty in the predictions. As expected, the past patrolling data (shown in Figure 5.13(a)) reveals that patrolling has been minimal in this region, due in part from few large mammals and lack of patrol posts. Since this region has the least historic data, the predictions are the least certain. This information is valuable for robust defender patrol planning which strategizes against the worst-case attack scenarios, and could also be used to plan patrol routes that explicitly target areas with high model uncertainty in order to reduce the existing data bias.

5.6 Patrol Planning Model Enhancement

5.6.1 Game Model

A wildlife conservation area can be discretized into a set of N grid cells (corresponding to 1×1 km regions) to form a graph $G(V,E)$ of nodes and edges. We model the problem of allocating park patrols as a game on this graph, played between a defender (park rangers) and a set of N adversaries (poachers), located at each of the grid cells. Each adversary may choose to attack their grid cell by placing snares to catch animals. The rangers attempt to thwart these attack by conducting patrols in order to detect the snares, and receive a payoff of 1 for every attack successfully thwarted. Rangers conduct their patrols in the conservation area over a period of T time steps, where a single time step corresponds to the minimum amount of time it would take to cross one grid cell (so that rangers must spend at least one time step in each grid cell they choose to visit). A ranger's patrol is then a path taken on a time unrolled graph $G(V',E')$, with a set of nodes indicating location and time. The set of possible paths on the time unrolled graph forms the pure strategy space of the defender.

Each patrol must begin and end at nodes in the graph designated as patrol posts, which we refer to as the source $s \in V$ for graph $G(V,E)$. In the time unrolled graph we designate the source of any patrol as $s_1 = (s,1) \in V'$ and the target $s_T = (s,T) \in V'$ corresponding to the patrol post visited in the first and last time steps. The ability of park rangers to detect poaching in any cell $v \in V$ depends on the level of patrol effort $c_{t,v}$ in that cell at time t. Rangers may increase the level of patrol effort at any cell by either spending multiple time steps at the cell in a single patrol or visiting that cell during multiple distinct patrols. We specify a set of constraints, so that a single feasible patrol corresponds to one unit of flow on the time unrolled graph; the sum of the total flow over all edges in G' is then equal to T. The pure strategy space is the set of such flows $\mathscr{F}$, given by

$$\mathscr{F} := \left\{ f_{u',v'} : \begin{array}{l} \sum_{u':(u',v')\in E'} f_{u',v'} = \sum_{u':(v',u')\in E'} f_{v',u'} \ \ \forall v' \in N' \\ \sum_{u':(s_1,u')\in E'} f_{s_1,u'} = \sum_{u':(u',s_T)\in E'} f_{u',s_T} = 1 \\ \sum_{(v',u')\in E'} f_{u',v'} = T \end{array} \right\}.$$

Let x_v be the defender coverage at cell v and a_v be the action of the adversary located at grid cell v. Each adversary responds to the mixed strategy coverage c_v of the defender at that cell and chooses at the beginning of the game either to attack $a_v(x_v) = A$ or not attack $a_v(x_v) = \neg A$ with some probability $\Pr[a_v(x_v)]$ such that $\Pr[a_v(x_v) = A] + \Pr[a_v(x_v) = \neg A] = 1$. The defender utility is conditioned on a successful detection of attack. Let $o_v(x_v) = O$ denote a successful detection of snares at grid cell v. Note that the detection success probability is also a function of the defender's mixed strategy coverage. The defender expected utility is then the probability of detecting snares at a grid cell, given that there is an attack at the cell, summed over all cells $U^d(a,x) = \sum_{v\in N} \Pr[o_v = O \mid a_v = A] \Pr[a_v = A]$. Note that we have omitted the dependence on x_v for ease of notation. We compute an equilibrium solution to this game where the defender attempts to maximize her utility function $U^d(a,x)$ and each of the adversaries maximizes their own utility function $U^a_v(a_v, x_v)$. If the adversaries were perfectly rational, this would correspond to a strong Stackelberg equilibrium with each of the adversaries as in standard Stackelberg security games. However, in Green Security Games, the adversaries are boundedly rational and we learn adversary behavior models from data, as described earlier in the study. Thus, we achieve equilibrium with non-rational attackers as in [71, 98].

5.6.2 Prescriptive Modeling with Certain Crime Predictions

The adversary utility and the probability of detecting snares given that the adversary chooses to attack are both unknown functions. We use past data to learn a predictive model of the adversary's response to the defender mixed strategy x as well as the defender detection probability. To compute solutions to the patrol planning game, we then need to optimize this predictive model, which we treat as a black box function. We use a similar framework as in [27] to perform this optimization, which allows us to not only plan with a black box objective function, but also to reason about continuous decision variables such as patrol effort. The model uses a piecewise linear objective function to approximate predictions of the model. This formulation was shown to have significant improvements in runtime compared to previous methods for patrol planning with a black box function, which could only reason about discrete levels of patrol effort.

The predictive model produces, for each cell, a function $g_v : c_v \to P_v$ which maps the total defender patrol effort at a particular grid cell $v \in N$ to a corresponding likelihood that there will be a detected attack P_v at that grid cell. The defender patrol effort is a function of the defender mixed strategy $c_v = x_v K$ where K is the number of patrols that the defender conducts. As in [26], piecewise linear (PWL) approximations to these functions g_v are constructed using $m \times N$ sampled points from the N functions g_v.

These define the optimization problem ($\mathscr{P}$) which can be expressed as a mixed integer linear program (MILP):

$$\begin{aligned} \max_{c,f} \quad & \sum_{v \in N} g_v^{\text{PWL}}(c_v) & & \\ & f_{u',v'} \in \mathscr{F} & \forall (u',v') \in E' & \\ & K \textstyle\sum_{u':(v',u') \in E'} f_{u',v'} = c_v & \forall v \in N, v' = (v,t) & \\ & \textstyle\sum_{v \in N} c_v = T \times K & & \end{aligned} \qquad (\mathscr{P})$$

Our objective function is given by the PWL approximation to the machine learning model predictions, which we refer to as g_v^{PWL}. The first constraints are the flow constraints on the time unrolled graph $G(N',E')$. The second constraint enforces that the patrol effort is equal to the amount of flow into a node v across all time (which gives the defender mixed strategy coverage x_v) times the number of patrols K. The last constraint enforces that the total patrol effort expended is equal to the length of each patrol T times the number of patrols K.

5.6.3 Prescriptive Modeling with Uncertain Crime Predictions

We use predictions from the GPB-iW model to plan patrol routes for park rangers. We only have black box access to the predictions as a function of the rangers patrol effort, which we need to be able to optimize in order to compute patrol paths. The planning model in [27] allows us to create paths, but does not account for uncertainty in the predictions and does not have a game theoretic component. To account for this, we augment the model by using the variance associated with each of the predictions from the GPB-iW model.

The new GPB-iW machine learning model now gives for each grid cell a variance function $v_v : c_v \rightarrow \mathscr{V}_v$, where $\mathscr{V}_v$ becomes the uncertainty score for each prediction $g_v(c_v)$, with likelihood P_v

that there will be a detected attack at that cell. We want to compute a series of patrols which not only maximize the probability of detecting attacks over the entire area, but also takes into account the uncertainty of each prediction. To do this, we take a robust approach by penalizing the expected probability of detection given by g_v by a scaled function of the uncertainty score ν_v. The utility of patrolling any grid cell v as a function of the patrol effort c_v is:

$$U_v(c_v) = g_v(c_v) - \alpha \nu_v(c_v) \ . \tag{5.4}$$

The uncertainty scores that we get from the GPB-iW model are scaled to the range $[0,1]$ through a logistic squashing function. We then choose $\alpha = \beta g_v(c_v)$, with $\beta \in [0,1]$ to rescale the uncertainty score and ensure that the objective function is always positive. Larger values of β allow us to optimize for plans which are more risk averse; β thus becomes a parameter that enables us to tune the robustness of our approach. $\beta > 0$ corresponds to a pessimistic approach where we will patrol less in cells where there is greater uncertainty. We can compute the optimal patrol by substituting $\sum_{v \in N} U_v^{\text{PWL}}(c_v) = \sum_{v \in N} g_v^{\text{PWL}}(c_v) - \alpha \nu_v^{\text{PWL}}(c_v)$ as our objective function in $\mathscr{P}$, where, as in [27] we construct PWL approximations to g_v and ν_v so that the optimization problem is expressible as a MILP.

5.6.4 Evaluation of the Prescriptive Model

To demonstrate the benefit of accounting for uncertainty when planning patrols, we compare the patrols computed with and without uncertainty scores by evaluating them on the ground truth given by the objective with uncertainty. We refer to the plan computed using uncertainty weighted with value β as $C_\beta = \text{argmax}_c \sum_v g_v(c) - \beta g_v \nu_v(c)$, such that $C_{\beta=0}$ is a plan which does not

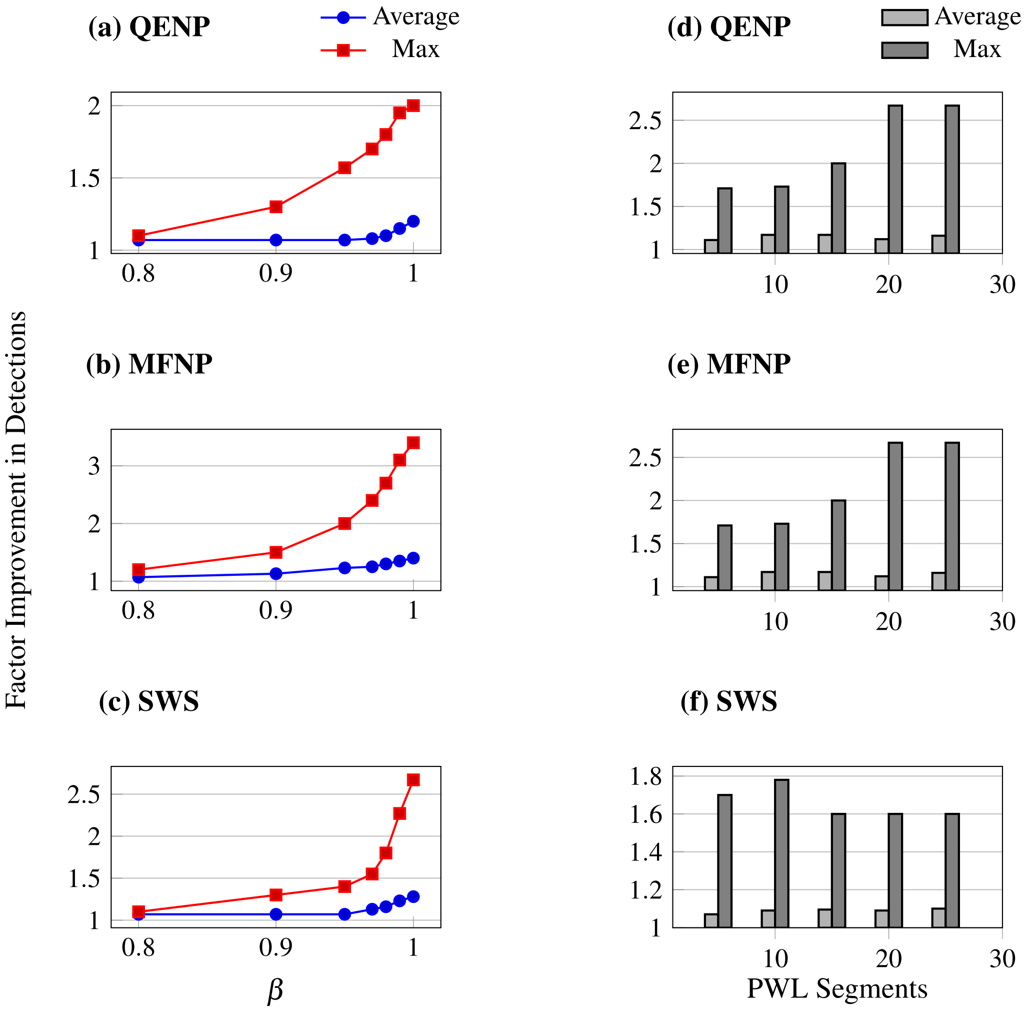


Figure 5.18: Improvement in detection of snares when accounting for uncertainty in patrol planning. Figures (a)–(c) show the improvement in solution quality measured by the ratio $U_\beta(C_\beta)/U_\beta(C_{\beta=0})$ as a function of the tuning parameter β which determines the robustness of the solutions through the weight on the uncertainty score. Figures (d)–(f) show improvement in solution quality with increasing segments in the PWL approximation to the GPB-iW predictions.

account for uncertainty and $C_{\beta=1}$ is a fully robust plan. We then evaluate each of the plans using a utility function $U_\beta(C)$ and compute the ratio of the solution quality of the plan at a given β to the baseline of $\beta = 0$, $U_\beta(C_\beta)/U_\beta(C_{\beta=0})$. For each dataset, we looked at the gain in solution quality for plans generated for each patrol post in the park, evaluated by varying the β parameter as well as the number of segments in the piecewise linear functional approximations to U_v. The results are shown in Figure 5.18, where we consider both the average gain over all patrol posts as well as the maximum gain achieved. Not only does this demonstrate the benefit of robust planning, but rangers may also use this knowledge to help decide where to collect more data. By looking at the ratio $U_\beta(C_\beta)/U_\beta(C_{\beta=0})$ we can determine whether reducing the uncertainty in the predictive model will affect the computed plans. When this ratio is close to 1 there is little benefit in collecting data in order to reduce uncertainty in predictions as it does not change the utility of the computed patrols. However, when the ratio is large, the computed patrols are highly dependent on the areas where the model is uncertain; therefore there can be significant gains in utility through the reduction of uncertainty through collecting more data. We also demonstrate the scalability of this approach in Figure 5.19a, which shows the runtime with increasing number of breakpoints in the PWL function approximation. Figure 5.19b plots the convergence of the utility of the computed solutions. We looked at the utility of the robust solutions $U_{\beta=1}(C_{\beta=1})$ with increasing number of segments in the PWL function and see that utility of the optimal solution converges around 20–25 segments.

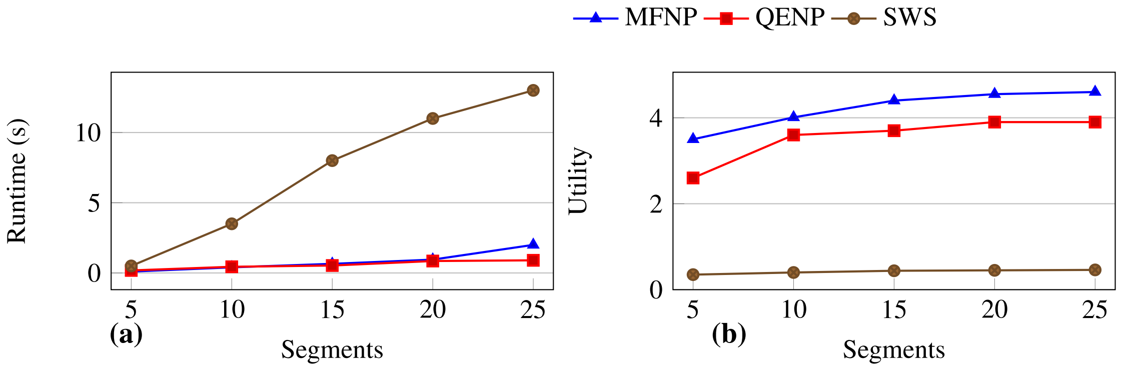


Figure 5.19: Prescriptive model runtime (a) and patrol plan utility (b) as a function of the number of segments in the PWL function approximation to the GPG-iW model.

5.7 Conclusion

To make an impact in wildlife protection, it is crucial to adopt predictive and prescriptive models in the real fields. Previous works suffer from addressing the major technical and application challenges in this domain. In this study, we present iWare-E, an efficient predictive model for wildlife protection, which accounts for imperfect crime information and uncertainty in wildlife data. This is the first time that this substantial challenge is addressed in data-driven adversarial reasoning in AI literature. Furthermore, we presented less computationally expensive fine-tuned generation of patrol routes based on the predictions of iWare-E to counteract poachers in the real-world more effectively (150% improvement in solution quality and 400 times higher speed). From domain perspective, previous works consider the coarse-grained temporal analysis of crime observations and they only evaluate on a single protected area. However, the predictive framework proposed in this study significantly improves accuracy and runtime even for fine-grained analysis of crime over multiple protected areas. To our knowledge, this is the first adversary behavior model for wildlife protection that has been developed and evaluated at this scale in two protected areas to prove country-wide reliability in prediction results. Such predictive and prescriptive analysis can be invaluable for assisting law enforcement agencies in protecting wildlife more intelligently.

Chapter 6

Patrol Route Planning with Imperfect Prior Knowledge

In this chapter, I discuss a game theoretical patrol planning algorithm when an imperfect adversarial model learned from real-world data is available. First, I propose MINION-sm, a novel online learning algorithm for Green Security Games, which does not rely on any prior error-prone model of attacker behavior, instead, it builds an implicit model of the attacker on-the-fly while simultaneously generating scheduling-constraint-aware patrols. MINION-sm achieves a sublinear regret against an optimal hindsight patrol strategy. Second, I propose MINION, a hybrid approach where our MINION-sm model and an ML model (based on historical data) are considered as two patrol planning experts and I obtain a balance between them based on their observed empirical performance. Third, I show that our online learning algorithms significantly outperform existing state-of-the-art solvers for Green Security Games.

6.1 Problem Domain

Security games are well known to be effective models of protecting valuable targets against an adversary and have been explored extensively at AAMAS [5,45,52,64]. Recently, there has been a lot of progress in the field of Green Security Games (GSGs), which has led to the development of several algorithms which serve as game-theoretic decision aids to optimize the use of limited human patrol resources to combat poaching [27, 42, 72]. The basic premise behind most of this work is that repeated interactions between patrollers and poachers provides the opportunity to gather data which can be used to learn models of poacher behavior [22]. Thus, most previous algorithms design patrol routes assuming poachers attack according to a fixed ”*learnable*” model (which could either have a functional form [22, 72], or it could be a black-box model [27, 95]). Most of these algorithms then try to solve a repeated Stackelberg game, where the patrollers (defenders) conduct randomized patrols against poachers (attackers) while balancing the priorities of different locations in the park. Unfortunately, this approach suffers from serious shortcomings, which impedes usability in the real-world.

In particular, the GSG approach can be expected to provide good results only if the collected historical data is a good representation of the actual poaching activities that occurred in the past (and those that will occur in the future), which would allow us to learn an accurate model for attacker behavior. Unfortunately, in the wildlife poaching domain, it is extremely difficult to know ahead of time whether the learned model of attacker behavior is accurate or not (over the entire protected area). Due to logistical issues, several patrollers only conduct patrols either close to their sparsely spread patrol posts, or in areas that are easily accessible by them. This issue is so prevalent that it has a special name in ecological research: the silent victim problem [56]. As a result, the poaching data collected in these domains may be highly biased (in a spatial sense). For example, Figure 6.1 shows the patrol coverage heatmap in Murchison Falls National Park in Uganda where the color shade indicates the intensity of coverage in the past (darker color correspond to higher patrol levels). Due to such biased data collection, the data sample might not fairly represent the entire space of the problem [53] and the learned model of the attacker behavior might have different prediction accuracy in the park areas that have high vs. low patrol densities in Figure 6.1. Thus, it may or may not be optimal to rely on learned models of attacker behavior in patrol planning, and there is no straightforward method to determine the optimal course of action prior to deployment, i.e., whether to use the learned model (or not) in patrol planning. Moreover, the sub-optimal choice may lead to arbitrary losses for the defender (as confirmed in our evaluation).

This chapter makes three significant contributions to address these shortcomings in the GSG approach. First, we propose a novel online learning algorithm, MINION-sm (a submodule of MultI-expert oNline model for constraIned patrol plaNning), which does not rely on any prior model of attacker behavior, instead it builds an implicit model of the attacker on-the-fly. MINION-sm frames the repeated security game as an adversarial combinatorial bandit problem and trades off exploitation of well-known high-reward patrol routes with exploration of untried patrol routes to provide an online policy for generating randomized patrols. It also takes into account scheduling constraints for defender. We prove that MINION-sm achieves sublinear regret against an optimal hindsight policy, which is the best that an adversarial bandit algorithm can hope for. Second, to model situations where the trained machine learning (ML) models may be a good representation of actual poacher behavior, we propose MINION (MultI-expert oNline model for constraIned patrol plaNning), an online learner which utilizes any benefits that can be achieved from exploitation of the learned ML models. Specifically, MINION considers our MINION-sm model and an ML model (based on historical data) as two patrol planning experts and dynamically combines the recommendations of both these experts to provide even better empirical performance. Finally, we evaluate our online learning algorithms and show that they outperform existing state-of-the-art GSG solvers by $\sim$100% on a variety of simulated game settings.

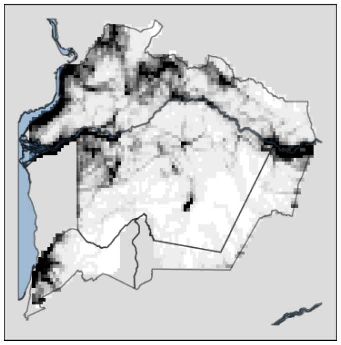
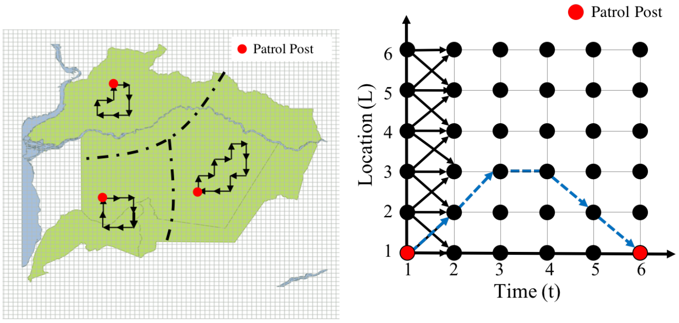


Figure 6.1: Non-uniform historical patrol coverage in Murchison Falls National Park implies biases in data collection by park rangers

6.2 Problem Formulation

Game Description We now describe the patrol route planning problem considered in this study. The entire wildlife park area is planned to be protected in within the patrol plan horizon of T and is divided into L distinct locations (grid cells). One of these locations is designated as the patrol post, w.l.o.g., we treat location 1 as the patrol post throughout the study. We cast the patrol route planning problem as a repeated game between a defender (having a single patrol team) and an attacker (having M poachers) on $L \times T$ targets (as we explain below). The game proceeds in D sequential rounds. We assume that both the defender and the attacker move simultaneously in each round of the game. However, consistent with the literature on repeated games, the defender (and the attacker) may use their opponent's actions in prior rounds to optimize the defender (and the attacker) strategy in the current round. In each round, the defender plans a patrol route $\{(l_1,t_1),\dots(l_T,t_T)\}$ for her patrol team, where l and t denote the location and time, respectively. On the other hand, the attacker chooses a set of M distinct $< l,t >$ pairs (or targets), i.e., a location



(a) Patrol Route Schema (b) Time-unrolled Graph

Figure 6.2: Patrol Planning in Green Security Games

and time pair for each of his M poachers to attack. As a result of choosing these actions, the defender gets a payoff U_i^c for each covered (patrolled) target which was attacked by the attacker, and a payoff of U_i^u ($U_i^u \leq U_i^c$) for each target which was uncovered (unpatrolled) but was attacked by a poacher. We assume that both U_i^c, $U_i^u \in [-0.5, 0.5]$ and their exact value is unknown to the defender. The goal of the defender is to design ”good” patrol routes (we formalize our exact objective later) against an adaptive attacker.

Note that our problem setup is slightly different from standard GSGs, where the primary goal of the defender is to uncover snares left by the poachers [22, 23]. As a result, a lot of emphasis is placed in prior GSG work on imperfect detection of snares by the defender when she patrols a location [71]. While this is an important real-world issue, we abstract away this complication by assuming that our defender can detect snares perfectly. In the real-world, this can be achieved by dividing the wildlife park into smaller-sized areas.

Defender’s Spatio-Temporal Constraints Due to real-world challenges, the patrol route (or pure strategy) chosen by the defender must satisfy certain spatio-temporal constraints. First, locations patrolled in consecutive time steps in the patrol route must correspond to geographically neighboring locations, otherwise it is not physically possible for the patrol team to implement that patrol. Second, any patrol route must originate from and return to the patrol post (i.e., location 1), as shown in Figure 6.2(a). We further assume that the patrol team can traverse at most T locations in each round of the game ($T << L$), and thus, the length of every patrol route must be exactly T. To simplify exposition, we model this problem using a time-unrolled graph $G(u,e)$, with LT nodes, as demonstrated in Figure 6.2(b). Each node u represents a pair $< l,t >$, i.e., location $l \in [L]$ at time $t \in [T]$ and each directed edge, e, connects a location at time t to another accessible location at time $t+1$. A defender pure strategy in this time-unrolled graph is a "feasible" path (i.e., path which satisfies spatio-temporal constraints) of length T, e.g., the blue dashed line in Figure 6.2(b) denotes a possible pure strategy for the defender. Similarly, an attacker pure strategy in this time-unrolled graph is a set of M graph nodes (note that each graph node is an $< l,t >$ pair).

Defender's Objective Note that the defender's payoffs in a given round depend only on whether her chosen pure strategy (patrol route) covers (time-unrolled) graph nodes which were chosen for attack by the attacker (this is by definition of the terms U_l^c and U_l^u). On the other hand, they do not depend on the exact ordering in which the graph nodes were patrolled. The time-indexed ordering of graph nodes (as required by the spatio-temporal constraints) in defender patrols is important only to ensure implementability of those patrols.

Thus, to formally define the defender's objective, we represent the defender's patrol route (pure strategy) as a binary vector $v \in \{0,1\}^{LT}$ s.t. $||v||_1 = T$, where each entry v_l is 1 if defender protects graph node l in that patrol route, and 0 otherwise. We reiterate that all such possible binary vectors v may not correspond to implementable patrol routes. However, corresponding to every feasible patrol route, there is exactly one binary vector v. We use $\mathscr{V}$ to denote the set of all such valid pure strategies for the defender. Similarly, we use $a \in \{0,1\}^{LT}$ s.t. $\|a\|_1 \leq M$ to denote an attacker pure strategy, and $\mathscr{A}$ to denote the set of all attacker pure strategies.

Given the defender and attacker pure strategies at round d, v^d and a^d, the defender's utility in round d is defined as $u(v_d, a_d) = \sum_{i\in[LT]} v_{d,i} a_{d,i} U_i^c + \sum_{i\in[LT]} (1 - v_{d,i}) a_{d,i} U_i^u$, which can be rewritten as $\sum_{i\in[LT]} v_{d,i} a_{d,i} [U_i^c - U_i^u] + \sum_{i\in[LT]} a_{d,i} U_i^u = v_d \cdot r_d(a_d) + C(a_d)$. Consistent with prior work, this utility equation indicates that defender needs to increase his utility by choosing strategy v_d at each round of the game. $r_d(a_d)$ denotes the reward that depends on the adversary actions.

We aim to maximize defender's expected utility over D rounds of the game $\mathbb{E}\big[\sum_{d=1}^{D} u(v_d, a_d)\big]$; where expectation is taken over randomness of the strategy. Alternatively, we want to minimize the defender's regret as computed in equation 6.1. The first term in equation 6.1 is the static optimal hindsight strategy, and is the benchmark that we compare our algorithm against. Specifically, it shows the utility of the best fixed hindsight strategy assuming all the $r_d(a_d)$ values chosen by the adversary are apriori known. This is the standard notion of regret computation used within adversarial bandit problems. This is because there are well known results that show that it is impossible to achieve sub-linear regret against a hindsight strategy which dynamically changes in every round [**?**]. Thus, the static optimal hindsight strategy is used as the benchmark, as it allows for greater computational tractability.

$$\begin{aligned} R_D &= \max_{v\in\mathscr{V}} \sum_{d=1}^{D} u(v, a_d) - \mathbb{E}\big[\sum_{d=1}^{D} u(v_d, a_d)\big] \\ &= \max_{v\in\mathscr{V}} \sum_{d=1}^{D} v \cdot r_d - \mathbb{E}\big[\sum_{d=1}^{D} v_d \cdot r_d\big] \end{aligned} \tag{6.1}$$

6.3 Patrol Route Planning with Imperfect Prior Knowledge

In GSG settings, attackers' behavior is usually represented by explicit models determined by machine learning methods that consume real-world historical data on illegal activities. These explicit models provide predictions on the likelihood of attacks on different targets based on the past adversarial actions detected by defenders who conduct patrols repeatedly to protect the targets. Consequently, if these historical data on illegal activities (collected by the defenders) are not a representative sample from the entire space, ML models might be inaccurate in estimation of attackers' behavior and pure exploitation of such attackers' model in patrol planning models can potentially result in underestimation of attacks in unexplored portions of the space [53] and be detrimental to the defender. Although there are settings that ML models could be beneficial for patrol planning, it is extremely difficult to guarantee the accuracy of the ML models for future deployments prior to the deployment. So to minimize the risk of undesirable exploitation of inaccurate (or insufficiently accurate) ML models, we propose a meta-learning approach that incorporates an online-learner along with an ML-based patrol planning model. Note that in this study, *prior knowledge* refers to the historical data about adversarial actions before the initial round of the game. In this section, (i) we propose an online learning approach for patrol planning when defender's strategy is constrained, no prior knowledge about past attacks are available and an implicit model of the attacker has to be learned on-the-fly, (ii) we discuss an ML-based patrol planning method where (potentially) imperfect prior knowledge is available, (iii) we outline our meta-learner approach which obtains the best patrol planning expert between the two previous methods based on their empirical performance.

6.3.1 Expert I: Patrol planning via online learning

To generate defender strategy based on an implicit model of the attackers, we propose an online patrol planning algorithm without any prior knowledge (i.e., historical data before the first round of the game) for constrained defender which builds upon the FPL-UE algorithm for repeated security games.

FPL-UE Algorithm The FPL-UE algorithm (follow-the-perturbed-leader with uniform exploration) proposed in [96] provides the best strategy in each round of the repeated security games by balancing exploration and exploitation. This algorithm assumes no scheduling constraints for defender and no prior knowledge about adversaries, i.e., reward $\tilde{r}_{1,i}$ in the initial round is 0 for all $i \in [N]$, where N is the number of the targets. In each round d of the experiments, a random coin is flipped to choose between exploration (with γ probability) and exploitation (with 1-γ probability) and then the defender strategy v_d is found as follows. They pick a predefined set of exploration strategies $E_{expl} = \{v_1, \dots, v_N\}$ such that target i is protected in pure strategy v_i. If the exploration phase is selected, the algorithm assures that a strategy is chosen uniformly random from set E_{expl} and each target is covered by $\frac{\gamma}{N}$ probability. If the exploitation phase is selected, v_d is the optimized strategy based on the current estimation of the rewards, $\tilde{r}_d$ and also a perturbation element that models the noise on the reward estimations. This noise is basically a random vector $z = (z_1, \dots, z_N)$, $z_i \sim \exp(\eta)$, independently drawn from the exponential distribution with parameter η. After the proposed strategy in each round is deployed, the reward estimation, $\tilde{r}_d$, is updated. The FPL-UE algorithm does not consider any constraints on the defender actions which makes the strategies impracticable for deployment in GSGs.

Algorithm 3: The MINION-sm Algorithm

```
   parameters: $\eta \in \mathbb{R}^{+}, W \in \mathbb{Z}^{+}, \gamma \in [0,1]$, $st \in [LT]$ , $ds \in [LT]$;
1  Initialize the estimated reward $\tilde{r}_d = 0 \in \mathbb{R}^{LT}$;
2  for $d = 1, \ldots, D$ do
3      sample $flag \in \{0,1\}$ such that $flag = 0$ with prob. $\gamma$;
4      if $flag == 0$ then
5          Let $j \in [LT]$ be a uniform randomly sampled target;
6          Draw $z_{d,i} \sim exp(\eta)$ independently for all $i \in [LT]$ and let $z = (z_1, \ldots, z_{LT})$;
7          Let $\alpha = 0$;
8          Let $v_d$ be $[\mathscr{P}(a = st, b = j),\ \mathscr{P}(a = j, b = ds)]$;
9      else
10         Draw $z_{d,i} \sim exp(\eta)$ independently for all $i \in [LT]$ and let $z = (z_1, \ldots, z_{LT})$;
11         Let $\alpha = 1$;
12         Let $v_d$ be $\mathscr{P}(a = st, b = ds)$ computed from the mathematical program 6.2;
13     Adversary picks $r_{d,i} \in [0,1]^{LT}$ and defender plays $v_d$;
14     Run $GR(\eta, w, \tilde{r}, d)$: estimate $\frac{1}{p_{d,i}}$ as $K_{d,i}$;
15     Update $\tilde{r}_{d,i} \leftarrow \tilde{r}_{d,i} + K_{d,i} r_{d,i} \mathbb{I}_{d,i}$; where $\mathbb{I}_{d,i} = 1$ for $v_{d,i} = 1$; $\mathbb{I}_{d,i} = 0$ otherwise;
```

MINION-sm Algorithm To overcome the limitation of the FPL-UE algorithm, we propose MINION-sm which recommends the best defender strategy in the repeated security games with scheduling constraints. Our MINION-sm algorithm outlined in Algorithm 3 assume no prior knowledge and initializes the estimation of the reward as 0 (line 1). At each round d of the game, MINION-sm conducts an exploration step with probability γ or plays an exploitative strategy with probability $1-\gamma$ (lines 4-13).

In the random exploration phase, we suggest a target-level sampling. In other words, we select target $i \in [LT]$ uniformly random and then we choose one route from a set of crossing routes at target i by solving two instances of mathematical program 6.2, [$\mathscr{P}(a = st, b = i)$, $\mathscr{P}(a = i, b = ds)$] in linear time (lines 5-8). The mathematical program $\mathscr{P}(a,b)$ in equation 6.2 gives the optimal path for the time-unrolled graph shown in Figure 6.2(b), from the starting node a to the destination node b (see the third constraint for the starting and the destination nodes). In our patrol route planning problem, *st* and *ds* denote the patrol post locations at the beginning and end of the patrol route. The weights in this graph are the estimated reward values. We add a random noise vector z to prevent the algorithm to choose a fixed route for all the times that a specific node j is selected in exploration phase (line 6). The mathematical program $\mathscr{P}(a,b)$ is equivalent to the problem of finding the longest path in a weighted directed acyclic graph, which can be solved in linear time. E in equation 6.2 represents the set of the edges in the time-unrolled graph $G(u,e)$ introduced in section 6.2. $\sigma^{+}(v_{d,i})$ denotes the in-going edges to the node $v_{d,i}$ and $\sigma^{-}(v_{d,i})$ denotes the out-going edges from the node $v_{d,i}$, in graph G. To find the longest path (the optimal defender strategy), we used a network flow approach. Thus, $f(e)$ represents the flow on each edge of the graph G. If a node is covered by defender, $f(e)$ will be 1 for one of the in-going edges and one of the out-going edges.

$$
\begin{aligned}
v_d = \operatorname*{arg\,max}_{v \in \mathscr{V}} \sum_{i=1}^{LT} v_{d,i}(\alpha \tilde{r}_{d,i} + z)& \\
\text{subject to}& \\
v_{d,i} = \textstyle\sum_{e \in \sigma^{+}(v_{d,i})} f(e) \qquad & \forall e \in E, \forall i \in [LT] \\
\textstyle\sum_{e \in \sigma^{+}(v_{d,i})} f(e) = \sum_{e \in \sigma^{-}(v_{d,i})} f(e) \qquad & \forall e \in E, \forall i \in [LT] \\
\textstyle\sum_{e \in \sigma^{-}(v_{1,a})} f(e) = \sum_{e \in \sigma^{+}(v_{D,b})} f(e) = 1 \qquad & \forall e \in E \\
f(e), v_{d,i} \in \{0,1\} \qquad & \forall i \in [LT], \forall e \in E
\end{aligned}
\tag{6.2}
$$

In our game, a pure strategy is defined as a feasible patrol route (i.e., a route in graph G) and the set of all possible strategies (all routes) are $O(L^T)$. Such set is computationally expensive to be generated for large-size graphs. Additionally, even if we generate such a large set for the exploration step, the algorithm would suffer from a slower convergence. So our target-level random sampling does not require generation of $O(L^T)$ routes and assures that each target i is covered by $p_i \geq \frac{\gamma}{LT}$, as opposed to the strategy-level uniform sampling which assures $p_i \geq \frac{\gamma}{L^T}$.

Hence, this approach facilitates scalability of the algorithm and demonstrates similar performance guarantee as FPL-UE without scheduling constraints.

In the exploitation phase, we choose an optimized patrol route computed by mathematical program 6.2 according to the current estimation of the rewards on all targets up to the current round (lines 10-12).

Once the defender strategy v_d is computed and deployed at round d, reward $r_{d,i}$ is observed for the targets visited by the defender (line 14). Then the probability $p_{d,i}$ that target i is chosen at round d by our algorithm is computed based on the algorithm 4 (line 15) and the reward estimations are adjusted and updated for visited targets as $\tilde{r}_{d+1,i} = \tilde{r}_{d,i} + \frac{r_{d,i}}{p_{d,i}} \mathbb{I}_{d,i}$ (line 16). $\mathbb{I}_{d,i}$ is the indicator function that indicates whether target i was chosen by the defender at round d. The term $\frac{r_{d,i}}{p_{d,i}} \mathbb{I}_{d,i}$ is an unbiased estimator of $r_{d,i}$ (i.e., $\mathbb{E}(\frac{r_{d,i}}{p_{d,i}}) = r_{d,i}$). This choice of the reward adjustment is for convenience of theoretical analysis. Since $p_{d,i}$ cannot be computed efficiently, we use the Geometric Resampling technique proposed by [67], outlined in Algorithm 4, where $K_{d,i} = \frac{1}{p_{d,i}}$ denotes the mean of the geometric distribution with success probability of $p_{d,i}$ for the first trial. W denotes number of the iteration that the algorithm 4 is run and is an input to the algorithm. The MINION-sm algorithm continues for D rounds.

Theorem 6.1 *The performance of MINION-sm follows the same theoretical properties as FPL-UE where the regret (i.e., the difference between the performance of MINION-sm and that of the best fixed patrolling strategy in hindsight) is upper bounded by:*

$$R_D \leq \gamma MD + 2DTe^{-W\frac{\gamma}{LT}} + \frac{T(\log LT + 1)}{\eta} + \eta MD \min(M,T)$$

By taking $\eta = \sqrt{\frac{T(\log LT+1)}{MD\min\{M,T\}}}$, $\gamma = \frac{\sqrt{T}}{\sqrt{MD}}$, $W = L\sqrt{TMD}\log(DT)$, we obtain the upper bound $\mathscr{O}\left(\sqrt{TMD\min\{M,T\}\log LT}\right)$.

Due to space limitations, the full proof of this theorem is omitted. However, it can be sketched as follows:

Proof 6.1 (Proof sketch) *A key step in the proof of is to bound below the probability that the chosen path will contain a particular node. By construction of MINION-sm, this value can be bounded below with γ/LT. By combining this bound with some ideas from the proof of Theorem 1 from [96] (tailored to our setting) and some further technical algebra, we can achieve the required regret bound.*

Algorithm 4: The GR Algorithm

```
input  : $\eta \in \mathbb{R}^+, W \in \mathbb{Z}^+, \tilde{r} \in \mathbb{R}^{LT}, d \in \mathbb{N}$
output: $K_d \in \mathbb{Z}^{LT}$
Initialize $\forall i \in [LT] : K_{d,i} = 0, k = 1$;
for $k = 1, \ldots, W$ do
    Execute steps $3-13$ in Algorithm 3 once just to produce $\tilde{v}$ as a simulation of $v_d$;
    for all $i \in [LT]$ do
        if $k < W$ and $\tilde{v}_i = 1$ and $K_{d,i} = 0$ then
            $K_{d,i} = k$
        else if $k = W$ and $K_{d,i} = 0$ then
            $K_{d,i} = W$
    if $K_{d,i} > 0$ for all $i \in [LT]$ then
        break
```

6.3.2 Expert II: Patrol planning via machine learning model

In green security games, the wildlife crime datasets are used for development of explicit attackers' model based on machine learning techniques. Since the ML modeling based on the real-world data is not the focus of this study, we skip the modeling details and we just briefly provide an overview of the inputs/outputs for such ML models and then we show how the outputs of such ML models are used for patrol planning purposes [26, 27].

ML Model Inputs In wildlife protection domain, the park rangers begin to conduct patrols from patrol posts located across the vast national parks and return to the same patrol posts every day as shown in Figure 6.2(a). So the wildlife crime datasets consist of several years of type, location, and date of the wildlife crime records detected by park rangers during the repeated patrols which is used for supervised ML modeling of attackers' behavior. Along with these historical observations, several environmental features such as terrain (e.g., slope), distance values (e.g., distance to the border, patrol posts, roads, and towns, rivers), and animal density along with past patrol coverage are considered as predictor features that influence the decision making process by adversaries. Such historical records are transformed into spatio-temporal data points to train a machine learning model as follows. The protected area is divided into grid cells $\tilde{l}$ (e.g., cells of size 1 sq. km) and the entire time span of the crime records, $\tilde{T}$, is divided into small time steps $\tilde{t}$ (e.g., 3 month or 12 month long due to sparsity of the data). Thus the dataset $\mathscr{D} = (\mathbf{X}, \mathbf{y})$, contains $\tilde{T}\tilde{L}$ of such spatio-temporal slices (usually tens of thousands) from all around the park over several years where $\mathbf{X} \in \mathbb{R}^{\tilde{T}\tilde{L} \times f}$ is a matrix of f predictor features and $\mathbf{y} \in \{0, 1\}^{\tilde{T}\tilde{L}}$ denotes the observation vector which represents the presence or absence of the attack.

ML Model Outputs Training a machine learning model based on $\mathscr{D} = (\mathbf{X}, \mathbf{y})$ gives predictions about probability scores (i.e., attack risk) $p(i) = h(\mathbf{x}_i)$ at each target i. Such predictions are used to generate optimized patrol strategies as shown by the following mathematical model $\mathscr{Q}(a, b)$, where a and b are starting and ending targets for patrolling.

$$
\begin{aligned}
& v_d = \arg\max_{v\in\mathscr{V}} \sum_{i=1}^{LT} v_{d,i}.p_i \\
& \text{subject to} \\
& v_{d,i} = \textstyle\sum_{e\in\sigma^+(v_{d,i})} f(e) && \forall e \in E, \forall i \in [LT] \\
& \textstyle\sum_{e\in\sigma^+(v_{d,i})} f(e) = \sum_{e\in\sigma^-(v_{d,i})} f(e) && \forall e \in E, \forall i \in [LT] \\
& \textstyle\sum_{e\in\sigma^-(v_{1,a})} f(e) = \sum_{e\in\sigma^+(v_{D,b})} f(e) = 1 && \forall e \in E \\
& f(e), v_{d,i} \in \{0,1\} && \forall i \in [LT], \forall e \in E
\end{aligned}
\tag{6.3}
$$

Due to the sparsity of the datasets, $\tilde{t}$, the smallest time resolution for ML model predictions is much larger than the smallest time horizon T required for fine-tuned patrol planning, i.e., $\tilde{t} \gg T$. Consequently, machine learning predictions for each location does not get updated real-time and remain nearly similar across time period T (i.e., stationary predictions) shown in time-unrolled graph in Figure 6.2(b).

6.3.3 Patrol planning via expert I and II

Algorithm 5 outlines our meta-learning approach to balance between two experts, i.e., (I) MINION-sm online learning algorithm with no prior knowledge and (II) an ML-based patrol planning model with potentially imperfect prior knowledge. This algorithm initializes the estimation of the reward as 0 (line 1) and then picks a set of exploration strategies to obtain an initial assessment about the performance of the experts; thus r_{ml} and r_{ol} are initialized for both patrol planning experts (line 2). At each round d of the game, the current collected rewards for each expert are perturbed (line 4) by drawing random noise for each expert from the exponential distribution with parameter β to model the noise on the current estimation of the rewards and then the best expert is chosen by the algorithm (line 5). If ML model is selected as the best expert, v_d

Algorithm 5: The MINION Algorithm

```
   parameters: $\eta \in \mathbb{R}^+, \beta \in \mathbb{R}^+, W \in \mathbb{Z}^+, \gamma \in [0,1]$, $e \in \mathbb{N}$, $st \in [LT]$ , $ds \in [LT]$;
1  Initialize the estimated reward $\tilde{r}_d = 0 \in \mathbb{R}^{LT}$, $r_{ml} = 0$, $r_{ol} = 0$, $n_{ml} = 0$, $n_{ol} = 0$;
2  Pick $e$ exploration strategies such that two experts ml (outlined in line 7) and ol (outlined
   in lines 10-16) are explored uniformly and $r_{ml}$ and $r_{ol}$ are initialized ;
3  for $d = 1, \dots, D$ do
4      Draw $c_{d,1} \sim exp(\beta)$ and $c_{d,2} \sim exp(\beta)$
5      if $\frac{r_{ml}}{n_{ml}} + c_{d,1} \geq \frac{r_{ol}}{n_{ol}} + c_{d,2}$ then
6          $n_{ml} \leftarrow n_{ml} + 1$;
7          $f = 0$;
8          Let $v_d$ be computed from the mathematical program $\mathscr{Q}(a = st, b = ds)$ in 6.3;
9      else
10         $n_{ol} \leftarrow n_{ol} + 1$;
11         $f = 1$;
12         Let $v_d$ be computed by following steps 3-13 in algorithm 3;
13     Adversary picks $r_{d,i} \in [0,1]^{LT}$ and defender plays $v_d$;
14     $r_{ml} \leftarrow r_{ml} + f v_d r_d$;
15     $r_{ol} \leftarrow r_{ol} + f v_d r_d$;
16     Run $GR(\eta, w, \tilde{r}, d)$: estimate $\frac{1}{p_{d,i}}$ as $K_{d,i}$;
17     Update $\tilde{r}_{d,i} \leftarrow \tilde{r}_{d,i} + K_{d,i} r_{d,i} \mathbb{I}_{d,i}$; where $\mathbb{I}_{d,i} = 1$ for $v_{d,i} = 1$; $\mathbb{I}_{d,i} = 0$ otherwise;
```

is computed based on the mathematical program $\mathscr{Q}(a = st, b = ds)$ presented by equations 6.3 (lines 6-8). Otherwise, the MINION-sm online learning approach is used (lines 10-22). Then the adversary picks the rewards r_d for the defender (line 24) and the collected rewards for each expert will be updated accordingly (lines 25-28). The MINION algorithm continues for *D* rounds.

The intuition behind MINION is that the algorithm will learn whether it is useful to rely on historical data. If yes, then it will use the ML model to predict the future payoffs, otherwise it will use MINION-sm to plan the patrolling strategy. In particular, we provide the following guarantee on the performance of MINION:

Theorem 6.2 *Let P_{ML} and P_{fixed} denote the expected performance of the ML model and the best fixed patrolling strategy in hindsight. The expected performance of MINION is at least as good* *as*

$\max\{P_{\text{ML}}, P_{\text{fixed}}\} - \mathscr{O}\left(\sqrt{TMD\min\{M,T\}\log LT}\right).$

Proof 6.2 (proof sketch) *If $P_{\text{ML}} > P_{\text{fixed}}$ then the meta-learner in MINION will learn this with $\mathscr{O}(\sqrt{D})$ regret (as the meta-learner a two-expert learning problem). Otherwise, it will converge to MINION-sm. This yields regret of $\mathscr{O}(\sqrt{D}) +$*

$\mathscr{O}\left(\sqrt{TMD\min\{M,T\}\log LT}\right) = \mathscr{O}\left(\sqrt{TMD\min\{M,T\}\log LT}\right).$

6.4 Numerical Evaluation

In this section, we evaluate the numerical performance of the MINION-sm and MINION against an ML-based patrol planning model (ML-exploit) and absolute exploratory defender strategies (pure-explore). We first evaluate our algorithms on a game with 25 locations ($L = 25$) and a patrol horizon of 6 time steps ($T = 6$) and then we show the average defender reward for all techniques by varying the patrol horizon (i.e., different time-unrolled graph sizes). The MINION-sm and pure-explore algorithm do not incorporate any explicit model for attackers' behavior. However, the MINION algorithm and also ML-exploit baseline algorithm require access to an ML model for attackers' behavior to solve the mathematical model 6.3 for patrol planning. We simulate the ML model predictions (ML outputs), $p(i) = h(\mathbf{x}_i)$, with three different scenario shown in Figures 6.4(a) to 6.4(c). These predictions are stationary for all locations across the patrol horizon. We assume two types of adversarial behavior: (i) STC- a Stochastic adversarial behavior where the likelihood of attack at each location can be defined by probability scores; Figures 6.4(d) to 6.4(f) shows our simulated cases for three different m values, where m indicates the expected number of the attackers. These probability scores represent the ground truth for adversarial behavior and are stationary across $T = 6$ time steps for all locations. We used them to pick rewards for the defender play in the game for all of the patrol planning methods. (ii) QR- a Quantal Response adversary where the attackers' behavior is non-stationary across the game rounds and the attackers respond to the empirical defender mixed strategy by a QR model [72].

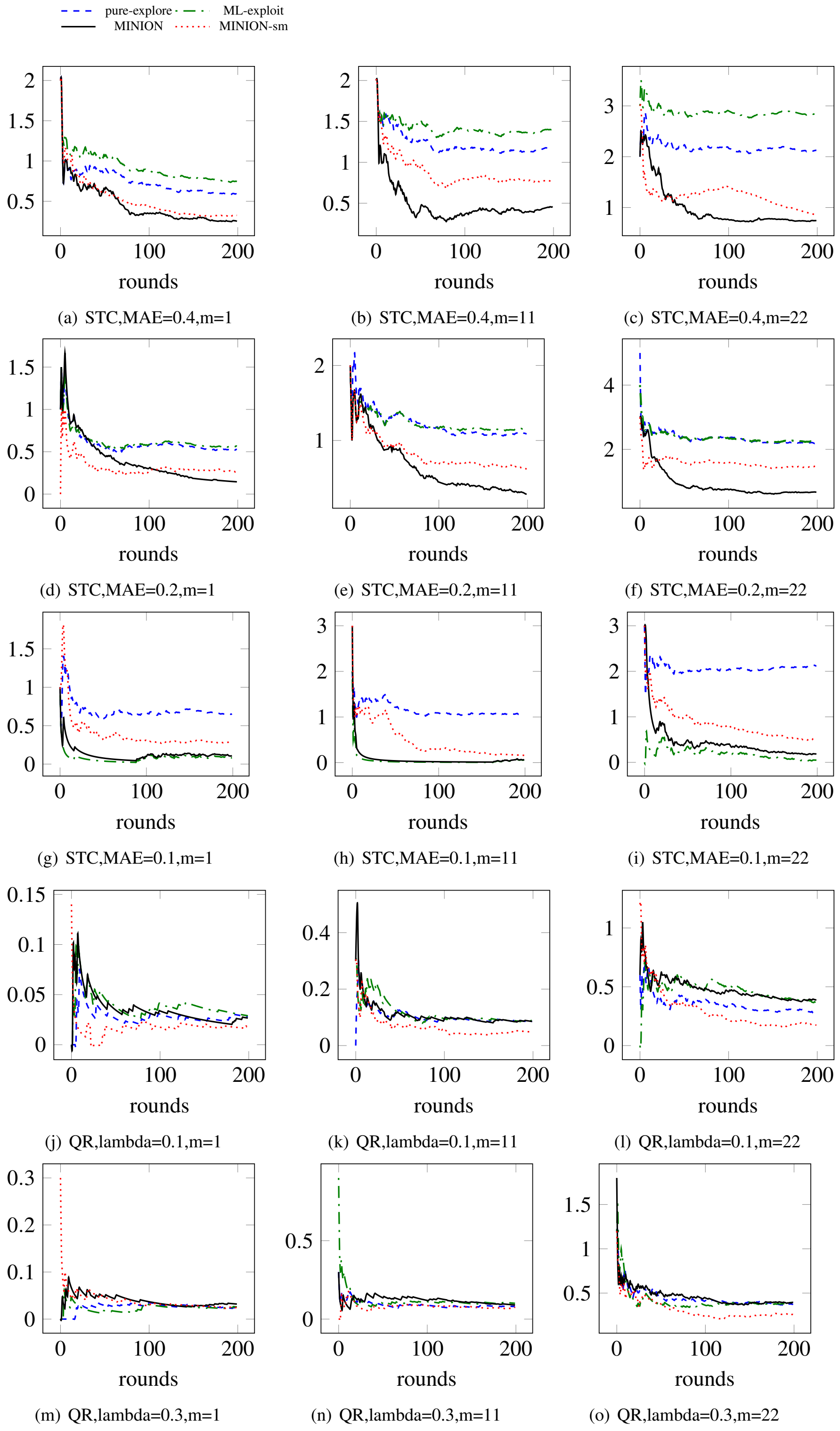
For each STC adversary represented by the ground truth (GT) probabilities shown in Figure 6.4(d) to 6.4(f), the ML simulations have different levels of inaccuracy. We quantify this difference via $MAE = \sum_{l=1}^{L} |p_{gt}(l) - p_{ml}(l)|/L$ which is the mean absolute error in predictions. In our simulated cases shown in Figure 6.4, MAE varies from 0.1 to 0.4. For QR adversaries, we do not have a fixed GT and adversaries' responses are updated according to the updated mixed strategy of the defender. We examine two λ values, i.e., 0.1 and 0.3 as the rationality parameters of the adversaries where the smaller values indicate more non-rational adversaries in QR model.

The regret values for all 9 scenarios for STC and 6 scenarios for QR are shown in Figures 6.3. The blue dashed lines and the green lines in the figures show the results for pure-explore and for ML-exploit baseline methods, respectively. The red dotted lines and black solid lines illustrate the results for MINION-sm and MINION algorithms proposed in this study. The regret values are shown along the y-axis and the game rounds are shown along the x-axis. In the top three rows, for the STC adversary, we show the performance loss when the accuracy of the ML model predictions (used in ML-exploit and MINION) increases from top to the bottom. From left to the right, we show the change in performance loss as the expected number of the adversary increases. In the first and the second rows in Figure 6.3 where ML model is relatively inaccurate, MINION-sm (the technique that uses no prior knowledge and balances exploration and exploitation), outperforms ML-exploit whereas in the third row the trend is reversed since the ML model is sufficiently accurate and informative for the patrol planning task. On the other hand, MINION outperforms all other methods in all cases since it obtains a balance between MINION-sm and ML-exploit and finds the best expert based on their empirical performance. In the bottom two rows, for QR adversaries, MINION-sm algorithm outperforms other techniques. When the relative number of the adversaries to the number of the defender resources is larger, this difference is more significant. The MINION is outperformed by MINION-sm against QR adversary, since it partially relies on an ML-based patrol planning expert for which the predictions are not updated accordingly over the game rounds and thus suffers from biases in prior knowledge.

Figure 6.5 shows the average defender utility over 200 rounds of the game on the y-axis vs. different patrol horizons. The game settings with different number of attackers are outlined across the different rows. For STC adversary (shown in the top three rows), MINION outperforms all other methods and for QR adversaries (shown in the bottom two rows), the MINION-sm algorithm outperforms other methods for all graph sizes. The key reason behind the poor performance of MINION vs. MINION-sm in QR scenario is that MINION incorporates an ML-based planner with stationary predictions about the attackers' behavior as an expert planner against the non-stationary (responsive and strategic) adversaries which is detrimental to the defender.

6.5 Conclusion

This chapter focuses on the important problem of game-theoretic patrol route selection for preventing poaching activities in wildlife parks. The main intellectual contribution of this study is that it shows that over-reliance on historical patrolling data (or ”prior knowledge”) in the patrol route generation process may lead to highly sub-optimal patrols, and that the optimal amount of reliance on prior knowledge can be learned effectively (by techniques put forth in the chapter). Specifically, this chapter makes the following methodological contributions: (I) we propose MINION-sm, an scalable online learning algorithm that learns an implicit model of the attacker when defender is spatio-temporally constrained, (II) we propose MINION, which is a scalable multi-expert patrol planning algorithm with spatio-temporal constraints for the defender that obtains a balance between the ML-based planners and MINION-sm based on their empirical performance. We showed that our algorithms outperformed other techniques in different game settings.



(a) STC,MAE=0.4,m=1

(b) STC,MAE=0.4,m=11

(c) STC,MAE=0.4,m=22

(d) STC,MAE=0.2,m=1

(e) STC,MAE=0.2,m=11

(f) STC,MAE=0.2,m=22

(g) STC,MAE=0.1,m=1

(h) STC,MAE=0.1,m=11

(i) STC,MAE=0.1,m=22

(j) QR,lambda=0.1,m=1

(k) QR,lambda=0.1,m=11

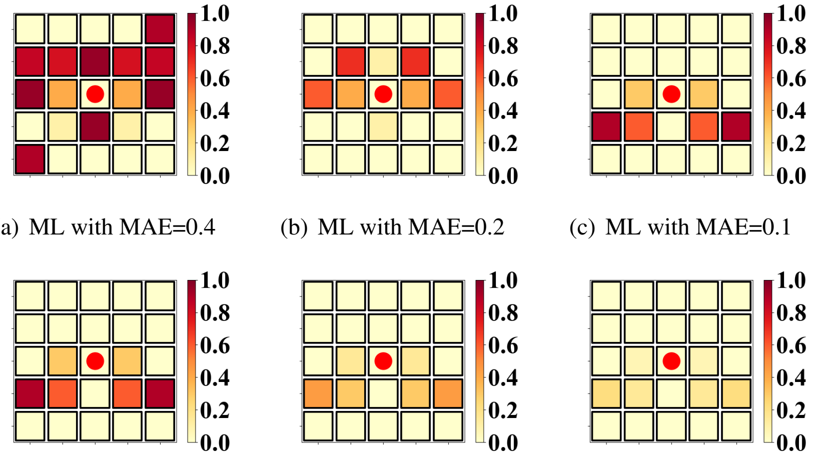
(l) QR,lambda=0.1,m=22

(m) QR,lambda=0.3,m=1

(n) QR,lambda=0.3,m=11

(o) QR,lambda=0.3,m=22

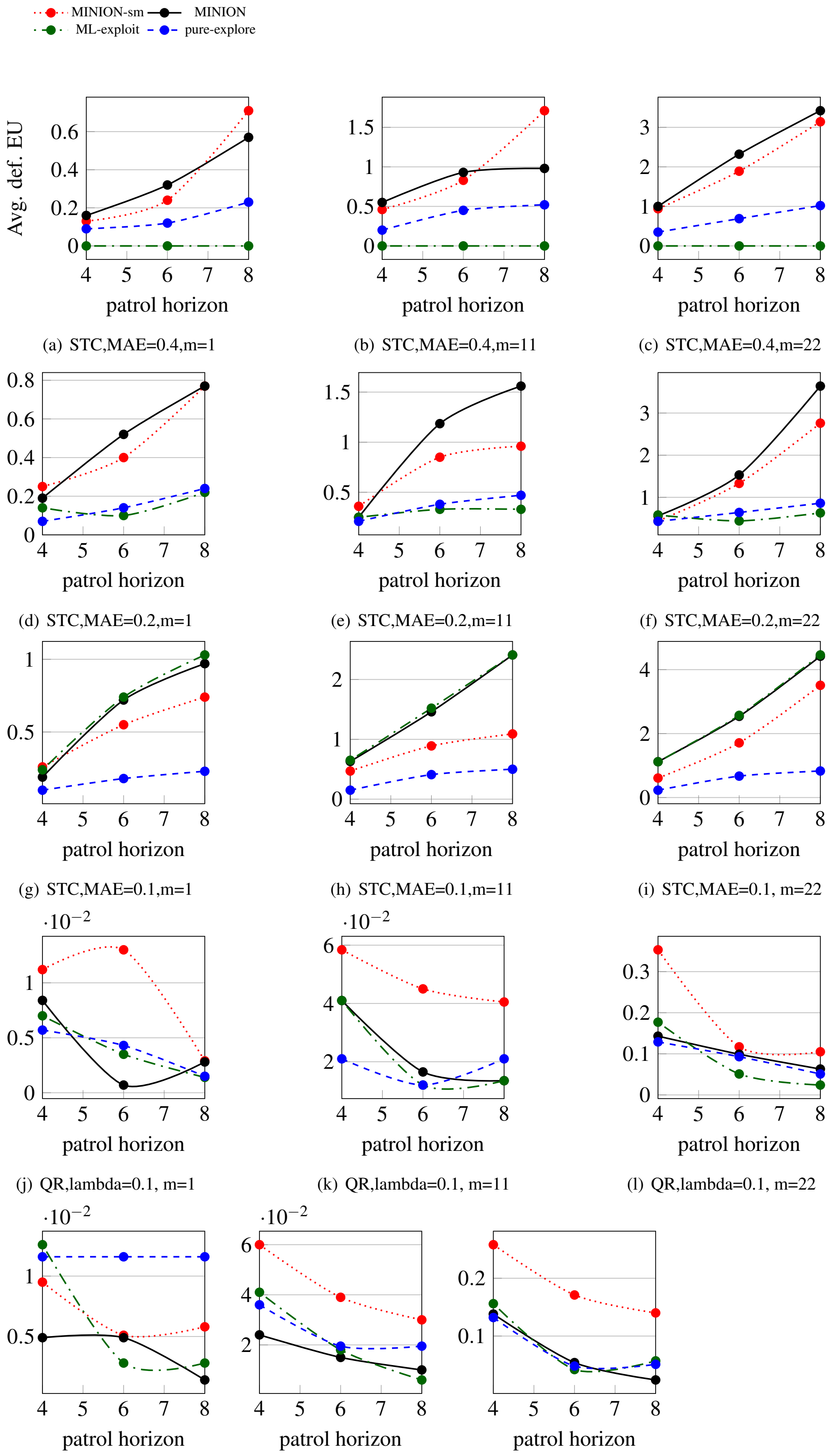
Figure 6.3: Regret for adversaries with stochastic (stationary) and Quantal response (non-stationary) behavior - $L = 25$, $T = 6$



(a) ML with MAE=0.4 (b) ML with MAE=0.2 (c) ML with MAE=0.1

(d) GT for m=22 (e) GT for m=11 (f) GT for m=1

Figure 6.4: top: Heat maps for attack probability predicted by different ML models, bottom: Heat maps for attack probability ground truth, red dot denotes patrol post location



(a) STC,MAE=0.4,m=1 (b) STC,MAE=0.4,m=11 (c) STC,MAE=0.4,m=22

(d) STC,MAE=0.2,m=1 (e) STC,MAE=0.2,m=11 (f) STC,MAE=0.2,m=22

(g) STC,MAE=0.1,m=1 (h) STC,MAE=0.1,m=11 (i) STC,MAE=0.1, m=22

(j) QR,lambda=0.1, m=1 (k) QR,lambda=0.1, m=11 (l) QR,lambda=0.1, m=22

(m) QR,lambda=0.3, m=1 (n) QR,lambda=0.3, m=11 (o) QR,lambda=0.3, m=22

Figure 6.5: Average defender utility over 200 rounds for adversaries with stochastic (stationary) and Quantal response (non-stationary) behavior for different graph sizes and different number of attackers m

Chapter 7

Collusive Security Games

In this chapter, I discuss a game theoretical technique to break up collusive actions between adversaries, I investigate algorithms for the defender assuming both rational and boundedly rational collusive adversaries. I propose collusive security games (COSGs), a model for security games involving potential collusion among adversaries and SPECTRE-R, an algorithm to solve COSGs and break collusion assuming rational adversaries. To model bounded rationality, I introduce a learned human behavioral model to predict when collusion will occur and SPECTRE-BR, an enhanced algorithm which optimizes against the learned behavior model to provide demonstrably better-performing defender strategies against human subjects compared to SPECTRE-R.

7.1 Problem Domain

Models and algorithms based on Stackelberg security games have been deployed by many security agencies including the US Coast Guard, the Federal Air Marshal Service, and Los Angeles International Airport [87] in order to protect against attacks by strategic adversaries in counter-terrorism settings. Recently, security games research has explored new domains such as wildlife protection, where effective strategies are needed to tackle sustainability problems such as illegal poaching and fishing [23].

Crucially, though, most previous work on security games assumes that different adversaries can be modeled independently [43, 47, 70]. However, there are many real-world security domains in which adversaries may collude in order to more effectively evade the defender. One example domain is wildlife protection. Trade in illicit wildlife products is growing rapidly, and poachers often collude both with fellow poachers and with middlemen who help move the product to customers [92]. These groups may coordinate to gain better access to information, reduce transportation costs, or reach new markets. This coordination can result in higher levels of poaching and damage to the environment. Additionally, connections have been observed between illicit wildlife trade and organized crime as well as terrorist organizations, and thus activities such as poaching can serve to indirectly threaten national security [94].

Another example domain is the illegal drug trade where international crime syndicates have increased collusive actions in order to facilitate drug trafficking, expand to distant markets, and evade local law enforcement [4]. In some cases, drug traders must collude with terrorist organizations to send drugs through particular areas. More broadly, expansion of global transportation networks and free trade has motivated collusion between criminal organizations across different countries [80]. A third example of a domain with collusive actions is the "rent-a-tribe" model in the payday lending industry. Authorities in the US attempt to regulate payday lenders which offer extremely high interest rates to low-income borrowers who cannot obtain loans from traditional banks. Recently, payday lenders have begun to operate in partnership with Native American tribes, which are exempt from state regulations. Thus, regulators seek policies which prevent collusion between payday lenders and Native American tribes [39].

Despite mounting evidence of the destructive influence of collusive behavior, strategies for preventing collusion have not been explored in the security games literature (there are some recent exceptions, which we discuss in Section 2). Furthermore, analysis of collusive adversary behaviors is complicated by the bounded rationality of human adversaries; such analysis with data from human players is also missing in the security games literature. To address these limitations and improve defender performance by combating collusion between adversaries, this paper (i) introduces the COllusive Security Game (COSG) model with three players: one defender and two adversaries with the potential to collude against the defender, (ii) provides a baseline algorithm, SPECTRE-R, which optimizes against collusive adversaries assuming them to be perfectly rational, (iii) analyzes data from an experiment involving 700 human subjects, (iv) proposes a data driven human behavioral model based on these factors to predict the level of collusion between human adversaries, and (v) develops a novel algorithm, SPECTRE-BR, which optimizes against the learned behavior model to better prevent collusion between adversaries (and as a result, outperforms SPECTRE-R). Indeed, we find that human adversaries are far from perfectly rational when deciding whether or not to collude. Our experiments show that defenders can improve their utility by modeling the subjective perceptions and attitudes which shape this decision and crafting strategies tuned to prevent collusion.

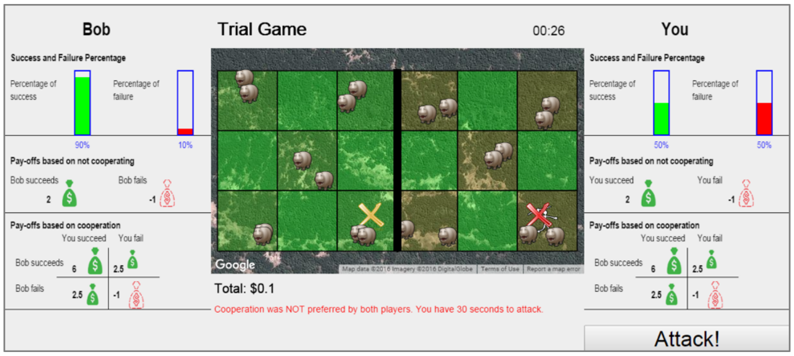
7.2 Illustrative Motivating Domain: Wildlife Poaching Game

As an illustrative motivating domain for the work reported in this paper, we focus on the challenge of wildlife poaching. Wildlife poaching poses a serious threat to the environment as well as national security in numerous countries around the world and is now estimated to be worth

$5 billion annually. The most common types of illicitly poached and traded wildlife products include elephant ivory, rhino horn, tiger parts, and caviar [66]. Biodiversity loss, invasive species introduction, and disease transmission resulting from illicit wildlife trade can all have disastrous impacts on the environment. Evidence [92] confirms that collusive actions (e.g., cost sharing for storage, handling, and transportation of goods) among adversaries can increase the rate of poaching and cause further damage to the environment. Modeling this as a security game, the defender is a ranger whose goal is to allocate patrolling resources optimally over the targets. The adversaries are poachers or illegal traders who execute attacks, possibly in collusion with one another. To better understand collusion in the wildlife poaching domain, we designed a game for human subjects to play on Amazon Mechanical Turk (AMT). Participants were asked to play our game in different settings and answer survey questions. Afterwards, their actions were analyzed using the theories explained above, allowing us to test assumptions about the rationality of human adversaries.

7.2.1 Game Overview

In our game, human subjects are asked to play the role of a poacher in a national park in Africa. The entire park area (see Figure 7.1) 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 obtain information about the other player's situation. To ensure repeatability of the experiments, the other side is played by a computer, not a real player. Since our goal is to study human adversaries, we do not reveal the identity of the other player to the human subjects. This creates a more realistic environment since the subjects believe that they are playing against another human.



Total: $0.1

Cooperation was NOT preferred by both players. You have 30 seconds to attack.

Figure 7.1: Poachers vs. Rangers game: Right side of the park is assigned to the player and the left side is assigned to Bob who is the other fellow poacher. Payoffs for each marked target are shown.

Each section of the park is divided into a 3 × 3 grid, giving each player nine potential targets to attack.

There are different numbers of hippopotamus distributed over the area which indicate the animal density at each target. The adversary's reward at each target is equal to the animal density at that target; hereafter, reward and animal density are used interchangeably. Players are able to view the probability of success and failure, as well as the reward and penalty, at any target on either section of the park as shown on the sides of the Figure 7.1. To help the human subjects better visualize the success/failure percentages (i.e., defender coverage) for each sub-regions, we overlaid a heat-map of the success probability on Google Maps imagery of the park. Also, to help the players understand the collusion mechanism, we provided a table that summarizes all possible payoffs for both colluding and not colluding. The human subjects may decide to attack "individually and independently" or "in collusion" with the other player. In both situations, they will attack different sections separately but if both agree to attack in collusion, they will share all of their payoffs with each other equally.

7.2.2 Experimental Procedure

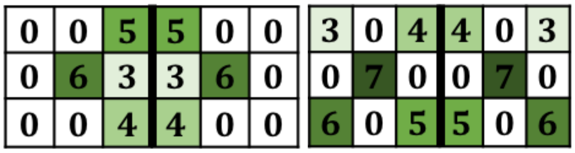
To enhance understanding of the game, participants were provided with a background story and detailed instructions about the game and then asked to play one trial game to become familiar with the game interface and procedures. After the trial game, participants played a validation game to ensure that had they read the instructions and were fully aware of the rules and options of the game. For our analysis, we included only players whose performance in the validation game passed a set of baseline criteria. Lastly, subjects played the main game for the analysis. After finishing all of the games, participants answered a set of survey questions.

In each individual game, the human player is given a set amount of time to explore the park and make decisions about: (i) whether to collude with the other player or not and (ii) which region of the park to place their snare. While the other player is a computer, it is suggested that they are actually another human. To make the first decision, a question appears on the screen which asks whether the human player is inclined to collude or not. After answering this question, a message appears on the screen that indicates whether collusion was preferred by both players or not. Collusion occurs only if it is preferred by both players. It is worth noting that the human participant has no opportunity to communicate with or learn about the other player. Next, players are asked to choose a target in their own region to attack. As before, players cannot communicate about which target to attack.

We analyze two situations: one where the human attacker is placed in an advantaged situation, with fewer defender resources protecting his side of the park than the other; and a disadvantaged situation, which is the reverse. In each situation, as we mentioned, we first check if the player is inclined to collude. Next, we designed a computer agent with rational behavior to play as the second adversary; thus there is an algorithm generating defender strategies, and two adversaries (one a human and one a computer agent). This computer agent seeks collusion when it is placed on the disadvantaged side and refuses collusion when it is in advantaged situation (Choosing a computer agent as a second player let us to avoid requiring coordination between two human players in the experiments). To simplify the analysis, we assume that the second stage of decision making (where each adversary chooses a target to attack) depends on his own inclination for collusion and does not depend on the attitude of the other adversary.

Consequently, there are four possible types of human adversaries in this game: (i) a disadvantaged attacker who decides to collude, DA-C, (ii) a disadvantaged attacker who decides not to collude, DA-NC, (iii) an advantaged attacker who decides to collude, A-C, and (iv) an advantaged attacker who decides not to collude, A-NC.

We tested different defender mixed strategies based on both the assumption of rationality and bounded rationality given by a behavioral model introduced in Section 6. For each strategy deployed on AMT, we recruited a new set of participants (50 people per setup) to remove any learning bias and to test against a wider population. Using the rational model for adversaries, four different defender strategies were deployed for each reward structure. The data sets collected from rational model deployments were used to learn the parameters of the bounded rationality model. This learning mimics the fact that in the real world, often data about past poaching incidents is available to build models of poacher behavior [70]. Players were given a base compensation of \$0.50 for participating in the experiment. In order to incentivize the players to perform well, we paid each player a performance bonus based on the utility that they obtained in each game. This bonus had a maximum total value of \$1.32 and a minimum of \$0.04.



0	0	5	5	0	0
0	6	3	3	6	0
0	0	4	4	0	0

(a) RS1

3	0	4	4	0	3
0	7	0	0	7	0
6	0	5	5	0	6

(b) RS2

Figure 7.2: Reward (animal density) structures deployed on AMT. Darker green shows higher reward.

7.2.3 Game Payoff Design

This "Poachers vs Rangers" game is a three-player security game with 9 targets available to each adversary. There is one defender with m resources to cover all the 18 targets (sub-regions in the park) and there are two adversaries that can attack a side of the park. An adversary's reward at each cell for an uncovered attack is equal to the animal density at that cell and the penalty at each cell for a covered attack is equal to -1. We deployed two different reward structures, $RS1$ and $RS2$, shown in Figures 7.2(a) and 7.2(b). In both of these symmetric structures, both players have an identical 3×3 reward distribution. In $RS1$ animal density is concentrated along the central axis of the park and is covered by 3 defender resources and in $RS2$ animal density is concentrated toward the center of each half of the park and is covered by 4 defender resources. We assumed a bonus of 1 for collusion in both set-ups; this bonus is added to the payoff for each successful attack if both attackers decide to collude. Section 4 gives further mathematical description and motivates the introduction of this bonus. This game is zero-sum, i.e., at each target the uncovered payoffs for the attacker and defender sum to zero.

7.3 Collusive security game model

In the collusive security game which we study in this study, there is one defender, Θ, and multiple adversaries, $\Psi_1,...,\Psi_N$, where N is the total number of attackers. Similarly to standard Stackelberg Security Games [87], the defender is the leader and the attackers are the followers. In this subsection, we focus on the games with one defender and two adversaries, such that adversaries can attack separate targets, but they have two options: i) attack their own targets individually and earn payoffs independently or ii) attack their own targets individually but collude with each other and share all of the payoffs equally. If the attackers decide to collude, the utility for a successful attack increases by ε. This reward models many of the example domains where adversaries operate in different geographic areas or portions of a supply chain, and so do not directly compete over the same targets. Instead, they choose to combine their operations or share information in some way which produces extra utility exogenous to the targets themselves.

To precisely define the model, let $T = \{t_1,...,t_n\}$ be a set of targets. T is partitioned into disjoint sets T_1 and T_2 which give the targets accessible to the first (resp. second) attacker. The defender has m resources, each of which can be assigned to cover one target. Since we consider games with no scheduling constraints [101], the set of defender pure strategies is all mappings from each of the m resources to a target. A mixed strategy is a probability distribution over such schedules, and can be compactly represented by a coverage vector C which gives the probability that each target is covered. Each attacker pure strategy is the combination of a choice of target to attack and the decision of whether or not to collude. Since the attackers choose their strategies after the defender, there is always an equilibrium in which they play only pure strategies [47].

Payoffs for individual attacks		Payoffs for collusive attacks	
Attackers: Ψ_1, Ψ_2	Defender: Θ	Each attacker: Ψ_1 or Ψ_2	Defender: Θ
$U^u_{\Psi_1}(t_1)$, $U^u_{\Psi_2}(t_2)$	$U^u_\Theta(t_1)+U^u_\Theta(t_2)$	$(U^u_{\Psi_1}(t_1) + U^u_{\Psi_2}(t_2) + 2\varepsilon)/2$	$U^u_\Theta(t_1)+U^u_\Theta(t_2) - 2\varepsilon$
$U^u_{\Psi_1}(t_1)$, $U^c_{\Psi_2}(t_2)$	$U^u_\Theta(t_1)+U^c_\Theta(t_2)$	$(U^u_{\Psi_1}(t_1) + U^c_{\Psi_2}(t_2) + \varepsilon)/2$	$U^u_\Theta(t_1)+U^c_\Theta(t_2) - \varepsilon$
$U^c_{\Psi_1}(t_1)$, $U^u_{\Psi_2}(t_2)$	$U^c_\Theta(t_1)+ U^u_\Theta(t_2)$	$(U^c_{\Psi_1}(t_1) + U^u_{\Psi_2}(t_2) + \varepsilon)/2$	$U^c_\Theta(t_1)+ U^u_\Theta(t_2) - \varepsilon$
$U^c_{\Psi_1}(t_1)$, $U^c_{\Psi_2}(t_2)$	$U^c_\Theta(t_1)+U^c_\Theta(t_2)$	$(U^c_{\Psi_1}(t_1)+U^c_{\Psi_2}(t_2))/2$	$U^c_\Theta(t_1)+U^c_\Theta(t_2)$

Table 7.1: Payoffs table for individual and collusive attacks

Hence, we encapsulate the targets which are attacked in a set of binary variables a_t, $t \in T$, where the variables corresponding to the targets which are attacked are set to 1.

We denote the utility that the defender receives when target t is attacked by $U^u_\Theta(t)$ if t is uncovered, and $U^c_\Theta(t)$ if t is covered. The payoffs for the ith attacker are analogously written $U^u_{\Psi_i}(t)$ and $U^c_{\Psi_i}(t)$. Suppose that the attackers select target $t_1 \in T_1$ and $t_2 \in T_2$. Since each may be covered or uncovered, four different outcomes are possible. Table 7.1 summarizes the players' payoffs in all possible cases when the attackers do not collude (the first two columns) and collude (the last two columns). In this table the first row indicates the payoffs when both targets are uncovered and both adversaries are successful. The second and third rows show the payoffs when only one attacker succeeds and the last row indicates the case of failure for both attackers.

If the attackers collude with each other, they share all of their utility equally. Additionally, they receive a bonus reward, ε, for any successful attack. As we focus on zero-sum games for the experiments, this bonus value is deducted from the defender's payoff. Further, while we assume that adversaries who choose to collude split their combined payoff equally, it is important to note that the algorithms we present are easily generalized to accommodate arbitrary payoff splits. There are two principal reasons as to why we specify a 50-50 split in this work. First, this division is motivated by inequity aversion theory, as outlined earlier. Second, our focus here is on the factors which lead individuals to collude in the first place, not on the bargaining process which decides their allocation of the rewards (a topic which is itself the subject of a great deal of work in game theory and psychology). Since the reward structures we consider are symmetric between the players, an equal distribution of rewards is a natural assumption. Thus, we can isolate the factors which lead subjects to enter into collusion instead of confounding the decision to collude with an additional bargaining process.

For a given coverage vector C defender's utility at each target t_i attacked individually by attacker i is defined by Equation 7.1. By replacing Θ with Ψ, the same notation applies for the expected utility of the attacker.

$$U_\Theta(t_i, C) = \quad c_{t_i} \cdot U^c_\Theta(t_i) + (1 - c_{t_i})U^u_\Theta(t_i) \tag{7.1}$$

Now we introduce our solution concept for COSGs, the Collusive Security Equilibrium (CSE), which generalizes the SSE to the case of multiple attackers. Let the defender's strategy be a coverage vector C, and the attackers' strategies g_1 and g_2 be functions from coverage vectors to $T \times \{collude, not\ collude\}$. Recall that a strategy profile forms an SSE if (1) the attacker and defender play mutual best responses and (2) the attacker breaks ties in favor of the defender. In COSGs, each attacker's best response depends on the other, since the decision of whether or not to collude depends on the utility the other attacker will obtain. Essentially, any fixed C induces a game between the attackers; the defender sets the attackers' payoff at each target via their resource allocation. The following conditions define a CSE:

1. C is a best response to g_1 and g_2.
2. $g_1(C)$ and $g_2(C)$ form a Nash equilibrium in the game where each target's utility is $U_\Psi(t,C)$.
3. Both attackers play *collude* if they obtain strictly greater utility in a (*collude, collude*) equilibrium than (*not collude, not collude*) equilibrium.
4. The attackers break ties between equilibria which satisfy (1)-(3) in favor of the defender.

The first two conditions are analogous to the best response conditions for SSE. In particular, when the followers play a Nash equilibrium (Condition 2), each is playing a best response to the fixed strategies of the other two players. Condition 3 removes the trivial equilibrium where neither attacker chooses to collude because they cannot gain unless the other attacker also decides to collude. Condition 4 enforces the normal SSE condition that remaining ties are broken in favor of the defender.

7.4 SPECTRE-R: Optimal defender strategy for rational adversaries

SPECTRE-R (Strategic Patrolling to Extinguish Collaborative ThREats from Rational adversaries) takes a COSG as input and solves for an optimal defender coverage vector corresponding to a CSE strategy through a mixed integer linear program (MILP). This MILP is based on the

ERASER formulation introduced by Kiekintveld et al. [47]. The original formulation was developed for SSGs with one defender and one adversary. We extend these ideas to handle collusion between two adversaries via the MILP in Equations 7.2-7.16. It is important to note that while the rewards structures we consider in the experiments are zero sum, the MILP we give applies to general sum games. Additionally, our methods are not restricted to the case of two adversaries. In the online appendix[1], we provide a generalization of this MILP to COSGs with *N* adversaries. Since a naive extension would entail a number of constraints which is exponential in *N*, we conduct more detailed analysis of the structure of the game, which allows us to formulate

[1] https://www.dropbox.com/s/kou5w6b8nbvm25o/nPlayerAppendix.pdf?dl=0

a MILP with only $O(N^3)$ constraints. However, this analysis is also deferred to the appendix as our experimental focus is on COSGs with two adversaries.

$$\max d \quad \text{s.t.} \tag{7.2}$$

$$a_t^{nc}, a_t^c, \alpha_1, \alpha_2, \beta \in \{0,1\} \tag{7.3}$$

$$c_t \in [0,1] \tag{7.4}$$

$$\sum_{t \in T} c_t \leq m \quad \sum_{t_i \in T_i} a_{t_i}^{nc} = 1 \quad \sum_{t_i \in T_i} a_{t_i}^c = 1 \tag{7.5}$$

$$U_\Theta^c(t_1, t_2, C) = U_\Theta(t_1, C) + U_\Theta(t_2, C) -$$

$$(1 - c_{t_1})\varepsilon - (1 - c_{t_2})\varepsilon \tag{7.6}$$

$$U_\Theta^{nc}(t_1, t_2, C) = U_\Theta(t_1, C) + U_\Theta(t_2, C) \tag{7.7}$$

$$d - U_\Theta^c(t_1, t_2, C) \leq (1 - a_{t_1}^c)Z + (1 - a_{t_2}^c)Z + (1 - \beta)Z \tag{7.8}$$

$$d - U_\Theta^{nc}(t_1, t_2, C) \leq (1 - a_{t_1}^{nc})Z + (1 - a_{t_2}^{nc})Z + \beta Z \tag{7.9}$$

$$U_{\Psi_i}^c(t_i, C) = U_{\Psi_i}(t_i, C) + (1 - c_{t_i})\varepsilon \tag{7.10}$$

$$U_{\Psi_i}^{nc}(t_i, C) = U_{\Psi_i}(t_i, C) \tag{7.11}$$

$$0 \leq k_i^c - U_{\Psi_i}^c(t_i, C) \leq (1 - a_{t_i}^c)Z \tag{7.12}$$

$$0 \leq k_i^{nc} - U_{\Psi_i}^{nc}(t_i, C) \leq (1 - a_{t_i}^{nc})Z \tag{7.13}$$

$$-\alpha_i Z \leq k_i^{nc} - \frac{1}{2}(k_1^c + k_2^c) \leq (1 - \alpha_i)Z \tag{7.14}$$

$$\beta \leq \alpha_i \tag{7.15}$$

$$(\alpha_1 + \alpha_2) \leq \beta + 1 \tag{7.16}$$

We now proceed to an explanation of the above MILP which is named as SPECTRE-R algorithm in this study and optimizes defender utility, d, against collusive adversaries. In all equations, nc stands for not colluding cases and c stands for colluding cases, and Z is a large constant. Additionally, constraints with free indices are repeated across all possible values, e.g. $i = 1,2$ or $t \in T$. Equation 7.3 defines the binary decision variables. a_t^c and a_t^{nc} whether each target would be attacked if the corresponding adversary chooses to collude or not collude, respectively. α_1 and α_2 indicate each adversary's decision of whether to collude. β is indicates whether collusion actually occurs; it is one if and only if both α_1 and α_2 are one. c_t, introduced in Equation 7.4 is the defender coverage probability at target t. Equation 7.5 enforces the defender resource constraint, and that the attackers each select exactly one target. Equations 7.6 and 7.7 calculate the defender expected utilities at each target in the case of collusion and no collusion. Equations 7.8 and 7.9 define the defender's final expected payoff based on which target is attacked in each case.

Equations 7.10 and 7.11 define the expected utility of the attackers in colluding and non-colluding situations. Equations 7.12 and 7.13 constrain the attackers to select a strategy in attack set of C in each situation. Equation 7.14 requires each attacker to collude whenever they obtain higher utility from doing so. Lastly, Equations 7.15 and 7.16 set $\beta = \alpha_1 \wedge \alpha_2$.

Proposition 7.1 *Any solution to the above MILP is a CSE.*

Proof 7.1 *We start by showing that the followers play a Nash equilibrium as required by condition (2). Let $(a^*_{t_i}, \alpha^*_i)$ be the action of one of the followers produced by the MILP where t_i is the target to attack and α_i is the decision of whether to collude. Let (a_{t_i}, α_i) be an alternative action. We need to show that the follower cannot obtain strictly higher utility by switching from $(a^*_{t_i}, \alpha^*_i)$ to (a_{t_i}, α_i). If $\alpha^*_i = \alpha_i$, then Equations 7.12 and 7.13 imply that a_{t_i} already maximizes* *the follower's utility. If $\alpha_i^* \neq \alpha_i$ then Equation 7.14 implies that $(a_{t_i}^*, \alpha_i^*)$ yields at least as much utility as $(a_{t_i}, 1-\alpha_i^*)$, for the a_{t_i} which maximizes the follower's utility given that they make the opposite decision about collusion. So, $(a_{t_i}^*, \alpha_i^*)$ yields at least as much utility as (a_{t_i}, α_i), and condition (2) is satisfied. For condition (3), note that in Equation 7.14, both followers compute the utility for collusion assuming that the other will also collude. So, if follower i would be best off with $\beta = 1$, the MILP requires that $\alpha_i = 1$. Thus, if both followers receive strictly highest utility in an equilibrium with $\beta = 1$, both will set $\alpha = 1$. In all other cases, the objective is simply maximizing d, which satisfies conditions (1) and (4) by construction.*

The following observations and propositions hold for the games with symmetric reward distribution between the two adversaries.

OBSERVATION 1. *The defender optimizes against rational adversaries by enforcing an imbalance in resource allocation between the sides and preventing collusion.*

In SPECTRE-R, the key idea for preventing collusion between two adversaries is to impose a resource imbalance between their situations. This places one adversary in an advantaged condition and the other in a disadvantaged condition. Assuming perfectly rational adversaries, we expect that the disadvantaged adversary will always seek to collude, and the advantaged attacker will always refuse (provided the imbalance outweighs the bonus ε). In other words, the optimal solution provided by SPECTRE-R satisfies $\theta \neq 0$ where $\theta = |x_1 - x_2|$, $x_i = \sum_{t_i \in T_i} c_{t_i}$ is difference in total resource allocation to the two sides. This approach incentivizes one attacker to refuse to collude by putting them in a better position than the other.

To analyze the effect of the imbalance in resource allocation on defender expected payoff, we added another constraint to the MILP formulation shown in Equation 7.17 forces a resource imbalance at an arbitrary level, δ. For the case of symmetric reward distribution, WLOG, we can fix the first attacker to be the one who receives higher payoff and simply linearize the following equation; however generally, we can divide the equation into two separate linear constraints.

$$|k_1^{nc} - k_2^{nc}| = \delta \tag{7.17}$$

OBSERVATION 2. *By varying δ, the following cases can occur:*

1. *For $\delta < \delta^*$, $k_i^{nc} - \frac{1}{2}(k_1^c + k_2^c) < 0$ for both attackers and consequently $\alpha_i = 1$ for $i = 1,2$. In other words, the defender is not able to prevent collusion between the attackers and $\beta = 1$.*

2. *For $\delta = \delta^*$, $k_1^{nc} - \frac{1}{2}(k_1^c + k_2^c) = 0$ for one of the attackers and $k_2^{nc} - \frac{1}{2}(k_1^c + k_2^c) < 0$ for the other one, so consequently α_1 can be either 0 or 1 and $\alpha_2 = 1$. In this case, the followers break ties in favor of the leader, so $\alpha_1 = 0$ and $\beta = 0$.*

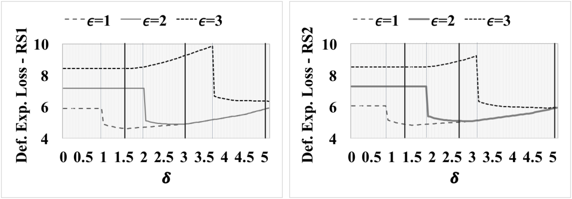
3. *For $\delta > \delta^*$, $k_1^{nc} - \frac{1}{2}(k_1^c + k_2^c) > 0$ for one of the attackers and consequently $\alpha_1 = 0$. For the other attacker $k_2^{nc} - \frac{1}{2}(k_1^c + k_2^c) < 0$ and $\alpha_2 = 1$. In other words, the defender is able to prevent collusion between the attackers and $\beta = 0$.*

Proposition 7.2 *The switch-over point, δ^*, introduced in the observation 2 is lower bounded by 0 and upper bounded by 2ε.*

Proof 7.2 *Using Equation 7.13, we know that at any target t_i, $k_i^{nc} \geq U_{\Psi_i}^{nc}(t_i, C)$. If we assume that the attacker attacks target t_i^c with coverage $c_{t_i}^c$ by adding and subtracting a term as $\varepsilon(1 - c_{t_i}^c)$, we can conclude that $k_i^{nc} \geq k_i^c - \varepsilon(1 - c_{t_i}^c)$. Consequently, $k_1^c + k_2^c \leq k_1^{nc} + k_2^{nc} + \varepsilon(1 - c_{t_1}^c) + \varepsilon(1 - c_{t_2}^c)$. On the other hand, according to observation 2.2, at $\delta = \delta^*$, we have $k_1^{nc} - \frac{1}{2}(k_1^c + k_2^c) = 0$. Combining these last two equations, we will get $(k_1^{nc} - k_2^{nc}) \leq \varepsilon(1 - c_{t_1}^c) + \varepsilon(1 - c_{t_2}^c)$. The LHS is* *equal to δ^* and the RHS can be rearranged as $2\varepsilon - \varepsilon(c^c_{t_1} + c^c_{t_2})$, so we will have $\delta^* \leq 2\varepsilon - \varepsilon(c^c_{t_1} + c^c_{t_2})$. Given the fact that coverage at each target is in range $[0,1]$, the upper bound for $-(c^c_{t_1} + c^c_{t_2})$ will be zero. Finally, by aggregating these results, we can conclude that $\delta^* \leq 2\varepsilon$. Following the same analysis, the lower bound for δ^* can be found starting from $k^c_1 + k^c_2 \geq k^{nc}_1 + k^{nc}_2 + \varepsilon(1 - c^{nc}_{t_1}) + \varepsilon(1 - c^{nc}_{t_2})$ and as a result, $0 \leq \delta^*$.*

Given the facts presented in Proposition 7.2, by enforcing an imbalance of maximum 2ε, the defender will be able to prevent collusion. These bounds can be tighter, if we have more information about the distribution of reward at targets. For instance, if reward distribution over targets is close enough to uniform distribution, then the average coverage on each side will be $\bar{c}_{t_1} = \frac{2x_1}{n}$ and $\bar{c}_{t_2} = \frac{2x_2}{n}$, where x_1 and x_2 are fraction of resources assigned to each side and there are $\frac{n}{2}$ targets on each side. As a result, $-(c^c_{t_1} + c^c_{t_2}) \approx -(\bar{c}_{t_1} + \bar{c}_{t_2})$. So we will be able to find an approximate upper bound of $2\varepsilon(1 - \frac{m}{n})$, where $m = x_1 + x_2$. This implies that when the ratio of $\frac{m}{n}$ is large, less imbalance in resource allocation is needed to prevent collusion. In the human subject experiments that will be discussed in the next section, we also observed that with a wider range of rewards ($RS2$ compared to $RS1$ in Figure 7.5(a) in OBSERVATION A) over targets, it becomes harder to prevent collusion between attackers.

SIMULATION 1. Simulation results of SPECTRE-R algorithm for the two games introduced in Section 3 are shown in Figure 7.3(a) and 7.3(b) for different values of the bonus ε. We vary δ along the x axis, and show the defender loss on the y axis. In all of the plots, *for each epsilon value*, there is a δ value (indicated with gray vertical lines) at which collusion breaks and also a δ^* value (which corresponds to an optimal resource imbalance θ^*) at which collusion is broken and defender loss is minimized (indicated with solid black vertical lines). The higher



(a) RS1: Def. Exp. Loss vs. δ vs. ε

(b) RS2: Def. Exp. Loss vs. δ vs. ε

Figure 7.3: Simulation results of SPECTRE-R: Defender Expected Loss vs. resource imbalance

the benefit of collusion, the larger the loss of the defender. Note that before collusion is broken, imposing a resource imbalance sometimes increases the defender's loss (see plots for $\varepsilon = 3$) because the defender deviates from the optimal coverage probabilities for a traditional SSG without reaping the benefit of reduced cooperation. Similarly, note that defender loss increases for $\delta > \delta^*$ since cooperation is already broken, so the defender only suffers by further reducing coverage on the advantaged player. This emphasizes the importance of precision in modeling and recognizing the optimal δ for allocating resources in real-world settings.

7.5 Human Behavioral Approach

7.5.1 COSG model for bounded rational adversaries

While for perfectly rational adversaries the calculations shown in Figure 7.3 would hold, our observations from human subject experiments did not match this expectation; the probability of collusion varied continuously with the level of asymmetry in the adversary's' situations. To address this problem, we propose a two layered model which is able to predict (i) the probability of collusion between the adversaries and (ii) the probability of attack over each target for each type of adversary. These layers account for ways in which human behavior experimentally differed from perfect rationality. We then use this model to generate the corresponding optimal patrol schedule.

Probability of attack over targets: We use a separate set of SUQR parameters for each adversary introduced in Section 3.1 to reflect differences in decision making. A generalized form of subjective expected utility is defined in Equation 7.18 which is a linear function of the modified defender coverage, $\hat{c}_{t_i}$ at target t_i, the uncovered payoff of the attacker, $U^u_{\Psi_i}(t_i)$, the bonus for collusion ε and the covered payoff of the attacker $U^c_{\Psi_i}(t_i)$. β is the attackers' decision variable about collusion. A vector of $\omega^\beta_i = (\omega^\beta_{i,1}, \omega^\beta_{i,2}, \omega^\beta_{i,3})$ is assigned to each adversary. Each component of ω^β_i indicates the relative weights that the adversary gives to each feature.

$$\hat{U}_{\Psi_i}(t_i,\beta) = \omega^\beta_{i,1}.\hat{c}_{t_i} + \omega^\beta_{i,2}.(U^u_{\Psi_i}(t_i) + \beta.\varepsilon) + \omega^\beta_{i,3}.U^c_{\Psi_i}(t_i) \tag{7.18}$$

The modified coverage probability, $\hat{c}_{t_i}$, is defined based on Prospect Theory mentioned in Section 2 and is related to the actual probability, c_{t_i}, via Equation 7.19, where γ and η determine the elevation and curvature of the S-shaped function [32], respectively. These functions are plotted in Section 7.3.

$$\hat{c}_{t_i} = \frac{\eta c^\gamma_{t_i}}{\eta c^\gamma_{t_i} + (1 - c_{t_i})^\gamma} \tag{7.19}$$

By the SUQR model mentioned in Section 2, the probability (conditioned on the decision about collusion) that the adversary, i, will attack target t_i is given by:

$$q_{t_i}(\hat{C} \mid \beta) = \frac{e^{\hat{U}_{\Psi_i}(t_i,\hat{C},\beta)}}{\sum\limits_{t_i \in T_i} e^{\hat{U}_{\Psi_i}(t_i,\hat{C},\beta)}} \tag{7.20}$$

For each attacker, the SUQR weight vector ω_i^β, and the probability perception parameters γ_i^β and η_i^β are estimated via maximum likelihood (MLE) using data collected from the human subject experiments. This resembles obtaining past data on poaching as mentioned in Section 3.2 to learn these parameters.

Probability of offering to collude: We propose a model which is intuitively based on SUQR to predict the probability of offering collusion by each adversary from a behavioral perspective. Different from the rational behavior model (see Figure 7.3) where collusion is deterministic, this model assumes that the attackers make stochastic decisions concerning collusion.

The probability of collusion for each adversary is calculated using Equation 7.21. Here, $\bar{U}^c_{\Psi_i} = \sum_{i \in N} \sum_{t_i \in T_i} \hat{U}_{\Psi_i}(t_i, \beta = 1)/(N.|T_i|)$ is the average adversary utility over all targets for a collusive attack and $\bar{U}^{nc}_{\Psi_i} = \sum_{t_i \in T_i} \hat{U}_{\Psi_i}(t_i, \beta = 0)/|T_i|$ is the average adversary utility over all targets for an individual attack.

$$q_i(\beta = 1) = \frac{e^{\bar{U}^c_{\Psi_i}}}{e^{\bar{U}^c_{\Psi_i}} + e^{\bar{U}^{nc}_{\Psi_i}}} \tag{7.21}$$

The coefficients in ω_i^β are learned for advantaged and disadvantaged attackers and $\beta = 0, 1$ using MLE and data collected from human subject experiments.

7.5.2 SPECTRE-BR: Optimal defender strategy for bounded rational adversaries

The two above mentioned models are incorporated in SPECTRE-BR (Strategic Patrolling to Extinguish Collaborative ThREats from Boundedly Rational adversaries) to generate the defender optimal strategy by maximizing the expected utility of the defender given in Equation 7.22 where the defender expected utility is computed as $U_\Theta(t_i, C, \beta) = c_{t_i} \cdot U^c_\Theta + (1 - c_{t_i})(U^u_\Theta + \beta\varepsilon)$ for target

t_i, mixed strategy C and the collusion variable β. In this equation, $\mathscr{C}$ represents the set of all possible coverage vectors. We define $q(\beta=1) = min(q_1(\beta=1), q_2(\beta=1))$ and $q(\beta=0) = 1-q(\beta=1)$. This assumption is supported by the fact that collusive attacks happen only when both parties are sufficiently inclined to collude, and the advantaged player will always be less inclined to offer collusion.

$$\max_{C\in\mathscr{C}} \left(\sum_{i=1}^{N} \sum_{t_i \in T_i} \sum_{\beta=0}^{1} U_{\Theta}(t_i, C, \beta) q_{t_i}(C \mid \beta) q(\beta) \right) \tag{7.22}$$

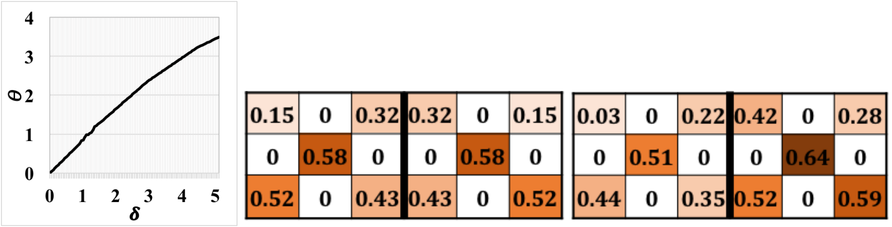
7.6 Human Subject Experiments

To determine how the behavior of human players differs from perfect rationality, we recruited participants from Amazon Mechanical Turk to play the game described in Section 3. Each experiment used 50 participants. Here we report on the results.

7.6.1 Resource imbalance effect on collusion

HYPOTHESIS A. *There exists a switch-over δ^* value, at which it is not rational for the adversaries to collude. Consequently, collusion will be broken completely.*

METHOD A. Given the intuition from the rational adversary model, the defender achieves higher expected utility by breaking collusion between the two adversaries. The main idea for preventing collusion was to place one adversary in the advantaged condition so he will avoid collusion. The corresponding optimal strategy results in an asymmetry between the maximum expected utilities on both sides which we referred to as δ. This δ is correlated with the difference between aggregated defender coverage on both sides, θ which is defined in OBSERVATION 2.



(a) θ vs δ

0.15	0	0.32	0.32	0	0.15
0	0.58	0	0	0.58	0
0.52	0	0.43	0.43	0	0.52

(b) $\delta = 0$, RS2

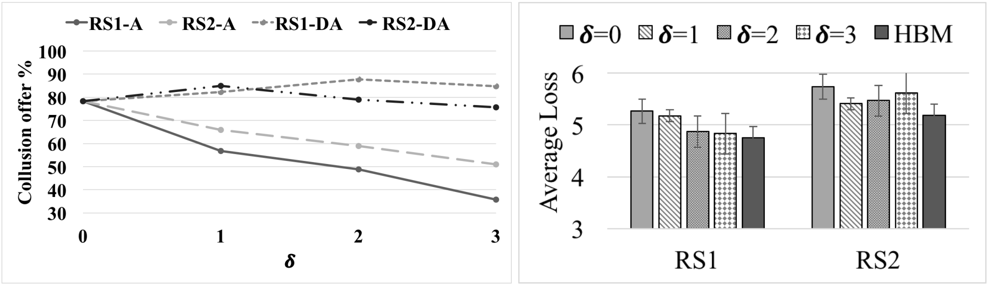
0.03	0	0.22	0.42	0	0.28
0	0.51	0	0	0.64	0
0.44	0	0.35	0.52	0	0.59

(c) $\delta = 1$, RS2

Figure 7.4: Defender strategy deployed on AMT and resource imbalance

Figure 7.4(a) illustrates this relationship by plotting δ on the x axis against the total resource imbalance on the y axis for $RS2$. As δ increases, the resource imbalance also increases. To see how deviating from balanced resource allocation affects human adversaries' decisions about collusion, we ran human subjects experiments on AMT for various δ values for two reward structures $RS1$ and $RS2$. Figures 7.4(b) and 7.4(c) illustrate two sample mixed strategy (defender coverage over targets) that we deployed on AMT for $RS2$. In Figure 4(b), resources are distributed symmetrically, while in Figure 4(c) δ was set equal to 1 and one side is covered more than the other. Next, as shown in Figure 5(a), for each reward structure, we tested 4 different coverage distribution i.e., $\delta \in \{0,1,2,3\}$. For each defender strategy we recruited 50 AMT workers. It is worth noting that the models introduced in this study are valid for both symmetric and asymmetric payoff structures; however, we show the simulation results and experiments for the symmetric case to hold the effect of other variables constant and focus mostly on the distribution of security resources.

OBSERVATION A. The experiments showed that for human adversaries, there is no switch-over point or sharp change in behavior as predicted in Figure 3 when assuming rational adversaries. Rather, the probability of offering collusion decreased smoothly as δ increased for both $RS1$ and $RS2$. This completely contradicts the results assuming a rational adversary as seen in



(a) Collusion level

(b) Average defender loss. SPECTRE-BR outperforms the rational model.

Figure 7.5: Collusion level and average defender loss

Figure 7.3. These results are shown in Figure 7.5(a). δ varies on the *x* axis while the *y* axis shows the probability of collusion. For advantaged attackers (denoted RS1-A and RS2-A in Figure 5(a)), we observe a smooth decline in collusion as δ increases. However, for disadvantaged attackers (RS1-DA and RS2-DA), we did not observe a significant change in the level of collusion; the disadvantaged attacker always offered to collude with high probability.

ANALYSIS A. The previous observation has several implications: i) for small values of δ there were a considerable number of human players in advantaged situations who refused to collude despite the fact that collusion was rational. ii) For large values of δ, there were a considerable number of human players in advantaged situations who chose to collude despite the fact that collusion was an irrational decision in that situation. This behavior might indicate that the bounded rationality model might be a better fit than the model assuming full rationality when modeling collusive adversaries.

7.6.2 SPECTRE-BR outperforms model assuming perfectly rational adversaries

HYPOTHESIS B. *A lower probability of collusion decreases defender loss.*

METHOD B. See method A.

RS	Rational Strategies (δ)			
	$\delta = 0$	$\delta = 1$	$\delta = 2$	$\delta = 3$
1	3.8×10^{-2}	6.6×10^{-4}	4.0×10^{-3}	4.6×10^{-3}
2	3.5×10^{-6}	1.9×10^{-3}	2.6×10^{-1}	5.1×10^{-2}

Table 7.2: Statistical Significance (t-Test p values for SPECTRE-BR and rational strategies)

OBSERVATION B. Figure 7.5(b) shows the average defender loss obtained by different strategies for both reward structures, $RS1$ and $RS2$. Strategies generated based on the human behavior model (SPECTRE-BR) are labeled ”HBM”, while the other bars represent strategies generated by the MILP from Section 4 using the specified δ. Figure 7.5(b) shows the empirical utility obtained by each strategy. We calculated the average loss from human players who were in the advantaged and disadvantaged position and who decided to collude and not collude. Figure 7.5(b) plots the average of these losses weighted according to the frequencies with which players decided to collude, observed in the experiments. We see that the human behavior model obtains uniformly lower loss than the perfect rationality model. In nearly all populations, the difference in utility between the strategies generated by the human behavioral model and those generated by the MILP is statistically significant ($p < 0.05$). Table 7.2 gives t-test results from comparing the utility obtained by the human behavioral model against each other strategy.

ANALYSIS B. Importantly, Figure 7.5(b) shows that breaking collusion does not always decrease defender loss. For example, in $RS2$, defender loss is lower at $\delta = 2$ compared to $\delta = 3$; however, the chance of collusion (as seen in Figure 5a) is higher for $\delta = 2$. Hence, simply decreasing the level of collusion (which is correlated with an increase in δ per OBSERVATION A.) may not always be optimal for the defender.

7.6.3 Defender coverage perception

HYPOTHESIS C. *Human adversaries' probability weightings follow S-shaped curves independent of their decision about collusion.*

METHOD C. Parameters of S-curves, γ and η in Equation 7.19 are learned for the data sets described in METHOD A using the techniques presented in Section 6.

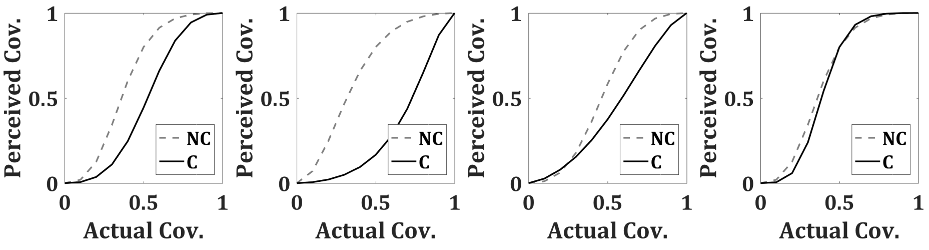
OBSERVATION C. Figures 7.6(a) and 7.6(b) show the probability weighting functions learned for the disadvantaged and advantaged adversaries for both groups who are colluding and not colluding for $RS1$. In these figures the defender coverage varies along the x axis, and the attackers' perceptions of defender coverage are shown along the y axis. Figures 7.6(c) and 7.6(d) show the same for $RS2$.

ANALYSIS C. There are two main points in these results: (i) probability weightings followed S-shaped curves, contradicting prospect theory [40, 90], i.e., low probabilities are underweighted and high probabilities are overweighted. (ii) Probability perceptions differed between those who decided to collude and not to collude. This analysis supports the use of SPECTRE-BR because humans' probability weightings are indeed nonlinear.

7.6.4 Individualism vs. collectivism

HYPOTHESIS D. *Human adversaries who are collectivists are more likely to collude than individualists in nearly all cases.*

METHOD D. All of the participants in our experiments were presented with a survey after playing the game. Eight questions were selected from the 16-item individualism-collectivism scale. Questions with the highest factor loading were selected because prior research shows that these are the most accurate indicators of individualism vs collectivism [83]. Players responded



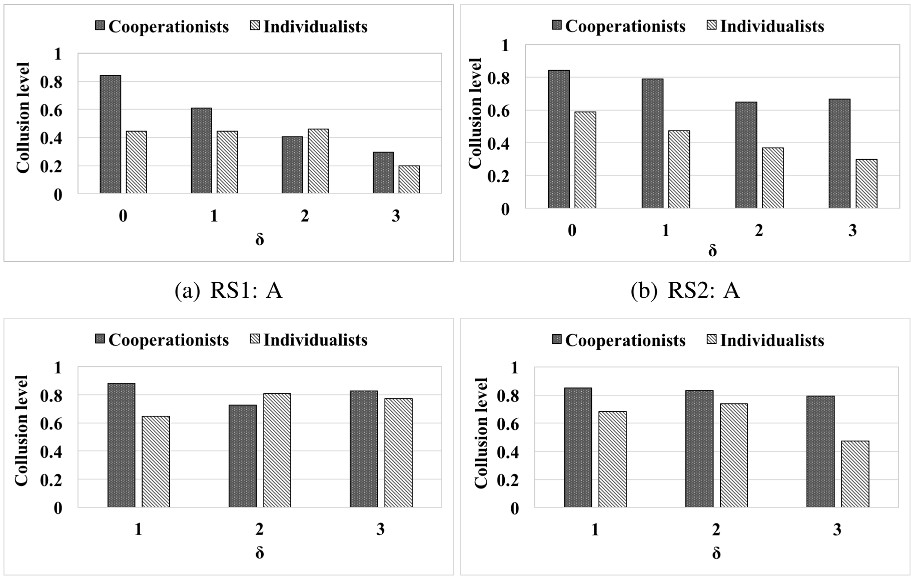
(a) RS1:DA (b) RS1:A (c) RS2:DA (d) RS2:A

Figure 7.6: Probability perception curves learned based on PT

on a scale from 1 (strongly disagree) to 7 (strongly agree). These responses were used to create a player's OI:OC (overall individualism to overall collectivism) ratio as follows. First, the sum of a player's collectivism responses, c, from collectivism-oriented questions, q_j and individualistic responses, i, from individualism-oriented questions, m_k were calculated as $c = \sum_{j=1}^{4} q_j, \left\{q_j \in \mathbb{R}^+ : 1 \leq q_j \leq 7\right\}$ and $i = \sum_{k=1}^{4} m_k, \left\{m_k \in \mathbb{R}^+ : 1 \leq m_k \leq 7\right\}$. A player's OI:OC ratio is simply i/c. A player is called an individualist if his OI:OC ratio falls above the median OI:OC score for all players, otherwise he is called a collectivist. We next explore how decisions differ between the two groups. Also please note that the order effect on individualism vs. collectivism analysis is discussed in the online appendix[2] due to space consideration.

OBSERVATION D. The data confirmed that regardless of setting, collectivists are more likely to collude than individualists. This principle was applicable regardless of a player's reward structure, the game's δ value, and whether a player was predetermined to play in an advantaged or disadvantaged state. Figure 7.7 shows the chance of collusion on the y axis versus δ on the x axis for our two reward structures and in situations where the human is in the advantaged and then disadvantaged situations; we see that the chance of offering collusion for collectivists is always higher than individualists. There is one exception in Figure 7.7(c), $\delta = 2$, where the chance of

[2] https://www.dropbox.com/s/uk9wqrdfq85vhk9/ICAppendix.pdf?dl=0



(a) RS1: A

(b) RS2: A

(c) RS1: DA

(d) RS2: DA

Figure 7.7: Cooperation level for collectivists and individualists. RS1 and RS2 indicate the reward structure, while A and DA indicate that a player was on the advantaged or disadvantaged side.

collusion for collectivists and individualists is approximately the same (a difference of less than 0.1 is observed). This single case can be considered an exception to the general rule.

ANALYSIS D. Due to factors like morality, social systems, cultural patterns, personality, etc. collectivists may prefer working with a fellow player [88] regardless of reward structure and delta value. However, the fact that collusion decreases as delta value increases has valuable implications. In security games, this means that adopting more rigorous defender strategies has the effect of dissolving collusion amongst attacker groups regardless of their OI:OC ratio. However, it is important to notice that if attackers have a relatively high OI:OC ratio (meaning they are individualists), the defender strategies given here are even more effective at preventing collusion. Please see the appendix for more individualism / collectivism analysis.

7.7 Conclusion

This chapter addresses the problem of collusion between adversaries in security domains from a game-theoretic and human behavioral perspective. Our contributions include: (i) the COSG model for security games with potential collusion among adversaries, (ii) SPECTRE-R to solve COSGs and break collusion assuming rational adversaries, (iii) observations and analyses of adversary behavior and the underlying factors including bounded rationality, imbalanced-resource-allocation effect, coverage perception, and individualism / collectivism attitudes within COSGs with data from 700 human subjects, (iv) a human behavioral model learned from the data which incorporates these underlying factors, and (v) SPECTRE-BR to optimize against the learned behavior model to provide better defender strategies against human subjects than SPECTRE-R.

Chapter 8

Conclusion

8.1 Contributions

In Green Security Games (GSG) domain, poaching is a serious threat to wildlife conservation around the world and can lead to the extinction of several important species and complete destruction of ecosystems [15]. Not only are the effects of poaching detrimental to animal species and the environmental sustainability, the illegal trade of wildlife also helps fund armed conflict by extremist groups around the world, and it has become a 213 billion dollar industry [2]. Security games are well known to be effective models of protecting valuable targets against an adversary and have been explored extensively in other domains including protection of critical infrastructure, suppressing urban crimes, or preventing cyber intrusions. However, a direct application of the existing security game models and algorithm to GSGs does not adequately consider the major challenges in this domain. The models and algorithms developed in this thesis advance the state of the art to a new generation of security games where adversarial behavior is presented by complex machine learning models which are aware of the uncertainty in past attack data, and the decision model is developed such that it benefits from several expert planners where insufficient or imperfect historical records of past attacks are available to learn adversarial behavior. The contributions of my thesis are as follows:

To model adversarial behavior based on the real-world data, I proposed two novel techniques. First, a hybrid model consists of two components: (i) an ensemble model which can work with the limited data common to the domain of environmental sustainability and (ii) a spatio-temporal model to boost the ensemble's predictions when sufficient data are available. When evaluated on real-world historical data from Queen Elizabeth National Park in Uganda, the hybrid model achieves better performance than the state-of-the-art approaches with either temporally-aware dynamic Bayesian networks or an ensemble of spatially-aware models. Second, I introduced a novel imperfect-observation aWare Ensemble (iWare-E) technique, which is designed to handle the uncertainty in crime information efficiently. This approach leads to superior accuracy for adversary behavior prediction compared to the previous state-of-the-art.

To evaluate the performance of the adversarial behavior models in the real field, I conducted a large-scale field test experiment in multiple national parks in Uganda (totaling about 7500 km^2). In these tests, several snares and snared animals were detected, and poachers were arrested, potentially more wildlife saved. The latest algorithm proposed in this thesis, that combines machine learning and game-theoretic patrol planning is planned to be deployed at 600 national parks around the world in the near future to combat poaching at a global scale.

To develop game-theoretical decision solutions based on adversarial models learned from insufficient or imperfect historical records of past attacks, I proposed a novel multi-expert online learning model for constrained patrol planning which benefits from several expert planners. Previous work in GSG literature relies on exploitation of error-prone machine learning (ML) models of poachers' behavior trained on (spatially) biased historical data; and online learning approaches for repeated security games (similar to GSGs) do not account for spatio-temporal scheduling constraints while planning patrols, potentially causing significant shortcomings in the effectiveness of the planned patrols. I proposed to integrate complex machine learning adversarial behavior along with an online learner to design efficient and feasible randomized defender strategies in Green Security Game.

To address collusive adversarial behavior, I introduced collusive security games, a model for security games involving potential collusion among adversaries and SPECTRE-R, an algorithm to solve COSGs and break collusion assuming rational adversaries. Furthermore, I proposed a learned human behavioral model that incorporates these factors to predict when collusion will occur and SPECTRE-BR, an enhanced algorithm which optimizes against the learned behavior model to provide demonstrably better performing defender strategies against human subjects compared to SPECTRE-R. I also provided observations and analyses of adversary behavior and the underlying factors including bounded rationality, imbalanced- resource-allocation effect, coverage perception, and individualism / collectivism attitudes within COSGs with data from 700 human subjects.

As future work, it is important to solve other difficult challenges arising for varying environmental sustainability and social good domains using game theory and machine learning. In the next section, I outline the future directions.

8.2 Future Work and Directions

In the future, it will be important to better understand the blind spots of machine learning models. Nowadays, machine learning models are trained based on the real-world data and are used for decision-making purposes in a variety of domains. However, the real-world data is usually insufficient, extremely noisy or collected in a way that is not representative of the entire space of the problem. Learning a predictive model based on such datasets can lead to the presence of blind spots and biases in predictions, and consequently designing a decision-making model based on such predictions might results in adverse long-term effects in the real domain. To that end, future research can focus on the development of models and decision solutions that balance exploration vs. exploitation to eliminate the blind spots of machine learning models and to mitigate the adverse effects of using biased machine learning models on decision solutions.

Another interesting area of research is the development of decision-making solutions that are designed in conjunction with machine learning models. The current approaches to design a data-to-decision pipeline are to develop each element (machine learning model and decision-making model) independently. More specifically, a fully trained machine learning model is fed into a decision-making model. However, the measures to predict the performance of predictive models based on the real-world data might not be completely aligned with the ultimate goals of the decision-making models. Thus, the key challenge is how to train a machine learning model in conjunction with the decision-making model to generate decision solutions with higher quality.

Additionally, another possible and important direction for future research is transfer learning for adversarial behavior reasoning. Data collection in some protected areas in the world has not been executed regularly and cautiously (e.g., protected areas in Cambodia). As a result, there is not a rich crime dataset available for training a predictive adversary model. The transfer learning and domain adaptation techniques are helpful to transfer knowledge across different domains. Conducting research to develop predictive models based on the richer data sets (e.g., protected areas in Uganda) and understanding how to transfer such models to the domain with less rich data sets (e.g., protected areas in Cambodia) can be a very challenging and interesting line of future work.

Furthermore, for future research, it will be valuable to augment foot patrolling data by other sources of data including imagery data via flying drones to obtain more data about animal movement and poacher's previous locations in order to conduct a multimodal analysis.

Lastly, building upon the work on AI for conservation presented in this thesis, one of the most interesting areas for future work beyond my thesis lies in the interdisciplinary research to address fundamental problems that arise from real-world challenges (e.g., the suicide prevention problem or the problem of influence maximization among homeless youths) in order to make a positive impact on society and the environment. The key common challenge in the AI for social good projects is the optimal allocation of limited intervention resources which can be posed as a game between human and nature. For example, in the suicide prevention domain interventionists (as the decision makers) aim to select and train gatekeepers (anyone who is strategically positioned to recognize and refer someone at risk of suicide) to maximize the coverage of their surveillance on a network of humans vulnerable to suicide commitment. This problem can be posed as a game against nature which intends to adversarially minimize the decision maker's surveillance coverage. I believe that artificial intelligence and machine learning have a tremendous potential to help humans improve society, empower low-resource communities and fight injustice. Although I have already dealt with environmental sustainability problems in my previous work, AI for social good projects that directly involve humans will raise fundamental new challenges in modeling approaches and designing decision-making solutions.

Reference List

[1] Noa Agmon, Sarit Kraus, and Gal A Kaminka. Multi-robot perimeter patrol in adversarial settings. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 2339–2345. IEEE, 2008.

[2] The Atlantic. Un warns that growing $213 billion poaching industry funds armed conflicts. https://www.theatlantic.com/international/archive/2014/06/un-warns-that-growing-213-billion-poaching-industry-funds-armed-conflicts/373324/, 2014.

[3] Maria-Florina Balcan, Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Commitment without regrets: Online learning in stackelberg security games. In *Proceedings of the sixteenth ACM conference on economics and computation*, pages 61–78. ACM, 2015.

[4] Horace A Bartilow and Kihong Eom. Free traders and drug smugglers: The effects of trade openness on states' ability to combat drug trafficking. *Lat. Am. Polit. Soc.*, 51(2):117–145, 2009.

[5] Nicola Basilico and Nicola Gatti. Strategic guard placement for optimal response toalarms in security games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1481–1482. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[6] Nicola Basilico, Nicola Gatti, and Francesco Amigoni. Leader-follower strategies for robotic patrolling in environments with arbitrary topologies. In *AAMAS*, 2009.

[7] Nathan Berg. Behavioral economics. *21st century economics: A reference handbook*, 2010.

[8] Christopher M Bishop. Pattern recognition. *Machine Learning*, 128:1–58, 2006.

[9] Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Learning optimal commitment to overcome insecurity. In *Advances in Neural Information Processing Systems*, pages 1826–1834, 2014.

[10] Allan G Bluman. *Elementary statistics: A step by step approach*. McGraw-Hill Higher Education New York, 2009.

[11] Elizabeth Bondi, Fei Fang, Mark Hamilton, Debarun Kar, Donnabell Dmello, Jongmoo Choi, Robert Hannaford, Arvind Iyer, Lucas Joppa, and Nevatia Ram Tambe, Milind. Spot poachers in action: Augmenting conservation drones with automatic detection in near real time. IAAI, 2018.

[12] Elizabeth Bondi, Fei Fang, Debarun Kar, Venil Noronha, Donnabell Dmello, Milind Tambe, Arvind Iyer, and Robert Hannaford. Viola: Video labeling application for security domains. In *International Conference on Decision and Game Theory for Security*, pages 377–396. Springer, 2017.

[13] Elizabeth Bondi, Hoon Oh, Haifeng Xu, Fei Fang, Bistra Dilkina, and Milind Tambe. Broken signals in security games: Coordinating patrollers and sensors in the real world. 2019.

[14] Colin Camerer. *Behavioral game theory*. Princeton University Press, 2003.

[15] Guillaume Chapron, Dale G Miquelle, Amaury Lambert, John M Goodrich, Stéphane Legendre, and Jean Clobert. The impact on tigers of poaching versus prey depletion. *Journal of Applied Ecology*, 45(6):1667–1674, 2008.

[16] Yuehui Chen, Bo Yang, and Ajith Abraham. Flexible neural trees ensemble for stock index modeling. *Neurocomputing*, 70(4):697–703, 2007.

[17] Vincent Conitzer and Tuomas Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM conference on Electronic commerce*, pages 82–90. ACM, 2006.

[18] Rosie Cooney, Dilys Roe, Holly Dublin, Jacob Phelps, David Wilkie, Aidan Keane, Henry Travers, Diane Skinner, Daniel WS Challender, James R Allan, et al. From poachers to protectors: engaging local communities in solutions to illegal wildlife trade. *Conservation Letters*, 10(3):367–374, 2017.

[19] R Critchlow, AJ Plumptre, M Driciru, A Rwetsiba, EJ Stokes, C Tumwesigye, F Wanyama, and CM Beale. Spatiotemporal trends of illegal activities from ranger-collected data in a ugandan national park. *Conservation Biology*, 29(5):1458–1470, 2015.

[20] Rob Critchlow, Andrew J Plumptre, Bazil Alidria, Mustapha Nsubuga, Margaret Driciru, Aggrey Rwetsiba, F Wanyama, and Colin M Beale. Improving law-enforcement effectiveness and efficiency in protected areas using ranger-collected monitoring data. *Conservation Letters*, 2016.

[21] Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 213–220. ACM, 2008.

[22] Fei Fang, Thanh H Nguyen, Rob Pickles, Wai Y Lam, Gopalasamy R Clements, Bo An, Amandeep Singh, Milind Tambe, and Andrew Lemieux. Deploying paws: Field optimization of the protection assistant for wildlife security. In *IAAI*, 2016.

[23] Fei Fang, Peter Stone, and Milind Tambe. When security games go green: Designing defender strategies to prevent poaching and illegal fishing. In *IJCAI*, 2015.

[24] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, pages 817–868, 1999.

[25] Nicola Gatti. Game theoretical insights in strategic patrolling: Model and algorithm in normal-form. In *ECAI*, pages 403–407, 2008.

[26] Shahrzad Gholami, Benjamin Ford, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, and Joshua Mabonga. Taking it for a test drive: a hybrid spatio-temporal model for wildlife poaching prediction evaluated through a controlled field test. In *Proceedings of the European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases, ECML PKDD*, 2017.

[27] Shahrzad Gholami, Sara Mc Carthy, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Joshua Mabonga, et al. Adversary models account for imperfect crime data: Forecasting and planning against real-world poachers. pages 823–831, 2018.

[28] Shahrzad Gholami, Bryan Wilder, Matthew Brown, Arunesh Sinha, Nicole Sintov, and Milind Tambe. A game theoretic approach on addressing cooperation among human adversaries. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*, 2016.

[29] Shahrzad Gholami, Bryan Wilder, Matthew Brown, Dana Thomas, Nicole Sintov, and Milind Tambe. Divide to defend: Collusive security games. In *GameSec*, pages 272–293. Springer, 2016.

[30] Shahrzad Gholami, Bryan Wilder, Matthew Brown, Dana Thomas, Nicole Sintov, and Milind Tambe. Toward addressing collusion among human adversaries in security games. In *ECAI*, pages 1750–1751, 2016.

[31] Shahrzad Gholami, Amulya Yadav, Long Tran-Thanh, Bistra Dilkina, and Milind Tambe. Don't put all your strategies in one basket: Playing green security games with imperfect prior knowledge (submitted). AAMAS, 2019.

[32] Richard Gonzalez and George Wu. On the shape of the probability weighting function. *Cognitive psychology*, 38(1):129–166, 1999.

[33] Thomas Gray, Rachel Crouthers, K Ramesh, J Vattakaven, Jimmy Borah, Mks Pasha, Thona Lim, Phan Channa, R Singh, Barney Long, S Chapman, O Keo, and M Baltzer. A framework for assessing readiness for tiger panthera tigris reintroduction: a case study from eastern cambodia. *Biodiversity and Conservation*, 05 2017.

[34] Qingyu Guo, Bo An, Yevgeniy Vorobeychik, Long Tran-Thanh, Jiarui Gan, and Chunyan Miao. Coalitional security games. In *Proceedings of AAMAS*, pages 159–167, 2016.

[35] Nika Haghtalab, Fei Fang, Thanh Hong Nguyen, Arunesh Sinha, Ariel D Procaccia, and Milind Tambe. Three strategies to success: Learning adversary models in security games. In *IJCAI*, volume 16, pages 308–314, 2016.

[36] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. Learning from class-imbalanced data: review of methods and applications. *Expert Systems with Applications*, 73:220–239, 2017.

[37] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9):1263–1284, 2008.

[38] Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991.

[39] Creola Johnson. America's first consumer financial watchdog is on a leash. *Cath. UL Rev.*, 61:381, 2011.

[40] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

[41] Nitin Kamra, Umang Gupta, Fei Fang, Yan Liu, and Milind Tambe. Policy learning for continuous space security games using neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[42] Debarun Kar, Fei Fang, Francesco Delle Fave, Nicole Sintov, and Milind Tambe. A game of thrones: when human behavior models compete in repeated stackelberg security games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1381–1390. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[43] Debarun Kar, Fei Fang, Francesco Delle Fave, Nicole Sintov, and Milind Tambe. "a game of thrones": When human behavior models compete in repeated stackelberg security games. In *AAMAS*, 2015.

[44] Debarun Kar, Benjamin Ford, Shahrzad Gholami, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, and Aggrey Rwetsiba. Cloudy with a chance of poaching: Adversary behavior modeling and forecasting with real-world poaching data. 2017.

[45] Christopher Kiekintveld, Towhidul Islam, and Vladik Kreinovich. Security games with interval uncertainty. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 231–238. International Foundation for Autonomous Agents and Multiagent Systems, 2013.

[46] Christopher Kiekintveld, Manish Jain, Jason Tsai, James Pita, Fernando Ordóñez, and Milind Tambe. Computing optimal randomized resource allocations for massive security games. AAMAS, 2009.

[47] Christopher Kiekintveld, Manish Jain, Jason Tsai, James Pita, Fernando Ordóñez, and Milind Tambe. Computing optimal randomized resource allocations for massive security games. In *AAMAS*, 2009.

[48] Richard Klíma, Christopher Kiekintveld, and Viliam Lisỳ. Online learning methods for border patrol resource allocation. In *International Conference on Decision and Game Theory for Security*, pages 340–349. Springer, 2014.

[49] Richard Klíma, Viliam Lisỳ, and Christopher Kiekintveld. Combining online learning and equilibrium computation in security games. In *International Conference on Decision and Game Theory for Security*, pages 130–149. Springer, 2015.

[50] Dmytro Korzhyk, Vincent Conitzer, and Ronald Parr. Complexity of computing optimal stackelberg strategies in security resource allocation games. In *AAAI*, 2010.

[51] Dmytro Korzhyk, Vincent Conitzer, and Ronald Parr. Security games with multiple attacker resources. In *IJCAI Proceedings*, volume 22, page 273, 2011.

[52] Dmytro Korzhyk, Vincent Conitzer, and Ronald Parr. Solving stackelberg games with uncertain observability. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1013–1020. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[53] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Eric Horvitz. Identifying unknown unknowns in the open world: Representations and policies for guided exploration. In *AAAI*, volume 1, page 2, 2017.

[54] Wee Sun Lee and Bing Liu. Learning with positive and unlabeled examples using weighted logistic regression. In *ICML*, volume 3, 2003.

[55] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

[56] Andrew M Lemieux. *Situational prevention of poaching*. Routledge, 2014.

[57] Xiao-Li Li and Bing Liu. Learning from positive and unlabeled examples with different data distributions. In *European Conference on Machine Learning*, pages 218–229. Springer, 2005.

[58] Bing Liu, Wee Sun Lee, Philip S Yu, and Xiaoli Li. Partially supervised classification of text documents. In *ICML*, volume 2, pages 387–394. Citeseer, 2002.

[59] Berendien Anna Lubbe, Elizabeth Ann du Preez, Anneli Douglas, and Felicite Fairer-Wessels. The impact of rhino poaching on tourist experiences and future visitation to national parks in south africa. *Current Issues in Tourism*, pages 1–8, 2017.

[60] Sara McCarthy, Milind Tambe, Christopher Kiekintveld, Meredith L. Gore, and Alex Killion. Preventing illegal logging: Simultaneous optimization of resource teams and tactics for security. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 3880–3886. AAAI Press, 2016.

[61] Daniel L McFadden. Quantal choice analaysis: A survey. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 363–390. NBER, 1976.

[62] Richard McKelvey and Thomas R. D.Palfrey. Quantal response equilibria for normal form games. In *Games and Economic Behavior*, 1995.

[63] Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

[64] Enrique Munoz de Cote, Ruben Stranders, Nicola Basilico, Nicola Gatti, and Nick Jennings. Introducing alarms in adversarial patrolling games. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 1275–1276. International Foundation for Autonomous Agents and Multiagent Systems, 2013.

[65] Robin Naidoo, Brendan Fisher, Andrea Manica, and Andrew Balmford. Estimating economic losses to tourism in africa from the illegal killing of elephants. *Nature communications*, 7, 2016.

[66] C Narrod, M Tiongco, and R Scott. Current and predicted trends in the production, consumption and trade of live animals and their products. *Rev. sci. tech. Off. int. Epiz.*, 30(1), 2011.

[67] Gergely Neu and Gábor Bartók. An efficient algorithm for learning with semi-bandit feedback. In *International Conference on Algorithmic Learning Theory*, pages 234–248. Springer, 2013.

[68] Thanh H Nguyen, Francesco M Delle Fave, Debarun Kar, Aravind S Lakshminarayanan, Amulya Yadav, Milind Tambe, Noa Agmon, Andrew J Plumptre, Margaret Driciru, Fred Wanyama, et al. Making the most of our regrets: Regret-based solutions to handle payoff uncertainty and elicitation in green security games. In *GameSec*, pages 170–191. Springer, 2015.

[69] Thanh H Nguyen, Debarun Kar, Matthew Brown, Arunesh Sinha, Milind Tambe, and Albert Xin Jiang. Towards a science of security games. *New Frontiers of Multi-Disciplinary Research in STEAM-H*, 2015.

[70] Thanh H. Nguyen, Arunesh Sinha, Shahrzad Gholami, Andrew Plumptre, Lucas Joppa, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Rob Critchlow, and Colin Beale. Capture: A new predictive anti-poaching tool for wildlife protection. In *AAMAS*, 2016.

[71] Thanh H Nguyen, Arunesh Sinha, Shahrzad Gholami, Andrew Plumptre, Lucas Joppa, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Rob Critchlow, et al. Capture: A new predictive anti-poaching tool for wildlife protection. pages 767–775. AAMAS, 2016.

[72] Thanh Hong Nguyen, Rong Yang, Amos Azaria, Sarit Kraus, and Milind Tambe. Analyzing the effectiveness of adversary modeling in security games. In *AAAI*, 2013.

[73] Steven Okamoto, Noam Hazon, and Katia Sycara. Solving non-zero sum multiagent network flow security games with attack costs. In *AAMAS*, pages 879–888, 2012.

[74] Hannah J O'Kelly. Monitoring conservation threats, interventions, and impacts on wildlife in a cambodian tropical forest. *Imperial College, London*, page 149, 2013.

[75] Praveen Paruchuri, Jonathan P Pearce, Janusz Marecki, Milind Tambe, Fernando Ordonez, and Sarit Kraus. Playing games for security: An efficient exact algorithm for solving bayesian stackelberg games. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 895–902. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[76] Nazneen Fatema Rajani and Raymond J Mooney. Supervised and unsupervised ensembling for knowledge base population. *arXiv preprint arXiv:1604.04802*, 2016.

[77] Parinaz Rashidi, Tiejun Wang, Andrew Skidmore, Hamed Mehdipoor, Roshanak Darvishzadeh, Shadrack Ngene, Anton Vrieling, and Albertus G Toxopeus. Elephant poaching risk assessed using spatial and non-spatial bayesian models. *Ecological Modelling*, 338:60–68, 2016.

[78] Parinaz Rashidi, Tiejun Wang, Andrew Skidmore, Anton Vrieling, Roshanak Darvishzadeh, Bert Toxopeus, Shadrack Ngene, and Patrick Omondi. Spatial and spatiotemporal clustering methods for detecting elephant poaching hotspots. *Ecological Modelling*, 297:180–186, 2015.

[79] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[80] Andres Lopez Restrepo and Álvaro Camacho Guizado. From smugglers to warlords: twentieth century colombian drug traffickers. *Can. J. Lat. Am. Caribb. Stud.*, 28(55-56):249–275, 2003.

[81] Rahul Savani and Bernhard Von Stengel. Exponentially many steps for finding a nash equilibrium in a bimatrix game. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 258–267. IEEE, 2004.

[82] Chris Seiffert, Taghi M Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE SMC-A: Systems and Humans*, 40(1):185–197, 2010.

[83] Theodore M. Singelis, Harry C. Triandis, Dharm P.S. Bhawuk, and Michele J. Gelfand. Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29(3):240–275, 1995.

[84] Arunesh Sinha, Debarun Kar, and Milind Tambe. Learning adversary behavior in security games: A pac model perspective. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 214–222. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

[85] SMART. Spatial monitoring and reporting tool. `http://smartconservationtools.org/`, 2013.

[86] Anne H Schistad Solberg, Torfinn Taxt, and Anil K Jain. A markov random field model for classification of multisource satellite imagery. *IEEE TGRS*, 34(1):100–113, 1996.

[87] Milind Tambe. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge University Press, 2011.

[88] Harry C Triandis and Michele J Gelfand. Converging measurement of horizontal and vertical individualism and collectivism. *Journal of personality and social psychology*, 74(1):118, 1998.

[89] Jason Tsai, Christopher Kiekintveld, Fernando Ordonez, Milind Tambe, and Shyamsunder Rathi. Iris-a tool for strategic security allocation in transportation networks. 2009.

[90] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

[91] Yufei Wang, Zheyuan Ryan Shi, Lantao Yu, Yi Wu, Rohit Singh, Lucas Joppa, and Fei Fang. Deep reinforcement learning for green security games with real-time information. *arXiv preprint arXiv:1811.02483*, 2018.

[92] Greg L Warchol, Linda L Zupan, and Willie Clack. Transnational criminality: An analysis of the illegal wildlife market in southern africa. *International Criminal Justice Review*, 13(1):1–27, 2003.

[93] Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, and Jiahang Chen. Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, 16(4):449–475, 2013.

[94] Liana S Wyler and Pervaze A Sheikh. International illegal trade in wildlife. DTIC Document, 2008.

[95] Haifeng Xu, Benjamin Ford, Fei Fang, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, et al. Optimal patrol planning for green security games with black-box attackers. In *International Conference on Decision and Game Theory for Security*, pages 458–477. Springer, 2017.

[96] Haifeng Xu, Long Tran-Thanh, and Nicholas R Jennings. Playing repeated security games with no prior knowledge. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 104–112. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

[97] Rong Yang. *Human Adversaries in Security Games: Integrating Models of Bounded Rationality and Fast Algorithms*. PhD thesis, University of Southern California, 2014.

[98] Rong Yang, Benjamin Ford, Milind Tambe, and Andrew Lemieux. Adaptive resource allocation for wildlife protection against illegal poachers. In *AAMAS*, 2014.

[99] Rong Yang, Christopher Kiekintveld, Fernando Ordonez, Milind Tambe, and Richard John. Improving resource allocation strategy against human adversaries in security games. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 458, 2011.

[100] Zhaozheng Yin and Robert Collins. Belief propagation in a 3d spatio-temporal mrf for moving object detection. In *IEEE CVPR*, pages 1–8. IEEE, 2007.

[101] Zhengyu Yin, Dmytro Korzhyk, Christopher Kiekintveld, Vincent Conitzer, , and Milind Tambe. Stackelberg vs. nash in security games: Interchangeability, equivalence, and uniqueness. In *AAMAS*, 2010.